CPSC 536C: Algorithms for Convex Optimization
University of British Columbia

Interior Point Methods

Akshay Ram
March 23, 2026

These notes are subject to change and may contain errors.

1 Introduction

In this note we will cover Interior Point Methods, which are currently among the fastest procedures to solve linear programs and general convex programs to high precision. Formally, they achieve $\log(1/\varepsilon)$ convergence, just like the Ellipsoid method, and each iteration can be implemented reasonably fast, though not as fast as first-order methods.

Our main illustrative example will be the following standard LP formulation:

$$\min\langle c, x\rangle \qquad \text{s.t.} \qquad Ax \geq b, \tag{1}$$

where $x \in \mathbb{R}^n, A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$. In order to solve this problem, we will solve a sequence of regularized problems of the form

$$\min_x f_\gamma(x) := \gamma\langle c, x\rangle + \phi(x).$$

Here ϕ is a so-called *barrier function*, which incorporates the constraint set $Ax \geq b$ into the objective (somewhat similar to the Lagrangian formulation). As $\gamma \to \infty$ we are increasing the importance of our original LP objective $\langle c, x\rangle$. Further, ϕ enjoys a property known as *self-concordance*, which allows us to use the *Newton method* to update our iterate to make very fast progress for f_γ and simultaneously increase $\gamma \to \infty$. This leads to $\log(1/\varepsilon)$ convergence for our original LP.

2 Outline of Interior Point Method

Informally, ϕ is a *barrier* for some domain D if $\phi \to \infty$ as $x \to \partial D$. We will construct a barrier for our feasible region with special properties for our interior point method.

Definition 1 (Central Path)**.** *For the standard LP given in eq.* (1) *and barrier function ϕ for the feasible region, the* central path $\{x_\gamma\}_{\gamma\geq 0}$ *is the sequence of solutions*

$$x_\gamma := \arg\min_x f_\gamma(x) = \gamma\langle c, x\rangle + \phi(x).$$

The barrier property of ϕ implies $f_\gamma(x) < \infty$ iff $x \in K$ is feasible. Also as $\gamma \to \infty$ the central path x_γ converges to the optimizer $x^* = \arg\min_{Ax\geq b}\langle c, x\rangle$.

To initiate the path following method, we require a strictly feasible x that is sufficiently close to the central path. This is discussed in section 6. In the following section, we show how the

Newton method gives an update scheme that allows us to improve our estimate for each f_γ very quickly if we are close enough to the central path. Then in the following section we study how to increase the value of γ while remaining close enough to the central path to apply the Newton method. Finally, in the last sections we discuss how to compute a good starting point to begin the path-following method; Finally, in section 5 we give a stopping criterion for the method which guarantees approximate optimality for our original LP eq. (1).

3 Newton Method

The Newton step is a natural descent step that attempts to correct for the slow convergence of gradient descent by incorporating higher-order information. Recall that we can reformulate gradient descent as

$$x - \eta \nabla f(x) = \arg\min_z f(x) + \langle \nabla f(x), z - x \rangle + \frac{1}{2\eta} \|z - x\|_2^2,$$

i.e. the update finds the optimal solution for a proxy for f which combines the linear approximation $f(x) + \langle \nabla f(x), z - x \rangle$ with a quadratic regularization term $\frac{1}{2\eta}\|z - x\|_2^2$. If the function f exactly matches the proxy, then gradient descent converges in a single step. On the other hand, even for slightly more complicated quadratic functions of the form $f(x) = \|Ax - b\|_2^2$, gradient descent could have very slow convergence. This is because the quadratic term A^*A could be very *ill-conditioned*, so the steepest local direction does not match the contours of f.

To fix this, we could try to incorporate Hessian information. This is naturally achieved by the Newton step:

Definition 2 (Newton Step). *For strictly convex $\phi : D \to \mathbb{R}$ and $x \in D$, the Newton step is*

$$x_+ := x + v_x := x - (\nabla^2 \phi(x))^{-1} \nabla \phi(x).$$

To motivative this update, we consider the following family of norms:

Definition 3. *For strictly convex $\phi : D \to \mathbb{R}$, the* local norms *at $x \in D$ are*

$$\|v\|_x^2 := \langle v, \nabla^2 \phi(x) v \rangle, \qquad \|u\|_{x,*}^2 := \langle u, (\nabla^2 \phi(x))^{-1} u \rangle.$$

Further, the Newton decrement is the local norm of the Newton step

$$\delta_\phi(x) := \|v_x\|_x = \|\nabla \phi(x)\|_{x,*}.$$

Proposition 4. *The following are equivalent characterizations of the Newton step and decrement*

1. $x + v_x := \arg\min_{z \in D} \langle \nabla \phi(x), z - x \rangle + \frac{1}{2}\langle (z - x), \nabla^2 \phi(x)(z - x) \rangle$;
2. $\langle \nabla \phi(x), v_x \rangle + \frac{1}{2}\langle v_x, \nabla^2 \phi(x) v_x \rangle = \frac{-\delta_\phi(x)^2}{2}$;
3. $\nabla \phi(x) + \nabla^2 \phi(x)(x_+ - x) = 0$;

Proof:

1. By strict convexity we have $\nabla^2 \phi(x) \succ 0$, so we can compute the optimizer by critical equation

$$\nabla \phi(x) + \nabla^2 \phi(x) v = 0 \iff v = -(\nabla^2 \phi(x))^{-1} \nabla \phi(x).$$

2. This follows by plugging in the Newton step and using the definition of Newton decrement:

$$\langle \nabla\phi(x), v_x\rangle + \frac{1}{2}\langle v_x, \nabla^2\phi(x)v_x\rangle = \langle \nabla\phi(x), (\nabla^2\phi(x))^{-1}\nabla\phi(x)\rangle(-1+1/2) = \frac{-\delta_\phi(x)^2}{2}.$$

3. This also follows by direct computation:

$$\nabla\phi(x) + \nabla^2\phi(x)(x_+ - x) = \nabla\phi(x) - \nabla^2\phi(x)(\nabla^2\phi(x))^{-1}\nabla\phi(x) = 0.$$

□

The above characterizations show that the Newton step is indeed a generalization of steepest descent, where the update step is taken step with respect to the local norm. The third expression is the first-order Taylor expression for the gradient, so this also motivates the Newton step as a natural update that attempts to solve the critical equation using its linear approximation.

For gradient and mirror descent, we used smoothness and strong convexity, which were upper and lower bounds on the Hessian, in order to prove convergence. Intuitively, smoothness bounds the Lipschitz constant of the gradient, so this says gradient descent works well if the gradient does not change too fast. In order to analyze the Newton step, which is a second-order method, we make an analogous assumption:

Definition 5 (Self-concordance)**.** *Let $\phi : D \to \mathbb{R}$ be a strictly convex function with domain D. Then ϕ is self-concordant if*

$$|D^3\phi(x)[u, v, w]| \leq 2\|u\|_x\|v\|_x\|w\|_x.$$

Remark 6. *The above definition is actually equivalent to the following simpler form:*

$$|\partial^3_{t=0}\phi(x + tv)| \leq 2\|v\|_x^{3/2}.$$

The proof requires some non-trivial facts about symmetric tensors, so we omit the proof and assume this fact for these notes.

The constant 2 is just a choice for convenience: it can be verified that $-\log(t)$ is self-concordant. We could also define α-self-concordance of ϕ as

$$|D^3\phi(x)[u, v, w]| \leq \frac{2}{\sqrt{\alpha}}\|u\|_x\|v\|_x\|w\|_x,$$

but notice that ϕ is self-concordant iff $\alpha\phi$ is α-self-concordant.

On the other hand, scaling inside the function preserves self-concordance. In fact the above definition is *affine invariant*, i.e. $\phi(x)$ is self-concordant iff $\phi(Ax)$ is self-concordant for any linear transform A. The proof of both these facts is left as an exercise.

The above definition can be used to bound the change in second-order information, just as smoothness bounds the change in first-order information.

Lemma 7. *For self-concordant $\phi : D \to \mathbb{R}$, consider $x, y \in D$ with $r := \|y - x\|_x < 1$. Then*

$$(1-r)\nabla^2\phi(x) \preceq \nabla^2\phi(y) \preceq \frac{\nabla^2\phi(x)}{1-r}.$$

We will prove this fundamental lemma using a functional inequality in the following subsections. Using this, we are able to prove convergence for the Newton Method:

Theorem 8. *Let ϕ be self-concordant and $x \in D$ with $\delta_\phi(x) < 1$. Then Newton update x_+ satisfies*

$$\delta_\phi(x_+) \le \left(\frac{\delta_\phi(x)}{1-\delta_\phi(x)}\right)^2.$$

Our notion of progress is the Newton decrement, which can be thought of as the size of the gradient in the local norm. We note that the analysis requires initial condition $\delta_\phi(x) < 1$. It can be shown that the Newton step always makes some progress (see [NN]), but the expression and analysis are more complicated, and are not required for our analysis of Interior Point Methods, so we omit them. Note that the convergence is extremely fast: δ error becomes $O(\delta^2)$ error in the next step, so the number of bits of precision doubles in each step; this is known as *quadratic convergence*. This is the main engine for our interior point method, as we can leverage self-concordance of our barrier to compute updates along the central path quickly.

3.1 Intuitive Analysis

In this subsection, we give a slightly weaker convergence bound for the Newton step using methods from the analysis of gradient and mirror descent. Specifically, we will adapt analysis for strongly convex and smooth functions to the *ϕ-local norm*. In the following subsection, we formally prove theorem 11 using more refined methods.

We recall some standard facts from the analysis of gradient and mirror descent, which we state without proof.

Proposition 9. *Let ϕ be α-strongly convex and β-smooth wrt norm $\|\cdot\|$. Then*

1. *Steepest descent step wrt $\|\cdot\|$ makes progress*

$$\phi(x_+) - \phi(x^*) \le (1-\alpha/\beta)(\phi(x) - \phi(x^*)).$$

2. *Function gap, gradient, and distance to optimality can be bounded*

$$\|x^* - x\| \lesssim \frac{\|\nabla\phi(x)\|_*}{\alpha}, \qquad \frac{\|\nabla\phi(x)\|_*^2}{\beta} \lesssim \phi(x) - \phi(x^*) \lesssim \min\{\beta\|x^*-x\|^2, \|\nabla\phi(x)\|_*^2/\alpha\}.$$

Next we claim that theorem 7 implies, for all $\|y-x\|_x \le \delta$ (where δ is at most a small enough constant), that ϕ is $(1+O(\delta))$-smooth and $(1-O(\delta))$-strongly convex, both wrt $\|\cdot\|_x$. We prove this more formally in the following section.

The Newton step is exactly the steepest descent step wrt local norm $\|\cdot\|_x$, so we can use the above analysis to bound convergence

$$\phi(x_+) - \phi(x^*) \le \left(1 - \frac{1-O(\delta)}{1+O(\delta)}\right)(\phi(x) - \phi(x^*)) \lesssim \delta(\phi(x) - \phi(x^*)),$$

where in the last step we used Taylor approximation $(1+\delta)^{-1} \approx 1-\delta$ for small δ. Therefore, if δ is at most a small constant, we make constant factor progress in each step.

We can also use this to bound all other measures of optimality:

$$\|\nabla\phi(x_+)\|_{x,*}^2 \lesssim \beta(\phi(x_+) - \phi(x^*)) \lesssim (1+O(\delta)) \cdot \delta(\phi(x) - \phi(x^*)) \lesssim \delta \cdot \frac{\delta^2}{\alpha} \lesssim \delta^3 = \|\nabla\phi(x)\|_{x,*}^3,$$

where we used smoothness and strong convexity from theorem 9(2) along with the previous analysis to bound the function gap. Similarly we can bound

$$\|x_+ - x^*\|_x^2 \leq \frac{\phi(x_+) - \phi(x^*)}{\alpha} \lesssim \delta^2.$$

Finally we can note that theorem 7 gives local norm approximation $\|\nabla\phi(x_+)\|_{x,*} \approx \|\nabla\phi(x_+)\|_{x_+,*} = \delta_\phi(x_+)$, so the above analysis in fact shows strong convergence of the Newton decrement.

3.2 Formal Analysis

Our proof will crucially use the Hessian approximation theorem 7 which we restate and prove below.

Lemma 10. *For self-concordant $\phi : D \to \mathbb{R}$, consider $x, y \in D$ with $r := \|y - x\|_x < 1$. Then*

$$\forall v : \quad (1 - r)\|v\|_x \leq \|v\|_y \leq \frac{\|v\|_x}{1 - r}.$$

Letting $H_x := \nabla^2\phi(x)$ for shorthand, this can be rewritten

$$(1 - r)H_x \preceq H_y \preceq \frac{H_x}{1 - r}.$$

Proof: We first prove the special case when $v = y - x$. Let $x_t := x + t(y - x)$ and let $h(t) := \|y - x\|_{x_t}^2$. Then we have

$$|\partial_t h(t)| = |\partial_t\langle(y - x), \nabla^2\phi(x_t)(y - x)\rangle| = |D^3\phi(x_t)[y - x, y - x, y - x]| \leq 2\|y - x\|_{x_t}^{3/2} = 2h(t)^{3/2},$$

where the first two steps were by definition, in the third we used self-concordance, and the final step was by definition of h. Rearranging this gives a functional inequality

$$\left|\partial_t \frac{1}{\sqrt{h(t)}}\right| = \left|\frac{-h'(t)}{2h(t)^{3/2}}\right| \leq 1 \quad \Longrightarrow \left|\frac{1}{\sqrt{h(1)}} - \frac{1}{\sqrt{h(0)}}\right| = \left|\int_0^1 \partial_t \frac{1}{\sqrt{h(t)}}\right| \leq 1.$$

Now we can note that $\sqrt{h(0)} = \|y - x\|_x$, $\sqrt{h(1)} = \|y - x\|_y$ so this gives

$$\frac{1}{\|y - x\|_x} - 1 \leq \frac{1}{\|y - x\|_y} \leq \frac{1}{\|y - x\|_x} + 1 \iff \quad \frac{\|y - x\|_x}{1 - r} \geq \|y - x\|_y \geq \frac{\|y - x\|_x}{1 + r}, \qquad (2)$$

where we substituted $r = \|y - x\|_x$ in the last step.

Now for general v we similarly consider $g(t) := \|v\|_{x_t}^2 = \langle v, \nabla^2\phi(x_t)v\rangle$ and note

$$|\partial_t g(t)| = |D^3\phi(x_t)[y - x, v, v]| \leq 2\|y - x\|_{x_t}\|v\|_{x_t}^2 = 2\sqrt{h(t)}g(t),$$

where again the first step is by definition of D^3, in the second we apply self-concordance. Rearranging gives functional inequality $|\partial_t \log g(t)| = |g'(t)/g(t)| \leq 2\sqrt{h(t)}$. We can bound

$$\sqrt{h(t)} = \|y - x\|_{x_t} \leq \frac{\|y - x\|_x}{1 - tr} = \frac{r}{1 - tr},$$

where the justification is either by performing the same calculations as above, or by noting $y - x = t^{-1}(x_t - x)$ so we can directly apply the previous bound for local norms with $\|x_t - x\|_x = t\|y - x\|_x$. Therefore we can finally relate our original norms

$$\left|\log\frac{g(1)}{g(0)}\right| = \left|\int_0^1 \partial_t \log g(t)\right| \leq \int_0^1 \frac{2r}{1 - tr} = \int_0^1 \partial_t(-2\log(1 - tr)) = \log\frac{1}{(1 - r)^2},$$

where we used the fundamental theorem of calculus and plugged in the bound above for g'/g in the second step. Finally noting $g(0) = \|v\|_x^2, g(1) = \|v\|_y^2$ and rearranging gives the lemma. □

We can now use this to analyze convergence of the Newton step.

Theorem 11. *Let ϕ be self-concordant and $x \in D$ with $\delta_\phi(x) < 1$. Then Newton update x_+ satisfies*

$$\delta_\phi(x_+) \leq \left(\frac{\delta_\phi(x)}{1-\delta_\phi(x)}\right)^2.$$

Proof: Our goal is to bound $\delta_\phi(x_+) = \|v_{x_+}\|_{x_+} = \|\nabla\phi(x_+)\|_{x_+,*}$ by theorem 3. As a first step, we use theorem 10 for $r := \|x_+ - x\|_x$ to note $\|\nabla\phi(x_+)\|_{x_+,*} \leq \frac{\|\nabla\phi(x_+)\|_{x,*}}{1-r}$. Therefore for the remainder we focus on bounding the gradient $\nabla\phi(x_+)$ according to the x-local norm.

Let $x_t := x + t(x_+ - x)$ and $H_t := \nabla^2\phi(x_t)$ for shorthand, then we have

$$\nabla\phi(x_+) - \nabla\phi(x) = \int_0^1 H_t(x_+ - x) = -\int_0^1 H_t H^{-1}\nabla\phi(x),$$

where the first step was by fundamental theorem of calculus, and the second was by definition of v_x. For the remainder, we use $\nabla := \nabla\phi(x), \nabla_+ := \nabla\phi(x_+), H_x := \nabla^2\phi(x)$ for shorthand, Recall by theorem 4(3) that the update is chosen to cancel out the first order approximation of the gradient, $\nabla + H(x_+ - x) = 0$. So we can bound

$$\begin{aligned}\|\nabla_+\|_{x,*} = \|\nabla_+ - (\nabla + H(x_+ - x))\|_{x,*} &= \left\|\int_0^1 (H_t - H)(x_+ - x)\right\|_{x,*} \\ = \left\|\int_0^1 (H_t - H)H^{-1}\nabla\right\|_{x,*} &= \left\|\int_0^1 H^{-1/2}(H_t - H)H^{-1/2}H^{-1/2}\nabla\right\|_2,\end{aligned}$$

where in the first step we used theorem 4(3) so the added term is zero, for the second we used the previous calculation for $\nabla_+ - \nabla$, in the third we substituted in the definition of $x_+ - x = -H^{-1}\nabla$, and the final step was by expanding out the definition of $\|\cdot\|_{x,*}$ in terms of $\|\cdot\|_2$.

Next, we use the local norm relation between x, x_t which shows $(1-tr)^2 H \preceq H_t \preceq (1-tr)^{-2}H$. Therefore we can bound

$$(-r + \frac{r^2}{3})I = (\int_0^1 (1-tr)^2 - 1)I \preceq \int_0^1 H^{-1/2}(H_t - H)H^{-1/2} \preceq \left(\int_0^1 \frac{1}{(1-tr)^2} - 1\right)I = \frac{r}{1-r}I$$

$$\implies \left\|\int_0^1 H^{-1/2}(H_t - H)H^{-1/2}H^{-1/2}\nabla\|_2 \leq \right\|\int_0^1 H^{-1/2}(H_t - H)H^{-1/2}\right\|_{op}\|H^{-1/2}\nabla\|_2 \leq \frac{r}{1-r}\|\nabla\|_{x,*},$$

Combining this with the first step gives the result

$$\delta_\phi(x_+) = \|\nabla_+\|_{x_+} \leq \frac{\|\nabla_+\|_x}{1-r} \leq \frac{1}{1-r}\frac{r}{1-r}\|H^{-1/2}\nabla\|_2 = \left(\frac{\delta_\phi(x)}{1-\delta_\phi(x)}\right)^2,$$

where in the last step we noted $r = \|x_+ - x\|_x = \|v_x\|_x = \|\nabla\|_{x,*} = \delta_\phi(x)$. □

4 Path Following

What happens when we move the interpolation parameter? Say we have x close to optimal for $f_\gamma := \gamma\langle c, \cdot\rangle + \phi$ and we move the parameter $\gamma \to \gamma + \Delta$. Then the gradient with respect to the new function is

$$\nabla f_{\gamma+\Delta}(x) = \Delta c + \nabla f_\gamma(x) = \Delta\frac{\nabla f_\gamma(x) - \nabla\phi(x)}{\gamma} + \nabla f_\gamma(x) = \nabla f_\gamma(x)(1 + \frac{\Delta}{\gamma}) + \frac{\Delta}{\gamma}\nabla\phi(x),$$

where in the second step we rearranged $\nabla f_\gamma = \gamma c + \nabla\phi$. This gives the following bound for the Newton decrement of the updated function:

$$\|\nabla f_{\gamma+\Delta}(x)\|_{x,*} \le (1+\frac{\Delta}{\gamma})\|\nabla f_\gamma(x)\|_{x,*} + \frac{\Delta}{\gamma}\|\nabla\phi(x)\|_{x,*}.$$

Therefore, in order to bound the increase in the Newton decrement when moving $\gamma \to \gamma+\Delta$, we need to control this last term $\|\nabla\phi(x)\|_{x,*}$. This is exactly captured by the following:

Definition 12. *Self-concordant $\phi : D \to \mathbb{R}$ has* barrier parameter

$$\theta := \sup_{x\in D} \|v_x\|_x = \|\nabla\phi(x)\|_{x,*}.$$

If $\|\nabla f_\gamma(x)\|_{x,*} leq \beta$ for some small enough constant β, and we choose $\Delta \approx \frac{\gamma}{\theta}$ then we have

$$\|\nabla f_{\gamma+\Delta}(x)\|_{x,*} \le (1+\frac{\Delta}{\gamma})\|\nabla f_\gamma(x)\|_{x,*} + \frac{\Delta}{\gamma}\|\nabla\phi(x)\|_{x,*} \le (1+1/\theta)\beta + O(1).$$

If these constants are small enough, then we can apply a single Newton step to decrease the Newton decrement and return close to the central path for the new function $f_{\gamma+\Delta}$.

Therefore the parameter θ controls how fast we can increase the interpolation γ, i.e. how fast we can move along the central path while remaining sufficiently close to it that Newton's method has quadratic convergence. We refer the reader to [NN] for a very refined analysis and note the convergence result below.

Theorem 13. *If starting point x is in the β-neighborhood of x_{γ_0} for some small enough constant β, then in $\theta\log(\gamma_0/\varepsilon)$ iterations we can compute a point in the β neighborhood of $x_{1/\varepsilon}$.*

5 Approximation Guarantee and Stopping Criterion

The above sections analyzed the Newton step and path following procedure. In this subsection we give guarantees for optimality for the original LP eq. (1) in terms of the central path parameter. We require the following lemma about self-concordant barriers which we prove in section 7:

Lemma 14. *Let ϕ be self-concordant barrier for domain D with barrier parameter θ. Then for any $x, y \in D^\circ$:*

$$\langle\nabla\phi(x), y-x\rangle < \theta^2. \tag{3}$$

With this we can relate the central path parameter to optimality guarantee.

Proposition 15 (Theorem 4.2.7 in Nesterov)**.** *If x is in the β-neighborhood of the central path at some value γ, then*

$$\langle c, x-x^*\rangle \le \frac{1}{\gamma}\left(\theta^2 + \frac{(\beta+\theta)\beta}{(1-\beta)}\right).$$

Proof: Let x be our current point which is in the β neighborhood of the central path at γ, i.e.

$$x_\gamma = \arg\min_z \gamma\langle c, z\rangle + \phi(z), \qquad \|\nabla f_\gamma(x)\|_{x,*} \le \beta.$$

Let x^* be the optimal solution for the original LP eq. (1), and we can bound

$$\langle c, x-x^*\rangle = \langle c, x-x_\gamma\rangle + \langle c, x_\gamma - x^*\rangle. \tag{4}$$

For the first term we can use $\gamma c = \nabla f_\gamma(x) - \nabla\phi(x)$ by definition, so we can bound

$$\langle c, x - x_\gamma\rangle = \frac{1}{\gamma}\langle \nabla f_\gamma(x) - \nabla\phi(x), x - x_\gamma\rangle \le \frac{1}{\gamma}(\|\nabla f_\gamma(x)\|_{x,*} + \|\nabla\phi(x)\|_{x,*})\|x - x_\gamma\|_x \le \frac{\beta+\theta}{\gamma}\frac{\beta}{1-O(\beta)},$$

where in the second step we used duality of norms $\|\cdot\|_{x,*}, \|\cdot\|_x$, in the third we bounded $\|\nabla f_\gamma(x)\|_{x,*} \le \beta$ by assumption and $\|\nabla\phi(x)\|_{x,*} \le \theta$ by the barrier parameter, and the remaining term we justify heuristically below: recall x_γ is the optimizer of f_γ and by assumption x has bounded gradient $\|\nabla f_\gamma\|_{x,*} \le \beta$; by Hessian approximation theorem 7 we can show this implies f_γ is $1-O(\beta)$-strongly convex wrt $\|\cdot\|_x$ for the neighborhood $\|z - x\|_x \le O(\beta)$, which contains the optimizer; therefore we can use strong convexity to bound the distance $\|x - x_\gamma\|_x \lesssim \|\nabla f_\gamma(x)\|_{x,*}/(1-O(\beta))$, and finally bound the numerator by β by assumption.

For the second term in eq. (4), we use the optimality condition $0 = \nabla f_\gamma(x_\gamma) = \gamma c + \nabla\phi(x_\gamma)$:

$$\langle c, x_\gamma - x^*\rangle = \frac{1}{\gamma}\langle\nabla\phi(x_\gamma), x_\gamma - x^*\rangle < \theta^2,$$

where the last step is by theorem 14 which we prove in theorem 22(1).

Putting these two together gives the bound

$$\langle c, x - x^*\rangle = \langle c, x - x_\gamma\rangle + \langle c, x_\gamma - x^*\rangle \le \frac{1}{\gamma}\left(\frac{(\beta+\theta)\beta}{1-O(\beta)} + \theta^2\right).$$

□

Therefore, in order to guarantee ε-optimality for the original eq. (1), it is enough to compute a point x near the central path at $\gamma \gtrsim \theta^2/\varepsilon$.

6 Initialization

The above analysis of Newton step required the initial point to start close to the central path for some value of the interpolation parameter. But how do we obtain such an initial point? In this section we show a trick that effectively allows us to reuse the path following analysis to compute this initial point.

A reasonable place to start would be to find the *analytic center*, which is the point on the central path for parameter $\gamma = 0$, i.e. $x_{ac} := \arg\min_{x\in D}\phi(x)$. Say we begin with some initial point x_{init}. We do not know how close it is to the central path. But note by first-order optimality we have

$$x_{\text{init}} = \arg\min_{x\in D}\phi(x) - \langle\nabla\phi(x_{\text{init}}), x\rangle.$$

So for cost $b := -\nabla\phi(x_{\text{init}})$, x_{init} is exactly on the central path $\tilde{f}_\gamma := \gamma\langle b, \cdot\rangle + \phi$ with $\gamma = 1$.

Given this, we would like to decrease γ in order to get closer to the analytic center $\gamma = 0$. But note that the path following analysis works verbatim for decreasing γ as it does for increasing. Therefore we get the same rate of convergence:

Theorem 16. *Given $x_{init} \in D$, let $b := -\nabla\phi(x_{init})$ and (backwards) central path*

$$x_\gamma := \arg\min_{x\in D}\phi(x) + \gamma\langle b, x\rangle.$$

We can compute a point in the β neighorhood of x_ε in $O(\theta\log(1/\varepsilon))$ iterations.

Here we used x_{init} is exactly on the central path for $\gamma = 1$.

The final requirement is to understand how small we need to take ε so that we are actually close to the analytic center $x_{ac} = x_0$. After this we can switch to the central path corresponding to our original cost and follow the path following analysis above. We bound this error below:

Lemma 17. *If x is in the $\beta/2$-neighborhood of x_ε for $\varepsilon \lesssim \beta/\|\nabla\phi(x_{init})\|_{x,*}$, then x is in the β-neighborhood of the analytic center $x_{ac} = x_0$.*

Proof: This follows by simple triangle inequality

$$\|\nabla f_0(x)\|_{x,*} \leq \|\nabla f_\varepsilon(x)\|_{x,*} + \varepsilon\|b\|_{x,*},$$

and this last term is exactly $b = -\nabla\phi(x_{\text{init}})$ by definition. □

Therefore we have reduced the problem of finding the analytic center to finding a good initialization point. It remains to find x_{init} so that $\nabla\phi(x_{\text{init}})$ is bounded in *every* dual norm. This part requires problem specific analysis. But in general a sufficiently interior point of the polytope is a good starting point. In general, the quantity $\|\nabla\phi(x_{\text{init}})\|_{x,*}$ can be bounded in terms of the bit complexity of the description of eq. (1).

7 Dikin Ellipsoid

In this section we cover some useful results about self-concordant barriers. Our main goal is to prove theorem 14 which is used for the optimality guarantee in section 5, and we prove some other results with similar arguments.

Definition 18. *Let $D := dom(\phi)$ for self-concordant function ϕ. Then for $x \in D^\circ$, the* Dikin Ellipsoid *at x is*

$$D_x := \{y \mid \|y - x\|_x < 1\}.$$

Lemma 19. *For any $x \in D^\circ$, the Dikin Ellipsoid is contained in the domain $D_x \subseteq D$.*

Proof: We want to show $\phi(y) < \infty$ for any $\|y - x\|_x =: r < 1$; note $x \in D^\circ$ so $\phi(x), \nabla\phi(x)$ are both bounded; further

$$\phi(y) - (\phi(x) + \langle\nabla\phi(x), y - x\rangle) = \int_0^1 \int_0^t \langle(y - x), \nabla^2\phi(x + t(y - x))(y - x)\rangle = \int_0^1 \|y - x\|_{x_t},$$

where we used $x_t := x + t(y - x)$ for shorthand. If can show this second derivative term is bounded along the line $[x, y]$ then we are done. For this we can use eq. (2) shown above:

$$\|y - x\|_{x_t} = t^{-1}\|x_t - x\|_{x_t} \leq \frac{t^{-1}\|x_t - x\|_x}{1 - \|x_t - x\|_x} = \frac{\|y - x\|_x}{1 - t\|y - x\|_x} < \infty$$

where we used $x_t - x = t(y - x)$ and in the last step we used $t \in [0, 1], \|y - x\|_x < 1$. □

Lemma 20. *Self concordant ϕ has barrier parameter θ iff any of the following conditions hold for any $x \in D$:*

$$\|\nabla\phi(x)\|_{x,*} \leq \theta \quad \iff \begin{pmatrix} \nabla^2\phi(x) & \nabla\phi(x) \\ (\nabla\phi(x))^* & \theta^2 \end{pmatrix} \succeq 0 \quad \iff \nabla^2\phi(x) \succeq \frac{1}{\theta^2}(\nabla\phi(x))(\nabla\phi(x))^*.$$

Proof: [Proof Sketch] The first is the definition of the barrier parameter, and the remaining follow by Schur complement lemma since $\nabla^2\phi \succ 0$ by assumption.

Definition 21. $\ell_x(y) := \sup\{t \mid x + t(y-x) \in D\}$. *And* $\pi_x(y) := \inf\{t > 0 \mid y - x \in t(D-x)\}$. *Note* $\ell_x(y)\pi_x(y) = 1$.

Proposition 22. *Let* $D := dom(\phi)$ *for self-concordant function* ϕ *with barrier parameter* θ *according to theorem 12. Then*

1. *For* $x, y \in D^\circ$, $\langle \nabla\phi(x), y - x\rangle < \theta^2$.

2. $\langle \nabla\phi(y), y - x\rangle \le \frac{\theta^2}{\ell_x(y)-1}$

3. *If* $\langle \nabla\phi(x), y - x\rangle \ge 0$ *then*
$$\ell_x(y) \le (1 + 2\theta^2)\|y - x\|_x.$$

4. *As a corollary, let* $x_{ac} := \arg\min \phi(x)$ *be the analytic center, and* $D_{ac} := \{x \mid \|x - x_{ac}\|_{x_{ac}} < 1\}$ *be its Dikin ellipsoid. Then*
$$D_{ac} \subseteq D \subseteq (1 + 2\theta^2)D_{ac}.$$

5. *Let* $z \in \partial D$ *and consider* $x \in D$ *such that* $\ell_z(x) \ge (1+\theta)^2$. *Then*
$$\langle \nabla\phi(x), z - x\rangle \ge 1 - \frac{(1+\theta)^2}{\ell_z(x)}.$$

Proof: The key to all these results will be a functional inequality involving $x_t := x + t(y - x)$ and $g(t) := \langle \nabla\phi(x_t, y - x\rangle$. Note that by the barrier parameter we have

$$g'(t) = \langle y - x, \nabla^2\phi(x_t), (y - x)\rangle \ge \frac{\langle \nabla\phi(x_t), y - x\rangle^2}{\theta^2} = \frac{g(t)^2}{\theta^2},$$

where the second step was by the equivalent characterization of the barrier parameter given in theorem 20. Rearranging this gives functional inequality $(-1/g)' = g'/g^2 \ge 1/\theta^2$. We apply this to various points below:

1. Our goal is to upper bound $g(0) < \theta^2$, so if $g(0) \le 0$ we are already done. Otherwise, we use the above functional inequality to show
$$\frac{1}{g(0)} = \frac{1}{g(1)} - \int_0^1 \left(\frac{1}{g(t)}\right)' \ge \frac{1}{g(1)} + \frac{1}{\theta^2}.$$
Next, since $y \in D^\circ$ we have that $\nabla\phi(y)$ must be bounded, so $g(1) < \infty$. By our assumption $g(0) > 0$ and $g'(t) > 0$ we get $1/g(1) > 0$. Therefore we can rearrange to get
$$g(0) \le \left(\frac{1}{g(1)} + \frac{1}{\theta^2}\right)^{-1} < \theta^2,$$
where we used $1/g(1) > 0$ in the last step.

2. In this case we want to show an upper bound on $g(1)$. Now if $\ell_x(y) \le 1$ then the inequality is vacuously true, so consider $t > 1$. If $g(t) \le 0$ then by the above we have that g is monotone increasing so $g(1) \le g(t) \le 0$ and we are done. Otherwise, for any $t < \ell_x(y)$ we can bound
$$\frac{1}{g(1)} = \frac{1}{g(t)} - \int_1^t \left(\frac{1}{g(s)}\right)' > 0 + \frac{t-1}{\theta^2},$$

where for the first term we used $t < \ell_x(y)$ so $x_t \in D^\circ$ which by the discussion gives $g(t) < \infty \iff 1/g(t) > 0$; and for the second term we used the functional inequality $(1/g)' \geq 1/\theta^2$. Rearranging this and taking a limit gives

$$\langle \nabla\phi(y), y - x\rangle = g(1) < \lim_{t \to \ell_x(y)} \frac{\theta^2}{t-1} = \frac{\theta^2}{\ell_x(y) - 1},$$

where we used the above inequality for every $t < \ell_x(y)$ so $x_t \in D^\circ$.

3. Rearranging (2) we get

$$\ell_x(y) \leq 1 + \frac{\theta^2}{\langle \nabla\phi(y), y - x\rangle}.$$

If $\langle \nabla\phi(x), y - x\rangle \geq 0$ then

$$\langle \nabla\phi(y), y - x\rangle \geq \langle \nabla\phi(y) - \nabla\phi(x), y - x\rangle \geq \frac{\|y - x\|_x^2}{1 + \|y - x\|_x}.$$

Now to show containment, note that the required inequality is homogeneous, so we can scale to the case $\|y - x\|_x = 1$, where the above expression reduces to $1/2$ and we can show

$$\ell_x(y) \leq 1 + \frac{\theta^2}{1/2} = 1 + 2\theta^2,$$

i.e. for any $y \in \partial D_x$ with $\langle \nabla\phi(x), y - x\rangle \geq 0$, $x + t(y - x) \notin D$ for $t > 1 + 2\theta^2$.

4. If we consider the analytic center $x_{ac} := \arg\min_x \phi(x)$, then by optimality we have $\nabla\phi(x_{ac}) = 0 \implies \forall y \in D : \langle \nabla\phi(x_{ac}), y - x\rangle \geq 0$. Therefore we can apply the above to show

$$D_{x_{ac}} \subseteq D \subseteq (1 + 2\theta^2) D_{x_{ac}}.$$

5. Let $x_t := z + t(x - z)$ and $g(t) := \langle \nabla\phi(x_t), x - z\rangle$, so our goal is to upper bound $g(1)$. Since $z \in \partial D$ and for every $y \in D$ we have $D_y \subseteq D$, we have $\|z - y\|_y \geq 1$. In terms of g this gives the bound

$$g'(t) = \langle x - z, \nabla^2\phi(x_t)(x - z)\rangle = \|x - z\|_{x_t}^2 = \frac{\|x_t - z\|_{x_t}^2}{t^2} \geq \frac{1}{t^2},$$

where the second step was by definition of $\|\cdot\|_{x_t}$, and in the third we used $x_t - z = t(x - z)$. From (2) we also have: $\langle \nabla\phi(y), y - z\rangle \leq \theta^2/(\ell_z(y) - 1)$, which applying to x_t gives

$$g(t) = \langle \nabla\phi(x_t), x - z\rangle = t^{-1}\langle \nabla\phi(x_t), x_t - z\rangle \leq t^{-1}\frac{\theta^2}{\ell_z(x_t) - 1} = \frac{t^{-1}\theta^2}{t^{-1}\ell_z(x) - 1},$$

where we again used $x_t - z = t(x - z)$ and in the last step this gives $\ell_z(x_t) = t^{-1}\ell_z(x)$. Putting these together for $t \geq 1$ we have the bound

$$g(1) = g(t) - \int_1^t g'(s) \leq \frac{t^{-1}\theta^2}{t^{-1}\ell_z(x) - 1} - \int_1^t \frac{1}{s^2} = \frac{\theta^2}{\ell_z(x) - t} - \left(\frac{1}{1} - \frac{1}{t}\right),$$

and finally choosing $t = \frac{\ell_z(x)}{1+\theta} \geq 1$ gives

$$g(1) + 1 \leq \frac{1}{t} + \frac{\theta^2}{\ell_z(x) - t} = \frac{1+\theta}{\ell_z(x)} + \frac{\theta^2}{\ell_z(x)(1 - 1/(1+\theta))} = \frac{1+\theta}{\ell_z(x)}\left(1 + \frac{\theta^2}{\theta/(1+\theta)}\right) = \frac{(1+\theta)^2}{\ell_z(x)},$$

which is exactly the required result.

□