# Convex functions

A supplementary note to Chapter 3 of *Convex Optimization* by S. Boyd and L. Vandenberghe

Optimization Lab.

IE department
Seoul National University

19th August 2009

## Definition

A function $f : \mathbb{R}^n \to \mathbb{R}$ is *convex* if $\mathrm{dom} f$ is convex and if for all $x, y \in \mathrm{dom} f$, and $0 \le \lambda \le 1$, we have

$$f((1-\lambda)x + \lambda y) \le (1-\lambda)f(x) + \lambda f(y).$$

- *Strictly convex* if strict inequality holds whenever $x \neq y$ and $0 < \lambda < 1$.
- We say $f$ is *concave* if $-f$ is convex. An affine function is both convex and concave.
- A function $f$ is convex if it is convex when restricted to any line intersecting its domain: for any $x \in \mathrm{dom} f$ and $v$, $g(x + tv)$ is convex on $\{t : x + tv \in \mathrm{dom} f\}$.

# Extended-value extensions

If $f$ is convex we define its extended-value extension,

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

With the extended reals, this can simplify notation, since we do not need to explicitly describe the domain.

**Example**

For a convex set $C$, its indicator function $\mathrm{I}_C$ is defined to be

$$\mathrm{I}_C(x) = \begin{cases} 0 & x \in C \\ \infty & x \notin C \end{cases}.$$

Suppose $\mathrm{dom} f = \mathbb{R}^n$. Then, $\min\{f(x) : x \in C\}$ is equivalent to minimizing $f + \mathrm{I}_C$.

# First-order conditions

**Theorem**

*Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable. Then $f$ is convex if and only if $\mathrm{dom} f$ is convex and*

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

**Proof for $n = 1$.** (Only if) Assume $f$ is convex and $x, y \in \mathrm{dom} f$. Since $\mathrm{dom} f$ is convex, we have for all $0 < \lambda \leq 1$, $x + \lambda(y - x) \in \mathrm{dom} f$, and by convexity of $f$, $f(x + \lambda(y - x)) \leq (1 - \lambda)f(x) + \lambda f(y)$.
Dividing both sides by $\lambda$, we obtain

$$f(y) \geq f(x) + \frac{f(x + \lambda(y - x)) - f(x)}{\lambda}.$$

Taking limit as $\lambda \to 0$, we get $f(y) \geq f(x) + f'(x)(y - x)$.

# First-order conditions (*cont'd*)

**Proof for** $n = 1$**.** (If) Choose any $x \neq y$ and $0 \leq \lambda \leq 1$, and let $z = \lambda x + (1-\lambda)y$. Then, by the above,

$$f(x) \geq f(z) + f'(z)(x-z), \qquad f(y) \geq f(z) + f'(z)(y-z).$$

Multiplying the first inequality by $\lambda$, the second by $1-\lambda$, and adding them yields

$$\lambda f(x) + (1-\lambda)f(y) \geq f(z) = f(\lambda x + (1-\lambda)y).$$

**Proof for** $n \geq 2$**.** Let $x, y \in \mathrm{dom} f$. Consider restriction of $f$ to the line through $x$ and $y$: $g(\lambda) := f(x + \lambda(y-x) = f((1-\lambda)x + \lambda y$, and apply the above case. $\square$

# Second-order conditions (*cont'd*)

**Proposition**

*Assume $f$ is twice differentiable on $\mathrm{dom} f$ which is open. Then $f$ is convex if and only if $\mathrm{dom} f$ is convex and its Hessian is positive semidefinite:* $\forall x \in \mathrm{dom} f$,

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

Remark that

> for $y \in \mathrm{dom} f$ and $z \in \mathbb{R}^n$, define $g(\lambda) := f(y + \lambda z)$. Then $g''(\lambda) = z^T \nabla^2 f(y + \lambda z) z$. Thus, $g''(\lambda) \geq 0$ on $\{\lambda | y + \lambda z \in \mathrm{dom} f\}$ if and only if $\nabla^2 f(x) \succeq 0 \ \forall x \in \mathrm{dom} f$.

Thus, it suffices to prove proposition on an open interval of the real line.

# Second-order conditions (*cont'd*)

**Case 1** $f : \mathbb{R} \to \mathbb{R}$
(Only if) If $f$ is convex, then $f(y) \geq f(x) + f'(x)(y-x)$ for all $x, y \in \mathrm{dom} f$, where $x < y$. Thus,

$$\frac{f(y) - f(x)}{y - x} \geq f'(x).$$

Taking limit as $x \to y$, we get $f'(y) \geq f'(x)$, which implies that $f'$ is monotone nondecreasing. Hence, $f''(x) \geq 0, \forall x \in \mathrm{dom} f$.
(If) For all $x, y \in \mathrm{dom} f$, there exists $z \in \mathrm{dom} f$ satisfying

$$f(y) = f(x) + f'(x)(y-x) + \frac{1}{2} f''(z)(y-x)^2 \geq f(x) + f'(x)(y-x).$$

The second inequality follows from the hypothesis. Hence $f$ is convex.

# Second-order conditions (*cont'd*)

**Case 2** $f : \mathbb{R}^n \to \mathbb{R}$
$f$ is convex if and only if $g(\lambda) = f(x + \lambda y)$ is convex on $\{\lambda | x + \lambda y \in \mathrm{dom} f\}$, $\forall x, y \in \mathrm{dom} f$. Then, by **Case 1**, the latter holds if and only if $g''(\lambda) \geq 0$ on $\{\lambda | x + \lambda y \in \mathrm{dom} f\}$:

$$\begin{aligned} g''(t) &= \frac{d}{dt} g'(t) = \frac{d}{dt}\left(\sum_{i=1}^{n} f_i'(x + ty) y_i\right) \\ &= \sum_{i=1}^{n} y_i \frac{d}{dt} f_i(x + \lambda y) = \sum_{i=1}^{n} y_i \nabla^2 f(x + \lambda y)_{i\cdot} y \\ &\geq y^T \nabla^2 f(x + \lambda y) y \geq 0, \end{aligned}$$

where $\nabla^2 f(x)_{i\cdot}$ is the $i$-th row of $\nabla^2 f(x)$. Therefore, $\nabla^2 f(x) \succeq 0$ for all $x \in \mathrm{dom} f$. $\square$

# Some simple examples

**Example**

- Exponential $e^{ax}$ is convex on $\mathbb{R}$ for $a \in \mathbb{R}$.
- Powers $x^a$ are convex on $\mathbb{R}_{++}$ for $a \geq 1$ 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 convex on $\mathbb{R}_{++}$.
- Negative entropy $x \log x$ is convex on $\mathbb{R}_{++}$. (Also on $\mathbb{R}_+$ if defined as $0$ for $x = 0$.)

# Norms

Every norm on $\mathbb{R}^n$ is convex.

**Proof** Remark that every norm function has the following properties:

- Positive homogeneity: $\|\lambda x\| = \lambda \|x\|$
- Triangle inequality: $\|x + y\| \leq \|x\| + \|y\|$
- Positive definiteness: $\|x\| = 0$ if and only if $x = 0$.

We will use triangle inequality and positive definiteness.
For $0 \leq \lambda \leq 1$, $\|\lambda x + (1 - \lambda)y)\| \leq \|\lambda x\| + \|(1 - \lambda)y\| = \lambda\|x\| + (1 - \lambda)\|y\|$. $\square$

# Max function

Max function, $f(x) = \max\{x_1, \ldots, x_n\}$ is convex on $\mathbb{R}^n$.

**Proof**

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

# Log-sum-exp

Log-sum-exp function $f(x) = \log(e^{x_1} + \cdots + e^{x_n})$ is convex on $\mathbb{R}^n$.

**Proof** The Hessian of the log-sum-exp function is

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

where $z = (e^{x_1}, \ldots, e^{x_n})$. We must show that for all $v$, $v^T \nabla^2 f(x) v \geq 0$, but

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

The inequality follows from the Cauchy-Schwarz inequality $(a^T a)(b^T b) \geq (a^T b)^2$ applied to $a_i = \sqrt{z_i}$ and $b_i = v_i\sqrt{z_i}$.

# Log-determinant

Log-determinant $f(X) = \log\det X$ is concave on $\mathrm{dom} f = S^n_{++}$.

**Proof** Consider restriction of $f$ to the line through $Z \in S^n_{++}$ to any direction $V \in S^n$:

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

where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of $Z^{-1/2}VZ^{-1/2}$.

$$g'(t) = \sum_{i=1}^{n} \frac{\lambda_i}{1 + t\lambda_i}, \quad g''(t) = -\sum_{i=1}^{n} \frac{\lambda_i^2}{(1 + t\lambda_i)^2}.$$

Since $g''(t) \leq 0$, we conclude that $f$ is concave.

# Sublevel sets and graphs

**Definition**

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

$$C_\alpha = \{x \in \mathrm{dom} f | f(x) \leq \alpha\}$$

Sublevel sets of a convex function are convex. (Converse is false.)

**Definition**

The *graph* of a function $f : \mathbb{R}^n \to \mathbb{R}$ is $\{(x, f(x)) | x \in \mathrm{dom} f\}$.
The *epigraph* of $f$ is $\mathrm{epi} f = \{(x, t) | x \in \mathrm{dom} f, f(x) \leq t\}$.
The *hypograph* of $f$ is $\mathrm{hyp} f = \{(x, t) | x \in \mathrm{dom} f, f(x) \geq t\}$.

A function is convex (concave) if and only if its epigraph (hypograph, resp.) is convex.

# Epigraph and convex function

Consider the first-order condition for convexity: $\forall\, x, y \in \mathrm{dom} f$, $f(y) \geq f(x) + \nabla f(x)^T(y - x)$. Thus, if $(y, t) \in \mathrm{epi} f$, then $t \geq f(y) \geq f(x) + \nabla f(x)^T(y - x)$. Hence $\nabla f(x)^T(y - x) - (t - f(x)) \leq 0$. Thus,

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

which means hyperplane in $\mathbb{R}^{n+1}$ defined by $(\nabla f(x), -1)$ supports $\mathrm{epi} f$ at the boundary point $(x, f(x))$.

# Extensions of Jensen's inequality

## Definition

Jensen's inequality: $\forall\ x, y \in \mathrm{dom} f$ and $0 \leq \lambda \leq 1$, $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$.

- To finite sums: $\forall\ x_1, \ldots, x_k \in \mathrm{dom} f$ and $\forall\ \lambda_1, \ldots, \lambda_k$ with $\lambda_i \geq 0$ and $\sum_i \lambda_i = 1$, we have

$$f(\lambda_1 x_1 + \ldots + \lambda_k x_k) \quad \leq \quad \lambda_1 f(x_1) + \cdots + \lambda_k f(x_k).$$

- To infinite sums:
- To integrals: $\forall p \geq 0$ such that $\int_S p = 1$ with $S \subseteq \mathrm{dom} f$,

$$f\left(\int_S p(x) x dx\right) \quad \leq \quad \int_S f(x) p(x) dx.$$

- To prob. measures: Let $x$ be a random variable with support in $\mathrm{dom} f$. Then, $f$ is convex if and only if $\forall$ probability measures of $x$ such that expectations exist, $f(\mathrm{E} x) \leq \mathrm{E} f(x)$.

# Hölder's inequality from Jensen's inequality

For $p > 1$, $1/p + 1/q = 1$, and $x, y \in \mathbb{R}^n$

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

**Proof** From convexity of $-\log x$, for $a, b \ge 0$ and $0 \le \lambda \le 1$, we can get,

$$a^\lambda b^{1-\lambda} \le \lambda a + (1-\lambda) b.$$

Applying this with $a = \frac{|x_i|^p}{\sum_j |x_j|^p}$, $b = \frac{|y_i|^q}{\sum_j |y_j|^q}$, and $\lambda = 1/p$ yields

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

Summing over $i$ yields the inequality. $\square$

# Nonnegative weighted sums

Convexity is preserved under nonnegative scaling.

**Proof** If $w \geq 0$ and $f$ is convex, we have

$$\mathrm{epi}(wf) = \begin{bmatrix} I & 0 \\ 0 & w \end{bmatrix} \mathrm{epi} f,$$

which is convex because the image of a convex set under a linear mapping is convex. $\square$

If $f_1, \ldots, f_m$ are convex functions, then $\forall\ w_i \geq 0$, $i = 1, \ldots, m$, $f = w_1 f_1 + \cdots + w_m f_m$ is convex.

# Composition with an affine mapping

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

# Pointwise maximum and supremum

If $f_1$ and $f_2$ are convex functions, then so is their *pointwise maximum*,

$$f(x) = \max\{f_1(x), f_2(x)\} \text{ with } \mathrm{dom} f = \mathrm{dom} f_1 \cap \mathrm{dom} f_2.$$

**Proof** $0 \le \lambda \le 1$ and $x, y \in \mathrm{dom} f$,

$$\begin{aligned}
f(\lambda x + (1-\lambda)y) &= \max\{f_1(\lambda x + (1-\lambda)y), f_2(\lambda x + (1-\lambda)y)\} \\
&\le \max\{\lambda f_1(x) + (1-\lambda)f_1(y), \lambda f_2(x) + (1-\lambda)f_2(y)\} \\
&\le \max\{\lambda f_1(x), \lambda f_2(x)\} + \max\{(1-\lambda)f_1(y), (1-\lambda)f_2(y)\} \\
&= \lambda \max\{f_1(x), f_2(x)\} + (1-\lambda)\max\{f_1(y), f_2(y)\} \\
&= \lambda f(x) + (1-\lambda)f(y). \ \square
\end{aligned}$$

# Pointwise maximum and supremum

If for each $y \in \mathcal{A}$, $f(x,y)$ is convex in $x$, then the function $g$, defined as

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

is convex in $x$. $(\mathrm{dom}g = \{x|\ (x,y) \in \mathrm{dom}f\ \forall y \in \mathcal{A},\ \sup_{y \in \mathcal{A}} f(x,y) < \infty\})$

## Application

- Support function of a set, $S_C(x) = \sup\{x^T y | y \in C\}$ is convex.
- Distance to farthest point of a set, $f(x) = \sup_{y \in C} \|x - y\|$ is convex.
- Least-squares as function of weights $g(w) = \inf_x \sum_{i=1}^n w_i(a_i^T x - b_i)^2$ with $\mathrm{dom}g = \{w|\inf_x \sum_{i=1}^n w_i(a_i^T x - b_i)^2 > -\infty\}$.
- Max eigenvalue of symm matrices $f(X) = \sup\{y^T X y|\ \|y\|_2 = 1\}$.
- Norm of a matrix

# Convex as pointwise affine supremum

If $f : \mathbb{R}^n \to \mathbb{R}$ is convex, with $\mathrm{dom} f = \mathbb{R}^n$, then we have

$$f(x) = \sup\{g(x) | g \text{ affine}, g(z) \leq f(z) \text{ for all } z\}.$$

**Proof** $(\geq)$ The inequality $\geq$ is clear.
$(\leq)$ For any $x$ we can find a supporting hyperplane of $\mathrm{epi} f$ at $(x, f(x))$:
$a \in \mathbb{R}^n$ and $b \in \mathbb{R}$ with $(a, b) \neq 0$ such that $\forall\ (z, t) \in \mathrm{epi} f$,

$$\begin{bmatrix} a \\ b \end{bmatrix}^T \begin{bmatrix} x - z \\ f(x) - t \end{bmatrix} \leq 0. \text{ Or, } \ a^T(x - z) + b(f(x) - f(z) - s) \leq 0,$$

for all $z \in \mathrm{dom} f = \mathbb{R}^n$ and all $s \geq 0$. This implies $b > 0$ as easily seen. Therefore,

$$g(z) = f(x) + (a/b)^T(x - z) \leq f(z)$$

for all $z$. The function $g$ is an affine underestimator of $f$ and satisfies $g(x) = f(x)$. $\square$

# Chain Rule: Review

Consider a twice differentiable $f : \mathbb{R}^n \to \mathbb{R}^m$ whose $\mathrm{dom} f$ is assumed to be open for simplicity.

- For $m = 1$, the *derivative* $Df : \mathbb{R}^n \to \mathbb{R}$ of $f$ at $x$ is defined to be

$$Df(x) = \left[\; D_1 f(x) \cdots D_n f(x) \;\right].$$

  A linear transformation from $\mathbb{R}^n$ to $\mathbb{R}$ which linearly approximates $f$ at $x$.

- For $m \geq 2$, the *derivative* of $f$ at $x$ is defined to be

$$Df(x) = \begin{bmatrix} Df_1(x) \\ \vdots \\ Df_m(x) \end{bmatrix}.$$

  A linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ which linearly approximates $f$ at $x$.

## Chain Rule: Review(*cont'd*)

- For $m = 1$, we define the *gradient* of $f$ is a column-wise representation of its derivative:

$$\nabla f(x) = \begin{bmatrix} D_1 f(x) \\ \vdots \\ D_n f(x) \end{bmatrix}.$$

  Thus, $\nabla f(x)$ is a function from $\mathbb{R}^n \to \mathbb{R}^n$.

- For $m = 1$, the *Hessian* $\nabla^2 f(x)$ of $f$ is defined to be the derivative of the gradient $\nabla f$

$$\nabla^2 f(x) = \begin{bmatrix} D_{11} f(x) & \cdots & D_{1n} f(x) \\ \vdots & \ddots & \vdots \\ D_{n1} f(x) & \cdots & D_{nn} f(x) \end{bmatrix}.$$

# Chain Rule: Review(*cont'd*)

Suppose that $h : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $x \in \mathrm{dom} h$, and that $g : \mathbb{R}^m \to \mathbb{R}^p$ is differentiable at $h(x) \in \mathrm{dom} g$. (Assume domains are open.) Let $f := g \circ h : \mathbb{R}^n \to \mathbb{R}^p$ by $(g \circ h)(x) = g(h(x))$. Then, $f$ is differentiable at $x$ and its derivative is

$$Df(x) = D(g \circ h)(x) = Dg(h(x))Dh(x).$$

# Hessian of composition

When $p = 1$, we have

$$Df(x) = Dg(h(x))Dh(x) = \begin{bmatrix} \nabla g(h(x))^T \end{bmatrix} \begin{bmatrix} \nabla h_1(x)^T \\ \vdots \\ \nabla h_m(x)^T \end{bmatrix}.$$

$$\begin{aligned} \text{Hence, } \nabla^2 f(x) &= D(\nabla f(x)) = D(Df(x)^T) \\ &= D\left( \begin{bmatrix} | & & | \\ \nabla h_1(x) & \cdots & \nabla h_m(x) \\ | & & | \end{bmatrix} \begin{bmatrix} | \\ \nabla g(h(x)) \\ | \end{bmatrix} \right). \end{aligned}$$

Let $D_j h(x) := \begin{bmatrix} D_j h_1(x) & \cdots & D_j h_m(x) \end{bmatrix}^T$, $j = 1, \ldots, n$. Then,

$$\nabla^2 f(x) = D\left( \begin{bmatrix} D_1 h(x)^T \\ \vdots \\ D_n h(x)^T \end{bmatrix} \begin{bmatrix} | \\ \nabla g(h(x)) \\ | \end{bmatrix} \right) = D\left( \begin{bmatrix} D_1 h(x)^T \nabla g(h(x)) \\ \vdots \\ D_n h(x)^T \nabla g(h(x)) \end{bmatrix} \right).$$

# Hessian of composition(*cont'd*)

The following holds for vector-valued functions, $a, b$: $\mathbb{R}^{\ltimes} \rightarrow \mathbb{R}^m$.

$$
\begin{aligned}
Da(x)^T b(x) &= D\left(\sum_{j=1}^{m} a_j(x) b_j(x)\right) = \sum_{j=1}^{m} D\left(a_j(x) b_j(x)\right) \\
&= \sum_{j=1}^{m} \left(b_j(x) \nabla a_j(x)^T + a_j(x) \nabla b_j(x)^T\right) \\
&= b(x)^T \begin{bmatrix} \nabla a_1(x)^T \\ \vdots \\ \nabla a_n(x)^T \end{bmatrix} + a(x)^T \begin{bmatrix} \nabla b_1(x)^T \\ \vdots \\ \nabla b_n(x)^T \end{bmatrix} \\
&= b(x)^T Da(x) + a(x)^T Db(x).
\end{aligned}
$$

# Hessian of composition(*cont'd*)

Therefore, taking $a(x) = D_j h(x)$ and $b(x) = \nabla g(h(x))$, for $j = 1, \ldots, n$, we have

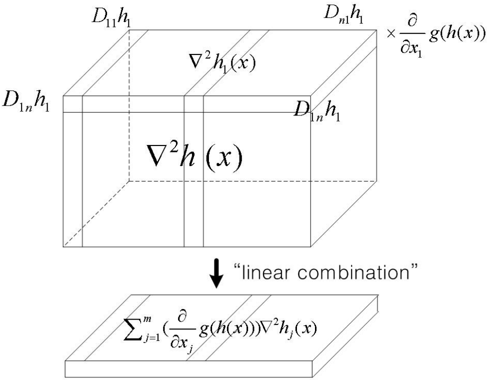
$$\begin{aligned} D(D_j h(x)^T \nabla g(h(x))) &= \nabla g(h(x))^T D(D_j h(x)) + D_j h(x)^T D(\nabla g(h(x))) \\ &= Dg(h(x)) D(D_j h(x)) + D_j h(x)^T \nabla^2 (g(h(x))) Dh(x). \end{aligned}$$

Hence,

$$\begin{aligned} \nabla^2 f(x) &= \begin{bmatrix} Dg(h(x)) D(D_1 h(x)) \\ \vdots \\ Dg(h(x)) D(D_n h(x)) \end{bmatrix} + Dh(x)^T \nabla^2 g(h(x)) Dh(x) \\ &:= Dg(h(x)) \nabla^2 h(x) + Dh(x)^T \nabla^2 g(h(x)) Dh(x). \end{aligned}$$

# Hessian of composition(*cont'd*)

We understand $\nabla^2 h(x)$ is a '3D' $m \times n \times n$ matrix whose $(k, i, j)$-th element is $D_{ij} h_k(x)$ and that $Dg(h(x))\nabla^2 h(x)$ is the linear combination of the $1 \times n \times n$ matrices $D_{ij} h_k(x)$ for fixed $k$'s with corresponding coefficients $D_k g(h(x))$'s.



# Hessian of composition(*cont'd*)

- The previous slides are rather for a mathematical practice.
- For the convexity conditions of composition, it suffices to consider one-dimensional cases: $n = 1$ and $m = 1$. Assume $g$, $h$ twice differ'ble, $\mathrm{dom} g = \mathrm{dom} h = \mathbb{R}^n$.

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

  - $g$ convex, nondecreasing, $h$ convex $\Rightarrow$ $f$ convex,
  - $g$ convex, nonincreasing, $h$ concave $\Rightarrow$ $f$ convex,
  - $g$ concave, nondecreasing, $h$ concave $\Rightarrow$ $f$ concave,
  - $g$ concave, nonincreasing, $h$ convex $\Rightarrow$ $f$ concave.

# Composition(*cont'd*)

**Example**

- $g(x) = \log(x)$, then $g$ concave, $\tilde{g}$ nondecreasing
- $g(x) = x^{1/2}$, then $g$ concave, $\tilde{g}$ nondecreasing
- $g(x) = x^{3/2}$, then $g$ convex, $\tilde{g}$ not nondecreasing
- $g(x) = x^{3/2}$ for $x \geq 0$, $= 0$ for $x < 0$ then $g$ convex, $\tilde{g}$ nondecreasing.

- In general,
  - $g$ convex, $\tilde{g}$ nondecreasing, $h$ convex $\Rightarrow$ $f$ convex,
  - $g$ convex, $\tilde{g}$ nonincreasing, $h$ concave $\Rightarrow$ $f$ convex,
  - $g$ concave, $\tilde{g}$ nondecreasing, $h$ concave $\Rightarrow$ $f$ concave,
  - $g$ concave, $\tilde{g}$ nonincreasing, $h$ convex $\Rightarrow$ $f$ concave.

# Composition(*cont'd*)

**Proposition**

$g$ *convex*, $\tilde{g}$ *nondecreasing*, $h$ *convex* $\Rightarrow f$ *convex.*

**Proof**: $\square$

# Composition(*cont'd*)

## Example

- $h$ convex $\Rightarrow \exp h$ convex.
- $h$ concave, positive $\Rightarrow \log h$ concave.
- $h$ concave, positive $\Rightarrow 1/h(x)$ concave.
- $h$ convex, nonnegative, and $p \geq 1 \Rightarrow h(x)^p$ convex.
- $h$ convex $\Rightarrow -\log(-g(x))$ convex on $\{x|g(x) < 0\}$.

# Composition(*cont'd*)

Consider $g : \mathbb{R}^m \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}^m$ with $\mathrm{dom} g = \mathbb{R}^m$ $\mathrm{dom} h = \mathbb{R}$.

$$\nabla^2 f(x) = Dg(h(x))\nabla^2 h(x) + Dh(x)^T \nabla^2 g(h(x)) Dh(x).$$

- $g$ convex, $\tilde{g}$ nondecreasing in each argument, $h_i$ convex $\Rightarrow$ $f$ convex,
- $g$ convex, $\tilde{g}$ nonincreasing in each argument, $h_i$ concave $\Rightarrow$ $f$ convex,
- $g$ concave, $\tilde{g}$ nondecreasing in each argument, $h_i$ concave $\Rightarrow$ $f$ concave.

# Composition(*cont'd*)

## Example

- $g(z) = z_{[1]} + \cdots + z_{[r]}$, sum of $r$ largest components of $z \in \mathbb{R}^m$. Then $g$ is convex and nondecreasing in each $z_i$. Therefore, if $h_1, \ldots, h_m$ convex functions on $\mathbb{R}^n$, $f := g \circ h$ is convex.
- $g(z) = \log(\sum_{i=1}^m e^{z_i})$ is convex and nondecreasing in each $z_i$. Hence if $h_i$ are convex, so is $g \circ h$.
- For $0 < p \leq 1$, $g(z) = (\sum_{i=1}^m z_i^p)^{1/p}$ is concave and its extension is nondecreasing in each $z_i$. Hence if $h_i$ are concave and nonnegative $g \circ h$ is concave.
- For $p \geq 1$, if $h_i$ are convex and nonnegative, $(\sum_{i=1}^m h_i(x)^p)^{1/p}$ is convex.
- $g(z) = (\prod_{i=1}^m z_i)^{1/m}$ on $\mathbb{R}_+^m$ is concave and its extension is nondecreasing in each $z_i$. If $h_i$ are nonnegative concave function, so is $(\prod_{i=1}^m g_i)^{1/m}$.

# Minimization

If $f$ is convex in $(x, y)$ and $C$ is nonempty and convex, then the function $g$, defined by

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

is convex in $x$ if $g(x) > -\infty$ for some $x$. Here, $\mathrm{dom} g = \{x | \ (x, y) \in \mathrm{dom} f$ for some $y \in C\}$.

**Proof**: $\square$

# Minimization(*cont'd*)

## Example

(Schur complement) Suppose for some $A$, $C$, $\in \mathbb{S}^n$

$$f(x, y) = x^T Ax + 2x^T Bx + x^T Cx,$$

is convex in $(x, y)$ so that $\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0$.
Consider $g(x) = \inf_x f(x, y)$ which is given by $g(x) = x^T (A - BC^{\dagger} B^T)x$. This is convex and hence $A - BC^{\dagger} B^T \succeq 0$. When $C$ is invertible, then $A - BC^{-1} B^T$ is called Schur complement of $\begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$.

# Minimization(*cont'd*)

## Example

(Distance to a set) Distance from $x$ to set $S$ w.r.t. $\|\cdot\|$ is

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

Function $\|x - y\|$ is convex in $(x, y)$, so if $S$ is convex, then $\mathrm{dist}(x, S)$ is convex in $x$.

# Perspective of a function

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

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

with domain

$$\mathrm{dom} g = \{(x, t) | x/t \in \mathrm{dom} f, t > 0\}$$

**Proposition**

*If $f$ is convex (concave, resp.), so is its perspective.*

**Proof**: $\square$

# Perspective of a function(*cont'd*)

**Example**

Suppose $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex, then is

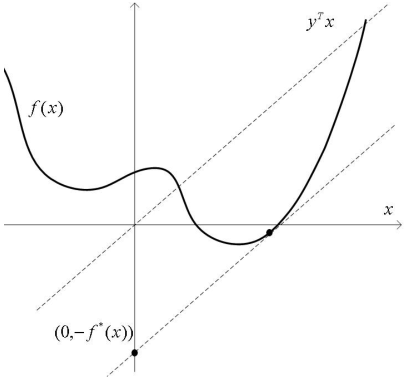
$$g(x) = (c^T x + d) f(Ax + b)/(c^T x + d),$$

with $\text{dom} g = \{x | c^T x + d > 0,\ Ax + b)/(c^T x + d) \in \text{dom} f\}$.

## Definition

Given $f : \mathbb{R}^n \to \mathbb{R}$, the conjugate $f^* : \mathbb{R}^n \to \mathbb{R}$ of $f$ is defined as:

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



- Affine functions $f(x) = a^T x + b$. Function $y^T x - a^T x - b$ is bounded only when $y = a$. Thus $\mathrm{dom} f^* = \{a\}$ and $f(a) = -b$.
- Negative logarithm $f(x) = -\log x$ with $\mathrm{dom} f = \mathbb{R}_{++}$. Function $y^T x + \log x$ is bounded only when $y < 0$ and attains its maximum when $x = -\frac{1}{y}$. Thus $\mathrm{dom} f^* = \mathbb{R}_{--}$ and $f^*(y) = -1 - \log(-y)$.
- Exponential $f(x) = e^x$. Function $yx - e^x$ attains its supremum only when $y > 0$ and then at $x = \log y$. Hence $f^*(y) = y \log(y) - y$ for $y > 0$. For $y = 0$, $f^*(y) = 0$.
- Negative entropy $f(x) = x \log x$ for $x \geq 0$ (defining $f(0) = 0$).
- Strictly convex quadratic $f(x) = \frac{1}{2} x^T Q x$ given $Q \succ 0$. As $y^T x - \frac{1}{2} x^T Q x$ is strictly concave, its unique maxima is attained when $x = Q^{-1} y$ for any $y$. Thus $f^*(y) = \frac{1}{2} y^T Q^{-1} y$.
- For any set $S \subseteq \mathbb{R}^n$, let $I_S(x)$ be its indicator function: $\mathrm{dom} I_S = S$ and $I_S(x) = 0$ for $x \in S$. Given $y$, $y^T x - I_S(x)$ is bounded only when $y^T x$ is bounded on $S$ and $f^*(y) = \sup \{y^T x | \ x \in S\}$ with $\mathrm{dom} f^* = \{y | \sup \{y^T x | \ x \in S\} < \infty\}$.

# Derivative of $f(X) = \log\det X$

For invertible $X \in \mathbb{R}^{n\times n}$, consider $f(X) = \log\det X$. From chain rule, $Df(X) = \frac{1}{\det X} D(\det X)$. Consider $\det X$ expanded w.r.t. $i$th row:

$$\det X = \sum_j X_{ij} \times (-1)^{i+j} \det \bar{X}_{ij},$$

where, $\bar{X}_{ij}$ is submatrix obtained by deleting row $i$ and column $j$ from $X$. Thus $\frac{\partial}{\partial X_{ij}} \det X = (-1)^{i+j} \det \bar{X}_{ij}$ and $D(\det X) = \mathrm{adj}(X)$, and hence

$$D(\log\det X) = (X^{-1})^T.$$

Thus if $X \in \mathbb{S}^n$, $D(\log\det X) = X^{-1}$. (See an alternative proof in Appendix of the textbook which seems more intuitive.)

- Log-determinant $f(X) = \log\det X^{-1}$ on $\mathbb{S}^n_{++}$. Then

$$f^*(Y) = \sup\{\operatorname{tr} YX + \log\det X | X \in \mathbb{S}^n_{++}\} \text{ for } Y \in \mathbb{S}^n.$$

First note that $f^*(Y) < \infty$ only when $Y \prec 0$. For, if $Y \not\prec 0$, $Y = \sum_{i=1}^n \lambda_i v_i v_i^T$ with $\|v_i\| = 1$ and $\lambda_r \geq 0$ for some $r$. Then let $X = I + t v_r v_r^T$. Then $X$ has $n-1$ 1's and $1+t$ as eigenvalues corresponding to $v_i$'s for $i \neq r$ and $v_r$, respectively. Thus $\operatorname{tr} YX + \log\det X = \operatorname{tr} Y + t\lambda + \log(1+t)$ which is unbounded on $t \geq 0$. When $Y \prec 0$, supremum attains when $D(\operatorname{tr} YX + \log\det X) = Y + X^{-1} = 0$, or $X = -Y^{-1}$. Hence,

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

- Norm $\|x\|$ and norm squared $\frac{1}{2}\|x\|^2$.

- As pointwise supremum of affine functions of $y$, $f^*$ is convex.
- From definition, we have Fenchel's inequality:

$$f(x) + f^*(y) \geq x^T y \; \forall \, x, y.$$

- We will see if $f$ is convex and closed, or $\mathrm{epi} f$ is closed, then $f^{**} = f$.

- For arbitrary $z \in \mathbb{R}^n$ define $y = \nabla f(z)$. Then we have

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

- For $a > 0$ and $b \in \mathbb{R}^n$, 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

$$g^*(y) = f^*(A^{-T}y) - b^T A^{-T} y$$

  with $\mathrm{dom}\ g^* = A^T \mathrm{dom}\ f^*$.

- 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

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

# Definition

## Definition

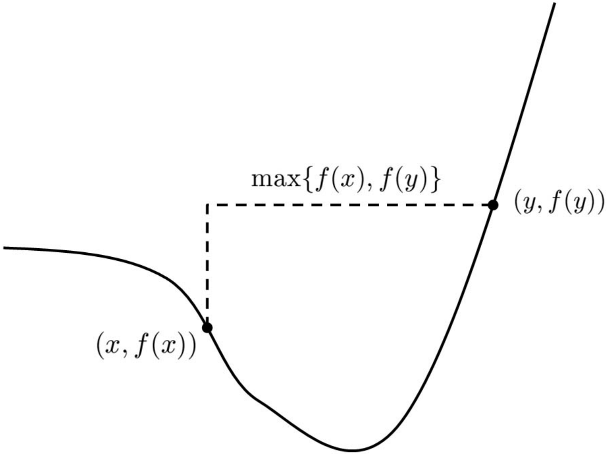
A function $f : \mathbb{R}^n \to \mathbb{R}$ is called quasiconvex if its domain and all its sublevel sets

$$S_\alpha = \{x \in \mathrm{dom} f \,|\, f(x) \leq \alpha\}$$

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

- A function is quasiconcave if $-f$ is quasiconvex.
- A function that is both quasiconvex and quasiconcave is called quasilinear.

# Definition



# Basic properties

### Theorem

*A function $f$ is quasiconvex if and only if* $\mathrm{dom} f$ *is convex and for any* $x, y \in \mathrm{dom} f$ *and* $0 \le \lambda \le 1$,

$$f(\lambda x + (1-\lambda)y) \le \max\{f(x), f(y)\}.$$

**Proof** $\square$

# Differentiable quasiconvex functions

- First-order conditions
  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 $x, y \in \mathbf{dom} f$

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

- Second-order conditions
  Supposer $f$ is twice differentiable. If $f$ is quasiconvex, then for all $x \in \mathrm{dom} f$, and all $y \in \mathbf{R}^n$, we have

$$y^T \nabla f(x) = 0 \Rightarrow y^T \nabla^2 f(x) y \geq 0.$$

  $f$ is quasiconvex if $f$ satisfies

$$y^T \nabla f(x) = 0 \Rightarrow y^T \nabla^2 f(x) y > 0$$

  for all $x \in \mathrm{dom} f$ and all nonzero $y \in \mathbb{R}^n$.

# Operations that preserve quasiconvexity

- Nonnegative weighted maximum
  A nonnegative weighted maximum of quasiconvex functions

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

  with $w_i \leq 0$ and $f_i$ quasiconvex, is quasiconvex.

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

# Operations that preserve quasiconvexity

- 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.

# Definition

### Definition

A function $f : \mathbb{R}^n \to \mathbb{R}$ is *logarithmically concave* or *log-concave* if $f(x) > 0$ for all $x \in \mathrm{dom} f$ and log $f$ is concave. It is said to be *logarithmically convex* or *log-convex* if log $f$ is convex.

### Definition

a function $f : \mathbb{R}^n \to \mathbb{R}$, with convex domain and $f(x) > 0$ for all $x \in \mathrm{dom} f$, is *log-concave* if and only if for all $x, y \in \mathrm{dom} f$ and $0 \leq \lambda \leq 1$, we have

$$f(\lambda x + (1-\lambda)y) \geq f(x)^{\lambda} f(y)^{1-\lambda}$$

# Properties

- Twice differentiable log-convex/concave functions
  Suppose $f$ is twice differentiable, with $\mathrm{dom} f$ convex, so

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

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

$$f(x) \nabla^2 f(x) \succeq \nabla f(x) \nabla f(x)^T$$

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

$$f(x) \nabla^2 f(x) \preceq \nabla f(x) \nabla f(x)^T$$

# Monotonicity and convexity w.r.t generalized inequality

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

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

and K-*increasing* if

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

- Suppose $K \subseteq \mathbb{R}^m$ is a proper cone with associated generalized inequality $\preceq_K$. We say $f : \mathbb{R}^n \to \mathbb{R}^m$ is K- *convex* if for all $x, y$, and $0 \leq \lambda \leq 1$,

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

# Homework

3.1, 3.4, 3.9, 3.14, 3.16, 3.18, 3.22, 3.26, 3.28(a) (b)*, 3.39(a-c), 3.43, 3.44, 3.49, 3.57,

* extra credit