Tutorial on Robust Interior Point

May 8, 2021

In this note, we give a short self-contained proof for interior point method and its robust version. Consider the primal linear program

$$\min_{\mathbf{A}x=b,x\in\mathbb{R}^n_{\geq 0}} c^\top x \tag{P}$$

and its dual

$$\max_{\mathbf{A}^\top y+s=c,s\in\mathbb{R}^n_{\geq 0}} b^\top y. \tag{D}$$

where $\mathbf{A}\in\mathbb{R}^{d\times n}$ and $\mathbb{R}_{\geq}=\{x\geq 0\}$. The feasible region for both programs are $\mathcal{P}=\{x\in\mathbb{R}^n_{\geq}:\mathbf{A}x=b\}$ and $\mathcal{D}=\{s\in\mathbb{R}^n_{\geq}:\mathbf{A}^\top y+s=c \text{ for some } y\}$. We define their interior $\mathcal{P}^\circ=\{x\in\mathbb{R}^n_{>}:\mathbf{A}x=b\}$ and $\mathcal{D}^\circ=\{s\in\mathbb{R}^n_{>}:\mathbf{A}^\top y+s=c \text{ for some } y\}$.

To motivate interior point methods, we recall the optimality condition for linear programs:

Theorem 1 (Complementary Slackness)**.** *For any $x\in\mathcal{P}$ and $s\in\mathcal{D}$, x and s are optimal if and only if $x^\top s=0$. Furthermore, if both $\mathcal{P}$ and $\mathcal{D}$ are non-empty, there are optimal $x^*\in\mathcal{P}$ and $s^*\in\mathcal{D}$ such that $(x^*)^\top s^*=0$ and $x^*+s^*>0$.*

More generally, the quantity $x^\top s$ measures the duality gap of the feasible solution:

Lemma 2 (Duality Gap)**.** *For any $x\in\mathcal{P}$ and $s\in\mathcal{D}$, the duality gap $c^\top x-b^\top y=x^\top s$. In particular $c^\top x\leq\min_{x\in\mathcal{P}}c^\top x+x^\top s$.*

Proof. Using $\mathbf{A}x=b$ and $\mathbf{A}^\top y+s=c$, we can compute the duality gap as follows

$$c^\top x-b^\top y=c^\top x-(\mathbf{A}x)^\top y=c^\top x-x^\top(\mathbf{A}y)=x^\top s.$$

By weak duality, we have

$$c^\top x=b^\top y+x^\top s\leq\max_{\mathbf{A}^\top y+s=c,s\in\mathbb{R}^n_+}b^\top y+x^\top s\leq\min_{x\in\mathcal{P}}c^\top x+x^\top s.$$

$\square$

The main implication of Lemma 2 is that any feasible (x,s) with small $x^\top s$ is a nearly optimal solution of the linear program. This leads us to the primal-dual algorithms in which we start with a feasible primal and dual solution (x,s) and iteratively update the solution to decrease the duality gap $x^\top s$.

1 Interior Point Method

In this note, we discuss the short-step interior point method. The algorithm maintains a pair $(x,s)\in\mathcal{P}^\circ\times\mathcal{D}^\circ$ and a scalar $t>0$ satisfying the invariant $\|\frac{xs}{t}-1\|_2\leq\frac{1}{4}$. Each step, it decreases t by a factor of $1-\Omega(n^{-1/2})$ while maintaining the invariant.

1.1 Basic Property of a Step

To see why there is a pair $(x,s)\in\mathcal{P}^\circ\times\mathcal{D}^\circ$ satisfying the invariant, we prove the following generalization:

Lemma 3 (Quadrant Representation of Primal-Dual)**.** *Suppose $\mathcal{P}$ is non-empty and bounded. For any positive vector $\mu \in \mathbb{R}^n_{>}$, there is an unique pair $(x_\mu, s_\mu) \in \mathcal{P}^\circ \times \mathcal{D}^\circ$ such that $x_\mu s_\mu = \mu$. Furthermore, $x_\mu = \min_{x\in\mathcal{P}} f_\mu(x)$ where*

$$f_\mu(x) = c^\top x - \sum_{i=1}^{n} \mu_i \ln x_i.$$

Proof. Fix $\mu \in \mathbb{R}^n_{>}$. We define $x_\mu = \arg\min_{x\in\mathcal{P}} f_\mu(x)$ and prove that $(x_\mu, s_\mu) \in \mathcal{P}^\circ \times \mathcal{D}^\circ$ with $x_\mu s_\mu = \mu$ for some s_μ. Since $\mathcal{P}$ is non-empty and bounded and since f_μ is strictly convex, such unique x exists. Furthermore, since $f_\mu(x) \to +\infty$ as $x_i \to 0$ for any i, we have that $x_\mu \in \mathcal{P}^\circ$.

By the optimality condition for f_μ, there is a vector y such that

$$c - \frac{\mu}{x} = \mathbf{A}^\top y.$$

Define $s_\mu = \frac{\mu}{x_\mu}$, then one can check that $s_\mu \in \mathcal{D}$ and $x_\mu s_\mu = \mu$.

For the uniqueness, if $(x, s) \in \mathcal{P}^\circ \times \mathcal{D}^\circ$ and $xs = \mu$, then x satisfies the optimality condition for f_μ. Since f_μ is strictly convex, such x is unique. ☐

Lemma 3 shows that any point in $\mathcal{P}^\circ \times \mathcal{D}^\circ$ is uniquely represented by a positive vector μ. Interior point methods move μ uniform to 0 while maintaining its corresponding x_μ. Now, we discuss how to find (x_μ, s_μ) given a nearby interior feasible point (x, s). Namely, how to move (x, s) to $(x + \delta_x, s + \delta_s)$ such that it satisfies the equation

$$\begin{aligned}
(x + \delta_x)(s + \delta_s) &= \mu, \\
\mathbf{A}(x + \delta_x) &= b, \\
\mathbf{A}^\top(y + \delta_y) + (s + \delta_s) &= c, \\
(x + \delta_x, s + \delta_s) &\in \mathbb{R}^{2n}_{>0}.
\end{aligned}$$

Although the equation above involves y, our approximate solution does not need to know y. By ignoring the second order term $\delta_x\delta_s$ on the equation above and the inequality constraint, we can simplify the formula a little bit by using $\mathbf{A}x = 0$ and $\mathbf{A}^\top y + s = c$:

$$\begin{aligned}
xs + \mathbf{S}\delta_x + \mathbf{X}\delta_s &= \mu, \\
\mathbf{A}\delta_x &= 0, \\
\mathbf{A}^\top\delta_y + \delta_s &= 0,
\end{aligned} \tag{1.1}$$

where $\mathbf{X}$ and $\mathbf{S}$ are the diagonal matrix with diagonal x and s. In the following Lemma, we show how to write the step above using a projection matrix.

Lemma 4. *Suppose that $\mathbf{A}$ has full row rank and $(x, s) \in \mathcal{P}^\circ \times \mathcal{D}^\circ$. Then, the unique solution for the linear system (1.1) is given by*

$$\begin{aligned}
\mathbf{X}^{-1}\delta_x &= (\mathbf{I} - \mathbf{P})(\delta_\mu/\mu), \\
\mathbf{S}^{-1}\delta_s &= \mathbf{P}(\delta_\mu/\mu)
\end{aligned}$$

where $\delta_\mu = \mu - xs$ and $\mathbf{P} = \mathbf{S}^{-1}\mathbf{A}^\top(\mathbf{A}\mathbf{S}^{-1}\mathbf{X}\mathbf{A}^\top)^{-1}\mathbf{A}\mathbf{X}$.

Proof. Note that the step satisfies $\mathbf{S}\delta_x + \mathbf{X}\delta_s = \delta_\mu$. Multiply both sides by $\mathbf{A}\mathbf{S}^{-1}$ and using $\mathbf{A}\delta_x = 0$, we have

$$\mathbf{A}\mathbf{S}^{-1}\mathbf{X}\delta_s = \mathbf{A}\mathbf{S}^{-1}\delta_\mu.$$

Now, we use that $\mathbf{A}^\top\delta_y + \delta_s = 0$ and get

$$\mathbf{A}\mathbf{S}^{-1}\mathbf{X}\mathbf{A}^\top\delta_y = -\mathbf{A}\mathbf{S}^{-1}\delta_\mu.$$

Since $\mathbf{A} \in \mathbb{R}^{m\times n}$ has full row rank and $\mathbf{S}^{-1}\mathbf{X}$ is invertible, we have that $\mathbf{A}\mathbf{S}^{-1}\mathbf{X}\mathbf{A}^\top$ is invertible and $\delta_y = -(\mathbf{A}\mathbf{S}^{-1}\mathbf{X}\mathbf{A}^\top)^{-1}\mathbf{A}\mathbf{S}^{-1}\delta_\mu$ and

$$\delta_s = \mathbf{A}^\top(\mathbf{A}\mathbf{S}^{-1}\mathbf{X}\mathbf{A}^\top)^{-1}\mathbf{A}\mathbf{S}^{-1}\delta_\mu.$$

Putting it into $\mathbf{S}\delta_x + \mathbf{X}\delta_s = \delta_\mu$, we have

$$\delta_x = \mathbf{S}^{-1}\delta_\mu - \mathbf{S}^{-1}\mathbf{X}\mathbf{A}^\top(\mathbf{A}\mathbf{S}^{-1}\mathbf{X}\mathbf{A}^\top)^{-1}\mathbf{A}\mathbf{S}^{-1}\delta_\mu.$$

The result follows from the definition of $\mathbf{P}$. ☐

1.2 Lower Bounding Step Size

The efficiency of interior point methods depends on how large the step we can take while stays within the domain. We first study the step operator $(\mathbf{I}-\mathbf{P})$ and $\mathbf{P}$. The following Lemma shows that $\mathbf{P}$ is a nearly-orthogonal projection matrix when μ is close to a constant vector. Hence, the relative changes of $\mathbf{X}^{-1}\delta_x$ and $\mathbf{S}^{-1}\delta_s$ are essentially the orthogonal decomposition of the relative step δ_μ/μ on μ.

Lemma 5. *Under the assumption in Lemma 4, $\mathbf{P}$ is a projection matrix such that $\|\mathbf{P}v\|_\mu \leq \|v\|_\mu$ for any $v \in \mathbb{R}^n$. Similarly, we have that $\|(\mathbf{I}-\mathbf{P})v\|_\mu \leq \|v\|_\mu$.*

Proof. $\mathbf{P}$ is a projection because $\mathbf{P}^2 = \mathbf{P}$. Define the orthogonal projection

$$\mathbf{P}_{\text{orth}} = \mathbf{S}^{-1/2}\mathbf{X}^{1/2}\mathbf{A}^\top(\mathbf{A}\mathbf{S}^{-1}\mathbf{X}\mathbf{A}^\top)^{-1}\mathbf{A}\mathbf{X}^{1/2}\mathbf{S}^{-1/2},$$

then we have

$$\begin{aligned}
\|\mathbf{P}v\|_\mu^2 &= v^\top\mathbf{X}\mathbf{A}^\top(\mathbf{A}\mathbf{S}^{-1}\mathbf{X}\mathbf{A}^\top)^{-1}\mathbf{A}\mathbf{S}^{-1}\mathbf{X}\mathbf{S}\mathbf{S}^{-1}\mathbf{A}^\top(\mathbf{A}\mathbf{S}^{-1}\mathbf{X}\mathbf{A}^\top)^{-1}\mathbf{A}\mathbf{X}v \\
&= v^\top\mathbf{S}^{1/2}\mathbf{X}^{1/2}\mathbf{P}_{\text{orth}}\mathbf{S}^{1/2}\mathbf{X}^{1/2}v \\
&\leq v^\top\mathbf{S}^{1/2}\mathbf{X}^{1/2}\mathbf{S}^{1/2}\mathbf{X}^{1/2}v = \|v\|_\mu^2.
\end{aligned}$$

The calculation for $\|(\mathbf{I}-\mathbf{P})v\|_\mu$ is similar. $\square$

Now, we give a lower bound of the largest feasible step size:

Lemma 6. *We have that $\|\mathbf{X}^{-1}\delta_x\|_\infty^2 \leq \frac{1}{\min_i \mu_i}\|\delta_\mu/\mu\|_\mu^2$ and $\|\mathbf{S}^{-1}\delta_s\|_\infty^2 \leq \frac{1}{\min_i \mu_i}\|\delta_\mu/\mu\|_\mu^2$. In particular, if $\|\delta_\mu/\mu\|_\mu^2 < \min_i \mu_i$, we have $(x+\delta_x, s+\delta_s) \in \mathcal{P}^\circ \times \mathcal{D}^\circ$*

Proof. For $\|\mathbf{X}^{-1}\delta_x\|_\infty$, we have $\min_i \mu_i \|\mathbf{X}^{-1}\delta_x\|_\infty^2 \leq \|\mathbf{X}^{-1}\delta_x\|_\mu^2$ and hence

$$\|\mathbf{X}^{-1}\delta_x\|_\infty^2 \leq \frac{1}{\min_i \mu_i}\|\mathbf{X}^{-1}\delta_x\|_\mu^2 = \frac{1}{\min_i \mu_i}\|(\mathbf{I}-\mathbf{P})(\delta_\mu/\mu)\|_\mu^2 \leq \frac{1}{\min_i \mu_i}\|\delta_\mu/\mu\|_\mu^2.$$

The proof for $\|\mathbf{S}^{-1}\delta_s\|_\infty$ is similar.

Hence, if $\|\delta_\mu/\mu\|_\mu^2 < \min_i \mu_i$, we have that $\|\mathbf{X}^{-1}\delta_x\|_\infty < 1$ and $\|\mathbf{S}^{-1}\delta_s\|_\infty < 1$. Therefore, $x+\delta_x$ and $s+\delta_s$ are feasible. $\square$

To decrease μ uniformly, we set $\delta_\mu = -h\mu$ for some step size h. To ensure the feasibility, we need $\|\delta_\mu/\mu\|_\mu^2 \leq \min_i \mu_i$ and this gives the maximum step size

$$h = \sqrt{\frac{\min_i \mu_i}{\sum_i \mu_i}}. \tag{1.2}$$

Note that the above quantity maximizes at $n^{-1/2}$ when μ is a constant vector.

Definition 7 (Central Path)**.** We define the central path $(x_t, s_t) \in \mathcal{P}^\circ \times \mathcal{D}^\circ$ by $x_t s_t = t$. We call x_t is on the central path of (P) at t.

1.3 Staying within ℓ_2 distance

Since the step size (1.2) maximizes when μ is a constant vector. A natural approach is to control μ vector ℓ_2 close to a constant t. This motivates the following algorithm:

Algorithm 1: `L2Step`$(\mathbf{A}, x, s, t_{\text{start}}, t_{\text{end}})$

Invariant: $(x,s) \in \mathcal{P}^\circ \times \mathcal{D}^\circ$ and $\|xs - t\|_2 \leq \frac{t}{4}$.
Define $\mathbf{P}_{x,s} = \mathbf{S}^{-1}\mathbf{A}^\top(\mathbf{A}\mathbf{S}^{-1}\mathbf{X}\mathbf{A}^\top)^{-1}\mathbf{A}\mathbf{X}$.
Let $t = t_{\text{start}}$, $h = 1/(16\sqrt{n})$ and n is the number of columns in $\mathbf{A}$.
while $t \neq t_{\text{end}}$ *and it is not the first step* **do**
- Let $t' = \begin{cases} \max((1-h)t, t_{\text{end}}) & \text{if } t_{\text{start}} \leq t_{\text{end}} \\ \min((1+h)t, t_{\text{end}}) & \text{else} \end{cases}$
- Let $\mu = xs$ and $\delta_\mu = t' - \mu$.
- Let $\delta_x = \mathbf{X}(\mathbf{I}-\mathbf{P}_{x,s})(\delta_\mu/\mu)$ and $\delta_s = \mathbf{S}\mathbf{P}_{x,s}(\delta_\mu/\mu)$.
- Set $x \leftarrow x + \delta_x$, $s \leftarrow s + \delta_s$ and $t \leftarrow t'$.

end
Return (x,s)

Note that the algorithm requires some initial point (x, s) close to the central path and we will show how to get it by changing the cost vector temporarily.

Theorem 8. *Suppose that the input satisfies $(x, s) \in \mathcal{P}^\circ \times \mathcal{D}^\circ$ and $\|xs - t_{\text{start}}\|_2 \le \frac{t_{\text{start}}}{4}$. Then throughout the algorithm* `L2Step` *(Algorithm 1), we have $\|xs - t_{\text{end}}\|_2 \le \frac{t_{\text{end}}}{6}$ (except the first step). Furthermore,* `L2Step` *takes $O(\sqrt{n}|\log(t_{\text{start}}/t_{\text{end}})|)$ Newton steps (defined in (1.1)).*

Proof. We prove by induction that $\|xs - t\|_2 \le \frac{t}{4}$. It holds for the first step by the assumption.

Let $x' = x + \delta_x$, $s' = s + \delta_s$ and t' defined in the algorithm. Note that

$$\begin{aligned} x's' - t' =& (x + \delta_x)(s + \delta_s) - t' \\ =& \mu + \mathbf{S}\delta_x + \mathbf{X}\delta_s + \delta_x\delta_s - t'. \end{aligned}$$

Lemma 4 shows that $\mathbf{S}\delta_x + \mathbf{X}\delta_s = t' - \mu$. Hence, we have

$$x's' - t' = \delta_x\delta_s = \mathbf{X}^{-1}\delta_x \cdot \mathbf{S}^{-1}\delta_s \cdot \mu$$

Using this, we have

$$\begin{aligned} \|x's' - t'\|_2 &\le \|\mu^{1/2}\mathbf{X}^{-1}\delta_x\|_2\|\mu^{1/2}\mathbf{S}^{-1}\delta_s\|_2 \\ &= \|\mathbf{X}^{-1}\delta_x\|_\mu\|\mathbf{S}^{-1}\delta_s\|_\mu \\ &\le \|\delta_\mu/\mu\|_\mu^2 \end{aligned}$$

where we used Lemma 5 at the end.

For the last term, using $t' - \mu = \frac{t'}{t}(t - \mu) + (\frac{t'}{t} - 1)\mu$, we have

$$\begin{aligned} \|\delta_\mu/\mu\|_\mu &= \|\frac{t'}{t}\frac{t - \mu}{\mu} + (\frac{t'}{t} - 1)\|_\mu \\ &\le \frac{t'}{t}\|xs - t\|_{\mu^{-1}} + \|\frac{t'}{t} - 1\|_\mu. \end{aligned}$$

Since $\|\mu - t\|_2 \le \frac{t}{4}$, we have $\min_i \mu_i \ge \frac{3t}{4}$ and $\max_i \mu_i \le \frac{5}{4}t$. Using $|\frac{t'}{t} - 1| \le h = \sqrt{n}/16$, we have

$$\|\delta_\mu/\mu\|_\mu \le \frac{t'}{t}\sqrt{\frac{4}{3t}}\|xs - t\|_2 + h\sqrt{\frac{5}{4}n} \le \frac{17}{16}\sqrt{\frac{t}{12}} + h\sqrt{\frac{5}{4}n} \le 0.38\sqrt{t}.$$

Hence, we have $\|x's' - t'\|_2 \le \|\delta_\mu/\mu\|_\mu^2 \le 0.15t \le t'/6$. Furthermore, $\|\delta_\mu/\mu\|_\mu^2 < \min_i \mu_i$ which implies (x, s) is feasible (Lemma 6).

Since `L2Step` decreases t by $1 - \Omega(n^{-1/2})$ factor each step and since we start with $t = t_{\text{start}}$ and ends with $t \approx t_{\text{end}}$, the number of iterations is $O(\sqrt{n}|\log(t_{\text{start}}/t_{\text{end}})|)$. ☐

1.4 Solving LP Approximately and Exactly

The runtime of interior point method depends on how degenerate is the linear program:

Definition 9. We define the following parameters for the linear program $\min_{\mathbf{A}x=b, x\ge 0} c^\top x$:

1. Inner radius r: There is $x \ge r$ with $\mathbf{A}x = b$.
2. Outer radius R: For any $x \ge 0$ with $\mathbf{A}x = b$, we have that $\|x\|_2 \le R$.
3. Lipschitz constant L: $\|c\|_2 \le L$.

Since `L2Step` requires an explicit central path, we modify the linear program to make it happens. To satisfy the constraint $\mathbf{A}x = b$, we start the algorithm by taking a least square solution of the constraint $\mathbf{A}x = b$. Since it can be negative, we write the variable $x = x^+ - x^-$ with both $x^+, x^- \ge 0$. We put a large cost vector on x^- to ensure the solution is roughly the same. The crux of the proof is that if we optimize this new program well enough, we will have $x^+ - x^- > 0$ and hence this gives a good starting point of the original program. Since the proof is a bit complicated and not interested, we defer both the proof and the algorithm in Appendix A.

Now, we state our main algorithm:

Algorithm 2: `L2LPApproximate`$(\mathbf{A}, b, c, x^{(0)}, \epsilon)$

Assume the linear program has diameter R and Lipschitz constant L.
$(x, s) =$ `L2Center`$(\mathbf{A}, b, c, x^{(0)})$
$(x, s) =$ `L2Step`$(\mathbf{A}, x, s, LR, t_{\text{end}})$ with $t_{\text{end}} = \epsilon LR/(2n)$.
Return x.

Theorem 10. *Consider a linear program* $\min_{\mathbf{A}x=b, x\geq 0} c^\top x$ *with* n *variables and* d *constraints. Assume the linear program has inner radius* r*, outer radius* R *and Lipschitz constant* L *(See Definition 9),* `L2LPApproximate` *outputs* x *such that*

$$\begin{aligned} c^\top x &\leq \min_{\mathbf{A}x=b, x\geq 0} c^\top x + \epsilon LR, \\ \mathbf{A}x &= b, \\ x &\geq 0. \end{aligned}$$

The algorithm takes $O(\sqrt{n}\log(nR/(\epsilon r)))$ *Newton steps (defined in (1.1)).*

Finally, if we assume that the solution $x^* = \arg\min_{\mathbf{A}x=b,x\geq 0} c^\top x$ *is unique and that* $c^\top x \geq c^\top x^* + \eta LR$ *for any other vertex* x *of* $\{\mathbf{A}x = b, x \geq 0\}$*, then, we have that* $\|x - x^*\|_2 \leq \frac{2\epsilon R}{\eta}$*.*

Proof. By Theorem ??, we can find (x, s) such that $xs \approx t$ with $t = O(LR)$. Now, Theorem 8 shows that `L2Step` returns (x, s) such that $xs \approx t_{\text{end}}$. Hence, Lemma 2 shows that

$$c^\top x \leq \min_{\mathbf{A}x=b,x\geq 0} c^\top x + x^\top s \leq \min_{\mathbf{A}x=b,x\geq 0} c^\top x + 2t_{\text{end}} n \leq \min_{\mathbf{A}x=b,x\geq 0} c^\top x + \epsilon LR.$$

For the runtime, `L2Center` takes $O(\sqrt{n}\log(nR/r))$ Newton steps and `L2Step`$(\mathbf{A}, x, s, t, t_{\text{end}})$ takes $O(\sqrt{n}\log(n/\epsilon))$ Newton steps.

For the last conclusion, we assume $\epsilon \leq \eta$ and let $\mathcal{P}_t = \mathcal{P} \cap \{c^\top x \leq c^\top x^* + tLR\}$. Note that $\mathcal{P}_\eta$ is a cone at x^* (because there is no vertex except x^* with value less than $c^\top x^* + tLR$). Hence, we have $\mathcal{P}_\epsilon - x^* = \frac{\epsilon}{\eta}(\mathcal{P}_\eta - x^*)$. Since $x \in \mathcal{P}_\epsilon$, we have that

$$\|x - x^*\|_2 \leq \frac{\epsilon}{\eta}\text{radius}(\mathcal{P}_\eta - x^*) \leq \frac{2\epsilon R}{\eta}.$$

□

If we know the solution of the linear program is integral, then getting a solution close enough to x^* allows us to round the solution to integral. Therefore, the last conclusion of last theorem gives us an exact linear program algorithm assuming $\mathbf{A}, b, c$ are integral and bounded. The uniqueness assumption can be achieved by perturbing the cost vector by a random vector via isolation lemma [1, Lemma 4].

References

[1] Adam R Klivans and Daniel Spielman. Randomness efficient identity testing of multivariate polynomials. In *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, pages 216–223, 2001.

A Finding a Point on Central Path

Due to technical reasons, we need to put an extra constraint $1^\top x^+ \leq \Lambda$ for some Λ to ensure the problem is bounded. See Algorithm 3 for the precise formulation of the modified linear program. The formulation is chosen such that we have an explicit interior point.

Algorithm 3: `L2Center`$(\mathbf{A}, b, c)$

Assume the linear program has inner radius r, outer radius R and Lipschitz constant L.
Let $x_c = \mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}b$.; `// We have` $x_c = \arg\min_{\mathbf{A}x=b} \|x\|_2$.
Let $t = 2^{30}n^3LR \cdot \frac{R}{r}$, $x^+ = \frac{t}{c+\frac{t}{60R}}$, $x^- = x^+ - x_c$, $\theta = 60R$.
Let $x^{(0)} = (x^+, x^-, \theta)$ and $s^{(0)} = t/x^{(0)}$, $\overline{\mathbf{A}} = \begin{bmatrix} \mathbf{A} & -\mathbf{A} & 0 \\ 1_d & 0_d & 1 \end{bmatrix}$.
`// Run IPM on the problem` $\min_{(x^+,x^-,\theta)\in\overline{\mathcal{P}}} c^\top x^+ + d^\top x^-$ `with the initial point` $(x^{(0)}, s^{(0)})$
`// where` $\overline{\mathcal{P}} = \{(x^+, x^-, \theta) \in \mathbb{R}^{2n+1}_{\geq 0} : \mathbf{A}(x^+ - x^-) = b, \sum_{i=1}^n x_i^+ + \theta = \Lambda\}$, $d = t/x^-$, $\Lambda = \sum_i x_i^+ + \theta$.
Let $(x^{(1)}, s^{(1)}) =$ `L2Step`$(\overline{\mathbf{A}}, x^{(0)}, s^{(0)}, t, t_{\text{end}})$ with $t_{\text{end}} = LR$.
Write $x^{(1)} = (\alpha, \beta, \gamma)$ and $s^{(1)} = (s_+^{(1)}, s_-^{(1)}, s_\theta^{(1)})$.
Let $x^{(2)} = \alpha - \beta$ and $s^{(2)} = s_+^{(1)} - s_\theta^{(1)}$.
Return $(x^{(2)}, s^{(2)})$.

First, we show that $x^{(0)}$ defined in `L2Center` is indeed on the central path of the modified linear program.

Lemma 11. *Let $\min_{(x^+,x^-,\theta)\in\overline{\mathcal{P}}} c^\top x^+ + d^\top x^-$ and $x^{(0)}$ be the linear program and the initial point defined in `L2Center` (Algorithm 3). Then, we have that $x^{(0)}$ is on the central path of the following problem at t (See Definition 7). Furthermore, we have $140nR \geq \Lambda \geq 45nR$ and $d_i \geq t/(80R)$ for all i.*

Proof. We say (x^+, x^-, θ) is on the central path at t if x^+, x^-, θ are positive and it satisfies the following equation

$$\begin{aligned} \mathbf{A}x^+ - \mathbf{A}x^- &= b, \\ \sum_{i=1}^n x_i^+ + \theta &= \Lambda, \\ \mathbf{A}^\top y + \lambda + s_1 &= c, \\ \mathbf{A}^\top y + s_2 &= d, \\ \lambda + s_3 &= 0, \end{aligned}$$

where for some $s_1, s_2 \in \mathbb{R}^n_{>0}$, $s_3 > 0$, $y \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$.

Now, we verify the solution $x^+ = \frac{t}{c+\frac{t}{60R}}$, $x^- = \frac{t}{c+\frac{t}{60R}} - x_c$, $\theta = 60R$, $y = 0$, $s_1 = \frac{t}{x^+}$, $s_2 = \frac{t}{x^-}$, $s_3 = \frac{t}{60R}$, $\lambda = -s_3$. Using $\mathbf{A}x_c = b$, one can check it satisfies all the equality constraints above.

For the inequality constraints, using $\|c\|_\infty \leq L$ and $t \geq 300LR$, we have

$$50R \leq \frac{t}{L + \frac{t}{60R}} \leq x_i^+ \leq \frac{t}{-L + \frac{t}{60R}} \leq 75R$$

and hence $x^+ > 0$ and so is s_1. Since $\|x_c\|_2 \leq R$ and $x_i^+ \geq 50R$ for all i, we have $x_i^- \geq 49R$ for all i. Hence, x^- and s_2 are positive. Finally, θ and s_3 are positive. Also, we have $\Lambda = \sum_i x_i^+ + \theta \geq (49n + 60)R \geq 45R$.

Since $x^+ \leq 75R$, we have $\Lambda = \sum_i x_i^+ + \theta \leq (75n + 60)R$. Also, since $x^- \leq 76R$ and hence $d \geq t/(76R)$. $\square$

Next, we show that the central path point $x^{(1)}$ found in `L2Center` is far from the constraints $x^+ \geq 0$ and is close to the constraints $x^- \geq 0$. The proof for both involves the same idea: use the optimality condition of $x^{(1)}$. By Theorem 8 shows that $x^{(1)}s^{(1)} = \mu$ for some $\|\mu - t_{end}\|_2 \leq \frac{t}{6}$. We write μ into three parts $(\mu^+, \mu^-, \mu_\theta)$. Similarly, we write $x^{(1)}$ into three parts $x^{(1)} = (\alpha, \beta, \gamma)$. By Lemma 3, we have that $x^{(1)} = (\alpha, \beta, \gamma)$ minimizes the function

$$f(\alpha, \beta, \gamma) \stackrel{\text{def}}{=} c^\top \alpha + d^\top \beta - \sum_{i=1}^n \mu_i^+ \log \alpha_i - \sum_{i=1}^n \mu_i^- \log \beta_i - \mu_\theta \log \gamma$$

over the domain $\overline{\mathcal{P}}$. The gradient of f is a bit complicated and we notice that we only need to consider the directional deriative at $x^{(1)}$ on the direction "$v - x^{(1)}$" where $v \geq r$ is the point such that $\mathbf{A}v = b$. Since our domain is in

$\overline{\mathcal{P}} \subset \mathbb{R}^{2n+1}$, we need to lift v to higher dimension. Now, we define the point

$$\begin{aligned} v^- &= \min(\beta, \frac{32 t_{\text{end}}}{t} R), \\ v^+ &= v + v^-, \\ v_\theta &= \Lambda - \sum_{i=1}^n v^+. \end{aligned}$$

The following Lemma shows that $(v^+, v^-, v_\theta) \in \overline{\mathcal{P}}$.

Lemma 12. *Under the assumptions in Lemma 11, we have that $(v^+, v^-, v_\theta) \in \overline{\mathcal{P}}$. Furthermore, we have $v_\theta \geq 40nR$.*

Proof. Note that (v^+, v^-, v_θ) satisfies the linear constraints in $\overline{\mathcal{P}}$ by construction. It suffices to prove the vector is positive. Since $\beta > 0$, we have $v^- > 0$. Since $v \geq r$, we also have $v^+ > 0$. For v_θ, we use $\Lambda \geq 45R$ (Lemma 11), $v \leq R$ and $v^- \leq \frac{32 t_{\text{end}}}{t} R \leq R$ to get

$$v_\theta \geq 45R - \sum_{i=1}^n (v_i + v_i^-) \geq 40nR.$$

□

Now, we define the path $p(t) = (1-t)(\alpha, \beta, \gamma) + t(v^+, v^-, v_\theta)$. Since $p(0)$ minimizes $f(\alpha, \beta, \gamma)$, we have that $\frac{d}{dt} f(p(t))|_{t=0} \geq 0$. In particular, we have

$$\begin{aligned} 0 &\leq \frac{d}{dt} f(p(t))|_{t=0} \\ &= c^\top (v^+ - \alpha) + d^\top (v^- - \beta) - \sum_{i=1}^n \frac{\mu_i^+}{\alpha_i}(v^+ - \alpha)_i - \sum_{i=1}^n \frac{\mu_i^-}{\beta_i}(v^- - \beta)_i - \frac{\mu_\theta}{\gamma}(v_\theta - \gamma) \\ &= \frac{\mu_\theta}{\gamma}(\gamma - v_\theta) + \sum_{i=1}^n (c_i - \frac{\mu_i^+}{\alpha_i})(v^+ - \alpha)_i + \sum_{i=1}^n (d_i - \frac{\mu_i^-}{\beta_i})(v^- - \beta)_i. \end{aligned} \tag{A.1}$$

Now, we bound each term one by one. For the first term, we note that

$$\frac{\mu_\theta}{\gamma}(\gamma - v_\theta) \leq \mu_\theta \leq 2 t_{\text{end}}. \tag{A.2}$$

For the second term, we have the following

Lemma 13. *Under the assumptions in Lemma 11, we have that*

$$\sum_{i=1}^n (c_i - \frac{\mu_i^+}{\alpha_i})(v^+ - \alpha)_i \leq 140nLR + 2n t_{\text{end}} - \frac{1}{2} \frac{t_{\text{end}} r}{\min_i \alpha_i}.$$

Proof. Note that

$$\begin{aligned} \sum_{i=1}^n (c_i - \frac{\mu_i^+}{\alpha_i})(v^+ - \alpha)_i &= \sum_{i=1}^n (c_i v_i^+ - \frac{\mu_i^+}{\alpha_i} v_i^+ - c_i \alpha_i + \mu_i^+) \\ &\leq \sum_{i=1}^n c_i v_i^+ + \sum_{i=1}^n \mu_i^+ - \sum_{i=1}^n \frac{\mu_i^+}{\alpha_i} v_i^+ \\ &\leq \|c\|_\infty \|v^+\|_1 + 2n t_{\text{end}} - \frac{1}{2} \sum_{i=1}^n \frac{t_{\text{end}} r}{\alpha_i} \end{aligned}$$

where we used $\mu_i^+ \in [\frac{t_{\text{end}}}{2}, 2t_{\text{end}}]$ and $v_i^+ \geq v_i \geq r$ at the end. The result follows from $\|c\|_\infty \leq L$, $\|v^+\|_1 \leq \Lambda \leq 140nR$ (Lemma 11). □

Lemma 14. *Under the assumptions in Lemma 11, we have that*

$$\sum_{i=1}^n (d_i - \frac{\mu_i^-}{\beta_i})(v^- - \beta)_i \leq 2 t_{\text{end}} - \frac{t}{16R} \max_i \beta_i.$$

Proof. Using $v^- = \min(\beta, \frac{32t_{\text{end}}}{t}R)$, we have $v_i^- \le \beta_i$. When $v_i^- < \beta_i$, we have $\beta_i \ge \frac{32t_{\text{end}}}{t}R$ and hence

$$d_i - \frac{\mu_i^-}{\beta_i} \ge d_i - \frac{\mu_i^-}{\frac{32t_{\text{end}}}{t}R} \ge d_i - \frac{2t_{\text{end}}}{\frac{32t_{\text{end}}}{t}R} = d_i - \frac{t}{16R} \ge \frac{t}{16R}.$$

Hence, we have

$$\sum_{i=1}^n (d_i - \frac{\mu_i^-}{\beta_i})(v^- - \beta)_i \le \frac{t}{16R}\sum_{i=1}^n (v^- - \beta)_i \le \frac{t}{16R}(\frac{32t_{\text{end}}}{t}R - \max_i \beta_i).$$

$\square$

Combining (A.1), (A.2), Lemma 13 and Lemma 14, we have

$$\begin{aligned}
0 &\le \frac{\mu_\theta}{\gamma}(\gamma - v_\theta) + \sum_{i=1}^n (c_i - \frac{\mu_i^+}{\alpha_i})(v^+ - \alpha)_i + \sum_{i=1}^n (d_i - \frac{\mu_i^-}{\beta_i})(v^- - \beta)_i \\
&\le 2t_{\text{end}} + 140nLR + 2nt_{\text{end}} - \frac{1}{2}\frac{t_{\text{end}}r}{\min_i \alpha_i} + 2t_{\text{end}} - \frac{t}{16R}\max_i \beta_i \\
&= 6nt_{\text{end}} + 140nLR - \frac{1}{2}\frac{t_{\text{end}}r}{\min_i \alpha_i} - \frac{t}{16R}\max_i \beta_i.
\end{aligned}$$

Setting $t_{\text{end}} = LR$, we have

$$\frac{1}{2}\frac{LRr}{\min_i \alpha_i} + \frac{t}{16R}\max_i \beta_i \le 150nLR.$$

In particular, this shows the following:

Lemma 15. *Under the assumptions in Lemma 11, we have that* $\min_i \alpha_i \ge \frac{r}{300n}$ *and* $\max_i \beta_i \le \frac{5000nLR}{t} \cdot R$.

Now, we are ready to prove the main result of this section.

Theorem 16. *Assume the linear program has inner radius r, outer radius R and Lipschitz constant L. The algorithm* `L2Center` *outputs $(x, s) \in \mathcal{P}^\circ \times \mathcal{D}^\circ$ such that $\|xs - t\|_2 \le \frac{t}{4}$ with $t = LR$. Furthermore,* `L2Center` *takes $O(\sqrt{n}\log(nR/r))$ Newton steps (defined in (1.1)).*

Proof. By the definition of $x^{(1)} = (\alpha, \beta, \gamma)$ and $s^{(1)} = (s_+^{(1)}, s_-^{(1)}, s_\theta^{(1)})$, we have that $x^{(1)}s^{(1)} = \mu$ with $\|\mu - t_{\text{end}}\|_2 \le \frac{t_{\text{end}}}{6}$ and

$$\begin{aligned}
\mathbf{A}(\alpha - \beta) &= b, \\
\sum_{i=1}^n \alpha_i + \theta &= \Lambda, \\
\mathbf{A}^\top y + \lambda + s_+^{(1)} &= c, \\
\mathbf{A}^\top y + s_-^{(1)} &= d, \\
\lambda + s_\theta^{(1)} &= 0.
\end{aligned}$$

By the choice of $t = 2^{30}n^3LR \cdot \frac{R}{r}$, Lemma 15 shows that

$$\max_i \beta_i \le \frac{r}{60000n^2} \le \frac{\min_i \alpha_i}{200n}.$$

Hence, we have $x^{(2)} = \alpha - \beta = (1 \pm \frac{1}{200n})\alpha > 0$ and that $\mathbf{A}x^{(2)} = b$.

Now, we prove that $s^{(2)} = s_+^{(1)} + \lambda$ is close to $s_+^{(1)}$. Since $x^{(2)} \in \mathcal{P}$, we have $x^{(2)} \le R$ and $\alpha \le \frac{5}{4}x^{(2)} \le \frac{5}{4}R$. Since $\alpha s_+^{(1)} = \mu_+ \ge \frac{5t_{\text{end}}}{6}$, we have

$$s_+^{(1)} \ge \frac{2t_{\text{end}}}{3R}.$$

On the other hand, we have $\theta = \Lambda - \sum_{i=1}^n \alpha_i \ge \Lambda - 2nR \ge 40nR$ (Lemma 11). Hence,

$$\lambda \le \frac{t_{\text{end}}}{20nR}.$$

Hence, we have $s^{(2)} = s_+^{(1)} - s_\theta^{(1)} = s_+^{(1)} + \lambda = (1 \pm \frac{3}{40n})s_+^{(1)} > 0$ and that $\mathbf{A}^\top y + x^{(2)} = c$.

Hence, we have $(x^{(2)}, s^{(2)}) \in \mathcal{P}^\circ \times \mathcal{D}^\circ$. Finally, we note that $\alpha s_+^{(1)} = \mu_+$ with $\|\mu_+ - t_{\text{end}}\|_2 \leq \frac{t_{\text{end}}}{6}$. Together with $x^{(2)} = (1 \pm \frac{1}{200n})\alpha$ and $s^{(2)} = (1 \pm \frac{3}{40n})s_+^{(1)}$ proved above, we have $\|x^{(2)}s^{(2)} - t_{\text{end}}\|_2 \leq \frac{t_{\text{end}}}{4}$. $\square$

Finally, we note that the modified linear program, the linear system is

$$\overline{\mathbf{A}} = \begin{bmatrix} \mathbf{A} & -\mathbf{A} & 0_d \\ 1_n^\top & 0_n & 1 \end{bmatrix}$$

For any diagonal matrices $\mathbf{W}_1, \mathbf{W}_2$ and any scalar α, we have

$$\mathbf{H} \stackrel{\text{def}}{=} \overline{\mathbf{A}} \begin{bmatrix} \mathbf{W}_1 & \mathbf{0} & 0_n \\ \mathbf{0} & \mathbf{W}_2 & 0_n \\ 0_n^\top & 0_n^\top & \alpha \end{bmatrix} \overline{\mathbf{A}}^\top = \begin{bmatrix} \mathbf{A}^\top(\mathbf{W}_1 + \mathbf{W}_2)\mathbf{A} & \mathbf{A}\mathbf{W}_1 1_n \\ (\mathbf{A}\mathbf{W}_1 1_n)^\top & 1_n^\top \mathbf{W}_1 1_n + \alpha \end{bmatrix}.$$

Note that the second row and column block has size 1. By block inverse formula, $\mathbf{H}^{-1}v$ is an explicit formula involving $(\mathbf{A}^\top(\mathbf{W}_1 + \mathbf{W}_2)\mathbf{A})^{-1}v_{1:n}$ and $(\mathbf{A}^\top(\mathbf{W}_1 + \mathbf{W}_2)\mathbf{A})^{-1}\mathbf{A}\mathbf{W}_1 1_n$. Hence, we can compute $\mathbf{H}^{-1}v$ by solving two linear systems of the form $\mathbf{A}^\top\mathbf{W}\mathbf{A}$ and some extra linear work.