Convex Functions (cont.)

Optimization - 10725

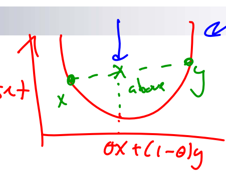
Carlos Guestrin

Carnegie Mellon University

March 3rd, 2008

Convex Functions

- Function $f: R^n \rightarrow R$ 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 ∈ 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

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

$x \neq y$

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

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"

$$\min_x f(x) \text{ convex}, \quad x \in C \text{ convex set}$$

$$\max_x f(x) \text{ concave}, \quad x \in C \text{ convex set}$$

Proving convexity for a very simple example

- $f(x)=x^2$

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

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

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

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

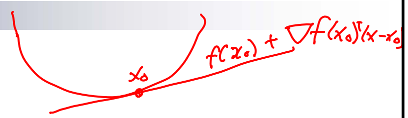
$$\underbrace{\theta x^2 + (1-\theta)y^2}_{\theta f(x) + (1-\theta)f(y)} + \underbrace{\theta(1-\theta)}_{\geq 0}\underbrace{(x-y)(y-x)}_{\leq 0}$$

done!!

boring!!

better way? Please...

First order condition (differentiable)



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

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

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

$\forall x \in dom f$
$\forall x_0 \in dom f$

$$f(x) \geq f(x_0) + \nabla f(x_0)^T (x - x_0)$$

$f(x) = x^2$ $\quad x^2 \geq x_0^2 + 2x_0(x - x_0)$

$$x^2 \geq 2x_0 x - x_0^2$$

$$(x - x_0)^2 \geq 0 \qquad \text{success!!}$$

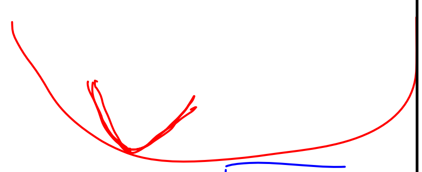
Second order condition (1D f)

for this rule, must be twice diff in entire dom f

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

$f(x) = x^2$

$f''(x) = 2 \geq 0 \quad$ success!!



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

$$f''(x) = \frac{\partial^2 f}{\partial x^2}(x) \geq 0 \quad \forall x \in dom f$$

not enough to check second derivative, also dom f convex!

- Note 1: Strictly convex if: $f''(x) > 0 \ \forall x \in dom f$
- Note 2: **dom** f must be convex
  - f(x)=1/x[2]
  - **dom** f = {x∈R|x≠0}

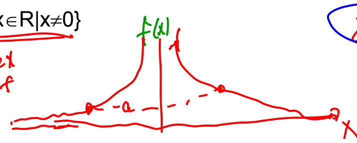
$f'(x) = -2\frac{1}{x^3}$; $f''(x) = 6\frac{1}{x^4} \geq 0$

$\forall x \in \mathbb{R}$ except for $\{0\}$

$dom f = \mathbb{R} - \{0\}$

not convex

not convex in dom f



Second order condition (general case)

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

$$\nabla^2 f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1^2} & \frac{\partial f}{\partial x_1 \partial x_2} & \cdots \\ \frac{\partial f}{\partial x_2 \partial x_1} & \frac{\partial f}{\partial x_2^2} & \cdots \end{pmatrix}$$

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

positive semi definite

$\forall x \quad \nabla f(x) \succeq 0$ convex

$\nabla^2 f(x) \preceq 0$ concave

- Note 1: Strictly convex if:

$\nabla^2 f(x) \succ 0$

positive definite

Quadratic programming

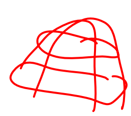
- $f(x) = (1/2)\, x^T A x + b^T x + c$

$\nabla f(x) = Ax + b$

$\nabla^2 f(x) = A$

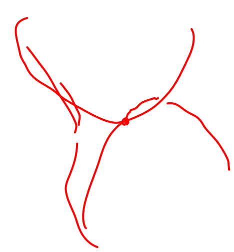
- Convex if: $A \succeq 0$
- Strictly convex if: $A \succ 0$

- Concave if: $A \preceq 0$
- Strictly concave if: $A \prec 0$



neither concave nor convex

$A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$



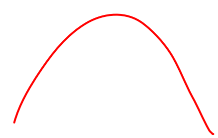
Simple examples

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

$f'(x) = -\log x - x \cdot \frac{1}{x}$

$f''(x) = \frac{-1}{x} \le 0 \quad \forall x \text{ dom} f$

Concave



A few important examples for ML

- Every norm on R^n is convex

  $\|x\|$

  regularization
- Log-sum-exp: $f(x) = \log \sum_i e^{a_i x_i}$
  - Convex in R^n

    maxent

    Logistic Regression

    partition function
- Log-det: $\log \det X = \log |X|$
  - ~~Convex~~ concave in S^n_{++}

p Gaussian with Covariance $X = \Sigma$: $H(p) = c + \frac{1}{2} \log \det X$

Extended-value extensions

- Convex function f over convex **dom** f

$-\log x$ convex in $\mathbb{R}_{++}$

- Extended-value extension:

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

$\mathbf{dom}\, \tilde{f} = \mathbb{R}^n$

- Still convex:

$f(x)$ convex $\Rightarrow \tilde{f}(x)$ convex ; $\tilde{f}(\theta x + (1-\theta) y) \leq \theta \tilde{f}(x) + (1-\theta)\tilde{f}(y)$

left side cannot if ∞ & right side not infinite, because dom f convex

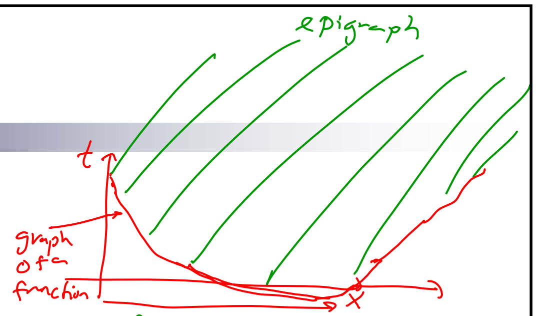
- For concave functions

$$\tilde{f}(x) = \begin{cases} f(x), & x \in \mathbf{dom} f \\ -\infty, & x \notin \mathbf{dom} f \end{cases}$$

- Very nice for notation, e.g.,
  - Minimization:
    $\min_{x \in C} f(x)$; $\min \tilde{f}(x)$
  - Sum: $g = f_1 + f_2$ $\quad \mathbf{dom}\, g = \mathbf{dom} f_1 \cap \mathbf{dom} f_2$
    - f_1 over convex **dom** f_1
    - f_2 over convex **dom** f_2

$\tilde{g} = \tilde{f}_1 + \tilde{f}_2$ automatically intersects domain

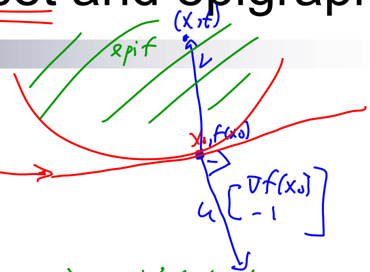
Epigraph



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

Support of a convex set and epigraph

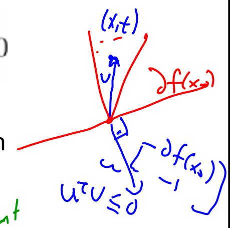
- If f is convex & differentiable
  - $f(x) \geq f(x_0) + \nabla f(x_0)^T(x - x_0)$



- For (x,t)∈**epi** f, t≥f(x), thus:
  - $t \geq f(x_0) + \nabla f(x_0)^T(x - x_0) \quad \forall x, x_0, t$
- Rewriting:

$u^T v \leq 0$

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

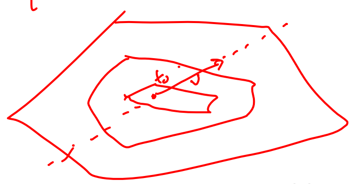


- Thus, if convex set is defined by epigraph of convex function
  - Obtain support of set by gradient!! $\begin{bmatrix}\nabla f(x_0) \\ -1\end{bmatrix}$
  - If f is not differentiable ? yes with sub gradient

Michelle Martin is new class admin & homeworks from me
Wean 4619

Restriction of a convex function to a line

- $f: \mathbb{R}^n \rightarrow \mathbb{R}$ convex if and only if $g(t)$ ($\mathbb{R} \rightarrow \mathbb{R}$) is convex in t
  - For all $x_0 \in$ **dom** f, $v \in \mathbb{R}^n$
  - $g(t) = f(x_0 + tv) \quad \forall v \in \mathbb{R}^n$
  - **dom** g = $\{t \mid x_0 + tv \in \text{dom } f\}$



$f(x) = x^T A x + x^T b + c$

arbitrary vector v

$g(t) = (x + tv)^T A (x + tv) + (x + tv)^T b + c$

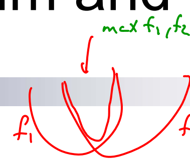
$g''(t) = 2v^T A v \quad \forall v \quad \geq 0$ if $A \succeq 0$

- Can make it much easier to check if f is convex, e.g.,
  - f(X) = log det X   concave
  - proof in the book…
  - $g(t) = f(X + tV)$ V is arbitrary symmetric matrix

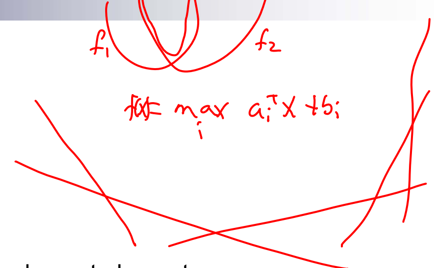
Operations that preserve convexity

- 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



- 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: $f(x) = \{\sum_{i=1}^{r} x_{l_i} \mid x_{l_1} \geq x_{l_2} \geq \dots \geq x_{l_n}\}$

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 $X \in R^{n \times n}$, $f:R^{n \times n} \rightarrow R$
  - $f(X) = \max_{v,\ \|v\|_2 = 1} v^T X v$ ← 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
- Convex functions can be written as supremum of (infinitely many) lower bounding hyperplanes:
  - f convex function, then

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

Composition: scalar differentiable, real domain case

- How do I prove convexity of log-sum-exp-positive-weighted-sum-monomials? :)
  -
- 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 $\in$ **dom** g| g(x) $\in$ **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, $h(y) = e^y$
- 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?
- 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}
- 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
- 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
  - 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?
- 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
  -
  - **dom** f = R^n_{++}, $c_i>0$, $a_i\geq 1$
  - log sum exp convex

Minimization

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

- Norm is convex: ||x-y||
  - minimum distance to a set C is convex:

Perspective function

- If f is convex (concave), then the perspective of f is convex (concave):
  - t>0, g(x,t) = t f(x/t)
- KL divergence:
  - f(x) = -log x is convex
  - Take the perspective:
  - Sum over many pairs (x_i,t_i)