IE 521: Convex Optimization	**Spring 2017, UIUC**

Lecture 4: Convex Functions, Part I – February 1

Instructor: Niao He *Scribe: Shuanglong Wang*

Courtesy warning: These notes do not necessarily cover everything discussed in the class. Please email TA (swang157@illinois.edu) if you find any typos or mistakes.

In this lecture, we cover the following topics

- Convex Functions
- Examples
- Convexity-preserving Operations

Reference: Boyd &Vandenberghe, Chapter 3.1-3.2

4.1 Convex Function

Let f be a function from $\mathbf{R}^n$ to $\mathbf{R}$. The domain of f is defined as $dom(f) = \{x \in \mathbf{R}^n : |f(x)| < \infty\}$. For example,

- $f(x) = \frac{1}{x}, \quad dom(f) = \mathbf{R}\setminus\{0\}$
- $f(x) = \sum_{i=1}^n x_i \ln(x_i), \quad dom(f) = R^n_{++} = \{x : x_i > 0, \forall i = 1, ..., n\}$

Definition 4.1 (Convex function) *A function $f(x) : \mathbf{R}^n \to \mathbf{R}$ is convex if*

(i) $dom(f) \subseteq \mathbf{R}^n$ s a convex set;

(ii) $\forall x, y \in dom(f)$ and $\lambda \in [0, 1]$, $f(\lambda x + (1 - \lambda)y) \le \lambda f(x) + (1 - \lambda)f(y)$.

Geometrically, the line segment between $(x, f(x))$, $(y, f(y))$ sits above the graph of f.

Definitions

- A function is called strictly convex if (ii) holds with strict sign, i.e. $f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y)$.
- A function is called α-strongly convex if $f(x) - \frac{\alpha}{2}\|x\|_2^2$ is convex.
- A function is called concave if $-f(x)$ is convex.

Note that strongly convex $\Longrightarrow$ strictly convex $\Longrightarrow$ convex

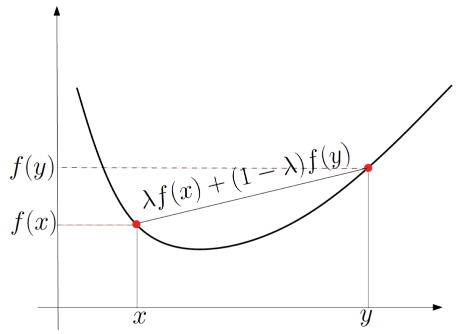

Figure 4.1: Convex function

4.2 Examples

1. **Simple univariate functions:**
   - Even powers: x^p, p is even
   - Exponential: $e^{ax}, \forall a \in \mathbf{R}$
   - Negative logarithmic: $-\log x$
   - Absolute value: $|x|$
   - Negative entropy: $x \log(x)$
2. **Affine functions:** $f(x) = a^T x + b$
   - both convex & concave, but not strctly convex/concave
3. **Some quadratic functions:** $f(x) = \frac{1}{2}x^T Qx + b^T x + c$
   - convex if and only if $Q \succeq 0$ is positive semi-definite
   - strictly convex if and only if $Q \succ 0$ is positive definite
   - special case: $f(x) = \| Ax - b \|_2^2$ is convex
4. **Norms**: A function $\pi(\cdot)$ is called a norm if
   (a) $\pi(x) \geq 0, \forall x$ and $\pi(x) = 0$ iff $x = 0$
   (b) $\pi(\alpha x) = |\alpha| \cdot \pi(x), \forall \alpha \in \mathbf{R}$
   (c) $\pi(x + y) \leq \pi(x) + \pi(y)$

   Note that norms are convex: $\forall \lambda \in [0, 1], \pi(\lambda x + (1 - \lambda)y) \leq \pi(\lambda x) + \pi((1 - \lambda)y) = \lambda\pi(x) + (1 - \lambda)\pi(y)$ where the inequality comes from (c) and the equality comes from (b).
   Examples of norms include:
   - l_p-norm on $\mathbf{R}^n$: $\| x \|_p := (\sum_{i=1}^n |x_i|^p)^{1/p}$, where $p \geq 1$
   - Q-norm on $\mathbf{R}^n$: $\| x \|_Q = \sqrt{x^T Qx}$, where $Q \succ 0$ is positive definite

- Frobenius norm on $\mathbf{R}^{m\times n}$: $\| A \|_F = (\sum_{i=1}^m \sum_{j=1}^n |A_{i,j}|^2)^{1/2}$
- spectral norm on S^n: $\| A \| = \max_{i=1,\ldots,n} |\lambda_i(A)|$, where λ_i's are the eigenvalues of A.

5. **Indicator function** $I_c(x) = \begin{cases} 0, x \in C \\ \infty, x \notin C \end{cases}$

   The indicator function $I_C(x)$ is convex if the set C is a convex set.

6. **Supporting function**: $I_C^*(x) = \sup_{y\in C} x^T y$
   The support function $I_C^*(x)$ is always convex for any set C.
   Proof: Note that $\sup_{y\in C} f(y) + g(y) \le \sup_{y\in C} f(y) + \sup_{y\in C} g(y)$
   Then $\forall x_1, x_2, \lambda \in [0,1]$

$$\begin{aligned} I_C^*(\lambda x_1 + (1-\lambda)x_2) &= \sup_{y\in C} \lambda x_1^T y + (1-\lambda)x_2^T y \\ &\le \sup_{y\in C} \lambda x_1^T y + \sup_{y\in C} (1-\lambda)x_2^T y \\ &= \lambda I_C^*(x_1) + (1-\lambda) I_C^*(x_2) \end{aligned}$$

■

7. **More examples**

   - Piecewise linear functions: $\max(a_1^T x + b_1, ..., a_k^T x + b_k)$
   - Log of exponential sums: $\log(\sum_{i=1}^k e^{a_i^T x + b_i})$
   - Negative log of determinant: $-\log(\det(X))$

   How to show convexity of these functions?

4.3 Convexity-Preserving Operators

1. **Taking conic combination**: If $f_i(x), i \in I$ are convex functions and $\alpha_i \ge 0, \forall i \in I$, then

$$g(x) = \sum_{i\in I} \alpha_i f_i(x)$$

   is a convex function.

   Proof: The domain of function g

$$dom(g) = \cap_{i:\alpha_i>0} dom(f_i)$$

   is convex. For any $x, y \in dom(g), \lambda \in [0,1]$

$$\begin{aligned} g(\lambda x + (1-\lambda)y) &= \sum \alpha_i f_i(\lambda x + (1-\lambda)y) \\ &\le \sum \alpha_i \left[\lambda f_i(x) + (1-\lambda) f_i(y)\right] \\ &= \lambda \sum \alpha_i f_i(x) + (1-\lambda) \sum \alpha_i f_i(y) \\ &= \lambda g(x) + (1-\lambda) g(y) \end{aligned}$$

■

Remark The property extend to infinite sums and integrals. If $f(x, \omega)$ is convex in x for any $\omega \in \Omega$ and $\alpha(\omega) \geq 0, \forall \omega \in \Omega$. then

$$g(x) = \int_{\Omega} \alpha(\omega) f(x, \omega) d\omega$$

is convex if well defined.

For example if $\eta = \eta(\omega)$ is a well-defined random variable on Ω, and $f(x, \eta(\omega))$ is convex, $\forall \omega \in \Omega$, then $\mathbf{E}_{\eta}[f(x, \eta)]$ is a convex function.

2. **Taking affine composition** If $f(x) : \mathbf{R}^n \to \mathbf{R}$ is convex and $\mathcal{A}(y) : y \mapsto Ay + b$ is an affine mapping from $\mathbf{R}^m$ to $\mathbf{R}^n$, then

$$g(y) := f(Ay + b)$$

is convex on $\mathbf{R}^m$.

Proof: $dom(g) = \{y : Ay + b \in dom(f)\}$ is convex.

$$\begin{aligned} \forall y_1, y_2 \in dom(g) : g(\lambda y_1 + (1 - \lambda) y_2) &= f(\lambda(Ay_1 + b) + (1 - \lambda)(Ay_2 + b)) \\ &\leq \lambda f(Ay_1 + b) + (1 - \lambda) f(Ay_2 + b) \\ &= \lambda g(y_1) + (1 - \lambda) g(y_2) \end{aligned}$$

■

For example, $\| Ax - b \|_2^2$, $\sum_i e^{a_i^T x - b_i}$ and $\sum_{i=1}^n \log(a_i^T x - b_i)$ are convex.

3. **Taking pointwise maximum and supremum** If $f_i(x), i \in I$ are convex, then

$$g(x) := max_{i \in I} f_i(x)$$

is also convex.

Proof: First of all, $dom(g) = \cap_{i \in I} dom(f_i)$ is convex, For any $x, y \in dom(g), \lambda \in [0, 1]$

$$\begin{aligned} g(\lambda x + (1 - \lambda) y) &= \max_{i \in I} f_i(\lambda x + (1 - \lambda) y) \\ &\leq \max_{i \in I} \{\lambda f_i(x) + (1 - \lambda) f_i(y)\} \\ &\leq \max_{i \in I} \lambda f_i(x) + \max_{i \in I} (1 - \lambda) f_i(y) \\ &= \lambda g(x) + (1 - \lambda) g(y) \end{aligned}$$

■

Remark The property extends to the pointwise supremum over a infinite set. If $f(x, \omega)$ is convex in x, for $\omega \in \Omega$, then

$$g(x) := \sup_{\omega \in \Omega} f(x, \omega)$$

is convex.
For example, the following functions are convex:

(a) piecewise linear functions: $f(x) = max(a_1^T x + b_1, ..., a_k^T x + b_k)$

(b) support function: $I_C^*(x) = \sup_{y \in C} x^T y$

(c) maximum distance to any set C: $d_{max}(x, C) = \max_{y \in C} \| y - x \|_2$

(d) maximum eigenvalue of a symmetric matrix: $\lambda_{max}(X) = \max_{\|y\|_2 = 1} y^T X y$

Indeed, almost every convex function can be expressed as the pointwise supremum of a family of affine functions!

4. **Taking convex monotone composition**:

- **scalar case** If f is a convex function on $\mathbf{R}^n$ and $F(\cdot)$ is a convex and non-decreasing function on $\mathbf{R}$, then $g(x) = F(f(x))$ is convex.
- **vector case** If $f_i(x), i = 1, \ldots, m$ are convex on $\mathbf{R}^n$ and $F(y_1, \ldots, y_m)$ is convex and non-decreasing (component-wise) in each argument, then

$$g(x) = F(f_1(x), \ldots, f_m(x))$$

is convex.

Proof: By convexity of f_i, we have

$$f_i(\lambda x + (1 - \lambda) y) \leq \lambda f_i(x) + (1 - \lambda) f_i(y), \forall i, \forall \lambda \in [0, 1].$$

Hence, we have for any $x, y \in dom(g)$, $\lambda \in [0, 1]$,

$$\begin{aligned} g(\lambda x + (1 - \lambda) y) &= F(f_1(\lambda x + (1 - \lambda) y), \ldots, f_m(\lambda x + (1 - \lambda) y)) \\ &\leq F(\lambda f_1(x) + (1 - \lambda) f_1(y), \ldots, \lambda f_m(x) + (1 - \lambda) f_m(y)) \quad \text{(by monotonicity of } F) \\ &\leq \lambda F(f_1(x), \ldots, f_m(x)) + (1 - \lambda) F(f_1(x), \ldots, f_m(x)) \quad \text{(by convexity of } F) \\ &= \lambda g(x) + (1 - \lambda) g(y) \quad \text{(by definition of } g) \end{aligned}$$

■

Remark Taking pointwise maximum is a special case of the above rule, by setting $F(y_1, ..., y_m) = \max(y_1, ..., y_m)$,

$$\max_{i=1,\ldots,m} f_i(x) = F(f_1(x), ..., f_m(x))$$

is convex.

For example:

(a) $e^{f(x)}$ is convex if f is convex

(b) $-\log f(x)$ is convex if f is concave

(c) $\log(\sum_{i=1}^k e^{f_i})$ is convex if f_i are convex.

5. **Taking Partial minimization**: If $f(x, y)$ is convex in $(x, y) \in \mathbf{R}^n$ and Y is a convex set, then

$$g(x) = \inf_{y \in Y} f(x, y)$$

is convex.

Proof: $dom(g) = \{x : (x, y) \in dom(f) \text{ and } y \in C\}$ is a projection of $dom(f)$, hence is convex. Given any x_1, x_2, by definition, for any $\epsilon > 0$, $\exists y_1 \in Y, y_2 \in Y$ s.t.

$$f(x_1, y_1) \leq g(x_1) + \epsilon/2$$
$$f(x_2, y_2) \leq g(x_2) + \epsilon/2$$

For any $\lambda \in [0, 1]$, adding the two equations, we have

$$\lambda f(x_1, y_1) + (1 - \lambda) f(x_2, y_2) \leq \lambda g(x_1) + (1 - \lambda) g(x_2) + \epsilon.$$

By convexity of $f(x, y)$, this implies

$$f(\lambda x_1 + (1 - \lambda)x_2, \lambda y_1 + (1 - \lambda)y_2) \leq \lambda g(x_1) + (1 - \lambda) g(x_2) + \epsilon.$$

Hence for any $\epsilon > 0$, $g(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda g(x_1) + (1 - \lambda) g(x_2) + \epsilon$. Letting $\epsilon \to 0$ leads to the convexity of g. ■

Examples

(a) Minimum distance to a convex set: $d(x, C) = \min_{y \in C} \| x - y \|_2$ where C is convex;

(b) Define

$$g(x) = \inf_y \{h(y) | Ay = x\}$$

is convex if h is convex. This is because $g(x) = \inf_y f(x, y)$, where

$$f(x, y) := \begin{cases} h(x) & Ay = x \\ \infty & o.w. \end{cases}$$

is convex in (x, y).