Ellipsoid method

- ellipsoid method
- convergence proof
- inequality constraints
Ellipsoid method

history

- developed by Shor, Nemirovski, Yudin in 1970s
- used in 1979 by Khachian to show polynomial solvability of LPs

properties

- each step requires cutting-plane or subgradient evaluation
- modest storage \((O(n^2))\)
- modest computation per step \((O(n^2))\), via analytical formula
- extremely simple to implement
- efficient in theory
- slow but steady in practice; rarely used
**Motivation**

**drawbacks of cutting-plane methods**

- serious computation needed to find next query point
typically, $O(n^2m)$ for analytic centering in ACCPM, with $m$ inequalities

- localization polyhedron grows in complexity as algorithm progresses
(with pruning, can keep $m$ proportional to $n$, e.g., $m = 4n$)

ellipsoid method addresses both issues, but retains theoretical efficiency
Ellipsoid algorithm for minimizing convex function

**idea:** localize $x^*$ in an *ellipsoid* instead of a *polyhedron*

**given** an initial ellipsoid $\mathcal{E}_0$ known to contain optimal set

**repeat** for $k = 1, 2, \ldots$

1. query oracle to get a neutral cut $a^T z \leq b$ at $x^{(k)}$, the center of $\mathcal{E}_{k-1}$
2. set $\mathcal{E}_k :=$ minimum volume ellipsoid covering $\mathcal{E}_{k-1} \cap \{z \mid a^T z \leq b\}$
**differences** with cutting-plane methods

- localization set doesn’t grow more complicated
- generating query point is trivial
- but, we add unnecessary points in step 2

**interpretation**

- reduces to bisection for $n = 1$
- can be viewed as an implementable version of the center-of-gravity cutting-plane method
Example
Updating the ellipsoid

\[ \mathcal{E} = \{ z | (z - x)^T P^{-1} (z - x) \leq 1 \} \]

\( \mathcal{E}^+ \) is min. volume ellipsoid covering

\[ \mathcal{E} \cap \{ z | g^T (z - x) \leq 0 \} \]

**update formula** (for \( n > 1 \)):

\[ \mathcal{E}^+ = \{ z | (z - x^+)^T (P^+)^{-1} (z - x^+) \leq 1 \} , \]

\[ x^+ = x - \frac{1}{n+1} P \tilde{g}, \quad P^+ = \frac{n^2}{n^2 - 1} \left( P - \frac{2}{n+1} P \tilde{g} \tilde{g}^T P \right) \]

where \( \tilde{g} = (1/\sqrt{g^T P g}) g \)

Ellipsoid method
Simple stopping criterion

for unconstrained problem of minimizing $f(x)$

lower bound on optimal value

$$
\begin{align*}
f(x^*) & \geq f(x^{(k)}) + g^{(k)T}(x^* - x^{(k)}) \\
& \geq f(x^{(k)}) + \inf_{z \in \mathcal{E}_{k-1}} g^{(k)T}(z - x^{(k)}) \\
& = f(x^{(k)}) - \sqrt{g^{(k)T}P^{(k-1)}g^{(k)}}
\end{align*}
$$

second inequality holds since $x^* \in \mathcal{E}_{k-1}$

simple stopping criterion to guarantee $f(x^{(k)}) - f(x^*) \leq \epsilon$:

$$
\sqrt{g^{(k)T}P^{(k-1)}g^{(k)}} \leq \epsilon
$$
Basic ellipsoid algorithm

ellipsoid described as

$$\mathcal{E}(x, P) = \{z \mid (z - x)^T P^{-1} (z - x) \leq 1\}$$

given ellipsoid $\mathcal{E}(x, P)$ containing $x^*$, accuracy $\epsilon > 0$

repeat

1. evaluate $g \in \partial f(x)$

2. if $\sqrt{g^T P g} \leq \epsilon$, return $x$; else, update ellipsoid

$$x := x - \frac{1}{n + 1} P \tilde{g}, \quad P := \frac{n^2}{n^2 - 1} \left( P - \frac{2}{n + 1} P \tilde{g} \tilde{g}^T P \right)$$

where $\tilde{g} = (1/\sqrt{g^T P g}) g$
Interpretation

• change coordinates

\[ \tilde{z} = P^{-1/2}z \]

so uncertainty is isotropic (same in all directions), \( i.e., E \) is unit ball

• take subgradient step with fixed length \( 1/(n + 1) \)

Shor calls ellipsoid method ‘gradient method with space dilation in
direction of gradient’
Improvements

• keep track of best upper and lower bounds:

\[ f_{\text{best}}^{(k)} = \min_{i=1,\ldots,k} f(x^{(i)}) \]

\[ l_{\text{best}}^{(k)} = \max_{i=1,\ldots,k} \left( f(x^{(i)}) - \sqrt{g(i)^T P(i-1) g(i)} \right) \]

stop when \( f_{\text{best}}^{(k)} - l_{\text{best}}^{(k)} \leq \epsilon \)

• propagate Cholesky factor of \( P \) (improves numerical stability)
Proof of convergence

assumptions: we consider the unconstrained problem

\[
\text{minimize } f(x)
\]

- \( f \) is Lipschitz: \(|f(y) - f(x)| \leq G\|y - x\|_2 \)
- \( \{x \mid f(x) \leq f^* + \epsilon\} \subseteq \mathcal{E}_0 \)
- \( \mathcal{E}_0 \) is ball with radius \( R \)

reduction of volume: can show that

\[
\text{vol } \mathcal{E}_{k+1} < e^{-\frac{1}{2n}} \text{vol } \mathcal{E}_k
\]

(reduction factor degrades rapidly with \( n \), compared to CG or MVE cutting-plane methods)
**proof.** suppose $f(x^{(i)}) > f^* + \epsilon$, $i = 1, \ldots, k$

- at iteration $i$ we only discard points with $f(z) \geq f(x^{(i)})$; therefore

$$\{z \mid f(z) \leq f^* + \epsilon\} \subseteq \mathcal{E}_k$$

- from Lipschitz condition, $\|z - x^*\|_2 \leq \epsilon/G$ implies $f(z) \leq f^* + \epsilon$; hence

$$B = \{z \mid \|z - x^*\|_2 \leq \epsilon/G\} \subseteq \mathcal{E}_k$$

- therefore $\text{vol } B \leq \text{vol } \mathcal{E}_k$, so

$$\alpha_n(\epsilon/G)^n \leq e^{-\frac{k}{2n}} \text{vol } \mathcal{E}_0 = e^{-\frac{k}{2n}} \alpha_n R^n$$

($\alpha_n$ is volume of unit ball in $\mathbb{R}^n$)

- this gives

$$k \leq 2n^2 \log(RG/\epsilon)$$
geometrical illustration

\[ B = \{ x \mid \|x - x^*\|_2 \leq \epsilon/G \} \]

**conclusion:** for \( k > 2n^2 \log(RG/\epsilon) \),

\[ f_{\text{best}}^{(k)} \leq f^* + \epsilon \]
Interpretation of complexity

• since $x^* \in \mathcal{E}_0 = \{x \mid \|x - x^{(1)}\|_2 \leq R\}$, our prior knowledge of $f^*$ is

$$f(x^{(1)}) - GR \leq f^* \leq f(x^{(1)})$$

our prior uncertainty in $f^*$ is $GR$

• after $k$ iterations our knowledge of $f^*$ is

$$f_{\text{best}}^{(k)} - \epsilon \leq f^* \leq f_{\text{best}}^{(k)}$$

posterior uncertainty in $f^*$ is $\leq \epsilon$

• iterations required:

$$2n^2 \log \frac{RG}{\epsilon} = 2n^2 \log_2 \left(\frac{RG}{\epsilon}\right) = 1.39 n^2 \log_2 \frac{\text{prior uncertainty}}{\text{posterior uncertainty}}$$

efficiency: $1/(1.39n^2) = 0.72/n^2$ bits per gradient evaluation
Example

\[
\text{minimize } \max_{i=1,\ldots,m} (a_i^T x + b_i)
\]

\[m = 100, \ n = 20, \ \|x^*\|_2 \approx 1.0, \ \text{start with } \mathcal{E} = \{x \mid \|x\|_2 \leq 10\}\]
Deep cut ellipsoid method

minimum volume ellipsoid containing ellipsoid intersected with halfspace

\[ \mathcal{E} \cap \{ z \mid g^T(z - x) + h \leq 0 \} \]

with \( h \geq 0 \), is given by

\[
\begin{align*}
x^+ &= x - \frac{1 + \alpha n}{n + 1} P \tilde{g} \\
P^+ &= \frac{n^2(1 - \alpha^2)}{n^2 - 1} \left( P - \frac{2(1 + \alpha n)}{(n + 1)(1 + \alpha)} P \tilde{g} \tilde{g}^T P \right)
\end{align*}
\]

where

\[
\tilde{g} = \frac{g}{\sqrt{g^T P g}}, \quad \alpha = \frac{h}{\sqrt{g^T P g}}
\]

(if \( \alpha > 1 \), intersection is empty)
Ellipsoid method with deep objective cuts

same example as on page 16

Ellipsoid method
Inequality constrained problems

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq 0, \quad i = 1, \ldots, m
\end{align*}
\]

- if \( x^{(k)} \) is feasible, update ellipsoid with objective cut

\[
g_0^T z \leq g_0^T x^{(k)} - f_0(x^{(k)}) + f^{(k)}_{\text{best}}, \quad g_0 \in \partial f_0(x^{(k)})
\]

\( f^{(k)}_{\text{best}} \) is best objective value of feasible iterates so far

- if \( x^{(k)} \) is infeasible (say, \( f_j(x^{(k)}) > 0 \)), use feasibility cut

\[
g_j^T z \leq g_j^T x^{(k)} - f_j(x^{(k)}), \quad g_j \in \partial f_j(x^{(k)})
\]
Stopping criterion

- if \( x^{(k)} \) is feasible, we have lower bound on \( p^* \) as before:

\[
p^* \geq f_0(x^{(k)}) - \sqrt{g_0^{(k)T} P^{(k-1)} g_0^{(k)}}
\]

- if \( x^{(k)} \) is infeasible, we have for all \( x \in \mathcal{E}_{k-1} \)

\[
f_j(x) \geq f_j(x^{(k)}) + g_j^{(k)T}(x - x^{(k)})
\]

\[
\geq f_j(x^{(k)}) + \inf_{z \in \mathcal{E}_{k-1}} g^{(k)T}(z - x^{(k)})
\]

\[
= f_j(x^{(k)}) - \sqrt{g_j^{(k)T} P^{(k-1)} g_j^{(k)}}
\]

hence, problem is infeasible if for some \( j \),

\[
f_j(x^{(k)}) - \sqrt{g_j^{(k)T} P^{(k-1)} g_j^{(k)}} > 0
\]
**stopping criteria:** terminate algorithm when

- $x^{(k)}$ is known to be $\epsilon$-suboptimal:

  $$x^{(k)} \text{ is feasible and } \sqrt{g_0^{(k)T} P^{(k-1)} g_0^{(k)}} \leq \epsilon$$

- or problem is shown to be infeasible:

  $$f_j(x^{(k)}) - \sqrt{g_j^{(k)T} P^{(k-1)} g_j^{(k)}} > 0 \text{ for some } j$$