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

$$P^{(k+1)} = \frac{n^2}{n^2-1}\left(P^{(k)} - \frac{2}{n+1} P^{(k)} \hat{g}\hat{g}^T P^{(k)}\right)$$

$k := k+1$

- Ellipsoid method improvement

① keep track of best upper and lower bound

$$U_k = \min_{i=1\cdots k} f(x^{(i)})$$

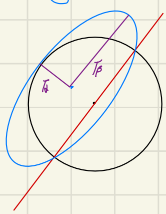
$$l_k = \max_{i=1\cdots k} \left(f(x^{(i)}) - \sqrt{g^T P^{(i)} g}\right.$$

Stop when $U_k - l_k \le \epsilon$

② can propagate cholesky factor of $P$

( avoid problem of $P \not\succ 0$ due to numerical roundoff

- Proof of convergence




$$\frac{\mathrm{vol}(\mathcal{E}^{(k+1)})}{\mathrm{vol}(\mathcal{E}^{(k)})} = (\sqrt{\beta})^{n-1} \cdot \sqrt{\alpha} = \left(\frac{n^2}{n^2-1}\right)^{\frac{n-1}{2}} \cdot \frac{n}{n+1}$$

$$= \left(\frac{n}{n+1}\right)^{\frac{n+1}{2}} \cdot \left(\frac{n}{n-1}\right)^{\frac{n-1}{2}} < e^{-\frac{1}{2n}}$$

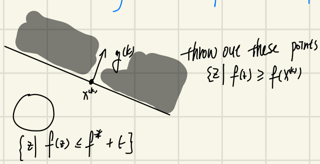
define $G = \max_{g \in \partial f(x),\ x \in \mathcal{E}^{(0)}} \|g\|$ ( $G$ is the Lipschitz constant over the initial ellipsoid )

$\mathcal{E}^{(0)}$ is ball with radius $R$

suppose $f(x^{(i)}) > f^* + \epsilon$ ( fail to find a $\epsilon$-suboptimal point )

then $f(x) \le f^* + \epsilon \Rightarrow x \in \mathcal{E}^{(i)}$

since at iteration $i$ we only discard point with $f \ge f(x^{(i)})$



from Lipschitz condition

$$\|x-x^*\| \le \epsilon/G \Rightarrow f(x) \le f(x^*)+\epsilon \Rightarrow x \in \mathcal{E}^{(k)}$$

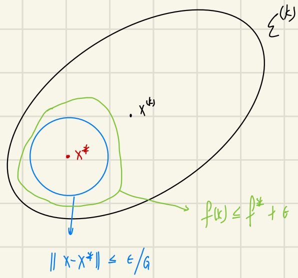
∴ the ball $\{z \mid \|z-x^*\| \le \epsilon/G\} \in \mathcal{E}^{(k)}$

∴ $vol(B) \le vol(\mathcal{E}^{(k)})$

∴ $a_n(\epsilon/G)^n \le e^{-\frac{k}{2n}} vol(\mathcal{E}^{(0)}) = e^{-\frac{k}{2n}} a_n \cdot R^n$

($a_n$ is the volume of unit ball in $\mathbb{R}^n$

∴ $k \le 2n^2 \log(RG/\epsilon)$




## · Interpretation of complexity

Since $x^* \in \mathcal{E}_0 = \{x \mid \|x-x^{(0)}\| \le R\}$, our prior knowledge is

$$f^* \in \{f(x^{(0)}) - GR,\ f(x^{(0)})\}$$

our prior knowledge uncertainty is $GR$ (the gap)

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

$$f^* \in \{\min_{i=0,\dots,k} f(x^{(i)}) - \epsilon \quad \min_{i=0,\dots,k} f(x^{(i)})\}$$

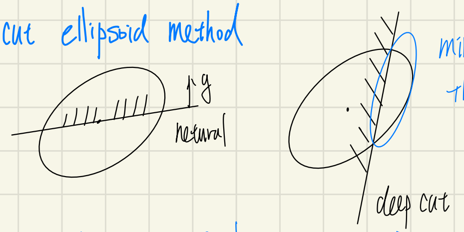
iterations required

$$2n^2 \log \frac{RG}{\epsilon} = 2n^2 \log \frac{\text{prior uncertainty}}{\text{posterior uncertainty}}$$

efficiency $\quad 0.72/n^2$ bits per subgradient evaluation

(like bisection)

## · Deep-cut ellipsoid method




minimum volume ellipsoid that covers a "less than half ellipsoid"

$g^T(z-x) + f(x) - f^{(k)}_{best} \le 0$

the half space $\{z \mid g^T(z-x) \le -h\}$ $\quad g^T(z-x) \le 0$ for neutral cut

$$x^+ = x - \frac{1-\alpha n}{n+1} P\tilde{g} \qquad \tilde{g} = g/\sqrt{g^T P g}$$

$$P^+ = \frac{n^2(1-\alpha^2)}{n^2-1}\left(P - \frac{2(1+\alpha n)}{(n+1)(1+\alpha)} P\tilde{g}\tilde{g}^T P\right) \qquad \alpha = h/\sqrt{g^T P g}$$

longer step length in the same direction as ellipsoid method

- Inequality constrained problems

$$\min.\ f_0(x)$$
$$\text{s.t.}\ f_i(x) \le 0$$

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

$$g_0^T(z - x^{(k)}) + f_0(x^{(k)}) - f_{best}^{(k)} \le 0$$

if $x^{(k)}$ is infeasible update ellipsoid with feasibility cut

$$g_i^T(z - x^{(k)}) + f_i(x^{(k)}) \le 0$$

everything that violate this inequality is guaranteed to be infeasible (cut off the infeasible set)

assuming $f_i(x^{(k)}) > 0$

- Stopping criterion

if $x^{(k)}$ is feasible, we have lower bound on $p^*$

$$f(x^*) \ge f(x^{(k)}) + g^{(k)T}(x^* - x^{(k)})$$
$$\ge f(x^{(k)}) + \inf_{z \in \mathcal{E}^{(k)}} g^{(k)}(z - x^{(k)})$$
$$= f(x^{(k)}) - \sqrt{g^{(k)} P g^{(k)}}$$

suboptimality

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

$$f_j(x) \ge f_j(x^{(k)}) + g_j^{(k)T}(x - x^{(k)})$$
$$\ge f_j(x^{(k)}) + \inf_{z \in \mathcal{E}^{(k)}} g^{(k)T}(z - x^{(k)})$$
$$= f_j(x^{(k)}) - \sqrt{g_j^{(k)} P g_j^{(k)}}$$

if $f_j(x^{(k)}) > \sqrt{g_j^{(k)} P g_j^{(k)}}$ problem is infeasible

# Decomposition method (encapsulation and layering)

- Separable problem

$$\min.\ f_1(x_1) + f_2(x_2)$$
$$\text{s.t.}\quad x_1 \in C_1 \quad x_2 \in C_2$$

( block-diagonal hessian saves much computation)

can solve for $x_1$ and $x_2$ separately (in parallel)
two problems are separable or trivially parallelizable
generalizes to any objective of form $\Psi(f_1\ f_2)$ with $\Psi$ non-decreasing

- Complicating variable

$$\min_{x_1, x_2, y}\ f(x) = f_1(x_1, y) + f_2(x_2, y)$$

$y$ is the complicating variable or coupling variable, when $y$ is fixed, the problem is separable in $x_1$ and $x_2$
$x_1$ $x_2$ are private or local variables, $y$ is a public or interface or boundary variable between the two subproblems

- Primal decomposition

fix $y$ and define

subproblem 1: $\underset{x_1}{\text{minimize}}\ f_1(x_1, y)$
subproblem 2: $\underset{x_2}{\text{minimize}}\ f_2(x_2, y)$

and optimal values $\phi_1(y)$ and $\phi_2(y)$

The original problem is equivalent to master problem

$$\min_y\ \phi_1(y) + \phi_2(y)$$

if $f$ is convex in $x$ and $y$, $C$ is a convex set
$g(x) = \inf_{y \in C} f(x, y)$ is convex

if the original problem is convex, so is the master problem
can solve the master problem using:
- bisection (if $y$ is scalar)
- gradient method (if $\phi$ is differentiable)
- subgradient, cutting-plane or ellipsoid method

each iteration of master problem requires solving the two subproblems

- Primal decomposition algorithm (using subgradient method)

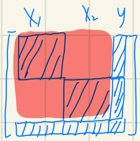
repeat

1. find $x_1$ that minimizes $f_1(x_1, y)$ and a subgradient $g_1 \in \partial \phi_1(y)$
   find $x_2$ that minimizes $f_2(x_2, y)$ and a subgradient $g_2 \in \partial \phi_2(y)$
2. update the complicating variable
   $y := y - \alpha_k(g_1 + g_2)$ $\alpha_k$ can be chosen in any of the standard ways

$f_1(x_1, y) + f_2(x_2, y)$

smart linear algebra

the hessian looks like




arrow - shape

the block-diagonal is easily-invertiable

- Dual decomposition

min. $f_1(x_1, y) + f_2(x_2, y)$

step 1: introducing new variable

min. $f_1(x_1, y_1) + f_2(x_2, y_2)$

s.t. $y_1 = y_2$

- $y_1, y_2$ are local versions of complicating variable
- $y_1 = y_2$ is consistency constraint

step 2: form dual problem

$\mathcal{L}(x_1, y_1, x_2, y_2, v) = f_1(x_1, y_1) + f_2(x_2, y_2) + v^T(y_1 - y_2)$

$g(v) = \inf_{x_1, y_1, x_2, y_2} \mathcal{L}$

separable: can minimize over $(x_1, y_1)$ and $(x_2, y_2)$ separately

$g_1(v) = \inf_{x_1, y_1} f_1(x_1, y_1) + v^T y_1$

$g_2(v) = \inf_{x_2, y_2} f_2(x_2, y_2) - v^T y_2$

dual problem is: max. $g(v) = g_1(v) + g_2(v)$

computing $g_i(v)$ are the dual subproblems

can be done in parallel

a subgradient of $-g(v)$ is $\underbrace{y_2 - y_1}_{\text{inconsistency}}$ ( $g(v)$ is concave $-g(v)$ is convex)

· Dual decomposition algorithm (using subgradient method)

repeat

1. Solve the dual subproblem

   Find $x_1$ $y_1$ that minimize $f_1(x_1, y_1) + v^T y_1$

   Find $x_2$ $y_2$ that minimize $f_2(x_2, y_2) - v^T y_2$

2. update dual variable (prices)

   $v := v - \alpha_k (y_2 - y_1)$

at each step we have a lower bound $g(v)$ on $p^*$

when consistency is achieved, $y_2 - y_1 = 0$, $v$ maximizes $g(v)$