19
The Centroid and Ellipsoid Algorithms

Our focus in this chapter is on the constrainted optimization problem:

> Given a convex function f, a convex set K, and a parameter $\varepsilon > 0$, find a point $\hat{x} \in K$ such that
>
> $$f(\hat{x}) \leq \min_{x \in K} f(x) + \varepsilon.$$

In previous sections, we saw gradient descent and mirror descent gave us algorithms whose dependence on ε was like $\text{poly}(1/\varepsilon)$. Moreover, we have examples that show algorithms based only on *local* gradient information need time at least polynomial in $1/\varepsilon$. Where? So can we do better?

The algorithms also had some dependence on f and K; e.g., if gradients $\|\nabla f(x)\|_2 \leq G$ for $x \in K$, and the diameter of K was at most D, then projected gradient descent ran in $O((GD/\varepsilon)^2)$ time.

In this chapter, we show how to use global information to get algorithms for convex programming that have $O(\log 1/\varepsilon)$-type convergence guarantees (under suitable assumptions). Specifically, we will examine the Centroid and Ellipsoid algorithms in depth. In turn, these will give us polynomial-time algorithms for Linear Programming problems.

19.1 The Centroid Algorithm

In this section, we discuss the Centroid Algorithm in the context of constrained convex minimization. Besides being interesting in its own right, it is a good lead-in to Ellipsoid, since it gives some intuition about high-dimensional bodies and their volumes.

Given a convex body $K \subseteq \mathbb{R}^n$ and a convex function $f : K \to \mathbb{R}$, we want to approximately minimize $f(x)$ over $x \in K$. As in previous sections, we assume a gradient oracle for f, one that returns the value $\nabla f(x)$ for any query point $x \in K$. We also assume that we can perform exact arithmetic over the reals; however, we will soon begin discussing issues that arise from using only finite-precision arithmetic.

As the name suggests, the algorithm is based on the notion of *centroid* for compact convex sets. The centroid of a set K is the point $c \in \mathbb{R}^n$ such that

$$c := \frac{\int_{x \in K} x\,dx}{\mathrm{vol}(K)} = \frac{\int_{x \in K} x\,dx}{\int_{x \in K} dx},$$

where $\mathrm{vol}(K)$ is the volume of the set K. Since c is the "average" of points in some convex set K, it also lies within K. The following result captures the crucial fact about the centroid that we use in our algorithm.

This is the analog of the centroid of a discrete set $S = \{x_1, x_2, \ldots, x_N\}$:

$$\mathrm{centroid}(S) := \frac{1}{|S|} \sum_i x_i.$$

Other names for the centroid are the center of gravity, and the barycenter.

B. Grünbaum (1960)

Theorem 19.1 (Grünbaum's Theorem)**.** *For any compact convex set $K \in \mathbb{R}^n$ with a centroid $c \in \mathbb{R}^n$, and any halfspace $H = \{x \mid a^\intercal(x - c) \geq 0\}$ whose supporing hyperplane passes through c,*

$$\frac{1}{e} \leq \frac{\mathrm{vol}(K \cap H)}{\mathrm{vol}(K)} \leq \left(1 - \frac{1}{e}\right).$$

This bound of $1/e$ in Grünbaum's Theorem is the best possible: e.g., consider the simplex $K = \{x \in [0,1]^n \mid \|x\|_1 \leq 1\}$ with centroid $\frac{1}{n+1}\mathbb{1}$. Defining the halfspace $H = \{x_1 \geq c\}$, we get that $K \cap H$ is a scaled-down copy of K, with volume

$$\left(1 - \frac{1}{n+1}\right)^n \to 1/e$$

as $n \to \infty$.

19.1.1 The Algorithm

In 1965, A. Ju. Levin and Donald Newman independently (and on opposite sides of the iron curtain) proposed the following algorithm.

A.Ju. Levin (1965)
D.J. Newman (1965)

For most of this chapter, we assume that we can perform *exact arithmetic on real numbers*. This assumption could be very restrictive loss in generality, since some of our algorithm take square-roots (e.g., when computing ellipsoids). Rounding numbers create all sorts of numerical problems, and a large part of the complication in the actual algorithms comes from these numerical issues.

Algorithm 18: Centroid(K, f, T)

```
18.1 $K_1 \leftarrow K$
18.2 for $t = 1, \ldots T$ do
18.3     at step $t$, let $c_t \leftarrow$ centroid of $K_t$
18.4     $K_{t+1} \leftarrow K_t \cap \{x \mid \langle \nabla f(c_t), x - c_t \rangle \leq 0\}$
18.5 return $\widehat{x} \leftarrow \arg\min_{t \in \{1,\ldots,T\}} f(c_t)$
```

The figure to the right shows a sample execution of the algorithm, where K is initially a ball. (Ignore the body K^ε for now.) We find the centroid c_1 and compute the gradient $\nabla f(c_1)$. Instead of moving in the direction opposite to the gradient, we consider the halfspace H_1 of vectors negatively correlated with the gradient, restrict our search to $K \leftarrow K \cap H_1$, and continue. We repeat this step some number of times, and then return the smallest of the function value at all the centroids seen by the algorithm. Note that the algorithm assumes:

1. Access to both a gradient oracle and a value oracle for the function f, and
2. access to a procedure that computes the centroid for any compact convex set K.

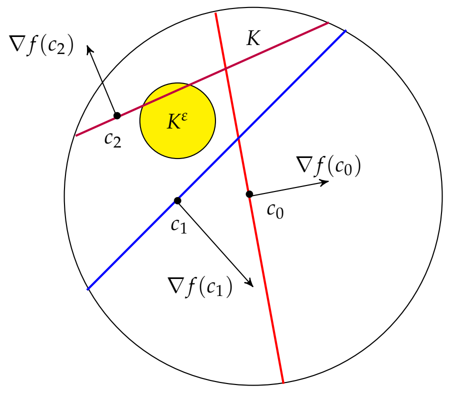

Figure 19.1: Sample execution of first three steps of the Centroid Algorithm.

Theorem 19.2. *Consider a convex set $K \subseteq (0, R) \subseteq \mathbb{R}^n$, and a convex function $f : K \to \mathbb{R}$ such that let $\|\nabla f(x)\| \leq G$ for all $x \in K$. If $\widehat{x}$ is the result of the algorithm, and $x^* = \arg\min_{x\in K} f(x)$, then*

$$f(\widehat{x}) - f(x^*) \leq 4GR \cdot \exp(-T/3n).$$

Hence, for any $\varepsilon \leq 1$, as long as $T \geq 3n \ln \frac{4GR}{\varepsilon}$,

$$f(\widehat{x}) - f(x^*) \leq \varepsilon.$$

Proof. For some $\delta \leq 1$, define the body

$$K^\delta := \{(1-\delta)x^* + \delta x \mid x \in K\}$$

as a scaled-down version of K centered at x^*. The following facts are immediate:

1. $\mathrm{vol}(K^\delta) = \delta^n \cdot \mathrm{vol}(K)$.
2. For any points $x, y \in K$, integrating along the path from x to y and using the fact that the gradients are bounded by G gives

$$\begin{aligned} f(x) - f(y) &= \int_{t=0}^{1} \langle f(y + t(x-y)), x - y\rangle\, dt \\ &\leq \int_{t=0}^{1} \|f(y+t(x-y))\| \|x-y\| dt \leq G\|x-y\| \leq G \cdot (2R). \end{aligned}$$

3. The value of f on any point $y = (1-\delta)x^* + \delta x \in K^\delta$ is

$$\begin{aligned} f(y) = f((1-\delta)x^* + \delta x) &\leq (1-\delta)f(x^*) + \delta f(x) \\ &\leq f(x^*) + \delta(f(x) - f(x^*)) \leq f(x^*) + 2\delta GR. \end{aligned}$$

Using Grünbaum's lemma, the volume falls by a constant factor in each iteration, so $\mathrm{vol}(K_t) \leq \mathrm{vol}(K) \cdot (1 - \frac{1}{e})^t$. If we define $\delta := 2(1-1/e)^{T/n}$, then after T steps the volume of K_T is smaller than that of K^δ, so some point of K^δ must have been cut off.

Consider such a step t such that $K^\delta \subseteq K_t$ but $K^\delta \not\subseteq K_{t+1}$. Let $y \in K^\delta \cap (K_t \setminus K_{t+1})$ be a point that is "cut off". By convexity we have

$$f(y) \geq f(c_t) + \langle \nabla f(c_t), y - c_t\rangle;$$

moreover, $\langle \nabla f(c_t), y - c_t\rangle > 0$ since the cut-off point $y \in K_t \setminus K_{t+1}$. Hence the corresponding centroid has value $f(c_t) < f(y) \leq f(x^*) + 2\delta GR$. Since $\widehat{x}$ is the centroid with the smallest function value, we get

$$f(\widehat{x}) - f(x^*) \leq 2GR \cdot 2\big(1 - 1/e\big)^{T/n} \leq 4GR \exp(-T/3n).$$

The second claim follows by substituting $T \geq 3n \ln \frac{4GR}{\varepsilon}$ into the first claim, and simplifying. □

19.1.2 *Comments on the Runtime*

The number of iterations T needed by the Centroid algorithm to get an error of ε is $O(n \log(GR/\varepsilon))$; compare this linear convergence to gradient descent requiring $O((GR/\varepsilon)^2)$ steps in the same setting. One downside with this approach is that the number of iterations explicitly depends on the number of dimensions n, whereas gradient descent does not.

Recall that *linear convergence* refers to a rate where we the *number of bits* of precision increases linearly: i.e., the number of iterations is logarithmic in $1/\varepsilon$.

Another all-important question is: *how do we compute the centroid*? This is a difficult problem—it is #P-hard to do exactly, which means it is at least as hard as counting the number of satisfying assignments to a SAT instance. (You will see this in a homework problem.) In 2002, Dimitris Bertsimas and Santosh Vempala suggested a way to find approximate centroids by sampling random points from convex bodies (which in turn is done via random walks). Combined with a robust version of Grünbaum's theorem gives us a polynomial-time version of the algorithm.

D. Bertsimas and S. Vempala (2006)

19.2 *Multi-Dimensional Binary Search*

Let us put the Centroid algorithm in a broader context. Given a convex body, one of the canonical ways of specifying it will be via a separation oracle.

An ε-weak separation oracle is one where we are just ensured that $\langle a, x\rangle > \langle a, y\rangle - \varepsilon$ for all $y \in K$. Specifying K via a weak separation oracle makes all our tasks much more challenging; in this course we restrict our discussions to strong separation, and defer the generalization to the GLS book.

Definition 19.3 (Strong Separation Oracle)**.** For a convex set $K \subseteq \mathbb{R}^n$, a *strong separation oracle* for K is an algorithm that takes a point $z \in \mathbb{R}^n$ and correctly outputs one of:

(i) `Yes` (i.e., $z \in K$), or

(ii) `No` (i.e., $z \notin K$), as well as a *separating hyperplane* given by $a \in \mathbb{R}^n, b \in \mathbb{R}$ such that $K \subseteq \{x \in \mathbb{R}^n \mid \langle a, x\rangle \leq b\}$ but $\langle a, z\rangle > b$.

The Hahn-Banach separation theorem ensures that exactly one of the two cases can hold for any x and K. Our goal now is to solve the following feasibility problem:

> Given access to a strong separation oracle for a convex body K, as well as positive values R, r such that (a) $K \subseteq B(0, R) \subseteq \mathbb{R}^n$, and (b) the body K is either empty, or else there is some unknown (full-dimensional) r-ball $B(c, r) \subseteq K$. If $K \neq \emptyset$, output a point $x \in K$, else say $K = \emptyset$.

19.2.1 *Feasibility using Centroids*

The ideas behind the Centroid algorithm also solves the feasibility problem:

Algorithm 19: CentroidFeasibility(K, R, r)

```
19.1 $\mathcal{E}_0 \leftarrow B(0, R)$
19.2 for $t = 0, 1, \ldots T := 3n \ln(R/r)$ do
19.3     query strong separation oracle on $c_t$, the centroid of $\mathcal{E}_t$
19.4     if $c_t \notin K$ then
19.5         $a_t \leftarrow$ direction from strong separation oracle
19.6         $\mathcal{E}_{t+1} \leftarrow \mathcal{E}_t \cap \{x \mid \langle \nabla a_t, x - c_t \rangle \leq 0\}$
19.7     else
19.8         output "$c_t \in K$" and stop
19.9     output "$K = \emptyset$"
```

The argument is nearly identical to the one we saw above:

1. By Grünbaum's theorem,

$$\text{vol}(\mathcal{E}_t) \leq (1 - 1/e)^t \, \text{vol}(\mathcal{E}_0) = (1 - 1/e)^t \, \text{vol}(B(0, R)),$$

which gives an *upper bound* on $\mathcal{E}_t$'s volume.

2. Suppose $K \neq \emptyset$. If none of the centers $c_{t'}$ for $t' < t$ belong to K, then $K \subseteq \mathcal{E}_t$ and hence $\text{vol}(\mathcal{E}_t) \geq \text{vol}(B(c, r))$. This gives a *lower bound* on the volume of $\mathcal{E}_t$, as long as none of the centroids fall within our target convex body K.

3. Putting the above statements togehter, $(1 - 1/e)^t \leq \frac{\text{vol}(B(0,R))}{\text{vol}(B(c,r))} = (R/r)^n$. This means that if we do not find a point in K within $3n \log(R/r)$ steps, K must be empty!

This approach is very flexible: we just need to (efficiently) maintain a sequence of bodies $\{\mathcal{E}_t\}_t$ such that for each step t:

(a) $\text{vol}(\mathcal{E}_{t+1}) \leq \text{vol}(\mathcal{E}_t) \cdot (1 - \delta)$ for some $\delta > 0$,

(b) if each of the "test points" $c_1, c_2, \ldots, c_t$ did not belong to K, then $K \subseteq \mathcal{E}_{t+1}$.

Following the same outline with these properties gives an iteration complexity of $O\big(\frac{n}{\delta} \log\left(\frac{R}{r}\right)\big)$. Maybe this more abstract view allows us to get an efficient algorithm (since computing centroids is #P-hard)?

19.2.2 The Ellipsoid Algorithm

Going from the Centroid to the Ellipsoid algorithm requires a remarkably small change. If our test point $c_t = \text{centroid}(\mathcal{E}_t)$ does not belong to K, the separation oracle returns a half-space $H := \{x \mid \langle a_t, x - c_t \rangle \leq 0\}$ that contains K. Now we don't just define

$$\mathcal{E}_{t+1} \leftarrow \mathcal{E}_t \cap H$$

but instead define:

$$\mathcal{E}_{t+1} \leftarrow \text{minimum volume ellipsoid containing } \mathcal{E}_t \cap H.$$

The minimum-volume ellipsoid containing a convex body K is often called the *John ellipsoid* for K, after Fritz John who proved several properties for it in 1948.

How can we compute this minimum-volume ellipsoid? And does the volume go down by a constant factor?

Since we start off with $\mathcal{E}_0$ being a ball (which is trivially an ellipsoid), it suffices to show how to compute the minimum-volume ellipsoid $\mathcal{E}_{t+1}$ of *half an ellipsoid* (the intersection of an ellipsoid $\mathcal{E}_t$ with a half-space passing through its center). We show how to do this in §19.5, and show that the ellipsoid $\mathcal{E}_{t+1} \supseteq \mathcal{E}_t \cap H_t$ has volume

Why not balls? Clearly, the smallest volume ball that contains half a ball is the ball itself. Interestingly, the same is true for boxes: the volume of the new box may not decrease. Thankfully, ellipsoids—and in fact, simplices—do have the volume-reduction property.

$$\frac{\text{vol}(\mathcal{E}_{t+1})}{\text{vol}(\mathcal{E}_t)} \leq e^{-\frac{1}{2(n+1)}} \approx \big(1 - O(1/n)\big).$$

Therefore, after $2(n+1)$ iterations, the ratio of the volumes falls by at least a factor of $\frac{1}{e}$. Hence, if after $O(n^2 \ln(R/r))$ steps, none of the ellipsoid centers have been inside K, we know that K must be empty.

This volume reduction is weaker by a factor of $\Theta(n)$ than that of the Centroid algorithm.

19.3 Ellipsoid for Convex Optimization

Let's go back to convex minimization: we want to solve $\min\{f(x) \mid x \in K\}$. Again, assume that K is given by a strong separation oracle, and we have numbers R, r such that $K \subseteq \text{Ball}(0, R)$, and K is either empty or contains a ball of radius r. The general structure is a one familiar by now, and combines ideas from both the previous sections.

1. Let the starting point $x_1 \leftarrow 0$, the starting ellipsoid be $\mathcal{E}_1 \leftarrow \text{Ball}(0, R)$, and the starting convex set $K_1 \leftarrow K$.

2. At time t, ask the separation oracle: "Is the center c_t of ellipsoid $\mathcal{E}_t$ in the convex body K_t?"

   Yes: Define half-space $H_t := \{x \mid \langle \nabla f(c_t), x - c_t \rangle \leq 0\}$. Observe that $K_t \cap H_t$ contains all points in K_t with value at most $f(c_t)$.

   No: In this case the separation oracle also gives us a separating hyperplane. This defines a half-space H_t such that $c_t \notin H_t$, but $K_t \subseteq H_t$.

   In both cases, set $K_{t+1} \leftarrow K_t \cap H_t$, and $\mathcal{E}_{t+1}$ to an ellipsoid containing $\mathcal{E}_t \cap H_t$. Since we knew that $K_t \subseteq \mathcal{E}_t$, we maintain that $K_{t+1} \subseteq \mathcal{E}_{t+1}$.

3. Finally, after $T = 2n(n+1)\ln(R/r)$ rounds either we have not seen any point in K—in which case we say "K is empty"—or else we output

$$\widehat{x} \leftarrow \arg\min\{f(c_t) \mid c_t \in K_t, t \in 1 \dots T\}.$$

One subtle issue: we make queries to a separation oracle for K_t, but we are promised only a separation oracle for $K_1 = K$. However, we can build separation oracles for H_t inductively: indeed, given strong separation oracle for K_{t-1}, we build one for $K_t = K_{t-1} \cap H_{t-1}$ as follows:

> Given $z \in \mathbb{R}^n$, query the oracle for K_{t-1} at z. If $z \notin K_{t-1}$, the separating hyperplane for K_{t-1} also works for K_t. Else, if $z \in K_{t-1}$, check if $z \in H_{t-1}$. If so, $z \in K_t = K_{t-1} \cap H_{t-1}$. Otherwise, the defining hyperplane for halfspace H_{t-1} is a separating hyperplane between z and K_t.

Now adapting the analysis from the previous sections gives us the following result (assuming exact arithmetic again):

Theorem 19.4 (Idealized Convex Minimization using Ellipsoid)**.** *Given K, r, R as above (and a strong separation oracle K), and a function f with gradients bounded by G, the Ellipsoid algorithm run for T steps either correctly reports that $K = \emptyset$, or else produces a point $\widehat{x}$ such that*

$$f(\widehat{x}) - f(x^*) \leq \frac{O(GR)}{r} \exp\left\{ -\frac{T}{2n(n+1)} \right\}.$$

Note the similarity to Theorem 19.2, as well as the differences: the exponential term is slower by a factor of $2(n+1)$. This is because the volume of the successive ellipsoids shrinks much slower than in Grünbaum's lemma. Also, we lose a factor of R/r because K is potentially smaller than the starting body by precisely this factor. (Again, this presentation ignores precision issues, and assumes we can do exact real arithmetic.)

19.4 The Ellipsoid Algorithm to Solve LPs

The Ellipsoid algorithm is usually attributed to Naum Shor; the fact that this algorithm gives a polynomial-time algorithm for linear programming was a breakthrough result due to Khachiyan, and was front page news at the time. A great source of information about this algorithm is the Grötschel-Lovász-Schrijver book. A historical perspective appears in this this survey by Bland, Goldfarb, and Todd.

N. Z. Šor and N. G. Žurbenko (1971)

L.G. Khachiyan (1979)

M. Grötschel, L. Lovász, and A. Schrijver (1988)

Let us mention some theorem statements about the Ellipsoid algorithm that are most useful in designing algorithms. The second-most important theorem is the following. Recall the notion of an extreme point or basic feasible solution (bfs) from §9.1.2. Let $\langle A \rangle, \langle b \rangle, \langle c \rangle$ denote the number of bits required to represent of A, b, c respectively.

Theorem 19.5 (Linear Programming in Polynomial Time)**.** *Given a linear program* $\min\{c^\intercal x \mid Ax \geq b\}$, *the Ellipsoid algorithm produces an optimal vertex solution for the LP, in time polynomial in $\langle A \rangle$, $\langle b \rangle$, $\langle c \rangle$.*

One may ask: does the runtime depend on the bit-complexity of the input because doing basic arithmetic on these numbers may require large amounts of time. Unfortunately, that is not the case. Even if we count the number of arithmetic operations we need to perform, the Ellipsoid algorithm performs $\text{poly}(\langle A\rangle + \langle b\rangle + \langle c\rangle)$ operations. A stronger guarantee would have been for the number of arithmetic operations to be $\text{poly}(m, n)$, where the matrix $A \in \mathbb{Q}^{m\times n}$: such an algorithm would be called a *strongly polynomial-time* algorithm. Obtaining such an algorithm remains a major open question.

19.4.1 Finding Vertex Solutions for LPs

There are several issues that we need to handle when solving LPs using this approach. For instance, the polytope may not be full-dimensional, and hence we do not have any non-trivial ball within K. Our separation oracles may only be approximate. Moreover, all the numerical calculations may only be approximate.

Even after we take care of these issues, we are working over the rationals so binary search-type techniques may not be able to get us to a vertex solution. So finally, when we have a solution x_t that is "close enough" to x^*, we need to "round" it and get a vertex solution. In a single dimension we can do the following (and this idea already appeared in a homework problem): we know that the optimal solution x^* is a rational whose denominator (when written in reduced terms) uses at most some b bits. So we find a solution within distance to x^* is smaller than some δ. Moreover δ is chosen to be small enough such that there is a unique rational with denominator smaller than 2^b in the δ-ball around x_t. This rational can only be x^*, so we can "round" x_t to it.

Consider the case where we perform binary-search over the interval $[0, 1]$ and want to find the point $1/3$: no number of steps will get us exactly to the answer.

In higher dimensions, the analog of this is a technique (due to Lovász) called *simultaneous Diophantine equations*.

19.4.2 Separation Implies Optimization

The most important theorem about Ellipsoid is the following:

Theorem 19.6 (Separation implies Optimization)**.** *Given an LP*

$$\min\{c^\intercal x \mid x \in K\}$$

for a polytope $K = \{x \mid Ax \geq b\} \subseteq \mathbb{R}^n$, and given access to a strong separation oracle for K, the Ellipsoid algorithm produces a vertex solution for the LP in time $\text{poly}(n, \max_i\langle a_i\rangle, \max_i\langle b_i\rangle, \langle c\rangle)$.

There is no dependence on the number of constraints in the LP; we can get a basic solution to any finite LP as long as each constraint has a reasonable bit complexity, and we can define a separation oracle for the polytope. This is often summarized by saying: *"separation implies optimization"*. Let us give two examples of exponential-sized LPs, for which we can give a separation oracles, and hence optimize over them.

19.5 Getting the New Ellipsoid

This brings us to the final missing piece: given a current ellipsoid $\mathcal{E}$ and a half-space H that does not contain its center, we want an ellipsoid $\mathcal{E}'$ that contains $\mathcal{E} \cap H$, and as small as possible. To start off, let us recall some basic facts about ellipsoids. The simplest ellipses in $\mathbb{R}^2$ are axis aligned, say with principal semi-axes having length a and b, and written as:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1.$$

Or in matrix notation we could also say

$$\begin{bmatrix} x \\ y \end{bmatrix}^{\intercal} \begin{bmatrix} 1/a^2 & 0 \\ 0 & 1/b^2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \leq 1$$

More generally, any ellipsoid $\mathcal{E}$ is perhaps best thought of as a invertible linear transformation L applied to the unit ball $B(0,1)$, and then it being shifted to the correct center c. The linear transformation yields:

$$\begin{aligned} L(\text{Ball}(0,1)) &= \{Lx : x^{\intercal}x \leq 1\} \\ &= \{y : (L^{-1}y)^{\intercal}(L^{-1}y) \leq 1\} \\ &= \{y : y^{\intercal}(LL^{\intercal})^{-1}y \leq 1\} \\ &= \{y : y^{\intercal}Q^{-1}y \leq 1\}, \end{aligned}$$

where $Q^{-1} := LL^{\intercal}$ is a positive semidefinite matrix. For an ellipsoid centered at c we simply write

$$\{y + 1 : y^{\intercal}Q^{-1}y \leq 1\} = \{y : (y-c)^{\intercal}Q^{-1}(y-c) \leq 1\}.$$

It is helpful to note that for any ball A,

$$\text{vol}(L(A)) = \text{vol}(A) \cdot |\det(L)| = \text{vol}(A)\ \sqrt{\det(Q)}$$

In the above problems, we are given an ellipsoid $\mathcal{E}_t$ and a half-space H_t that does not contain the center of $\mathcal{E}_t$. We want to find a matrix Q_{t+1} and a center c_{t+1} such that the resulting ellipsoid $\mathcal{E}_{t+1}$ contains $\mathcal{E}_t \cap H_t$, and satisfies

$$\frac{\text{vol}(\mathcal{E}_{t+1})}{\text{vol}(\mathcal{E}_t)} \leq e^{-1/2(n+1)}.$$

Given the above discussion, it suffices to do this when $\mathcal{E}_t$ is a unit ball: indeed, when $\mathcal{E}_t$ is a general ellipsoid, we apply the inverse linear transformation to convert it to a ball, find the smaller ellipsoid for it, and then apply the transformation to get the final smaller ellipsoid. (The volume changes due to the two transformations cancel each other out.)

We give the construction for the unit ball below, but first let us record the claim for general ellipsoids:

Theorem 19.7. *Given an ellipsoid $\mathcal{E}_t$ given by (c_t, Q_t) and a separating hyperplane $a_t^\intercal(x - c_t) \leq 0$ through its center, the new ellipsoid $\mathcal{E}_{t+1}$ with center c_{t+1} and psd matrix $Q_{t+1})$ is found by taking*

$$c_{t+1} := c_t - \frac{1}{n+1} h$$

and

$$Q_{t+1} = \frac{n^2}{n^2-1}\left(Q_k - \frac{2}{n+1} hh^\intercal\right)$$

where $h = \sqrt{a_t^\intercal Q_t a_t}$.

Note that the construction requires us to take square-roots: this may result in irrational numbers which we then have to either truncate, or represent implicitly. In either case, we face numerical issues; ensuring that these issues are not real problems lies at the heart of the formal analysis. We refer to the GLS book, or other textbooks for details and references.

19.5.1 Halving a Ball

Before we end, we show that the problem of finding a smaller ellipsoid that contains half a ball is, in fact, completely straight-forward. By rotational symmetry, we might as well find a small ellipsoid that contains

$$K = \text{Ball}(0, 1) \cap \{x \mid x_1 \geq 0\}.$$

By symmetry, it makes sense that the center of this new ellipsoid $\mathcal{E}$ should be of the form

$$c = (c_1, 0, \ldots, 0).$$

Again by symmetry, the ellipsoid can be axis-aligned, with semi-axes of length a along e_1, and $b > a$ along all the other coordinate axes. Moreover, for $\mathcal{E}$ to contain the unit ball, it should contain the points $(1, 0)$ and $(0, 1)$, say. So

$$\frac{(1-c_1)^2}{a^2} \leq 1 \quad \text{and} \quad \frac{c_1^2}{a^2} + \frac{1}{b^2} \leq 1.$$

Suppose these two inequalities are tight, then we get

$$a = 1 - c_1, \qquad b = \sqrt{\frac{(1-c_1)^2}{(1-c_1)^2 - c_1^2}} = \sqrt{\frac{(1-c_1)^2}{(1-2c_1}},$$

and moreover the ratio of volume of the ellipsoid to that of the ball is

$$ab^{n-1} = (1-c_1) \cdot \left(\frac{(1-c_1)^2}{1-2c_1}\right)^{(n-1)/2}.$$

This is minimized by setting $c_1 = \frac{1}{n+1}$ gives us

$$\frac{\mathrm{vol}(\mathcal{E})}{\mathrm{vol}(\mathrm{Ball}(0,1))} = \cdots \leq e^{-\frac{1}{2(n+1)}}.$$

For a more detailed description and proof of this process, see these notes from our LP/SDP course for details.

In fact, we can view the question of finding the minimum-volume ellipsoid that contains the half-ball K: this is a convex program, and looking at the optimality conditions for this gives us the same construction above (without having to make the assumptions of symmetry).

19.6 *Algorithms for Solving LPs*

We have now seen two different classes of algorithms to solve linear programs: the first approach using multiplicative weights gave us solutions which violate the constraints by ε and take $O(1/\varepsilon^2)$ steps (ignoring terms that depend on the other input parameters for now). Next we saw the Centroid and Ellipsoid algorithms for convex programming which require only $O(\log 1/\varepsilon)$ steps. However, they are typically not used to solve LPs in practice. There are several other algorithms: let us mention them in passing. Let $K := \{x \mid Ax \geq b\} \subseteq \mathbb{R}^n$, and we want to minimize $\{c^\intercal x \mid x \in K\}$.

Simplex: This is perhaps the first algorithm for solving LPs that most of us see. It was also the first general-purpose linear program solver known, having been developed by George Dantzig in 1947. This is a local-search algorithm: it maintains a vertex of the polyhedron K, and at each step it moves to a neighboring vertex without decreasing the objective function value, until it reaches an optimal vertex. (The convexity of K ensures that such a sequence of steps is possible.) The strategy to choose the next vertex is called the *pivot rule*. Unfortunately, for most known pivot rules, there are examples on which the following the pivot rule takes exponential (or at least a super-polynomial) number of steps. Despite that, it is often used in practice: e.g., the Excel software contains an implementation of simplex.

G.B. Dantzig (1990)

Interior Point: A different approach to get algorithms for LPs is via interior-point algorithms: these happen to be good both in theory and in practice. The first polynomial-time interior-point algorithm was proposed by Karmarkar in 1984. We discuss this in the next chapter.

Geometric Algorithms for LPs: These approaches are geared towards solving LPs fast when the number of dimensions n is small. If m is the number of constraints, these algorithms often allow a poor runtime in n, at the expense of getting a good dependence on m. As an example, a randomized algorithm of Raimund Seidel's has a runtime of $O(m \cdot n!) = O(m \cdot n^{n/2})$; a different algorithm of Ken Clarkson (based on the multiplicative weights approach!) has a runtime of $O(n^2 m) + n^{O(n)} O(\log m)^{O(\log n)}$. One of the fastest such algorithm is by Jiri Matoušek, Micha Sharir, and Emo Welzl, and has a runtime of

$$O(n^2 m) + e^{O(\sqrt{n \log n})}.$$

For details and references, see this survey by Martin Dyer, Nimrod Megiddo, and Emo Welzl.

Naturally, there are other approaches to solve linear programs as well: write more here.