10-725/36-725: Convex Optimization **Spring 2015**

Lecture 2: January 14

Lecturer: Ryan Tibshirani *Scribes: Qi Guo, Sangwon Hyun (Justin)*

2.1 Outline

Last time, we discussed optimization problems and why convexity is our friend. In short, it is because convex optimization problems are definitely solvable with known methods, among other desirable properties. It also sometimes provides insight about the statistical properties of the problem in hand. Today, we cover: Convex sets, convexity-preserving operations, convex functions and examples.

2.2 Convex Sets

1. Definition

   A **convex set** is defined as $C \subseteq \mathbb{R}^n$ such that $x, y \in C \implies tx + (1-t)y \in C$ for all $0 \le t \le 1$. In other words, a line segment joining any two elements lies entirely in the set. Informally, every point in a convex set can 'see' every other point in the set.

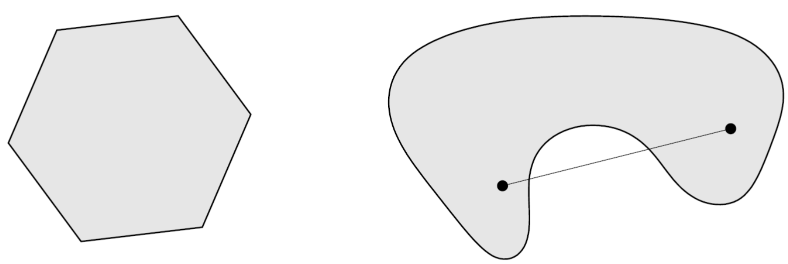

Figure 2.1: Examples of convex sets

A **convex combination** of $x_1, \cdots, x_k \in \mathbb{R}^n$ is any linear combination:

$$\sum_{i=1}^{k} \theta_i x_i = \theta_1 x_1 + \cdots + \theta_k x_k$$

with $\theta_i \ge 0, i = 1, \cdots, k$ and $\sum_{i=1}^{k} \theta_1 = 1$.

A **convex hull** of a set C is the set of all convex combination of its elements. A convex hull is always convex, and any convex combination of points in conv(C) is also

$$\text{conv}(C) = \{\sum_{i=1}^{k} \theta_i x_i : k \in \{1, 2, \cdots\},\ \theta_i \ge 0,\ \sum_{i=1}^{k} \theta_i = 1,\ x_i \in C\}$$

2. Examples
   Some examples are:

   - **Norm ball:** $\{x : \|x\| \le r\}$ for a given norm $\|\cdot\|$, radius r.
   - **Hyperplane:** $\{x : a^T x = b\}$ for given vector a, b.
   - **Halfspace:** $\{x : a^T x \le b\}$ for given vector a, b.
   - **Affine Space:** $\{x : Ax = b\}$ for a given matrix A and vector b
   - **Polyhedron:** $\{x : Ax \le b\}$ for matrix A and vector b. You can visualize every row of A as a normal vector for each hyperplane involved! Also, $\{x : Ax \le b, Cx = d\}$ is also a polyhedron because the equality $Cx = d$ can be made into two inequalities $Cx \ge d$ and $Cx \le d$.

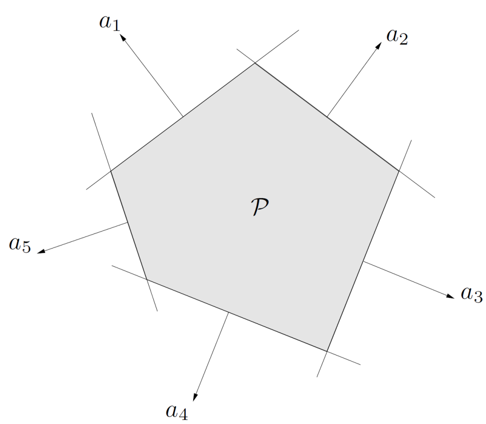

Figure 2.2: Polyhedron, with each row of A equals to $a_1, \cdots, a_n$.

   - **Simplex:** is a special case of polyhedra, given by the convex hull of a set of affinely independent points $x_0, \cdots, x_k$ (i.e. $\text{conv}\{x_0, \cdots x_k\}$). Affinely independent means that $x_1 - x_0, \cdots, x_k - x_0$ are linearly independent. A canonical example is the probability simplex

$$\text{conv}\{e_1, \cdots, e_n\} = \{\omega : \omega \ge 0, 1^T \omega = 1\}$$

   Tip: these are all easy to think about in $\mathbb{R}^2$! and after $\mathbb{R}^3$, the geometrical intuition is all similar!

   - **Convex Cones**:
     A cone is $C \in \mathbb{R}^n$ such that

$$x \in C \Longrightarrow tx \in C \text{ for all } t \ge 0$$

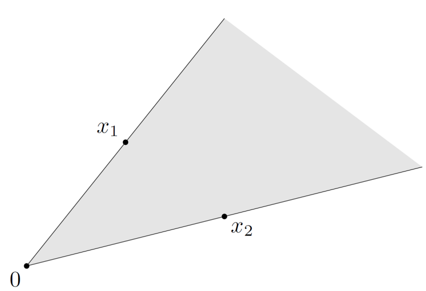

Figure 2.3: Example of cone.

A convex cone is a cone that is also convex i.e.,

$$x_1, x_2 \in C \Longrightarrow t_1x_1 + t_2x_2 \in C \text{ for all } t_1, t_2 \geq 0$$

A conic combination of points $x_1, \cdots, x_k \in \mathbb{R}^n$ is, for any $\theta_i \geq 0, i = 1, \cdots, k$, any linear combination

$$\theta_1 x_1 + \cdots + \theta_k x_k$$

A conic hull collects all conic combinations of $x_1, \cdots, x_k$ (or a general set C

$$\text{conic}(\{x_1, \cdots, x_k\}) = \{\theta_1 x_1 + \cdots + \theta_k x_k,\ \theta_i \geq 0,\ i = 1, \cdots, k\}$$

Some examples of convex cones are of special interest, because they appear frequently.

- **Norm Cone** A norm cone is $\{(x, t) : \|x\| \leq t\}$. Under the ℓ_2 norm $\|\cdot\|_2$, this is called a second-order cone.

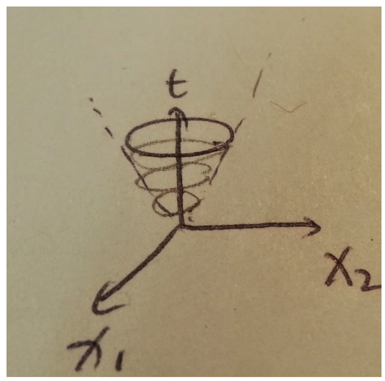

Figure 2.4: Example of second order cone.

- **Normal Cone** Given set C and point $x \in C$, a normal cone is

$$\mathcal{N}_C(x) = \{g : g^T x \geq g^T y, \text{ for all } y \in C\}$$

In the figure, for the corner point x depicted here, the points g in the shaded region have larger dot product with x than with any other point.

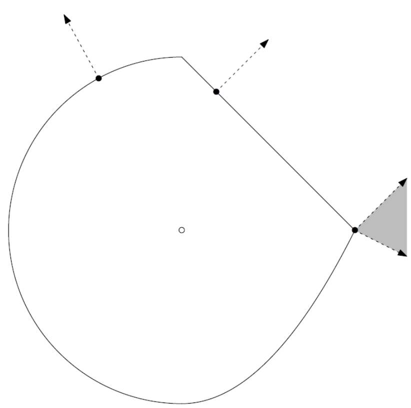

Figure 2.5: Normal Cone

  - **PSD cone**: A positive semidefinite cone is the set of positive definite symmetric matrices. ($\mathbb{S}^n$ are $n \times n$ symmetric matrices)

$$\mathbb{S}^n_+ = \{X \in \mathbb{S}^n : X \succeq 0\}$$

  You can verify that all of these are convex sets.

3. Properties of Convex Sets
   We now cover some useful properties of Convex Sets. The first one is the **separating hyperplane theorem**, which states that two convex sets have a separating hyperplane between them:

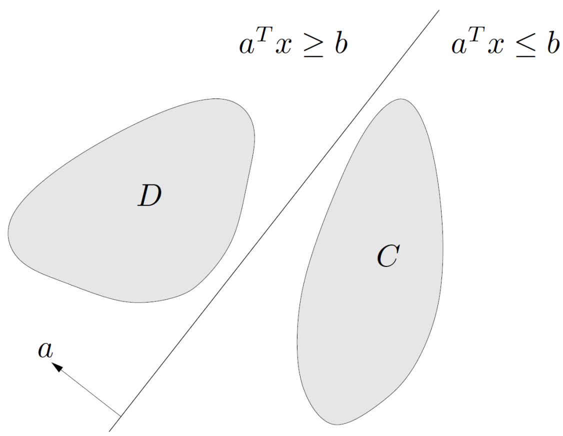

Figure 2.6: Separating hyperplane

   Formally, if C, D are nonempty disjoint convex sets, then there exists a, b such that

$$C \subseteq \{x : a^T x \leq b\}$$
$$D \subseteq \{x : a^T x \geq b\}$$

   Similarly, the **supporting hyperplane theorem** states that any boundary point of a convex set has a supporting hyperplane passing through it:

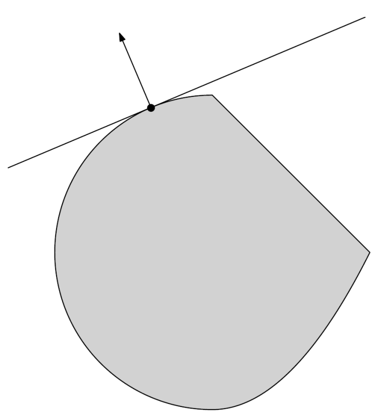

Figure 2.7: Supporting hyperplane

Formally, given a nonempty convex set C, for every point $x_0 \in bd(C)$, there exists a such that

$$C \subseteq \{x : a^T x \leq a^T x_0\}$$

See 2.5 of BV for more.

2.3 Operations Preserving Convexity

There are several operations that can be done on vectors, matrices or sets that preserve convexity.

- **Intersection** of convex sets is convex.
- **Scaling and Translation**: for convex set C, $aC + b = \{ax + b : x \in C\}$ is also convex.
- **Affine image** for convex set C and affine function $f(x) = Ax + b$, the following is convex:

$$f(C) = \{f(x) : x \in C\}$$

- **Affine Preimages** similarly, for convex set D and affine function f (not necessarily invertable), the following is convex:

$$f^{-1}(D) = \{x : f(x) \in D\}$$

- **Perspective image** and **preimage**
For function $P : \mathbb{R}^n \times \mathbb{R}_{++} \to \mathbb{R}^n$ (where $\mathbb{R}_{++}$ is positive reals),

$$P(x, z) = x/z$$

for $z > 0$ is a perspective function. If $C \subseteq \mathrm{dom}(f)$ is convex, then so is $P(C)$, and if D is convex, so is $P^{-1}(D)$.

- **Linear-fractional image and preimage** A linear fractional function is a perspective map composed with an affine function, defined on $c^T x + d > 0$:

$$f(x) = \frac{Ax + b}{c^T x + d}$$

The image and preimage of this function are both convex.

2.4 Examples of Convex Sets + Operations

Three major examples are given: Linear matrix inequality solution set, the Fantope, and Conditional probability set

- **Example 1)** Linear matrix inequality solution set

Given symmetric matrices $A_1, \cdots, A_k, B \in \mathbb{S}^n$, the set of points satisfying a **linear matrix inequality** is:

$$C = \{x : x_1 A_1 + \cdots + x_k A_k \preceq B\} = \{x : B - \sum_i x_i A_i \succeq 0\} \tag{2.1}$$

There are two ways to prove that C is convex. The first approach is to directly verify that $x, y \in C \Rightarrow tx + (1-t)y \in C$. This follows by checking that, for any v,

$$\begin{aligned} v^T(B - \sum_{i=1}^{k}(tx_i + (1-t)y_i)A_i)v &= v^T(tB + (1-t)B - \sum_{i=1}^{k}(tx_i + (1-t)y_i)A_i)v \\ &= t(v^T(B - \sum_i x_iA_i)v) + (1-t)(v^T(B - \sum_i y_iA_i)v) \\ &\geq 0 \end{aligned}$$

Another (smarter) approach is to let $f : \mathbb{R}^k \to \mathbb{S}^n$, $f(x) = B - \sum_i x_iA_i$, and note that this is the affine preimage of a convex set $C = f^{-1}(\mathcal{S}^n_+)$, from the right side of 2.1.

- **Example 2)** Fantope

  A fantope of order k for some integer $k \geq 0$ is:

$$\begin{aligned} \mathcal{F} &= \{Z \in \mathbb{S}^n : 0 \preceq Z \preceq I, \mathrm{tr}(Z) = k\} \\ &= \{Z \in \mathbb{S}^n : 0 \leq \lambda_1(Z) \leq \cdots \lambda_n(Z) \leq 1, \sum \lambda_i = k\} \end{aligned}$$

  One approach to proving this is convex is the usual way: to take two matrices $0 \preceq Z, W \preceq I$ and $\mathrm{tr}(Z) = \mathrm{tr}(W) = k$ implies the same for $tZ + (1-t)W$. A smarter approach is to recognize that this fantope is:

$$\mathcal{F} = \{Z \in \mathbb{S}^n : Z \succeq 0\} \cup \{Z \in \mathbb{S}^n : Z \preceq I\} \cup \{Z \in \mathbb{S}^n : \mathrm{tr}(Z) = k\}$$

  which is an intersection of linear inequality and equality constraints, hence like a polyhedron but for matrices. (the last set is a linear equality)

- **Example 3)** Conditional probability set
  Let U, V be random variables over $\{1, \cdots, n\}$ and $\{1, \cdots, m\}$. Let $C \in \mathbb{R}^{n \times m}$ be a set of joint distributions (probabilities) for U, V. i.e.:

$$p_{ij} = \mathbb{P}(U = i, V = j)$$

  and D contain conditional distributions (probabilities):

$$q_{ij} = \mathbb{P}(U = i | V = j)$$

  Assume C is convex. Let's prove that D is convex. The set D can be rewritten as an image of a linear fractional function:

$$D = \{q \in \mathbb{R}^{n \times m} : q_{ij} = \frac{p_{ij}}{\sum_{k=1}^{n} p_{kj}} \text{ for some } p \in C\} = f(C)$$

  Hence it is convex.

2.5 Convex Functions

If we know sets really well, we can always derive thing for functions.

2.5.1 Definition

Definition 2.1. *A **convex function** sis a function $f : \mathbb{R}^n \to \mathbb{R}$, such that $dom(f) \subseteq \mathbb{R}^n$ is convex, and*

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$$

, for $0 \leq t \leq 1$.

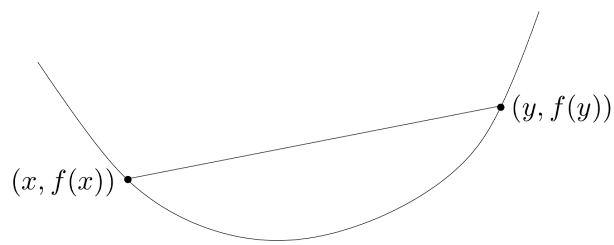

Figure 2.8: Graph of a convex function. The line segment betwen any two points on the graph lies above the graph.

In words, f lies below the line segment joining $f(x)$, $f(y)$.

A **concave function** is where the reverse is true, i.e.,

$$f \text{ concave} \iff -f \text{ convex}$$

There are some important modifiers:

- **Strictly Convex**: A function f is strictly convex if $f(tx + (1-t)y) < tf(x) + (1-t)f(y)$ for $x \neq y$ and $0 < t < 1$. In words, f is convex and has greater curvature than a linear function. With the strict inequality, the line segmement joing $f(x)$, $f(y)$ is strictly above the function inside the interval.
- **Strongly Convex**: A function f is strongly convex with parameter $m > 0$ if $f - \frac{m}{2}\|x\|_2^2$ is convex. In words, f is at least as convex as a quadratic fuction.

From the above definition, we can conclude that

$$\text{strong convexity} \implies \text{strict convexity} \implies \text{convexity}$$

. There are definitions analogously for concave functions, just plugin $-f$.

2.5.2 Examples of convex function

Following are some exmaples:

- **Some univariate functions**:

  - exponential function e^{ax} is convex for any a,
  - power function x^a is convex for $a \geq 1$ or $a \leq 0$, power function is concave for $0 \leq a \leq 1$
  - logrithmic function $\log x$ is concave.

- **Affine function**: $a^T + b$ is both convex and concave,
- **Quadratic function**: $\frac{1}{2}x^T Qx + b^T x + c$ is convex provided that $Q \succeq 0$ (positive semidefinite).
- **Least square loss**: $\|y - Ax\|_2^2$ is always convex (since ATA is always positive semidefinite)

$$\|y - Ax\|_2^2 = x^T A^T Ax - 2y^T Ax + y^T y$$

  where $y^T y$ is constant, and $A^T A$ is always positive semidefinite. So it always satisfies the convexity condition for quadratic function.

- **Norm**:$\|x\|$ is convex for any norm; e.g., l_p norms.

$$\|x\|_p = \left(\sum_{i=1}^{n} x_i^p\right)^{1/p} \text{ for } p \geq 1, \|x\|_\infty = \max_{i=1,\ldots,n} |x_i|$$

  and also operator (spectral) norm and trace (nuclear) norm:

$$\|X\|_{op} = \sigma_1(X), \quad \|X\|_{tr} = \sum_{i=1}^{r} \sigma_r(X)$$

  where $\sigma_1(X) \geq \cdots \geq \sigma_r(X) \geq 0$ are the sigular values of the matrix X.

  Note that $l_0 = \#(i|x_i \neq 0)$ norm is **not** really a norm, is just what we call it.

- **Indicator function**: if C is convex, then its indicator funtion I_C is convex, where

$$I_C = \begin{cases} 0 & x \in C \\ \infty & x \notin C \end{cases}$$

- **Supporting function**: for any set C (convex or not), its support function

$$I_C^*(x) = \max_{y \in C} x^T y$$

  is convex. There is a reason for the notation which we will talk about when we discuss duality.

- **Max function**: $f(x) = \max\{x_1, \ldots, x_n\}$ is convex.

2.5.3 Key properties of convex functions

- A function is convex if and only if its restriction to any line is convex. For $f : \mathbb{R}^n \to \mathbb{R}$, we observe how f behaves over a line in $\mathbb{R}^n$, that is convex for any line if and only if f is convex.
- **Epigraph characterization**: An epigraph of a function $f : \mathbb{R}^n \to \mathbb{R}$ is the set of points lying on or above its graph:

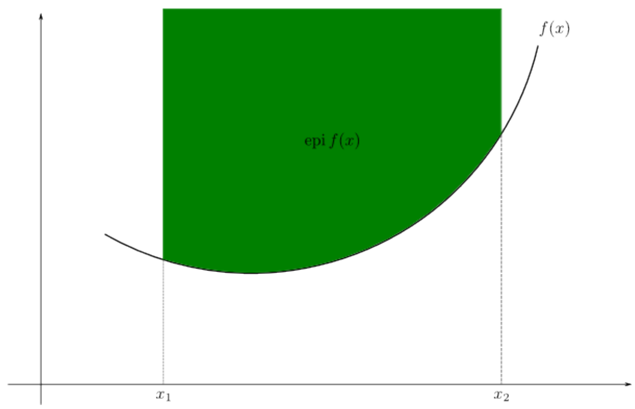

Figure 2.9: Epigraph of f. From wikipedia.

A function f is convex if and only if its epigraph

$$\mathrm{epi}(f) = \{(x,t) \in \mathrm{dom}(f) \times \mathbb{R} : f(x) \leq t\}$$

is a convex set. This property connects functions and sets in terms of convexity. With this property, we can derive everything about convex functions from sets, including KKT conditions etc.

- **Convex Sublevel sets**: if f is convex, then its sublevel sets

$$\{x \in \mathrm{dom}(f) : f(x) \leq t$$

  are convex, for all $t \in \mathbb{R}$. The converse is not true.A function with convex sublevel sets is called a quasiconvex function.

- **First-order characterization**: if f is differentiable, then f is convex if and only if $\mathrm{dom}(f)$ is convex, and

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

  for all $x, y \in \mathrm{dom}(f)$. Therefore for a differentiable convex function $\nabla f(x) = 0 \Longleftrightarrow x$ minimizes f In words, the tangent to the function at x, is an under approximator for the function, i.e., the function lies above its tangent line. The above equation is also called first-order Taylor expansion.

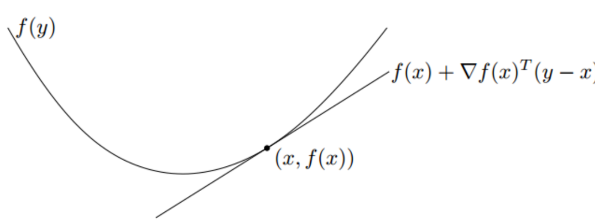

Figure 2.10: If f is differentiable, then f is convex i.f.f. $\mathrm{dom}(f)$ is convex, and $f(y) \geq f(x) + \nabla f(x)^T (y - x)$ for all $x, y \in \mathrm{dom}(f)$

- **Second-order characterization**: if f is twice differentiable, then f is convex if and only if $\mathrm{dom}(f)$ is convex, and $\nabla^2 f(x) \succeq 0$ for all $x \in \mathrm{dom}(f)$

- **Jensen's inequality**: if f is convex, and X is a random variable supported on $\text{dom}(f)$, then $f(\mathbb{E}[X]) \leq \mathbb{E}[f(x)]$

Note: We will use the first-order characterization and second-order characterization a lot.

2.5.4 Operations perserving convexity

Following are some operations perserving convexity.

- **Nonnegative linear combination**: $f_1, \ldots, f_m$ convex implies $a_1 f_1 + \cdots + a_m f_m$ convex for any $a_1, \ldots, a_m \geq 0$.
- **Pointwise maximization**: if f_s is convex for any $s \in S$, then $f(x) = max_{s \in S} f_s(x)$ is convex. Note that the set S here (number of functions f_s) can be infinite.

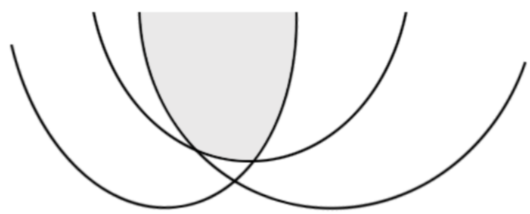

Figure 2.11: The point-wise maximum of convex functions is convex.

- **Partial minimization**: if $g(x, y)$ is convex in x, y, and C is convex, then $f(x) = min_{y \in C} g(x, y) isconvex.$

Examples: *Distance to a set* Let C be an artibrary set, and consider the maximum distance to C under an arbitrary norm $\|\cdot\|$:

$$f(x) = max_{y \in C} \|x - y\|$$

Let's check this is convex: $f_y(x) = \|x - y\|$ is convex for any fixed y, so by point wise maximization rule, f is convex.

Now let C be convex, and consider the minimum distance of C:

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

Let's check this is convex: $g(x, y) = \|x - y\|$ is convex in x, y jointly, and C is assumed convex, so apply partial minimization rule.

More operations preserving convexity.

- **Affine composition**: f convex implies g(x)=f(Ax+b) convex This is really useful when you want to prove some function convex, and you realize there is a affine transformation in there. Because affine transformation mess up sometimes things like taking gradients or Hessians, making it more complicated. Just do not bother with that, just do it for the case there is no affine transformation, and claim when affine transformation is in there, we will still have convexity.
- **General composition**: suppose $f = h \cdot g$, i.e., $f(x) = h(g(x))$, where $g : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R} \to \mathbb{R}$. Then:

  - f in convex if h is convex and **nondecreasing**, g is convex
  - f in convex if h is convex and nonincreasing, g is concave
  - f in concave if h is concave and nondecreasing, g is concave
  - f in concave if h is concave and nonincreasing, g is convex

  How to remember these? You can think of the chain rule when $n = 1$:

$$f''(x) = h''(g(x))g'(x)^2 + h'(g(x))g''(x)$$

- **Vector composition**: suppose that

$$f(x) = h(g(x)) = h(g_1(x), \dots, g_k(x))$$

  , where $g : \mathbb{R}^n \to \mathbb{R}^k$, $h : \mathbb{R}^k \to \mathbb{R}$, $f : \mathbb{R}^n \to \mathbb{R}$. Then:

  - f is convex if h is convex and nondecreasing in each argument, g is convex
  - f is convex if h is convex and nonincreasing in each argument, g is concave
  - f is concave if h is concave and nondecreasing in each argument, g is concave
  - f is concave if h is concave and nonincreasing in each argument, g is convex

Example: *log-sum-exp function* Log-sum-exp function: $g(x) = log(\sum_{i=1}^k e^{a_i^T x + b_i})$, for fixed a_i,b_i,$i = 1, \dots, k$. Ofen called "soft max", as it smoothly approximates $max_{i=1,\dots,k}(a_i^T x + b_i)$.

How to show convexity? First, note it suffices to prove convexity of $f(x) = \log(\sum_{i=1}^n e^{x_i})$(affine composition rule)

Now use second-order characterization. Calculate

$$\nabla_i f(x) = \frac{e^{x_i}}{\sum_{l=1}^n e^{x_l}}$$

$$\nabla_{ij}^2 f(x) = \frac{e^{x_i}}{\sum_{l=1}^n e^{x_l}} 1\{i = j\} - \frac{e^{x_i} e^{x_j}}{(\sum_{l=1}^n e^{x_l})^2}$$

Write $\nabla^2 f(x) = \text{diag}(z) - zz^T$, where $z_i = e^{x_i}/(\sum_{l=1}^n e^{x_l})$. This matrix is diagonally dominant, hence positive semidefinite.

2.6 References and further reading

- S. Boyd and L. Vandenberghe (2004), "Convex optimization", Chapters 2 and 3
- J.P. Hiriart-Urruty and C. Lemarechal (1993), "Fundamentals of convex analysis", Chapters A and B
- R. T. Rockafellar (1970), "Convex analysis", Chapters 1-10