When Does Primal Interior Point Method Beat Primal-dual in Linear Optimization?

Wenzhi Gao[*1], Huikang Liu[†2], Yinyu Ye[‡1,3], and Madeleine Udell[§1,3]

[1]ICME, Stanford University
[2]Antai School of Economics and Management, Shanghai Jiaotong University
[3]Management Science and Engineering, Stanford University

November 26, 2024

Abstract

The primal-dual interior point method (IPM) is widely regarded as the most efficient IPM variant for linear optimization. In this paper, we demonstrate that the improved stability of the pure primal IPM can allow speedups relative to a primal-dual solver, particularly as the IPM approaches convergence. The stability of the primal scaling matrix makes it possible to accelerate each primal IPM step using fast preconditioned iterative solvers for the normal equations. Crucially, we identify properties of the central path that make it possible to stabilize the normal equations. Experiments on benchmark datasets demonstrate the efficiency of primal IPM and showcase its potential for practical applications in linear optimization and beyond.

1 Introduction

We consider the standard-form primal and dual linear programs (LP)

$$
\begin{array}{ll}
\text{(P)} \quad \underset{\mathbf{x}}{\text{minimize}} & \langle \mathbf{c}, \mathbf{x} \rangle \\
\text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b} \\
& \mathbf{x} \geq \mathbf{0}
\end{array}
\qquad
\begin{array}{ll}
\text{(D)} \quad \underset{(\mathbf{y},\mathbf{s})}{\text{maximize}} & \langle \mathbf{b}, \mathbf{y} \rangle \\
\text{subject to} & \mathbf{A}^\top \mathbf{y} + \mathbf{s} = \mathbf{c} \\
& \mathbf{s} \geq \mathbf{0}
\end{array}
$$

with primal and dual feasible region $\mathcal{F}_p := \{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$ and $\mathcal{F}_d := \{(\mathbf{y},\mathbf{s}) : \mathbf{A}^\top \mathbf{y} + \mathbf{s} = \mathbf{c}, \mathbf{s} \geq \mathbf{0}\}$.

Since Karmarkar developed the first polynomial time interior point method [24], a substantial body of literature has explored both the theoretical [42, 44, 32] and practical aspects [2, 1, 25, 29] of IPMs. Over the years, IPM has evolved into a mature technology applicable to both linear and nonlinear optimization problems [27].

IPM was first introduced in its primal form, known as primal potential reduction [24], which operates solely in the primal space. However, with the evolution of IPM, several variants have emerged [32] that demonstrate improved robustness and efficiency compared to the original primal methods. Among these, the primal-dual IPM has stood the test of time and is widely regarded as the most efficient variant for linear and conic problems [18]. One of the few exceptions is in semidefinite programming (SDP), where the dual potential

---

[*]gwz@stanford.edu
[†]hkliu2014@gmail.com
[‡]yyye@stanford.edu
[§]udell@stanford.edu

reduction algorithm [5, 16] outperforms primal-dual methods in sparse SDPs. Despite its initial significance, primal IPM now receives less attention in linear optimization. It is, therefore, interesting to ask:

Are there problems on which primal IPM outperforms primal-dual IPM?

This paper provides a positive answer to the question by clarifying the spectral properties of the primal normal matrices: in particular, its stability near convergence. We demonstrate that this enhanced stability can be exploited to speed up later iterations of an IPM solver using the primal IPM iteration.

Contributions.

- We show that the normal matrix of the primal IPM is stable: when primal IPM converges, the normal matrix converges and need not become degenerate. This property reduces the cost of solving the normal equations using a preconditioned iterative method such as conjugate gradient. As a by-product, we obtain an $\mathcal{O}(n^{3.5} + n^{2.5} \log^2(1/\varepsilon))$ complexity result for primal non-degenerate LPs.

- We develop a hybrid primal/primal-dual IPM solver, using primal IPM to speed up later iterations of a primal-dual IPM solver. Extensive numerical experiments verify the effectiveness of our approach for linear optimization. We also studied the convergence pattern of IPM and established the empirical conditions under which primal IPM is likely to outperform primal-dual IPM.

1.1 Related Work

Iterative solver-based IPM. Solving the normal equations $(\mathbf{A}\mathbf{D}^2\mathbf{A}^\top)\Delta\mathbf{y} = \mathbf{r}$, where $\mathbf{D}$ is a diagonal matrix, often presents the most serious computational obstacle for large-scale LPs. One popular idea is to avoid factorizations by instead solving the normal equations using preconditioned iterative methods, leading to so-called matrix-free IPMs that do not require entrywise access to the matrix [19]. Practically, IPMs using iterative solvers have shown empirical success in many applications [45, 11, 40, 36, 41]. On the theory side, literature on these methods must analyze the convergence of IPMs with inexact subproblem solves [3, 15, 18, 4], as iterative solvers generally cannot reach the same accuracy as factorization-based solvers. An important branch of research focuses on the design of preconditioners for the IPM normal matrices [6, 20, 10, 9, 8, 7].

Arithmetic complexity of IPM. Improving the arithmetic complexity of IPM has been a longstanding goal in the IPM literature [26, 39, 13], and often builds on tools from theoretical computer science. State-of-the-art complexity results for LP have been improved to the same order of matrix multiplication [12, 38]. These results generally ensure the stability of the primal-dual scaled projection matrix through a carefully designed data structure, which reduces the cost of solving for the IPM search directions. This paper similarly relies on the stability of the normal matrix but exploits the natural stability of the primal normal matrix instead of requiring extra machinery.

1.2 Notations

Throughout the paper we use $\langle \cdot, \cdot \rangle$ to denote Euclidean inner product and $\| \cdot \|$ to denote Euclidean norm or matrix operator norm. Boldface letters $\mathbf{A}, \mathbf{a}$ denote matrices and vectors. $\mathbf{0}$ and $\mathbf{e}$ denote all-zero and all-one vectors and $\mathbf{I}$ denotes the identity matrix of proper size. Given a vector $\mathbf{x} \in \mathbb{R}^n$, $\mathrm{diag}(\mathbf{x})$ is an $n \times n$ matrix with $\mathbf{x}$ on the diagonal. Following the convention of IPM literature, we use $\mathbf{X}, \mathbf{S}, \mathbf{D}$ to denote $\mathrm{diag}(\mathbf{x})$, $\mathrm{diag}(\mathbf{s})$ and $\mathrm{diag}(\mathbf{d})$ respectively. We denote by $\mathcal{F}_p^0 := \{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} > \mathbf{0}\}$ and $\mathcal{F}_d^0 := \{(\mathbf{y}, \mathbf{s}) : \mathbf{A}^\top \mathbf{y} + \mathbf{s} = \mathbf{c}, \mathbf{s} > 0\}$ the primal and dual relative interior, respectively. Given a symmetric positive definite matrix $\mathbf{M}$, $\kappa(\mathbf{M}) := \|\mathbf{M}\|\|\mathbf{M}^{-1}\| = \lambda_{\max}(\mathbf{M})/\lambda_{\min}(\mathbf{M})$ denotes its Euclidean condition number and $\mathbf{M}^{1/2}$ denotes its square-root. Given two symmetric matrices $\mathbf{A}, \mathbf{B}$, we define the semidefinite order $\mathbf{A} \succeq \mathbf{B}$ if and only if $\mathbf{A} - \mathbf{B}$ is positive semidefinite. Given any $\mathbf{x} \in \mathbb{R}_{++}^n$, we define $\|\mathbf{x}\|_{-\infty} := \min_{1 \leq j \leq n}\{x_j\}$ and $\|\mathbf{a}\|_\mathbf{x} := \|\mathbf{X}^{-1}\mathbf{a}\|$. Given an index set $\mathcal{B}$, we use $\mathbf{A}_\mathcal{B}, \mathbf{x}_\mathcal{B}$ to denote slicing of columns of $\mathbf{A}$, or slicing of coordinates in $\mathbf{x}$ over indices $\mathcal{B}$. We define gradient projection operator for a full row-rank matrix $\mathbf{A}$ as $\mathbf{P}_\mathbf{A} := \mathbf{I} - \mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}\mathbf{A}$. Unlike its definition

in standard linear algebra, the term *normal matrix* in this paper will refer to the left-hand-side matrix of the normal equation.

2 Interior Point Method and Scaling Matrix

This section describes the background and notation for interior point methods and introduces the primal IPM.

2.1 Scaling Matrix for Interior Point Methods

Given a primal-dual feasible pair $(\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \mathcal{F}_p^0 \times \mathcal{F}_d^0$, an interior point search direction is determined by the optimality condition with linearized complementarity condition $\mathbf{XSe} = \mu\mathbf{e}$:

$$\begin{pmatrix} \mathbf{A} & & \\ & \mathbf{A}^\top & \mathbf{I} \\ \mathbf{D_X} & & \mathbf{D_S} \end{pmatrix} \begin{pmatrix} \Delta\mathbf{x} \\ \Delta\mathbf{y} \\ \Delta\mathbf{s} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mu\mathbf{e} - \mathbf{XSe} \end{pmatrix}. \tag{1}$$

The first two equations ensure primal-dual feasibility, while the last one linearizes $\mathbf{XSe} = \mu\mathbf{e}$. There are several different ways of linearization which correspond to different *scaling matrices* [37]:

- *Primal scaling* linearizes $\mathbf{s} - \mu\mathbf{X}^{-1}\mathbf{e} = \mathbf{0}$: $(\mathbf{s} + \Delta\mathbf{s}) - \mu(\mathbf{X}^{-1}\mathbf{e} - \mathbf{X}^{-2}\Delta\mathbf{x}) = \mathbf{0}$, resulting in the system

$$\mu\mathbf{X}^{-1}\Delta\mathbf{x} + \mathbf{X}\Delta\mathbf{s} = \mu\mathbf{e} - \mathbf{XSe}. \tag{2}$$

- *Primal-dual scaling* linearizes $\mathbf{XSe} = \mu\mathbf{e}$, resulting in the system

$$\mathbf{S}\Delta\mathbf{x} + \mathbf{X}\Delta\mathbf{s} = \mu\mathbf{e} - \mathbf{XSe}. \tag{3}$$

- *Dual scaling* linearizes $\mathbf{x} - \mu\mathbf{S}^{-1}\mathbf{e} = \mathbf{0}$: $(\mathbf{x} + \Delta\mathbf{x}) - \mu(\mathbf{S}^{-1}\mathbf{e} - \mathbf{S}^{-2}\Delta\mathbf{s}) = \mathbf{0}$, resulting in the system

$$\mathbf{S}\Delta\mathbf{x} + \mu\mathbf{S}^{-1}\Delta\mathbf{s} = \mu\mathbf{e} - \mathbf{XSe}. \tag{4}$$

All three ways coincide when $\mathbf{x}$ and $\mathbf{s}$ exactly satisfy $\mathbf{XSe} = \mu\mathbf{e}$; otherwise, they define different linear systems for IPM search directions. IPM software factorizes either the augmented matrix

$$\begin{pmatrix} \mathbf{D_S^{-1}D_X} & \mathbf{A}^\top \\ \mathbf{A} & \end{pmatrix}$$

(formed by by eliminating $\Delta\mathbf{s}$ from (1)), or the normal matrix $\mathbf{M} = \mathbf{A}\mathbf{D_X^{-1}D_S}\mathbf{A}^\top$ (formed by eliminating $\Delta\mathbf{x}$ as well). The *primal, primal-dual, and dual normal matrices* are defined respectively as

$$\mathbf{M_X} := \mathbf{A}\mathbf{X}^2\mathbf{A}^\top \qquad \mathbf{M_{X,S}} := \mathbf{A}\mathbf{X}\mathbf{S}^{-1}\mathbf{A}^\top \qquad \mathbf{M_S} := \mathbf{A}\mathbf{S}^{-2}\mathbf{A}^\top.$$

Most IPM software uses primal-dual scaling, as years of computational experience shows that primal-dual scaling performs best for linear optimization [28, 42]. When the IPM converges, some elements of $\mathbf{XS}^{-1}$ approach $+\infty$, resulting in ill-conditioning of the primal-dual normal matrix $\mathbf{M_{X,S}}$. Although the solution to the normal equations remains bounded so long as $\mathbf{x}$ and $\mathbf{s}$ are in the interior [39], the primal-dual normal matrix $\mathbf{M_{X,S}}$ itself is unstable along IPM iterations. In this respect, the primal normal matrix $\mathbf{M_X}$ has better properties: convergence of IPM implies stability of $\mathbf{M_X}$. Moreover, assuming the convergence of interior point iterates $\{\mathbf{x}_k\}$ to $\mathbf{x}^\star$, primal IPM solves a sequence of converging linear systems $\mathbf{M_{X_k}} \to \mathbf{M_{X^\star}}$, which are not necessarily ill-conditioned [22].

2.2 Primal Scaling Interior Point Method

The first polynomial-time IPM [24] is a primal-only method based on projective transformation and reducing a potential function. Since then, many variants of IPM have emerged. The two best-known variants, using a primal scaling matrix, are primal potential reduction [43] and primal logarithmic barrier methods [34]. This paper uses the primal logarithmic barrier method [34] as the working algorithm. Primal IPM converges with minimal assumptions on the problem data:

A1: $\mathcal{F}_p^0$ and $\mathcal{F}_d^0$ are nonempty and $\mathbf{A}$ has full row rank.

The assumption **A1**, well-known as the interior point condition, is pervasive in the IPM literature. Given $\mathbf{x} \in \mathcal{F}_p^0$, we can derive the primal IPM by perturbing the KKT system, replacing complementary slackness $\mathbf{XSe} = \mathbf{0}$ by $\mathbf{XSe} = \mu\mathbf{e}$, or equivalently, $\mu\mathbf{X}^{-1}\mathbf{e} = \mathbf{s}$, resulting in the perturbed KKT system

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \qquad \mathbf{A}^\top\mathbf{y} + \mathbf{s} = \mathbf{c}, \qquad \mathbf{s} - \mu\mathbf{X}^{-1}\mathbf{e} = \mathbf{0}. \tag{KKT}$$

Then we can expand the nonlinear equation to first order to obtain the system (1) with primal scaling (2)

$$\begin{aligned} \mathbf{A}\Delta\mathbf{x} &= \mathbf{0} \\ \mathbf{A}^\top\Delta\mathbf{y} + \Delta\mathbf{s} &= \mathbf{0} \\ \Delta\mathbf{s} + \mu\mathbf{X}^{-2}\Delta\mathbf{x} &= \mu\mathbf{X}^{-1}\mathbf{e} - \mathbf{s}. \end{aligned} \tag{5}$$

While (KKT) aligns with the derivation of primal-dual IPM [42], another natural motivation for primal IPM is through the logarithmic barrier: we replace $\mathbf{x} \geq 0$ by the log-barrier penalty

$$\varphi_\mu(\mathbf{x}) := \langle \mathbf{c}, \mathbf{x} \rangle - \mu \sum_{i=1}^n \log x_i.$$

At each iteration, we minimize the first-order expansion of φ_μ over an ellipsoid $\mathcal{E}_\beta(\mathbf{x}) := \{\mathbf{d} : \|\mathbf{d}\|_\mathbf{x} \leq \beta\}$:

$$\Delta\mathbf{x} = \underset{\mathbf{A}\mathbf{d}=\mathbf{0}}{\arg\min} \quad \varphi_\mu(\mathbf{x}) + \langle \nabla\varphi_\mu(\mathbf{x}), \mathbf{d} \rangle \quad \text{subject to} \quad \|\mathbf{d}\|_\mathbf{x} \leq \beta, \tag{6}$$

where the scaled distance $\|\mathbf{d}\|_\mathbf{x} = \|\mathbf{X}^{-1}\mathbf{d}\|$. The ellipsoid, known as Dikin's ellipsoid [14], guarantees the feasibility of $\mathbf{x} + \mathbf{d}$ if $\beta \leq 1$. With the same choice of μ, (5) and (6) yield the same projected Newton direction

$$\Delta\mathbf{x} = -\mathbf{X}(\mathbf{I} - \mathbf{X}\mathbf{A}^\top(\mathbf{A}\mathbf{X}^2\mathbf{A}^\top)^{-1}\mathbf{A}\mathbf{X})(\tfrac{1}{\mu}\mathbf{X}\mathbf{c} - \mathbf{e}) = -\mathbf{X}\mathbf{P}_{\mathbf{A}\mathbf{X}}(\tfrac{1}{\mu}\mathbf{X}\mathbf{c} - \mathbf{e}) \tag{7}$$

up to a multiplicative constant. Primal IPM starts from some $\mu = \mu_1$, takes a projected Newton direction, and slightly reduces μ in every iteration (**Algorithm 1**).

Algorithm 1: Primal interior point method

input Initial point $\mathbf{x}_1 \in \mathcal{F}_p^0$, barrier parameter $\mu_1 > 0$, barrier reduction parameter $\tau \in (0,1)$

for $k = 1, 2,...$ **do**

    compute $\Delta\mathbf{x}_k = -\mathbf{X}_k\mathbf{P}_{\mathbf{A}\mathbf{X}_k}(\tfrac{1}{\mu_k}\mathbf{X}_k\mathbf{c} - \mathbf{e})$

    choose $\alpha_k \in (0,1)$ such that $\mathbf{x}_k + \alpha_k\Delta\mathbf{x}_k \in \mathcal{F}_p^0$

    update $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k\Delta\mathbf{x}_k$ and $\mu_{k+1} = (1 - \tau)\mu_k$

end

With proper initialization of $\mathbf{x}_1$ and configuration of $\{\mu_k\}$, it takes **Algorithm 1** $\mathcal{O}(\sqrt{n}\log(1/\varepsilon))$ iterations to find an ε-optimal $\mathbf{x}_\varepsilon$ such that $\mathbf{x}_\varepsilon \in \mathcal{F}_p$ and $\langle \mathbf{c}, \mathbf{x}_\varepsilon \rangle \leq \langle \mathbf{c}, \mathbf{x}^\star \rangle + \varepsilon$ [44, 34]. Moreover, an ε-optimal dual solution $(\mathbf{y}_\varepsilon, \mathbf{s}_\varepsilon) \in \mathcal{F}_d$ such that $\langle \mathbf{b}, \mathbf{y}_\varepsilon \rangle \geq \langle \mathbf{b}, \mathbf{y}^\star \rangle - \varepsilon$ can be computed by a projection subproblem [34, 43].

2.3 Primal Central Path

To make the discussion self-contained, we review a few definitions concerning the primal central path [21].

Definition 1 (Primal central path [33]). *The primal central path is defined by*

$$\mathcal{C}_p := \{\mathbf{x}_\mu : \mu > 0\}, \qquad \mathbf{x}_\mu := \operatorname*{arg\,min}_{\mathbf{x} \in \mathcal{F}_p} \{\langle \mathbf{c}, \mathbf{x} \rangle - \mu \sum_{i=1}^n \log x_i\}.$$

Under **A1**, the primal central path is guaranteed to exist [33]. Primal IPM (**Algorithm 1**) explicitly follows $\mathcal{C}_p$: the algorithm generates iterates that stay near $\mathcal{C}_p$ in the sense of the following proximity measure.

Definition 2 (Proximity to central path [33]). *Given $\mathbf{x} \in \mathcal{F}_p^0$, its centrality proximity measure $\delta(\mathbf{x}, \mu)$ with respect to barrier parameter $\mu > 0$ is defined by*

$$\delta(\mathbf{x}, \mu) := \min_{(\mathbf{y}, \mathbf{s}) \in \mathcal{F}_d} \left\| \tfrac{1}{\mu}\mathbf{Xs} - \mathbf{e} \right\| = \|\mathbf{P}_{\mathbf{AX}}(\tfrac{1}{\mu}\mathbf{Xc} - \mathbf{e})\|, \tag{8}$$

where $\mathbf{P}_{\mathbf{AX}}$ is the same as in (7).

Evaluating central path proximity is essentially a projection problem in the dual space, which minimizes the distance to $\mathcal{C}_p$. The proximity measure $\delta(\mathbf{x}, \mu)$ is a "primal-only" version of the ℓ_2-neighborhood widely used in primal-dual path-following IPM [42]. Finally, we introduce the concept of maximal complementarity pair.

Definition 3 (Maximal complementarity pair [23]). *Under **A1**, there exists some partition $(\mathcal{B}, \mathcal{N})$ satisfying $\mathcal{B} \cup \mathcal{N} = \{1, \ldots, n\}$ and $\mathcal{B} \cap \mathcal{N} = \varnothing$ such that*

$$\mathcal{B} = \{j : x_j^\star > 0 \ \text{for some} \ \mathbf{x}^\star\}, \qquad \mathcal{N} = \{j : s_j^\star > 0 \ \text{for some} \ \mathbf{s}^\star\},$$

where $(\mathbf{x}^\star, \mathbf{s}^\star)$ is some optimal primal-dual solution pair.

Partition $(\mathcal{B}, \mathcal{N})$ is known as the maximal complementarity pair. Moreover, $\mathbf{x}_\mathcal{N}^\star = \mathbf{0}$ and $\mathbf{s}_\mathcal{B}^\star = \mathbf{0}$ hold for any optimal solution $(\mathbf{x}^\star, \mathbf{s}^\star)$, and there exists some strictly complementary solution $(\mathbf{x}^\star, \mathbf{s}^\star)$ such that $\mathbf{x}^\star + \mathbf{s}^\star > \mathbf{0}$. We define constant

$$\gamma := \tfrac{n}{\|\mathbf{s}_\mathcal{N}^\star\|_{-\infty}} > 0 \tag{9}$$

for an arbitrarily chosen strictly complementary solution $(\mathbf{x}^\star, \mathbf{s}^\star)$.

3 Interior Point Method in the Primal Form

The bottleneck of primal IPM lies in computing the IPM direction (7), where a normal equation with left-hand-side normal matrix $\mathbf{M}_{\mathbf{X}_k}$ has to be solved. As previously remarked, convergence of the IPM iterates $\{\mathbf{x}_k\}$ implies the convergence of $\{\mathbf{M}_{\mathbf{X}_k}\}$. In this section, we show it is possible to reduce the cost of solving normal equations by exploiting the fact that solving a sequence of linear systems is easier.

3.1 Convergence of the Scaling Matrix

We start by characterizing the convergence behavior of primal IPM under primal non-degeneracy:

A2: The primal problem is non-degenerate; the maximal complementarity partition satisfies $|\mathcal{B}| = m$.

The primal optimal solution $\mathbf{x}^\star$ is unique given **A2**. The proof is simple: we show that for a fixed proximity parameter λ, the diameter of the proximity set $\mathcal{P}_\mu := \{\mathbf{x} : \delta(\mathbf{x}, \mu) \le \lambda\}$ is bounded by $\mathcal{O}(\mu)$. Since primal IPM decreases μ linearly and its iterates stay within the proximity set, we obtain convergence of $\{\mathbf{x}_k\}$ as $\mu \to 0$.

Lemma 3.1. *Under **A1** and **A2**, for any $\lambda \in [0, \frac{\sqrt{2}}{2}]$ and $\mu > 0$, the central path proximity set $\mathcal{P}_\mu = \{\mathbf{x} : \delta(\mathbf{x}, \mu) \le \lambda\}$, satisfies $\operatorname{diam}(\mathcal{P}_\mu) = \mathcal{O}(\mu)$.*

With **Lemma 3.1**, we obtain a global iteration convergence rate for primal IPM.

Theorem 3.1. *Under* **A1** *and* **A2**, *suppose an algorithm maintains* $\delta(\mathbf{x}_k, \mu_k) \leq \frac{\sqrt{2}}{2}$ *for all* $k = 1, 2, \ldots$ *and that* $\mu_{k+1} = (1 - \tau)\mu_k$ *for* $\tau \in (0, 1/\sqrt{2n}]$. *Then* $\|\mathbf{x}_i - \mathbf{x}_j\| = \mathcal{O}(\mu_k)$ *for all* $i, j \geq k$.

Since the primal normal matrix $\mathbf{M_X} = \mathbf{A}\mathbf{X}^2\mathbf{A}^\top$ consists of quadratic terms with respect to $\mathbf{x}$, convergence of $\{\mathbf{x}_k\}$ implies convergence of $\{\mathbf{M}_{\mathbf{X}_k}\}$. Therefore, the final iterations of the primal IPM solve linear systems with converging coefficient matrices: normal matrices of primal IPM stay in proximity to each other. This observation opens a way for preconditioned iterative solvers in primal IPM.

3.2 Proximity in Euclidean Distance

Intuitively, two nearby non-degenerate matrices have similar spectra. **Lemma 3.2** quantifies this intuition by analyzing the conditioning of the preconditioned systems.

Lemma 3.2 (Conditioning within the Euclidean ball). *Let* $\lambda_{\mathbf{x}} := \lambda_{\min}(\mathbf{M_X}) > 0$. *For any* $\mathbf{x} \in \mathbb{R}^n_{++}$, $\beta \leq 1$ *and* $\Delta\mathbf{x}$ *such that* $\|\Delta\mathbf{x}\| \leq \frac{\sqrt{\lambda_{\mathbf{x}}}}{\|\mathbf{A}\|}\beta$, *we have* $(1 - \beta)^2 \cdot \mathbf{I} \preceq \mathbf{M_X}^{-1/2}(\mathbf{M}_{\mathbf{X}+\Delta\mathbf{x}})\mathbf{M_X}^{-1/2} \preceq (1 + \beta)^2 \cdot \mathbf{I}$ *and*

$$\kappa(\mathbf{M_X}^{-1/2}(\mathbf{M}_{\mathbf{X}+\Delta\mathbf{x}})\mathbf{M_X}^{-1/2}) \leq \left(\tfrac{\beta+1}{\beta-1}\right)^2.$$

Lemma 3.2 shows that when $\mathbf{M_X}$ is far from singular and $\Delta\mathbf{x}$ is small, we can solve $\mathbf{M}_{\mathbf{X}+\Delta\mathbf{x}}$ using $\mathbf{M_X}$ as a preconditioner: proximity in Euclidean distance implies effective preconditioning. In particular, **Lemma 3.2** and **Theorem 3.1** together imply that when the primal IPM converges, good conditioning is guaranteed for all the normal matrices generated by the primal IPM iteration:

Theorem 3.2. *Under* **A1** *and* **A2**, *let* $\{\mathbf{M}_{\mathbf{X}_k}\}$ *be the sequence of normal matrices generated by the primal IPM specified in* **Algorithm 1** *and suppose the conditions in* **Theorem 3.1** *hold. Then*

$$\kappa(\mathbf{M}_{\mathbf{X}_k}^{-1/2}\mathbf{M}_{\mathbf{X}_j}\mathbf{M}_{\mathbf{X}_k}^{-1/2}) \leq 9$$

for all $j \geq k \geq k^\star := \min\left\{k : \mu_k \leq \frac{\sqrt{\lambda^\star}\tau}{16\gamma\|\mathbf{A}\|} \cdot (1 + \frac{\|\mathbf{A}_{\mathcal{N}}\|}{\sqrt{\lambda_{\min}(\mathbf{A}_{\mathcal{B}}^\top\mathbf{A}_{\mathcal{B}})}})^{-1}\right\}$, *where* $\lambda^\star := \lambda_{\min}(\mathbf{M}_{\mathbf{X}^\star})$. *In particular, choosing* $\tau = 1/\sqrt{2n}$ *gives*

$$k^\star = \mathcal{O}\left(\sqrt{n}\left[\log(\tfrac{\gamma\|\mathbf{A}\|}{\sqrt{\lambda^\star}}) + \log\left(1 + \tfrac{\|\mathbf{A}_{\mathcal{N}}\|}{\sqrt{\lambda_{\min}(\mathbf{A}_{\mathcal{B}}^\top\mathbf{A}_{\mathcal{B}})}}\right)\right]\right).$$

Given a linear system with a bounded condition number after preconditioning, it is well-known that preconditioned iterative solvers, such as conjugate gradient, can solve the linear system to ε accuracy in $\mathcal{O}(n^2 \log(1/\varepsilon))$ arithmetic complexity [35]. **Theorem 3.2** itself provides an algorithm: when the primal IPM iterates converge, we reuse the factorization of past normal matrices as preconditioners to solve future normal equations. Combining **Theorem 3.2** with the IPM inexact analysis yeilds the arithmetic complexity for **Algorithm 2**.

Theorem 3.3 (Informal). *The total complexity of* **Algorithm 2** *is* $\mathcal{O}(n^3 k^\star + n^{2.5}\log^2(\frac{1}{\varepsilon}))$, *where* $k^\star$ *is defined in* **Theorem 3.2**.

The proof of **Theorem 3.3** reveals the algorithm has two distinct phases. The $n^3 k^\star$ complexity follows from the standard analysis of primal IPM [44]. When μ_k is small enough that all future normal matrices $\{\mathbf{M}_k\}_{k\geq k^\star}$ are close to $\mathbf{M}_{k^\star}$, using the fixed preconditioner $\mathbf{M}_{k^\star}$ decreases the cost per IPM iteration from $\mathcal{O}(n^3)$ to $\mathcal{O}(n^2)$. Proof of **Theorem 3.3** is deferred to **Appendix C**.

Algorithm 2 is simple and straight-forward to implement. Yet in practice, $\mu_{k^\star}$ can be extremely small, so the algorithm may terminate before the second, faster, phase of convergence begins. To make primal IPM practically useful, we must introduce one more type of proximity defined with respect to the scaled distance.

Algorithm 2: Preconditioned primal interior point method

> **input** Initial points $\mathbf{z} = \mathbf{x}_1 \in \mathcal{F}_p^0$, barrier parameter $\mu_1 > 0$, barrier reduction parameter $\tau \in (0, 1)$,
>   cached factorization of $\mathbf{M_Z} = \mathbf{A}\mathbf{Z}^2\mathbf{A}^\top$, factorization update tolerance $\theta > 0$
> **for** $k = 1, 2,...$ **do**
>   **if** $\|\mathbf{x}_k - \mathbf{z}\| \geq \theta$ **then**
>   **update $\mathbf{z} = \mathbf{x}_k$ and factorize $\mathbf{M_Z} = \mathbf{A}\mathbf{X}_k^2\mathbf{A}^\top$**
>   **end**
>   **compute** $\Delta\hat{\mathbf{x}}_k$ with iterative solver and preconditioner $\mathbf{M_Z}$
>   **choose** $\alpha_k \in (0, 1]$ such that $\mathbf{x}_k + \alpha_k \Delta\hat{\mathbf{x}}_k \in \mathcal{F}_p^0$
>   **update** $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \Delta\hat{\mathbf{x}}_k$ and $\mu_{k+1} = (1 - \tau)\mu_k$
> **end**

3.3 Proximity in the Scaled Distance

Given $\mathbf{x}, \mathbf{w} \in \mathbb{R}_{++}^n$, we say $\mathbf{w}$ is close to $\mathbf{x}$ in the scaled distance if $\|\mathbf{w} - \mathbf{x}\|_{\mathbf{x}}$ is small. An important consequence of proximity in the scaled distance is **Lemma 3.3**, known as the *shifted scaling theorem* in the IPM literature. It shows that given $\mathbf{w} = \mathbf{x} + \Delta\mathbf{x}$ that is close to $\mathbf{x}$ in the scaled distance, the scaled projection operator $\mathbf{P_{AX}}$ can be well approximated by $\mathbf{P_{AW}}$.

Lemma 3.3 (Shifted scaling [21]). *For any* $\mathbf{x} \in \mathbb{R}_{++}^n, \beta \leq \frac{1}{4}$ *and* $\Delta\mathbf{x}$ *such that* $\|\Delta\mathbf{x}\|_{\mathbf{x}} \leq \beta$,

$$\left\| \left[\mathbf{X}^{-1}(\mathbf{X} + \Delta\mathbf{X})\mathbf{P}_{\mathbf{A}(\mathbf{X}+\Delta\mathbf{X})}\mathbf{X}^{-1}(\mathbf{X} + \Delta\mathbf{X}) - \mathbf{P_{AX}} \right] \mathbf{v} \right\| \leq 3\beta\|\mathbf{P_{AX}}\mathbf{v}\| \;\; \textit{for all} \;\; \mathbf{v} \in \mathbb{R}^n.$$

Lemma 3.3 was initially proposed as a theoretical tool to analyze the primal central path [21]. However, it will be a critical tool for our algorithm design: suppose we are at some $\mathbf{x} \in \mathcal{F}_p^0$. Then, according to (7), the primal IPM direction at $\mathbf{x}$ is

$$-\mathbf{X}\mathbf{P_{AX}}(\tfrac{1}{\mu}\mathbf{X}\mathbf{c} - \mathbf{e}).$$

Suppose instead of applying $\mathbf{P_{AX}} = \mathbf{I} - \mathbf{X}\mathbf{A}^\top(\mathbf{A}\mathbf{X}^2\mathbf{A}^\top)^{-1}\mathbf{A}\mathbf{X}$, which involves $(\mathbf{A}\mathbf{X}^2\mathbf{A}^\top)^{-1}$, we only have access to the operator $\mathbf{P_{AW}}$ at another point $\mathbf{w}$ such that $\|\mathbf{w} - \mathbf{x}\|_{\mathbf{x}}$ is small. Then, using **Lemma 3.3** with $\mathbf{x} + \Delta\mathbf{x} = \mathbf{w}$ and $\mathbf{v} = \frac{1}{\mu}\mathbf{X}\mathbf{c} - \mathbf{e}$, we notice that

$$\left\| \left[-\mathbf{X}\mathbf{P_{AX}}(\tfrac{1}{\mu}\mathbf{X}\mathbf{c} - \mathbf{e}) \right] - \left[-\mathbf{W}\mathbf{P_{AW}}\mathbf{X}^{-1}\mathbf{W}(\tfrac{1}{\mu}\mathbf{X}\mathbf{c} - \mathbf{e}) \right] \right\|_{\mathbf{x}}$$

$$= \left\| \mathbf{X}^{-1}\mathbf{W}\mathbf{P_{AW}}\mathbf{X}^{-1}\mathbf{W}(\tfrac{1}{\mu}\mathbf{X}\mathbf{c} - \mathbf{e}) - \mathbf{P_{AX}}(\tfrac{1}{\mu}\mathbf{X}\mathbf{c} - \mathbf{e}) \right\| \tag{10}$$

$$= \left\| \left[\mathbf{X}^{-1}\mathbf{W}\mathbf{P_{AW}}\mathbf{X}^{-1}\mathbf{W} - \mathbf{P_{AX}} \right] (\tfrac{1}{\mu}\mathbf{X}\mathbf{c} - \mathbf{e}) \right\| \leq 3\beta\|\mathbf{P_{AX}}(\tfrac{1}{\mu}\mathbf{X}\mathbf{c} - \mathbf{e})\| = 3\beta\delta(\mathbf{x}, \mu), \tag{11}$$

where (10) uses the definition $\|\mathbf{a}\|_{\mathbf{x}} = \|\mathbf{X}^{-1}\mathbf{a}\|$ and (11) uses $\delta(\mathbf{x}, \mu) = \|\mathbf{P_{AX}}(\tfrac{1}{\mu}\mathbf{X}\mathbf{c} - \mathbf{e})\|$ from (8). Hence, if $\delta(\mathbf{x}, \mu)$ is small and $\mathbf{w}$ is close to $\mathbf{x}$ in the scaled distance, the shifted scaling theorem provides a good surrogate direction

$$-\mathbf{W}\mathbf{P_{AW}}\mathbf{X}^{-1}\mathbf{W}(\tfrac{1}{\mu}\mathbf{X}\mathbf{c} - \mathbf{e}) \tag{12}$$

without even computing $\mathbf{P_{AX}}$. This observation is significant in making primal IPM practically useful.

In the next section, we put all the things together and show how these two types of proximity can be combined to make primal IPM competitive in solving real LPs.

3.4 Combining two Types of Proximity

We have shown that proximity in Euclidean and scaled distance both allow us to reduce the cost of solving the normal equations. However, as the following example shows, neither of them *alone* may be usable in practice.

Example 3.1. *Let $\mathbf{z} = (10^{10}, 10^{-5})^{\top}$ and $\mathbf{x} = (10^{10} - 10^5, 10^{-10})^{\top}$. Then*

$$\|\mathbf{z} - \mathbf{x}\| \approx 10^5 \quad and \quad \|\mathbf{z} - \mathbf{x}\|_{\mathbf{x}} \approx 10^5.$$

$\mathbf{x}$ *and* $\mathbf{z}$ *are far away in both Euclidean and the scaled distance.*

Example 3.1 can indeed happen in a real algorithm: the first coordinate corresponds to a basic variable converging to a large value, and the second is a non-basic variable converging to 0. Although $\mathbf{x}$ and $\mathbf{z}$ are both near convergence, neither Euclidean nor the scaled distance is small. The issue arises due to the misalignment of two types of distances: as the primal iterates converge, the coordinates converging to 0 will be close in Euclidean distance but not the scaled distance. In contrast, the coordinates converging to a (large) nonzero value will be close in the scaled distance long before they are close in Euclidean distance. Two geometries complement each other. Hence, we define a combined metric to reap the benefits of both.

Definition 4 (Thresholded scaled distance). *The ν-thresholded scaled distance between $\mathbf{y}, \mathbf{z} \in \mathbb{R}^n$ with respect to $\mathbf{x} \in \mathbb{R}^n_{++}$ is given by*

$$\|\mathbf{y} - \mathbf{z}\|_{\mathbf{x},\nu} = \sqrt{\|\mathbf{X}_{\mathcal{L}}^{-1}(\mathbf{y}_{\mathcal{L}} - \mathbf{z}_{\mathcal{L}})\|^2 + \|\mathbf{y}_{\mathcal{S}} - \mathbf{z}_{\mathcal{S}}\|^2},$$

where $\mathcal{L} := \{j : x_j \geq \nu\}$ and $\mathcal{S} = \{j : x_j < \nu\}$.

The distance $\|\mathbf{y} - \mathbf{z}\|_{\mathbf{x},\nu}$ combines Euclidean and the scaled distance with a thresholding parameter ν: distance with respect to large elements $\mathcal{L}$ are scaled down, while the small elements $\mathcal{S}$ are not. For example, in **Example 3.1**, taking $\nu = 1$ gives $\|\mathbf{z} - \mathbf{x}\|_{\mathbf{x},1} \approx 10^{-5}$. This new metric motivates a modification of primal IPM that takes advantage of both geometries.

3.4.1 Primal Interior Point Method with Delayed Scaling

We are ready to leverage the thresholded scaled distance to motivate a more practical primal IPM. Given $\mathbf{x}, \mathbf{z} \in \mathcal{F}_p^0$ and a factorization for $\mathbf{M_Z}$, suppose we want to compute the primal search direction

$$\Delta \mathbf{x} = -\mathbf{X}\mathbf{P_{AX}}(\tfrac{1}{\mu}\mathbf{X}\mathbf{c} - \mathbf{e}).$$

Following our previous argument, we hope the factorization of $\mathbf{M_Z}$ can accelerate computing $\Delta\mathbf{x}$. But even if $\|\mathbf{z} - \mathbf{x}\|_{\mathbf{x},\nu}$ is small, there is no guarantee that either $\|\mathbf{z} - \mathbf{x}\|$ or $\|\mathbf{z} - \mathbf{x}\|_{\mathbf{x}}$ is small. Instead, our algorithm (**Algorithm 3**) constructs an intermediate point $\mathbf{w}$ so that both $\|\mathbf{w} - \mathbf{z}\|$ and $\|\mathbf{w} - \mathbf{x}\|_{\mathbf{x}}$ are small. Then we compute a surrogate search direction $\Delta\hat{\mathbf{x}} = -\mathbf{W}\mathbf{P_{AW}}(\tfrac{1}{\mu}\mathbf{W}\mathbf{c} - \mathbf{W}\mathbf{X}^{-1}\mathbf{e})$ in (12) with preconditioner $\mathbf{M_Z}$.

Algorithm 3: Preconditioned primal interior point method with delayed scaling matrix

input Initial point $\mathbf{z} = \mathbf{x}_1 \in \mathcal{F}_p^0$, barrier parameter $\mu_1 > 0$, barrier reduction parameter $\tau \in (0, 1)$,
cached factorization of $\mathbf{M_Z} = \mathbf{A}\mathbf{Z}^2\mathbf{A}^{\top}$, threshold $\nu > 0$, factorization update tolerance $\theta > 0$
for $k = 1, 2,...$ **do**
  if $\|\mathbf{x}_k - \mathbf{z}\|_{\mathbf{x}_k,\nu} \geq \theta$ **then**
    update $\mathbf{z} = \mathbf{x}_k$ and factorize $\mathbf{M_Z} = \mathbf{A}\mathbf{Z}^2\mathbf{A}^{\top}$
  end
  choose delayed scaling matrix $\mathbf{w}$, $\mathbf{w}_{\mathcal{S}} = \mathbf{x}_{k,\mathcal{S}}, \mathbf{w}_{\mathcal{L}} = \mathbf{z}_{\mathcal{L}}$
  compute $\Delta\hat{\mathbf{x}}_k = -\mathbf{W}\mathbf{P_{AW}}(\tfrac{1}{\mu_k}\mathbf{W}\mathbf{c} - \mathbf{W}\mathbf{X}_k^{-1}\mathbf{e})$ with preconditioner $\mathbf{M_Z}$
  choose $\quad \alpha_k \in (0, 1)$ such that $\mathbf{x}_k + \alpha_k\Delta\hat{\mathbf{x}}_k \in \mathcal{F}_p^0$
  update $\quad \mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k\Delta\hat{\mathbf{x}}_k \quad$ and $\quad \mu_{k+1} = (1 - \tau)\mu_k$
end

The validity of **Algorithm 3** is guaranteed by **Theorem 3.4**.

Theorem 3.4. *Fix $\nu \in (0, 1]$. Suppose $\mu > 0$, $\mathbf{x}, \mathbf{z} \in \mathcal{F}_p^0$ and $\mathbf{M_Z} \succeq \lambda_{\mathbf{z}} \cdot \mathbf{I}$ is already factorized and that*

$$\|\mathbf{z} - \mathbf{x}\|_{\mathbf{x},\nu} \le \min\big\{ \tfrac{\sqrt{\lambda_{\mathbf{z}}}}{2\|\mathbf{A}\|}, \tfrac{1}{4}, \tfrac{\varepsilon}{6\delta(\mathbf{x},\mu)} \big\}.$$

*Then **Algorithm 3** computes $\Delta\hat{\mathbf{x}}$ such that $\|\Delta\mathbf{x} - \Delta\hat{\mathbf{x}}\|_{\mathbf{x}} \le \varepsilon$ in $\mathcal{O}(n^2 \log(1/\varepsilon))$ arithmetic complexity with the cached factorization of $\mathbf{M_Z}$.*

To illustrate the choice of $\mathbf{w}$, again in **Example 3.1**, we have

$$\mathbf{x} = \begin{pmatrix} 10^{10} - 10^5 \\ 10^{-10} \end{pmatrix} \stackrel{\text{Scaled}}{\approx} \mathbf{w} = \begin{pmatrix} 10^{10} \\ 10^{-10} \end{pmatrix} \stackrel{\text{Euclidean}}{\approx} \begin{pmatrix} 10^{10} \\ 10^{-5} \end{pmatrix} = \mathbf{z}$$

with $\|\mathbf{w} - \mathbf{z}\| \approx 10^{-5}$, $\|\mathbf{w} - \mathbf{x}\|_{\mathbf{x}} \approx 10^{-5}$. More generally, good preconditioning requires some scaling matrix $\mathbf{W}$ such that **1)** $\mathbf{w}$ is close to $\mathbf{z}$ in Euclidean distance so that **Lemma 3.2** guarantees good preconditioning. **2)** $\mathbf{w}$ is close to $\mathbf{x}$ in the scaled distance so that **Lemma 3.3** guarantees that the search direction is still accurate. Such a $\mathbf{w}$ exists in primal IPM, and we call it a *delayed* scaling matrix: the small coordinates in $\mathbf{x}$ are updated to form $\mathbf{w}$, but the update of the large coordinates in $\mathbf{x}$ is delayed to maximize the stability of the normal matrices, by **Lemma 3.2**.

4 Primal Interior Point Method in Practice

This section describes how to implement the primal IPM, including an infeasible start extension and dualization, and shows how to generalize primal IPM to SDP.

4.1 Infeasible Start Primal IPM

Most modern IPM solvers use an infeasible start variant of primal-dual IPM. Infeasible start IPM only requires iterates to lie in $\mathbb{R}_{++}^n$, while primal and dual feasibility can be violated. We derive an infeasible start variant of primal IPM through the KKT interpretation (5) in **Section 2**. Given a primal-dual solution pair $(\mathbf{x}, \mathbf{y}, \mathbf{s})$, primal, dual and complementary residuals are, respectively,

$$\mathbf{r}_p = \mathbf{A}\mathbf{x} - \mathbf{b}, \quad \mathbf{r}_d = \mathbf{A}^\top \mathbf{y} + \mathbf{s} - \mathbf{c}, \quad \mathbf{r}_\mu = \mathbf{s} - \mu \mathbf{X}^{-1}\mathbf{e}.$$

Linearizing these relations, we obtain

$$\mathbf{A}\Delta\mathbf{x} = -\mathbf{r}_p, \quad \mathbf{A}^\top\Delta\mathbf{y} + \Delta\mathbf{s} = -\mathbf{r}_d, \quad \Delta\mathbf{s} + \mu\mathbf{X}^{-2}\Delta\mathbf{x} = -\mathbf{r}_\mu,$$

where $(\mathbf{X} + \Delta\mathbf{X})^{-1}\mathbf{e}$ is linearized as $\mathbf{X}^{-1}\mathbf{e} - \mathbf{X}^{-2}\Delta\mathbf{x}$. After eliminating $\Delta\mathbf{x}$ and $\Delta\mathbf{s}$, we find

$$
\begin{aligned}
\mathbf{A}\mathbf{X}^2\mathbf{A}^\top\Delta\mathbf{y} &= -\mu\mathbf{r}_p + \mathbf{A}\mathbf{X}^2(\mathbf{r}_\mu - \mathbf{r}_d) \\
\Delta\mathbf{s} &= -\mathbf{r}_d - \mathbf{A}^\top\Delta\mathbf{y} \\
\Delta\mathbf{x} &= -\tfrac{1}{\mu}\mathbf{X}^2(\mathbf{r}_\mu + \Delta\mathbf{s}).
\end{aligned}
\tag{13}
$$

Similar to primal-dual IPM, after obtaining $\Delta\mathbf{y}$ from the normal equation, we do backward substitution to recover $\Delta\mathbf{s}$ and $\Delta\mathbf{x}$. Then, primal IPM performs a ratio test, decreases μ, and goes to the next iteration. One advantage of (13) is that we explicitly maintain dual variables $(\mathbf{y}, \mathbf{s})$, and primal IPM behaves like a primal-dual IPM. We conclude this section with a discussion of potential numerical issues arising from (13): although mathematically correct, solving (13) directly sometimes fails due to floating point error: error in $\Delta\mathbf{x}$ is magnified by $1/\mu$. To resolve this issue, we can rearrange and divide the normal equation by μ to obtain

$$\mathbf{A}\mathbf{X}^2\mathbf{A}^\top\big(\tfrac{1}{\mu}\Delta\mathbf{y}\big) = -\mathbf{r}_p + \tfrac{1}{\mu}\mathbf{A}\mathbf{X}(\mathbf{X}\mathbf{S}\mathbf{e} - \mu\mathbf{e}) - \tfrac{1}{\mu}\mathbf{A}\mathbf{X}^2\mathbf{r}_d,$$

and solve for $\mu^{-1}\Delta\mathbf{y}$ instead. Notice that $\frac{1}{\mu}(\mathbf{XSe} - \mu\mathbf{e})$ corresponds to the proximity measure and is bounded for primal IPM. Hence it is possible to obtain $\mu^{-1}\Delta\mathbf{y}$ with high accuracy so long as $\mathbf{r}_d$ is not large. This strategy provides a more stable solution for the primal IPM direction.

4.2 Dualization, Symmetric Standard Form and Variable Bounds

Convergence of the normal matrix is only available for primal IPM, but our results extend to the dual problem in symmetric primal-dual standard form:

$$\operatorname*{minimize}_{\mathbf{x}\geq\mathbf{0}} \quad \langle\mathbf{c},\mathbf{x}\rangle \quad \text{subject to} \quad \mathbf{Ax}\geq\mathbf{b} \quad \text{and} \quad \operatorname*{maximize}_{\mathbf{y}\geq\mathbf{0}} \quad \langle\mathbf{b},\mathbf{y}\rangle \quad \text{subject to} \quad \mathbf{A}^\top\mathbf{y}\leq\mathbf{c}. \tag{14}$$

(14) ensures the primal and dual problems have the same form. Therefore, primal IPM can be applied after dualization. By introducing primal and dual slack $\mathbf{w},\mathbf{z}\geq\mathbf{0}$ such that $\mathbf{Ax}-\mathbf{w}=\mathbf{b}, \mathbf{A}^\top\mathbf{y}+\mathbf{z}=\mathbf{c}$,

$$\mathbf{Ax}-\mathbf{w}=\mathbf{b}, \quad \mathbf{ZXe}=\mu\mathbf{e}, \quad (\mathbf{x},\mathbf{w})\geq\mathbf{0}$$
$$\mathbf{A}^\top\mathbf{y}+\mathbf{z}=\mathbf{c}, \quad \mathbf{WYe}=\mu\mathbf{e}, \quad (\mathbf{y},\mathbf{z})\geq\mathbf{0}$$

In the primal form, $\mathbf{z}=\mu\mathbf{X}^{-1}\mathbf{e}$ and $\mathbf{y}=\mu\mathbf{W}^{-1}\mathbf{e}$ are linearized:

$$\Delta\mathbf{z}+\mu\mathbf{X}^{-2}\Delta\mathbf{x}=\mu\mathbf{X}^{-1}\mathbf{e}-\mathbf{z} \quad \text{and} \quad \Delta\mathbf{y}+\mu\mathbf{W}^{-2}\Delta\mathbf{w}=\mu\mathbf{W}^{-1}\mathbf{e}-\mathbf{y}.$$

In this case, the normal matrix is $\mathbf{AX}^2\mathbf{A}^\top+\mathbf{W}^2$. In the dual form, $\mathbf{x}=\mu\mathbf{Z}^{-1}\mathbf{e}, \mathbf{w}=\mathbf{Y}^{-1}\mathbf{e}$ are linearized:

$$\Delta\mathbf{x}+\mu\mathbf{Z}^{-2}\Delta\mathbf{z}=\mu\mathbf{Z}^{-1}\mathbf{e}-\mathbf{x} \quad \text{and} \quad \Delta\mathbf{w}+\mu\mathbf{Y}^{-2}\Delta\mathbf{y}=\mu\mathbf{Y}^{-1}\mathbf{e}-\mathbf{w}.$$

The normal matrix becomes $\mathbf{A}^\top\mathbf{Y}^2\mathbf{A}+\mathbf{Z}^2$. The symmetric form allows switching between primal and dual problems to exploit sparsity or (non-)degeneracy.

Variable Bound. Another case of interest is the standard form with bounds,

$$\operatorname*{minimize}_{\mathbf{x}} \quad \langle\mathbf{c},\mathbf{x}\rangle \quad \text{subject to} \quad \mathbf{Ax}=\mathbf{b} \quad \mathbf{0}\leq\mathbf{x}\leq\mathbf{u} \tag{15}$$

where $u_j\in[0,\infty)\cup\{+\infty\}$. The barrier function becomes $-\log(u-x)-\log x$ for upper-bounded variables, and the barrier Hessian gives the scaling matrix $\mathbf{D}^2=\mathbf{X}^{-2}+(\mathbf{U}-\mathbf{X})^{-2}$ with diagonals

$$\sqrt{\frac{1}{x^{-2}+(u-x)^{-2}}}=\frac{x(u-x)}{\sqrt{x^2+(u-x)^2}}.$$

The diagonal of the scaling matrix also remains stable when x converges.

4.3 Semidefinite Programming

This section discusses how a preconditioned primal IPM generalizes to SDP. Consider standard form SDP:

$$\operatorname*{minimize}_{\mathbf{X}} \quad \langle\mathbf{C},\mathbf{X}\rangle \quad \text{subject to} \quad \mathcal{A}(\mathbf{X})=\mathbf{b}, \quad \mathbf{X}\succeq\mathbf{0},$$

where $\mathcal{A}(\mathbf{X})=(\langle\mathbf{A}_1,\mathbf{X}\rangle,\dots,\langle\mathbf{A}_m,\mathbf{X}\rangle)^\top\in\mathbb{R}^m$. The primal normal matrix is

$$\mathbf{M}_\mathbf{X}=\begin{pmatrix} \langle\mathbf{A}_1,\mathbf{XA}_1\mathbf{X}\rangle & \cdots & \langle\mathbf{A}_1,\mathbf{XA}_m\mathbf{X}\rangle \\ \vdots & \ddots & \vdots \\ \langle\mathbf{A}_m,\mathbf{XA}_1\mathbf{X}\rangle & \cdots & \langle\mathbf{A}_m,\mathbf{XA}_m\mathbf{X}\rangle \end{pmatrix}.$$

When m is large, setting up and factorizing $\mathbf{M_X}$ becomes a bottleneck in IPM for SDP. Suppose $\mathbf{M}_{\mathbf{X}_k}$ is applied as a preconditioner. The action of the normal matrix $\mathbf{M_X}$ on $\Delta \mathbf{y}$ can be written as

$$\mathbf{M_X} \Delta \mathbf{y} = \mathcal{A}(\mathbf{X}(\mathcal{A}^* \Delta \mathbf{y})\mathbf{X}).$$

Hence, we can solve the normal equation at each iterate $\mathbf{x}$ using an iterative solver without constructing $\mathbf{M_X}$ explicitly. Note that forming the preconditioner still requires (occasional) explicit construction and factorization of the primal normal matrix.

4.4 Combining the Benefits of Primal IPM with Primal-Dual IPM

We conclude with a few remarks on how primal IPM can benefit existing primal-dual solvers. First of all, we advocate that primal IPM should not be used as a replacement for primal-dual IPM but rather as a complement: the advantage of primal IPM only appears near convergence, while the initial iterations needed for primal IPM to ensure proximity in the (thresholded) distance can be slower than the time required for primal-dual to solve the whole problem. Therefore, it works best to let primal-dual IPM iterate until a medium-accuracy solution is found and then to invoke primal IPM to complete the final iterations. Last, we discuss the complexity of implementation. On an algorithmic level, infeasible-start primal IPM is similar to primal-dual IPM; the main difference is simply the use of an iterative solver. However, given that primal-dual IPM software often uses an iterative refinement procedure (a particular type of iterative method), primal IPM can be implemented in a primal-dual solver without much additional effort.

5 Numerical Experiments

In this section, we conduct numerical experiments to validate the performance of primal IPM.

5.1 Experiment Setup

Our experiment contains three parts.

- **Section 5.2** demonstrates the practical utility of a combined primal-dual and primal IPM for solving LPs.

- **Section 5.3** demonstrates the effectiveness of primal IPM for SDP with many constraints.

- **Section 5.4** studies the convergence behavior of $\{\mathbf{x}_k\}$ in primal-dual IPM.

Testing environment. Qualitative results in **Section 5.4** are obtained with `Macbook Pro` with `Apple Silicon 32G`. The rest of the results are obtained with `AMD Ryzen 9 5900X 12-Core 128G`.

Dataset. In LP-related tests, we use 924 LP instances from the LP relaxation of MIPLIB 2017 [17]. Each instance is presolved by `Gurobi` and converted to standard form. We use the instances from Hans Mittelmann's SDP benchmark [31] for SDP-related tests.

IPM solver. We compare three primal-dual IPM implementations:

- `Gurobi` and `COPT` for IPM convergence behavior analysis in **Section 5.4**.

- A textbook version of a primal-dual IPM with Mehrotra type corrector step in `Matlab` for **Section 5.4**.

- A textbook version of a primal-dual IPM within the open-source software `HDSDP` [16] for **Section 5.2**.

We compare two primal IPM implementations:

- The same primal IPM as in `HDSDP`. We use either conjugate gradient method as the iterative solver. This version is used for **Section 5.2**.

- SDP version of primal IPM in `HDSDP`. The conjugate gradient method is used as the iterative solver. This version is used for **Section 5.3**.

Convergence metric. For LPs, our experiments use standard convergence metrics: primal infeasibility, dual infeasibility, and duality gap:

$$e_p := \frac{\|\mathbf{Ax-b}\|}{1+\|\mathbf{b}\|}, \quad e_d := \frac{\|\mathbf{A}^\top \mathbf{y+s-c}\|}{1+\|\mathbf{c}\|}, \quad e_g := \frac{|\langle\mathbf{c,x}\rangle-\langle\mathbf{b,y}\rangle|}{1+|\langle\mathbf{c,x}\rangle|+|\langle\mathbf{b,y}\rangle|}.$$

For SDPs, our experiments use DIMACS error [30].

5.2 Primal IPM for LPs

In this section, we report the performance of primal IPM applied to solving LPs.

Additional experiment details. For each LP instance, we run

- Standard primal-dual IPM with predictor-corrector step.

- Hybrid primal-dual-primal IPM. The method starts with primal-dual IPM. When primal-dual IPM runs, the solver estimates the ratio between the time factorizing the normal matrix and forward/backward substitution. The hybrid solver switches to primal IPM if $\|\mathbf{x}_k - \mathbf{x}_{k-1}\|_{\mathbf{x}_k,\nu} \leq 10^{-1}$ and the time ratio is greater than 30. For convenience, we call this hybrid algorithm primal IPM.

until one of the three conditions is satisfied: **1)** $\max\{e_p, e_d, e_g\} \leq 10^{-10}$, in which case we say the solver succeeds on the instance; **2)** the iteration number reaches the maximum limit of 100; **3)** matrix factorization fails due to insufficient memory or numerical error. In the latter two cases, we say the solver fails the test on the instance. There are four possible cases:

C1. Primal-dual IPM fails the test.

C2. The hybrid solver does not invoke Primal IPM.

C3. Primal-dual IPM passes the test, while primal IPM fails the test.

C4. Both primal-dual IPM and primal IPM pass the test.

Let T_{PD} and T_P be the solution time of primal-dual and primal IPM in some instance, respectively. We define the speedup of primal IPM by $s = \frac{T_{PD}-T_P}{T_{PD}}$. If $T_{PD} \leq 15$, we repeat the experiment 10 times and record the average running time. Detailed time statistics are given in the **Appendix A**.

Experiment results. With the setup above, 744 instances fall into **C1** and **C2**. Among the remaining 180 instances, 41 are in **C3** and 139 are in **C4**. Primal IPM shows benefits in **C4**. **Figure 1** (Left) plots the distribution of speedup from primal IPM in **C4**. Only six instances slow down when primal IPM is invoked and passes the test, and more than 30% of instances speed up by at least 30%. Around 10% of instances speed up by at least 50%. This result suggests primal IPM can be useful for a broad class of LPs.

5.3 Primal IPM for SDPs

We implement preconditioned primal IPM in `HDSDP` (called `PSDP`) and test on 21 instances with $m \geq 10n$ from the SDP benchmark [31] and the DIMACS challenge [30]. Both `HDSDP` and `PSDP` solve to 10^{-6} accuracy in DIMACS error. **Figure 1** (Right) illustrates the speed up using primal IPM when it is combined with a dual IPM method. When $m \gg n$, primal IPM significantly reduces the solution time for the SDP solver (up to 10 times faster). This observation aligns with our intuition since, in this case, most of the solution time is spent on forming and factorizing the normal matrix. Moreover, since SDP generally does not have superlinear convergence, the potential speedup due to switching to the primal IPM is larger.
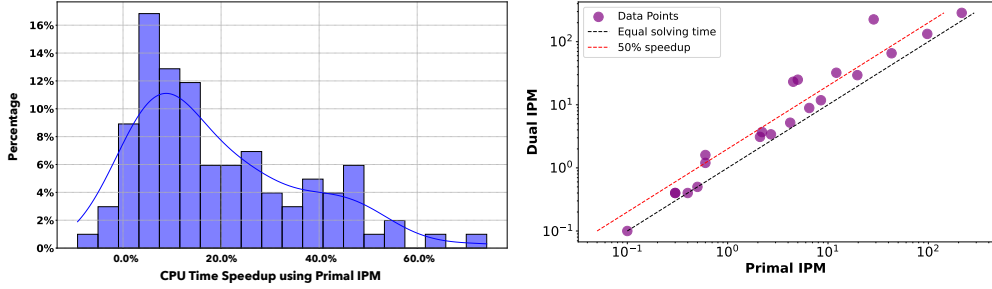
Figure 1: Left: distribution of primal IPM speedup over 139 LPs. Right: Dual vs hybrid dual-primal IPM on 21 SDPs. The hybrid approach is always faster, with at least 50% speedup on 6 instances.

5.4 Convergence Behavior of the Primal-Dual IPM Iteration Sequence

The convergence behavior of primal-dual IPM determines the performance improvement from switching to primal IPM: switching to primal IPM offers the largest benefit when primal-dual IPM converges slowly in the last iterations. Therefore, in this section, we discuss the convergence behavior of the primal-dual IPM iteration sequence. We take 924 instances from the MIPLIB and record $\{\|\mathbf{x}_{k+1} - \mathbf{x}_k\|\}, \{\|\mathbf{x}_{k+1} - \mathbf{x}_k\|_{\mathbf{x}_k,1}\}$ from primal-dual. We summarize the general convergence patterns of IPM as follows.

Consistent linear convergence (with plateau). All distance metrics except scaled distance decrease linearly (**Figure 2**) since early stage of IPM. There is no noticeable superlinear convergence behavior. IPM iterations may or may not reach a plateau region. Problems with zero objective $\mathbf{c} = \mathbf{0}$ are more likely to exhibit this type of convergence pattern. Primal IPM speeds up convergence by 30% to 60% on these problems.

Rapid convergence only at the end. **Figure 3** shows that for some problems, all the distance metrics decrease at the slow linear or sublinear rate in the initial iterations of IPM. In the last 2-5 iterations, (super)linear convergence appears, and iterates converge fast. More than half of the tested problems exhibit this convergence pattern. Whether primal IPM hastens convergence is determined by **1)** the convergence speed of thresholded distance; **2)** the duration of the superlinear convergence phase; **3)** the cost of factorizing the normal equations compared to forward/backward solve. The speedup from primal IPM is between 5% to 15% in this case. There is no benefit in applying primal IPM if superlinear convergence reaches the optimal solution in one or two iterations.

No clear convergence. For the rest of the problems (less than 5%), like the examples shown in **Figure 4**, the iteration sequence does not indicate clear convergence behavior. This case is less common and generally often happens when the primal solution has large-scale elements: the objective value converges while the solution sequence oscillates. Our theory excludes the use of primal IPM for such problems.
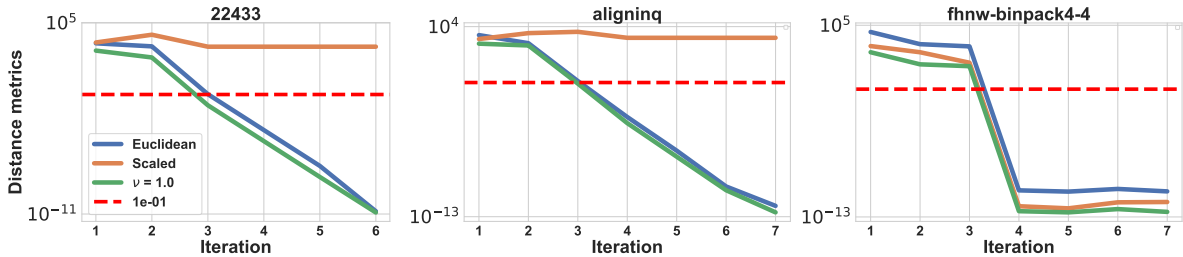

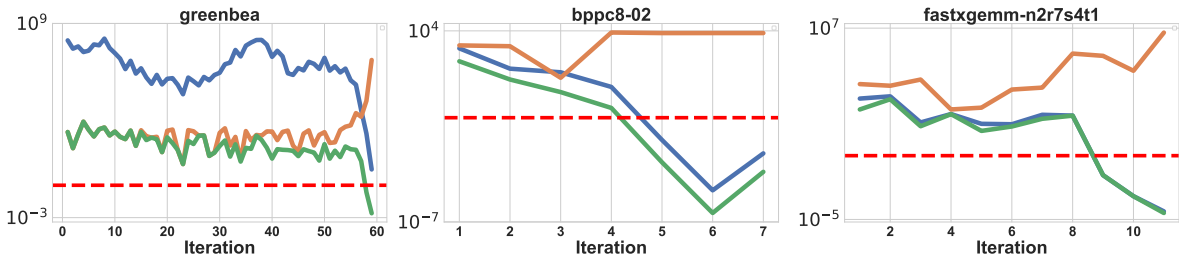Figure 2: Problems with convergence pattern: linear convergence

Figure 3: Problems with convergence pattern: fast convergence at the end. For problems like the leftmost figure, primal IPM is not useful. In contrast, for the second and third figures, the number of iterations after entering the fast convergence phase is comparable to the total iteration count, so primal IPM offers a speedup.
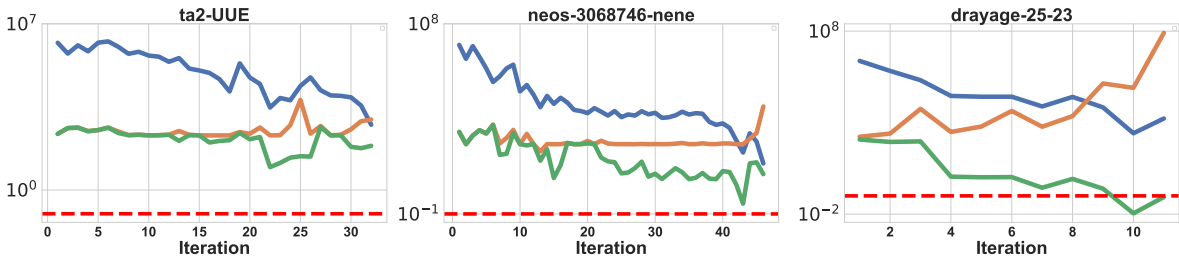

Figure 4: Problems with convergence pattern: no clear convergence. Often this pattern results when the solution **x** has extremely imbalanced coordinates.

6   Conclusions

In this paper, we revisit the primal interior point method and explore its potential advantages over the widely-used primal-dual IPM. We leverage the stability of the primal normal matrix to enhance computational efficiency in the primal IPM, particularly as the algorithm approaches convergence. Extensive numerical experiments demonstrate the practical efficiency of primal IPM on a broad class of LPs.

Acknowledgement

References

[1] Ilan Adler, Mauricio GC Resende, Geraldo Veiga, and Narendra Karmarkar. An implementation of kar-markar's algorithm for linear programming. *Mathematical programming*, 44:297–335, 1989.

[2] Erling D Andersen, Jacek Gondzio, Csaba Mészáros, Xiaojie Xu, et al. *Implementation of interior point methods for large scale linear programming*. HEC/Université de Geneve, 1996.

[3] Stefania Bellavia. Inexact interior-point method. *Journal of Optimization Theory and Applications*, 96:109–121, 1998.

[4] Stefania Bellavia, Jacek Gondzio, and Margherita Porcelli. An inexact dual logarithmic barrier method for solving sparse semidefinite programs. *Mathematical Programming*, 178:109–143, 2019.

[5] Steven J Benson and Yinyu Ye. Algorithm 875: Dsdp5-software for semidefinite programming. *ACM Transactions on Mathematical Software (TOMS)*, 34(3):1–20, 2008.

[6] Luca Bergamaschi, Jacek Gondzio, Ángeles Martínez, John W Pearson, and Spyridon Pougkakiotis. A new preconditioning approach for an interior point-proximal method of multipliers for linear and convex quadratic programming. *Numerical Linear Algebra with Applications*, 28(4):e2361, 2021.

[7] Luca Bergamaschi, Jacek Gondzio, Manolo Venturin, and Giovanni Zilli. Inexact constraint preconditioners for linear systems arising in interior point methods. *Computational Optimization and Applications*, 36:137–147, 2007.

[8] Luca Bergamaschi, Jacek Gondzio, and Giovanni Zilli. Preconditioning indefinite systems in interior point methods for optimization. *Computational Optimization and Applications*, 28:149–171, 2004.

[9] Agniva Chowdhury, Palma London, Haim Avron, and Petros Drineas. Faster randomized infeasible interior point methods for tall/wide linear programs. *Advances in Neural Information Processing Systems*, 33:8704–8715, 2020.

[10] Ya-Chi Chu, Luiz-Rafael Santos, and Madeleine Udell. Randomized nystr\" om preconditioned interior point-proximal method of multipliers. *arXiv preprint arXiv:2404.14524*, 2024.

[11] Stefano Cipolla and Jacek Gondzio. Proximal stabilized interior point methods and low-frequency-update preconditioning techniques. *Journal of Optimization Theory and Applications*, 197(3):1061–1103, 2023.

[12] Michael B Cohen, Yin Tat Lee, and Zhao Song. Solving linear programs in the current matrix multiplication time. *Journal of the ACM (JACM)*, 68(1):1–39, 2021.

[13] Daniel Dadush, Sophie Huiberts, Bento Natura, and László A Végh. A scaling-invariant algorithm for linear programming whose running time depends only on the constraint matrix. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pages 761–774, 2020.

[14] II Dikin. Iterative solution of problems of linear and quadratic programming. In *Doklady Akademii Nauk*, volume 174, pages 747–748. Russian Academy of Sciences, 1967.

[15] Roland W Freund, Florian Jarre, and Shinji Mizuno. Convergence of a class of inexact interior-point algorithms for linear programs. *Mathematics of Operations Research*, 24(1):50–71, 1999.

[16] Wenzhi Gao, Dongdong Ge, and Yinyu Ye. Hdsdp: Software for semidefinite programming. *arXiv preprint arXiv:2207.13862*, 2022.

[17] Ambros Gleixner, Gregor Hendel, Gerald Gamrath, Tobias Achterberg, Michael Bastubbe, Timo Berthold, Philipp M. Christophel, Kati Jarck, Thorsten Koch, Jeff Linderoth, Marco Lübbecke, Hans D. Mittelmann, Derya Ozyurt, Ted K. Ralphs, Domenico Salvagnin, and Yuji Shinano. MIPLIB 2017: Data-Driven Compilation of the 6th Mixed-Integer Programming Library. *Mathematical Programming Computation*, 2021.

[18] Jacek Gondzio. Interior point methods 25 years later. *European Journal of Operational Research*, 218(3):587–601, 2012.

[19] Jacek Gondzio. Matrix-free interior point method. *Computational Optimization and Applications*, 51:457–480, 2012.

[20] Jacek Gondzio, Spyridon Pougkakiotis, and John W Pearson. General-purpose preconditioning for regularized interior point methods. *Computational Optimization and Applications*, 83(3):727–757, 2022.

[21] Clovis C Gonzaga and Marli Cardia. Properties of the central points in linear programming problems. *Numerical Algorithms*, 35:185–204, 2004.

[22] Osman Güler, Dick den Hertog, Cornelis Roos, Tamas Terlaky, and Takashi Tsuchiya. Degeneracy in interior point methods for linear programming: a survey. *Annals of Operations Research*, 46:107–138, 1993.

[23] Osman Güler and Yinyu Ye. Convergence behavior of interior-point algorithms. *Mathematical Programming*, 60(1-3):215–228, 1993.

[24] Narendra Karmarkar. A new polynomial-time algorithm for linear programming. In *Proceedings of the sixteenth annual ACM symposium on Theory of computing*, pages 302–311, 1984.

[25] NK Karmarkar and KG Ramakrishnan. Computational results of an interior point algorithm for large scale linear programming. *Mathematical Programming*, 52:555–586, 1991.

[26] Yin Tat Lee and Aaron Sidford. Path finding methods for linear programming: Solving linear programs in o (vrank) iterations and faster algorithms for maximum flow. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, pages 424–433. IEEE, 2014.

[27] David G Luenberger, Yinyu Ye, et al. *Linear and nonlinear programming*, volume 2. Springer, 1984.

[28] Irvin J Lustig, Roy E Marsten, and David F Shanno. Interior point methods for linear programming: Computational state of the art. *ORSA Journal on Computing*, 6(1):1–14, 1994.

[29] Sanjay Mehrotra. On the implementation of a primal-dual interior point method. *SIAM Journal on optimization*, 2(4):575–601, 1992.

[30] Hans D Mittelmann. An independent benchmarking of sdp and socp solvers. *Mathematical Programming*, 95(2):407–430, 2003.

[31] Hans D Mittelmann. Benchmarking optimization software-a (hi) story. In *SN operations research forum*, volume 1, page 2. Springer, 2020.

[32] Cornelis Roos, Tamás Terlaky, and J-Ph Vial. Interior point methods for linear optimization. 2005.

[33] Cornelis Roos, Tamás Terlaky, and Jean-Philippe Vial. *Theory and algorithms for linear optimization: an interior point approach*. Wiley Chichester, 1997.

[34] Cornelis Roos and J Ph Vial. A polynomial method of approximate centers for linear programming. *Mathematical Programming*, 54:295–305, 1992.

[35] Yousef Saad. *Iterative methods for sparse linear systems*. SIAM, 2003.

[36] Lukas Schork and Jacek Gondzio. Implementation of an interior point method with basis preconditioning. *Mathematical Programming Computation*, 12(4):603–635, 2020.

[37] Michael J Todd. Scaling, shifting and weighting in interior-point methods. *Computational Optimization and Applications*, 3(4):305–315, 1994.

[38] Jan van den Brand. A deterministic linear program solver in current matrix multiplication time. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 259–278. SIAM, 2020.

[39] Stephen A Vavasis and Yinyu Ye. A primal-dual interior point method whose running time depends only on the constraint matrix. *Mathematical Programming*, 74(1):79–120, 1996.

[40] Marta Inez Velazco, Aurelio RL Oliveira, and Frederico Ferreira Campos. A note on hybrid preconditioners for large-scale normal equations arising from interior-point methods. *Optimization Methods & Software*, 25(2):321–332, 2010.

[41] Weichung Wang and Dianne P O'Leary. Adaptive use of iterative methods in predictor–corrector interior point methods for linear programming. *Numerical Algorithms*, 25:387–406, 2000.

[42] Stephen J Wright. *Primal-dual interior-point methods*. SIAM, 1997.

[43] Yinyu Ye. An $o(n^3l)$ potential reduction algorithm for linear programming. *Mathematical programming*, 50(1-3):239–258, 1991.

[44] Yinyu Ye. *Interior point algorithms: theory and analysis*. John Wiley & Sons, 2011.

[45] Filippo Zanetti and Jacek Gondzio. A new stopping criterion for krylov solvers applied in interior point methods. *SIAM Journal on Scientific Computing*, 45(2):A703–A728, 2023.

Appendix

Table of Contents

Structure of the Appendix   The appendix is organized as follows. In **Section B**, we prove the main results in our paper. **Section C** is a self-contained part that leverages our main results to establish convergence of inexact primal IPM.

A Detailed statistics on MIPLIB instances

Instance	Primal Time	Primal-dual Time	Speedup	Instance	Primal Time	Primal-dual Time	Speedup
allcolor58	1003.678	1045.208	0.041378	neos-4338804-snowy	0.187	0.223	0.193895
berlin_5_8_0	0.182	0.224	0.231925	neos-4797081-pakoka	1.585	2.116	0.334746
bnatt400	0.505	0.619	0.225389	neos-4971100-bodrog	5.441	5.660	0.040281
bnatt500	0.911	1.120	0.229308	neos-498623	0.574	0.587	0.023318
cdc7-4-3-2	144.080	164.713	0.143203	neos-873061	148.944	153.862	0.033016
chromaticindex128-5	13.771	15.711	0.140903	neos-885524	1.579	1.433	-0.092729
chromaticindex256-8	103.370	118.173	0.143205	neos-950242	2034.914	2143.839	0.053528
chromaticindex32-8	0.220	0.280	0.273078	neos-960392	7.323	9.232	0.260649
chromaticindex512-7	805.685	922.078	0.144465	neos9	64.425	74.273	0.152860
CMS750_4	19.897	24.745	0.243658	nexp-150-20-8-5	1.208	1.254	0.037911
cod105	0.113	0.160	0.413499	no-ip-64999	0.273	0.386	0.415453
cryptanalysiskb128n5obj14	27.533	38.303	0.391161	no-ip-65059	0.260	0.371	0.426903
cryptanalysiskb128n5obj16	28.689	39.961	0.392899	ns1208400	0.361	0.449	0.244861
cvrpa-n64k9vrpi	431.431	479.280	0.110908	ns1828997	3332.151	4068.890	0.221100
cvrpb-n45k5vrpi	51.542	56.801	0.102025	osorio-cta	0.608	0.685	0.125941
dale-cta	19.820	20.415	0.030033	probportfolio	0.028	0.030	0.067266
disctom	0.073	0.101	0.377128	pythago7824	4.204	6.261	0.489257
genus-g31-8	492.335	601.367	0.221459	pythago7825	4.618	6.877	0.489196
genus-g61-25	1050.716	1311.999	0.248671	queens-30	0.197	0.224	0.139101
graph20-80-1rand	7.089	7.496	0.057381	ramos3	0.494	0.656	0.327558
graph40-20-1rand	68.218	73.883	0.083047	scpj4scip	1.025	1.087	0.060606
graph40-40-1rand	958.515	956.266	-0.002346	scpk4	2.280	2.262	-0.008056
lectsched-2	2.234	2.859	0.279880	shipsched	2.680	3.053	0.139111
lectsched-4-obj	0.572	0.606	0.059112	shipschedule6shipsmixi	0.446	0.599	0.343217
lectsched-5-obj	11.697	13.078	0.118042	shipschedule8shipsmixuci	0.547	0.753	0.376145
mspsphard01i	0.356	0.467	0.310907	supportcase18	0.191	0.228	0.195104
neos-1324574	0.555	0.606	0.092235	supportcase21i	60.069	66.351	0.104575
neos-2978193-inde	1.100	1.204	0.094842	supportcase29	7.555	7.529	-0.003391
neos-2978205-isar	97.132	97.086	-0.000471	supportcase30	0.057	0.084	0.471918
neos-3004026-krka	13.471	20.117	0.493343	supportcase3	2.014	2.991	0.485151
neos-3135526-osun	0.655	0.715	0.091298	supportcase4	0.334	0.495	0.481269
neos-3211096-shag	1.335	1.903	0.425468	traininstance2	1.520	1.594	0.048841
neos-3355323-arnon	13.661	21.152	0.548381	tw-myciel4	5.853	6.203	0.059735
neos-3426132-dieze	0.171	0.203	0.189664	v150d30-2hopcds	46.815	50.653	0.081980
neos-3572885-glina	7.999	9.097	0.137252	van	20.796	22.234	0.069160
neos-3603137-hoteo	2.329	3.593	0.542789				

Table 1: Speedup on instances where primal IPM is triggered

B Proof of Results in Section 3

B.1 Auxiliary Results

This section uses the following auxiliary results on primal central path solution $\mathbf{x}^\mu$.

Lemma B.1 ([33]). *Under* **A1**, *if* $\delta(\mathbf{x}, \mu) \leq \lambda \leq \frac{\sqrt{2}}{2}$, *then*

$$\|\mathbf{x} - \mathbf{x}^\mu\|_{\mathbf{x}^\mu} \leq \sqrt{1 - \sqrt{1 - 2\lambda^2}}. \tag{16}$$

Lemma B.2. *Under* **A1**, *there exist* $\alpha, \gamma > 0$ *such that*

$$\gamma\mu \geq x_j^\mu, \quad \text{for all } j \in \mathcal{N},$$

where $\mathcal{B}, \mathcal{N}$ *is the maximal complementarity pair from* **Definition 3**.

Proof. Let $(\mathbf{x}^\star, \mathbf{s}^\star)$ be a strictly complementary solution. By definition,

$$\mathbf{x}_\mathcal{B}^\star > \mathbf{0} \quad \text{and} \quad \mathbf{s}_\mathcal{N}^\star > \mathbf{0}.$$

Given any $\mathbf{x}^\mu \in \mathcal{C}_p$, , there exists some dual feasible solution $(\mathbf{s}^\mu, \mathbf{y}^\mu)$ such that $\mathbf{X}^\mu \mathbf{s}^\mu = \mu\mathbf{e}$ [34]. Then,

$$\langle \mathbf{x}^\mu - \mathbf{x}^\star, \mathbf{s}^\mu - \mathbf{s}^\star \rangle = \langle \mathbf{x}^\mu - \mathbf{x}^\star, \mathbf{A}^\top(\mathbf{y}^\star - \mathbf{y}^\mu) \rangle = \langle \mathbf{A}(\mathbf{x}^\mu - \mathbf{x}^\star), \mathbf{y}^\star - \mathbf{y}^\mu \rangle = 0,$$

where the first equality holds since $\mathbf{s}^\mu = \mathbf{c} - \mathbf{A}^\top\mathbf{y}^\mu$ and $\mathbf{s}^\star = \mathbf{c} - \mathbf{A}^\top\mathbf{y}^\star$, and the last equality holds since $\mathbf{A}\mathbf{x}^\star = \mathbf{A}\mathbf{x}^\mu = \mathbf{b}$. On the other hand,

$$0 = \langle \mathbf{x}^\mu - \mathbf{x}^\star, \mathbf{s}^\mu - \mathbf{s}^\star \rangle = \mu n - \langle \mathbf{x}^\mu, \mathbf{s}^\star \rangle - \langle \mathbf{s}^\mu, \mathbf{x}^\star \rangle = \mu n - \langle \mathbf{x}_\mathcal{N}^\mu, \mathbf{s}_\mathcal{N}^\star \rangle - \langle \mathbf{s}_\mathcal{B}^\mu, \mathbf{x}_\mathcal{B}^\star \rangle, \tag{17}$$

where the second equality holds due to the fact that $\mathbf{X}^\mu \mathbf{s}^\mu = \mu\mathbf{e}$ and $\langle \mathbf{x}^\star, \mathbf{s}^\star \rangle = 0$, and the last one holds since $\mathbf{x}_\mathcal{N}^\star = \mathbf{0}$ and $\mathbf{s}_\mathcal{B}^\star = \mathbf{0}$. Next we rearrange the terms to get $\langle \mathbf{x}_\mathcal{N}^\mu, \mathbf{s}_\mathcal{N}^\star \rangle + \langle \mathbf{s}_\mathcal{B}^\mu, \mathbf{x}_\mathcal{B}^\star \rangle = \mu n$. The nonnegativity of all the entries implies that $x_j^\mu s_j^\star \leq \mu n$ for all $j \in \mathcal{N}$, which implies

$$x_j^\mu \leq \mu n(s_j^\star)^{-1} \leq \mu n(\|\mathbf{s}_\mathcal{N}^\star\|_{-\infty})^{-1}, \quad \text{for all } j \in \mathcal{N}.$$

Using (9), $\gamma = \frac{n}{\|\mathbf{s}_\mathcal{N}^\star\|_{-\infty}}$, this completes the proof. $\qquad\square$

Lemma B.3. *Given* $\mathbf{x} \in \mathcal{F}_p^0, \mu \in (0, 1]$ *such that* $\delta(\mathbf{x}, \mu) \leq \lambda = \frac{1}{2}$, *there exists some* $\alpha > 0$ *such that*

$$\|\mathbf{x}\|_{-\infty} \geq \alpha\mu$$

Proof. Again we use $\mathbf{x}^\mu, \mathbf{s}^\mu$ to denote primal and dual solutions on the central path with barrier parameter μ. Now take any $\mathbf{x}' \in \mathcal{F}_p^0, \mathbf{s}' \in \mathcal{F}_d^0$ and

$$0 = \langle \mathbf{x}^\mu - \mathbf{x}', \mathbf{s}^\mu - \mathbf{s}' \rangle = \mu n - \langle \mathbf{x}^\mu, \mathbf{s}' \rangle - \langle \mathbf{s}^\mu, \mathbf{x}' \rangle + \langle \mathbf{x}', \mathbf{s}' \rangle.$$

Next rearrangement gives $\langle \mathbf{s}^\mu, \mathbf{x}' \rangle + \langle \mathbf{x}^\mu, \mathbf{s}' \rangle \leq \mu n + \langle \mathbf{x}', \mathbf{s}' \rangle$ and

$$\mu(x_j^\mu)^{-1}x_j' = s_j^\mu x_j' \leq \langle \mathbf{s}^\mu, \mathbf{x}' \rangle + \langle \mathbf{x}^\mu, \mathbf{s}' \rangle \leq \mu n + \langle \mathbf{x}', \mathbf{s}' \rangle,$$

for all j, which implies $x_j^\mu \geq \frac{\mu x_j'}{\mu n + \langle \mathbf{x}', \mathbf{s}' \rangle} \geq \frac{\mu x_j'}{n + \langle \mathbf{x}', \mathbf{s}' \rangle}$ since $\mu \leq 1$. Now take some $\mathbf{x} \in \mathcal{F}_p^0$ and by **Lemma B.1**,

$$\max_{1 \leq j \leq n} \left| \frac{x_j - x_j^\mu}{x_j^\mu} \right| = \|(\mathbf{X}^\mu)^{-1}(\mathbf{x} - \mathbf{x}^\mu)\|_\infty \leq \|\mathbf{x} - \mathbf{x}^\mu\|_{\mathbf{x}^\mu} \leq \sqrt{1 - \sqrt{1 - \frac{1}{2}}} < 0.9.$$

Hence for each j,

$$x_j \geq (1 - 0.9) \cdot x_j^\mu \geq \frac{0.1 x_j'}{n + \langle \mathbf{x}', \mathbf{s}' \rangle} \mu$$

and taking $\alpha = \frac{0.1 \|\mathbf{x}'\|_{-\infty}}{n + \langle \mathbf{x}', \mathbf{s}' \rangle} > 0$ completes the proof. $\qquad \square$

Lemma B.4 (ℓ_∞ version of shifted scaling [21]). *Suppose* $\mathbf{x}, \mathbf{z} \in \mathcal{F}_p^0$ *and* $\|\mathbf{X}^{-1}(\mathbf{z} - \mathbf{x})\|_\infty \leq \rho \in [0, \frac{1}{4}]$. *Then*

$$\|(\mathbf{X}^{-1}\mathbf{Z}\mathbf{P_{AZ}}\mathbf{X}^{-1}\mathbf{Z} - \mathbf{P_{AX}})\mathbf{v}\| \leq 3\rho \|\mathbf{P_{AX}}\mathbf{v}\|$$

for all $\mathbf{v} \in \mathbb{R}^n$

B.2   Proof of Lemma 3.1

Denote $\Delta := \sqrt{1 - \sqrt{1 - 2\lambda^2}} \leq 1$. For any $\mathbf{x}^1, \mathbf{x}^2 \in \mathcal{P}_\mu$, (16) of **Lemma B.1** suggests that

$$\|\mathbf{x}^1 - \mathbf{x}^\mu\|_{\mathbf{x}^\mu} \leq \Delta \quad \text{and} \quad \|\mathbf{x}^2 - \mathbf{x}^\mu\|_{\mathbf{x}^\mu} \leq \Delta. \tag{18}$$

By triangle inequality, the scaled distance between $\mathbf{x}^1$ and $\mathbf{x}^2$ is bounded by

$$\|\mathbf{x}^1 - \mathbf{x}^2\|_{\mathbf{x}^\mu} \leq \|\mathbf{x}^1 - \mathbf{x}^\mu\|_{\mathbf{x}^\mu} + \|\mathbf{x}^2 - \mathbf{x}^\mu\|_{\mathbf{x}^\mu} \leq 2\Delta.$$

Squaring both sides and partitioning the coordinates by $\mathcal{B}, \mathcal{N}$,

$$\|\mathbf{x}^1 - \mathbf{x}^2\|_{\mathbf{x}^\mu}^2 = \|(\mathbf{X}_\mathcal{N}^\mu)^{-1}(\mathbf{x}_\mathcal{N}^1 - \mathbf{x}_\mathcal{N}^2)\|^2 + \|(\mathbf{X}_\mathcal{B}^\mu)^{-1}(\mathbf{x}_\mathcal{B}^1 - \mathbf{x}_\mathcal{B}^2)\|^2 \leq 4\Delta^2$$

where $\mathbf{X}_\mathcal{N}^\mu = \text{diag}(\mathbf{x}_\mathcal{N}^\mu)$ and $\mathbf{X}_\mathcal{B}^\mu = \text{diag}(\mathbf{x}_\mathcal{B}^\mu)$. Invoke the lower bound from **Lemma B.2**,

$$\frac{1}{\gamma^2 \mu^2} \|\mathbf{x}_\mathcal{N}^1 - \mathbf{x}_\mathcal{N}^2\|^2 \leq \|(\mathbf{X}_\mathcal{N}^\mu)^{-1}(\mathbf{x}_\mathcal{N}^1 - \mathbf{x}_\mathcal{N}^2)\|^2 \leq 4\Delta^2 \leq 4 \tag{19}$$

and thus $\|\mathbf{x}_\mathcal{N}^1 - \mathbf{x}_\mathcal{N}^2\| \leq 2\gamma\mu$. Next, we bound the term $\|\mathbf{x}_\mathcal{B}^1 - \mathbf{x}_\mathcal{B}^2\|$. Since

$$\mathbf{A}_\mathcal{B}\mathbf{x}_\mathcal{B}^1 + \mathbf{A}_\mathcal{N}\mathbf{x}_\mathcal{N}^1 = \mathbf{b} \quad \text{and} \quad \mathbf{A}_\mathcal{B}\mathbf{x}_\mathcal{B}^2 + \mathbf{A}_\mathcal{N}\mathbf{x}_\mathcal{N}^2 = \mathbf{b}, \tag{20}$$

we subtract them to get $\mathbf{A}_\mathcal{B}(\mathbf{x}_\mathcal{B}^1 - \mathbf{x}_\mathcal{B}^2) = -\mathbf{A}_\mathcal{N}(\mathbf{x}_\mathcal{N}^1 - \mathbf{x}_\mathcal{N}^2)$, which, associated with **A2**, implies that

$$\|\mathbf{x}_\mathcal{B}^1 - \mathbf{x}_\mathcal{B}^2\| = \|(\mathbf{A}_\mathcal{B}^\top \mathbf{A}_\mathcal{B})^{-1}\mathbf{A}_\mathcal{B}^\top \mathbf{A}_\mathcal{N}(\mathbf{x}_\mathcal{N}^1 - \mathbf{x}_\mathcal{N}^2)\| \leq \|(\mathbf{A}_\mathcal{B}^\top \mathbf{A}_\mathcal{B})^{-1}\mathbf{A}_\mathcal{B}^\top\| \cdot \|\mathbf{A}_\mathcal{N}\| \cdot \|\mathbf{x}_\mathcal{N}^1 - \mathbf{x}_\mathcal{N}^2\|.$$

Using $\|\mathbf{U}\|^2 = \|\mathbf{U}\mathbf{U}^\top\|$,

$$\|(\mathbf{A}_\mathcal{B}^\top \mathbf{A}_\mathcal{B})^{-1}\mathbf{A}_\mathcal{B}^\top\|^2 = \|(\mathbf{A}_\mathcal{B}^\top \mathbf{A}_\mathcal{B})^{-1}\mathbf{A}_\mathcal{B}^\top \mathbf{A}_\mathcal{B}(\mathbf{A}_\mathcal{B}^\top \mathbf{A}_\mathcal{B})^{-1}\| = \|(\mathbf{A}_\mathcal{B}^\top \mathbf{A}_\mathcal{B})^{-1}\| = \frac{1}{\lambda_{\min}(\mathbf{A}_\mathcal{B}^\top \mathbf{A}_\mathcal{B})}.$$

Finally, we put things together to get

$$\|\mathbf{x}^1 - \mathbf{x}^2\| \leq \|\mathbf{x}_\mathcal{N}^1 - \mathbf{x}_\mathcal{N}^2\| + \|\mathbf{x}_\mathcal{B}^1 - \mathbf{x}_\mathcal{B}^2\| \leq 2\gamma\mu\left(1 + \frac{\|\mathbf{A}_\mathcal{N}\|}{\sqrt{\lambda_{\min}(\mathbf{A}_\mathcal{B}^\top \mathbf{A}_\mathcal{B})}}\right),$$

and this completes the proof.

B.3   Proof of Theorem 3.1

Consider $\mathbf{x}^{\mu_k}, \mathbf{x}^{\mu_{k+1}}$ and let $\mathbf{s}^{\mu_k}, \mathbf{s}^{\mu_{k+1}}$ denote the corresponding central path dual solution. We have

$$\delta(\mathbf{x}^{\mu_{k+1}}, \mu_k) = \min_{(\mathbf{y}, \mathbf{s}) \in \mathcal{F}_d} \|\frac{1}{\mu_k}\mathbf{X}^{\mu_{k+1}}\mathbf{s} - \mathbf{e}\| \leq \|\frac{1}{\mu_k}\mathbf{X}^{\mu_{k+1}}\mathbf{s}^{\mu_{k+1}} - \mathbf{e}\| = \|\frac{\mu_{k+1}}{\mu_k}\mathbf{e} - \mathbf{e}\| = \tau\sqrt{n} \leq \frac{\sqrt{2}}{2}$$

and $\mathbf{x}_{\mu_{k+1}} \in \mathcal{P}_{\mu_k}$. Since $\delta(\mathbf{x}_k, \mu_k) \leq \frac{\sqrt{2}}{2}$ for all $\mathbf{x}_k$ and μ_k, for all $i \geq k$, the fact that $\mathbf{x}_i \in \mathcal{P}_{\mu_i}$ implies that $\mathbf{x}_i \in \bigcup_{l=k}^{\infty} \mathcal{P}_{\mu_l}$. Then, for any $i, j \geq k$,

$$\|\mathbf{x}_i - \mathbf{x}_j\| \leq \operatorname{diam}\left(\bigcup_{l=k}^{\infty} \mathcal{P}_{\mu_l}\right) \leq \sum_{l=k}^{\infty} \operatorname{diam}(\mathcal{P}_{\mu_l})$$

where the last inequality holds because $\mathbf{x}_{\mu_{k+1}} \in \mathcal{P}_{\mu_k}$ implies that $\mathcal{P}_{\mu_k} \cap \mathcal{P}_{\mu_{k+1}} \neq \varnothing$ holds for all k. Based on **Lemma 3.1**, we have

$$\|\mathbf{x}_i - \mathbf{x}_j\| \leq 2\gamma(1 + \frac{\|\mathbf{A}_{\mathcal{N}}\|}{\sqrt{\lambda_{\min}(\mathbf{A}_{\mathcal{B}}^{\top}\mathbf{A}_{\mathcal{B}})}})\sum_{l=k}^{\infty} \mu_l = 2\gamma(1 + \frac{\|\mathbf{A}_{\mathcal{N}}\|}{\sqrt{\lambda_{\min}(\mathbf{A}_{\mathcal{B}}^{\top}\mathbf{A}_{\mathcal{B}})}}) \cdot \mu_k \tau^{-1} \tag{21}$$

and this completes the proof.

B.4   Proof of Lemma 3.2

Using $|2xy| \leq \gamma x^2 + \gamma^{-1} y^2$, for any $0 \leq \gamma \leq 1$,

$$(1 - \gamma)\mathbf{X}^2 + (1 - \gamma^{-1})\Delta\mathbf{X}^2 \preceq (\mathbf{X} + \Delta\mathbf{X})^2 \preceq (1 + \gamma)\mathbf{X}^2 + (1 + \gamma^{-1})\Delta\mathbf{X}^2$$

Now, since $\mathbf{A}$ is full rank,

$$(1 - \gamma)\mathbf{M}_{\mathbf{X}} + (1 - \gamma^{-1})\mathbf{A}\Delta\mathbf{X}^2\mathbf{A}^{\top} \preceq \mathbf{M}_{\mathbf{X}+\Delta\mathbf{X}} \preceq (1 + \gamma)\mathbf{M}_{\mathbf{X}} + (1 + \gamma^{-1})\mathbf{A}\Delta\mathbf{X}^2\mathbf{A}^{\top}.$$

Multiplying $\mathbf{M}_{\mathbf{X}}^{-1/2}$ on both sides and using $\mathbf{M}_{\mathbf{X}} \succeq \lambda_{\mathbf{x}} \cdot \mathbf{I}$,

$$\left[(1 - \gamma) + (1 - \gamma^{-1})\frac{\|\mathbf{A}\|^2\|\Delta\mathbf{x}\|_{\infty}^2}{\lambda_{\mathbf{x}}}\right] \cdot \mathbf{I} \preceq \mathbf{M}_{\mathbf{X}}^{-1/2}(\mathbf{M}_{\mathbf{X}+\Delta\mathbf{x}})\mathbf{M}_{\mathbf{X}}^{-1/2} \preceq \left[(1 + \gamma) + (1 + \gamma^{-1})\frac{\|\mathbf{A}\|^2\|\Delta\mathbf{x}\|_{\infty}^2}{\lambda_{\mathbf{x}}}\right] \cdot \mathbf{I}.$$

Since $\|\Delta\mathbf{x}\|_{\infty} \leq \frac{\sqrt{\lambda_{\mathbf{x}}}}{\|\mathbf{A}\|}\beta$, this implies $\frac{1}{\lambda_{\mathbf{x}}}\|\mathbf{A}\|^2\|\Delta\mathbf{x}\|_{\infty}^2 \leq \beta^2 \leq 1$ and

$$(1 - \gamma) + (1 - \gamma^{-1})\frac{\|\mathbf{A}\|^2\|\Delta\mathbf{x}\|_{\infty}^2}{\lambda_{\mathbf{x}}} = (1 - \gamma^{-1})\left(\gamma - \frac{\|\mathbf{A}\|^2\|\Delta\mathbf{x}\|_{\infty}^2}{\lambda_{\mathbf{x}}}\right) \geq 0$$

for all $\frac{1}{\lambda_{\mathbf{x}}}\|\mathbf{A}\|^2\|\Delta\mathbf{x}\|_{\infty}^2 \leq \gamma \leq 1$. Last taking $\gamma = \sqrt{\frac{1}{\lambda_{\mathbf{x}}}\|\mathbf{A}\|^2\|\Delta\mathbf{x}\|_{\infty}^2} \geq \frac{1}{\lambda_{\mathbf{x}}}\|\mathbf{A}\|^2\|\Delta\mathbf{x}\|_{\infty}^2$, we get

$$(\gamma - 1)^2 \cdot \mathbf{I} \preceq \mathbf{M}_{\mathbf{X}}^{-1/2}(\mathbf{M}_{\mathbf{X}+\Delta\mathbf{x}})\mathbf{M}_{\mathbf{X}}^{-1/2} \preceq (\gamma + 1)^2 \cdot \mathbf{I}$$

Notice that $\gamma = \frac{1}{\sqrt{\lambda_{\mathbf{x}}}}\|\mathbf{A}\|\|\Delta\mathbf{x}\|_{\infty} \leq \frac{1}{\sqrt{\lambda_{\mathbf{x}}}}\|\mathbf{A}\|\|\Delta\mathbf{x}\| \leq \beta \leq 1$, we have

$$(1 - \beta)^2 \cdot \mathbf{I} \preceq \mathbf{M}_{\mathbf{X}}^{-1/2}(\mathbf{M}_{\mathbf{X}+\Delta\mathbf{x}})\mathbf{M}_{\mathbf{X}}^{-1/2} \preceq (\beta + 1)^2 \cdot \mathbf{I}$$

and thus

$$\kappa(\mathbf{M}_{\mathbf{X}}^{-1/2}(\mathbf{M}_{\mathbf{X}+\Delta\mathbf{x}})\mathbf{M}_{\mathbf{X}}^{-1/2}) \leq (\frac{\beta+1}{\beta-1})^2,$$

which completes the proof.

B.5   Proof of Theorem 3.2

By (21) in the proof of **Theorem 3.1**, for all $j \geq k \geq k^{\star}$, we have $\mu_k \leq \frac{\sqrt{\lambda^{\star}}\tau}{16\gamma\|\mathbf{A}\|} \cdot (1 + \frac{\|\mathbf{A}_{\mathcal{N}}\|}{\lambda_{\min}(\mathbf{A}_{\mathcal{B}}^{\top}\mathbf{A}_{\mathcal{B}})})^{-1}$ and

$$\|\mathbf{x}_j - \mathbf{x}_k\| \leq \frac{2\gamma\mu_k}{\tau}(1 + \frac{\|\mathbf{A}_{\mathcal{N}}\|}{\sqrt{\lambda_{\min}(\mathbf{A}_{\mathcal{B}}^{\top}\mathbf{A}_{\mathcal{B}})}}) \leq \frac{\sqrt{\lambda^{\star}}}{8\|\mathbf{A}\|}.$$

In particular, $\|\mathbf{x}_k - \mathbf{x}^{\star}\| \leq \frac{\sqrt{\lambda^{\star}}}{8\|\mathbf{A}\|}$ and using **Lemma 3.2**, $\mathbf{M}_{\mathbf{X}_k}^{-1/2}\mathbf{M}_{\mathbf{X}^{\star}}\mathbf{M}_{\mathbf{X}_k}^{-1/2} \preceq (\beta+1)^2 \cdot \mathbf{I}$, we have $\mathbf{M}_{\mathbf{X}_k} \succeq \frac{1}{4}\lambda^{\star}$. Finally, invoking **Lemma 3.2** again with $\beta = \frac{1}{2}, \mathbf{x} = \mathbf{x}_k, \lambda_{\mathbf{x}} = \frac{1}{4}\lambda^{\star}$ and $\Delta\mathbf{x} = \mathbf{x}_k - \mathbf{x}_j$ completes the proof.

B.6 Proof of Theorem 3.4

Let $\mathbf{v} = \mathbf{X}(\frac{1}{\mu}\mathbf{c} - \mathbf{X}^{-1}\mathbf{e})$, $\theta = \|\mathbf{z} - \mathbf{x}\|_{\mathbf{x},\nu}$, and define the partition $\mathcal{L} := \{j : x_j \geq \nu\}, \mathcal{S} := \{j : x_j < \nu\}$. Recall that

$$\Delta\mathbf{x} = -\mathbf{X}(\mathbf{I} - \mathbf{X}\mathbf{A}^\top(\mathbf{A}\mathbf{X}^2\mathbf{A}^\top)^{-1}\mathbf{A}\mathbf{X})\mathbf{v}.$$

We show how to find $\Delta\hat{\mathbf{x}}$ combining shifted scaling and preconditioning. Let $\mathbf{w} \in \mathbb{R}^n$ be such that $\mathbf{w}_\mathcal{S} = \mathbf{x}_\mathcal{S}$ and $\mathbf{w}_\mathcal{L} = \mathbf{z}_\mathcal{L}$, and define

$$\Delta\tilde{\mathbf{x}} = -\mathbf{W}(\mathbf{I} - \mathbf{W}\mathbf{A}^\top(\mathbf{A}\mathbf{W}^2\mathbf{A}^\top)^{-1}\mathbf{A}\mathbf{W})\mathbf{W}\mathbf{X}^{-1}\mathbf{v}.$$

By the definition of ν-thresholded scaled distance,

$$\|\mathbf{X}^{-1}(\mathbf{x} - \mathbf{w})\|_\infty = \|\mathbf{X}_\mathcal{L}^{-1}(\mathbf{x}_\mathcal{L} - \mathbf{z}_\mathcal{L})\|_\infty \leq \|\mathbf{z} - \mathbf{x}\|_{\mathbf{x},\nu} \leq \tfrac{1}{4}.$$

Then, we apply **Lemma B.4** and deduce that

$$\begin{aligned}
&\|\mathbf{X}^{-1}(\Delta\mathbf{x} - \Delta\tilde{\mathbf{x}})\| \\
&= \|(\mathbf{I} - \mathbf{X}\mathbf{A}^\top(\mathbf{A}\mathbf{X}^2\mathbf{A}^\top)^{-1}\mathbf{A}\mathbf{X})\mathbf{v} - \mathbf{X}^{-1}\mathbf{W}(\mathbf{I} - \mathbf{W}\mathbf{A}^\top(\mathbf{A}\mathbf{W}^2\mathbf{A}^\top)^{-1}\mathbf{A}\mathbf{W})\mathbf{W}\mathbf{X}^{-1}\mathbf{v}\| \\
&\leq 3\theta\|(\mathbf{I} - \mathbf{X}\mathbf{A}^\top(\mathbf{A}\mathbf{X}^2\mathbf{A}^\top)^{-1}\mathbf{A}\mathbf{X})\mathbf{v}\| \\
&= 3\theta\delta(\mathbf{x},\mu),
\end{aligned}$$

where the last equality uses the fact $\delta(\mathbf{x},\mu) = \|(\mathbf{I} - \mathbf{X}\mathbf{A}^\top(\mathbf{A}\mathbf{X}^2\mathbf{A}^\top)^{-1}\mathbf{A}\mathbf{X})\mathbf{X}(\frac{1}{\mu}\mathbf{c} - \mathbf{X}^{-1}\mathbf{e})\|$ from **Definition 2**. Plugging in condition $\theta \leq \frac{\varepsilon}{6\delta(\mathbf{x},\mu)}$ gives $\|\mathbf{X}^{-1}(\Delta\mathbf{x} - \Delta\tilde{\mathbf{x}})\| \leq \frac{\varepsilon}{2}$.

Next consider solving $(\mathbf{A}\mathbf{W}^2\mathbf{A}^\top)^{-1}\mathbf{A}\mathbf{W}(\mathbf{W}\mathbf{X}^{-1})\mathbf{v}$. We have $\|\mathbf{w} - \mathbf{z}\| = \|\mathbf{w}_\mathcal{S} - \mathbf{z}_\mathcal{S}\| \leq \theta$ and for $\theta \leq \frac{\sqrt{\lambda_\mathbf{z}}}{2\|\mathbf{A}\|}$, we invoke **Lemma 3.2** with $\beta = \frac{1}{2}$ to get $\mathbf{A}\mathbf{W}^2\mathbf{A}^\top \succeq \frac{1}{4}\mathbf{A}\mathbf{Z}^2\mathbf{A}^\top \succeq \frac{\lambda_\mathbf{z}}{4}$ and

$$\kappa((\mathbf{A}\mathbf{Z}^2\mathbf{A}^\top)^{-1/2}\mathbf{A}\mathbf{W}^2\mathbf{A}^\top(\mathbf{A}\mathbf{Z}^2\mathbf{A}^\top)^{-1/2}) \leq 9,$$

Finally, we analyze the effect of inexactness. Suppose we obtain $\hat{\mathbf{d}}$ satisfying

$$\mathbf{A}\mathbf{W}^2\mathbf{A}^\top\hat{\mathbf{d}} = \mathbf{A}\mathbf{W}(\mathbf{W}\mathbf{X}^{-1})\mathbf{v} + \boldsymbol{\zeta}_\mathbf{w}$$

and the inexact solution $\Delta\hat{\mathbf{x}}$ is chosen as

$$\begin{aligned}
\Delta\hat{\mathbf{x}} &= \mathbf{W}^2\mathbf{X}^{-1}\mathbf{v} - \mathbf{W}^2\mathbf{A}^\top\hat{\mathbf{d}} \\
&= \mathbf{W}^2\mathbf{X}^{-1}\mathbf{v} - \mathbf{W}^2\mathbf{A}^\top(\mathbf{A}\mathbf{W}^2\mathbf{A}^\top)^{-1}(\mathbf{A}\mathbf{W}(\mathbf{W}\mathbf{X}^{-1})\mathbf{v} + \boldsymbol{\zeta}_\mathbf{w}),
\end{aligned}$$

which implies

$$\|\mathbf{X}^{-1}(\Delta\tilde{\mathbf{x}} - \Delta\hat{\mathbf{x}})\| = \|\mathbf{X}^{-1}\mathbf{W}^2\mathbf{A}^\top(\mathbf{A}\mathbf{W}^2\mathbf{A}^\top)^{-1}\boldsymbol{\zeta}_\mathbf{w}\| \leq \frac{\|\boldsymbol{\zeta}_\mathbf{w}\| \cdot \|\mathbf{X}^{-1}\mathbf{w}\|_\infty}{\sqrt{\lambda_{\min}(\mathbf{A}\mathbf{W}^2\mathbf{A}^\top)}} \leq \frac{2(\theta+1)}{\sqrt{\lambda_\mathbf{z}}}\|\boldsymbol{\zeta}_\mathbf{w}\|$$

where the last inequality uses $\|\mathbf{X}^{-1}\mathbf{w}\|_\infty = \|\mathbf{X}^{-1}(\mathbf{w} - \mathbf{x}) + \mathbf{e}\|_\infty \leq \theta + 1$. Using an iterative solver, such as conjugate gradient, and preconditioned by $\mathbf{A}\mathbf{Z}^2\mathbf{A}^\top$, we can achieve $\|\mathbf{X}^{-1}(\Delta\tilde{\mathbf{x}} - \Delta\hat{\mathbf{x}})\| \leq \frac{\varepsilon}{2}$ in $\mathcal{O}(\log(\frac{4(\theta+1)}{\varepsilon\sqrt{\lambda_\mathbf{z}}})) = \mathcal{O}(\log(\frac{1}{\varepsilon}))$ iterations. Finally, we put things together to get

$$\|\mathbf{X}^{-1}(\Delta\mathbf{x} - \Delta\hat{\mathbf{x}})\| \leq \|\mathbf{X}^{-1}(\Delta\tilde{\mathbf{x}} - \Delta\hat{\mathbf{x}})\| + \|\mathbf{X}^{-1}(\Delta\mathbf{x} - \Delta\tilde{\mathbf{x}})\| \leq \varepsilon.$$

This completes the proof.

Remark 1. We remark that $\Delta\tilde{\mathbf{x}}$ may not satisfy $\mathbf{A}\Delta\tilde{\mathbf{x}} = \mathbf{0}$ due to inexactness and will violate primal feasibility. We address this issue in **Section C**.

C   Inexact Primal Interior Point Method and Proof of Theorem 3.3

C.1   Algorithm Design

This section provides convergence analysis of **Algorithm 2**. We start by defining inexactness. Note that inexact direction $\Delta \mathbf{x}$ can be re-written as

$$
\begin{aligned}
\Delta \mathbf{x} &= -\mathbf{X}(\mathbf{I} - \mathbf{X}\mathbf{A}^\top(\mathbf{A}\mathbf{X}^2\mathbf{A}^\top)^{-1}\mathbf{A}\mathbf{X})(\tfrac{1}{\mu}\mathbf{X}\mathbf{c} - \mathbf{e}) \\
&= -(\tfrac{1}{\mu}\mathbf{X}^2\mathbf{c} - \mathbf{x}) + \mathbf{X}^2\mathbf{A}^\top(\mathbf{A}\mathbf{X}^2\mathbf{A}^\top)^{-1}(\tfrac{1}{\mu}\mathbf{A}\mathbf{X}^2\mathbf{c} - \mathbf{A}\mathbf{x}) \\
&= -(\tfrac{1}{\mu}\mathbf{X}^2\mathbf{c} - \mathbf{x}) + \mathbf{X}^2\mathbf{A}^\top\mathbf{d},
\end{aligned}
$$

where the normal equation $\mathbf{A}\mathbf{X}^2\mathbf{A}^\top\mathbf{d} = \tfrac{1}{\mu}\mathbf{A}\mathbf{X}^2\mathbf{c} - \mathbf{A}\mathbf{x}$ is solved inexactly. Suppose an error $\boldsymbol{\zeta}_\mathbf{x}$ is incurred:

$$
\mathbf{A}\mathbf{X}^2\mathbf{A}^\top\hat{\mathbf{d}} = \tfrac{1}{\mu}\mathbf{A}\mathbf{X}^2\mathbf{c} - \mathbf{A}\mathbf{x} + \boldsymbol{\zeta}_\mathbf{x} \tag{22}
$$

If we directly substitute back, $\Delta\hat{\mathbf{x}} = -(\tfrac{1}{\mu}\mathbf{X}^2\mathbf{c} - \mathbf{x}) + \mathbf{X}^2\mathbf{A}^\top\hat{\mathbf{d}}$ and

$$
\begin{aligned}
\mathbf{A}\Delta\hat{\mathbf{x}} &= -\mathbf{A}(\tfrac{1}{\mu}\mathbf{X}^2\mathbf{c} - \mathbf{x}) + \mathbf{A}\mathbf{X}^2\mathbf{A}^\top\hat{\mathbf{d}} \\
&= -\mathbf{A}(\tfrac{1}{\mu}\mathbf{X}^2\mathbf{c} - \mathbf{x}) + \tfrac{1}{\mu}\mathbf{A}\mathbf{X}^2\mathbf{c} - \mathbf{A}\mathbf{x} + \boldsymbol{\zeta}_\mathbf{x} = \boldsymbol{\zeta}_\mathbf{x}.
\end{aligned}
$$

In other words, the error in solving the normal equation propagates and results in primal infeasibility, which is undesirable. There is a simple fix: take arbitrary basis matrix $\mathbf{B}$ of $\mathbf{A}$ and denote the column partition of $\mathbf{A}$ as (B, N). Then we can take $\boldsymbol{\lambda}_\mathbf{x}$ such that $\boldsymbol{\lambda}_{\mathbf{x},B} = \mathbf{B}^{-1}\boldsymbol{\zeta}_\mathbf{x}$ and $\boldsymbol{\lambda}_{\mathbf{x},N} = \mathbf{0}$. Taking $\Delta\hat{\mathbf{x}} = -(\tfrac{1}{\mu}\mathbf{X}^2\mathbf{c} - \mathbf{x}) + \mathbf{X}^2\mathbf{A}^\top\hat{\mathbf{d}} + \boldsymbol{\lambda}_\mathbf{x}$, this gives $\mathbf{A}\Delta\hat{\mathbf{x}} = \mathbf{0}$. For brevity of analysis, we take

$$
\boldsymbol{\lambda}_\mathbf{x} := \operatorname*{arg\,min}_{\mathbf{A}\boldsymbol{\lambda} = -\boldsymbol{\zeta}_\mathbf{x}} \|\boldsymbol{\lambda}\| = -\mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}\boldsymbol{\zeta}_\mathbf{x}.
$$

From now on we formally define $\Delta\hat{\mathbf{x}} := -(\tfrac{1}{\mu}\mathbf{X}^2\mathbf{c} - \mathbf{x}) + \mathbf{X}^2\mathbf{A}^\top\hat{\mathbf{d}} + \boldsymbol{\lambda}_\mathbf{x}$.

To analyze the convergence of **Algorithm 2**, we first present auxiliary results in **Section C.2**. We prove the complexity of **Algorithm 2** in **Section C.3**. Without loss of generality, in this section, we drop the iteration index k and only consider two consecutive IPM iterations $\mathbf{x}, \mathbf{x}^+$ with barrier parameters μ, μ^+ respectively. Given $\mathbf{x} \in \mathcal{F}_p^0$, we define the projected dual solution

$$
\mathbf{s}(\mathbf{x}, \mu) := \operatorname*{arg\,min}_{(\mathbf{y}, \mathbf{s}) \in \mathcal{F}_d} \|\tfrac{1}{\mu}\mathbf{X}\mathbf{s} - \mathbf{e}\|
$$

and let $\mathbf{z} := \tfrac{1}{\mu}\mathbf{X}\mathbf{s}(\mathbf{x}, \mu)$.

C.2   Auxiliary Results

Lemma C.1 (Primal step [34])**.** *The following identities hold*

$$
\mathbf{P}_{\mathbf{A}\mathbf{X}}(\tfrac{1}{\mu}\mathbf{X}\mathbf{c} - \mathbf{e}) = \tfrac{1}{\mu}\mathbf{X}\mathbf{s}(\mathbf{x}, \mu) - \mathbf{e} \tag{23}
$$

$$
\Delta\mathbf{x} = \mathbf{x} - (\tfrac{1}{\mu}\mathbf{X}^2\mathbf{s}(\mathbf{x}, \mu) - \mathbf{x}) = 2\mathbf{x} - \mathbf{X}\mathbf{z} \tag{24}
$$

Lemma C.2 (Path-following [34])**.** *If $\mu^+ = (1 - \tau)\mu, \tau \in (0, 1)$, then $\delta(\mathbf{x}, \mu^+) \leq \frac{1}{1-\tau}(\delta(\mathbf{x}, \mu) + \tau\sqrt{n})$.*

C.3 Inexact Analysis

Lemma C.3. *Suppose $\delta(\mathbf{x}, \mu) \leq \frac{1}{2}$, we have*

$$\mathbf{x}^+ = \mathbf{X}(2\mathbf{e} - \mathbf{z} + \boldsymbol{\psi}_\mathbf{x}) \quad and \quad \delta(\mathbf{x}^+, \mu) \leq \sqrt{2}\delta(\mathbf{x}, \mu)^2 + \sqrt{2n}\|\boldsymbol{\psi}_\mathbf{x}\|,$$

where $\boldsymbol{\psi}_\mathbf{x} := \mathbf{X}\mathbf{A}^\top(\mathbf{A}\mathbf{X}^2\mathbf{A}^\top)^{-1}\boldsymbol{\zeta}_\mathbf{x} + \mathbf{X}^{-1}\boldsymbol{\lambda}_\mathbf{x}$.

Proof. By definition of the inexact primal IPM step, we successively deduce that

$$
\begin{aligned}
\mathbf{x}^+ &= \mathbf{x} + \Delta\hat{\mathbf{x}} \\
&= \mathbf{x} + \Delta\mathbf{x} + \mathbf{X}^2\mathbf{A}^\top(\mathbf{A}\mathbf{X}^2\mathbf{A}^\top)^{-1}\boldsymbol{\zeta}_\mathbf{x} + \boldsymbol{\lambda}_\mathbf{x} \\
&= 2\mathbf{x} - \mathbf{X}\mathbf{z} + \mathbf{X}^2\mathbf{A}^\top(\mathbf{A}\mathbf{X}^2\mathbf{A}^\top)^{-1}\boldsymbol{\zeta}_\mathbf{x} + \boldsymbol{\lambda}_\mathbf{x} \\
&= \mathbf{X}(2\mathbf{e} - \mathbf{z} + \mathbf{X}\mathbf{A}^\top(\mathbf{A}\mathbf{X}^2\mathbf{A}^\top)^{-1}\boldsymbol{\zeta}_\mathbf{x} + \mathbf{X}^{-1}\boldsymbol{\lambda}_\mathbf{x}) \\
&= \mathbf{X}(2\mathbf{e} - \mathbf{z} + \boldsymbol{\psi}_\mathbf{x}),
\end{aligned}
\tag{25}
$$

where (25) applies (24). In view of the proximity measure, we have

$$\delta(\mathbf{x}^+, \mu) = \|\tfrac{1}{\mu}\mathbf{X}^+\mathbf{s}(\mathbf{x}^+, \mu) - \mathbf{e}\| \leq \|\tfrac{1}{\mu}\mathbf{X}^+\mathbf{s}(\mathbf{x}, \mu) - \mathbf{e}\| = \|\mathbf{X}^+\mathbf{X}^{-1}\mathbf{z} - \mathbf{e}\|.$$

Plugging in $\mathbf{x}^+ = \mathbf{X}(2\mathbf{e} - \mathbf{z} + \boldsymbol{\psi}_\mathbf{x})$,

$$\mathbf{X}^+\mathbf{X}^{-1}\mathbf{z} = \mathbf{X}^{-1}\mathbf{X}(2\mathbf{I} - \mathbf{Z} + \boldsymbol{\Psi}_\mathbf{x})\mathbf{z} = 2\mathbf{z} - \mathbf{Z}\mathbf{z} + \boldsymbol{\Psi}_\mathbf{x}\mathbf{z},$$

where $\boldsymbol{\Psi}_\mathbf{x} := \mathrm{diag}(\boldsymbol{\psi}_\mathbf{x})$. Hence, we deduce that

$$
\begin{aligned}
\delta(\mathbf{x}^+, \mu)^2 &= \|\mathbf{X}^+\mathbf{X}^{-1}\mathbf{z} - \mathbf{e}\|^2 \\
&= \|2\mathbf{z} - \mathbf{Z}\mathbf{z} - \mathbf{e} + \boldsymbol{\Psi}_\mathbf{x}\mathbf{z}\|^2 \\
&= \|2\mathbf{z} - \mathbf{Z}\mathbf{z} - \mathbf{e}\|^2 + \langle 2\boldsymbol{\Psi}_\mathbf{x}\mathbf{z}, 2\mathbf{z} - \mathbf{Z}\mathbf{z} - \mathbf{e}\rangle + \|\boldsymbol{\Psi}_\mathbf{x}\mathbf{z}\|^2 \\
&\leq 2\delta(\mathbf{x}, \mu)^4 + 2\|\boldsymbol{\Psi}_\mathbf{x}\mathbf{z}\|^2,
\end{aligned}
\tag{26}
$$

where (26) uses Cauchy's inequality. Taking square root on both sides and using $\sqrt{x + y} \leq \sqrt{x} + \sqrt{y}$,

$$\delta(\mathbf{x}^+, \mu) \leq \sqrt{2}\delta(\mathbf{x}^+, \mu)^2 + \sqrt{2}\|\boldsymbol{\Psi}_\mathbf{x}\mathbf{z}\| \leq \sqrt{2}\delta(\mathbf{x}^+, \mu)^2 + (\sqrt{2n} + 1)\|\boldsymbol{\psi}_\mathbf{x}\|,$$

where we use the relation $\|\mathbf{z} - \mathbf{e}\| \leq \frac{1}{2}$ and $\|\mathbf{z}\| \leq \|\mathbf{e}\| + \frac{1}{2} = \sqrt{n} + \frac{1}{2}$. This completes the proof. $\square$

Lemma C.4. *Suppose $\delta(\mathbf{x}, \mu) \leq \frac{1}{2}$, then for all $\mu \leq \frac{\sqrt{\lambda^\star}\tau}{16\gamma} \cdot \left(\frac{\|\mathbf{A}_\mathcal{N}\|}{\sqrt{\lambda_{\min}(\mathbf{A}_\mathcal{B}^\top\mathbf{A}_\mathcal{B})}} + 1\right)^{-1}$*

$$\|\boldsymbol{\psi}_\mathbf{x}\| \leq \left(\frac{1}{\alpha\mu\sigma_{\min}(\mathbf{A})} + \frac{2}{\sqrt{\lambda^\star}}\right)\frac{1}{\sigma_{\min}(\mathbf{A})}\|\boldsymbol{\zeta}_\mathbf{x}\|.$$

Proof. Recall that $\boldsymbol{\psi}_\mathbf{x} = \mathbf{X}\mathbf{A}^\top(\mathbf{A}\mathbf{X}^2\mathbf{A}^\top)^{-1}\boldsymbol{\zeta}_\mathbf{x} + \mathbf{X}^{-1}\boldsymbol{\lambda}_\mathbf{x}$ and we first consider $\boldsymbol{\lambda}_\mathbf{x}$. First note that

$$\|\boldsymbol{\lambda}_\mathbf{x}\| = \sqrt{\langle \boldsymbol{\zeta}_\mathbf{x}, (\mathbf{A}\mathbf{A}^\top)^{-1}\boldsymbol{\zeta}_\mathbf{x}\rangle} \leq \frac{1}{\sqrt{\lambda_{\min}(\mathbf{A}\mathbf{A}^\top)}}\|\boldsymbol{\zeta}_\mathbf{x}\| = \frac{1}{\sigma_{\min}(\mathbf{A})}\|\boldsymbol{\zeta}_\mathbf{x}\|$$

Then using $\delta(\mathbf{x}, \mu) \leq \frac{1}{2}$ and **Lemma B.3**, $\|\mathbf{x}\|_{-\infty} \geq \alpha\mu$ and $\|\mathbf{X}^{-1}\boldsymbol{\lambda}_\mathbf{x}\| \leq \frac{1}{\alpha\mu\sigma_{\min}(\mathbf{A})}\|\boldsymbol{\zeta}_\mathbf{x}\|$. Next we consider $\mathbf{X}\mathbf{A}^\top(\mathbf{A}\mathbf{X}^2\mathbf{A}^\top)^{-1}\boldsymbol{\zeta}_\mathbf{x}$. Similarly, we have

$$\|\mathbf{X}\mathbf{A}^\top(\mathbf{A}\mathbf{X}^2\mathbf{A}^\top)^{-1}\boldsymbol{\zeta}_\mathbf{x}\| = \sqrt{\langle \boldsymbol{\zeta}_\mathbf{x}, (\mathbf{A}\mathbf{X}^2\mathbf{A}^\top)^{-1}\boldsymbol{\zeta}_\mathbf{x}\rangle} \leq \frac{1}{\sqrt{\lambda_{\min}(\mathbf{A}\mathbf{X}^2\mathbf{A}^\top)}}\|\boldsymbol{\zeta}_\mathbf{x}\| \leq \frac{2}{\sqrt{\lambda^\star}}\|\boldsymbol{\zeta}_\mathbf{x}\|,$$

where we again invoked **Lemma 3.2** with $\beta = \frac{1}{2}$ to obtain $\lambda_{\min}(\mathbf{A}\mathbf{X}^2\mathbf{A}^\top) \succeq \frac{\lambda^\star}{4}\mathbf{I}$. Putting things together,

$$\|\boldsymbol{\psi}_{\mathbf{x}}\| = \|\mathbf{X}\mathbf{A}^\top(\mathbf{A}\mathbf{X}^2\mathbf{A}^\top)^{-1}\boldsymbol{\zeta}_{\mathbf{x}} + \mathbf{X}^{-1}\boldsymbol{\lambda}_{\mathbf{x}}\| \leq \left(\frac{1}{\alpha\mu\sigma_{\min}(\mathbf{A})} + \frac{2}{\sqrt{\lambda^\star}}\right)\|\boldsymbol{\zeta}_{\mathbf{x}}\|$$

and this completes the proof. $\qquad\square$

C.4   Proof of Theorem 3.3

Theorem C.1 (Theorem 3.3). *Under* **A1** *and* **A2** *and assume* $n > 100$, *suppose primal IPM starts from*

$$\mu_1 < \frac{\sqrt{\lambda^\star}}{40\sqrt{n}\gamma}\left(\frac{\|\mathbf{A}_{\mathcal{N}}\|}{\sqrt{\lambda_{\min}(\mathbf{A}_{\mathcal{B}}^\top\mathbf{A}_{\mathcal{B}})}} + 1\right)^{-1}$$

such that $\delta(\mathbf{x}_1, \mu_1) \leq \frac{1}{2}$ *and suppose* $\tau = \frac{1}{10\sqrt{n}}$. *Then* $\delta(\mathbf{x}_k, \mu_k) \leq \frac{1}{2}$ *for all* k *and the arithmetic complexity of finding an* ε-*optimal primal solution is* $\mathcal{O}(n^{2.5}\log^2(\frac{1}{\varepsilon}))$.

Proof. Putting all the results together,

$$\mathbf{x}^+ = \mathbf{X}(2\mathbf{e} - \mathbf{z} + \boldsymbol{\psi}_{\mathbf{x}}) \tag{27}$$

$$\|\boldsymbol{\psi}_{\mathbf{x}}\| \leq \left(\frac{1}{\alpha\mu\sigma_{\min}(\mathbf{A})} + \frac{2}{\sqrt{\lambda^\star}}\right)\|\boldsymbol{\zeta}_{\mathbf{x}}\| \tag{28}$$

$$\delta(\mathbf{x}^+, \mu) \leq \sqrt{2}\delta(\mathbf{x}, \mu)^2 + (\sqrt{2n} + 1)\|\boldsymbol{\psi}_{\mathbf{x}}\| \tag{29}$$

$$\delta(\mathbf{x}, \mu^+) \leq \frac{1}{1-\tau}(\delta(\mathbf{x}, \mu) + \tau\sqrt{n}) \tag{30}$$

Assume we let

$$\|\boldsymbol{\zeta}_{\mathbf{x}}\| \leq \frac{\alpha\mu\sigma_{\min}(\mathbf{A})\sqrt{\lambda^\star}}{40\sqrt{3n}(\sqrt{\lambda^\star} + 2\alpha\mu\sigma_{\min}(\mathbf{A}))}$$

and by (28) $\|\boldsymbol{\psi}_{\mathbf{x}}\| \leq \frac{1}{40\sqrt{3n}}$. With **Theorem 3.2** it takes $\mathcal{O}(\log\frac{1}{\mu})$ iterations for an iterative solver to converge. Since $\mathbf{x}, \mu$ satisfies $\delta(\mathbf{x}, \mu) \leq \frac{1}{2}$. We have $\|\mathbf{z} - \mathbf{e}\|_\infty \leq \|\mathbf{z} - \mathbf{e}\| \leq \frac{1}{2}$ and from (27) we deduce that

$$2\mathbf{e} - \mathbf{z} + \boldsymbol{\psi}_{\mathbf{x}} \geq \left(\frac{1}{2} - \|\boldsymbol{\psi}_{\mathbf{x}}\|\right)\mathbf{e} \geq 0.1\mathbf{e}.$$

Therefore $\mathbf{x}^+ \in \mathcal{F}_p^0$. Next fix $\tau = 1/(10\sqrt{n})$ and for $n \geq 100$

$$\begin{aligned}
\delta(\mathbf{x}^+, \mu^+) &\leq \frac{1}{1-\tau}(\delta(\mathbf{x}^+, \mu) + \tau\sqrt{n}) \\
&\leq \frac{1}{1-\tau}(\sqrt{2}\delta(\mathbf{x}, \mu)^2 + (\sqrt{2n} + 1)\|\boldsymbol{\psi}_{\mathbf{x}}\|) + \frac{\tau\sqrt{n}}{1-\tau} \\
&= \frac{\sqrt{2}}{1-\tau}\delta(\mathbf{x}, \mu)^2 + \frac{\sqrt{2n}+1}{1-\tau}\|\boldsymbol{\psi}_{\mathbf{x}}\| + \frac{\tau\sqrt{n}}{1-\tau} \\
&\leq \frac{3}{2}\delta(\mathbf{x}, \mu)^2 + \sqrt{3n}\|\boldsymbol{\psi}_{\mathbf{x}}\| + \frac{1}{10}
\end{aligned}$$

Now that $\delta(\mathbf{x}, \mu) \leq \frac{1}{2}$, we have

$$\delta(\mathbf{x}^+, \mu^+) \leq \frac{3}{8} + \sqrt{3n}\|\boldsymbol{\psi}_{\mathbf{x}}\| + \frac{1}{10} \leq \frac{19}{40} + \sqrt{3n}\|\boldsymbol{\psi}_{\mathbf{x}}\| \leq \frac{19}{40} + \frac{1}{40} = \frac{1}{2}.$$

According to **Theorem 3.1**, we have $\|\mathbf{x}_k - \mathbf{x}^\star\| = \mathcal{O}(\mu_k) = \mathcal{O}((1-\tau)^k)$, and the total outer iterations are

$$\mathcal{O}(\log(1-\tau) \cdot \log\varepsilon) = \mathcal{O}(\sqrt{n}\log(\tfrac{1}{\varepsilon})).$$

Besides, in each iteration, the preconditioned iterative algorithm needs to take $\log(\frac{n}{\varepsilon})$ iterations to find an inexact solution with accuracy ε, and the cost of each iteration is the matrix-vector product and is thus n^2. This completes the proof. $\qquad\square$