Ellipsoid method

- ACCPM

given an initial polyhedron P_0 known to contain X

$k := 0$

repeat {

compute $x^{(k+1)} = A_C(P_k)$

Query cutting-plane oracle at $x^{(k+1)}$

if $x^{(k+1)} \in X$ quit

else add returned cutting plane inequality to P

$P_{k+1} = P_k \cap \{z \mid a^T z \le b\}$

if $P_{k+1} = \phi$ quit

$k := k+1$

- Simplex and polyhedron

define a bounded polyhedron in R^{20}

when we have 8 inequalities $a_i^T x \le b_i$

we have a 12-dimensional subspace that's orthogonal to those a_i

the answer is 21

A simplex is the simplest polyhedron that's bounded (n dimension → n+1 inequalities)

- Ellipsoid method

each requires cutting-plane and subgradient evaluation

modest storage $O(n^2)$ and modest computation per step $O(n^2)$

efficient in theory, slow but steady and robust in practice

- Motivation

in cutting-plane method

Serious computation is needed to find next query point

(typically $O(n^2 m)$, with not small constant)

localization polyhedron grows in complexity as algorithm progress

(not a problem in practice by pruning constraints)

Ellipsoid algorithm for minimizing convex function

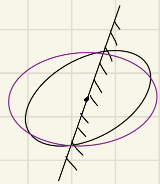
idea: localize x^* in a ellipsoid instead of a polyhedron

1. At iteration k we know $x^* \in \mathcal{E}^{(k)}$
2. Set $x^{(k+1)} := \text{center}(\mathcal{E}^{(k)})$, evaluate $g^{(k)} \in \partial f(x^{(k+1)})$

   we have cutting plane $g^T(z-x) \le 0$ $\quad f(x^*) \ge f(x) + g^T(x^* - x)$ $\quad g^T(x-z) \ge 0 \Rightarrow f(z) \ge f(x)$
3. we know (ignorance reduction by a factor of 2)

   $x^* \in \mathcal{E}^{(k)} \cap \{z \mid g^{(k)T}(z-x) \le 0\}$
4. Set $\mathcal{E}^{(k+1)} =$ minimum volume ellipsoid covering ignorance increasing

   $\mathcal{E}^{(k)} \cap \{z \mid g^{(k+1)T}(z-x) \le 0\}$



$\{z \mid g^T(z-x) \le 0\}$

compared to cutting-plane method

- ☺ localization set (ellipsoid) doesn't grow more complicated
- ☺ easy to compute query point
- ☹ add unnecessary points in step 4

Properties of ellipsoid method

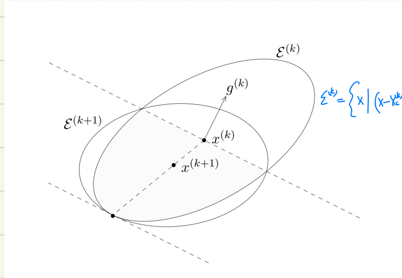
reduce to bisection for $n=1$

simple formula for $\mathcal{E}^{(k+1)}$ given $\mathcal{E}^{(k)}$, $g^{(k+1)}$

$\mathcal{E}^{(k+1)}$ can be larger than $\mathcal{E}^{(k)}$ in diameter (max semi-axis length), but is always smaller in volume

$\text{vol}(\mathcal{E}^{(k+1)}) < e^{-\frac{1}{2n}} \text{vol}(\mathcal{E}^{(k)})$ volume reduction factor degrades rapidly with n

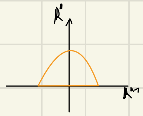
· Updating the ellipsoid



$\mathcal{E}^{(k)} = \{x \mid (x-x_c^{(k)})^T P^{(k)-1} (x-x_c^{(k)}) = 1\}$

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

can affinely change coordinates with the problem staying the same; don't have to consider the general problem, can transform any half-ellipsoid to a half unit ball pointing upwards



completely symmetric and circular in remaining n-1 dimensions

∴ the minimum volume ellipsoid that covers the half ellipsoid has to be symmetric in these n-1 dimensions

in convex optimization, if the problem has a symmetry, the solution must have the same symmetry (absolutely false for non-convex problems)

In ellipsoid method, each step we compute a cutting plane at the center of localization ellipsoid

$\mathcal{H} = \{x \mid g^T(x-x_c) \le 0\}$ $\quad \mathcal{E} = \{x \in \mathbb{R}^n \mid (x-x_c) P^{-1} (x-x_c) \le 1\}$

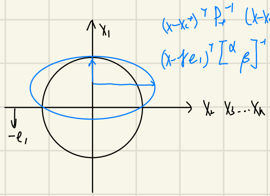
and update the localization ellipsoid to the minimum volume ellipsoid that covers $\mathcal{E} \cap \mathcal{H}$

(a) we first consider the special case:

$\mathcal{E}$ is the unit ball centered at the origin ($P = I$, $x_c = 0$) and $g = -e_1$ $\quad e_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$

$(x-x_c^+)^T P_+^{-1} (x-x_c) = 1$

$(x-\gamma e_1)^T \begin{bmatrix} \alpha & \\ & \beta \end{bmatrix}^{-1} (x-\gamma e_1) = 1$



Since $\mathcal{E} \cap \mathcal{H}$ (a half sphere) is symmetric about e_1, the minimum volume ellipsoid covers $\mathcal{E} \cap \mathcal{H}$ will have the same symmetry.

∴ P^+ is diagonal, as has n-1 same diagonal entries

symmetry: $P^+ = \mathrm{diag}(\alpha, \beta, \beta, \ldots)$ $\quad x_c^+ = \gamma e_1$

⇓ form convex optimization problem

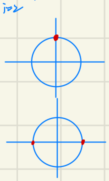
$\min.\ \alpha \cdot \beta^{n-1}$

s.t. $(x-\gamma e_1)^T \begin{bmatrix} \alpha & \\ & \beta \end{bmatrix}^{-1} (x-\gamma e_1) \le 1$ for all $x \in \{x \mid x_1 \ge 0,\ x^T x = 1\}$

$(x-\gamma e_1)^T \begin{bmatrix} \alpha & \\ & \beta \end{bmatrix}^{-1} (x-\gamma e_1) = \frac{(x_1-\gamma)^2}{\alpha} + \frac{1}{\beta} \sum_{i=2}^{n} x_i^2 \le 1$ for all $x \in \{x \mid x_1 \ge 0,\ x^T x = 1\}$

① when $x_1 = 1$, $x_2 = x_3 = \cdots = x_n = 0$: $\frac{(1-\gamma)^2}{\alpha} \le 1$

② when $x_1 = 0$, $x_2^2 + \cdots + x_n^2 = 1$: $\frac{\gamma^2}{\alpha} + \frac{1}{\beta} \le 1$



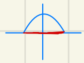
2 necessary conditions for the inequality constraints.

when ① and ② hold:

$$\frac{(x_1-\gamma)^2}{\alpha} + \frac{1}{\beta}\sum_{i=2}^{n} x_i^2 \le \frac{(x_1-\gamma)^2}{\alpha} + \left(1-\frac{\gamma^2}{\alpha}\right)\sum_{i=2}^{n} x_i^2$$

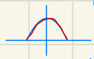
consider the boundary. (when the boundary is in the covering ellipsoid, the interior is)

1° $x_1=0$ $\sum_{i=2}^{n} x_i^2 \le 1$



$$\frac{(x_1-\gamma)^2}{\alpha} + \frac{1}{\beta}\sum_{i=2}^{n} x_i^2 = \frac{\gamma^2}{\alpha} + \frac{1}{\beta}\sum_{i=2}^{n} x_i^2 = \frac{\gamma^2}{\alpha} + \frac{1}{\beta}\sum_{i=2}^{n} x_i^2 \le \frac{\gamma^2}{\alpha} + \frac{1}{\beta} \le 1$$

2° $x_1 \ge 0$ $x^Tx=1$



$$\frac{(x_1-\gamma)^2}{\alpha} + \frac{1}{\beta}\sum_{i=2}^{n} x_i^2 \le \frac{(x_1-\gamma)^2}{\alpha} + \left(1-\frac{\gamma^2}{\alpha}\right)(1-x_1^2)$$

$$= \frac{x_1^2-2\gamma x_1+\gamma^2}{\alpha} - \frac{\gamma^2(1-x_1^2)}{\alpha} + (1-x_1^2)$$

$$= \frac{x_1^2-2\gamma x_1+\gamma^2 x_1^2}{\alpha} + (1-x_1^2)$$

$$\le \frac{x_1^2(1-2\gamma+\gamma^2)}{\alpha} + (1-x_1^2) = \frac{(1-\gamma)^2}{\alpha}x_1^2 + 1 - x_1^2 \le 1$$

$\{x \mid x_1=1\ x_2^2+\cdots+x_n^2=0\}$ and $\{x \mid x_1=0\ x^Tx=1\}$ satisfy $(x-x_c)^T\begin{bmatrix}\alpha & \\ & \beta\end{bmatrix}^{-1}(x-x_c)\le 1$ $\Leftrightarrow$ $\frac{(1-\gamma)^2}{\alpha}\le 1$ $\frac{\gamma^2}{\alpha}+\frac{1}{\beta}\le 1$

$\Updownarrow$

$\{x \mid x_1\ge 0\ x^Tx\}$ satisfy $(x-x_c)^T\begin{bmatrix}\alpha & \\ & \beta\end{bmatrix}^{-1}(x-x_c)\le 1$

produce an equivalent optimization problem

min. $\alpha\beta^{n-1}$

s.t. $(1-\gamma)^2/\alpha \le 1$

$\frac{\gamma^2}{\alpha}+\frac{1}{\beta}\le 1$

when $\frac{\gamma^2}{\alpha}+\frac{1}{\beta}<1$ can take $\tilde\beta<\beta$ s.t. $\frac{\gamma^2}{\alpha}+\frac{1}{\tilde\beta}=1$ $\alpha\tilde\beta^{n-1}<\alpha\beta^{n-1}$

∴ when $\alpha\beta^{n-1}$ achieves optimum, $\frac{\gamma^2}{\alpha}+\frac{1}{\beta}=1$ i.e. constraint 2 is tight

when $\frac{(1-\gamma)^2}{\alpha}<1$ and $\frac{\gamma^2}{\alpha}+\frac{1}{\beta}=1$

take $\tilde\gamma=\sqrt{c}\,\gamma$ $\tilde\alpha=c\alpha$ st. $\frac{(1-\tilde\gamma)^2}{\tilde\alpha}=1$

$1-2\sqrt{c}\,\gamma+c\gamma^2-c\alpha=0$

$(\gamma^2-\alpha)c-2\sqrt{c}\,\gamma+1=0$

$\sqrt{c}=\frac{2\gamma\pm\sqrt{4\gamma^2-4(\gamma^2-\alpha)}}{2(\gamma^2-\alpha)}=\frac{\gamma\pm\sqrt{\alpha}}{\gamma^2-\alpha}=\frac{1}{\gamma\pm\sqrt{\alpha}}$

∵ $\frac{\gamma^2}{\alpha}+\frac{1}{\beta}=1$ ∴ $\gamma^2<\alpha$

∴ $\sqrt{c}=\frac{1}{\gamma+\sqrt{\alpha}}>0$

∵ $\frac{(1-\gamma)^2}{\alpha}<1$ $-\sqrt{\alpha}<1-\gamma<\sqrt{\alpha}$ $\sqrt{\alpha}+\gamma>1$

∴ $\sqrt{c}<1$ $c<1$

$\tilde\alpha=c\alpha<\alpha$

$\tilde\gamma=\sqrt{c}\,\gamma<\gamma$

$\tilde\alpha\beta^{n-1}<\alpha\beta^{n-1}$

∴ when $\alpha\beta^{n-1}$ achieves optimum $\frac{(1-\gamma)^2}{\alpha}=1$

when the optimization achieves optimum, the inequalitys are tight min. $\alpha\cdot\beta^{n-1}$

s.t. $\frac{(1-\gamma)^2}{\alpha}=1 \quad \frac{\gamma^2}{\alpha}+\frac{1}{\beta}=1$

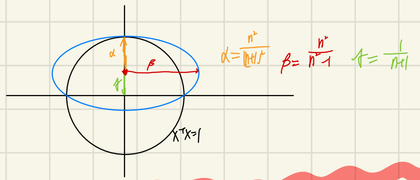
$$\alpha=(1-\gamma)^2 \qquad \beta=\frac{1}{1-\frac{\gamma^2}{\alpha}}=\frac{(1-\gamma)^2}{(1-\gamma)^2-\gamma^2}=\frac{(1-\gamma)^2}{1-2\gamma}$$

$$\alpha\cdot\beta^{n-1}=(1-\gamma)^2\cdot\frac{(1-\gamma)^{2n-2}}{(1-2\gamma)^{n-1}}=\frac{(1-\gamma)^{2n}}{(1-2\gamma)^{n-1}}$$

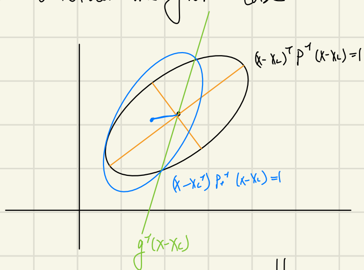
$$\log\alpha\cdot\beta^{n-1}=2n\cdot\log(1-\gamma)-(n-1)\cdot\log(1-2\gamma)$$

$$\frac{d}{d\gamma}\left(\log\alpha\cdot\beta^{n-1}\right)=\frac{2n}{\gamma-1}+\frac{2(n-1)}{1-2\gamma}:=0$$

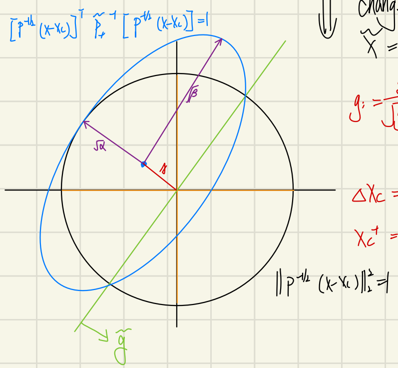
$$\gamma=\frac{1}{n+1} \qquad \alpha=(1-\gamma)^2=\frac{n^2}{(n+1)^2} \qquad \beta=\frac{(1-\gamma)^2}{1-2\gamma}=\frac{n^2}{n^2-1}$$



(b). consider the general case



$\Downarrow$ change coordinate $\tilde{x}=P^{-1/2}(x-x_c)$



$$\begin{cases}\tilde{g}^T P^{-1/2}(x-x_c)=0\\ g^T(x-x_c)=0\end{cases} \qquad \tilde{g}=P^{1/2}g$$

$g:=\frac{g}{\sqrt{g^TPg}}$ so that $\|P^{1/2}g\|=1$

$$\Delta x_c=-\gamma P^{1/2}\tilde{g}=-\gamma Pg$$

$$x_c^+=x_c+\Delta x_c=x_c-\frac{1}{n+1}Pg$$

$$\begin{cases}(x-x_c^+)^T P^{-1/2}\,\tilde{P}_+^{-1}\,P^{-1/2}(x-x_c^+)=1\\ (x-x_c^+)\,P_+^{-1}\,(x-x_c^+)=1\end{cases}$$

$$P_+=P^{1/2}\tilde{P}_+P^{1/2}$$

$\tilde{P}_+$ has $n-1$ eigenvalues β, 1 eigenvalues α with eigenvector $\tilde{g}$

$$v^T\tilde{P}_+v=\beta v^Tv-\beta\|\tilde{g}^Tv\|_2^2+\alpha\|\tilde{g}^Tv\|_2^2$$
$$=v^T\beta I v-v^T\beta\tilde{g}\tilde{g}^Tv+v^T\alpha\tilde{g}\tilde{g}^Tv$$
$$=v^T[\beta I-\beta\tilde{g}\tilde{g}^T+\alpha\tilde{g}\tilde{g}^T]v$$

$$\tilde{P}_+=\beta I-\beta\tilde{g}\tilde{g}^T+\alpha\tilde{g}\tilde{g}^T$$
$$=\beta I-(\beta-\alpha)P^{1/2}gg^TP^{1/2}$$

$$P_+=P^{1/2}\tilde{P}_+P^{1/2}=\frac{n^2}{n^2-1}\left(P-\frac{2}{n+1}Pgg^TP\right)$$

- Basic ellipsoid method

given an initial ellipsoid $(P^{(0)}, x^{(0)})$ containing a minimizer of f

$k=0$

repeat:

compute a subgradient $g \in \partial f(x^{(k)})$

normalize the subgradient $\hat{g} := \frac{1}{\sqrt{g^T P^{(k)} g}} g$

update ellipsoid: $x^{(k+1)} = x^{(k)} - \frac{1}{n+1} P^{(k)} \hat{g}$

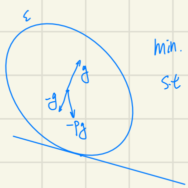
$$P^{(k+1)} = \underbrace{\frac{n^2}{n^2-1}}_{\text{where ellipsoid get bigger}} \left(P^{(k)} - \underbrace{\frac{2}{n+1} P^{(k)} \hat{g} \hat{g}^T P^{(k)}}_{} \right)$$ rank 1 update

$k := k+1$

p shrink only in direction $\hat{g}$.

that's where the ellipsoid get smaller } interpretation 1

(noisy measurement)



min. $g^T v$
s.t $v \in E$ } interpretation 2
(like newton's method)

gradient method $\Longrightarrow$ newton's method

subgradient method $\Rightarrow$ ellipsoid method

- Simple stopping criterion

$$f(x^*) \geq f(x^{(k)}) + g^{(k)T}(x^* - x^{(k)})$$

$$\geq f(x^{(k)}) + \inf_{z \in E^{(k)}} g^{(k)T}(z - x^{(k)})$$

$$= f(x^{(k)}) - \sqrt{g^{(k)T} P^{(k)} g^{(k)}}$$ suboptimality