Mathematical methods in communication July 5th, 2009

Appendix: Convex functions

Lecturer: Haim Permuter *Scribe: Koby Todros and Assaf Levanon*

I. Notation

- $\mathbb{R}$: The set of real numbers.
- $\mathbb{R}_+$: The set of nonnegative real numbers.
- $\mathbb{R}_{++}$: The set of positive real numbers.
- $\mathbb{S}^k$: The set of symmetric $k \times k$ matrices.
- $\mathbb{S}_+^k$: The set of symmetric positive semi-definite $k \times k$ matrices.
- $\mathbb{S}_{++}^k$: The set of symmetric positive definite $k \times k$ matrices.
- $\text{dom}f$: The domain of the function f. Let $f : \mathbb{R}^n \to \mathbb{R}^m$, then $\text{dom}f \triangleq \{x \in \mathbb{R}^n : f(x) \text{ exists}\}$. For example, $\text{dom}\log = \mathbb{R}_{++}$

II. Definitions

Definition 1 (Convex set.) A set $C \in \mathbb{R}^n$ is convex if the line segment between any two points in C lies in C, i.e. $\forall x_1, x_2 \in C$ and any $0 \leq \theta \leq 1$ we have $\theta x_1 + (1-\theta) x_2 \in C$.

Definition 2 (Convex function.) $f : \mathbb{R}^n \to \mathbb{R}$ is convex if $\text{dom}f$ is a convex set and if $\forall x, y \in \text{dom}f$ and any $0 \leq \theta \leq 1$

$$f(\theta x + (1-\theta) y) \leq \theta f(x) + (1-\theta) f(y). \tag{1}$$

Geometrically, this means that the line segment between $(x, f(x))$ and $(y, f(y))$ lies above the graph of f. An illustration of convex function is given in Fig. 1.

Definition 3 (Strictly convex function.) $f : \mathbb{R}^n \to \mathbb{R}$ is strictly convex if $\text{dom}f$ is a convex set and if $\forall x, y \in \text{dom}f$ and any $0 \leq \theta \leq 1$

$$f(\theta x + (1-\theta) y) < \theta f(x) + (1-\theta) f(y). \tag{2}$$

Definition 4 (Concave function.) $f : \mathbb{R}^n \to \mathbb{R}$ is concave if $-f$ is convex.

Definition 5 (Strictly concave function.) $f : \mathbb{R}^n \to \mathbb{R}$ is strictly concave if $-f$ is strictly convex.

Definition 6 (Sublevel set.) Let $f : \mathbb{R}^n \to \mathbb{R}$. The α-sublevel set of f is defined as

$$C_\alpha \triangleq \{x \in \text{dom}f : \ f(x) \leq \alpha\}. \tag{3}$$

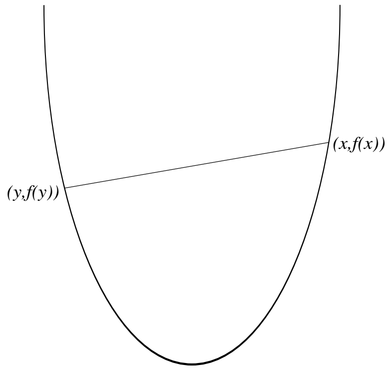

Fig. 1. Graph of a convex function. The chord between any two points on the graph lies above the graph.

Sublevel sets of convex functions are convex (converse is false).

Definition 7 (Epigraph.) Let $f : \mathbb{R}^n \to \mathbb{R}$. The epigraph of f is defined as

$$\mathrm{epi} f \triangleq \left\{(x, t) \in \mathbb{R}^{n+1} : x \in \mathrm{dom},\ f(x) \le t\right\}. \tag{4}$$

The function f is convex iff $\mathrm{epi} f$ is a convex set.

Definition 8 (Jensen's inequality.) Let $f : \mathbb{R}^n \to \mathbb{R}$, and let $z \in \mathbb{R}^n$ denote a random variable, such that $\Pr\{z \in \mathrm{dom} f\} = 1$. If f is convex, then

$$f(\mathbb{E}z) \le \mathbb{E}f(z). \tag{5}$$

Proof: Convexity of f implies that it is the upper envelope of the set of linear functions lying below it, i.e.,

$$f(z) = \sup_{L \in \mathcal{L}} L(z), \tag{6}$$

where

$$\mathcal{L} \triangleq \{L : L(z) = az + b,\ \forall \infty < z < \infty\}. \tag{7}$$

Thus,

$$\mathbb{E}[f(z)] = \mathbb{E}\left[\sup_{L \in \mathcal{L}} L(z)\right]. \tag{8}$$

Since $\sup_{L \in \mathcal{L}} L(z) \ge L(z)$, then by monotonicity of the expectation operator,

$$\mathbb{E}\left[\sup_{L \in \mathcal{L}} L(z)\right] \ge \mathbb{E}[L(z)]. \tag{9}$$

Taking supremum on both sides of (9) w.r.t. $L(z)$ implies that

$$\mathbb{E}\left[\sup_{L \in \mathcal{L}} L(z)\right] \ge \sup_{L \in \mathcal{L}} \mathbb{E}[L(z)]. \tag{10}$$

Therefore, according to (8) and (10)

$$\mathbb{E}\left[f\left(z\right)\right] \geq \sup_{L\in\mathcal{L}} \mathbb{E}\left[L\left(z\right)\right] \quad (11)$$

$$= \sup_{L\in\mathcal{L}} L\left(\mathbb{E}\left[z\right]\right) \quad (12)$$

$$= f\left(\mathbb{E}\left[z\right]\right), \quad (13)$$

where the equalities in (12) and (12) stem from the linearity of $L\left(\right)$ and (6), respectively. ■

The basic inequality in (1) is a special case of (5), whenever $z \in \{x, y\}$, $\Pr\left(z = x\right) = \theta$ and $\Pr\left(z = y\right) = 1 - \theta$.

III. Examples

A. Examples on $\mathbb{R}$

- Convex functions:
  - Affine: $f\left(x\right) = ax + b$ on $\mathbb{R}$, for any $a, b \in \mathbb{R}$.
  - Exponential: $f\left(x\right) = \exp\left(ax\right)$ on $\mathbb{R}$, for any $a \in \mathbb{R}$.
  - Powers: $f\left(x\right) = x^{\alpha}$ on $\mathbb{R}_{++}$, for $\alpha \geq 1$ or $\alpha \leq 0$.
  - Powers of absolute values: $|x|^p$ on $\mathbb{R}$, for $p \geq 1$.
  - Negative entropy: $x \log x$ on $\mathbb{R}_{++}$.
- Cocave functions
  - Affine: $f\left(x\right) = ax + b$ on $\mathbb{R}$, for any $a, b \in \mathbb{R}$.
  - Powers: $f\left(x\right) = x^{\alpha}$ on $\mathbb{R}_{++}$, for $0 \leq \alpha \leq 1$.
  - Logarithm: $\log x$ on $\mathbb{R}_{++}$.

B. Examples on $\mathbb{R}^n$ and $\mathbb{R}^{m\times n}$

Affine functions are both concave and convex. All norms are convex.

- Examples on $\mathbb{R}^n$
  - Affine: $f\left(x\right) = a^T x + b$, where $a, b, x \in \mathbb{R}^n$.
  - Norms:
    * l_p norm: $\|x\|_p = \left(\sum_{n=1}^{p} |x_i|^p\right)^{1/p}$, for $p \geq 1$.
    * l_∞ norm $\|x\|_\infty = \max_i |x_i|$.
- Examples on $\mathbb{R}^{m\times n}$
  - Affine function: $f\left(X\right) = \mathrm{tr}\left[A^T X\right] + b = \sum_{i=1}^{m}\sum_{i=1}^{n} A_{i,j}X_{i,j} + b$, where $\mathrm{tr}\left[\cdot\right]$ denotes the trace operator, $A, X \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}$.
  - Spectral norm: $f\left(X\right) = \|X\|_2 = \left(\lambda_{\max}\left(X^T X\right)\right)^{1/2}$, where $\lambda_{\max}\left(A\right)$ is the maximum eigenvalue of A, and $X \in \mathbb{R}^{m\times n}$.

IV. VERIFYING CONVEXITY OF A FUNCTION

Convexity of a function can be verified via the following manners:

- Using the definition of convex function (refer to definition 2).
- Applying some special criteria.
  - Restriction of a convex function to a line.
  - First order conditions.
  - Second order conditions.
- Showing that the function under inspection is obtained through operations that preserve convexity.

A. Restriction of a convex function to a line

The function $f : \mathbb{R}^n \to \mathbb{R}$ is convex iff the function $g : \mathbb{R} \to \mathbb{R}$,

$$g(t) = f(x + tv), \ \mathrm{dom} g = \{t \in \mathbb{R} : x + tv \in \mathrm{dom} f\}, \tag{14}$$

is convex in t for any $x \in \mathrm{dom} f$ and $v \in \mathbb{R}^n$. Therefore, checking convexity of multivariate functions can be carried out by checking convexity of univariate functions.

Example 1 Let $f : \mathbb{S}^n \to \mathbb{R}$ with

$$f(X) = -\log\det X, \ \mathrm{dom} f = \mathbb{S}^n_{++}. \tag{15}$$

Then

$$\begin{aligned} g(t) = -\log\det(X + tV) &= -\log\det X - \log\det\left(I + tX^{-1/2}VX^{-1/2}\right) \\ &= -\log\det X - \sum_{i=1}^{n} \log(1 + t\lambda_i), \end{aligned} \tag{16}$$

where I is the identity matrix and λ_i, $i = 1, \ldots, n$ are the eigenvalues of the matrix $X^{-1/2}VX^{-1/2}$. Since g is convex in t for any choice of V and any $X \in \mathrm{domf}$, then f is convex.

B. First order condition

Let $f : \mathbb{R}^n \to \mathbb{R}$ denote a differentiable function, i.e. $\mathrm{dom} f$ is open and $\forall x \in \mathrm{dom} f$ the gradient vector

$$\nabla f(x) \triangleq \left[\frac{\partial f(x)}{\partial x_1}, \ldots, \frac{\partial f(x)}{\partial x_n}\right]^T \tag{17}$$

exists. Then f is convex iff $\mathrm{dom} f$ is convex and $\forall x, y \in \mathrm{dom} f$

$$f(y) \geq f(x) + \nabla f(x)^T (y - x). \tag{18}$$

The r.h.s. of (18) is the first order taylor approximation of $f(y)$ in the vicinity of x. According to (18), the first order taylor approximation in case where f is convex, is a global underestimate of f. This is a very important property used in algorithm designs and performance analysis. The inequality in (18) is illustrated in Fig. 2.

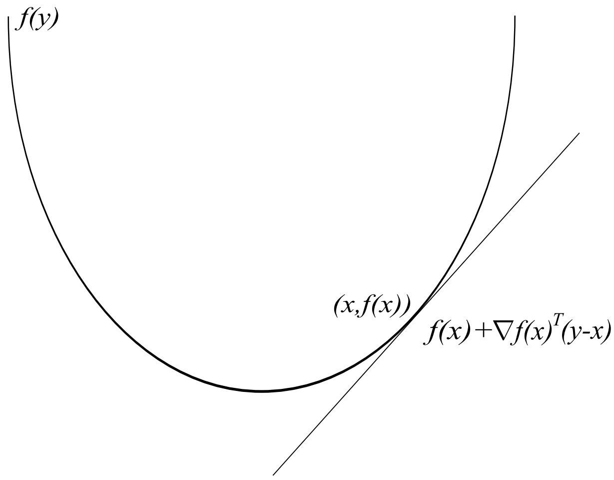

Fig. 2. If f is convex and differentiable, then $f(y) \geq f(x) + \nabla f(x)^T (y - x)$ for all $x, y \in \mathrm{dom} f$.

C. *Second order condition*

Let $f : \mathbb{R}^n \to \mathbb{R}$ denote a twice differentiable function, i.e. $\mathrm{dom} f$ is open and $\forall x \in \mathrm{dom} f$ the Hessian matrix, $\nabla^2 f(x) \in \mathbb{S}^n$,

$$\nabla^2 f(x)_{i,j} \triangleq \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \tag{19}$$

exists. Then f is convex iff $\mathrm{dom} f$ is convex and $\nabla^2 f(x) \succcurlyeq 0$, $\forall x \in \mathrm{dom} f$.

Examples for the use of the second order condition:

- Quadratic function: Let $f : \mathbb{R}^n \to \mathbb{R}$, such that

$$f(x) = \frac{1}{2} x^T P x + q^T x + r, \tag{20}$$

  $q, r \in \mathbb{R}^n$ and $P \in \mathbb{S}^n$. Since

$$\nabla f(x) = Px + q, \tag{21}$$

  then

$$\nabla^2 f(x) = P. \tag{22}$$

  Therefore, if $P \succcurlyeq 0$ then f is convex.

- Least-squares objective function: Let $f : \mathbb{R}^n \to \mathbb{R}$, such that

$$f(x) = \|Ax - b\|_2^2, \tag{23}$$

  $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m \times n}$. Since

$$\nabla f(x) = 2A^T (Ax - b), \tag{24}$$

then

$$\nabla^2 f(x) = 2A^T A. \tag{25}$$

Therefore, f is convex for any $A \in \mathbb{R}^{m\times n}$.

- Quadratic-over-linear function: Let $f : \mathbb{R}^2 \to \mathbb{R}$, such that

$$f(x, y) = \frac{x^2}{y}. \tag{26}$$

Then

$$\nabla^2 f(x, y) = \frac{2}{y^2}\begin{bmatrix} y & -x \\ -x & \frac{x^2}{y} \end{bmatrix}. \tag{27}$$

Therefore, f is convex for any $y > 0$.

V. Operations that preserve convexity

- Positive scaling
- Sum
- Composition with affine functions
- Pointwise maximum
- Pointwise supremum
- Composition with scalar functions
- Composition with vector functions
- Minimization
- Perspective

A. Positive scaling

Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex, then λf is convex $\forall \lambda > 0$.

B. Sum

Let $f_1, f_2 : \mathbb{R}^n \to \mathbb{R}$ be convex, then $f_1 + f_2$ is convex. This property can be extended to infinite sums and integrals.

C. Composition with affine function

Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex, and let $g : \mathbb{R}^m \to \mathbb{R}$ be affine, i.e. $g(x) = Ax + b$, where $x \in \mathbb{R}^m$, $b \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n\times m}$. The composition

$$(f \circ g)(x) = f(Ax + b) \tag{28}$$

is convex. For example, using the sum, and composition with affine function properties, along with the fact that $-\log(\cdot)$ is convex, it is concluded that

$$f(x) = -\sum_{i=1}^{n} \log\left(b_i - a_i^T x\right), \ \mathrm{dom} f = \left\{x : a_i^T x < b_i\right\}, \ i = 1, \ldots, n, \tag{29}$$

is convex. In addition, convexity of the norm implies that then any norm of affine function is convex, i.e.

$$f(x) = \|Ax + b\| \tag{30}$$

is convex.

D. Pointwise maximum

Let $f_1, \ldots, f_m : \mathbb{R}^n \to \mathbb{R}$ be convex. Then

$$F(x) = \max_{i=1,\ldots,m} \{f_1(x), \ldots, f_m(x)\}, \ \mathrm{dom} F = \bigcap_{i=1}^{m} \mathrm{dom} f_i, \tag{31}$$

is convex.

Examples:

- Piecewise-linear fucntion: $f(x) = \max_{i=1,\ldots,m} \left(a_i^T x + b_i\right)$ is convex.
- Sum of r largest components of a vector $x \in \mathbb{R}^n$:

$$f(x) = x_{[1]} + x_{[2]} + \ldots + x_{[r]} \tag{32}$$

  is convex, where $x_{[i]}$ is the i-th largest component of x. Proof:

$$f(x) = \max_{i_1,\ldots,i_r \in I_r} \{x_{i_1} + \ldots + x_{i_r}\}, \tag{33}$$

  where $I_r \triangleq \{(i_1, \ldots, i_r) : i_1 < \ldots, i_r, i_j \in \{1, \ldots, m\}, j = 1, \ldots, n\}$.

E. Pointwise supremum

Let $\mathcal{A} \subseteq \mathbb{R}^p$ and $f : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}$. Let $f(x, y)$ be convex in x for each $y \in \mathcal{A}$. Then the supremum function over the set $\mathcal{A}$ is convex, i.e.

$$g(x) = \sup_{y \in \mathcal{A}} f(x, y) \tag{34}$$

is convex.

Examples:

- Support function of a set C:

$$S_C(x) = \sup_{y \in C} y^T x. \tag{35}$$

- Distance to farthest point in a set C:

$$f(x) = \sup_{y \in C} \|x - y\|. \tag{36}$$

- Maximum eigenvalue of symmetric matrix: for $X \in \mathbb{S}^n$,

$$\lambda_{\max}(X) = \sup_{\|y\|_2=1} y^T X y. \tag{37}$$

F. Composition with scalar functions

Let $g : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R} \to \mathbb{R}$ with $\mathrm{dom} g = \mathbb{R}^n$ and $\mathrm{dom} h = \mathbb{R}$. Then

$$f(x) = h(g(x)) \tag{38}$$

is convex if

1) g is convex, h is nondecreasing and convex. For example, $\exp(g(x))$ is convex if g is convex.
2) g is concave, h is nonincreasing and convex. For example, $\frac{1}{g(x)}$ is convex if g is concave and positive.

G. Composition with vector functions

Let $g : \mathbb{R}^n \to \mathbb{R}^p$, $h : \mathbb{R}^p \to \mathbb{R}$ with $\mathrm{dom} g = \mathbb{R}^n$ and $\mathrm{dom} h = \mathbb{R}^p$. Then

$$f(x) = h(g(x)) = h(g_1(x), \ldots, g_p(x)) \tag{39}$$

is convex if

1) Each g_i is convex, h is nondecreasing and convex in each argument. For example, $\sum_{i=1}^{m} \exp(g_i(x))$ is convex if g_i, $i = 1, \ldots, m$, are convex.
2) Each g_i is concave, h is nonincreasing and convex in each argument. For example, $-\sum_{i=1}^{m} \log g_i(x)$ is convex if g_i, $i = 1, \ldots, m$, are concave and positive.

H. Minimization

Let $C \subseteq \mathbb{R}^n \times \mathbb{R}^p$ be a nonempty convex set, $f : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}$ be convex (in $(x, y \in \mathbb{R}^n \times \mathbb{R}^p)$). Then

$$g(x) = \inf_{y \in C} f(x, y) \tag{40}$$

is convex in x. For example, for a nonempty convex set, $C \subset \mathbb{R}^n$, since $f(x, y) = \|x - y\|$ is convex in (x, y) then

$$\inf_{y \in C} \|x - y\| \tag{41}$$

is convex in x.

I. Perspective

The perspective of a function $f:\mathbb{R}^n \to \mathbb{R}$ is the function $g:\mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$, such that

$$g\left(x,t\right) \triangleq tf\left(\frac{x}{t}\right),\ \mathrm{dom}g = \left\{(x,t) : \frac{x}{t} \in \mathrm{dom}f,\ t>0\right\}. \tag{42}$$

Hence, g is convex in (x,t) if f is convex.

Example 2 Since $f\left(x\right) = x^T x$ is convex, then $g\left(x,t\right) = \frac{x^T x}{t}$ is convex in (x,t).

VI. Convexity and information measures

In this section, the properties of convex functions, shown above are used for proving convexity/convavity of information measures such as entropy and relative entropy:

A. Concavity of entropy of discrete random variable

Let $p_X\left(x\right)$ denote probability mass functions of a discrete random variable X with alphabet $\mathcal{X}$. Let $f\left(p_X\left(x\right)\right) = p_X\left(x\right)\log p_X\left(x\right)$. Since $p_X \geq 0$, then $\frac{d^2 f}{dp_X^2} \geq 0$. Hence by the second-order condition for verification of convexity it is implied that $f\left(p_X\right)$ is convex in p_X. Now, since convexity is sum invariant, then the negative entropy, $-H\left(p\right) \triangleq \sum\limits_x p_X\left(x\right)\log p_X\left(x\right)$, is convex in p_X. Thus, $H\left(p_X\right)$ is concave in p_X.

B. Convexity of relative entropy

Let $p_X\left(x\right), q_X\left(x\right)$ denote probability mass functions of a random variable X with alphabet $\mathcal{X}$. The negative logarithm, $f\left(p_X\left(x\right)\right) = -\log p_X\left(x\right)$ is convex. Hence, the perspective function $g\left(p_X\left(x\right), q_X\left(x\right)\right) = q_X\left(x\right)\log\frac{q_X(x)}{p_X(x)}$ is convex in (p,q). Since convexity is sum invariant, then the relative entropy $D\left(q||p\right) \triangleq \sum\limits_{x\in\mathcal{X}} q_X\left(x\right)\log\frac{q_X(x)}{p_X(x)}$ is convex in (p,q).

References

[1] S. Boyd, *Convex Optimization.* Cambridge University Press , 2004.

[2] J.M. Steele, *Stochastic Calculus and Financial Applications.* Springer, 2001.