An Interior Point-Proximal Method of Multipliers for Linear Positive Semi-Definite Programming

Spyridon Pougkakiotis[1] (iD) · **Jacek Gondzio**[1]

Abstract

In this paper we generalize the Interior Point-Proximal Method of Multipliers (IP-PMM) presented in Pougkakiotis and Gondzio (Comput Optim Appl 78:307–351, 2021. https://doi.org/10.1007/s10589-020-00240-9) for the solution of linear positive Semi-Definite Programming (SDP) problems, allowing inexactness in the solution of the associated Newton systems. In particular, we combine an infeasible Interior Point Method (IPM) with the Proximal Method of Multipliers (PMM) and interpret the algorithm (IP-PMM) as a primal-dual regularized IPM, suitable for solving SDP problems. We apply some iterations of an IPM to each sub-problem of the PMM until a satisfactory solution is found. We then update the PMM parameters, form a new IPM neighbourhood, and repeat this process. Given this framework, we prove polynomial complexity of the algorithm, under mild assumptions, and without requiring exact computations for the Newton directions. We furthermore provide a necessary condition for lack of strong duality, which can be used as a basis for constructing detection mechanisms for identifying pathological cases within IP-PMM.

1 Introduction

Positive Semidefinite Programming (SDP) problems have attracted a lot of attention in the literature for more than two decades, and have been used to model a plethora of

different problems arising from control theory [3, Chapter 14], power systems [13], stochastic optimization [6], truss optimization [28], and many other application areas (e.g. see [3,27]). More recently, SDP has been extensively used for building tight convex relaxations of NP-hard combinatorial optimization problems (see [3, Chapter 12], and the references therein).

As a result of the seemingly unlimited applicability of SDP, numerous contributions have been made to optimization techniques suitable for solving such problems. The most remarkable milestone was achieved by Nesterov and Nemirovskii [17], who designed a polynomially convergent Interior Point Method (IPM) for the class of SDP problems. This led to the development of numerous successful IPM variants for SDP; some of theoretical (e.g. [15,29,31]) and others of practical nature (e.g. [4,5,16]). While IPMs enjoy fast convergence, in theory and in practice, each IPM iteration requires the solution of a very large-scale linear system, even for small-scale SDP problems. What is worse, such linear systems are inherently ill-conditioned. A viable and successful alternative to IPMs for SDP problems (e.g. see [30]), which circumvents the issue of ill-conditioning without significantly compromising convergence speed, is based on the so-called Augmented Lagrangian method (ALM), which can be seen as the dual application of the proximal point method (as shown in [20]). The issue with ALMs is that, unlike IPMs, a consistent strategy for tuning the algorithm parameters is not known. Furthermore, polynomial complexity is lost, and is replaced with merely a finite termination. An IPM scheme combined with the Proximal Method of Multipliers (PMM) for solving SDP problems was proposed in [9], and was interpreted as a primal-dual regularized IPM. The authors established global convergence, and numerically demonstrated the efficiency of the method. However, the latter is not guaranteed to converge to an ϵ-optimal solution in a polynomial number of iterations, or even to find a global optimum in a finite number of steps. Finally, viable alternatives based on proximal splitting algorithms have been studied in [12,25]. Such schemes are very efficient and require significantly less computations and memory per iteration, as compared to IPM or ALM. However, as they are first-order methods, their convergence to high accuracy might be slow. Hence, such methodologies are only suitable for finding approximate solutions with low-accuracy.

In this paper, we are extending the Interior Point-Proximal Method of Multipliers (IP-PMM) presented in [19]. In particular, the algorithm in [19] was developed for convex quadratic programming problems and assumed that the resulting linear systems are solved exactly. Under this framework, it was proved that IP-PMM converges in a polynomial number of iterations, under mild assumptions, and an infeasibility detection mechanism was established. An important feature of this method is that it provides a reliable tuning for the penalty parameters of the PMM; indeed, the reliability of the algorithm is established numerically in a wide variety of convex problems in [7,8,10,19]. In particular, the IP-PMMs proposed in [7,8,10] use preconditioned iterative methods for the solution of the resulting linear systems, and are very robust despite the use of inexact Newton directions. In what follows, we develop and analyze an IP-PMM for linear SDP problems, which furthermore allows for inexactness in the solution of the linear systems that have to be solved at every iteration. We show that the method converges polynomially under standard assumptions. Subsequently, we provide a necessary condition for lack of strong duality, which can serve as a

basis for constructing implementable detection mechanisms for pathological cases (following the developments in [19]). As is verified in [19], IP-PMM is competitive with standard non-regularized IPM schemes, and is significantly more robust. This is because the introduction of regularization prevents severe ill-conditioning and rank deficiency of the associated linear systems solved within standard IPMs, which can hinder their convergence and numerical stability. For detailed discussions on the effectiveness of regularization within IPMs, the reader is referred to [1,2,18], and the references therein. A particularly important benefit of using regularization, is that the resulting Newton systems can be preconditioned effectively (e.g. see the developments in [7,10]), allowing for more efficient implementations, with significantly lowered memory requirements. We note that the paper is focused on the theoretical aspects of the method, and an efficient, scalable, and reliable implementation would require a separate study. Nevertheless, the practical effectiveness of IP-PMM (both in terms of efficiency, scalability, and robustness) has already been demonstrated for linear, convex quadratic [7,10,19], and nonlinear convex problems [8].

The rest of the paper is organized as follows. In Sect. 2, we provide some preliminary background and introduce our notation. Then, in Sect. 3, we provide the algorithmic framework of the method. In Sect. 4, we prove polynomial complexity of the algorithm, and establish its global convergence. In Sect. 5, a necessary condition for lack of strong duality is derived, and we discuss how it can be used to construct an implementable detection mechanism for pathological cases. Finally, we derive some conclusions in Sect. 6.

2 Preliminaries and Notation

2.1 Primal-Dual Pair of SDP Problems

Let the vector space $\mathcal{S}^n := \{B \in \mathbb{R}^{n \times n} \colon B = B^\top\}$ be given, endowed with the inner product $\langle A, B \rangle = \text{Tr}(AB)$, where $\text{Tr}(\cdot)$ denotes the trace of a matrix. In this paper, we consider the following primal-dual pair of linear positive semi-definite programming problems, in the standard form:

$$\min_{X \in \mathcal{S}^n} \langle C, X \rangle, \quad \text{s.t. } \mathcal{A}X = b, \ X \in \mathcal{S}_+^n, \tag{P}$$

$$\max_{y \in \mathbb{R}^m, \ Z \in \mathcal{S}^n} b^\top y, \quad \text{s.t. } \mathcal{A}^* y + Z = C, \ Z \in \mathcal{S}_+^n, \tag{D}$$

where $\mathcal{S}_+^n := \{B \in \mathcal{S}^n \colon B \succeq 0\}$, $C, X, Z \in \mathcal{S}^n$, $b, y \in \mathbb{R}^m$, $\mathcal{A}$ is a linear operator on $\mathcal{S}^n$, $\mathcal{A}^*$ is the adjoint of $\mathcal{A}$, and $X \succeq 0$ denotes that X is positive semi-definite. We note that the norm induced by the inner product $\langle A, B \rangle = \text{Tr}(AB)$ is in fact the *Frobenius norm*, denoted by $\| \cdot \|_F$. Furthermore, the adjoint $\mathcal{A}^* \colon \mathbb{R}^m \mapsto \mathcal{S}^n$ is such that $y^\top \mathcal{A}X = \langle \mathcal{A}^* y, X \rangle$, $\forall \, y \in \mathbb{R}^m$, $\forall \, X \in \mathcal{S}^n$.

For the rest of this paper, except for Sect. 5, we will assume that the linear operator $\mathcal{A}$ is onto and that problems (P) and (D) are both strictly feasible (that is, Slater's constraint qualification holds for both problems). It is well-known that under the

previous assumptions, the primal-dual pair (P)–(D) is guaranteed to have optimal solution for which strong duality holds (see [17]). Such a solution can be found by solving the Karush–Kuhn–Tucker (KKT) optimality conditions for (P)–(D), which read as follows:

$$
\begin{bmatrix} \mathcal{A}^*y + Z - C \\ \mathcal{A}X - b \\ XZ \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad X, \ Z \in \mathcal{S}_+^n. \tag{1}
$$

2.2 A Proximal Method of Multipliers

The author in [20] presented for the first time the *Proximal Method of Multipliers* (PMM), in order to solve general convex programming problems. Let us derive this method for the pair (P)–(D). Given arbitrary starting point $(X_0, y_0) \in \mathcal{S}_+^n \times \mathbb{R}^m$, the PMM can be summarized by the following iteration:

$$
X_{k+1} = \underset{X \in \mathcal{S}_+^n}{\arg\min} \left\{ \langle C, X \rangle - y_k^\top (\mathcal{A}X - b) + \frac{\mu_k}{2} \|X - X_k\|_F^2 + \frac{1}{2\mu_k} \|\mathcal{A}X - b\|_2^2 \right\},
$$
$$
y_{k+1} = y_k - \frac{1}{\mu_k} (\mathcal{A}X_{k+1} - b), \tag{2}
$$

where $\{\mu_k\}$ is a sequence of positive penalty parameters. The previous iteration admits a unique solution, for all k.

We can write (2) equivalently by making use of the *maximal monotone* operator $T_{\mathcal{L}} \colon \mathbb{R}^m \times \mathcal{S}^n \rightrightarrows \mathbb{R}^m \times \mathcal{S}^n$ (see [20,21]), whose graph is defined as:

$$
T_{\mathcal{L}}(X, y) := \{ (V, u) : V \in C - \mathcal{A}^*y + \partial \delta_{\mathcal{S}_+^n}(X), \ u = \mathcal{A}X - b \}, \tag{3}
$$

where $\delta_{S_+^n}(\cdot)$ is an indicator function defined as:

$$
\delta_{S_+^n}(X) := \begin{cases} 0, & \text{if } X \in \mathcal{S}_+^n, \\ \infty, & \text{otherwise,} \end{cases} \tag{4}
$$

and $\partial(\cdot)$ denotes the sub-differential of a function, hence (from [22, Corollary 23.5.4]):

$$
Z \in \partial \delta_{S_+^n}(X) \Leftrightarrow -Z \in \mathcal{S}_+^n, \ \langle X, Z \rangle = 0.
$$

By convention, we have that $\partial \delta_{\mathcal{S}_+^n(X^*)} = \emptyset$ if $X^* \notin \mathcal{S}_+^n$. Given a bounded pair (X^*, y^*) such that $(0, 0) \in T_{\mathcal{L}}(X^*, y^*)$, we can retrieve a matrix $Z^* \in \partial \delta_{S_+^n}(X^*)$, using which $(X^*, y^*, -Z^*)$ is an optimal solution for (P)–(D). By defining the *proximal operator*:

$$
\mathcal{P}_k := \left(I_{n+m} + \frac{1}{\mu_k} T_{\mathcal{L}} \right)^{-1}, \tag{5}
$$

where I_{n+m} is the identity operator of size $n + m$, and describes the direct sum of the idenity operators of $\mathcal{S}_n$ and $\mathbb{R}^m$, we can express (2) as:

$$(X_{k+1}, y_{k+1}) = \mathcal{P}_k(X_k, y_k), \tag{6}$$

and it can be shown that $\mathcal{P}_k$ is single valued and firmly *non-expansive* (see [21]).

2.3 An Infeasible Interior Point Method

In what follows we present a basic infeasible IPM suitable for solving the primal-dual pair (P)–(D). Such methods handle the conic constraints by introducing a suitable logarithmic barrier in the objective (for an extensive study of logarithmic barriers, the reader is referred to [17]). At each iteration, we choose a *barrier parameter* $\mu > 0$ and form the logarithmic *barrier primal-dual pair:*

$$\min_{X \in \mathcal{S}^n} \langle C, X \rangle - \mu \ln(\det(X)), \quad \text{s.t. } \mathcal{A}X = b, \tag{7}$$

$$\max_{y \in \mathbb{R}^m, \, Z \in \mathcal{S}^n} b^\top y + \mu \ln(\det(Z)), \quad \text{s.t. } \mathcal{A}^* y + Z = C. \tag{8}$$

The first-order (barrier) optimality conditions of (7)–(8) read as follows:

$$\begin{bmatrix} \mathcal{A}^* y + Z - C \\ \mathcal{A}X - b \\ XZ - \mu I_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad X, \, Z \in \mathcal{S}_{++}^n, \tag{9}$$

where $\mathcal{S}_{++}^n := \{B \in \mathcal{S}^n : B \succ 0\}$. For every chosen value of μ, we want to approximately solve the following nonlinear system of equations:

$$F_{\sigma,\mu}^{IPM}(w) := \begin{bmatrix} \mathcal{A}^* y + Z - C \\ \mathcal{A}X - b \\ XZ - \sigma \mu I_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

where, with a slight abuse of notation, we set $w = (X, y, Z)$. Notice that $F_{\sigma,\mu}^{IPM}(w) = 0$ is a perturbed form of the barrier optimality conditions. In particular, $\sigma \in (0, 1)$ is a *centering parameter* which determines how fast μ will be forced to decrease at the next IPM iteration. For $\sigma = 1$ we recover the barrier optimality conditions in (9), while for $\sigma = 0$ we recover the optimality conditions in (1).

In IPM literature it is common to apply Newton method to approximately solve the system of nonlinear equations $F_{\sigma,\mu}^{IPM}(w) = 0$. Newton method is favored for systems of this form due to the *self-concordance* of the logarithmic barrier (see [17]). However, a well-known issue in the literature is that the matrix XZ is not necessarily symmetric. A common approach to tackle this issue is to employ a symmetrization operator $H_P : \mathbb{R}^{n \times n} \mapsto \mathcal{S}^n$, such that $H_P(XZ) = \mu I$ if and only if $XZ = \mu I$,

given that $X, Z \in \mathcal{S}_+^n$. Following Zhang ([29]), we employ the following operator: $H_P : \mathbb{R}^{n \times n} \mapsto \mathcal{S}^n$:

$$H_P(B) := \frac{1}{2}(PBP^{-1} + (PBP^{-1})^\top), \tag{10}$$

where P is a non-singular matrix. It can be shown that the central path (a key notion used in IPMs—see [17]) can be equivalently defined as $H_P(XZ) = \mu I$, for any non-singular matrix P. In this paper, we will make use of the choice $P_k = Z_k^{-\frac{1}{2}}$. For a plethora of alternative choices, the reader is referred to [26]. We should note that the analysis in this paper can be tailored to different symmetrization strategies, and this choice is made for simplicity of exposition.

At the beginning of the k-th iteration, we have $w_k = (X_k, y_k, Z_k)$ and μ_k available. The latter is defined as $\mu_k = \frac{\langle X_k, Z_k \rangle}{n}$. By substituting the symmetrized complementarity in the last block equation and applying Newton method, we obtain the following system of equations:

$$\begin{bmatrix} 0 & \mathcal{A}^* & I_n \\ \mathcal{A} & 0 & 0 \\ \mathcal{E}_k & 0 & \mathcal{F}_k \end{bmatrix} \begin{bmatrix} \Delta X \\ \Delta y \\ \Delta Z \end{bmatrix} = \begin{bmatrix} C - \mathcal{A}^* y - Z_k \\ b - \mathcal{A} X_k \\ \mu I_n - H_{P_k}(X_k Z_k) \end{bmatrix}, \tag{11}$$

where $\mathcal{E}_k := \nabla_X H_{P_k}(X_k Z_k)$, and $\mathcal{F}_k := \nabla_Z H_{P_k}(X_k Z_k)$.

2.4 Vectorized Format

In what follows we vectorize the associated operators, in order to work with matrices. In particular, given any matrix $B \in \mathbb{R}^{m \times n}$, we denote its vectorized form as $\boldsymbol{B}$, which is a vector of size mn, obtained by stacking the columns of B, from the first to the last. For the rest of this manuscript, any boldface letter denotes a vectorized matrix. Furthermore, if $\mathcal{A} : \mathcal{S}^n \mapsto \mathbb{R}^m$ is a linear operator, we can define it component-wise as $(\mathcal{A}X)_i := \langle A_i, X \rangle$, for $i = 1, \ldots, m$, and any $X \in \mathcal{S}^n$, where $A_i \in \mathcal{S}^n$. Furthermore, the adjoint of this operator, that is $\mathcal{A}^* : \mathbb{R}^m \mapsto \mathcal{S}^n$ is defined as $\mathcal{A}^* y := \sum_{i=1}^m y_i A_i$, for all $y \in \mathbb{R}^m$. Using this notation, we can equivalently write (P)–(D) in the following form:

$$\min_{X \in \mathcal{S}^n} \langle C, X \rangle, \quad \text{s.t. } \langle A_i, X \rangle = b_i, \quad i = 1, \ldots, m, \quad X \in \mathcal{S}_+^n, \tag{12}$$

$$\max_{y \in \mathbb{R}^m, \, Z \in \mathcal{S}^n} b^\top y, \quad \text{s.t. } \sum_{i=1}^m y_i A_i + Z = C, \quad Z \in \mathcal{S}_+^n. \tag{13}$$

The first-order optimality conditions can be re-written as:

$$\begin{bmatrix} A^\top y + \boldsymbol{Z} - \boldsymbol{C} \\ A\boldsymbol{X} - b \\ XZ \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad X, \, Z \in \mathcal{S}_+^n,$$

where $A^\top = [A_1 \; A_2 \; \cdots \; A_m]$.

Below we summarize any additional notation that is used later in the paper. An iteration of the algorithm is denoted by $k \in \mathbb{N}$. Given an arbitrary matrix A (resp., vector x), A_k (resp., x_k) denotes that the matrix (resp., vector) depends on the iteration k. An optimal solution to the pair (P)–(D) will be denoted as (X^*, y^*, Z^*). Optimal solutions of different primal-dual pairs will be denoted using an appropriate subscript, in order to distinguish them (e.g. (X_r^*, y_r^*, Z_r^*) will denote an optimal solution for a PMM sub-problem). Any norm (resp., semi-norm) is denoted by $\|\cdot\|_\chi$, where χ is used to distinguish between different norms (e.g. $\|\cdot\|_2$ denotes the Euclidean norm). Given two matrices $X, \, Y \in \mathcal{S}_+^n$, we write $X \succeq Y$ when X is larger than Y with respect to the Loewner ordering. Given two logical statements T_1, T_2, the condition $T_1 \wedge T_2$ is true only when both T_1 and T_2 are true. Given two real-valued positive increasing functions $T(\cdot)$ and $f(\cdot)$, we say that $T(x) = O(f(x))$ (resp., $T(x) = \Omega(f(x))$) if there exist $x_0 \geq 0$, $c_1 > 0$, such that $T(x) \leq c_1 f(x)$ (resp., $c_2 > 0$ such that $T(x) \geq c_2 f(x)$), for all $x \geq x_0$. We write $T(x) = \Theta(f(x))$ if and only if $T(x) = O(f(x))$ and $T(x) = \Omega(f(x))$. Finally, let an arbitrary matrix A be given. The maximum (resp., minimum) singular value of A is denoted by $\eta_{\max}(A)$ (resp., $\eta_{\min}(A)$). Similarly, the maximum (resp., minimum) eigenvalue of a square matrix A is denoted by $\nu_{\max}(A)$ (resp., $\nu_{\min}(A)$).

3 An Interior Point-Proximal Method of Multipliers for SDP

In this section we present an inexact extension of IP-PMM presented in [19], suitable for solving problems of the form of (P)–(D). Assume that we have available an estimate λ_k for a Lagrange multiplier vector at iteration k. Similarly, denote by $\Xi_k \in \mathcal{S}_+^n$ an estimate of a primal solution. As we discuss later, these estimate sequences (i.e. $\{\lambda_k\}$, $\{\Xi_k\}$) are produced by the algorithm, and represent the dual and primal proximal estimates, respectively. During the k-th iteration of the PMM, applied to (P), the following proximal penalty function has to be minimized:

$$\mathcal{L}_{\mu_k}^{PMM}(X; \Xi_k, \lambda_k) := \langle C, X \rangle - \lambda_k^\top (\mathcal{A}X - b) + \frac{1}{2\mu_k}\|\mathcal{A}X - b\|_2^2 + \frac{\mu_k}{2}\|X - \Xi_k\|_F^2, \quad (14)$$

with $\{\mu_k\}$ being some non-increasing sequence of positive penalty parameters. Notice that this is equivalent to the iteration (2). We approximately minimize (14) by applying one (or a few) iterations of the previously presented infeasible IPM. We alter (14) by adding a logarithmic barrier:

$$\mathcal{L}_{\mu_k}^{IP-PMM}(X; \Xi_k, \lambda_k) := \mathcal{L}_{\mu_k}^{PMM}(X; \Xi_k, \lambda_k) - \mu_k \log(\det(X)), \quad (15)$$

and we treat μ_k as the barrier parameter. In order to form the optimality conditions of this sub-problem, we equate the gradient of $\mathcal{L}_{\mu_k}^{IP-PMM}(\cdot; \Xi_k, \lambda_k)$ to the zero vector, i.e.:

$$C - \mathcal{A}^*\lambda_k + \frac{1}{\mu_k}\mathcal{A}^*(\mathcal{A}X - b) + \mu_k(X - \Xi_k) - \mu_k X^{-1} = 0.$$

Introducing the variables $y = \lambda_k - \frac{1}{\mu_k}(\mathcal{A}X - b)$ and $Z = \mu_k X^{-1}$, yields:

$$
\begin{bmatrix} C - \mathcal{A}^* y - Z + \mu_k(X - \varXi_k) \\ \mathcal{A}X + \mu_k(y - \lambda_k) - b \\ XZ - \mu_k I_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Leftrightarrow \begin{bmatrix} C - A^\top y - Z + \mu_k(X - \varXi_k) \\ AX + \mu_k(y - \lambda_k) - b \\ H_{P_k}(XZ) - \mu_k I_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad (16)
$$

where the second system is obtained by introducing the symmetrization in (10), and by vectorizing the associated matrices and operators.

Given an arbitrary vector $b \in \mathbb{R}^m$, and matrix $C \in \mathbb{R}^{n \times n}$, we define the semi-norm:

$$
\|(b, C)\|_{\mathcal{S}} := \min_{X,y,Z} \left\{ \|(X, Z)\|_2 \ : \ \begin{matrix} AX = b, \\ A^\top y + Z = C \end{matrix} \right\}. \quad (17)
$$

A similar semi-norm was used before in [15], as a way to measure infeasibility in the case of linear programming problem. For a discussion of the properties of the aforementioned semi-norm, as well as how to evaluate it (using an appropriate QR factorization, which can be computed in a polynomial time), the reader is referred to [15, Section 4].

Starting Point Let us define the starting point for IP-PMM. For that, we set $(X_0, Z_0) = \rho(I_n, I_n)$, for some $\rho > 0$. We also set y_0 to some arbitrary value (e.g. $y_0 = 0$), and $\mu_0 = \frac{\langle X_0, Z_0 \rangle}{n}$. Using the aforementioned triple, we have:

$$
AX_0 = b + \bar{b}, \ A^\top y_0 + Z_0 = C + \bar{C}, \ \varXi_0 = X_0, \ \lambda_0 = y_0. \quad (18)
$$

for some $\bar{b} \in \mathbb{R}^m$, and $\bar{C} \in \mathcal{S}^n$.

Neighbourhood Below, we describe a neighbourhood in which the iterations of the method should lie. Unlike most path-following methods, we have to define a family of neighbourhoods that depend on the PMM sub-problem parameters.

Given (18), some μ_k, λ_k, and $\varXi_k$, we define the *regularized set of centers*:

$$
\mathscr{P}_{\mu_k}(\varXi_k, \lambda_k) := \{(X, y, Z) \in \mathscr{C}_{\mu_k}(\varXi_k, \lambda_k) \ : \ X \in \mathcal{S}_{++}^n, \ Z \in \mathcal{S}_{++}^n, \ XZ = \mu_k I_n\},
$$

$$
\mathscr{C}_{\mu_k}(\varXi_k, \lambda_k) := \left\{ (X, y, Z) \ : \ \begin{matrix} AX + \mu_k(y - \lambda_k) = b + \frac{\mu_k}{\mu_0}\bar{b}, \\ A^\top y + Z - \mu_k(X - \varXi_k) = C + \frac{\mu_k}{\mu_0}\bar{C} \end{matrix} \right\},
$$

where $\bar{b}$, $\bar{C}$ are as in (18). The term set of centers originates from [15].

We enlarge the previous set, by defining the following set:

$$
\tilde{\mathscr{C}}_{\mu_k}(\varXi_k, \lambda_k) := \left\{ (X, y, Z) \ : \ \begin{matrix} AX + \mu_k(y - \lambda_k) = b + \frac{\mu_k}{\mu_0}(\bar{b} + \tilde{b}_k), \\ A^\top y + Z - \mu_k(X - \varXi_k) = C + \frac{\mu_k}{\mu_0}(\bar{C} + \tilde{C}_k) \\ \|(\tilde{b}_k, \tilde{C}_k)\|_2 \le K_N, \ \|(\tilde{b}_k, \tilde{C}_k)\|_{\mathcal{S}} \le \gamma_{\mathcal{S}}\rho \end{matrix} \right\},
$$

where $K_N > 0$ is a constant, $\gamma_{\mathcal{S}} \in (0, 1)$ and $\rho > 0$ is as defined in the starting point. The vector $\tilde{b}_k$ and the matrix $\tilde{C}_k$ represent the current scaled (by $\frac{\mu_0}{\mu_k}$) infeasibilities, and will vary depending on the iteration k. While these can be defined recursively,

it is not necessary. Instead it suffices to know that they satisfy the bounds given in the definition of the previous set. In essence, the previous set requires these scaled infeasibilities to be bounded above by some constants, with respect to the 2-norm as well as the semi-norm defined in (17). We can now define a family of neighbourhoods:

$$\mathcal{N}_{\mu_k}(\Xi_k, \lambda_k) := \left\{ (X, y, Z) \in \tilde{\mathcal{C}}_{\mu_k}(\Xi_k, \lambda_k) \; : \; \begin{array}{c} X \in \mathcal{S}_{++}^n, \; Z \in \mathcal{S}_{++}^n, \\ \|H_P(XZ) - \mu I_n\|_F \le \gamma_\mu \mu_k \end{array} \right\}, \quad (19)$$

where $\gamma_\mu \in (0, 1)$ is a constant restricting the symmetrized complementarity products. Obviously, the starting point defined in (18) belongs to the neighbourhood $\mathcal{N}_{\mu_0}(\Xi_0, \lambda_0)$, with $(\tilde{b}_0, \tilde{C}_0) = (0, 0)$. Notice that the neighbourhood depends on the choice of the constants $K_N, \gamma_S, \gamma_\mu$. However, as it also depends on the parameters μ_k, λ_k, Ξ_k, the dependence on the constants is omitted for simplicity of notation.

Newton System As discussed earlier, we employ the Newton method for approximately solving a perturbed form of system (16), for all k. In particular, we perturb (16) in order to take into consideration the target reduction of the barrier parameter μ_k (by introducing the centering parameter σ_k), as well as to incorporate the initial infeasibility, given our starting point in (18). In particular, we would like to approximately solve the following system:

$$\begin{bmatrix} -(\mu_k I_n) & \mathcal{A}^* & I_n \\ \mathcal{A} & \mu_k I_m & 0 \\ Z_k & 0 & X_k \end{bmatrix} \begin{bmatrix} \Delta X_k \\ \Delta y_k \\ \Delta Z_k \end{bmatrix} = \begin{bmatrix} (C + \frac{\sigma_k \mu_k}{\mu_0} \bar{C}) - \mathcal{A}^* y_k - Z_k + \sigma_k \mu_k (X_k - \Xi_k) \\ -\mathcal{A} X_k - \sigma_k \mu_k (y_k - \lambda_k) + (b + \frac{\sigma_k \mu_k}{\mu_0} \bar{b}) \\ -X_k Z_k + \sigma_k \mu_k I_n \end{bmatrix}, \quad (20)$$

where $\bar{b}, \bar{C}$ are as in (18). We note that we could either first linearize the last block equation of (16) and then apply the symmetrization, defined in (10), or first apply the symmetrization directly to the last block equation of (16) and then linearize it. Both approaches are equivalent. Hence, following the former approach, we obtain the vectorized Newton system, that has to be solved at every iteration of IP-PMM:

$$\begin{bmatrix} -(\mu_k I_{n^2}) & A^\top & I_{n^2} \\ A & \mu_k I_m & 0 \\ E_k & 0 & F_k \end{bmatrix} \begin{bmatrix} \boldsymbol{\Delta X}_k \\ \Delta y_k \\ \boldsymbol{\Delta Z}_k \end{bmatrix}$$
$$= \begin{bmatrix} (\boldsymbol{C} + \frac{\sigma_k \mu_k}{\mu_0} \bar{\boldsymbol{C}}) - A^\top y_k - \boldsymbol{Z}_k + \sigma_k \mu_k (\boldsymbol{X}_k - \boldsymbol{\Xi}_k) \\ -A \boldsymbol{X}_k - \sigma_k \mu_k (y_k - \lambda_k) + (b + \frac{\sigma_k \mu_k}{\mu_0} \bar{b}) \\ -(Z_k^{\frac{1}{2}} \otimes Z_k^{\frac{1}{2}}) \boldsymbol{X}_k + \sigma_k \mu_k \boldsymbol{I}_n \end{bmatrix} + \begin{bmatrix} \mathsf{E}_{d,k} \\ \epsilon_{p,k} \\ \mathsf{E}_{\mu,k} \end{bmatrix}, \quad (21)$$

where $E_k = (Z_k^{\frac{1}{2}} \otimes Z_k^{\frac{1}{2}})$, $F_k = \frac{1}{2} \big(Z_k^{\frac{1}{2}} X_k \otimes Z_k^{-\frac{1}{2}} + Z^{-\frac{1}{2}} \otimes Z_k^{\frac{1}{2}} X_k \big)$, and $(\mathsf{E}_{d,k}, \epsilon_{p,k}, \mathsf{E}_{\mu,k})$ models potential errors, occurring by solving the symmetrized version of system (20) inexactly (e.g. by using a Krylov subspace method). In order to make sure that the computed direction is accurate enough, we impose the following accuracy conditions:

$$\|\mathsf{E}_{\mu,k}\|_2 = 0, \qquad \|(\epsilon_{p,k}, \mathsf{E}_{d,k})\|_2 \le \frac{\sigma_{\min}}{4\mu_0} K_N \mu_k, \qquad \|(\epsilon_{p,k}, \mathsf{E}_{d,k})\|_S \le \frac{\sigma_{\min}}{4\mu_0} \gamma_S \rho \mu_k, \quad (22)$$

where $\sigma_{\min}$ is the minimum allowed value for σ_k, K_N, $\gamma_{\mathcal{S}}$ are constants defined by the neighbourhood in (19), and ρ is defined in the starting point in (18). Notice that the condition $\|\mathsf{E}_{\mu,k}\|_2 = 0$ is imposed without loss of generality, since it can be easily satisfied in practice. For more on this, see the discussions in [31, Section 3] and [11, Lemma 4.1]. Furthermore, as we will observe in Sect. 4, the bound on the error with respect to the semi-norm defined in (17) is required to ensure polynomial complexity of the method. While evaluating this semi-norm is not particularly practical (and is never performed in practice, e.g. see [7,8,19]), it can be done in a polynomial time (see [15, Section 4]), and hence does not affect the polynomial nature of the algorithm. The algorithmic scheme of the method is summarized in Algorithm IP–PMM.

Algorithm IP–PMM Interior Point-Proximal Method of Multipliers

---

Input: $\mathcal{A}$, b, C, tol.
Parameters: $0 < \sigma_{\min} \leq \sigma_{\max} \leq 0.5$, $K_N > 0$, $0 < \gamma_{\mathcal{S}} < 1$, $0 < \gamma_\mu < 1$.
Starting point: Set as in (18).
  for $(k = 0, 1, 2, \cdots)$ **do**
    if $\left(\left(\|AX_k - b\|_2 < \text{tol} \right) \wedge \left(\|C - A^\top y_k - Z_k\|_2 < \text{tol} \right) \wedge \left(\frac{\langle X_k, Z_k \rangle}{n} < \text{tol} \right) \right)$ **then**
      return (X_k, y_k, Z_k).
    else
      Choose $\sigma_k \in [\sigma_{\min}, \sigma_{\max}]$ and solve (21) so that (22) holds.
      Choose α_k, as the largest $\alpha \in (0, 1]$, s.t. $\mu_k(\alpha) \leq (1 - 0.01\alpha)\mu_k$, and:

$$(X_k + \alpha_k \Delta X_k, \, y_k + \alpha_k \Delta y_k, \, Z_k + \alpha_k \Delta Z_k) \in \mathcal{N}_{\mu_k(\alpha)}(\Xi_k, \lambda_k),$$
$$\text{where,} \quad \mu_k(\alpha) = \frac{\langle X_k + \alpha_k \Delta X_k, \, Z_k + \alpha_k \Delta Z_k \rangle}{n}.$$

      Set $(X_{k+1}, y_{k+1}, Z_{k+1}) = (X_k + \alpha_k \Delta X_k, \, y_k + \alpha_k \Delta y_k, \, Z_k + \alpha_k \Delta Z_k)$.
      Set $\mu_{k+1} = \frac{\langle X_{k+1}, Z_{k+1} \rangle}{n}$.
      Let $r_p = AX_{k+1} - (b + \frac{\mu_{k+1}}{\mu_0}\bar{b})$, $R_d = (C + \frac{\mu_{k+1}}{\mu_0}\bar{C}) - A^\top y_{k+1} - Z_{k+1}$.
      if $\left(\left(\|(r_p, R_d)\|_2 \leq K_N \frac{\mu_{k+1}}{\mu_0} \right) \wedge \left(\|(r_p, R_d)\|_{\mathcal{S}} \leq \gamma_{\mathcal{S}} \rho \frac{\mu_{k+1}}{\mu_0} \right) \right)$ **then**
        $(\Xi_{k+1}, \lambda_{k+1}) = (X_{k+1}, y_{k+1})$.
      else
        $(\Xi_{k+1}, \lambda_{k+1}) = (\Xi_k, \lambda_k)$.
      end if
    end if
  end for

Algorithm IP–PMM deviates from standard IPM schemes due to the solution of a different Newton system, as well as due to the possible updates of the proximal estimates, i.e. Ξ_k and λ_k. Notice that when these estimates are updated, the neighbourhood in (19) changes as well, since it is parametrized by them. Intuitively, when this happens, the algorithm accepts the current iterate as a sufficiently accurate solution to the associated PMM sub-problem. However, as we will see in Sect. 4, it is not necessary for these estimates to converge to a primal-dual solution, for Algorithm IP–PMM to converge. Instead, it suffices to ensure that these estimates will remain bounded. In light of this, Algorithm IP–PMM is not studied as an inner-outer scheme, but rather as a standard IPM scheme. We will return to this point at the end of Sect. 4.

4 Convergence Analysis

In this section we prove polynomial complexity of Algorithm IP–PMM, and establish its global convergence. The analysis is modeled after that in [19]. We make use of the following two standard assumptions, generalizing those employed in [19] to the SDP case.

Assumption 1 The problems (P) and (D) are strictly feasible, that is, Slater's constraint qualification holds for both problems. Furthermore, there exists an optimal solution (X^*, y^*, Z^*) and a constant $K_* > 0$ independent of n and m such that $\|(X^*, y^*, Z^*)\|_F \le K_* \sqrt{n}$.

Assumption 2 The vectorized constraint matrix A of (P) has full row rank, that is rank$(A) = m$. Moreover, there exist constants $K_{A,1} > 0$, $K_{A,2} > 0$, $K_{r,1} > 0$, and $K_{r,2} > 0$, independent of n and m, such that:

$$\eta_{\min}(A) \ge K_{A,1}, \quad \eta_{\max}(A) \le K_{A,2}, \quad \|b\|_\infty \le K_{r,1}, \quad \|C\|_2 \le K_{r,2}\sqrt{n}.$$

Remark 1 Note that the independence of the previous constants from n and m is assumed for simplicity of exposition. In particular, as long as these constants depend polynomially on n (or m), the analysis still holds, simply by altering the worst-case polynomial bound for the number of iterations of the algorithm (given later in Theorem 4.2).

Remark 2 Assumption 1 is a direct extension of that in [19, Assumption 1]. The positive semi-definiteness of X^* and Z^* implies that Tr(X^*) + Tr$(Z^*) \le 2K_* n$ (from equivalence of the norms $\|\cdot\|_1$ and $\|\cdot\|_2$), which is one of the assumptions employed in [29,31]. Notice that we assume $n > m$, without loss of generality. The theory in this section would hold if $m > n$, simply by replacing n by m in the upper bound of the norm of the optimal solution as well as of the problem data.

Before proceeding with the convergence analysis, we briefly provide an outline of it, for the convenience of the reader. Firstly, it should be noted that polynomial complexity as well as global convergence of Algorithm IP–PMM is proven by induction on the iterations k of the method. To that end, we provide some necessary technical results in Lemmas 4.1–4.3. Then, in Lemma 4.4 we are able to show that the iterates (X_k, y_k, Z_k) of Algorithm IP–PMM will remain bounded for all k. Subsequently, we provide some additional technical results in Lemmas 4.5–4.7, which are then used in Lemma 4.8, where we show that the Newton direction computed at every iteration k is also bounded. All the previous are utilized in Lemmas 4.9–4.10, where we provide a lower bound for the step-length α_k chosen by Algorithm IP–PMM at every iteration k. Then, Q-linear convergence of μ_k (with R-linear convergence of the regularized residuals) is shown in Theorem 4.1. Polynomial complexity is proven in Theorem 4.2, and finally, global convergence is established in Theorem 4.3.

Let us now use the properties of the proximal operator defined in (5).

Lemma 4.1 *Given Assumption* 1, *and for all* $\lambda \in \mathbb{R}^m$, $\varXi \in \mathcal{S}_+^n$ *and* $0 \le \mu < \infty$, *there exists a unique pair* (X_r^*, y_r^*), *such that* $(X_r^*, y_r^*) = \mathcal{P}(\varXi, \lambda)$, $X_r^* \in \mathcal{S}_+^n$, *and*

$$\|(X_r^*, y_r^*) - (X^*, y^*)\|_2 \le \|(\varXi, \lambda) - (X^*, y^*)\|_2, \tag{23}$$

where $\mathcal{P}(\cdot)$ *is defined as in* (5), *and* (X^*, y^*) *is such that* $(0, 0) \in T_{\mathcal{L}}(X^*, y^*)$.

Proof The thesis follows from the developments in [21, Proposition 1]. □

In the following lemma, we bound the solution of every PMM sub-problem encountered by Algorithm IP–PMM, while establishing bounds for the proximal estimates $\varXi_k$, and λ_k.

Lemma 4.2 *Given Assumptions* 1, 2, *there exists a triple* $(X_{r_k}^*, y_{r_k}^*, Z_{r_k}^*)$, *satisfying:*

$$
\begin{aligned}
A X_{r_k}^* + \mu(y_{r_k}^* - \lambda_k) - b &= 0, \\
-C + A^\top y_{r_k}^* + Z_{r_k}^* - \mu(X_{r_k}^* - \varXi_k) &= 0, \\
\langle X_{r_k}^*, Z_{r_k}^* \rangle &= 0,
\end{aligned}
\tag{24}
$$

with $X_{r_k}^*, Z_{r_k}^* \in \mathcal{S}_+^n$, *and* $\|(X_{r_k}^*, y_{r_k}^*, Z_{r_k}^*)\|_2 = O(\sqrt{n})$, *for all* $\lambda_k \in \mathbb{R}^m$, $\varXi_k \in \mathcal{S}_+^n$, *produced by Algorithm* IP–PMM, *and any* $\mu \in [0, \infty)$. *Moreover,* $\|(\varXi_k, \lambda_k)\|_2 = O(\sqrt{n})$, *for all* $k \ge 0$.

Proof We prove the claim by induction on the iterates, $k \ge 0$, of Algorithm IP–PMM. At iteration $k = 0$, we have that $\lambda_0 = y_0$ and $\varXi_0 = X_0$. But from the construction of the starting point in (18), we know that $\|(X_0, y_0)\|_2 = O(\sqrt{n})$. Hence, $\|(\varXi_0, \lambda_0)\|_2 = O(\sqrt{n})$ (assuming $n > m$). Invoking Lemma 4.1, there exists a unique pair $(X_{r_0}^*, y_{r_0}^*)$ satisfying:

$$(X_{r_0}^*, y_{r_0}^*) = \mathcal{P}_0(\varXi_0, \lambda_0), \qquad \|(X_{r_0}^*, y_{r_0}^*) - (X^*, y^*)\|_2 \le \|(\varXi_0, \lambda_0) - (X^*, y^*)\|_2,$$

where (X^*, y^*, Z^*) solves (P)–(D), and from Assumption 1, is such that $\|X^*, y^*, Z^*\|_2 = O(\sqrt{n})$. Using the triangular inequality, and combining the latter inequality with our previous observations, yields that $\|(X_{r_0}^*, y_{r_0}^*)\|_2 = O(\sqrt{n})$. From the definition of the operator in (6), we know that:

$$-C + \mathcal{A}^* y_{r_0}^* - \mu(X_{r_0}^* - \varXi_0) \in \partial \delta_{\mathcal{S}_+^n}(X_{r_0}^*), \qquad \mathcal{A} X_{r_0}^* + \mu(y_{r_0}^* - \lambda_0) - b = 0,$$

where $\partial(\delta_{\mathcal{S}_+^n}(\cdot))$ is the sub-differential of the indicator function defined in (4). Hence, there must exist $-Z_{r_0}^* \in \partial \delta_{\mathcal{S}_+^n}(X_{r_0}^*)$ (and thus, $Z_{r_0}^* \in \mathcal{S}_+^n$, $\langle X_{r_0}, Z_{r_0} \rangle = 0$), such that:

$$Z_{r_0}^* = C - \mathcal{A}^* y_{r_0}^* + \mu(X_{r_0}^* - \varXi_0), \quad \langle X_{r_0}^*, Z_{r_0}^* \rangle = 0, \quad \|Z_{r_0}^*\|_2 = O(\sqrt{n}),$$

where $\|Z_{r_0}^*\|_2 = O(\sqrt{n})$ follows from Assumption 2, combined with $\|(X_{r_0}^*, y_{r_0}^*)\|_2 = O(\sqrt{n})$.

Let us now assume that at some iteration k of Algorithm IP–PMM, we have $\|(\varXi_k, \lambda_k)\|_2 = O(\sqrt{n})$. There are two cases for the subsequent iterations:

1. The proximal estimates are updated, that is $(\Xi_{k+1}, \lambda_{k+1}) = (X_{k+1}, y_{k+1})$, or
2. the proximal estimates stay the same, i.e. $(\Xi_{k+1}, \lambda_{k+1}) = (\Xi_k, \lambda_k)$.

Case 1 We know by construction that this occurs only if the following is satisfied:

$$\|(r_p, \boldsymbol{R}_d)\|_2 \le K_N \frac{\mu_{k+1}}{\mu_0},$$

where r_p, $\boldsymbol{R}_d$ are defined in Algorithm IP–PMM. However, from the neighbourhood conditions in (19), we know that:

$$\|\big(r_p + \mu_{k+1}(y_{k+1} - \lambda_k), \boldsymbol{R}_d + \mu_{k+1}(X_{k+1} - \boldsymbol{\Xi}_k)\big)\|_2 \le K_N \frac{\mu_{k+1}}{\mu_0}.$$

Combining the last two inequalities by applying the triangular inequality, and using the inductive hypothesis ($\|(\boldsymbol{\Xi}_k, \lambda_k)\|_2 = O(\sqrt{n})$), yields that

$$\|(X_{k+1}, y_{k+1})\|_2 \le \frac{2K_N}{\mu_0} + \|(\boldsymbol{\Xi}_k, \lambda_k)\|_2 = O(\sqrt{n}).$$

Hence, $\|(\boldsymbol{\Xi}_{k+1}, \lambda_{k+1})\|_2 = O(\sqrt{n})$. Then, we can invoke Lemma 4.1, with $\lambda = \lambda_{k+1}$, $\Xi = \Xi_{k+1}$ and any $\mu \ge 0$, which gives

$$\|(X^*_{r_{k+1}}, y^*_{r_{k+1}}) - (X^*, y^*)\|_2 \le \|(\boldsymbol{\Xi}_{k+1}, \lambda_{k+1}) - (X^*, y^*)\|_2.$$

A simple manipulation shows that $\|(X^*_{r_{k+1}}, y^*_{r_{k+1}})\|_2 = O(\sqrt{n})$. As before, we use (6) alongside Assumption 2 to show the existence of $-Z^*_{r_{k+1}} \in \partial \delta_{\mathcal{S}^n_+}(X^*_{r_{k+1}})$, such that the triple $(X^*_{r_{k+1}}, y^*_{r_{k+1}}, Z^*_{r_{k+1}})$ satisfies (24) with $\|Z^*_{r_{k+1}}\|_2 = O(\sqrt{n})$.

Case 2 In this case, we have $(\Xi_{k+1}, \lambda_{k+1}) = (\Xi_k, \lambda_k)$, and hence the inductive hypothesis gives us directly that $\|(\boldsymbol{\Xi}_{k+1}, \lambda_{k+1})\|_2 = O(\sqrt{n})$. As before, there exists a triple $(X^*_{r_{k+1}}, y^*_{r_{k+1}}, Z^*_{r_{k+1}})$ satisfying (24), with $\|(X^*_{r_{k+1}}, y^*_{r_{k+1}}, \boldsymbol{Z}^*_{r_{k+1}})\|_2 = O(\sqrt{n})$. □

In the next lemma we define and bound a triple solving a particular parametrized nonlinear system of equations, which is then used in Lemma 4.4 in order to prove boundedness of the iterates of Algorithm IP–PMM.

Lemma 4.3 *Given Assumptions 1, 2, a pair (Ξ_k, λ_k), produced at an arbitrary iteration $k \ge 0$ of Algorithm IP–PMM, and any $\mu \in [0, \infty)$, there exists a triple $(\tilde{X}, \tilde{y}, \tilde{Z})$ which satisfies the following system of equations:*

$$A\tilde{X} + \mu\tilde{y} = b + \bar{b} + \mu\lambda_k + \tilde{b}_k,$$
$$A^\top\tilde{y} + \tilde{Z} - \mu\tilde{X} = C + \bar{C} - \mu\Xi_k + \tilde{C}_k, \qquad (25)$$
$$\tilde{X}\tilde{Z} = \theta I_n,$$

for some arbitrary $\theta > 0$ ($\theta = \Theta(1)$), with $\tilde{X}, \tilde{Z} \in \mathcal{S}^n_{++}$ and $\|(\tilde{X}, \tilde{y}, \tilde{Z})\|_2 = O(\sqrt{n})$, where $\tilde{b}_k$, $\tilde{C}_k$ are defined in (19), while $\bar{b}$, $\bar{C}$ are defined with the starting point in (18). Furthermore, $v_{\min}(\tilde{X}) \ge \xi$ and $v_{\min}(\tilde{Z}) \ge \xi$, for some positive $\xi = \Theta(1)$.

Proof Let $k \geq 0$ denote an arbitrary iteration of Algorithm IP–PMM. Let also $\bar{b}$, $\bar{C}$ as defined in (18), and $\tilde{b}_k$, $\tilde{C}_k$, as defined in the neighbourhood conditions in (19). Given an arbitrary positive constant $\theta > 0$, we consider the following barrier primal-dual pair:

$$\min_{X \in \mathcal{S}^n} \left(\langle C + \bar{C} + \tilde{C}_k \rangle, X \rangle - \theta \ln(\det(X)) \right), \quad \text{s.t. } \mathcal{A}X = b + \bar{b} + \tilde{b}_k, \tag{26}$$

$$\max_{y \in \mathbb{R}^m, Z \in \mathcal{S}^n} \left((b + \bar{b} + \tilde{b}_k)^\top y + \theta \ln(\det(Z)) \right), \quad \text{s.t. } \mathcal{A}^* y + Z = C + \bar{C} + \tilde{C}_k. \tag{27}$$

Let us now define the following triple:

$$(\hat{X}, \hat{y}, \hat{Z}) := \arg \min_{(X, y, Z)} \left\{ \|(X, Z)\|_2 : AX = \tilde{b}_k, \ A^\top y + Z = \tilde{C}_k \right\}.$$

From the neighbourhood conditions (19), we know that $\|(\tilde{b}_k, \tilde{C}_k)\|_{\mathcal{S}} \leq \gamma_{\mathcal{S}} \rho$, and from the definition of the semi-norm in (17), we have that $\|(\hat{X}, \hat{Z})\|_2 \leq \gamma_{\mathcal{S}} \rho$. Using (17) alongside Assumption 2, we can also show that $\|\hat{y}\|_2 = \Theta(\|(\hat{X}, \hat{Z})\|_2)$. On the other hand, from the definition of the starting point, we have that $(X_0, Z_0) = \rho(I_n, I_n)$. By defining the following auxiliary point:

$$(\bar{X}, \bar{y}, \bar{Z}) = (X_0, y_0, Z_0) + (\hat{X}, \hat{y}, \hat{Z}),$$

we have that $(1 + \gamma_{\mathcal{S}})\rho(I_n, I_n) \succeq (\bar{X}, \bar{Z}) \succeq (1 - \gamma_{\mathcal{S}})\rho(I_n, I_n)$, that is, the eigenvalues of these matrices are bounded by constants that are independent of the problem under consideration. By construction, the triple $(\bar{X}, \bar{y}, \bar{Z})$ is a feasible solution for the primal-dual pair in (26)–(27), giving bounded primal and dual objective values, respectively. This, alongside Weierstrass's theorem on a potential function, can be used to show that the solution of problem (26)–(27) is bounded. In other words, for any choice of $\theta > 0$, there must exist a bounded triple (X_s^*, y_s^*, Z_s^*) solving (26)–(27), i.e.:

$$AX_s^* = b + \bar{b} + \tilde{b}_k, \quad A^\top y_s^* + Z_s^* = C + \bar{C} + \tilde{C}_k, \quad X_s^* Z_s^* = \theta I_n,$$

such that $\nu_{\max}(X_{s*}) \leq K_{s*}$ and $\nu_{\max}(Z_{s*}) \leq K_{s*}$, where $K_{s*} > 0$ is a positive constant. In turn, combining this with Assumption 2 implies that $\|(X_s^*, y_s^*, Z_s^*)\|_2 = O(\sqrt{n})$.

Let us now apply the PMM to (26)–(27), given the estimates Ξ_k, λ_k. We should note at this point that the proximal operator used here is different from that in (5), since it is based on a different maximal monotone operator from that in (3). In particular, we associate a single-valued maximal monotone operator to (26)–(27), with graph:

$$\tilde{T}_{\mathcal{L}}(X, y) := \left\{ (V, u) : V = (C + \bar{C} + \tilde{C}_k) - \mathcal{A}^* y - \theta X^{-1}, u = \mathcal{A}X - (b + \bar{b} + \tilde{b}_k) \right\}.$$

As before, the proximal operator is defined as $\tilde{\mathcal{P}} := \left(I_{n+m} + \frac{1}{\mu}\tilde{T}_{\mathcal{L}}\right)^{-1}$, and is single-valued and non-expansive. We let any $\mu \in [0, \infty)$ and define the following penalty function:

$$\tilde{\mathcal{L}}_{\mu,\theta}(X; \Xi_k, \lambda_k) := \langle C + \bar{C} + \tilde{C}_k, X \rangle + \frac{1}{2}\mu\|X - \Xi_k\|_F^2 + \frac{1}{2\mu}\|\mathcal{A}X - (b + \bar{b} + \tilde{b}_k)\|_2^2$$
$$- (\lambda_k)^\top(\mathcal{A}X - (b + \bar{b} + \tilde{b}_k)) - \theta \ln(\det(X)).$$

By defining the variables $y = \lambda_k - \frac{1}{\mu}(\mathcal{A}X - (b + \bar{b} + \tilde{b}_k))$ and $Z = \theta X^{-1}$, we can see that the optimality conditions of this PMM sub-problem are exactly those stated in (25). Equivalently, we can find a pair $(\tilde{X}, \tilde{y})$ such that $(\tilde{X}, \tilde{y}) = \tilde{\mathcal{P}}(\Xi_k, \lambda_k)$ and set $\tilde{Z} = \theta \tilde{X}^{-1}$. We can now use the non-expansiveness of $\tilde{\mathcal{P}}$, as in Lemma 4.1, to obtain:

$$\|(\tilde{X}, \tilde{y}) - (X_s^*, y_s^*)\|_2 \le \|(\Xi_k, \lambda_k) - (X_s^*, y_s^*)\|_2.$$

But we know, from Lemma 4.2, that $\|(\Xi_k, \lambda_k)\|_2 = O(\sqrt{n}), \forall\, k \ge 0$. Combining this with our previous observations, yields that $\|(\tilde{X}, \tilde{y})\|_2 = O(\sqrt{n})$. Setting $\tilde{Z} = \theta\tilde{X}^{-1}$, gives a triple $(\tilde{X}, \tilde{y}, \tilde{Z})$ that satisfies (25), while $\|(\tilde{X}, \tilde{y}, \tilde{Z})\|_2 = O(\sqrt{n})$ (from dual feasibility).

To conclude the proof, let us notice that the value of $\tilde{\mathcal{L}}_{\mu,\theta}(X; \Xi_k, \lambda_k)$ will grow unbounded as $v_{\min}(X) \to 0$ or $v_{\max}(X) \to \infty$. Hence, there must exist a constant $\tilde{K} > 0$, such that the minimizer of this function satisfies $\frac{1}{\tilde{K}} \le v_{\min}(\tilde{X}) \le v_{\max}(\tilde{X}) \le \tilde{K}$. The relation $\tilde{X}\tilde{Z} = \theta I_n$ then implies that $\frac{\theta}{\tilde{K}} \le v_{\min}(\tilde{Z}) \le v_{\max}(\tilde{Z}) \le \theta\tilde{K}$. Hence, there exists some $\xi = \Theta(1)$ such that $v_{\min}(\tilde{X}) \ge \xi$ and $v_{\min}(\tilde{Z}) \ge \xi$. $\qquad\square$

In the following lemma, we derive boundedness of the iterates of Algorithm IP–PMM.

Lemma 4.4 *Given Assumptions 1 and 2, the iterates (X_k, y_k, Z_k) produced by Algorithm IP–PMM, for all $k \ge 0$, are such that:*

$$Tr(X_k) = O(n), \qquad Tr(Z_k) = O(n), \qquad \|(X_k, y_k, Z_k)\|_2 = O(n).$$

Proof Let an iterate $(X_k, y_k, Z_k) \in \mathcal{N}_{\mu_k}(\Xi_k, \lambda_k)$, produced by Algorithm IP–PMM during an arbitrary iteration $k \ge 0$, be given. Firstly, we invoke Lemma 4.3, from which we have a triple $(\tilde{X}, \tilde{y}, \tilde{Z})$ satisfying (25), for $\mu = \mu_k$. Similarly, by invoking Lemma 4.2, we know that there exists a triple $(X_{r_k}^*, y_{r_k}^*, Z_{r_k}^*)$ satisfying (24), with $\mu = \mu_k$. Consider the following auxiliary point:

$$\left(\left(1 - \frac{\mu_k}{\mu_0}\right)X_{r_k}^* + \frac{\mu_k}{\mu_0}\tilde{X} - X_k, \left(1 - \frac{\mu_k}{\mu_0}\right)y_{r_k}^* + \frac{\mu_k}{\mu_0}\tilde{y} - y_k, \left(1 - \frac{\mu_k}{\mu_0}\right)Z_{r_k}^* + \frac{\mu_k}{\mu_0}\tilde{Z} - Z_k\right).$$
$$\tag{28}$$

Using (28) and (24)–(25) (for $\mu = \mu_k$), one can observe that:

$$
\begin{aligned}
A&\left(\left(1 - \frac{\mu_k}{\mu_0}\right) X_{r_k}^* + \frac{\mu_k}{\mu_0} \tilde{X} - X_k\right) + \mu_k\left(\left(1 - \frac{\mu_k}{\mu_0}\right) y_{r_k}^* + \frac{\mu_k}{\mu_0}\tilde{y} - y_k\right) \\
&= \left(1 - \frac{\mu_k}{\mu_0}\right)(AX_{r_k}^* + \mu_k y_{r_k}^*) + \frac{\mu_k}{\mu_0}(A\tilde{X} + \mu_k\tilde{y}) - AX_k - \mu_k y_k \\
&= \left(1 - \frac{\mu_k}{\mu_0}\right)(b + \mu_k\lambda_k) + \frac{\mu_k}{\mu_0}(b + \mu_k\lambda_k + \tilde{b}_k + \bar{b}) - AX_k - \mu_k y_k \\
&= b + \mu_k\lambda_k + \frac{\mu_k}{\mu_0}(\tilde{b}_k + \bar{b}) - AX_k - \mu_k y_k = 0,
\end{aligned}
$$

where the last equality follows from the definition of the neighbourhood $\mathcal{N}_{\mu_k}(\Xi_k, \lambda_k)$. Similarly, one can show that:

$$
\begin{aligned}
-\mu_k&\left(\left(1 - \frac{\mu_k}{\mu_0}\right) X_{r_k}^* + \frac{\mu_k}{\mu_0}\tilde{X} - X_k\right) + A^\top\left(\left(1 - \frac{\mu_k}{\mu_0}\right) y_{r_k}^* + \frac{\mu_k}{\mu_0}\tilde{y} - y_k\right) \\
&+ \left(\left(1 - \frac{\mu_k}{\mu_0}\right) Z_{r_k}^* + \frac{\mu_k}{\mu_0}\tilde{Z} - Z_k\right) = 0.
\end{aligned}
$$

By combining the previous two relations, we obtain:

$$
\begin{aligned}
&\left(\left(1 - \frac{\mu_k}{\mu_0}\right) X_{r_k}^* + \frac{\mu_k}{\mu_0}\tilde{X} - X_k\right)^\top\left(\left(1 - \frac{\mu_k}{\mu_0}\right) Z_{r_k}^* + \frac{\mu_k}{\mu_0}\tilde{Z} - Z_k\right) \\
&= \mu_k\left(\left(1 - \frac{\mu_k}{\mu_0}\right) X_{r_k}^* + \frac{\mu_k}{\mu_0}\tilde{X} - X_k\right)^\top\left(\left(1 - \frac{\mu_k}{\mu_0}\right) X_{r_k}^* + \frac{\mu_k}{\mu_0}\tilde{X} - X_k\right) \quad (29) \\
&+ \mu_k\left(\left(1 - \frac{\mu_k}{\mu_0}\right) y_{r_k}^* + \frac{\mu_k}{\mu_0}\tilde{y} - y_k\right)^\top\left(\left(1 - \frac{\mu_k}{\mu_0}\right) y_{r_k}^* + \frac{\mu_k}{\mu_0}\tilde{y} - y_k\right) \geq 0.
\end{aligned}
$$

Observe that (29) can equivalently be written as:

$$
\begin{aligned}
&\left\langle\left(1 - \frac{\mu_k}{\mu_0}\right) X_{r_k}^* + \frac{\mu_k}{\mu_0}\tilde{X}, Z_k\right\rangle + \left\langle\left(1 - \frac{\mu_k}{\mu_0}\right) Z_{r_k}^* + \frac{\mu_k}{\mu_0}\tilde{Z}, X_k\right\rangle \leq \\
&\left\langle\left(1 - \frac{\mu_k}{\mu_0}\right) X_{r_k}^* + \frac{\mu_k}{\mu_0}\tilde{X}, \left(1 - \frac{\mu_k}{\mu_0}\right) Z_{r_k}^* + \frac{\mu_k}{\mu_0}\tilde{Z}\right\rangle + \langle X_k, Z_k\rangle.
\end{aligned}
$$

However, from Lemmas 4.2 and 4.3, we have that $\tilde{X} \succeq \xi I_n$ and $\tilde{Z} \succeq \xi I_n$, for some positive constant $\xi = \Theta(1)$, $\langle X_{r_k}^*, Z_k\rangle \geq 0$, $\langle Z_{r_k}^*, X_k\rangle \geq 0$, while $\|(X_{r_k}^*, Z_{r_k}^*)\|_F = O(\sqrt{n})$, and $\|(\tilde{X}, \tilde{Z})\|_F = O(\sqrt{n})$. Furthermore, by definition we have that $n\mu_k =$

$\langle X_k, Z_k \rangle$. By combining all the previous, we obtain:

$$
\begin{aligned}
&\frac{\mu_k}{\mu_0} \xi \big(\mathrm{Tr}(X_k) + \mathrm{Tr}(Z_k) \big) \\
&= \frac{\mu_k}{\mu_0} \xi \big(\langle I_n, X_k \rangle + \langle I_n, Z_k \rangle \big) \\
&\leq \left\langle \left(1 - \frac{\mu_k}{\mu_0}\right) X_{r_k}^* + \frac{\mu_k}{\mu_0} \tilde{X}, Z_k \right\rangle + \left\langle \left(1 - \frac{\mu_k}{\mu_0}\right) Z_{r_k}^* + \frac{\mu_k}{\mu_0} \tilde{Z}, X_k \right\rangle \\
&\leq \left\langle \left(1 - \frac{\mu_k}{\mu_0}\right) X_{r_k}^* + \frac{\mu_k}{\mu_0} \tilde{X}, \left(1 - \frac{\mu_k}{\mu_0}\right) Z_{r_k}^* + \frac{\mu_k}{\mu_0} \tilde{Z} \right\rangle + \langle X_k, Z_k \rangle \quad (30) \\
&= \frac{\mu_k}{\mu_0} \left(1 - \frac{\mu_k}{\mu_0}\right) \langle X_{r_k}^*, \tilde{Z} \rangle + \frac{\mu_k}{\mu_0} \left(1 - \frac{\mu_k}{\mu_0}\right) \langle \tilde{X}, Z_r^* \rangle \\
&\quad + \left(\frac{\mu_k}{\mu_0}\right)^2 \langle \tilde{X}, \tilde{Z} \rangle + \langle X_k, Z_k \rangle \\
&= O(n\mu_k),
\end{aligned}
$$

where the first inequality follows since $X_{r_k}^*$, $Z_{r_k}^*$, $\tilde{X}$, $\tilde{Z} \in \mathcal{S}_+^n$ and $(\tilde{X}, \tilde{Z}) \succeq \xi(I_n, I_n)$. In the penultimate equality we used (24) (i.e. $\langle X_{r_k}^*, Z_{r_k}^* \rangle = 0$). Hence, (30) implies that:

$$
\mathrm{Tr}(X_k) = O(n), \qquad \mathrm{Tr}(Z_k) = O(n).
$$

From positive definiteness we have that $\|(X_k, Z_k)\|_F = O(n)$. Finally, from the neighbourhood conditions we know that:

$$
\boldsymbol{C} - A^\top y_k - \boldsymbol{Z}_k + \mu_k (X_k - \boldsymbol{\Xi}_k) + \frac{\mu_k}{\mu_0} (\tilde{\boldsymbol{C}}_k + \bar{\boldsymbol{C}}) = 0.
$$

All terms above (except for y_k) have a 2-norm that is bounded by some quantity that is $O(n)$ (note that $\|(\bar{C}, \bar{b})\|_2 = O(\sqrt{n})$ using Assumption 2 and the definition in (18)). Hence, using again Assumption 2 (i.e. A is full rank, with singular values independent of n and m) yields that $\|y_k\|_2 = O(n)$, and completes the proof. □

In what follows, we provide Lemmas 4.5–4.7, which we use to prove boundedness of the Newton direction computed at every iteration of Algorithm IP–PMM, in Lemma 4.8.

Lemma 4.5 *Let* $D_k = S_k^{-\frac{1}{2}} F_k = S_k^{\frac{1}{2}} E_k^{-1}$, *where* $S_k = E_k F_k$, *and* E_k, F_k *are defined as in the Newton system in* (21). *Then, for any* $M \in \mathbb{R}^{n \times n}$,

$$
\|D_k^{-T} \boldsymbol{M}\|_2^2 \leq \frac{1}{(1 - \gamma_\mu)\mu_k} \|Z_k^{\frac{1}{2}} M Z_k^{\frac{1}{2}}\|_F^2, \quad \|D_k \boldsymbol{M}\|_2^2 \leq \frac{1}{(1 - \gamma_\mu)\mu_k} \|X_k^{\frac{1}{2}} M X_k^{\frac{1}{2}}\|_F^2,
$$

where γ_μ is defined in (19). Moreover, we have that:

$$\|D_k^{-T}\|_2^2 \le \frac{1}{(1-\gamma_\mu)\mu_k}\|Z_k\|_F^2 = O\left(\frac{n^2}{\mu_k}\right), \quad \|D_k\|_2^2 \le \frac{1}{(1-\gamma_\mu)\mu_k}\|X_k\|_F^2 = O\left(\frac{n^2}{\mu_k}\right).$$

Proof The proof of the first two inequalities follows exactly the developments in [31, Lemma 5]. The bound on the 2-norm of the matrix D_k^{-T} follows by choosing M such that M is a unit eigenvector, corresponding to the largest eigenvalue of D_k^{-T}. Then, $\|D_k^{-T}M\|_2^2 = \|D_k^{-T}\|_2^2$. But, we have that:

$$\begin{aligned}
\|D_k^{-T}M\|_2^2 &\le \frac{1}{(1-\gamma_\mu)\mu_k}\|Z_k^{\frac{1}{2}}MZ_k^{\frac{1}{2}}\|_F^2 \\
&= \frac{1}{(1-\gamma_\mu)\mu_k}\mathrm{Tr}(Z_kM^\top Z_kM) \\
&\le \frac{1}{(1-\gamma_\mu)\mu_k}\|Z_k\|_F^2 = O\left(\frac{n^2}{\mu_k}\right)
\end{aligned}$$

where we used the cyclic property of the trace as well as Lemma 4.4. The same reasoning applies to deriving the bound for $\|D_k\|_2^2$. $\qquad\square$

Lemma 4.6 *Let D_k and S_k be defined as in Lemma 4.5. Then, we have that:*

$$\|D_k^{-T}\boldsymbol{\Delta X}_k\|_2^2 + \|D_k\boldsymbol{\Delta Z}_k\|_2^2 + 2\langle\Delta X_k, \Delta Z_k\rangle = \|S_k^{-\frac{1}{2}}\boldsymbol{R}_{\mu,k}\|_2^2,$$

where $R_{\mu,k} = \sigma_k\mu_k I_n - Z_k^{\frac{1}{2}}X_kZ_k^{\frac{1}{2}}$. Furthermore,

$$\|H_{P_k}(\Delta X_k\Delta Z_k)\|_F \le \frac{\sqrt{\frac{1+\gamma_\mu}{1-\gamma_\mu}}}{2}\left(\|D_k^{-T}\boldsymbol{\Delta X}_k\|_2^2 + \|D_k\boldsymbol{\Delta Z}_k\|_2^2\right),$$

where γ_μ is defined in (19).

Proof The equality follows directly by pre-multiplying by $S^{-\frac{1}{2}}$ on both sides of the third block equation of the Newton system in (21) and by then taking the 2-norm (see [29, Lemma 3.1]). For a proof of the inequality, we refer the reader to [29, Lemma 3.3]. $\qquad\square$

Lemma 4.7 *Let S_k be as defined in Lemma 4.5, and $R_{\mu,k}$ as defined in Lemma 4.6. Then,*

$$\|S_k^{-\frac{1}{2}}\boldsymbol{R}_{\mu,k}\|_2^2 = O(n\mu_k).$$

Proof The proof is omitted since it follows exactly the developments in [31, Lemma 7]. $\qquad\square$

We are now ready to derive bounds for the Newton direction computed at every iteration of Algorithm IP–PMM.

Lemma 4.8 *Given Assumptions* 1 *and* 2*, and the Newton direction* $(\Delta X_k, \Delta y_k, \Delta Z_k)$ *obtained by solving system* (21) *during an arbitrary iteration* $k \geq 0$ *of Algorithm* IP–PMM*, we have that:*

$$\|H_{P_k}(\Delta X_k \Delta Z_k)\|_F = O(n^4\mu), \qquad \|(\boldsymbol{\Delta X}_k, \Delta y_k, \boldsymbol{\Delta Z}_k)\|_2 = O(n^3).$$

Proof Consider an arbitrary iteration k of Algorithm IP–PMM. We invoke Lemmas 4.2, 4.3, for $\mu = \sigma_k\mu_k$. That is, there exists a triple $(X_{r_k}^*, y_{r_k}^*, Z_{r_k}^*)$ satisfying (24), and a triple $(\tilde{X}, \tilde{y}, \tilde{Z})$ satisfying (25), for $\mu = \sigma_k\mu_k$. Using the centering parameter σ_k, define:

$$
\begin{aligned}
\hat{C} &= -\left(\frac{\sigma_k}{\mu_0}\bar{C} - (1-\sigma_k)\left(X_k - \boldsymbol{\Xi}_k + \frac{\mu_k}{\mu_0}(\tilde{X} - X_{r_k}^*)\right) + \frac{1}{\mu_k}\mathsf{E}_{d,k}\right), \\
\hat{b} &= -\left(\frac{\sigma_k}{\mu_0}\bar{b} + (1-\sigma_k)\left(y_k - \lambda_k + \frac{\mu_k}{\mu_0}(\tilde{y} - y_{r_k}^*)\right) + \frac{1}{\mu_k}\epsilon_{p,k}\right),
\end{aligned}
\tag{31}
$$

where $\bar{b}, \bar{C}, \mu_0$ are given by (18) and $\epsilon_{p,k}, \mathsf{E}_{d,k}$ model the errors which occur when system (20) is solved inexactly. Notice that these errors are required to satisfy (22) at every iteration k. Using Lemmas 4.2, 4.3, 4.4, relation (22), and Assumption 2, we know that $\|(\hat{C}, \hat{b})\|_2 = O(n)$. Then, by applying again Assumption 2, we know that there must exist a matrix $\hat{X} \in \mathbb{R}^{n \times n}$ such that $\mathcal{A}\hat{X} = \hat{b}$, $\|\hat{X}\|_F = O(n)$, and by setting $\hat{Z} = \hat{C} + \mu\hat{X}$, we have that $\|\hat{Z}\|_F = O(n)$ and:

$$\mathcal{A}\hat{X} = \hat{b}, \qquad \hat{Z} - \mu_k\hat{X} = \hat{C}. \tag{32}$$

Using $(X_{r_k}^*, y_{r_k}^*, Z_{r_k}^*)$, $(\tilde{X}, \tilde{y}, \tilde{Z})$, as well as the triple $(\hat{X}, 0, \hat{Z})$, where $(\hat{X}, \hat{Z})$ is defined in (32), we define the following auxiliary triple:

$$(\bar{X}, \bar{y}, \bar{Z}) = (\Delta X_k, \Delta y_k, \Delta Z_k) + \frac{\mu_k}{\mu_0}(\tilde{X}, \tilde{y}, \tilde{Z}) - \frac{\mu_k}{\mu_0}(X_{r_k}^*, y_{r_k}^*, Z_{r_k}^*) + \mu_k(\hat{X}, 0, \hat{Z}). \tag{33}$$

Using (33), (31), and the second block equation of (21):

$$
\begin{aligned}
A\bar{X} + \mu_k\bar{y} &= (A\boldsymbol{\Delta X}_k + \mu_k\Delta y_k) + \frac{\mu_k}{\mu_0}((A\tilde{X} + \mu_k\tilde{y}) - (AX_{r_k}^* + \mu_k y_{r_k}^*)) + \mu_k A\hat{X} \\
&= \left(b + \sigma_k\frac{\mu_k}{\mu_0}\bar{b} - AX_k - \sigma_k\mu_k(y_k - \lambda_k) + \epsilon_{p,k}\right) \\
&\quad + \frac{\mu_k}{\mu_0}((A\tilde{X} + \mu_k\tilde{y}) - (AX_{r_k}^* + \mu_k y_{r_k}^*)) \\
&\quad - \mu_k\left(\sigma_k\frac{\bar{b}}{\mu_0} + (1-\sigma_k)(y_k - \lambda_k)\right) - \frac{\mu_k}{\mu_0}(1-\sigma_k)\mu_k(\tilde{y} - y_{r_k}^*) - \epsilon_{p,k}.
\end{aligned}
$$

Then, by deleting opposite terms in the right-hand side, and employing (24)–(25) (evaluated at $\mu = \sigma_k \mu_k$ from the definition of $(X_{r_k}^*, y_{r_k}^*, Z_{r_k}^*)$ and $(\tilde{X}, \tilde{y}, \tilde{Z})$), we have

$$
\begin{aligned}
A\bar{X} + \mu_k \bar{y} &= \left(b + \sigma_k \frac{\mu_k}{\mu_0} \bar{b} - AX_k - \sigma_k \mu_k (y_k - \lambda_k) \right) + \frac{\mu_k}{\mu_0} (b + \sigma_k \mu_k \lambda_k + \bar{b} + \tilde{b}_k) \\
&\quad - \frac{\mu_k}{\mu_0}(\sigma_k \mu_k \lambda_k + b) - \mu_k \left(\sigma_k \frac{\bar{b}}{\mu_0} + (1 - \sigma_k)(y_k - \lambda_k) \right) \\
&= b + \frac{\mu_k}{\mu_0}(\bar{b} + \tilde{b}_k) - AX_k - \mu_k (y_k - \lambda_k) \\
&= 0,
\end{aligned}
$$

where the last equation follows from the neighbourhood conditions (i.e. $(X_k, y_k, Z_k) \in \mathcal{N}_{\mu_k}(\Xi_k, \lambda_k)$). Similarly, we can show that:

$$
A^\top \bar{y} + \bar{Z} - \mu_k \bar{X} = 0.
$$

The previous two equalities imply that:

$$
\langle \bar{X}, \bar{Z} \rangle = \langle \bar{X}, -\mathcal{A}^* \bar{y} + \mu_k \bar{X} \rangle = \mu_k \langle \bar{X}, \bar{X} \rangle + \mu_k \bar{y}^\top \bar{y} \geq 0. \tag{34}
$$

On the other hand, using the last block equation of the Newton system (21), we have:

$$
E_k \bar{X} + F_k \bar{Z} = R_{\mu,k} + \frac{\mu_k}{\mu_0} E_k (\tilde{X} - X_{r_k}^* + \mu_0 \hat{X}) + \frac{\mu_k}{\mu_0} F_k (\tilde{Z} - Z_{r_k}^* + \mu_0 \hat{Z}),
$$

where $R_{\mu,k}$ is defined as in Lemma 4.6. Let S_k be defined as in Lemma 4.5. By multiplying both sides of the previous equation by $S_k^{-\frac{1}{2}}$, we get:

$$
D_k^{-T} \bar{X} + D_k \bar{Z} = S_k^{-\frac{1}{2}} R_{\mu,k} + \frac{\mu_k}{\mu_0} \left(D_k^{-T} (\tilde{X} - X_{r_k}^* + \mu_0 \hat{X}) + D_k (\tilde{Z} - Z_{r_k}^* + \mu_0 \hat{Z}) \right). \tag{35}
$$

But from (34) we know that $\langle \bar{X}, \bar{Z} \rangle \geq 0$ and hence:

$$
\| D_k^{-T} \bar{X} + D_k \bar{Z} \|_2^2 \geq \| D_k^{-T} \bar{X} \|_2^2 + \| D_k \bar{Z} \|_2^2.
$$

Combining (35) with the previous inequality, gives:

$$
\begin{aligned}
\| D_k^{-T} \bar{X} \|_2^2 \leq \Bigg\{ &\| S_k^{-\frac{1}{2}} R_{\mu,k} \|_2 + \frac{\mu_k}{\mu_0} \bigg(\| D_k^{-T} (\tilde{X} - X_{r_k}^* + \mu_0 \hat{X}) \|_2 \\
&+ \| D_k (\tilde{Z} - Z_{r_k}^* + \mu_0 \hat{Z}) \|_2 \bigg) \Bigg\}^2.
\end{aligned}
$$

We take square roots, use (33) and apply the triangular inequality, to get:

$$\|D_k^{-T}\boldsymbol{\Delta X}_k\|_2 \le \|S_k^{-\frac{1}{2}}\boldsymbol{R}_{\mu,k}\|_2 + \frac{\mu_k}{\mu_0}\Big(2\|D_k^{-T}(\tilde{\boldsymbol{X}} - \boldsymbol{X}_{r_k}^* + \mu_0\hat{\boldsymbol{X}})\|_2$$
$$+ \|D_k(\tilde{\boldsymbol{Z}} - \boldsymbol{Z}_{r_k}^* + \mu_0\hat{\boldsymbol{Z}})\|_2\Big). \tag{36}$$

We now proceed to bounding the terms in the right hand side of (36). A bound for the first term of the right hand side is given by Lemma 4.7, that is:

$$\|S_k^{-\frac{1}{2}}\boldsymbol{R}_{\mu,k}\|_2 = O(n^{\frac{1}{2}}\mu_k^{\frac{1}{2}}).$$

On the other hand, we have (from Lemma 4.5) that

$$\|D_k^{-T}\|_2 = O\left(\frac{n}{\mu_k^{\frac{1}{2}}}\right), \qquad \|D_k\|_2 = O\left(\frac{n}{\mu_k^{\frac{1}{2}}}\right).$$

Hence, using the previous bounds, as well as Lemmas 4.2, 4.3, and (32), we obtain:

$$2\frac{\mu_k}{\mu_0}\|D_k^{-T}(\tilde{\boldsymbol{X}} - \boldsymbol{X}_{r_k}^* + \mu_0\hat{\boldsymbol{X}})\|_2 + \frac{\mu_k}{\mu_0}\|D_k(\tilde{\boldsymbol{Z}} - \boldsymbol{Z}_{r_k}^* + \mu_0\hat{\boldsymbol{Z}})\|_2 = O(n^2\mu_k^{\frac{1}{2}}),$$

Combining all the previous bounds yields that $\|D_k^{-T}\boldsymbol{\Delta X}_k\|_2 = O(n^2\mu_k^{\frac{1}{2}})$. One can bound $\|D_k\boldsymbol{\Delta Z}_k\|_2$ in the same way. The latter is omitted for ease of presentation.

Furthermore, we have that:

$$\|\boldsymbol{\Delta X}_k\|_2 = \|D_kD_k^{-T}\boldsymbol{\Delta X}_k\|_2 \le \|D_k\|_2\|D_k^{-T}\boldsymbol{\Delta X}_k\|_2 = O(n^3).$$

Similarly, we can show that $\|\boldsymbol{\Delta Z}_k\|_2 = O(n^3)$. From the first block equation of the Newton system in (21), alongside Assumption 2, we can show that $\|\Delta y_k\|_2 = O(n^3)$.

Finally, using the previous bounds, as well as Lemma 4.6, we obtain the desired bound on $\|H_{P_k}(\Delta X_k\Delta Z_k)\|_F$, that is:

$$\|H_{P_k}(\Delta X_k\Delta Z_k)\|_F = O(n^4\mu_k),$$

which completes the proof. □

We can now prove (Lemmas 4.9–4.10) that at every iteration of Algorithm IP–PMM there exists a step-length $\alpha_k > 0$, using which, the new iterate satisfies the conditions required by the algorithm. The lower bound on any such step-length will later determine the polynomial complexity of the method. To that end, we assume the following notation:

$$\big(X_k(\alpha), y_k(\alpha), Z_k(\alpha)\big) \equiv (X_k + \alpha\Delta X_k, y_k + \alpha\Delta y_k, Z_k + \alpha\Delta Z_k).$$

Lemma 4.9 *Given Assumptions* 1, 2, *and by letting* $P_k(\alpha) = Z_k(\alpha)^{\frac{1}{2}}$, *there exists a step-length* $\alpha^* \in (0, 1)$, *such that for all* $\alpha \in [0, \alpha^*]$ *and for all iterations* $k \geq 0$ *of Algorithm* IP–PMM, *the following relations hold:*

$$\langle X_k + \alpha \Delta X_k, Z_k + \alpha \Delta Z_k \rangle \geq (1 - \alpha(1 - \beta_1))\langle X_k, Z_k \rangle, \tag{37}$$

$$\|H_{P_k(\alpha)}(X_k(\alpha)Z_k(\alpha)) - \mu_k(\alpha)\|_F \leq \gamma_\mu \mu_k(\alpha), \tag{38}$$

$$\langle X_k + \alpha \Delta X_k, Z_k + \alpha \Delta Z_k \rangle \leq (1 - \alpha(1 - \beta_2))\langle X_k, Z_k \rangle, \tag{39}$$

where, without loss of generality, $\beta_1 = \frac{\sigma_{\min}}{2}$ *and* $\beta_2 = 0.99$. *Moreover,* $\alpha^* \geq \frac{\kappa^*}{n^4}$ *for all* $k \geq 0$, *where* $\kappa^* > 0$ *is independent of* n, m.

Proof From Lemma 4.8, there exist constants $K_\Delta > 0$ and $K_{H\Delta} > 0$, independent of n and m, such that:

$$\langle \Delta X_k, \Delta Z_k \rangle = (D_k^{-T} \boldsymbol{\Delta X}_k)^\top (D_k \boldsymbol{\Delta Z}_k) \leq \|D_k^{-T} \boldsymbol{\Delta X}_k\|_2 \|D_k \boldsymbol{\Delta Z}_k\|_2 \leq K_\Delta^2 n^4 \mu_k,$$

$$\|H_{P_k}(\Delta X_k \Delta Z_k)\|_F \leq K_{H\Delta} n^4 \mu_k.$$

From the last block equation of the Newton system (20), we can show that:

$$\langle Z_k, \Delta X_k \rangle + \langle X_k, \Delta Z_k \rangle = (\sigma_k - 1)\langle X_k, Z_k \rangle. \tag{40}$$

The latter can also be obtained from (21), since we require $\mathsf{E}_{\mu,k} = 0$. Furthermore:

$$H_{P_k}(X_k(\alpha)Z_k(\alpha)) = (1 - \alpha)H_{P_k}(X_k Z_k) + \alpha \sigma_k \mu_k I_n + \alpha^2 H_{P_k}(\Delta X_k \Delta Z_k), \tag{41}$$

where $(X_{k+1}, y_{k+1}, Z_{k+1}) = (X_k + \alpha \Delta X_k, y_k + \alpha \Delta y_k, Z_k + \alpha \Delta Z_k)$.

We proceed by proving (37). Using (40), we have:

$$\langle X_k + \alpha \Delta X_k, Z_k + \alpha \Delta Z_k \rangle - (1 - \alpha(1 - \beta_1))\langle X_k, Z_k \rangle$$
$$= \langle X_k, Z_k \rangle + \alpha(\sigma_k - 1)\langle X_k, Z_k \rangle + \alpha^2 \langle \Delta X_k, \Delta Z_k \rangle - (1 - \alpha)\langle X_k, Z_k \rangle - \alpha\beta_1\langle X_k, Z_k \rangle$$
$$\geq \alpha(\sigma_k - \beta_1)\langle X_k, Z_k \rangle - \alpha^2 K_\Delta^2 n^4 \mu_k \geq \alpha \left(\frac{\sigma_{\min}}{2}\right) n\mu_k - \alpha^2 K_\Delta^2 n^4 \mu_k,$$

where we set (without loss of generality) $\beta_1 = \frac{\sigma_{\min}}{2}$. The most-right hand side of the previous inequality will be non-negative for every α satisfying:

$$\alpha \leq \frac{\sigma_{\min}}{2K_\Delta^2 n^3}.$$

In order to prove (38), we will use (41) and the fact that from the neighbourhood conditions we have that $\|H_{P_k}(X_k Z_k) - \mu_k\|_F \leq \gamma_\mu \mu_k$. For that, we use the result in [29, Lemma 4.2], stating that:

$$\|H_{P_k(\alpha)}(X_k(\alpha)Z_k(\alpha)) - \mu_k(\alpha)I_n\|_F \leq \|H_{P_k}(X_k(\alpha)Z_k(\alpha)) - \mu_k(\alpha)I_n\|_F.$$

By combining all the previous, we have:

$$
\begin{aligned}
&\|H_{P_k(\alpha)}(X_k(\alpha)Z_k(\alpha)) - \mu_k(\alpha)I_n\|_F - \gamma_\mu \mu_k(\alpha) \\
&\quad \leq \|H_{P_k}(X_k(\alpha)Z_k(\alpha)) - \mu_k(\alpha)I_n\|_F - \gamma_\mu \mu_k(\alpha) \\
&\quad = \|(1-\alpha)(H_{P_k}(X_k Z_k) - \mu_k I_n) + \alpha^2 H_{P_k}(\Delta X_k, \Delta Z_k) \\
&\qquad - \frac{\alpha^2}{n}\langle \Delta X_k, \Delta Z_k\rangle I_n\|_F - \gamma_\mu \mu_k(\alpha) \\
&\quad \leq (1-\alpha)\|H_{P_k}(X_k Z_k) - \mu_k I_n\|_F + \alpha^2 \mu_k\left(\frac{K_\Delta^2}{n} + K_{H\Delta}\right)n^4 \\
&\qquad - \gamma_\mu\left((1-\alpha)\mu_k + \alpha\sigma_k\mu_k + \frac{\alpha^2}{n}\langle \Delta X_k, \Delta Z_k\rangle\right) \\
&\quad \leq -\gamma_\mu \alpha \sigma_{\min}\mu_k + \alpha^2 \mu_k\left(\frac{2K_\Delta^2}{n} + K_{H\Delta}\right)n^4,
\end{aligned}
$$

where we used the neighbourhood conditions in (19), the equality $\mu_k(\alpha) = (1-\alpha)\mu_k + \alpha\sigma_k\mu_k + \frac{\alpha^2}{n}\langle \Delta X_k, \Delta Z_k\rangle$ (which can be derived from (40)), and the third block equation of the Newton system (21). The most-right hand side of the previous is non-positive for every α satisfying:

$$
\alpha \leq \frac{\sigma_{\min}\gamma_\mu}{\left(\frac{2K_\Delta^2}{n} + K_{H\Delta}\right)n^4}.
$$

Finally, to prove (39), we set (without loss of generality) $\beta_2 = 0.99$. We know, from Algorithm IP–PMM, that $\sigma_{\max} \leq 0.5$. With the previous two remarks in mind, we have:

$$
\begin{aligned}
&\frac{1}{n}\langle X_k + \alpha\Delta X_k, Z_k + \alpha\Delta Z_k\rangle - (1-0.01\alpha)\mu_k \\
&\quad \leq (1-\alpha)\mu_k + \alpha\sigma_k\mu_k + \alpha^2\frac{K_\Delta^2 n^4}{n}\mu_k - (1-0.01\alpha)\mu_k \\
&\quad \leq -0.99\alpha\mu_k + 0.5\alpha\mu_k + \alpha^2\frac{K_\Delta^2 n^4}{n}\mu_k \\
&\quad = -0.49\alpha\mu_k + \alpha^2\frac{K_\Delta^2 n^4}{n}\mu_k.
\end{aligned}
$$

The last term will be non-positive for every α satisfying:

$$
\alpha \leq \frac{0.49}{K_\Delta^2 n^3}.
$$

By combining all the previous bounds on the step-length, we have that (37)–(39) hold for every $\alpha \in (0, \alpha^*)$, where:

$$\alpha^* := \min \left\{ \frac{\sigma_{\min}}{2K_\Delta^2 n^3}, \; \frac{\sigma_{\min} \gamma_\mu}{\left(\frac{2K_\Delta^2}{n} + K_{H\Delta} \right) n^4}, \; \frac{0.49}{K_\Delta^2 n^3}, \; 1 \right\}. \tag{42}$$

Since $\alpha^* = \Omega\left(\frac{1}{n^4} \right)$, we know that there must exist a constant $\kappa^* > 0$, independent of n, m and of the iteration k, such that $\alpha^* \geq \frac{\kappa}{n^4}$, for all $k \geq 0$, and this completes the proof. $\qquad \square$

Lemma 4.10 *Given Assumptions 1, 2, and by letting $P_k(\alpha) = Z_k(\alpha)^{\frac{1}{2}}$, there exists a step-length $\bar{\alpha} \geq \frac{\bar{\kappa}}{n^4} \in (0, 1)$, where $\bar{\kappa} > 0$ is independent of n, m, such that for all $\alpha \in [0, \bar{\alpha}]$ and for all iterations $k \geq 0$ of Algorithm IP–PMM, if $(X_k, y_k, Z_k) \in \mathcal{N}_{\mu_k}(\Xi_k, \lambda_k)$, then letting:*

$$(X_{k+1}, y_{k+1}, Z_{k+1}) = (X_k + \alpha \Delta X_k, y_k + \alpha \Delta y_k, Z_k + \alpha \Delta Z_k), \; \mu_{k+1} = \frac{\langle X_{k+1}, Z_{k+1} \rangle}{n},$$

for any $\alpha \in (0, \bar{\alpha}]$, gives $(X_{k+1}, y_{k+1}, Z_{k+1}) \in \mathcal{N}_{\mu_{k+1}}(\Xi_{k+1}, \lambda_{k+1})$, where Ξ_k, and λ_k are updated as in Algorithm IP–PMM.

Proof Let α^* be given as in Lemma 4.9 (i.e. in (42)) such that (37)–(39) are satisfied. We would like to find the maximum $\bar{\alpha} \in (0, \alpha^*)$, such that:

$$(X_k(\alpha), y_k(\alpha), Z_k(\alpha)) \in \mathcal{N}_{\mu_k(\alpha)}(\Xi_k, \lambda_k), \; \text{for all } \alpha \in (0, \bar{\alpha}),$$

where $\mu_k(\alpha) = \frac{\langle X_k(\alpha), Z_k(\alpha) \rangle}{n}$. Let:

$$\tilde{r}_p(\alpha) = AX_k(\alpha) + \mu_k(\alpha)(y_k(\alpha) - \lambda_k) - \left(b + \frac{\mu_k(\alpha)}{\mu_0} \bar{b} \right), \tag{43}$$

and

$$\tilde{R}_d(\alpha) = A^\top y_k(\alpha) + Z_k(\alpha) - \mu_k(\alpha)(X_k(\alpha) - \Xi_k) - \left(C + \frac{\mu_k(\alpha)}{\mu_0} \bar{C} \right). \tag{44}$$

In other words, we need to find the maximum $\bar{\alpha} \in (0, \alpha^*)$, such that:

$$\|\tilde{r}_p(\alpha), \tilde{R}_d(\alpha)\|_2 \leq K_N \frac{\mu_k(\alpha)}{\mu_0}, \; \|\tilde{r}_p(\alpha), \tilde{R}_d(\alpha)\|_{\mathcal{S}} \leq \gamma_{\mathcal{S}} \rho \frac{\mu_k(\alpha)}{\mu_0}, \; \text{for all } \alpha \in (0, \bar{\alpha}). \tag{45}$$

If the latter two conditions hold, then $(X_k(\alpha), y_k(\alpha), Z_k(\alpha)) \in \mathcal{N}_{\mu_k(\alpha)}(\Xi_k, \lambda_k)$, for all $\alpha \in (0, \bar{\alpha})$. Then, if Algorithm IP–PMM updates Ξ_k, and λ_k, it does so only when similar conditions (as in (45)) hold for the new parameters. Indeed, notice that the estimates Ξ_k and λ_k are only updated if the last conditional of Algorithm IP–PMM is satisfied. But this is equivalent to saying that (45) is satisfied after setting $\Xi_k = X_k(\alpha)$

and $\lambda_k = y_k(\alpha)$. On the other hand, if the parameters are not updated, the new iterate lies in the desired neighbourhood because of (45), alongside (37)–(39).

We start by rearranging $\tilde{r}_p(\alpha)$. Specifically, we have that:

$$
\begin{aligned}
\tilde{r}_p(\alpha) &= A(X_k + \alpha \boldsymbol{\Delta X}_k) + \Big(\mu_k + \alpha(\sigma_k - 1)\mu_k \\
&\quad + \frac{\alpha^2}{n}\langle \Delta X_k, \Delta Z_k \rangle\Big)\Big((y_k + \alpha \Delta y_k - \lambda_k) - \frac{\bar{b}}{\mu_0}\Big) - b \\
&= \Big(AX_k + \mu_k(y_k - \lambda_k) - b - \frac{\mu_k}{\mu_0}\bar{b}\Big) + \alpha(A\boldsymbol{\Delta X}_k + \mu_k \Delta y_k) \\
&\quad + \Big(\alpha(\sigma_k - 1)\mu_k + \frac{\alpha^2}{n}\langle \Delta X_k, \Delta Z_k\rangle\Big)\Big((y_k - \lambda_k + \alpha \Delta y_k) - \frac{\bar{b}}{\mu_0}\Big) \\
&= \frac{\mu_k}{\mu_0}\tilde{b}_k + \alpha\Big(b - AX_k - \sigma_k \mu_k \Big((y_k - \lambda_k) - \frac{\bar{b}}{\mu_0}\Big) + \epsilon_{p,k} \\
&\quad + \mu_k\big((y_k - \lambda_k) - \frac{\bar{b}}{\mu_0}\big) - \mu_k\Big((y_k - \lambda_k) - \frac{\bar{b}}{\mu_0}\Big)\Big) \\
&\quad + \Big(\alpha(\sigma_k - 1)\mu_k + \frac{\alpha^2}{n}\langle \Delta X_k, \Delta Z_k\rangle\Big)\Big((y_k - \lambda_k + \alpha \Delta y_k) - \frac{\bar{b}}{\mu_0}\Big),
\end{aligned}
$$

where we used the definition of $\tilde{b}_k$ in the neighbourhood conditions in (19), and the second block equation in (21). By using again the neighbourhood conditions, and then by deleting the opposite terms in the previous equation, we obtain:

$$
\begin{aligned}
\tilde{r}_p(\alpha) = \;& (1 - \alpha)\frac{\mu_k}{\mu_0}\tilde{b}_k + \alpha\epsilon_{p,k} + \alpha^2(\sigma_k - 1)\mu_k \Delta y_k \\
& + \frac{\alpha^2}{n}\langle \Delta X_k, \Delta Z_k\rangle\Big(y_k - \lambda_k + \alpha \Delta y_k - \frac{\bar{b}}{\mu_0}\Big).
\end{aligned}
\tag{46}
$$

Similarly, we can show that:

$$
\begin{aligned}
\tilde{\boldsymbol{R}}_d(\alpha) = \;& (1 - \alpha)\frac{\mu_k}{\mu_0}\tilde{\boldsymbol{C}}_k + \alpha\mathsf{E}_{d,k} - \alpha^2(\sigma_k - 1)\mu_k \boldsymbol{\Delta X}_k \\
& - \frac{\alpha^2}{n}\langle \Delta X_k, \Delta Z_k\rangle\Big(X_k - \boldsymbol{\Xi}_k + \alpha \boldsymbol{\Delta X}_k + \frac{1}{\mu_0}\bar{\boldsymbol{C}}\Big).
\end{aligned}
\tag{47}
$$

Recall (Lemma 4.8) that $\langle \Delta X_k, \Delta Z_k \rangle \leq K_\Delta^2 n^4 \mu_k$, and define the following quantities

$$
\begin{aligned}
\xi_2 =\; & \mu_k \|(\Delta y_k, \boldsymbol{\Delta X}_k)\|_2 + K_\Delta^2 n^3 \mu_k \bigg(\|(y_k - \lambda_k, X_k - \boldsymbol{\Xi}_k)\|_2 \\
& + \alpha^* \|(\Delta y_k, \boldsymbol{\Delta X}_k)\|_2 + \frac{1}{\mu_0} \|(\bar{b}, \bar{C})\|_2 \bigg), \\
\xi_{\mathcal{S}} =\; & \mu_k \|(\Delta y_k, \boldsymbol{\Delta X}_k)\|_{\mathcal{S}} + K_\Delta^2 n^3 \mu_k \bigg(\|(y_k - \lambda_k, X_k - \boldsymbol{\Xi}_k)\|_{\mathcal{S}} \\
& + \alpha^* \|(\Delta y_k, \boldsymbol{\Delta X}_k)\|_{\mathcal{S}} + \frac{1}{\mu_0} \|(\bar{b}, \bar{C})\|_{\mathcal{S}} \bigg),
\end{aligned}
\tag{48}
$$

where α^* is given by (42). Using the definition of the starting point in (18), as well as results in Lemmas 4.4, 4.8, we can observe that $\xi_2 = O(n^4 \mu_k)$. On the other hand, using Assumption 2, we know that for every pair $(r_1, \boldsymbol{R}_2) \in \mathbb{R}^{m+n^2}$ (where $R_2 \in \mathbb{R}^{n \times n}$ is an arbitrary matrix), if $\|(r_1, \boldsymbol{R}_2)\|_2 = \Theta(f(n))$, where $f(\cdot)$ is a positive polynomial function of n, then $\|(r_1, R_2)\|_{\mathcal{S}} = \Theta(f(n))$. Hence, we have that $\xi_{\mathcal{S}} = O(n^4 \mu_k)$. Using the quantities in (48), equations (46), (47), as well as the neighbourhood conditions, we have that:

$$
\begin{aligned}
\|\tilde{r}_p(\alpha), \tilde{\boldsymbol{R}}_d(\alpha)\|_2 &\leq (1 - \alpha) K_N \frac{\mu_k}{\mu_0} + \alpha \mu_k \|(\epsilon_{p,k}, \mathsf{E}_{d,k})\|_2 + \alpha^2 \mu_k \xi_2, \\
\|\tilde{r}_p(\alpha), \tilde{\boldsymbol{R}}_d(\alpha)\|_{\mathcal{S}} &\leq (1 - \alpha) \gamma_{\mathcal{S}} \rho \frac{\mu_k}{\mu_0} + \alpha \mu_k \|(\epsilon_{p,k}, \mathsf{E}_{d,k})\|_{\mathcal{S}} + \alpha^2 \mu_k \xi_{\mathcal{S}},
\end{aligned}
$$

for all $\alpha \in (0, \alpha^*)$, where α^* is given by (42) and the error occurring from the inexact solution of (20), $(\epsilon_{p,k}, \mathsf{E}_{d,k})$, satisfies (22). From (37), we know that:

$$
\mu_k(\alpha) \geq (1 - \alpha(1 - \beta_1)) \mu_k, \text{ for all } \alpha \in (0, \alpha^*).
$$

By combining the last three inequalities, using (22) and setting $\beta_1 = \frac{\sigma_{\min}}{2}$, we obtain

$$
\|\tilde{r}_p(\alpha), \tilde{\boldsymbol{R}}_d(\alpha)\|_2 \leq \frac{\mu_k(\alpha)}{\mu_0} K_N, \text{ for all } \alpha \in \left(0, \min\left\{\alpha^*, \frac{\sigma_{\min} K_N}{4 \xi_2 \mu_0}\right\} \right].
$$

Similarly,

$$
\|\tilde{r}_p(\alpha), \tilde{\boldsymbol{R}}_d(\alpha)\|_{\mathcal{S}} \leq \frac{\mu_k(\alpha)}{\mu_0} \gamma_{\mathcal{S}} \rho, \text{ for all } \alpha \in \left(0, \min\left\{\alpha^*, \frac{\sigma_{\min} \gamma_{\mathcal{S}} \rho}{4 \xi_{\mathcal{S}} \mu_0}\right\} \right].
$$

Hence, we have:

$$
\bar{\alpha} := \min \left\{ \alpha^*, \frac{\sigma_{\min} K_N}{4 \xi_2 \mu_0}, \frac{\sigma_{\min} \gamma_{\mathcal{S}} \rho}{4 \xi_{\mathcal{S}} \mu_0} \right\}.
\tag{49}
$$

Since $\bar{\alpha} = \Omega\left(\frac{1}{n^4}\right)$, we know that there must exist a constant $\bar{\kappa} > 0$, independent of n, m and of the iteration k, such that $\bar{\alpha} \geq \frac{\kappa}{n^4}$, for all $k \geq 0$, and this completes the proof. $\quad\square$

The following theorem summarizes our results.

Theorem 4.1 *Given Assumptions* 1, 2, *the sequence $\{\mu_k\}$ generated by Algorithm* IP–PMM *converges Q-linearly to zero, and the sequences of regularized residual norms*

$$\left\{\left\|AX_k + \mu_k(y_k - \lambda_k) - b - \frac{\mu_k}{\mu_0}\bar{b}\right\|_2\right\} \; and \; \left\{\left\|A^\top y_k + Z_k - \mu_k(X_k - \Xi_k) - C - \frac{\mu_k}{\mu_0}\bar{C}\right\|_2\right\}$$

converge R-linearly to zero.

Proof From (39) we have that:

$$\mu_{k+1} \leq (1 - 0.01\alpha_k)\mu_k,$$

while, from (49), we know that $\forall\, k \geq 0, \exists\, \bar{\alpha} \geq \frac{\bar{\kappa}}{n^4}$ such that $\alpha_k \geq \bar{\alpha}$. Hence, we can easily see that $\mu_k \to 0$. On the other hand, from the neighbourhood conditions, we know that for all $k \geq 0$:

$$\left\|AX_k + \mu_k(y_k - \lambda_k) - b - \frac{\mu_k}{\mu_0}\bar{b}\right\|_2 \leq K_N \frac{\mu_k}{\mu_0}$$

and

$$\left\|A^\top y_k + Z_k - \mu_k(X_k - \Xi_k) - C - \frac{\mu_k}{\mu_0}\bar{C}\right\|_2 \leq K_N \frac{\mu_k}{\mu_0}.$$

This completes the proof. $\quad\square$

The polynomial complexity of Algorithm IP–PMM is established in the following theorem.

Theorem 4.2 *Let $\varepsilon \in (0, 1)$ be a given error tolerance. Choose a starting point for Algorithm* IP–PMM *as in* (18)*, such that $\mu_0 \leq \frac{K}{\varepsilon^\omega}$ for some positive constants K, ω. Given Assumptions* 1 *and* 2*, there exists an index $k_0 \geq 0$ with:*

$$k_0 = O\left(n^4 \left|\log \frac{1}{\varepsilon}\right|\right)$$

such that the iterates $\{(X_k, y_k, Z_k)\}$ generated from Algorithm IP–PMM *satisfy:*

$$\mu_k \leq \varepsilon, \quad \text{for all } k \geq k_0.$$

Proof The proof can be found in [19, Theorem 3.8]. $\quad\square$

Finally, we present the global convergence guarantee of Algorithm IP–PMM.

Theorem 4.3 *Suppose that Algorithm* IP–PMM *terminates when a limit point is reached. Then, if Assumptions 1 and 2 hold, every limit point of* $\{(X_k, y_k, Z_k)\}$ *determines a primal-dual solution of the non-regularized pair* (P)–(D).

Proof From Theorem 4.1, we know that $\{\mu_k\} \to 0$, and hence, there exists a subsequence $\mathcal{K} \subseteq \mathbb{N}$, such that:

$$\left\{ AX_k + \mu_k(y_k - \lambda_k) - b - \frac{\mu_k}{\mu_0}\bar{b} \right\}_{\mathcal{K}} \to 0,$$

$$\left\{ A^\top y_k + Z_k - \mu_k(X_k - \Xi_k) - C - \frac{\mu_k}{\mu_0}\bar{C} \right\}_{\mathcal{K}} \to 0.$$

However, since Assumptions 1 and 2 hold, we know from Lemma 4.4 that $\{(X_k, y_k, Z_k)\}$ is a bounded sequence. Hence, we obtain that:

$$\{AX_k - b\}_{\mathcal{K}} \to 0, \quad \{A^\top y_k + Z_k - C\}_{\mathcal{K}} \to 0.$$

One can readily observe that the limit point of the algorithm satisfies the optimality conditions of (P)–(D), since $\langle X_k, Z_k \rangle \to 0$ and $X_k, \ Z_k \in \mathcal{S}_+^n$. $\qquad \square$

Remark 4.1 As mentioned at the end of Sect. 3, we do not study the conditions under which one can guarantee that $X_k - \Xi_k \to 0$ and $y_k - \lambda_k \to 0$, although this could be possible. This is because the method is shown to converge globally even if this is not the case. Indeed, notice that if one were to choose $X_0 = 0$ and $\lambda_0 = 0$, and simply ignore the last conditional statement of Algorithm IP–PMM, the convergence analysis established in this section would still hold. In this case, the method would be interpreted as an interior point-quadratic penalty method, and we could consider the regularization as a diminishing primal-dual Tikhonov regularizer (i.e. a variant of the regularization proposed in [23]).

5 A Sufficient Condition for Strong Duality

We now drop Assumptions 1, 2, in order to analyze the behaviour of the algorithm when solving problems that are strongly (or weakly) infeasible, problems for which strong duality does not hold (weakly feasible), or problems for which the primal or the dual solution is not attained. For a formal definition and a comprehensive study of the previous types of problems we refer the reader to [14], and the references therein. Below we provide a well-known result, stating that strong duality holds if and only if there exists a KKT point.

Proposition 1 *Let* (P)–(D) *be given. Then,* val(P) $\geq$ val(D), *where* val($\cdot$) *denotes the optimal objective value of a problem. Moreover,* val(P) $=$ val(D) *and* (X^*, y^*, Z^*) *is an optimal solution for* (P)–(D), *if and only if* (X^*, y^*, Z^*) *satisfies the KKT optimality conditions in* (1).

Proof This is a well-known fact, the proof of which can be found in [24, Proposition 2.1]. □

Let us employ the following two premises:

Premise 1 *During the iterations of Algorithm* IP–PMM*, the sequences* $\{\|y_k - \lambda_k\|_2\}$ *and* $\{\|X_k - \Xi_k\|_F\}$*, remain bounded.*

Premise 2 *There does not exist a primal-dual triple, satisfying the KKT conditions in* (1) *associated with the primal-dual pair* (P)–(D)*.*

The following analysis extends the result presented in [19, Section 4], and is based on the developments in [9, Sections 10 & 11]. In what follows, we show that Premises 1 and 2 are contradictory. In other words, if Premise 2 holds (which means that strong duality does not hold for the problem under consideration), then Premise 1 cannot hold, and hence Premise 1 is a sufficient condition for strong duality (and its negation is a necessary condition for Premise 2). We show that if Premise 1 holds, then the algorithm converges to an optimal solution. If not, however, it does not necessarily mean that the problem under consideration is infeasible. For example, this could happen if either (P) or (D) is strongly infeasible, weakly infeasible, and in some cases even if either of the problems is weakly feasible (e.g. see [14,24]). As we discuss later, the knowledge that Premise 1 does not hold could be useful in detecting pathological problems.

Lemma 5.1 *Given Premise* 1*, and by assuming that* $\langle X_k, Z_k \rangle > \varepsilon$*, for some* $\varepsilon > 0$*, for all iterations k of Algorithm* IP–PMM*, the Newton direction produced by* (21) *is uniformly bounded by a constant dependent only on n and/or m.*

Proof The proof is omitted since it follows exactly the developments in [9, Lemma 10.1]. We notice that the regularization terms (blocks (1,1) and (2,2) in the Jacobian matrix in (21)) depend on μ_k which by assumption is always bounded away from zero: $\mu_k \geq \frac{\epsilon}{n}$. □

In the following Lemma, we prove by contradiction that the parameter μ_k of Algorithm IP–PMM converges to zero, given that Premise 1 holds.

Lemma 5.2 *Given Premise* 1*, and a sequence* $(X_k, y_k, Z_k) \in \mathcal{N}_{\mu_k}(\Xi_k, \lambda_k)$ *produced by Algorithm* IP–PMM*, the sequence* $\{\mu_k\}$ *converges to zero.*

Proof Assume, by virtue of contradiction, that $\mu_k > \varepsilon > 0$, for all $k \geq 0$. Then, we know (from Lemma 5.1) that the Newton direction obtained by the algorithm at every iteration, after solving (21), will be uniformly bounded by a constant dependent only on n, that is, there exists a positive constant $K^\dagger$, such that $\|(\Delta X_k, \Delta y_k, \Delta Z_k)\|_2 \leq K^\dagger$. We define $\tilde{r}_p(\alpha)$ and $\tilde{\boldsymbol{R}}_d(\alpha)$ as in (43) and (44), respectively, for which we know that equalities (46) and (47) hold, respectively. Take any $k \geq 0$ and define the following functions:

$$f_1(\alpha) := \langle X_k(\alpha), Z_k(\alpha) \rangle - \left(1 - \alpha \left(1 - \frac{\sigma_{\min}}{2}\right)\right) \langle X_k, Z_k \rangle,$$

$$f_2(\alpha) := \gamma_\mu \mu_k(\alpha) - \|H_{P_k(\alpha)}(X_k(\alpha)Z_k(\alpha)) - \mu_k(\alpha)\|_F$$

$$f_3(\alpha) := (1 - 0.01\alpha)\langle X_k, Z_k \rangle - \langle X_k(\alpha), Z_k(\alpha) \rangle,$$

$$g_2(\alpha) := \frac{\mu_k(\alpha)}{\mu_0} K_N - \|(\tilde{r}_p(\alpha), \tilde{\boldsymbol{R}}_d(\alpha))\|_2,$$

where $\mu_k(\alpha) = \frac{\langle X_k + \alpha \Delta X_k, Z_k + \alpha \Delta Z_k \rangle}{n}$, $(X_k(\alpha), y_k(\alpha), Z_k(\alpha)) = (X_k + \alpha \Delta X_k, y_k + \alpha \Delta y_k, Z_k + \alpha \Delta Z_k)$. We would like to show that there exists $\alpha^* > 0$, such that:

$$f_1(\alpha) \geq 0, \quad f_2(\alpha) \geq 0, \quad f_3(\alpha) \geq 0, \quad g_2(\alpha) \geq 0, \text{ for all } \alpha \in (0, \alpha^*].$$

These conditions model the requirement that the next iteration of Algorithm IP–PMM must lie in the updated neighbourhood $\mathcal{N}_{\mu_{k+1}}(\varXi_k, \lambda_k)$ (notice however that the restriction with respect to the semi-norm defined in (17) is not required here, and indeed it cannot be incorporated unless $\mathrm{rank}(A) = m$). Since Algorithm IP–PMM updates the parameters λ_k, $\varXi_k$ only if the selected new iterate belongs to the new neighbourhood, defined using the updated parameters (again, ignoring the restrictions with respect to the semi-norm), it suffices to show that $(X_{k+1}, y_{k+1}, Z_{k+1}) \in \mathcal{N}_{\mu_{k+1}}(\varXi_k, \lambda_k)$.

Proving the existence of $\alpha^* > 0$, such that each of the aforementioned functions is positive, follows exactly the developments in Lemmas 4.9–4.10, with the only difference being that the bounds on the directions are not explicitly specified in this case. Using the same methodology as in aforementioned lemmas, while keeping in mind our assumption, namely $\langle X_k, Z_k \rangle > \varepsilon$, we can show that:

$$\alpha^* := \min \left\{ 1, \frac{\sigma_{\min}\epsilon}{2(K^\dagger)^2}, \frac{(1 - \gamma_\mu)\sigma_{\min}\gamma_\mu\epsilon}{2n(K^\dagger)^2}, \frac{0.49\epsilon}{2(K^\dagger)^2}, \frac{\sigma_{\min}K_N\epsilon}{4\mu_0(\xi_2)} \right\}, \tag{50}$$

where ξ_2 is a bounded constant, defined as in (48), and dependent on $K^\dagger$. However, using the inequality:

$$\mu_{k+1} \leq (1 - 0.01\alpha)\mu_k, \text{ for all } \alpha \in [0, \alpha^*]$$

we get that $\mu_k \to 0$, which contradicts our assumption that $\mu_k > \varepsilon$, $\forall\, k \geq 0$, and completes the proof. $\qquad\square$

Finally, using the following Theorem, we derive a necessary condition for lack of strong duality.

Theorem 5.1 *Given Premise* 2*, i.e. there does not exist a KKT triple for the pair* (P)–(D)*, then Premise* 1 *fails to hold.*

Proof By virtue of contradiction, let Premise 1 hold. In Lemma 5.2, we proved that given Premise 1, Algorithm IP–PMM produces iterates that belong to the neighbourhood (19) and $\mu_k \to 0$. But from the neighbourhood conditions we can observe that:

$$\left\| AX_k + \mu_k(y_k - \lambda_k) - b - \frac{\mu_k}{\mu_0}\bar{b} \right\|_2 \leq K_N \frac{\mu_k}{\mu_0},$$

and

$$\left\| A^\top y_k + Z_k - \mu_k(X_k - \varXi_k) - C - \frac{\mu_k}{\mu_0}\bar{C} \right\|_2 \leq K_N \frac{\mu_k}{\mu_0}.$$

Hence, we can choose a sub-sequence $\mathcal{K} \subseteq \mathbb{N}$, for which:

$$\{A\boldsymbol{X}_k + \mu_k(y_k - \lambda_k) - b - \frac{\mu_k}{\mu_0}\bar{b}\}_{\mathcal{K}} \to 0, \text{ and}$$

$$\{A^\top y_k + \boldsymbol{Z}_k - \mu_k(\boldsymbol{X}_k - \boldsymbol{\Xi}_k) - \boldsymbol{C} - \frac{\mu_k}{\mu_0}\bar{\boldsymbol{C}}\}_{\mathcal{K}} \to 0.$$

But since $\|y_k - \lambda_k\|_2$ and $\|\boldsymbol{X}_k - \boldsymbol{\Xi}_k\|_F$ are bounded, while $\mu_k \to 0$, we have that:

$$\{A\boldsymbol{X}_k - b\}_{\mathcal{K}} \to 0, \ \{\boldsymbol{C} - A^\top y_k - \boldsymbol{Z}_k\}_{\mathcal{K}} \to 0, \text{ and } \{\langle \boldsymbol{X}_k, \boldsymbol{Z}_k\rangle\}_{\mathcal{K}} \to 0.$$

This contradicts Premise 2, i.e. that the pair (P)–(D) does not have a KKT triple, and completes the proof.   □

In the previous Theorem, we proved that the negation of Premise 1 is a necessary condition for Premise 2. Nevertheless, this does not mean that the condition is also sufficient. In order to obtain a more reliable algorithmic test for lack of strong duality, we have to use the properties of Algorithm IP–PMM. In particular, we can notice that if there does not exist a KKT point, then the PMM sub-problems will stop being updated after a finite number of iterations. In that case, we know from Theorem 5.1 that the sequence $\|(\boldsymbol{X}_k - \boldsymbol{\Xi}_k, y_k - \lambda_k)\|_2$ will grow unbounded. Hence, we can define a maximum number of iterations per PMM sub-problem, say $k_\dagger > 0$, as well as a very large constant $K_\dagger$. Then, if $\|(\boldsymbol{X}_k - \boldsymbol{\Xi}_k, y_k - \lambda_k)\|_2 > K_\dagger$ and $k_{in} \geq k_\dagger$ (where k_{in} counts the number of IPM iterations per PMM sub-problem), the algorithm is terminated with a guess that there does not exist a KKT point for (P)–(D).

Remark 5.1 Let us notice that the analysis in Sect. 4 employs the standard assumptions used when analyzing a non-regularized IPM. However, the method could still be useful if these assumptions were not met. Indeed, if for example the constraint matrix was not of full row rank, one could still prove global convergence of the method, using the methodology employed in this section by assuming that Premise 1 holds and Premise 2 does not (or by following the developments in [9]). Furthermore, in practice the method would not encounter any numerical issues with the inversion of the Newton system (see [2]). Nevertheless, showing polynomial complexity in this case is still an open problem. The aim of this work is to show that under the standard Assumptions 1, 2, Algorithm IP–PMM is able to enjoy polynomial complexity, while having to solve better conditioned systems than those solved by standard IPMs at each iteration, thus ensuring better stability (and as a result better robustness and potentially better efficiency).

6 Conclusions

In this paper we developed and analyzed an interior point-proximal method of multipliers, suitable for solving linear positive semi-definite programs, without requiring the exact solution of the associated Newton systems. By generalizing appropriately some previous results on convex quadratic programming, we show that IP-PMM inherits

the polynomial complexity of standard non-regularized IPM schemes when applied to SDP problems, under standard assumptions, while having to approximately solve better-conditioned Newton systems, compared to their non-regularized counterparts. Furthermore, we provide a tuning for the proximal penalty parameters based on the well-studied barrier parameter, which can be used to guide any subsequent implementation of the method. Finally, we study the behaviour of the algorithm when applied to problems for which no KKT point exists, and give a necessary condition which can be used to construct detection mechanisms for identifying such pathological cases.

A future research direction would be to construct a robust and efficient implementation of the method, which should utilize some Krylov subspace solver alongside an appropriate preconditioner for the solution of the associated linear systems. Given previous implementations of IP-PMM for other classes of problems appearing in the literature, we expect that the theory can successfully guide the implementation, yielding a competitive, robust, and efficient method.

References

1. Altman, A., Gondzio, J.: Regularized symmetric indefinite systems in interior point methods for linear and quadratic optimization. Optim. Methods Softw. **11**(1–4), 275–302 (1999). https://doi.org/10.1080/10556789908805754
2. Armand, P., Benoist, J.: Uniform boundedness of the inverse of a Jacobian matrix arising in regularized interior-point methods. Math. Program. **137**, 587–592 (2013). https://doi.org/10.1007/s10107-011-0498-3
3. Balakrishnan, V., Wang, F.: Handbook of Semidefinite Programming. In: Wolkowicz, H., Saigal, R., Vandenberghe, L. (eds.) Handbook of Semidefinite Programming: Theory, Algorithms and Applications. International Series in Operations Research & Management Science, vol. 27, pp. 421–441. Springer, Boston (2000). https://doi.org/10.1007/978-1-4615-4381-7_14
4. Bellavia, S., Gondzio, J., Porcelli, M.: A relaxed interior point method for low-rank semidefinite programming problems. J. Sci. Comput. (2019). https://doi.org/10.1007/s10915-021-01654-1
5. Bellavia, S., Gondzio, J., Porcelli, M.: An inexact dual logarithmic barrier method for solving sparse semidefinite programs. Math. Program. **178**, 109–143 (2019). https://doi.org/10.1007/s10107-018-1281-5
6. Ben-Tal, A., Nemirovski, A.: Robust convex optimization. Math. Oper. Res. **23**(4), 769–805 (1998). https://doi.org/10.1287/moor.23.4.769
7. Bergamaschi, L., Gondzio, J., Martínez, A., Pearson, J.W., Pougkakiotis, S.: A new preconditioning approach for an interior point-proximal method of multipliers for linear and convex quadratic programming. Numer. Linear Algebra Appl. **28**(4), e2361 (2021). https://doi.org/10.1002/nla.2361
8. De Simone, V., di Serafino, D., Gondzio, J., Pougkakiotis, S., Viola, M.: Sparse approximations with interior point methods. arXiv:2102.13608v2 [math.OC] (2021)
9. Dehghani, A., Goffin, J.L., Orban, D.: A Primal-dual regularized interior-point method for semidefinite programming. Optim. Methods Softw. **32**(1), 193–219 (2017). https://doi.org/10.1080/10556788.2016.1235708

10. Gondzio, J., Pougkakiotis, S., Pearson, J.W.: General-purpose preconditioning for regularized interior point methods. arXiv:2107.06822 [math.OC] (2021)
11. Gu, M.: Primal-dual interior-point methods for semidefinite progamming in finite precision. SIAM J. Optim. **10**, 462–502 (2000). https://doi.org/10.1137/S105262349731950X
12. Jiang, X., Vandenberghe, L.: Bregman primal-dual first-order method and application to sparse semidefinite programming. http://www.optimization-online.org/DB_HTML/2020/03/7702.html (2020)
13. Lavaei, J., Low, S.H.: Zero duality gap in optimal power flow problem. IEEE Trans. Power Syst. **27**(1), 92–107 (2012). https://doi.org/10.1109/TPWRS.2011.2160974
14. Liu, M., Pataki, G.: Exact duals and short certificates of infeasibility and weak infeasibility in conic linear programming. Math. Program. **167**, 435–480 (2018). https://doi.org/10.1007/s10107-017-1136-5
15. Mizuno, S., Jarre, F.: Global and polynomial-time convergence of an infeasible-interior-point algorithm using inexact computation. Math. Program. **84**, 105–122 (1999). https://doi.org/10.1007/s10107980020a
16. MOSEK: The MOSEK Optimization Software. http://www.mosek.com, v9.2 (2020)
17. Nesterov, Y., Nemirovskii, A.S.: Interior Point Polynomial Methods in Convex Programming: Theory and Algorithms. SIAM Publications, Philadelphia (1994). https://doi.org/10.1137/1.9781611970791
18. Pougkakiotis, S., Gondzio, J.: Dynamic non-diagonal regularization in interior point methods for linear and convex quadratic programming. J. Optim. Theory Appl. **181**, 905–945 (2019). https://doi.org/10.1007/s10957-019-01491-1
19. Pougkakiotis, S., Gondzio, J.: An interior point-proximal method of multipliers for convex quadratic programming. Comput. Optim. Appl. **78**, 307–351 (2021). https://doi.org/10.1007/s10589-020-00240-9
20. Rockafellar, R.T.: Augmented Lagrangians and applications of the proximal point algorithm in convex programming. Math. Oper. Res. **1**(2), 97–116 (1976). https://doi.org/10.1287/moor.1.2.97
21. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**(5), 877–898 (1976). https://doi.org/10.1137/0314056
22. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1996). https://doi.org/10.1515/9781400873173
23. Saunders, M., Tomlin, J.A.: Solving regularized linear programs using barrier methods and KKT systems. Technical Report SOL 96-4, Systems Optimization Laboratory, Department of Operational Research, Stanford University, Stanford, CA 94305, USA (1996)
24. Shapiro, A., Scheinberg, K.: Duality and optimality conditions. In: Wolkowicz, H., Saigal, R., Vandenberghe, L. (eds.) Handbook of Semidefinite Programming: Theory, Algorithms and Applications. International Series in Operations Research & Management Science, vol. 27, pp. 67–110. Springer, Boston (2000)
25. Souto, M., Garcia, J.D., Veiga, A.: Exploiting low-rank structure in semidefinite programming by approximate operator splitting. Optimization (2020). https://doi.org/10.1080/02331934.2020.1823387
26. Todd, M.J.: A study of search directions in primal-dual interior-point methods for semidefinite programming. Optim. Methods Softw. **11**, 545–581 (1999). https://doi.org/10.1080/10556789908805745
27. Vandenberghe, L., Boyd, S.: Applications of semidefinite programming. Appl. Numer. Math. **29**(3), 283–299 (1999). https://doi.org/10.1016/S0168-9274(98)00098-1
28. Weldeyesus, A.G., Gondzio, J., He, L., Gilbert, M., Sheperd, P., Tyas, A.: Adaptive solution of truss layout optimization problems with global stability constraints. Struct. Multidiscip. Optim. **60**, 2093–2111 (2019). https://doi.org/10.1007/s00158-019-02312-9
29. Zhang, Y.: On extending some primal-dual interior-point algorithms from linear programming to semidefinite programming. SIAM J. Optim. **8**, 365–386 (1998). https://doi.org/10.1137/S1052623495296115
30. Zhao, X.Y., Sun, D., Toh, K.C.: A Newton-CG augmented Lagrangian method for semidefinite programming. SIAM J. Optim. **20**(4), 1737–1765 (2010). https://doi.org/10.1137/080718206
31. Zhou, G., Toh, K.C.: Polynomiality of an inexact infeasible interior point algorithm for semidefinite programming. Math. Program. **99**(A), 261–282 (2004). https://doi.org/10.1007/s10107-003-0431-5

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