Let us study the dual, slaters condition, strong duality, KKT conditions & application of all this through the example of LASSO

$$\min_{x\in\mathbb{R}^n} \|Ax-b\|_2^2 + \lambda\|x\|_1$$

where λ is fixed (not a Lagrange variable)

Recap

For Lasso, it can be shown that for every θ there exists a $\lambda \geq 0$ s.t following two problems are equivalent:

$$\min_x \|Ax-b\|_2^2 + \lambda\|x\|_1$$

① ... say soln is x^* & $\|x^*\|_1 = \beta$

$$\min_x \|Ax-b\|_2^2 \quad \text{s.t } \|x\|_1 \leq \theta$$

② ... say solution is $\hat{x}$

Solution to ② with $\theta = \beta = x^*$ is also x^*!

Solution to ① with λ as soln to $A^T(b-Ax) = \lambda g_{\hat{x}}$ is also $\hat{x}$!

$g_{\hat{x}} \in \partial\|\hat{x}\|_1$

What about dual of Lasso?

Dual of $\min_{x \in \mathbb{R}^n} \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_1$ is a constant!

Redefine primal as:

$$\min_{\substack{x \in \mathbb{R}^n \\ z \in \mathbb{R}^m}} \frac{1}{2}\|y - z\|_2^2 + \lambda\|x\|_1$$

$$\text{s.t. } z = Ax$$

$$L^*(u) = \min_{\substack{x \in \mathbb{R}^n \\ z \in \mathbb{R}^m}} \frac{1}{2}\|y - z\|^2 + \lambda\|x\|_1 + u^T(z - Ax)$$

$$= \frac{1}{2}\|y\|_2^2 - \frac{1}{2}\|y - u\|_2^2 - I\left(\left\|\frac{A^T u}{\lambda}\right\|_\infty \le 1\right)$$

$\therefore$ Lasso dual problem is

$$\max_{u \in \mathbb{R}^m} \frac{1}{2}\left(\|b\|^2 - \|b-u\|^2\right)$$

$$\underline{\underline{s.t}} \quad \|A^T u\|_\infty \leq \lambda$$

$$\underline{\underline{ie}} \quad \max_{u \in \mathbb{R}^m} \frac{1}{2}\left(\|u-b\|_2^2\right)$$

$$s.t \quad \|A^T u\|_\infty \leq \lambda$$

<u>Recall:</u> Dual norm for p is q s.t

$$\frac{1}{p} + \frac{1}{q} = 1$$

Here $p = 1$ & $q = \infty$

$\therefore$ KKT conditions are necessary & sufficient
(since Slater's condition satisfied)

① Primal constraints: $\hat{z} = A\hat{x}$

② Dual constraints: NONE

③ Complementary slackness: NONE

④ Subdiff wrt primal variables $= 0$:

$$\partial_z L(\hat{x}, \hat{z}, \hat{u}) = \hat{z} - b - \hat{u} = 0$$

$$\partial_x L(\hat{x}, \hat{z}, \hat{u}) = \lambda \begin{bmatrix} \operatorname{sgn}(\hat{x}_i) \text{ if } \hat{x}_i \neq 0 \\ \theta_i \in [-1, 1] \text{ if } \hat{x}_i = 0 \end{bmatrix} + A^T \hat{u} = 0$$

Q: How to use the dual formulation & KKT?

Ans: We know by Slater's condition & KKT that if $\hat{x}, \hat{u}, \hat{z}$ satisfy KKT then

$$\frac{1}{2}\|A\hat{x}-b\|^2 + \lambda\|\hat{x}\|_1 = \min_{\substack{x\in\mathbb{R}^n \\ z\in\mathbb{R}^m}} \frac{1}{2}\|z-b\|_2^2 + \lambda\|x\|_1 \quad \text{s.t. } Ax = z$$

$$= \max_{u\in\mathbb{R}^m} \frac{1}{2}\|u-b\|_2^2 \quad \text{s.t. } \|A^T u\|_\infty \le \lambda$$

$$= \frac{1}{2}\|\hat{u}-b\|_2^2$$

Iterative algo for primal

Claim: If $\bar{a}_j = 2\sum_{i=1}^{n} [A_{ij}]^2$ & $\bar{b}_j = 2\sum_{i=1}^{n} A_{ij}(y_i - x_j^{T} A_j)$

Then:

$$x_j^* = \begin{cases} (\bar{b}_j + \lambda)/\bar{a}_j & \text{if } \bar{b}_j < -\lambda \\ 0 & \text{if } \bar{b}_j \in [-\lambda, \lambda] \\ (\bar{b}_j - \lambda)/\bar{a}_j & \text{if } \bar{b}_j > \lambda \end{cases}$$

So to satisfy this, Lasso iterates on x^K as follows: $x^{(0)} \to \bar{b}^{(0)} \to x^{(1)}$ .... until convergence ... We can understand through following simplification where $A = I$

Iterative algo for primal ...

Stopping criterion:

$x^{(k)} \to$ solve for $u^{(k)}$ & $z^{(k)}$ using KKT

$\downarrow$

Check gap $\left\{ \frac{1}{2}\|Ax^{(k)} - b\|_2^2 + \lambda \|x^{(k)}\|_1 - \frac{1}{2}\|u^{(k)} - b\|_2^2 \right\} \le \epsilon$

$$\min_{x \in D} f(x)$$

s.t $g_i(x) \le 0 \quad i=1 \ldots m$

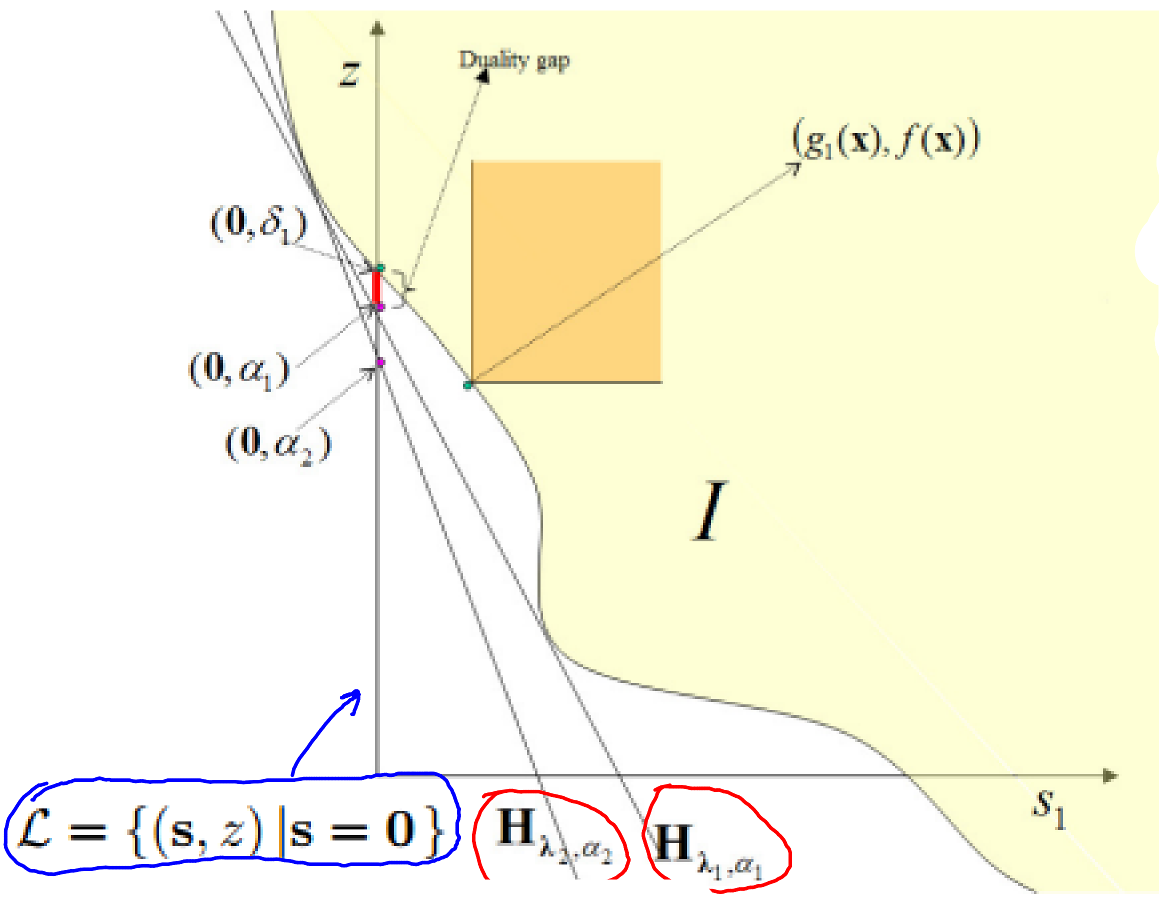
:tr

(The general dual problem & its geometric interpretation)

pg 292, sec 4.4.3 of http://www.cse.iitb.ac.in/~cs709/notes/BasicsOfConvexOptimization.pdf

Consider the set:

$$\mathcal{I} = \{(\mathbf{s}, z) \mid \mathbf{s} \in \Re^m,\ z \in \Re,\ \exists \mathbf{x} \in \mathcal{D} \text{ with } g_i(\mathbf{x}) \le s_i\ \forall 1 \le i \le m,\ f(\mathbf{x}) \le z\}$$



Kind of join between epigraphs of f & g_i's

$$H_{\lambda,\alpha} = \{(\mathbf{s}, z) \mid \lambda^T.\mathbf{s} + z = \alpha\}$$

$$\min_{x \in \mathcal{D}} f(x)$$

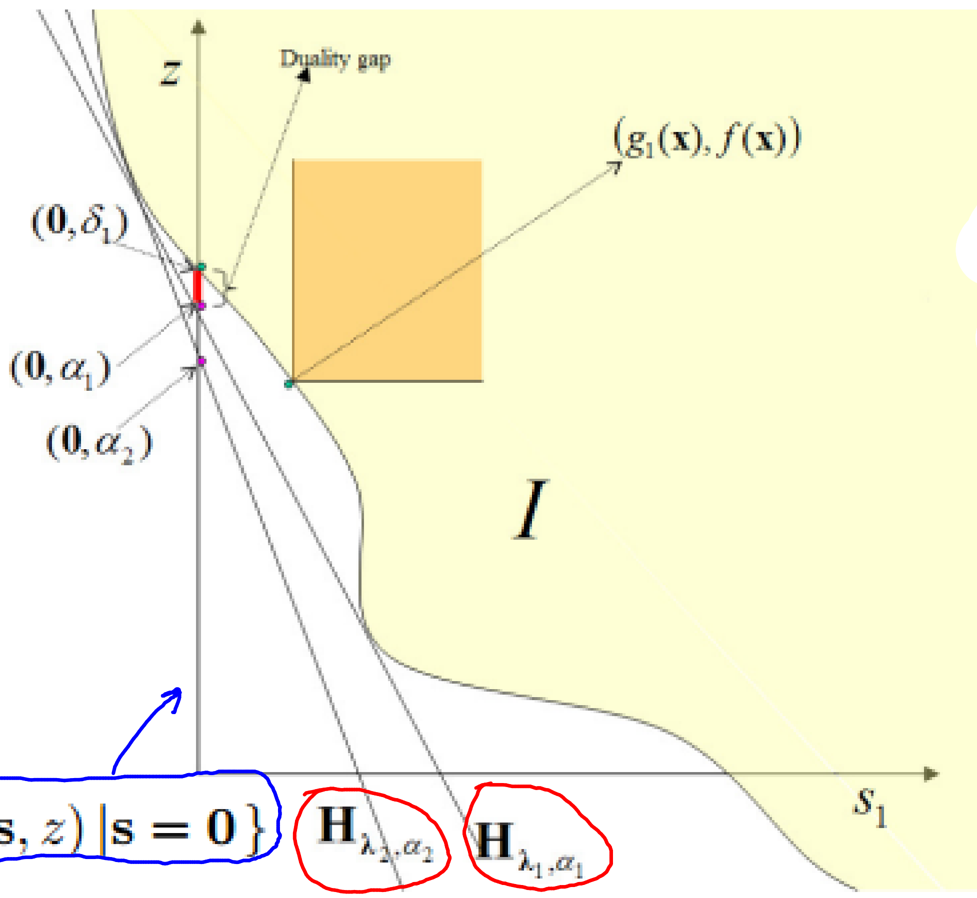
$$\text{s.t. } g_i(x) \le 0 \quad i = 1 \ldots m$$

(The general dual problem & its geometric interpretation)

pg 292, sec 4.4.3 of http://www.cse.iitb.ac.in/~cs709/notes/BasicsOfConvexOptimization.pdf

Consider the set:

$$\mathcal{I} = \{(\mathbf{s}, z) \mid \mathbf{s} \in \Re^m,\ z \in \Re,\ \exists \mathbf{x} \in \mathcal{D} \textit{ with } g_i(\mathbf{x}) \le s_i\ \forall 1 \le i \le m,\ f(\mathbf{x}) \le z\}$$



$$\mathcal{L} = \{(\mathbf{s}, z) \mid \mathbf{s} = \mathbf{0}\} \qquad \mathbf{H}_{\lambda_2, \alpha_2} \qquad \mathbf{H}_{\lambda_1, \alpha_1}$$

$$\mathcal{H}_{\lambda,\alpha} = \{(\mathbf{s}, z) \mid \lambda^T.\mathbf{s} + z = \alpha\}$$

smallest z value in $\mathcal{I}$ for $s_1 \le 0$ will be for $s_1 = 0$

since $(s_1, z_1) \in \mathcal{I}$

$\Rightarrow (s_2, z_2) \in \mathcal{I}$

$\forall s_2 \ge s_1$ & $z_2 \ge z_1$

Thus $\Rightarrow$

is not possible!

$$
\begin{array}{ll}
\text{max} & \alpha \\
\text{subject to} & \mathcal{H}^+_{\lambda,\alpha} \supseteq \mathcal{I}
\end{array}
$$

$$
\begin{array}{ll}
\text{max} & \alpha \\
\text{subject to} & \lambda^T.\mathbf{s} + z \geq \alpha \ \forall (\mathbf{s}, z) \in \mathcal{I}
\end{array}
$$

$$
\begin{array}{ll}
\text{max} & \alpha \\
\text{subject to} & \lambda^T.\mathbf{s} + z \geq \alpha \ \forall (\mathbf{s}, z) \in \mathcal{I} \\
 & \lambda \geq \mathbf{0}
\end{array}
$$

a

If $\exists$ α_b soln to b,
s.t. $\lambda^T g(x_b) + f(x_b) = \alpha_b$
then $s = g(x_b)$ $z = f(x_b)$
is soln to a
complete proof of equivalence

$$
\begin{array}{ll}
\text{max} & \alpha \\
\text{subject to} & \lambda^T.\mathbf{g}(\mathbf{x}) + f(\mathbf{x}) \geq \alpha \ \forall \mathbf{x} \in \mathcal{D} \\
 & \lambda \geq \mathbf{0}
\end{array}
$$

b

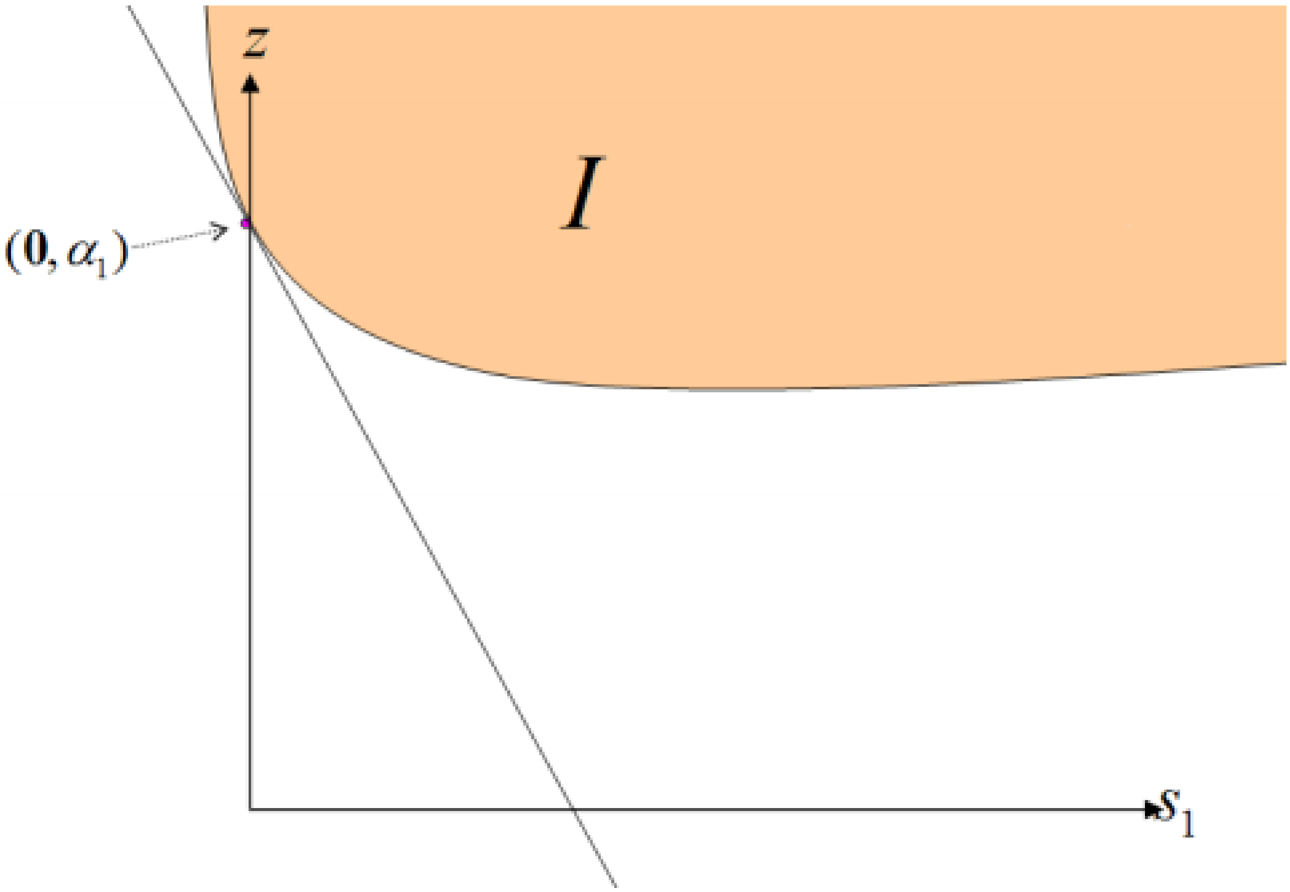
$$
\begin{array}{ll}
\max & \alpha \\
\text{subject to} & L(\mathbf{x}, \lambda) \geq \alpha \ \forall \mathbf{x} \in \mathcal{D} \\
 & \lambda \geq \mathbf{0}
\end{array}
$$

Since, $L^*(\lambda) = \min_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \lambda)$, we can deal with the equivalent

$$
\begin{array}{ll}
\max & \alpha \\
\text{subject to} & L^*(\lambda) \geq \alpha \\
 & \lambda \geq \mathbf{0}
\end{array}
$$

This problem can be restated as

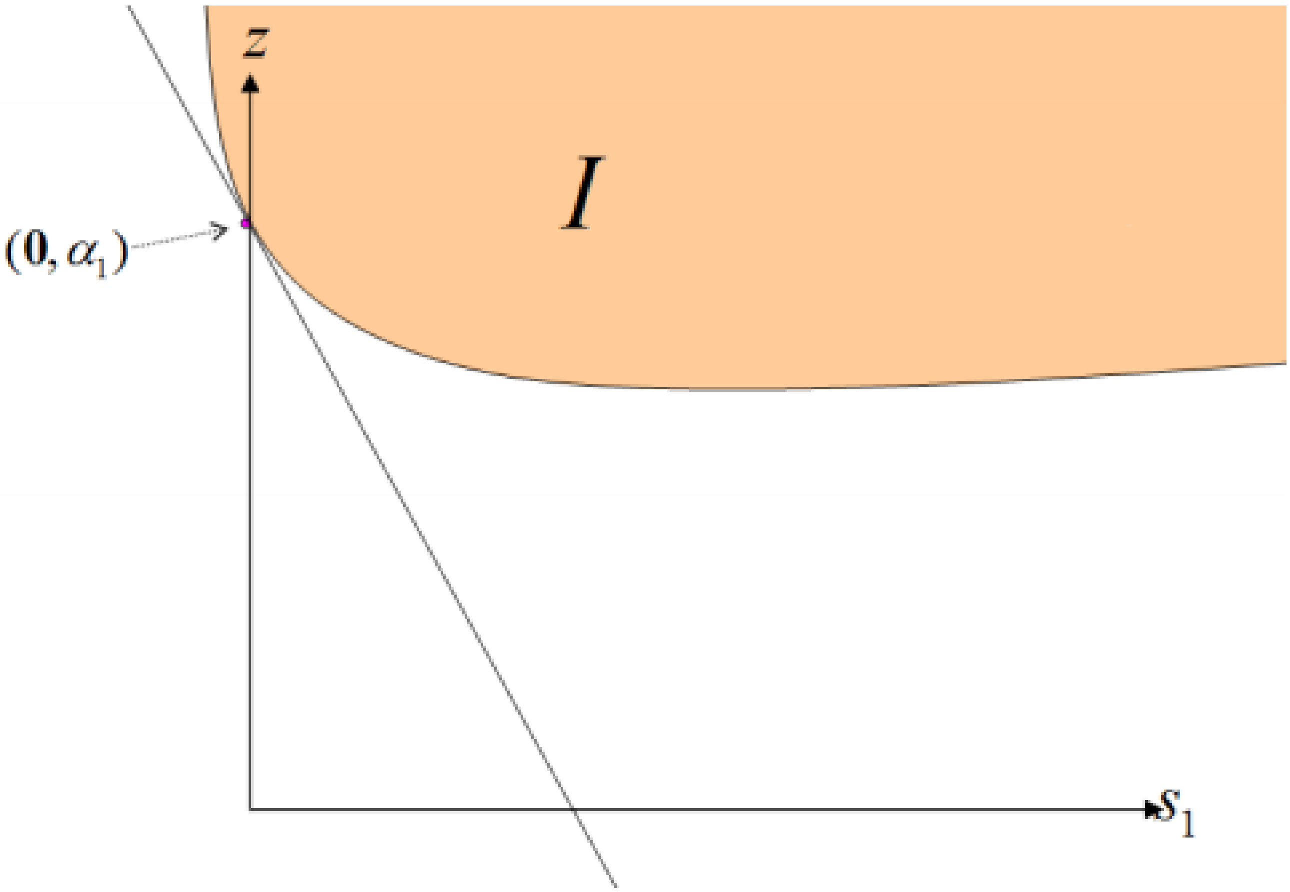
$$
\begin{array}{ll}
\max & L^*(\lambda) \\
\text{subject to} & \lambda \geq \mathbf{0}
\end{array}
$$



Q: What is desirable of the set I for zero duality gap?

① I should be convex ie f & g_i's are convex

② $(0, \delta_1)$ should exist ie I's intersection with $s=0$ should be closed



Q: What is desirable of the set I for zero duality gap?

Ans: $\exists\ (0,\alpha) \in I$ and λ s.t $\lambda^T s + z \geq \alpha\ \forall\ (s,z) \in I$
& Intersection of I with z axis is closed below

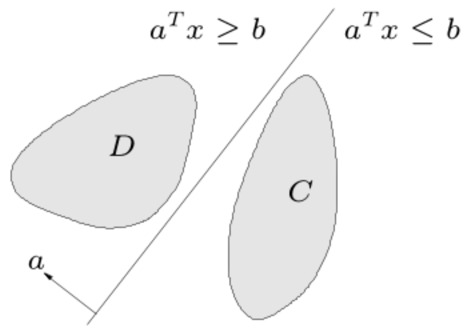
$\Leftrightarrow$ $\exists$ a supporting hyperplane to I at $(0,\alpha)$
& Intersection of I with z axis is closed below (with $(0,\alpha)$ being boundary pt)

$\Leftarrow$ I is closed & $\exists$ a supporting hyperplane to I at every boundary point

Separating hyperplane theorem

if C and D are disjoint convex sets, then there exists $a \neq 0$, b such that

$$a^T x \leq b \text{ for } x \in C, \qquad a^T x \geq b \text{ for } x \in D$$



the hyperplane $\{x \mid a^T x = b\}$ separates C and D

strict separation requires additional assumptions ($e.g.$, C is closed, D is a singleton)

Consequence

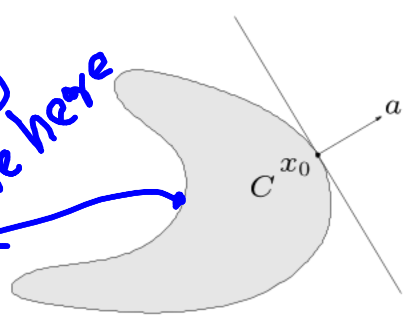
Supporting hyperplane theorem

supporting hyperplane to set C at boundary point x_0:

$$\{x \mid a^T x = a^T x_0\}$$

where $a \neq 0$ and $a^T x \leq a^T x_0$ for all $x \in C$

no supporting hyperplane here



supporting hyperplane theorem: if C is convex, then there exists a supporting hyperplane at every boundary point of C

Proof (from separating hyperplane theorem):

(a) interior$(C) \neq \phi$ (so that $\text{int}(C) \cap \{x_0\} = \phi$)

Apply separating hyperplane theorem to sets $C' = \{x_0\}$ and $D' = \text{interior}(C)$

Strict separation not applicable since int(C) is open

$\exists\, a, b$ s.t. $a^T x \geq b \;\forall x \in D'$

$a^T x_0 \leq b$

This inequality extends to all boundary (limit pts) leading to $a^T x_0 \geq b$

combined: $a^T x_0 = b$

(b) $\text{int}(C) = \phi$

$\Rightarrow$ Any hyperplane containing that affine set contains C & x_0 (H/w: Prove)

$\Rightarrow$ This hyperplane is a trivial supporting hyperplane

Some topological concepts: [Topological set is set with concept of nbrhood]

① A set U is called an open set if it does not contain any of its boundary pts. If S is a metric space (eg an inner product space) with distance metric $d(x,y)$, then a subset U of S is called open if, given any $x \in U$, $\exists\ \epsilon > 0$ such that given any $y \in S$ with $d(x,y) < \epsilon$, $y \in U$

② A set $V \subseteq S$ is called closed if its complement $S \setminus V$ is an open set

③ $x \in S$ is called an interior point of S if there exists a neighborhood of x contained in S. If S is a metric space, then $x \in S$ is an interior pt if $\exists \epsilon > 0$ s.t $\forall y$ s.t $d(x,y) < \epsilon$, $y \in S$

The set of all interior pts of S form the interior of S. Thus, if S is a metric space:

$$\underbrace{\text{int}(S)}_{\text{interior of } S} = \{x \mid \exists \epsilon > 0 \text{ s.t } \forall y \text{ s.t } d(x,y) < \epsilon, y \in S\}$$

What can I say if interior$(C) = \emptyset$

eg: sufficient conditions

① $C \subseteq$ hyperplane. In particular, if S is affine this is necessary & sufficient

② eg: a shell

③ eg: $C = \partial K$ in topological space S (see next page for defn of boundary of set K denoted by ∂K)

④ The set of pts of a set S s.t every neighborhood of a point from the set consists of atleast one point in S and one point not in S is called the boundary ∂S of S. If S is a metric space

$$\partial S = \{ x \in S \mid \forall \epsilon > 0, \exists y \text{ s.t } d(x,y) < \epsilon \text{ \& } y \in S \text{ and } \exists y' \text{ s.t } d(x,y') < \epsilon \text{ \& } y' \notin S \}$$

⑤ Let S be a subset of a topological space X. A point $x \in X$ is a limit point of S if every neighborhood of x contains at least one point of S different from x itself.

→ can be relaxed to open neighborhoods without loss

If S happens to have an associated metric d, and $A \subseteq S$, then $\alpha \in S$ is a limit point of A iff:

$$\forall \epsilon > 0 : \{ x \in A \text{ s.t } 0 < d(x, \alpha) < \epsilon \} \neq \phi$$

[Note: α need not belong to A]

Informally speaking, is a limit point of A if there are points in A that are different from but arbitrarily close to it

⑥ Closure of S [cl(S)] = S ∪ {limit points of S}

Some standard results that we will regularly invoke for topological spaces

① Intersection of (even uncountable) closed sets is closed

② Union of (even uncountable) open sets is open

③ Intersection of finite number of open sets is open

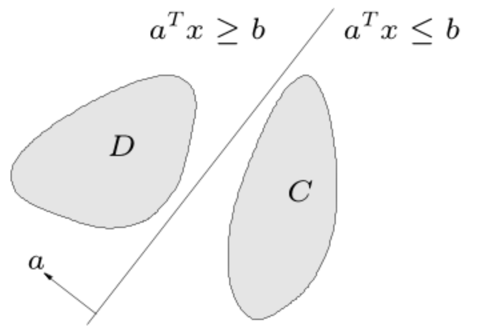
④ Union of finite number of closed sets is closed

⑤ S is closed iff S^c is open

Separating hyperplane theorem

if C and D are disjoint convex sets, then there exists $a \neq 0$, b such that

$$a^T x \leq b \text{ for } x \in C, \qquad a^T x \geq b \text{ for } x \in D$$



the hyperplane $\{x \mid a^T x = b\}$ separates C and D

strict separation requires additional assumptions ($e.g.$, C is closed, D is a singleton)

$\boxed{\text{Proof:}}$ Let $S = \{x - y \mid x \in C, y \in D\}$.

Now we can prove (see http://www.cse.iitb.ac.in/~cs709/notes/eNotes/ExtraProblems-1.pdf, Q1) that S, being a sum of two convex sets, is convex.

Since $C \cap D = \phi$, $0 \notin S$.

(a) Suppose $0 \notin cl(S)$: Consider the sets $\{0\}$ and $cl(S)$. We will prove that $\exists\, a \neq 0$ s.t

$$\underbrace{a^T z > 0 \;\; \forall\, z \in cl(S)}_{\text{Q: How to choose } a\text{?}} \;\; \& \;\; \underbrace{a^T w = 0 \;\; \text{for } w \in \{0\}}_{\text{obvious}}$$

complete proof given in class: H/W

i.e $\exists a$ s.t $a^T(x-y) > 0 \ \forall \ x-y \in S$

i.e $a^T x > a^T y \ \forall \ x \in C \ \& \ y \in D$

Let $b = \inf_{x \in C} a^T x$. Then we proved existence of a & b s.t

$$a^T x \geq b \ \forall \ x \in C \ \& \ a^T y \leq b \ \forall \ y \in D$$

(b) suppose $0 \in cl(S)$. Since $0 \notin S$, $0 \in bndry(S)$

If interior $(S) = \phi$ (empty), S must be $\subseteq \{z \mid a^T z = b\}$ A hyperplane

& the hyperplane must includ 0 on bndry (S) &

$\Rightarrow b = 0$. i.e $a^T x = a^T y \ \forall x \in C$ & $y \in D$

$\Rightarrow$ we have a trivial separating hyperplane limit by $\bar{a}$, we have

$$\boxed{a(\epsilon_k)^T z > 0} \quad \forall\, z \in S_{-\epsilon_k}$$

for all k & therefore

$$\boxed{\bar{a}^T z > 0 \quad \forall\, z \in \text{interior}(S)}$$

proof by contradiction

and

$$\bar{a}^T z \geq 0 \quad \forall\, z \in S$$

(use the property that a convex set is connected?)

that is

$$\bar{a}^T x \geq \bar{a}^T y$$

$$\forall\, x \in C \;\&\; y \in D$$

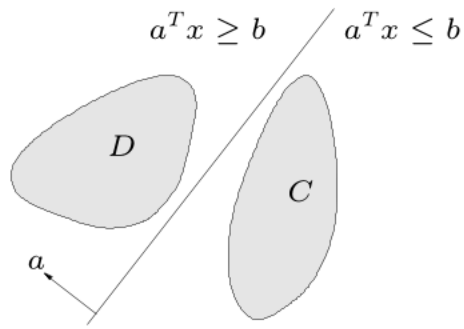
Hence proved!

Thus

Separating hyperplane theorem

if C and D are disjoint convex sets, then there exists $a \neq 0$, b such that

$$a^T x \leq b \text{ for } x \in C, \qquad a^T x \geq b \text{ for } x \in D$$



the hyperplane $\{x \mid a^T x = b\}$ separates C and D

strict separation requires additional assumptions ($e.g.$, C is closed, D is a singleton)

Consequence

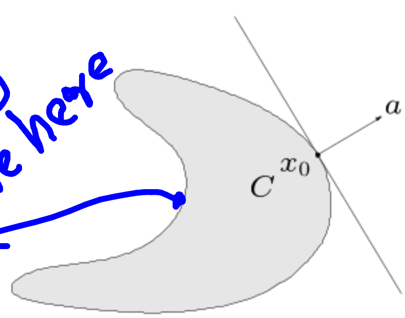
Supporting hyperplane theorem

supporting hyperplane to set C at boundary point x_0:

$$\{x \mid a^T x = a^T x_0\}$$

where $a \neq 0$ and $a^T x \leq a^T x_0$ for all $x \in C$

no supporting hyperplane here



supporting hyperplane theorem: if C is convex, then there exists a supporting hyperplane at every boundary point of C