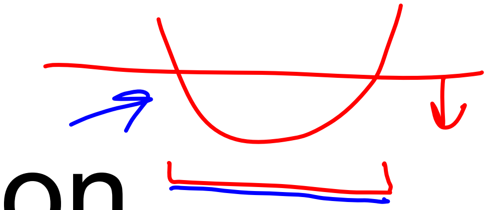


Convex optimization

minimize $\overset{\text{convex}}{f(\overset{\mathbb{R}^n}{x})}$ subject to $\overset{\mathbb{R}^{m\times n}}{A}x = \overset{\mathbb{R}^m}{b}$

$\overset{\text{convex}}{g_i(x)} \le 0 \quad \forall i \in I$

e.g., min x^2+y^2 s.t. $\overbrace{1-x-y}^{g_1(x,y)} \le 0$

Linear inequalities: OK

Positivity: OK

} just include in g_i's

Optimality

- Optimal value:

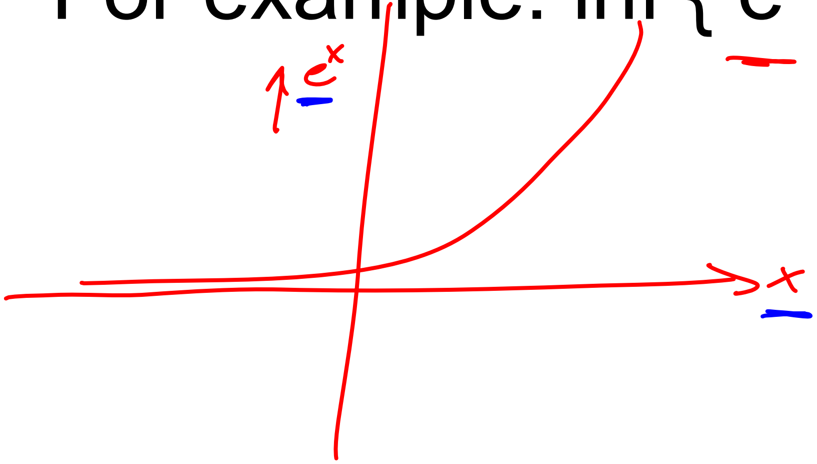
  $v^* = \inf \{ f(x) \mid g_i(x) \le 0 \ (\forall i), Ax = b \}$

Also, $v^* \le x$ $\forall x \in S$

- Definition of $v^* = \inf S$: $\subseteq \mathbb{R}$ infimum

$\iff \forall \epsilon > 0 \ \exists x_\epsilon \in S. \ v^* + \epsilon \ge x_\epsilon$

- For example: $\inf \{ e^x \mid x \text{ real} \} =$



Sup (supremum)

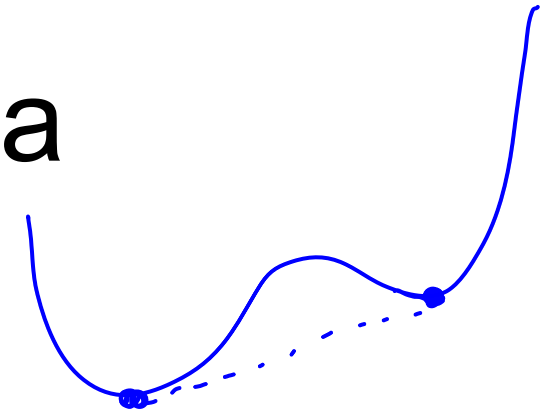
tightest upper bound

Optimal point

- $\underline{x^*}$ is optimal iff: $f(x) = v^*$ $Ax^* = b$ $\overset{\forall i}{g_i(x^*) \le 0}$

- Is there always an optimal point?
  - ex. 1: $\max \log(x)$ s.t. $x \geq 0$
  - ex. 2: $\min e^{-x}$ s.t. $x \geq 0$

Local optima



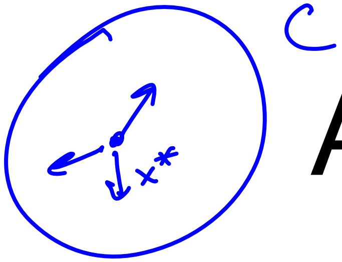
- min $\underline{f(x)}$ s.t. $\underline{g(x)} \le 0$
- $\underline{x_0}$ is a local optimum iff

x_0 optimal for min $f(x)$ s.t. $g(x) \le 0$
$\|x - x_0\| \le R$

for some $R > 0$

- In a convex program,

all local optima are also global optima



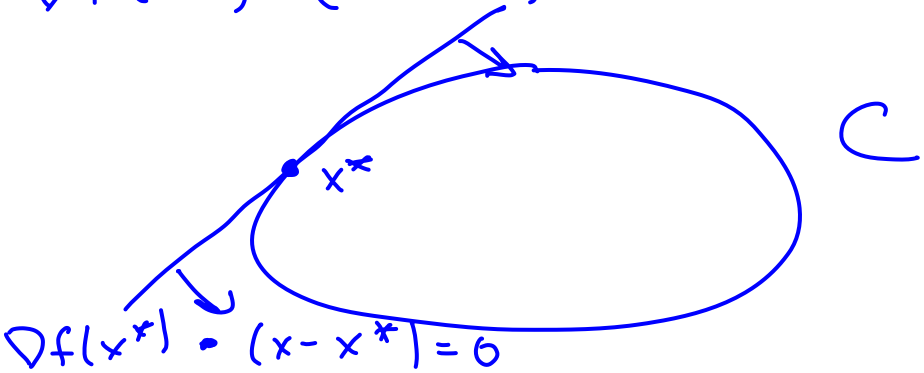
An optimality criterion

- Suppose f(x) is differentiable, convex
- Then x* = arg min f(x) iff $\nabla f(x^*) = 0$
- What about x* = arg min f(x) s.t. x ∈ C?

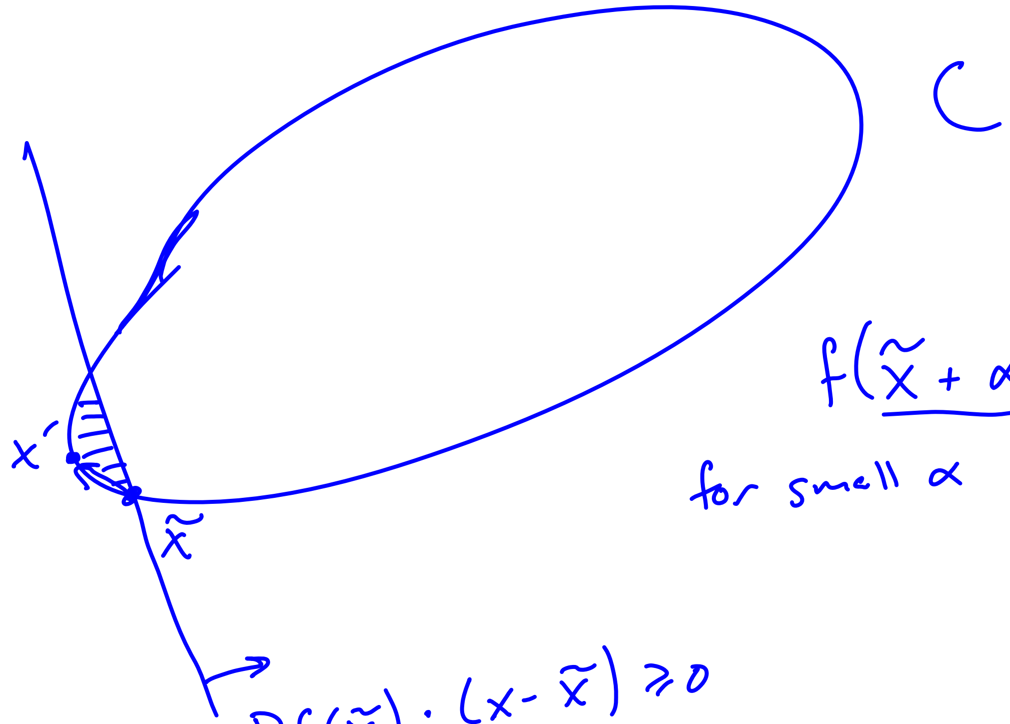
convex

$\rightarrow x^* \in \text{int } C \Rightarrow \nabla f(x^*) = 0 \Leftarrow$

$\rightarrow x^* \in \delta C \Rightarrow \nabla f(x^*) \cdot (x - x^*) \geq 0 \ \forall x \in C$



Proof of criterion



$Df(\tilde{x}) \cdot (x - \tilde{x}) \geq 0$

$$\underline{f(\tilde{x} + \alpha(x' - \tilde{x}))} \simeq f(\tilde{x}) + \nabla f(\tilde{x}) \cdot \underbrace{\alpha(x' - \tilde{x})}_{\leq 0}$$

for small α

LP “<” QP “<” SOCP “<” SDP

Types of convex program

$\text{SOC} = \{(x,t) \mid \|x\| \le t\}$

- Linear program, quadratic program
- Second-order cone program

$$\min \underline{f^T x} \quad \text{s.t.} \quad Fx = g$$

$$\|\underline{A_i x + b_i}\| < \underline{c_i \cdot x + d_i} \quad \forall i \in I$$

$y_i = A_i x + b_i$

$t_i = c_i \cdot x + d_i$

$\|y_i\| \le t_i$

- Semidefinite program

$$\min c^T x \quad \text{s.t.}$$

$$\underline{F_1 x_1 + F_2 x_2 + \dots F_k x_k} \succeq 0$$

$$Ax = b$$

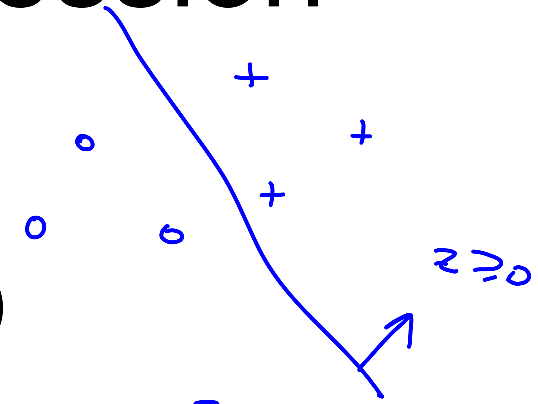
$F_i \in S^{n \times n}$

$$F_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$F_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

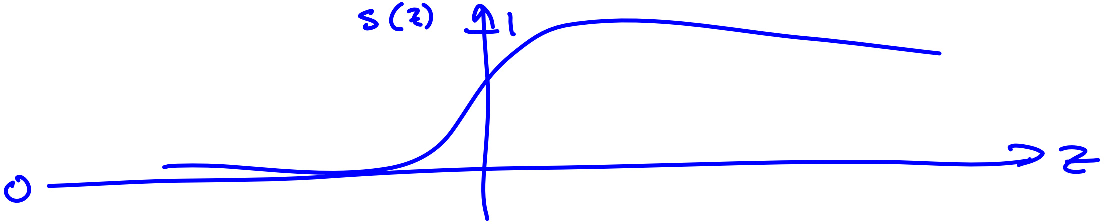
$$\vdots$$

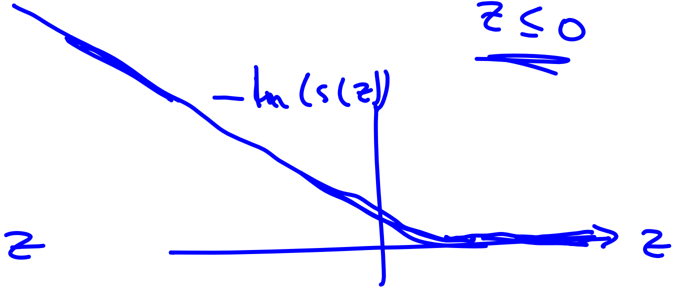
Example: logistic regression



- Data $x_i \in R^n$, $y_i \in \{0,1\}$
- $P(y_i = 1 \mid x_i, w, b) = s(w^T x_i + b)$

  $s(z) = 1/(1+\exp(-z))$





$-\ln(s(z)) = \ln(1+e^{-z})$

- $\max_{wb} P(w, b)\, P(y \mid x, w, b) =$

  $\max_{wb} P(w, b) \prod_i P(y_i \mid x_i, w, b) =$

$\max_{w,b} \ln P(w,b) + \sum_i (\ln s(x_i \cdot w + b)) y_i + (1-y_i) \ln(1 - s(-))$

Ex: minimize top eigenvalue

- Suppose $\underline{A_i} \in \underline{S^{n*n}}$ for i = {1, 2, …}
- $\min_{\alpha,A}$ $\underline{\lambda_1(A)}$ s.t.

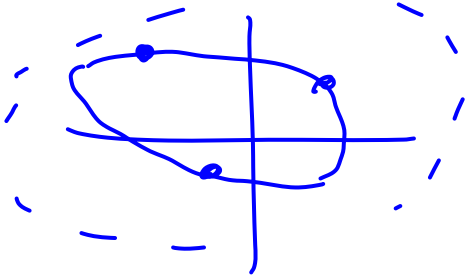
  $\underline{A} = \underline{\alpha_1 A_1 + \alpha_2 A_2 + \ldots}$

equivalent

$$\min_{\lambda,\ \alpha_i,\ A} \quad \lambda \quad \text{s.t.} \quad \lambda \geq x^T A x \quad \forall x \mid \|x\| \leq 1$$

$$\forall x \mid \|x\| = 1$$

Ex: minimum volume ellipsoid



- Given points $x_1, x_2, \ldots, x_k$
- $\min\limits_{A, x_C}$ vol(A) s.t.

  $(x_i - x_C)^T A \, (x_i - x_C) \le 1 \quad i = 1, \ldots, k$

  $A \in S^{n*n}$

  $A \succcurlyeq 0$

$$\text{vol}(A) \propto \frac{1}{\sqrt{|A|}}$$

$$\ln \text{vol}(A) = -\frac{1}{2} \ln |A| + \text{const}$$

$\min -\ln|A|$ s.t.

{constraints}

Schur complement

- Symmetric block matrix M = $\begin{pmatrix} A & B \\ B^T & C \end{pmatrix}$
- Schur complement is S = $C - B^T A^{-1} B$
- → M $\overset{\succ}{\cancel{\succeq}}$ 0 iff $A \succ 0$ and $S \succ 0$

$M \succeq 0$ iff $\min_{x} x^T M x = 0$

$x = \begin{pmatrix} u \\ v \end{pmatrix}$ $\quad x^T M x = \underline{u^T A u} + v^T C v + \underline{\underline{2u^T B v}}$

min w.r.t. u exists iff $A \succeq 0$ $\quad$ suppose $A \succ 0$

$0 = 2Au + 2Bv \Rightarrow u = -A^{-1} B v$

$\min_{x} x^T M x = \min_{v} (v^T B^T \cancel{A^{-T}} \cancel{A} A^{-1} B v + v^T C v - 2 v^T B^T A^{-1} B v)$

$= \min_{v} v^T C v - v^T B^T A^{-1} B v = \min_{v} v^T S v$

exists iff $S \succeq 0$

$\Rightarrow$ want: $x_i^T A x_i - 2 x_i^T A x_c + x_c^T A x_c \le 1$

Back to min-volume ellipsoid

- $\max_{A, x_C}$ $\underline{\log |A|}$ s.t.

$\rightarrow$ $\underline{(x_i - x_C)^T A (x_i - x_C) \le 1}$ $\underline{i = 1, \ldots, k}$

$\underline{A = A^T}$, $\underline{A \succcurlyeq 0}$

$y_i \equiv \begin{pmatrix} x_i \\ 1 \end{pmatrix}$

- $\max_{A, B, x_C, z}$ $\log |A|$ s.t.

$B = \begin{pmatrix} A & x_c \\ x_c^T & z \end{pmatrix}$

$B = B^T$ $B \succeq 0 \iff A \succeq 0 \wedge \underbrace{z - x_c^T A^{-1} x_c \succeq 0}$

$z - x_c^T A^{-1} x_c \ge 0$

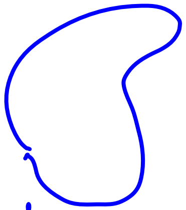
$z \le x_c^T A^{-1} x_c$

$y_i^T B y_i \le 1 \quad \forall i$

$\hookrightarrow$ $x_i^T A x_i - 2 x_i^T x_c + z \le 1$

Ex: soap bubbles

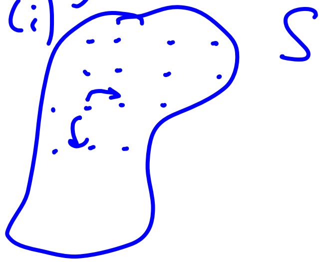
- Q: Dip a bent paper clip into soapy water. What shape film will it make?
- A: minimum area surface
- Write h(x,y) for height
- Suppose (x,y) in S →



- Area is: $\int_S \sqrt{1+\nabla h(x,y)^2}\, dxdy$

Soap bubbles

$\left.\begin{matrix} R(i) \\ D(i) \end{matrix}\right\} \rightarrow$ neighbors of i



$h_i = h(x_i, y_i)$

(x_i, y_i)

$i \in I$

area =

$$\sum_i \frac{1}{\epsilon^2}\sqrt{1+\left(\frac{h_i - h_{R(i)}}{\epsilon}\right)^2 + \left(\frac{h_{D(i)} - h_i}{\epsilon}\right)^2}$$

$x_{R(i)} = x_i + \epsilon$ ← grid width

$y_{R(i)} = y_i$

$x_{D(i)} = x_i$

$y_{D(i)} = y_i - \epsilon$

$$\min \sum_i \frac{1}{\epsilon^2} t_i$$

$$t_i \geq \sqrt{\quad \sim\!\sim\!\sim \quad} = \left\| \begin{matrix} 1 \\ (h_i - h_{R(i)})/\epsilon \\ (h_{D(i)} - h_i)/\epsilon \end{matrix} \right\|_2$$

Ex: manifold learning

- Given points $x_1, \ldots, x_m$ $\in R^{n} \leftarrow$ big
- Find points $y_1, \ldots, y_m$ $\in R^{d} \leftarrow$ $d \ll n$
- Preserving **local geometry**
  - – neighbor edges N $(i,j) \in N \iff$ preserve geometry btwn x_i, x_j
  - – distances $\|x_i - x_j\| = \|y_i - y_j\| \quad (i,j) \in N$
- If we preserve distances among x_i, x_j, x_ℓ

  we also preserve angles $(x_i - x_j) \cdot (x_i - x_\ell)$

Step 1: “embed” R^n into R^n

- While preserving local distances, move points to make manifold as flat as possible
- max

  s.t.

Step 2: reduce to R^d

- Now that manifold is flat, just use PCA: