Lecture 3: Convex Functions

Xiugang Wu

University of Delaware

Fall 2019

Outline

- Basic Properties and Examples
- Operations that Preserve Convexity
- The Conjugate Function
- Quasiconvex Functions
- Log-concave and Log-convex Functions
- Convexity with respect to Generalized Inequalities

Outline

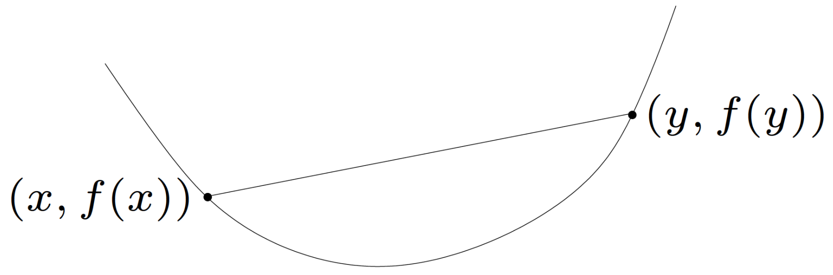
- Basic Properties and Examples
- Operations that Preserve Convexity
- The Conjugate Function
- Quasiconvex Functions
- Log-concave and Log-convex Functions
- Convexity with respect to Generalized Inequalities

Convex Function

A function $f : \mathbf{R}^n \to \mathbf{R}$ is convex if $\mathbf{dom}\ f$ is a convex set and

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

for all $x, y \in \mathbf{dom}\ f$ and $\theta \in [0,1]$



- f is strictly convex if the above holds with "$\leq$" replaced by "$<$"

- f is concave if $-f$ is convex

- affine functions are both convex and concave; conversely, if a function is both convex and concave, then it is affine

- f is convex iff it is convex when restricted to any line that intersects its domain, i.e. iff for all $x \in \mathbf{dom}\ f$ and all v, the function $g(t) = f(x + tv)$ on its domain $\{t | x + tv \in \mathbf{dom}\ f\}$ is convex

Extended-Value Extension

- The extended-value extension $\tilde{f}$ of f is defined by

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

- This simplifies notation as f is convex iff

$$\tilde{f}(\theta x + (1-\theta)y) \leq \theta \tilde{f}(x) + (1-\theta)\tilde{f}(y),\ \forall x, y \in \mathbf{R}^n,\ \theta \in [0,1]$$

- We can recover the domain of f from the extension by taking $\textbf{dom } f = \{x | \tilde{f}(x) < \infty\}$

First Order Condition

- f is differentiable if **dom** f is open and the gradient

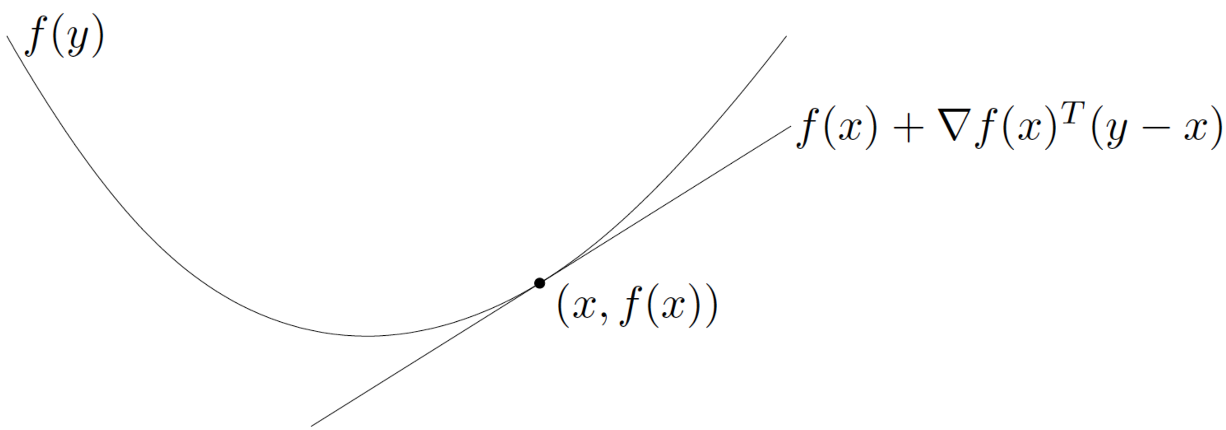
$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n} \right)$$

exists at each $x \in \textbf{dom } f$

- differentiable f with convex domain is convex iff

$$f(y) \geq f(x) + \nabla f(x)^T (y - x), \ \forall x, y \in \textbf{dom } f$$

- Note that the affine function $g(y) = f(x) + \nabla f(x)^T (y - x)$ is the first-order Taylor approximation of f near x. The above says that f is convex iff the first-order Taylor approximation is always an underestimator of f.



Second Order Condition

- f is twice differentiable if $\textbf{dom}\ f$ is open and the Hessian $\nabla^2 f(x) \in \mathbf{S}^n$,

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j},\ \ i,j \in [1:n]$$

exists at each $x \in \textbf{dom}\ f$

- twice differentiable f with convex domain is convex iff the Hessian $\nabla^2 f(x)$ is PSD for all $x \in \textbf{dom}\ f$

- twice differentiable f with convex domain is strictly convex if the Hessian $\nabla^2 f(x)$ is PD for all $x \in \textbf{dom}\ f$

- Quadratic function $f(x) = \frac{1}{2}x^T P x + q^T x + r$ is convex iff its Hessian P is PSD [Note that $\nabla(x^T P x) = 2Px$ and $\nabla^2(x^T P x) = 2P$]

- Least-squares objective $f(x) = \|Ax - b\|_2^2$ is convex because $\nabla f = 2A^T(Ax-b)$ and $\nabla^2 f = 2A^T A \succeq 0$ for any A

Examples in One Dimension

Convex

- affine: $ax + b$ on $\mathbf{R}$, for any $a, b \in \mathbf{R}$
- exponential: e^{ax} on $\mathbf{R}$, for any $a \in \mathbf{R}$
- powers: x^a on $\mathbf{R}_{++}$, for $a \geq 1$ or $a \leq 0$
- powers of absolute value: $|x|^a$ on $\mathbf{R}$, for $a \geq 1$
- negative entropy: $x \log x$ on $\mathbf{R}_{++}$

Concave

- affine: $ax + b$ on $\mathbf{R}$, for any $a, b \in \mathbf{R}$
- powers: x^a on $\mathbf{R}_{++}$, for $a \in [0, 1]$
- logarithm: $\log x$ on $\mathbf{R}_{++}$

Examples in High Dimensions

affine functions are convex and concave; all norms are convex

Examples on $\mathbf{R}^n$

- affine function: $f(x) = a^T x + b$

- norms: e.g., $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$ for $p \geq 1$; $\|x\|_\infty = \max_{i \in [1:n]} |x_i|$

Examples on $\mathbf{R}^{m \times n}$

- affine function:

$$f(X) = \mathbf{tr}(A^T X) + b = \sum_{i \in [1:m]} \sum_{j \in [1:n]} A_{ij} X_{ij} + b$$

- spectral (maximum singlar value) norm:

$$f(X) = \|X\|_2 = \sigma_{\max}(X) = (\lambda_{\max}(X^T X))^{1/2}$$

Examples in High Dimensions

- max function: $f(x) = \max_i x_i$ is convex

- Quadratic over linear function: $f(x, y) = x^2/y$ is convex for $y > 0$

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

- Log-sum-exponential: $f(x) = \log(e^{x_1} + \cdots + e^{x_n})$ is convex; also called softmax

- Geometric mean: $f(x) = (\prod_i x_i)^{1/n}$ is concave

- Log determinant: $f(x) = \log\det X, \mathbf{dom}\ f = \mathbf{S}^n_{++}$ is concave

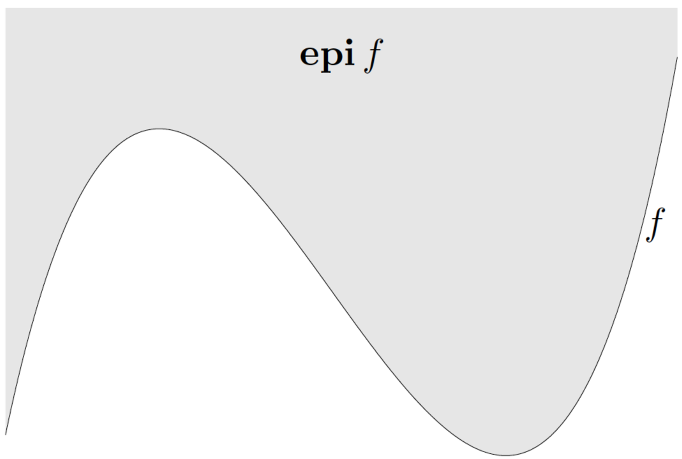
$$\begin{aligned} g(t) &= \log\det(Z + tV) \\ &= \log\det(Z^{1/2}(I + tZ^{-1/2}VZ^{-1/2})Z^{1/2}) \\ &= \log\det Z + \log\det(I + tZ^{-1/2}VZ^{-1/2}) \\ &= \log\det Z + \sum_{i=1}^{n} \log(1 + t\lambda_i) \end{aligned}$$

where $\lambda_i, i \in [1:n]$ are the eigenvalues of $Z^{-1/2}VZ^{-1/2}$

Sublevel Set and Epigraph

- The α-sublevel set of f is $\{x \in \textbf{dom} f | f(x) \leq \alpha\}$. If f is convex, then its sublevel set is convex. Converse is not true: $-e^x$ is concave, but its sublevel sets are convex.

- The graph of function f is defined as $\{(x, f(x) | x \in \textbf{dom } f)\}$, which is a subset in $\mathbf{R}^{n+1}$. The epigraph is defined as $\textbf{epi} f = \{(x, t) | x \in \textbf{dom } f, t \geq f(x)\}$. Function f is convex iff its epigraph is a convex set.



Jensen's Inequality

- If f is convex, then $f(\mathbf{E}x) \leq \mathbf{E}f(x)$ for any random variable x with $x \in \mathbf{dom}\ f$ w.p. 1. If f is non-convex, then there is a random variable x with $x \in \mathbf{dom} f$ w.p. 1, such that $f(\mathbf{E}x) > \mathbf{E}f(x)$.

- An intepretation: Consider convex function f. For any $x \in \mathbf{dom}\ f$ and random variable z with $\mathbf{E}z = 0$, we have

$$\mathbf{E}f(x+z) \geq f(\mathbf{E}[x+z]) = f(x),$$

which means that randomization or dithering (adding a zero mean random vector to the argument) only increases the value of convex function on average.

Outline

- Basic Properties and Examples
- Operations that Preserve Convexity
- The Conjugate Function
- Quasiconvex Functions
- Log-concave and Log-convex Functions
- Convexity with respect to Generalized Inequalities

A Calculus of Convex Functions

Practical methods for establishing convexity of a function:

- verify definition (often simplified by restricting to a line)

- for twice differentiable function, show its Hessian is PSD

- show that f is obtained from simple convex functions by operations that preserve convexity

    – nonnegative weighted sum
    – composition with affine function
    – pointwise maximum and supremum
    – composition
    – minimization
    – perspective

Positive Weighted Sum and Composition with Affine

Nonnegative weighted sum: If $f_1, \ldots, f_m$ are convex then $f = w_1 f_1 + \cdots + w_m f_m$ is also convex, where $w_1, \ldots, w_m$ are nonnegative

Composition with affine function: $f(Ax + b)$ is convex if f is convex

- log barrier for linear inequalities:

$$f(x) = -\sum_{i=1}^{m} \log(b_i - a_i^T x), \mathbf{dom}\ f = \{x | a_i^T x < b_i, i \in [1:m]\}$$

- any norm of affine function: $f(x) = \|Ax + b\|$

Pointwise Maximum

If $f_1, \ldots, f_m$ are convex then $f(x) = \max\{f_1(x), \ldots, f_m(x)\}$ is also convex

- piecewise-linear function: $f(x) = \max_{i \in [1:m]}(a_i^T x + b)$ is convex

- sum of r largest components of $x \in \mathbf{R}^n$:

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

is convex, where $x_{[i]}$ is the ith largest component of x; this is because

$$f(x) = \max_{I \subseteq [1:n], |I| = r} \sum_{i \in I} x_i$$

Pointwise Supremum

If $f(x, y)$ is convex in x for each $y \in A$ then

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

is also convex

- support function of a set C: $S_C(x) = \sup_{y \in C} y^T x$ is convex

- distance to farthest point in a set C:

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

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

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

Composition with Scalar Functions

Composition of $h : \mathbf{R} \to \mathbf{R}$ with scalar function $g : \mathbf{R}^n \to \mathbf{R}$:

$$f(x) = h(g(x))$$

f is convex if:

- g convex, h convex, $\tilde{h}$ nondecreasing
- g concave, h convex, $\tilde{h}$ nonincreasing

Remark:

- proof (for $n = 1$, differentiable g, h)

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

- monotonicity must hold for extended-value extension $\tilde{h}$

Examples:

- $\exp g(x)$ is convex if g is convex
- $1/g(x)$ is convex if g is concave and positive

Composition with Vector Functions

Composition of $h : \mathbf{R}^k \to \mathbf{R}$ with scalar function $g : \mathbf{R}^n \to \mathbf{R}^k$:

$$f(x) = h(g_1(x), g_2(x), \ldots, g_k(x))$$

f is convex if:

- g_i convex, h convex, $\tilde{h}$ nondecreasing in each argument
- g_i concave, h convex, $\tilde{h}$ nonincreasing in each argument

proof (for $n = 1$, differentiable g, h)

$$f''(x) = g'(x)^T \nabla^2 h(g(x)) g'(x) + \nabla h(g(x))^T g''(x)$$

Examples:

- $\sum_{i=1}^m \log g_i(x)$ is concave if g_i are concave and positive
- $\log \sum_{i=1}^m \exp g_i(x)$ is convex if g are convex

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 also convex

E.g., distance to a set: $d(x, S) = \inf_{y \in S} \|x - y\|$ is convex if S is convex [function $\|x - y\|$ is convex in (x, y) so if S is convex then $d(x, S)$ is convex in x]

Perspective

The perspective of $f : \mathbf{R}^n \to \mathbf{R}$ is the function $g : \mathbf{R}^{n+1} \to \mathbf{R}$ given by

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

with $\mathbf{dom} g = \{(x,t) \mid x/t \in \mathbf{dom} f, t > 0\}$. g is convex if f is convex.

E.g.: Euclidean norm squared. The perspective of the convex function $f(x) = x^T x$ is $g(x,t) = t(x/t)^T(x/t) = x^T x/t$, which is convex in (x,t) for $t > 0$

E.g.: Consider the convex function $f(x) = -\log x$. Its perspective is $g(x,t) = -t \log x/t = t \log t/x = t \log t - t \log x$ which is convex in (x,t). The function g is called the relative entropy of t and x. Therefore, the relative entropy of two vectors u, v defined as

$$\sum_i u_i \log(u_i/v_i)$$

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

Outline

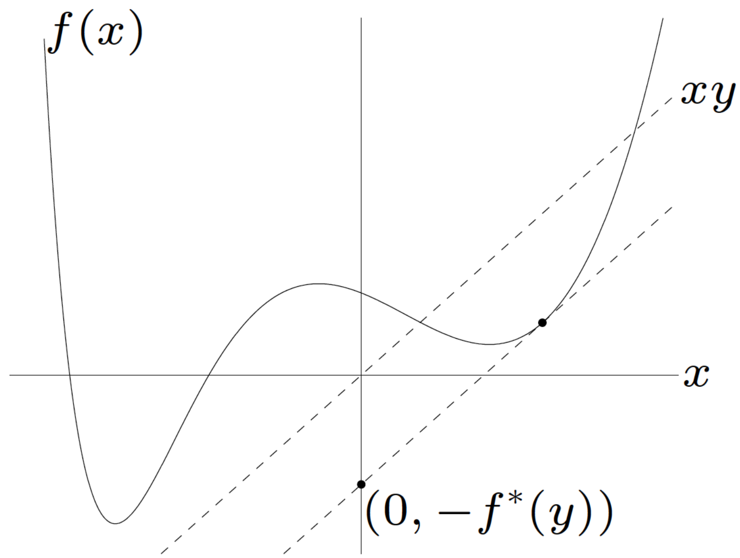
- Basic Properties and Examples
- Operations that Preserve Convexity
- The Conjugate Function
- Quasiconvex Functions
- Log-concave and Log-convex Functions
- Convexity with respect to Generalized Inequalities

Conjugate Function

For a function $f : \mathbf{R}^n \to \mathbf{R}$, its conjugate function $f^* : \mathbf{R}^n \to \mathbf{R}$ is defined as

$$f^*(y) = \sup_{x \in \mathbf{dom} f} (y^T x - f(x)).$$

- The domain of f^* consists of $y \in \mathbf{R}^n$ for which the supremum if finite
- f^* is convex (even if f is not)



Examples

- Affine functions: $f(x) = ax + b$. As a function of x, $yx - ax - b$ is bounded iff $y = a$, in which case it is constant. Therefore, the domain of the conjugate function f^* is the singleton $\{a\}$, and $f^*(a) = -b$

- Negative logarithm. $f(x) = -\log x$. The function $xy + \log x$ is unbounded for $y \geq 0$ and for $y < 0$, reaches its maximum $f^*(y) = -\log(-y) - 1$ at $x = -1/y$

- Strictly convex quadratic function. $f(x) = \frac{1}{2}x^T Q x$, with $Q \in \mathbf{S}^n_{++}$. The function $y^T x - \frac{1}{2}x^T Q x$ is bounded above as a function of x for all y, and attains the maximum at $x = Q^{-1}y$, so $f^*(y) = \frac{1}{2}y^T Q^{-1} y$

Outline

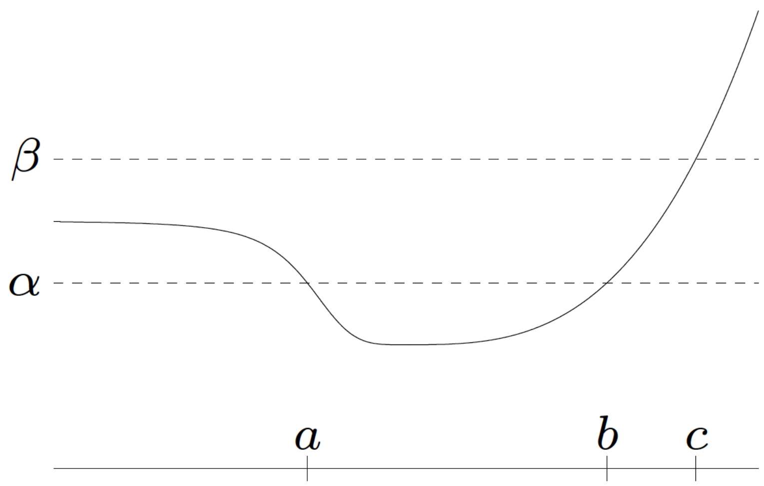
- Basic Properties and Examples
- Operations that Preserve Convexity
- The Conjugate Function
- Quasiconvex Functions
- Log-concave and Log-convex Functions
- Convexity with respect to Generalized Inequalities

Quasiconvex Functions

- A function f is quasiconvex if its domain and all its sublevel sets are convex

- f is quasiconcave if $-f$ is quasiconvex, i.e. all its superlevel sets are convex

- f is quasilinear if it is both quasiconvex and quasiconcave



Examples

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

- $\log x$ is quasilinear on $\mathbf{R}_{++}$

- $f(x_1, x_2) = x_1 x_2$ is quasiconcave on $\mathbf{R}^2_{++}$

- linear-fractional function

$$f(x) = \frac{a^T x + b}{c^T x + d}, \quad \mathbf{dom}\ f = \{x | c^T x + d > 0\}$$

is quasilinear

Properties

- Modified Jensen's inequality for quasiconvex function f: for any x, y in the domain and $\theta \in [0, 1]$,

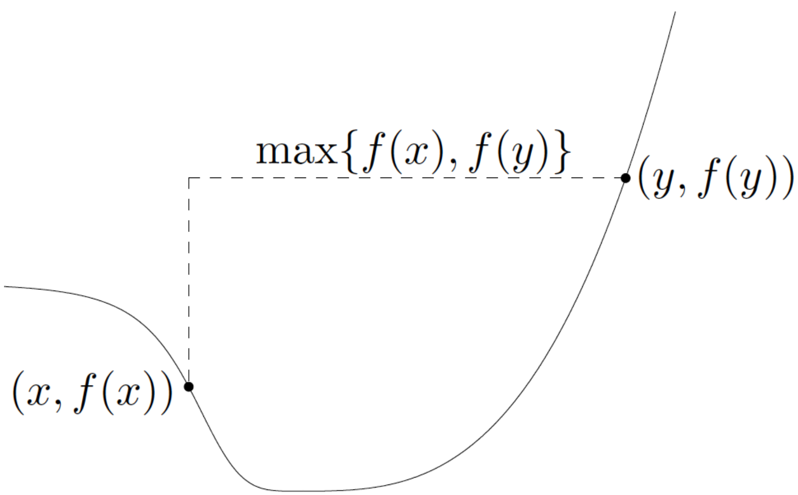
$$f(\theta x + (1-\theta)y) \leq \max\{f(x), f(y)\},$$

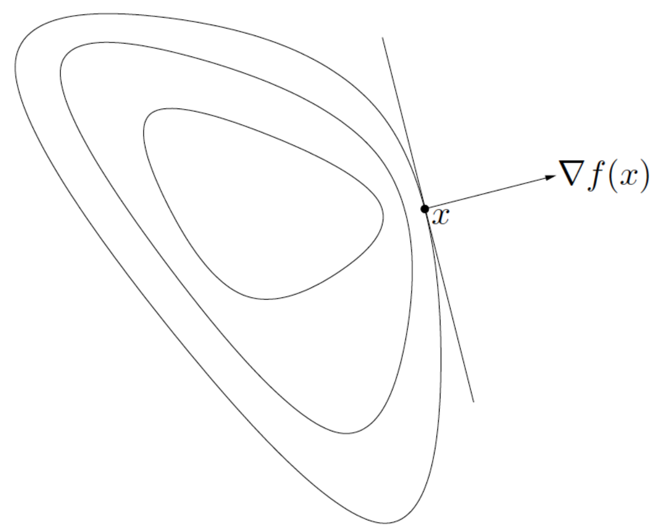
i.e. the value of the function on a segment doesn't exceed the maximum of the values at the endpoints.

- First-order condition: differentiable f with convex domain is quasiconvex iff

$$f(y) \leq f(x) \Rightarrow \nabla f(x)^T (y - x) \leq 0$$

- sums of quasiconvex functions are not necessarily quasiconvex





Outline

- Basic Properties and Examples
- Operations that Preserve Convexity
- The Conjugate Function
- Quasiconvex Functions
- Log-concave and Log-convex Functions
- Convexity with respect to Generalized Inequalities

Log-concave and Log-convex Functions

a positive function f is log-concave if $\log f$ is concave:

$$f(\theta x + (1-\theta)y) \geq f(x)^{\theta} f(y)^{1-\theta}, \forall \theta \in [0,1]$$

f is log-convex if $\log f$ is convex

- powers: x^a on $\mathbf{R}_{++}$ is log-convex for $a \leq 0$, log-concave for $a \geq 0$

- many common probability densities are log-concave, e.g., Gaussian:

$$f(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \exp\left(-\frac{1}{2}(x-\bar{x})^T \Sigma^{-1} (x-\bar{x})\right)$$

- cdf of Gaussian:

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

- determinant $\det X$

Properties

- twice differentiable f with convex domain is log-concave iff

$$f(x)\nabla^2 f(x) \preceq \nabla f(x)\nabla f(x)^T,\ \forall x \in \mathbf{dom} f$$

f is log-convex if $\log f$ is convex

- product of log-concave functions is log-concave

- sum of log-concave functions is not necessarily log-concave

- integration: if $f : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}$ is log-concave, then

$$g(x) = \int f(x, y)dy$$

is log-concave

Outline

- Basic Properties and Examples
- Operations that Preserve Convexity
- The Conjugate Function
- Quasiconvex Functions
- Log-concave and Log-convex Functions
- Convexity with respect to Generalized Inequalities

Convexity with respect to Generalized Inequalities

$f : \mathbf{R}^n \to \mathbf{R}^m$ is K-convex if **dom** f is convex and for any $x, y \in$ **dom** f and $\theta \in [0, 1]$,

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