Semi-infinite programming: What happens when you want to find

$$\min_x \max_y f(x,y)?$$

One can form $\varphi(x) = \min_y f(x,y)$, but this function might fail to be smooth:

Also, what if one has a continuous family of constraints:

$$g(x,z) \le 0 \quad (\text{before, } z \in \text{finite set}).$$

Typical examples:

- Distance between 2 sets $\quad d(X,Y) = \max_x \min_y |x-y| +$ symmetric term.

- Find ellipsoid of smallest area containing some body B: $\det X \to \max$: $X \succeq 0$,

$$(X\bar{z},\bar{z}) - 1 \le 0 \quad \text{for } z \in B.$$

Recall

Lemma If $K \subset \mathbb{R}^n$ is compact, and $\{h : (h,x) < 0 \text{ for all } x \in K\}$ (what we used to call K^*) is empty, then $0 \in \operatorname{conv}(K)$.

Thm: If in $\boxed{\begin{array}{l} \min_x \max_{y \in Y} f(x,y) \text{ subj to} \\ g(x,z) \le 0, \ z \in Z \end{array}}$ ← semi-infinite program

(here $x \in X \leftarrow$ open; Y, Z are compact, $f, g, \nabla_x f(x,y)$ $\nabla_x g(x,z)$ are continuous), and x_* is a local minimizer, then there is a <u>finite number</u> of y_j, z_k's at most $n+1$ altogether:

$$f(x_*, y_j) = \max_y f(x_*, y) \text{ for all } j; \quad g(x_*, z_k) = 0 \text{ for all } k$$

and $\quad \sum \lambda_j \nabla_x f(x_*, y_j) + \sum \mu_k \nabla_x g(x_*, y_j) = 0$ for some $\lambda, \mu \ge 0$, not all $= 0$ ...

Proof: $\varphi(x) = \max_{y \in Y} f(x,y)$; $\gamma(x) = \max_{z \in Z} g(x,z)$

$$\Theta(x) = \max\left(\varphi(x) - \varphi(x_*), \gamma(x)\right)$$

If x_* is local minimizer (i.e. in a vicinity of x_*, for any x: $g(x,z) \le 0$ for all $z \Rightarrow \varphi(x) \ge \varphi(x_*)$) then $\Theta(x) \ge 0$ near x_* (and $\Theta(x_*) = 0$).

Define active sets $Y_* = \{y : f(x_*, y) = \max_{y \in Y} f(x_*, y)\}$

$Z_* = \{z : g(x_*, z) = 0\}$.

If for some $h \in \mathbb{R}^n$, and any $y \in Y_*, z \in Z_*$

$(\nabla_x f(x_*, y), h) < 0$ &

$(\nabla_x g(x_*, z), h) < 0$, then $x_* + th$ will have $f(x_* + th, y) < \varphi(x_*)$ for small t, and all y (this is true over some $[0, t(y))$) for all y, whether in Y_*, or $Y - Y_*$; hence true for some close y's; by compactness can choose finite over, and smallest of $t(y_\alpha) > 0$), and same way $g(x + th, z) \le 0$ for small t, all z. This contradicts assumption on $x_* \Rightarrow$

$$\bigcap_{y \in Y_*} \{h : (\nabla_x f(x_*, y), h) < 0\} \bigcap$$

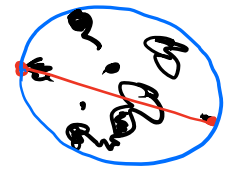
$$\bigcap_{z \in Z_*} \{h : (\nabla_x g(x_*, z), h) < 0\} = \emptyset.$$

So: $0 \in$ conv. hull of gradients $\nabla_x f(x_*, y)$ $y \in Y_*$, $\nabla_x g(x_*, y)$ $z \in Z_*$

But, by Radon's theorem, one can find $\le n+1$ vectors among those, whose convex hull is $0 \Rightarrow$ conclusion $\square$

Example: Jung's Inequality.

Let $S \subset \mathbb{R}^n$ be a compact set.



Define $D(S) = \sup_{x,y \in S} \|x-y\|$.

$R(S)$ = smallest radius of ball containing S.

i.e. $R^2(S) = \min_x \max_{y \in S} \|x-y\|^2$

Thm $D(S) \geq \sqrt{\frac{2(n+1)}{n}}\, R(S)$.

Take $\varphi(x) = \max_{y \in S} \|x-y\|^2$. φ is max of strictly convex fns; coercive $\Rightarrow$ unique minimizer x_*.

By thm above, there are $\leq n+1$ points in S:

$\|x-y_j\|^2 = R^2(S)$ and for some $\lambda_j \geq 0$

$$\sum \lambda_j \nabla_x \|x_*-y_j\|^2 = 2\sum \lambda_j (x_*-y_j) = 0$$

Can rescale and shift, so that $x_* = 0$, $\sum \lambda_j = 1$, $\lambda_j \geq 0$, $\sum \lambda_j y_j = 0$

Then $\sum_{j,k} \lambda_j \lambda_k \|y_j - y_k\|^2 = \sum_{j,k} \lambda_k \lambda_j \|y_j\|^2 + \sum_{j,k} \lambda_k \lambda_j \|y_k\|^2 - 2\sum_{j,k} \lambda_j \lambda_k (y_j, y_k)$

$$= 2\sum \lambda_k \|y_k\|^2 - 2\Big(\overbrace{\sum \lambda_j y_j, \sum \lambda_k y_k}^{0}\Big) = 2R^2(S)$$

$$\Rightarrow 2R^2(S) = \sum_{j \neq k} \lambda_j \lambda_k \|y_j - y_k\|^2 \leq \sum_{j \neq k} \lambda_j \lambda_k \cdot \max \|y_j - y_k\|^2 \leq$$

$$\leq \sum_{j \neq k} \lambda_j \lambda_k \cdot D^2(S) \quad \&$$

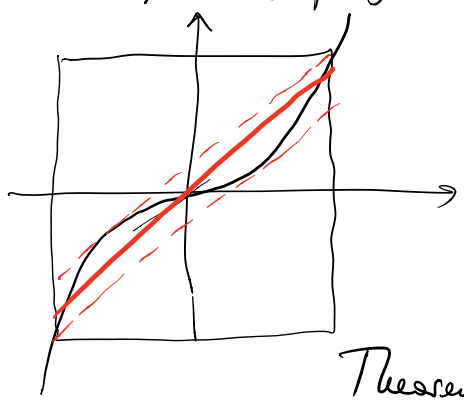
$\sum \lambda_k^2 + \sum_{j \neq k} \lambda_j \lambda_k = 1$ so that $\sum_{j \neq k} \lambda_j \lambda_k = 1 - \sum \lambda_k^2 \leq 1 - \frac{1}{k}$ by Cauchy-Schwartz

$\left(\sum \lambda_k^2 \cdot \sum 1^2 \geq \left(\sum \lambda_k\right)^2 = 1 \Rightarrow \sum \lambda_k^2 \geq 1/k \,;\, 1 - \sum \lambda_k^2 \leq 1 - 1/k \right)$.

So, $R^2(S) \leq \left(\frac{k-1}{k}\right) D^2(S)$ $\square$

Inequality exact: see regular simplex ...

<u>Last example</u>: Best approximation for x^n on $[-1,1]$ in L_∞ norm, among polynomials of degree $< n$.

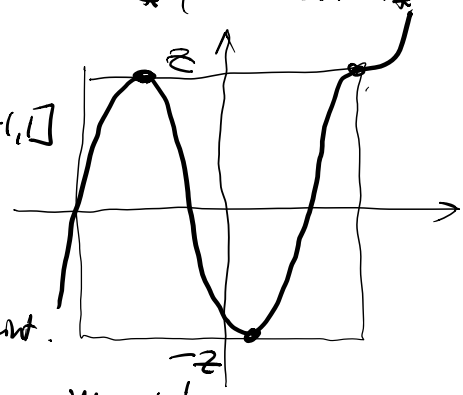


In other words, we solve

$$z \to \min \text{ subj to}$$

$$-z \le x^n + \sum_{k=1}^{n} a_k x^{n-k} \le z \quad \text{for } x \in [-1,1].$$

Theorem implies there are $\le n+1$ points, where the constraints are active, for optimal $\bar{a}_*$ (we call P_* corresponding polynomial).



Define those points $t_k, t'_\ell \in I = [-1,1]$ where $P_*(t_k) = z$; $P_*(t'_\ell) = -z$.

① If t_k or $t'_\ell \in (-1,1)$, then multiplicity of $(P_* \pm z)$ at that point is even. Call these multiplicities m_k, m'_ℓ.

② Sum of m_ℓ is n: otherwise, $P - \varepsilon \prod (t - t_\ell)^{m_\ell} - \varepsilon^2$ fits into a smaller z-box. Can use $\min(m_\ell, 2) \to$ to prove $m_\ell = 2$

③ One has -1 and $+1$ among $\{t_k, t'_\ell\}$ ... (Otherwise, taking $P_*(t-a)$ can be used to show that $\sum m_\ell$ or $\sum m'_k$ is less than n, contradiction.

OK: All that implies that $\tilde{P} = \frac{1}{z} P_*$ is such that the leading coeff. is x^n/z; and $\tilde{P} \pm 1$ has n root in $[-1,1]$. Take $(\tilde{P}')^2 (x^2-1)$ — it has same roots as $(\tilde{P}^2 - 1)$, and lead. coefficients are n^2/z^2 vs $1/z^2$. So

$$n\left(\tilde{p}^2-1\right)=\left(\tilde{p}'\right)^2\left(x^2-1\right)$$, i.e. $\tilde{p}$ solves

$$\frac{p'}{\sqrt{1-p^2}}=\frac{n}{\sqrt{1-x^2}} \Rightarrow \int\frac{dp}{\sqrt{1-p^2}}=n\int\frac{dx}{\sqrt{1-x^2}}$$

$$\cos p = n\cos x + c \quad \ldots \quad \square$$