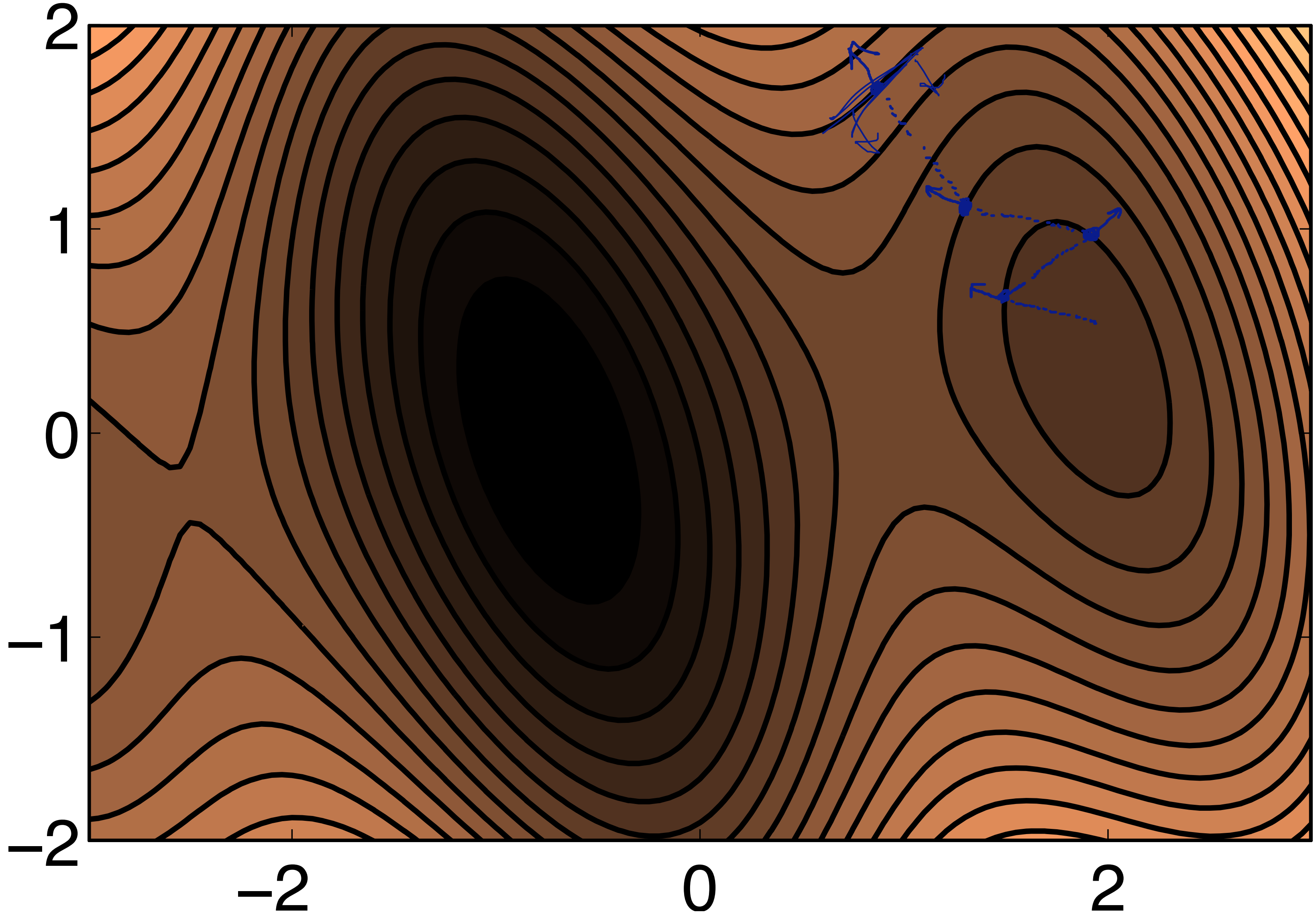
First-order methods
Convexity

10-725 Optimization
Geoff Gordon
Ryan Tibshirani

Gradient descent



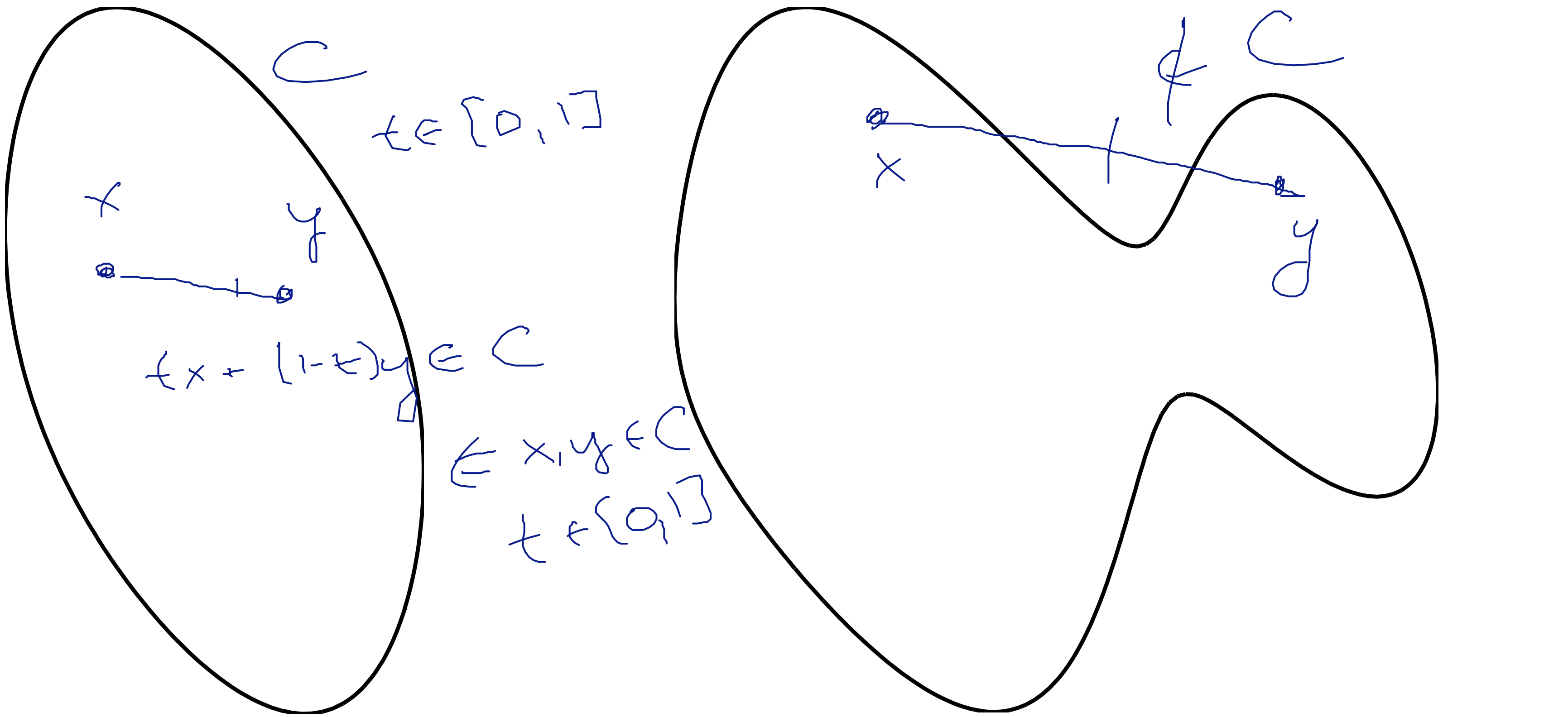
When do we stop?

- Using holdout set, if $f(x) = E(f_i(x) \mid i \sim P)$

$$E_{a,b}(a \cdot x - b)^2$$

- Using convergence bounds (later)
  - usual form is:
    - $K_f\,(f(x_0) - f(x^*))$ [some fn of $1/\epsilon$]
  - need estimates of first two terms
- For $f(x^*)$, duality (later); for K_f, properties of f:
  - convex? strongly convex? Lipschitz?

Convex sets



For all x, y ∈ C, for all t ∈ [0, 1]:

$tx + (1-t)y \in C$

$P(x) \in C$

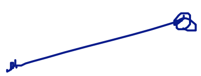
$E_{x \sim P}(x) \in C$

Examples

point

$\emptyset$

$\mathbb{R}^n$



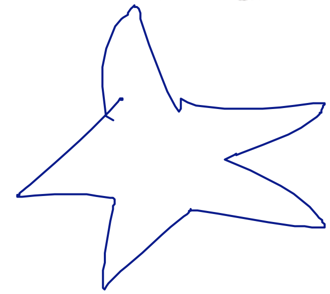
cube

$\|x\|_p \le 1$

$p \ge 1$

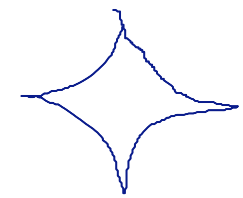
yes

no



$\|x\|_p = \sqrt[p]{\sum_i |x_i|^p}$

$\ell_p \quad p < 1$



$B_p(x, \epsilon) = \{y \mid \|x-y\|_p < \epsilon\}$

take $p=2$ below

Boundaries

- x on boundary of C (∂C) if:

  $B_p(x,\epsilon) \cap C \neq \emptyset$ $B_p(x,\epsilon) \cap C^c \neq \emptyset$

- x in interior of C if:

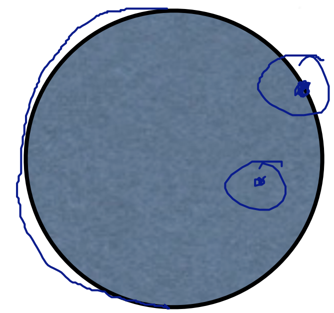
  $B_p(x,\epsilon) \subseteq C$ for small ϵ

- x in *relative* interior (rel int C) if:

  subspace $S \supseteq C$ restrict to S

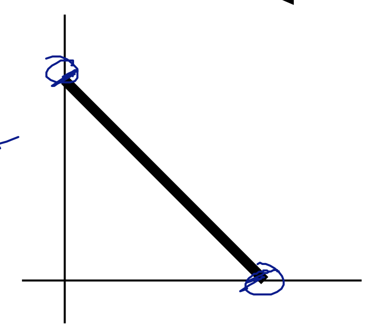
  int C restrict S

- C closed if: $\partial C \subseteq C$
- C open if: $\partial C \cap C = \emptyset$
- C compact if: closed bounded



$\{x \mid \|x\| \leq 1\}$

$<$



$\{x \mid x \geq 0, x_1+x_2=1\}$

$x_1 \geq 0 \wedge x_2 \geq 0$

R^n

$\emptyset$

Convex hull

$$\left\{ \sum_i a_i x_i \mid x_i \in C,\ a_i \geq 0,\ \sum a_i = 1 \right\}$$

Any closed, convex C

$= \text{hull}(X)$

$\uparrow$

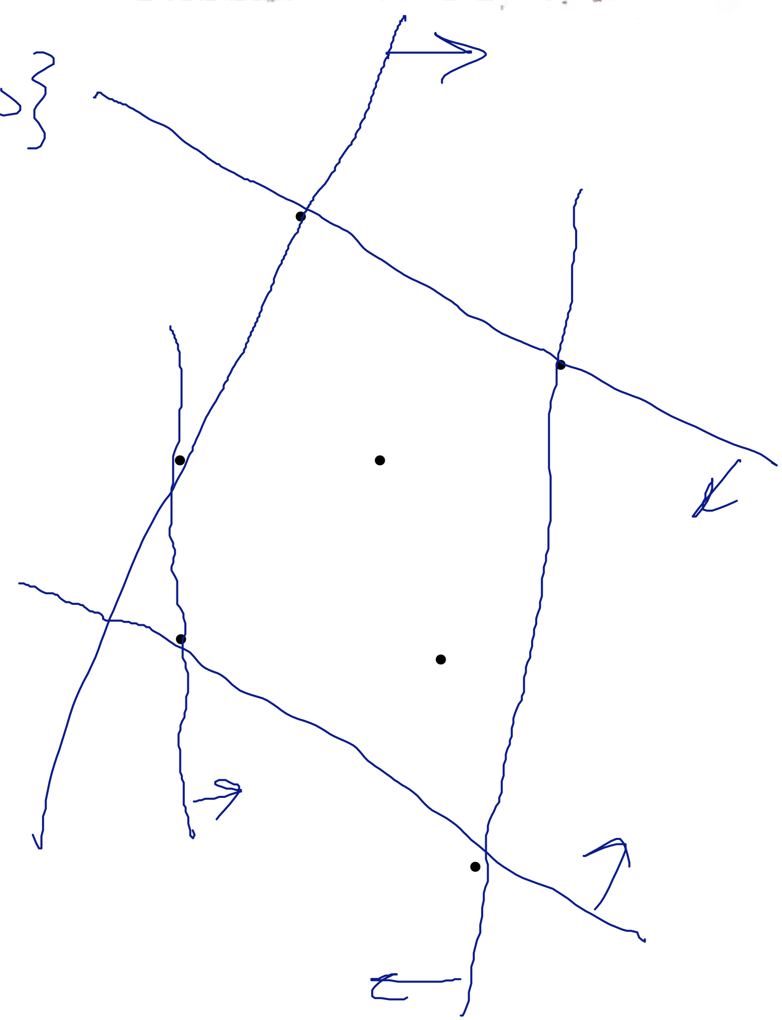
$|X| = \infty$

Dual representation

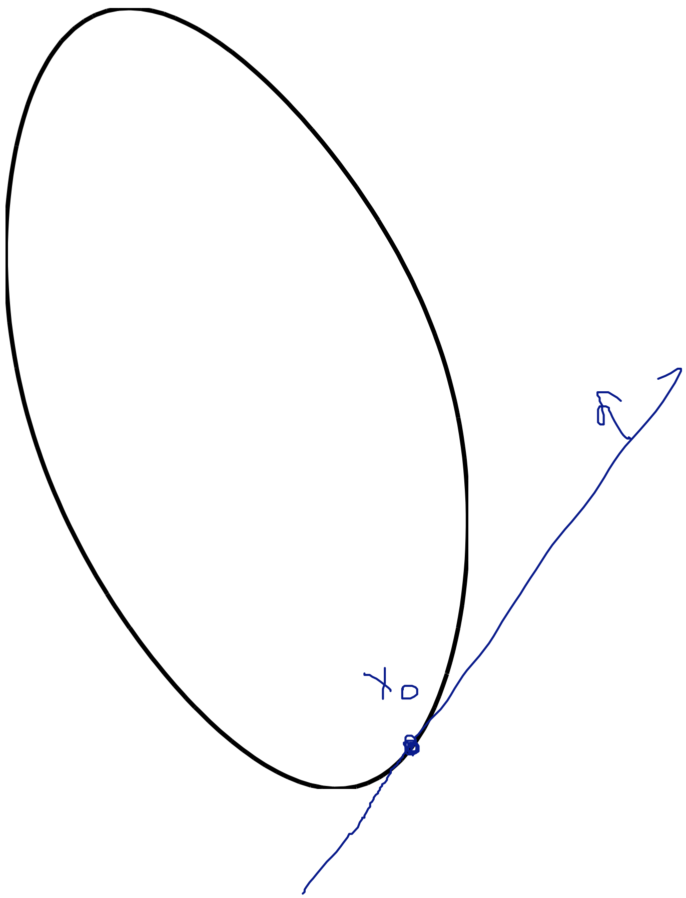
$$C = \bigcap_{a_i, b_i} \{ x \mid a_i \cdot x + b_i \le 0 \}$$

Any closed convex C

$=$ this



Supporting hyperplane thm

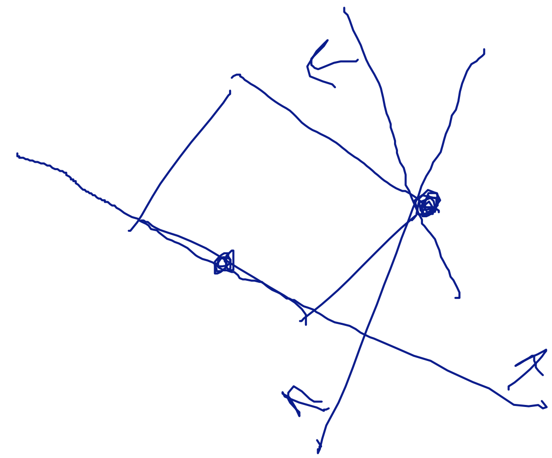


- For any point x_0 on boundary of convex C:
  - exist (w, b) with
  - $x_0 \cdot w = b$
  - $x \cdot w \le b \quad \forall x \in C$

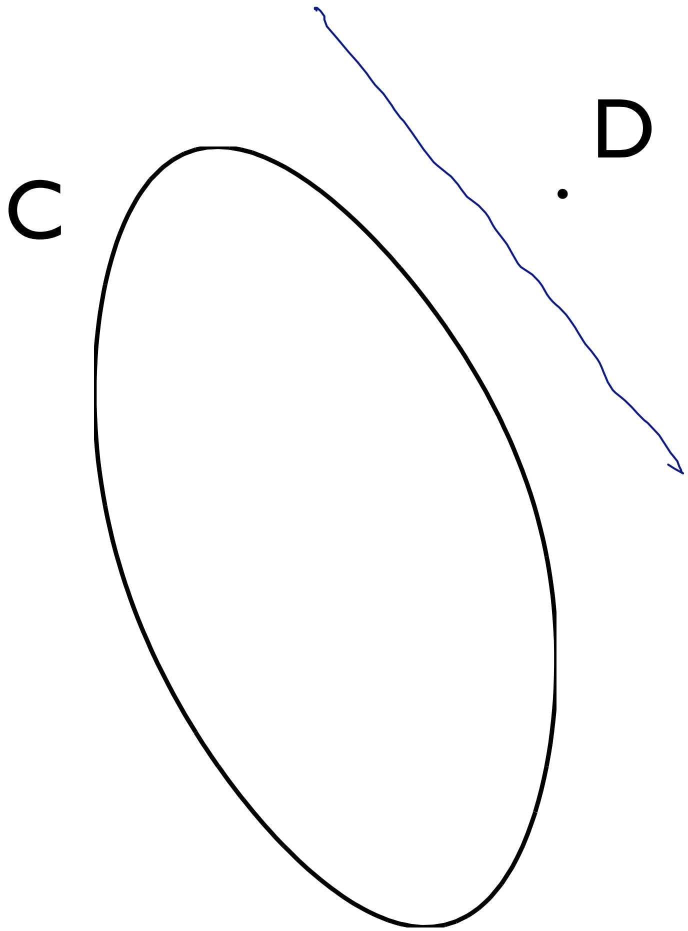
Supporting hyperplane exs

$x^2 + y^2 < 1$

$|x| + |y| \leq 1$



Separating hyperplane thm



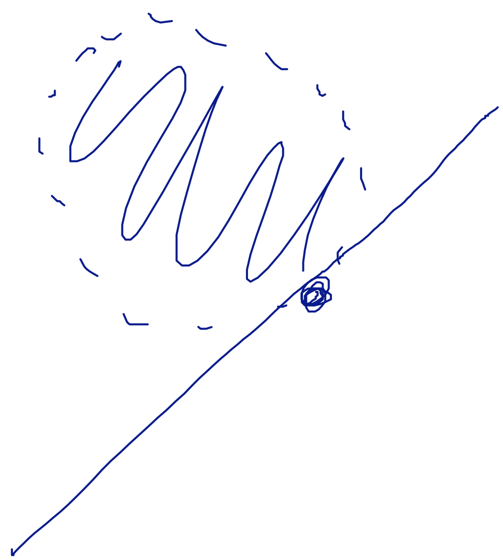
- For any convex C and D with $C \cap D = \emptyset$
  - exist (w, b) with
  - $w \cdot x \le b \quad x \in C$
  - $w \cdot x \ge b \quad x \in D$
- If both C, D are closed, and at least one compact:

strict

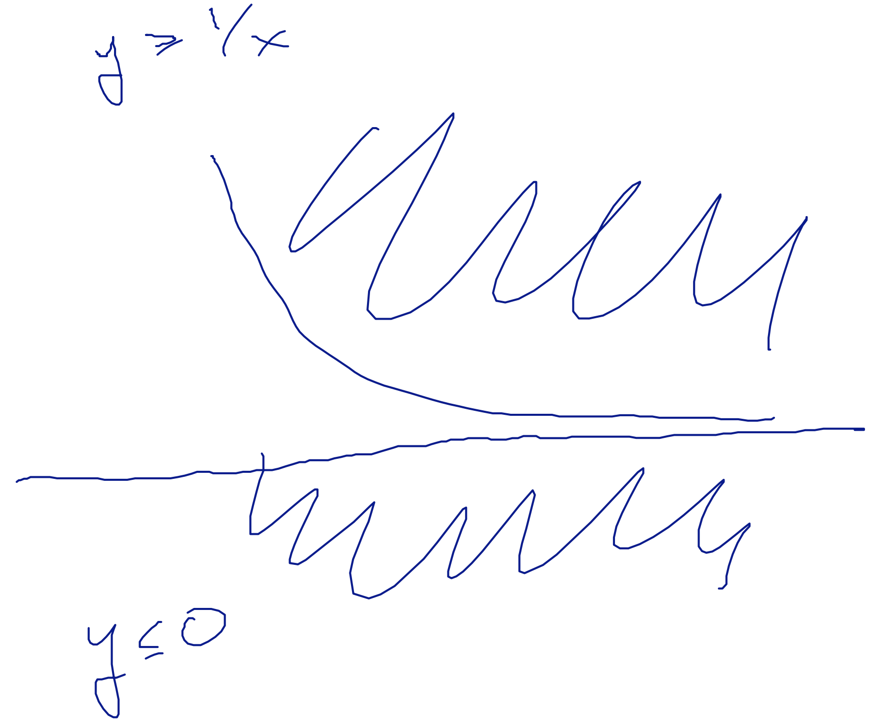
$w \cdot x < b$

$w \cdot x > b$

Separating hyperplane exs

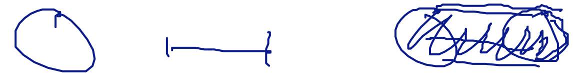


halfspace & complement



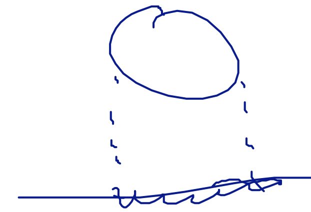
Proving a set convex

- Use definition directly $tx + (1-t)y$
- Represent as convex hull or ∩ of halfspaces
- Supporting hyperplane partial converse
  - C closed, nonempty interior, has supporting hyperplane at all boundary points ⇒ C convex
- Build C up from simpler sets using convexity-preserving operations

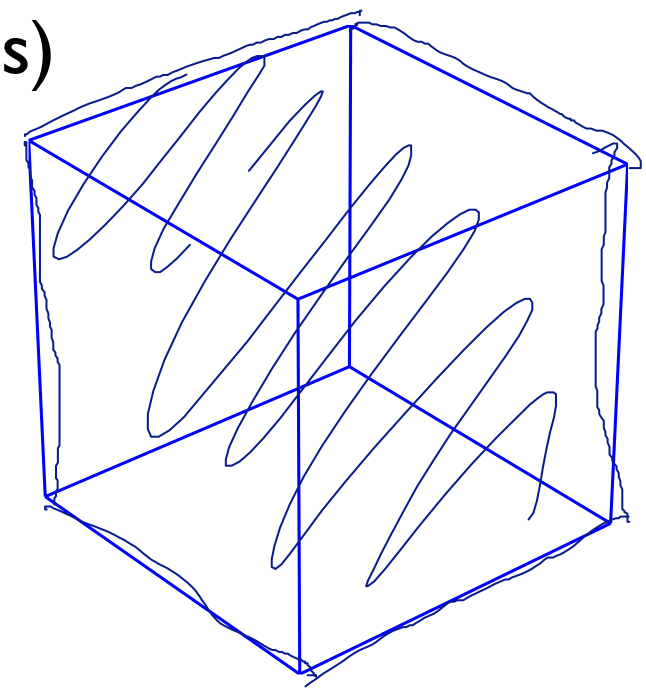


Convexity-preserving set ops

- Translation $\{x+b \mid x \in C\} = C+b$
- Scaling $\{\lambda x \mid x \in C\}$ λC



- Affine fn $\{Ax+b \mid x \in C\}$ $AC+b$
    - projection (e.g., dropping coords)



- Intersection $C \cap D$
- Set sum $C+D = \{x+y \mid x \in C, y \in D\}$
- Perspective $(x,y,z) \in C$ $z>0$ $(x/z, y/z)$

Ex: symmetric PSD matrices

$x^T A x \geq 0 \quad \forall x$

- Two proofs that $\{A \mid A = A^T, \underline{A \succeq 0}\}$ is convex
  - $x^T (tA + (1-t)B)\, x =$

    $t\, \underline{x^T A x} + (1-t)\, \underline{x^T B x}$
  - $x^TAx =$

    $\sum_i \sum_j x_i A_{ij} x_j$ ← linear fn of A

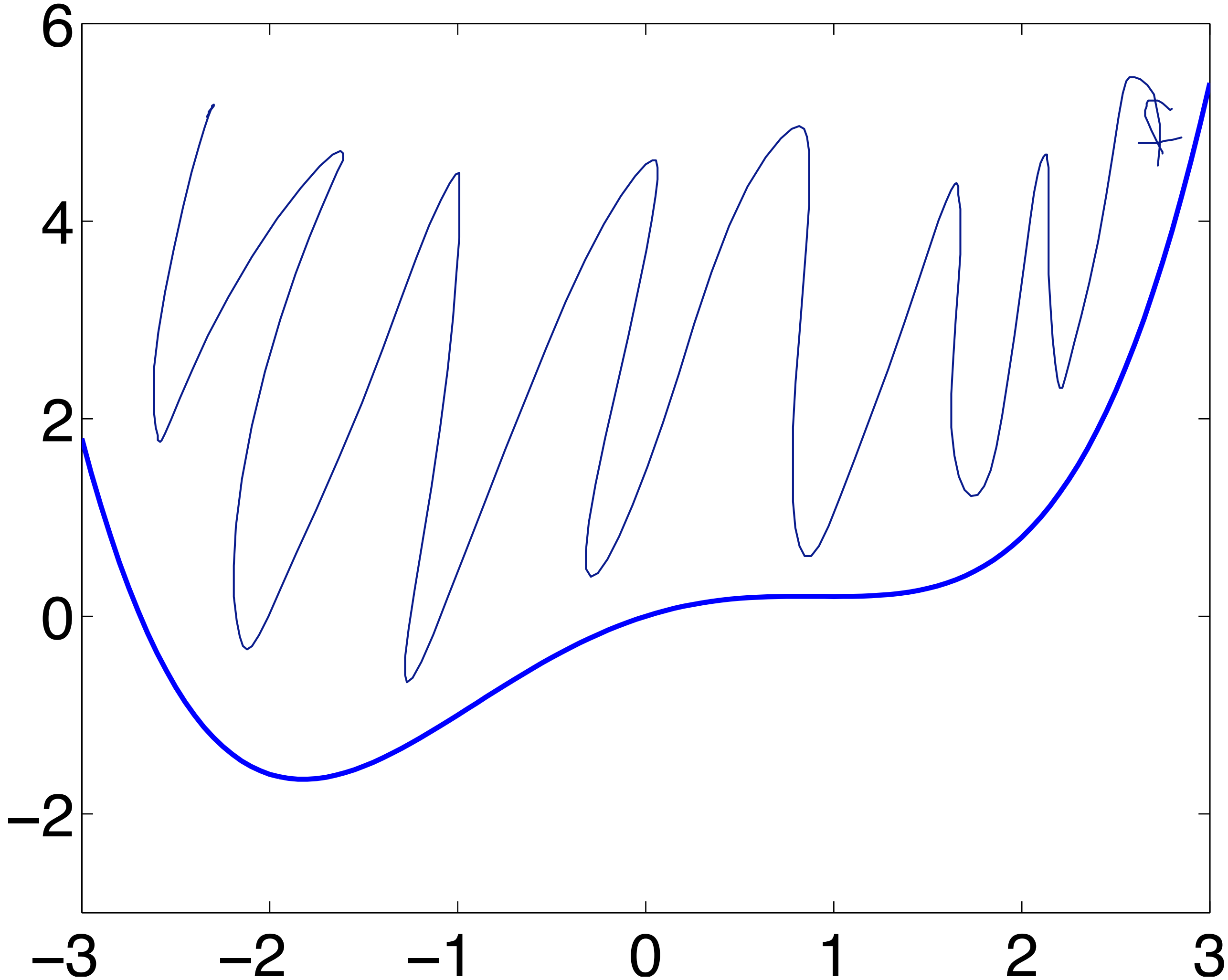
    ≥ 0

Ex: conditionals

$x \in \{0,1\}^7$

- Given a convex set of dist'ns $P(x_{1:7})$,
  - $P(x_{1:5} \mid x_{6:7}) =$ $P(x_{1:7}) / P(x_{6:7})$
  - numerator:
  - denominator:
- Convex? ✓

Epigraph



$\{(x,t) \mid t \geq f(x)\}$

Domain

- dom f = $\{x \mid f(x) \text{ exists } f(x) < \infty\}$

- dom 1/x = $\mathbb{R} - \{0\}$
- dom ln(x) = $x > 0$

f convex $\iff$ epi f convex

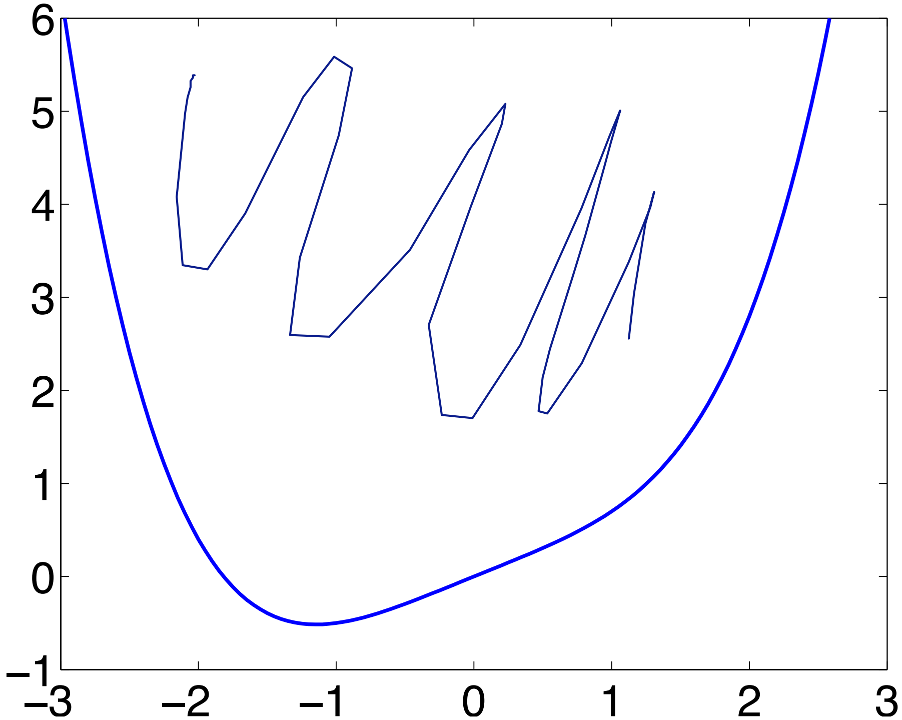
Convex functions

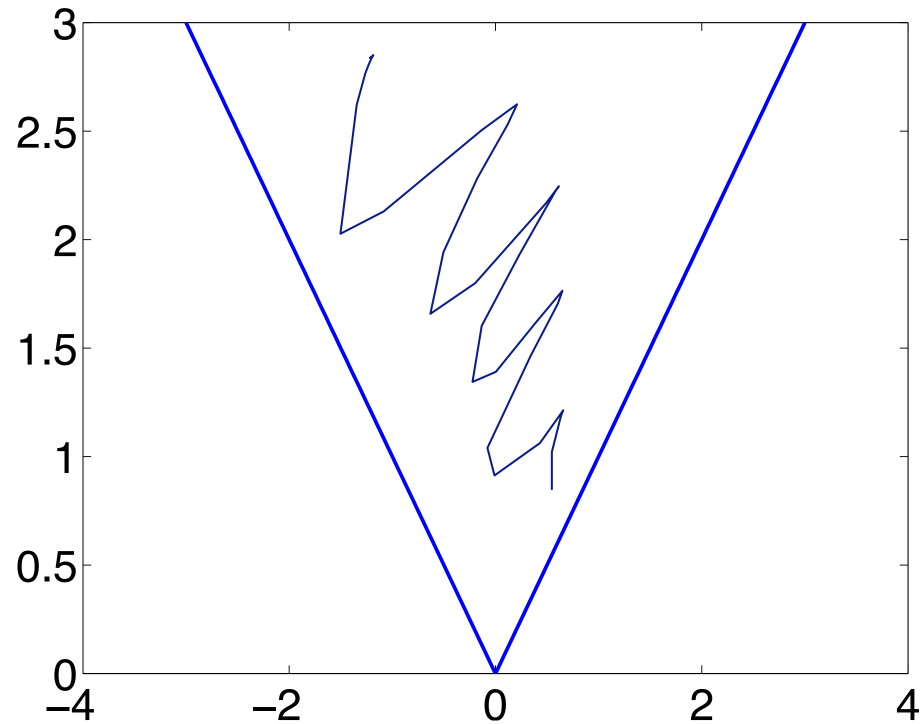
dom f convex

$t f(x) + (1-t) f(y) \geq$

$f(tx + (1-t) y)$

$\forall x, y \in \mathrm{dom}\, f$





$P(x)$ $x \in \mathrm{dom} f$

$E_{x\sim P} f(x) \geq f(E_{x\sim P}\, x)$

Relating convex sets and fns

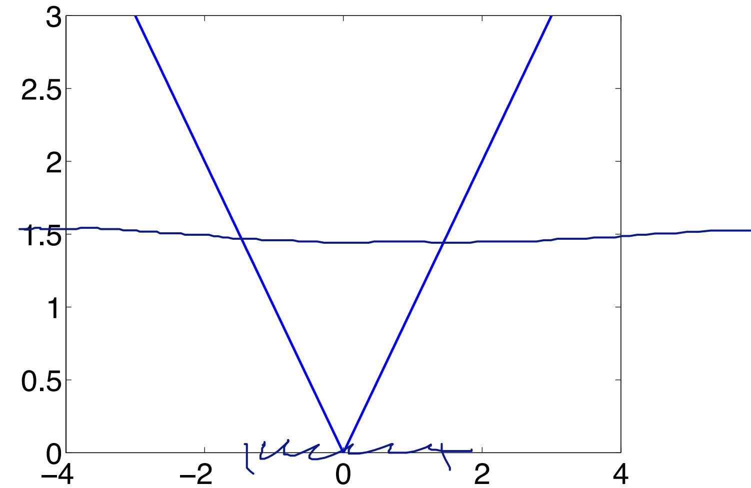
sublevel set

- f(x) convex ⇒ { x | $f(x) \le k$ } convex

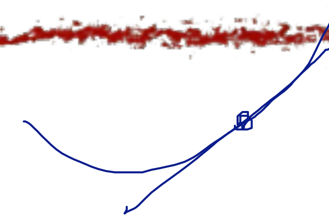
- Converse?

$f(x) = \sqrt{|x|}$

quasi convex



Proving a function convex



- Use definition directly
- Prove that epigraph is convex via set methods
  - e.g., supporting hyperplanes: for all x, y, $f(x) \geq f(y) + (x-y)f'(y)$
  - this is first-order convexity condition for fns
- 2nd order: $\frac{d^2}{dx^2} f(x) \geq 0$
- Construct f from simpler convex fns using convexity-preserving ops

Convexity-preserving fn ops

- Nonnegative weighted sum $(af + bg)(x) = af(x) + bg(x)$
- Pointwise max/sup $g(x) = \sup_{i \in I} f_i(x)$
- Composition w/ affine $f(Ax + b)$
- Composition w/ monotone convex g monotone convex $g(f(x))$
- Perspective $\tilde{f}(x,t) = f(x/t)\,t$, $t > 0$
- f(x, y) convex in (x, y), set C convex:
    - g(y) = $\min_{x \in C}$ f(x,y) is convex if g(y) > –∞

epi g

epi f $= \{(x,y,t) \mid t \geq f(x,y)\}$

Example: $f(x) = |x|$

$$|x| \geq x$$
$$|x| \geq -x$$

$$|tx + (1-t)y|$$

if $tx + (1-t)y \geq 0$:

$$|tx + (1-t)y| = tx + (1-t)y \leq t|x| + (1-t)|y|$$

o/w

$$|tx + (1-t)y| = -tx - (1-t)y$$
$$\leq t|x| + (1-t)|y|$$

In 2 or more dimensions

$$f(x) \geq f(y) + (x-y) \cdot \nabla f(y)$$

- All the above, but for 2nd order:

$$H = \frac{d^2}{dx^2} f(x)$$

$$\begin{pmatrix} \frac{d^2}{dx^2} f & \frac{d^2}{dx\,dy} f \\ \frac{d^2}{dx\,dy} f & \frac{d^2}{dy^2} f \end{pmatrix}$$

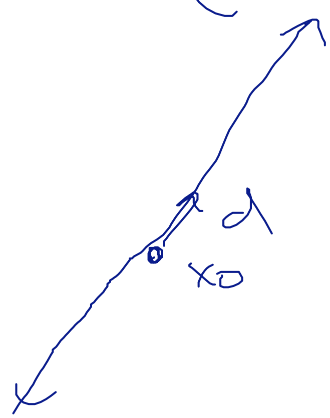
$$H \succeq 0$$

$$\mathrm{eig}(H)$$

- Or: reduce to 1D

$$x_0 + td$$

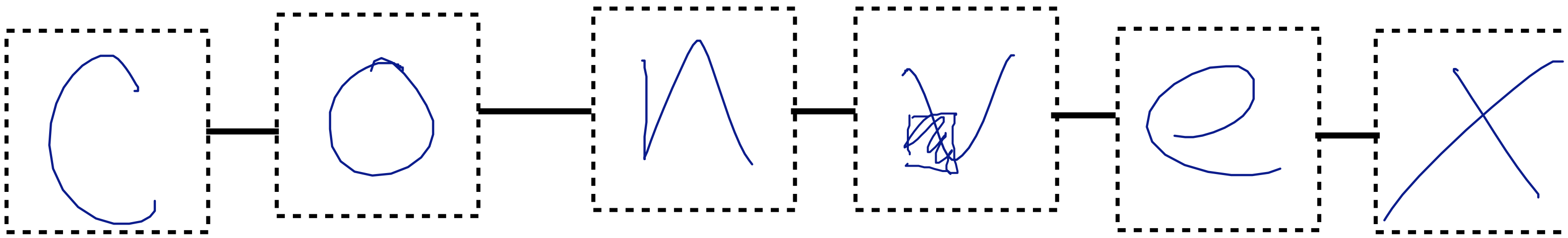
$$f(x_0 + td)$$



Ex: structured classifier

c o n v e x

Ex: structured classifier



x_i pixels of char i

y_i $a \dots z$

$\varphi_j(x_i)$ feature of a char

$\psi_{ijk}(x_i, y_i)$ $\phi_j(x_i)\,\delta(y_i = k)$ $\leftarrow w_{ijk}$

$\chi_{ikl}(y_i, y_{i+1})$ $\delta(y_i = k)\,\delta(y_{i+1} = l)$ $\leftarrow v_{ikl}$

$$L(x,y;v,w) = \sum_{ijk} \psi_{ijk} w_{ijk} + \sum_{ikl} \chi_{ikl}(y_i, y_{i+1})\, v_{ikl}$$

Classifier: $y = \arg\max_y L(x,y;v,w)$

Learning structured classifier

- Get it right if: $L(x, y; v, w) > L(x, y', v, w) \quad \forall y' \neq y$
- So, want: $L(x, y; v, w) \geq \max_{y'} \left(L(x, y'; v, w) + \pi(y, y') \right)$
- Where π(y,y') = $\begin{cases} 0 & y = y' \\ > 0 & y \neq y' \end{cases}$
- RHS: convex in v, w
- RHS – LHS: convex
- Train: lots of pairs (x^t, y^t) $\quad \min_{v,w} \sum_t (RHS^t - LHS^t)$