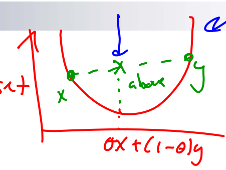
Convex Functions (cont. 2)

Optimization - 10725
Carlos Guestrin
Carnegie Mellon University

March 5th, 2008

Convex Functions



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

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

- 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., ∪ ,   convex non-strict

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

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

Operations that preserve convexity

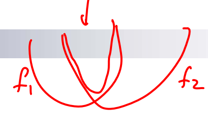
- Many operations preserve convexity
  - Knowing them will make your life much easier when you want to show that something is convex
  - Examples in next few slides
- Simplest: Non-negative weighted sum:
  - $f = \sum_i w_i f_i \quad w_i \geq 0$
  - If all f_i's are convex, then f is convex
  - If all f_i's are concave, then f is concave
  - Example: integral of f(x,y) convex x, y , $g(x) = \int_{y \in C} f(x,y)\, dy$ → convex
- Affine mapping: $f: R^n \rightarrow R$, $A \in R^{n \times m}$, $b \in R^m$ e.g., A, b come from PCA
  - $g(x) = f(Ax+b)$
  - **dom** g = $\{x \mid Ax+b \in \text{dom} f\}$ ← always convex if dom f convex
  - If f is convex, then g is convex
  - If f is concave, then g is concave

Pointwise maximum and supremum

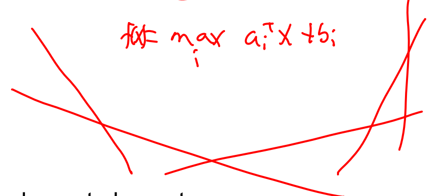
$f(X) = tr(Y^T X) = \sum_{ij} Y_{ij} X_{ij}$

$Xv = \lambda v$
$v^T X v = \lambda v^T v$
$\|v\|_2 = 1 \quad 1$

max f_1, f_2



- If f_i's are convex, then $f(x) = \max_i f_i(x)$
- Piecewise linear convex functions: $f(x) = \max_i a_i^T x + b_i$
  - Fundamental for POMDPs



- For x in a convex set C, sum of the r largest elements:
  - Sort x, pick r largest components, sum them: $f(x) = \{\sum_{i=1}^{r} x_{\ell_i} \mid x_{\ell_1} \geq x_{\ell_2} \geq \dots \geq x_{\ell_n}\}$

max of $\binom{n}{r}$ different functions $f(x) = \max_{A \subseteq \{1,\dots,n\}, |A| = r} x_{A_1} + x_{A_2} + \dots + x_{A_r}$

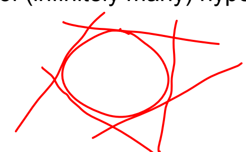
- Maximum eigenvalue of symmetric matrix $X \in R^{n \times n}$, $f: R^{n \times n} \rightarrow R$
  - $f(X) = \max_{v, \|v\|_2 = 1} v^T X v = \sum_{ij} X_{ij} v_i v_j$ linear in X

λ_{max} ← max of linear functions of X

Pointwise maximum of affine functions: general representation

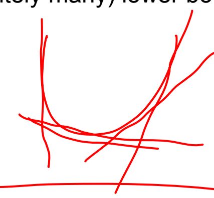
- We saw: convex set can be written as intersection of (infinitely many) hyperplanes:
  - C convex, then

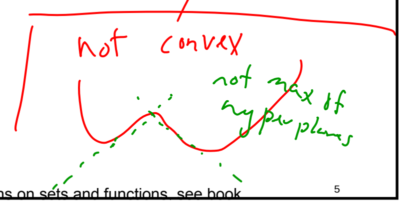
$$C = \bigcap_i \left(a_i^T x + b_i \geq 0\right)$$



- Convex functions can be written as supremum of (infinitely many) lower bounding hyperplanes:
  - f convex function, then

$$f(x) = \max_i a_i^T x + b_i$$





Discussion on this slide subject to mild conditions on sets and functions, see book

Composition: scalar differentiable, real domain case

$f'(x) = h'(g(x))\, g'(x)$

- How do I prove convexity of log-sum-exp-positive-weighted-sum-monomials? :)
  - $f(x) = \log \sum_i e^{\sum_j c_{ij} x^{a_{ij}}}$ $c_{ij} \geq 0$, $a_{ij} \geq 1$
- If $h:R^k \rightarrow R$ and $g:R^n \rightarrow R^k$, when is f(x) = h(g(x)) convex (concave)?
  - **dom** f = {x ∈ **dom** g | g(x) ∈ **dom** h}
- Simple case: $h:R \rightarrow R$ and $g:R^n \rightarrow R$, **dom** g = **dom** h = R, g and h differentiable
  - E.g., $g(x)=x^T\Sigma x$, Σ psd, $h(y) = e^y$ $f(x) = e^{x^T \Sigma x}$ ← is this convex?
- Second derivative:
  - $f''(x) = h''(g(x))g'(x)^2 + h'(g(x))g''(x)$
    - When is $f''(x)\geq 0$ (or $f''(x)\leq 0$) for all x?

when is $f''(x) \geq 0 \ \forall x$
e.g., $h''(g(x)) \geq 0$, h is convex
e.g., $h'(g(x)) \cdot g''(x)$, if g convex, $h'(g(x)) \geq 0$
h is ~~increasing~~ nondecreasing

- Example of sufficient (but not necessary) conditions:
  - f convex if h is convex and nondecreasing and g is convex
  - f convex if h is convex and nonincreasing and g is concave
  - f concave if h is concave and nondecreasing and g is concave
  - f concave if h is concave and nonincreasing and g is convex

Composition: scalar, general case

- If $h:R^k \rightarrow R$ and $g:R^n \rightarrow R^k$, when is $f(x) = h(g(x))$ convex (concave)?
  - **dom** f = {x ∈ **dom** g| g(x) ∈ **dom** h}

not diff

dom is not R

$\subset R$

- Simple case: $h:R \rightarrow R$ and $g:R^n \rightarrow R$, general domain and non-differentiable
  - Example of sufficient (but not necessary) conditions:
    - f convex if h is convex and $\tilde{h}$ nondecreasing and g is convex
    - f convex if h is convex and $\tilde{h}$ nonincreasing and g is concave
    - f concave if h is concave and $\tilde{h}$ nondecreasing and g is concave
    - f concave if h is concave and $\tilde{h}$ nonincreasing and g is convex

$$\tilde{h}(x) = \begin{cases} h(x) & \text{if } x \in \text{dom } h \\ \infty & \text{if } x \notin \text{dom } h \end{cases}$$

convex case

- nondecreasing or nonincreasing condition on extend value extension of h is fundamental
  - counter example in the book if nondecreasing property holds for h but not for $\tilde{h}$, the composition no longer convex
  - If $h(x)=x^{3/2}$ with **dom** h = R_+, convex but extension is not nondecreasing

$$\tilde{h}(x) = \begin{cases} x^{3/2} & , x \geq 0 \\ \infty & , x < 0 \end{cases}$$

  - If $h(x)=x^{3/2}$ for $x \geq 0$, and $h(x)=0$ for $x<0$, **dom** h = R, convex and extension is nondecreasing

Vector composition: differentiable

- If $h:R^k \rightarrow R$ and $g:R^n \rightarrow R^k$, when is $f(x) = h(g(x))$ convex (concave)?
  - **dom** f = {x ∈ **dom** g| g(x) ∈ **dom** h}
- Focus on $f(x) = h(g(x)) = h(g_1(x), g_2(x),\ldots, g_k(x))$
- Second derivative:
  - $f''(x) = g'(x)^T \nabla^2 h(g(x))g'(x) + \nabla h(g(x))\, g''(x)$
    - When is $f''(x) \geq 0$ (or $f''(x) \leq 0$) for all x?

$f''(x) \geq 0$ convex

e.g., if h convex ≥ 0, because $\nabla^2 h(x) \succeq 0$ $\forall x$

- Example of sufficient (but not necessary) conditions:
  - f convex if h is convex and nondecreasing in each argument, and g_i are convex
  - f convex if h is convex and nonincreasing in each argument, and g_i are concave
  - f concave if h is concave and nondecreasing in each argument, and g_i are concave
  - f concave if h is concave and nonincreasing in each argument, and g_i are convex
- Back to log-sum-exp-positive-weighted-sum-monomials
  - $f(x) = \log \sum e^{\sum_j c_{ij} x^{a_{ij}}}$
  - **dom** f = R^n_{++}, $c_{ij}>0$, $a_{ij} \geq 1$
  - log sum exp convex $h(x) = \log \sum_i e^{x_i}$ nondecreasing in each argument! :)

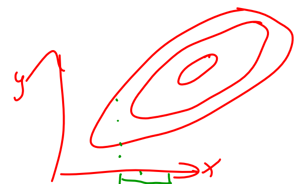
$x^{a_{ij}}$, $a_{ij} \geq 1$ convex

$g_i(x) = \sum_j c_{ij} x^{a_{ij}}$ convex (positive sum of convex fns)

Minimization

- If f(x,y) is convex in (x,y) and C is a convex set, then:

$g(y) = \min_{x \in C} f(x,y)$



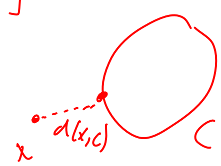
- Norm is convex: ||x-y||

$(x-y) = A\begin{bmatrix} x \\ y \end{bmatrix}$

  - minimum distance to a set C is convex:

$d(x, C) = \min_{y \in C} \|x-y\|$

convex



Perspective function

$KL(p\|q) = \sum_x p(x) \log p(x) - \sum_x p(x) \log q(x)$

- If f is convex (concave), then the perspective of f is convex (concave):
  - t>0, g(x,t) = t f(x/t)

Computer vision

f convex in x, g is convex in both x, t

- KL divergence:
  - f(x) = -log x is convex
  - Take the perspective: $g(x,t) = -t \log \frac{x}{t} = -t \log x + t \log t$
  - Sum over many pairs (x_i,t_i) $h(X,T) = \sum_i g(x_i,t_i)$

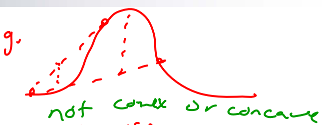
$= \sum_i t_i \log t_i - \sum_i t_i \log x_i$

KL divergence convex in p & q !!

Quasiconvex functions

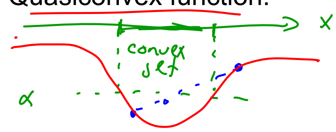
optimize: find smallest α s.t. $S_\alpha \neq \emptyset$

- Unimodal functions are not always convex , e.g.



- But they are (usually) still easy to optimize: Quasiconvex function:
  - All sublevel sets are convex, for all $\alpha \in R$:

$$S_\alpha = \{x \mid f(x) \le \alpha\}$$



- Equivalent definition: max of extremes is higher than function

convex:

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

quasi convex

$$f(\theta x + (1-\theta)y) \le \max(f(x), f(y))$$

- Applications include computer vision (geometric reconstruction) [Ke & Kanade '05]

Log-convex functions

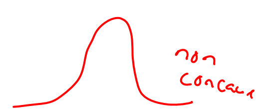
- Function $f: R^n \rightarrow R$, with $f(x)>0$ in all (convex) **dom** f
  - f log-convex if and only if:

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

- Or equivalently:

$$f(\theta x + (1-\theta)y) \le f(x)^{\theta} f(y)^{(1-\theta)}$$

- Examples
  - Gaussian $f(x) = \frac{1}{\sqrt{(2\pi)^n det\Sigma}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$



$$\log f(x) = \underbrace{\log \frac{1}{\sqrt{(2\pi)^n|\Sigma|}}}_{\text{doesn't depend on } x} \underbrace{- \frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}_{\text{concave}}$$

log-concave in x not μ, Σ

What should you know: Convex fns

- definition
- showing that a function is convex/concave
  - first principle
  - first and second order condition
  - epigraph
  - operations that preserve convexity
- quasiconvexity
- log-convexity