Convex Sets (cont.) Convex Functions

Optimization - 10725

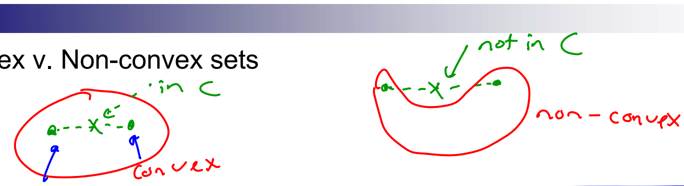
Carlos Guestrin

Carnegie Mellon University

February 27th, 2008

Definitions of convex sets

- Convex v. Non-convex sets

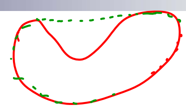


- Line segment definition: $\forall x_1, x_2 \in C, \ \forall \theta \in [0,1]$, $z = \theta x_1 + (1-\theta) x_2 \Rightarrow z \in C$
- Convex combination definition:
  $\theta_1, \dots \theta_k \geq 0$, $\sum_i \theta_i = 1$, if $x_1 \dots x_k \in C \Rightarrow \sum_i \theta_i x_i \in C$
- Probabilistic interpretation:
  - If $C \subseteq R^n$ is convex
  - Define a*ny* probability distribution $p(x)$: $p(x) \geq 0$ $\int_{x \in C} p(x)dx = 1$
  - Then $E[X] \in C$

see slides for detail ? $E[x]$ exist → fancy math

General convex hull

- Given some set C



- Convex hull of C, **conv** C

  $\mathrm{conv}\, C = \{x \mid x = \sum_i \theta_i x_i,\ x_i \in C,\ \theta_i \geq 0, \sum \theta_i = 1\}$

- Properties of convex hull:
  - Idempotency: $C \in$ convex $\ \mathrm{conv}\, C = C$, $\mathrm{conv}\, C = \mathrm{conv}\, \mathrm{conv}\, C$
  - Convexity:

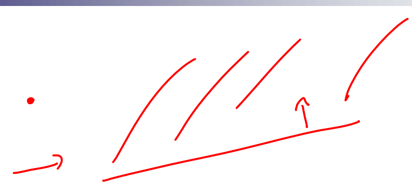
- Usefulness: obtain a lower bound on non-convex problem

  $\min_x f(x)$ ← convex s.t. $x \in C$ ← non convex) $\min_x f(x)$ $x \in \mathrm{conv}\, C$

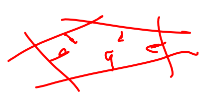
3

Examples of convex sets we have already seen…

- R^n
- point
- half space



- polyhedron



- line
- line segment
- linear subspace $\quad x \mid \overset{n}{_m A} x = \overset{1}{b}$ ← $m < n$ $\quad \mathbb{R}^n$, 0

  basis for subspace $v_1 \ldots v_k$

  $x_0 \in$ subspace $\quad x = \sum_i \lambda_i v_i + x_0 \in$ subspace

4

First non-linear example: Euclidean balls and Ellipsoids

- $B(x_c,r)$ - ball centered at x_c centered at r:

$$B(x_c,r) = \{x \mid \|x-x_c\|_2 \le r\}$$
$$= \{x \mid \sqrt{(x-x_c)^T(x-x_c)} \le r\}$$

- Convexity:

$x_1, x_2 \in B(x_c,r)$, $\theta x_1 + (1-\theta)x_2 \in B(x_c,r)$

$$\|\theta x_1 + (1-\theta)x_2 - x_c\|_2 = \|\theta x_1 + (1-\theta)x_2 - \theta x_c - (1-\theta)x_c\|_2$$
$$\le \|\theta x_1 - \theta x_c\|_2 + \|(1-\theta)x_2 - (1-\theta)x_c\|_2 \quad \text{triangle inequality}$$
$$= \theta\|x_1 - x_c\|_2 + (1-\theta)\|x_2 - x_c\|_2 \le r \quad \text{cool!!}$$

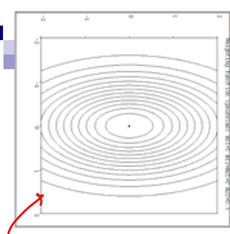
$\le r$ $\le r$

scaling of norms

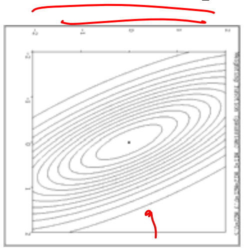
- Ellipsoid:
  - $(x-x_c)^T\Sigma^{-1}(x-x_c) \le 1$
  - Σ is positive semidefinite

convex.

Examples of Norm Balls



Scaled Euclidian (L_2)



Mahalanobis

$\|x\|$

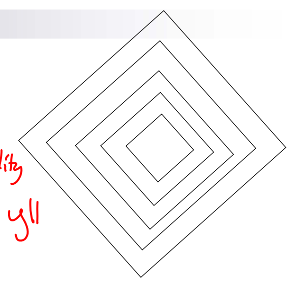
triangle inequality

$\|x+y\| \le \|x\| + \|y\|$

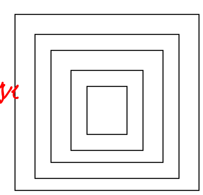
scaling

$\|\theta x\| = |\theta|\,\|x\|$

only need for positive θ



L_1 norm (absolute)



L_∞ *(max) norm*

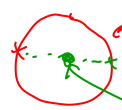
Norm balls

- Convexity of norm balls
  - Properties of norms:
    - Scaling
    - Triangle inequality
    - $\|0\|=0$

      $\|x\|=0 \Leftrightarrow x=0$
- Norm balls are extremely important in ML
  - e.g., SVMs: $\|w\|_2 \le 1$

    L1 norms (Lasso) $\|w\|_1 \le 1$
- What about achieving a norm with equality? ... not convex!!

  $\|w\|_2 = 1$

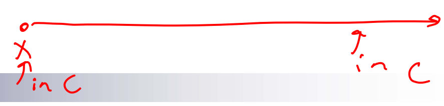


only on shell

not on shell

Cones

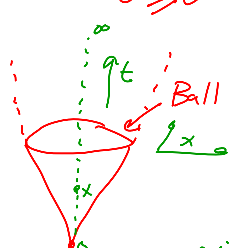
non convex cone



- Set C is a cone if set is invariant to non-negative scaling

  in a cone C, if $x \in C$

  $\Rightarrow \theta x \in C \quad \forall \theta \geq 0$
- If the cone is convex, we call it: Convex Cone
  - extremely important in ML (as we'll see)
- A cool cone: The ice cream cone
  - a.k.a. second order cone



$\{(x,t) \mid \|x\|_2 \le t\}$

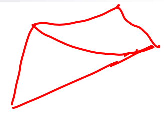
$x^T x \le t^2, \; t \ge 0$

also a second order cone if

$x^T \Sigma^{-1} x \le t^2, \quad \Sigma \succ 0$

positive semidefinite

Positive semidefinite cone



- Positive semidefinite matrices: $\Sigma^{n} \in S_+^n$ $\quad \Sigma \succeq 0$
  $\Rightarrow$ symmetric : $\Sigma = \Sigma^T$
  $\forall x \quad x^T \Sigma x \geq 0$
  positive definite: $x^T \Sigma x > 0$, $\Sigma \succ 0$, $\Sigma \in S_{++}^n$
- Positive semidefinite cone:
  $\Sigma = \Sigma^T$, $\Sigma \in \mathbb{R}^{n \times n}$, $\forall x, x \neq 0$
  $\{\Sigma \mid \forall x \; x^T \Sigma x \geq 0\}$
  cone: $\forall \theta \geq 0 \quad \theta\Sigma \in C$
  $x^T \Sigma x \geq 0 \Rightarrow x^T \theta \Sigma x = \theta x^T \Sigma x \geq 0$

QuickTime™ and a TIFF (LZW) decompressor are needed to see this picture.

- Alternate definition: Eigenvalues
  $\Sigma \succeq 0 \Rightarrow \forall i, \lambda_i \geq 0$, λ_i: eigenvalues, $\Sigma \succ 0 \Rightarrow \Sigma^{-1} \succ 0$
- Convexity:
  $\Sigma_1 \in C, \Sigma_2 \in C \Rightarrow \theta\Sigma_1 + (1-\theta)\Sigma_2 \in C$, $\theta \in [0,1]$
  $x^T[\theta\Sigma_1 + (1-\theta)\Sigma_2]x = \underbrace{\theta x^T \Sigma_1 x}_{\geq 0} + \underbrace{(1-\theta) x^T \Sigma_2 x}_{\geq 0} \geq 0$ convex

$$X = \begin{bmatrix} x & y \\ y & z \end{bmatrix}$$

$X \succeq 0$
$xz - y^2 \geq 0$
$x \geq 0$
$z \geq 0$

- Examples in ML:
  - covariance matrix Σ
  - Kernel matrix
  $$K = \begin{pmatrix} K(x_1,x_1) & K(x_1,x_2) & \cdots \\ K(x_2,x_1) & \ddots & \\ \vdots & & \ddots \end{pmatrix}$$
- A fundamental convex set
  - Useful in a huge number of applications
  - Basis for very cool approximation algorithms
  - Generalizes pretty many "named" convex optimization problems

clustering
bounds

Operations that preserve convexity 1: Intersection

- Intersection of convex sets is convex
  C_i are convex $\Rightarrow \bigcap_i C_i$ convex
- Examples:
  - Polyhedron $Ax \geq b$

intersection of half spaces $\Rightarrow$ convex

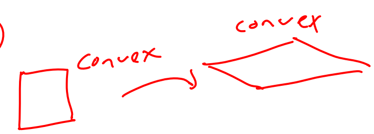
  - Robust linear regression
    $\varepsilon \geq \sum_i \alpha_i \left(t_i - \sum_j w_j f_j(x_i)\right)^2 \quad \forall w$
    Convex in ε, α, because intersection of half spaces, infinitely many
  - Positive semidefinite cone
    $\{\Sigma \mid \forall x \; x^T \Sigma x \geq 0, \; \Sigma = \Sigma^T\}$
    for each x, linear function of Σ, i.e., a half space
    $\Sigma = \Sigma^T$ linear subspace $\sigma_{ij} = \sigma_{ji}$
    intersection of half spaces with a linear subspace

Operations that preserve convexity 2: Affine functions

any A any b

decreasing dimensionality is OK

- Affine function: $f(x) = Ax + b$
- Set S is convex
  - Image of S under f is convex $\{f(x) \mid x \in S\} = \{Ax+b \mid x \in S\}$
- Translation: $x \in S$ is convex, $x + \delta$ convex
- Scaling: $x \in S$, ax is convex
- General affine transformation: $Ax + b$



- Why is ellipsoid convex? S convex $\{y \mid y^T y \le 1\}$
  - $(x-x_c)^T \Sigma^{-1} (x-x_c) \le 1$
  - Σ is positive semidefinite

$$y = \Sigma^{-1/2}(x - x_c) = \Sigma^{-1/2} x - \Sigma^{-1/2} x_c \quad \text{affine function}$$

$$\Sigma^{-1/2} : \Sigma^{-1} = (\Sigma^{-1/2})^T \Sigma^{-1/2} \quad \Rightarrow x = \Sigma^{1/2} y + x_c \quad \Rightarrow \text{ellipsoid convex}$$

Operations that preserve convexity 3: Linear-fractional functions

- Linear fractional functions:
  - $f(x) = \dfrac{Ax + b}{c^T x + d}$ $c^T x + d > 0$
  - Closely related to perspective projections (useful in computer vision)
- Given convex set C, image according to linear fractional function:

$$C \text{ convex} \Rightarrow \left\{ \frac{Ax+b}{c^T x + d} \;\middle|\; x \in C \right\} \text{ convex}$$

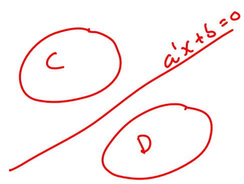
- Example: $P(X=i, Y=j) = p_{ij}$ $p_{ij} \ge 0$, $\sum_{ij} p_{ij} = 1$ convex set

$$P(X=i \mid Y=j) = \frac{p_{ij}}{\sum_i p_{ij}} \quad \text{convex set}$$

Separating hyperplane theorem

- **Theorem**: Every two non-intersecting convex sets *C* and *D* have a separating hyperplane:

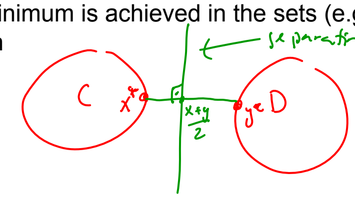
$C \cap D = \emptyset$



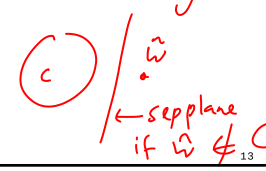
$\forall x \in C \quad a'x+b \geq 0$

$\forall x \in D \quad a'x+b \leq 0$

- Intuition of proof (for special case)
  - Minimum distance between sets: $d(C,D) = \min_{x,y} \|x-y\|_2$, $x \in C,\ y \in D$
  - If minimum is achieved in the sets (e.g., both sets closed, and one is bounded), then

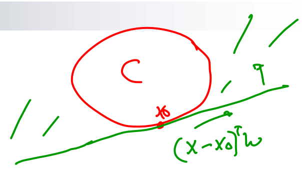


constraint generation



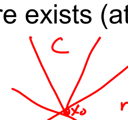
Supporting hyperplane

- General definition: Some set $C \subseteq R^n$
  - Point x_0 on boundary
    - Boundary is the closure of the set minus its interior
  - Supporting hyperplane:
    - Geometrically: a tangent at x_0
    - Half-space contains *C*:



$\forall y \in C, \quad (y-x_0)^T w \geq 0$

- **Theorem**: for any non-empty convex set *C*, and any point x_0 in the boundary of *C*, there exists (at least one) supporting hyperplane at x_0



many supporting hyperplanes

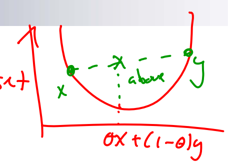
- (One) **Converse**: If set *C* is closed with non-empty interior, and there is a supporting hyperplane at every boundary point, then *C* is convex

What you need to know

- Definitions of convex sets
  - Main examples of convex sets
- Proving a set is convex
- Operations that preserve convexity
  - There are many many many other operations that preserve convexity
    - See book for several more examples
- Separating and supporting hyperplanes

Convex Functions

- Function $f:R^n \rightarrow R$ is convex if
  - Domain is convex   dom f convex set
  - $\forall x, y \in \text{dom} f$, $\theta \in [0,1]$



$$f(\theta x + (1-\theta) y) \leq \theta f(x) + (1-\theta) f(y)$$

- Generalization: Jensen's inequality:

prob dist x $\in$ convex domain of f

$$f[E[x]] \leq E_x[f(x)]$$

useful in ML
e.g., EM

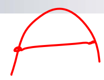
- Strictly convex function: e.g., $\cup$,   convex non-strict

$\forall x, y \in \text{dom} f$, $\theta \in (0,1)$
$x \neq y$

$$f(\theta x + (1-\theta) y) < \theta f(x) + (1-\theta) f(y)$$

Concave functions

- Function f is concave if
  - dom f is convex
  - −f is convex



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

- Strictly concave: −f is strictly convex

- We will be able to optimize: "easy"

$\min_x f(x)$ convex, $x \in C$ convex set

$\max_x f(x)$ concave, $x \in C$ convex set

Proving convexity for a very simple example

- $f(x)=x^2$

$$f(\theta x + (1-\theta) y) \overset{?}{\leq} \theta f(x) + (1-\theta) f(y)$$

$$(\theta x + (1-\theta) y)^2 \overset{?}{\leq} \theta x^2 + (1-\theta) y^2$$

$$\theta x (\theta x + (1-\theta) y) + (1-\theta) y (\theta x + (1-\theta) y)$$

+x−x

$$\theta x^2 + \theta x ((\theta - 1) x + (1-\theta) y) + (1-\theta) y^2 + (1-\theta) y (\theta x - \theta y)$$

$$\underbrace{\theta x^2 + (1-\theta) y^2}_{\theta f(x) + (1-\theta) f(y)} + \underbrace{\theta (1-\theta)}_{\geq 0} \underbrace{(x-y)(y-x)}_{\leq 0}$$

$\theta f(x) + (1-\theta) f(y) +$

done!! :)

boring!!

better way?

please...

First order condition

- If f is differentiable in all **dom** f
- Then f convex if and only if **dom** f is convex and

Second order condition (1D f)

- If f is twice differentiable in **dom** f
- Then f convex if and only if **dom** f is convex and
- Note 1: Strictly convex if:
- Note 2: **dom** f must be convex
  - $f(x)=1/x^2$
  - **dom** f = $\{x \in R | x \neq 0\}$

Second order condition (general case)

- If f is twice differentiable in **dom** f

- Then f convex if and only if **dom** f is convex and

- Note 1: Strictly convex if:

Quadratic programming

- $f(x) = (1/2)\, x^T A x + b^T x + c$

- Convex if:
- Strictly convex if:

- Concave if:
- Strictly concave if:

Simple examples

- Exponentiation: e^{ax}
  - convex on R, any $a \in R$
- Powers: x^a on R_{++}
  - Convex for $a \leq 0$ or $a \geq 1$
  - Concave for $0 \leq a \leq 1$
- Logarithm: log x
  - Concave on R_{++}
- Entropy: -x log x
  - Concave on R_+
  - (0 log 0 = 0)

A few important examples for ML

- Every norm on R^n is convex
- Log-sum-exp:
  - Convex in R^n
- Log-det:
  - Convex in S^n_{++}

Extended-value extensions

- Convex function f over convex **dom** f
- Extended-value extension:
- Still convex:
- Very nice for notation, e.g.,
  - Minimization:
  - Sum:
    - f_1 over convex **dom** f_1
    - f_2 over convex **dom** f_2

Epigraph

- Graph of a function $f: R^n \rightarrow R$
  - $\{(x,t) \mid x \in$ **dom** $f, f(x)=t\}$
- Epigraph:
  - **epi** f =
- **Theorem**: f is convex if and only if

Support of a convex set and epigraph

- If f is convex & differentiable
  - $f(x) \geq f(x_0) + \nabla f(x_0)(x - x_0)$

- For $(x,t) \in$ **epi** f, $t \geq f(x)$, thus:
  -

- Rewriting:

$$(x,t) \in \mathbf{epi} f \Rightarrow \begin{bmatrix} \nabla f(x_0) \\ -1 \end{bmatrix}^T \left(\begin{bmatrix} x \\ t \end{bmatrix} - \begin{bmatrix} x_0 \\ f(x_0) \end{bmatrix} \right) \leq 0$$

- Thus, if convex set is defined by epigraph of convex function
  - Obtain support of set by gradient!!
  - If f is not differentiable