

A SINGLE LOOP METHOD FOR QUADRATIC MINMAX OPTIMIZATION*

STEFANO CIPOLLA [†], OLIVER STEIN [‡], AND ALAIN ZEMKOHO [§]

Abstract. We consider a quadratic minmax problem with coupled inner constraints and propose a method to compute a class of stationary points. To motivate the need to compute such stationary points, we first show that they are meaningful, in the sense that they can be locally optimal for our problem under suitable non-degeneracy conditions. Then based on a suitable log barrier function, we build an infeasible interior point-type *single loop method* (which does not explicitly distinguish between the outer and inner problem) and prove that a non-degenerate stationary point is an attraction point as the algorithm moves along the designed central path. We show in particular that our method is polynomial in the special case where the inner feasible set of our constrained minmax problem is independent from outer variables. Our numerical experiments, on both synthetic data and a class of min cost flow problems, showcase the behavior of our method and how it outperforms existing algorithms from the literature in terms of the quality of the computed stationary points.

Key words. minmax optimization; stationarity; non-degeneracy; interior point method.

AMS subject classifications. 90C20; 90C31; 90C33; 90C47; 90C51.

1 Introduction. We consider the quadratic minmax optimization problem

$$(P) \quad \min_{x \in X} \max_{y \in Y(x)} f(x, y),$$

where the convex-concave objective function $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ is given by

$$(1.1) \quad f(x, y) := \frac{1}{2} \begin{bmatrix} x^\top & y^\top \end{bmatrix} \underbrace{\begin{bmatrix} Q_{11} & Q_{12} \\ Q_{12}^\top & -Q_{22} \end{bmatrix}}_{=: Q} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} c_x^\top & c_y^\top \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

with the data vectors $c_x \in \mathbb{R}^n$, $c_y \in \mathbb{R}^m$, a matrix $Q_{12} \in \mathbb{R}^{n \times m}$ as well as symmetric positive semi-definite matrices $Q_{11} \in \mathbb{R}^{n \times n}$ and $Q_{22} \in \mathbb{R}^{m \times m}$. The *min* and *max* operators represent the *outer* and *inner* optimization problems, respectively. Analogously, the *outer* and *inner* feasible sets of problem (P) are respectively defined by

$$(1.2) \quad \begin{aligned} X &:= \{x \in \mathbb{R}^n : A_O x = b_O, \ x \geq 0\}, \\ Y(x) &:= \{y \in \mathbb{R}^m : A_I x + B_I y = b_I, \ y \geq 0\}, \end{aligned}$$

where inequalities between vectors are meant in the component-wise sense. The corresponding data matrices are such that $A_O \in \mathbb{R}^{p \times n}$, $b_O \in \mathbb{R}^p$, $A_I \in \mathbb{R}^{q \times n}$, $B_I \in \mathbb{R}^{q \times m}$, and $b_I \in \mathbb{R}^q$ with the dimensions $p < n$ and $q < m$.

There has been a regained interest in minmax optimization in recent years, thanks to a wide range of applications discovered in artificial intelligence; see, e.g., [31, 11, 20]. The specific model in (P) has many applications, including in support vector machines, adversarial learning, interdiction games, and robust quadratic programming problems [24, 32, 36].

Clearly, (P) is a special class of convex-concave minmax problems. However, it is still quite difficult to solve due to the presence of the outer variable in the inner feasible set; more precisely, $A_I x + B_I y = b_I$ corresponds to the coupled constraint. Hence, overall, the problem is a quadratic version of the minmax problem with coupled inner constraint. Problem (P) was shown in [36] to be NP-hard. One of the key challenges in addressing the problem is that the famous von Neumann minmax theorem cannot hold in this case, as viewing (P) as a game problem, the order of play becomes crucial; the outer player has to play first by selecting $x \in X$, and then, the inner player can react by selecting $y \in Y(x)$. This is the reason why some papers (see, e.g., [20, 13]) refer to such problems where the minmax theorem cannot hold as Stackelberg games. Problem (P) is strictly speaking not a Stackelberg game problem. However, interpreting it as that of finding

$$(1.3) \quad x \in \arg \min_{x \in X} f(x, y(x)) \quad \text{with} \quad y(x) \in \arg \max_{y \in Y(x)} f(x, y)$$

*Submitted to the editors DATE.

Funding: The work of the third author is partly funded by an Alexander von Humboldt Research Fellowship for Experienced Researchers within the Continuous Optimization Chair at the Institute for Operations Research, Karlsruhe Institute of Technology (KIT).

[†]School of Mathematical Sciences, University of Southampton (s.cipolla@soton.ac.uk).

[‡]Institute for Operations Research (IOR), Karlsruhe Institute of Technology (KIT) (stein@kit.edu).

[§]School of Mathematical Sciences, University of Southampton (a.b.zemkoho@soton.ac.uk).

can be seen as a special class of Stackelberg (or *bilevel optimization*) problem, where the leader and follower have the same objective function, but which are optimized in opposite directions; i.e., w.r.t. x (resp. y) for the upper-level (resp. lower-level) player. We do not use the specific concept of solution (1.3) in this paper. To introduce the notion of optimal solution used here, consider the optimal value function

$$(1.4) \quad \varphi(x) := \max_{y \in Y(x)} f(x, y)$$

associated to the inner optimization problem (parameterized by $x \in X$)

$$(P_I(x)) \quad \max_y f(x, y) \quad \text{s.t.} \quad A_I x + B_I y - b_I = 0, \quad y \geq 0.$$

To focus our attention on the main ideas, we assume throughout that the corresponding inner optimal solution set is non-empty for all outer variables; i.e.,

$$(1.5) \quad S(x) := \arg \max_{y \in Y(x)} f(x, y) \neq \emptyset \quad \text{for all } x \in X.$$

In view of (1.5), φ is a real-valued function on X . Based on this interpretation, the outer problem in (P) can be equivalently rewritten as

$$(P_O) \quad \min_x \varphi(x) \quad \text{s.t.} \quad A_O x - b_O = 0, \quad x \geq 0.$$

Hence, we use the following concept of optimal solution throughout this paper.

DEFINITION 1.1. *A point $\bar{x}$ will be said to be a local optimal solution of problem (P) if there exists a neighborhood U of the point such that*

$$(1.6) \quad \varphi(\bar{x}) \leq \varphi(x) \quad \text{for all } x \in U \cap X.$$

It goes without saying that if U is the whole space $\mathbb{R}^n$ in (1.6), then the point $\bar{x}$ will be said to be a global optimal solution of problem (P). Other concepts of optimal solutions for minmax problems, where the order of play is important, can be found in [20, 10].

The following concept of stationarity, which will be at the center of the study in this paper, is tightly linked to the notion of local optimal solution in Definition 1.1. In the next definition, the symbol “ $\circ$ ” represents the Hadamard product for two vectors; i.e., $a \circ b := [a_1 b_1, \dots, a_n b_n]^\top$ for $a, b \in \mathbb{R}^n$, which we use for reasons to become apparent later.

DEFINITION 1.2. *A point $\bar{x} \in X$ is called stationary for problem (P) if for some vector $\bar{y} \in S(\bar{x})$, there exist outer and inner multipliers $(\bar{\lambda}_O, \bar{s}_O) \in \mathbb{R}^p \times \mathbb{R}^n$ and $(\bar{\lambda}_I, \bar{s}_I) \in \mathbb{R}^q \times \mathbb{R}^m$, respectively, such that the vector $(\bar{x}, \bar{y}, \bar{\lambda}_O, \bar{s}_O, \bar{\lambda}_I, \bar{s}_I)$ solves the joint outer and inner KKT systems*

$$(1.7) \quad \left. \begin{aligned} Q_{11}x + Q_{12}y + c_x + A_I^\top \lambda_I + A_O^\top \lambda_O - s_O &= 0, \\ A_O x - b_O &= 0, \\ x \circ s_O &= 0, \\ x, s_O &\geq 0, \end{aligned} \right\} \text{outer KKT system}$$

$$\left. \begin{aligned} -Q_{22}y + Q_{12}^\top x + c_y + B_I^\top \lambda_I + s_I &= 0, \\ A_I x + B_I y - b_I &= 0, \\ y \circ s_I &= 0, \\ y, s_I &\geq 0. \end{aligned} \right\} \text{inner KKT system}$$

The inner part of (1.7) corresponds to the KKT conditions of the inner problem $(P_I(x))$ for given $x \in X$. Indeed, in view of our concavity assumption on the objective function of $(P_I(x))$, and the polyhedrality of its feasible set, the condition $y \in S(x)$ is equivalently characterized by y being a KKT point of $(P_I(x))$, i.e., satisfying the inner KKT system with corresponding multipliers λ_I, s_I . In the following we shall abbreviate the set of points (y, λ_I, s_I) solving the inner KKT system for given $x \in X$ by $KKT_I(x)$. Secondly, apart from the pair (x, y) , the outer and inner KKT systems are only connected by the Lagrange multiplier λ_I associated to the coupled constraint in (P). If there are no coupling constraint (i.e., $A_I = 0$) and no coupling variables (i.e., $Q_{12} = 0$), both KKT systems are completely independent from each other.

We shall frequently return to the following running example to illustrate our results.

EXAMPLE 1.3. *The problem data*

$$Q_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad Q_{22} = \begin{bmatrix} 4/3 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad Q_{12} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad c_x = \begin{bmatrix} -\delta \\ 0 \end{bmatrix}, \quad c_y = \begin{bmatrix} 8/3 \\ 0 \\ 0 \end{bmatrix},$$

$$A_O = \begin{bmatrix} 1 & 1 \end{bmatrix}, \quad b_O = 4, \quad A_I = \begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix}, \quad B_I = \begin{bmatrix} -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \quad b_I = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

with parameter $\delta > 4/3$ yield

$$\begin{aligned} X &= \{x \in \mathbb{R}^2 \mid x_1 + x_2 = 4, x \geq 0\}, \\ Y(x) &= \{y \in \mathbb{R}^3 \mid x_1 - y_1 + y_2 = 0, -x_1 + y_1 + y_3 = 1, y \geq 0\} \text{ for } x \in X, \\ f(x, y) &= \frac{1}{2}x_1^2 - \delta x_1 - \frac{2}{3}y_1^2 + \frac{8}{3}y_1. \end{aligned}$$

The point $\bar{x} = [4, 0]^\top$ is stationary for (P), given that $(\bar{x}, \bar{y}, \bar{\lambda}_O, \bar{s}_O, \bar{\lambda}_I, \bar{s}_I)$, with the following values, solves the KKT system (1.7):

$$\bar{x} = \begin{bmatrix} 4 \\ 0 \end{bmatrix}, \quad \bar{y} = \begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix}, \quad \bar{\lambda}_O = \delta - 4/3, \quad \bar{s}_O = \begin{bmatrix} 0 \\ \delta - 4/3 \end{bmatrix}, \quad \bar{\lambda}_I = \begin{bmatrix} -8/3 \\ 0 \end{bmatrix}, \quad \bar{s}_I = \begin{bmatrix} 0 \\ 8/3 \\ 0 \end{bmatrix}.$$

The joint KKT system (1.7) is not new, as it has been derived in various papers, including [42], where some relevant references can be found. More recently, these conditions have been obtained in [8] and have also been used for the development of algorithms as it will become clear in Subsection 1.1. Before proceeding, it is important to note that the primary objective of this paper is to develop a tractable and robust algorithm to compute stationary points in the sense of Definition 1.2. On the other hand, before discussing any algorithmic approach, we first address the following key questions, which do not seem to have been rigorously addressed for the quadratic minmax program in (P): (i) How can we show that a local optimal solution of problem (P), in the sense of (1.6), is a stationary point? (ii) Conversely, under what conditions can a stationary point be locally optimal for problem (P), in the sense of (1.6)? (iii) How is our concept of stationarity (in Definition 1.2) connected to a notion of saddle point associated to the minmax problem (P)? We address these questions in this paper, and subsequently provide a framework to ensure that a stationary point is *non-degenerate* (definition to be provided). We subsequently establish that the non-degeneracy of a stationary point in the sense of Definition 1.2 ensures that such a point serves as an attraction point for our interior point method as it traverses the central path.

1.1 Related work. Problem (P) is a special case of the problem

$$(1.8) \quad \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}(x)} f(x, y),$$

where f is a general real-valued function and $\mathcal{X}$ and $\mathcal{Y}(x)$ (for $x \in X$) are general subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. Various special cases of this problem have been considered in the literature in recent years, including [43, 9, 10, 8, 36, 13, 19, 26, 25, 41, 44], which represent, to the best of our knowledge, the main contributions in the existing literature where the inner feasible set effectively depends on the outer variable. For instance, in [8, 44], $\mathcal{Y}(x)$ is defined by an equality constraint of the form $A_I x + B_I y = b_I$, while it is an inequality $A_I x + B_I y \leq b_I$ in [43]. The constraint is a general inequality $g(x, y) \leq 0$ in [13], and even further generalized in [19], where we have $g(x, y) \in \mathcal{K}$, where $\mathcal{K}$ is closed convex cone in $\mathbb{R}^m$.

Apart from [13, 41], the main idea in all these papers consists of building a surrogate unconstrained minmax problem for (1.8) using a Lagrangian [43, 9, 10, 8, 44] or some augmented Lagrangian-type function [19, 26, 25]. The resulting problem is then solved by some form of gradient descent ascent-type algorithm, building on techniques to solve unconstrained minmax problems. The article [36] is rather mainly focused on using the Lagrangian function to build duality results for a class of problem (1.8). As for [19], it is dedicated to the construction of a special parametric minimization problem using the concept of partial forward-backward envelope. The main effort there is to show that the new problem and the corresponding version of (1.8) can possess the same *first-order minmax point*. It turns out that this concept, which is the main focus of the analysis there, is simply a compact form of our stationarity notion in Definition 1.1.

For the papers [43, 10, 8, 44, 26, 25], which focus on the development of numerical methods, the main effort is concentrated on estimating the iteration complexity for the corresponding first order method. All these algorithms also aim to compute the corresponding approximate versions of the stationarity system (1.7). However, for the articles [13, 41], they use completely different paths and solution concepts to develop algorithms for special versions of problem (1.8). Similarly to the previous papers, [13] uses a first order method, but instead to compute a type of solution point $(\bar{x}, \bar{y})$ called (ϵ, δ) -Stackelberg equilibrium, which is defined by the fulfillment of the conditions

$$\bar{y} \in \mathcal{Y}(\bar{x}) \quad \text{and} \quad \psi(\bar{x}) - \delta \leq f(\bar{x}, \bar{y}) \leq \min_{x \in \mathcal{X}} \psi(x) + \epsilon,$$

where ψ is the optimal value function of associated to the corresponding inner problem of (1.8). Clearly, such a point appears to be some form of approximation of the solution concept in Definition 1.1. As for [41], the algorithm developed there, is, in some sense, the closest to the one proposed and studied in this paper. This is mainly for two reasons: first, the method in [41] is a second order-type algorithm, similarly to our approach. Secondly, it builds from some concept of central path and interior point-type idea. However, the notion of solution there as well as the overall algorithmic process is completely different from what we do here.

1.2 Main contributions. The main contributions of this paper can be summarized as addressing the following two questions:

(Q1) How is the stationary concept in Definition 1.2 related to the notion of local optimal solution in Definition 1.1?

(Q2) How to tractably compute the stationary points in the sense of Definition 1.2?

The motivation for (Q1) is that computing stationary points is important, but it is useful to know whether such points are meaningful with regards to their potential of being locally optimal for problem (P_O) . In particular, we show that if a stationary point is such that anon-degeneracy condition tailored to (P_O) holds, then this point is a local optimal solution for the problem; cf. Section 2.

With regards to (Q2), we introduce log barrier functions tailored to problem (P) that we use to build a central path. We use such a central path to design an infeasible interior point method converging, under a suitable framework provided in Sections 2 and 3, to a stationary point in the sense Definition 1.2. A key feature of our algorithm is that, unlike in descent-ascent methods, where the outer and inner iterates are computed alternately, we do not distinguish them, and therefore calculate our iterates in a *single loop*. In particular, we show that when the inner feasible set is decoupled (i.e., when $A_I = 0$), our method is polynomial; more precisely, we establish the complexity of our algorithm with iteration count $O((n+m)^2 |\log(\epsilon)|)$ to achieve an ϵ -approximation for a solution of the stationarity conditions in Definition 1.2. Moreover, we apply our method to both synthetic data and a class of min cost flow problems, and we demonstrate favorable performance when compared to related methods from the literature; cf. Section 4.

1.3 Organization of the paper. In Section 2, we establish a link between the stationarity concept, the local optimality, and a usual notion of saddle points tailored to problem (P) . For the latter analysis, a nondegeneracy concept is introduced and studied, which also prepares the ground for the convergence analysis of our algorithm, which is introduced in the subsequent section. In particular, in Section 3, we develop an infeasible interior point method that we show to be polynomial in the decoupled case, where Y is independent from x . After conducting our numerical experiments in Section 4, some concluding remarks are provided in Section 5.

2 Stationary and saddle points. The first step in this section is to formally link the stationarity concept in Definition 1.2 and the notion of local optimality for problem (P) in Definition 1.1; cf. Subsection 2.1. In Subsection 2.2, we construct a sufficient condition for a stationary point to be a *strict* local optimal solution for (P) in the sense of (1.6). Subsequently, Subsection 2.3 discusses an appropriate non-degeneracy concept for (P) and, based on this, Subsection 2.4 presents a relationship between a stationary point and a local saddle point concept for (P) .

2.1 An auxiliary quadratic problem and stationarity derivation. In order to be able to derive the stationarity in Definition 1.2 (as necessary optimality condition of (P)) for *free*, as one would expect for any quadratic optimization problem (with affine linear constraints), we exploit

the connection between our problem and the following single-level minimization problem:

$$(QP) \quad \begin{aligned} \min_{x, y, \lambda_I, s_I} \quad & \frac{1}{2} y^\top Q_{22} y + \lambda_I^\top (A_I x - b_I) + \frac{1}{2} x^\top Q_{11} x + c_x^\top x \\ \text{s.t.} \quad & A_O x - b_O = 0, \quad x \geq 0, \\ & -Q_{22} y + c_y + Q_{12}^\top x + B_I^\top \lambda_I + s_I = 0, \quad s_I \geq 0. \end{aligned}$$

The global and local relationships between this problem and problem (P_O) rely on a Wolfe duality result for the inner problem $(P_I(x))$. To establish it, we introduce the following notation for each $x \in X$:

$$\begin{aligned} q(x, y, \lambda_I, s_I) &:= \frac{1}{2} y^\top Q_{22} y + \lambda_I^\top (A_I x - b_I) + \frac{1}{2} x^\top Q_{11} x + c_x^\top x, \\ F(x) &:= \{(y, \lambda_I, s_I) \in \mathbb{R}^m \times \mathbb{R}^q \times \mathbb{R}^m : \\ &\quad -Q_{22} y + c_y + Q_{12}^\top x + B_I^\top \lambda_I + s_I = 0, \quad s_I \geq 0\}, \\ S_D(x) &:= \arg \min_{(y, \lambda_I, s_I) \in F(x)} q(x, y, \lambda_I, s_I), \\ \varphi_D(x) &:= \min_{(y, \lambda_I, s_I) \in F(x)} q(x, y, \lambda_I, s_I). \end{aligned}$$

LEMMA 2.1. *The following assertions hold for each $x \in X$:*

- (a) $\emptyset \neq KKT_I(x) \subseteq S_D(x)$.
- (b) $\varphi(x) = \varphi_D(x)$.

Proof. By (1.5) and the polyhedrality of the inner feasible set $Y(x)$, for each $x \in X$, the set $KKT_I(x)$ is non-empty. Furthermore, observe that with the Lagrangian

$$\begin{aligned} L_I^x(y, \lambda_I, s_I) &= -\frac{1}{2} y^\top Q_{22} y + c_y^\top y + x^\top Q_{12} y + \lambda_I^\top (A_I x + B_I y - b_I) + s_I^\top y + \frac{1}{2} x^\top Q_{11} x + c_x^\top x \\ &= y^\top (-Q_{22} y + c_y + Q_{12}^\top x + B_I^\top \lambda_I + s_I) + \frac{1}{2} y^\top Q_{22} y + \lambda_I^\top (A_I x - b_I) + \frac{1}{2} x^\top Q_{11} x + c_x^\top x \end{aligned}$$

of $(P_I(x))$, the feasible set of its Wolfe dual problem is defined by the stationarity condition

$$0 = \nabla_y L_I^x(y, \lambda_I, s_I) = -Q_{22} y + c_y + Q_{12}^\top x + B_I^\top \lambda_I + s_I$$

and $s_I \geq 0$, and thus its objective function reduces to

$$L_I^x(y, \lambda_I, s_I) = \frac{1}{2} y^\top Q_{22} y + \lambda_I^\top (A_I x - b_I) + \frac{1}{2} x^\top Q_{11} x + c_x^\top x.$$

Therefore the Wolfe dual problem of $(P_I(x))$ can be written as

$$(D_I(x)) \quad \min_{y, \lambda_I, s_I} q(x, y, \lambda_I, s_I) \quad \text{s.t.} \quad (y, \lambda_I, s_I) \in F(x),$$

and weak Wolfe duality [39] yields $\varphi_D(x) \geq \varphi(x)$. Moreover, for any $(y, \lambda_I, s_I) \in KKT_I(x)$ one has $(y, \lambda_I, s_I) \in F(x)$ and $y \in Y(x)$, which yields

$$\varphi_D(x) \leq q(x, y, \lambda_I, s_I) = L_I^x(y, \lambda_I, s_I) = f(x, y) \leq \varphi(x) \leq \varphi_D(x).$$

This shows the strong duality result (b) as well as the second assertion of (a). $\square$

We remark that equality of $KKT_I(x)$ and $S_D(x)$ in Lemma 2.1(a) cannot be expected, since the primal feasibility condition $y \in Y(x)$ is not part of the definition of $F(x)$. Nevertheless, Lemma 2.1(a) shows that, like the optimal solution mapping $S : X \rightrightarrows \mathbb{R}^m$ of $(P_I(x))$, also the optimal solution mapping $S_D : X \rightrightarrows \mathbb{R}^m \times \mathbb{R}^q \times \mathbb{R}^m$ of $(D_I(x))$ has non-empty images for all $x \in X$, so that with φ also φ_D is real-valued on X .

Hence, problem (P_O) can be equivalently written as

$$(P_D) \quad \min_x \varphi_D(x) \quad \text{s.t.} \quad x \in X.$$

Thanks to this dual representation of problem (P_O) , we can now establish the following global equivalence between it and problem (QP) .

PROPOSITION 2.2. *The following statements hold true:*

- (a) Let $\bar{x}$ be a global optimal solution of problem (P_O) . Then, for all $(\bar{y}, \bar{\lambda}_I, \bar{s}_I) \in S_D(\bar{x})$, the quadruple $(\bar{x}, \bar{y}, \bar{\lambda}_I, \bar{s}_I)$ is a global optimal solution of problem (QP) .
- (b) Let $(\bar{x}, \bar{y}, \bar{\lambda}_I, \bar{s}_I)$ be a global optimal solution of problem (QP) . Then the point $\bar{x}$ is a global optimal solution of problem (P_O) .

Proof. For (a), note that $\bar{x}$ is also a global optimal solution of problem (P_D) . Hence, let $(\bar{y}, \bar{\lambda}_I, \bar{s}_I) \in S_D(\bar{x})$, then it follows that for any $x \in X$ and $(y, \lambda_I, s_I) \in F(x)$, we have

$$q(\bar{x}, \bar{y}, \bar{\lambda}_I, \bar{s}_I) = \varphi_D(\bar{x}) \leq \varphi_D(x) \leq q(x, y, \lambda_I, s_I).$$

Hence, $(\bar{x}, \bar{y}, \bar{\lambda}_I, \bar{s}_I)$ is a global optimal solution of problem (QP) .

As for (b), considering the hypothesis, it automatically holds that $(\bar{y}, \bar{\lambda}_I, \bar{s}_I) \in S_D(\bar{x})$. Otherwise, we can find $(\tilde{y}, \tilde{\lambda}_I, \tilde{s}_I) \in F(\bar{x})$ such that

$$q(\bar{x}, \bar{y}, \bar{\lambda}_I, \bar{s}_I) > q(\bar{x}, \tilde{y}, \tilde{\lambda}_I, \tilde{s}_I),$$

and since $(\bar{x}, \tilde{y}, \tilde{\lambda}_I, \tilde{s}_I)$ is a feasible point to problem (QP) , we have a contradiction. Therefore, for any $x \in X$ and $(y, \lambda_I, s_I) \in F(x)$,

$$\varphi_D(\bar{x}) = q(\bar{x}, \bar{y}, \bar{\lambda}_I, \bar{s}_I) \leq q(x, y, \lambda_I, s_I).$$

By Lemma 2.1, we have $S_D(x) \neq \emptyset$, and hence, it follows in particular that

$$\varphi_D(\bar{x}) = q(\bar{x}, \bar{y}, \bar{\lambda}_I, \bar{s}_I) \leq \min_{(y, \lambda_I, s_I) \in F(x)} q(x, y, \lambda_I, s_I) = \varphi_D(x).$$

Therefore, $\bar{x}$ is a global optimal solution of problem (P_D) and hence for problem (P_O) . $\square$

For one direction of the local relationship, we need the inner semicontinuity of the set-valued mapping S_D , which will be said to hold at a point $(\bar{x}, \bar{y}, \bar{\lambda}_I, \bar{s}_I)$ if for any sequence (x^k) converging to $\bar{x}$, there is a sequence $(y^k, \lambda_I^k, s_I^k) \in S_D(x^k)$ that converges to $(\bar{y}, \bar{\lambda}_I, \bar{s}_I)$ as $k \rightarrow \infty$. We will state a sufficient condition for inner semicontinuity of S_D in Remark 2.10.

PROPOSITION 2.3. *The following statements hold true:*

- (a) Let $\bar{x}$ be a local optimal solution of (P_O) . Then, for all $(\bar{y}, \bar{\lambda}_I, \bar{s}_I) \in S_D(\bar{x})$, the quadruple $(\bar{x}, \bar{y}, \bar{\lambda}_I, \bar{s}_I)$ is a local optimal solution of problem (QP) .
- (b) Let the point $(\bar{x}, \bar{y}, \bar{\lambda}_I, \bar{s}_I)$, where S_D is inner semicontinuous, be a local optimal solution of (QP) . Then $\bar{x}$ is a local optimal solution of problem (P_O) .

Proof. For (a), $\bar{x}$ also being a local optimal solution of problem (P_D) , we claim that for any $(\bar{y}, \bar{\lambda}_I, \bar{s}_I) \in S_D(\bar{x})$, the point $(\bar{x}, \bar{y}, \bar{\lambda}_I, \bar{s}_I)$ is automatically a local optimal solution of problem (QP) . Otherwise, consider a point $(\bar{y}, \bar{\lambda}_I, \bar{s}_I) \in S_D(\bar{x})$ such that $(\bar{x}, \bar{y}, \bar{\lambda}_I, \bar{s}_I)$ is not a local optimal solution of problem (QP) . Then we can find a sequence $(x^k, y^k, \lambda_I^k, s_I^k)$ from the graph of F on X with $x^k \rightarrow \bar{x}$, $y^k \rightarrow \bar{y}$, $\lambda_I^k \rightarrow \bar{\lambda}_I$, and $s_I^k \rightarrow \bar{s}_I$ such that we have

$$q(x^k, y^k, \lambda_I^k, s_I^k) < q(\bar{x}, \bar{y}, \bar{\lambda}_I, \bar{s}_I) = \varphi_D(\bar{x}) \text{ for all } k.$$

Then considering the definition of φ_D , it follows that

$$\varphi_D(x^k) \leq q(x^k, y^k, \lambda_I^k, s_I^k) < q(\bar{x}, \bar{y}, \bar{\lambda}_I, \bar{s}_I) = \varphi_D(\bar{x}) \text{ for all } k.$$

This clearly contradicts the fact $\bar{x}$ is a local optimal solution of problem (P_D) given that $x^k \in X$ for all k . Now, given that $S(\bar{x}) \neq \emptyset$, it follows that $S_D(\bar{x}) \neq \emptyset$ by the strong Wolfe duality result. Thus, there exists some $(\bar{y}, \bar{\lambda}_I, \bar{s}_I) \in S_D(\bar{x})$ such that $(\bar{x}, \bar{y}, \bar{\lambda}_I, \bar{s}_I)$ is locally optimal for (QP) .

As for (b), if $\bar{x}$ is not a local optimal solution of problem (P_D) , then we can find a feasible sequence $x^k \rightarrow \bar{x}$ such that $\varphi_D(\bar{x}) > \varphi_D(x^k)$ for all k . As the set-valued mapping S_D is inner semicontinuous at $(\bar{x}, \bar{y}, \bar{\lambda}_I, \bar{s}_I)$, we can find a sequence $(y^k, \lambda_I^k, s_I^k) \in S_D(x^k)$ that converges to $(\bar{y}, \bar{\lambda}_I, \bar{s}_I)$. It follows by the construction that

$$q(\bar{x}, \bar{y}, \bar{\lambda}_I, \bar{s}_I) = \varphi_D(\bar{x}) > \varphi_D(x^k) = q(x^k, y^k, \lambda_I^k, s_I^k)$$

with $x^k \in X$, $(y^k, \lambda_I^k, s_I^k) \in F(x^k)$ for all k , which contradicts the assumed local minimality of the point $(\bar{x}, \bar{y}, \bar{\lambda}_I, \bar{s}_I)$ for problem (QP) . $\square$

Remark 2.4. In the special case $A_I = 0$, when no coupling constraints in the inner problem $(P_I(x))$ exist and the outer and inner problems are only coupled via the matrix Q_{12} , the optimal value function φ (1.4) is convex as the pointwise maximum of convex functions. Therefore, (P_O)

and (P_D) are convex, possibly nonsmooth, optimization problems. Similarly, in this case, also (QP) is a convex problem. Therefore, for all these problems local minimal points are necessarily global minimal points. Hence, in the case $A_I = 0$, by Proposition 2.2 we can state that these problems are equivalent without distinguishing global and local minimal points. In particular, in this case, the inner semicontinuity of S_D required in Proposition 2.3 is not necessary.

Finally, thanks to Proposition 2.3(a), we can derive the main result of this subsection.

COROLLARY 2.5. *If $\bar{x}$ is locally optimal for problem (P_O) , then it is stationary.*

Proof. Since $S_D(\bar{x}) \neq \emptyset$, it follows from Proposition 2.3(a) that there exists $(\bar{y}, \bar{\lambda}_I, \bar{s}_I) \in S_D(\bar{x})$ such that $(\bar{x}, \bar{y}, \bar{\lambda}_I, \bar{s}_I)$ is a locally optimal for (QP) . Since the feasible set of the latter problem is defined only by affine linear constraints, with its Lagrangian

$$\begin{aligned} \mathcal{L}(x, y, \lambda_I, s_I, \lambda_O, s_O, \gamma, \tau) := & \frac{1}{2} y^\top Q_{22} y + \lambda_I^\top (A_I x - b_I) + \frac{1}{2} x^\top Q_{11} x + c_x^\top x \\ & + \lambda_O^\top (A_O x - b_O) - s_O^\top x + \gamma^\top (-Q_{22} y + c_y + Q_{12}^\top x + B_I^\top \lambda_I + s_I) - \tau^\top s_I \end{aligned}$$

we can find a multiplier vector $(\lambda_O, s_O, \gamma, \tau)$ such that the KKT conditions

$$\begin{aligned} \nabla_{(x, y, \lambda_I, s_I)} \mathcal{L}(\bar{x}, \bar{y}, \lambda_I, s_I, \lambda_O, s_O, \gamma, \tau) &= 0, \\ A_O \bar{x} - b_O &= 0, \\ \bar{x} \circ s_O &= 0, \\ \bar{x}, s_O &\geq 0, \\ -Q_{22} \bar{y} + c_y + Q_{12}^\top \bar{x} + B_I^\top \lambda_I + s_I &= 0, \\ \tau \circ s_I &= 0, \\ \tau, s_I &\geq 0 \end{aligned}$$

are satisfied. The equations corresponding to $\nabla_y \mathcal{L}$ and $\nabla_{s_I} \mathcal{L}$ in the first line of the system particularly yield $Q_{22} \bar{y} = Q_{22} \gamma$ and $\tau = \gamma$, respectively. Using these identities to write τ and $Q_{22} \bar{y}$ in terms of γ , and calculating $\nabla_x \mathcal{L}$ and $\nabla_{\lambda_I} \mathcal{L}$, leads to an equivalent system which coincides with (1.7), where γ solves the inner KKT system with multipliers λ_I and s_I . This implies $\gamma \in S(\bar{x})$ and concludes the proof. $\square$

Remark 2.6. We point out that in the above proof of Corollary 2.5 the firstly chosen point $\bar{y}$ is not the one which necessarily leads to the stationarity condition, but the possibly different point $\gamma \in S(\bar{x})$. In particular, the primal feasibility conditions $A_I \bar{x} + B_I y = b_I$, $y \geq 0$ do not necessarily hold for $\bar{y}$, but they do hold for γ . Also note that γ is a dual variable corresponding to an equation from the Wolfe dual of the inner problem and, thus, a “bi-dual variable”. Therefore it makes sense that it lies in the minimal point set $S(\bar{x})$ of the primal inner problem.

To conclude this subsection, note that the stationarity system in (1.7) can be derived by multiple other ways. For instance, if the optimal solution set-valued mapping S reduces to a single-valued and continuously differentiable function, the stationarity conditions in Definition 1.2 could also be obtained from the Lagrange multiplier rule of the minimization of the function $f(x, y(x))$ subject to $x \in X$, under the non-degeneracy of the outer and inner points of interest $\bar{x}$ and $\bar{y} \in S(\bar{x})$, respectively, cf. Subsection 2.2.

2.2 On the meaningfulness of stationarity. As the algorithm that we introduce in this paper aims to compute stationary points in the sense of Definition 1.2, an interesting question is to know whether such points are actually meaningful for problem (P) . The *meaningfulness* here is in terms of whether such points are connected to a solution concept associated to problem (P) . Indeed, this is the question that we address in this subsection. More precisely, we construct here a framework that ensures that a stationary point, in the sense of Definition 1.2, is a local optimal solution of problem (P) , in the sense of Definition 1.1.

This framework uses inner and outer non-degeneracy conditions. Let

$$\mathcal{I}_{\bar{x}} := \{i = 1, \dots, n : \bar{x}_i = 0\} \text{ and } \mathcal{I}_{\bar{y}} := \{j = 1, \dots, m : \bar{y}_j = 0\}$$

denote the active index sets of the inequality constraints $x \geq 0$ at $\bar{x}$ and of $y \geq 0$ at $\bar{y}$, respectively. Furthermore, $\mathbb{1}_{\bar{x}}$ and $\mathbb{1}_{\bar{y}}$ stand for the matrices with row vectors $(e_n^i)^\top$, $i \in \mathcal{I}_{\bar{x}}$, and $(e_m^k)^\top$, $j \in \mathcal{I}_{\bar{y}}$, respectively. We say that the outer linear independence constraint qualification (OLICQ) holds at the point $\bar{x} \in X$ if the matrix

$$(\text{OLICQ}) \quad [A_O^\top, \mathbb{1}_{\bar{x}}^\top] \text{ possesses full column rank,}$$

and similarly, the inner linear independence constraint qualification (ILICQ) holds at $\bar{y} \in Y(\bar{x})$ if the matrix

$$(\text{ILICQ}) \quad [B_I^\top, \mathbf{1}_{\bar{y}}^\top] \text{ possesses full column rank.}$$

Given a stationary point $\bar{x}$ of (P) with some $\bar{y} \in S(\bar{x})$, the (ILICQ) implies that the corresponding inner multipliers $\bar{\lambda}_I$, $\bar{s}_I$ are uniquely determined, and therefore, under the (ILICQ), the (OLICQ) implies the same for the outer multipliers $\bar{\lambda}_O$, $\bar{s}_O$.

Under the (ILICQ), inner strict complementary slackness (ISCS) is said to hold in the case

$$(\text{ISCS}) \quad \min_{i=1, \dots, m} ([\bar{s}_I]_i + [\bar{y}]_i) > 0,$$

and under the (ILICQ) and (OLICQ), outer strict complementary slackness (OSCS) if

$$(\text{OSCS}) \quad \min_{i=1, \dots, n} ([\bar{s}_O]_i + [\bar{x}]_i) > 0.$$

Finally, we state inner and outer second order sufficiency conditions. To this end, for $\bar{x} \in X$, let

$$(2.1) \quad T(\bar{y}, Y(\bar{x})) := \ker \begin{bmatrix} B_I \\ \mathbf{1}_{\bar{y}} \end{bmatrix}$$

denote the tangent space to $Y(\bar{x})$ at $\bar{y}$. Then the inner second order sufficiency condition (ISOSC) is said to hold if

$$(\text{ISOSC}) \quad \eta^\top (-Q_{22}) \eta < 0 \quad \text{for all } \eta \in T(\bar{y}, Y(\bar{x})) \setminus \{0\}.$$

Analogously, let

$$(2.2) \quad T(\bar{x}, X) = \ker \begin{bmatrix} A_O \\ \mathbf{1}_{\bar{x}} \end{bmatrix}$$

denote the tangent space to X at $\bar{x}$. Under (ILICQ), (ISCS) and (ISOSC) the outer second order sufficiency condition (OSOSC) is said to hold if

$$(\text{OSOSC}) \quad d^\top Q'_{11} d > 0 \quad \text{for all } d \in T(\bar{x}, X) \setminus \{0\},$$

where

$$(2.3) \quad Q'_{11} := Q_{11} - \begin{bmatrix} Q_{12} & A_I^\top & 0 \end{bmatrix} \begin{bmatrix} -Q_{22} & B_I^\top & \mathbf{1}_{\bar{y}}^\top \\ B_I & 0 & 0 \\ \mathbf{1}_{\bar{y}} & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} Q_{12}^\top \\ A_I \\ 0 \end{bmatrix}.$$

The inverted matrix in (2.3) is indeed non-singular, which can be shown with the techniques from [12, 21]. For completeness, and since these results also play a crucial role in Subsection 2.3, we briefly recall the corresponding arguments. More details are given in Subsection 2.3.

Let us abbreviate the system of equations in the inner KKT system in (1.7) by

$$(2.4) \quad 0 = \Phi_I(x, y, \lambda_I, s_I) := \begin{bmatrix} -Q_{22}y + Q_{12}^\top x + c_y + B_I^\top \lambda_I + s_I \\ A_I x + B_I y - b_I \\ s_I \circ y \end{bmatrix}.$$

The Jacobian with respect to the inner variables and multipliers (y, λ_I, s_I) of Φ_I at a stationary point $(\bar{x}, \bar{y}, \bar{\lambda}_I, \bar{s}_I)$ is

$$(2.5) \quad D_{(y, \lambda_I, s_I)} \Phi_I(\bar{x}, \bar{y}, \bar{\lambda}_I, \bar{s}_I) = \begin{bmatrix} -Q_{22} & B_I^\top & \mathbf{1} \\ B_I & 0 & 0 \\ \text{diag}(\bar{s}_I) & 0 & \text{diag}(\bar{y}) \end{bmatrix},$$

where, e.g., $\text{diag}(\bar{s}_I)$ stands for the diagonal matrix with diagonal entries from $\bar{s}_I$ and results as the partial Jacobian of the Hadamard product $s_I \circ y$ with respect to y . Moreover, here and in the following $\mathbf{1}$ stands for the identity matrix of appropriate dimension. $D_{(y, \lambda_I, s_I)} \Phi_I(\bar{x}, \bar{y}, \bar{\lambda}_I, \bar{s}_I)$ is non-singular iff (ISCS) holds and the matrix

$$(2.6) \quad \begin{bmatrix} -Q_{22} & B_I^\top & \mathbf{1}_{\bar{y}}^\top \\ B_I & 0 & 0 \\ \mathbf{1}_{\bar{y}} & 0 & 0 \end{bmatrix}$$

is non-singular. The latter is the case under **(ILICQ)** and **(ISOSC)** (cf. Lemma 2.13).

Under **(ILICQ)**, **(ISCS)** and **(ISOSC)**, this result allows us to apply the Implicit Function Theorem to the system $\Phi_I(x, y, \lambda_I, s_I) = 0$ at $(\bar{x}, \bar{y}, \bar{\lambda}_I, \bar{s}_I)$, which yields the existence of a unique smooth function $(y(x), \lambda_I(x), s_I(x))$ with $(y(\bar{x}), \lambda_I(\bar{x}), s_I(\bar{x})) = (\bar{y}, \bar{\lambda}_I, \bar{s}_I)$ and $\Phi_I(x, y(x), \lambda_I(x), s_I(x)) \equiv 0$ for all x from some neighborhood U of $\bar{x}$. The **(ISCS)** and continuity arguments also imply $y(x) \geq 0$ and $s_I(x) \geq 0$ for sufficiently small U , so that $y(x) \in S(x)$ holds for all $x \in U$. This yields the local description $\varphi(x) = f(x, y(x))$ of the inner maximal value function for all $x \in U$. In particular, the function φ is smooth around $\bar{x}$ and, therefore, local optimality conditions for (P_O) can be obtained from conditions for smooth problems (cf. also [18, 38]). To formulate them, some computations based on the Implicit Function Theorem yield $\nabla\varphi(\bar{x}) = Q_{11}\bar{x} + Q_{12}\bar{y} + c_x + A_I^\top \bar{\lambda}_I$ and $D^2\varphi(\bar{x}) = Q'_{11}$ with the matrix from (2.3).

THEOREM 2.7. *Let $\bar{x}$ be a stationary point in the sense of Definition 1.2 such that the corresponding pair $(\bar{x}, \bar{y})$ with $\bar{y} \in S(\bar{x})$ satisfies **(ILICQ)**, **(ISCS)**, **(ISOSC)**, **(OLICQ)**, **(OSCS)** and **(OSOSC)**. Then $\bar{x}$ is a strict local minimal point for problem (P_O) .*

Proof. As seen above, under **(ILICQ)**, **(ISCS)**, **(ISOSC)**, with the implicitly defined function $y(x)$ the problem (P_O) is locally equivalent to the minimization of $\varphi(x) = f(x, y(x))$ over X . In particular, the strict local optimal solutions of both problems coincide.

A standard second order sufficiency condition for $\bar{x} \in X$ to be a strict local optimal solution of $\varphi(x) = f(x, y(x))$ on X consists of the KKT condition

$$\begin{aligned} \nabla\varphi(\bar{x}) + A_O^\top \bar{\lambda}_O - \bar{s}_O &= 0, \\ A_O \bar{x} - b_O &= 0, \\ \bar{x} \circ \bar{s}_O &= 0, \\ \bar{x}, \bar{s}_O &\geq 0 \end{aligned}$$

with multipliers $\bar{\lambda}_O, \bar{s}_O$, the **(OLICQ)**, **(OSCS)** as well as the second order condition

$$d^\top D^2\varphi(\bar{x})d > 0 \quad \text{for all } d \in T(\bar{x}, X) \setminus \{0\}.$$

In view of the above formulas for $\nabla\varphi(\bar{x})$ and $D^2\varphi(\bar{x})$ this means that the triple $(\bar{x}, \bar{\lambda}_O, \bar{s}_O)$ must satisfy the outer KKT system in (1.7), **(OLICQ)**, **(OSCS)** and **(OSOSC)**. Since all these conditions are covered by the assumptions, $\bar{x}$ is a strict local optimal solution of $\varphi(x) = f(x, y(x))$ on X , which shows the assertion. $\square$

Remark 2.8. The simultaneous satisfaction of the inner and outer non-degeneracy assumptions **(ILICQ)**, **(ISCS)**, **(ISOSC)**, **(OLICQ)**, **(OSCS)** and **(OSOSC)** is a mild requirement in a topological sense. In fact, the epigraph reformulation [34] of (P) yields a generalized semi-infinite optimization problem (GSIP), and the corresponding non-degeneracy conditions are known to hold generically at each solution point of a GSIP (cf. [35] for the linear and [16] for the nonlinear case).

EXAMPLE 2.9. *In the situation of Example 1.3 one obtains $\mathcal{I}_{\bar{x}} = \mathcal{I}_{\bar{y}} = \{2\}$ and*

$$[A_O^\top, \mathbb{1}_{\bar{x}}^\top] = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \quad [B_I^\top, \mathbb{1}_{\bar{y}}^\top] = \begin{pmatrix} -1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$

so that **(OLICQ)** and **(ILICQ)** hold at $\bar{x}$ and $\bar{y} \in Y(\bar{x})$, respectively. Moreover, one obtains $T(\bar{x}, X) \setminus \{0\} = \emptyset$ and $T(\bar{y}, Y(\bar{x})) \setminus \{0\} = \emptyset$, so that **(OSOSC)** and **(ISOSC)** are trivially satisfied. Hence, Theorem 2.7 ensures that $\bar{x}$ is a strict local optimal solution of (P_O) .

Remark 2.10. As seen above, at a point $(\bar{x}, \bar{y}, \bar{\lambda}_I, \bar{s}_I)$ with $(\bar{y}, \bar{\lambda}_I, \bar{s}_I) \in KKT_I(\bar{x})$ the conditions **(ILICQ)**, **(ISCS)** and **(ISOSC)** imply the existence of a unique locally defined smooth function $(y(x), \lambda_I(x), s_I(x))$ which satisfies $(y(\bar{x}), \lambda_I(\bar{x}), s_I(\bar{x})) = (\bar{y}, \bar{\lambda}_I, \bar{s}_I)$ as well as $KKT_I(x) = \{(y(x), \lambda_I(x), s_I(x))\}$ for all x from some neighborhood U of $\bar{x}$. Therefore, in view of $KKT_I(x) \subseteq S_D(x)$ (Lemma 2.1), the inner semicontinuity of S_D required in Proposition 2.3(b) holds at every such point $(\bar{x}, \bar{y}, \bar{\lambda}_I, \bar{s}_I)$.

Remark 2.11. The assertion of Theorem 2.7 remains valid in the absence of **(OLICQ)** and **(OSCS)**, if in **(OSOSC)** the tangent space $T(\bar{x}, X)$ is replaced by the outer critical cone

$$\{d \in \mathbb{R}^n \mid (Q_{11}\bar{x} + Q_{12}\bar{y} + c_x + A_I^\top \bar{\lambda}_I)^\top d = 0, A_O d = 0, d_i \geq 0, i \in \mathcal{I}_{\bar{x}}\}.$$

To close this subsection, it might also be useful to note that a meaningfulness result (i.e., ensuring the same conclusion as in Theorem 2.7 for a stationary point in the sense of Definition 1.2) can be obtained without the need of the inner or outer strict complementarity condition; see Section A for an alternative result to Theorem 2.7 with the corresponding proof.

2.3 Non-degenerate stationary points. In Section 3, we will devise an interior point method (IPM) to approximate a stationary point of problem (P) in the sense of Definition 1.2. The convergence proof will require the non-singularity of the Jacobian of the system of equations appearing in (1.7), which we abbreviate by

$$0 = \Phi(x, \lambda_O, s_O, y, \lambda_I, s_I) := \begin{bmatrix} Q_{11}x + Q_{12}y + c_x + A_I^\top \lambda_I + A_O^\top \lambda_O - s_O \\ A_O x - b_O \\ s_O \circ x \\ -Q_{22}y + Q_{12}^\top x + c_y + B_I^\top \lambda_I + s_I \\ A_I x + B_I y - b_I \\ s_I \circ y \end{bmatrix}.$$

The Jacobian of Φ at $(\bar{x}, \bar{\lambda}_O, \bar{s}_O, \bar{y}, \bar{\lambda}_I, \bar{s}_I)$ is

$$(2.7) \quad D\Phi(\bar{x}, \bar{\lambda}_O, \bar{s}_O, \bar{y}, \bar{\lambda}_I, \bar{s}_I) = \begin{bmatrix} Q_{11} & A_O^\top & -\mathbf{1} & Q_{12} & A_I^\top & 0 \\ A_O & 0 & 0 & 0 & 0 & 0 \\ \text{diag}(\bar{s}_O) & 0 & \text{diag}(\bar{x}) & 0 & 0 & 0 \\ Q_{12}^\top & 0 & 0 & -Q_{22} & B_I^\top & \mathbf{1} \\ A_I & 0 & 0 & B_I & 0 & 0 \\ 0 & 0 & 0 & \text{diag}(\bar{s}_I) & 0 & \text{diag}(\bar{y}) \end{bmatrix}.$$

Note that for the analysis of the matrix $D\Phi(\bar{x}, \bar{\lambda}_O, \bar{s}_O, \bar{y}, \bar{\lambda}_I, \bar{s}_I)$ at a solution point of (1.7) we may employ the additional conditions $\bar{x}, \bar{s}_O, \bar{y}, \bar{s}_I \geq 0$ as well as $\bar{y} \in S(\bar{x})$.

DEFINITION 2.12. For a stationary point $\bar{x}$ of (P) with $\bar{y} \in S(\bar{x})$ and multipliers $(\bar{\lambda}_O, \bar{s}_O, \bar{\lambda}_I, \bar{s}_I)$ we call $(\bar{x}, \bar{y}, \bar{\lambda}_O, \bar{\lambda}_I, \bar{s}_O, \bar{s}_I)$ non-degenerate if $D\Phi(\bar{x}, \bar{\lambda}_O, \bar{s}_O, \bar{y}, \bar{\lambda}_I, \bar{s}_I)$ is non-singular. We call $\bar{x}$ with $\bar{y} \in S(\bar{x})$ a non-degenerate stationary point if it is stationary and if multipliers $(\bar{\lambda}_O, \bar{s}_O, \bar{\lambda}_I, \bar{s}_I)$ exist such that $(\bar{x}, \bar{y}, \bar{\lambda}_O, \bar{\lambda}_I, \bar{s}_O, \bar{s}_I)$ is non-degenerate.

A natural sufficient condition for the non-singularity of $D\Phi(\bar{x}, \bar{\lambda}_O, \bar{s}_O, \bar{y}, \bar{\lambda}_I, \bar{s}_I)$ at a solution point of (1.7) arises from its block structure. It is strongly related to the notions of (ILICQ), (ISCS), (ISOSC), (OLICQ), (OSCS) and (OSOSC) introduced above. To see this, we start with a discussion of the non-singularity of the lower right 3-by-3 block of $D\Phi(\bar{x}, \bar{\lambda}_O, \bar{s}_O, \bar{y}, \bar{\lambda}_I, \bar{s}_I)$, namely the Jacobian $D_{(y, \lambda_I, s_I)} \Phi_I(\bar{x}, \bar{y}, \bar{\lambda}_I, \bar{s}_I)$ of Φ_I from (2.4) (see (2.5)). By Laplace expansion of its determinant it is not hard to see that it is non-singular iff (ISCS) holds and the matrix from (2.6) is non-singular.

The latter can be characterized by the following result from [15, Lemma 3.4] (see also [22]), where the inertia $\text{In}(C)$ of a square matrix C denotes the triple of the numbers of its positive, negative, and zero eigenvalues. Furthermore, for $A = A^\top \in \mathbb{R}^{q \times q}$ and $B \in \mathbb{R}^{t \times q}$ let the columns of N form a basis of $\ker B$. Then $A|_{\ker B} := N^\top A N$ is called the restriction of A to $\ker B$.

LEMMA 2.13. Let $A = A^\top \in \mathbb{R}^{q \times q}$ and let $B \in \mathbb{R}^{t \times q}$ possess rank r . Then the matrix

$$C = \begin{bmatrix} A & B^\top \\ B & 0 \end{bmatrix}$$

satisfies

$$(2.8) \quad \text{In}(C) = \text{In}(A|_{\ker B}) + (r, r, t - r).$$

Remark 2.14. In the statement of Lemma 2.13, $t = q = r$ leads to $\ker B = \{0\}$, so that no matrix N can be chosen. This case can be formally encompassed by considering a $(q, 0)$ -matrix N , which makes $N^\top A N$ a $(0, 0)$ -matrix with inertia $(0, 0, 0)$. With this convention, the inertia formula of the lemma holds also in this case. In particular, for $A = 0$ and $t = q = r$ not only the matrix C is non-singular, but also the empty matrix $0|_{\{0\}}$ is, due the absence of vanishing eigenvalues.

With the notation from Lemma 2.13, by (2.8) C is non-singular iff $A|_{\ker B}$ is non-singular and B possesses full row rank. Therefore, the non-singularity of the matrix in (2.6) is equivalent to the non-singularity of $-Q_{22}$ restricted to the tangent space (cf. (2.1))

$$T(\bar{y}, Y(\bar{x})) = \ker \begin{bmatrix} B_I \\ \mathbf{1}_{\bar{y}} \end{bmatrix}$$

and to the full row rank of the matrix $\begin{bmatrix} B_I \\ \mathbf{1}_{\bar{y}} \end{bmatrix}$, which coincides with **(ILICQ)**. Since we assume $-Q_{22}$ to be negative semi-definite, the non-singularity of $-Q_{22}|_{T(\bar{y}, Y(\bar{x}))}$ is equivalent to its negative definiteness, which coincides with **(ISOSC)**. This motivates the following definition.

DEFINITION 2.15. For $\bar{x} \in X$, a point $\bar{y} \in S(\bar{x})$ is called *inner non-degenerate* if **(ILICQ)**, **(ISCS)** and **(ISOSC)** are satisfied.

The above derivation shows the following result.

LEMMA 2.16. For $\bar{x} \in X$ a point $\bar{y} \in S(\bar{x})$ is inner non-degenerate if and only if the Jacobian $D_{(y, \lambda_I, s_I)} \Phi_I(\bar{x}, \bar{y}, \bar{\lambda}_I, \bar{s}_I)$ is non-singular.

Remark 2.17. We emphasize that even for a linear inner problem $(P_I(\bar{x}))$ a point $\bar{y} \in S(\bar{x})$ may be inner non-degenerate. Linearity implies $Q_{22} = 0$, so that negative definiteness of $-Q_{22}|_{T(\bar{x}, Y(\bar{x}))}$ requires $T(\bar{x}, Y(\bar{x})) = \{0\}$ (see Remark 2.14). For an inner non-degenerate point $\bar{y}$ this means that **(ILICQ)** must hold with $|\mathcal{I}_{\bar{y}}| = m - q$, so that in linear programming terminology $\bar{y}$ is a non-degenerate vertex of $Y(\bar{x})$. Moreover, **(ISCS)** yields $[s_I]_{\mathcal{I}_{\bar{y}}} > 0$ in this case, which in linear programming is sometimes called dual non-degeneracy of $\bar{y}$ for $(P_I(\bar{x}))$.

Next we discuss the non-singularity of the full Jacobian matrix given in (2.7). Again, by Laplace expansion of the determinant of $D\Phi(\bar{x}, \bar{\lambda}_O, \bar{s}_O, \bar{y}, \bar{\lambda}_I, \bar{s}_I)$, one sees that its non-singularity is equivalent to **(OSCS)**, **(ISCS)** and the non-singularity of

$$(2.9) \quad K := \begin{bmatrix} Q_{11} & A_O^\top & -\mathbf{1}_{\bar{x}}^\top & Q_{12} & A_I^\top & 0 \\ A_O & 0 & 0 & 0 & 0 & 0 \\ -\mathbf{1}_{\bar{x}} & 0 & 0 & 0 & 0 & 0 \\ Q_{12}^\top & 0 & 0 & -Q_{22} & B_I^\top & \mathbf{1}_{\bar{y}}^\top \\ A_I & 0 & 0 & B_I & 0 & 0 \\ 0 & 0 & 0 & \mathbf{1}_{\bar{y}} & 0 & 0 \end{bmatrix} = \begin{bmatrix} K_{11} & K_{12} \\ K_{12}^\top & K_{22} \end{bmatrix}$$

with

$$K_{11} := \begin{bmatrix} Q_{11} & A_O^\top & -\mathbf{1}_{\bar{x}}^\top \\ A_O & 0 & 0 \\ -\mathbf{1}_{\bar{x}} & 0 & 0 \end{bmatrix}, \quad K_{22} := \begin{bmatrix} -Q_{22} & B_I^\top & \mathbf{1}_{\bar{y}}^\top \\ B_I & 0 & 0 \\ \mathbf{1}_{\bar{y}} & 0 & 0 \end{bmatrix}, \quad K_{12} := \begin{bmatrix} Q_{12} & A_I^\top & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

and where K_{22} is the matrix from (2.6). Hence, if we assume $\bar{y}$ to be inner non-degenerate, then K_{22} is non-singular by Lemma 2.16, and K is non-singular if and only if the Schur complement

$$K/K_{22} = K_{11} - K_{12}K_{22}^{-1}K_{12}^\top = \begin{bmatrix} Q'_{11} & A_O^\top & -\mathbf{1}_{\bar{x}}^\top \\ A_O & 0 & 0 \\ -\mathbf{1}_{\bar{x}} & 0 & 0 \end{bmatrix}$$

with Q'_{11} from (2.3) is. By Lemma 2.13, and since the sign of $\mathbf{1}_{\bar{x}}$ is irrelevant for this argument, the non-singularity of K/K_{22} is equivalent to **(OLICQ)** at $\bar{x}$ in X and the non-singularity of $Q'_{11}|_{T(\bar{x}, X)}$, where $T(\bar{x}, X)$ is the tangent space from (2.2). While above we used that the non-singularity of $-Q_{22}|_{T(\bar{y}, Y(\bar{x}))}$ is equivalent to its negative definiteness and, thus, to **(ISOSC)** at any solution of (1.7), an analogous argument cannot be made for $Q'_{11}|_{T(\bar{x}, X)}$. Therefore we introduce the outer second order *regularity* condition

$$(OSORC) \quad Q'_{11}|_{T(\bar{x}, X)} \text{ is non-singular,}$$

which is weaker than **(OSOSC)**.

DEFINITION 2.18. A stationary point $\bar{x}$ of (P) with inner non-degenerate $\bar{y} \in S(\bar{x})$ is called *outer non-degenerate*, if **(OLICQ)**, **(OSCS)**, and **(OSORC)** hold. If instead of **(OSORC)** even **(OSOSC)** holds, then $\bar{x}$ is called a *non-degenerate local minimal point* (cf. Theorem 2.7).

The above derivation proves the following result.

THEOREM 2.19. For $\bar{x} \in X$, let $\bar{y} \in S(\bar{x})$ be inner non-degenerate. Then $\bar{x}$ is a non-degenerate stationary point of (P) if and only if $\bar{x}$ is stationary and outer non-degenerate.

EXAMPLE 2.20. In the situation of Example 1.3 the considerations from Example 2.9 yield that **(ISOSC)** and **(OSORC)** are trivially satisfied (Remark 2.14). Hence, by Theorem 2.19 $\bar{x}$ is a non-degenerate stationary point of (P) . This illustrates, in particular, that the singularity of Q_{11} and Q_{22} does not prevent non-degenerate stationarity. In addition, this situation is stable under variations of the parameter $\delta > 4/3$.

2.4 Non-degenerate stationarity and local saddle points. A main idea of the algorithmic approach presented in Section 3 is to treat the variables x and y not separately, but as a joint variable $z = (x, y)$. Indeed, standard solution methods for minmax problems alternately solve the outer problem in x and the inner problem in y , while our proposal is a single loop method in the sense that it solves a joint problem in the joint variable z . In this joint problem, the stationary points of (P) in the sense of Definition 1.2 correspond to local saddle points. The corresponding results need a reordering of the variable vector as well as of the equations in (1.7). Since the Jacobian matrix of the reordered system will turn out to be congruent to $D\Phi(\bar{x}, \bar{\lambda}_O, \bar{s}_O, \bar{y}, \bar{\lambda}_I, \bar{s}_I)$, it is also non-singular under the assumptions of Theorem 2.19, i.e., under the conditions derived for the separate variable model.

In the following we will study a local saddle point property of the function f on the combined outer and inner feasible set

$$\begin{aligned}\Omega &= \{(x, y) \in X \times \mathbb{R}^m : y \in Y(x)\} \\ &= \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m : A_O x = b_O, A_I x + B_I y = b_I, x, y \geq 0\}.\end{aligned}$$

In fact, Ω is the graph of the set-valued mapping $Y : X \rightrightarrows \mathbb{R}^m$. Furthermore, let

$$(2.10) \quad z := \begin{bmatrix} x \\ y \end{bmatrix}, \quad Q := \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{12}^\top & -Q_{22} \end{bmatrix}, \quad c := \begin{bmatrix} c_x \\ c_y \end{bmatrix}, \quad A := \begin{bmatrix} A_O & 0 \\ A_I & B_I \end{bmatrix}, \quad b := \begin{bmatrix} b_O \\ b_I \end{bmatrix},$$

and consider the function $f(z) = \frac{1}{2} z^\top Q z + c^\top z$ on the set Ω , which can now be written compactly as $\Omega = \{z \in \mathbb{R}^{n+m} : Az = b, z \geq 0\}$.

Recall that for an unconstrained differentiable function a point is called a local saddle point if it is a critical point, but neither a local minimal nor a local maximal point. Therefore, each neighborhood of the critical point contains points with smaller as well as points with larger function values than at the critical point. For a detailed comparison of the concepts of local and global saddle points we refer to [33, Subsection 3.5].

The generalization of this concept to constrained functions needs a stationarity concept which encompasses, both, local minimal and local maximal points. Since this is not the case for KKT points, following the terminology from [21] we will call $\bar{z}$ a critical point of f on Ω if a relaxed KKT system is satisfied which does not require the non-negativity of multipliers corresponding to inequality constraints.

DEFINITION 2.21. A point $\bar{z} \in \Omega$ is called

(a) *critical point of f on Ω if multipliers $\bar{\lambda} \in \mathbb{R}^{p+q}$, $\bar{\sigma} \in \mathbb{R}^{n+m}$ exist such that $(\bar{z}, \bar{\lambda}, \bar{\sigma})$ solves the system*

$$(2.11) \quad 0 = \Psi(z, \lambda, \sigma) := \begin{bmatrix} Qz + c + A^\top \lambda + \sigma \\ Az - b \\ z \circ \sigma \end{bmatrix},$$

(b) *local saddle point of f on Ω if $\bar{z}$ is a critical point, but neither a local minimal nor a local maximal point of f on Ω .*

LEMMA 2.22. A point $\bar{x} \in X$ with $\bar{y} \in S(\bar{x})$ is stationary for (P) with multipliers $\bar{\lambda}_O, \bar{s}_O, \bar{\lambda}_I, \bar{s}_I$ if and only if $\bar{z} = (\bar{x}, \bar{y}) \in \Omega$ is a critical point of f on Ω with multipliers

$$\bar{\lambda} = \begin{bmatrix} \bar{\lambda}_O \\ \bar{\lambda}_I \end{bmatrix} \quad \text{and} \quad \bar{\sigma} = \begin{bmatrix} -\bar{s}_O \\ \bar{s}_I \end{bmatrix}, \quad \bar{s}_O, \bar{s}_I \geq 0.$$

Proof. In more detailed form the system $0 = \Psi(z, \lambda, \sigma)$ with $\sigma := \begin{bmatrix} -s_O \\ s_I \end{bmatrix}$ reads

$$0 = \Psi(x, y, \lambda_O, \lambda_I, -s_O, s_I) = \begin{bmatrix} Q_{11}x + Q_{12}y + c_x + A_O^\top \lambda_O + A_I^\top \lambda_I - s_O \\ Q_{12}^\top x - Q_{22}y + c_y + B_I^\top \lambda_I + s_I \\ A_O x - b_O \\ A_I x + B_I y - b_I \\ x \circ (-s_O) \\ y \circ s_I \end{bmatrix}$$

and is thus, up to reordering and a flip of sign in the fifth equation, identical to the equations from the stationarity system (1.7) in Definition 1.2. Hence the “only if” part of the assertion is clear, and to see the “if” part, recall that the assumption implies that $\bar{y}$ is a KKT point of the inner problem and, thus, an element of $S(\bar{x})$. $\square$

To show that the stationary points of (P) are typically not only critical points, but even saddle points of f on Ω , we introduce an appropriate non-degeneracy concept for critical points of f on Ω .

DEFINITION 2.23. A point $\bar{z} \in \Omega$ is called a non-degenerate critical point of f on Ω if $\bar{z}$ is a critical point with some multipliers $\bar{\lambda}, \bar{\sigma}$ such that the Jacobian

$$D\Psi(\bar{z}, \bar{\lambda}, \bar{\sigma}) = \begin{bmatrix} Q & A^\top & \mathbf{1} \\ A & 0 & 0 \\ \text{diag}(\bar{\sigma}) & 0 & \text{diag}(\bar{z}) \end{bmatrix}$$

of Ψ from (2.11) is non-singular.

LEMMA 2.24. A point $\bar{x} \in X$ with $\bar{y} \in S(\bar{x})$ is a non-degenerate stationary point of (P) with multipliers $\bar{\lambda}_O, \bar{s}_O, \bar{\lambda}_I, \bar{s}_I$ if and only if $\bar{z} = (\bar{x}, \bar{y}) \in \Omega$ is a non-degenerate critical point of f on Ω with multipliers $\bar{\lambda} = \begin{bmatrix} \bar{\lambda}_O \\ \bar{\lambda}_I \end{bmatrix}$ and $\bar{\sigma} := \begin{bmatrix} -\bar{s}_O \\ \bar{s}_I \end{bmatrix}$, $\bar{s}_O, \bar{s}_I \geq 0$.

Proof. The equivalence of stationarity and criticality holds by Lemma 2.22. Hence, it remains to show that

$$D\Phi(\bar{x}, \bar{\lambda}_O, \bar{s}_O, \bar{y}, \bar{\lambda}_I, \bar{s}_I) = \begin{bmatrix} Q_{11} & A_O^\top & -\mathbf{1} & Q_{12} & A_I^\top & 0 \\ A_O & 0 & 0 & 0 & 0 & 0 \\ \text{diag}(\bar{s}_O) & 0 & \text{diag}(\bar{x}) & 0 & 0 & 0 \\ Q_{12}^\top & 0 & 0 & -Q_{22} & B_I^\top & \mathbf{1} \\ A_I & 0 & 0 & B_I & 0 & 0 \\ 0 & 0 & 0 & \text{diag}(\bar{s}_I) & 0 & \text{diag}(\bar{y}) \end{bmatrix}$$

is non-singular iff

$$(2.12) \quad D\Psi(\bar{z}, \bar{\lambda}, \bar{\sigma}) = D\Psi(\bar{x}, \bar{y}, \bar{\lambda}_O, \bar{\lambda}_I, -\bar{s}_O, \bar{s}_I) \\ = \begin{bmatrix} Q_{11} & Q_{12} & A_O^\top & A_I^\top & \mathbf{1} & 0 \\ Q_{12}^\top & -Q_{22} & 0 & B_I^\top & 0 & \mathbf{1} \\ A_O & 0 & 0 & 0 & 0 & 0 \\ A_I & B_I & 0 & 0 & 0 & 0 \\ -\text{diag}(\bar{s}_O) & 0 & 0 & 0 & \text{diag}(\bar{x}) & 0 \\ 0 & \text{diag}(\bar{s}_I) & 0 & 0 & 0 & \text{diag}(\bar{y}) \end{bmatrix}$$

is. This readily follows from the congruence $D\Psi(\bar{z}, \bar{\lambda}, \bar{\sigma}) = V^\top D\Phi(\bar{x}, \bar{\lambda}_O, \bar{s}_O, \bar{y}, \bar{\lambda}_I, \bar{s}_I) V$ with

$$V = \begin{bmatrix} \mathbf{1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\mathbf{1} & 0 \\ 0 & \mathbf{1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{1} \end{bmatrix}.$$

□

Lemma 2.24 and Theorem 2.19 entail the following characterization of non-degenerate criticality.

COROLLARY 2.25. For $\bar{x} \in X$ let $\bar{y} \in S(\bar{x})$ be inner non-degenerate. Then $\bar{z} = (\bar{x}, \bar{y})$ is a non-degenerate critical point of (P) if and only if $\bar{z}$ is a critical point and $\bar{x}$ is outer non-degenerate.

First- and second-order properties of a non-degenerate critical point $\bar{z}$ of f on Ω are readily encoded in its linear and quadratic indices, which we introduce next. First observe that, by Laplace expansion of its determinant, the Jacobian $D\Psi(\bar{z}, \bar{\lambda}, \bar{\sigma})$ is non-singular iff the absolute strict complementary slackness (ASCS) $\min_{i=1, \dots, n+m} (|\bar{\sigma}_i| + \bar{z}_i) > 0$ holds and

$$(2.13) \quad C := \begin{bmatrix} Q & A^\top & \mathbf{1}_{\bar{z}}^\top \\ A & 0 & 0 \\ \mathbf{1}_{\bar{z}} & 0 & 0 \end{bmatrix}$$

is non-singular. Note that a non-negativity constraint on $\bar{\sigma}$ is not needed for the latter argument, leading to the ASCS in place of the usual strict complementary slackness condition.

Full row rank of $\begin{bmatrix} A \\ \mathbf{1}_{\bar{z}} \end{bmatrix}$ means that the linear independence constraint qualification (LICQ)

holds at $\bar{z}$ in Ω . Furthermore, $T(\bar{z}, \Omega) := \ker \begin{bmatrix} A \\ \mathbf{1}_{\bar{z}} \end{bmatrix}$ is the tangent space to Ω at $\bar{z}$. Therefore, by

Lemma 2.13 the matrix C from (2.13) is non-singular if and only if the LICQ holds and $Q|_{T(\bar{z}, \Omega)}$ is non-singular. Altogether, this shows that $\bar{z} \in \Omega$ is a non-degenerate critical point of f on Ω iff multipliers $\bar{\lambda}, \bar{\sigma}$ exist which satisfy $\Psi(\bar{z}, \bar{\lambda}, \bar{\sigma}) = 0$, ASCS, LICQ and non-singularity of $Q|_{T(\bar{z}, \Omega)}$.

Following the terminology from [21], we call the number of negative entries of $\bar{\sigma}_{\mathcal{I}_{\bar{z}}}$ linear index (LI) and the number of negative eigenvalues of $Q|_{T(\bar{z}, \Omega)}$ quadratic index (QI) of $\bar{z}$. Standard first- and second-order optimality conditions imply that a non-degenerate critical point is a local minimal point iff LI and QI both vanish, a local maximal point iff LI and QI attain their maximal values $|\mathcal{I}_{\bar{z}}|$ and $(n + m) - (p + q + |\mathcal{I}_{\bar{z}}|)$, respectively, and otherwise a local saddle point.

For the statement of the following result recall that outer non-degeneracy of a stationary point $\bar{x}$ of (P) in the sense of Definition 2.18 entails that the $(n - (p + |\mathcal{I}_{\bar{x}}|), (n - (p + |\mathcal{I}_{\bar{x}}|))$ -matrix $Q'_{11}|_{T(\bar{x}, X)}$ is non-singular. Hence, assuming that it possesses ϱ negative eigenvalues is equivalent to its inertia being $(n - p - |\mathcal{I}_{\bar{x}}| - \varrho, \varrho, 0)$.

THEOREM 2.26. *For $\bar{x} \in X$ let $\bar{y} \in S(\bar{x})$ be inner non-degenerate, and let $\bar{x}$ be a stationary and outer non-degenerate point of (P), where $Q'_{11}|_{T(\bar{x}, X)}$ possesses ϱ negative eigenvalues. Then $\bar{z} = (\bar{x}, \bar{y})$ is a non-degenerate critical point of f on Ω with $LI = |\mathcal{I}_{\bar{y}}|$ and $QI = m - q - |\mathcal{I}_{\bar{y}}| + \varrho$.*

Proof. Let $\bar{\lambda}_O, \bar{s}_O, \bar{\lambda}_I, \bar{s}_I$ be multipliers from the stationarity condition of $\bar{x}$. By Theorem 2.19 $\bar{z} = (\bar{x}, \bar{y})$ is a non-degenerate stationary point of (P), and by Lemma 2.24 also a non-degenerate critical point with multipliers $\bar{\lambda} = \begin{bmatrix} \bar{\lambda}_O \\ \bar{\lambda}_I \end{bmatrix}$ and $\bar{\sigma} = \begin{bmatrix} -\bar{s}_O \\ \bar{s}_I \end{bmatrix}$. Furthermore, (OSCS) and (ISCS) imply that the first $|\mathcal{I}_{\bar{x}}|$ entries of $\bar{\sigma}_{\mathcal{I}_{\bar{x}}}$ are positive, whereas the remaining $|\mathcal{I}_{\bar{y}}|$ entries are negative. Therefore the linear index of $\bar{z}$ is $|\mathcal{I}_{\bar{y}}|$.

To determine the quadratic index QI of $\bar{z}$, we compute the inertia $\text{In}(Q|_{T(\bar{z}, \Omega)})$. Lemma 2.13 and Sylvester's law of inertia for the congruent matrices C from (2.13) and K from (2.9) imply

$$\text{In}(Q|_{T(\bar{z}, \Omega)}) = \text{In}(C) - (p + q + |\mathcal{I}_{\bar{z}}|, p + q + |\mathcal{I}_{\bar{z}}|, 0) = \text{In}(K) - (p + q + |\mathcal{I}_{\bar{z}}|, p + q + |\mathcal{I}_{\bar{z}}|, 0).$$

Furthermore, by the additivity formula for the inertia of Schur complements from [17] we have

$$\text{In}(K) = \text{In}(K/K_{22}) + \text{In}(K_{22}) = \text{In} \begin{bmatrix} Q'_{11} & A_O^\top & -\mathbf{1}_{\bar{x}}^\top \\ A_O & 0 & 0 \\ -\mathbf{1}_{\bar{x}} & 0 & 0 \end{bmatrix} + \text{In} \begin{bmatrix} -Q_{22} & B_I^\top & \mathbf{1}_{\bar{y}}^\top \\ B_I & 0 & 0 \\ \mathbf{1}_{\bar{y}} & 0 & 0 \end{bmatrix}$$

with Q'_{11} from (2.3). Another application of Sylvester's law of inertia and Lemma 2.13 yield

$$\text{In}(K) = \text{In}(Q'_{11}|_{T(\bar{x}, X)}) + (p + |\mathcal{I}_{\bar{x}}|, p + |\mathcal{I}_{\bar{x}}|, 0) + \text{In}(-Q_{22}|_{T(\bar{y}, Y(\bar{x}))}) + (q + |\mathcal{I}_{\bar{y}}|, q + |\mathcal{I}_{\bar{y}}|, 0).$$

The assumption on the number of negative eigenvalues of $Q'_{11}|_{T(\bar{x}, X)}$ and the non-degeneracy of $\bar{y} \in S(\bar{x})$, respectively, imply

$$\begin{aligned} \text{In}(Q'_{11}|_{T(\bar{x}, X)}) &= (n - p - |\mathcal{I}_{\bar{x}}| - \varrho, \varrho, 0), \\ \text{In}(-Q_{22}|_{T(\bar{y}, Y(\bar{x}))}) &= (0, m - q - |\mathcal{I}_{\bar{y}}|, 0) \end{aligned}$$

and thus

$$\text{In}(K) = (n - \varrho, p + |\mathcal{I}_{\bar{x}}| + \varrho, 0) + (q + |\mathcal{I}_{\bar{y}}|, m, 0) = (n + q + |\mathcal{I}_{\bar{y}}| - \varrho, m + p + |\mathcal{I}_{\bar{x}}| + \varrho, 0)$$

as well as

$$\begin{aligned} \text{In}(Q|_{T(\bar{z}, \Omega)}) &= (n + q + |\mathcal{I}_{\bar{y}}| - \varrho, m + p + |\mathcal{I}_{\bar{x}}| + \varrho, 0) - (p + q + |\mathcal{I}_{\bar{z}}|, p + q + |\mathcal{I}_{\bar{z}}|, 0) \\ &= (n - p - |\mathcal{I}_{\bar{x}}| - \varrho, m - q - |\mathcal{I}_{\bar{y}}| + \varrho, 0). \end{aligned}$$

Since QI is the second entry of the inertia triple, this shows the assertion. $\square$

COROLLARY 2.27. *Let the assumptions of Theorem 2.26 hold and let $Q'_{11}|_{T(\bar{x}, X)}$ be positive definite. Then $\bar{z}$ is a local saddle point of f on Ω .*

Proof. The positive definiteness of $Q'_{11}|_{T(\bar{x}, X)}$ is equivalent to $\varrho = 0$. Assume that the non-degenerate critical point $\bar{z}$ is a local minimal point of f on Ω . As mentioned before the statement of Theorem 2.26, this is characterized by the vanishing of both LI and QI. In view of Theorem 2.26 the latter is equivalent to $\mathcal{I}_{\bar{y}} = \emptyset$ and $0 = m - q - |\mathcal{I}_{\bar{y}}| + \varrho = m - q$. This, however, contradicts the blanket assumption $q < m$.

Analogously, assume that $\bar{z}$ is a local maximal point of f on Ω . This is characterized by LI and QI attaining their maximal values $|\mathcal{I}_{\bar{x}}| + |\mathcal{I}_{\bar{y}}|$ and $n - p - |\mathcal{I}_{\bar{x}}| + m - q - |\mathcal{I}_{\bar{y}}|$, respectively. By Theorem 2.26 this is equivalent to $\mathcal{I}_{\bar{x}} = \emptyset$ and $0 = n - p - |\mathcal{I}_{\bar{x}}| - \varrho = n - p$. This contradicts the blanket assumption $p < n$. Therefore the critical point $\bar{z}$ is neither a local minimal nor a local maximal point of f on Ω , hence a local saddle point. $\square$

The assumptions of Corollary 2.27 mean that $\bar{x}$ is a non-degenerate local minimal point of (P) , that is, the type of outer non-degenerate point $\bar{x}$ which we wish to find algorithmically.

EXAMPLE 2.28. In the situation of Example 1.3 we have seen in Example 2.20 that the assumptions of Corollary 2.27 are satisfied. In particular, (OSOSC) holds trivially. Hence, $(\bar{x}, \bar{y})$ is a local saddle point of f on Ω .

2.5 Summary of the relationships. In this section, we pictorially summarize the relationships intervening between the stationarity concept presented in Definition 1.2 and the notions of local optimality and saddle point studied throughout this section; cf. Fig. 1.

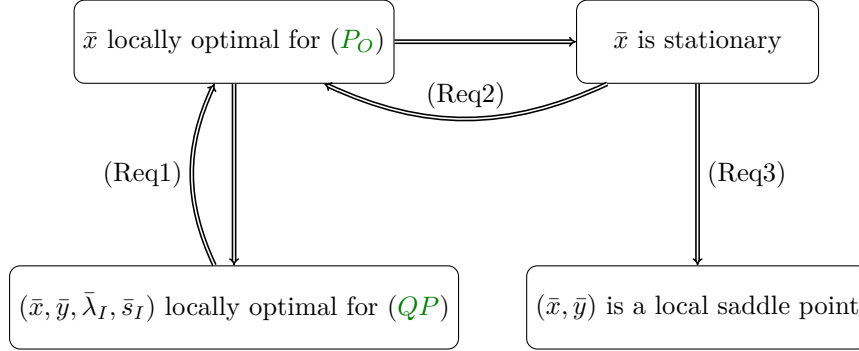


Fig. 1: Summary of relationships. The local optimality in the right hand side of the implication $\bar{x}$ is stationary $\implies \bar{x}$ locally optimal for (P_O) is, in fact, strict.

In Fig. 1, the first requirement (Req1) corresponds to the inner semicontinuity of S_D at the point $(\bar{x}, \bar{y}, \bar{\lambda}_I, \bar{s}_I)$ from Proposition 2.3(b), while the second requirement (Req2) represents all the assumptions made in Theorem 2.7. As for the third requirement (Req3), it collects all the assumptions in Corollary 2.27.

3 An interior point method. In this section, we introduce an interior point method to compute a solution of the stationarity condition (1.7). To this aim, let us consider, given a parameter $\mu > 0$, the barrier minmax problem

$$(P_\mu) \quad \min_{x \in \bar{X}} \max_{y \in \bar{Y}(x)} f(x, y) - \mu \sum_{i=1}^n \log(x_i) + \mu \sum_{j=1}^m \log(y_j),$$

where f is defined in (1.1) and $\bar{X}$ and $\bar{Y}(x)$ are given by

$$\bar{X} := \{x \in \mathbb{R}^n : A_O x = b_O\} \quad \text{and} \quad \bar{Y}(x) := \{y \in \mathbb{R}^m : A_I x + B_I y = b_I\},$$

respectively. Obviously, problem (P_μ) can be written as the minimization of the function $\varphi_\mu(x)$ over the outer feasibility set $\bar{X}$ with $\varphi_\mu(x) := \tilde{\varphi}_\mu(x) - \hat{\varphi}_\mu(x)$, where

$$\begin{cases} \tilde{\varphi}_\mu(x) &:= \frac{1}{2} x^\top (Q_{11}) x + c_x^\top x - \mu \sum_{i=1}^n \log(x_i), \\ \hat{\varphi}_\mu(x) &:= \min_{y \in \bar{Y}(x)} \frac{1}{2} y^\top (Q_{22}) y - x^\top Q_{12} y - c_y^\top y - \mu \sum_{j=1}^m \log(y_j). \end{cases}$$

Now, let $\bar{x}$ with $\bar{x}_i > 0$, $i = 1, \dots, n$, be a local optimal solution of (P_μ) , in a sense similar to (1.6); i.e., for some neighborhood U of $\bar{x}$,

$$\bar{x} \in \operatorname{argmin}_{x \in \bar{X} \cap U} \varphi_\mu(x).$$

If we assume that the optimal solution set-valued mapping

$$S_\mu(x) := \operatorname{argmin}_{y \in \bar{Y}(x)} \frac{1}{2} y^\top (Q_{22}) y - x^\top Q_{12} y - c_y^\top y - \mu \sum_{j=1}^m \log(y_j)$$

is inner semicontinuous at $(\bar{x}, \bar{y})$ with $\bar{y} \in S_\mu(\bar{x})$ and the matrix $B_I^\top$ possesses full column rank, then the function $\hat{\varphi}_\mu$ is Lipschitz continuous around $\bar{x}$ and its Clarke subdifferential at this point

can be estimated as

$$\partial\hat{\varphi}_\mu(\bar{x}) \subset \left\{ -Q_{12}\bar{y} + A_I^\top \lambda_I : Q_{22}\bar{y} - Q_{12}\bar{y} + B_I^\top \lambda_I - c_y - \begin{bmatrix} \frac{\mu}{\bar{y}_1} \\ \vdots \\ \frac{\mu}{\bar{y}_m} \end{bmatrix} = 0, \lambda_I \in \mathbb{R}^q \right\};$$

see, e.g., [30]. Therefore, as $\bar{x}$ is locally optimal for problem (P_μ) ,

$$(3.1) \quad 0 \in \partial\varphi_\mu(\bar{x}) + N_{\bar{X}}(\bar{x}),$$

where $N_{\bar{X}}$ stands for the normal cone to $\bar{X}$ in the sense of convex analysis. Then observing that we have the formulas

$$\nabla\tilde{\varphi}_\mu(\bar{x}) = Q_{11}\bar{x} + c_x - \begin{bmatrix} \frac{\mu}{\bar{x}_1} \\ \vdots \\ \frac{\mu}{\bar{x}_n} \end{bmatrix} \quad \text{and} \quad N_{\bar{X}}(\bar{x}) = A_O^\top \mathbb{R}^p,$$

it follows from (3.1) and the upper estimate of $\partial\hat{\varphi}_\mu(\bar{x})$ given above that there exist Lagrange multipliers $\lambda_O \in \mathbb{R}^p$ and $\lambda_I \in \mathbb{R}^q$ such that we have the stationary conditions

$$(3.2) \quad \begin{aligned} Qz + c + A^\top \lambda - \begin{bmatrix} \mathbb{1} & 0 \\ 0 & -\mathbb{1} \end{bmatrix} s &= 0, \\ Az &= b, \\ z \circ s &= \mu e, \\ z, s &> 0, \end{aligned}$$

while considering the notation in (2.10) and setting $(s_O)_i := \frac{\mu}{\bar{x}_i}$ for $i = 1, \dots, n$ and $(s_I)_j := \frac{\mu}{\bar{y}_j}$ for $j = 1, \dots, m$. A direct comparison of equations (1.7) and (3.2) reveals that the former differs from the latter solely in the presence of the parameter μ in the complementarity conditions and the positivity of z and s .

Interpreting the system (3.2) as a perturbation of (1.7) constitutes the classical foundation for devising interior point methods (see, e.g., [14]), and indeed, intuitively, the stationarity system (3.2) approximates (1.7) more and more closely when μ converges to 0. Motivated by the barrier interpretation presented above, we analyze in the subsequent sections how classical IPM solvers can be adapted to solve (1.7) and establish their theoretical guarantees.

Note that (1.7) represents a particular instance of a mixed linear complementarity problem and, more generally, of a constrained equation problem, both of which are amenable to specialized IPM-type solvers (see, e.g., [7, 37]). However, to the best of our knowledge, adopting such a perspective does not yield substantially improved convergence or complexity and guarantees. Hence, we employ here a standard infeasible IPM framework; see, e.g., [40, Chapter 6].

Before presenting the details of the IPM method, let us present a series of notational changes or additions to enable our presentation to be closer to standard IPM notation. In particular, in the following, we will set for any $z \in \mathbb{R}^{n+m}$, $Z := \text{diag}(z)$ and hence, $z \circ s = ZSe$, whereas any inequalities involving a vector and a scalar are meant to hold component-wise. Moreover, aiming at devising IPM-type solvers for the system (1.7), we introduce the map

$$(3.3) \quad F(z, \lambda, s) := \begin{bmatrix} Qz + c + A^\top \lambda - \begin{bmatrix} \mathbb{1} & 0 \\ 0 & -\mathbb{1} \end{bmatrix} s \\ Az - b \\ ZSe \end{bmatrix} \quad \text{with } z, s \geq 0,$$

as a version of the mapping $\Psi(z, \lambda, (-s_O, s_I))$ from (2.11) in standard IPM notation.

Multiplying the first block row of the function F by $\begin{bmatrix} \mathbb{1} & 0 \\ 0 & -\mathbb{1} \end{bmatrix}$, we get the function

$$\hat{F}(z, \lambda, s) := \begin{bmatrix} \begin{bmatrix} \mathbb{1} & 0 \\ 0 & -\mathbb{1} \end{bmatrix} (Qz + c + A^\top \lambda) - s \\ Az - b \\ ZSe \end{bmatrix} \quad \text{with } z, s \geq 0,$$

which can be employed to equivalently write the stationarity system (1.7). We can hence succinctly re-write the stationarity system (1.7) as

$$(3.4) \quad \hat{F}(z, \lambda, s) := \begin{bmatrix} \hat{Q}z + \hat{c} + \hat{A}^\top \lambda - s \\ Az - b \\ ZSe \end{bmatrix} = 0, \text{ with } z, s \geq 0,$$

where the expression $\hat{F}(z, \lambda, s)$ is written in the compact form using the notation

$$\begin{aligned} \hat{Q} &:= \begin{bmatrix} \mathbf{1} & 0 \\ 0 & -\mathbf{1} \end{bmatrix} Q = \begin{bmatrix} Q_{11} & Q_{12} \\ -Q_{12}^\top & Q_{22} \end{bmatrix}, \\ \hat{c} &:= \begin{bmatrix} \mathbf{1} & 0 \\ 0 & -\mathbf{1} \end{bmatrix} c = \begin{bmatrix} c_x \\ -c_y \end{bmatrix}, \\ \hat{A}^\top &:= \begin{bmatrix} \mathbf{1} & 0 \\ 0 & -\mathbf{1} \end{bmatrix} A^\top = \begin{bmatrix} A_O^\top & A_I^\top \\ 0 & -B_I^\top \end{bmatrix}. \end{aligned}$$

From now on, we will focus only on $\hat{F}$ and hence, on the computation of stationary points. Additionally, in the sequel, we will use the notation

$$B(v_0, r) := \{v \mid \|v - v_0\| < r\} \quad \text{and} \quad \overline{B(v_0, r)} := \{v \mid \|v - v_0\| \leq r\},$$

while employing, when beneficial for readability, the shorthand $v := (z, \lambda, s)$.

3.1 On the local existence of the central path. Of utmost importance for the development of any kind of IPM is the concept of *central path*. To this aim, for any $\mu > 0$, we introduce the perturbation of the function $\hat{F}$ in (3.4), i.e.,

$$(3.5) \quad H(z, \lambda, s, \mu) := \begin{bmatrix} \hat{Q}z + \hat{c} + \hat{A}^\top \lambda - s \\ Az - b \\ ZSe - \mu e \end{bmatrix} = \hat{F}(z, \lambda, s) - \mu \begin{bmatrix} 0 \\ 0 \\ e \end{bmatrix},$$

with $H : \mathbb{R}^{N+1} \rightarrow \mathbb{R}^N$ with $H \in C^1(\mathbb{R}^{N+1}, \mathbb{R}^N)$, where $N := 2(n+m) + p + q$. Subsequently, we define the *central path* by

$$\mathcal{C} := \{(z_\mu, \lambda_\mu, s_\mu) \in \mathbb{R}_{>0}^{n+m} \times \mathbb{R}^{p+q} \times \mathbb{R}_{>0}^{n+m} \mid H(z, \lambda, s, \mu) = 0 \text{ for } \mu > 0\}.$$

The introduced notation will allow us to establish the local existence of the central path and other properties of the Newton matrices by means of the Implicit Function Theorem. In this regard, and throughout this section, we will make the crucial assumption:

ASSUMPTION 3.1. *The point $v^* = (z^*, \lambda^*, s^*)$ satisfies $z^*, s^* \geq 0$, $H(v^*, 0) = 0$, and $\partial_v H(v^*, 0)$ is invertible.*

The next lemma ensures that Theorem 2.19 provides a sufficient condition for Assumption 3.1 in terms of the mild inner and outer non-degeneracy requirements for (P) from Section 2.

LEMMA 3.2. *Assumption 3.1 holds at $v^* = (z^*, \lambda^*, s^*)$ with $z^* := (x^*, y^*)$, $\lambda^* := (\lambda_O^*, \lambda_I^*)$, and $s^* := (s_O^*, s_I^*)$ if and only if x^* is a non-degenerate stationary point of (P) in the sense of Definition 2.12 with $y^* \in S(x^*)$ and multipliers $(\lambda_O^*, s_O^*, \lambda_I^*, s_I^*)$.*

Proof. By (3.4) stationarity of x^* with corresponding $y^* \in S(x^*)$ and $(\lambda_O^*, s_O^*, \lambda_I^*, s_I^*)$ is equivalent to $H(v^*, 0) = 0$, $z^*, s^* \geq 0$. Moreover, the matrix

$$\partial_v H(v^*, 0) = \begin{bmatrix} \hat{Q} & \hat{A}^\top & -\mathbf{1} \\ A & 0 & 0 \\ S^* & 0 & Z^* \end{bmatrix} = \begin{bmatrix} Q_{11} & Q_{12} & A_O^\top & A_I^\top & -\mathbf{1} & 0 \\ -Q_{12}^\top & Q_{22} & 0 & -B_I^\top & 0 & -\mathbf{1} \\ A_O & 0 & 0 & 0 & 0 & 0 \\ A_I & B_I & 0 & 0 & 0 & 0 \\ \text{diag}(s_O^*) & 0 & 0 & 0 & \text{diag}(x^*) & 0 \\ 0 & \text{diag}(s_I^*) & 0 & 0 & 0 & \text{diag}(y^*) \end{bmatrix}$$

may be written as

$$\begin{bmatrix} \mathbf{1} & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{1} & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{1} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{1} & 0 & 0 \\ 0 & 0 & 0 & 0 & -\mathbf{1} & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{1} \end{bmatrix} D\Psi(x^*, y^*, \lambda_O^*, \lambda_I^*, -s_O^*, s_I^*) \begin{bmatrix} \mathbf{1} & 0 & 0 & 0 & 0 & 0 \\ 0 & -\mathbf{1} & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{1} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{1} & 0 & 0 \\ 0 & 0 & 0 & 0 & -\mathbf{1} & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{1} \end{bmatrix}$$

with $D\Psi$ from (2.12). Therefore the non-singularity of $\partial_v H(v^*, 0)$ means that $z^* \geq 0$ is a non-degenerate critical point in the sense of Definition 2.23 with $s^* \geq 0$, and Lemma 2.24 shows the assertion. $\square$

Given the form of $\hat{F}$ and Assumption 3.1, we can state the following lemma, propaedeutic for the specific version of the Implicit Function Theorem that we will consider (see Theorem 3.4), but also for developments in Subsection 3.4.

LEMMA 3.3. *There exist $\delta_1, L, K > 0$ such that for all $v, v' \in B(v^*, \delta_1)$,*

$$\|D\hat{F}(v) - D\hat{F}(v')\| \leq L\|v - v'\| \quad \text{and} \quad \|D\hat{F}(v)^{-1}\| \leq K.$$

The specific version of the Implicit Function Theorem here considered, see Theorem 3.4, will be instrumental in proving the local existence of the central path. For the sake of completeness, we report the proof of such a theorem in Section B.

THEOREM 3.4. *Consider*

$$V_\delta := \{(v, \mu) \mid \|v - v^*\| \leq \delta, |\mu| \leq \delta\}$$

and select $\delta > 0$ such that

$$\sup_{(v, \mu) \in V_\delta} \|I - \partial_v H(v^*, 0)^{-1} \partial_v H(v, \mu)\| \leq \frac{1}{2}$$

and $\partial_v H(v, \mu)^{-1}$ exists for all $(v, \mu) \in V_\delta$. Furthermore, let

$$B_\delta := \max \left\{ 1, \sup_{(v, \mu) \in V_\delta} \|\partial_\mu H(v, \mu)\| \right\} \quad \text{and} \quad M := \max \{1, \|\partial_v H(v^*, 0)^{-1}\|\}.$$

Define $\bar{\mu} := (2MB_\delta)^{-1}\delta$ and $\Lambda_{\bar{\mu}} := \{\mu \in \mathbb{R} \mid |\mu| < \bar{\mu}\}$. Then there exists a function $g \in C^1(\Lambda_{\bar{\mu}}, \mathbb{R}^N)$ such that all the solutions of the equation $H(v, \mu) = 0$ in the set $\overline{B(v^*, \delta)} \times \Lambda_{\bar{\mu}}$ are given by $(g(\mu), \mu)$. In addition, there exists $C > 0$ such that

$$\|g(\mu_1) - g(\mu_2)\| \leq C|\mu_1 - \mu_2| \quad \text{for all } \mu_1, \mu_2 \in \Lambda_{\bar{\mu}}$$

and

$$\frac{d}{d\mu} g(\mu) = -(\partial_v H(g(\mu), \mu))^{-1} \partial_\mu H(g(\mu), \mu).$$

This remark presents the details about the local existence of the central path.

Remark 3.5. Using the differentiability of g at $\mu = 0$, we have that

$$(3.6) \quad g(\mu) = \underbrace{v^* - (\partial_v H(v^*, 0))^{-1} \begin{bmatrix} 0 \\ 0 \\ -e \end{bmatrix}}_{\frac{d}{d\mu} g(0) =: v'(0)} \mu + o(\mu).$$

That is,

$$(\partial_v H(v^*, 0))v'(0) = \begin{bmatrix} 0 \\ 0 \\ e \end{bmatrix} \quad \text{with} \quad \partial_v H(v^*, 0) = \begin{bmatrix} \hat{Q} & \hat{A}^\top & -\mathbf{1} \\ A & 0 & 0 \\ S^* & 0 & Z^* \end{bmatrix}.$$

Hence, looking at the last set of block equations $[s^*]_i [z'(0)]_i + [z^*]_i [s'(0)]_i = 1$ for $i = 1, \dots, m+n$ and recalling that $Z^* S^* e = 0$, we have that it must hold

$$(3.7) \quad a := \min_i \{[s^*]_i + [z^*]_i\} > 0;$$

i.e., s_i^* and z_i^* can not be zero at the same time. Hence, that $s_i^* = 0$ and $z_i^* = 0$ imply $s'(0)_i > 0$ and $z'(0)_i > 0$, respectively. On the other hand, when $s_i^* > 0$ or $z_i^* > 0$ we do not have information about the signs of $s'(0)_i$ or $z'(0)_i$. This does not represent a problem, because we can still conclude

that for μ small enough, $z(\mu), s(\mu) > 0$. It is important to note that not only the quantity a defined (3.7) is a measure of the radius where $D\hat{F}(v)$ is invertible, it must be indeed $a < \delta_1$, see [1, Section 4], where δ_1 is from Lemma 3.3, but also, as it appears from (3.6), a measure of *how far* one can extend, in general, the central path from the solution v^* .

Similarly to [1], with reference to the quantities defined in Lemma 3.3 and Theorem 3.4, let

$$(3.8) \quad \delta^* := \min \left\{ \delta, \delta_1, \frac{1}{2KL} \right\} \quad \text{and} \quad \mu^* := \min \left\{ \bar{\mu}, \frac{\delta^*}{C} \left(1 + \sqrt{1 + 2KL\delta^*} \right)^{-1} \right\}.$$

3.2 An infeasible interior point method. In Algorithm 3.1, we present the main computational scheme considered in this work. To this aim, we need the following notation:

$$(3.9) \quad r_b := Az - b,$$

$$(3.10) \quad r_c := \hat{Q}z + \hat{A}^\top \lambda + \hat{c} - s,$$

and

$$\begin{aligned} (z(\alpha), \lambda(\alpha), s(\alpha)) &:= (z, \lambda, s) + \alpha(\Delta z, \Delta \lambda, \Delta s), \\ \mu(\alpha) &:= \frac{1}{n+m} z(\alpha)^\top s(\alpha). \end{aligned}$$

When r_b and r_c are computed at the point $(z, \lambda, s) = (z^k, \lambda^k, s^k)$, we denote them by r_b^k and r_c^k . The algorithm here considered, see [40, Chapter 6], works with the central path neighborhood $\mathcal{N}_{-\infty}(\underline{\gamma}, \beta)$ defined by

$$(3.11) \quad \begin{aligned} \mathcal{N}_{-\infty}(\underline{\gamma}, \beta) &= \left\{ (z, \lambda, s) \in \mathbb{R}_{>0}^{n+m} \times \mathbb{R}^{p+q} \times \mathbb{R}_{>0}^{n+m} \right. \\ &\quad \left. | \| (r_b, r_c) \| \leq \left[\frac{\| (r_b^0, r_c^0) \|}{\mu_0} \right] \beta \mu, \quad z_i s_i \geq \underline{\gamma} \mu, \quad i = 1, 2, \dots, n \right\}, \end{aligned}$$

where $\underline{\gamma} \in (0, 1)$ and $\beta > 1$ are given parameters and (r_b^0, r_c^0) and μ_0 are evaluated at the starting point (z^0, λ^0, s^0) . Note that we must have

$$\|r_b^0\| \leq \beta \mu_0 \quad \text{and} \quad \|r_c^0\| \leq \beta \mu_0$$

to ensure that the initial point (z^0, λ^0, s^0) belongs to $\mathcal{N}_{-\infty}(\underline{\gamma}, \beta)$.

Algorithm 3.1 Interior Point Algorithm (IPM)

Require: $\gamma \in (0, 1)$, $\beta \geq 1$, $\sigma_{\min} < \sigma_{\max} < 1$.

1: Choose (z^0, λ^0, s^0) with $(z^0, s^0) > 0$.

2: **for** $k = 0, 1, 2, \dots$ **do**

3: Choose $\sigma_k \in [\sigma_{\min}, \sigma_{\max}]$ and solve

$$(3.12) \quad \begin{bmatrix} \hat{Q} & \hat{A}^\top & -\mathbf{1} \\ A & 0 & 0 \\ S^k & 0 & Z^k \end{bmatrix} \begin{bmatrix} \Delta z \\ \Delta \lambda \\ \Delta s \end{bmatrix} = \begin{bmatrix} -r_b^k \\ -r_c^k \\ -Z^k S^k e + \sigma_k \mu_k e \end{bmatrix}.$$

4: Choose α_k as the largest value of $\alpha \in [0, 1]$ such that

$$(3.13) \quad (z^k(\alpha), \lambda^k(\alpha), s^k(\alpha)) \in \mathcal{N}_{-\infty}(\underline{\gamma}, \beta)$$

and the Armijo condition holds:

$$(3.14) \quad \mu(\alpha) \leq (1 - (1 - \sigma_{\max})\alpha)\mu_k.$$

5: Set $(z^{k+1}, \lambda^{k+1}, s^{k+1}) = (z^k(\alpha_k), \lambda^k(\alpha_k), s^k(\alpha_k))$.

6: **end for**

LEMMA 3.6. *It holds that:*

$$\begin{aligned} Az(\alpha) - b &= (1 - \alpha)r_b, \\ \hat{Q}z(\alpha) + \hat{A}^\top \lambda(\alpha) - s(\alpha) + \hat{c} &= (1 - \alpha)r_c, \\ (z + \alpha\Delta z)^\top (s + \alpha\Delta s) &= z^\top s(1 + \alpha(\sigma - 1)) + \alpha^2(\Delta z)^\top \Delta s, \\ s_i \Delta z_i + z_i \Delta s_i &= \sigma \frac{z_i^\top s_i}{n+m} - z_i s_i. \end{aligned}$$

Proof. We prove the statement only for r_c , the proof for r_b is analogous. We have,

$$\begin{aligned} \hat{Q}(z + \alpha\Delta z) + \hat{A}^\top (\lambda + \alpha\Delta \lambda) - (s + \alpha\Delta s) + \hat{c} \\ = \hat{Q}z + \hat{A}^\top \lambda - s + \hat{c} + \alpha(\hat{Q}\Delta z + \hat{A}^\top \Delta - \Delta s) = (1 - \alpha)r_c. \end{aligned}$$

The last part of the statement follows by direct inspection of the last block equation in (3.12). In particular, $s^\top \Delta z + z^\top \Delta s = -z^\top s + \sigma(n + m)\mu = (\sigma - 1)z^\top s$. $\square$

Remark 3.7 (Convergence and computational complexity per iteration). For all points in the neighborhood $\mathcal{N}_\infty(\underline{\gamma}, \beta)$, the infeasibility is bounded by some multiple of the duality measure μ . By forcing μ to zero monotonically – see Algorithm 3.1 Line 3.14 – and restricting all iterates (z^k, λ^k, s^k) to the neighborhood $\mathcal{N}_\infty(\underline{\gamma}, \beta)$, the interior point algorithm presented in Algorithm 3.1 ensures primal-dual convergence, i.e., that $r_b^k \rightarrow 0$ and $r_c^k \rightarrow 0$ as $k \rightarrow \infty$. The other condition in (3.11) (i.e., $z_i s_i \geq \underline{\gamma}\mu$) keeps the pairwise products $z_i s_i$ roughly in balance and prevents the Newton-like search directions from being distorted by components of (z, s) that approach zero prematurely.

From a computational point of view, the dominant cost of each iteration is the solution of the primal-dual Newton system (3.12). For a direct linear algebra implementation, this cost is governed by the factorization of the corresponding saddle-point/KKT matrix. In the worst case, if the system is treated as a dense linear system of dimension equal to the total number of primal-dual variables, the cost is cubic in that dimension. In practice, however, interior point solvers exploit the symmetry, sparsity, and block structure of the Newton system, often reducing the computational burden substantially. Thus, the per-iteration cost should be understood as the cost of forming and solving the structured Newton system, rather than as the cost of the line-search procedure. And indeed, the Armijo condition in (3.14) should not be interpreted as introducing an additional optimization loop comparable to the Newton step itself. Its role is only to select an admissible stepsize along the already computed Newton direction. This requires inexpensive scalar operations, such as checking positivity, membership in the neighborhood, and the decrease of the duality measure. These operations are negligible compared with the solution of (3.12). Moreover, in many practical implementations of primal-dual interior point methods, including the implementation considered here, see Section 4, the Armijo backtracking procedure is not explicitly used. Instead, the stepsize is chosen by a fraction-to-the-boundary rule, which takes a fixed fraction of the largest step preserving positivity of the primal-dual variables. This choice retains the main purpose of the line-search safeguard while avoiding an inner backtracking loop.

3.3 Convergence analysis: the case $A_I = 0$. In this section, we conduct the convergence and complexity analysis of Algorithm 3.1 for the particular case $A_I = 0$ (cf. Remark 2.4). While this case does not fully capture the generality of the original minmax problem (P), it remains an important version of such problem for several reasons. Firstly, it provides valuable insights as a motivational example. Secondly, it establishes that stationary points of *minmax problems with decoupled constraints* – that is, problems of the form (P) where $Y(x) \equiv Y$ – can be computed in polynomial time. This result is further corroborated by the equivalent formulation (QP) of problem (P), which reduces to a convex programming problem when $A_I = 0$, and was also preliminarily observed in [36, Section 2] in the strongly convex-concave case. The analysis carried out in this section will generalize such an observation without requiring strong convexity-concavity. It is important to note, moreover, that such an extension is particularly relevant as it postulates the existence of polynomial algorithms for the resolution of problems of a rich class of problems

$$\min_{x \in X} \max_{0 \leq y \leq p} y^\top x$$

(with X given in (1.2)), which can be used, for example, to model applications where the objective function of a linear program is subject to uncertainty. Finally, we observe that the polynomial

solvability of this kind of problems is well understood by the robust optimization community, see [2], but usually overlooked by the min-max one.

LEMMA 3.8. *If $A_I = 0$, then $\bar{z}^\top \bar{s} \geq 0$ for any point $(\bar{z}, \bar{\lambda}, \bar{s})$ such that*

$$A\bar{z} = 0 \quad \text{and} \quad \hat{Q}\bar{z} + \hat{A}^\top \bar{\lambda} - \bar{s} = 0.$$

Proof. We have $\bar{z}^\top \bar{s} = \bar{z}^\top \hat{Q}\bar{z} + \bar{z}^\top \hat{A}^\top \bar{\lambda}$. Since $A_I = 0$, equality $A\bar{z} = 0$ implies $\hat{A}\bar{z} = 0$; hence, $\bar{z}^\top \hat{A}^\top \bar{\lambda} = 0$. The result follows by observing that

$$\bar{z}^\top \hat{Q}\bar{z} = \frac{1}{2} \bar{z}^\top (\hat{Q} + \hat{Q}^\top) \bar{z} \geq 0. \quad \square$$

It is important to note that Lemma 3.8 also guarantees that for every $X, S > 0$ the IPM matrices

$$\begin{bmatrix} \hat{Q} & \hat{A}^\top & -\mathbf{1} \\ A & 0 & 0 \\ S & 0 & Z \end{bmatrix}$$

are invertible. The proof of such a statement can be easily obtained by observing that Lemma 3.8 implies that the matrix $\begin{bmatrix} \hat{Q} & \hat{A}^\top & -\mathbf{1} \\ A & 0 & 0 \end{bmatrix}$ has the *mixed- P_0 property* by an application of [29, Lemma 2]. Moreover, using Assumption 3.1 for (z^*, λ^*, s^*) , Lemma 3.8, the *mixed- P_0 property* in the interior, it can easily be shown that the central path, i.e., the solutions of (3.5), exists for every $\mu \in [0, +\infty)$.

To end this section, in the following Theorem 3.9, we summarize convergence and complexity guarantees for Algorithm 3.1 in the case $A_I = 0$.

THEOREM 3.9. *Suppose that $\varepsilon > 0$ is given and let $(z^0, \lambda^0, s^0) = (\zeta e, 0, \zeta e)$ be such that ζ satisfies*

$$\|(z^*, s^*)\| \leq \zeta \quad \text{and} \quad \zeta^2 \leq \frac{C}{\varepsilon^\kappa}$$

for suitable positive constants C, κ . There exists $K \in \mathbb{N}$ with

$$K = O((n+m)^2 |\log(\varepsilon)|)$$

such that the iterates $\{(z^k, \lambda^k, s^k)\}_{k \in \mathbb{N}}$, generated by Algorithm 3.1, for solving the system (3.4) when $A_I = 0$, satisfy

$$\mu^k \leq \varepsilon \quad \text{for all } k \geq K.$$

An outline of the proof is reported in Section C.

Remark 3.10. It is important to note that by definition of $\mathcal{N}_{-\infty}(\gamma, \beta)$, see (3.11), if $\mu^k \leq \varepsilon$, then $\|(r_c, r_b)\| \in O(\varepsilon)$, i.e., (z, λ, s) is an ε -approximate solution of the stationary conditions (3.3).

3.4 Convergence analysis: the case $A_I \neq 0$. In this section, we analyze the local convergence of Algorithm (3.1) when $A_I \neq 0$. We emphasize that, as outlined in Remark 2.4, problem (P) is nonconvex in this case. Hence, polynomial complexity can not be expected in this more general case. To prove such a local convergence result, we will use a slightly different neighborhood of the infeasible central path w.r.t. the one used in the previous section, i.e., we consider

$$(3.15) \quad \begin{aligned} \mathcal{N}(\bar{\gamma}, \underline{\gamma}, \gamma_p, \gamma_d) &:= \{(z, \lambda, s) \in \mathbb{R}_{>0}^{n+m} \times \mathbb{R}^{p+q} \times \mathbb{R}_{>0}^{n+m} \mid \\ &\bar{\gamma}\mu \geq z_i s_i \geq \underline{\gamma}\mu \quad \text{for } i = 1, \dots, n+m, \quad z^\top s \geq \gamma_p \|r_b\|, \quad z^\top s \geq \gamma_d \|r_c\|\}, \end{aligned}$$

where $\bar{\gamma} > 1 > \underline{\gamma} > 0$ and $(\gamma_p, \gamma_d) > 0$. Therefore, condition (3.13) in Algorithm 3.1 has to be replaced with

$$(3.16) \quad (z^k(\alpha), \lambda^k(\alpha), s^k(\alpha)) \in \mathcal{N}(\bar{\gamma}, \underline{\gamma}, \gamma_p, \gamma_d).$$

Before continuing, we would like to remind the reader that the constants δ^* and μ^* used in the remainder of this section are defined in (3.8) and ensure, thanks to Theorem 3.4, that the Newton

system remains in a suitable neighborhood of a solution of 3.4. The main idea of the proof is based on the observation that when an initial point is chosen from

$$N^* := B(v^*, \delta^*) \cap \mathcal{N}(\bar{\gamma}, \underline{\gamma}, \gamma_p, \gamma_d),$$

all the other iterates of Algorithm 3.1 remain in this set. The convergence will essentially follow by observing that in N^* , the IPM matrices and the right-hand side in (3.12) are uniformly bounded, see proof of Corollary 3.14, and by the definition $\mathcal{N}(\bar{\gamma}, \underline{\gamma}, \gamma_p, \gamma_d)$, see proof Theorem 3.15. All the missing proofs of the results presented in this section can be found in Section D.

To establish the local convergence of Algorithm 3.1, we need the following preliminary results.

LEMMA 3.11 (see Lemma 2 in [1]). *For all $z \in B(v^*, \delta^*)$ and for all $\mu^+ \in [0, \mu^*)$, the full-Newton iterate defined by*

$$(3.17) \quad v^+ = v - D\hat{F}(v)^{-1}(\hat{F}(v) - \mu^+ \tilde{e})$$

is well defined and satisfies

$$\|v^+ - g(\mu^+)\| \leq \frac{KL}{2} \|v - g(\mu^+)\|^2$$

and

$$(3.18) \quad \|v^+ - v^*\| \leq \frac{1}{2} \|v - v^*\| + \frac{\mu^+ \delta^*}{\mu^* 2}.$$

In particular, $v^+ \in B(v^, \delta^*)$.*

LEMMA 3.12 (Damped Newton step). *For all $v \in B(v^*, \delta^*)$ and all $\mu^+ \in [0, \mu^*)$, let v^+ be the Newton iterate given by (3.17). For $\alpha \in [0, 1]$, define the damped iterate*

$$v_\alpha = v + \alpha(v^+ - v).$$

Then the following bounds hold:

$$\|v_\alpha - g(\mu^+)\| \leq (1 - \alpha) \|v - g(\mu^+)\| + \alpha \frac{KL}{2} \|v - g(\mu^+)\|^2$$

and

$$(3.19) \quad \|v_\alpha - v^*\| \leq \left(1 - \frac{\alpha}{2}\right) \|v - v^*\| + \alpha \frac{\mu^+ \delta^*}{\mu^* 2}.$$

In particular, $v_\alpha \in B(v^, \delta^*)$.*

THEOREM 3.13. *Suppose that $(z^k, \lambda^k, s^k) \in \mathcal{N}(\bar{\gamma}, \underline{\gamma}, \gamma_p, \gamma_d)$ is a point satisfying $(z^k)^\top s^k > 0$. Moreover, define*

$$\begin{aligned} \alpha_p^{*,k} &:= \sup \{ \alpha \in [0, 1] \mid z^k(\alpha) \geq 0 \}, \\ \alpha_d^{*,k} &:= \sup \{ \alpha \in [0, 1] \mid s^k(\alpha) \geq 0 \}, \\ \alpha^{*,k} &:= \min \{ \alpha_p^{*,k}, \alpha_d^{*,k} \}. \end{aligned}$$

If $(z(\alpha^{,k}), \lambda(\alpha^{*,k}), s(\alpha^{*,k}))$ is not a solution of problem (3.4), then there exists a scalar $\hat{\alpha}^k$ with $0 < \hat{\alpha}^k < \alpha^{*,k}$ such that (3.16) and (3.14) hold for all $\alpha \in [0, \hat{\alpha}^k]$.*

Proof. In this proof, we omit the IPM iterate counter k ; i.e., $(z^k, \lambda^k, s^k) \equiv (z, \lambda, s)$. Let us define the following functions, for all $i = 1, \dots, n$:

$$\begin{aligned} f_i(\alpha) &:= (z_i + \alpha \Delta z_i)(s_i + \alpha \Delta s_i) - \underline{\gamma}(z + \alpha \Delta z)^\top (s + \alpha \Delta s)/(n + m), \\ \bar{f}_i(\alpha) &:= \bar{\gamma}(z + \alpha \Delta z)^\top (s + \alpha \Delta s)/(n + m) - (z_i + \alpha \Delta z_i)(s_i + \alpha \Delta s_i), \\ h(\alpha) &:= (1 - (1 - \sigma_{\max})\alpha)z^\top s - (z + \alpha \Delta z)^\top (s + \alpha \Delta s), \\ g_d(\alpha) &:= (z + \alpha \Delta z)^\top (s + \alpha \Delta s) \\ &\quad - \gamma_d \|\hat{Q}(z + \alpha \Delta z) - \hat{A}^\top(\lambda + \alpha \Delta \lambda) + \hat{c} - (s + \alpha \Delta s)\|, \\ g_p(\alpha) &:= (z + \alpha \Delta z)^\top (s + \alpha \Delta s) - \gamma_p \|A(z + \alpha \Delta z) - b\|. \end{aligned}$$

Using Lemma 3.6 in the expressions of $g_d(\alpha)$ and $g_p(\alpha)$, we have

$$\begin{aligned} g_d(\alpha) &= (z + \alpha \Delta z)^\top (s + \alpha \Delta s) - \gamma_d(1 - \alpha) \|r_c\|, \\ g_p(\alpha) &\geq (z + \alpha \Delta z)^\top (s + \alpha \Delta s) - \gamma_p(1 - \alpha) \|r_b\|. \end{aligned}$$

We start by proving that there exists $\hat{\alpha}^k > 0$ such that

$$f_i(\alpha) \geq 0, \quad \bar{f}_i(\alpha) \geq 0, \quad h(\alpha) \geq 0, \quad g_p(\alpha) \geq 0, \quad g_d(\alpha) \geq 0$$

for all $i = 1, \dots, n+m$ and for all $\alpha \in [0, \hat{\alpha}^k]$. In the following, we will extensively use the identities in Lemma 3.6. We have

$$\begin{aligned} (3.20) \quad f_i(\alpha) &= \underbrace{(1 - \alpha) \left(z_i s_i - \underline{\gamma} \frac{z^\top s}{n+m} \right)}_{\geq 0} \\ &\quad + \alpha^2 \left(\Delta z_i \Delta s_i - \underline{\gamma} \frac{(\Delta z)^\top \Delta s}{n+m} \right) + \alpha \sigma (1 - \underline{\gamma}) \frac{z^\top s}{n+m}, \\ &\geq \alpha^2 \left(\Delta z_i \Delta s_i - \underline{\gamma} \frac{(\Delta z)^\top \Delta s}{n+m} \right) + \alpha \sigma (1 - \underline{\gamma}) \frac{z^\top s}{n+m}. \end{aligned}$$

Since $z^\top s > 0$, using a simple continuity argument, we can infer the existence of a small enough $\underline{f}_i > 0$ such that $f_i(\alpha) \geq 0$ for all $\alpha \in [0, \underline{f}_i]$. Reasoning analogously,

$$\begin{aligned} (3.21) \quad \bar{f}_i(\alpha) &= \underbrace{(1 - \alpha) \left(\bar{\gamma} \frac{z^\top s}{n+m} - z_i s_i \right)}_{\geq 0} \\ &\quad + \alpha^2 \left(\bar{\gamma} \frac{(\Delta z)^\top \Delta s}{n+m} - \Delta z_i \Delta s_i \right) + \alpha \sigma (\bar{\gamma} - 1) \frac{z^\top s}{n+m}, \\ &\geq \alpha^2 \left(\bar{\gamma} \frac{(\Delta z)^\top \Delta s}{n+m} - \Delta z_i \Delta s_i \right) + \alpha \sigma (\bar{\gamma} - 1) \frac{z^\top s}{n+m}, \end{aligned}$$

and hence, there exists a small enough $\bar{f}_i > 0$ such that $\bar{f}_i(\alpha) \geq 0$ for all $\alpha \in [0, \bar{f}_i]$.

Concerning $h(\alpha)$, we have

$$(3.22) \quad h(\alpha) = z^\top s (\sigma_{max} - \sigma) \alpha - \alpha^2 (\Delta z)^\top \Delta s,$$

and since $z^\top s (\sigma_{max} - \sigma) > 0$, there exists $\hat{h} > 0$ small enough such that we have $h(\alpha) \geq 0$ for all $\alpha \in [0, \hat{h}]$.

Concerning $g_d(\alpha)$, we have

$$\begin{aligned} (3.23) \quad g_d(\alpha) &= \underbrace{(1 - \alpha) (z^\top s - \gamma_d \|r_c\|)}_{\geq 0} + \alpha \sigma z^\top s + \alpha^2 (\Delta z)^\top \Delta s \\ &\geq \alpha \sigma z^\top s + \alpha^2 (\Delta z)^\top \Delta s, \end{aligned}$$

and hence, there exists $\hat{g}_d > 0$ small enough such that $g_d(\alpha) \geq 0$ for all $\alpha \in [0, \hat{g}_d]$.

Finally, concerning $g_p(\alpha)$, we have

$$\begin{aligned} (3.24) \quad g_p(\alpha) &\geq \underbrace{(1 - \alpha) (z^\top s - \gamma_p \|r_b\|)}_{\geq 0} + \alpha \sigma z^\top s + \alpha^2 (\Delta z)^\top \Delta s \\ &\geq \alpha \sigma z^\top s + \alpha^2 (\Delta z)^\top \Delta s, \end{aligned}$$

and hence there exists $\hat{g}_p > 0$ small enough such that $g_p(\alpha) \geq 0$ for all $\alpha \in [0, \hat{g}_p]$.

Let us define

$$\hat{\alpha}^k = \min \left\{ \min_i \underline{f}_i, \min_i \bar{f}_i, \hat{h}, \hat{g}_d, \hat{g}_p, 1 \right\} > 0.$$

To prove the thesis, it remains that

$$(z(\alpha), \lambda(\alpha), s(\alpha)) \in \mathbb{R}_{>0}^n \times \mathbb{R}^m \times \mathbb{R}_{>0}^n \text{ for all } \alpha \in [0, \hat{\alpha}^k].$$

To show the last part of the thesis, it is enough to prove that $\alpha^{*,k} > \hat{\alpha}^k$. To this aim, let us suppose by contradiction that $\alpha^{*,k} \leq \hat{\alpha}^k$. By definition of $\alpha^{*,k}$, there exists $\bar{\ell} \in \{1, \dots, n+m\}$ such that $(z_{\bar{\ell}} + \alpha^{*,k} \Delta z_{\bar{\ell}})(s_{\bar{\ell}} + \alpha^{*,k} \Delta s_{\bar{\ell}}) = 0$. Hence,

$$f_{\bar{\ell}}(\alpha^{*,k}) = -\underline{\gamma} (z(\alpha^{*,k}))^\top \frac{s(\alpha^{*,k})}{n+m} \geq 0 \implies (z(\alpha^{*,k}))^\top s(\alpha^{*,k}) = 0.$$

From this implication, using $g_d(\alpha^{*,k})$ and $g_p(\alpha^{*,k})$, we obtain that

$$\begin{aligned} Az(\alpha^{*,k}) - b &= 0, \\ \hat{Q}z(\alpha^{*,k}) + \hat{A}^\top \lambda(\alpha^{*,k}) - s(\alpha^{*,k}) + \hat{c} &= 0, \end{aligned}$$

i.e., $(z(\alpha^{*,k}), \lambda(\alpha^{*,k}), s(\alpha^{*,k}))$ is a solution of problem (3.4). $\square$

The next two results establish that Algorithm 3.1 converges to a solution of problem (3.4). This is done by establishing that the right-hand sides of the Newton systems are uniformly bounded, see Corollary 3.14, and by showing that the complementarity product $(z^k)^\top s^k$ cannot be bounded away from zero.

COROLLARY 3.14. *The right-hand sides of the Newton systems (3.12) are uniformly bounded.*

Proof. The proof of Corollary 3.14 is somehow standard, see Section D for more details. $\square$

THEOREM 3.15. *Suppose that $(z^0, \lambda^0, s^0) \in B(v^*, \delta^*) \cap \mathcal{N}(\bar{\gamma}, \underline{\gamma}, \gamma_p, \gamma_d)$. Then Algorithm 3.1 for solving the system (3.4) when $A_I \neq 0$, produces a sequence of iterates in $B(v^*, \delta^*) \cap \mathcal{N}(\bar{\gamma}, \underline{\gamma}, \gamma_p, \gamma_d)$ such that we have $\lim(z^k)^\top s^k = 0$, i.e., using the definition of $\mathcal{N}(\bar{\gamma}, \underline{\gamma}, \gamma_p, \gamma_d)$, (z^k, λ^k, s^k) converges to a solution of problem (3.4).*

Proof. From Lemma 3.12 and Theorem 3.13, we can assume that for all $k \in \mathbb{N}$, we have $(z^k, \lambda^k, s^k) \in B(v^*, \delta^*) \cap \mathcal{N}(\bar{\gamma}, \underline{\gamma}, \gamma_p, \gamma_d)$. Let us argue by contradiction supposing that there exists $\varepsilon^* > 0$ such that we have $(z^k)^\top s^k > \varepsilon^*$ for all $k \in \mathbb{N}$.

Claim 1 There exists a constant C_1 such that

$$\|[\Delta z^k, \Delta \lambda^k, \Delta s^k]^\top\| \leq C_1 \text{ for all } k \in \mathbb{N}.$$

The proof of this fact follows observing that in $B(v^*, \delta^*)$, the Newton matrices have a uniformly bounded inverse, and that the right-hand sides of (3.12) are uniformly bounded; see Lemma 3.3 and Corollary 3.14. As a consequence, there exists another constant C_2 such that

$$\begin{aligned} (3.25) \quad & \left| \Delta z_i^k \Delta s_i^k - \underline{\gamma} \frac{(\Delta z^k)^\top \Delta s^k}{n+m} \right| \leq C_2, \\ & \left| \bar{\gamma} \frac{(\Delta z^k)^\top \Delta s^k}{n+m} - \Delta z_i^k \Delta s_i^k \right| \leq C_2, \\ & |(\Delta z^k)^\top \Delta s^k| \leq C_2. \end{aligned}$$

Claim 2 There exists $\alpha^* > 0$ such that $\alpha^k \geq \alpha^*$ for all $j \in \mathbb{N}$.

Using (3.25) in equations (3.20), (3.21), (3.22), (3.23), and (3.24) we have

$$\begin{aligned} f_i(\alpha) &\geq \alpha^2 \left(\Delta z_i^k \Delta s_i^k - \underline{\gamma} \frac{(\Delta z^k)^\top \Delta s^k}{n+m} \right) + \alpha \sigma (1 - \underline{\gamma}) \frac{(z^k)^\top s^k}{n+m}, \\ &\geq -C_2 \alpha^2 + \alpha \sigma (1 - \underline{\gamma}) \frac{\varepsilon^*}{n+m}, \\ \bar{f}_i(\alpha) &\geq \left(\bar{\gamma} \frac{(\Delta z^k)^\top \Delta s^k}{n+m} - \Delta z_i^k \Delta s_i^k \right) + \alpha \sigma (\bar{\gamma} - 1) \frac{(z^k)^\top s^k}{n+m}, \\ &\geq -C_2 \alpha^2 + \alpha \sigma (\bar{\gamma} - 1) \frac{\varepsilon^*}{n+m}, \\ h(\alpha) &= (z^k)^\top s^k (\sigma_{max} - \sigma) \alpha - \alpha^2 (\Delta z^k)^\top \Delta s^k, \\ &\geq \varepsilon^* (\sigma_{max} - \sigma) \alpha - C_2 \alpha^2, \\ g_d(\alpha) &\geq \alpha \sigma (z^k)^\top s^k + \alpha^2 (\Delta z^k)^\top \Delta s^k \geq \varepsilon^* \sigma \alpha - C_2 \alpha^2, \\ g_p(\alpha) &\geq \alpha (\sigma - q \gamma_p) (z^k)^\top s^k + \alpha^2 (\Delta z^k)^\top \Delta s^k \geq \varepsilon^* \alpha \sigma - C_2 \alpha^2. \end{aligned}$$

Hence, $\alpha^k \geq \alpha^*$ with

$$\alpha^* := \min \left\{ 1, \frac{\sigma(1 - \underline{\gamma})\varepsilon^*}{C_2(n+m)}, \frac{\sigma(\bar{\gamma} - 1)\varepsilon^*}{C_2(n+m)}, \frac{(\sigma_{max} - \sigma)\varepsilon^*}{C_2}, \frac{\sigma\varepsilon^*}{C_2} \right\}.$$

The convergence claim follows by observing that the inequality

$$\varepsilon^* \leq (z^k)^\top s^k \leq (1 - (1 - \sigma_{max})\alpha^*)^k (z^0)^\top s^0$$

leads to a contradiction when $k \rightarrow \infty$. $\square$

Before concluding this section, we would like to emphasize that the above proof complements and expands [23, Remark 3.1] and [6, Section 3].

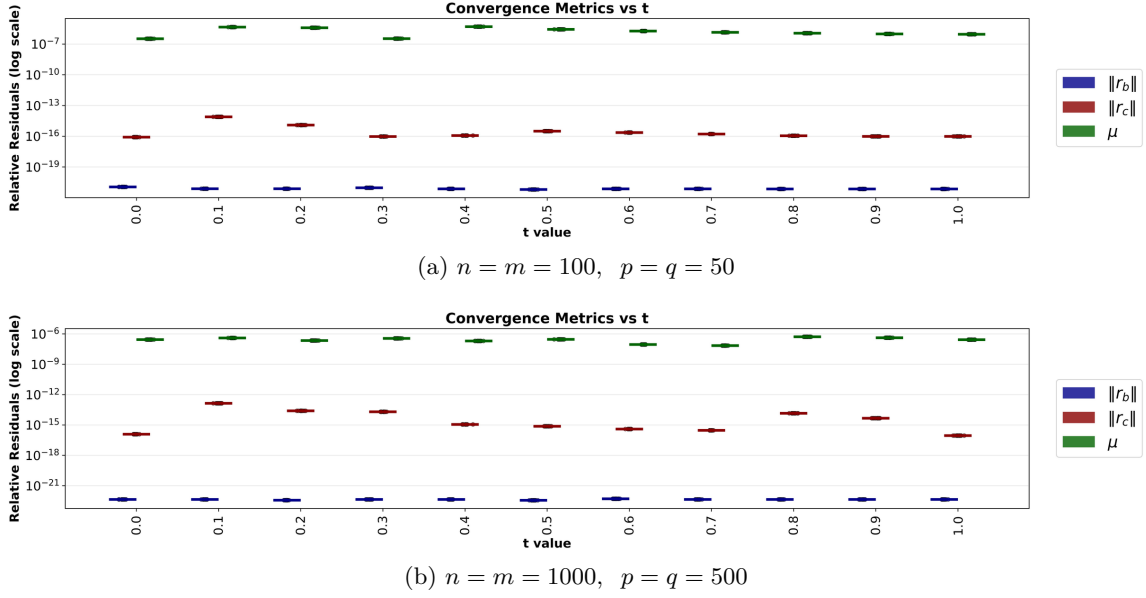


Fig. 2: Convergence metrics versus problem scaling parameter t . Grouped boxplots (three per t value) show primal feasibility residual $\|r_b\|$ (blue), dual feasibility residual $\|r_c\|$ (red), and complementarity measure μ (green) across 15 independent trials for each $t \in [0, 1]$.

4 Numerical experiments. All the numerical results are obtained using Python 3 on a MacBook Pro using Apple M3 Pro CPU with 18 GB of RAM. The code is publicly available at <https://github.com/StefanoCipolla/MinMax>.

4.1 Implementation Details. As is customary in the IPM literature, the implementation usually deviates from the theoretical developments in order to gain computational efficiency. In particular, we do not require the iterates of Algorithm 3.1 to lie in any of the neighborhoods previously defined, using, instead, a well-known Predictor-Corrector strategy [28]. Newton's directions are computed by solving (3.12) by a well-known 2×2 block reduction, see, e.g. [14]. Moreover, following a standard practice in IPM literature, stopping criteria are chosen as

$$\|\hat{Q}z + \hat{A}^\top \lambda + \hat{c} - s\|_\infty \leq R \cdot \text{tol} \quad \wedge \quad \|Az - b\|_1 \leq R \cdot \text{tol} \quad \wedge \quad c_{z,s} \leq \text{tol},$$

where

$$\text{tol} = 10^{-6}, \quad R := \max\{\|Q\|_\infty, \|A\|_\infty, \|b\|_1, \|c\|_1\},$$

and

$$c_{z,s} := \max_i \{\min\{|(z_i s_i)|, |z_i|, |s_i|\}\}.$$

For the initial point, we adapt the technique from [28].

4.2 Dataset. The random instances, used for the experiments in Subsections 4.3 and 4.4, are generated by adapting the procedure used in [5] for SDP problems. In particular, such random instances are generated with varying dimensions (n, m, p, q) . The vector (z^*, λ^*, s^*) is drawn uniformly from $[10^{-3}, 10^3]$ with strict complementarity enforced by setting z to zero at $k = \min\{10, \lfloor (n+m)/5 \rfloor\}$ random positions and ensuring s is non-zero only at these positions. Blocks Q_{11}, Q_{22} are constructed via spectral decomposition $UDU^\top$ with controlled eigenvalue spectra to ensure positive definiteness. Constraint matrices have standard Gaussian entries, with $A_I = 0$ for the case with decoupled constraints. The routines to generate the random instances are publicly available at <https://github.com/StefanoCipolla/MinMax>.

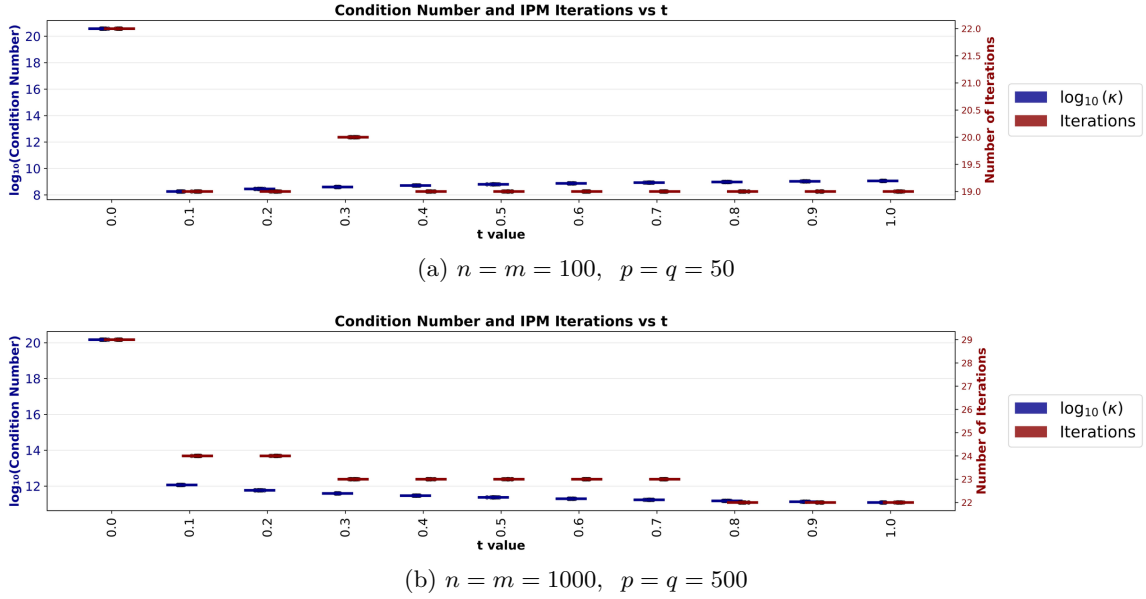


Fig. 3: Stationarity matrix condition number and iteration count versus problem scaling parameter t . Grouped boxplots (two per t value) display $\log_{10}(\kappa)$ (blue) and IPM iteration count (red) across 15 trials. Despite the significant increase in condition number at $t = 0$ (loss of strong convexity-concavity), the iteration count remains stable at approximately 20 iterations.

4.3 Decoupled constraints case. The objective of this section is to demonstrate the robustness of the proposed IPM with respect to the degradation of strong convexity and strong concavity properties in the decoupled case ($A_I = 0$). To systematically assess this robustness, we generate random instances of the problem data ($\hat{Q}, \hat{c}, b, A_O, B_I$) and parametrize the diagonal blocks of the matrix $\hat{Q}$ as tQ_{11} and tQ_{22} for $t \in [0, 1]$. As t approaches zero, the strong convexity of the maximization objective and the strong concavity of the minimization objective are progressively eliminated, allowing us to evaluate algorithmic performance as the problem structure degenerates from strongly convex-strongly concave to merely convex-concave.

Figure 2 presents the convergence behavior of the IPM across this parametrized family of problems. The upper panel displays results for problem dimensions $n = m = 100$ and $p = q = 50$, while the lower panel presents results for the larger-scale setting with $n = m = 1000$ and $p = q = 500$. Similarly, Figure 3 illustrates the iteration count and numerical conditioning of the Jacobian of the stationarity system $D\hat{F}(z^*, \lambda^*, s^*)$ for both problem sizes, with the upper and lower panels corresponding to the small-scale and larger-scale instances, respectively. The numerical experiments reveal several important findings that underscore the robustness of the proposed method:

Convergence accuracy. As expected from a well-designed IPM, the magnitude of the relative residuals remains below the selected tolerance $\text{tol} = 1e - 6$, see Figure 2, demonstrating that the algorithm consistently delivers solutions meeting the specified accuracy requirements regardless of the degree of strong convexity or strong concavity.

Iteration stability. Perhaps most remarkably, the iteration count exhibits striking consistency across the entire parameter range $t \in [0, 1]$. Despite the substantial degradation in problem conditioning as $t \rightarrow 0$ – where the diagonal blocks Q_{11} and Q_{22} vanish entirely – the number of iterations required for convergence shows only negligible variation. This stability persists even as the condition number κ of the Jacobian $D\hat{F}(z^*, \lambda^*, s^*)$ increases significantly, as evidenced in Figure 3.

Dimension scalability. The comparison of the upper and lower panels of Figure 3 reveals the algorithm’s favorable scaling properties. Increasing the problem dimension by one order of magnitude – from $n = m = 100$ to $n = m = 1000$ – results in virtually no increase in iteration count when $t \rightarrow 1$ (strongly convex-concave regime). Even in the most challenging scenario where $t \rightarrow 0$ (complete loss of strong convexity-concavity), the average iteration count increases by only approximately 5 iterations. This mild growth in iteration counts w.r.t. the problem dimension provides strong empirical support for the polynomial complexity established theoretically in Subsection 3.3.

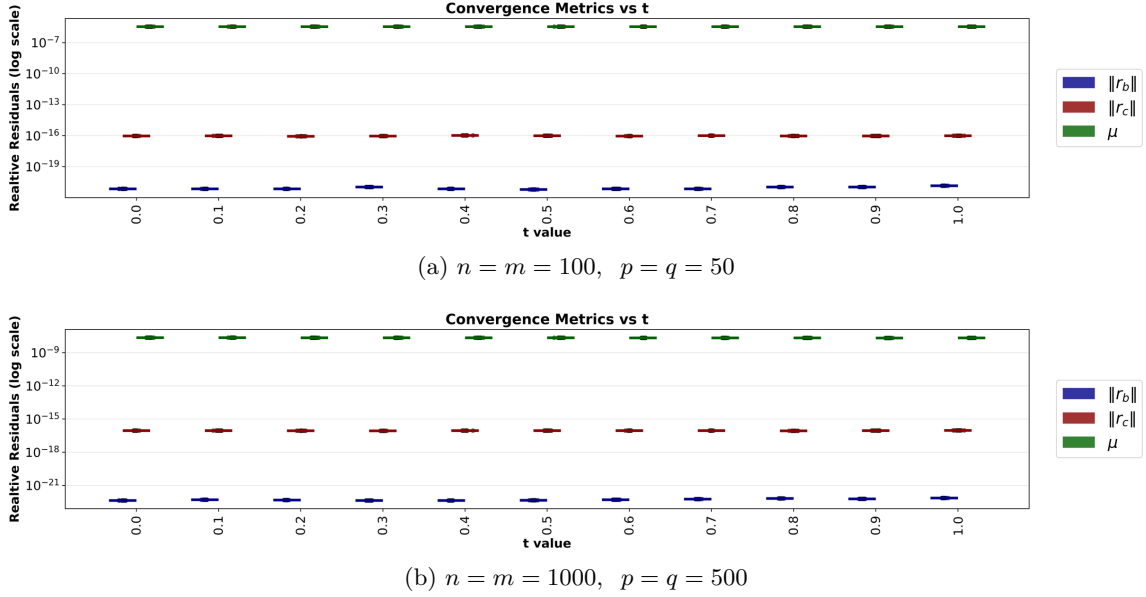


Fig. 4: Convergence metrics versus constraints coupling parameter t . Grouped boxplots (three per t value) show primal feasibility residual $\|r_b\|$ (blue), dual feasibility residual $\|r_c\|$ (red), and complementarity measure μ (green) across 15 independent trials for each $t \in [0, 1]$.

4.4 Coupled constraints case. The objective of this section is to demonstrate the behavior of the proposed IPM framework in the *coupled constraints* case. Analogous to the experiments in the previous section, we examine this behavior by *calibrating the strength* of the coupling matrix A_I . Specifically, we consider tA_I for $t \in [0, 1]$, noting that the case $t = 0$ recovers the decoupled constraints setting.

It is important to emphasize that the numerical experiments presented in this section employ initial points generated as random perturbations of the solution of each problem instance. We deliberately select the perturbation magnitude to be 0.3, which is two orders of magnitude larger than the quantity a defined in (3.7) – in our case, $a \in \mathcal{O}(10^{-3})$. As discussed previously, this quantity provides an estimate for the radius of local convergence of Algorithm 3.1. Our choice of the perturbation magnitude aims to demonstrate that this theoretical estimate may be overly conservative for certain problem classes, such as the randomly generated instances considered here with dimensions as in the previous section.

Figure 4 presents the convergence behavior of the IPM across this parametrized family of problems. Similarly, Figure 5 illustrates the iteration count and numerical conditioning of the Jacobian of the stationarity system $D\hat{F}(z^*, \lambda^*, s^*)$, with the upper and lower panels corresponding to the small-scale and larger-scale instances, respectively. The numerical experiments presented here confirm the robustness, stability and scalability observed in the previous case, showcasing, as well, the theoretical local convergence results proved in Subsection 3.4. Indeed, Figure 4 supports the local convergence result presented in Subsection 3.4, as Algorithm 3.1 is able to deliver the required accuracy across all the t when initialized sufficiently close to the solution. Figure 5 showcases, instead, how the number of IPM iterations remains roughly constant when the *strength* of the coupling constraint term increases ($t \rightarrow 1$) while the conditioning of the Jacobian of the stationarity system remains almost unchanged.

4.5 Min cost flow problems and adversarial attacks. The aim of this section is to showcase the *quality* of the computed stationary points using our proposed Algorithm 3.1. For this purpose, we consider a well-established class of adversarial attacks on networks that can be formulated using the *min-max-with-coupled-constraints* analyzed in this work, see [36, 44, 19]. Such a problem can be formulated as

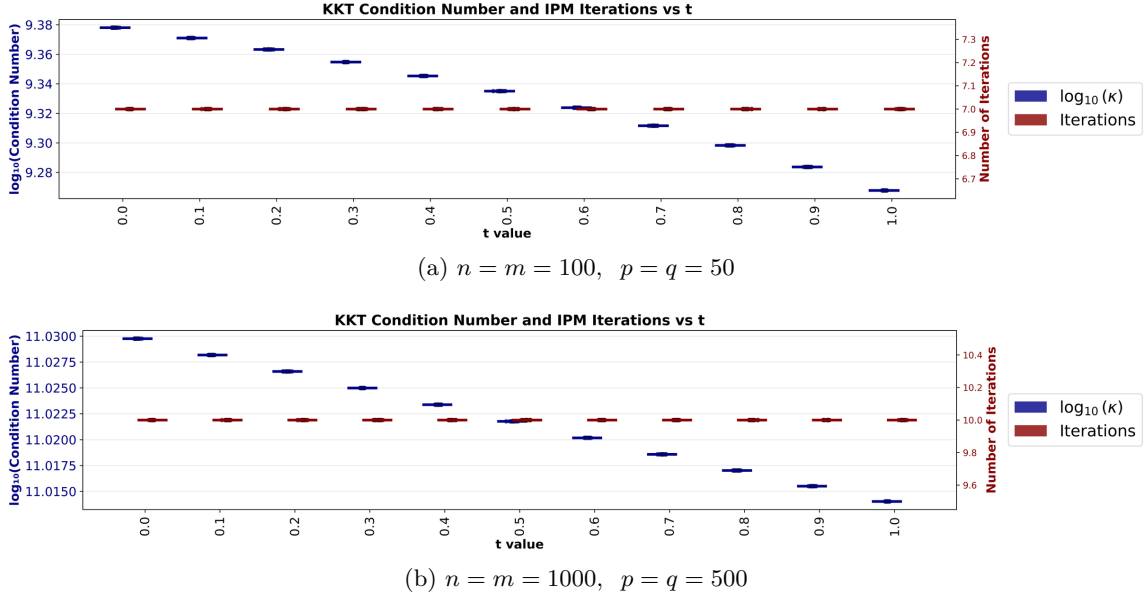


Fig. 5: Stationarity matrix condition number and iteration count versus constraints coupling parameter t . Grouped boxplots (two per t value) display $\log_{10}(\kappa)$ (blue) and IPM iteration count (red) across 15 trials. Despite the significantly increasing strength of constraints coupling at $t = 1$, the iteration count remains stable at approximately 20 iterations.

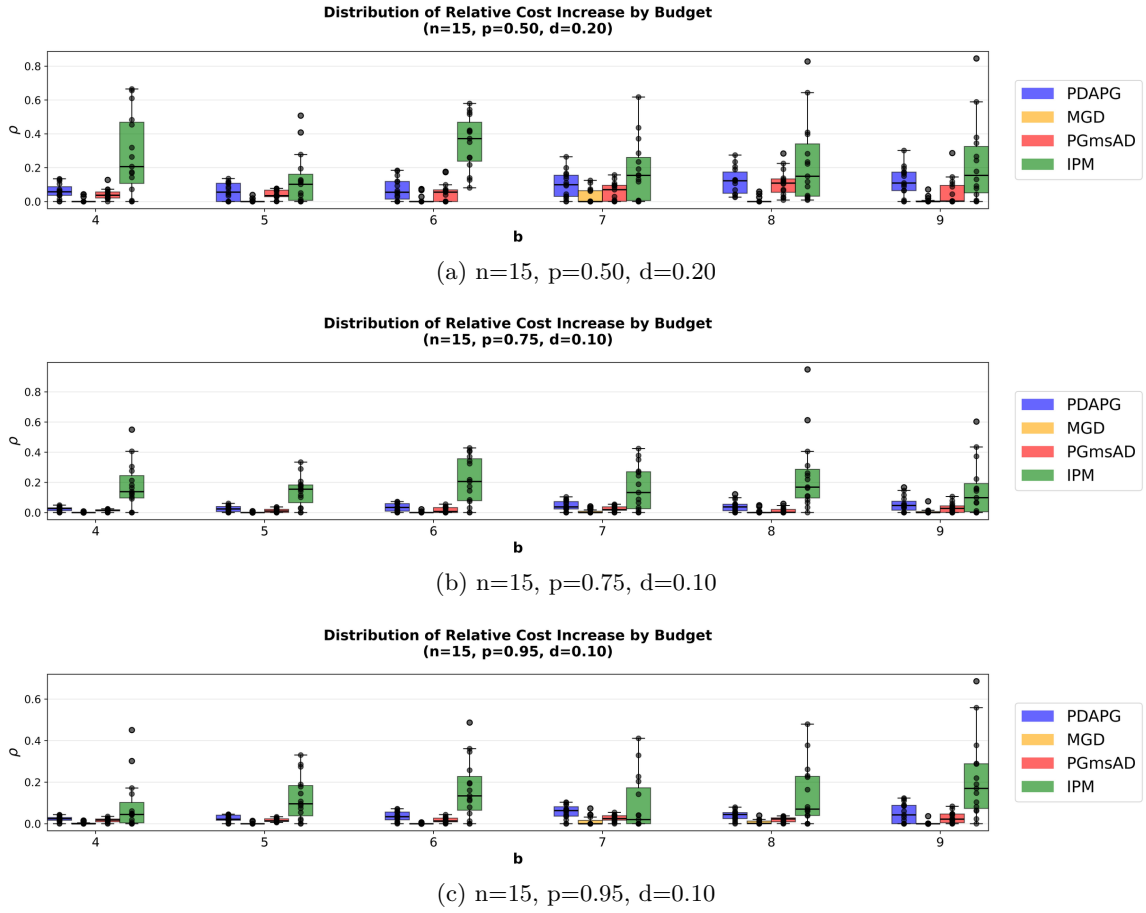


Fig. 6: Comparison of algorithms for network attack: relative cost increase vs. budget parameter.

$$\begin{aligned}
(4.1) \quad & \min_{\substack{(x,z) \geq 0 \\ e^\top x = a_b}} \max_{y \geq 0} \sum_{e \in E} -w^\top y - y^\top \text{diag}(w)y - y^\top \text{diag}(w)x + \frac{\eta}{2}\|x\|^2 + \frac{\eta}{2}\|z\|^2 \\
& x + y + z = p, \\
& \sum_{e \in E: e=(v,t)} y_e = r_t, \\
& P^{s,t}y = 0.
\end{aligned}$$

The notation details for problem (4.1) and a detailed problem description are reported Section E. It can be readily seen that problem (4.1) can be recast in the form (1), observing that the inner can be written as

$$(4.2) \quad \begin{bmatrix} e^\top & 0 & 0 \\ 0 & 0 & P_{s,t} \\ 0 & 0 & \hat{e}^\top \\ \mathbf{1} & \mathbf{1} & \mathbf{1} \end{bmatrix} \begin{bmatrix} x \\ z \\ y \end{bmatrix} = \begin{bmatrix} a_b \\ 0 \\ r_t \\ p \end{bmatrix},$$

i.e., $A_O := [e^\top, 0]$, $b_O := a_b$, $A_I := \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \mathbf{1} & \mathbf{1} \end{bmatrix}$, $B_I := \begin{bmatrix} P_{s,t} \\ \hat{e}^\top \\ \mathbf{1} \end{bmatrix}$, and $b_I := \begin{bmatrix} 0 \\ r_t \\ p \end{bmatrix}$. Concerning the objective function, we have

$$\begin{aligned}
Q_{11} &:= \begin{bmatrix} \eta I & 0 \\ 0 & \eta I \end{bmatrix}, \quad Q_{12} := \begin{bmatrix} -\frac{1}{2} \text{diag}(w) \\ 0 \end{bmatrix}, \quad Q_{22} =: \text{diag}(w), \\
c_x &= [0, 0], \quad c_y = -w^\top.
\end{aligned}$$

Having computed a solution (x^*, z^*, y^*) of problem (4.1), we recompute the min-cost flow

$$\begin{aligned}
\rho_{att} &:= \min \sum_{e \in E} c_e(y_e)y_e \\
&\text{s.t. } Py = b, \\
&0 \leq y \leq p - x^*,
\end{aligned}$$

and, as a measure of the quality of the computed solution, we consider

$$\rho = \frac{\rho_{att} - \rho_{cf}}{\rho_{cf}},$$

where ρ_{cf} and ρ_{att} are, as defined above, the cost flow before and after the attack. Clearly, bigger ρ s correspond to *more effective attacks*. In the following, the total flow r_t is $100 * d\%$ of the sum of capacities of the edges exiting the source, with d being a parameter to be chosen, whereas p is the probability that an edge appears in the Erdos-Renyi model random graph. In Figure 6, we report numerical results for selected d and p , of the approach proposed here, IPM, when compared with PGmsAD and MGD, see [36], and PDAPG, see [44]. The implementation of the three baseline methods and the selection of hyperparameters are based on the code kindly provided by the authors of [44]. In particular, we choose as regularization parameter $\eta = 1e - 2$ whereas we redirect the interested reader to [44] for all other details, including how the instances are generated.

The results presented in Figure 6 showcase how Algorithm 3.1, when applied to solve problem (4.1), is able to generate, typically, more effective *attacks* w.r.t. the obtained ρ , than the baseline terms of comparison. Moreover, since Algorithm 3.1 is an infeasible one, in contrast to PGmsAD, MGD and PDAPG, we complement Figure 6 with Figures 7, 8 and 9 in Section E, which explore the correlation of $\|r_b\|$, i.e., the residual associated to (4.2), with ρ . Such figures clearly show, indeed, how the effectiveness of the attack is not positively correlated with the infeasibility of the computed point and, in general, effective attacks are obtained also when such infeasibility measure is very small.

5 Conclusions. This paper presents a comprehensive framework for computing stationary points of quadratic minmax optimization problems with coupled linear constraints. Our main contributions span theoretical characterizations, algorithmic development, and practical validation.

We developed an infeasible interior point method (Algorithm 3.1) that computes stationary points by following a local central path associated with the perturbed stationarity system (3.2). For the important special case of decoupled constraints ($A_I = 0$, cf. Remark 2.4), we established polynomial complexity with iteration count $O((n + m)^2 |\log(\varepsilon)|)$ to achieve ε -approximate solutions without assuming strong convexity-concavity (Theorem 3.9), as it is the case in the existing literature (see [36, Section 2]). This result demonstrates that the stationary points of minmax problems with decoupled constraints can be computed efficiently. For the general case with coupled constraints ($A_I \neq 0$), while polynomial complexity cannot be expected due to the inherent non-convexity of the equivalent formulation (QP), we proved local convergence under the assumption that the initial point lies in a neighborhood of a non-degenerate stationary point (Theorem 3.15).

Our computational experiments on randomly generated instances demonstrate remarkable robustness across a spectrum of problem structures. The algorithm maintains consistent iteration counts even as strong convexity-concavity properties degrade completely, and as problem dimensions grow. These empirical findings strongly support our polynomial complexity theory for the decoupled case and demonstrate practical local convergence for the coupled case. Applications to adversarial network flow interdiction problems (4.1) showcase the algorithm's ability to compute high-quality solutions that significantly increase network costs under optimal attacks.

It might be important to note that the quadratic setting of the minmax problem studied in this paper enables the derivation of the necessary optimality conditions of problem (P_O) for free (cf. Corollary 2.5) and construction of the polynomiality results highlighted above. As far as the other main results of the paper are concerned, they could be extended to more general smooth convex-concave minmax problems under suitable versions of our assumptions. Of course, if nonlinear outer and/or inner constraints are involved in this general settings, the validity of a constraint qualifications will also become important in the corresponding extensions. Also, in the extension of our methods to such general settings, multivariate first and second order derivatives will be necessary, rather than the more explicit matrix-vector expressions used in this paper.

Acknowledgments. This work was initiated while the 3rd author was visiting the Continuous Optimization Chair at the Institute for Operations Research, Karlsruhe Institute of Technology (KIT), as part of his Alexander von Humboldt Research Fellowship for Experienced Researchers.

REFERENCES

- [1] P. ARMAND AND J. BENOIST, *A local convergence property of primal-dual methods for nonlinear programming*, Math. Program., 115 (2008), pp. 199–222, <https://doi.org/10.1007/s10107-007-0136-2>.
- [2] A. BEN-TAL, L. EL GHAOU, AND A. NEMIROVSKI, *Robust optimization*, Princeton Series in Applied Mathematics, Princeton University Press, Princeton, NJ, 2009, <https://doi.org/10.1515/9781400831050>.
- [3] A. BEN-TAL AND J. ZOWE, *Necessary and sufficient optimality conditions for a class of nonsmooth minimization problems*, Math. Program., 24 (1982), pp. 70–91, <https://doi.org/10.1007/BF01585095>.
- [4] J. F. BONNANS AND A. SHAPIRO, *Perturbation analysis of optimization problems*, Springer, 2000.
- [5] S. CIPOLLA AND J. GONDZIO, *Proximal-stabilized semidefinite programming*, Comput. Optim. Appl., 91 (2025), pp. 573–616, <https://doi.org/10.1007/s10589-024-00614-3>.
- [6] S. CIPOLLA, J. GONDZIO, AND F. ZANETTI, *A regularized interior point method for sparse optimal transport on graphs*, European J. Oper. Res., 319 (2024), pp. 413–426, <https://doi.org/10.1016/j.ejor.2023.11.027>.
- [7] R. W. COTTLE, J.-S. PANG, AND R. E. STONE, *The linear complementarity problem*, vol. 60 of Classics in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2009, <https://doi.org/10.1137/1.9780898719000.ch1>.
- [8] Y.-H. DAI, J. WANG, AND L. ZHANG, *Optimality conditions and numerical algorithms for a class of linearly constrained minimax optimization problems*, SIAM J. Optim., 34 (2024), pp. 2883–2916, <https://doi.org/10.1137/22M1535243>.
- [9] Y.-H. DAI AND L. ZHANG, *Optimality conditions for constrained minimax optimization*, CSIAM Transactions on Applied Mathematics, 1 (2020), pp. 296–315, <https://doi.org/10.4208/csiam-am.2020-0014>.
- [10] Y.-H. DAI AND L.-W. ZHANG, *The rate of convergence of augmented Lagrangian method for minimax optimization problems with equality constraints*, J. Oper. Res. Soc. China, 12 (2024), pp. 265–297, <https://doi.org/10.1007/s40305-022-00439-z>.
- [11] F. FARNIA AND D. TSE, *A minimax approach to supervised learning*, Advances in Neural Information Processing Systems, 29 (2016).
- [12] A. V. FIACCO AND G. P. MCCORMICK, *Nonlinear programming*, vol. 4 of Classics in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second ed., 1990, <https://doi.org/10.1137/1.9781611971316>. Sequential unconstrained minimization techniques.
- [13] D. GOKTAS AND A. GREENWALD, *Convex-concave min-max stackelberg games*, Advances in Neural Information Processing Systems, 34 (2021), pp. 2991–3003.
- [14] J. GONDZIO, *Interior point methods in the year 2025*, EURO J. Comput. Optim., 13 (2025), pp. Paper No. 100105, 23, <https://doi.org/10.1016/j.ejco.2025.100105>.

- [15] N. I. M. GOULD, *On practical conditions for the existence and uniqueness of solutions to the general equality quadratic programming problem*, Math. Program., 32 (1985), pp. 90–99, <https://doi.org/10.1007/BF01585660>.
- [16] H. GÜNZEL, H. T. JONGEN, AND O. STEIN, *Generalized semi-infinite programming: on generic local minimizers*, Journal of Global Optimization, 42 (2008), pp. 413–421.
- [17] E. V. HAYNSWORTH, *Determination of the inertia of a partitioned Hermitian matrix*, Linear Algebra Appl., 1 (1968), pp. 73–81, [https://doi.org/10.1016/0024-3795\(68\)90050-5](https://doi.org/10.1016/0024-3795(68)90050-5).
- [18] R. P. HETTICH AND H. T. JONGEN, *Semi-infinite programming: conditions of optimality and applications*, in Optimization techniques (Proc. 8th IFIP Conf., Würzburg, 1977), Part 2, vol. Vol. 7 of Lect. Notes Control Inf. Sci., Springer, Berlin-New York, 1978, pp. 1–11.
- [19] X. HU, K.-C. TOH, S. WANG, AND N. XIAO, *A minimization approach for minimax optimization with coupled constraints*, 2024, <https://doi.org/https://arxiv.org/abs/2408.17213>.
- [20] C. JIN, P. NETRAPALLI, AND M. JORDAN, *What is local optimality in nonconvex-nonconcave minimax optimization?*, in International conference on machine learning, PMLR, 2020, pp. 4880–4889.
- [21] H. T. JONGEN, P. JONKER, AND F. TWILT, *Critical sets in parametric optimization*, Math. Program., 34 (1986), pp. 333–353, <https://doi.org/10.1007/BF01582234>.
- [22] H. T. JONGEN, T. MÖBERT, J. RÜCKMANN, AND K. TAMMER, *On inertia and Schur complement in optimization*, Linear Algebra Appl., 95 (1987), pp. 97–109, [https://doi.org/10.1016/0024-3795\(87\)90028-0](https://doi.org/10.1016/0024-3795(87)90028-0).
- [23] M. KOJIMA, N. MEGIDDO, AND S. MIZUNO, *A primal-dual infeasible-interior-point algorithm for linear programming*, Math. Program., 61 (1993), pp. 263–280, <https://doi.org/10.1007/BF01582151>.
- [24] Q. LIU AND J. WANG, *A projection neural network for constrained quadratic minimax optimization*, IEEE Transactions on Neural Networks and Learning Systems, 26 (2015), pp. 2891–2900.
- [25] Z. LU AND S. MEI, *A first-order method for nonconvex-strongly-concave constrained minimax optimization*, <https://zhaosong-lu.github.io/ResearchPapers/strongly-cvx-minimax.pdf>, (2023).
- [26] Z. LU AND S. MEI, *A first-order augmented Lagrangian method for constrained minimax optimization*, Math. Program., 213 (2025), pp. 1063–1104, <https://doi.org/10.1007/s10107-024-02163-3>.
- [27] P. MEHLITZ AND A. B. ZEMKOHO, *Sufficient optimality conditions in bilevel programming*, Math. Oper. Res., 46 (2021), pp. 1573–1598, <https://doi.org/10.1287/moor.2021.1122>.
- [28] S. MEHROTRA, *On the implementation of a primal-dual interior point method*, SIAM J. Optim., 2 (1992), pp. 575–601, <https://doi.org/10.1137/0802028>.
- [29] R. D. C. MONTEIRO AND J.-S. PANG, *Properties of an interior-point mapping for mixed complementarity problems*, Math. Oper. Res., 21 (1996), pp. 629–654, <https://doi.org/10.1287/moor.21.3.629>.
- [30] B. S. MORDUKHOVICH, N. M. NAM, AND H. M. PHAN, *Variational analysis of marginal functions with applications to bilevel programming*, J. Optim. Theory Appl., 152 (2012), pp. 557–586, <https://doi.org/10.1007/s10957-011-9940-1>.
- [31] H. RAFIQUE, M. LIU, Q. LIN, AND T. YANG, *Weakly-convex-concave min-max optimization: provable algorithms and applications in machine learning*, Optim. Methods Softw., 37 (2022), pp. 1087–1121, <https://doi.org/10.1080/10556788.2021.1895152>.
- [32] M. RAZAVIYAYN, T. HUANG, S. LU, M. NOUIEHED, M. SANJABI, AND M. HONG, *Nonconvex min-max optimization: Applications, challenges, and recent theoretical advances*, IEEE Signal Processing Magazine, 37 (2020), pp. 55–66.
- [33] O. STEIN, *A guided tour through the zoo of paired optimization problems*, preprint, Optimization Online, 2025, <https://optimization-online.org/?p=32394>.
- [34] O. STEIN, *A tutorial on properties of the epigraph reformulation*, EURO Journal on Computational Optimization, 13 (2025), p. 100109.
- [35] O. STEIN AND G. STILL, *On generalized semi-infinite optimization and bilevel optimization*, European Journal of Operational Research, 142 (2002), pp. 444–462.
- [36] I. TSAKNAKIS, M. HONG, AND S. ZHANG, *Minimax problems with coupled linear constraints: computational complexity and duality*, SIAM J. Optim., 33 (2023), pp. 2675–2702, <https://doi.org/10.1137/21M1462428>.
- [37] T. WANG, R. D. C. MONTEIRO, AND J.-S. PANG, *An interior point potential reduction method for constrained equations*, Math. Program., 74 (1996), pp. 159–195.
- [38] W. WETTERLING, *Definitheitsbedingungen für relative Extrema bei Optimierungs- und Approximationsaufgaben*, Numer. Math., 15 (1970), pp. 122–136, <https://doi.org/10.1007/BF02165377>.
- [39] P. WOLFE, *A duality theorem for non-linear programming*, Quart. Appl. Math., 19 (1961), pp. 239–244, <https://doi.org/10.1090/qam/135625>.
- [40] S. J. WRIGHT, *Primal-dual interior-point methods*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997, <https://doi.org/10.1137/1.9781611971453>.
- [41] S. ZAKOVIĆ, C. PANTELIDES, AND B. RUSTEM, *An interior point algorithm for computing saddle points of constrained continuous minimax*, vol. 99, 2000, pp. 59–77, <https://doi.org/10.1023/A:1019284715657>. Applied mathematical programming and modeling, IV (Limassol, 1998).
- [42] A. B. ZEMKOHO, *A simple approach to optimality conditions in minmax programming*, Optimization, 63 (2014), pp. 385–401, <https://doi.org/10.1080/02331934.2011.653788>.
- [43] H. ZHANG, J. WANG, Z. XU, AND Y.-H. DAI, *Primal dual alternating proximal gradient algorithms for nonsmooth nonconvex minimax problems with coupled linear constraints*, arXiv:2212.04672v3, (2024).
- [44] H. ZHANG, Z. XU, AND Y.-H. DAI, *Iterative minimax games with coupled linear constraints*, 2025, <https://doi.org/https://arxiv.org/abs/2212.04672>.

Appendix A. An Alternative Meaningfulness Result. Here, we present an alternative meaningfulness result to Theorem 2.7, where strict complementarity slackness is not necessary. The approach is completely different from the one in Subsection 2.2, and will mainly consists of showing that under appropriate conditions, a stationarity point in the sense of (1.7) is a strict local optimal solution, in the sense that it satisfies the *second order growth condition*. The second order growth condition will be said to hold at a point $\bar{x} \in C \subset \mathbb{R}^n$ for an optimization problem to minimize $\psi(x)$ subject to $x \in C$, if there exist a neighborhood U of $\bar{x}$ and a constant $\alpha > 0$ such that

$$(A.1) \quad \psi(\bar{x}) + \alpha \|x - \bar{x}\|^2 \leq \psi(x) \quad \text{for all } x \in C \cap U.$$

To proceed, we need a few other concepts, including first and second order directional derivatives, as well as the notion of second order epi-regularity. The *directional derivative* (in the sense of Gâteaux) of a function $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$ at a point $\bar{x} \in \text{dom } \psi$ in the direction $d \in \mathbb{R}^n$ is the following limit if it exists:

$$\psi'(\bar{x}; d) = \lim_{t \downarrow 0} \frac{\psi(\bar{x} + td) - \psi(\bar{x})}{t}.$$

Based on this first order notion, we can introduce the second order directional derivative concept that seems to have originated from [3], and which will play a key role in establishing the results in this section. If the function ψ is directional differentiable at $\bar{x}$, its *second order directional derivative* at this point, in the directions $d \in \mathbb{R}^n$ and $w \in \mathbb{R}^n$, is the following limit if it exists:

$$\psi''(\bar{x}; d, w) = \lim_{t \downarrow 0} \frac{\psi(\bar{x} + td + \frac{1}{2}t^2w) - \psi(\bar{x}) - t\psi'(\bar{x}; d)}{\frac{1}{2}t^2}.$$

By means of the Taylor expansion of order two, we can easily check that if ψ is twice continuously differentiable at $\bar{x}$, then we have $\psi''(\bar{x}; d, w) = \nabla \psi(\bar{x})^\top w + d^\top \nabla^2 \psi(\bar{x}) d$.

The concept of epi-regularity will be associated to that of second order directional differentiability to generate our main result here. So, we next introduce the former. Let ψ be an extended real-valued function that is second order directionally differentiable at a point $\bar{x} \in \text{dom } \psi$. It will be said to be *second-order epi-regular* at $\bar{x}$ in direction $d \in \mathbb{R}^n$ if for each path $w : \mathbb{R}_+ \rightarrow \mathbb{R}^n$ such that $tw(t) \rightarrow 0$ as $t \downarrow 0$,

$$\psi(\bar{x} + td + \frac{1}{2}w(t)) \geq \psi(\bar{x}) + t\psi'(\bar{x}; d) + \frac{1}{2}t^2\psi''(\bar{x}; w(t)) + o(t).$$

The function ψ will be said to be second-order epi-regular at $\bar{x}$ if it is second-order epi-regular at $\bar{x}$ in each direction from $\mathbb{R}^n$.

The following lemma will serve as base to establish the alternative meaningfulness result here.

LEMMA A.1 ([4]). *The second order growth condition holds at $\bar{x} \in \text{dom } \psi$ for the minimization of the function $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$ if the following conditions hold:*

- (i) ψ is second order directionally differentiable and second order epi-regular;
- (ii) $\psi'(\bar{x}; d) \geq 0$ for all $d \in \mathbb{R}^n$;
- (iii) $\inf_{\omega \in \mathbb{R}^n} \psi''(\bar{x}; d, \omega) > 0$ for all $d \in \{r \in \mathbb{R}^n \mid \psi'(\bar{x}; r) = 0\}$.

Before stating our result, we introduce the suitable versions of the inner and outer second order sufficient conditions that will be used here. To proceed, note that the *inner critical cone* at $\bar{y}$ is defined by

$$\mathcal{C}_{Y(\bar{x})}(\bar{y}) := \left\{ d \in \mathbb{R}^m \mid (Q_{22}\bar{y} - Q_{12}^\top \bar{x} - c_y)^\top d = 0, \ B_I d = 0, \ d_j \geq 0, \ j \in \mathcal{I}_{\bar{y}} \right\}$$

for a fixed outer variable $x := \bar{x}$. Thanks to this notion, we can introduce the alternative inner second order sufficient condition, which will be said to hold at $(\bar{x}, \bar{y})$ if

$$(ISOSC') \quad d^\top Q_{22} d > 0 \quad \text{for all } d \in \mathcal{C}_{Y(\bar{x})}(\bar{y}) \setminus \{0\}.$$

It is well-known that under the (ISOSC'), it holds that $\{\bar{y}\} = S(\bar{x})$. Subsequently, under the fulfillment of the (ILICQ) and (ISOSC') at $\bar{y}$ and $(\bar{x}, \bar{y})$ with $\{\bar{y}\} = S(\bar{x})$, respectively, we can introduce the outer critical cone

$$(A.2) \quad \mathcal{C}_X(\bar{x}) := \left\{ d \in \mathbb{R}^n \mid (Q_{11}\bar{x} + Q_{12}\bar{y} + c_x + A_I^\top \lambda_I)^\top d = 0, \right. \\ \left. A_O d = 0, \ d_i \geq 0, \ i \in \mathcal{I}_{\bar{x}} \right\},$$

where λ_I is such that the pair (λ_I, s_I) is the unique Lagrange multiplier pair corresponding to the point $(\bar{x}, \bar{y})$ with $\{\bar{y}\} = S(\bar{x})$. Analogously, based on the cone (A.2), we can introduce the outer second order sufficient condition to be used here, which will be said to hold at $(\bar{x}, \bar{y})$ if

$$(\text{OSOSC}') \quad \forall d \in \mathcal{C}_X(\bar{x}) \setminus \{0\} : \quad \sup_e \left\{ \left[d^\top \ e^\top \right] Q \begin{bmatrix} d \\ e \end{bmatrix} \mid \begin{array}{l} e_j = 0, \ j \in \mathcal{I}_{\bar{y}}, \ (s_I)_j > 0 \\ e_j \geq 0, \ j \in \mathcal{I}_{\bar{y}}, \ (s_I)_j = 0 \\ A_I d + B_I e = 0 \end{array} \right\} > 0.$$

We are now ready to state and prove our alternative meaningfulness result, in the absence of the strict complementarity slackness.

THEOREM A.2. *Let $\bar{x}$ be a stationary point in the sense of Definition 1.2 such that the corresponding pair $(\bar{x}, \bar{y})$ with $\bar{y} \in S(\bar{x})$ satisfies (ILICQ), (ISOSC'), (OLICQ), and (OSOSC'). Then $\bar{x}$ is a strict local optimal solution for problem (P_O).*

Proof. The assertion follows from the second-order growth condition at $\bar{x}$ for problem (P). The proof will first ensure that the latter follows from the second-order growth condition at $\bar{x}$ for the auxiliary unconstrained optimization problem

$$(A.3) \quad \min_{x \in \mathbb{R}^n} \psi(x) \quad := \quad \max \left\{ \phi(x) - \phi(\bar{x}), \right. \\ \left. -x_1, \dots, -x_n, \pm [(a_O^1)^\top x - b_1], \dots, \pm [(a_O^p)^\top x - b_p] \right\}.$$

Subsequently, the rest of the proof will consists of showing that the requirements for the application of Lemma A.1 to the function ψ (A.3) are satisfied. Note that ψ is an extended real-valued function with ϕ defined by

$$(A.4) \quad \phi(x) := \sup_{y \in Y(x)} f(x, y) \quad \text{for all } x \in \mathbb{R}^n.$$

This obviously means that $\phi(x) = \varphi(x)$ for all $x \in X$ considering assumption (1.5). Also note that for a vector a , we use the notation $\pm a := \{+a, -a\}$.

Now, assuming that the second order growth condition holds for problem (A.3) at the point $\bar{x}$, we can find $\epsilon > 0$ and $\alpha > 0$ such that

$$(A.5) \quad \psi(x) \geq \alpha \|x - \bar{x}\|^2 \quad \text{for all } x \in \mathbb{B}_\epsilon(\bar{x}),$$

considering the fact that $\psi(\bar{x}) = 0$, as $\bar{x} \in X$. Suppose for a moment now that the second order growth condition does not hold at $\bar{x}$ for problem (P); i.e., for the problem

$$(A.6) \quad \min_{x \in X} \varphi(x),$$

where φ is defined in (1.4). This implies that we can find a sequence $\{x^k\}_{k \in \mathbb{N}} \subseteq X$ converging to $\bar{x}$ such that we have

$$\varphi(x^k) < \varphi(\bar{x}) + \frac{1}{k} \|x^k - \bar{x}\|^2 \quad \text{for all } k \in \mathbb{N}.$$

Subsequently, this leads to a contradiction of condition (A.5), given that

$$\psi(x^k) \leq \max \{ \varphi(x^k) - \varphi(\bar{x}), 0 \} < \frac{1}{k} \|x^k - \bar{x}\|^2 \quad \text{for all } k \in \mathbb{N}.$$

In other words, this means that to show that the point $\bar{x}$ is a strict local optimal solution of problem (P) in the sense of (1.6), it will suffice to show that problem (A.3) satisfies the second order growth condition at this point. To prove this, it suffices to show that the properties (i)–(iii) in Lemma A.1 hold.

Next, observe that the function ϕ (A.4) can be written as $\phi(x) := \tilde{\phi}(x) - \hat{\phi}(x)$ for all $\mathbb{R}^n$, where

$$(A.7) \quad \begin{cases} \tilde{\phi}(x) &:= \frac{1}{2} x^\top Q_{11} x + c_x^\top x, \\ \hat{\phi}(x) &:= \inf_{y \in Y(x)} \frac{1}{2} y^\top Q_{22} y - x^\top Q_{12}^\top y - c_y^\top y. \end{cases}$$

Hence, based on [27, Proposition 4], it follows from the fulfillment of (ILICQ) and (ISOSC') that ϕ is continuously differentiable at $\bar{x}$ with

$$(A.8) \quad \nabla \phi(\bar{x}) = Q_{11} \bar{x} + Q_{12} \bar{y} + c_x + A_I^\top \lambda_I,$$

where λ_I is the unique (thanks to the fulfillment of **(ILICQ)**) Lagrange multiplier associated to the inner equality constraint, and moreover, $\hat{\phi}$ and $-\hat{\phi}$ are second order directionally differentiable and epi-regular at $\bar{x}$, and we have

$$\hat{\phi}''(\bar{x}; d, \omega) = (\lambda_I^\top A_I - \bar{y}^\top Q_{12}^\top) \omega - \sup_e \left\{ \zeta^\top \hat{Q} \zeta \mid e \in \mathcal{Z}_d(\bar{x}, \bar{y}, s_I, \lambda_I) \right\},$$

where $\{(s_I, \lambda_I)\} = \Lambda(\bar{x}, \bar{y})$ thanks to **(ILICQ)**, while $\zeta := (d, e) \in \mathbb{R}^n \times \mathbb{R}^m$ and the matrix $\hat{Q}$ and set $\mathcal{Z}_d(\bar{x}, \bar{y}, s_I, \lambda_I)$ are respectively defined by

$$\hat{Q} := \begin{bmatrix} O & Q_{12} \\ Q_{12}^\top & -Q_{22} \end{bmatrix} \quad \text{and} \quad \mathcal{Z}_d(\bar{x}, \bar{y}, s_I, \lambda_I) := \left\{ e \mid \begin{array}{l} e_j = 0, \quad j \in \mathcal{I}_{\bar{y}}, \quad (s_I)_j > 0 \\ e_j \geq 0, \quad j \in \mathcal{I}_{\bar{y}}, \quad (s_I)_j = 0 \\ A_I d + B_I e = 0 \end{array} \right\}.$$

And subsequently, considering the expression of ϕ as decomposed in **(A.7)**,

$$\begin{aligned} \phi''(\bar{x}; d, \omega) &= d^\top Q_{11} d + \bar{z}^\top \bar{Q} \omega + \sup_e \left\{ \zeta^\top \hat{Q} \zeta \mid e \in \mathcal{Z}_d(\bar{x}, \bar{y}, s_I, \lambda_I) \right\} \\ (A.9) \quad &= \bar{z}^\top \bar{Q} \omega + \sup_e \left\{ \zeta^\top Q \zeta \mid e \in \mathcal{Z}_d(\bar{x}, \bar{y}, s_I, \lambda_I) \right\}, \end{aligned}$$

where $\bar{z}^\top$ and $\bar{Q}$ are respectively defined by

$$\bar{z}^\top := [\bar{x}^\top, \bar{y}^\top, \lambda_I^\top, 1] \quad \text{and} \quad \bar{Q} := [Q_{11} \quad Q_{12} \quad -A_I^\top \quad c_x]^\top.$$

To prove the fulfillment of condition (ii) in Lemma **A.1**, we will show that the stationarity of the point $\bar{x}$ ensures

$$(A.10) \quad \phi'(\bar{x}; d) \geq 0 \quad \text{for all} \quad d \in \mathcal{L}_X(\bar{x}),$$

where $\mathcal{L}_X(\bar{x})$ corresponds to the linearized cone of X at the point $\bar{x} \in X$:

$$(A.11) \quad \mathcal{L}_X(\bar{x}) := \{d \in \mathbb{R}^n \mid A_O d = 0, \quad d_i \geq 0, \quad i \in \mathcal{I}_{\bar{x}}\}.$$

To see why the fulfillment of **(A.10)** ensures (ii) for ψ , observe that for all $d \in \mathbb{R}^n$,

$$(A.12) \quad \begin{aligned} \psi'(\bar{x}; d) &= \max \{ \phi'(\bar{x}; d), \max \{ -d_i \mid i \in \mathcal{I}_{\bar{x}} \}, \\ &\quad \pm (a_O^1)^\top d, \dots, \pm (a_O^p)^\top d \}, \end{aligned}$$

while accounting for the fact that since ϕ is continuously differentiable at $\bar{x}$, $\phi'(\bar{x}; d)$ exists and based on **(A.8)**, we have

$$(A.13) \quad \phi'(\bar{x}; d) = (Q_{11} \bar{x} + Q_{12} \bar{y} + c_x + A_I^\top \lambda_I)^\top d.$$

It therefore follows that based on the formula **(A.12)**, if condition **(A.10)** is satisfied, then assumption (ii) in Lemma **A.1** holds.

Now, observe that based on **(A.13)**, we have $\phi'(\bar{x}; 0) = 0$. Additionally, with the linearized cone $\mathcal{L}_X(\bar{x})$ being convex, condition **(A.10)** is equivalent to having 0 as an optimal solution to problem

$$(A.14) \quad \min_d \phi'(\bar{x}; d) \quad \text{s.t.} \quad d \in \mathcal{L}_X(\bar{x}),$$

which is a linear program, considering the expressions in **(A.11)** and **(A.13)**. Hence, the 0 being an optimal solution of problem **(A.14)** is equivalent to the existence of $(\lambda_O, s_O) \in \mathbb{R}^p \times \mathbb{R}^n$ such that

$$(A.15) \quad \begin{aligned} \nabla \phi(\bar{x}) + A_O^\top \lambda_O - (s_O)_{\mathcal{I}_{\bar{x}}} &= 0, \\ (s_O)_i &\geq 0, \quad i \in \mathcal{I}_{\bar{x}}, \end{aligned}$$

while taking into account that $\nabla_d \phi'(\bar{x}; 0) = \nabla \phi(\bar{x})$. Since $\bar{x}$ is a stationary point in the sense of Definition **1.2**, it follows that from the *outer KKT system* component of **(1.7)** and the formula in **(A.8)** that the condition **(A.15)** is satisfied. Therefore **(A.14)** holds, and hence, **(A.10)** is satisfied. So, condition (ii) in Lemma **A.1** holds.

Finally, to establish the fulfillment of assumption (iii) in Lemma A.1, start by observing that based on (A.13), we can easily check that the cone (A.2) can be rewritten as

$$\mathcal{C}_X(\bar{x}) = \{d \in \mathbb{R}^n \mid \psi'(\bar{x}; d) = 0\}.$$

Now, let $d \in \mathcal{C}_X(\bar{x})$. Then considering (A.13) and the formula (A.2) it holds that $\phi'(\bar{x}; d) = 0$. Hence, for all $\omega \in \mathbb{R}^n$,

$$\begin{aligned} \psi''(\bar{x}; d, \omega) &= \max \{ \phi''(\bar{x}; d, \omega), \{-\omega_i, i \in \mathcal{I}_{\bar{x}}(d)\}, \\ &\quad \{\pm(a_O^k)^\top w, j = 1, \dots, q\} \} \end{aligned}$$

with the expression of $\phi''(\bar{x}; d, \omega)$ given in (A.9), while $\mathcal{I}_{\bar{x}}(d)$ is given by

$$\mathcal{I}_{\bar{x}}(d) := \{i \in \mathcal{I}_{\bar{x}} \mid d_i = 0\}.$$

Based on this formula, (OSOSC') ensures that (iii) in Lemma A.1 is satisfied if

$$(A.16) \quad \min_{\omega, \alpha} \left\{ \alpha \left| \begin{array}{l} \phi''(\bar{x}; d, \omega) \leq \alpha \\ -w_i \leq \alpha, \quad i \in \mathcal{I}_{\bar{x}}(d) \\ \pm(a_O^k)^\top w \leq \alpha, \quad j = 1, \dots, q \end{array} \right. \right\} > 0 \text{ for all } d \in \mathcal{C}_X(\bar{x}) \setminus \{0\}$$

with a_O^k denotes the j^{th} row of the matrix A_O .

Based on the formula of $\phi''(\bar{x}; d, \omega)$ given in (A.9), condition (A.16) reduces to

$$(A.17) \quad \min_{\omega, \alpha} \left\{ \alpha \left| \begin{array}{l} \alpha - \bar{z}^\top \bar{Q} \omega \geq \rho_d(\bar{x}, \bar{y}, s_I, \lambda_I) \\ \alpha + e_i^\top \omega \geq 0, \quad i \in \mathcal{I}_{\bar{x}}(d) \\ \alpha - (a_O^k)^\top w \geq 0, \quad j = 1, \dots, q \\ \alpha + (a_O^k)^\top w \geq 0, \quad j = 1, \dots, q \end{array} \right. \right\} > 0,$$

where we have

$$\rho_d(\bar{x}, \bar{y}, s_I, \lambda_I) := \sup_e \{ \zeta^\top Q \zeta \mid e \in \mathcal{Z}_d(\bar{x}, \bar{y}, s_I, \lambda_I) \}.$$

The dual of the linear program in the left-hand-side of equation (A.17) is

$$(A.18) \quad \max_{\kappa, s_O, \lambda_O} \{ \kappa \rho_d(\bar{x}, \bar{y}, s_I, \lambda_I) \mid (\kappa, s_O, \lambda_O) \in \Delta_d(\lambda_I) \}$$

with the set $\Delta_d(\lambda_I)$ defined by

$$(A.19) \quad \Delta_d(\lambda_I) := \left\{ (\kappa, s_O, \lambda_O) \left| \begin{array}{l} \kappa \bar{Q}^\top \bar{z} - \sum_{i \in \mathcal{I}_{\bar{x}}(d)} (s_O)_i e_n^i - \sum_{j=1}^q (\lambda_O)_j a_O^k = 0 \\ \kappa + \sum_{i \in \mathcal{I}_{\bar{x}}(d)} (s_O)_i + \sum_{j=1}^q (\lambda_O)_j = 1 \\ \kappa \geq 0, (s_O)_i \geq 0, i \in \mathcal{I}_{\bar{x}}(d), \lambda_O \in \mathbb{R}^q \end{array} \right. \right\}.$$

Considering the fulfillment of (OLICQ) at $\bar{x}$, we must have $\kappa \neq 0$ and hence, we get the result from the combination of condition (A.17) and the dual of the left-hand-side linear program, which is provided in (A.18)–(A.19). $\square$

Appendix B. Proof of Theorem 3.4. Consider $(v, \mu) \in B(v^*, \delta) \times \Lambda_{\bar{\mu}} \subset V_\delta$ and define the function

$$\Theta_\mu(v) := v - \partial_v H(v^*, 0)^{-1} H(v, \mu).$$

Clearly, we have $\Theta_\mu(v) = v$ if and only if $H(v, \mu) = 0$. Moreover,

$$(B.1) \quad \|\Theta_\mu v^* - v^*\| = \|\Theta_\mu v^* - \Theta_0 v^*\| \leq M \|H(v^*, \mu) - H(v^*, 0)\| \leq M B_\delta \bar{\mu}.$$

We want to use the fixed point theorem to prove that for all $\mu \in \Lambda_{\bar{\mu}}$, there exists $v \in \overline{B(v^*, \delta)}$ such that $\Theta_\mu(v) = v$. To proceed, we need to show the following:

1. $\Theta_\mu(B(v^*, \delta)) \subset B(v^*, \delta)$;
2. $\exists \sigma \in (0, 1)$: $\|\Theta_\mu(v_1) - \Theta_\mu(v_2)\| \leq \sigma \|v_1 - v_2\|$ for all $v_1, v_2 \in \overline{B(v^*, \delta)}$.

For the first assertion, observe that for $v \in \overline{B(v^*, \delta)}$, we have

$$\begin{aligned} \|\Theta_\mu(v) - v^*\| &\leq \|\Theta_\mu(v) - \Theta_\mu(v^*)\| + \|\Theta_\mu(v^*) - v^*\| \\ &\leq \sup_{(v, \mu) \in V_\delta} \|\partial_v \Theta_\mu(v)\| \|v - v^*\| + MB_\delta \bar{\mu} \\ &\leq \frac{1}{2} \|v - v^*\| + \frac{1}{2} \delta \\ &\leq \delta, \end{aligned}$$

where the last but one inequality follows by the definition of V_δ , by (B.1) and the definition of $\bar{\mu}$. To prove 2., observe that $\sigma = 1/2$ since

$$\|\Theta_\mu(v_1) - \Theta_\mu(v_2)\| \leq \frac{1}{2} \|v_1 - v_2\|,$$

using again the definition of V_δ . Hence, the function $g : \Lambda_{\bar{\mu}} \rightarrow \overline{B(v^*, \delta)} \subset \mathbb{R}^N$ defined by $g(\mu) := v_\mu$ such that $\Theta_\mu(v_\mu) = v_\mu$ and $H(g(\mu), \mu) = 0$ is well defined. To prove regularity, observe that for $\mu_1, \mu_2 \in \Lambda_{\bar{\mu}}$, we have

$$\begin{aligned} \|g(\mu_1) - g(\mu_2)\| &= \|\Theta_{\mu_1}(v_{\mu_1}) - \Theta_{\mu_2}(v_{\mu_2})\| \\ &= \|\Theta_{\mu_1}(v_{\mu_1}) - \Theta_{\mu_1}(v_{\mu_2}) + \Theta_{\mu_1}(v_{\mu_2}) - \Theta_{\mu_2}(v_{\mu_2})\| \\ &\leq \frac{1}{2} \|v_{\mu_1} - v_{\mu_2}\| + M \|H(v_{\mu_2}, \mu_1) - H(v_{\mu_2}, \mu_2)\| \\ &\leq \frac{1}{2} \|g(\mu_1) - g(\mu_2)\| + MB_\delta |\mu_1 - \mu_2| \end{aligned}$$

from which Lipschitz continuity of g on $\Lambda_{\bar{\mu}}$ follows. By the differentiability of $\hat{F}$ and the Lipschitz-continuity of g , given $\mu \in \Lambda_{\bar{\mu}}$ and h sufficiently small, we have that

$$\begin{aligned} &\|H(g(\mu + h), \mu + h) - H(g(\mu), \mu) \\ &\quad + \partial_v H(g(\mu), \mu) [g(\mu + h) - g(\mu)] + \partial_\mu H(g(\mu), \mu) h\| = o(|h|), \end{aligned}$$

which implies that

$$\|\partial_v H(g(\mu), \mu) [g(\mu + h) - g(\mu)] + \partial_\mu H(g(\mu), \mu) h\| = o(|h|)$$

given that $H(g(\mu + h), \mu + h) = H(g(\mu), \mu) = 0$. Hence,

$$\lim_{|h| \rightarrow 0} \|g(\mu + h) - g(\mu) + \partial_v H(g(\mu), \mu)^{-1} \partial_\mu H(g(\mu), \mu) h\| = 0,$$

and the continuity of $\frac{d}{d\mu} g(\lambda)$ follows from the given expression of the derivative.

Appendix C. Proof of Theorem 3.9. The proof of Theorem 3.9 follows closely the one from [40, Theorem 6.2] and using Lemma 3.8. For the sake of completeness, in the remainder of this section, we will outline the main steps. Before, observe that defining $\nu_k := \prod_{j=0}^{k-1} (1 - \alpha_j)$ with $\nu_0 = 1$, and using the linearity of the first two equations in (3.2), we have that

$$(r_b^k, r_c^k) = \nu_k (r_b^0, r_c^0).$$

Coming back to the Polynomial Convergence proof, let us start observing that also in the case here considered the following estimations for $\|(z^k, s^k)\|_1$ are true:

(E1) There exists a constant C_1 such that

$$\nu_k \|(z^k, s^k)\|_1 \leq C_1 \mu^k \text{ for all } k \geq 0.$$

For the proof, our Lemma 3.8 can be applied in the context of the proof of [40, Lemma 6.3], see in particular [40, Eq. 6.19];

(E2) If the initial point is chosen as in Theorem 3.9, then

$$\zeta \nu_k \|(z^k, s^k)\|_1 \leq 4\beta(n + m) \mu^k \text{ for all } k \geq 0.$$

The proof follows using (E1) and an analogous proof of [40, Lemma 6.4].

Moreover, also in the case here considered, defining $D^k := (Z^k)^{1/2}(S^k)^{-1/2}$, the following estimations about the norm of the Newton directions hold:

(N1) There exists a constant C_2 such that

$$\|(D^k)^{-1}\Delta z^k\| \leq C_2(\mu^k)^{1/2}, \quad \|D^k\Delta s^k\| \leq C_2(\mu^k)^{1/2} \text{ for all } k \geq 0.$$

For the proof use our Lemma 3.8 in the proof of [40, Lemma 6.5]; in particular in [40, Eq. (6.30)] the sign “=” becomes “ $\leq$ ”;

(N2) If the initial point is chosen as in Theorem 3.9, then there exists a constant ω independent from $m+n$ such that

$$\|(D^k)^{-1}\Delta z^k\| \leq \omega(m+n)(\mu^k)^{1/2}, \quad \|D^k\Delta s^k\| \leq \omega(m+n)(\mu^k)^{1/2} \text{ for all } k \geq 0.$$

The proof follows using (N1) and an analogous proof of [40, Lemma 6.6].

Finally, the proof of Theorem 3.9 follows by observing that using (E1), (E2), (N1), and (N2), the statement of [40, Lemma 6.7] holds also in our case.

Appendix D. Proofs from Subsection 3.4.

D.1 Proof of Lemma 3.12 (a) Using the triangle inequality and (3.11), it holds that

$$\begin{aligned} \|v_\alpha - g(\mu^+)\| &= \|(1-\alpha)(v - g(\mu^+)) + \alpha(v^+ - g(\mu^+))\|, \\ &\leq (1-\alpha)\|v - g(\mu^+)\| + \alpha\|v^+ - g(\mu^+)\|, \\ &\leq (1-\alpha)\|v - g(\mu^+)\| + \alpha\frac{KL}{2}\|v - g(\mu^+)\|^2. \end{aligned}$$

(b) Similarly, using (3.18), we have

$$\begin{aligned} \|v_\alpha - v^*\| &= \|(1-\alpha)(v - v^*) + \alpha(v^+ - v^*)\|, \\ &\leq (1-\alpha)\|v - v^*\| + \alpha\|v^+ - v^*\|, \\ &\leq (1-\alpha)\|v - v^*\| + \alpha\left(\frac{1}{2}\|v - v^*\| + \frac{\mu^+}{\mu^*}\frac{\delta^*}{2}\right), \\ &= \left(1 - \frac{\alpha}{2}\right)\|v - v^*\| + \alpha\frac{\mu^+}{\mu^*}\frac{\delta^*}{2}. \end{aligned}$$

The fact that $v_\alpha \in B(v^*, \delta^*)$ follows from the fact that the right-hand side of (3.19) is maximized at $\alpha = 1$ (the full step), which by Lemma 3.11 yields a value $< \delta^*$.

D.2 Proof of Corollary 3.14 As a consequence of Theorem 3.13, we can suppose the existence of a sequence of iterates $\{(z^k, \lambda^k, s^k)\}_{k \in \mathbb{N}}$ produced by Algorithm 3.1 such that

$$(z^k, \lambda^k, s^k) \in \mathcal{N}(\bar{\gamma}, \underline{\gamma}, \gamma_p, \gamma_d).$$

Since by construction $(z^k)^\top s^k \leq (z^0)^\top s^0$, we have from (3.15) that

$$\begin{aligned} \|Az^k - b\| &\leq (z^0)^\top s^0 / \gamma_p, \\ \|\hat{Q}z^k + \hat{A}^\top \lambda^k - s^k + \hat{c}\| &\leq (z^0)^\top s^0 / \gamma_d. \end{aligned}$$

Moreover, we have

$$\begin{aligned} \|S^k Z^k \mathbf{e} - \sigma \mu^k \mathbf{e}\| &\leq \|S^k Z^k \mathbf{e}\| + \sigma \mu^k \|\mathbf{e}\|, \\ &\leq \frac{\bar{\gamma}}{\sqrt{n+m}}(z^k)^\top s^k + \frac{\sigma}{\sqrt{n+m}}(z^k)^\top s^k, \\ &\leq \frac{\bar{\gamma} + \sigma}{\sqrt{n+m}}(z^0)^\top s^0. \end{aligned}$$

Appendix E. Min-Cost Flow Adversarial Attack. For the sake of clarity, we will briefly present how the formulation (4.1) is obtained and the related notational details. To this aim, consider a directed graph $G = (V, E)$, where V is the set of vertices and E is the set of edges. Define, moreover, the incidence matrix P of G as the matrix $P \in \mathbb{R}^{|V| \times |E|}$ with entries defined as

$$(E.1) \quad P_{ve} = \begin{cases} +1 & \text{if } e = (w, v) \text{ for some } w \in V, \\ -1 & \text{if } e = (v, w) \text{ for some } w \in V, \\ 0 & \text{otherwise.} \end{cases}$$

Given $b \in \mathbb{R}^{|V|}$ such that $\sum_{v \in V} b_v = 0$, where for $v \in V$, $b_v > 0$ is an exogenous source whereas $b_v < 0$ is exogenous demand, and a capacity vector $u \in \mathbb{R}^{|E|}$, the min-cost flow problem on G can be formulated as

$$\begin{aligned} \rho_{cf} &:= \min \sum_{e \in E} c_e(y_e) y_e \\ \text{s.t. } &Py = b, \\ &0 \leq y \leq p, \end{aligned}$$

where $y := \{y_e\}_{e \in E} \in \mathbb{R}^{|E|}$ and $c_e(y_e)$ denote the cost of moving one unit of flow across edge e . In the following, we fix two nodes, source and sink, s, t , and a sink demand $b_t = r_t \Rightarrow b_s = -r_t$. Let us now consider an extension to such a problem, where, in addition to the regular network user, *the follower*, there is an adversary, *the leader*, who will inject flows to the network in order to force the regular user to use more expensive paths. Let $x := \{x_e\}_{e \in E}$ denote flows controlled by the adversary, with a total attack-budget $a_b > 0$. Then the problem of the adversary can be formulated as

$$\begin{aligned} \max_{\substack{0 \leq x \leq p \\ e^\top x = a_b}} \quad & \min_{0 \leq y \leq p} \sum_{e \in E} c_e(y_e, x_e) y_e \\ & x + y \leq p, \\ & \sum_{e \in E: e=(v,t)} y_e = r_t, \\ & P^{s,t} y = b, \end{aligned}$$

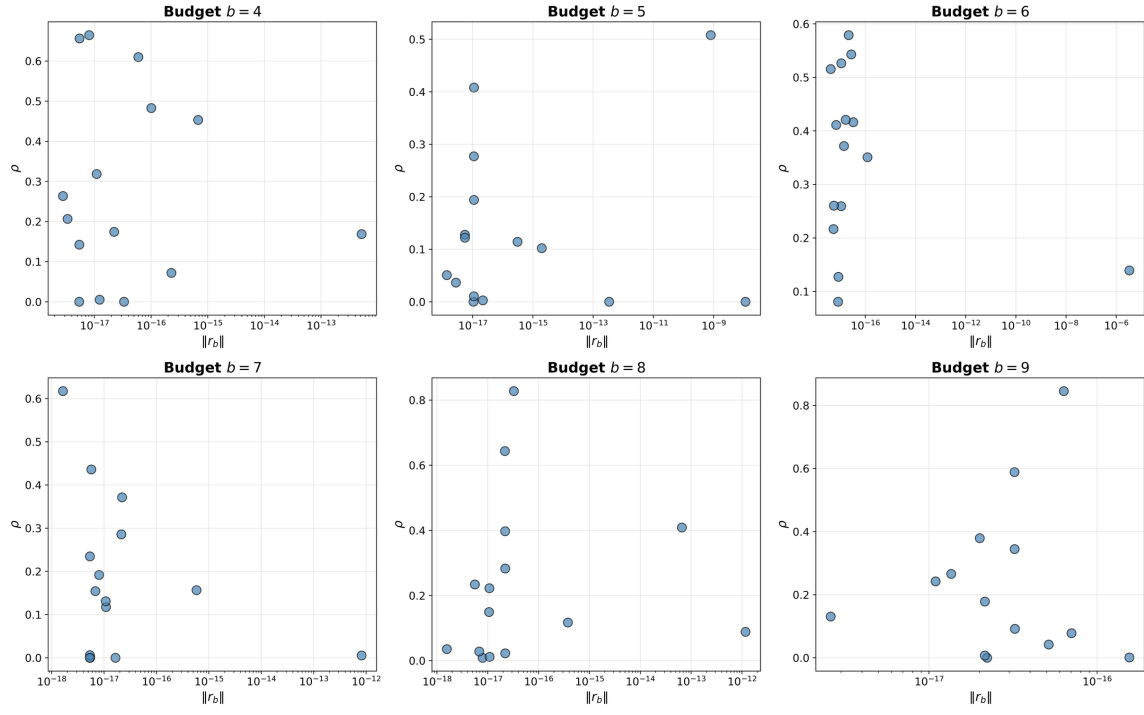
where, for $e \in E$, $c_e(y_e, x_e)$ is related to the cost of routing both x_e and y_e onto the network and $P^{s,t}$ is the incidence matrix defined in (E.1), but having had removed the rows s, t . In the following, we will consider $c_e(y_e, x_e) = w_e(y_e + x_e)$, i.e., the cost of routing the flow is a function of the amount of flow there routed. It is important to note that, using the particular form of c_e and eliminating redundant constraints, the above problem can be written as

$$\begin{aligned} \max_{\substack{(x,z) \geq 0 \\ e^\top x = a_b}} \quad & \min_{y \geq 0} \sum_{e \in E} w^\top y + y^\top \text{diag}(w)y + y^\top \text{diag}(w)x - \frac{\eta}{2} \|x\|^2 - \frac{\eta}{2} \|z\|^2 \\ & x + y + z = p, \\ & \sum_{e \in E: e=(v,t)} y_e = r_t, \\ & P^{s,t} y = 0, \end{aligned}$$

where we note that the variable z belongs to the attacker, and we added the term $\frac{\eta}{2} \|x\|^2 + \frac{\eta}{2} \|z\|^2$ to ensure strong convexity for the outer problem as done in [36, 44, 19].

E.1 Correlation figures. Figures 7, 8, and 9 show the correlations among $\|r_b\|$ and ρ .

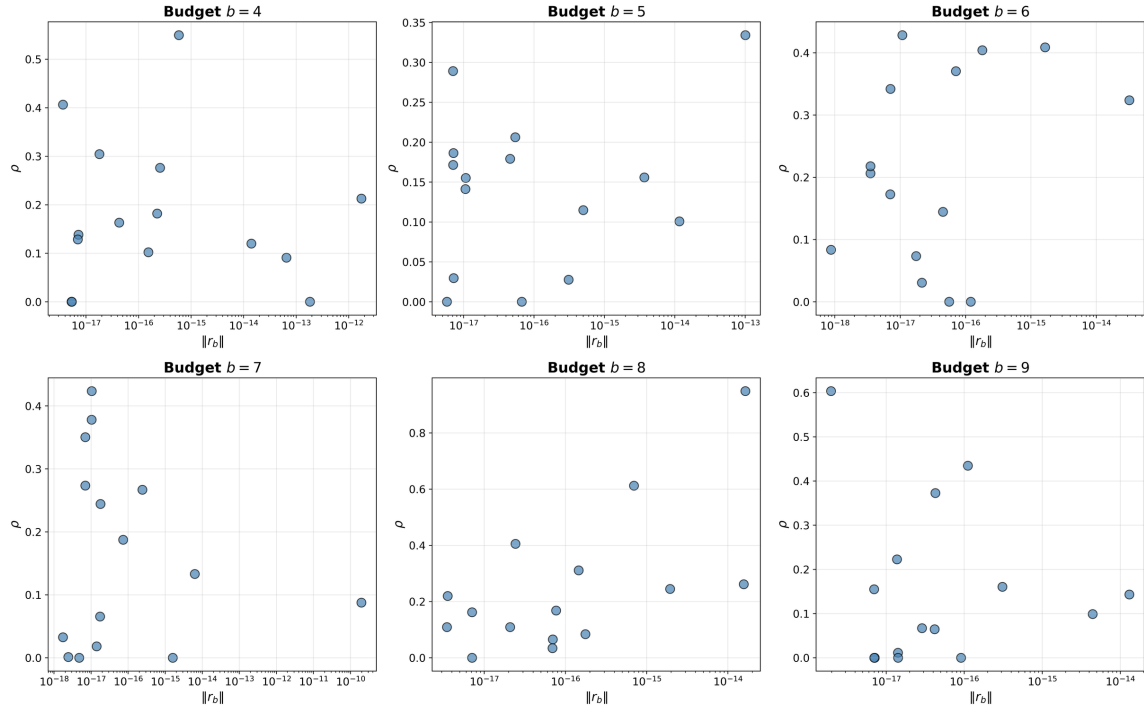
IPM Residual vs Attack Effectiveness by Budget
($n=15, p=0.50, d=0.20$)



(a) $n=15, p=0.50, d=0.20$

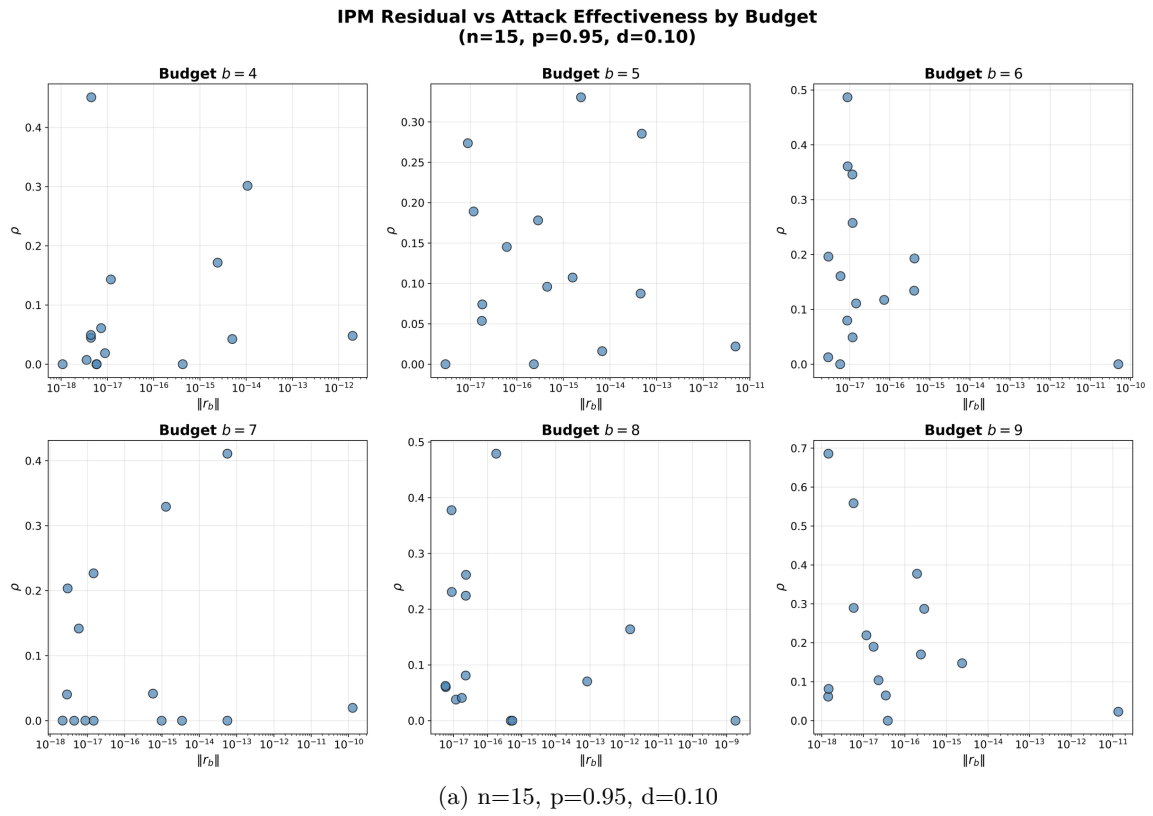
Fig. 7: Correlation among $\|r_b\|$ and ρ .

IPM Residual vs Attack Effectiveness by Budget
($n=15, p=0.75, d=0.10$)



(a) $n=15, p=0.75, d=0.10$

Fig. 8: Correlation among $\|r_b\|$ and ρ .

Fig. 9: Correlation among $\|r_b\|$ and ρ .