AA Lecture 20 3/25

Ellipsoid Algorithm

Feasibility problem: $F \triangleq \{ x \in \mathbb{R}^n : Ax \le b \}$

find $x \in F$. (or report $F = \emptyset$).

Approx: given $\varepsilon > 0$, either find $x \in F$

or report $\mathrm{vol}(F) < \varepsilon$.

For orig. problem enough $\varepsilon = \exp(-\mathrm{poly}(n,m)B)$.

Solving Approx. Feasibility:

Def: fix $r > 0$, ball $B_r(y) = \{ x \in \mathbb{R}^n : \|x - y\| \le r \}$

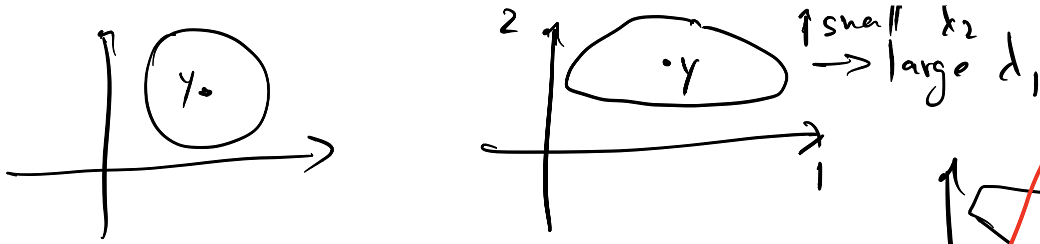
Def: axis-aligned ellipsoids

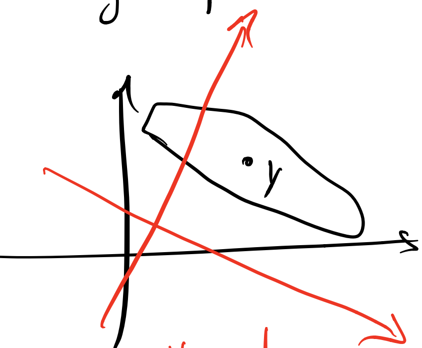
fix $\lambda_1, \dots, \lambda_n$, $r > 0$:

$$E_{r,\lambda}(y) = \left\{ x : \sum_{i=1}^{n} \left(\frac{x_i - y_i}{\lambda_i} \right)^2 \le r^2 \right\}.$$

Def: general ellipsoids fix $r > 0$, rank-n matrix $A \in \mathbb{R}^{n \times n}$:

$$E_{r,A}(y) = \{ x + y : x^T A^T A x \le r^2 \}$$
$$= \{ x + y : \|Ax\|_2^2 \le r^2 \}.$$





$$E_{r,\Lambda}(y) = \{ x+y : \| \Lambda R x \|^2 \le r^2 \}$$

rotation (unitary)

diagonal matrix.

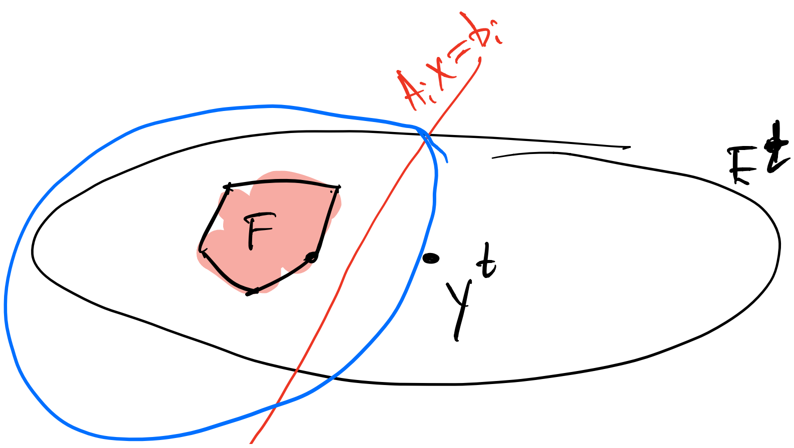
Ellipsoid Algos

Idea: – start with E^0 = some starting ellipsoid.

s.t. $F \subseteq E^0$.

– iterate, and compute E^t at step t s.t.: 1) $F \subseteq E^t$.

2) $\mathrm{vol}(E^t) < \mathrm{vol}(E^{t-1})$.



Algo: 1) start with $E^0 \supseteq F$. $E^0 = B_R$ with R large enough.

$y^0 =$ center of E^0.

2) for $t = 0 \ldots T$:

3) if $y^t \in F$, done: report $x = y^t$.

4) if not, $\exists$ constraint violated:

$$A_i y^t > b_i.$$

$$\Rightarrow F \subseteq E^t \cap \{A_i x \le b_i\} \subsetneq E^t.$$

5) define $E^{t+1}, y^{t+1} =$ smallest ellipsoid containing $E^t \cap \{A_i x \le b_i\}$

6) terminate if $\mathrm{vol}(E^{t+1}) < \varepsilon$.

Easy check: if algo terminates it's correct.

Main question: how many iterations? (what is T).

Claim: $\mathrm{vol}(E^{t+1}) \le \mathrm{vol}(E^t) \cdot \left(1 - \frac{1}{2n}\right)$.

$$\Rightarrow \mathrm{vol}(E^T) \le \mathrm{vol}(E^0)\cdot\left(1-\frac{1}{2n}\right)^T$$

$$\le (2R)^n \cdot \left(1-\frac{1}{2n}\right)^T.$$

when $\uparrow \le \varepsilon$?

$$(2R)^n \cdot \left(1-\frac{1}{2n}\right)^T \le \varepsilon$$

$$-n \cdot \lg 2R + T \cdot \frac{\Theta(1)}{2n} \ge \lg 1/\varepsilon$$

$$\Rightarrow T \ge \Theta(n) \cdot \left[\lg 1/\varepsilon + n \cdot \mathrm{poly}(n,m) \cdot B\right].$$

$\Rightarrow$ # iterations is $\simeq \Theta(n \cdot \lg 1/\varepsilon + \mathrm{poly}(n,m)B)$.

$$\varepsilon = \exp(-\mathrm{poly}(n,m)B).$$

$\Rightarrow$ runtime overall is $\mathrm{poly}(n,m,B)$.

<u>Remark:</u> – the only "access" to input constraints $Ax \le b$ was through "separation oracle":

SO_F: given $y \in \mathbb{R}^n$
– either report $y \in F$.

– or return some constraint hyperplane $Ax = b$ s.t. F & y are on opposite sides.

$\#$ oracle calls $= \Theta(n) \cdot \lceil \lg 1/\varepsilon + n \cdot \lg R \rceil$.

Applicable to situations where # constraints may be exponential!

Gradient Descent:

unconstrained opt: $\min\limits_{x \in \mathbb{R}^n} f(x)$.

Constrained opt: $\min\limits_{\substack{x \in \mathbb{R}^n \\ x \in F}} f(x)$.

Reduction from const. opt. to unconstr. o.

$$g(x) \triangleq \begin{cases} f(x), & x \in F \\ +\infty, & x \notin F. \end{cases}$$

Enough to solve $\min_{x \in \mathbb{R}^n} g(x)$.

GD: - start with $x^0 \in \mathbb{R}^n$.

- at time $t = 0..T$:

compute x^{t+1} = function of x^t, $f(x^t)$...

Main questionss 1) how to define x^{t+1} from x^t.

2) #iteration T to find global min

$$\min_{x \in \mathbb{R}^n} f(x).$$

Assume that f is "nice" function, so that we can approx/write Taylor expansion of f:

$$f(\underset{\mathbb{R}^n}{x} + \underset{\mathbb{R}^n}{\delta}) = f(x) + \nabla f(x)^T \cdot \delta + \delta^T \nabla^2 f(y) \cdot \delta.$$

$y \in [x, x+\delta]$.

Def: $\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), \dots \right) \in \mathbb{R}^n$

Hessian: $\nabla^2 f(x) =$ matrix $n \times n$ with

$$\text{entry}_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}(x).$$

Idea of GD: 1) $x^{t+1} = x^t + \delta$.

2) think of δ as being small and compute best δ s.t.

$$\min_{\delta} f(x^t + \delta).$$

Suppose $\|\delta\| \le \varepsilon$. How to choose δ?

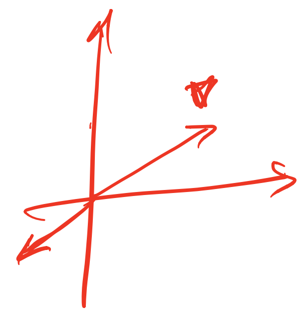
Would like: $\delta = \underset{\|\delta\| \le \varepsilon}{\arg\min} \; f(x) + \nabla f^T(x) \cdot \delta + \delta^T \cdot \nabla^2 f(x) \cdot \delta$

Option 1: note that:

$$f(x+\delta) \approx f(x) + \nabla f(x)^T \cdot \delta + O(\varepsilon^2).$$

ignore $+ O(\varepsilon^2)$ term, $\ll \nabla f(x)^T \cdot \delta$.

$$\delta = \underset{\|\delta\| \le \varepsilon}{\arg\min} \; f(x) + \nabla f(x)^T \cdot \delta$$



$$= \underset{\|\delta\| \le \varepsilon}{\arg\min} \; \nabla f(x)^T \cdot \delta$$

$$= \eta \cdot (-\nabla f(x)) \quad \text{s.t. } \|\eta \cdot \nabla f(x)\| = \varepsilon.$$

$$\Leftrightarrow \quad \eta = \frac{\varepsilon}{\|\nabla f(x)\|}.$$

Basic GD algo: $x^{t+1} = x^t - \underbrace{\eta \cdot \nabla f(x)}_{\delta}$.

Def: $\beta > 0$. Call f β-smooth if:

$\forall x, y \in \mathbb{R}^n$: $\|\nabla f(x) - \nabla f(y)\| \le \beta \cdot \|x - y\|$

$\Leftrightarrow$ $\forall \delta \in \mathbb{R}^n$: $\delta^T \nabla^2 f(x) \delta \le \beta \cdot \|\delta\|^2$

$\Leftrightarrow$ max eigenvalue of $\nabla^2 f(x) \le \beta$.

Rederive opt. δ for GD for β-sm.

Taylor exp:

$$f(x+\delta) = f(x) + \nabla f(x)^T \cdot \delta + \delta^T \nabla^2 f(y)\, \delta / 2$$

$$\leq f(x) + \nabla f(x)^T \cdot \delta + \beta \cdot \|\delta\|^2 / 2$$

<u>Option 2</u>: find best step δ:

$$\delta = \operatorname*{argmin}_{\delta} f(x) + \nabla f(x)^T \cdot \delta + \beta \cdot \|\delta\|^2 / 2$$

$$= \operatorname*{argmin}_{\delta} \nabla f(x)^T \cdot \delta + \frac{\beta}{2} \cdot \|\delta\|^2$$

$$= \operatorname*{argmin}_{\eta:\, \delta = -\eta \cdot \nabla f(x)} -\eta \cdot \|\nabla f(x)\|^2 + \frac{\beta}{2} \cdot \eta^2 \cdot \|\nabla f(x)\|^2$$

$$= \operatorname*{argmin}_{\delta = -\eta \cdot \nabla f(x)} \eta\left(\frac{\beta}{2} \cdot \eta - 1\right)$$

opt. sol. $\eta = 1/\beta$.

opt $\delta = -\frac{1}{\beta} \nabla f(x)$.

<u>Conclusion:</u> $f(x+\delta) \leq$

$$f(x) - \frac{1}{\beta} \cdot \|\nabla f(x)\|^2 + \frac{\beta}{2} \cdot \frac{1}{\beta^2} \|\nabla f(x)\|^2$$

$$= f(x) - \frac{1}{2\beta} \cdot \|\nabla f(x)\|^2.$$

If $\nabla f(y) \neq 0 \Rightarrow$ make progress!

Will be able to bound #it. T to get a "good" solution.