Smoothness

TODO

Defn: A differentiable function $f : \mathbb{R}^d \to \mathbb{R}$ is β-smooth if its gradient is β-Lipschitz.

$$\|\nabla f(v) - \nabla f(w)\| \leq \beta \|v - w\| \quad \forall v, w$$

Note smoothness implies

$$f(v) \leq \underbrace{f(w) + \langle \nabla f(w), v - w \rangle}_{A_w(v)} + \frac{\beta}{2} \|v - w\|^2$$

convexity:

$$A_w(v) \leq f(v)$$

Both $\Rightarrow$

$$A_w(v) \leq f(v) \leq A_w(v) + \frac{\beta}{2} \|v - w\|^2$$

means: upper & lower bounds on affine approximation

TODO

Note

$$f(v) \le A_w(v) + \frac{\beta}{2}\|v-w\|^2$$

set $v = w - \frac{1}{\beta}\nabla f(w)$

$$f(v) \le f(w) + \left\langle \nabla f(w), \frac{1}{\beta}\nabla f(w) \right\rangle - \frac{\beta}{2}\frac{1}{\beta^2}\|\nabla f\|^2$$

$$\Rightarrow \frac{1}{2\beta}\|\nabla f(w)\|^2 \le f(w) - f(v)$$

if, in addition $f(v) \ge 0 \quad \forall v$, then

$$\|\nabla f(w)\|^2 \le 2\beta f(w) \quad \forall w$$

"self bounded" function

Sept 20

Boyd 2.5 Separating & Supporting Hyperplanes

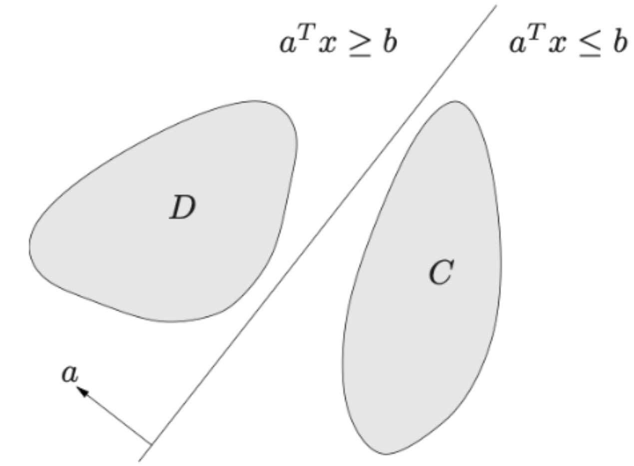


Figure 2.19 The hyperplane $\{x \mid a^T x = b\}$ separates the disjoint convex sets C and D. The affine function $a^T x - b$ is nonpositive on C and nonnegative on D.

Theorem: C, D nonempty disjoint convex sets

there exists $A(x) = ax - b$ s.t.

$$\begin{cases} A(x) \leq 0 & \forall x \in C \\ A(x) \geq 0 & \forall x \in D \end{cases}$$

$H = \{A(x) = 0\}$ "separating hyperplane"

Defn: $\mathrm{dist}(C,D) = \inf\{\|u - v\|_2 \mid u \in C, v \in D\}$

$\|u\|_2 = (u_1^2 + \dots + u_n^2)^{1/2}$

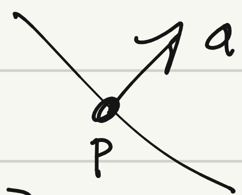
(*) Suppose $\mathrm{dist}(C,D) > 0$ & $\exists\, c \in C, d \in D$ s.t.

$\|c - d\|_2 = \mathrm{dist}(C,D)$

construction:

$A(x) = ax - b.$

Recall formula.



$A(x) = a(x-p)$

point slope formula for a plane

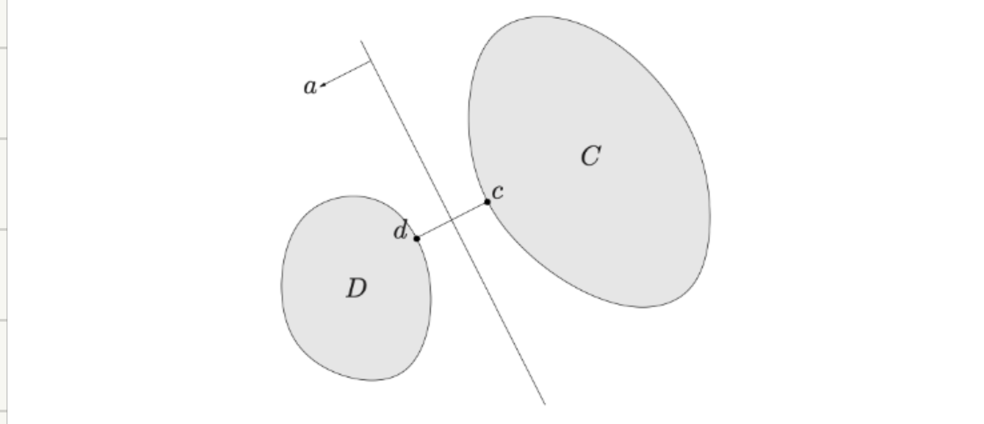


Figure 2.20 Construction of a separating hyperplane between two convex sets. The points $c \in C$ and $d \in D$ are the pair of points in the two sets that are closest to each other. The separating hyperplane is orthogonal to, and bisects, the line segment between c and d.

In this example

$p = \frac{d+c}{2} \quad a = d-c \quad \Rightarrow f(x) = (d-c)^T\left(x - \frac{d+c}{2}\right)$

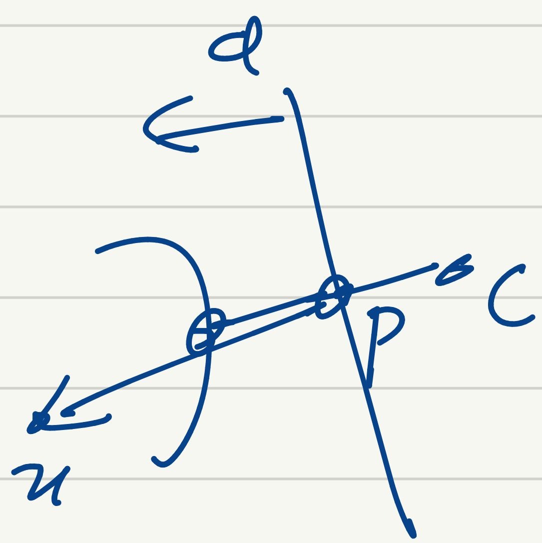
rewrite $f(x) = (d-c)\left(x - d + \frac{d-c}{2}\right)$

$= (d-c)(x-d) + \frac{1}{2}\|d-c\|^2$

& Note $f(d) = 0 + \frac{1}{2}\|d-c\|^2$

$f(c) = -\frac{1}{2}\|d-c\|^2$

$f(u) \geq 0$ since a & $u-p$ point in same direction.



proof of thm: in Boyd notes & In Class

Convex functions

Log-Sum-exp.

$LSE(x) = f(x) = \log(e^{x_1} + \dots + e^{x_n})$

approximates $\max(x_1, \dots, x_n) =: x_m(x)$

$\max(x_1, \dots, x_n) \le f(x) \le \max(x_1, \dots, x_n) + \log n$

Note $f(x,t) = \log(e^{tx_1} + \dots + e^{tx_n})/t \to \max$ as $t \to \infty$

$\max(x_i) = \log(\exp(x_{max})) \le LSE(x)$
since log is increasing & $e^y \ge 0$.

check $LSE(x) \le \log(e^{x_m} \cdot n) = x_m + \log n$
since $\sum e^{x_i} \le \sum e^{x_m}$ //

f Convex !!!

Jensen's Ineq:

① $f\left(\frac{x+y}{2}\right) \leq \frac{f(x)+f(y)}{2}$ when f convex

More general: θ weight vector: $\theta_1, \ldots, \theta_n \geq 0$
$\sum \theta_i = 1$

$$\theta \cdot x = \sum_{i=1}^{n} \theta_i x_i.$$

$f(\theta \cdot x) \leq$ " $\theta \cdot f(X)$ " where $f(X) := (f(x_1), \ldots f(x_n))$

② i.e. $f\left(\sum \theta_i x_i\right) \leq \sum \theta_i f(x_i)$

③ $f(\mathbb{E}x) \leq \mathbb{E}f(x)$

where $\mathbb{E}[x] = \int x \, p(x) dx$ for $\left[\begin{array}{l} \int p(x) dx = 1 \\ p(x) \geq 0 \end{array}\right]$ prob density

$\mathbb{E}[f(x)] = \int f(x) p(x) dx$