Chapter X: Interior Point Methods

This is a crash course on interior point methods for convex optimization. We will not make rigorous proofs for all statements in this handout, as it would take too much time, but we will try to understand the general behaviour of these optimization algorithms. This material is based on the following references [3, Chapters 2–4], [1, Chapters 9–11], [2, Lectures 16–17].

1 General Idea

Consider a convex optimization problem

$$\min_{\substack{\boldsymbol{x}\in\mathcal{X}\\ A\boldsymbol{x}=\boldsymbol{b}}} \quad f(\boldsymbol{x}) \tag{P}$$

where $\mathcal{X} \subseteq \mathbb{R}^n$ is a closed convex set with a nonempty interior (Note: we can always reduce to this case).

The idea of *interior point methods* (IPMs) is to equip $\mathbf{int}\,\mathcal{X}$ with a *barrier function* F satisfying:

(i) F is smooth and *strongly convex**, i.e., $\exists \nu > 0 : \nabla^2 F(\boldsymbol{x}) \succeq \nu I$ for all $\boldsymbol{x} \in \mathcal{X}$.

(ii) $F(\boldsymbol{x}_k) \to \infty$ for all sequences of points $\boldsymbol{x}_k \in \mathbf{int}\,\mathcal{X}$ converging to a boundary point $\bar{\boldsymbol{x}} \in \partial X$.

Then, for all $t \geq 0$, the *equality-constrained* penalized problem

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & t\, f(\boldsymbol{x}) + F(\boldsymbol{x}) \\ s.t. \quad & A\boldsymbol{x} = \boldsymbol{b} \end{aligned} \tag{P_t}$$

has a unique solution, satisfying $\boldsymbol{x}^*(t) \in \mathbf{int}\,\mathcal{X}$ and $A\boldsymbol{x}^*(t) = \boldsymbol{b}$. The set of points $\{\boldsymbol{x}^*(t) : t \geq 0\}$ is called the *central path*, and (under mild conditions) it converges to a point in the optimal set of (P) as $t \to \infty$.

Path-following algorithms typically work as follows: Given a current iterate ($t_k > 0, \boldsymbol{x}^{(k)} \in \mathbf{int}\,\mathcal{X}$), with $A\boldsymbol{x}^{(k)} = \boldsymbol{b}$ and $\boldsymbol{x}^{(k)}$ *reasonably close* to $\boldsymbol{x}^*(t_k)$,

1. replace the current value of t_k by a new, larger value $t_{k+1} > t_k$;
2. use an equality-constrained minimization algorithm to solve (approximately) Problem $(P_{t_{k+1}})$, starting from the initial guess $\boldsymbol{x}^{(k)}$, to obtain a point $\boldsymbol{x}^{(k+1)}$ *reasonably close* to $\boldsymbol{x}^*(t_{k+1})$

Advandages:

- Can use methods of unconstrained (or equality-constrained) optimization, such as Newton's method;
- For certain class of barrier functions (theory of *self-concordance*), we can prove that the above path-following scheme returns an $\epsilon-$suboptimal solution in polynomial time (w.r.t. input size of the problem and $\log(1/\epsilon)$);
- IPMs are good for both theoretical and practical purpose.

We are next going to study how to solve unconstrained optimization problems (and the same method can easily be generalized to equality-constrained optimization, see Section 4), and give some details on the steps 1. and 2. above, in order to control the speed of convergence.

*the definition of strong-convexity can be extended to non-smooth functions: we say that F is ν-strictly convex iff $\boldsymbol{x} \mapsto F(\boldsymbol{x}) - \nu/2\|\boldsymbol{x}\|^2$ is convex.

2 Newton's Method for unconstrained optimization

We consider an unconstrained optimization problem $p^* = \inf_{\boldsymbol{x}} f(\boldsymbol{x})$, where f is twice differentiable (in particular, $\mathbf{dom}\, f$ is open), and strongly convex (in particular, the Hessian matrix $\nabla^2 f(\boldsymbol{x})$ is always invertible).

One way to present Newton's method is via successive quadratic approximations: Given a current iterate $\boldsymbol{x}^{(k)} \in \mathbf{dom}\, f$, we can use Taylor expansion to obtain a 2d order approximation of f around $\boldsymbol{x}^{(k)}$: $f(\boldsymbol{x}^{(k)} + \boldsymbol{u}) \simeq \hat{f}(\boldsymbol{x}^{(k)} + \boldsymbol{u}) := f(\boldsymbol{x}^{(k)}) + \nabla f(\boldsymbol{x}^{(k)})^T \boldsymbol{u} + \frac{1}{2} \boldsymbol{u}^T \nabla^2 f(\boldsymbol{x}^{(k)}) \boldsymbol{u}$. The basic Newton's method sets $\boldsymbol{x}^{(k+1)}$ equal to the unique minimizer of $\hat{f}$:

$$\boldsymbol{x}^{(k+1)} := \boldsymbol{x}^{(k)} + \Delta \boldsymbol{x}, \quad \text{where } \Delta \boldsymbol{x} = -\nabla^2 f(\boldsymbol{x}^{(k)})^{-1} \nabla f(\boldsymbol{x}^{(k)}).$$

In practice however, this method can fail to converge, in particular one can have $\boldsymbol{x}^{(k+1)} \notin \mathbf{dom}\, f$. To circumvent this issue, one can take *damped* Newton steps, $\boldsymbol{x}^{(k+1)} := \boldsymbol{x}^{(k)} + \delta \Delta \boldsymbol{x}$, where the step size $0 < \delta \leq 1$ is found by performing a *backtracking line search*: Given search parameters $0 < \alpha \leq \frac{1}{2}$ and $0 < \beta < 1$, we start at $\delta = 1$, and we shrink $\delta := \beta \delta$ until the criterion $f(\boldsymbol{x}^{(k+1)}) \leq f(\boldsymbol{x}^{(k)}) + \alpha \delta \nabla f(\boldsymbol{x}^{(k)})^T \Delta \boldsymbol{x}$ is fulfilled.

As a stopping criterion, one can use $\frac{1}{2} \lambda(\boldsymbol{x}^{(k)})^2 \leq \epsilon$, where ϵ is a tolerance parameter. Here, the function $\lambda(\boldsymbol{x})$ is the *Newton decrement*:

$$\lambda(\boldsymbol{x}) = \sqrt{\nabla f(\boldsymbol{x})^T \ \nabla^2 f(\boldsymbol{x})^{-1} \ \nabla f(\boldsymbol{x})}.$$

The quantity $\frac{1}{2} \lambda(\boldsymbol{x})^2$ is equal to $f(\boldsymbol{x}) - \inf_{\boldsymbol{y}} \hat{f}(\boldsymbol{y})$, where $\hat{f}$ is the quadratic approximation of f at $\boldsymbol{x}$. Hence $\frac{1}{2} \lambda(\boldsymbol{x}^{(k)})^2$ is an estimate of $f(\boldsymbol{x}^{(k)}) - p^*$.

3 Convergence Analysis of Newton's Method

There are two main frameworks that can be used to control the convergence of Newton's method. The first one relies on regularity parameters of the function f. The first approach has the disavantage that it depends on some parameters m, L and H that are, in general, unknwon. Instead, a parameter-free analysis has been proposed by Nesterov and Nemirovski for *self-concordant* functions. The class of functions of this second approach is thus restricted, but this is exactly the class of functions we need to analyze LP, SOCP, and SDPs.

We give below the main result for these two approaches, starting with the standard analysis based on strong convexity.

3.1 Convergence analysis for regular functions

Assume that f is strongly convex with parameter ν, ∇f is Lipschitz with parameter M, and $\nabla^2 f$ is Lipschitz with parameter L, i.e., $\|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\| \leq M \|\boldsymbol{x} - \boldsymbol{y}\|$, $\|\nabla^2 f(\boldsymbol{x}) - \nabla^2 f(\boldsymbol{y})\|_F \leq L \|\boldsymbol{x} - \boldsymbol{y}\|$, and we use Newton's method with backtracking line search (with parameters α and β). Then, there exists 2 numbers η and γ, which depend on the parameters α, β, ν, M and L, such that:

- In a first phase (as long as $\|\nabla f(\boldsymbol{x}^{(k)})\| \geq \eta$), the criterion is decreased by at least $\gamma \geq 0$ at each iteration:

$$f(\boldsymbol{x}^{(k)}) - f(\boldsymbol{x}^{(k+1)}) \geq \gamma. \tag{1}$$

- In a second phase (for all iterations $k \geq \ell$, where $k = \ell$ is the first iterate satisfying $\|\nabla f(\boldsymbol{x}^{(k)})\| < \eta$), the backtracking line search always returns the step size $\delta = 1$, and it holds:

$$f(\boldsymbol{x}^{(k)}) - p^* \leq \frac{2\nu^3}{L^2} \left(\frac{1}{2}\right)^{2^{k-\ell+1}}. \tag{2}$$

We can use (1) to bound the number of iterations in the first phase by $\frac{f(\boldsymbol{x}^{(0)})-p^*}{\gamma}$, and Equation (2) shows that the convergence is *locally quadratic*, that is, when we are close to the solution, the number of accurate digits typically doubles at each iteration. For all practical purposes, 5 or 6 iterations of the second phase ensure that we have found an extremely accurate solution.

One drawback of the above approach is that the analysis is not *affine invariant*: For example, the parameters M and L change if we make a change of variable $\boldsymbol{y} = \alpha\boldsymbol{x} + \beta$, although the Newton method *is affine invariant*. The analysis of Newton's method based on self-concordance circumvents this issue.

3.2 Convergence analysis for self-concordant functions

We introduce a class of functions that satisfy certain regularity assumptions. The next definition may look a bit strange, but it turns out that when working with conic programming problems, we can use self-concordant penalty functions to follow the central path (cf. Section 5.2).

Definition 1 (Self-concordant function)**.** A function $f : \mathbb{R}^n \to \mathbb{R}$ is called self-concordant if

- $\mathbf{dom}\, f$ is an open convex set;
- $\nabla^2 f(\boldsymbol{x}) \succ 0$ for all $\boldsymbol{x} \in \mathbf{dom}\, f$ (in particular, f is strictly convex);
- $f(\boldsymbol{x}_i) \to \infty$ along every sequence $\{\boldsymbol{x}_i \in \mathbf{dom}\, f\}$ converging (as $i \to \infty$) to a boundary point of $\mathbf{dom}\, f$ (we call this the *barrier property*);
- f is C^3 and for all $\boldsymbol{x} \in \mathbf{dom}\, f$, $\boldsymbol{h} \in \mathbb{R}^n$, the restriction F to the line $\boldsymbol{x} + t\boldsymbol{h}$, i.e., $F : t \mapsto f(\boldsymbol{x} + t\boldsymbol{h})$ satisfies
$$|F'''(t)| \leq 2F''(t)^{3/2}, \qquad \forall t \in \mathbf{dom}\, F. \tag{3}$$

Recall that in the standard case, $\frac{1}{2}\lambda(\boldsymbol{x})^2$ is only an *estimate* of $f(\boldsymbol{x}) - p^*$, but here (when f is self-concordant), multiplying this estimate by 2 yields a provable upper bound on the gap to optimality.

Proposition 1. *If f is self-concordant and the Newton decrement is $\lambda(\boldsymbol{x}) \leq 0.68$, then it holds*
$$f(\boldsymbol{x}) - p^* \leq \lambda(\boldsymbol{x})^2.$$

Proof. First, we can show that the self-concordance inequality (3) is equivalent to
$$\left|\frac{d}{dt}F''(t)^{-1/2}\right| \leq 1,\ \forall t \in \mathbf{dom}\, F. \tag{4}$$
This implies: $F''(t)^{-1/2} \leq F''(0)^{-1/2} + t$ for all $t \geq 0$, i.e.,
$$F''(t) \geq \frac{1}{(F''(0)^{-1/2} + t)^2} = \frac{F''(0)}{(1 + tF''(0)^{1/2})^2}.$$
Now, we consider a semi-line $\boldsymbol{x} + t\boldsymbol{h}, t \geq 0$ of $\mathbb{R}^n$, and we consider the function $F : t \mapsto f(\boldsymbol{x} + t\boldsymbol{h})$, which is self-concordant by definition. Integrating the above inequality, we obtain:
$$F'(t) \geq F'(0) + \frac{F''(0)t}{1 + tF''(0)^{1/2}} = F'(0) + F''(0)^{1/2} - \frac{F''(0)^{1/2}}{1 + tF''(0)^{1/2}}.$$
Then, we integrate a second time, to obtain:
$$F(t) \geq F(0) + t(F'(0) + F''(0)^{1/2}) - \log(1 + tF''(0)^{1/2}).$$

The right-hand side of the above expression is a convex function, which reaches its minimum at $t^* = -F'(0)/[F''(0) + F''(0)^{1/2}F'(0)]$. Evaluating this expression at t^*, we obtain:

$$\begin{aligned}\forall t \in \mathbf{dom}\, F \cap \mathbb{R}_+,\ F(t) &\geq F(0) + t^*(F'(0) + F''(0)^{1/2}) - \log(1 + t^* F''(0)^{1/2})\\ &= F(0) - F'(0)F''(0)^{-1/2} + \log\left(1 + F'(0)F''(0)^{-1/2}\right)\\ &= F(0) + \rho\Big(-F'(0)F''(0)^{-1/2}\Big), \end{aligned} \tag{5}$$

where $\rho(u) := u + \log(1-u) = -(x^2/2 + x^3/3 + x^4/4 + \ldots)$.

Now, we replace the values of F and its derivatives at $t=0$ by their respective expressions, i.e.,

$$F(0) = f(\boldsymbol{x}), \quad F'(0) = \boldsymbol{h}^T\nabla f(\boldsymbol{x}), \quad F''(0) = \boldsymbol{h}^T\nabla^2 f(\boldsymbol{x})\boldsymbol{h}.$$

In particular, this yields

$$F'(0)F''(0)^{-1/2} = \frac{\boldsymbol{h}^T\nabla f(\boldsymbol{x})}{\sqrt{\boldsymbol{h}^T\nabla^2 f(\boldsymbol{x})\boldsymbol{h}}} = \frac{\Big\langle \boldsymbol{h}, \left(\nabla^2 f(\boldsymbol{x})\right)^{-1}\, \nabla f(\boldsymbol{x})\Big\rangle_{\boldsymbol{x}}}{\sqrt{\langle \boldsymbol{h}, \boldsymbol{h}\rangle_{\boldsymbol{x}}}},$$

where the inner product is $\langle \boldsymbol{a}, \boldsymbol{b}\rangle_{\boldsymbol{x}} = \boldsymbol{a}^T\nabla^2 f(\boldsymbol{x})\boldsymbol{b}$. Hence, Cauchy-Schwartz inequality shows that $|F'(0)F''(0)^{-1/2}| \leq \sqrt{\nabla f(\boldsymbol{x})^T\ \nabla^2 f(\boldsymbol{x})^{-1}\ \nabla f(\boldsymbol{x})} = \lambda(\boldsymbol{x})$. In particular, this shows that the expression (5) is well defined for $\lambda(\boldsymbol{x}) < 1$. In this case, since one can show that $\rho(u) \geq \rho(a)$ for all $|u| \leq a$, we obtain:

$$\forall t \in \mathbf{dom}\, F \cap \mathbb{R}^+,\ F(t) \geq f(\boldsymbol{x}) + \rho(\lambda(\boldsymbol{x}))$$

Finally, since the direction $\boldsymbol{h}$ was chosen arbitrarily, it holds

$$p^* \geq f(\boldsymbol{x}) + \rho(\lambda(\boldsymbol{x})) \iff f(\boldsymbol{x}) - p^* \leq -\rho(\lambda(\boldsymbol{x})),$$

and a straightforward analysis shows that $-\rho(\lambda) \leq \lambda^2$ whenever $0 \leq \lambda \leq 0.683803$. □

Now, we consider damped Newton steps $\boldsymbol{x}^+ := \boldsymbol{x} + \delta\Delta\boldsymbol{x}$, where the step sizes $\delta = \frac{1}{1+\lambda(\boldsymbol{x}^{(k)})}$ are controlled by the Newton decrement $\lambda(\boldsymbol{x})$.

Proposition 2. *Let $\boldsymbol{x} \in \mathbf{int\,dom}\, f$. Then, the iterate $\boldsymbol{x}^+$ of the damped Newton step satisfies:*

1. $\boldsymbol{x}^+ \in \mathbf{int\,dom}\, f$;
2. $f(\boldsymbol{x}) - f(\boldsymbol{x}^+) \geq \Big(\lambda(\boldsymbol{x}) - \log\left(1 + \lambda(\boldsymbol{x})\right)\Big)$.
   In particular, if $\lambda(\boldsymbol{x}) \geq \frac{1}{4}$, then $f(\boldsymbol{x}) - f(\boldsymbol{x}^+) \geq (\frac{1}{4} - \log\frac{5}{4}) = 0.026856$;
3. *If $\lambda(\boldsymbol{x}) \leq \frac{1}{4}$, then we are in the region of quadratic convergence, and $2\lambda(\boldsymbol{x}^+) \leq (2\lambda(\boldsymbol{x}))^2$.*

Proof. 1. Consider the function $F : t \mapsto f(\boldsymbol{x} + t\boldsymbol{h})$ on the half line $\boldsymbol{x} + t\boldsymbol{h}, t \geq 0$, where $\boldsymbol{h} = \Delta\boldsymbol{x}$ is the Newton direction. This function satisfies $F'(0) = -\lambda(\boldsymbol{x})^2$ and $F''(0) = \lambda(\boldsymbol{x})^2$. From the self-concordance of F, we have $F''(t)^{-1/2} \geq F''(0)^{-1/2} - t$ for $t \geq 0$, cf. (4), that is,

$$\forall t \in \mathbf{dom}\, F \cap [0, F''(0)^{-1/2}), \quad F''(t) \leq \frac{1}{(F''(0)^{-1/2} - t)^2} = \frac{F''(0)}{(1 - tF''(0)^{1/2})^2}.$$

We integrate this relation twice:

$$F'(t) \leq F'(0) + \frac{F''(0)t}{1 - tF''(0)^{1/2}}.$$

$$F(t) \leq F(0) + tF'(0) - F''(0)^{1/2}t - \log(1 - tF''(0)^{1/2}).$$

Substituting the values of F and its derivatives at $t=0$,

$$F(t) = f(\boldsymbol{x} + t\boldsymbol{h}) \leq f(\boldsymbol{x}) - t\lambda(\boldsymbol{x})^2 - t\lambda(\boldsymbol{x}) - \log(1 - t\lambda(\boldsymbol{x})), \tag{6}$$

which is valid for all $t \in \mathbf{dom}\, F$ such that $0 \leq t < F''(0)^{-1/2} = \frac{1}{\lambda(\boldsymbol{x})}$. Assume (by contradiction) that $\delta \notin \mathbf{int\,dom}\, F$. Then, the segment $[\boldsymbol{x}, \boldsymbol{x}+\delta\boldsymbol{h}]$ intersects the boundary of $\mathbf{dom}\, f$. Hence, there is a sequence $\{t_i\}$ such that $\boldsymbol{x}_i = \boldsymbol{x}+t_i\boldsymbol{h}$ converges to a boundary point of $\mathbf{dom}\, f$, with $t_i \in \mathbf{dom}\, F$ and $0 \leq t_i \leq \delta < \frac{1}{\lambda(\boldsymbol{x})}$, $\forall i \in \mathbb{N}$. By (6), the sequence $\{F(t_i)\}$ is bounded, but this contradicts the barrier property of the self-concordant function f. This shows $\delta \in \mathbf{int\,dom}\, F$, that is, $\boldsymbol{x}^+ \in \mathbf{int\,dom}\, f$.

2. Now, let us evaluate the inequality (6) at $t = \delta = \frac{1}{1+\lambda(\boldsymbol{x})}$:

$$\begin{aligned} f(\boldsymbol{x}^+) &\leq f(\boldsymbol{x}) - \frac{\lambda(\boldsymbol{x})^2}{1+\lambda(\boldsymbol{x})} - \frac{\lambda(\boldsymbol{x})}{1+\lambda(\boldsymbol{x})} - \log(1 - \frac{\lambda(\boldsymbol{x})}{1+\lambda(\boldsymbol{x})}) \\ &= f(\boldsymbol{x}) - \lambda(\boldsymbol{x}) + \log(1+\lambda(\boldsymbol{x})). \end{aligned}$$

3. We skip the proof of this fact, which requires additional technical lemmas. The idea is to show that self-concordance implies a bound on the Hessian matrix at $\boldsymbol{x}^+$, and as usual we must integrate this bound twice. □

We are now ready to prove the main result on the convergence of the damped Newton's method for self-concordant functions:

Theorem 3. *Given an initial guess $\boldsymbol{x}^{(0)} \in \mathbf{int\,dom}\, f$, the number of damped Newton steps*

$$\boldsymbol{x}^{(k+1)} \leftarrow \boldsymbol{x}^{(k)} + \frac{1}{1+\lambda(\boldsymbol{x}^{(k)})}\ \nabla^2 f(\boldsymbol{x}^{(k)})^{-1} \nabla f(\boldsymbol{x}^{(k)})$$

required to find an $\epsilon-$suboptimal solution to the problem of minimizing $f(\boldsymbol{x})$ is upper bounded by

$$O(1)[f(\boldsymbol{x}_0) - p^*] + \log_2 \log_2 \frac{1}{\epsilon},$$

where the hidden constant does not depend on f, and is less than 38.

Proof. By Proposition 2 (2.), the number of iteration of the first phase, during which $\lambda(\boldsymbol{x}^{(k)}) \geq \frac{1}{4}$, is upper bounded by $C(f(\boldsymbol{x}_0) - p^*)$, where $C \simeq \frac{1}{0.026856}$, hence $C^{-1} < 38$.

Let ℓ be the index of the first iterate such that $\lambda(\boldsymbol{x}^{(\ell)}) < \frac{1}{4}$. By using the results of Propositions 1 and Proposition 2 (3.), an induction over $k \geq \ell$ shows that

$$f(\boldsymbol{x}^{(k)}) - p^* \leq \lambda(\boldsymbol{x}^{(k)})^2 \leq \left(\frac{1}{2}\right)^{2^{k-\ell+1}}, \quad \forall k \geq \ell.$$

Hence, we reach the desired precision after at most $\log_2 \log_2 \frac{1}{\epsilon}$ iterations of the second phase. □

4 Newton's method for equality constrained problems

Newton's method can be adapted to the case of equality constrained problems $p^* = \inf\{f(\boldsymbol{x}) : A\boldsymbol{x} = \boldsymbol{b}\}$. To this end, we must chose *feasible* Newton directions. To this end, given a current iterate $\boldsymbol{x}$ satisfying $A\boldsymbol{x} = \boldsymbol{b}$, we minimize the quadratic approximation $\hat{f}$ of f around $\boldsymbol{x}$:

$$\begin{aligned} \min_{u} \quad & \hat{f}(\boldsymbol{x} + \boldsymbol{u}) := f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T \boldsymbol{u} + \frac{1}{2} \boldsymbol{u}^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{u} \\ s.t. \quad & A(\boldsymbol{x} + \boldsymbol{u}) = \boldsymbol{b}. \end{aligned}$$

The constraint of this problem rewrites $A\boldsymbol{u} = 0$, and the (necessary and sufficient) KKT conditions for this problem are: $\exists \boldsymbol{\mu}$:

$$\begin{aligned} A\boldsymbol{u} &= 0 \qquad \text{(primal feasibility)} \\ \nabla f(\boldsymbol{x}) + \nabla^2 f(\boldsymbol{x})\boldsymbol{u} + A^T\boldsymbol{\mu} &= 0 \qquad \text{(gradient of Lagrangian vanishes).} \end{aligned}$$

Hence, at iteration k, the Newton direction $\boldsymbol{\Delta x}$ is found by solving the following KKT system:

$$\begin{bmatrix} \nabla^2 f(\boldsymbol{x}^{(k)}) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Delta x} \\ \boldsymbol{\mu} \end{bmatrix} = \begin{bmatrix} -\nabla f(\boldsymbol{x}^{(k)}) \\ 0 \end{bmatrix}$$

Then, the convergence analysis of damped newton method is similar as in the unconstrained case.

Remark 4. In fact, we know that equality constrained can be removed for a problem, by expressing the feasible affine space as

$$A\boldsymbol{x} = \boldsymbol{b} \iff \exists \boldsymbol{z} \in \mathbb{R}^r : \boldsymbol{x} = U\boldsymbol{z} + \boldsymbol{x}_0,$$

for some matrix $U \in \mathbb{R}^{n\times r}$ and a vector $\boldsymbol{x}_0 \in \mathbb{R}^n$. Then, it can be shown that the Newton method for the equivalent unconstrained problem

$$\underset{\boldsymbol{z}\in\mathbb{R}^r}{\textbf{minimize}} \qquad f(U\boldsymbol{z} + \boldsymbol{x}_0)$$

produces the same iterates as the Newton method for the above equality constrained problem.

5 Path following algorithm

As stated in the introduction, at each iteration of the basic IPM algorithm, we increasse the penalty parameter t, and we then perform a *centering*, which consists in using Newton's method to come close to the central path $\boldsymbol{x}^*(t)$.

In this section, we make the simplifying assumptions that (i) an interior point $\boldsymbol{x}^{(0)}$ is given; and (ii) *exact centering* is used, that is, at iteration k, we have a penalty parameter t_k and we compute $\boldsymbol{x}^{(k)} = \boldsymbol{x}^*(t_k)$ by solving Problem (P_{t_k}) exactly, but we point out that the method can be modified to allow inexact centering, by imposing a bound on the Newton decrement $\lambda_{t_k}(\boldsymbol{x}^{(k)})$ of the centering Problem (P_{t_k}).

Concerning assumption (i), the barrier method can be applied to another optimization problem, called *phase I optimization problem*, for which a strictly feasible solution can easily be found, and whose solution yields either a strictly feasible point, or a certificate that Problem (7) is not strictly feasible.

5.1 Barrier method for a convex problem with m inequality constraints

We consider a convex optimization problem of the form

$$\begin{aligned} p^* = \min \quad & f_0(\boldsymbol{x}) \\ s.t. \quad & f_i(\boldsymbol{x}) \leq 0 \quad (i = 1, \ldots, m) \\ & A\boldsymbol{x} = \boldsymbol{b}, \end{aligned} \tag{7}$$

We assume that the f_i's are twice differentiable, and that Problem (7) is strictly feasible and solvable, so there exists a pair of primal and dual optimal solution : $p^* = f(\boldsymbol{x}^*) = g(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$

We study the central path $\{\boldsymbol{x}^*(t) : t \geq 0\}$, where $\boldsymbol{x}^*(t)$ is the unique minimizer (uniqueness can be guaranteed if at least one of the f_i's is strongly convex) of:

$$\begin{aligned} \min \quad & F_t(\boldsymbol{x}) := t f_0(\boldsymbol{x}) + \sum_{i=1}^m - \log(-f_i(\boldsymbol{x})) \\ & A\boldsymbol{x} = \boldsymbol{b}, \end{aligned} \tag{Q_t}$$

Intuitively, to justify this approach, observe that if we divide the objective by t, we obtain $f_0(\boldsymbol{x}) + \frac{1}{t}\sum_{i=1}^m - \log(-f_i(\boldsymbol{x}))$, which can be interpreted as an approximation (as $t \to \infty$) of the function

$$\begin{cases} f_0(\boldsymbol{x}) & \text{if } \forall i \geq 1,\ f_i(\boldsymbol{x}) \leq 0 \\ +\infty & \text{otherwise.} \end{cases}$$

We now show that the central path actually converges to an optimal solution of Problem (7). In fact, we even obtain a bound on suboptimality of $\boldsymbol{x}^*(t)$ which depends on t.

Proposition 5. *Let $\boldsymbol{x}^*(t)$ solve Problem (Q_t). Then, $\boldsymbol{x}^*(t)$ is feasible for Problem (7), and*

$$f_0(\boldsymbol{x}^*(t)) - p^* \leq \frac{m}{t}.$$

Proof. When we use Newton's method to solve the equality constrained problem (Q_t), we obtain $\boldsymbol{x}^*(t)$ and also the optimal multiplier $\boldsymbol{\mu}^*(t)$ for the equality constraints, which satisfies $\nabla F_t(\boldsymbol{x}^*(t)) + A^T\boldsymbol{\mu}^*(t) = 0$.

Now, we will prove the following: if we set

$$\lambda_i(t) = -\frac{1}{t f_i(\boldsymbol{x}^*(t))}\ (\forall i = 1, \ldots, m), \quad \boldsymbol{\mu}(t) = \frac{\boldsymbol{\mu}^*(t)}{t},$$

then $(\boldsymbol{\lambda}(t), \boldsymbol{\mu}(t))$ is dual feasible for the original problem (7), and hence provide a bound. The fact that $\boldsymbol{\lambda}(t) \geq \boldsymbol{0}$ is clear, because $f_i(\boldsymbol{x}^*(t)) < 0$. So the only thing to show is that $(\boldsymbol{\lambda}(t), \boldsymbol{\mu}(t)) \in \mathbf{dom}\, g$, that is, $g(\boldsymbol{\lambda}(t), \boldsymbol{\mu}(t)) > -\infty$, where g is the Lagrange dual function for Problem (7). The optimality condition $\nabla F_t(\boldsymbol{x}^*(t)) + A^T\boldsymbol{\mu}^*(t) = 0$ can be rewritten as

$$\begin{aligned} & t\nabla f_0(\boldsymbol{x}^*(t)) - \sum_{i=1}^m \frac{1}{f_i(\boldsymbol{x}^*(t))} \nabla f_i(\boldsymbol{x}^*(t)) + A^T\boldsymbol{\mu}^*(t) = \boldsymbol{0} \\ \iff & \nabla f_0(\boldsymbol{x}^*(t)) + \sum_{i=1}^m \lambda_i(t)\ \nabla f_i(\boldsymbol{x}^*(t)) + A^T\boldsymbol{\mu}(t) = \boldsymbol{0} \\ & \iff \nabla_{\boldsymbol{x}}\ \mathcal{L}(\boldsymbol{x}^*(t), \boldsymbol{\lambda}(t), \boldsymbol{\mu}(t)) = \boldsymbol{0}, \end{aligned}$$

where $\mathcal{L}$ is the lagrangian of Problem (7). This shows that $\boldsymbol{x}^*(t)$ is a minimizer of $\boldsymbol{x} \mapsto \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}(t), \boldsymbol{\mu}(t))$. So:

$$p^* \geq d^* \geq g(\boldsymbol{\lambda}(t), \boldsymbol{\mu}(t)) = \mathcal{L}(\boldsymbol{x}^*(t), \boldsymbol{\lambda}(t), \boldsymbol{\mu}(t)) = f_0(\boldsymbol{x}^*(t)) + \underbrace{\sum_{i=1}^m \lambda_i(t)\ f_i(\boldsymbol{x}^*(t))}_{=-\frac{m}{t}} + \underbrace{\boldsymbol{\mu}(t)^T(A\boldsymbol{x}^*(t) - \boldsymbol{b})}_{=0}.$$

This shows that $f_0(\boldsymbol{x}^*(t)) - p^* \leq \frac{m}{t}$, and hence $\boldsymbol{x}^*(t)$ converges to an optimal solution of Problem (7). □

The bound $f_0(\boldsymbol{x}^*(t)) - p^* \leq \frac{m}{t}$ shows that to solve the original problem approximately (within an error ϵ), it suffices to solve the equality-constrained problem (Q_t) for a $t \geq \frac{m}{\epsilon}$. This approach does not work well in practice, because the function F_t becomes *nasty* when t grows large. The path following method is an alternative, in which we minimize $F_t(\boldsymbol{x})$ for a sequence of increasing values of t, until $t \geq \frac{m}{\epsilon}$, by using the last point found as a starting point for the next equality-constrained problem.

The basic algorithm starts at $t = t_0$ and increases t as $t := \omega t$ at each iteration, so that the accuracy $\frac{m}{t_0\omega^k}$ is achieved after k iterations. This shows that the desired accuracy level ϵ is reached after exactly

$k = \left\lceil \frac{\log m/(\epsilon t_0)}{\log \omega} \right\rceil$. Clearly, there is a tradeoff in the choice of the parameters t_0 and ω. Typically, a large ω decreases the number of centering iterations, but each iteration will require a larger amount of Newton steps, because $\boldsymbol{x}^*(t)$ is not necessarily a good starting point for Problem $(Q_{\omega t})$.

In practice, this method works pretty well for values of ω around 10 or 20. However, with this approach it is hard to prove that the centering problems does not become prohibitively difficult when t becomes large. But the theory of self-concordance gives an elegant framework, which allows one to bound the total number of Newton steps needed to reach the accuracy ϵ.

5.2 Conic Program with a self-concordant barrier

We consider a pair of primal and dual cone programs

$$\begin{aligned} \min \quad & \boldsymbol{c}^T\boldsymbol{x} & (P) \\ s.t. \quad & F\boldsymbol{x} = \boldsymbol{g} \\ & A\boldsymbol{x} \succeq \boldsymbol{b} \end{aligned} \qquad\qquad \begin{aligned} \max \quad & \boldsymbol{g}^T\boldsymbol{y} + \boldsymbol{b}^T\boldsymbol{z} & (D) \\ s.t. \quad & F^T\boldsymbol{y} + A^T\boldsymbol{z} = \boldsymbol{c} \\ & \boldsymbol{z} \succeq_* 0 \end{aligned}$$

The generalized inequalities $\succeq$ and $\succeq_*$ are with respect to a proper cone K and its dual K^*.

Definition 2 ($\theta-$normal barrier)**.** We say that ϕ is a $\theta-$normal barrier for the cone K if:

- ϕ is self-concordant on $\mathbf{int}\, K$
- ϕ is logarithmically homogeneous with parameter θ:

$$\phi(t\boldsymbol{x}) = \phi(\boldsymbol{x}) - \theta \log t \qquad \forall \boldsymbol{x} \in \mathbf{int}\, K,\ t > 0.$$

Example:
A normal barrier ϕ can be interpreted as a kind of generalization of the logarithm defined over K. For example, the following functions are normal barriers for the main cones studied in this lecture:

- $\phi(\boldsymbol{x}) = \sum -\log x_i$ is a normal barrier for $K = \mathbb{R}^m_+$, with $\theta = m$.
- $\phi(\boldsymbol{x}) = -\log(t^2 - \|\boldsymbol{x}\|_2^2)$ is a normal barrier for the Lorentz cone $K = \mathbb{L}^m_+ := \{(\boldsymbol{x}, t) \in \mathbb{R}^{m-1} \times \mathbb{R} : \|\boldsymbol{x}\|_2 \leq t\}$, with $\theta = 2$. #1
- $\phi(X) = -\log\det X$ is a normal barrier for $K = \mathbb{S}^m_+$, with $\theta = m$.
- $\phi(x, y, z) = -\log(y \log z/y - x) - \log z - \log y$ is a normal barrier for the exponential cone $K_{\exp} = \mathrm{cl}\{(x, y, z) |\ y > 0,\ ye^{x/y} \leq z\}$, with $\theta = 3$.

Also, it is easy to see that if ϕ_1 and ϕ_2 are normal barriers for the cones K_1 and K_2, of respective parameters θ_1 and θ_2, then $\phi : \mathbf{int}\, K_1 \times \mathbf{int}\, K_2 \to \mathbb{R}$, $\phi(\boldsymbol{x}, \boldsymbol{y}) = \phi_1(\boldsymbol{x}) + \phi_2(\boldsymbol{y})$ is a $(\theta_1 + \theta_2)-$normal barrier for the cone $K_1 \times K_2$. Further properties of calculus with $\theta-$normal barriers follow:

Proposition 6. *Let ϕ be a $\theta-$normal barrier function for a proper cone K. Let $t > 0$, $\boldsymbol{x} \in \mathbf{int}\, K$. Then,*

i. $\nabla\phi(t\boldsymbol{x}) = \frac{1}{t}\nabla\phi(\boldsymbol{x})$;

ii. $\boldsymbol{x}^T\nabla\phi(\boldsymbol{x}) = -\theta$;

iii. $\nabla\phi(\boldsymbol{x})^T\nabla^2\phi(x)^{-1}\nabla\phi(\boldsymbol{x}) = \theta$

iv. $-\nabla\phi(\boldsymbol{x}) \in \operatorname{\mathbf{int}} K^*$;

Proof. *i.* We differentiate the equality $\phi(t\boldsymbol{x}) = \phi(\boldsymbol{x}) - \theta\log(t)$ with respect to $\boldsymbol{x}$. This gives $t\nabla\phi(t\boldsymbol{x}) = \nabla\phi(\boldsymbol{x})$.

ii. Now we differentiate with respect to t, which gives $\boldsymbol{x}^T\nabla\phi(t\boldsymbol{x}) = -\theta/t$. The desired equality is obtained by setting $t = 1$.

iii. We differentiate the equality of i with respect to t: $\nabla^2\phi(t\boldsymbol{x})\boldsymbol{x} = -\frac{1}{t^2}\nabla\phi(\boldsymbol{x})$. At $t = 1$, this gives

$$\nabla^2\phi(\boldsymbol{x})\boldsymbol{x} = -\nabla\phi(\boldsymbol{x}).$$

Combining this with the property *ii.*, we get $\boldsymbol{x}^T\nabla^2\phi(\boldsymbol{x})\boldsymbol{x} = \theta$, and then $\boldsymbol{x}^T\nabla^2\phi(\boldsymbol{x})\boldsymbol{x} = \nabla\phi(\boldsymbol{x})^T(\nabla^2\phi(\boldsymbol{x}))^{-1}\nabla\phi(\boldsymbol{x})$.

iv. We use the first order condition of convexity for ϕ at points $t\boldsymbol{x}$ and $\boldsymbol{y} \in \operatorname{\mathbf{int}} K$:

$$\phi(\boldsymbol{y}) \geq \phi(t\boldsymbol{x}) + \langle\nabla\phi(t\boldsymbol{x}), \boldsymbol{y} - t\boldsymbol{x}\rangle.$$

By using the properties i and ii above, this gives

$$\phi(\boldsymbol{y}) \geq \phi(\boldsymbol{x}) - \theta\log(t) + \frac{1}{t}\nabla\phi(\boldsymbol{x})^T\boldsymbol{y} + \theta.$$

This implies $\nabla\phi(\boldsymbol{x})^T\boldsymbol{y} < 0$, as otherwise $-\theta\log(t) + \frac{1}{t}\nabla\phi(\boldsymbol{x})^T\boldsymbol{y}$ would tend to $+\infty$ as $t \to 0$, a contradiction. Hence, we have: $\forall\boldsymbol{y} \in \operatorname{\mathbf{int}} K, -\nabla\phi(\boldsymbol{x})^T\boldsymbol{y} > 0$, that is, $-\nabla\phi(\boldsymbol{x}) \in K^*$. To conclude, assume (by contradiction) that $\nabla\phi(\boldsymbol{x})$ lies on the boundary of $-K^*$. Then, there exists $\boldsymbol{v} \neq \boldsymbol{0}$ such that $\nabla\phi(\boldsymbol{x}) + t\boldsymbol{v} \notin -K^*, \forall t > 0$. But we know that $\nabla^2\phi(\boldsymbol{x}) \succ 0$, so we can define $\boldsymbol{u} := \nabla^2\phi(\boldsymbol{x})^{-1}\boldsymbol{v}$. Then, the first order expansion of $\nabla\phi$ yields

$$\nabla\phi(\boldsymbol{x} + t\boldsymbol{u}) = \nabla\phi(\boldsymbol{x}) + t\nabla^2\phi(\boldsymbol{x})\boldsymbol{u} + o(t) = \nabla\phi(\boldsymbol{x}) + t(\boldsymbol{v} + o(1)).$$

The right hand side does not belong to $-K^*$ for $t > 0$ small enough, and the existence of a $t > 0$ such that $\nabla\phi(\boldsymbol{x} + t\boldsymbol{u}) \notin -K^*$ is a contradiction. □

Given a $\theta-$normal barrier for the cone K, we can show that $\psi(\boldsymbol{x}) := \phi(A\boldsymbol{x} - \boldsymbol{b})$ is self-concordant on $\{\boldsymbol{x} : A\boldsymbol{x} \succ \boldsymbol{b}\}$. Then, to solve Problem (P), we use the path-following algorithm along the central path $\{\boldsymbol{x}^*(t) : t \geq 0\}$, where $\boldsymbol{x}^*(t)$ solves the equality-constrained problem

$$\begin{aligned} \min \quad & t\,c^T\boldsymbol{x} + \phi(A\boldsymbol{x} - \boldsymbol{b}) \\ s.t. \quad & F\boldsymbol{x} = \boldsymbol{g}. \end{aligned} \tag{P_t}$$

We want to prove an analog of Proposition 5 for the case of a conic programming problem with a normal barrier. This time, we will obtain a bound on suboptimality based on the parameter of the barrier ϕ: the parameter θ plays the role of the number m of inequalities in the general case.

Proposition 7. *Let $\boldsymbol{x}^*(t)$ solve Problem (P_t). Then, $\boldsymbol{x}^*(t)$ is feasible for Problem (P), and*

$$\boldsymbol{c}^T\boldsymbol{x}^*(t) - p^* \leq \frac{\theta}{t}.$$

Proof. We consider a point $\boldsymbol{x}^*(t)$ on the central path, associated with a Lagrange multiplier $\boldsymbol{\mu}^*(t)$ for the equality constraint $F\boldsymbol{x} = \boldsymbol{g}$. The KKT optimality conditions for Problem (P_t) read

$$\begin{aligned} t\boldsymbol{c} + A^T\nabla\phi(A\boldsymbol{x}^*(t) - \boldsymbol{b}) + F^T\boldsymbol{\mu}^*(t) = \boldsymbol{0}, &\qquad \text{[stationarity]};\\ F\boldsymbol{x}^*(t) = \boldsymbol{g}, &\qquad \text{[primal feasibility]}. \end{aligned}$$

Now, define $\boldsymbol{y} = -\frac{1}{t}\boldsymbol{\mu}^*(t),\ \boldsymbol{z} = -\frac{1}{t}\nabla\phi(A\boldsymbol{x}^*(t) - \boldsymbol{b})$. We will show that $(\boldsymbol{y}, \boldsymbol{z})$ is feasible for (D). Indeed, dividing the stationarity equation by t yields:

$$\boldsymbol{c} - A^T\boldsymbol{z} - F^T\boldsymbol{y} = \boldsymbol{0},$$

and $\boldsymbol{z} \succ_* \mathbf{0}$ follows from $A\boldsymbol{x}^*(t) - \boldsymbol{b} \succ \mathbf{0}$ and the property iv. of Proposition 6. Hence, we obtain a dual bound for p^*:

$$\begin{aligned} p^* \geq d^* \geq \boldsymbol{g}^T\boldsymbol{y} + \boldsymbol{b}^T\boldsymbol{z} &= \langle F\boldsymbol{x}^*(t), \boldsymbol{y}\rangle + \boldsymbol{b}^T\boldsymbol{z}. \\ &= \langle F\boldsymbol{x}^*(t), \boldsymbol{y}\rangle + \langle \boldsymbol{x}^*(t), A^T\boldsymbol{z}\rangle - \langle \boldsymbol{x}^*(t), A^T\boldsymbol{z}\rangle + \boldsymbol{b}^T\boldsymbol{z} \\ &= \langle \boldsymbol{x}^*(t), F^T\boldsymbol{y} + A^T\boldsymbol{z}\rangle - \langle \boldsymbol{z}, A\boldsymbol{x}^*(t) - \boldsymbol{b}\rangle, \end{aligned}$$

where the first equality follows from primal feasibility. Then, the first term above is equal to $\boldsymbol{c}^T\boldsymbol{x}^*$, and the second term is

$$-\boldsymbol{z}^T(A\boldsymbol{x}^* - \boldsymbol{b}) = \frac{1}{t}(A\boldsymbol{x}^* - \boldsymbol{b})^T\nabla\phi(A\boldsymbol{x}^* - \boldsymbol{b}) = -\frac{\theta}{t},$$

by property ii. of Proposition 6. This shows: $p^* \geq \boldsymbol{c}^T\boldsymbol{x}^* - \frac{\theta}{t}$, as claimed. □

To prove polynomial-time convergence of the path-following method, the idea is to take *short-steps*, i.e., we set $t_{k+1} = \omega t_k$ at iteration $k+1$, where ω is sufficiently small, so $\boldsymbol{x}^*(t)$ lies in the quadratic convergence region of Problem $(P_{\omega t})$. Then, 5 or 6 Newton steps are sufficient to compute the point $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^*(t_{k+1})$. A parameter ω which satisfies this property is $\omega = 1 + \frac{1}{4\sqrt{\theta}}$.

Proposition 8. *Let $\boldsymbol{x}^*(t)$ be a solution of (P_t), and let $\omega = 1 + \frac{1}{4\sqrt{\theta}}$. Then $\boldsymbol{x}^*(t)$ is in the region of quadratic convergence for problem $(P_{\omega t})$, that is, $\lambda_{\omega t}(\boldsymbol{x}^*(t)) \leq \frac{1}{4}$, where λ_τ is the Newton decrement for Problem (P_τ).*

Proof. It suffices to prove the result for unconstrained problems, as every constrained problem can be reduced to an unconstrained problem, by making a change of variable, and the Newton method for equality constrained problems produces the same iterates as the Newton method for the reduced unconstrained problem (cf. Remark 4). Hence, the central point $\boldsymbol{x}^*(t)$ satisfies $\nabla f_t(\boldsymbol{x}^*(t)) = 0$, where $f_t(\boldsymbol{x}) = t\boldsymbol{c}\boldsymbol{x} + \psi(\boldsymbol{x})$ and $\psi(\boldsymbol{x}) = \phi(A\boldsymbol{x} - \boldsymbol{b})$, that is,

$$\nabla f_t(\boldsymbol{x}^*(t)) = t\boldsymbol{c} + \nabla\psi(\boldsymbol{x}^*(t)) = 0.$$

Then, recall that the Newton decrement for Problem (P_τ) at $\boldsymbol{x}$ can be expressed as $\lambda_\tau(\boldsymbol{x}) = \Big(\nabla f_\tau(\boldsymbol{x})\, \nabla^2 f_\tau(\boldsymbol{x})^{-1}\, \nabla f_\tau(\boldsymbol{x})\Big)^{1/2}$. So, for all $\omega > 1$ it holds:

$$\begin{aligned} \lambda_{\omega t}(\boldsymbol{x}^*(t)) &= \Big(\, \big(\omega t\boldsymbol{c} + \nabla\psi(\boldsymbol{x}^*(t))\big)^T\, \nabla^2\psi(\boldsymbol{x}^*(t))^{-1}\, \big(\omega t\boldsymbol{c} + \nabla\psi(\boldsymbol{x}^*(t))\big) \, \Big)^{1/2} \\ &= \|\omega t\boldsymbol{c} + \nabla\psi(\boldsymbol{x}^*(t))\|_*, \end{aligned}$$

where $\|\boldsymbol{u}\|_*^2 = \boldsymbol{u}^T\nabla^2\psi(\boldsymbol{x}^*(t))^{-1}\boldsymbol{u}$.

Now, we claim the following: $\|\nabla\psi(\boldsymbol{x}^*(t))\|_* \leq \Big(\nabla\phi(\boldsymbol{s})^T\nabla^2\phi(\boldsymbol{s})^{-1}\nabla\phi(\boldsymbol{s})\Big)^{1/2} = \sqrt{\theta}$, where $\boldsymbol{s} = A\boldsymbol{x}^*(t) - \boldsymbol{b}$. The last equality is nothing but property iii. of Proposition 6. To prove the inequality, we observe that $\nabla\psi(\boldsymbol{x}^*(t)) = A^T\nabla\phi(\boldsymbol{s})$ and $\nabla^2\psi(\boldsymbol{x}^*(t)) = A^T\nabla^2\phi(\boldsymbol{s})A$. So, if we set $\boldsymbol{v} := \nabla\phi(\boldsymbol{s})$ and $H = \nabla^2\phi(\boldsymbol{s}) \succ 0$, the inequality to show is equivalent to

$$\boldsymbol{v}^T A(A^T H A)^{-1} A^T \boldsymbol{v} \leq \boldsymbol{v}^T H^{-1}\boldsymbol{v}.$$

In fact, this is true for all vectors $\boldsymbol{v}$, as we can show the linear matrix inequality $A(A^THA)^{-1}A^T \preceq H^{-1}$ using a Schur complement:

$$\begin{bmatrix} H^{-1} & A \\ A^T & A^THA \end{bmatrix} = \begin{bmatrix} H^{-1/2} \\ A^TH^{1/2} \end{bmatrix} \begin{bmatrix} H^{-1/2} \\ A^TH^{1/2} \end{bmatrix}^T \succeq 0 \iff H^{-1} - A(A^THA)^{-1}A^T \succeq 0.$$

This proves our claim: $\|\nabla\psi(\boldsymbol{x}^*(t))\|_* \leq \sqrt{\theta}$.

Now, we have

$$\begin{aligned}
\lambda_{\omega t}(\boldsymbol{x}^*(t)) &= \|\omega t\boldsymbol{c} + \nabla\psi(\boldsymbol{x}^*(t))\|_* \\
&= \left\| \omega \underbrace{(t\boldsymbol{c} + \nabla\psi(\boldsymbol{x}^*(t)))}_{=0} - (\omega - 1)\nabla\psi(\boldsymbol{x}^*(t)) \right\|_* \\
&= (\omega - 1)\|\nabla\psi(\boldsymbol{x}^*(t))\|_* \\
&\leq (\omega - 1)\sqrt{\theta}.
\end{aligned}$$

Finally, for the value $\omega = 1 + \frac{1}{4\sqrt{\theta}}$, it holds:

$$\lambda_{\omega t}(\boldsymbol{x}^*(t)) \leq \frac{1}{4\sqrt{\theta}}\sqrt{\theta} = \frac{1}{4},$$

so $\boldsymbol{x}^*(t)$ is in the region of quadratic convergence for Problem $(P_{\omega t})$. □

To summarize, (assuming we are given a point $\boldsymbol{x}^*(t_0)$ on the central path), the algorithm works as follows: For $k = 1, 2, \ldots,$

- $t_k = (1 + \frac{1}{4\sqrt{\theta}})t_{k-1}$;
- Compute $\boldsymbol{x}^*(t_k)$ by using a few (say ≤ 6) Newton steps to solve Problem (P_{t_k}), starting at $\boldsymbol{x}^*(t_{k-1})$;
- Stop if $\frac{\theta}{t_k} \leq \epsilon$

Hence, the total number of Newton steps needed to solve Problem (P) within tolerance ϵ is:

$$N_{\text{newton}} = O(1)\frac{\log\theta/(\epsilon t_0)}{\log\omega} = O(1)\sqrt{\theta}\log\frac{\theta}{\epsilon t_0}.$$

Multiplying by the number of flops per Newton step, we obtain a polynomial-time worst-case complexity result.

Theorem 9. *If a point $\boldsymbol{x}^*(t_0)$ on the central path is given, the above path following algorithm computes an $\epsilon-$suboptimal solution to Problem (P) in time polynomial with respect to the input size and* $\log\frac{1}{\epsilon}$.

Again, in practice the value $\omega = (1 + \frac{1}{4\sqrt{\theta}})$ is too small to obtain a quick convergence, this is just a convenient value for the theoretical analysis of the algorithm. There are many variants of the path following method presented here. In particular,

- The analysis can be done by using *inexact centering*, that is, it is sufficient to ensure that the point $\boldsymbol{x}^{(k)}$ is close enough to the central path $\boldsymbol{x}^*(t_k)$, by imposing a bound on $\lambda_{t_k}(\boldsymbol{x}^{(k)})$;
- It is also possible to take larger steps, while still guaranteeing convergence. To this end, predictor-corrector algorithms linearize the central path around the current iterate;
- Last but not least, when the cone K is symmetric (in particular, for LP, SOCP and SDP), state-of-the art solvers use some primal-dual interior point methods which take Newton steps simultaneously on both the primal and the dual problem.

References

[1] Boyd, S., & Vandenberghe, L. (2004). Convex optimization. Cambridge university press. `https://stanford.edu/~boyd/cvxbook/`

[2] Vandenberghe, L. (2016). Lecture notes for EE236C, UCLA. `http://www.seas.ucla.edu/~vandenbe/ee236c.html`

[3] Nemirovski, A. (2004). Interior point polynomial time methods in convex programming. Lecture Notes ISYE 8813, Georgia Tech. `http://www2.isye.gatech.edu/~nemirovs/Lect_IPM.pdf`