Primal-Dual Interior Point Methods

In this lecture series, we will analyze the geometry of linear programs (LP) from the perspective of interior point methods (IPM). This will eventually lead to a new combinatorial measure of the complexity for linear programs, called *straight-line complexity*, which has been used to provide the first *strongly-polynomial* algorithm for linear programs having at most two variables per inequality [ADL+22, DKN+24], as well as the provide examples of linear programs for which IPMs require a number of iterations that is exponential in the dimension [ABGJ18, AGV22].

In the present lecture, we will present a primal-dual interior point method (IPM) for solving linear programs, which together with the simplex method, are the principal methods for solving linear programs in practice. The primal-dual linear programs we will consider throughout is:

$$\begin{aligned} \min\ & \langle c, x\rangle \\ \mathbf{A}x &= b \\ x &\geq 0_n\,, \end{aligned} \tag{Primal LP}$$

$$\begin{aligned} \max\ & \langle b, y\rangle \\ \mathbf{A}^\top y + s &= c \\ s &\geq 0_n. \end{aligned} \tag{Dual LP}$$

Notation. The instance data is the constraint matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$, and $\mathrm{rk}(\mathbf{A}) = m \leq n$. We use the notation $\langle x, y\rangle := x^\top y = \sum_{i=1}^n x_i y_i$ to denote the standard inner product on $\mathbb{R}^n$, $\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2} = \sqrt{\langle x, x\rangle}$ for the ℓ_2 norm, $\|x\|_\infty := \max_{i\in[n]} |x_i|$ for the ℓ_∞ norm. The primal and dual feasible regions are denoted $\mathcal{P} := \{x \in \mathbb{R}^n \mid \mathbf{A}x = b, x \geq 0_n\}$ and $\mathcal{D} := \{(s, y) \in \mathbb{R}^{n+m} \mid \mathbf{A}^\top y + s = c, s \geq 0_n\}$, with the strictly feasible solutions denoted by $\mathcal{P}_{++} := \{x \in \mathcal{P} \mid x > 0_n\}$ and $\mathcal{D}_{++} := \{(s, y) \in \mathcal{D} \mid s > 0_n\}$.

A Very Brief History of LP Algorithms. The algorithmic theory of LP solving began with Dantzig's development of the simplex algorithm in 1947 [Dan90]. The first polynomial time algorithm (which solves the LPs exactly), is due to Khachiyan [Kha79], who relied on the ellipsoid method of Yudin and Nemirovski [Yud76] (see also Shor [Sho77]), to achieve a running time that scales as $\mathrm{poly}(n, \log(\langle A, b, c\rangle))$, where $\langle A, b, c\rangle$ denotes the bit-complexity. While polynomial, the ellipsoid method was known for having very poor practical performance. The first polynomial time interior point method for LP was designed by Karmarkar [Kar84], which had the benefit of being effective in practice. Renegar [Ren88] gave an improved IPM which followed a *central path*, and relied on the use the classical Newton's method. The first *primal-dual* path following algorithm was first developed by Kojima, Mizuno and Yoshise [KMY89], variants of which are implemented in all commercial solvers (see [YTM94, Gon96, Wri97]).

A major open question is whether LPs can be solved in *strongly polynomial* time. Stated slightly simplistically, the question is whether there is an algorithm which solves (Primal LP) and (Dual LP) using only $\mathrm{poly}(n, m)$ basic arithmetic operations on real numbers (i.e., $+, \times, /, \geq$) [1]. All known polynomial time algorithms require a number of operations that depend on $\langle A, b, c\rangle$,

[1] There are important considerations relating to the size of the numbers used during the arithmetic computations which we ignore.

and hence are not strongly polynomial. This question was explicitly asked by Megiddo [Meg83] and highlighted by fields medalist Steven Smale [Sma98] as one of the mathematical challenges for the 21st century.

LP Duality. The fundamental result which underlies the polynomial solvability of linear programming is the linear programming duality:

THEOREM 1 (STRONG DUALITY) *If* (Primal LP) *and* (Dual LP) *are both feasible, they have the same value and both admit optimal solutions. Furthermore, for $x \in \mathcal{P}$ and $(s, y) \in \mathcal{D}$, we have that:*

$$\langle c, x\rangle - \langle y, b\rangle = \langle c, x\rangle - \langle y, \mathbf{A}x\rangle = \left\langle c - \mathbf{A}^\top y, x\right\rangle = \langle s, x\rangle \geq 0. \qquad \text{(Gap Formula)}$$

We now recall the complementary slackness condition for optimality. Assuming both programs are feasible, we may let x^*, s^*, y^* be optimal primal-dual solutions. Since the primal and dual have the same value, we see that

$$0 = \langle c, x^*\rangle - \langle y^*, b\rangle = \langle x^*, s^*\rangle .$$

Since x^*, s^* are non-negative vectors, we must have x^* and s^* have *disjoint supports*. That is, $x_i^* s_i^* = 0$, for all $i \in [n]$. As we show next, interior point methods will solve the LP by relaxing $x_i s_i = 0$ constraint to $x_i s_i = \mu > 0$, for all $i \in [n]$, and slowly driving $\mu \to 0$.

1 Primal-Dual Central Path

We will solve the primal-dual system above, by following a primal-dual *central path* that converges towards primal and dual optimal solutions. The method will require the primal and dual be *strictly feasible*, and that we have knowledge of a good starting point. A way to initialize the method without this requirement will be discussed in Section 4.

We define the primal-dual central path

$$\text{CP} := \{z(\mu) := (x(\mu), s(\mu), y(\mu)) \mid \mu > 0\},$$

as the optimal solutions to the following sequence of parametric optimization problems:

$$x(\mu) := \operatorname{argmin}\{\langle c, x\rangle - \mu \sum_{i=1}^n \ln(x_i) \mid \mathbf{A}x = b, x > 0_n\}, \qquad \text{(Primal Path Program)}$$

$$(s(\mu), y(\mu)) := \operatorname{argmax}\{\langle y, b\rangle + \mu \sum_{i=1}^n \ln(s_i) \mid \mathbf{A}^\top y + s = c, s > 0_n\}. \qquad \text{(Dual Path Program)}$$

The logarithmic term in the objectives is called the *logarithmic barrier*, which blows up on the boundary of the feasible region. The barrier encourages the path to stay as far away from the boundary as possible. The optimal solutions $x(\mu)$ and $(s(\mu), y(\mu))$ are in fact unique, which follows from the strict convexity of the objective for the primal and strict concavity in the dual.

While the primal and dual central path programs are seemingly independent, their optimal solutions are intimately linked via the *central path equations*, which we explain below. For this purpose, we will need the Hadamard product of two vectors $x, y \in \mathbb{R}^n$:

$$x \circ y := (x_1 y_1, x_2 y_2, \ldots, x_n y_n). \qquad (1)$$

For simplicity of notation, we simply write xy for $x \circ y$. We generalize this notation to expressions of the form $x^\alpha/y^\beta = (x_1^\alpha/y_1^\beta, \dots, x_n^\alpha/y_n^\beta)$ for $\alpha, \beta \in \mathbb{R}$. We also use the notation X^α, where $X := \operatorname{diag}(x_1, \dots, x_n)$ is defined to be diagonal matrix having x on its diagonal.

The relation between the paths follows from the fact that up to an additive constant, the respective programs are Lagrangian duals of each other. Starting with the primal program, we may Lagrangify out the equality constraints using multipliers $y \in \mathbb{R}^m$ as follows:

$$\begin{aligned} \inf_{\mathbf{A}x=b, x>0_n} \langle c, x\rangle - \mu \sum_{i=1}^n \ln(x_i) &\geq \inf_{x>0_n} \langle c, x\rangle + \langle y, b - \mathbf{A}x\rangle - \mu \sum_{i=1}^n \ln(x_i) \\ &= \inf_{x>0_n} \langle y, b\rangle + \left\langle c - \mathbf{A}^\top y, x \right\rangle - \mu \sum_{i=1}^n \ln(x_i). \end{aligned} \tag{2}$$

Letting $s := c - \mathbf{A}^\top y$, it is easy to verify that the relaxed program has value $> -\infty$ (a non-trivial lower bound) if and only if $s > 0_n$, i.e., if and only if $(s, y) \in \mathcal{D}_{++}$. Assuming $s > 0_n$, the unique choice of x which minimizes $\langle s, x\rangle - \mu \sum_{i=1}^n \ln(x_i)$ is precisely $x = \mu s^{-1} > 0_n$, as the objective function is convex and this choice sets the gradient to 0_n (recall that $(\ln t)' = 1/t$). For $(s, y) \in \mathcal{D}_{++}$, the value of the relaxed program becomes

$$\begin{aligned} \inf_{x>0_n} \langle y, b\rangle + \langle s, x\rangle - \mu \sum_{i=1}^n \ln(x_i) &= \langle y, b\rangle + \left\langle s, \mu s^{-1} \right\rangle - \mu \sum_{i=1}^n \ln(\mu/s_i) \\ &= \langle y, b\rangle + \mu \sum_{i=1}^n \ln(s_i) + \mu(1 - \ln \mu)n. \end{aligned} \tag{3}$$

The problem of maximizing the value of this lower bound over $(s, y) \in \mathcal{D}_{++}$ is, up to the additive constant $\mu(1 - \ln \mu)n$, precisely the problem (Dual Path Program) (one may also verify that dualizing (Dual Path Program) yields an equivalent program to (Primal Path Program)).

As with strong LP duality, assuming that $\mathcal{P}_{++}$ and $\mathcal{D}_{++}$ are non-empty, strong Lagrangian duality[2] implies that both programs have optimal solutions and that the value of (Primal Path Program) equals the value of (Dual Path Program) plus $\mu(1 - \ln(\mu))n$. Letting $x(\mu)$ and $(s(\mu), y(\mu))$ denote the respective optimal solutions, we must have that all the inequalities derived in (2) and (3) must be tight when starting from $y(\mu)$. One can verify that the inequalities can only be tight when $x(\mu) = \mu s(\mu)^{-1} \Leftrightarrow x(\mu)s(\mu) = \mu 1_n$. This yields the following characterization of the central path.

Lemma 2 *Assuming the primal and dual are strictly feasible, the central path is well-defined for any $\mu > 0$. Furthermore, the tuple $(x(\mu), s(\mu), y(\mu))$ is uniquely characterized by the following equations:*

$$\begin{aligned} \mathbf{A}x(\mu) &= b, x(\mu) > 0_n && \text{(Strict Primal Feasibility)} \\ \mathbf{A}^\top y(\mu) + s(\mu) &= c, s(\mu) > 0_n && \text{(Strict Dual Feasibility)} \\ x(\mu)s(\mu) &= \mu 1_n, && \text{(Centrality Equation)} \end{aligned}$$

where 1_n denote the vector of all ones.

From the above equations, the optimality gap between the primal solution $x(\mu)$ and the dual solution $(s(\mu), y(\mu))$ is precisely

$$\langle c, x(\mu)\rangle - \langle y(\mu), b\rangle = \langle x(\mu), s(\mu)\rangle = n\mu.$$

[2] We will not require in-depth knowledge of Lagrangian duality. We only use it here to help explain the central path equations.

Thus, as $\mu \to 0$, the optimality gap also shrinks to 0. We now define the limit optimal solution $z^* := z(0) := \lim_{\mu \to 0} z(\mu) \in \overline{\mathrm{CP}}$, which lives in the closure of the central path. The limit optimal solution $z^* := (x^*, s^*, y^*)$ will correspond the "most interior" among optimal solutions (note that there may be an entire face of optimal solutions). In particular, z^* will be *strictly complementary*: for each $i \in [n]$, we will have either $x_i^* > 0$ or $s_i^* > 0$. This strict complementarity induces a (unique) *optimality partition* $B^* \cup N^* = [n]$ defined by $B^* := \mathrm{support}(x^*)$ and $N^* := \mathrm{support}(s^*)$.

2 Predictor-Corrector Algorithm for Primal-Dual Path Following

We now explain a method that produces feasible iterates $\{(x^k, s^k, y^k)\}_{k \geq 0}$ that *closely track* the central path, and whose gap $\langle x^k, s^k \rangle$ decrease *geometrically*. Specifically, we will show that every $O(\sqrt{n})$ iterations the gap decreases by a factor 2 (in fact, this is a rather pessimistic estimate).

To measure our distance to the central path, we will utilize a multiplicative notion of distance from centrality.

DEFINITION 3 (NORMALIZED GAP, CENTRALITY DISTANCE AND ℓ_2 NEIGHBORHOOD) *For $z := (x, s, y) \in \mathcal{P} \times \mathcal{D}$ primal-dual feasible, the* **normalized gap** *is*

$$\mu(z) := \langle x, s \rangle / n.$$

When $\mu(z) > 0$, we define the **distance to centrality** *of z by*

$$\mathrm{dist_c}(z) := \left\| \frac{xs}{\mu(z)} - 1_n \right\|_2.$$

If $z = (x, s, y)$ is on the primal-dual central path then $\mu := x_1 s_1 = \cdots = x_n s_n$, and hence $\mu = \sum_{i=1}^n x_i s_i / n := \mu(z)$ and $\mathrm{dist_c}(z) = 0$.
The ℓ_2 **neighborhood** *$N_2(\beta)$ of the central path of width $\beta \in (0,1)$ is then defined by*

$$N_2(\beta) := \{z \in \mathcal{P}_{++} \times \mathcal{D}_{++} : \mathrm{dist_c}(z) \leq \beta\}.$$

We further define $\overline{N}_2(\beta)$ to be the closure of the ℓ_2 neighborhood. The closure will contain the limit optimal solution $z^* := z(0)$ at the end of the path as defined in the last section.

The width parameter β is a tunable parameter that controls how close we wish to stay to the path. The method we present will slow down predictably as β gets smaller. The constant β can in fact be fixed to $1/2$ in the present analysis, however a variant of the algorithm we will analyze later will need β to be a bit smaller for provable correctness (e.g., $\beta = 1/100$).

The algorithm we present below, due to Mizuno, Todd and Ye [MTY93], alternates between two types of steps. The first type are *corrector steps*, which bring us closer to the central path by decreasing our centrality distance. The second type are *predictor steps*, which try to move us down the path, by decreasing the duality gap as fast as possible while keeping the centrality distance below the β-threshold.

Corrector Step. Given $z = (x, s, y) \in N_2(\beta)$, we would like to compute an update $\Delta z^c := (\Delta x^c, \Delta s^c, \Delta y^c)$ such that $(x + \Delta x^c)(s + \Delta s^c)$ moves closer to $\mu(z) 1_n$. Specifically, we will aim for

$z + \Delta z \in N_2(\beta')$ for $\beta' \leq \beta/2$. We will compute the update by using a simple linear approximation of the corresponding quadratic equation:

$$\mu(z)1_n = (x + \Delta x)(s + \Delta s) = xs + s\Delta x + x\Delta s + \Delta x \Delta s \approx xs + s\Delta x + x\Delta s,$$

where we optimistically assume that the quadratic term $\Delta x \Delta s$ is "small". Heuristically, if we are β-close to the path, we will hope that the corresponding quadratic term will have size $\beta^2 \ll \beta$, assuming $\beta \ll 1$.

Using the linear approximation, the corrector direction at $z = (x, s, y)$ is defined by the following linear system of equations:

$$\begin{aligned} s\Delta x^c + x\Delta s^c &= \mu(z)1_n - xs, \\ \mathbf{A}\Delta x^c &= 0_m, \\ \mathbf{A}^\top \Delta y^c + \Delta s^c &= 0_n. \end{aligned} \qquad \text{(Corrector Direction)}$$

We then take a corrector step by updating z to $z + \Delta z^c$.

Predictor Step. Given $z = (x, s, y) \in N_2(\beta')$, for $\beta' \leq \beta/2$, we would like to substantially decrease $\mu(z)$ while staying inside the $N_2(\beta)$ neighborhood. Our goal will be to compute a *predictor direction* $\Delta z^p := (\Delta x^p, \Delta s^p, \Delta y^p)$ and a largest possible step-size $\alpha \in [0, 1]$ such that $(x + \alpha\Delta x^p)(s + \alpha\Delta s^p) \approx (1 - \alpha)xs$ and $z + \alpha\Delta z^p \in \overline{N}_2(\beta)$. This will decrease $\mu(z)$ by a $(1 - \alpha)$ factor.

The predictor direction is computed using a linear approximation of the equation

$$0_n = (x + \Delta x^p)(s + \Delta s^p) = xs + s\Delta x^p + s\Delta s^p + \Delta x^p \Delta s^p \approx xs + s\Delta x^p + s\Delta s^p,$$

where we again hope that the contribution of the quadratic term $\Delta x^p \Delta s^p$ is small. The above equation is motivated by the fact that any optimal primal-dual solution (x^*, s^*, y^*) satisfies $x^* s^* = 0_n$. The predictor direction at $z = (x, s, y)$ is then defined as follows:

$$\begin{aligned} s\Delta x^p + x\Delta s^p &= -xs, \\ \mathbf{A}\Delta x^p &= 0_m, \\ \mathbf{A}^\top \Delta y^p + \Delta s^p &= 0_n. \end{aligned} \qquad \text{(Predictor Direction)}$$

We then take a predictor step by moving to $z + \alpha\Delta z^p$, where $\alpha \in [0, 1]$ is chosen as large as possible to ensure the resulting iterates lies in $\overline{N}_2(\beta)$ (the closure ensures we can reach the end of the path).

REMARK 1 The linear system defining the predictor and corrector direction differ only in the $\mu 1_n$ term in the first equation. Importantly, the linear systems will always admit a solution. In the literature, the predictor direction defined above is known as the *affine scaling* direction.

Predictor-Corrector Algorithm. The algorithm is given below. Importantly, the algorithm assumes as input an initial iterate $z^0 := (x^0, s^0, y^0)$ in β-neighborhood. Note the initial iterate determines the right-hand side $b = \mathbf{A}x^0$ and objective $c = \mathbf{A}^\top y^0 + s^0$.

Algorithm 1: PREDICTOR-CORRECTOR IPM

Input : Constraint matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\text{rank}(\mathbf{A}) = m$, and initial iterate $z^0 := (x^0, s^0, y^0) \in N_2(\beta)$, $\beta \in (0, 1/2]$, error parameter $\varepsilon \geq 0$.

Output: $z = (x, s, y) \in \overline{N}_2(\beta)$, $\langle x, s \rangle \leq \varepsilon$.

$z = (x, s, y) \leftarrow (x^0, s^0, y^0)$;

while $\langle x, s \rangle > \varepsilon$ **do**

  Compute Δz^c at z according to (Corrector Direction);

  $z \leftarrow z + \Delta z^c$;

  Compute Δz^p at z according to (Predictor Direction);

  `// See Lemma 5 part (3) for the precise choice of` α^p

  Choose $\alpha^p \in [0, 1]$ as large as possible so that $z + \alpha^p \Delta z^p \in \overline{N}_2(\beta)$;

  $z \leftarrow z + \alpha^p \Delta z^p$;

return z;

2.1 Projection Characterization of Step Equations

In this subsection, show how to interpret the step equations in the form of orthogonal projections:

$$s \Delta x + x \Delta s = t \qquad \text{(Target Eq)}$$

$$\mathbf{A} \Delta x = 0_m \qquad \text{(Kernel Eq)}$$

$$\mathbf{A}^\top \Delta y + \Delta s = 0_n \qquad \text{(Image Eq)}$$

for $x, s > 0_n$ and $t \in \mathbb{R}^n$. The equations (Kernel Eq) and (Image Eq) are *subspace equations* that can be concisely expressed as $\Delta x \in \ker(\mathbf{A}) = \{w \in \mathbb{R}^n \mid \mathbf{A}w = 0_m\}$, the *kernel* of $\mathbf{A}$, and $\Delta s \in \text{im}(\mathbf{A}^\top) := \{\mathbf{A}^\top z \mid z \in \mathbb{R}^m\}$, the *image* of $\mathbf{A}^\top$. Under the assumption that $\Delta s \in \text{im}(\mathbf{A}^\top)$, note that Δy is uniquely determined by the equation (Image Eq) since the rows of $\mathbf{A}$ are linearly independent.

For a linear subspace $W \subseteq \mathbb{R}^n$, the *orthogonal complement* of W is defined by $W^\perp := \{z \in \mathbb{R}^n : \langle z, w \rangle = 0, \forall w \in W\}$. With this definition, we have the $\ker(\mathbf{A}) := \text{im}(\mathbf{A}^\top)^\perp$ since

$$\mathbf{A}\mathbf{x} = 0_m \Leftrightarrow \langle y, \mathbf{A}\mathbf{x} \rangle = 0, \forall y \in \mathbb{R}^m \Leftrightarrow \left\langle \mathbf{A}^\top y, \mathbf{x} \right\rangle = 0, \forall y \in \mathbb{R}^m \Leftrightarrow \langle w, \mathbf{x} \rangle = 0, \forall w \in \text{im}(\mathbf{A}^\top). \qquad (4)$$

Let us multiply (Target Eq) by the diagonal matrix $(XS)^{-1/2}$, recalling that $X := \text{diag}(x)$ and $S := \text{diag}(s)$, which yields

$$\sqrt{\frac{s}{x}} \Delta x + \sqrt{\frac{x}{s}} \Delta s = \frac{t}{\sqrt{xs}}. \qquad (5)$$

Letting $D = X^{1/2} S^{-1/2}$, $\sqrt{\frac{x}{s}} \Delta s = D \Delta s \in D \, \text{im}(\mathbf{A}^\top) = \text{im}((\mathbf{A}D)^\top)$. Similarly, $\sqrt{\frac{s}{x}} \Delta x = D^{-1} \Delta x \in D^{-1} \ker(\mathbf{A}) = \ker(\mathbf{A}D)$.

Letting $W := \text{im}(\mathbf{A}D)^\top$, by (4) we know that $W^\perp = \ker(\mathbf{A}D)$. We may thus interpret (5) as an orthogonal decomposition of $\frac{t}{\sqrt{xs}}$ into its component $\sqrt{\frac{x}{s}} \Delta s$ on W and its component $\sqrt{\frac{s}{x}} \Delta x$ on $W^\perp$. We conclude that the system of equations indexed by (Target Eq) always admits a solution. More precisely, letting $\Pi_W, \Pi_{W^\perp} \in \mathbb{R}^{n \times n}$ denote the orthogonal projections onto $W, W^\perp$ respectively we get that:

$$\left(\sqrt{\frac{s}{x}} \Delta x, \sqrt{\frac{x}{s}} \Delta s \right) = \left(\Pi_W \left(\frac{t}{\sqrt{xs}} \right), \Pi_{W^\perp} \left(\frac{t}{\sqrt{xs}} \right) \right),$$

which follows since $z = \Pi_W(z) + \Pi_{W^\perp}(z)$[3] for any $z \in \mathbb{R}^n$.

A useful property of orthogonal decomposition is the Pythagorean identity:

$$\|z\|_2^2 = \|\Pi_W z + \Pi_{W^\perp} z\|_2^2 = \langle \Pi_W z, \Pi_W z\rangle + 2\underbrace{\langle \Pi_W z, \Pi_{W^\perp} z\rangle}_{=0} + \langle \Pi_{W^\perp} z, \Pi_{W^\perp} z\rangle \tag{6}$$

$$= \|\Pi_W z\|^2 + \|\Pi_{W^\perp} z\|^2. \tag{7}$$

Specializing to (5), we derive the corresponding identity which we will use repeatedly to analyze the IPM:

$$\left\|\sqrt{\frac{s}{x}}\Delta x\right\|_2^2 + \left\|\sqrt{\frac{x}{s}}\Delta s\right\|_2^2 = \left\|\frac{t}{\sqrt{xs}}\right\|_2^2. \tag{8}$$

3 Analysis of Predictor-Corrector Algorithm

We analyze the properties of corrector and predictor steps separately. In the two lemmas below, recall the definition of normalize gap $\mu(z) := \langle x, s\rangle / n$ for an iterate $z := (x, s, y) \in \mathcal{P} \times \mathcal{D}$.

LEMMA 4 (CORRECTOR ANALYSIS) *Let $z \in N_2(\beta)$ for $\beta \in (0, 1/2]$. The corrector direction Δz^c at z satisfies:*

1. $\mu(z + \Delta z^c) = \mu(z)$.

2. $z + \Delta z^c \in N_2(\frac{\beta^2}{2(1-\beta)}) \subseteq N_2(\beta/2)$.

REMARK 2 For $\beta \in (0, 1/2]$, $\frac{\beta^2}{2(1-\beta)} \leq \beta^2 \leq \beta/2$.

LEMMA 5 (PREDICTOR ANALYSIS) *Let $z \in N_2(\beta/2)$, $\beta \in (0, 1/2]$, $\Delta z^p = (\Delta x^p, \Delta s^p, \Delta y^p)$ be the predictor direction at z and $q_z := \left\|\frac{\Delta x^p \Delta s^p}{\mu(z)}\right\|_2$ be the normalized quadratic error. Then the following holds:*

1. $\mu(z + \alpha\Delta z^p) = (1 - \alpha)\mu(z)$, *for* $\alpha \in \mathbb{R}$.

2. $q_z \leq n/2$.

3. *Define* $h(\alpha) = \frac{1-\alpha}{\alpha^2}$ *for* $\alpha \in (0, 1]$. *Then,* $\alpha^p := h^{-1}(\frac{2q_z}{\beta}) \geq \frac{1}{2}\sqrt{\frac{\beta}{n}}$ *satisfies* $z + \alpha\Delta z^p \in N(\beta)$, $\forall \alpha \in [0, \alpha^p)$. *In particular,* $z + \alpha^p \Delta z^p \in \overline{N}(\beta)$.

Combining the above two lemmas, we directly obtain the following estimate on the worst-case convergence rate on PREDICTOR-CORRECTOR IPM.

THEOREM 6 *After $T \geq 0$ iterations of the while loop, PREDICTOR-CORRECTOR IPM produces an iterate $z = (x, s, y) \in \overline{N}_2(\beta)$, $\beta \in (0, 1/2]$, satisfying $\langle x, s\rangle \leq \langle x^0, s^0\rangle e^{-\frac{1}{2}\sqrt{\frac{\beta}{n}}T}$. Furthermore, if $\langle x^0, s^0\rangle \geq \varepsilon > 0$, the algorithm terminates after at most $2\sqrt{\frac{n}{\beta}}\ln(\langle x^0, s^0\rangle / \varepsilon) + 1$ iterations.*

[3] One may in fact define Π_W as the unique linear map satisfying $\Pi_W(z) \in W$ and $z - \Pi_W(z) \in W^\perp$, $\forall z \in \mathbb{R}^n$. By symmetry, note then $\mathbf{I}_n - \Pi_W = \Pi_{W^\perp}$, where $\mathbf{I}_n$ is the $n \times n$ identity.

Proof: We prove the first claim by induction on $T \geq 0$. Note that the claim is trivially true for $T = 0$ (i.e., before we enter the while loop), since by assumption $z^0 \in N_2(\beta) \subseteq \overline{N}_2(\beta)$. Assuming that the induction hypothesis holds after iteration T, we prove it for iteration $T+1$. If we terminate before iteration $T+1$, there is nothing to prove and the statement holds. So assume we enter the while loop at iteration $T+1$. Then at the beginning of iteration $T+1$, by our induction hypothesis, we have that

$$z = (x, s, y) \in \overline{N}_2(\beta), \qquad \text{and} \qquad n\mu(z) = \langle x, s \rangle \leq \left\langle x^0, s^0 \right\rangle e^{-\frac{1}{2}\sqrt{\frac{\beta}{n}}T}.$$

Given that we entered the while loop, we must have that $n\mu(z) = \langle x, s \rangle > \varepsilon \geq 0$. In this case, $z \in \overline{N}_2(\beta)$ implies that $z \in N_2(\beta)$, since $\overline{N}_2(\beta) \setminus N_2(\beta)$ only contains points with $\mu(z) = 0$.

Starting from z, we now take one corrector and one predictor step. For the corrector step, we update z to z_c where $z_c = z + \Delta z^c$. By Lemma 4, we have that $\mu(z) = \mu(z_c)$ and that $z_c \in N_2(\beta/2)$ since $\beta \in (0, 1/2]$. For the predictor step, we update z_c to $z_p = z_c + \alpha^p \Delta z_c^p$ where α^p is chosen according to Lemma 5 part (3), which satisfies $z_p \in \overline{N}(\beta)$ and

$$\begin{aligned} \mu(z_p) &\overset{\text{Lem 5 p (1)}}{=} (1-\alpha^p)\mu(z_c) = (1-\alpha^p)\mu(z) \\ &\overset{\text{Lem 5 p (3)}}{\leq} (1 - \frac{1}{2}\sqrt{\frac{\beta}{n}})\mu(z) \leq e^{-\frac{1}{2}\sqrt{\frac{\beta}{n}}}\mu(z) \overset{\text{ind. hyp.}}{\leq} e^{-\frac{1}{2}\sqrt{\frac{\beta}{n}}(T+1)}\mu(z^0). \end{aligned}$$

Recalling that $z_p = (x_p, s_p, y_p)$ is the iterate produced after $T+1$ iterations, that $n\mu(z_p) = \langle x_p, s_p \rangle$ and $n\mu(z^0) = \langle x^0, s^0 \rangle$, we conclude that the induction hypothesis continues to hold after $T+1$ iterations.

For furthermore, let $\overline{T}$ be the number of iterations performed by Predictor-Corrector IPM with target error $\langle x^0, s^0 \rangle \geq \varepsilon > 0$. Since $\langle x_0, s_0 \rangle / \varepsilon \geq 1$, note that $2\sqrt{\frac{n}{\beta}} \ln(\langle x^0, s^0 \rangle / \varepsilon) + 1 \geq 1$. Therefore, we may assume that $\overline{T} \geq 1$ since otherwise the statement holds trivially. Then, by definition of $\overline{T}$, it must be the case that the duality gap after iteration $\overline{T} - 1 \geq 0$ is greater than ε. By the invariant proved above, we conclude that $\langle x^0, s^0 \rangle e^{-\frac{1}{2}\sqrt{\frac{\beta}{n}}(\overline{T}-1)} > \varepsilon$. In particular, $\overline{T} < 2\sqrt{\frac{n}{\beta}} \ln(\langle x^0, s^0 \rangle / \varepsilon) + 1$, as needed. □

3.1 Helper Propositions

We begin with the following simple propositions.

Proposition 7 *For $z = (x, s, y) \in \mathcal{P}_{++} \times \mathcal{D}_{++}$. If $\mathrm{dist_c}(z) \leq \beta$, then $(1-\beta)1_n \leq \frac{xs}{\mu(z)} \leq (1+\beta)1_n$.*

Proof: By definition $\mathrm{dist_c}(z) \leq \beta$ if only and only we can write $\frac{xs}{\mu} = 1_n + \xi$, where $\|\xi\|_2 \leq \beta$. Therefore, $\|\xi\|_\infty := \max_{i \in [n]} |\xi_i| \leq \|\xi\|_2 \leq \beta$. In particular, $(1-\beta)1_n \leq 1_n + \xi \leq (1+\beta)1_n$, as needed. □

Proposition 8 *Let $u, v \in \mathbb{R}^n$. Then $\|uv\|_2 \leq \frac{1}{2}(\|u\|_2^2 + \|v\|_2^2)$.*

PROOF: Recall that for $a, b \geq 0$, then $a + b \geq 2\sqrt{ab}$ since $a + b - 2\sqrt{ab} = (\sqrt{a} - \sqrt{b})^2 \geq 0$. Using this inequality with $a = \|u\|_2^2$ and $b = \|v\|_2^2$, we derive the desired inequality as follows

$$\begin{aligned}\frac{1}{2}(\|u\|_2^2 + \|v\|_2^2) \geq \|u\|_2 \|v\|_2 &= \sqrt{\sum_{i\in[n]} u_i^2}\sqrt{\sum_{j\in[n]} v_j^2} = \sqrt{\sum_{i,j\in[n]} u_i^2 v_j^2} \\ &\geq \sqrt{\sum_{i\in[n]} u_i^2 v_i^2} = \|uv\|_2 .\end{aligned}$$

□

PROPOSITION 9 *Define* $h(\alpha) = \frac{1-\alpha}{\alpha^2}$ *for* $\alpha \in (0, 1]$*. Then for* $\nu \geq 0$*,*

$$h^{-1}(\nu) \geq \begin{cases} 1 - \nu & : 0 \leq \nu \leq 1/2 \\ \frac{1}{2} & : 1/2 \leq \nu \leq 1 \\ \frac{1}{2\sqrt{\nu}} & : \nu \geq 1 \end{cases}$$

PROOF: Since $h(\alpha) = \frac{1-\alpha}{\alpha^2}$ is monotone decreasing in $\alpha \in (0, 1]$, we have that $h^{-1}(\nu) \geq \alpha \Leftrightarrow h(\alpha) \geq \nu$. Using this, we verify the desired lower bounds on h^{-1}. For $0 \leq \nu \leq 1/2$, we have $h(1 - \nu) = \frac{1-(1-\nu)}{(1-\nu)^2} \geq \nu$. For $1/2 \leq \nu \leq 1$, we have $h(1/2) = \frac{1-1/2}{(1/2)^2} = 2 \geq \nu$. For $\nu \geq 1$, we have $h((2\sqrt{\nu})^{-1}) = (1 - \frac{1}{2\sqrt{\nu}})(2\sqrt{\nu})^2 \geq (1 - 1/2)4\nu \geq \nu$. □

3.2 Proof of Lemma 4

Let $\Delta z^c = (\Delta x^c, \Delta s^c, \Delta y^c)$ be the corrector at $z \in N_2(\beta)$, which satisfies $s\Delta x^c + x\Delta s^c = \mu 1_n - xs$ where $\mu := \mu(z)$, $\Delta x^c \in \ker(\mathbf{A})$, $\Delta s^c \in \operatorname{im}(\mathbf{A}^\top)$. In particular, $\langle \Delta x^c, \Delta s^c \rangle = 0$.

Proof of Part (1).

$$\begin{aligned}n\mu(z + \Delta z^c) &= \langle x + \Delta x^c, s + \Delta s^c \rangle = \langle x, s \rangle + \langle x, \Delta s^c \rangle + \langle \Delta x^c, s \rangle + \underbrace{\langle \Delta x^c, \Delta s^c \rangle}_{=0} \\ &= \left\langle \underbrace{xs + x\Delta s^c + s\Delta x^c}_{=\mu 1_n}, 1_n \right\rangle = \langle \mu 1_n, 1_n \rangle := n\mu(z).\end{aligned}$$

Proof of Part (2). We start with an exact expression of the centrality distance of $z + \Delta z^c$:

$$\begin{aligned}\operatorname{dist}_c(z + \Delta z^c) &= \left\| \frac{(x + \Delta x^c)(s + \Delta s^c)}{\mu(z + \Delta z^c)} - 1_n \right\|_2 \\ &= \left\| \frac{xs + s\Delta x^c + x\Delta s^c + \Delta x^c \Delta s^c}{\mu} - 1_n \right\|_2 \\ &= \left\| \frac{\mu 1_n + \Delta x^c \Delta s^c}{\mu} - 1_n \right\|_2 = \left\| \frac{\Delta x^c \Delta s^c}{\mu} \right\|_2 .\end{aligned}$$

It suffices to show $\left\|\frac{\Delta x^c \Delta s^c}{\mu}\right\|_2 \leq \frac{\beta^2}{2(1-\beta)}$. This is derived as follows:

$$\begin{aligned}
\left\|\frac{\Delta x^c \Delta s^c}{\mu}\right\|_2 &= \left\|\left(\sqrt{\frac{s}{x\mu}}\Delta x^c\right)\left(\sqrt{\frac{x}{s\mu}}\Delta s^c\right)\right\|_2 \\
&\overset{\text{Prop. (8)}}{\leq} \frac{1}{2}\left(\left\|\sqrt{\frac{s}{x\mu}}\Delta x^c\right\|_2^2 + \left\|\sqrt{\frac{x}{s\mu}}\Delta s^c\right\|_2^2\right) \\
&\overset{\text{Eq. (8)}}{=} \frac{1}{2}\left\|\frac{\mu 1_n - xs}{\sqrt{\mu xs}}\right\|_2^2 = \frac{1}{2}\left\|\frac{\mu 1_n - xs}{\mu}\frac{\sqrt{\mu}}{\sqrt{xs}}\right\|_2^2 \\
&\leq \frac{1}{2}\left\|\frac{\mu 1_n - xs}{\mu}\right\|_2^2 \left\|\frac{\mu}{xs}\right\|_\infty \\
&\overset{z\in N_2(\beta)}{\leq} \frac{\beta^2}{2}\max_{i\in[n]}\frac{\mu}{x_i s_i} \\
&\overset{\text{Prop. (7)}}{\leq} \frac{\beta^2}{2(1-\beta)}.
\end{aligned}$$

3.3 Proof of Lemma 5

Let $\Delta z^p = (\Delta x^p, \Delta s^p, \Delta y^p)$ be the corrector at $z \in N_2(\beta/2)$, which satisfies $s\Delta x^p + x\Delta s^p = -xs$ where $\mu := \mu(z)$, $\Delta x^p \in \ker(\mathbf{A})$, $\Delta s^p \in \operatorname{im}(\mathbf{A}^\top)$. In particular, $\langle \Delta x^p, \Delta s^p\rangle = 0$.

Proof of Part (1).

$$n\mu(z+\alpha\Delta z^p) = \langle x+\alpha\Delta x^p, s+\alpha\Delta s^p\rangle = \langle x, s\rangle + \alpha\langle x, \Delta s^p\rangle + \alpha\langle \Delta x^p, s\rangle + \alpha^2 \underbrace{\langle \Delta x^p, \Delta s^p\rangle}_{=0}$$

$$= \left\langle \underbrace{xs + \alpha x\Delta s^p + \alpha s\Delta x^p}_{=(1-\alpha)xs}, 1_n \right\rangle = (1-\alpha)\langle xs, 1_n\rangle := (1-\alpha)n\mu(z).$$

Proof of Part (2). Recall that $q_z := \left\|\frac{\Delta x^p \Delta s^p}{\mu}\right\|_2$. The bound on q_z is derived as follows:

$$\begin{aligned}
\left\|\frac{\Delta x^p \Delta s^p}{\mu}\right\|_2 &= \left\|\left(\sqrt{\frac{s}{x\mu}}\Delta x^p\right)\left(\sqrt{\frac{x}{s\mu}}\Delta s^p\right)\right\|_2 \\
&\overset{\text{Prop. (8)}}{\leq} \frac{1}{2}\left(\left\|\sqrt{\frac{s}{x\mu}}\Delta x^p\right\|_2^2 + \left\|\sqrt{\frac{x}{s\mu}}\Delta s^p\right\|_2^2\right) \\
&\overset{\text{Eq. (8)}}{=} \frac{1}{2}\left\|\frac{-xs}{\sqrt{\mu xs}}\right\|_2^2 = \frac{1}{2}\sum_{i=1}^n \frac{x_i s_i}{\mu} = \frac{1}{2}n.
\end{aligned}$$

Proof of Part (3). Let $\alpha \in [0,1)$. We bound the centrality distance as follows:

$$\begin{aligned}
\operatorname{dist}_c(z+\alpha\Delta z^p) &= \left\| \frac{(x+\alpha\Delta x^p)(s+\Delta s^p)}{\mu(z+\alpha\Delta z^p)} - 1_n \right\|_2 \\
&= \left\| \frac{xs+\alpha s\Delta x^p+\alpha x\Delta s^p+\alpha^2\Delta x^p\Delta s^p}{(1-\alpha)\mu} - 1_n \right\|_2 \\
&= \left\| \frac{(1-\alpha)xs+\alpha^2\Delta x^c\Delta s^c}{(1-\alpha)\mu} - 1_n \right\|_2 \\
&\le \left\| \frac{xs}{\mu} - 1_n \right\|_2 + \frac{\alpha^2}{1-\alpha}\left\| \frac{\Delta x^p\Delta s^p}{\mu} \right\|_2 \\
&\overset{z\in N_2(\beta/2)}{\le} \frac{\beta}{2} + \frac{\alpha^2}{1-\alpha} q_z.
\end{aligned}$$

For $\alpha^p := h^{-1}(\frac{2q_z}{\beta})$, we must show that $z+\alpha\Delta z^p \in N_2(\beta)$ for $\alpha \in [0,\alpha^p)$. By the above, it suffices to show that $\frac{\alpha^2}{1-\alpha}q_z \le \frac{\beta}{2}, \forall\alpha \in [0,\alpha^p) \Leftrightarrow \frac{2q_z}{\beta} \le \frac{1-\alpha}{\alpha^2}, \forall \alpha \in [0,\alpha^p)$. This follows directly from the definition of h^{-1} using the fact that h is monotone decreasing. From here, we have

$$\alpha^p = h^{-1}\left(\frac{2q_z}{\beta}\right) \overset{\text{P (2)}}{\ge} h^{-1}\left(\frac{n}{\beta}\right) \ge \frac{1}{2}\sqrt{\frac{\beta}{n}},$$

where the last inequality follows from Proposition 9 and $\frac{n}{\beta} \ge 1$.

4 Initialization

In this section, we present the (magical) homogeneous self-dual initialization due to Ye, Todd and Mizuno [YTM94], which works in an extended space.

Self-Dual Optimality System. We start with the following homogeneous feasibility system (SDOS):

$$\begin{aligned}
&\mathbf{A}x = \tau b \\
&\mathbf{A}^\top y + s = \tau c \\
&\langle c,x\rangle - \langle b,y\rangle + \kappa = 0 \\
&x \ge 0_n, s \ge 0_n, \tau \ge 0, \kappa \ge 0, y \in \mathbb{R}^m.
\end{aligned} \tag{SDOS}$$

Note that any solution above with $\tau > 0$ satisfies that $(x/\tau, s/\tau, y/\tau) \in \mathcal{P}\times\mathcal{D}$. As in the computation of (Gap Formula), the equality constraints of (SDOS) imply

$$-\kappa\tau = (\langle c,x\rangle - \langle b,y\rangle)\tau = \langle \tau c,x\rangle - \langle \tau b,y\rangle = \langle \tau c,x\rangle - \langle Ax,y\rangle = \left\langle \tau c - \mathbf{A}^\top y, x\right\rangle = \langle s,x\rangle.$$

In particular, for any feasible solution, we have

$$\underbrace{\langle s,x\rangle}_{\ge 0} + \underbrace{\kappa\tau}_{\ge 0} = 0. \tag{9}$$

If $\tau > 0$, then by the above we must have $\kappa = 0$. Therefore, if $\tau > 0$, then $(x/\tau, s/\tau, y/\tau)$ form an optimal primal-dual pair of solutions. Similarly, if $\kappa > 0$, then $\tau = 0$, and then $\langle c,x\rangle - \langle b,y\rangle + \kappa =$

0 implies that either $\langle c, x\rangle < 0$ or $\langle b, y\rangle > 0$. In the former case, we get $\langle c, x\rangle < 0$, $\mathbf{A}x = 0_n$ (since $\tau = 0$) and $x \geq 0_n$, and thus x certifies that the primal is unbounded from below (i.e., the dual is infeasible). Similarly in the latter case, we get $\langle y, b\rangle > 0$, $\mathbf{A}^\top y + s = 0_n$ (since $\tau = 0$) and $s \geq 0_n$, which certifies that the dual is unbounded from below (i.e., the primal is infeasible). Linear programming duality indeed guarantees that there always exists a solution to (SDOS) with either τ or κ positive.

Self-Dual Initialization System. From the above discussion, finding a solution either with $\tau > 0$ or $\kappa > 0$ immediately solves the primal-dual linear program (it provides optimal solutions or an infeasibility certificate for one side). As of yet however, it is not clear how one could initialize a path following scheme to find such a solutions.

The insight of Ye, Todd and Mizuno [YTM94] is that one can add an additional variable and constraint which allows one to violate the non-negativity constraints, and where one can explicitly initialize the central path for minimizing infeasibility. The corresponding self-dual initialization system (SDIS) is given by

$$\begin{aligned}
&\min \quad (n+1)\theta \\
&\mathbf{A}(x - \theta 1_n) = (\tau - \theta) b \\
&\mathbf{A}^\top y + (s - \theta 1_n) = (\tau - \theta) c \\
&\langle c, x - \theta 1_n\rangle - \langle b, y\rangle + (\kappa - \theta) = 0 \\
&\langle 1_n, x\rangle + \langle 1_n, s\rangle + \tau + \kappa - (n+1)\theta = n+1 \\
&x \geq 0_n, s \geq 0_n, \tau \geq 0, \kappa \geq 0, y \in \mathbb{R}^m, \theta \in \mathbb{R}.
\end{aligned} \tag{SDIS}$$

If $(x, s, \tau, \kappa, y, \theta)$ is feasible for (SDIS) only if $(x - \theta 1_n, s - \theta 1_n, \tau - \theta, \kappa - \theta, y)$ is feasible for (SDOS) where the non-negativity constraints are relaxed from ≥ 0 to $\geq -\theta$. Using this interpretation combined with (9), we see that $\theta \geq 0$ for any feasible solution to (SDIS). In particular, the optimal value of the above program is 0, and any feasible solution with $\theta = 0$ is a feasible solution to (SDOS) with $\langle 1_n, x\rangle + \langle 1_n, s\rangle + \tau + \kappa = n + 1$ (in particular, the solution is non-zero).

The above program is self-dual in the sense that the dual program is equivalent to the primal problem. The dual variables are also $(x, s, \tau, \kappa, y, \theta)$ having the same constraints, where the dual objective is $\max -(n+1)\theta$. Most important in this duality, is which variables are dual to each other, which we require to define the primal-dual gap. In particular, the x and s variables are dual, and the variables κ and τ are dual.

This self-duality yields the following symmetric central path equations:
$(x(\mu), s(\mu), \tau(\mu), \kappa(\mu), y(\mu), \theta(\mu))$ is the primal central path point at parameter $\mu > 0$ if it is feasible for (SDIS) and

$$x(\mu)s(\mu) = \mu 1_n \text{ and } \tau(\mu)\kappa(\mu) = \mu.$$

The dual central path is the same except that we swap $x(\mu)$ with $s(\mu)$ as well as $\tau(\mu)$ with $\kappa(\mu)$. With these equations, it is easy to see that the primal central path point at parameter $\mu = 1$ is given by

$$(x(1), s(1), \tau(1), \kappa(1), y(1), \theta(1)) = (1_n, 1_n, 1, 1, 0_m, 1).$$

With this explicit initialization, we can now directly apply Predictor-Corrector IPM starting from $\mu = 1$. As the IPM ensures that the limit optimal solutions are strictly complementary, we will have that either $\tau(\mu)$ or $\kappa(\mu)$ will converge to something non-zero as $\mu \to 0$. We note that the duality gap in this initialization roughly corresponds to "approximate" feasibility and optimality in the original system (assuming the starting primal and dual LPs are feasible).

As a last remark, we note that a naive implementation of the predictor-corrector IPM would explicitly keep track of both primal and dual iterates, which corresponds to two times $2(n+1)+m+1$ many variables. Due to the symmetry, one can design an optimized implementation that only directly keeps track of one side.

References

[ABGJ18] Xavier Allamigeon, Pascal Benchimol, Stéphane Gaubert, and Michael Joswig. Log-barrier interior point methods are not strongly polynomial. *SIAM Journal on Applied Algebra and Geometry*, 2(1):140–178, 2018.

[ADL+22] Xavier Allamigeon, Daniel Dadush, Georg Loho, Bento Natura, and László A Végh. Interior point methods are not worse than simplex. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 267–277. IEEE, 2022.

[AGV22] Xavier Allamigeon, Stéphane Gaubert, and Nicolas Vandame. No self-concordant barrier interior point method is strongly polynomial. In *Proceedings of the 54th Annual ACM Symposium on Theory of Computing (STOC)*, pages 515–528, 2022.

[Dan90] George B. Dantzig. *Origins of the simplex method*, page 141–151. Association for Computing Machinery, New York, NY, USA, 1990.

[DKN+24] Daniel Dadush, Zhuan Khye Koh, Bento Natura, Neil Olver, and László A Végh. A strongly polynomial algorithm for linear programs with at most two nonzero entries per row or column. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1561–1572, 2024.

[Gon96] Jacek Gondzio. Multiple centrality corrections in a primal-dual method for linear programming. *Computational optimization and applications*, 6(2):137–156, 1996.

[Kar84] Narendra Karmarkar. A new polynomial-time algorithm for linear programming. In *Proceedings of the 16th Annual ACM Symposium on Theory of Computing (STOC)*, pages 302–311, 1984.

[Kha79] Leonid G Khachiyan. A polynomial algorithm in linear programming. In *Doklady Academii Nauk SSSR*, volume 244, pages 1093–1096, 1979.

[KMY89] Masakazu Kojima, Shinji Mizuno, and Akiko Yoshise. *A Primal-Dual Interior Point Algorithm for Linear Programming*, pages 29–47. Springer New York, New York, NY, 1989.

[Meg83] Nimrod Megiddo. Towards a genuinely polynomial algorithm for linear programming. *SIAM Journal on Computing*, 12(2):347–353, 1983.

[MTY93] Shinji Mizuno, Michael Todd, and Yinyu Ye. On adaptive-step primal-dual interior-point algorithms for linear programming. *Mathematics of Operations Research*, 18:964–981, 11 1993.

[Ren88] James Renegar. A polynomial-time algorithm, based on Newton's method, for linear programming. *Mathematical Programming*, 40(1-3):59–93, 1988.

[Sho77] N.Z. Shor. Cut-off method with space extension in convex programming problems. *Kibernetika*, 13(1):94–95, 1977. Translated in Cybernetics 13(1), 94-96.

[Sma98] Steve Smale. Mathematical problems for the next century. *The Mathematical Intelligencer*, 20:7–15, 1998.

[Wri97] Stephen J Wright. *Primal-dual interior-point methods*. SIAM, 1997.

[YTM94] Yinyu Ye, Michael J. Todd, and Shinji Mizuno. An $O(\sqrt{n}L)$-iteration homogeneous and self-dual linear programming algorithm. *Mathematics of Operations Research*, 19(1):53–67, 1994.

[Yud76] A.S Yudin, D.B. Nemirovski. Informational complexity and effective methods of solution for convex extremal problem. *Ekonomika i Matematicheskie Metody*, 12:357–359, 1976. Translated in Matekon: Translations of Russian and East European Math. Economics 13, 25-45, Spring '77.