Introduction to Convex Optimization

Lecture 4: Convex Function

Silin DU

Department of Management Science and Engineering
Tsinghua University
dsl21@mails.tsinghua.edu.cn

August 16, 2022

Contents

Lecture Overview

In this lecture, we focus on some definitions and properties in convex functions, which are the foundation of convex optimization.

1. Convex functions.
2. Operations that preserve convexity.
3. Conjugate functions.
4. Quasiconvex functions.
5. Operations that preserve quasiconvexity.
6. Log-concave functions.
7. Convexity by generalized inequality.

We put some proofs in appendix.

Contents

Definition I

Definition 1 (Convex Function)

A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if $\textbf{dom} f$ is a convex set and if for all x, $y \in \textbf{dom} f$, and θ with $0 \leq \theta \leq 1$, we have

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

Fig. 1: Graph of a convex function.

- Geometrically, the line segment between $(x, f(x))$ and $(y, f(y))$, which is the chord from x to y, lies above the graph of f.

Definition II

- A function f is strictly convex if strict inequality holds in Eq. 1 whenever $x \neq y, 0 < \theta < 1$. We say f is concave if $-f$ is convex, and strictly concave if $-f$ is strictly convex.
- It is often convenient to extend a convex function to all of $\mathbb{R}^n$ by defining its value to be ∞ outside its domain. If f is convex we define its extended-value extension $\tilde{f} : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ by

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

  The extension $\tilde{f}$ is defined on all $\mathbb{R}^n$, and takes values in $\mathbb{R} \cup \{\infty\}$. We can recover the domain of the original function f from the extension $\tilde{f}$ as $\operatorname{\mathbf{dom}} f = \{x \mid \tilde{f}(x) < \infty\}$.

First-order Conditions I

Property 1 (First-order Conditions, FOC)

Suppose f is differentiable (i.e., its gradient ∇f exists at each point in dom f, which is open). Then f is convex if and only if dom f is convex and

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x) \tag{2}$$

holds for all $x, y \in \mathbf{dom} f$.

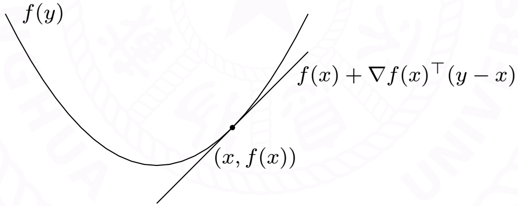

Fig. 2: If f is convex and differentiable, then $f(x) + \nabla f(x)^\top (y - x) \leq f(y)$ for all $x, y \in \mathbf{dom} f$.

First-order Conditions II

- The affine function of y given by $f(x) + \nabla f(x)^\top$ is the first-order Taylor approximation of f near x.
- Eq. 2 states that for a convex function, the first-order Taylor approximation is in fact a global underestimator of the function.
- From local information about a convex function (i.e., its value and derivative at a point) we can derive global information (i.e., a global underestimator of it).
- Eq. 2 shows that if $\nabla f(x) = 0$, then for all $y \in \mathbf{dom} f$, $f(y) \geq f(x)$, i.e., x is a global minimizer of f.
- f is strictly convex if and only if $\mathbf{dom} f$ is convex and for $x, y \in \mathbf{dom} f, x \neq y$, we have

$$f(y) > f(x) + \nabla f(x)^\top (y - x)$$

- f is concave if and only if $\mathbf{dom} f$ is convex and

$$f(y) \leq f(x) + \nabla f(x)^\top (y - x)$$

- The proof of first-order convexity condition is shown in Appendix 1.

Second-order Conditions I

Property 2 (Second-order Conditions, SOC)

Assume that f is twice differentiable, that is, its Hessian or second derivative $\nabla^2 f$ exists at each point in $\mathbf{dom} f$, which is open. Then f is convex if and only if $\mathbf{dom} f$ is convex and its Hessian is positive semidefinite: for all $x \in \mathbf{dom} f$,

$$\nabla^2 f(x) \succeq 0.$$

- For a function on $\mathbb{R}$, this reduces to the simple condition $f''(x) \geq 0$ (and $\mathbf{dom} f$ convex, i.e., an interval), which means that the derivative is nondecreasing.
- Similarly, f is concave if and only if $\mathbf{dom} f$ is convex and $\nabla^2 f(x) \preceq 0$ for all $x \in \mathbf{dom} f$.
- Strict convexity can be partially characterized by second-order conidtions. If $\nabla^2 f(x) \prec 0$ for all $x \in \mathbf{dom} f$, then f is strictly convex.
- The converse is not true: for example, the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^4$ is strictly convex but has zero second order derivative at $x = 0$.
- The proof of second-order convexity condition is shown in Appendix 2.

Second-order Conditions II

- Consider the quadratic function $f : \mathbb{R}^n \to \mathbf{dom} f = \mathbb{R}^n$, given by

$$f(x) = \frac{1}{2}x^\top Px + q^\top x + r,$$

  with $P \in \mathcal{S}^n, q \in \mathbb{R}^n$, and $r \in \mathbb{R}$. Since $\nabla^2 f(x) = P$ for all x, f is convex if and only if $P \succeq 0$ (and concave if and only if $P \preceq 0$).
- The separate requirement that $\mathbf{dom} f$ be convex cannot be dropped from the first- or second-order characterizations of convexity and concavity. For example, the function $f(x) = 1/x^2$, with $\mathbf{dom} f = \{x \mid x \neq 0, x \in \mathbb{R}\}$, satisfies $f''(x) > 0$ for all $x \in \mathbf{dom} f$, but is not a convex function.

Jensen's Inequalities I

Property 3 (Jensen's Inequalities)

Let $f : \mathbb{R}^n \to \mathbb{R}$, then f is convex if and only if

$$f\left(\sum_{i=1}^{m} \lambda_i x_i\right) \le \sum_{i=1}^{m} \lambda_i f(x_i)$$

where $\lambda_i \ge 0, \sum_{i=1}^{m} \lambda_i = 1, x_i \in \mathbf{dom} f$.

- The inequality can extend to infinite sums. If $p(x) \ge 0$ on $S \subseteq \mathbf{dom} f$, and $\int_S p(x)\,\mathrm{d}x = 1$, then

$$f\left(\int_S p(x) x \,\mathrm{d}x\right) \le \int_S f(x) p(x)\,\mathrm{d}x$$

  provided the integrals exist and f is convex.

- If x is a random variable such that $x \in \mathbf{dom} f$ with probability one, and f is convex, then we have

$$f\left(\mathbf{E}(x)\right) \le \mathbf{E}(f(x))$$

Jensen's Inequalities II

- Consider the arithmetic-geometric mean inequality:

$$a^{\theta} b^{1-\theta} \leq \theta a + (1-\theta) b \tag{3}$$

  for $a, b \geq 0$. The function $-\log x$ is convex; Jensen's inequality with $0 \leq \theta \leq 1$ yields

$$-\log(\theta a + (1-\theta) b) \leq -\theta \log a - (1-\theta) \log b$$

  Taking the exponential of both sides yields Eq. 3.

- It can extent to a more general cases. Consider $x_i \geq 0, \lambda_i \geq 0, i = 1 \cdots n$. By the convexity of $-\log x$, we have

$$\begin{aligned}
-\log\left(\sum_{i=1}^{n} \lambda_i x_i\right) &\leq -\sum_{i=1}^{n} \lambda_i \log x_i \\
&= -\sum_{i=1}^{n} \log\left(x_i^{\lambda_i}\right) \\
&= -\log\left(\prod_{i=1}^{n} x_i^{\lambda_i}\right)
\end{aligned}$$

Jensen's Inequalities III

Taking the exponential of both sides yields

$$\sum_{i=1}^{n} \lambda_i x_i \geq \prod_{i=1}^{n} x_i^{\lambda_i}$$

$\lambda_i = \frac{1}{n}$ yields a more general arithmetic-geometric mean inequality:

$$\frac{1}{n} \sum_{i=1}^{n} x_i \geq \sqrt[n]{\prod_{i=1}^{n} x_i}$$

- Hölder's inequality: for $p > 1, 1/p + 1/q = 1$, and $x, y \in \mathbb{R}^n$,

$$\sum_{i=1}^{n} x_i y_i \leq \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} \left(\sum_{i=1}^{n} |y_i|^q \right)^{1/q}$$

We know that

$$a^\theta b^{1-\theta} \leq \theta a + (1-\theta) b$$

Jensen's Inequalities IV

valid for $a, b \geq 0, 0 \leq \theta \leq 1$. Applying this with

$$a = \frac{|x_i|^p}{\sum_{j=1}^n |x_j|^p}, \quad b = \frac{|y_i|^q}{\sum_{j=1}^n |y_j|^q}, \quad \theta = 1/p$$

yields

$$\left(\frac{|x_i|^p}{\sum_{j=1}^n |x_j|^p} \right)^{1/p} \left(\frac{|y_i|^q}{\sum_{j=1}^n |y_j|^q} \right)^{1/q} \leq \frac{|x_i|^p}{p \sum_{j=1}^n |x_j|^p} + \frac{|y_i|^q}{q \sum_{j=1}^n |y_j|^q}.$$

Summing over i yields

$$\sum_{i=1}^n \frac{x_i y_i}{\left(\sum_{j=1}^n |x_j|^p\right)^{1/p} \left(\sum_{j=1}^n |y_j|^q\right)^{1/q}} \leq 1$$

which is the Hölder's inequality.

Property of Convex Functions I

Property 4

$f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if the function $g : \mathbb{R} \to \mathbb{R}, g(t) = f(x + tv)$, $\mathbf{dom} g = \{t \mid x + tv \in \text{dom g}\}$ is convex in t.

- Log-determinate function. $f : \mathcal{S}^n \to \mathbb{R}, f(x) = \log(\det(X)), \mathbf{dom} f = \mathcal{S}^n_{++}$. Is f convex or concave?
- Let $g(t) = f(X + tV), \mathbf{dom} g = \left\{t \mid X + tV \in \mathcal{S}^n_{++}\right\}$

$$
\begin{aligned}
g(t) &= \log(\det(X + tV)), X \in \mathcal{S}^n_{++}, V \in \mathcal{S}^n \\
&= \log\left(\det X^{1/2}\left(I + tX^{-1/2}VX^{-1/2}\right)X^{1/2}\right) \\
&= \log\left(\det X^{1/2}\left(UU^\top + tU\Lambda U^\top\right)X^{1/2}\right) && \text{(Eigendecomposition)} \\
&= 2\log\left(\det X^{1/2}\right) + \log\left(\det\left(U(I + t\Lambda)U^\top\right)\right) \\
&= \underbrace{\log(\det X)}_{\text{constant}} + \underbrace{\sum_{i=1}^{n} \log\left(1 + t\lambda_i\right)}_{\text{concave in } t}
\end{aligned}
$$

Property of Convex Functions II

- All affine (and therefore also linear) functions are both convex and concave. Conversely, any function that is convex and concave is affine.
- A convex function is continuous on the relative interior of its domain; it can have discontinuities only on its relative boundary.

Sublevel Sets I

Definition 2 (α-sublevel set)

The α-sublevel set of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as

$$C_\alpha = \{x \in \mathbf{dom} f \mid f(x) \leq \alpha\}$$

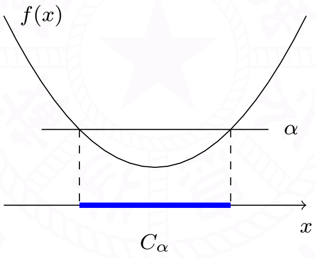

Fig. 3: α-sublevel set

Sublevel Sets II

- Sublevel sets of a convex function are convex, for any value of α. If $x, y \in C_\alpha$, then $f(x) \leq \alpha, f(y) \leq \alpha$, and so $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y) = \alpha$ for $0 \leq \theta \leq 1$, and hence $\theta x + (1-\theta)y \in C_\alpha$.
- The converse is not true. A function can have all its sublevel sets convex but not be a convex function. For example, $f(x) = e^{-x}$ is not convex (indeed, it is strictly concave) but all its sublevel sets are convex.

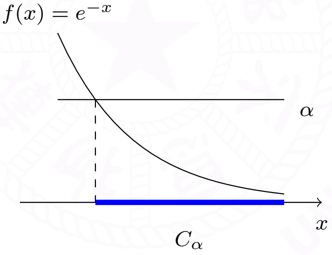

Fig. 4: $f(x) = e^{-x}$ os concave but all its sublevel sets are convex.

Sublevel Sets III

- If f is concave, then its α-superlevel set, given by $\{x \in \mathbf{dom} f \mid f(x) \geq \alpha\}$, is a convex set.

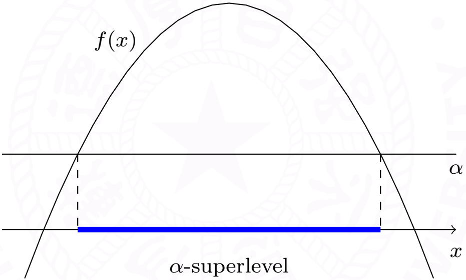

Fig. 5: The α-superlevel set of a concave function.

Sublevel Sets IV

- The sublevel set property is often a good way to establish convexity of a set.
- The geometric and arithmetic means of $x \in \mathbb{R}_+^n$ are, respectively,

$$G(x) = \left(\prod_{i=1}^{n} x_i\right)^{1/n}, \quad A(x) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

(where we take $0^{1/n} = 0$ in our definition of G). The arithmetic-geometric mean inequality states that $G(x) \leq A(x)$ Suppose $0 \leq \alpha \leq 1$, and consider the set

$$\left\{x \in \mathbb{R}_+^n \mid G(x) \geq \alpha A(x)\right\}$$

i.e., the set of vectors with geometric mean at least as large as a factor α times the arithmetic mean. This set is convex, since it is the 0-superlevel set of the function $G(x) - \alpha A(x)$, which is concave. In fact, the set is positively homogeneous, so it is a convex cone.

Sublevel Sets V

Theorem 1

Let f be a function: $\mathbb{R}^n \to \mathbb{R}$. Then f is convex if and only if

$$f((1-\lambda)x + \lambda y) < (1-\lambda)\alpha + \lambda\beta, \lambda \in (0,1)$$

whenever $f(x) < \alpha, f(y) < \beta$.

Epigraph I

Definition 3 (Epigraph)

The graph of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as

$$\{(x, f(x)) \mid x \in \mathbf{dom} f\}$$

which is a subset of $\mathbb{R}^{n+1}$. The epigraph of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as

$$\mathbf{epi}(f) = \{(x, t) \mid x \in \mathbf{dom} f, f(x) \leq t\}$$

which is a subset of $\mathbb{R}^{n+1}$. ('Epi' means 'above' so epigraph means 'above the graph'.)

Epigraph II

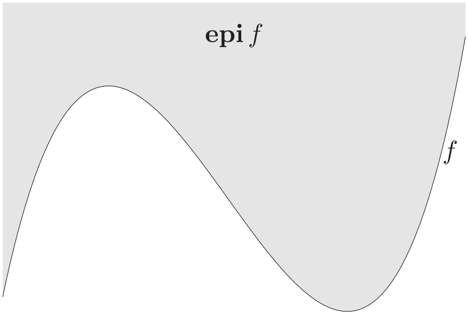

Fig. 6: Epigraph of a function f, shown shaded. The lower boundary shown darker, is the graph of f.

- A function is convex if and only if its epigraph is a convex set. A function is concave if and only if its hypograph, defined as

$$\mathbf{hypo}(f) = \{(x,t) \mid t \leq f(x)\}$$

is a convex set.

- Consider the first-order condition for convexity:

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

where f is convex and $x, y \in \mathbf{dom} f$. We can interpret this basic inequality geometrically in terms of $\mathbf{epi}(f)$. If $(y,t) \in \mathbf{epi}(f)$, then

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

We can express this as

$$(y,t) \in \mathbf{epi}(f) \Longrightarrow \begin{bmatrix} \nabla f(x) \\ -1 \end{bmatrix}^\top \left(\begin{bmatrix} y \\ t \end{bmatrix} - \begin{bmatrix} x \\ f(x) \end{bmatrix} \right) \leq 0$$

This means that the hyperplane defined by $(\nabla f(x), -1)$ supports $\mathbf{epi}(f)$ at the boundary point $(x, f(x))$; see Figure 7 .

Epigraph IV

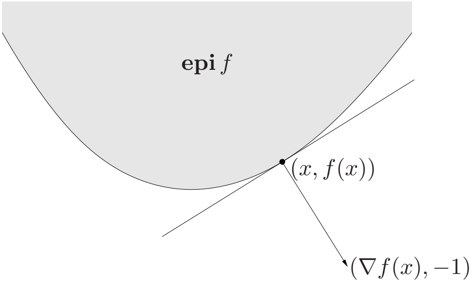

Fig. 7: For a differentiable convex function f, the vector $(\nabla f(x), -1)$ defines a supporting hyperplane to the epigraph of f at x.

Epigraph V

Definition 4 (Effective Domain)

The effective domain of a convex function f on $\mathcal{S}$, denoted by $\mathbf{dom} f$ is a projection of the epigraph of f on $\mathbb{R}^n$,

$$\mathbf{dom} f = \{x \mid \exists \mu, (x, \mu) \in \mathbf{epi}(f)\} = \{x \mid f(x) < +\infty\}$$

Definition 5 (Proper)

A convex function f is said to be proper if its epigraph is non-empty and contains no lines.

Eaxmples of Convex Functions I

We start with some functions on $\mathbb{R}$.

- Exponential. e^{ax} is convex on $\mathbb{R}$, for any $a \in \mathbb{R}$.
- Powers. x^a is convex on $\mathbb{R}_{++}$ when $a \geq 1$ or $a \leq 0$, and concave for $0 \leq a \leq 1$.
- Powers of absolute value. $|x|^p$, for $p \geq 1$, is convex on $\mathbb{R}$.
- Logarithm. $\log x$ is concave on $\mathbb{R}_{++}$.
- Negative entropy. $x \log x$ is convex on $\mathbb{R}_{++}$.

Convexity or concavity of these examples can be shown by verifying the basic inequality (Eq. 1), or by checking that the second derivative is nonnegative or nonpositive.

Eaxmples of Convex Functions II

We now provide some examples on $\mathbb{R}^n$.

- Norms. Every norm on $\mathbb{R}$ is convex. If $f : \mathbb{R}^n \to \mathbb{R}$ is a norm, and $0 \leq \theta \leq 1$, then

$$f(\theta x + (1-\theta)y) \leq f(\theta x) + f((1-\theta)y) = \theta f(x) + (1-\theta)f(y)$$

  The inequality follows from the triangle inequality, and the equality follows from homogeneity of a norm.

- Max function. $f(x) = \max x_1, x_2, ..., x_n$ is convex on $\mathbb{R}^n$. For $0 \leq \theta \leq 1$, we have

$$\begin{aligned} f(\theta x + (1-\theta)y) &= \max_i(\theta x_i + (1-\theta)y_i) \\ &\leq \theta \max_i x_i + (1-\theta)\max_i y_i \\ &= \theta f(x) + (1-\theta)f(y) \end{aligned}$$

Eaxmples of Convex Functions III

- Quadratic-over-linear function. The function $f(x, y) = x^2/y$, with

$$\mathbf{dom} f = \mathbb{R} \times \mathbb{R}_{++} = \left\{(x, y) \in \mathbb{R}^2 \mid y > 0\right\}$$

  is convex (Figure 8).

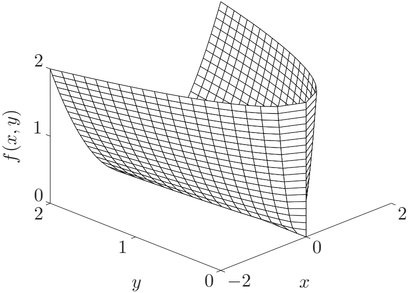

Fig. 8: Graph of $f(x, y) = x^2/y$.

We note that for $y > 0$,

$$\nabla^2 f(x,y) = \frac{2}{y^3}\begin{bmatrix} y^2 & -xy \\ -xy & x^2 \end{bmatrix} = \frac{2}{y^3}\begin{bmatrix} y \\ -x \end{bmatrix}\begin{bmatrix} y \\ -x \end{bmatrix}^\top \succeq 0$$

- Log-sum-exp. The function $f(x) = \log\left(e^{x_1} + \cdots + e^{x_n}\right)$ is convex on $\mathbb{R}^n$. This function can be interpreted as a differentiable (in fact, analytic) approximation of the max function, since

$$\max\{x_1, \cdots, x_n\} \leq f(x) \leq \max\{x_1, \cdots, x_n\} + \log n$$

Figure 9 shows f for $n = 2$. The Hessian of the log-sum-exp function is

$$\nabla^2 f(x) = \frac{1}{\left(\mathbf{1}^\top z\right)^2}\left(\left(\mathbf{1}^\top z\right)\mathbf{diag}(z) - zz^\top\right)$$

where $z = (e^{x_1}, \ldots, e^{x_n})$. To verify that $\nabla^2 f(x) \succeq 0$ we must show that for all v, $v^\top \nabla^2 f(x) v \geq 0$, i.e.,

$$v^\top \nabla^2 f(x) v = \frac{1}{\left(\mathbf{1}^\top z\right)^2}\left(\left(\sum_{i=1}^n z_i\right)\left(\sum_{i=1}^n v_i^2 z_i\right) - \left(\sum_{i=1}^n v_i z_i\right)^2\right) \geq 0.$$

But this follows from the Cauchy-Schwarz inequality $\left(a^\top a\right)\left(b^\top b\right) \geq \left(a^\top b\right)^2$ applied to the vectors with components $a_i = v_i\sqrt{z_i}, b_i = \sqrt{z_i}$.

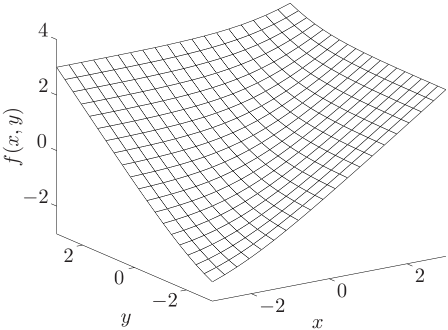

Fig. 9: Graph of $f(x,y) = \log\left(e^x + e^y\right)$.

Eaxmples of Convex Functions VI

- Geometric mean. The geometric mean

$$f(x)=\left(\prod_{i=1}^{n} x_i\right)^{1/n}$$

is concave on $\mathbf{dom} f=\mathbb{R}_{++}^{n}$. Its Hessian $\nabla^2 f(x)$ is given by

$$\frac{\partial^2 f(x)}{\partial x_k^2}=-(n-1)\frac{\left(\prod_{i=1}^{n} x_i\right)^{1/n}}{n^2 x_k^2}, \quad \frac{\partial^2 f(x)}{\partial x_k \partial x_l}=\frac{\left(\prod_{i=1}^{n} x_i\right)^{1/n}}{n^2 x_k x_l} \quad \text{for } k \neq l$$

and can be expressed as

$$\nabla^2 f(x)=-\frac{\prod_{i=1}^{n} x_i^{1/n}}{n^2}\left(n\,\mathbf{diag}\left(1/x_1^2,\ldots,1/x_n^2\right)-qq^{\top}\right)$$

where $q_i=1/x_i$. We must show that $\nabla^2 f(x) \preceq 0$, i.e., that

$$v^{\top} \nabla^2 f(x) v=-\frac{\prod_{i=1}^{n} x_i^{1/n}}{n^2}\left(n \sum_{i=1}^{n} v_i^2/x_i^2-\left(\sum_{i=1}^{n} v_i/x_i\right)^2\right) \leq 0$$

for all v. Again this follows from the Cauchy-Schwarz inequality $\left(a^{\top} a\right)\left(b^{\top} b\right) \geq \left(a^{T} b\right)^2$, applied to the vectors $a=\mathbf{1}$ and $b_i=v_i/x_i$.

Eaxmples of Convex Functions VII

- Matrix fractional function. The function $f : \mathbb{R}^n \times \mathcal{S}^n \to \mathbb{R}$, defined as

$$f(x, Y) = x^\top Y^{-1} x$$

is convex on $\mathbf{dom} f = \mathbb{R}^n \times \mathcal{S}^n_{++}$. (This generalizes the quadratic-over-linear function $f(x, y) = x^2/y$, with $\mathbf{dom} f = \mathbb{R} \times \mathbb{R}_{++}$). One easy way to establish convexity of f is via its epigraph:

$$\begin{aligned} \text{epi } f &= \left\{(x, Y, t) \mid Y \succ 0, x^\top Y^{-1} x \le t\right\} \\ &= \left\{(x, Y, t) \mid \begin{bmatrix} Y & x \\ x^\top & t \end{bmatrix} \succeq 0, Y \succ 0\right\} \end{aligned}$$

using the Schur complement condition for positive semidefiniteness of a block matrix. The last condition is a linear matrix inequality in (x, Y, t), and therefore $\mathbf{epi}(f)$ is convex.

For the special case $n = 1$, the matrix fractional function reduces to the quadraticover-linear function x^2/y, and the associated LMI representation is

$$\begin{bmatrix} y & x \\ x & t \end{bmatrix} \succeq 0, \quad y > 0$$

Contents

Nonnegative Weighted Sums

Proposition 1 (Nonnegative Weighted Sums)

- If f is convex, αf is convex for $\forall \alpha \geq 0$;
- A nonnegative weighted sum of convex functions,

$$f = w_1 f_1 + \cdots + w_m f_m, w_i \geq 0, i = 1, \cdots, m$$

  is convex. Similarly, a nonnegative weighted sum of concave functions is concave. A nonnegative, nonzero weighted sum of strictly convex (concave) functions is strictly convex (concave).
- These properties extend to infinite sums and integrals. If $f(x, y)$ is convex in x for each $y \in \mathcal{A}$, and $w(y) \geq 0$ for each $y \in \mathcal{A}$, then

$$g(x) = \int_{\mathcal{A}} w(y) f(x, y) \mathrm{d}y$$

  is convex in x (provided the integral exists).

Composition with Affine Functions

Property 5 (Composition with Affine Functions)

Suppose $f : \mathbb{R}^n \to \mathbb{R}, A \in \mathbb{R}^{n \times m}$, and $b \in \mathbb{R}^n$. Define $g : \mathbb{R}^m \to \mathbb{R}$ by

$$g(x) = f(Ax + b)$$

with $\mathbf{dom} g = \{x \mid Ax + b \in \mathbf{dom} f\}$. Then if f is convex, so is g; if f is concave, so is g.

- Log barrier for linear inequalities. $f(x) = -\sum_{i=1}^{m} \log\left(b_i - a_i^\top x\right), \mathbf{dom} f = \left\{x \mid a_i^\top x < b_i, \forall i\right\}, a_i \in \mathbb{R}^n, b_i \in \mathbb{R}, x_i \in \mathbb{R}^n$. f is convex.
- Norm of affine functions. $f(x) = \|Ax + b\|$

Property 6 (Pointwise Maximum and Supremum)

(1)

$$f(x) = \max_i \{f_1(x), f_2(x), \cdots, f_m(x)\}$$

is convex, if $f_i(x)$ is convex for all $i = \{1, 2, ..., m\}$, with $\mathbf{dom} f = \mathbf{dom} f_1 \cap \mathbf{dom} f_2 \cap \cdots \cap \mathbf{dom} f_m$.

(2) If $f(x, y)$ is convex in x for $\forall y \in \mathcal{A}$, then $g(x) = \sup_{y \in \mathcal{A}} f(x, y)$ is convex in x, where $\mathbf{dom} g = \left\{x \mid (x, y) \in \mathbf{dom} f, \forall y \in \mathcal{A}, \sup_{y \in \mathcal{A}} f(x, y) < \infty\right\}$. Note that $\mathcal{A}$ is not necessarily convex. Similarly, the pointwise infimum of a set of concave functions is a concave function.

- Piecewise-linear functions. The function

$$f(x) = \max\left\{a_1^\top x + b_1, \ldots, a_L^\top x + b_L\right\}$$

defines a piecewise-linear (or really, affine) function (with L or fewer regions). It is convex since it is the pointwise maximum of affine functions.

- Sum of r largest components. For $x \in \mathbb{R}^n$ we denote by $x_{[i]}$ the i-th largest component of x, i.e., $x_{[1]} \geq x_{[2]} \geq ... \geq x_{[n]}$. Then the function

$$f(x) = \sum_{i=1}^{r} x_{[i]}$$

i.e., the sum of the r largest elements of x, is a convex function. Let $A \in \mathbb{R}^{\begin{pmatrix} n \\ r \end{pmatrix} \times n}$ $\left(\begin{pmatrix} n \\ r \end{pmatrix} = \frac{n!}{r!(n-r)!} \right)$ where each row of A has r componenets equal to 1 , call others 0 , and no-two rows of A are identical. Then $f(x)$ can be represented as

$$f(x) = \max_i \left\{ a_i^\top x, i = 1, 2, \dots, \begin{pmatrix} n \\ r \end{pmatrix} \right\}$$

which is the point-wise maximum of $\frac{n!}{r!(n-r)!}$ linear functions, so it is convex.

Pointwise Maximum and Supremum III

- Support function of a set. Let $C \subseteq \mathbb{R}^n$, with $C \neq \emptyset$. The support function S_C associated with the set C is defined as

$$S_C(x) = \sup \left\{ x^\top y \mid y \in C \right\}$$

(and, naturally, $\mathbf{dom} S_C = \left\{ x \mid \sup_{y \in C} x^\top y < \infty \right\}$). For each $y \in C, x^\top y$ is a linear function of x, so S_C is the pointwise supremum of a family of linear functions, hence convex.

- Distance to farthest point of a set. Let $C \subseteq \mathbb{R}^n$. The distance (in any norm) to the farthest point of C,

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

is convex. For any y, the function $\|x - y\|$ is convex in x. Since f is the pointwise supremum of a family of convex functions (indexed by $y \in C$), it is a convex function of x.

- Maximum eigenvalue of a symmetric matrix. The function $f(X) = \lambda_{\max}(X)$, with $\mathbf{dom} f = \mathcal{S}^m$, is convex. To see this, we express f as

$$f(X) = \sup \left\{ y^\top X y \mid \|y\|_2 = 1 \right\},$$

  i.e., as the pointwise supremum of a family of linear functions of X (i.e., $y^\top X y$) indexed by $y \in \mathbb{R}^m$).

- Norm of a matrix. As a generalization suppose $\|\cdot\|_a$ and $\|\cdot\|_b$ are norms on $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively. The induced norm of a matrix $X \in \mathbb{R}^{p\times q}$ is defined as

$$\|X\|_{a,b} = \sup_{v\neq 0} \frac{\|Xv\|_a}{\|v\|_b}.$$

  (This reduces to the spectral norm when both norms are Euclidean.) The induced norm can be expressed as

$$\begin{aligned} \|X\|_{a,b} &= \sup \{\|Xv\|_a \mid \|v\|_b = 1\} \\ &= \sup \left\{ u^\top X v \mid \|u\|_{a*} = 1, \|v\|_b = 1 \right\}, \end{aligned}$$

where $\|\cdot\|_{a*}$ is the dual norm of $\|\cdot\|_a$, and we use the fact that

$$\|z\|_a = \sup\left\{u^\top z \mid \|u\|_{a*} = 1\right\}.$$

Since we have expressed $\|X\|_{a,b}$ as a supremum of linear functions of X, it is a convex function.

Composition I

We examine conditions on $h : \mathbb{R}^k \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^k$ that guarantee convexity or concavity of their composition $f = h \circ g : \mathbb{R}^n \to \mathbb{R}$, defined by

$$f(x) = h(g(x)), \quad \mathbf{dom} f = \{x \in \mathbf{dom} g \mid g(x) \in \mathbf{dom} h\}.$$

$\mathbf{K = 1}$

Property 7 (Scalar Composition)

$f(x) = h(g(x))$, where $g : \mathbb{R}^n \to \mathbb{R}, h : \mathbb{R} \to \mathbb{R}$.s Then

f is convex if h is convex and nondecreasing, and g is convex,

f is convex if h is convex and nonincreasing, and g is concave,

f is concave if h is concave and nondecreasing, and g is concave,

f is concave if h is concave and nonincreasing, and g is convex.

Composition II

- The second derivative of the composition function $f = h \circ g$ is given by

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

  Suppose that g is convex ($g'' \geq 0$) and h is convex and nondecreasing ($h'' \geq 0$ and $h'0$). Therefore $f'' \geq 0$, i.e., f is convex.

- It turns out that very similar composition rules hold in the general case $n > 1$, without assuming differentiability of h and g, or that $\mathbf{dom}g = \mathbb{R}^n$ and $\mathbf{dom}h = \mathbb{R}$:

$$\begin{aligned}
&f \text{ is convex if } h \text{ is convex, } \tilde{h} \text{ is nondecreasing, and } g \text{ is convex,}\\
&f \text{ is convex if } h \text{ is convex, } \tilde{h} \text{ is nonincreasing, and } g \text{ is concave,}\\
&f \text{ is concave if } h \text{ is concave, } \tilde{h} \text{ is nondecreasing, and } g \text{ is concave,}\\
&f \text{ is concave if } h \text{ is concave, } \tilde{h} \text{ is nonincreasing, and } g \text{ is convex.}
\end{aligned}$$

  where $\tilde{h}$ denotes the extended-value extension of the function h, which assigns the value ∞ ($-\infty$) to points not in $\mathbf{dom}h$ for h convex (concave).

Composition III

- To say that $\tilde{h}$ is nondecreasing means that for any $x, y \in \mathbb{R}$, with $x < y$, we have $\tilde{h}(x) \leq \tilde{h}(y)$. In particular, this means if $y \in \textbf{dom}h$, then $x \in \textbf{dom}h$. In other words, In other words, the domain of h extends infinitely in the negative direction; it is either $\mathbb{R}$, or an interval of the form (∞, a) or $(\infty, a]$. This is illustrated in Figure 10.

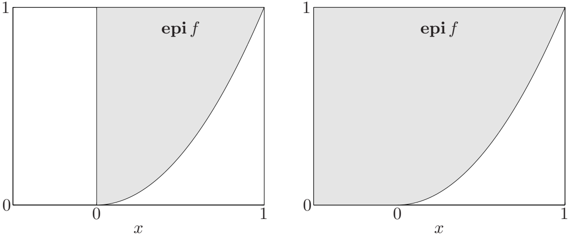

Fig. 10: Left. The function x^2, with domain $\mathbb{R}_+$, is convex and nondecreasing on its domain, but its extended-value extension is not nondecreasing. Right. The function $\max\{x, 0\}^2$, with domain $\mathbb{R}$, is convex, and its extended-value extension is nondecreasing.

Composition IV

Some examples of h in the composition rules.

- $h(x) = \log x$, with $\mathbf{dom} h = \mathbb{R}_{++}$, is concave and satisfies $\tilde{h}$ nondecreasing.
- $h(x) = x^{1/2}$, with $\mathbf{dom} h = \mathbb{R}_{+}$, is concave and satisfies $\tilde{h}$ nondecreasing.
- $h(x) = x^{3/2}$, with $\mathbf{dom} h = \mathbb{R}_{+}$ is convex but does not satisfy the condition $\tilde{h}$ nondecreasing. $\tilde{h}(-1) = \infty, \tilde{h}(1) = 1$.

Simple composition results.

- If g is convex then $\exp g(x)$ is convex.
- If g is concave and positive, then $\log g(x)$ is concave.
- If g is concave and positive, then $1/g(x)$ is convex.
- If g is convex and nonnegative and $p \geq 1$, then $g(x)^p$ is convex.
- If g is convex then $-\log(-g(x))$ is convex on $\{x \mid g(x) < 0\}$.

Composition V

Property 8 (Vector Composition)

Suppose

$$f(x) = h(g(x)) = h(g_1(x), ..., g_k(x))$$

where $g_i : \mathbb{R} \to \mathbb{R}, h : \mathbb{R}^k \to \mathbb{R}$. Then f is convex

if h is convex and nondecreasing in each argument, and g_i are convex,

if h is convex and nonincreasing in each argument, and g_i are concave,

f is concave

if h is concave and nondecreasing in each argument, and g_i are concave

- For the general results $g_i : \mathbb{R}^n \to \mathbb{R}$, the monotonicity condition on h must hold for the extended-value extension $\tilde{h}$.

Composition VI

- Let $h(z) = z_{[1]} + \cdots + z_{[r]}$, the sum of the r largest components of $z \in \mathbb{R}^k$. Then h is convex and nondecreasing in each argument. Suppose $g_1, \ldots, g_k$ are convex functions on $\mathbb{R}^n$. Then the composition function $f = h \circ g$, i.e., the pointwise sum of the r largest g_i 's, is convex.
- The function $h(z) = \log\left(\sum_{i=1}^k e^{z_i}\right)$ is convex and nondecreasing in each argument, so $\log\left(\sum_{i=1}^k e^{g_i}\right)$ is convex if g_i are convex.
- For $0 < p \leq 1$, the function $h(z) = \left(\sum_{i=1}^k z_i^p\right)^{1/p}$ on $\mathbb{R}_+^k$ is concave, and its extension (which has the value $-\infty$ for $z \not\succeq 0$) is nondecreasing in each component. So if g_i are concave and nonnegative, we conclude that $f(x) = \left(\sum_{i=1}^k g_i(x)^p\right)^{1/p}$ is concave.

Composition VII

- Suppose $p \geq 1$, and $g_1, \ldots, g_k$ are convex and nonnegative. Then the function $\left(\sum_{i=1}^k g_i(x)^p\right)^{1/p}$ is convex.
  To show this, we consider the function $h : \mathbb{R}^k \to \mathbb{R}$ defined as
  $$h(z) = \left(\sum_{i=1}^k \max\{z_i, 0\}^p\right)^{1/p}$$
  with $\mathbf{dom} h = \mathbb{R}^k$, so $h = \tilde{h}$. This function is convex, and nondecreasing, so we conclude $h(g(x))$ is a convex function of x. For $z \succeq 0$, we have $h(z) = \left(\sum_{i=1}^k z_i^p\right)^{1/p}$, so our conclusion is that $\left(\sum_{i=1}^k g_i(x)^p\right)^{1/p}$ is convex.
- The geometric mean $h(z) = \left(\prod_{i=1}^k z_i\right)^{1/k}$ on $\mathbb{R}_+^k$ is concave and its extension is nondecreasing in each argument. It follows that if $g_1, \ldots, g_k$ are nonnegative concave functions, then so is their geometric mean, $\left(\prod_{i=1}^k g_i\right)^{1/k}$.

Property 9 (Minimization)

If f is convex in (x, y), and C is a convex nonempty set, then the function

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

is convex in x, provided $g(x) > -\infty$ for all x. The domain of g is the projection of $\mathbf{dom} f$ on its x-coordinate, i.e.,

$$\mathbf{dom} g = \{x \mid (x, y) \in \mathbf{dom} f, \text{for some } y \in C\}$$

- Schur complement. Suppose the quadratic function

$$f(x, y) = x^\top Ax + 2x^\top By + y^\top Cy,$$

(where A and C are symmetric) is convex in (x, y), which means

$$\begin{bmatrix} A & B \\ B^\top & C \end{bmatrix} \succeq 0$$

Minimization II

We can express $g(x) = \inf_y f(x, y)$ as

$$g(x) = x^\top \left(A - BC^\dagger B^\top\right) x,$$

where $C^\dagger$ is the pseudo-inverse of C . By the minimization rule, g is convex, so we conclude that $A - BC^\dagger B^\top \succeq 0$.

If C is invertible, i.e., $C \succ 0$, then the matrix $A - BC^{-1}B^\top$ is called the *Schur complement* of C in the matrix

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$$

- Distance to a set. The distance of a point x to a set $S \subseteq \mathbb{R}^n$, in the norm $\|\cdot\|$, is defined as

$$\textbf{dist}(x, S) = \inf_{y \in S} \|x - y\|.$$

The function $\|x - y\|$ is convex in (x, y), so if the set S is convex, the distance function $\textbf{dist}(x, S)$ is a convex function of x.

Minimization III

- Suppose h is convex. Then the function g defined as

$$g(x) = \inf\{h(y) \mid Ay = x\}$$

is convex. To see this, we define f by

$$f(x, y) = \begin{cases} h(y) & \text{if } Ay = x \\ \infty & \text{otherwise} \end{cases}$$

which is convex in (x, y). Then g is the minimum of f over y, and hence is convex.

Property 10 (Perspective of a Function)

If $f : \mathbb{R}^n \to \mathbb{R}$, then the perspective of f is the function $g : \mathbb{R}^{n+1} \to \mathbb{R}$ defined by

$$g(x, t) = tf(x/t),$$

with domain

$$\textbf{dom} g = \{(x, t) \mid x/t \in \text{dom}\, f, t > 0\}.$$

The perspective operation preserves convexity: If f is a convex function, then so is its perspective function g. Similarly, if f is concave, then so is g.

- Euclidean norm squared. The perspective of the convex function $f(x) = x^\top x$ on $\mathbb{R}^n$ is

$$g(x, t) = t(x/t)^\top (x/t) = \frac{x^\top x}{t},$$

which is convex in (x, t) for $t > 0$.

Perspective of a Function II

- Negative logarithm. Consider the convex function $f(x) = -\log x$ on $\mathbb{R}_{++}$. Its perspective is

$$g(x,t) = -t\log(x/t) = t\log(t/x) = t\log t - t\log x,$$

  and is convex on $\mathbb{R}^2_{++}$. The function g is called the relative entropy of t and x. For $x = 1, g$ reduces to the negative entropy function.

- The relative entropy of two vectors $u, v \in \mathbb{R}^n_{++}$, defined as

$$\sum_{i=1}^{n} u_i \log\left(u_i/v_i\right)$$

  is convex in (u, v), since it is a sum of relative entropies of u_i, v_i.

- The Kullback-Leibler divergence between $u, v \in \mathbb{R}^n_{++}$, given by

$$D_{\mathrm{kl}}(u,v) = \sum_{i=1}^{n} \left(u_i \log\left(u_i/v_i\right) - u_i + v_i\right)$$

  is convex, since it is the relative entropy plus a linear function of (u, v).

Perspective of a Function III

- Example 3.20 Suppose $f : \mathbb{R}^m \to \mathbb{R}$ is convex, and $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, c \in \mathbb{R}^n$, and $d \in \mathbb{R}$. We define

$$g(x) = \left(c^\top x + d\right) f\left((Ax+b) / \left(c^\top x + d\right)\right)$$

with

$$\mathbf{dom} g = \left\{x \mid c^\top x + d > 0, (Ax+b) / \left(c^\top x + d\right) \in \operatorname{dom} f\right\}$$

Then g is convex.

Contents

Definition and Examples I

Definition 6 (Conjugate Function)

Let $f : \mathbb{R}^n \to \mathbb{R}$. The function $f^* : \mathbb{R}^n \to \mathbb{R}$, defined as

$$f^*(y) = \sup_{x \in \mathbf{dom} f} \left(y^\top x - f(x) \right)$$

is called the conjugate of the function f.

- The domain of the conjugate function consists of $y \in \mathbb{R}^n$ for which the supremum is finite, i.e., for which the difference $y^\top x - f(x)$ is bounded above on $\mathbf{dom} f$.
- The conjugate function $f^*(y)$ is convex in y, since it is the pointwise supremum of a family of convex (indeed, affine) functions of y.
- The conjugate of the conjugate of f is not necessarily f. If f is convex and closed, then $f^{**} = f$.

We derive the conjugates of some convex functions.

- Affine function. $f(x) = ax + b$. As a function of x, $yx - ax - b$ is bounded if and only if $y = a$, in which case it is constant. Therefore, the domain of the conjugate function f^* is the singleton $\{a\}$, and $f^*(a) = -b$.
- Negative logarithm. $f(x) = -\log x$, with $\mathbf{dom} f = \mathbb{R}_{++}$. The function $xy + \log x$ is unbounded above if $y \geq 0$ and reaches its maximum at $x = -1/y$ otherwise. Therefore, $\mathbf{dom} f^* = \{y \mid y < 0\} = -\mathbb{R}_{++}$and $f^*(y) = -\log(-y) - 1$ for $y < 0$.
- Exponential. $f(x) = e^x . xy - e^x$ is unbounded if $y < 0$. For $y > 0, xy - e^x$ reaches its maximum at $x = \log y$, so we have $f^*(y) = y \log y - y$. For $y = 0$, $f^*(y) = \sup_x -e^x = 0$. In summary, $\mathbf{dom} f^* = \mathbb{R}_+$and $f^*(y) = y \log y - y$ (with the interpretation $0 \log 0 = 0$).

- Negative entropy. $f(x) = x \log x$, with $\textbf{dom} f = \mathbb{R}_+$ (and $f(0) = 0$). The function $xy - x \log x$ is bounded above on $\mathbb{R}_+$for all y, hence $\textbf{dom} f^* = \mathbb{R}$. It attains its maximum at $x = e^{y-1}$, and substituting we find $f^*(y) = e^{y-1}$.
- Inverse. $f(x) = 1/x$ on $\mathbb{R}_{++}$. For $y > 0, yx - 1/x$ is unbounded above. For $y = 0$ this function has supremum 0 ; for $y < 0$ the supremum is attained at $x = (-y)^{-1/2}$. Therefore we have $f^*(y) = 2y^{1/2}$, with $\textbf{dom} f^* = -\mathbb{R}_+$.
- Strictly convex quadratic function. Consider $f(x) = \frac{1}{2} x^\top Q x$, with $Q \in \mathcal{S}^n_{++}$. The function $y^\top x - \frac{1}{2} x^\top Q x$ is bounded above as a function of x for all y. It attains its maximum at $x = Q^{-1} y$, so

$$f^*(y) = \frac{1}{2} y^\top Q^{-1} y.$$

Definition and Examples IV

- Log-determinant. We consider $f(X) = \log\det X^{-1}$ on $\mathcal{S}^n_{++}$ The conjugate function is defined as

$$f^*(Y) = \sup_{X \succ 0} (\operatorname{tr}(YX) + \log\det X)$$

since $\operatorname{tr}(YX)$ is the standard inner product on $\mathcal{S}^n$. We first show that $\operatorname{tr}(YX) + \log\det X$ is unbounded above unless $Y \prec 0$. If $Y \not\prec 0$, then Y has an eigenvector v, with $\|v\|_2 = 1$, and eigenvalue $\lambda \geq 0$. Taking $X = I + tvv^\top$ we find that

$$\operatorname{tr}(YX) + \log\det X = \operatorname{tr} Y + t\lambda + \log\det\left(I + tvv^T\right) = \operatorname{tr} Y + t\lambda + \log(1+t)$$

which is unbounded above as $t \to \infty$. Now consider the case $Y \prec 0$. We can find the maximizing X by setting the gradient with respect to X equal to zero:

$$\nabla_X(\operatorname{tr}(YX) + \log\det X) = Y + X^{-1} = 0$$

, which yields $X = -Y^{-1}$ (which is, indeed, positive definite). Therefore we have

$$f^*(Y) = \log\det(-Y)^{-1} - n,$$

with $\mathbf{dom} f^* = -\mathcal{S}^n_{++}$

- Indicator function. Let I_S be the indicator function of a (not necessarily convex) set $S \subseteq \mathbb{R}^n$, i.e., $I_S(x) = 0$ on $\operatorname{dom} I_S = S$. Its conjugate is

$$I_S^*(y) = \sup_{x \in S} y^\top x,$$

which is the support function of the set S.

Theorem 2 (Fenchel's Inequality)

From the definition of conjugate function, we immediately obtain the inequality

$$f(x) + f^*(y) \geq x^\top y$$

for all x, y. This is called Fenchel's inequality (or Young's inequality when f is differentiable).

$f(x) = \frac{1}{2}x^\top Qx$, where $Q \in \mathcal{S}^n_{++}$, we obtain the inequality

$$x^\top y \leq \frac{1}{2}x^\top Qx + \frac{1}{2}y^\top Q^{-1}y$$

The conjugate of a differentiable function f is also called the Legendre transform of f. Suppose f is convex and differentiable, with $\mathbf{dom} f = \mathbb{R}^n$. Any maximizer x^* of $y^\top x - f(x)$ satisfies $y = \nabla f(x^*)$, and conversely, if x^* satisfies $y = \nabla f(x^*)$, then x^* maximizes $y^\top x - f(x)$. Therefore, if $y = \nabla f(x^*)$, we have

$$f^*(y) = x^{*\top} \nabla f(x^*) - f(x^*).$$

This allows us to determine $f^*(y)$ for any y for which we can solve the gradient equation $y = \nabla f(z)$ for z.
We can express this another way. Let $z \in \mathbb{R}^n$ be arbitrary and define $y = \nabla f(z)$. Then we have

$$f^*(y) = z^\top \nabla f(z) - f(z).$$

- For $a > 0$ and $b \in \mathbb{R}$, the conjugate of $g(x) = af(x) + b$ is $g^*(y) = af^*(y/a) - b$.
- Suppose $A \in \mathbb{R}^{n \times n}$ is nonsingular and $b \in \mathbb{R}^n$. Then the conjugate of $g(x) = f(Ax + b)$ is
$$g^*(y) = f^*\left(A^{-\top} y\right) - b^\top A^{-\top} y,$$
with $\mathbf{dom} g^* = A^\top \mathbf{dom} f^*$.

Sums of Independent Functions

If $f(u,v) = f_1(u) + f_2(v)$, where f_1 and f_2 are convex functions with conjugates f_1^* and f_2^*, respectively, then

$$f^*(w,z) = f_1^*(w) + f_2^*(z).$$

In other words, the conjugate of the sum of independent convex functions is the sum of the conjugates. ('Independent' means they are functions of different variables.)

Contents

Definition and Examples I

Definition 7 (Quasiconvex, Quasiconcave and Quasilinear Functions)

1. A function $f : \mathbb{R}^n \to \mathbb{R}$ is called quasiconvex (or unimodal) if its domain and all its sublevel sets

$$S_\alpha = \{x \in \mathbf{dom} f \mid f(x) \leq \alpha\},$$

for $\alpha \in \mathbb{R}$, are convex.

2. A function is quasiconcave if $-f$ is quasiconvex, i.e., every superlevel set $\{x \mid f(x) \geq \alpha\}$ is convex.
3. If a function f is quasilinear, then its domain, and every level set $\{x \mid f(x) = \alpha\}$ is convex.

- Convex functions have convex sublevel sets, and so are quasiconvex. But the converse is not true.
- Logarithm. $\log x$ on $\mathbb{R}_{++}$ is quasiconvex and quasiconcave, hence quasilinear.
- Ceiling function. $\mathrm{ceil}(x) = \inf z \in \mathbb{Z} \mid z \geq x$ is quasiconvex and quasiconcave.
- These examples show that quasiconvex functions can be concave, or discontinuous.

Definition and Examples II

We now give some examples on $\mathbb{R}^n$.

- Length of a vector. We define the length of $x \in \mathbb{R}^n$ as the largest index of a nonzero component, i.e.,

$$f(x) = \max\{i \mid x_i \neq 0\}.$$

(We define the length of the zero vector to be zero.) This function is quasiconvex on $\mathbb{R}^n$, since its sublevel sets are subspaces:

$$f(x) \leq \alpha \Longleftrightarrow x_i = 0 \text{ for } i = \lfloor \alpha \rfloor + 1, \ldots, n.$$

- Consider $f : \mathbb{R}^2 \to \mathbb{R}$, with $\mathbf{dom} f = \mathbb{R}_+^2$ and $f(x_1, x_2) = x_1 x_2$. This function is neither convex nor concave since its Hessian

$$\nabla^2 f(x) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

is indefinite; it has one positive and one negative eigenvalue, The function f is quasiconcave, however, since the superlevel neta

$$\left\{x \in \mathbb{R}_+^2 \mid x_1 x_2 \geq \alpha\right\}$$

are convex sets for all α. (Note, however, that f is not quasiconcave on $\mathbb{R}^2$)

- Linear-fractional function. The function

$$f(x) = \frac{a^\top x + b}{c^\top x + d}$$

with dom $f = \left\{x \mid c^\top x + d > 0\right\}$, is quasiconvex, and quasiconcave, 1.e, quasilinear. Its $\alpha-$ sublevel set is

$$\begin{aligned} S_a &= \left\{x \mid c^\top x + d > 0, \left(a^\top x + b\right) / \left(c^\top x + d\right) \leq \alpha\right\} \\ &= \left\{x \mid c^\top x + d > 0, a^T x + b \leq \alpha \left(c^\top x + d\right)\right\} \end{aligned}$$

which is convex, since it is the intersection of an open halfspace and a closed halfspace. (The same method can be used to show its superlevel sets are convex.)

- Distance rutio function. Suppose $a, b \in \mathbb{R}^n$, and define

$$f(x)=\frac{\|x-a\|_2}{\|x-b\|_2},$$

i.e., the ratio of the Euclidean distance to a to the distance to b. Then f is quasiconvex on the halfepace $\{x \mid \|x-a\|_2 \leq \|x-b\|_2\}$. To see this, we consider the a-sublevel set of f, with $\alpha \leq 1$ since $f(x) \leq 1$ on the halfspace $\{x \mid \|x-a\|_2 \leq \|x-b\|_2\}$. This sublevel bet is the set of points satisfying

$$\|x-a\|_2 \leq \alpha\|x-b\|_2 .$$

Squaring both sides, and rearranging terms, we see that this is equivalent to

$$\left(1-a^2\right) x^{\top} x-2\left(a-a^2 b\right)^{\top} x+a^{\top} a-a^2 b^{\top} b \leq 0 .$$

This describes a convex set (in fact a Euclidean ball) if $a \leq 1$.

Definition and Examples V

- Intemal rate of return. Let $x = (x_0, x_1, \ldots, x_n)$ denote a cash flow sequence over n periods, where $x_i > 0$ means a payment to us in period i, and $x_i < 0$ means a payment by us in period i. We define the present value of a cash flow, with interest rate $r \geq 0$, to be

$$\mathrm{PV}(x, r) = \sum_{i=0}^{n} (1+r)^{-i} x_i$$

(The factor $(1+r)^{-i}$ is a discount factor for a payment by or to us in period i.) Now we consider cash flows for which $x_0 < 0$ and $x_0 + x_1 + \cdots + x_n > 0$. This means that we start with an investment of $|x_0|$ in period 0 , and that the total of the remaining cash flow, $x_1 + \cdots + x_n$, (not taking any discount factors into account) exceeds our initial investment.
For such a cash flow, $\mathrm{PV}(x, 0) > 0$ and $\mathrm{PV}(x, r) \to x_0 < 0$ as $r \to \infty$, so it follows that for at least one $r \geq 0$, we have $\mathrm{PV}(x, r) = 0$. We define the internal rate of return of the cash flow as the smallest interest rate $r \geq 0$ for which the present value is zero:

$$\mathrm{IRR}(x) = \inf\{r \geq 0 \mid \mathrm{PV}(x, r) = 0\}.$$

Internal rate of return is a quasiconcave function of x (restricted to $x_0 < 0, x_1 + \cdots + x_n > 0$). To see this, we note that

$$\operatorname{IRR}(x) \geq R \Longleftrightarrow \operatorname{PV}(x, r) > 0 \text{ for } 0 \leq r < R$$

The lefthand side defines the R-superlevel set of IRR. The righthand side is the intersection of the sets $\{x \mid \operatorname{PV}(x, r) > 0\}$, indexed by r, over the range $0 \leq r < R$. For each $r, \operatorname{PV}(x, r) > 0$ defines an open halfspace, so the righthand side defines a convex set.

Basic Properties I

Property 11 (Modified Jensen's Inequalities)

A function f is quasiconvex if and only if $\mathbf{dom} f$ is convex and for any $x, y \in \mathbf{dom} f$ and $0 \leq \theta \leq 1$,

$$f(\theta x + (1-\theta)x) \leq \max\{f(x), f(y)\}$$

i.e., the value of the function on a segment does not exceed the maximum of its values at the endpoints.

- Cardinality of a nonnegative vector. The cardinality or size of a vector $x \in \mathbb{R}^n$ is the number of nonzero components, and denoted $\text{card}(x)$. The function card is quasiconcave on $\mathbb{R}_+^n$ (but not $\mathbb{R}^n$). This follows immediately from the modified Jensen inequality

$$\text{card}(x+y) \geq \min\{\text{card}(x), \text{card}(y)\}$$

which holds for $x, y \succeq 0$.

Basic Properties II

- Rank of positive semidefinite matrix. The function rank X is quasiconcave on $\mathcal{S}_+^n$. This follows from the modified Jensen inequality (3.19),

$$\operatorname{rank}(X+Y) \geq \min\{\operatorname{rank} X, \operatorname{rank} Y\}$$

which holds for $X, Y \in \mathcal{S}_+^n$. (This can be considered an extension of the previous example, since $\operatorname{rank}(\operatorname{diag}(x)) = \operatorname{card}(x)$ for $x \succeq 0$.)

Property 12

$f : \mathbb{R}^n \to \mathbb{R}$ is quasiconvex if and only if the function $g : \mathbb{R} \to \mathbb{R}, g(t) = f(x+tv)$, $\mathbf{dom} g = \{t \mid x + tv \in \text{dom g}\}$ is quasiconvex in t.

Basic Properties III

Property 13 (Quasiconvex Functions on $\mathbb{R}$)

A continuous function $f : \mathbf{R} \to \mathbf{R}$ is quasiconvex if and only if at least one of the following conditions holds:

- f is nondecreasing
- f is nonincreasing
- there is a point $c \in \mathbf{dom} f$ such that for $t \leq c$ (and $t \in \mathbf{dom} f$), f is nonincreasing, and for $t \geq c$ (and $t \in \mathbf{dom} f$), f is nondecreasing.

The point c can be chosen as any point which is a global minimizer of f.

- $\sqrt{|x|}$ on $\mathbb{R}$ is quasiconvex.

First-Order Condition I

Property 14 (First-Order Condition)

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable. Then f is quasiconvex if and only if $\mathbf{dom} f$ is convex and for all $x, y \in \mathbf{dom} f$

$$f(y) \leq f(x) \Longrightarrow \nabla f(x)^\top (y - x) \leq 0.$$

- The condition states that $\nabla f(x)$ (when $\nabla f(x) \neq= 0$) defines a supporting hyperplane to the sublevel set $\{y \mid f(y) \leq f(x)\}$, at the point x, as illustrated in Figure 11.
- Note that if f is convex and $\nabla f(x) = 0$, then x is a global minimizer of f. But this statement is false for quasiconvex functions: it is possible that $\nabla f(x) = 0$, but x is not a global minimizer of f.
- The proof of the first-order condition is in Appendix 3.

First-Order Condition II

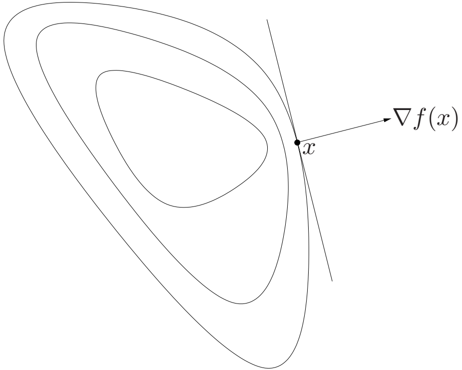

Fig. 11: Three level curves of a quasiconvex function f are shown. The vector $\nabla f(x)$ defines a supporting hyperplane to the sublevel set $\{z \mid f(z) \le f(x)\}$ at x.

Second-Order Condition I

Property 15 (Second-Order Conditions)

Now suppose f is twice differentiable. If f is quasiconvex, then for all $x \in$ $\mathbf{dom} f$, and all $y \in \mathbb{R}^n$, we have

$$y^\top \nabla f(x) = 0 \Longrightarrow y^\top \nabla^2 f(x) y \geq 0.$$

For a quasiconvex function on $\mathbb{R}$, this reduces to the simple condition

$$f'(x) = 0 \Longrightarrow f''(x) \geq 0,$$

i.e., at any point with zero slope, the second derivative is nonnegative.

Contents

Property 16 (Nonnegative Weighted Maximum)

A nonnegative weighted maximum of quasiconvex functions, i.e.,

$$f = \max\{w_1 f_1, \cdots, w_m f_m\}$$

with $w_i \geq 0$ and f_i quasiconvex, is quasiconvex. The property extends to the general pointwise supremum

$$f(x) = \sup_{y \in C} (w(y) g(x,y))$$

where $w(y) \geq 0$ and $g(x,y)$ is quasiconvex in x for each y.

- This can be easily verified: $f(x) \leq \alpha$ if and only if

$$w(y)g(x,y) \leq \alpha, \forall y \in C$$

i.e., the α-sublevel sets of f is the intersection of the α-sublevel sets of the functions $w(y)g(x,y)$ in the variable x.

Nonnegative Weighted Maximum II

- Generalized eigenvalue. The maximum generalized eigenvalue of a pair of symmetric matrices (X, Y), with $Y \succ 0$, is defined as

$$\lambda_{\max}(X, Y) = \sup_{u \neq 0} \frac{u^T X u}{u^T Y u} = \sup\{\lambda \mid \det(\lambda Y - X) = 0\}.$$

  This function is quasiconvex on $\mathbf{dom} f = \mathcal{S}^n \times \mathcal{S}^n_{++}$. To see this we consider the expression

$$\lambda_{\max}(X, Y) = \sup_{u \neq 0} \frac{u^T X u}{u^T Y u}.$$

  For each $u \neq 0$, the function $u^T X u / u^T Y u$ is linear-fractional in (X, Y), hence a quasiconvex function of (X, Y). We conclude that $\lambda_{\max}$ is quasiconvex, since it is the supremum of a family of quasiconvex functions.

Composition I

Property 17 (Composition)

If $g : \mathbb{R}^n \to \mathbb{R}$ is quasiconvex and $h : \mathbb{R} \to \mathbb{R}$ is nondecreasing, then $f = h \circ g$ is quasiconvex.

- The composition of a quasiconvex function with an affine or linear-fractional transformation yields a quasiconvex function.
- If f is quasiconvex, then $g(x) = f(Ax + b)$ is quasiconvex, and $\tilde{g}(x) = f\left((Ax+b)/\left(c^\top x + d\right)\right)$ is quasiconvex on the set

$$\left\{x \mid c^\top x + d > 0, (Ax+b)/\left(c^\top x + d\right) \in \mathbf{dom} f\right\}$$

Property 18 (Minimization)

If $f(x, y)$ is quasiconvex jointly in x and y and C is a convex set, then the function

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

is quasiconvex.

- We need to show that $\{x \mid g(x) \leq \alpha\}$ is convex for any $\alpha \in \mathbb{R}$.
- $g(x) \leq \alpha$ iff for any $\epsilon>0$ there exists a $y \in C$ with $f(x, y) \leq \alpha+\epsilon$.
- Now let x_1 and x_2 be two points in the α-sublevel set of g. Then for any $\epsilon>0$, there exists $y_1, y_2 \in C$ with

  $$f\left(x_1, y_1\right) \leq \alpha+\epsilon, \quad f\left(x_2, y_2\right) \leq \alpha+\epsilon$$

  and since f is quasiconvex in x and y, we also have

  $$f\left(\theta x_1+(1-\theta) x_2, \theta y_1+(1-\theta) y_2\right) \leq \alpha+\epsilon$$

  for $0 \leq \theta \leq 1$. Hence $g\left(\theta x_1+(1-\theta) x_2\right) \leq \alpha$, which proves that $\{x \mid g(x) \leq \alpha\}$ is convex.

Contents

Definitions and Examples I

Definition 8 (Log-Concave (Convex) Function)

A positive-valued function $f : \mathbb{R}^n \to \mathbb{R}$ is log-concave if $\log(f)$ is concave, that is

$$f(\theta x + (1-\theta)y) \geq f(x)^{\theta} f(y)^{1-\theta}, 0 \leq \theta \leq 1, x, y \in \mathbf{dom} f$$

Similarly, f is log-convex if $\log(f)$ is convex.

- e^h is convex if h is convex, so a log-convex function is convex.
- A nonnegative concave function is log-concave.
- A log-convex function is quasiconvex and a log-concave function is quasiconcave, since the logarithm is monotone increasing.

Definitions and Examples II

- Affine function. $f(x) = a^T x + b$ is log-concave on $\{x \mid a^T x + b > 0\}$.
- Powers. $f(x) = x^a$, on $\mathbb{R}_{++}$, is log-convex for $a \leq 0$, and log-concave for $a \geq 0$.
- Exponentials. $f(x) = e^{ax}$ is log-convex and log-concave.
- The cumulative distribution function of a Gaussian density,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du$$

  is log-concave (see Appendix 4).
- Gamma function. The Gamma function,

$$\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du$$

  is log-convex for $x \geq 1$ (see Appendix 4).
- Determinant. $\det X$ is log concave on $\mathcal{S}^n_{++}$.
- Determinant over trace. $\det X / \operatorname{tr} X$ is log concave on $\mathcal{S}^n_{++}$ (see Appendix 4).

Definitions and Examples III

Many common probability density functions are log-concave.

- Multivariate normal distribution

$$f(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} e^{-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)}$$

  where $\mu \in \mathbb{R}^n$ and $\Sigma \in \mathcal{S}^n_{++}$.

- Exponential distribution on $\mathbb{R}^n_+$

$$f(x) = \left(\prod_{i=1}^n \lambda_i\right) e^{-\lambda^\top x}$$

  where $\lambda \succ 0$.

Property 19 (Twice Differentiable Log-Concave (Convex) Function)

Suppose f is twice differentiable, with $\mathbf{dom} f$ convex, so

$$\nabla^2 \log f(x) = \frac{1}{f(x)} \nabla^2 f(x) - \frac{1}{f(x)^2} \nabla f(x) \nabla f(x)^\top.$$

We conclude that f is log-convex if and only if for all $x \in \mathbf{dom} f$,

$$f(x) \nabla^2 f(x) \succeq_{\mathcal{S}^n_+} \nabla f(x) \nabla f(x)^\top,$$

and log-concave if and only if for all $x \in \mathbf{dom} f$,

$$f(x) \nabla^2 f(x) \preceq_{\mathcal{S}^n_+} \nabla f(x) \nabla f(x)^\top.$$

Properties of Log-Concave Functions II

Property 20 (Multiplication, Addition and Integration)

1. If f and g are log-concave, then so is the pointwise product $h(x) = f(x)g(x)$, since $\log h(x) = \log f(x) + \log g(x)$, and $\log f$ and $\log g$ are convex functions.
2. The sum of log-concave functions is not, in general, log-concave. Log-convexity, however, is preserved under sums. Let f and g be logconvex functions, i.e., $F = \log f$ and $G = \log g$ are convex. From the composition rules for convex functions, it follows that

$$\log\left(\exp F + \exp G\right) = \log(f + g)$$

   is convex.
3. If $f(x, y)$ is log-convex in x for each $y \in C$ then

$$g(x) = \int_C f(x, y)\mathrm{d}y$$

   is log-convex.

- Laplace transform of a nonnegative function and the moment generating functions.
  Suppose $p : \mathbb{R}^n \to \mathbb{R}$ satisfies $p(x) \geq 0$ for all x. The Laplace transform of p

  $$P(z) = \int p(x) e^{-z^\top x} \mathrm{d}x,$$

  is log-convex on $\mathbb{R}^n$. (Here $\mathbf{dom} P$ is, naturally, $\{z \mid P(z) < \infty\}$.)
  Now suppose p is a density, i.e., satisfies $\int p(x) \mathrm{d}x = 1$. The function $M(z) = P(-z)$ is called the moment generating function of the density. It gets its name from the fact that the moments of the density can be found from the derivatives of the moment generating function, evaluated at $z = 0$, e.g.,

  $$\nabla M(0) = \mathbf{E} v, \quad \nabla^2 M(0) = \mathbf{E} v v^\top$$

  where v is a random variable with density p.

Properties of Log-Concave Functions IV

Property 21 (Integration of Log-Concave Functions)

In some special cases log-concavity is preserved by integration. If $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is log-concave, then

$$g(x) = \int f(x, y)\mathrm{d}y$$

is a log-concave function of x (on $\mathbb{R}^n$). The integration here is over $\mathbb{R}^m$.

- Marginal distributions of log-concave probability densities are log-concave.
- Log-concavity is closed under convolution. If f and g are log-concave on $\mathbb{R}^n$, then so is the convolution

  $$(f * g)(x) = \int f(x - y)g(y)\mathrm{d}y$$

  Note that $g(y)$ and $f(x - y)$ are log-concave in (x, y), hence the product $f(x - y)g(y)$ is.

- Suppose $C \subseteq \mathbb{R}^n$ is a convex set and w is a random vector in $\mathbf{R}^n$ with logconcave probability density p. Then the function

$$f(x) = \mathbf{prob}(x + w \in C)$$

is log-concave in x. To see this, express f as

$$f(x) = \int g(x+w)p(w)\mathrm{d}w$$

where g is defined as

$$g(u) = \begin{cases} 1 & u \in C \\ 0 & u \notin C \end{cases}$$

(which is log-concave) and apply the integration result.

Properties of Log-Concave Functions VI

- The cumulative distribution function of a probability density function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as

$$F(x) = \mathbf{prob}(w \preceq x) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f(z) \mathrm{d}z_1 \cdots \mathrm{d}z_n$$

where w is a random variable with density f. If f is log-concave, then F is log concave. For example, the cumulative distribution function of a Gaussian random variable,

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} \mathrm{d}t$$

is log-concave.

- Volume of polyhedron. Let $A \in \mathbb{R}^{m \times n}$. Define

$$P_u = \{x \in \mathbb{R}^n \mid Ax \preceq u\}.$$

Then its volume $\mathbf{vol} P_u$ is a log-concave function of u.
Note that the function

$$\Psi(x, u) = \begin{cases} 1 & Ax \preceq u \\ 0 & \text{otherwise} \end{cases}$$

is log-concave. By the integration result, we conclude that

$$\int \Psi(x,u)dx = \mathbf{vol} P_u$$

is log-concave.

- The distribution of the sum of two independent log-concave random variables is log-concave.

Contents

Monotonicity I

Definition 9 (Monotonicity with respect to a Generalized Inequality)

Suppose $K \subseteq \mathbb{R}^n$ is a proper cone with associated generalized inequality $\preceq$. A function $f : \mathbb{R}^n \to \mathbb{R}$ is called K-nondecreasing if

$$x \preceq_K y \Longrightarrow f(x) \leq f(y),$$

and K-increasing if

$$x \preceq_K y, x \neq y \Longrightarrow f(x) < f(y).$$

We define K-nonincreasing and K-decreasing functions in a similar way.

- Monotone vector functions. A function $f : \mathbb{R}^n \to \mathbb{R}$ is nondecreasing with respect to $\mathbb{R}^n_+$ if and only if

$$x_1 \leq y_1, \ldots, x_n \leq y_n \Longrightarrow f(x) \leq f(y)$$

for all x, y. This is the same as saying that f, when restricted to any component x_i (i.e., x_i is considered the variable while x_j for $j \neq i$ are fixed), is nondecreasing.

Monotonicity II

- Matrix monotone functions. A function $f : \mathcal{S}^n \to \mathbb{R}$ is called matrix monotone (increasing, decreasing) if it is monotone with respect to the positive semidefinite cone.
- $\operatorname{tr}(WX)$, where $W \in \mathcal{S}^n$, is matrix nondecreasing if $W \succeq 0$, and matrix increasing if $W \succ 0$ (it is matrix nonincreasing if $W \preceq 0$, and matrix decreasing if $W \prec 0$).
- $\operatorname{tr}\left(X^{-1}\right)$ is matrix decreasing on $\mathcal{S}_{++}^n$.
- $\det X$ is matrix increasing on $\mathcal{S}_{++}^n$, and matrix nondecreasing on $\mathcal{S}_{+}^n$.

Monotonicity III

Property 22 (Gradient Conditions for Monotonicity)

A differentiable function f, with convex domain, is K-nondecreasing if and only if

$$\nabla f(x) \succeq_{K^*} 0$$

for all $x \in \mathbf{dom} f$, where K^* is the dual cone of K. Note the graident must be nonnegative in the dual inequality.

If

$$\nabla f(x) \succ_{K^*} 0$$

for all $x \in \mathbf{dom} f$, then f is K-increasing. The converse is note true.

Convexity with respect to a Generalized Inequality I

Definition 10 (Convexity with respect to a Generalized Inequality)

Let $K \subseteq \mathbb{R}^m$ be a proper cone with associated generalized inequality $\prec_K$. $f : \mathbb{R}^n \to \mathbb{R}^m$ is K-convex if:

1. $\mathbf{dom} f$ is convex.
2. For all x, y and $0 \leq \theta \leq 1$,

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

The function is strictly K-convex if

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

for all $x \neq y$ and $0 < \theta < 1$. These definitions reduce to ordinary convexity and strict convexity when $m = 1$ and $K = \mathbb{R}_+$.

Convexity with respect to a Generalized Inequality II

- Convexity with respect to componentwise inequality. A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is convex with respect to componentwise inequality (i.e., the generalized inequality induced by $\mathbb{R}^m_+$) if and only if for all x, y and $0 \leq \theta \leq 1$,

$$f(\theta x + (1-\theta)y) \preceq \theta f(x) + (1-\theta)f(y),$$

  i.e., each component f_i is a convex function.

- Matrix convexity. Suppose f is a symmetric matrix valued function, i.e., $f : \mathbb{R}^n \to \mathcal{S}^m$. The function f is convex with respect to matrix inequality if

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

  for any x and y, and for $\theta \in [0,1]$. This is sometimes called matrix convexity. An equivalent definition is that the scalar function $z^\top f(x) z$ is convex for all vectors z. (This is often a good way to prove matrix convexity).

Convexity with respect to a Generalized Inequality III

- $f(X) = XX^\top$ where $X \in \mathbb{R}^{n\times m}$ is matrix convex, since for fixed z the function $z^\top XX^\top z = \left\|X^\top z\right\|_2^2$ is a convex quadratic function of (the components of) X. For the same reason, $f(X) = X^2$ is matrix convex on $\mathcal{S}^n$.
- X^p is matrix convex on $\mathcal{S}^n_{++}$ for $1 \le p \le 2$ or $-1 \le p \le 0$, and matrix concave for $0 \le p \le 1$.
- The function $f(X) = e^X$ is not matrix convex on $\mathcal{S}^n$, for $n \ge 2$.

Contents

Appendix 1: First-order Convexity Condition I

Appendix 2: Second-order Convexity Condition I

Appendix 3: FOC of Quasiconvex Functions I

Appendix 4: Log-Convexity of Several Functions I