Mathematical Programming	Lecture 23
OR 630 Fall 2005	November 17, 2005
Instructor: Mike Todd	Scribe: Chris Provan

Deeper Into the Ellipsoid Method

Given the polyhedron $Q := \{x \in \mathbf{R}^m : A^T x \leq b\}$ where we assume:
(1) $Q \subseteq B(0, R) = \{x \in \mathbf{R}^m : \|x\| \leq R\}$ (Q possibly empty) and
(2) if $Q \neq \emptyset$ then $Q \supseteq B(\hat{x}, r) = \{x \in \mathbf{R}^m : \|x - \hat{x}\| \leq r\}$ for some (unknown) $\hat{x} \in \mathbf{R}^m$ and some (known) $0 < r < R$,
our goal is to find a feasible point in Q or show that $Q = \emptyset$.

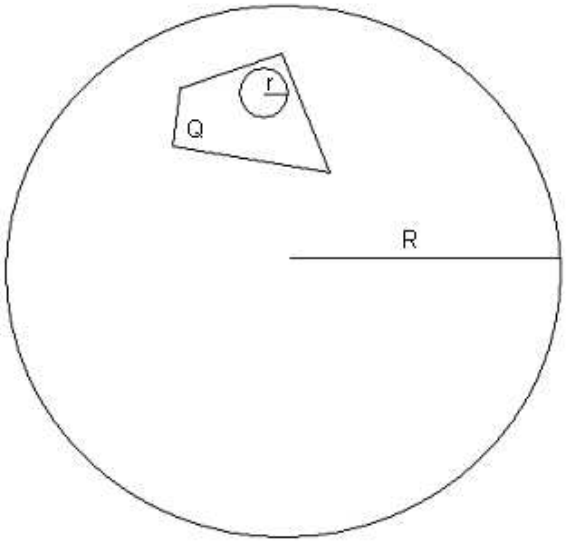

Figure 1: *A two-dimensional polyhedron Q satisfying the initial assumptions for the ellipsoid method*

To close in on a feasible point, we will construct a shrinking sequence of ellipsoids expressed as $E(z, B) := \{x \in \mathbf{R}^m : (x-z)^T B^{-1}(x-z) \leq 1\}$ for some center $z \in \mathbf{R}^m$ and some $B \in \mathbf{R}^{m \times m}$ with B symmetric and positive definite (i.e. $\forall v \neq 0 : v^T B v > 0 \Leftrightarrow$ the eigenvalues of B are all positive) in the following manner:

Step 0: Set $z_0 = 0$ and $B_0 = R^2 I$. Then $Q \subseteq E(z_0, B_0) =: E_0$.
Step k + 1, k ≥ 0: Given $Q \subseteq E(z_k, B_k) =: E_k$, if $z_k \in Q$ then we have a feasible point - **STOP**. If $z_k \notin Q$ and k is "large enough" (see below) then we can conclude that Q is empty - **STOP**. Otherwise, generate $E_{k+1} := E(z_{k+1}, B_{k+1})$ by choosing some $a = a_i \neq 0$ where $a_i^T z_k > b_i$ (i.e. z_k violates the ith constraint of Q), and letting E_{k+1} be the minimum volume ellipsoid such that $E_{k+1} \supseteq E_k^{1/2} := \{x \in E_k : a^T x \leq a^T z_k\}$. Clearly we have $Q \subseteq E_k^{1/2} \subseteq E_{k+1}$,

so we move on to the next iteration.

How large does k have to be for us to determine that Q is empty? If $vol(\cdot)$ is the m-dimensional volume then let $Vol(\cdot)$ be the scaled m-dimensional volume, i.e. $Vol(\cdot) = \frac{vol(\cdot)}{vol(B(0,1))}$. Then $Vol(E_0) = R^m Vol(B(0,1)) = R^m$, and similarly, $Vol(B(\hat{x}, r)) = r^m$. So if $Vol(E_k) < r^m$ then E_k cannot contain $B(\hat{x}, r)$ and thus Q must be empty. We will show later that $Vol(E_{k+1}) < exp(\frac{-1}{2(m+1)})VolE_k$, and therefore we can conclude that Q is empty when

$$\begin{aligned} Vol(E_k) < exp(\tfrac{-k}{2(m+1)})Vol(E_0) \leq r^m \quad &\Rightarrow exp(\tfrac{-k}{2(m+1)})R^m \leq r^m \\ &\Rightarrow k \geq 2(m+1)m\ln(\tfrac{R}{r}). \end{aligned}$$

We now prove some intermediate results that will help to establish the algorithm for an iteration. To simplify the notation going forward, when we refer to a specific iteration going from E_k to E_{k+1}, we will drop any subscript k and use $+$ in place of $k+1$ ($E_+ := E_{k+1}$, etc).

Lemma 1 *If B is symmetric and positive definite then it has a symmetric, positive definite square root $B^{1/2}$ with $B^{1/2}B^{1/2} = B$. Moreover, $Vol(E(z, B)) = \sqrt{det(B)}$.*

Proof: We can leverage what we know about numbers and apply this first to diagonal matrices and then to symmetric matrices. B can be factorized into $B = QDQ^T$ where Q is an orthogonal matrix ($Q^TQ = I$) and D is a diagonal matrix. Moreover, the columns of Q are the eigenvectors of B and the diagonal entries of D are its eigenvalues, and thus all d_{jj} are positive.

Then setting

$$D^{1/2} = \begin{pmatrix} \sqrt{d_{11}} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \sqrt{d_{mm}} \end{pmatrix}$$

and $B^{1/2} = QD^{1/2}Q^T$ gives

$$B^{1/2}B^{1/2} = (QD^{1/2}Q^T)(QD^{1/2}Q^T) = Q(D^{1/2}ID^{1/2})Q^T = QDQ^T = B.$$

Now note that

$$x \in E(z, B) \Leftrightarrow (x-z)^T B^{-1}(x-z) \leq 1 \Leftrightarrow (B^{-1/2}(x-z))^T(B^{-1/2}(x-z)) \leq 1 \Leftrightarrow x = z + B^{1/2}w$$

for some $w \in B(0,1)$. So defining $X(w) := z + B^{1/2}w$ gives

$$\begin{aligned} Vol(E(z,B)) &= \tfrac{\int_{x \in E(z,B)} 1dx}{vol(B(0,1))} \\ &= \tfrac{\int_{w \in B(0,1)} 1|det(B^{1/2})|dw}{vol(B(0,1))} \\ &= |det(B^{1/2})|\tfrac{vol(B(0,1))}{vol(B(0,1))} \\ &= |det(B^{1/2})| \\ &= \sqrt{det(B)} \qquad .\square \end{aligned}$$

Lemma 1 confirms our earlier comment that $Vol(E_0) = R^m$.

Lemma 2 *If $B \in \mathbb{R}^{m\times m}$ is symmetric and positive definite and $a \in \mathbf{R}^m$ is non-zero then*

$$\bar{B} := B - \sigma \frac{Baa^TB}{a^TBa}$$

is symmetric and positive definite for $\sigma < 1$ with $det(\bar{B}) = (1-\sigma)det(B)$ and

$$\bar{B}^{-1} = B^{-1} + (\tfrac{\sigma}{1-\sigma})(\tfrac{aa^T}{a^TBa}).$$

Proof: $a^TBa > 0$, so $\bar{B}$ is well-defined and clearly symmetric. We also have

$$\begin{aligned}\bar{B} &= B^{1/2}(I - \sigma \tfrac{B^{1/2}aa^TB^{1/2}}{(B^{1/2}a)^T(B^{1/2}a)})B^{1/2} \\ &= B^{1/2}(I - \sigma \tfrac{uu^T}{u^Tu})B^{1/2}\end{aligned}$$

for $u = B^{1/2}a$. $(I - \sigma \frac{uu^T}{u^Tu})$ has an eigenvalue of $(1-\sigma) > 0$ associated with the eigenvector u and eigenvalue $1 > 0$ with multiplicity $(m-1)$ associated with the $(m-1)$-dimensional subspace orthogonal to u, so it is positive definite with determinant $(1-\sigma)$. Then $\forall v \neq 0$, $v^T\bar{B}v = (B^{1/2}v)^T(I - \sigma \frac{uu^T}{u^Tu})(B^{1/2}v) > 0$, so $\bar{B}$ is positive definite, and

$$\begin{aligned}det(\bar{B}) &= det(B^{1/2})det(I - \sigma \tfrac{uu^T}{u^Tu})det(B^{1/2}) \\ &= (1-\sigma)det(B).\end{aligned}$$

Finally, $\bar{B} = B + vw^T$ for $v = -\sigma \frac{Ba}{a^TBa}$ and $w = Ba$, thus

$$\begin{aligned}\bar{B}^{-1} &= B^{-1} - \tfrac{B^{-1}vw^TB^{-1}}{1+w^TB^{-1}v} \\ &= B^{-1} + \frac{\sigma \frac{B^{-1}Baa^TBB^{-1}}{a^TBa}}{1-\sigma \frac{a^TBB^{-1}Ba}{a^TBa}} \\ &= B^{-1} + (\tfrac{\sigma}{1-\sigma})(\tfrac{aa^T}{a^TBa}).\end{aligned}$$

□

Lemma 3 *For any $a \neq 0$, the minimum of a^Tx over $x \in E(z,B)$ is $a^Tz - \sqrt{a^TBa}$ and is attained by $x = z - \frac{Ba}{\sqrt{a^TBa}}$.*

Proof: By Cauchy-Schwartz, if we minimize $(B^{1/2}a)^Tw$ over the unit ball $B(0,1)$ then the optimal solution is $-\sqrt{a^TBa}$ and is attained by $w = -\frac{B^{1/2}a}{\sqrt{a^TBa}}$. If we apply the transformation $X(w)$ as in the proof of Lemma 1 then we have the desired result. □

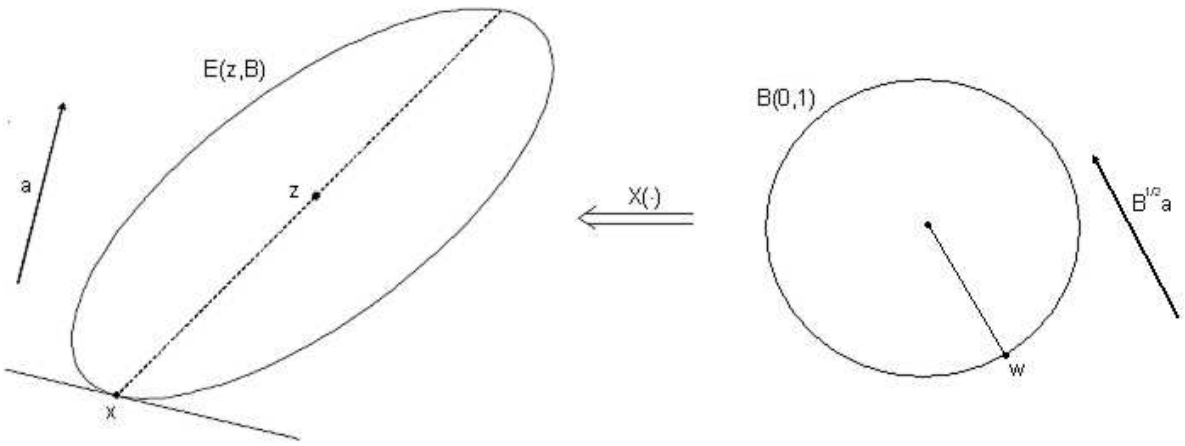

Figure 2: *The transformation used in the proof of Lemma 3.*

We are now ready to look at the algorithm for performing an iteration.

Theorem 1 *Given an ellipsoid $E = E(z, B)$ and $0 \neq a \in \mathbf{R}^m$, the minimum volume ellipsoid containing $E^{1/2} = \{x \in E : a^T x \leq a^T z\}$ is $E_+ = E(z_+, B_+)$ for*

$$z_+ := z - \tau \frac{Ba}{\sqrt{a^T B a}} \text{and}$$

$$B_+ := \delta(B - \sigma \frac{Baa^T B}{a^T B a})$$

where $\tau = \frac{1}{m+1}$, $\delta = \frac{m^2}{m^2-1}$, and $\sigma = \frac{2}{m+1}$. Moreover,

$$\frac{Vol(E_+)}{Vol(E)} < exp(\frac{-1}{2(m+1)}).$$

But this theorem is just the special case where $\alpha = 0$ of:

Theorem 2 *Given an ellipsoid $E = E(z, B)$ and $0 \neq a \in \mathbf{R}^m$, the minimum volume ellipsoid containing $E_\alpha = \{x \in E : a^T x \leq a^T z - \alpha\sqrt{a^T B a}\}$ for $-\frac{1}{m} \leq \alpha < 1$ is $E_+ = E(z_+, B_+)$ for*

$$z_+ := z - \tau \frac{Ba}{\sqrt{a^T B a}} \text{and}$$

$$B_+ := \delta(B - \sigma \frac{Baa^T B}{a^T B a})$$

where $\tau = \frac{1+m\alpha}{m+1}$, $\delta = \frac{(1-\alpha^2)m^2}{m^2-1}$, and $\sigma = \frac{2(1+m\alpha)}{(m+1)(1+\alpha)}$. Also,

$$\frac{Vol(E_+)}{Vol(E)} = (\frac{m}{m+1})(\frac{m^2}{m^2-1})^{\frac{m-1}{2}}(1-\alpha)(1-\alpha^2)^{\frac{m-1}{2}}.$$

For $\alpha = 0$, this ratio is less than $exp(\frac{-1}{2(m+1)})$.

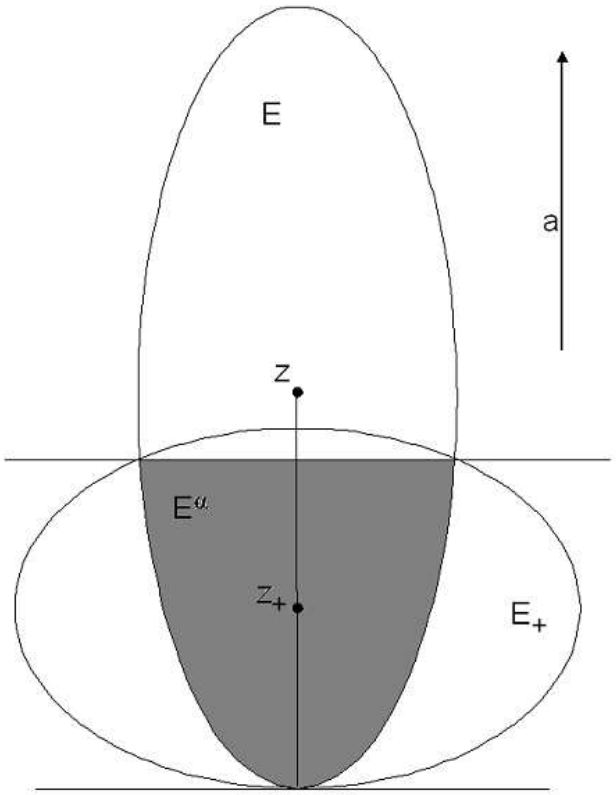

Figure 3: *An illustration of Theorem 2. Note that the parallel lines are the linear constraints discussed in the proof.*

Proof: We will not prove that E_+ is the minimum volume ellipsoid containing E_α, but we will show that the volume ratio between E and E_+ for $\alpha = 0$ holds as in the theorem, and thus our earlier conclusions regarding the necessary number of iterations are valid.

We begin by noting that $x \in E_\alpha$ implies that x satisfies the quadratic constraint $(x-z)^T B^{-1}(x-z) \leq 1$ and the linear constraint $a^T x \leq a^T z - \alpha\sqrt{a^T B a}$. But from Lemma 3, we know that any $x \in E$ also satisfies the linear constraint $a^T x \geq a^T z - \sqrt{a^T B a}$. Setting $\bar{a} = \frac{a}{\sqrt{a^T B a}}$, combining the two linear inequalities, and writing in terms of $(x-z)$ gives us a second quadratic inequality, $(\bar{a}^T(x-z)+\alpha)(\bar{a}^T(x-z)+1) \leq 0$. Now take a weighted combination of the two quadratic constraints, multiplying the first by $(1-\sigma)$ and the second by σ. We are left with

$$(x-z)^T((1-\sigma)B^{-1} + \sigma\bar{a}\bar{a^T})(x-z) + \sigma(1+\alpha)\bar{a}^T(x-z) \leq (1-\sigma) - \sigma\alpha \ .$$

We will complete this proof in the next lecture.