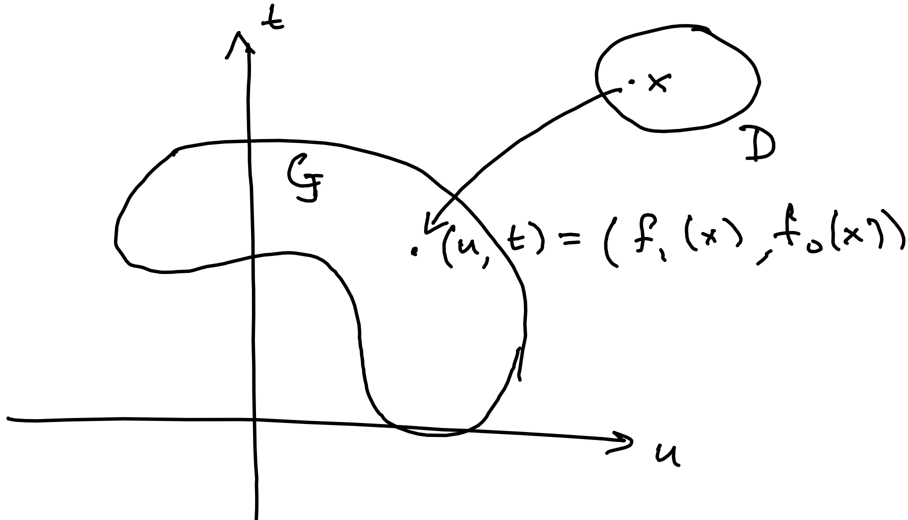
Consider a standard form problem w/ one constraint.

$$\min_{x \in D} \; f_0(x)$$
$$\text{st} \quad f_1(x) \leq 0$$

Define the set $\mathcal{G} = \{(f_1(x), f_0(x)) \mid x \in D\}$.

This can readily be extend to higher dimensions for more constraints.

Suppose we have a highly nonconvex problem, where D is a closed connected set in $\mathbb{R}^2$, yielding

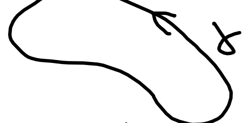


It's hard to write but easy to see that there's some diffeomorphism (f_1, f_2) from D to $\mathcal{G}$.

Side Note: Analytically parametrizing the "boomerang"

It's interesting to consider how to construct such a $\mathcal{G}$.

Given a closed ccw-oriented curve γ

e.g.  γ, we can define the interior with the winding number by Residue

$$w(x) = \frac{1}{2\pi i} \oint_\gamma \frac{dz}{z-x},$$ interpreting $\mathbb{R}^2$ as $\mathbb{C}$.

(this does not scale to larger dimensions, where one would use the ray test). Then $\mathcal{G} = \{x \mid w(x) > 0\}$. But how to get f_0, f_1? With this representation it's clear that fixing the second term is sufficient. E.g. take

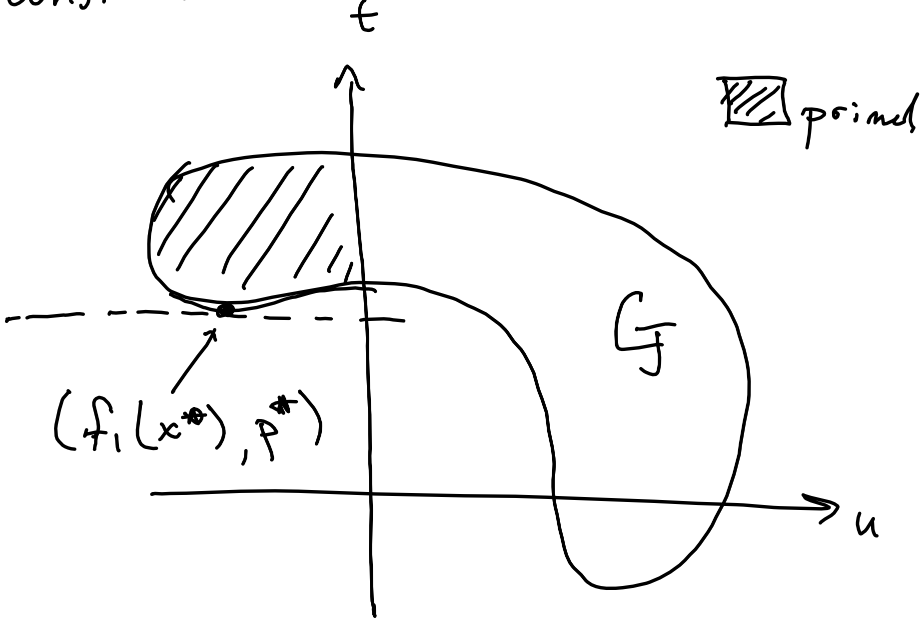
$$D = \mathbb{R}^2 \quad \min_{a,b} \; b \;\text{ st }\; w((a,b)) \geq \varepsilon$$

for any $\varepsilon \in (0,1)$.

Coming back to our picture now. Notice $\mathcal{G}$ encodes more than "vanilla info" of

* Primal feasible set
* Primal objective.

In particular, we see new primal values for infeasible constraint values (which are feasible in perturbed problems, which have relaxed constraints).



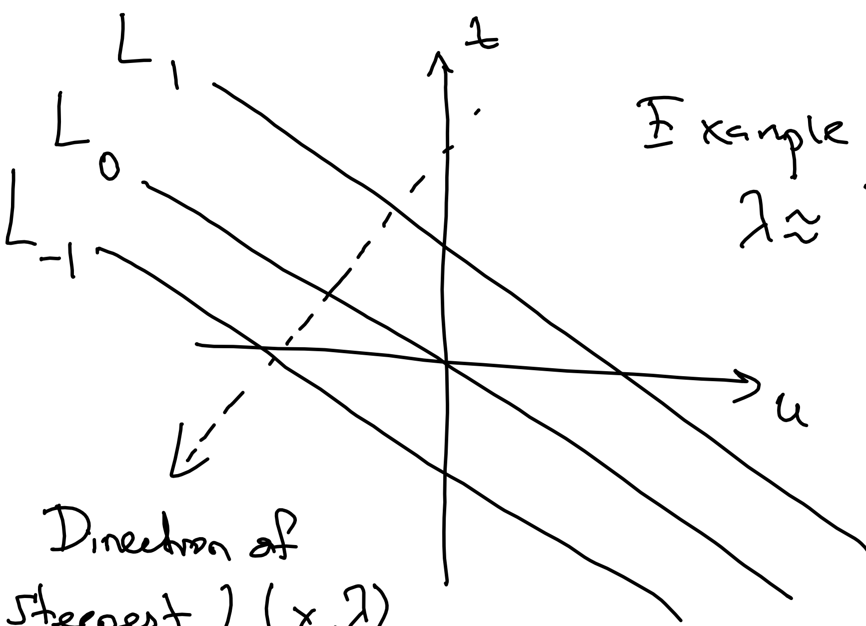
Clearly, the optimal point is the lowest one on the left half space of the plane.

How does the Lagrangian fit into this picture?

Recall the parameterization $t = f_0(x)$, $u = f_1(x)$.

Then
$$L(x,\lambda) = f_0(x) + \lambda f_1(x)$$
$$= t + \lambda u.$$

For fixed values $V = t + \lambda u$, we can define level sets $L_V = \{(u,t) \mid V = t + \lambda u\}$.



Example w/ $\lambda \approx 1/2$

Direction of steepest $L(x,\lambda)$ descent in (u,t) space.

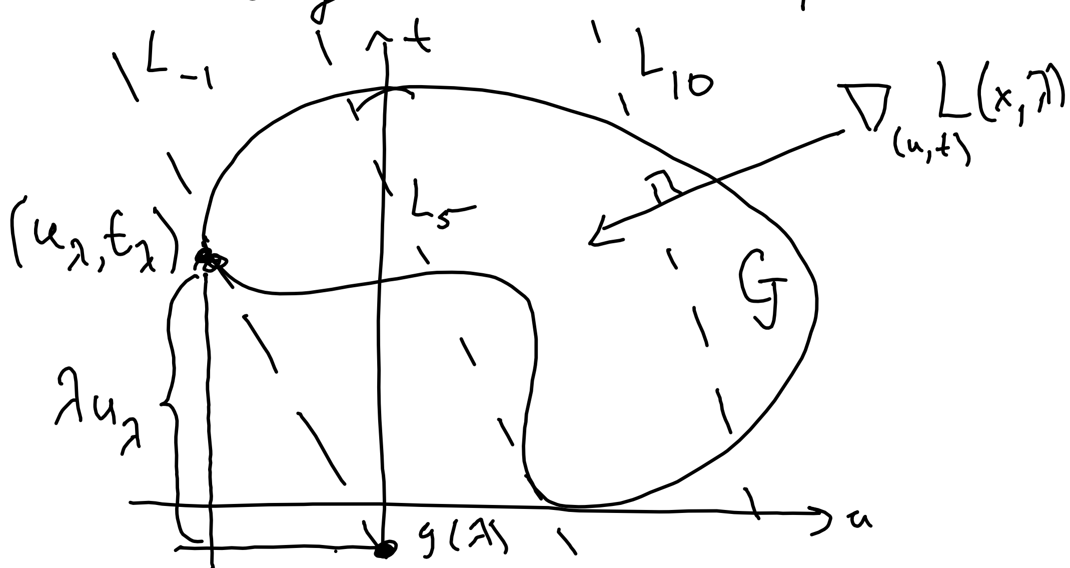
Recall the dual function over $\lambda \geq 0$

$$g(\lambda) = \inf_{x \in D} L(x, \lambda)$$

$$= \inf_{x \in D} f_0(x) + \lambda f_1(x)$$

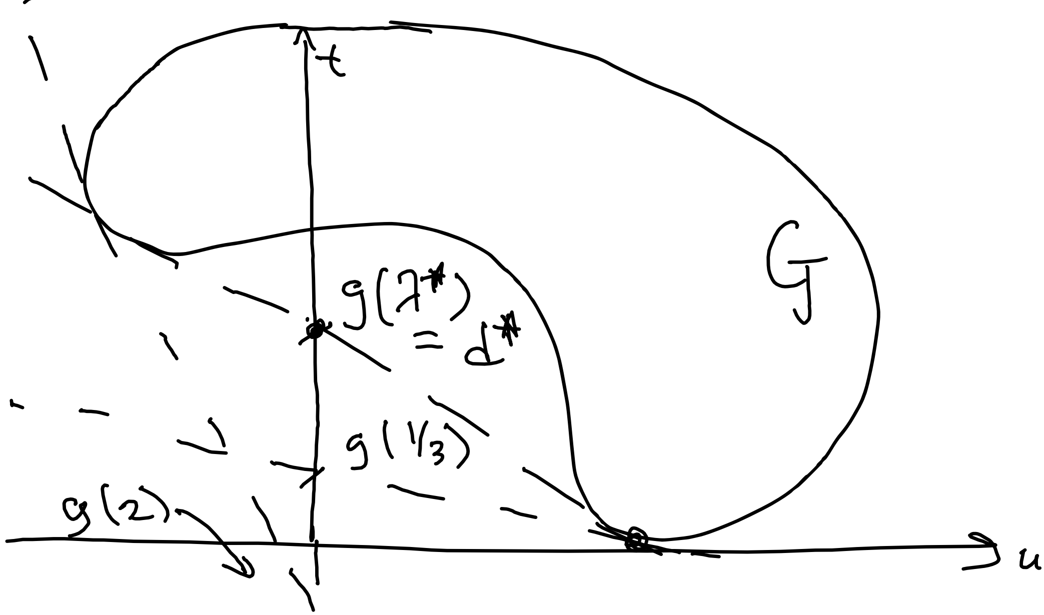
$$= \inf_{(u,t) \in \mathcal{G}} t + \lambda u$$

Thus, for each λ, if $g(\lambda) > -\infty$ then $\exists\, (u_\lambda, t_\lambda) \in \overline{\mathcal{G}}$, assuming continuity, which b/c of $\lambda \geq 0$ are the "lowest tangents of $\mathcal{G}$ for slope $-\lambda$"

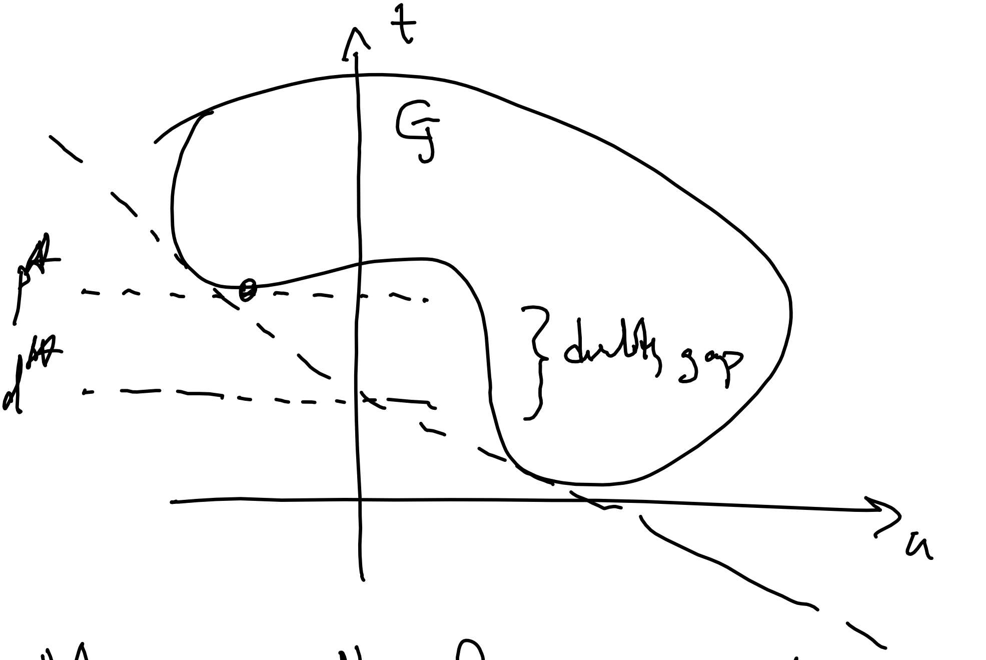


Per the previous illustration:
the dual function is the shift/intercept necessary to make the slope-$(-\lambda)$ hyperplane support $\mathcal{G}$.

Thus, rotating majorized supporting hyperplanes around $\mathcal{G}$ and tracing out the highest intercept we get ~~gets~~ us to



This makes it clear what the <u>duality gap</u> is: the vertical difference b/w the lowest feasible point in $\mathcal{G}$ and the highest feasible (maximized) supporting hyperplane intercept of $\mathcal{G}$.

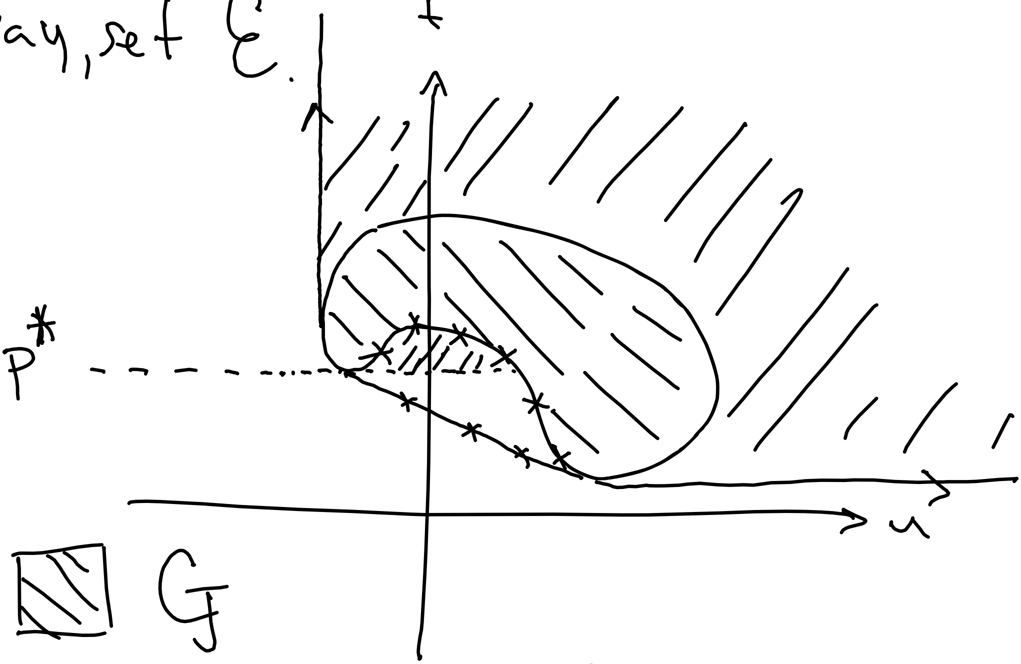


More generally, for m inequalities and p equalities, define

$$\mathcal{A} = \mathcal{G} + (\mathbb{R}^m_+ \times \{0\} \times \mathbb{R}^p)$$

So $\mathcal{A} = \{(u,t) \mid \exists x \in D \;\; f_0(x) \le t, \; f_1(x) \le u\}$

then the duality gap is more visible as the vertical t-axis spread between $\mathcal{A}$ and the intersection of all supporting (feasible, so negative-slope) half planes of $\mathcal{G}$, say, set $\mathcal{E}$.



$\mathcal{G}$

$\mathcal{A} \setminus \mathcal{G}$ (note $\mathcal{G} \subseteq \mathcal{A}$)

$\mathcal{E} \setminus \mathcal{A}$ (note $\mathcal{A} \subseteq \mathcal{E}$)

(CVX). A <u>convex problem</u> has f_i cvx for $i \in [m] \cup \{0\}$ and h_j affine for $j \in [p]$.

(SF). A standard form program is <u>strictly feasible</u> if $\exists\, \tilde{x} \in \text{relint}\, D$ s.t. $f_i(\tilde{x}) < 0$ for all non-affine f_i.

<u>Refined Slater's Condition.</u>

A CVX, SF problem is strongly dual.

Pf. (concept, 1-D)

Define $\mathcal{A}$ as before.

Let $B = \{(0,t) \mid t < p^*\}$.

Notice: $\mathcal{A}$ is cvx b/c f_0, f_1 are.

B is cvx. (SHT)

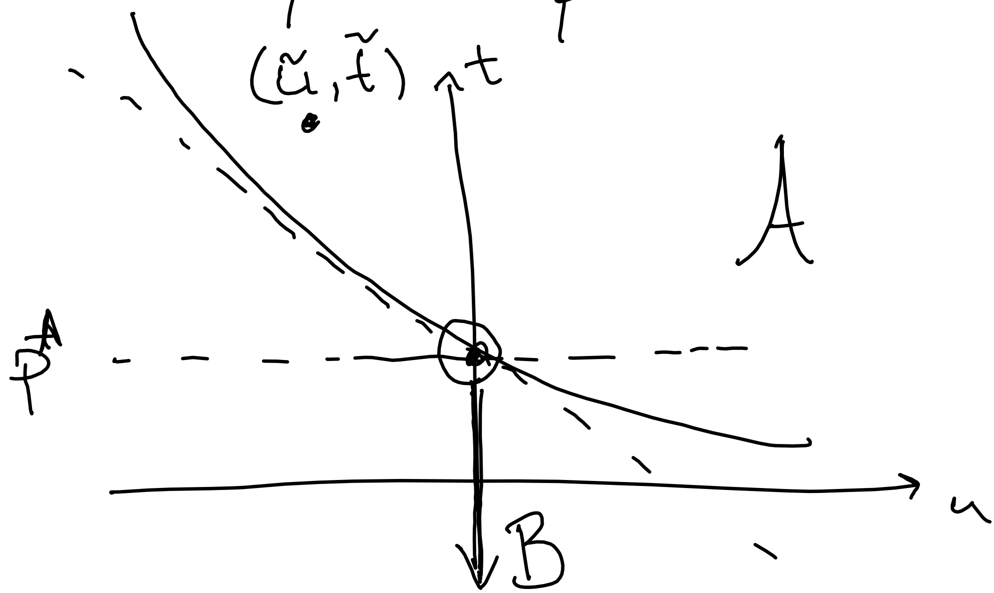
Since $\mathcal{A} \cap B = \emptyset$, by separating hyperplane theorem, there exists a line with slope $-\tilde{\lambda}$ separating $\mathcal{A}$ & B.

By SF, $\exists\ \tilde{x}$ with $(\tilde{u}, \tilde{t}) = (f_1(\tilde{x}), f_0(\tilde{x}))$ in negative half plane $\{(u,t) \mid u<0\}$.

since $\tilde{t} \geq p^* \geq_{\pi, B}$, $\tilde{\lambda} \geq 0$;

$\uparrow$ primal optimality $\qquad$ $\uparrow$ def. B.

i.e. the plane L_{p^*} is nonvertical:



$$L_{p^*} = \{(u,t) \mid p^* = t + \tilde{\lambda} u\}$$

But then consider the dual function $g(\tilde{\lambda})$, which by SHT must be above B and thus $g(\tilde{\lambda}) \geq p^*$. Then with weak duality we have $g(\tilde{\lambda}) = p^*$, ie Strong Duality.

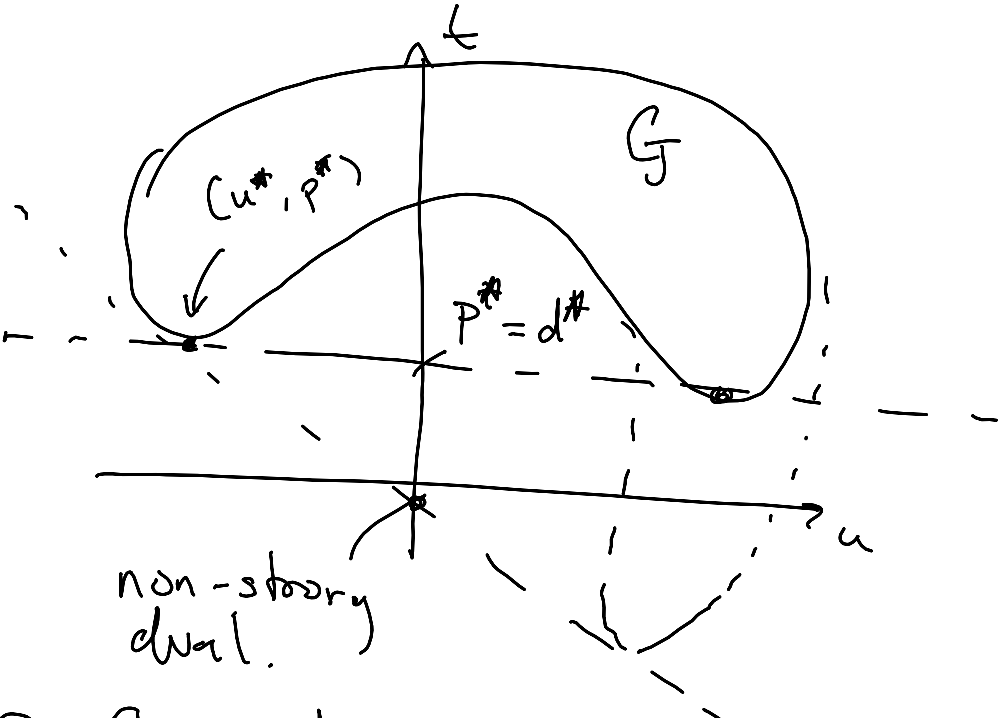
This proof gives us intuition why cvx problems *typically* have SD (strong duality) and vice-versa for non-cvx.

There are counterexamples, see the end for exercises.

When SD holds, the KKT conditions are *necessary* for all optima.

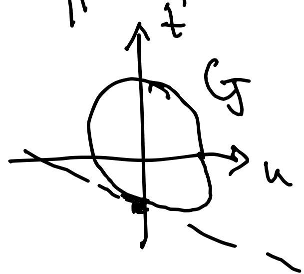
The geometric intuition shows this clearly, for instance:

Complementary Slackness, the condition that $\lambda_i^* f_i(x^*) = 0$ for optimal x^*, λ^* of a standard form problem & λ its inequality dual.



SD: $\mathcal{G}$ cannot dip below p^* plane.

So either: the optimum $(u^*, p^*) = (f_1(x^*), f_0(x^*))$ is, as the lowest point, going to admit $\lambda^* = 0$ a flat supporting hyperplane, or, in the $\lambda^* > 0$ case:



Examples (5.22 & 5.21)

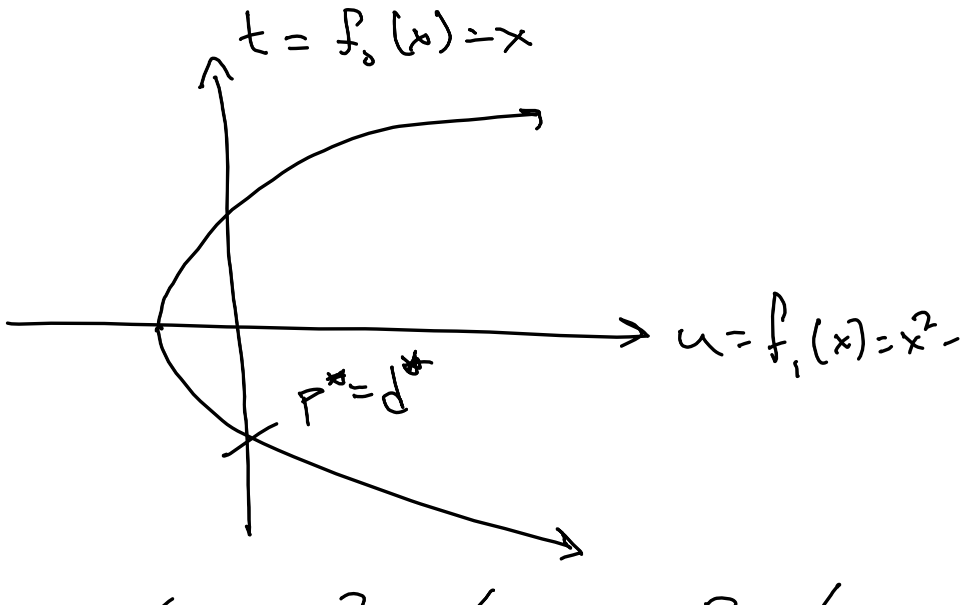
$D = \mathbb{R}$ unless specified o/w

$$P = \begin{array}{l} \min \ x \\ \text{st } x^2 \le 1 \end{array} \quad (p^* = -1 \ @ \ x^* = p^*)$$

$$D = \max_{\lambda \ge 0} \inf_x \ x + \lambda(x^2 - 1)$$

$$= \max_{\lambda \ge 0} -\left(\lambda + \frac{1}{4\lambda}\right)$$

$$\Rightarrow \lambda^* = \frac{1}{2} \qquad d^* = \underline{1}$$

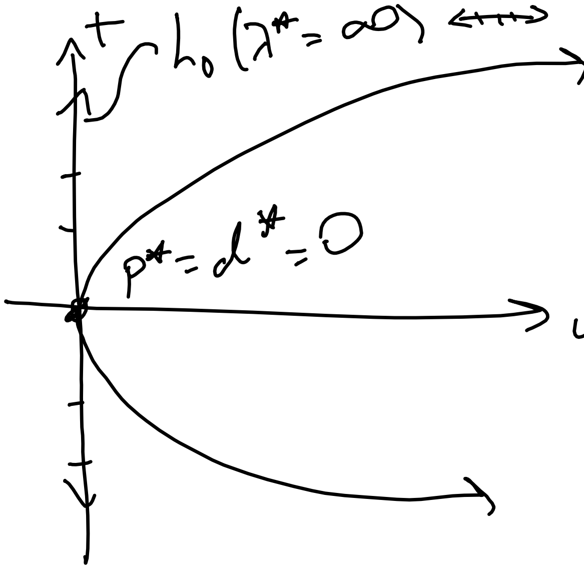


CVX? ✓ SF? ✓ SD? ✓

$$P = \begin{array}{l} \min x \\ \text{s.t. } x^2 \le 0 \end{array} \quad (p^* = x^* = 0)$$

$$D = \max_{\lambda \ge 0} \inf_x \left(x + \lambda x^2\right)$$

$$= \max_{\lambda \ge 0} -1/4\lambda \Rightarrow \left(d^* = 0 \ @ \ \lambda^* \to \infty\right)$$



u

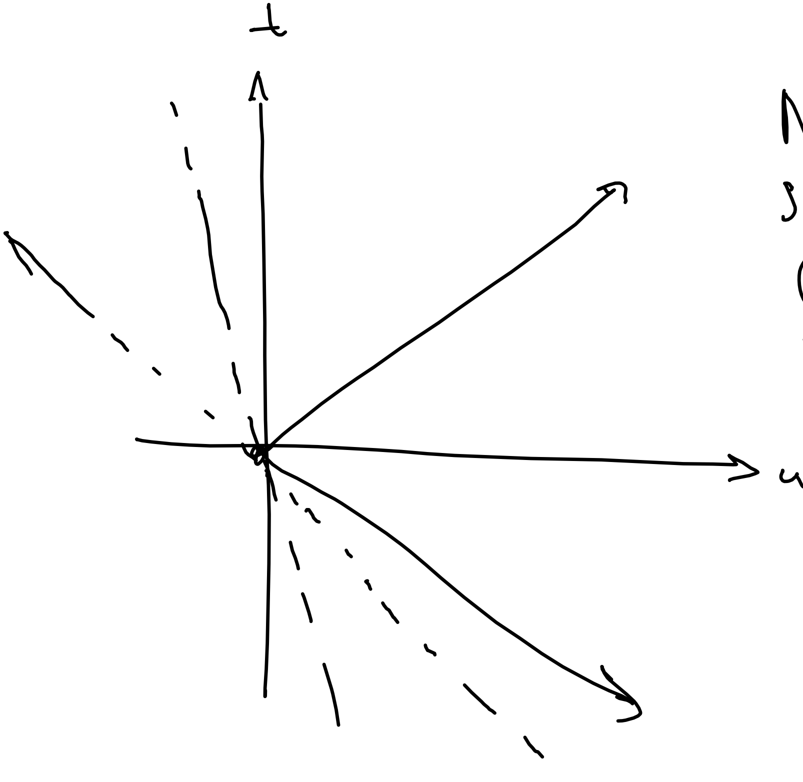
CVX? ✓ SF? ✗ SD? ✓

$$P = \begin{array}{c} \min\ x \\ \text{st } |x| \leq 0 \end{array} \quad (p^* = x^* = 0)$$

$$D = \max_{\lambda \geq 0} \inf_x \ x + \lambda |x|$$

$$= \max_{\lambda \geq 0} \begin{cases} 0 & \lambda \geq 1 \\ -\infty & \lambda < 1 \end{cases}$$

$$\Rightarrow (d^* = 0 \ @ \ \lambda^* \in [1, \infty)$$



Note any subdifferential @ optimum differs to support.

CVX? ✓ SF? ✗ SD? ✓

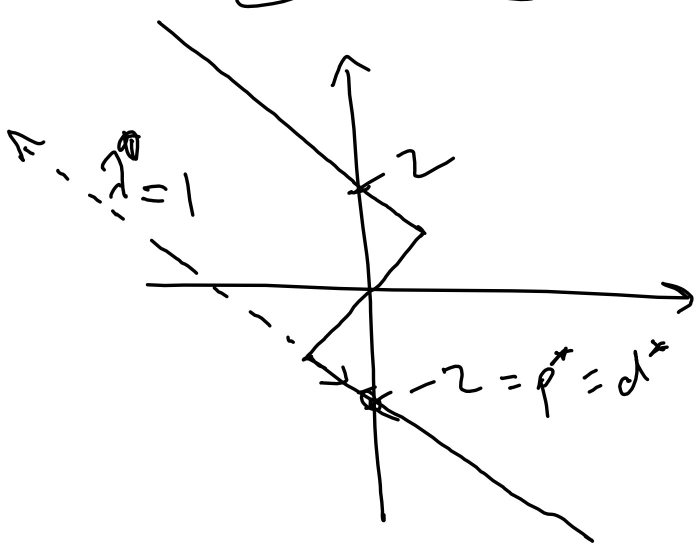
$$P = \begin{array}{l} \min\ x \\ \text{st} \quad \begin{cases} -x+2 & x \geq 1 \\ x & -1 \leq x < 1 \\ -x-2 & x < -1 \end{cases} \leq 0 \end{array}$$

$$p^* = x^* = -2$$

$$D = \max_{\lambda \geq 0} \inf_x x + \lambda \begin{cases} -x+2 & x \geq 1 \\ x & -1 \leq x < 1 \\ -x-2 & x < -1 \end{cases}$$

$$= \max_{\lambda \geq 0} \begin{cases} -\infty & \lambda > 1 \\ -2 & \lambda = 1 \\ -\infty & \lambda < 1 \end{cases}$$

$$\Rightarrow d^* = -2 \ @ \ \lambda^* = 1$$



CVX? X

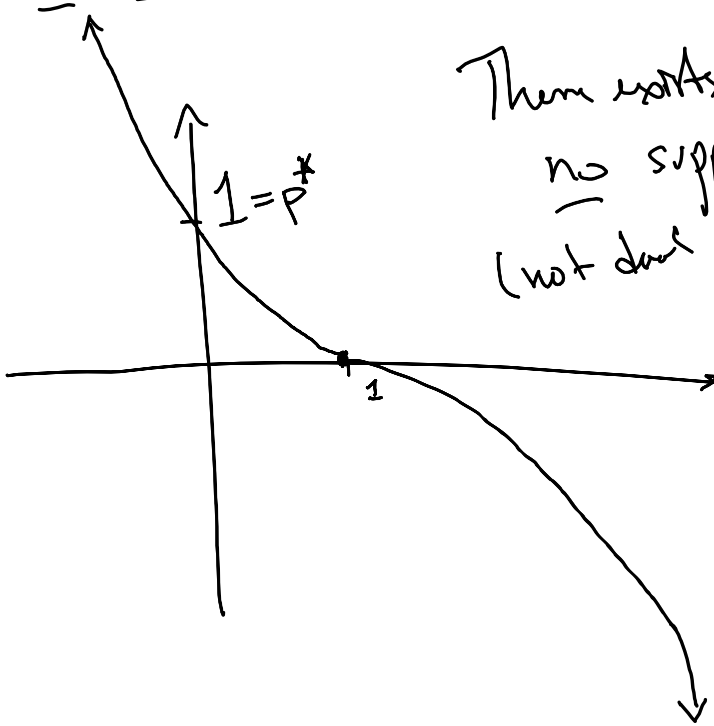
SF? ✓

SD? ✓

$$P = \min_{x \in D} x^3 \quad \text{s.t. } -x+1 \le 0 \qquad p^* = x^* = -1$$

$$D = \max_{\lambda \geq 0} \inf_{x} x^3 + \lambda(1-x) \qquad D = \mathbb{R}$$

$$= -\infty$$



There exists <u>no</u> supporting hyperplane
(not dual feasible).

CVX? ✗ SF? ✓ SD? ✗

$D = \mathbb{R}_+$

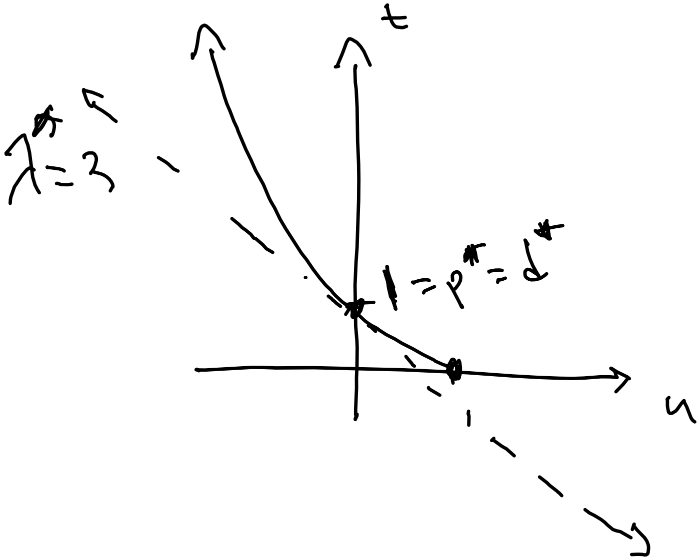
$$P = \min_{x \in D} \; x^3 \qquad p^* = x^* = -1$$

$$\text{st} \quad -x + 1 \leq 0$$

$$D = \max_{\lambda \geq 0} \inf_{x \in D} \; x^3 + \lambda(1-x)$$

$$= \max_{\lambda \geq 0} \; \lambda - \frac{2\lambda^{3/2}}{3\sqrt{3}}$$

$$\Rightarrow d^* = 1 \;\; @ \;\; \lambda^* = 3$$

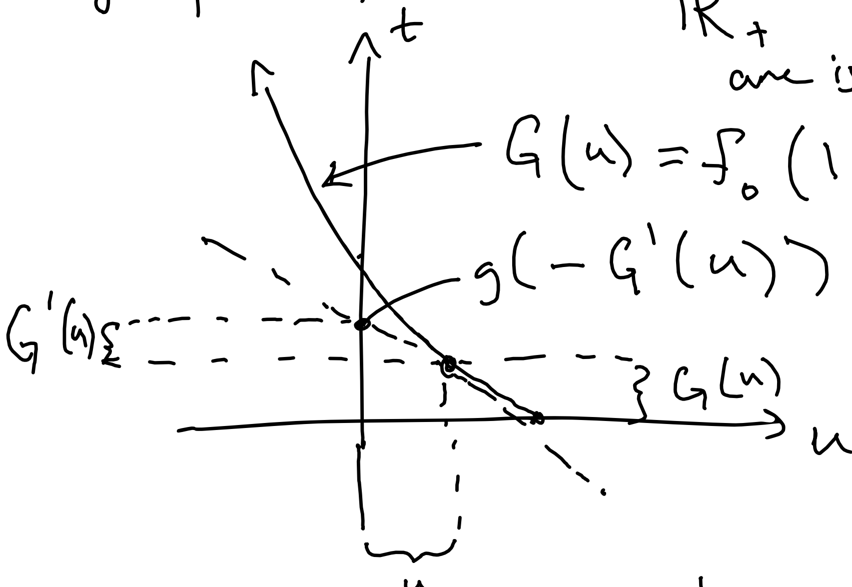


Notice instead of solving

$$\max_{\lambda \geq 0} \inf_{x \in D} x^3 + \lambda(1-x)$$

we can instead solve for $g(\lambda)$ "graphically". As $x \longleftrightarrow 1-u$

$\mathbb{R}_+$ $1-\mathbb{R}_+$ are isomorphic,



$$G(u) = (1-u)^3 \quad G'(u) = 3(1-u)^2$$

For $\lambda = -G'(u)$, $g(\lambda) = uG'(u) + G(u)$

$$= 3u(1-u)^2 + (1-u)^3$$

B/c $x \longleftrightarrow u$ (really, we don't need full isomorphism, but just that Image(G') is the range of dens. λ),

$$\max_{\lambda \geq 0} \inf_{x \in D} x^3 + \lambda(1-x)$$

$$= \max_{\lambda \geq 0} g(\lambda)$$

$$= \max_{u \in 1-\mathbb{R}_+} 3u(1-u)^2 + (1-u)^3$$

Which gets rid of the $\inf\limits_{x \in D}$ step!

(and indeed $d^* = 1$ @ $u^* = 0$)

CVX? ✓ SF? ✓ SD? ✓

5.21

$$\min_{x,y} \; e^{-x}$$

$$\text{st} \quad x^2/y \le 0$$

$$D = \{(x,y) \mid y > 0\}$$

(a) Why is this cvx?

$$(e^{-x})'' = e^{-x} > 0 \quad ✓$$

$$f(x,y) = x^2/y$$

$$\nabla^2 f = \frac{2}{y}(1 \;\; -x/y)\begin{pmatrix} 1 \\ -x/y \end{pmatrix} \succeq 0 \text{ on } D$$

(b) $$\max_{\lambda \geq 0} \inf_{\substack{x \\ y>0}} e^{-x} + \lambda(x^2/y)$$

the sequence (n, n^3) as $n \to \infty$

shows $d^* = 0$ for all λ.

(c) No SF

(d) For perturbed $p^*(u)$

$$\min_D e^{-x}$$

$$\text{st } x^2/y \le u$$

Choosing limit sequence from before yields

$$p^*(u) = \begin{cases} 0 & u > 0 \\ 1 & u = 0 \\ \infty & u < 0 \end{cases}$$

Thus global sensitivity inequality

$$p^*(u) \not\geq p^*(0) - \lambda^* u$$

does not hold.

(Note SD is required for non-trivial global sensitivity inequality).

Picture of 5.21,

non-SD

non-SF

CVX problem