10-801: Advanced Optimization and Randomized Methods

Lecture 2: Convex functions

(Jan 15, 2014)

Lecturer: *Suvrit Sra* ***Scribes:*** *Avinava Dubey, Ahmed Hefny*
Addr: *Carnegie Mellon University, Spring 2014*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

2.1 Review

The lecture began with a review of convexity in metric spaces, a topic that we could not cover fully in Lecture 1.

Definition 2.1. A metric space $(\mathcal{X}, d)$ is called *complete* if any Cauchy sequence within the space converges to a point in the space.

Definition 2.2. A metric space is called *locally compact* if every point in the space has a compact neighborhood.

Note: Sets that have empty interiors do not have this property (since having empty interiors, they cannot be neighborhoods).

Definition 2.3. A *geodesic* is a continuous path between two points in $(\mathcal{X}, d)$ denoted by:

$$\gamma(t) = [x, y]_t, \quad t \in [0, 1],$$

s.t. $\gamma(1) = y,\ \gamma(0) = x$ satisfying:

$$d(\gamma(t_1), \gamma(t_2)) = |t_1 - t_2| d(x, y) \quad \forall t_1, t_2 \in [0, 1]$$

The latter condition implies that a geodesic is a shortest path.

Theorem 2.4. *Suppose $(\mathcal{X}, d)$ is a complete, locally compact metric space. Then, the following are equivalent:*

(i) $(\mathcal{X}, d)$ *is Menger-convex*

(ii) $(\mathcal{X}, d)$ *has "midpoints", i.e.,*

$$\forall x, y \in \mathcal{X}, \exists m \in \mathcal{X} \ s.t.\ d(x, m) = d(y, m) = \tfrac{1}{2} d(x, y),$$

(iii) $(\mathcal{X}, d)$ *is a* geodesic space, *i.e.,*

$$\forall x, y \in \mathcal{X}, \quad \text{there exists a geodesic } [x, y]_t.$$

2.2 Convex Functions

2.2.1 Notations and Conventions: Extended Reals

Before we start the topic of convex functions, we need to introduce notation that will be useful when we discuss functions whose values are infinite. We define the set of *extended reals* as $\bar{\mathbb{R}} := \mathbb{R} \cup \{\infty, -\infty\}$. By convention, the value of infinity has the following properties:

$$\begin{aligned}
&x + \infty = \infty,\ -\infty < x \leq \infty \\
&0 \cdot \infty = \infty \cdot 0 = 0 \\
&-(-\infty) = +\infty \\
&\inf \phi = +\infty,\ \sup \phi = -\infty
\end{aligned}$$

This convention allows us to talk about convex functions on $\mathbb{R}$ without always having to worry about their domains. For example,

$$f(x) = \begin{cases} \frac{1}{x} & \text{for } x > 0 \\ \infty & \text{for } x \leq 0. \end{cases}$$

The domain of f, denoted $\operatorname{dom} f$ is defined to be the (convex) set on which f assumes values smaller than $+\infty$.

2.2.2 Convexity and Midpoint Convexity

Definition 2.5. A function f is *midpoint convex* if

$$f\left(\frac{x+y}{2}\right) \leq \frac{f(x)+f(y)}{2} \quad \forall x, y \in \operatorname{dom}(f)$$

Definition 2.6. A function f is *convex* if

$$f((1-\alpha)x + \alpha y) \leq (1-\alpha)f(x) + \alpha f(y) \quad \forall x, y \in \operatorname{dom}(f), \alpha \in (0,1)$$

This condition is called *Jensen's inequality*.

Definition 2.7. A function f defined on a metric space is *convex* if

$$f((1-\alpha)x \oplus \alpha y) \leq (1-\alpha)f(x) + \alpha f(y) \quad \forall x, y \in \operatorname{dom}(f), \alpha \in (0,1),$$

where $(1-\alpha)x \oplus \alpha y := \gamma(\alpha)$, γ represents the geodesic as defined in 2.3.

Theorem 2.8. *(Jensen, 1905) If f is a continuous midpoint convex function then f is convex.*

Proof. By contradiction: Suppose f is a continuous midpoint convex function that does not satisfy Jensen's inequality at some choice of x, y. Define:

$$g(\alpha) := f((1-\alpha)x + \alpha y) - (1-\alpha)f(x) - \alpha f(y)$$

Then, by our assumption

$$\exists \quad \alpha \in (0,1) \text{ s.t. } g(\alpha) > 0.$$

Then

$$\max_{\alpha \in (0,1)} g(\alpha) = M > 0.$$

Let α_0 be the smallest value of $\alpha \in (0,1)$ satisfying $g(\alpha_0) = M$. Also, let $\delta > 0$ be small enough such that $(\alpha_0 - \delta, \alpha_0 + \delta) \subset (0,1)$. Define:

$$\bar{x} := (1 - \alpha_0 - \delta)x + (\alpha_0 + \delta)y$$
$$\bar{y} := (1 - \alpha_0 + \delta)x + (\alpha_0 - \delta)y$$

Then the midpoint convexity assumption implies that

$$f\left(\frac{\bar{x}+\bar{y}}{2}\right) \leq \frac{f(\bar{x})+f(\bar{y})}{2}$$

Note that

$$\begin{aligned} g(\alpha_0+\delta) + g(\alpha_0-\delta) &= f(\bar{x}) + f(\bar{y}) - 2[(1-\alpha_0)f(x) + \alpha_0 f(y)] \\ &= f(\bar{x}) + f(\bar{y}) + 2[g(\alpha_0) - f((1-\alpha_0)x + \alpha_0 y)] \\ &= 2g(\alpha_0) + f(\bar{x}) + f(\bar{y}) - 2f\left(\frac{\bar{x}+\bar{y}}{2}\right) \geq 2g(\alpha_0) \end{aligned}$$

Therefore,

$$g(\alpha_0) \leq \frac{g(\alpha_0+\delta)+g(\alpha_0-\delta)}{2} < \frac{M+M}{2} < M$$

which contradicts our premise that $g(\alpha_0) = M$. $\square$

Theorem 2.8 shows that, in oder to prove the convexity of a continuous function, it is sufficient to prove midpoint convexity. This simplification can be helpful, as can be seen in the following example.

Example 2.9. The function $f(x) := \log\det(X)$ $(X \in \mathbb{P}^n_+)$ is concave.
Note: $\mathbb{P}^n_+$ is the set of symmetric positive definite matrices.

Proof. Since the function is continuous, it is sufficient to prove midpoint concavity, that is we need to show that

$$\left|\frac{X+Y}{2}\right| \geq |X|^{\frac{1}{2}}|Y|^{\frac{1}{2}}$$

By dividing each side by $|X|$, we obtain

$$\left|\frac{I+X^{-1}Y}{2}\right| \geq |X^{-1}Y|^{\frac{1}{2}}$$

$$\Leftrightarrow \prod_i \left(\frac{1+\lambda_i(X^{-1}Y)}{2}\right) \geq \prod_i \sqrt{\lambda_i(X^{-1}Y)},$$

where $\lambda_i(X)$ is the i^{th} eigenvalue of X.
Note: the i^{th} factor in the LHS is the arithmetic mean of λ_i and 1, while the i^{th} factor in the RHS is their geometric mean. Therefore it suffices to prove that $\lambda_i \geq 0$. X is symmetric positive definite, therefore so are its inverse and square root. Hence

$$\lambda_i(X^{-1}Y) = \lambda_i(X^{1/2}X^{-1}YX^{-1/2}) = \lambda_i(X^{-1/2}YX^{-1/2}) \geq 0,$$

since $X^{-1/2}$ is symmetric and $Y \succ 0$, whereby

$$z^T X^{-1/2} Y X^{-1/2} z = \tilde{z}^T Y \tilde{z} \geq 0 \qquad \forall \tilde{z} \neq 0.$$

□

Exercise 2.1. ***[CHALLENGE]*** *Let $x_1, x_2, \ldots, x_n > 0$ be a sequence of real variables. Define*

$$\begin{aligned}
h_1(x_1) &:= \frac{1}{x_1} \\
h_2(x_1, x_2) &:= \frac{1}{x_1} + \frac{1}{x_2} - \frac{1}{x_1+x_2} \\
h_3(x_1, x_2, x_3) &:= \frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} - \frac{1}{x_1+x_2} - \frac{1}{x_2+x_3} - \frac{1}{x_1+x_3} + \frac{1}{x_1+x_2+x_3} \\
&\vdots
\end{aligned}$$

Prove that h_n is convex.

Definition 2.10. The *epigraph* of a function f is defined as

$$epi(f) := \{(x,t) \in \mathbb{R}^n \times \mathbb{R} | f(x) \leq t\}$$

A convex function f is called *closed* if its epigraph $epi(f)$ is closed and convex.

An example of an epigraph is shown in figure 2.1.

2.2.3 Important Convex Functions

Example 2.11. Let $C \subset R^n$ be a non-empty set. Define

$$\delta_C(x) := \begin{cases} 0 & \text{if } x \in C \\ \infty & \text{if } x \notin C \end{cases}$$

If C is closed and convex then δ is closed and convex.

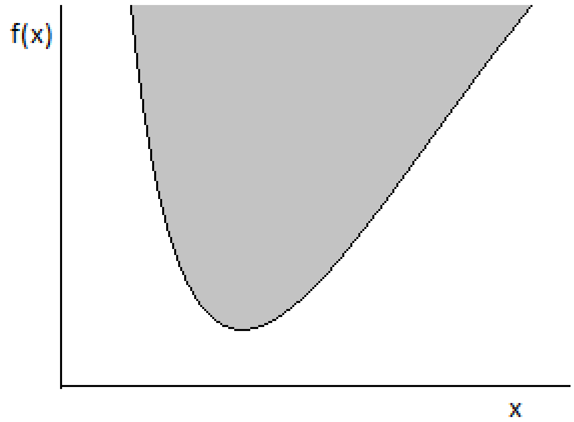

Figure 2.1: A function of a single variable with the epigraph indicated by the shaded region.

Example 2.12. Let $C \subset R^n$ be a non-empty closed set. The *support function* of C, defined as

$$\sigma_C(z) := \sup \langle z, x \rangle,$$

is a convex function.

Example 2.13. Let $h(x, y)$ be a family of functions indexed by $y \in \mathcal{Y}$. If $h(x, y)$ is convex in x for each y then

$$f(x) := \sup_{y \in \mathcal{Y}} h(x, y)$$

is also convex.

Definition 2.14. Let $f : \mathbb{R}^n \to \bar{\mathbb{R}}$. The *Fenchel conjugate* of function f is defined as

$$f^*(z) := \sup_{x \in \mathbb{R}^n} \langle z, x \rangle - f(x)$$

Note that f^* is a special case of example 2.13 and hence is convex regardless of the convexity of f.

2.3 Norms

Definition 2.15. A function $f : \mathbb{R}^n \to \mathbb{R}$ is a norm on $\mathbb{R}^n$ if it satisfies the following conditions:

1. $f(x) \geq 0$, $f(x) = 0$ iff $x = 0$
2. $f(\lambda x) = |\lambda| f(x), \quad \forall \lambda \in \mathbb{R}$
3. $f(x + y) \leq f(x) + f(y)$

If f is a norm then f is convex.

Example 2.16. The ℓ_p norm on $\mathbb{R}^n$ for $1 \leq p < \infty$ is defined as:

$$\|x\|_p := \left(\sum_i |x_i|^p \right)^{\frac{1}{p}}$$

For $p = \infty$ the norm is defined as

$$\|x\|_\infty := \max_{1 \leq i \leq n} |x_i|$$

Example 2.17. Let $x = (\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k) \in \mathbb{R}^{n_1 + n_2 + \cdots + n_k}$. The $\ell_{p,q}$ norm is defined as

$$\|x\|_{p,q} := \left(\sum_i \|\vec{x}_i\|_q^p \right)^{1/p}$$

2.3.1 Matrix Norms

There are different ways to define matrix norms depending on how we look at a matrix. We can view a matrix as a linear operator, which leads to the operator norm.

Definition 2.18. The *operator norm* of a matrix A is defined as

$$\|A\|_u := \sup_{x \neq 0} \frac{\|Ax\|_u}{\|x\|_u}$$

Exercise 2.2. *Show that the following matrix functions are norms:*

- $\|A\|_2 = \sigma_1(A)$, *the largest singular value of* A *(also known as the spectral norm)*
- $\|A\|_1$, *the largest absolute column sum*
- $\|A\|_\infty$, *the largest absolute row sum*

It is worth noting that, for a generic p, $\|A\|_p$ *is NP-hard to compute.*

Alternatively, we can treat the elements of an $m \times n$ matrix as a vector and use any vector norm. Using the ℓ_2 norm on the matrix viewed as a vector results in the so-called Frobenius norm.

Definition 2.19. The *Frobenius norm* of a matrix $A \in \mathbb{C}^{m \times n}$ is defined as

$$\|A\|_{\mathrm{F}} := \sqrt{\mathrm{tr}(A^* A)} = \sqrt{\sum_{ij} |a_{ij}|^2}$$

Definition 2.20. The *Schatten p-norm* of a matrix $A \in \mathbb{C}^{m \times n}$ is defined as

$$\|A\|_p = \|\sigma(A)\|_p,$$

where $\sigma(A)$ is a vector of singular values. Setting p to 1 yields the *trace norm* $\|A\|_*$, also know as the *nuclear norm*.

Note: While the largest singular value is a valid matrix norm, the largest eigenvalue is not. The largest eigenvalue is, however, a valid norm over positive definite matrices.

2.3.2 Dual Norm and Hölder's Inequality

Definition 2.21. For a norm $\|\cdot\|$ over $\mathbb{R}^n$, the *dual norm* is defined as:

$$\|u\|_* := \sup_x \{u^T x \mid \|x\| \leq 1\}$$

Theorem 2.22 (Hölders's inequality)**.** *For any* $x, y \in \mathbb{R}^n$,

$$x^T y \leq \|x\| \|y\|_*$$

Exercise 2.3. *Let* $\|x\|$ *be a norm. Show that* $f^*(z) = \delta_{\|\cdot\|_* \leq 1}(z)$. *[Hint: Consider separately the cases where* $\|z\|_* > 1$ *and* $\|z\|_* \leq 1$. *You might also need to use Hölder's inequality].*