# Chapter 8

# Interior Point Methods

This chapter on *path-following interior point methods* is loosely based on Chapter 5.3 "*Convex Optimization: Algorithms and Complexity*" by Bubeck[1] but less abstract and including essentially all proofs. We will describe a polynomial time algorithm that solves the optimization problem $\min\{\boldsymbol{c}^T\boldsymbol{x} \mid \boldsymbol{x} \in P\}$ where $P = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}\}$ is a polytope with a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$ and $\boldsymbol{c} \in \mathbb{R}^n$ are vectors. The main result in a simplified form is as follows:

**Theorem 8.1.** *For $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$ and $\boldsymbol{c} \in \mathbb{R}^n$ with $m \leq O(n)$ one can solve the LP $\min\{\boldsymbol{c}^T\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}\}$ in time $O(n^{3.5}L)$, where $L$ is the number of bits needed to represent $\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{c}$.*

We assume that $P$ is a full-dimensional polytope and hence the *interior* $\mathrm{int}(P) := \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} < \boldsymbol{b}\}$ is non-empty. We also assume that we know at least one point in $\mathrm{int}(P)$. Let $s_i(\boldsymbol{x}) := b_i - \boldsymbol{A}_i\boldsymbol{x}$ be the *slack* that $\boldsymbol{x}$ has with respect to the $i$th constraint. It is not further important, but if one likes, one can normalized the rows so that $\|\boldsymbol{A}_i\|_2 = 1$ for $i = 1, \ldots, m$ and then $s_i(\boldsymbol{x})$ gives the geometric distance of $\boldsymbol{x}$ to the $i$th hyperplane. For some parameter $t \geq 0$ consider the convex *log-barrier function*

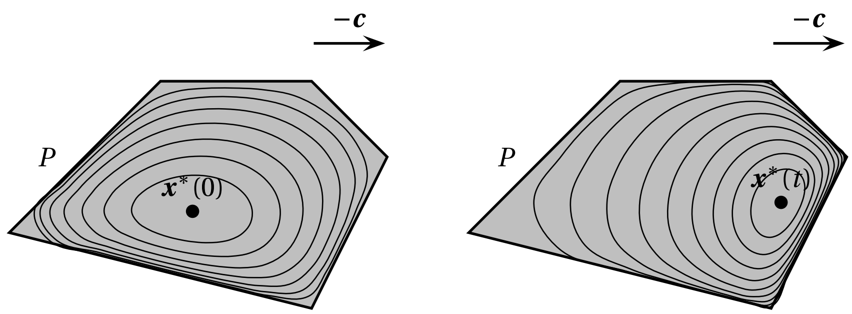
$$F_t(\boldsymbol{x}) := t \cdot \boldsymbol{c}^T\boldsymbol{x} + \sum_{i=1}^m \ln\Big(\frac{1}{s_i(\boldsymbol{x})}\Big).$$

Later we will occasionally use $F_0(\boldsymbol{x})$ which is only the log-barrier term. We define

$$\boldsymbol{x}^*(t) := \mathrm{argmin}\{F_t(\boldsymbol{x}) \mid \boldsymbol{x} \in \mathbb{R}^n\}$$

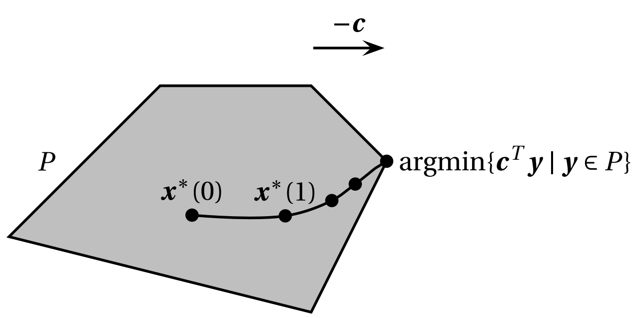
as the unique minimizer, where we interpret $F_t(\boldsymbol{x}) = \infty$ if $\boldsymbol{x} \notin \mathrm{int}(P)$. It is not hard to imagine that if $t \to \infty$, more weight is put on the linear term and $\boldsymbol{x}^*(t)$ will converge to the optimum solution of $\min\{\boldsymbol{c}^T\boldsymbol{x} \mid \boldsymbol{x} \in P\}$.

[1] see `https://arxiv.org/abs/1405.4980`




level curves for $F_t$ for $t = 0$ level curves for $F_t$ for $t \gg 0$

The points $\{\boldsymbol{x}^*(t)\}_{t\geq 0}$ define a curve inside $P$ that is called the *central path*.




In fact, the interior point method will approximately follow the central path to converge to the optimum. A simple calculation yields the *gradient* of the barrier function as

$$\nabla F_t(\boldsymbol{x}) = t \cdot \boldsymbol{c} - \sum_{i=1}^{m} \frac{\boldsymbol{A}_i}{s_i(\boldsymbol{x})}$$

Moreover, the *Hessian* of the barrier function as

$$\nabla^2 F_t(\boldsymbol{x}) = \sum_{i=1}^{m} \frac{\boldsymbol{A}_i \boldsymbol{A}_i^T}{s_i(\boldsymbol{x})^2}$$

Observe that the Hessian is an $n \times n$ symmetric, positive definite matrix; in our case it is independent of $t$ and $\boldsymbol{c}$. We define the *Dikin ellipsoid* of radius $R$ around $\boldsymbol{x}$ as

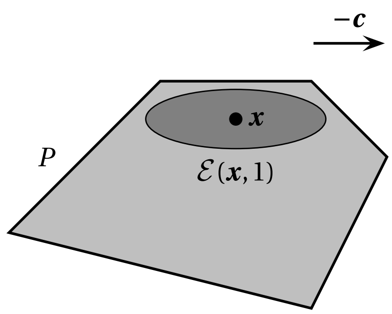
$$\mathcal{E}(\boldsymbol{x}, R) := \left\{\boldsymbol{y} \in \mathbb{R}^n \mid \sum_{i=1}^{m} \frac{(s_i(\boldsymbol{y}) - s_i(\boldsymbol{x}))^2}{s_i(\boldsymbol{x})^2} \leq R^2\right\} = \left\{\boldsymbol{y} \in \mathbb{R}^n \mid (\boldsymbol{y} - \boldsymbol{x})^T [\nabla^2 F_t(\boldsymbol{x})](\boldsymbol{y} - \boldsymbol{x}) \leq R^2\right\}.$$

As we can see, the Hessian of $F_t$ is also the matrix that defines the ellipsoid.

**Lemma 8.2.** *For any $\boldsymbol{x} \in int(P)$ one has $\mathcal{E}(\boldsymbol{x}, 1) \subseteq P$.*

*Proof.* Let $\boldsymbol{y} \in \mathcal{E}(\boldsymbol{x}, 1)$. Then in particular $(s_i(\boldsymbol{y}) - s_i(\boldsymbol{x}))^2 \leq s_i(\boldsymbol{x})^2$ which implies $s_i(\boldsymbol{y}) \geq 0$ for all $i$. $\square$

This means that starting at a point $\boldsymbol{x}$, it is safe to move to another point $\boldsymbol{x}' \in \mathcal{E}(\boldsymbol{x}, 1)$ without the danger of leaving $P$.




Geometrically one can see that if $\boldsymbol{x}$ is very close to the $i$th boundary constraint, then the ellipsoid $\mathcal{E}(\boldsymbol{x}, 1)$ is getting very thin in direction $\boldsymbol{A}_i$.

## 8.1 The algorithm

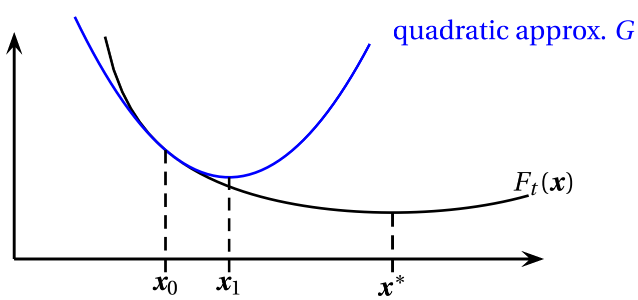
Suppose we have a starting point $\boldsymbol{x}_0 \in \text{int}(P)$ and some fixed $t$ and we try to move closer to the current optimum $\boldsymbol{x}^* := \boldsymbol{x}^*(t)$. Consider the *quadratic approximation*

$$G(\boldsymbol{x}) = F_t(\boldsymbol{x}_0) + (\nabla F_t(\boldsymbol{x}_0))^T(\boldsymbol{x} - \boldsymbol{x}_0) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_0)[\nabla^2 F_t(\boldsymbol{x}_0)](\boldsymbol{x} - \boldsymbol{x}_0)$$

at this point. Of course, we can obtain an explicit optimum solution for any quadratic function. The *first order optimality condition* tells us that the minimizer $\boldsymbol{x}_1$ of $G$ satisfies

$$\nabla G(\boldsymbol{x}_1) = \nabla F_t(\boldsymbol{x}_0) + [\nabla^2 F_t(\boldsymbol{x}_0)](\boldsymbol{x}_1 - \boldsymbol{x}_0) \stackrel{!}{=} \boldsymbol{0} \quad \Rightarrow \quad \boldsymbol{x}_1 = \boldsymbol{x}_0 - [\nabla^2 F_t(\boldsymbol{x}_0)]^{-1}(\nabla F_t(\boldsymbol{x}_0))$$

A 1-dimensional visualization would like like this:




Replacing $\boldsymbol{x}_0$ by the point $\boldsymbol{x}_1$ that minimizes the quadratic approximation is also called a *Newton step*. It turns out that the distance to the optimum point $\boldsymbol{x}^*$ decreases quadratically if the starting point $\boldsymbol{x}_0$ is close enough to $\boldsymbol{x}^*$. Back to our interior point method, this means that applying a Newton iteration to a point $\boldsymbol{x} \in \mathcal{E}(\boldsymbol{x}^*(t), R)$ moves it closer to $\boldsymbol{x}^*(t)$ in terms of the Dikin radius, assuming that $R$ was small enough. Once the current point is close enough to $\boldsymbol{x}^*(t)$ we can then increase the value of $t$ by a factor $1 + \Theta(\frac{1}{\sqrt{m}})$. The full algorithm is as follows:

**Path Following Interior Point Method**

- **Input:** LP $\min\{\boldsymbol{c}^T\boldsymbol{x} \mid \boldsymbol{x} \in P\}$ and $\boldsymbol{x}_0 \in \mathcal{E}(\boldsymbol{x}^*(t_0), \frac{1}{12})$ for some $t_0 > 0$.
- **Output:** Sequence $\{\boldsymbol{x}_k\}_{k\geq 0}$ converging to $\boldsymbol{y}^* = \mathrm{argmin}\{\boldsymbol{c}^T\boldsymbol{y} \mid \boldsymbol{y} \in P\}$

(1) FOR $k = 0$ TO $\infty$ DO

  (2) Perform Newton Step $\boldsymbol{x}_{k+1} := \boldsymbol{x}_k - [\nabla^2 F_{t_k}(\boldsymbol{x}_k)]^{-1}(\nabla F_{t_k}(\boldsymbol{x}_k))$

  (3) Update $t_{k+1} := t_k \cdot (1 + \frac{1}{100\sqrt{m}})$

The analysis breaks down in the following main steps:

(1) First we will analyze the Newton step and prove that $\boldsymbol{x}_k \in \mathcal{E}(\boldsymbol{x}^*(t_k), R) \Rightarrow \boldsymbol{x}_{k+1} \in \mathcal{E}(\boldsymbol{x}^*(t_k), 6R^2)$. Setting $R = \frac{1}{12}$ means that $\boldsymbol{x}_{k+1} \in \mathcal{E}(\boldsymbol{x}^*(t_k), \frac{1}{24})$. We could in fact iterate the Newton step to get arbitrarily close to $\boldsymbol{x}^*(t_k)$ — but that would not give an asymptotic advantage.

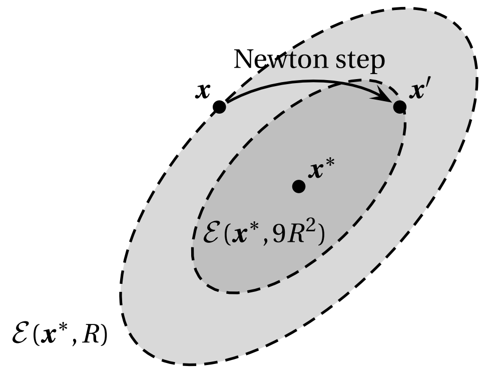
(2) For step (3), we need to show that the point on the central curve does not move too quickly when increasing the parameter $t$. In fact we show that $\boldsymbol{x}^*(t \cdot (1+\varepsilon)) \in \mathcal{E}(\boldsymbol{x}^*(t), \varepsilon\sqrt{m})$. Setting $\varepsilon := \frac{1}{200\sqrt{m}}$ is enough to guarantee that $\boldsymbol{x}^*(t_{k+1}) \in \mathcal{E}(\boldsymbol{x}^*(t_k), \frac{1}{25})$. As the ellipsoids change only slowly, one has $\left(\boldsymbol{x}_{k+1} \in \mathcal{E}(\boldsymbol{x}^*(t_k), \frac{1}{24}) \;\&\; \boldsymbol{x}^*(t_{k+1}) \in \mathcal{E}(\boldsymbol{x}^*(t_k), \frac{1}{25})\right) \Rightarrow \boldsymbol{x}_{k+1} \in \mathcal{E}(\boldsymbol{x}^*(t_{k+1}), \frac{1}{12})$.

(3) We did not describe yet how to obtain a starting point $\boldsymbol{x}_0$ that is close enough to $\boldsymbol{x}^*(t_0)$ for some $t_0 > 0$. It turns out that one can run a "reverse path following algorithm" move from any point $\boldsymbol{x} \in \mathrm{int}(P)$ close to the analytic center.

(4) Finally we can bound the distance from any intermediate point to the optimum $\boldsymbol{y}^* := \mathrm{argmin}\{\boldsymbol{c}^T\boldsymbol{y} \mid \boldsymbol{y} \in P\}$ by proving that $\boldsymbol{x} \in \mathcal{E}(\boldsymbol{x}^*(t), 1) \Rightarrow \boldsymbol{c}^T\boldsymbol{x} \leq \boldsymbol{c}^T\boldsymbol{y}^* + O(\frac{m}{t})$.

## 8.2 Analysis of a Newton Step

For a symmetric matrix $\boldsymbol{H} \in \mathbb{R}^{n\times n}$ with Eigen decomposition $\boldsymbol{H} = \sum_{i=1}^n \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^T$, let $|\boldsymbol{H}| := \sum_{i=1}^n |\lambda_i| \boldsymbol{u}_i \boldsymbol{u}_i^T$ be the matrix where the absolute value function has been applied to all Eigenvalues. In particular $|\boldsymbol{H}| \succeq 0$. For a $\boldsymbol{H} \in \mathbb{R}^{n\times n}$ that is not necessarily symmetric, let $\|\boldsymbol{H}\|_{\mathrm{op}}$ be the largest singular value. In particular we will use that $\|\boldsymbol{H}\boldsymbol{x}\|_2 \leq \|\boldsymbol{H}\|_{\mathrm{op}} \cdot \|\boldsymbol{x}\|_2$ for any vector $\boldsymbol{x}$. We will now show that a Newton step shrinks the distance from the optimum (if the distance is measured in the norm induced by the Hessian).




**Lemma 8.3.** *Fix a value of* $t \geq 0$ *and let* $\boldsymbol{x}^* := \boldsymbol{x}^*(t)$. *For* $\boldsymbol{x} \in \mathcal{E}(\boldsymbol{x}^*, R)$ *with* $R \leq \frac{1}{8}$, *set*

$$\boldsymbol{x}' := \boldsymbol{x} - [\nabla^2 F_t(\boldsymbol{x})]^{-1}(\nabla F_t(\boldsymbol{x}))$$

*Then* $\boldsymbol{x}' \in \mathcal{E}(\boldsymbol{x}^*, 9R^2)$.

*Proof.* We begin with proving two useful claims. First we show that the Hessian of $F_t$ only changes slowly:
**Claim I.** *For any vector* $\boldsymbol{y}$ *be on the line segment between* $\boldsymbol{x}$ *and* $\boldsymbol{x}^*$ *one has* $(1-3R) \cdot \nabla^2 F_t(\boldsymbol{x}) \preceq \nabla^2 F_t(\boldsymbol{y}) \preceq (1+3R) \cdot \nabla^2 F_t(\boldsymbol{x})$.
**Proof of claim.** We have

$$\frac{|s_i(\boldsymbol{y}) - s_i(\boldsymbol{x})|}{s_i(\boldsymbol{x})} \leq \frac{|s_i(\boldsymbol{x}^*) - s_i(\boldsymbol{x})|}{s_i(\boldsymbol{x})} = \underbrace{\frac{|s_i(\boldsymbol{x}^*) - s_i(\boldsymbol{x})|}{s_i(\boldsymbol{x}^*)}}_{\leq R} \cdot \underbrace{\frac{s_i(\boldsymbol{x}^*)}{s_i(\boldsymbol{x})}}_{\leq 1+R \leq \frac{9}{8}} \leq \frac{9}{8}R$$

Inverting and squaring the inequality $(1-\frac{9}{8}R)s_i(\boldsymbol{x}) \leq s_i(\boldsymbol{y}) \leq (1+\frac{9}{8}R)s_i(\boldsymbol{x})$ gives the claim since $\frac{1}{(1+\frac{9}{8}R)^2} \geq 1-3R$ and $\frac{1}{(1-\frac{9}{8}R)^2} \leq 1+3R$ for $R \leq \frac{1}{8}$. □
**Claim II.** *One can write* $\nabla F_t(\boldsymbol{x}) = ([\nabla^2 F_t(\boldsymbol{x})] + \boldsymbol{E})(\boldsymbol{x} - \boldsymbol{x}^*)$ *where* $|\boldsymbol{E}| \preceq 3R \cdot \nabla^2 F_t(\boldsymbol{x})$.
**Proof of claim.** We apply the fundamental theorem of calculus to obtain

$$\nabla F_t(\boldsymbol{x}) = \nabla F_t(\boldsymbol{x}) - \underbrace{\nabla F_t(\boldsymbol{x}^*)}_{=\boldsymbol{0}} = \underbrace{\Big[\int_0^1 \nabla^2 F_t(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{x}^*)]d\lambda\Big]}_{=:\nabla^2 F_t(\boldsymbol{x}) + \boldsymbol{E}}(\boldsymbol{x} - \boldsymbol{x}^*)$$

with $|\boldsymbol{E}| \preceq 3R \cdot \nabla^2 F_t(\boldsymbol{x})$ as we can derive from Claim I. □

Now we can write

$$\begin{aligned} \boldsymbol{x}' - \boldsymbol{x}^* &\overset{\text{Def. } \boldsymbol{x}'}{=} (\boldsymbol{x} - \boldsymbol{x}^*) - [\nabla^2 F_t(\boldsymbol{x})]^{-1}(\nabla F_t(\boldsymbol{x})) \qquad (*) \\ &\overset{\text{Claim II}}{=} (\boldsymbol{x} - \boldsymbol{x}^*) - [\nabla^2 F_t(\boldsymbol{x})]^{-1}([\nabla^2 F_t(\boldsymbol{x})] + \boldsymbol{E})(\boldsymbol{x} - \boldsymbol{x}^*) \\ &= -[\nabla^2 F_t(\boldsymbol{x})]^{-1}\boldsymbol{E}(\boldsymbol{x} - \boldsymbol{x}^*) \end{aligned}$$

**Claim III.** *One has* $\boldsymbol{x}' \in \mathcal{E}(\boldsymbol{x}^*, 9R^2)$.
**Proof of claim.** Instead of a careful (and annoying) calculation with matrices that have

different Eigenvectors, we can use another trick. We apply a linear transformation to $P$ so that $\nabla^2 F_t(\boldsymbol{x}^*) = \boldsymbol{I}$. Then $\boldsymbol{x} \in \mathcal{E}(\boldsymbol{x}^*, R) \Leftrightarrow \|\boldsymbol{x}-\boldsymbol{x}^*\|_2 \le R$. Moreover, $(1-3R)\boldsymbol{I} \preceq \nabla^2 F_t(\boldsymbol{x}) \preceq (1+3R)I$ and $-6R\cdot \boldsymbol{I} \preceq \boldsymbol{E} \preceq 6R\cdot\boldsymbol{I}$. Then

$$\|\boldsymbol{x}'-\boldsymbol{x}\|_2 = \|\nabla^2 F_t(\boldsymbol{x})^{-1}\boldsymbol{E}(\boldsymbol{x}-\boldsymbol{x}^*)\|_2 \le \underbrace{\|[\nabla^2 F_t(\boldsymbol{x})]^{-1}\|_{\text{op}}}_{\le \frac{1}{1-3R}\le\frac{3}{2}} \cdot \underbrace{\|\boldsymbol{E}\|_{\text{op}}}_{\le 6R} \cdot \underbrace{\|\boldsymbol{x}-\boldsymbol{x}^*\|_2}_{\le R} \le 9R^2$$

and hence $\boldsymbol{x}' \in \mathcal{E}(\boldsymbol{x}^*, 9R^2)$. □

## 8.3 Bounding the movement of $x^*(t)$

One of the main arguments is that the parameter $t$ can be increased by a $(1+\Theta(\frac{1}{\sqrt{m}}))$-factor while the new optimum $\boldsymbol{x}^*(t')$ still lies in the Dikin ellipsoid around $\boldsymbol{x}^*(t)$.

First, we observe that around an optimum point $\boldsymbol{x}^*(t)$ the function $F_t$ is well approximated by the square of the Dikin radius. We can give a general approximation result (here the term $\pm R^3$ means that the equation holds up to an error that lies in $[-R^3, R^3]$):

**Lemma 8.4.** *Let $\boldsymbol{x} \in int(P)$ and $\boldsymbol{x}+\boldsymbol{h}$ on the boundary of $\mathcal{E}(\boldsymbol{x}, R)$ for $0 \le R \le \frac{1}{2}$. Then*

$$F_t(\boldsymbol{x}+\boldsymbol{h}) = F_t(\boldsymbol{x}) + \langle \nabla F_t(\boldsymbol{x}), \boldsymbol{h}\rangle + \frac{R^2}{2} \quad \pm R^3$$

*Proof.* We write

$$\begin{aligned}
F_t(\boldsymbol{x}+\boldsymbol{h}) - F_t(\boldsymbol{x}) &= t\cdot\boldsymbol{c}^T\boldsymbol{h} - \sum_{i=1}^m \ln\Big(\frac{s_i(\boldsymbol{x}+\boldsymbol{h})}{s_i(\boldsymbol{x})}\Big) = t\cdot\boldsymbol{c}^T\boldsymbol{h} - \sum_{i=1}^m \ln\Big(1+\frac{\langle \boldsymbol{A}_i, \boldsymbol{h}\rangle}{s_i(\boldsymbol{x})}\Big) \\
&= t\cdot\boldsymbol{c}^T\boldsymbol{h} \underbrace{- \sum_{i=1}^m \frac{\langle \boldsymbol{A}_i, \boldsymbol{h}\rangle}{s_i(\boldsymbol{x})}}_{=\langle \nabla F_t(\boldsymbol{x}), \boldsymbol{h}\rangle} + \frac{1}{2} \underbrace{\sum_{i=1}^m \frac{\langle \boldsymbol{A}_i, \boldsymbol{h}\rangle^2}{s_i(\boldsymbol{x})^2}}_{=\boldsymbol{h}^T\nabla^2 F_t(\boldsymbol{x})\boldsymbol{h}=R^2} \pm \sum_{i=1}^m \frac{|\langle \boldsymbol{A}_i, \boldsymbol{h}\rangle|^3}{s_i(\boldsymbol{x})^3} \\
&= t\cdot\boldsymbol{c}^T\boldsymbol{h} + \langle \nabla F_t(\boldsymbol{x}), \boldsymbol{h}\rangle + \frac{R^2}{2} \pm \max_{i=1,\ldots,m}\Big\{\underbrace{\frac{|\langle \boldsymbol{A}_i, \boldsymbol{h}\rangle|}{s_i(\boldsymbol{x})}}_{\le R}\Big\} \cdot \underbrace{\sum_{i=1}^m \frac{\langle \boldsymbol{A}_i, \boldsymbol{h}\rangle^2}{s_i(\boldsymbol{x})^2}}_{=R^2}
\end{aligned}$$

using that $\ln(1+z) = z - \frac{1}{2}z^2 \pm |z|^3$ for $|z| \le \frac{1}{2}$. □

For an arbitrary point $\boldsymbol{x} \in \text{int}(P)$, the function $t\cdot\boldsymbol{c}^T\boldsymbol{x}$ might vary arbitrarily over the Dikin ellipsoid $\mathcal{E}(\boldsymbol{x}, R)$. Interestingly the function can only vary by $\sqrt{m}$ if $\boldsymbol{x}$ is an optimum point to $F_t$.

**Lemma 8.5.** *One has* $\max\{t\cdot\boldsymbol{c}^T(\boldsymbol{x}-\boldsymbol{x}^*(t)) \mid \boldsymbol{x} \in \mathcal{E}(\boldsymbol{x}^*(t), R)\} \le R\sqrt{m}$.

*Proof.* Consider the ratios $r_i := \frac{s_i(\boldsymbol{x})-s_i(\boldsymbol{x}^*(t))}{s_i(\boldsymbol{x}^*(t))}$. By first order optimality

$$\nabla F_t(\boldsymbol{x}^*(t)) = t\cdot\boldsymbol{c} - \sum_{i=1}^m \frac{\boldsymbol{A}_i}{s_i(\boldsymbol{x}^*(t))} = \boldsymbol{0}$$

Multiplying this vector equation with $\boldsymbol{x} - \boldsymbol{x}^*(t)$ reveals that

$$t \cdot \boldsymbol{c}^T(\boldsymbol{x} - \boldsymbol{x}^*(t)) = \sum_{i=1}^m \frac{\boldsymbol{A}_i(\boldsymbol{x} - \boldsymbol{x}^*(t))}{s_i(\boldsymbol{x}^*(t))} = \sum_{i=1}^m r_i \leq \|\boldsymbol{r}\|_1 \overset{\boldsymbol{r}\in\mathbb{R}^m}{\leq} \sqrt{m} \cdot \underbrace{\|\boldsymbol{r}\|_2}_{\leq R} \leq R \cdot \sqrt{m}$$

□

Finally we can proof an important fact:

**Lemma 8.6.** *Fix values of $t, t' > 0$ so that $\boldsymbol{x}^*(t')$ lies on the boundary of $\mathcal{E}(\boldsymbol{x}^*(t), R)$ for $0 \leq R \leq \frac{1}{4}$. Then $\frac{t'}{t} \geq 1 + \frac{R}{4\sqrt{m}}$.*

*Proof.* As $\boldsymbol{x}^*(t')$ lies on the boundary of $\mathcal{E}(\boldsymbol{x}^*(t), R)$ we can apply Lemma 8.4 to get $F_t(\boldsymbol{x}^*(t')) = F_t(\boldsymbol{x}^*(t)) + \frac{R^2}{2} \pm R^3$ as $\nabla F_t(\boldsymbol{x}^*(t)) = \boldsymbol{0}$. Abbreviate $t' = t \cdot (1+\varepsilon)$. Then

$$0 \overset{\text{optimality}}{\leq} F_{t'}(\boldsymbol{x}^*(t)) - F_{t'}(\boldsymbol{x}^*(t')) = \varepsilon \cdot \underbrace{t \cdot \boldsymbol{c}^T(\boldsymbol{x}^*(t) - \boldsymbol{x}^*(t'))}_{\leq R\sqrt{m}} + \underbrace{(F_t(\boldsymbol{x}^*(t)) - F_t(\boldsymbol{x}^*(t')))}_{\leq -\frac{R^2}{2} + R^3 \leq -\frac{R^2}{4}} \leq \varepsilon R\sqrt{m} - \frac{R^2}{4}$$

Rearranging gives $\varepsilon \geq \frac{R}{4\sqrt{m}}$. □

## 8.4 Distance from the optimum

Finally we prove an upper bound on the optimality gap as $t$ grows:

**Lemma 8.7.** *For any $t > 0$ and $\boldsymbol{x} \in \mathcal{E}(\boldsymbol{x}^*(t), 1)$ one has $c^T\boldsymbol{x} - \min\{\boldsymbol{c}^T\boldsymbol{y} \mid \boldsymbol{y} \in P\} \leq \frac{3m}{t}$.*

*Proof.* Let $\boldsymbol{y}^*$ be the point minimizing $\boldsymbol{c}^T\boldsymbol{y}$ over $P$. The function value $\boldsymbol{c}^T\boldsymbol{x}$ differs only by an additive $\frac{\sqrt{m}}{t} \leq \frac{m}{t}$ term over points in $\mathcal{E}(\boldsymbol{x}^*(t), 1)$ as we have seen in Lemma 8.5. Hence it suffices to show that $\boldsymbol{c}^T\boldsymbol{x}^*(t) \leq \boldsymbol{c}^T\boldsymbol{y}^* + \frac{2m}{t}$.

Consider the midpoint $\boldsymbol{x}' := \frac{1}{2}\boldsymbol{y}^* + \frac{1}{2}\boldsymbol{x}^*(t)$. We know that $s_i(\boldsymbol{x}') \geq \frac{1}{2}s_i(\boldsymbol{x}^*(t))$ for each $i \in [m]$. Hence

$$\begin{aligned} 0 &\overset{\text{optimality}}{\leq} F_t(\boldsymbol{x}') - F_t(\boldsymbol{x}^*(t)) = t \cdot \underbrace{(\boldsymbol{c}^T\boldsymbol{x}' - \boldsymbol{c}^T\boldsymbol{x}^*(t))}_{=\frac{1}{2}(\boldsymbol{c}^T\boldsymbol{y}^* - \boldsymbol{c}^T\boldsymbol{x}^*(t))} + \sum_{i=1}^m \underbrace{\Big(\ln\Big(\frac{1}{s_i(\boldsymbol{x}')}\Big) - \ln\Big(\frac{1}{s_i(\boldsymbol{x}^*(t))}\Big)\Big)}_{\leq 1} \\ &\leq \frac{t}{2}(\boldsymbol{c}^T\boldsymbol{y}^* - \boldsymbol{c}^T\boldsymbol{x}^*(t)) + m \end{aligned}$$

Rearranging gives the claim. □

## 8.5 Finding the analytical center

In the Interior Point Method that we stated above, we do assume that we know the *analytical center* $\boldsymbol{x}^*(0)$ (or at least a very close point). We want to quickly argue how that center can be found — assuming that we know an arbitrary point $\boldsymbol{y} \in \text{int}(P)$. First, we define an auxiliary function

$$\tilde{F}_t(\boldsymbol{x}) := t \cdot (-\nabla F_0(\boldsymbol{y}))^T \boldsymbol{x} + \sum_{i=1}^{m} \ln\left(\frac{1}{s_i(\boldsymbol{x})}\right)$$

that differs from $F_t$ only in the linear part. Let $\tilde{\boldsymbol{x}}^*(t) := \text{argmin}\{\tilde{F}_t(\boldsymbol{x}) \mid \boldsymbol{x} \in \mathbb{R}^n\}$ be the central path with respect to $\tilde{F}_t$. In particular we have defined the linear part in $\tilde{F}_t(\boldsymbol{x})$ so that $\nabla \tilde{F}_1(\boldsymbol{y}) = \boldsymbol{0}$, which implies that $\tilde{\boldsymbol{x}}^*(1) = \boldsymbol{y}$. The 2nd trick is that the analysis (in particular Lemma 8.6) also applies if we *decrease* $t$ by a factor $1 - \frac{1}{100\sqrt{m}}$ and the interior point will stay close the the central path of $\tilde{F}_t$. Then for $t \to 0$ we obtain a sequence of points that converge to the anaytical center $\tilde{\boldsymbol{x}}^*(0)$ which is also $\boldsymbol{x}^*(0)$. We can summarize the algorithm as follows:

**Interior Point Method for finding Analytical Center**

- **Input:** Polytope $P = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}\}$ and point $\boldsymbol{y} \in \text{int}(P)$
- **Output:** Sequence $\{\boldsymbol{x}_k\}_{k \geq 0}$ converging towards analytical center

(1) Set $F(\boldsymbol{x}) := \sum_{i=1}^m \ln\left(\frac{1}{s_i(\boldsymbol{x})}\right)$
(2) Set $\tilde{F}_t(\boldsymbol{x}) := t \cdot (-\nabla F(\boldsymbol{y}))^T \boldsymbol{x} + F(\boldsymbol{x})$.
(3) Set $\boldsymbol{x}_0 := \boldsymbol{y}$ and $t_0 := 1$
(4) FOR $k = 0$ TO $\infty$ DO
    (5) Perform Newton Step $\boldsymbol{x}_{k+1} := \boldsymbol{x}_k - [\nabla^2 \tilde{F}_{t_k}(\boldsymbol{x}_k)]^{-1}(\nabla \tilde{F}_{t_k}(\boldsymbol{x}_k))$
    (6) Update $t_{k+1} := t_k \cdot (1 - \frac{1}{100\sqrt{m}})$

Obviously this raises the question how to find a feasible point in $P$ after all. But one can simply solve $\min\{\lambda \mid \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} + \lambda\boldsymbol{1}\}$ which has a strictly feasible solution of $(\boldsymbol{x}, \lambda) = (\boldsymbol{0}, \|\boldsymbol{b}\|_\infty + 1)$ and an optimum solution is contained in $\text{int}(P)$, given that $\text{int}(P) \neq \emptyset$.

## 8.6 Running time

Finally we want to discuss the running time of the method in terms of number of *arithmetic operations*. Let $L$ be the *number of bits* needed to encode the linear program. Suppose that $\boldsymbol{A} \in \mathbb{Z}^{m \times n}$, $\boldsymbol{b} \in \mathbb{Z}^m$ and $\boldsymbol{c} \in \mathbb{Z}^n$, then a safe definition is $L := \sum_{i,j}(1 + \log(|A_{ij}| + 1)) + \sum_{i=1}^m (1 + \log(|b_i| + 1)) + \sum_{j=1}^n (1 + \log(|c_j| + 1))$. To keep the calculations simple, we assume that $m = \Theta(n)$ and we will express the running time in terms of $n$ and $L$.

A somewhat technical calculation shows that it suffices to find a point $\boldsymbol{x} \in \text{int}(P)$ that is within a $2^{-\Theta(L)}$ term of the optimum $\boldsymbol{y}^*$. Then one can select indices $I := \{i \in [m] \mid s_i(\boldsymbol{x}) \leq 2^{-\Theta(L)}\}$ and the projection of $\boldsymbol{x}$ onto the subspace $\{\boldsymbol{y} \in \mathbb{R}^n \mid \boldsymbol{A}_i \boldsymbol{y} = b_i \ \forall i \in I\}$ will recover an optimum solution to $\min\{\boldsymbol{c}^T \boldsymbol{y} \mid \boldsymbol{y} \in P\}$. Similarly, the starting value of $t$ can be of the form $t_0 \geq 2^{-\Theta(L)}$. Hence, the number iterations of the interior point method can be bounded by $O(\sqrt{n} \cdot L)$. Each iteration is dominated by the time that it takes to solve the *linear system* $[\nabla^2 F_{t_k}(\boldsymbol{x}_k)]\boldsymbol{y} = \nabla F_{t_k}(\boldsymbol{x}_k)$ for $\boldsymbol{y}$. Using Gaussian elimination, solving a linear system takes time $O(n^3)$ which results in a total running time of $O(n^{3.5}L)$ for solving a linear program.

On the other hand, matrices can be multiplied/inverted in time $O(n^\omega)$ where the best known value for the exponent is currently $\omega < 2.3729$. That means using *fast matrix multiplication*, linear programs can be solved in a total running time of $O(n^{\omega+1/2}L) \leq O(n^{2.8729}L)$.

However, it seems that fast matrix multiplication is not used in practice, so we want to describe a different speed-up that is based on *low rank updates*. The idea uses the basic fact that for a symmetric matrix $\boldsymbol{S} \in \mathbb{R}^{n\times n}$, a vector $\boldsymbol{v} \in \mathbb{R}^n$ and a scalar $\lambda \in \mathbb{R}$ one has the *Sherman-Morrison formula*

$$(\boldsymbol{S} + \lambda \boldsymbol{v}\boldsymbol{v}^T)^{-1} = \boldsymbol{S}^{-1} - \underbrace{\frac{\lambda}{1+\lambda \boldsymbol{v}^T\boldsymbol{S}^{-1}\boldsymbol{v}}}_{\in\mathbb{R}} \cdot \underbrace{(\boldsymbol{S}^{-1}\boldsymbol{v})(\boldsymbol{S}^{-1}\boldsymbol{v})^T}_{\text{rank-1 matrix}}$$

(assuming that both $\boldsymbol{S}$ and $\boldsymbol{S} + \lambda\boldsymbol{v}\boldsymbol{v}^T$ are invertible). In particular if $\boldsymbol{S}^{-1}$ is known, then $(\boldsymbol{S} + \lambda\boldsymbol{v}\boldsymbol{v}^T)^{-1}$ can be computed in time $O(n^2)$.

Now suppose that instead of performing a Newton step $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - [\nabla^2 F_t(\boldsymbol{x}_k)]^{-1}(\nabla F_t(\boldsymbol{x}_k))$ with the exact inverse of the Hessian, we maintain the inverse $\boldsymbol{S}_k^{-1}$ for a matrix $\boldsymbol{S}_k \in \mathbb{R}^{n\times n}$ satisfying $\frac{1}{1+R}\boldsymbol{S}_k \preceq \nabla^2 F_t(\boldsymbol{x}_k) \preceq (1+R)\boldsymbol{S}_k$. Then one can still prove the implication

$$\boldsymbol{x}_k \in \mathcal{E}(\boldsymbol{x}^*(t), R) \quad \Rightarrow \quad \boldsymbol{x}_{k+1} \in \mathcal{E}(\boldsymbol{x}^*(t), O(R^2))$$

by slightly modifying the calculations in Lemma 8.3. Then again, choosing $R > 0$ as a small enough constant suffices for our purpose. Recall that $\nabla^2 F_t(\boldsymbol{x}) = \sum_{i=1}^m \frac{1}{s_i(\boldsymbol{x})^2}\boldsymbol{A}_i\boldsymbol{A}_i^T$. The natural idea is to choose $\boldsymbol{S}_k = \sum_{i=1}^m \frac{1}{d_k(i)^2}\boldsymbol{A}_i\boldsymbol{A}_i^T$ but only update $d_{k+1}(i) := s_i(\boldsymbol{x}_{k+1})$ when the distances have changed by more than a $1 \pm \frac{R}{4}$ factor from the last update. If $\boldsymbol{S}_k^{-1}$ is known and only $q$ distances have been updated, then computing $\boldsymbol{S}_{k+1}^{-1}$ takes time $O(qn^2)$.

Now take consecutive points $\boldsymbol{x}_{k+1} \in \mathcal{E}(\boldsymbol{x}_k, R)$ and consider the ratio $r_i = \frac{s_i(\boldsymbol{x}_{k+1}) - s_i(\boldsymbol{x}_k)}{s_i(\boldsymbol{x}_k)}$. Then the amortized number of rank-1 updates caused by this Newton step are $O(\|\boldsymbol{r}\|_1) \leq O(\sqrt{n}\cdot\|\boldsymbol{r}\|_2) \leq O(\sqrt{n})$. That means the amortized time per iteration is $O(n^{2.5})$ and hence solving the linear program takes time $O(n^3 L)$ even without fast matrix multiplication.

A further improvement can be made by combining low rank updates and fast matrix multiplication. In fact, Lee and Sidford show that the amortized running time per iteration can be brought down to $\tilde{O}(n^2)$ (where the $\tilde{O}$-notation hides some lower order terms), which results in a total running time of $O(n^{2.5}L)$ to solve linear programs $\min\{\boldsymbol{c}^T\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}\}$ where $\boldsymbol{A} \in \mathbb{Z}^{O(n)\times n}$.

## 8.7 Exercises

**Exercise 1.** *Consider the cube $P := [0,1]^n = \{\boldsymbol{x} \in \mathbb{R}^n \mid 0 \leq x_i \leq 1$ for $i = 1,\dots,n\}$. Consider a sequence of points $\{\boldsymbol{x}_k\}_{k\geq 0}$ with $\boldsymbol{x}_0 = (\frac{1}{2},\dots,\frac{1}{2})$ and with the only restriction that $\boldsymbol{x}_{k+1} \in$*

$\mathcal{E}(\boldsymbol{x}_k, \frac{1}{2})$. *Prove that it takes at least* $\Omega(\sqrt{n} \cdot \log(\frac{1}{\delta}))$ *iterations until* $\boldsymbol{x}_k$ *can be within a* $\|\cdot\|_\infty$*-distance of* $\delta$ *from the vertex* $\mathbf{0}$.

**Exercise 2.** *Recall that the presented interior point method takes* $O(L\sqrt{m})$ *iterations to get within an additive* $2^{-L}$ *distance to the optimum for a polytope* $P = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}\}$ *with* $\boldsymbol{A} \in \mathbb{R}^{m \times n}$. *There is indeed a way of bringing the number of iterations down to* $O(L\sqrt{n})$. *A deep result of Nesterov and Nemirovsky says that there is a convex function* $\phi : \mathbb{R}^n \to \mathbb{R}$ *that is* self-concordant *which means it satisfies the following properties for some universal constant* $C > 0$*:*

*(A)* *For any* $0 < R \leq \frac{1}{C}$ *and* $\boldsymbol{x} \in \mathcal{E}(\boldsymbol{x}^*, R)$ *one has* $(1-2R)\nabla^2\phi(\boldsymbol{x}) \preceq \nabla^2\phi(\boldsymbol{x}^*) \preceq (1+2R)\nabla^2\phi(\boldsymbol{x})$ *where we redefine the ellipsoid* $\mathcal{E}(\boldsymbol{x}^*, R) := \{\boldsymbol{x} \in \mathbb{R}^n \mid (\boldsymbol{x}-\boldsymbol{x}^*)^T[\nabla^2\phi(\boldsymbol{x}^*)](\boldsymbol{x}-\boldsymbol{x}^*) \leq R^2\}$

*(B)* *One has* $\nabla\phi(\boldsymbol{x})\nabla\phi(\boldsymbol{x})^T \preceq Cn \cdot \nabla^2\phi(\boldsymbol{x})$ *for all* $\boldsymbol{x} \in$ *int*$(P)$.

*(C)* *If* $\boldsymbol{x} \to \partial P$, *then* $\phi(\boldsymbol{x}) \to \infty$.

*For* $t \geq 0$ *we modify the barrier function to* $F_t(\boldsymbol{x}) := t \cdot \boldsymbol{c}^T\boldsymbol{x} + \phi(\boldsymbol{x})$. *Prove the following (where the* $O$*-notation is allowed to hide dependence on* $C$*):*

*(1)* *Show that for* $\boldsymbol{x} \in \mathcal{E}(\boldsymbol{x}^*, R)$ *with* $\boldsymbol{x}^* := \boldsymbol{x}^*(t)$ *and* $\boldsymbol{x}' := \boldsymbol{x} - [\nabla^2 F_t(\boldsymbol{x})]^{-1}\nabla F_t(\boldsymbol{x})$ *one has* $\boldsymbol{x}' \in \mathcal{E}(\boldsymbol{x}^*, O(R^2))$ *assuming* $R$ *is small enough.*

*(2)* *One has* $\max\{t \cdot \boldsymbol{c}^T(\boldsymbol{x} - \boldsymbol{x}^*(t)) \mid \boldsymbol{x} \in \mathcal{E}(\boldsymbol{x}^*(t), R)\} \leq O(R\sqrt{n})$ *for all* $t > 0$ *and* $R > 0$ *small enough.*