Tutorial on Robust Interior Point

May 15, 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 regions for the two 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

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

□

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 $\mathbf{A}^\top y + s = c$:

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

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. □

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. □

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 small 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`$(\mathbf{A}, x, s, t_{\text{start}}, t_{\text{end}})$

Define $\mathbf{P}_{x,s} = \mathbf{S}^{-1}\mathbf{A}^\top(\mathbf{A}\mathbf{S}^{-1}\mathbf{X}\mathbf{A}^\top)^{-1}\mathbf{A}\mathbf{X}$.
Invariant: $(x,s) \in \mathcal{P}^\circ \times \mathcal{D}^\circ$ and $\|xs - t\|_2 \leq \frac{t}{4}$.
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 $t' = \max((1-h)t, t_{\text{end}})$
    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 $\|xs - t_{\text{start}}\|_2 \leq \frac{t_{\text{start}}}{4}$. Then the algorithm* `L2Step` *(Algorithm 1) outputs (x, s) such that $\|xs - t_{\text{end}}\|_2 \leq \frac{t_{\text{end}}}{6}$. 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 $\|xs - t\|_2 \leq \frac{t}{4}$. 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

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

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

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

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

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

Hence, we have $\|x's' - t'\|_2 \leq \|\delta_\mu/\mu\|_\mu^2 \leq 0.15t \leq t'/6$. 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 $O(\sqrt{n}\log(t_{\text{start}}/t_{\text{end}}))$. $\square$

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 $\min_{\mathbf{A}x=b,x\geq 0} c^\top x$:

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

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 $x^+, x^- \geq 0$. 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 19, 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

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

□

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].

2 Robust Interior Point Method

To improve the interior point method, one can either improve the number of steps $\widetilde{O}(\sqrt{n})$ or the cost per step. In this note, we focus on the later question. Recall from (1.1) that the linear system we solve in each step is of the form

$$
\begin{aligned}
\mathbf{S}\delta_x + \mathbf{X}\delta_s &= \delta_\mu, \\
\mathbf{A}\delta_x &= 0, \\
\mathbf{A}^\top\delta_y + \delta_s &= 0. \qquad (2.1)
\end{aligned}
$$

In each step, x, s and δ_μ in the equation above changes relatively by a vector with bounded ℓ_2 norm. So, only few coordinates change a lot in each step. To take advantage of this, the robust interior point method contains two new components: 1) Analyze the convergence when we only solve the linear system approximately (Section 2.1). 2) Show how to maintain the solution throughout the iteration (Section 2.2).

2.1 Staying within small ℓ_∞ distance

To make the interior point method robust against the noise in x and s, we need the interior point method works under a larger neighborhood than $\|xs - t\|_2 \leq \frac{t}{4}$. One natural choice of distance and potential would be $\|xs - t\|_q^q$. However, the step δ_μ that minimizes $\|\mu + \delta_\mu - t\|_q^q$ involves many cases. Instead, we pick the potential

$$\Phi(v) = \sum_{i=1}^{n} \cosh(\lambda v_i). \tag{2.2}$$

for some scalar $\lambda > 0$ where $\cosh(v) = (e^v + e^{-v})/2$. This potential induces the following algorithm:

Algorithm 3: `RobustStep`$(\mathbf{A}, x, s, t_{\text{start}}, t_{\text{end}})$

Define $r = (xs - t)/t$ and Φ according to (2.2) with $\lambda = 16 \log 40n$.
Invariant: $(x, s) \in \mathcal{P}^\circ \times \mathcal{D}^\circ$ and $\Phi(r) \leq 16n$.
Let $t = t_{\text{start}}$, $h = 1/(48\lambda\sqrt{n})$ and n is the number of columns in $\mathbf{A}$.
while $t \neq t_{\text{end}}$ *and* $\Phi(r) \geq 8n$ **do**
  Pick $\overline{x}$, $\overline{s}$ and $\overline{r}$ such that $\|\frac{\overline{x}-x}{x}\|_\infty \leq \frac{1}{48}$, $\|\frac{\overline{s}-s}{s}\|_\infty \leq \frac{1}{48}$ and $\|\overline{r} - r\|_\infty \leq \frac{1}{48\lambda}$.
  Let $t' = \max((1-h)t, t_{\text{end}})$, $\overline{\delta}_\mu = t' - t - \frac{t'}{32\lambda} \frac{\overline{g}}{\|\overline{g}\|_2}$, $\overline{g} = \nabla\Phi(\overline{r})$.
  Find δ_x, δ_s such that

$$\begin{aligned} \overline{\mathbf{S}}\delta_x + \overline{\mathbf{X}}\delta_s &= \overline{\delta}_\mu, \\ \mathbf{A}\delta_x &= 0, \\ \mathbf{A}^\top \delta_y + \delta_s &= 0. \end{aligned} \tag{2.3}$$

  Set $x \leftarrow x + \delta_x$, $s \leftarrow s + \delta_s$ and $t \leftarrow t'$.
end
Return (x, s)

Here, we prove the key facts we use about Φ:

Lemma 11. *Define Φ according to (2.2). For any $v \in \mathbb{R}^n$, we have that $\|v\|_\infty \leq \frac{\log 2\Phi(v)}{\lambda}$ and $\|\nabla\Phi(v)\|_2 \geq \frac{\lambda}{\sqrt{n}}(\Phi(v) - n)$. Furthermore, if $\Phi(v) \geq 4n$ and $\|\delta\|_\infty \leq \frac{1}{5\lambda}$, we have*

$$\|\nabla\Phi(v+\delta) - \nabla\Phi(v)\|_2 \leq \frac{1}{3}\|\nabla\Phi(v)\|_2.$$

Proof. For the first result: we have $\Phi(v) \geq \frac{1}{2}\min_i e^{\lambda|v_i|}$ and hence $\|v\|_\infty \leq \frac{\log 2\Phi(v)}{\lambda}$.

For the second result: we have

$$\begin{aligned} \|\nabla\Phi(v)\|_2 &= \lambda\sqrt{\sum_{i=1}^{n}\sinh^2(\lambda v_i)} = \lambda\sqrt{\sum_{i=1}^{n}(\cosh^2(\lambda v_i) - 1)} \\ &\geq \frac{\lambda}{\sqrt{n}}\sum_{i=1}^{n}\sqrt{\cosh^2(\lambda v_i) - 1} \geq \frac{\lambda}{\sqrt{n}}\sum_{i=1}^{n}(\cosh(\lambda v_i) - 1) \\ &= \frac{\lambda}{\sqrt{n}}(\Phi(v) - n). \end{aligned}$$

For the last result, using $\sinh(v+\delta) = \sinh v\cosh\delta + \cosh v\sinh\delta$ and $|\cosh v - \sinh v| \leq 1$, for $|\delta| \leq \frac{1}{5}$, we have

$$\begin{aligned} |\sinh(v+\delta) - \sinh(v)| &\leq |\cosh\delta - 1| \cdot |\sinh v| + |\sinh\delta| \cdot \cosh v \\ &\leq (|\cosh\delta - 1| + |\sinh\delta|) \cdot |\sinh v| + |\sinh\delta| \\ &\leq \frac{1}{4}|\sinh v| + \frac{1}{4}. \end{aligned}$$

Using that $\nabla\Phi(v) = \sum_{i=1}^{n}\lambda\sinh(\lambda v_i)$, for $\|\delta\|_\infty \leq \frac{1}{5\lambda}$, we have

$$\|\nabla\Phi(v+\delta) - \nabla\Phi(v)\|_2 \leq \frac{1}{4}\|\nabla\Phi(v)\|_2 + \frac{\sqrt{n}\lambda}{4}. \tag{2.4}$$

Since $\Phi(v) \geq 4n$, we have that $\|\nabla\Phi(v)\|_2 \geq 3\sqrt{n}\lambda$ and hence (2.4) shows that

$$\|\nabla\Phi(v+\delta) - \nabla\Phi(v)\|_2 \leq (\frac{1}{4} + \frac{1}{12})\|\nabla\Phi(v)\|_2 = \frac{1}{3}\|\nabla\Phi(v)\|_2.$$

$\square$

We collect some basic bounds on the step in the following Lemma:

Lemma 12. *Using the notation in* `RobustStep` *(Algorithm 3). Under the invariant $\Phi((xs-t)/t) \leq 16n$, we have $\|xs - t\|_\infty \leq \frac{t}{16}$, $\|\delta_x/x\|_2 \leq \frac{1}{16\lambda}$, $\|\delta_s/s\|_2 \leq \frac{1}{16\lambda}$, and $\|r' - r\|_\infty \leq \frac{1}{6\lambda}$ where $r' = ((x+\delta_x)(s+\delta_s) - t')/t'$.*

Proof. Using $\Phi((xs-t)/t) \leq 16n$ and Lemma 11, we have

$$\|xs - t\|_\infty \leq \frac{t \log 32n}{\lambda} \leq \frac{t}{16}$$

By Lemma 4, we have

$$\mathbf{X}^{-1}\delta_x = (\mathbf{I} - \mathbf{P})(\overline{\delta}_\mu/\overline{\mu})$$

where $\overline{\mu} = \overline{xs}$ and $\mathbf{P} = \overline{\mathbf{S}}^{-1}\mathbf{A}^\top(\mathbf{A}\overline{\mathbf{S}}^{-1}\overline{\mathbf{X}}\mathbf{A}^\top)^{-1}\mathbf{A}\overline{\mathbf{X}}$. By Lemma 5, we have

$$\|\delta_x/x\|_{\overline{\mu}} = \|(\mathbf{I} - \mathbf{P})v\|_{\overline{\mu}} \leq \|\overline{\delta}_\mu/\overline{\mu}\|_{\overline{\mu}}.$$

Using that $\|xs - t\|_\infty \leq \frac{t}{16}$, $\|\frac{x-\overline{x}}{x}\|_\infty \leq \frac{1}{48}$, $\|\frac{s-\overline{s}}{s}\|_\infty \leq \frac{1}{48}$, we have $\overline{\mu} \geq \frac{10}{11}t$ and hence

$$\|\delta_x/x\|_2 \leq \sqrt{\frac{11}{10t}}\|\delta_x/x\|_{\overline{\mu}} \leq \sqrt{\frac{11}{10t}}\|\overline{\delta}_\mu\|_{\overline{\mu}^{-1}} \leq \frac{11}{10t}\|\overline{\delta}_\mu\|_2$$

Using the formula $\overline{\delta}_\mu = t' - t - \frac{t'}{32\lambda}\frac{\overline{g}}{\|\overline{g}\|_2}$ and $|\frac{t'-t}{t}| \leq \frac{1}{48\lambda\sqrt{n}}$, we have

$$\|\delta_x/x\|_2 \leq \frac{11}{10}\left(\|\frac{t'-t}{t}\|_2 + \frac{t'}{32\lambda t}\right) \leq \frac{1}{16\lambda}.$$

Same proof gives $\|\delta_s/s\|_2 \leq \frac{1}{16\lambda}$.

Finally, we have

$$\begin{aligned} r' - r &= \frac{(x+\delta_x)(s+\delta_s)}{t'} - \frac{xs}{t} \\ &= (\frac{t}{t'} - 1)\frac{xs}{t} + \frac{xs}{t'}(\delta_x/x + \delta_s/s + (\delta_x/x)(\delta_s/s)). \end{aligned}$$

Using $|\frac{t'-t}{t}| \leq \frac{1}{48\lambda\sqrt{n}}$,$\|xs - t\|_\infty \leq \frac{t}{16}$, $\|\delta_x/x\|_2 \leq \frac{1}{16\lambda}$, $\|\delta_s/s\|_2 \leq \frac{1}{16\lambda}$, we have

$$\|r' - r\|_\infty \leq \frac{1}{48\lambda}\frac{17}{16} + \frac{17}{16}\frac{49}{48}(\frac{1}{16\lambda} + \frac{1}{16\lambda} + \frac{1}{16\lambda}\frac{1}{16\lambda}) \leq \frac{1}{6\lambda}.$$

$\square$

Now, we are already to prove the main theorem.

Theorem 13. *Suppose that the input satisfies $(x,s) \in \mathcal{P}^\circ \times \mathcal{D}^\circ$ and $\Phi((xs - t_{\text{start}})/t_{\text{start}}) \leq 16n$. Then the algorithm* `RobustStep` *(Algorithm 3) outputs x and s such that $\Phi((xs - t_{\text{end}})/t_{\text{end}}) \leq 8n$. Furthermore,* `RobustStep` *takes $O(\sqrt{n}\log n \log(t_{\text{start}}/t_{\text{end}}))$ Newton steps (defined in (1.1)).*

Proof. First, we prove the invariant $\Phi(r) \leq 16n$ by induction. It holds for the first step by the assumption. Let $x' = x + \delta_x$, $s' = s + \delta_s$ and t' defined in the algorithm and $r' = \frac{x's' - t'}{t'}$. Our goal is to prove $\Phi(r') \leq 16n$.

If $\Phi(r) \leq 8n$, we simply use the fact that $\|r - r'\|_\infty \leq \frac{1}{16\lambda}$ (Lemma 12) and hence $\Phi(r') \leq 2\Phi(r) \leq 16n$. Hence, we only need to prove for the case $\Phi(r) \geq 8n$. By the definition of δ_x and δ_s, we have

$$\overline{\mathbf{S}}\delta_x + \overline{\mathbf{X}}\delta_s = \overline{\delta}_\mu = t' - t - \frac{t'}{32\lambda}\frac{\overline{g}}{\|\overline{g}\|_2}$$

and hence

$$\begin{aligned}\frac{x's'-t'}{t'} &= \frac{(x+\delta_x)(s+\delta_s)-t'}{t'}\\ &= \frac{xs+\mathbf{S}\delta_x+\mathbf{X}\delta_s+\delta_x\delta_s-t'}{t'}\\ &= \frac{xs-t'+\overline{\mathbf{S}}\delta_x+\overline{\mathbf{X}}\delta_s+(s-\overline{s})\delta_x+(x-\overline{x})\delta_s+\delta_x\delta_s}{t'}\\ &= \frac{xs-t}{t}-\frac{1}{32\lambda}\cdot\frac{\overline{g}}{\|\overline{g}\|_2}+\eta \end{aligned} \tag{2.5}$$

where the error term

$$\eta=(\frac{t}{t'}-1)\frac{xs-t}{t}+\frac{(s-\overline{s})\delta_x+(x-\overline{x})\delta_s+\delta_x\delta_s}{t'}.$$

Now, we bound the error term η. Using Lemma 12 ($\|\delta_x/x\|_2\leq\frac{1}{16\lambda}$, $\|\delta_s/s\|_2\leq\frac{1}{16\lambda}$, $\|xs-t\|_\infty\leq\frac{t}{16}$) and the definition of the algorithm ($\lambda\geq 16$, $|t-t'|\leq\frac{t'}{48\lambda\sqrt{n}}$, $\|\frac{x-\overline{x}}{x}\|_\infty\leq\frac{1}{48}$, $\|\frac{s-\overline{s}}{s}\|_\infty\leq\frac{1}{48}$), we have

$$\begin{aligned}\|\eta\|_2 \leq &|\frac{t}{t'}-1|\|\frac{xs-t}{t}\|_\infty\sqrt{n}+\|\frac{xs}{t'}\|_\infty\|\frac{s-\overline{s}}{s}\|_\infty\|\frac{\delta_x}{x}\|_2\\ &+\|\frac{xs}{t'}\|_\infty\|\frac{x-\overline{x}}{x}\|_\infty\|\frac{\delta_s}{s}\|_2+\|\frac{xs}{t'}\|_\infty\|\frac{\delta_x}{x}\|_2\|\frac{\delta_s}{s}\|_2\\ \leq &\frac{1}{48\lambda}\frac{1}{16}+\frac{9}{8}\frac{1}{48}(\frac{1}{16\lambda}+\frac{1}{16\lambda})+\frac{9}{8}(\frac{1}{16\lambda}\frac{1}{16\lambda})\leq\frac{1}{200\lambda}.\end{aligned} \tag{2.6}$$

Mean value theorem shows there is $\widetilde{r}$ between r and r' such that

$$\begin{aligned}\Phi(r') &= \Phi(r)+\langle\nabla\Phi(\widetilde{r}),r'-r\rangle\\ &= \Phi(r)+\left\langle\nabla\Phi(\widetilde{r}),-\frac{1}{32\lambda}\frac{\overline{g}}{\|\overline{g}\|_2}+\eta\right\rangle\end{aligned}$$

where we used (2.5) at the end. Using Lemma 12, we have $\|r-r'\|_\infty\leq\frac{1}{6\lambda}$ and $\|\overline{r}-r\|_\infty\leq\frac{1}{48\lambda}$. Hence, we have $\|\overline{r}-\widetilde{r}\|_\infty\leq\frac{1}{5\lambda}$. Since $\Phi(r)\geq 8n$, we have $\Phi(\overline{r})\geq 4n$ and hence Lemma 11 shows that

$$\|\nabla\Phi(\widetilde{r})-\nabla\Phi(\overline{r})\|_2\leq\frac{1}{3}\|\nabla\Phi(\overline{r})\|_2.$$

Using $\overline{g}=\nabla\Phi(\overline{r})$ and letting $\eta_2=\nabla\Phi(\widetilde{r})-\nabla\Phi(\overline{r})$, we have

$$\begin{aligned}\Phi(r')-\Phi(r) &= \left\langle\overline{g}+\eta_2,-\frac{1}{32\lambda}\frac{\overline{g}}{\|\overline{g}\|_2}+\eta\right\rangle\\ &= -\frac{1}{32\lambda}\|\overline{g}\|_2-\frac{1}{32\lambda}\eta_2^\top\frac{\overline{g}}{\|\overline{g}\|_2}+\overline{g}^\top\eta+\eta_2^\top\eta.\end{aligned}$$

Using $\|\eta_2\|_2\leq\frac{1}{3}\|\overline{g}\|_2$ and $\|\eta\|_2\leq\frac{1}{200\lambda}$ (2.6), we have

$$\begin{aligned}\Phi(r')-\Phi(r) &\leq -\frac{1}{32\lambda}\|\overline{g}\|_2+\frac{1}{32\lambda}\cdot\frac{1}{3}\|\overline{g}\|_2+\|\overline{g}\|_2\cdot\frac{1}{200\lambda}+\frac{1}{3}\|\overline{g}\|_2\cdot\frac{1}{200\lambda}\\ &\leq -\frac{1}{100\lambda}\|\overline{g}\|_2\end{aligned}$$

Using Lemma 11, we have $\|\overline{g}\|_2\geq\frac{\lambda}{\sqrt{n}}(\Phi(\overline{r})-n)\geq 3\lambda\sqrt{n}$. Hence, we have

$$\Phi(r')\leq\Phi(r)-\frac{\sqrt{n}}{40}\leq 16n. \tag{2.7}$$

So, in both cases, we have that $\Phi(r')\leq 16n$ and this finishes the induction.

Since `RobustStep` decreases t by $1-\Omega(n^{-1/2}/\lambda)$ factor each step and since we start with $t=t_{\text{start}}$ and ends with $t\approx t_{\text{end}}$, it takes $O(\sqrt{n}\log n\cdot\log(t_{\text{start}}/t_{\text{end}}))$ iterations for t moves to t_{end}. Finally, if $\Phi(r)\geq 8n$, (2.7) shows that Φ decreases by at least $\frac{\sqrt{n}}{40}$ per steps. Hence, it takes at most $O(\sqrt{n})$ extra iterations to decrease Φ back from $16n$ to $8n$. □

2.2 Inverse Maintenance

....

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 4 for the precise formulation of the modified linear program. The formulation is chosen such that we have an explicit interior point.

Algorithm 4: `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 14. *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 4). 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 15. *Under the assumptions in Lemma 14, 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 14), $v \leq R$ and $v^- \leq \frac{32 t_{\text{end}}}{t} R \leq R$ to get

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

$\square$

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

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

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

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

For the second term, we have the following

Lemma 16. *Under the assumptions in Lemma 14, 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 14). $\square$

Lemma 17. *Under the assumptions in Lemma 14, we have that*

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

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

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

Hence, we have

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

$\square$

Combining (A.1), (A.2), Lemma 16 and Lemma 17, we have

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

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

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

In particular, this shows the following:

Lemma 18. *Under the assumptions in Lemma 14, we have that* $\min_i \alpha_i \geq \frac{r}{300n}$ *and* $\max_i \beta_i \leq \frac{5000nLR}{t} \cdot R$.

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

Theorem 19. *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 $\|xs - t\|_2 \leq \frac{t}{4}$ 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 $\|\mu - t_{\text{end}}\|_2 \leq \frac{t_{\text{end}}}{6}$ 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 18 shows that

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

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 $x^{(2)} \leq R$ and $\alpha \leq \frac{5}{4}x^{(2)} \leq \frac{5}{4}R$. Since $\alpha s_+^{(1)} = \mu_+ \geq \frac{5t_{\text{end}}}{6}$, we have

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

On the other hand, we have $\theta = \Lambda - \sum_{i=1}^n \alpha_i \geq \Lambda - 2nR \geq 40nR$ (Lemma 14). Hence,

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

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.