Introduction to Convexity

George Voutsadakis[1]

[1]Mathematics and Computer Science
Lake Superior State University

LSSU Math 500

Subsection 1

Convex Functions on the Real Line

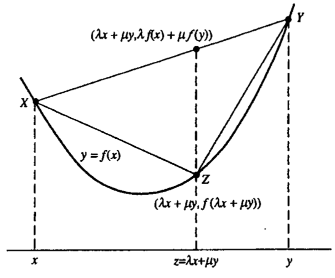
Convex and Concave Functions

- We will be concerned with a real-valued function $f : I \to \mathbb{R}$ defined on a non-degenerate (i.e., contains more than one point) interval I of the real line.
- Such a function f is said to be **convex** if

$$f(\lambda x + \mu y) \leq \lambda f(x) + \mu f(y),$$

whenever $x, y \in I$ and $\lambda, \mu \geq 0$ with $\lambda + \mu = 1$.

- Geometrically, f is convex if every chord joining two points on its graph lies on or above the graph.
- If $-f : I \to \mathbb{R}$ is convex, then $f : I \to \mathbb{R}$ is said to be **concave**.



Example

- We show that the square function $f:\mathbb{R}\to\mathbb{R}$ defined for real x by the equation
$$f(x)=x^2$$
is convex.
- Let $x,y\in\mathbb{R}$ and let $\lambda,\mu\geq 0$ with $\lambda+\mu=1$.
- Then
$$\begin{aligned}\lambda f(x)+\mu f(y)-f(\lambda x+\mu y) &= \lambda x^2+\mu y^2-(\lambda x+\mu y)^2\\ &= \lambda x^2+\mu y^2-\lambda^2x^2-2\lambda\mu xy-\mu^2y^2\\ &= \lambda(1-\lambda)x^2-2\lambda\mu xy+\mu(1-\mu)y^2\\ &= \lambda\mu x^2-2\lambda\mu xy+\lambda\mu y^2\\ &= \lambda\mu(x^2-2xy+y^2)\\ &= \lambda\mu(x-y)^2\geq 0.\end{aligned}$$
- This establishes the convexity of the square function.

The Three Chords Lemma

Theorem (Three Chords Lemma)

Let $f : I \to \mathbb{R}$ be a convex function and let $x, y, z \in I$ satisfy $x < z < y$. Then

$$\frac{f(z) - f(x)}{z - x} \leq \frac{f(y) - f(x)}{y - x} \leq \frac{f(y) - f(z)}{y - z}.$$

- We express z as a convex combination of x, y: $z = \frac{y - z}{y - x} x + \frac{z - x}{y - x} y$.

  By the convexity of f, $f(z) \leq \frac{y - z}{y - x} f(x) + \frac{z - x}{y - x} f(y)$. Thus,

  $$f(z) - f(x) \leq \frac{y - z - y + x}{y - x} f(x) + \frac{z - x}{y - x} f(y) = \frac{z - x}{y - x}(f(y) - f(x)).$$

  So, we get $\frac{f(z) - f(x)}{z - x} \leq \frac{f(y) - f(x)}{y - x}$.

  The other inequality follows similarly.

The Slope Function

Corollary

Let $f : I \to \mathbb{R}$ be a convex function and let $a \in I$. Then the function $g : I \backslash \{a\} \to \mathbb{R}$ defined by the equation

$$g(x) = \frac{f(x) - f(a)}{x - a}, \quad x \in I \backslash \{a\},$$

is increasing.

- If $b, c \in I \backslash \{a\}$ with $b < c$, then we must show that $g(b) \leq g(c)$.

  Either $b < c < a$, $b < a < c$, or $a < b < c$. Suppose that $b < c < a$. Then the theorem with $x = b$, $y = a$, $z = c$ shows that $g(b) \leq g(c)$.

  The other cases can be proved in a similar fashion.

Convexity and Differentiability

Theorem

Let $f : I \to \mathbb{R}$ be a convex function. Then f possesses left and right derivatives at each interior point of I. Moreover, if a, b are interior points of I with $a < b$, then

$$f'_-(a) \le f'_+(a) \le \frac{f(b) - f(a)}{b - a} \le f'_-(b) \le f'_+(b).$$

- Let c be an interior point of f and let x, y be points of I such that $x < c < y$. The corollary shows that, as x increases to c from below, $\frac{f(x)-f(c)}{x-c}$ increases and is bounded above by $\frac{f(y)-f(c)}{y-c}$. Thus, the left derivative $f'_-(c)$ exists and satisfies the inequality

$$f'_-(c) \le \frac{f(y) - f(c)}{y - c}.$$

Convexity and Differentiability (Cont'd)

- Letting y decrease to c in this inequality, we see that the right derivative $f'_+(c)$ exists and satisfies the inequality $f'_-(c) \le f'_+(c)$.

  Thus, if a, b are interior points of I, then

  $$f'_-(a) \le f'_+(a) \quad \text{and} \quad f'_-(b) \le f'_+(b).$$

  By the corollary, for $a < x < b$,

  $$\frac{f(x)-f(a)}{x-a} \le \frac{f(b)-f(a)}{b-a} \quad \text{and} \quad \frac{f(b)-f(a)}{b-a} \le \frac{f(b)-f(x)}{b-x}.$$

  Letting $x \to a^+$ in the first and $x \to b^-$ in the second, we get

  $$f'_+(a) \le \frac{f(b)-f(a)}{b-a} \le f'_-(b).$$

Convexity and Continuity

Corollary

Let $f : I \to \mathbb{R}$ be a convex function. Then, on the interior of I, f is continuous and f'_-, f'_+ are increasing.

- At each interior point of I, f has both left and right derivatives, and so is continuous from the left and from the right.

  Hence it is continuous.

  That f'_-, f'_+ are increasing on the interior of f follows immediately from the theorem.

Behavior at the Boundary

- A convex function need not be continuous at the boundary points of its domain.

  Example: The convex function $f : [0,1] \to \mathbb{R}$ defined by the equations

  $$f(x) = \begin{cases} 0, & \text{if } 0 < x < 1, \\ 1, & \text{if } x = 0, 1. \end{cases}$$

  is not continuous at 0 and 1.

- Also a convex function need not be differentiable, even at an interior point of its domain.

  Example: The modulus (absolute value) function is not differentiable at the origin. There its left and right derivatives are -1 and 1, respectively.

Points of NonDifferentiability

Corollary

Let $f : I \to \mathbb{R}$ be a convex function. Then the set of those points of I at which f is not differentiable is countable.

- Let C be the set of points of int$I$ at which f is not differentiable. With each c in C, we associate a rational r_c such that $f'_-(c) < r_c < f'_+(c)$. It follows from the theorem that, if $c, d \in C$ with $c < d$, then

$$f'_-(c) < r_c < f'_+(c) < f'_-(d) < r_d < f'_+(d),$$

whence $r_c < r_d$. This shows immediately that the set of points of int$I$, and hence of I, at which I is not differentiable is countable.

Criterion for Convexity

Theorem

Let $f : I \to \mathbb{R}$ be a differentiable function. Then f is convex if and only if f' is increasing.

- Suppose first that f is convex. Let $a, b \in I$ with $a < b$. Then a previous corollary shows that

$$f'(a) = \lim_{x \to a^+} \frac{f(x) - f(a)}{x - a} \le \frac{f(b) - f(a)}{b - a} \le \lim_{x \to b^-} \frac{f(x) - f(b)}{x - b} = f'(b).$$

  Hence $f'(a) < f'(b)$ and f' is increasing.

Criterion for Convexity (Converse)

- Suppose next that f' is increasing. Let $a, b \in I$ with $a < b$ and let $\lambda, \mu > 0$ with $\lambda + \mu = 1$. By the first Mean Value Theorem, there exist real numbers, c, d with $a < c < \lambda a + \mu b < d < b$, such that
$$\frac{f(\lambda a + \mu b) - f(a)}{\lambda a + \mu b - a} = f'(c) \le f'(d) = \frac{f(b) - f(\lambda a + \mu b)}{b - \lambda a - \mu b}.$$
So we get
$$\frac{f(\lambda a + \mu b) - f(a)}{\mu(b - a)} \le \frac{f(b) - f(\lambda a + \mu b)}{\lambda(b - a)}$$
$$\lambda f(\lambda a + \mu b) - \lambda f(a) \le \mu f(b) - \mu f(\lambda a + \mu b)$$
$$f(\lambda a + \mu b) \le \lambda f(a) + \mu f(b).$$
Hence, f is convex.

Corollary

Let $f : I \to \mathbb{R}$ be a twice differentiable function. Then f is convex if and only if $f''(x) \ge 0$ for all x in I.

Example

- The function e^x is convex on $\mathbb{R}$.

$$(e^x)'' = (e^x)' = e^x > 0.$$

- The function $-\log x$ is convex on $(0, +\infty)$.

$$(-\log x)'' = \left(-\frac{1}{x}\right)' = \frac{1}{x^2} > 0.$$

- The function $x \log x$ is convex on $(0, +\infty)$.

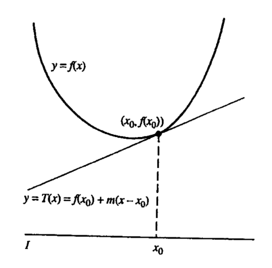
$$(x \log x)'' = \left(\log x + x\frac{1}{x}\right)' = \frac{1}{x} > 0.$$

- The function x^p, $p \geq 1$, is convex on $[0, \infty)$.

$$(x^p)'' = (px^{p-1})' = p(p-1)x^{p-2} \geq 0.$$

Support

- Suppose that $f : I \to \mathbb{R}$ is a real-valued function defined on an open interval I of the real line and that $x_0 \in I$.
- Then an affine transformation $T : \mathbb{R} \to \mathbb{R}$ is said to **support** f at x_0 if $T(x_0) = f(x_0)$ and $T(x) \leq f(x)$, for all $x \in I$.
  We say that f **has support** T at x_0.



- Such an affine transformation T can be expressed in the form $T(x) = f(x_0) + m(x - x_0)$ for some real number m.
- $y = f(x_0) + m(x - x_0)$ is the equation of the line with slope m passing through the point $(x_0, f(x_0))$ on the graph of f.
- The condition $T(x) \leq f(x)$ means that this line lies on or below the graph of f.

Convexity and Support

Theorem

Let $f : I \to \mathbb{R}$ be a real-valued function defined on an open interval I of $\mathbb{R}$. Then f is convex if and only if it has support at each point of I.

- Suppose first that f has support at each point of I. Let $x, y \in I$ and let $\lambda, \mu \geq 0$ with $\lambda + \mu = 1$. Let T support f at $\lambda x + \mu y$. Then

$$f(\lambda x + \mu y) = T(\lambda x + \mu y) = \lambda T(x) + \mu T(y) \leq \lambda f(x) + \mu f(y).$$

So f is convex.

Convexity and Support (Converse)

- Suppose next that f is convex. Let $x_0 \in I$ and let m be a real number satisfying the inequalities $f'_-(x_0) \le m \le f'_+(x_0)$. Define an affine transformation $T : \mathbb{R} \to \mathbb{R}$ by the equation

$$T(x) = f(x_0) + m(x - x_0), \quad x \in \mathbb{R}.$$

Let $y, z \in I$ be such that $y < x_0 < z$. Then, by a previous theorem,

$$\begin{aligned} \frac{f(y)-f(x_0)}{y-x_0} &\le f'_-(x_0) \\ &\le \frac{T(y)-T(x_0)}{y-x_0} = m = \frac{T(z)-T(x_0)}{z-x_0} \\ &\le f'_+(x_0) \\ &\le \frac{f(z)-f(x_0)}{z-x_0}. \end{aligned}$$

Hence $T(y) \le f(y)$ and $T(z) \le f(z)$. Thus T supports f at x_0.

Differentiability and Support

Theorem

Let $f : I \to \mathbb{R}$ be a convex function defined on an open interval I of $\mathbb{R}$. Then f is differentiable at a point x_0 of I if and only if it has unique support at x_0.

- Suppose first that f is differentiable at x_0. Let $T : \mathbb{R} \to \mathbb{R}$ support f at x_0; say

  $$T(x) = f(x_0) + m(x - x_0), \text{ for } x \in \mathbb{R},$$

  where m is a real number. Let $y, z \in I$ be such that $y < x_0 < z$. Then

  $$\frac{f(y) - f(x_0)}{y - x_0} \le \frac{T(y) - T(x_0)}{y - x_0} = m = \frac{T(z) - T(x_0)}{z - x_0} \le \frac{f(z) - f(x_0)}{z - x_0}.$$

  Thus, letting $y \to x_0^-$, $z \to x_0^+$, we deduce that $m = f'(x_0)$. Hence, f has unique support at x_0.

Differentiability and Support (Converse)

- Suppose next that f has unique support at x_0.

  Let the real number m satisfy $f'_-(x_0) \leq m \leq f'_+(x_0)$.

  Then, as in the proof of the preceding theorem, the affine transformation T defined by the equation

$$T(x) = f(x_0) + m(x - x_0)$$

  supports f at x_0. But f has unique support at x_0.

  Hence, m is unique and $f'_-(x_0) = f'_+(x_0)$.

  So f is differentiable at x_0.

Subsection 2

Classical Inequalities

Jensen's Inequality

Theorem (Jensen's Inequality)

Let $f: I \to \mathbb{R}$ be a convex function. Let $x_1, \ldots, x_m \in I$ and let $\lambda_1, \ldots, \lambda_m \geq 0$ with $\lambda_1 + \cdots + \lambda_m = 1$. Then

$$f(\lambda_1 x_1 + \cdots + \lambda_m x_m) \leq \lambda_1 f(x_1) + \cdots + \lambda_m f(x_m).$$

- We argue by induction on m.

  The inequality is trivially true when $m = 1$.

  Assume, then, that it is true when $m = k$, where $k \geq 1$.

  Let a real number x be defined by the equation

  $$x = \lambda_1 x_1 + \cdots + \lambda_{k+1} x_{k+1},$$

  where $x_1, \ldots, x_{k+1} \in I$ and $\lambda_1, \ldots, \lambda_{k+1} \geq 0$ with $\lambda_1 + \cdots + \lambda_{k+1} = 1$.

  At least one of $\lambda_1, \ldots, \lambda_{k+1}$ must be less than 1, say $\lambda_{k+1} < 1$.

Jensen's Inequality (Cont'd)

- Write
$$\lambda = \lambda_1 + \cdots + \lambda_k = 1 - \lambda_{k+1}.$$
Then $\lambda > 0$. Write
$$y = \frac{\lambda_1}{\lambda} x_1 + \cdots + \frac{\lambda_k}{\lambda} x_k.$$
The induction hypothesis shows that
$$f(y) \le \frac{\lambda_1}{\lambda} f(x_1) + \cdots + \frac{\lambda_k}{\lambda} f(x_k).$$
Since f is convex,
$$\begin{array}{rcl} f(x) & = & f(\lambda y + \lambda_{k+1} x_{k+1}) \\ & \le & \lambda f(y) + \lambda_{k+1} f(x_{k+1}) \\ & \le & \lambda_1 f(x_1) + \cdots + \lambda_{k+1} f(x_{k+1}). \end{array}$$
This establishes the inequality for $m = k + 1$.

Arithmetic and Geometric Means

- In this section the word number will be used exclusively to mean *positive real number*.
- The **arithmetic mean** and the **geometric mean** of numbers x_1 and x_2 are defined to be

$$\frac{1}{2}(x_1+x_2) \quad \text{and} \quad \sqrt{x_1x_2}.$$

- The basic inequality between these means is that the geometric mean never exceeds the arithmetic mean, i.e., $\sqrt{x_1x_2} \le \frac{1}{2}(x_1+x_2)$.
- This follows immediately from the fact that $(\sqrt{x_1}-\sqrt{x_2})^2 \ge 0$.
- The **arithmetic mean** and the **geometric mean** of numbers $x_1,\ldots,x_m$ are defined, respectively, to be

$$\frac{1}{m}(x_1+\cdots+x_m) \quad \text{and} \quad (x_1\cdots x_m)^{1/m}.$$

- Once again the geometric mean never exceeds the arithmetic mean, although the proof is appreciably more difficult than when $m=2$.

Weighted Arithmetic and Geometric Means

- The concepts of arithmetic and geometric means can be generalized by attaching **weights** $\alpha_1, \ldots, \alpha_m$ to the numbers as follows.
- Let $\alpha_1, \ldots, \alpha_m$ be numbers whose sum is 1.
- Then the numbers

$$\alpha_1 x_1 + \cdots + \alpha_m x_m \quad \text{and} \quad x_1^{\alpha_1} \cdots x_m^{\alpha_m}$$

  are called, respectively, the **weighted arithmetic mean** and the **weighted geometric mean** of the numbers $x_1, \ldots, x_m$ with respect to the weights $\alpha_1, \ldots, \alpha_m$.
- These weighted means reduce to the usual means when each of the weights $\alpha_1, \ldots, \alpha_m$ is $\frac{1}{m}$.

Relations Between Weighted Means

Theorem

Let $x_1,\ldots,x_m,\alpha_1,\ldots,\alpha_m > 0$ with $\alpha_1+\cdots+\alpha_m = 1$. Then

$$x_1^{\alpha_1}\cdots x_m^{\alpha_m} \le \alpha_1 x_1+\cdots+\alpha_m x_m.$$

- The function $-\log$ is convex on $(0,\infty)$. Hence, by Jensen's inequality,

$$\begin{aligned} -\log(\alpha_1 x_1+\cdots+\alpha_m x_m) &\le -(\alpha_1\log x_1+\cdots+\alpha_m\log x_m) \\ &= -\log(x_1^{\alpha_1}\cdots x_m^{\alpha_m}). \end{aligned}$$

  Since $\log$ is a strictly increasing function, we can deduce that

$$x_1^{\alpha_1}\cdots x_m^{\alpha_m} \le \alpha_1 x_1+\cdots+\alpha_m x_m.$$

Corollary

Let $x_1,\ldots,x_m > 0$. Then

$$(x_1\cdots x_m)^{1/m} \le \frac{1}{m}(x_1+\cdots+x_m).$$

A very General Inequality

Theorem

Let $a_{ij} > 0$ $(i = 1, \ldots, m; j = 1, \ldots, n)$ and $\alpha_1, \ldots, \alpha_m > 0$ with $\alpha_1 + \cdots + \alpha_m = 1$. Then

$$a_{11}^{\alpha_1} \cdots a_{m1}^{\alpha_m} + \cdots + a_{1n}^{\alpha_1} \cdots a_{mn}^{\alpha_m} \leq (a_{11} + \cdots + a_{1n})^{\alpha_1} \cdots (a_{m1} + \cdots + a_{mn})^{\alpha_m}.$$

- We use the inequality between weighted means to deduce that, for each $j = 1, \ldots, n$,

$$\frac{a_{1j}^{\alpha_1} \cdots a_{mj}^{\alpha_m}}{(a_{11} + \cdots + a_{1n})^{\alpha_1} \cdots (a_{m1} + \cdots + a_{mn})^{\alpha_m}} \leq \frac{\alpha_1 a_{1j}}{a_{11} + \cdots + a_{1n}} + \cdots + \frac{\alpha_m a_{mj}}{a_{m1} + \cdots + a_{mn}}.$$

Adding these n inequalities together, we deduce that

$$\sum_{j=1}^{n} \frac{a_{ij}^{\alpha_1} \cdots a_{mj}^{\alpha_m}}{(a_{11} + \cdots + a_{1n})^{\alpha_1} \cdots (a_{m1} + \cdots + a_{mn})^{\alpha_m}} \leq \alpha_1 + \cdots + \alpha_m = 1.$$

The desired result follows immediately.

Hölder's Inequality

Corollary

Let $x_1,\ldots,x_m,\ y_1,\ldots,y_m>0$. Then

$$(x_1\cdots x_m)^{1/m}+(y_1\cdots y_m)^{1/m}\le(x_1+y_1)^{1/m}\cdots(x_m+y_m)^{1/m}.$$

- Let $n=2$, $\alpha_1=\frac{1}{m},\ldots,\alpha_m=\frac{1}{m}$, $a_{i1}=x_i$ and $a_{i2}=y_i$ in the theorem.

Corollary (Hölder's Inequality)

Let $x_1,\ldots,x_n,y_1,\ldots,y_n>0$. Suppose that $p,q>0$ satisfy $\frac{1}{p}+\frac{1}{q}=1$. Then

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

- Let $m=2$, $\alpha_1=\frac{1}{p}$, $\alpha_2=\frac{1}{q}$ and let $a_{1j}=x_j^p$, $a_{2j}=y_j^q$ in the above theorem.

Minkowski's Inequality

Theorem (Minkowski's Inequality)

Let $x_1,\ldots,x_n,y_1,\ldots,y_n>0$ and let $p\geq 1$. Then

$$\left(\sum_{i=1}^{n}(x_1+y_i)^p\right)^{1/p}\leq\left(\sum_{i=1}^{n}x_i^p\right)^{1/p}+\left(\sum_{i=1}^{n}y_i^p\right)^{1/p}.$$

- Write $a=(\sum_{i=1}^{n}x_i^p)^{1/p}$ and $b=(\sum_{i=1}^{n}y_i^p)^{1/p}$. Since x^p $(p\geq 1)$ is convex on $(0,\infty)$, we can deduce that, for $i=1,\ldots,n$,

$$\left(\frac{x_i+y_i}{a+b}\right)^p=\left(\frac{a}{a+b}\left(\frac{x_i}{a}\right)+\frac{b}{a+b}\left(\frac{y_i}{b}\right)\right)^p\leq\frac{a}{a+b}\left(\frac{x_i}{a}\right)^p+\frac{b}{a+b}\left(\frac{y_i}{b}\right)^p.$$

  Adding these n inequalities together, we deduce

$$\sum_{i=1}^{n}\left(\frac{x_i+y_i}{a+b}\right)^p\leq\frac{a}{a+b}\left(\frac{\sum_{i=1}^{n}x_i^p}{a^p}\right)+\frac{b}{a+b}\left(\frac{\sum_{i=1}^{n}y_i^p}{b^p}\right)=\frac{a}{a+b}+\frac{b}{a+b}=1.$$

  Thus, $\sum_{i=1}^{n}(x_1+y_i)^p\leq(a+b)^p=((\sum_{i=1}^{n}x_i^p)^{1/p}+(\sum_{i=1}^{n}y_i^p)^{1/p})^p$.

Harmonic Mean and Root Mean Square

- Given the numbers $x_1, \ldots, x_m$, their **harmonic mean** is defined to be

$$\frac{1}{\frac{1}{m}\left(\frac{1}{x_1} + \cdots + \frac{1}{x_m}\right)}.$$

- Their **root mean square** is defined to be

$$\sqrt{\frac{x_1^2 + \cdots + x_m^2}{m}}.$$

- The basic inequalities connecting the four means are:

$$\begin{aligned} \text{harmonic mean} &\leq \text{geometric mean} \\ &\leq \text{arithmetic mean} \\ &\leq \text{root mean square.} \end{aligned}$$

Weighted Harmonic Mean and Root Mean Square

- The harmonic mean and the root mean square are generalized in the natural way to the corresponding weighted means.
- Let $\alpha_1, \ldots, \alpha_m > 0$ with $\alpha_1 + \cdots + \alpha_m = 1$.
- Then the numbers

$$\frac{1}{\frac{\alpha_1}{x_1} + \cdots + \frac{\alpha_m}{x_m}} \quad \text{and} \quad \sqrt{\alpha_1 x_1^2 + \cdots + \alpha_m x_m^2}$$

  are called, respectively, the **weighted harmonic mean** and the **weighted root mean square** of the numbers $x_1, \ldots, x_m$ with respect to the weights $\alpha_1, \ldots, \alpha_m$.
- We will see that the basic inequalities stated above connecting the four unweighted means continue to hold for the weighted means.

Mean of Order t

- The four means so far introduced are special cases of the mean of order t:
- Let $\alpha = (\alpha_1, \ldots, \alpha_m)$, $x = (x_1, \ldots, x_m)$, where $\alpha_1, \ldots, \alpha_m$, $x_1, \ldots, x_m > 0$ and $\alpha_1 + \cdots + \alpha_m = 1$.
- Then for each non-zero real number t, the **mean** $M_t(\boldsymbol{x}; \boldsymbol{\alpha})$ **of order** t is defined by the equation

$$M_t(\boldsymbol{x}; \boldsymbol{\alpha}) = (\alpha_1 x_1^t + \cdots + \alpha_m x_m^t)^{1/t}.$$

- The values $t = -1, 1, 2$ give rise, respectively, to the weighted harmonic mean, the weighted arithmetic mean and the weighted root mean square.
- The weighted geometric mean is not the mean of order t for any non-zero real number t.

The Mean of Order Zero

- We consider the limit of $M_t(\boldsymbol{x};\boldsymbol{\alpha})$ as t tends to zero.
- Taking logarithms on both sides of the defining equation of $M_t(\boldsymbol{x};\boldsymbol{\alpha})$,

$$\log M_t(\boldsymbol{x};\boldsymbol{\alpha}) = \frac{\log(\alpha_1 x_1^t + \cdots + \alpha_m x_m^t)}{t}.$$

- By definition, $\lim_{t\to 0}\frac{\log(\alpha_1 x_1^t + \cdots + \alpha_m x_m^t)}{t}$ is the derivative of $\log(\alpha_1 x_1^t + \cdots + \alpha_m x_m^t)$ at $t=0$.
- We calculate

$$[\log(\alpha_1 x_1^t + \cdots \alpha_m x_m^t)]' = \frac{\alpha_1 x_1^t \log x_1 + \cdots + \alpha_m x_m^t \log x_m}{\alpha_1 x_1^t + \cdots + \alpha_m x_m^t}.$$

- Therefore,

$$\begin{aligned}\lim_{t\to 0}\frac{\log(\alpha_1 x_1^t + \cdots + \alpha_m x_m^t)}{t} &= \alpha_1 \log x_1 + \cdots \alpha_m \log x_m \\ &= \log(x_1^{\alpha_1} \cdots x_m^{\alpha_m}).\end{aligned}$$

- Hence, $\lim_{t\to 0}\log M_t(\boldsymbol{x};\boldsymbol{\alpha}) = \log(x_1^{\alpha_1}\cdots x_m^{\alpha_m})$.

The Mean of Order Zero (Cont'd)

- We calculated
$$\lim_{t\to 0}\log M_t(\boldsymbol{x};\boldsymbol{\alpha}) = \log(x_1^{\alpha_1}\cdots x_m^{\alpha_m}).$$
- Thus
$$\begin{aligned}\lim_{t\to 0} M_t(\boldsymbol{x};\boldsymbol{\alpha}) &= \lim_{t\to 0} e^{\log M_t(\boldsymbol{x};\boldsymbol{\alpha})}\\ &= e^{\lim_{t\to 0}\log M_t(\boldsymbol{x};\boldsymbol{\alpha})}\\ &= e^{\log(x_1^{\alpha_1}\cdots x_m^{\alpha_m})}\\ &= x_1^{\alpha_1}\cdots x_m^{\alpha_m}.\end{aligned}$$
- So $M_t(\boldsymbol{x};\boldsymbol{\alpha}) \stackrel{t\to 0}{\longrightarrow} x_1^{\alpha_1}\cdots x_m^{\alpha_m}$.
- We define the **mean of order zero**
$$M_0(\boldsymbol{x};\boldsymbol{\alpha}) := x_1^{\alpha_1}\cdots x_m^{\alpha_m}.$$
- $M_t(\boldsymbol{x};\boldsymbol{\alpha})$ is now defined for every real number t and is continuous on the whole of $\mathbb{R}$, in particular at $t=0$.

Monotonicity of Mean of Order t

Theorem

Let $\boldsymbol{x} = (x_1, \ldots, x_m)$, $\alpha = (\alpha_1, \ldots, \alpha_m)$, where $x_1, \ldots, x_m$, $\alpha_1, \ldots, \alpha_m > 0$ and $\alpha_1 + \cdots + \alpha_m = 1$. Then $M_t(\boldsymbol{x}; \boldsymbol{\alpha})$ is an increasing function of t.

- Since $\boldsymbol{x}$ and $\boldsymbol{\alpha}$ are fixed, we write $M_t(\boldsymbol{x}; \boldsymbol{\alpha})$ simply as $M(t)$.
  We show that $M'(t) \geq 0$ for all non-zero real numbers t.
  Since M is continuous at 0, this shows that M is increasing on $\mathbb{R}$.
  We have $t \log M(t) = \log(\alpha_1 x_1^t + \cdots + \alpha_m x_m^t)$.
  So, by differentiating,

$$t \frac{M'(t)}{M(t)} + \log M(t) = \frac{\alpha_1 x_1^t \log x_1 + \cdots + \alpha_m x_m^t \log x_m}{\alpha_1 x_1^t + \cdots + \alpha_m x_m^t}, \quad t \neq 0.$$

  Thus, for $t \neq 0$,

$$t^2 \frac{M'(t)}{M(t)} + t \log M(t) = \frac{\alpha_1 x_1^t \log x_1^t + \cdots + \alpha_m x_m^t \log x_m^t}{\alpha_1 x_1^t + \cdots + \alpha_m x_m^t}.$$

Monotonicity of Mean of Order t (Cont'd)

- We get

$$\frac{t^2 M'(t)(\alpha_1 x_1^t + \cdots + \alpha_m x_m^t)}{M(t)} = \alpha_1 x_1^t \log x_1^t + \cdots + \alpha_m x_m^t \log x_m^t - (\alpha_1 x_1^t + \cdots + \alpha_m x_m^t) \log (\alpha_1 x_1^t + \cdots + \alpha_m x_m^t).$$

Jensen's inequality, applied to the convex function $y \log y$ on $(0, \infty)$, shows that, for all $y_1, \ldots, y_m > 0$,

$$(\alpha_1 y_1 + \cdots + \alpha_m y_m) \log (\alpha_1 y_1 + \cdots + \alpha_m y_m) \leq \alpha_1 y_1 \log y_1 + \cdots + \alpha_m y_m \log y_m.$$

If we put $y_i = x_i^t$ for $i = 1, \ldots, m$ in this inequality, we deduce from the equality previously stated that $M'(t) \geq 0$ for $t \neq 0$.

Corollary

Let $\boldsymbol{x} = (x_1, \ldots, x_m)$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)$, where $x_1, \ldots, x_m$, $\alpha_1, \ldots, \alpha_m > 0$ and $\alpha_1 + \cdots + \alpha_m = 1$. Then $M_{-1}(\boldsymbol{x}; \boldsymbol{\alpha}) \leq M_0(\boldsymbol{x}; \boldsymbol{\alpha}) \leq M_1(\boldsymbol{x}; \boldsymbol{\alpha}) \leq M_2(\boldsymbol{x}; \boldsymbol{\alpha})$.

Subsection 3

The Gamma and Beta Functions

Hölder's Inequality for Intervals

- $\int_I f$ denotes the (Riemann) integral of a continuous function $f : I \to \mathbb{R}$ over an interval I of the real line.

Theorem (Hölder's Inequality for Integrals)

Let $f, g : I \to \mathbb{R}$ be continuous non-negative functions for which the integrals $\int_I f$, $\int_I g$ are positive. Let $\lambda, \mu \geq 0$ with $\lambda + \mu = 1$. Then

$$\int_I f^\lambda g^\mu \leq \left(\int_I f\right)^\lambda \left(\int_I g\right)^\mu.$$

- By a previous theorem, for $t \in I$, $\left(\frac{f(t)}{\int_I f}\right)^\lambda \left(\frac{g(t)}{\int_I g}\right)^\mu \leq \lambda \frac{f(t)}{\int_I f} + \mu \frac{g(t)}{\int_I g}$. We integrate both sides of this inequality to deduce that

$$\frac{\int_I f^\lambda(t) g^\mu(t) dt}{(\int_I f)^\lambda (\int_I g)^\mu} \leq \lambda \frac{\int_I f}{\int_I f} + \mu \frac{\int_I g}{\int_I g} = \lambda + \mu = 1.$$

Hence $\int_I f^\lambda f^\mu \leq (\int_I f)^\lambda (\int_I g)^\mu$.

Log-Convex Functions

- Let $f : I \to \mathbb{R}$ be a function defined on an interval I of the real line.
- Then I is said to be **log-convex** if it is positive and its logarithm, $\log f : I \to \mathbb{R}$, is convex.
- Thus a positive function f is log-convex on an interval I if and only if, whenever $x, y \in I$ and $\lambda, \mu \geq 0$ with $\lambda + \mu = 1$, we have

$$\log f(\lambda x + \mu y) \leq \lambda \log f(x) + \mu \log f(y) = \log f^{\lambda}(x) f^{\mu}(y).$$

- This amounts to

$$f(\lambda x + \mu y) \leq f^{\lambda}(x) f^{\mu}(y).$$

- Since $f^{\lambda}(x) f^{\mu}(y) \leq \lambda f(x) + \mu f(y)$, it follows that every log-convex function is convex.
- On the other hand, on the interval $(0, \infty)$, the positive function x is convex but not log-convex.
- For any positive number a, the function a^x is log-convex on $\mathbb{R}$.

Closure Under Addition and Multiplication

- The class of functions which are log-convex on some interval I is closed under addition and multiplication.
- Suppose that the functions $f,g:I\to\mathbb{R}$ are log-convex.
  Let $x,y\in I$ and let $\lambda,\mu\geq 0$ with $\lambda+\mu=1$.
  By a previous theorem,

$$\begin{array}{rcl}
(f+g)(\lambda x+\mu y) & = & f(\lambda x+\mu y)+g(\lambda x+\mu y)\\
& \leq & f^{\lambda}(x)f^{\mu}(y)+g^{\lambda}(x)g^{\mu}(y)\\
& \leq & (f(x)+g(x))^{\lambda}(f(y)+g(y))^{\mu}\\
& = & (f+g)^{\lambda}(x)+(f+g)^{\mu}(y);\\
(fg)(\lambda x+\mu y) & = & f(\lambda x+\mu y)g(\lambda x+\mu y)\\
& \leq & f^{\lambda}(x)f^{\mu}(y)g^{\lambda}(x)g^{\mu}(y)\\
& = & (fg)^{\lambda}(x)(fg)^{\mu}(y).
\end{array}$$

The Gamma Function

- The **gamma function** $\Gamma:(0,\infty)\to\mathbb{R}$ is defined by the equation

$$\Gamma(x)=\int_0^\infty t^{x-1}e^{-t}dt,\quad x>0.$$

- Elementary analysis shows that, for each $x>0$, $\Gamma(x)$ is a well-defined positive number.

Theorem

The gamma function has the following properties:

(i) $\Gamma(x+1)=x\Gamma(x)$ for $x>0$;

(ii) $\Gamma(1)=1$;

(iii) Γ is log-convex.

Proofs of the Properties

(i) For $x > 0$,

$$\Gamma(x+1) = \int_0^\infty t^x e^{-t} dt = [-t^x e^{-t}]_0^\infty + x\int_0^\infty t^{x-1} e^{-t} dt = x\Gamma(x).$$

(ii) $\Gamma(1) = \int_0^\infty e^{-t} dt = \lim_{A\to\infty}[1 - e^{-A}] = 1.$

(iii) Let $x, y > 0$. Let $\lambda, \mu \geq 0$ with $\lambda + \mu = 1$. Then, by the preceding theorem,

$$\begin{aligned} \Gamma(\lambda x + \mu y) &= \int_0^\infty t^{\lambda x + \mu y - 1} e^{-t} dt \\ &= \int_0^\infty (t^{x-1} e^{-t})^\lambda (t^{y-1} e^{-t})^\mu dt \\ &\leq (\int_0^\infty t^{x-1} e^{-t} dt)^\lambda (\int_0^\infty t^{y-1} e^{-t} dt)^\mu \\ &= \Gamma^\lambda(x)\Gamma^\mu(y). \end{aligned}$$

Value on Integers and Limit Properties

Corollary

For $n = 0, 1, 2, \ldots$, $\Gamma(n+1) = n!$.

- By the theorem, $\Gamma(1) = 1$. Hence, for $n = 1, 2, \ldots$,

$$\Gamma(n+1) = n\Gamma(n) = n(n-1)\Gamma(n-1) = n(n-1)\cdots\Gamma(1) = n!.$$

Corollary

The gamma function is convex, continuous, and $x\Gamma(x) \to 1$, $\Gamma(x) \to \infty$ as $x \to 0^+$.

- The gamma function is log-convex. So it is convex. By a previous corollary, it must also be continuous. The continuity of Γ at 1 shows that

$$x\Gamma(x) = \Gamma(x+1) \xrightarrow{x \to 0^+} \Gamma(1) = 1.$$

Hence $\Gamma(x) \to \infty$ as $x \to 0^+$.

The Gamma Function and the Factorial Function

- Since $\Gamma(n+1) = n!$ for $n = 0, 1, 2, \ldots$, the gamma function can be considered to be an extension of the factorial function, even if the two functions are one unit out of phase with each other.
- There are, of course, infinitely many functions $f : (0, \infty) \to \mathbb{R}$ satisfying $f(n+1) = n!$ for $n = 0, 1, 2, \ldots$.
- The natural question that arises is:

  Is there some sense in which the gamma function is a unique extension of the factorial function?
- One answer is given by Artin's Characterization.

Artin's Characterization of the Gamma Function

Theorem (Artin's Characterization of the Gamma Function)

Let the function $f:(0,\infty)\to\mathbb{R}$ satisfy:

(i) $f(x+1)=xf(x)$ for $x>0$;

(ii) $f(1)=1$;

(iii) f is log-convex.

Then $f=\Gamma$.

- Conditions (i), (ii) imply that $f(n+1)=n!$ for $n=0,1,2,\ldots$.
  Let $0<x\leq 1$ and let n be any positive integer. Then the log-convexity of f and condition (i) show that

$$\begin{array}{rcl} f(n+1+x) & = & f((1-x)(n+1)+x(n+2)) \\ & \leq & f^{1-x}(n+1)f^{x}(n+2) \\ & = & f^{1-x}(n+1)((n+1)f(n+1))^{x} \\ & = & (n+1)^{x}f(n+1)=(n+1)^{x}n!. \end{array}$$

Artin's Characterization of the Gamma Function (Cont'd)

- We also have

$$\begin{array}{rcl} n! = f(n+1) & = & f(x(n+x)+(1-x)(n+1+x)) \\ & \le & f^x(n+x)f^{1-x}(n+1+x) \\ & = & (n+x)^{-x}f^x(n+1+x)f^{1-x}(n+1+x) \\ & = & (n+x)^{-x}f(n+1+x). \end{array}$$

But $f(n+1+x) = (n+x)(n-1+x)\cdots xf(x)$.
Therefore,

$$\left(1+\frac{x}{n}\right)^x \le \frac{(n+x)(n-1+x)\cdots xf(x)}{n!n^x} \le \left(1+\frac{1}{n}\right)^x.$$

Hence

$$f(x) = \lim_{n\to\infty} \frac{n!n^x}{(n+x)(n-1+x)\cdots x}, \quad \text{for } 0 < x \le 1.$$

Artin's Characterization of the Gamma Function (Cont'd)

- Suppose that $x>1$. Let m be the positive integer such that $0<x-m\leq 1$. Then, by condition (i) and what we have just proved,

$$\begin{aligned}
f(x) &= (x-1)\cdots(x-m)f(x-m)\\
&= (x-1)\cdots(x-m)\lim_{n\to\infty}\frac{n!n^{x-m}}{(n+x-m)(n-1+x-m)\cdots(x-m)}\\
&= \lim_{n\to\infty}\left(\frac{n!n^x}{(n+x)(n-1+x)\cdots x}\cdot\frac{(n+x)(n+x-1)\cdots(n+x-(m-1))}{n^m}\right)\\
&= \lim_{n\to\infty}\frac{n!n^x}{(n+x)(n-1+x)\cdots x}\cdot\\
&\qquad \lim_{n\to\infty}((1+\tfrac{x}{n})(1+\tfrac{x-1}{n})\cdots(1+\tfrac{1+x-m}{n}))\\
&= \lim_{n\to\infty}\frac{n!n^x}{(n+x)(n-1+x)\cdots x}.
\end{aligned}$$

Thus, for all $x>0$, $f(x)=\lim_{n\to\infty}\frac{n!n^x}{(n+x)(n-1+x)\cdots x}$.

This is a remarkable conclusion, since it shows that f is uniquely determined by conditions (i), (ii), and (iii).

Since Γ itself satisfies these three conditions, we must have $f=\Gamma$.

Gamma and Sine

Theorem

For every real x with $0 < x < 1$,

$$\Gamma(x)\Gamma(1-x) = \frac{\pi}{\sin \pi x}.$$

- Artin's Theorem shows that for $0 < x < 1$,

$$\begin{aligned}
\Gamma(x)\Gamma(1-x) &= \lim_{n\to\infty} \frac{n!n^x n!n^{1-x}}{(n+x)\cdots x(n+1-x)\cdots(1-x)} \\
&= \lim_{n\to\infty} \frac{n}{(n+1-x)x\frac{1}{1^2 2^2 \cdots n^2}(1+x)(1-x)\cdots(n+x)(n-x)} \\
&= \frac{1}{x\prod_{k=1}^{\infty}(1-\frac{x^2}{k^2})} \\
&= \frac{\pi}{\sin \pi x}. \qquad (\sin x = x\prod_{k=1}^{\infty}(1-\frac{x^2}{k^2\pi^2}))
\end{aligned}$$

- From the Theorem, we get $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

Legendre's Duplication Formula

Theorem (Legendre's Duplication Formula)

$$\Gamma\left(\frac{x}{2}\right)\Gamma\left(\frac{x+1}{2}\right)=\frac{\sqrt{\pi}}{2^{x-1}}\Gamma(x), \text{ for } x>0.$$

- Define a function $f:(0,\infty)\to\mathbb{R}$ by

$$f(x)=\frac{2^{x-1}}{\sqrt{\pi}}\Gamma\left(\frac{x}{2}\right)\Gamma\left(\frac{x+1}{2}\right), \text{ for } x>0.$$

  Then f is a product of log-convex functions. So it is itself log-convex. We also have, for all $x>0$:
  - $f(x+1)=\frac{2^x}{\sqrt{\pi}}\Gamma(\frac{x+1}{2})\Gamma(\frac{x+2}{2})=2\frac{2^{x-1}}{\sqrt{\pi}}\Gamma(\frac{x+1}{2})\frac{x}{2}\Gamma(\frac{x}{2})=xf(x)$;
  - $f(1)=\frac{1}{\sqrt{\pi}}\Gamma(\frac{1}{2})\Gamma(1)=\frac{1}{\sqrt{\pi}}\sqrt{\pi}1=1$.

  Thus, by Artin's Theorem, $f=\Gamma$.

Lemma for Stirling's Formula

Lemma

The sequence whose nth term is $\log n! - (n + \frac{1}{2})\log n + n$ converges.

- Let $a_n = \log n! - (n + \frac{1}{2})\log n + n$. First we show that the sequence (a_n) is decreasing. Then we show that it is bounded below. We note that, for $n = 1, 2, \ldots$, $a_n - a_{n+1} = (n + \frac{1}{2})\log(1 + \frac{1}{n}) - 1$. Since $\frac{1}{x}$ is convex on $(0, \infty)$, the area bounded by the graph of $y = \frac{1}{x}$, the x-axis, and the lines $x = n$, $x = n + 1$ exceeds that of the trapezoid bounded by the tangent to $y = \frac{1}{x}$ at the point $\left(n + \frac{1}{2}, \frac{1}{n + \frac{1}{2}}\right)$, the x-axis, and the lines $x = n$, $x = n + 1$; i.e.,

$$\log\left(1 + \frac{1}{n}\right) = \int_n^{n+1} \frac{dx}{x} > \frac{1}{n + \frac{1}{2}}.$$

  It now follows from the preceding formula, that $a_n - a_{n+1} > 0$. Hence the sequence (a_n) is decreasing.

Lemma for Stirling's Formula (Cont'd)

- Since $\log x$ is concave on $(0,\infty)$, the area bounded by the graph of $y=\log x$, the x-axis, and the lines $x=r-\frac{1}{2}$, $x=r+\frac{1}{2}$ for $r=1,2,\ldots$, is less than that of the trapezoid bounded by the tangent to $y=\log x$ at the point $(r,\log r)$, the x-axis, and the lines $x=r-\frac{1}{2}$, $x=r+\frac{1}{2}$, i.e., $\int_{r-\frac{1}{2}}^{r+\frac{1}{2}}\log x dx<\log r$. It follows easily that, for $n\geq 3$,

$$\begin{array}{rcl}\int_1^n\log x dx & = & \int_1^{1\frac{1}{2}}\log x dx+\int_{1\frac{1}{2}}^{2\frac{1}{2}}\log x dx+\cdots+\int_{n-\frac{3}{2}}^{n-\frac{1}{2}}\log x dx+\int_{n-\frac{1}{2}}^{n}\log x dx\\ & < & \frac{1}{2}\log 1\frac{1}{2}+\log 2+\cdots+\log(n-1)+\frac{1}{2}\log n\\ & < & \frac{1}{2}+\log(n!)-\frac{1}{2}\log n.\end{array}$$

Thus,

$$n\log n-n+1=\int_1^n\log x dx<\frac{1}{2}+\log n!-\frac{1}{2}\log n.$$

Hence $a_n=\log n!-(n+\frac{1}{2})\log n+n>\frac{1}{2}$. Thus, the decreasing sequence (a_n) is bounded below by $\frac{1}{2}$. So it converges.

Lemma for Stirling's Formula

Theorem (Stirling's Formula)

$n! \sim \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n}$.

- In the notation of the proof of the lemma, let for $n = 1, 2, \ldots$, $b_n = e^{a_n} = \frac{n!}{n^{n+\frac{1}{2}} e^{-n}}$. Then the sequence (b_n) converges to some $b > 0$. Thus,

$$\frac{(b_n)^2}{b_{2n}} = \frac{(n!)^2 (2n)^{2n+\frac{1}{2}} e^{-2n}}{n^{2n+1} e^{-2n} (2n)!} = \frac{2^{2n+\frac{1}{2}} (n!)^2}{n^{\frac{1}{2}} (2n)!} \to \frac{b^2}{b} = b, \text{ as } n \to \infty.$$

For $n = 1, 2, \ldots$, let $c_n = \frac{n! n^{\frac{1}{2}}}{(n+\frac{1}{2}) \cdots \frac{3}{2} \frac{1}{2}}$. Then $c_n \stackrel{n\to\infty}{\longrightarrow} \Gamma(\frac{1}{2}) = \sqrt{\pi}$. So

$$\frac{(b_n)^2}{b_{2n}} = \frac{n! n^{1/2} (2n+1)\sqrt{2}}{2n \frac{(2n+1)!}{2^{n+1} 2^n n!}} = c_n \left(1 + \frac{1}{2n}\right) \sqrt{2} \stackrel{n\to\infty}{\longrightarrow} \sqrt{2\pi}.$$

Hence, $b = \sqrt{2\pi}$. So $b_n = \frac{n!}{n^{n+\frac{1}{2}} e^{-n}} \stackrel{n\to\infty}{\longrightarrow} \sqrt{2\pi}$.

The Beta Function

- The **beta function** B is the real function of two variables defined by the equation

$$B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt, \text{ for } x,y > 0.$$

Theorem

The beta function has the following properties:

(i) $B(x+1,y) = \frac{x}{x+y}B(x,y)$ for $x,y > 0$;

(ii) $B(x,y)$ is a log-convex function of x for each fixed $y > 0$;

(iii) $B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$, for $x,y > 0$.

The Beta Function (Part (i))

(i) We have

$$\begin{aligned}
B(x+1,y) &= \int_0^1 t^x(1-t)^{y-1}dt \\
&= \int_0^1 \frac{t^x}{(1-t)^x}(1-t)^x(1-t)^{y-1}dt \\
&= \int_0^1 (1-t)^{x+y-1}\left(\frac{t}{1-t}\right)^x dt \\
&= \left[\frac{-(1-t)^{x+y}}{x+y}\left(\frac{t}{1-t}\right)^x\right]_0^1 - \int_0^1 \frac{-(1-t)^{x+y}}{x+y}\left[\left(\frac{t}{1-t}\right)^x\right]' dt \\
&= \left[\frac{-(1-t)^{x+y}}{x+y}\left(\frac{t}{1-t}\right)^x\right]_0^1 - \int_0^1 \frac{-(1-t)^{x+y}}{x+y}\left[x\frac{t^{x-1}}{(1-t)^{x-1}}\frac{1}{(1-t)^2}\right] dt \\
&= \left[\frac{-(1-t)^{x+y}}{x+y}\left(\frac{t}{1-t}\right)^x\right]_0^1 + \int_0^1 \frac{x}{x+y}t^{x-1}(1-t)^{y-1}dt \\
&= \frac{x}{x+y}B(x,y).
\end{aligned}$$

The Beta Function (Part (ii))

(ii) Let $a, b, y > 0$. Let $\lambda, \mu \geq 0$, with $\lambda + \mu = 1$.

By a previous theorem,

$$\begin{aligned} B(\lambda a + \mu b, y) &= \int_0^1 (t^{\lambda a + \mu b - 1}(1-t)^{y-1}) dt \\ &= \int_0^1 (t^{a-1}(1-t)^{y-1})^{\lambda} (t^{b-1}(1-t)^{y-1})^{\mu} dt \\ &\leq (\int_0^1 t^{a-1}(1-t)^{y-1} dt)^{\lambda} (\int_0^1 t^{b-1}(1-t)^{y-1} dt)^{\mu} \\ &= B^{\lambda}(a,y) B^{\mu}(b,y). \end{aligned}$$

Thus $B(x, y)$ is a log-convex function of x, for fixed y.

The Beta Function (Part (iii))

(iii) Let $y > 0$. Define a function $f_y : (0,\infty) \to \mathbb{R}$ by

$$f_y(x) = \frac{\Gamma(x+y)B(x,y)}{\Gamma(y)}, \text{ for } x > 0.$$

Then f_y is a product of log-convex functions. So it is log-convex. For $x > 0$,

$$\begin{aligned} f_y(x+1) &= \frac{\Gamma(x+y+1)B(x+1,y)}{\Gamma(y)} \\ &= \frac{[(x+y)\Gamma(x+y)]\frac{x}{x+y}B(x,y)}{\Gamma(y)} = xf_y(x); \\ f_y(1) &= \frac{\Gamma(1+y)B(1,y)}{\Gamma(y)} \\ &= y\int_0^1 (1+t)^{y-1}dt = 1. \end{aligned}$$

Thus, $f_y = \Gamma$ by Artin's Theorem. So $B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$, for $x, y > 0$.

An Integral Formula for B

- According to the definition,

$$B\left(\frac{n+1}{2},\frac{n+1}{2}\right)=\int_0^1 t^{\frac{n-1}{2}}(1-t)^{\frac{n-1}{2}}dt.$$

- Set $u=2t-1$. Then $dt=\frac{1}{2}du$, $t=\frac{1+u}{2}$, $1-t=\frac{1-u}{2}$ and $t=0,1$ correspond to $u=-1,1$, respectively.

  Thus, we get

$$\begin{array}{rcl} B(\frac{n+1}{2},\frac{n+1}{2}) & = & \int_{-1}^1 (\frac{1+u}{2})^{\frac{n-1}{2}}(\frac{1-u}{2})^{\frac{n-1}{2}}\frac{1}{2}du \\ & = & \frac{1}{2}\int_{-1}^1 \frac{1}{2^{n-1}}(1-u^2)^{\frac{n-1}{2}}du \\ & = & \frac{1}{2^{n-1}}\int_0^1 (1-u^2)^{\frac{n-1}{2}}du. \end{array}$$

A Recursive Formula for B

- We prove by induction on n that $B(\frac{n+1}{2},\frac{n+1}{2})=\frac{1}{2^n}B(\frac{1}{2},\frac{n+1}{2})$.
- For the base case, we prove the formula for $n=0$ and $n=1$.
  - For $n=0$, $B(\frac{0+1}{2},\frac{0+1}{2})=B(\frac{1}{2},\frac{1}{2})=\frac{1}{2^0}B(\frac{1}{2},\frac{0+1}{2})$.
  - For $n=1$, noting that $B(1,y)=\frac{1}{y}$, we get $B(\frac{1+1}{2},\frac{1+1}{2})=B(1,1)=1=\frac{1}{2}\frac{1}{1/2}=\frac{1}{2}B(\frac{1}{2},1)=\frac{1}{2^1}B(\frac{1}{2},\frac{1+1}{2})$.
- Assume the formula holds for some n.
- Then, recalling $B(x+1,y)=\frac{x}{x+y}B(x,y)$, we get

$$\begin{aligned}
B(\tfrac{(n+2)+1}{2},\tfrac{(n+2)+1}{2}) &= \frac{\frac{n+1}{2}}{\frac{n+1+n+3}{2}}B(\tfrac{n+1}{2},\tfrac{(n+2)+1}{2}) \\
&= \frac{n+1}{2(n+2)}\frac{n+1}{2(n+1)}B(\tfrac{n+1}{2},\tfrac{n+1}{2}) \\
&= \frac{n+1}{2^2(n+2)}\frac{1}{2^n}B(\tfrac{1}{2},\tfrac{n+1}{2}) \\
&= \frac{1}{2^{n+2}}\frac{\frac{n+1}{2}}{\frac{n+2}{2}}B(\tfrac{1}{2},\tfrac{n+1}{2}) \\
&= \frac{1}{2^{n+2}}B(\tfrac{1}{2},\tfrac{(n+2)+1}{2}).
\end{aligned}$$

Subsection 4

Convex Functions on $\mathbb{R}^n$

Convex Function on $\mathbb{R}^n$

- A real-valued function f defined on a non-empty convex set X in $\mathbb{R}^n$ is said to be **convex** if

$$f(\lambda \boldsymbol{x} + \mu \boldsymbol{y}) \le \lambda f(\boldsymbol{x}) + \mu f(\boldsymbol{y})$$

  whenever $\boldsymbol{x}, \boldsymbol{y} \in X$ and $\lambda, \mu \ge 0$ with $\lambda + \mu = 1$.
- The convexity of X ensures that $\lambda \boldsymbol{x} + \mu \boldsymbol{y} \in X$.
- A **concave function** is one whose negative is convex.
- Exactly as in the case of a convex function of a single real variable, each convex function $f : X \to \mathbb{R}^n$ satisfies *Jensen's inequality*:

$$f(\lambda_1 \boldsymbol{x}_1 + \cdots + \lambda_m \boldsymbol{x}_m) \le \lambda_1 f(\boldsymbol{x}_1) + \cdots + \lambda_m f(\boldsymbol{x}_m),$$

  whenever $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \in X$ and $\lambda_1, \ldots, \lambda_m \ge 0$ with $\lambda_1 + \cdots + \lambda_m = 1$.
- Affine transformations $f : \mathbb{R}^n \to \mathbb{R}$ and their restrictions to non-empty convex subsets of $\mathbb{R}^n$ provide important examples of convex functions.

Convexity of Distance of Convex Sets

- The distance function $d_X : \mathbb{R}^n \to \mathbb{R}$ of a non-empty set X in $\mathbb{R}^n$ was defined by the equation

$$d_X(\boldsymbol{u}) = \inf\{\|\boldsymbol{u} - \boldsymbol{x}\| : \boldsymbol{x} \in X\}, \text{ for } \boldsymbol{u} \in \mathbb{R}^n.$$

- We now assume that X is convex and show that in this case the resulting distance function d_X is convex.
- Let $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ and let $\lambda, \mu \geq 0$ with $\lambda + \mu = 1$. Then, for each $\varepsilon > 0$, there exist points $\boldsymbol{x}, \boldsymbol{y} \in X$ such that

$$\|\boldsymbol{u} - \boldsymbol{x}\| \leq d_X(\boldsymbol{u}) + \varepsilon \quad \text{and} \quad \|\boldsymbol{v} - \boldsymbol{y}\| \leq d_X(\boldsymbol{v}) + \varepsilon.$$

Since X is convex, $\lambda\boldsymbol{x} + \mu\boldsymbol{y} \in X$. So

$$\begin{array}{rcl} d_X(\lambda\boldsymbol{u} + \mu\boldsymbol{v}) & \leq & \|\lambda\boldsymbol{u} + \mu\boldsymbol{v} - (\lambda\boldsymbol{x} + \mu\boldsymbol{y})\| \\ & \leq & \lambda\|\boldsymbol{u} - \boldsymbol{x}\| + \mu\|\boldsymbol{v} - \boldsymbol{y}\| \\ & \leq & \lambda d_X(\boldsymbol{u}) + \mu d_x(\boldsymbol{v}) + \varepsilon. \end{array}$$

But $\varepsilon > 0$ is arbitrary. Hence, $d_X(\lambda\boldsymbol{u} + \mu\boldsymbol{v}) \leq \lambda d_X(\boldsymbol{u}) + \mu d_X(\boldsymbol{v})$.

Example: Introducing Graphs

- Consider the convex function $f(x_1) = x_1^2$ defined on $\mathbb{R}^1$.
- The graph of f is the parabola $\{(x_1, x_1^2) : x_1 \in \mathbb{R}\}$ in $\mathbb{R}^2$, which is clearly not convex.
- The set of points $\{(x_1, x) : x_1 \in \mathbb{R}, x \geq x_1^2\}$ in $\mathbb{R}^2$ which lie on or above the graph of f, however, is convex.
- Thus with this particular convex function of a single variable, we have associated a convex set in $\mathbb{R}^2$.
- We will show how the convexity of a real-valued function of n variables is equivalent to the convexity of a certain subset of $\mathbb{R}^{n+1}$.

Graphs and Epigraphs

- Let f be a real-valued function defined on a non-empty convex set X in $\mathbb{R}^n$.
- Then the **graph** of f is defined to be the subset

$$\{(x_1,\ldots,x_n,f(x_1,\ldots,x_n)):(x_1,\ldots,x_n)\in X\}$$

  of $\mathbb{R}^{n+1}$.
- The **epigraph** of f, denoted $\text{epi} f$, is defined to be the subset

$$\{(x_1,\ldots,x_n,x):(x_1,\ldots,x_n)\in X, x\geq f(x_1,\ldots,x_n)\}$$

  of $\mathbb{R}^{n+1}$.

Convex Functions and Their Epigraphs

Theorem

Let f be a real-valued function defined on a non-empty convex set X in $\mathbb{R}^n$. Then f is convex if and only if its epigraph is convex.

- For each point $\boldsymbol{x} = (x_1, \ldots, x_n)$ of $\mathbb{R}^n$ and for each scalar x, we denote by $(\boldsymbol{x}, x)$ the point $(x_1, \ldots, x_n, x)$ of $\mathbb{R}^{n+1}$.

  Suppose that f is convex. Let $(\boldsymbol{x}, x), (\boldsymbol{y}, y) \in \text{epi} f$. So $\boldsymbol{x}, \boldsymbol{y} \in X$ and $x \geq f(\boldsymbol{x})$, $y \geq f(\boldsymbol{y})$. Let $\lambda, \mu \geq 0$ with $\lambda + \mu = 1$. Then the convexity of f shows that

$$f(\lambda \boldsymbol{x} + \mu \boldsymbol{y}) \leq \lambda f(\boldsymbol{x}) + \mu f(\boldsymbol{y}) \leq \lambda x + \mu y.$$

  Thus the point $\lambda(\boldsymbol{x}, x) + \mu(\boldsymbol{y}, y) = (\lambda \boldsymbol{x} + \mu \boldsymbol{y}, \lambda x + \mu y)$ belongs to $\text{epi} f$.

  So $\text{epi} f$ is convex.

Convex Functions and Their Epigraphs (Converse)

- Conversely, suppose that epi$f$ is convex.

  Let $\boldsymbol{x}, \boldsymbol{y} \in X$ and let $\lambda, \mu \geq 0$ with $\lambda + \mu = 1$.

  Since epi$f$ is convex, the point

$$\lambda(\boldsymbol{x}, f(\boldsymbol{x})) + \mu(\boldsymbol{y}, f(\boldsymbol{y})) = (\lambda\boldsymbol{x} + \mu\boldsymbol{y}, \lambda f(\boldsymbol{x}) + \mu f(\boldsymbol{y}))$$

  belongs to epi$f$.

  Hence

$$f(\lambda\boldsymbol{x} + \mu\boldsymbol{y}) \leq \lambda f(\boldsymbol{x}) + \mu f(\boldsymbol{y}).$$

  This shows that f is a convex function.

Properties of Convex Functions and of Convex Sets

Theorem

Let $(f_i : i \in I)$ be a non-empty family of convex functions defined on a non-empty convex set X in $\mathbb{R}^n$ such that, for each $\boldsymbol{x}$ in X, the set $\{f_i(\boldsymbol{x}) : i \in I\}$ of real numbers is bounded above. Then the function $f : X \to \mathbb{R}$ defined by the equation $f(\boldsymbol{x}) = \sup\{f_i(\boldsymbol{x}) : i \in I\}$, for $\boldsymbol{x} \in X$, is convex.

- We observe that

$$\begin{aligned} \operatorname{epi} f &= \{(x_1,\ldots,x_n,x) : (x_1,\ldots,x_n) \in X, x \geq f(x_1,\ldots,x_n)\} \\ &= \{(x_1,\ldots,x_n,x) : (x_1,\ldots,x_n) \in X, x \geq f_i(x_1,\ldots,x_n) \text{ for } i \in I\} \\ &= \textstyle\bigcap_{i\in I}\{(x_1,\ldots,x_n,x) : (x_1,\ldots,x_n) \in X, x \geq f_i(x_1,\ldots,x_n)\} \\ &= \textstyle\bigcap_{i\in I} \operatorname{epi} f_i. \end{aligned}$$

  The preceding theorem shows that all of the sets $\operatorname{epi} f_i$ are convex. Hence so is their intersection $\operatorname{epi} f$. Thus, by the same theorem f is a convex function.

Linear Combinations of Convex Functions

Theorem

Let f, g be convex functions defined on a non-empty convex subset X of $\mathbb{R}^n$ and let $\alpha, \beta \geq 0$. Then the function $\alpha f + \beta g$ is convex.

- Let $\boldsymbol{x}, \boldsymbol{y} \in X$ and let $\lambda, \mu \geq 0$ with $\lambda + \mu = 1$.

  Then

$$\begin{array}{rcl}(\alpha f + \beta g)(\lambda \boldsymbol{x} + \mu \boldsymbol{y}) & = & \alpha f(\lambda \boldsymbol{x} + \mu \boldsymbol{y}) + \beta g(\lambda \boldsymbol{x} + \mu \boldsymbol{y}) \\ & \leq & \alpha(\lambda f(\boldsymbol{x}) + \mu f(\boldsymbol{y})) + \beta(\lambda g(\boldsymbol{x}) + \mu g(\boldsymbol{y})) \\ & = & \lambda(\alpha f + \beta g)(\boldsymbol{x}) + \mu(\alpha f + \beta g)(\boldsymbol{y}).\end{array}$$

Composition of Convex and Increasing Convex Functions

Theorem

Let f be a convex function defined on a non-empty convex set X in $\mathbb{R}^n$ and let $g : I \to \mathbb{R}$ be an increasing convex function defined on an interval I of $\mathbb{R}$ which contains the image $f(X)$ of X under f. Then the composite function $g \circ f : X \to \mathbb{R}$ is convex.

- Let $\boldsymbol{x}, \boldsymbol{y} \in X$ and let $\lambda, \mu \geq 0$, with $\lambda + \mu = 1$.

  Then

$$\begin{aligned} (g \circ f)(\lambda \boldsymbol{x} + \mu \boldsymbol{y}) &= g(f(\lambda \boldsymbol{x} + \mu \boldsymbol{y})) \\ &\leq g(\lambda f(\boldsymbol{x}) + \mu f(\boldsymbol{y})) \\ &\leq \lambda g(f(\boldsymbol{x})) + \mu g(f(\boldsymbol{y})) \\ &= \lambda (g \circ f)(\boldsymbol{x}) + \mu (g \circ f)(\boldsymbol{y}). \end{aligned}$$

Supporting Affine Transformations

- Let f be a real-valued function defined on a convex set X in $\mathbb{R}^n$ and let $\boldsymbol{x}_0 \in X$.
- Then an affine transformation $T : \mathbb{R}^n \to \mathbb{R}$ is said to **support f at $\boldsymbol{x}_0$** if $T(\boldsymbol{x}_0) = f(\boldsymbol{x}_0)$ and $T(\boldsymbol{x}) \le f(\boldsymbol{x})$ for all $\boldsymbol{x} \in X$.
- The geometrical interpretation of T supporting f at $\boldsymbol{x}_0$ is clear.

  The set
  $$\{(x_1, \ldots, x_n, T(x_1, \ldots, x_n)) : (x_1, \ldots, x_n) \in \mathbb{R}^n\}$$
  is a hyperplane in $\mathbb{R}^{n+1}$ that passes through the point $(\boldsymbol{x}_0, f(\boldsymbol{x}_0))$ and lies on or below the graph
  $$\{(x_1, \ldots, x_n, f(x_1, \ldots, x_n)) : (x_1, \ldots, x_n) \in X\}$$
  of f.

Convexity and Support

Theorem

Let f be a real-valued function defined on a non-empty open convex set X in $\mathbb{R}^n$. Then f is convex if and only if it has support at each point of X.

- Suppose that f has support at each point of X. Let $\mathbf{x},\mathbf{y}\in X$ and let $\lambda,\mu\geq 0$ with $\lambda+\mu=1$. Then there is an affine transformation $T:\mathbb{R}^n\to\mathbb{R}$ which supports f at $\lambda\mathbf{x}+\mu\mathbf{y}$. Hence

$$f(\lambda\mathbf{x}+\mu\mathbf{y})=T(\lambda\mathbf{x}+\mu\mathbf{y})=\lambda T(\mathbf{x})+\mu T(\mathbf{y})\leq\lambda f(\mathbf{x})+\mu f(\mathbf{y}).$$

  This shows that f is convex.

  Conversely, suppose that f is convex and that $\mathbf{x}_0\in X$. Since f is convex, its epigraph epi$f$ is a convex set in $\mathbb{R}^{n+1}$. Now $(\mathbf{x}_0,f(\mathbf{x}_0))$ is a boundary point of epi$f$. So there exists a support hyperplane H to epi$f$ at $(\mathbf{x}_0,f(\mathbf{x}_0))$.

Convexity and Support (Converse)

- Suppose that H has equation $a_1x_1 + \cdots + a_nx_n + a_{n+1}x_{n+1} = a_0$. Suppose, also, that $a_1x_1 + \cdots + a_nx_n + a_{n+1}x_{n+1} \geq a_0$, whenever $(x_1,\ldots,x_n) \in X$ and $x_{n+1} \geq f(x_1,\ldots,x_n)$.
  - We have $a_{n+1} \neq 0$. Otherwise, the hyperplane in $\mathbb{R}^n$ with equation $a_1x_1 + \cdots + a_nx_n = a_0$ supports X at $\boldsymbol{x}_0$. This is impossible because $\boldsymbol{x}_0$ is an interior point of X.
  - For each $(x_1,\ldots,x_n) \in X$, $a_1x_1 + \cdots + a_nx_n + a_{n+1}\lambda \geq a_0$ for all $\lambda \geq f(x_1,\ldots,x_n)$. Hence, $a_{n+1} > 0$.

  Define an affine transformation $T : \mathbb{R}^n \to \mathbb{R}$ by the equation

  $$T(x_1,\ldots,x_n) = \frac{1}{a_{n+1}}(a_0 - a_1x_1 - \cdots - a_nx_n), \text{ for } (x_1,\ldots,x_n) \in \mathbb{R}^n.$$

  Since H supports epi$f$ at $(\boldsymbol{x}_0, f(\boldsymbol{x}_0))$ and $a_{n+1} > 0$,
  - $T(\boldsymbol{x}_0) = \frac{1}{a_{n+1}}(a_0 - a_1x_1^0 - \cdots - a_nx_n^0) = \frac{a_{n+1}x_{n+1}^0}{a_{n+1}} = x_{n+1}^0 = f(\boldsymbol{x}_0)$;
  - For all $\boldsymbol{x} \in X$, $T(\boldsymbol{x}) = \frac{1}{a_{n+1}}(a_0 - a_1x_1 - \cdots - a_nx_n) \leq \frac{a_{n+1}x_{n+1}}{a_{n+1}} = f(\boldsymbol{x})$.

  Thus, T supports f at $\boldsymbol{x}_0$.

Positively Homogeneous Functions

- Many of the functions which arise naturally in convexity are real-valued functions f defined on a convex cone X in $\mathbb{R}^n$ (often $\mathbb{R}^n$ itself) that satisfy the equation

$$f(\lambda \boldsymbol{x}) = \lambda f(\boldsymbol{x}), \text{ for all } \boldsymbol{x} \in X \text{ and all } \lambda \geq 0.$$

- Such functions are called **positively homogeneous**.
- The most important example of such a function is the norm mapping $\|\cdot\|$, which is defined on the whole of $\mathbb{R}^n$.

Positive Homogeneous vs. Convex Functions

Theorem

Let f be a positively homogeneous function defined on a convex cone X in $\mathbb{R}^n$. Then f is convex if and only if $f(\boldsymbol{x}+\boldsymbol{y}) \le f(\boldsymbol{x}) + f(\boldsymbol{y})$ for all $\boldsymbol{x}, \boldsymbol{y} \in X$.

- Suppose that f is convex. Let $\boldsymbol{x}, \boldsymbol{y} \in X$. Then

$$\frac{1}{2} f(\boldsymbol{x}+\boldsymbol{y}) = f\left(\frac{1}{2}\boldsymbol{x} + \frac{1}{2}\boldsymbol{y}\right) \le \frac{1}{2} f(\boldsymbol{x}) + \frac{1}{2} f(\boldsymbol{y}).$$

So $f(\boldsymbol{x}+\boldsymbol{y}) \le f(\boldsymbol{x}) + f(\boldsymbol{y})$.

Conversely, suppose that $f(\boldsymbol{x}+\boldsymbol{y}) \le f(\boldsymbol{x}) + f(\boldsymbol{y})$ for all $\boldsymbol{x}, \boldsymbol{y} \in X$. Then, for all $\boldsymbol{x}, \boldsymbol{y} \in X$ and for all $\lambda, \mu \ge 0$ with $\lambda + \mu = 1$,

$$f(\lambda \boldsymbol{x} + \mu \boldsymbol{y}) \le f(\lambda \boldsymbol{x}) + f(\mu \boldsymbol{y}) = \lambda f(\boldsymbol{x}) + \mu f(\boldsymbol{y}).$$

This shows that f is convex.

The Level Sets of a Function

- Let f be a real-valued function defined on a non-empty convex set X in $\mathbb{R}^n$.
- Then, for each scalar α, the **level set L_α of f at height α** is the set defined by the equation

$$L_\alpha = \{\mathbf{x} \in X : f(\mathbf{x}) \leq \alpha\}.$$

- We show that each level set L_α of a convex function $f : X \to \mathbb{R}$ is convex.

  Let $\mathbf{x}, \mathbf{y} \in L_\alpha$ and let $\lambda, \mu \geq 0$ with $\lambda + \mu = 1$. Then, since f is convex,

$$f(\lambda \mathbf{x} + \mu \mathbf{y}) \leq \lambda f(\mathbf{x}) + \mu f(\mathbf{y}) \leq \lambda \alpha + \mu \alpha = \alpha.$$

  Thus $\lambda \mathbf{x} + \mu \mathbf{y} \in L_\alpha$ and L_α is convex.
- There exist non-convex functions all of whose level sets are convex.

  An example is the cube function defined on the real line.

Non-Negative Positive Homogeneous Case

Theorem

Let f be a non-negative positively homogeneous function defined on a convex cone X in $\mathbb{R}^n$ such that the level set $\{\boldsymbol{x} \in X : f(\boldsymbol{x}) \leq 1\}$ is convex. Then f is a convex function.

- We use the criterion of the preceding theorem to show that f is convex. Let $\boldsymbol{x}, \boldsymbol{y} \in X$. Choose scalars α, β such that $\alpha > f(\boldsymbol{x})$, $\beta > f(\boldsymbol{y})$. Since f is non-negative and positively homogeneous, $f(\frac{\boldsymbol{x}}{\alpha}) \leq 1$ and $f(\frac{\boldsymbol{y}}{\beta}) \leq 1$. Thus $\frac{\boldsymbol{x}}{\alpha}$ and $\frac{\boldsymbol{y}}{\beta}$ lie in the level set of f at height 1. The assumed convexity of this level set shows that

$$\begin{aligned} \frac{1}{\alpha+\beta} f(\boldsymbol{x}+\boldsymbol{y}) &= f\left(\frac{\boldsymbol{x}+\boldsymbol{y}}{\alpha+\beta}\right) = f\left(\frac{\alpha}{\alpha+\beta}\frac{\boldsymbol{x}}{\alpha} + \frac{\beta}{\alpha+\beta}\frac{\boldsymbol{y}}{\beta}\right) \\ &\leq \frac{\alpha}{\alpha+\beta} f\left(\frac{\boldsymbol{x}}{\alpha}\right) + \frac{\beta}{\alpha+\beta} f\left(\frac{\boldsymbol{y}}{\beta}\right) \leq \frac{\alpha}{\alpha+\beta} + \frac{\beta}{\alpha+\beta} = 1. \end{aligned}$$

  Hence $f(\boldsymbol{x}+\boldsymbol{y}) \leq \alpha+\beta$ whenever $\alpha > f(\boldsymbol{x})$, $\beta > f(\boldsymbol{y})$. So $f(\boldsymbol{x}+\boldsymbol{y}) \leq f(\boldsymbol{x}) + f(\boldsymbol{y})$. This shows that f is convex.

Example

- Let $p \geq 1$. Define a function f on the nonnegative orthant X of $\mathbb{R}^n$ by the equation
$$f(x_1,\ldots,x_n) = (x_1^p + \cdots + x_n^p)^{1/p}, \text{ for } x_1,\ldots,x_n \geq 0.$$
Then f is non-negative and positively homogeneous.
It follows from a previous theorem and the fact that the function x^p is convex on the interval $[0,\infty)$, that the function $f^p : X \to \mathbb{R}$ is convex.
Hence the level set $\{\boldsymbol{x} \in X : f^p(\boldsymbol{x}) \leq 1\} = \{\boldsymbol{x} \in X : f(\boldsymbol{x}) \leq 1\}$ is convex.
By the preceding theorem, f is convex.
Let $\boldsymbol{x} = (x_1,\ldots,x_n)$, $\boldsymbol{y} = (y_1,\ldots,y_n)$ belong to X.
Then, by a previous theorem, $f(\boldsymbol{x}+\boldsymbol{y}) \leq f(\boldsymbol{x}) + f(\boldsymbol{y})$. That is,
$$((x_1+y_1)^p + \cdots + (x_n+y_n)^p)^{1/p} \leq (x_1^p + \cdots + x_n^p)^{1/p} + (y_1^p + \cdots + y_n^p)^{1/p}.$$
We have re-proved Minkowski's inequality.

Subsection 5

Continuity and Differentiability

Convex Functions on Open Convex Sets

- Let f be a convex function defined on an open convex set X in $\mathbb{R}^n$.
- Let $\boldsymbol{x} \in X$ and $\boldsymbol{y} \in \mathbb{R}^n$.
- Then the set $I = \{\lambda \in \mathbb{R} : \boldsymbol{x} + \lambda \boldsymbol{y} \in X\}$ is an open interval of $\mathbb{R}$ which contains the origin.
- The function $g : I \to \mathbb{R}$ defined by the equation

$$g(\lambda) = f(\boldsymbol{x} + \lambda \boldsymbol{y}), \text{ for } \lambda \in I,$$

is convex.

To see that g is convex, let $a, b \in I$ and let $\lambda, \mu \geq 0$ with $\lambda + \mu = 1$. Then

$$\begin{aligned} g(\lambda a + \mu b) &= f(\boldsymbol{x} + (\lambda a + \mu b)\boldsymbol{y}) \\ &= f(\lambda(\boldsymbol{x} + a\boldsymbol{y}) + \mu(\boldsymbol{x} + b\boldsymbol{y})) \\ &\leq \lambda f(\boldsymbol{x} + a\boldsymbol{y}) + \mu f(\boldsymbol{x} + b\boldsymbol{y}) \\ &= \lambda g(a) + \mu g(b). \end{aligned}$$

- Thus $g'_+(0) = \lim_{\lambda \to 0^+} \frac{g(\lambda) - g(0)}{\lambda} = \lim_{\lambda \to 0^+} \frac{f(\boldsymbol{x} + \lambda \boldsymbol{y}) - f(\boldsymbol{x})}{\lambda}$ exists.

Continuity

Theorem

Let f be a convex function defined on a non-empty open convex set X in $\mathbb{R}^n$. Then f is continuous on X.

- Let $\boldsymbol{x}_0 \in X$ and let $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m$ be the vertices of some polytope P which is contained in X and has $\boldsymbol{x}_0$ as an interior point. Choose $r > 0$ such that $B[\boldsymbol{x}_0; r] \subseteq P$. Each point $\boldsymbol{x}$ of $B[\boldsymbol{x}_0; r]$ can be expressed in the form $\boldsymbol{x} = \lambda_1 \boldsymbol{y}_1 + \cdots + \lambda_m \boldsymbol{y}_m$ for some $\lambda_1, \ldots, \lambda_m \geq 0$ with $\lambda_1 + \cdots + \lambda_m = 1$. Setting $M = \max\{f(\boldsymbol{y}_1), \ldots, f(\boldsymbol{y}_m)\}$ and applying Jensen's inequality to f, we get

$$\begin{aligned} f(\boldsymbol{x}) &= f(\lambda_1 \boldsymbol{y}_1 + \cdots + \lambda_m \boldsymbol{y}_m) \\ &\leq \lambda_1 f(\boldsymbol{y}_1) + \cdots + \lambda_m f(\boldsymbol{y}_m) \\ &\leq \lambda_1 M + \cdots + \lambda_m M = M. \end{aligned}$$

  Hence f is bounded above by M on the closed ball $B[\boldsymbol{x}_0; r]$.

Continuity (Cont'd)

- Let $\boldsymbol{x} \in \mathbb{R}^n$ satisfy the inequalities $0 < \|\boldsymbol{x} - \boldsymbol{x}_0\| \le r$. Then the function $g : [-r, r] \to \mathbb{R}$ defined by the equation

$$g(t) = f\left(\boldsymbol{x}_0 + t\frac{\boldsymbol{x} - \boldsymbol{x}_0}{\|\boldsymbol{x} - \boldsymbol{x}_0\|}\right), \text{ for } -r \le t \le r,$$

is convex, and $g(t) \le M$ for $-r \le t \le r$. By a previous corollary,

$$\begin{aligned} -\frac{M - g(0)}{r} \le \frac{g(-r) - g(0)}{-r} &\le \frac{g(\|\boldsymbol{x} - \boldsymbol{x}_0\|) - g(0)}{\|\boldsymbol{x} - \boldsymbol{x}_0\|} \\ &\le \frac{g(r) - g(0)}{r} \le \frac{M - g(0)}{r}. \end{aligned}$$

Hence

$$|f(\boldsymbol{x}) - f(\boldsymbol{x}_0)| = |g(\|\boldsymbol{x} - \boldsymbol{x}_0\|) - g(0)| \le \frac{M - f(\boldsymbol{x}_0)}{r}\|\boldsymbol{x} - \boldsymbol{x}_0\|.$$

Thus, if $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, \ldots$ is a sequence of points of X that converges to $\boldsymbol{x}_0$, then $f(\boldsymbol{x}_k) \to f(\boldsymbol{x}_0)$ as $k \to \infty$. So f is continuous at $\boldsymbol{x}_0$.

Partial Derivatives

- Let f be a real-valued function defined on an open set X in $\mathbb{R}^n$ and let $\boldsymbol{x} = (x_1, \ldots, x_n)$ be a point of X.
- Recall that the i**th partial derivative** $\frac{\partial f}{\partial x_i}$ of f at $\boldsymbol{x}$, when it exists, is the derivative at x_i of the function of a single variable obtained by regarding f as a function of its ith variable only, the remaining $n-1$ variables being held fixed to their values at $\boldsymbol{x}$.
- Thus, for $i = 1, \ldots, n$,

$$\frac{\partial f}{\partial x_i} = \lim_{\lambda \to 0} \frac{f(x_1, \ldots, x_{i-1}, x_i + \lambda, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, x_n)}{\lambda}.$$

- More succinctly,

$$\frac{\partial f}{\partial x_i}(\boldsymbol{x}) = \lim_{\lambda \to 0} \frac{f(\boldsymbol{x} + \lambda \boldsymbol{e}_i) - f(\boldsymbol{x})}{\lambda},$$

  where $\boldsymbol{e}_i$ denotes the ith elementary vector in $\mathbb{R}^n$.

Directional Derivatives

- For the directional derivative, which is a natural generalization of a partial derivative, we simply consider the above limit with an arbitrary vector $\boldsymbol{y}$ in $\mathbb{R}^n$ replacing the vector $\boldsymbol{e}_i$.
- The **directional derivative of f at $\boldsymbol{x}$ relative to $\boldsymbol{y}$** is defined to be the limit
$$\lim_{\lambda \to 0} \frac{f(\boldsymbol{x}+\lambda \boldsymbol{y})-f(\boldsymbol{x})}{\lambda},$$
whenever this limit exists.
- Thus the partial derivative $\frac{\partial f}{\partial x_i}$ is simply the directional derivative of f relative to $\boldsymbol{e}_i$.

One-Sided Directional Derivatives

- A convex function defined on an open interval of $\mathbb{R}$ need not be differentiable, but it always possesses both one-sided derivatives.
- The **one-sided directional derivative of f at $\boldsymbol{x}$ relative to $\boldsymbol{y}$** is defined to be the limit

$$f'(\boldsymbol{x};\boldsymbol{y}) = \lim_{\lambda \to 0^+} \frac{f(\boldsymbol{x}+\lambda \boldsymbol{y}) - f(\boldsymbol{x})}{\lambda},$$

  provided that this limit exists.
- We have

$$-f'(\boldsymbol{x};-\boldsymbol{y}) = \lim_{\lambda \to 0^-} \frac{f(\boldsymbol{x}+\lambda \boldsymbol{y}) - f(\boldsymbol{x})}{\lambda}.$$

- So the directional derivative of f at $\boldsymbol{x}$ relative to $\boldsymbol{y}$ exists if and only if both of the one-sided directional derivatives $f'(\boldsymbol{x};\boldsymbol{y})$ and $f'(\boldsymbol{x};-\boldsymbol{y})$ exist and satisfy the relation $f'(\boldsymbol{x};\boldsymbol{y}) = -f'(\boldsymbol{x};-\boldsymbol{y})$.

Notation and Remark

- If, for some $\boldsymbol{x} \in X$, the one-sided directional derivative $f'(\boldsymbol{x};\boldsymbol{y})$ exists for each $\boldsymbol{y} \in \mathbb{R}^n$, we write $f'(\boldsymbol{x};\ )$ to denote the function $f'(\boldsymbol{x};\ ) : \mathbb{R}^n \to \mathbb{R}$ whose value at $\boldsymbol{y}$ is $f'(\boldsymbol{x};\boldsymbol{y})$.
- The remarks before the preceding theorem show that, for each convex function $f : X \to \mathbb{R}^n$, the one-sided directional derivative $f'(\boldsymbol{x};\boldsymbol{y})$ exists for every $\boldsymbol{x}$ in the interior of X and for all $\boldsymbol{y}$ in $\mathbb{R}^n$.

Example

- Consider the convex function $f : \mathbb{R}^n \to \mathbb{R}$ defined, for each $\boldsymbol{x} = (x_1, \ldots, x_n)$, by

$$f(\boldsymbol{x}) = \|\boldsymbol{x}\|^2 = x_1^2 + \cdots + x_n^2.$$

- Then, for each $\boldsymbol{y} = (y_1, \ldots, y_n)$ in $\mathbb{R}^n$,

$$\begin{aligned} f'(\boldsymbol{x}; \boldsymbol{y}) &= \lim_{\lambda \to 0^+} \frac{f(\boldsymbol{x} + \lambda \boldsymbol{y}) - f(\boldsymbol{x})}{\lambda} \\ &= \lim_{\lambda \to 0^+} \frac{2\lambda(x_1 y_1 + \cdots + x_n y_n) + \lambda^2 (y_1^2 + \cdots + y_n^2)}{\lambda} \\ &= 2x_1 y_1 + \cdots + 2x_n y_n \\ &= 2\boldsymbol{x} \cdot \boldsymbol{y}. \end{aligned}$$

- Thus $f'(\boldsymbol{x}; \boldsymbol{y})$ exists and equals $2\boldsymbol{x} \cdot \boldsymbol{y}$.
- For this particular function, the (two-sided) directional derivative of f at $\boldsymbol{x}$ relative to $\boldsymbol{y}$ exists.
- The one-sided derivative $f'(\boldsymbol{x};\ ) : \mathbb{R}^n \to \mathbb{R}$ is linear for each $\boldsymbol{x}$ in $\mathbb{R}^n$.

Properties of Directional Derivative Function

Theorem

Let f be a convex function defined on an open convex set X in $\mathbb{R}^n$ and let $\boldsymbol{x} \in X$. Then $f'(\boldsymbol{x};\ )$ is a positively homogeneous convex function such that $f'(\boldsymbol{x};\boldsymbol{y}) \geq -f'(\boldsymbol{x};-\boldsymbol{y})$ for all $\boldsymbol{y}$ in $\mathbb{R}^n$. If f has a directional derivative at $\boldsymbol{x}$ relative to $\boldsymbol{y}$, then $f'(\boldsymbol{x};\lambda\boldsymbol{y}) = \lambda f'(\boldsymbol{x};\boldsymbol{y})$ for all scalars λ.

- Let $\mu > 0$ and let $\boldsymbol{y} \in \mathbb{R}^n$. Then

$$f'(\boldsymbol{x};\mu\boldsymbol{y}) = \lim_{\lambda\to 0^+} \frac{f(\boldsymbol{x}+\lambda\mu\boldsymbol{y}) - f(\boldsymbol{x})}{\lambda} = \lim_{\lambda\to 0^+} \mu \frac{f(\boldsymbol{x}+\lambda\mu\boldsymbol{y}) - f(\boldsymbol{x})}{\lambda\mu} = \mu f'(\boldsymbol{x};\boldsymbol{y}).$$

  This shows that $f'(\boldsymbol{x};\ )$ is positively homogeneous.

  Let $\boldsymbol{y},\boldsymbol{z} \in \mathbb{R}^n$. By the convexity of f,

$$\begin{aligned} f'(\boldsymbol{x};\boldsymbol{y}+\boldsymbol{z}) &= \lim_{\lambda\to 0^+} \frac{f(\boldsymbol{x}+\lambda(\boldsymbol{y}+\boldsymbol{z}))-f(\boldsymbol{x})}{\lambda} \\ &\leq \lim_{\lambda\to 0^+} \left(\frac{1}{2}\frac{f(\boldsymbol{x}+2\lambda\boldsymbol{y})-f(\boldsymbol{x})}{\lambda} + \frac{1}{2}\frac{f(\boldsymbol{x}+2\lambda\boldsymbol{z})-f(\boldsymbol{x})}{\lambda}\right) \\ &= f'(\boldsymbol{x};\boldsymbol{y}) + f'(\boldsymbol{x};\boldsymbol{z}). \end{aligned}$$

Properties of Directional Derivative Function (Cont'd)

- A previous theorem shows that $f'(\boldsymbol{x};\ )$ is convex.

  By what we have just proved, for each $\boldsymbol{y}$ in $\mathbb{R}^n$,

  $$0 = f'(\boldsymbol{x};\boldsymbol{0}) = f'(\boldsymbol{x};\boldsymbol{y}-\boldsymbol{y}) \le f'(\boldsymbol{x};\boldsymbol{y}) + f'(\boldsymbol{x};-\boldsymbol{y}).$$

  Hence $f'(\boldsymbol{x};\boldsymbol{y}) \ge -f'(\boldsymbol{x};-\boldsymbol{y})$.

  Suppose, finally, that f has a directional derivative at $\boldsymbol{x}$ relative to $\boldsymbol{y}$. Then $f'(\boldsymbol{x};\boldsymbol{y}) = -f'(\boldsymbol{x};\boldsymbol{y})$. If $\lambda < 0$, then, since f is positively homogeneous,

  $$f'(\boldsymbol{x};\lambda\boldsymbol{y}) = f'(\boldsymbol{x};(-\lambda)(-\boldsymbol{y})) = -\lambda f'(\boldsymbol{x};-\boldsymbol{y}) = \lambda f'(\boldsymbol{x};\boldsymbol{y}).$$

  Hence $f'(\boldsymbol{x};\lambda\boldsymbol{y}) = \lambda f'(\boldsymbol{x};\boldsymbol{y})$ for all scalars λ.

Differentiability and Gradient

- Suppose now that f is a real-valued function defined on an open set X in $\mathbb{R}^n$ and that $\boldsymbol{x}$ is a point of X.
- Recall that f is **differentiable at $\boldsymbol{x}$** if there exists a vector $\boldsymbol{x}'$ (necessarily unique) such that

$$\lim_{\boldsymbol{u}\to\boldsymbol{0}} \frac{f(\boldsymbol{x}+\boldsymbol{u})-f(\boldsymbol{x})-\boldsymbol{x}'\cdot\boldsymbol{u}}{\|\boldsymbol{u}\|}=0.$$

- When such an $\boldsymbol{x}'$ exists it is called the **gradient** of f at $\boldsymbol{x}$.

Gradient and Directional Derivatives

- Suppose that f is a real-valued function defined on an open set X in $\mathbb{R}^n$ and that $\boldsymbol{x}$ is a point of X.
- Let f be differentiable at $\boldsymbol{x}$ with gradient $\boldsymbol{x}'$ there.
- Then, for any non-zero vector $\boldsymbol{y}$ in $\mathbb{R}^n$,

$$0 = \lim_{\lambda \to 0} \frac{|f(\boldsymbol{x}+\lambda\boldsymbol{y}) - f(\boldsymbol{x}) - \boldsymbol{x}' \cdot (\lambda\boldsymbol{y})|}{\|\lambda\boldsymbol{y}\|}$$
$$= \lim_{\lambda \to 0} \frac{1}{\|\boldsymbol{y}\|} \left| \frac{f(\boldsymbol{x}+\lambda\boldsymbol{y}) - f(\boldsymbol{x})}{\lambda} - \boldsymbol{x}' \cdot \boldsymbol{y} \right|.$$

- This shows that f possesses a directional derivative at $\boldsymbol{x}$ relative to $\boldsymbol{y}$ and that $f'(\boldsymbol{x};\boldsymbol{y}) = \boldsymbol{x}' \cdot \boldsymbol{y}$.
- So $f'(\boldsymbol{x};\ )$ is linear.

Directional Derivatives and Differentiability

- The existence of the directional derivatives of f at $\boldsymbol{x}$ relative to all points $\boldsymbol{y}$ in $\mathbb{R}^n$ neither guarantees that f is differentiable nor that $f'(\boldsymbol{x};\ )$ is linear.

Theorem

Suppose that a convex function f defined on an open convex set X in $\mathbb{R}^n$ possesses all its partial derivatives $\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}$ at some point $\boldsymbol{x}$ of X. Then f is differentiable at $\boldsymbol{x}$.

- Let $r > 0$ be such that $B(\boldsymbol{x}; r) \subseteq X$. For each $\boldsymbol{u} = (u_1, \ldots, u_n)$ in $B(\mathbf{0}; r)$, let

$$\psi(\boldsymbol{u}) = f(\boldsymbol{x} + \boldsymbol{u}) - f(\boldsymbol{x}) - \left(\frac{\partial f}{\partial x_1} u_1 + \cdots + \frac{\partial f}{\partial x_n} u_n\right).$$

  Then ψ is convex on $B(\mathbf{0}; r)$.

Directional Derivatives and Differentiability (Cont'd)

- For each $i=1,\ldots,n$, define a function θ_i on $B(\mathbf{0};r)$ at a point $\boldsymbol{u}=(u_1,\ldots,u_n)$ of $B(\mathbf{0};r)$ as follows:

$$\theta_i(\boldsymbol{u})=\begin{cases}\frac{\psi(u_i\boldsymbol{e}_i)}{u_i}, & \text{for } u_i\neq 0,\\ 0, & \text{for } u_i=0.\end{cases}$$

Then $\theta_i(\boldsymbol{u})\to 0$ as $\boldsymbol{u}\to\mathbf{0}$. For each $\boldsymbol{u}=(u_1,\ldots,u_n)$ such that $n\|\boldsymbol{u}\|<r$, Jensen's inequality applied to the convex function ψ shows that

$$\begin{aligned}\psi(\boldsymbol{u}) &= \psi\left(\tfrac{1}{n}(nu_1\boldsymbol{e}_1)+\cdots+\tfrac{1}{n}(nu_n\boldsymbol{e}_n)\right)\leq\tfrac{1}{n}\psi(nu_1\boldsymbol{e}_1)+\cdots+\tfrac{1}{n}\psi(nu_n\boldsymbol{e}_n)\\ &= u_1\theta_1(n\boldsymbol{u})+\cdots+u_n\theta_n(n\boldsymbol{u})\leq\|\boldsymbol{u}\|(|\theta_1(n\boldsymbol{u})|+\cdots+|\theta_n(n\boldsymbol{u})|).\end{aligned}$$

But $0=\psi\left(\frac{1}{2}\boldsymbol{u}+\frac{1}{2}(-\boldsymbol{u})\right)\leq\frac{1}{2}\psi(\boldsymbol{u})+\frac{1}{2}\psi(-\boldsymbol{u})$. So $\psi(\boldsymbol{u})\geq-\psi(-\boldsymbol{u})$. Thus,

$$-\|\boldsymbol{u}\|(|\theta_1(-n\boldsymbol{u})|+\cdots+|\theta_n(-n\boldsymbol{u})|)\leq\psi(\boldsymbol{u})\leq\|\boldsymbol{u}\|(|\theta_1(n\boldsymbol{u})|+\cdots+|\theta_n(n\boldsymbol{u})|).$$

So $\frac{\psi(\boldsymbol{u})}{\|\boldsymbol{u}\|}\to 0$ as $\boldsymbol{u}\to\mathbf{0}$. Hence f has gradient $(\frac{\partial f}{\partial x_1},\ldots,\frac{\partial f}{\partial x_n})$ at $\boldsymbol{x}$.

Differentiability and Uniqueness of Support

Theorem

Let f be a convex function defined on an open convex set X in $\mathbb{R}^n$. Then f is differentiable at a point $\boldsymbol{x}_0$ of X if and only if it has unique support at $\boldsymbol{x}_0$.

- Suppose that f is differentiable at $\boldsymbol{x}_0$. Let $T : \mathbb{R}^n \to \mathbb{R}$ be a support for f at $\boldsymbol{x}_0$. Then there exists $\boldsymbol{x}' \in \mathbb{R}^n$ such that, for all $\boldsymbol{x} \in \mathbb{R}^n$, $T(\boldsymbol{x}_0 + \boldsymbol{x}) = f(\boldsymbol{x}_0) + \boldsymbol{x}' \cdot \boldsymbol{x}$. Let $\boldsymbol{y} \in \mathbb{R}^n$. Then, for all sufficiently small $\lambda > 0$,

  $$f(\boldsymbol{x}_0 + \lambda \boldsymbol{y}) - f(\boldsymbol{x}_0) \geq \lambda \boldsymbol{x}' \cdot \boldsymbol{y}.$$

  Hence $f'(\boldsymbol{x}_0; \boldsymbol{y}) \geq \boldsymbol{x}' \cdot \boldsymbol{y}$. Replacing $\boldsymbol{y}$ by $-\boldsymbol{y}$ in this last inequality and using the fact that f is differentiable at $\boldsymbol{x}_0$, we deduce that

  $$-f'(\boldsymbol{x}_0; \boldsymbol{y}) = f'(\boldsymbol{x}_0; -\boldsymbol{y}) \geq -\boldsymbol{x}' \cdot \boldsymbol{y}.$$

  Hence $f'(\boldsymbol{x}_0; \boldsymbol{y}) = \boldsymbol{x}' \cdot \boldsymbol{y}$. It follows that $\boldsymbol{x}' = (f'(\boldsymbol{x}_0; \boldsymbol{e}_1), \ldots, f'(\boldsymbol{x}_0; \boldsymbol{e}_n))$. So f has unique support T at $\boldsymbol{x}_0$.

Differentiability and Uniqueness of Support (Cont'd)

- Suppose next that f has unique support $T:\mathbb{R}^n\to\mathbb{R}$ at $\boldsymbol{x}_0$. Let m be any real number satisfying $-f'(\boldsymbol{x}_0;-\boldsymbol{e}_1)\le m\le f'(\boldsymbol{x}_0;\boldsymbol{e}_1)$. Let L be the line in $\mathbb{R}^{n+1}$ defined by the equation

$$L=\{(\boldsymbol{x}_0+t\boldsymbol{e}_1,f(\boldsymbol{x}_0)+mt):t\in\mathbb{R}\}.$$

It can be shown that $f(\boldsymbol{x}_0)+mt\le f(\boldsymbol{x}_0+t\boldsymbol{e}_1)$, for $\boldsymbol{x}_0+t\boldsymbol{e}_1\in X$.

Thus, L meets the epigraph of f at $(\boldsymbol{x}_0,f(\boldsymbol{x}_0))$ but does not meet its interior. A previous corollary shows that there is a support hyperplane to the epigraph of f at $(\boldsymbol{x}_0,f(\boldsymbol{x}_0))$ which contains L.

The uniqueness of the support to f at $\boldsymbol{x}_0$ shows that this support hyperplane must be the graph of T. Hence

$$T(\boldsymbol{x}_0+t\boldsymbol{e}_1)=f(\boldsymbol{x}_0)+mt=T(\boldsymbol{x}_0)+mt,\ \text{for } t\in\mathbb{R}.$$

Differentiability and Uniqueness of Support (Cont'd)

- Thus, m is uniquely determined by T.

  Thus, by the choice of m,

  $$-f'(\boldsymbol{x}_0; -\boldsymbol{e}_1) = f'(\boldsymbol{x}_0; \boldsymbol{e}_1).$$

  This shows that the partial derivative $\frac{\partial f}{\partial x_1}$ at $\boldsymbol{x}_0$ exists.

  Similarly, the partial derivatives $\frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}$ exist.

  By the preceding theorem, f is differentiable.

Criterion for Convexity

Theorem

Let f be a real-valued function which is defined and has continuous second-order partial derivatives on a non-empty convex set X in $\mathbb{R}^n$. Then f is convex if and only if, for every $\boldsymbol{x} \in X$,

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \left[\frac{\partial^2 f}{\partial x_i \partial x_j} \right]_{\boldsymbol{x}} z_i z_j \geq 0,$$

for all $(z_1, \ldots, z_n) \in \mathbb{R}^n$.

- Let $\boldsymbol{y} \in X$ and $\boldsymbol{z} = (z_1, \ldots, z_n) \in \mathbb{R}^n$. Let f be the open interval of $\mathbb{R}$ defined by the equation $I = \{\lambda \in \mathbb{R} : \boldsymbol{y} + \lambda \boldsymbol{z} \in X\}$. We have already seen that the function $g : I \to \mathbb{R}$ defined by the equation $g(\lambda) = f(\boldsymbol{y} + \lambda \boldsymbol{z})$ for $\lambda \in I$ is convex when f is. Conversely, suppose that each such function g is convex. We show that this implies that f is convex.

Criterion for Convexity (Cont'd)

- Let $\boldsymbol{x}, \boldsymbol{y} \in X$ and let $0 \leq \lambda \leq 1$. Write $\boldsymbol{z} = \boldsymbol{x} - \boldsymbol{y}$. Since g is convex,

$$\begin{aligned} f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) &= f(\boldsymbol{y} + \lambda(\boldsymbol{x} - \boldsymbol{y})) \\ &= g((1-\lambda)0 + \lambda 1) \\ &\leq (1-\lambda)g(0) + \lambda g(1) \\ &= \lambda f(\boldsymbol{x}) + (1-\lambda)f(\boldsymbol{y}). \end{aligned}$$

  This shows that f is convex. Thus f is convex on X if and only if each function g (as above) is convex on f. Since f has continuous second-order partial derivatives on X, each function g is differentiable twice on f. The first two derivatives of g can be calculated from the chain rule for functions of n variables:

$$g'(\lambda) = \sum_{j=1}^{n} \left[\frac{\partial f}{\partial x_j} \right]_{\boldsymbol{x}} z_j, \quad g''(\lambda) = \sum_{i=1}^{n} \sum_{j=1}^{n} \left[\frac{\partial^2 f}{\partial x_i \partial x_j} \right]_{\boldsymbol{x}} z_i z_j,$$

  where $\lambda \in I$ and the partial derivatives are evaluated at the point $\boldsymbol{x} = \boldsymbol{y} + \lambda \boldsymbol{z}$. The desired result follows by a previous corollary.

The Hessian

- Suppose that f is as in the last theorem.
- Then the $n \times n$ matrix whose (i,j)th element is $\frac{\partial^2 f}{\partial x_i \partial x_j}$ evaluated at a point $\boldsymbol{x}$ of X is called the **Hessian matrix of f at $\boldsymbol{x}$**.
- The conditions which we have imposed upon f ensure that this matrix is symmetric.
- We have thus proved that:

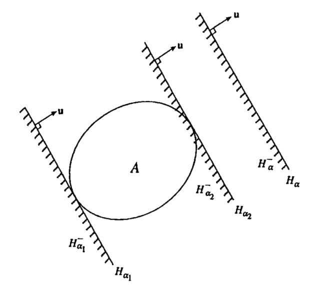
  f is convex on X if and only if its Hessian matrix is *non-negative semidefinite* at each point of X.

Subsection 6

Support Functions

Family of Parallel Hyperplanes

- Let A be a non-empty compact convex set in $\mathbb{R}^n$ and let $\boldsymbol{u}$ be a nonzero vector in $\mathbb{R}^n$.
- For each real number α, denote by H_α the hyperplane defined by the equation
$$H_\alpha = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{u} \cdot \boldsymbol{x} = \alpha\}.$$
- Denote by H_α^- the closed halfspace defined by the equation
$$H_\alpha^- = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{u} \cdot \boldsymbol{x} \leq \alpha\}.$$
- As α increases, the hyperplane H_α describes a family of parallel hyperplanes each having $\boldsymbol{u}$ as a normal vector.



Family of Parallel Hyperplanes (Cont'd)

- In general, there will be two values of α for which the hyperplane H_α supports A.
- These values are α_1 and α_2 in the figure.
- Only one of these, α_2 in the figure, will be such that $A \subseteq H_\alpha^-$.
- Clearly $A \subseteq H_\alpha^-$ if and only if $\boldsymbol{u} \cdot \boldsymbol{a} \leq \alpha$ for all $\boldsymbol{a}$ in A, i.e., if and only if
$$\sup\{\boldsymbol{u} \cdot \boldsymbol{a} : \boldsymbol{a} \in A\} \leq \alpha.$$
- If, in addition to the requirement $A \subseteq H_\alpha^-$, it is also demanded that H_α supports A, then, for some point $\boldsymbol{a}_0$ of A, $\boldsymbol{u} \cdot \boldsymbol{a}_0 = \alpha$.
- Thus H_α is a support hyperplane to A such that $A \subseteq H_\alpha^-$ if and only if
$$\alpha = \sup\{\boldsymbol{u} \cdot \boldsymbol{a} : \boldsymbol{a} \in A\}.$$

The Support Function of a Nonempty Compact Convex Set

- The **support function** h, or more precisely h_A, of a non-empty compact convex set A in $\mathbb{R}^n$ is defined by the equation

$$h(\boldsymbol{u}) = \sup\{\boldsymbol{u} \cdot \boldsymbol{a} : \boldsymbol{a} \in A\}, \text{ for each } \boldsymbol{u} \text{ in } \mathbb{R}^n.$$

- Since A is non-empty and bounded, for each $\boldsymbol{u}$ in $\mathbb{R}^n$, the subset $\{\boldsymbol{u} \cdot \boldsymbol{a} : \boldsymbol{a} \in A\}$ of $\mathbb{R}$ is non-empty and bounded.

  Hence $h(\boldsymbol{u})$ is well defined.
- The above definition of h makes sense even if A is only assumed to be non-empty and bounded.
- For our purposes, it will suffice to consider the restricted case when A is a non-empty compact convex set.

Example

- We find the support function h of the regular n-crosspolytope A defined by the equation

$$A = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : |x_1| + \cdots + |x_n| \le 1\}.$$

- Let $\boldsymbol{u} = (u_1, \ldots, u_n)$.
- Then

$$\begin{array}{rcl} h(\boldsymbol{u}) & = & \sup\{\boldsymbol{u} \cdot \boldsymbol{a} : \boldsymbol{a} \in A\} \\ & = & \sup\{u_1 a_1 + \cdots + u_n a_n : |a_1| + \cdots + |a_n| \le 1\} \\ & \le & \sup\{|u_1||a_1| + \cdots + |u_n||a_n| : |a_1| + \cdots + |a_n| \le 1\} \\ & \le & \sup\{(\max\{|u_1|, \ldots, |u_n|\})(|a_1| + \cdots + |a_n|) : \\ & & \qquad |a_1| + \cdots + |a_n| \le 1\} \\ & = & \max\{|u_1| \ldots, |u_n|\}. \end{array}$$

Example (Cont'd)

- Let $m \in \{1, \ldots, n\}$ be such that $|u_m| = \max\{|u_1|, \ldots, |u_n|\}$.
- Define a point $\boldsymbol{a} = (a_1, \ldots, a_n)$ of A by the conditions $a_i = 0$ when $i \neq m$ and a_m is 1 or -1 according as u_m is non-negative or negative.
- Then
$$\boldsymbol{u} \cdot \boldsymbol{a} = |u_m| = \max\{|u_1|, \ldots, |u_n|\}.$$
- Hence $h(\boldsymbol{u}) \geq \max\{|u_1|, \ldots, |u_n|\}$.
- We have thus shown that
$$h(\boldsymbol{u}) = \max\{|u_1|, \ldots, |u_n|\}.$$
- We note that this support function is positively homogeneous and convex.

Positive Homogeneity and Convexity

Theorem

The support function of a non-empty compact convex set in $\mathbb{R}^n$ is positively homogeneous and convex.

- Let h be the support function of a non-empty compact convex set A in $\mathbb{R}^n$. Let $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ and let $\lambda > 0$. Then

$$h(\lambda \boldsymbol{u}) = \sup\{(\lambda \boldsymbol{u}) \cdot \boldsymbol{a} : \boldsymbol{a} \in A\} = \lambda \sup\{\boldsymbol{u} \cdot \boldsymbol{a} : \boldsymbol{a} \in A\} = \lambda h(\boldsymbol{u}).$$

This shows that h is positively homogeneous.

Also

$$\begin{aligned} h(\boldsymbol{u}+\boldsymbol{v}) &= \sup\{(\boldsymbol{u}+\boldsymbol{v}) \cdot \boldsymbol{a} : \boldsymbol{a} \in A\} \\ &= \sup\{\boldsymbol{u} \cdot \boldsymbol{a} + \boldsymbol{v} \cdot \boldsymbol{a} : \boldsymbol{a} \in A\} \\ &\le \sup\{\boldsymbol{u} \cdot \boldsymbol{a} : \boldsymbol{a} \in A\} + \sup\{\boldsymbol{v} \cdot \boldsymbol{a} : \boldsymbol{a} \in A\} \\ &= h(\boldsymbol{u}) + h(\boldsymbol{v}). \end{aligned}$$

The convexity of h now follows from a previous theorem.

Exposed Face and Outward Normal

- Suppose that h is the support function of a non-empty compact convex set A in $\mathbb{R}^n$, and that $\boldsymbol{u}$ is a non-zero vector in $\mathbb{R}^n$.
- By the definition of h, $\boldsymbol{u}\cdot\boldsymbol{a} \le h(\boldsymbol{u})$ for each $\boldsymbol{a}$ in A, whence $A \subseteq \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{u}\cdot\boldsymbol{x} \le h(\boldsymbol{u})\}$.
- Consider the function $f : A \to \mathbb{R}$ defined by the rule $f(\boldsymbol{a}) = \boldsymbol{u}\cdot\boldsymbol{a}$ for each point $\boldsymbol{a}$ in A.
- Then f is continuous, and so is bounded and attains its bounds on the compact set A.
- In particular, there exists a point $\boldsymbol{a}_0$ in A such that

$$\boldsymbol{u}\cdot\boldsymbol{a}_0 = \sup\{\boldsymbol{u}\cdot\boldsymbol{a} : \boldsymbol{a} \in A\} = h(\boldsymbol{u}).$$

- So the hyperplane with equation $\boldsymbol{u}\cdot\boldsymbol{x} = h(\boldsymbol{u})$ supports A at $\boldsymbol{a}_0$.

Exposed Face and Outward Normal (Cont'd)

- The distance of this support hyperplane from the origin is $\frac{|h(\boldsymbol{u})|}{\|\boldsymbol{u}\|}$, which simplifies to $h(\boldsymbol{u})$ when $\boldsymbol{u}$ is a unit vector and the origin is a point of A.
- The earlier discussion shows that the set

$$\{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{u} \cdot \boldsymbol{x} = h(\boldsymbol{u})\} \cap A = \{\boldsymbol{x} \in A : \boldsymbol{u} \cdot \boldsymbol{x} = h(\boldsymbol{u})\}$$

  is a non-empty exposed face of A.
- It is called the **exposed face** of A with **outward normal $\boldsymbol{u}$** and is denoted by $A^{\boldsymbol{u}}$.
- Since h is positively homogeneous, for $\lambda > 0$,

$$\begin{aligned} A^{\lambda \boldsymbol{u}} &= \{\boldsymbol{x} \in A : (\lambda \boldsymbol{u}) \cdot \boldsymbol{x} = h(\lambda \boldsymbol{u})\} \\ &= \{\boldsymbol{x} \in A : \boldsymbol{u} \cdot \boldsymbol{x} = h(\boldsymbol{u})\} \\ &= A^{\boldsymbol{u}}. \end{aligned}$$

Properties of the Support Function

Theorem

Let A, B be non-empty compact convex sets in $\mathbb{R}^n$ with support functions h_A, h_B, respectively. Then the support functions h_{A+B} of $A+B$ and $h_{\lambda A}$ of λA, where $\lambda \geq 0$, are given by the equations $h_{A+B} = h_A + h_B$ and $h_{\lambda A} = \lambda h_A$.

- Let $\boldsymbol{u} \in \mathbb{R}^n$. Then

$$\begin{aligned} h_{A+B}(\boldsymbol{u}) &= \sup\{\boldsymbol{u} \cdot (\boldsymbol{a}+\boldsymbol{b}) : \boldsymbol{a} \in A, \boldsymbol{b} \in B\} \\ &= \sup\{\boldsymbol{u} \cdot \boldsymbol{a} : \boldsymbol{a} \in A\} + \sup\{\boldsymbol{u} \cdot \boldsymbol{b} : \boldsymbol{b} \in B\} \\ &= h_A(\boldsymbol{u}) + h_B(\boldsymbol{u}). \end{aligned}$$

  Hence $h_{A+B} = h_A + h_B$. Also

$$h_{\lambda A}(\boldsymbol{u}) = \sup\{\boldsymbol{u} \cdot (\lambda \boldsymbol{a}) : \boldsymbol{a} \in A\} = \lambda \sup\{\boldsymbol{u} \cdot \boldsymbol{a} : \boldsymbol{a} \in A\} = \lambda h_A(\boldsymbol{u}).$$

  Hence $h_{\lambda A} = \lambda h_A$.

Properties of the Exposed Face

Theorem

Let A, B be non-empty compact convex sets in $\mathbb{R}^n$. Then, for each non-zero vector $\boldsymbol{u}$ in $\mathbb{R}^n$ and for each $\lambda \geq 0$, $(A+B)^{\boldsymbol{u}} = A^{\boldsymbol{u}} + B^{\boldsymbol{u}}$ and $(\lambda A)^{\boldsymbol{u}} = \lambda A^{\boldsymbol{u}}$.

- We note that

$$\begin{array}{rcl} (A+B)^{\boldsymbol{u}} & = & \{\boldsymbol{a}+\boldsymbol{b} : \boldsymbol{a} \in A, \boldsymbol{b} \in B, h_{A+B}(\boldsymbol{u}) = \boldsymbol{u}\cdot(\boldsymbol{a}+\boldsymbol{b})\} \\ & = & \{\boldsymbol{a}+\boldsymbol{b} : \boldsymbol{a} \in A, \boldsymbol{b} \in B, h_A(\boldsymbol{u}) + h_B(\boldsymbol{u}) = \boldsymbol{u}\cdot\boldsymbol{a} + \boldsymbol{u}\cdot\boldsymbol{b}\} \\ & = & \{\boldsymbol{a}+\boldsymbol{b} : \boldsymbol{a} \in A, \boldsymbol{b} \in B, h_A(\boldsymbol{u}) = \boldsymbol{u}\cdot\boldsymbol{a}, h_B(\boldsymbol{u}) = \boldsymbol{u}\cdot\boldsymbol{b}\} \\ & = & \{\boldsymbol{a} \in A : h_A(\boldsymbol{u}) = \boldsymbol{u}\cdot\boldsymbol{a}\} + \{\boldsymbol{b} \in B : h_B(\boldsymbol{u}) = \boldsymbol{u}\cdot\boldsymbol{b}\} \\ & = & A^{\boldsymbol{u}} + B^{\boldsymbol{u}}. \end{array}$$

  We also have, for $\lambda \geq 0$,

$$\begin{array}{rcl} (\lambda A)^{\boldsymbol{u}} & = & \{\lambda\boldsymbol{a} : \boldsymbol{a} \in A, h_{\lambda A}(\boldsymbol{u}) = \boldsymbol{u}\cdot(\lambda\boldsymbol{a})\} \\ & = & \lambda\{\boldsymbol{a} \in A : \lambda h_A(\boldsymbol{u}) = \lambda\boldsymbol{u}\cdot\boldsymbol{a}\} \\ & = & \lambda\{\boldsymbol{a} \in A : h_A(\boldsymbol{u}) = \boldsymbol{u}\cdot\boldsymbol{a}\} \\ & = & \lambda A^{\boldsymbol{u}}. \end{array}$$

Convex Sets Determined By Support Functions

Theorem

Let h be the support function of a non-empty compact convex set A in $\mathbb{R}^n$. Then $A = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{u} \cdot \boldsymbol{x} \le h(\boldsymbol{u}) \text{ for all } \boldsymbol{u} \in \mathbb{R}^n\}$.

- We prove the theorem by showing that:
  - (i) If $\boldsymbol{a} \in A, \boldsymbol{u} \in \mathbb{R}^n$, then $\boldsymbol{u} \cdot \boldsymbol{a} \le h(\boldsymbol{u})$;
  - (ii) If $\boldsymbol{a}_0 \in \mathbb{R}^n \backslash A$, then $\boldsymbol{u} \cdot \boldsymbol{a}_0 > h(\boldsymbol{u})$ for some $\boldsymbol{u} \in \mathbb{R}^n$.

  Statement (i) follows immediately from the definition of h.

  Suppose that $\boldsymbol{a}_0 \in \mathbb{R}^n \backslash A$. Then $\{\boldsymbol{a}_0\}$ and A can be strictly separated by a hyperplane. Thus there exists $\boldsymbol{u} \in \mathbb{R}^n$ such that

$$h(\boldsymbol{u}) = \sup\{\boldsymbol{u} \cdot \boldsymbol{a} : \boldsymbol{a} \in A\} < \boldsymbol{u} \cdot \boldsymbol{a}_0.$$

  This verifies Statement (ii).

Positively Homogeneous Convex Functions as Supports

Theorem

Let $g : \mathbb{R}^n \to \mathbb{R}$ be a positively homogeneous convex function. Then the set A defined by the equation

$$A = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{u} \cdot \boldsymbol{x} \le g(\boldsymbol{u}) \text{ for all } \boldsymbol{u} \in \mathbb{R}^n\}$$

is non-empty, compact, convex, and has support function g.

- Let $\boldsymbol{u} \in \mathbb{R}^n$. Since g is convex, it has support at $\boldsymbol{u}$. So there exist $a_0 \in \mathbb{R}$, $\boldsymbol{a} \in \mathbb{R}^n$ such that $a_0 + \boldsymbol{a} \cdot \boldsymbol{u} = g(\boldsymbol{u})$ and $a_0 + \boldsymbol{a} \cdot \boldsymbol{v} \le g(\boldsymbol{v})$, for $\boldsymbol{v} \in \mathbb{R}^n$. Putting $\boldsymbol{v} = \lambda \boldsymbol{u}$, we get, for all $\lambda \ge 0$,

$$a_0 + \lambda(\boldsymbol{a} \cdot \boldsymbol{u}) \le g(\lambda \boldsymbol{u}) = \lambda g(\boldsymbol{u}) = \lambda a_0 + \lambda(\boldsymbol{a} \cdot \boldsymbol{u}).$$

Thus, $a_0 \le \lambda a_0$ for all $\lambda \ge 0$. Hence, $a_0 = 0$. Putting $a_0 = 0$ in the same relations, we find that $\boldsymbol{a} \cdot \boldsymbol{u} = g(\boldsymbol{u})$ and $\boldsymbol{a} \in A$.

Positively Homogeneous Convex Functions (Cont'd)

- We have just shown that A is non-empty.

  From its definition, A is an intersection of closed halfspaces, and so is closed and convex.

  For each $\boldsymbol{a} = (a_1, \ldots, a_n)$ in A, and $i = 1, \ldots, n$,

  $$-g(-\boldsymbol{e}_i) \leq \boldsymbol{a} \cdot \boldsymbol{e}_i = a_i \leq g(\boldsymbol{e}_i).$$

  This shows that A is bounded.

  Thus A is a non-empty compact convex set.

  Denote by h the support function of A. Let $\boldsymbol{u} \in \mathbb{R}^n$.

  By the first part of this proof, there is $\boldsymbol{a} \in A$ for which $\boldsymbol{a} \cdot \boldsymbol{u} = g(\boldsymbol{u})$. Hence, $g(\boldsymbol{u}) \leq h(\boldsymbol{u})$. For each $\boldsymbol{a}$ in A, $\boldsymbol{a} \cdot \boldsymbol{u} \leq g(\boldsymbol{u})$. So $h(\boldsymbol{u}) \leq g(\boldsymbol{u})$.

  Thus $g = h$ and g is the support function of A.

The Gauge Function

- Let A be a closed convex set in $\mathbb{R}^n$ having the origin as an interior point.
- Then it follows easily that $\lambda A \subseteq \mu A$ whenever $0 \leq \lambda \leq \mu$.
- Moreover, for each $\boldsymbol{x}$ in $\mathbb{R}^n$, there is some $\lambda \geq 0$ such that $\boldsymbol{x} \in \lambda A$.
- Thus $\mathbb{R}^n$ can be expressed as an increasing union of convex sets as follows:
$$\mathbb{R}^n = \bigcup (\lambda A : \lambda \geq 0).$$
- The **gauge function** g, or more precisely g_A, of A is the function $g : \mathbb{R}^n \to \mathbb{R}$ defined, for each $\boldsymbol{x}$ in $\mathbb{R}^n$, by the equation
$$g(\boldsymbol{x}) = \inf \{\lambda \geq 0 : \boldsymbol{x} \in \lambda A\}.$$
- In view of the earlier comments, g is well defined.

Properties of the Gauge Function

- Some immediate consequences of the definition are:
  (i) $g(\mathbf{0}) = 0$ and $g(\boldsymbol{x}) \geq 0$ for $\boldsymbol{x} \in \mathbb{R}^n$;
  (ii) $g(\boldsymbol{x}) \leq 1$ when $\boldsymbol{x} \in A$;
  (iii) If $g(\boldsymbol{x}) = 0$, then $\{\mu \boldsymbol{x} : \mu \geq 0\} \subseteq A$;
  (iv) $g(\boldsymbol{x}) = 0$ for all $\boldsymbol{x} \in \mathbb{R}^n$ if and only if $A = \mathbb{R}^n$.
- Suppose now that $g(\boldsymbol{x}) > 0$. Then, for each $\varepsilon > 0$, $\boldsymbol{x} \in (g(\boldsymbol{x}) + \varepsilon)A$. Hence $\frac{\boldsymbol{x}}{g(\boldsymbol{x})+\varepsilon} \in A$. Letting $\varepsilon \to 0^+$ and using our assumption that A is closed, we deduce that $\frac{\boldsymbol{x}}{g(\boldsymbol{x})} \in A$. Hence $\boldsymbol{x} \in g(\boldsymbol{x})A$. In particular, if $0 < g(\boldsymbol{x}) \leq 1$, then $\boldsymbol{x} \in g(\boldsymbol{x})A \subseteq A$. We have thus established:
  (v) $A = \{\boldsymbol{x} \in \mathbb{R}^n : g(\boldsymbol{x}) \leq 1\}$.

Positive Homogeneity and Convexity of the Gauge Function

Theorem

The gauge function of a closed convex set having the origin as an interior point is positively homogeneous and convex.

- Let g be the gauge function of a closed convex set A in $\mathbb{R}^n$ which contains the origin in its interior.

  Let $\boldsymbol{x} \in \mathbb{R}^n$ and let $\lambda > 0$. Then $\lambda \boldsymbol{x} \in \mu A$ if and only if $\boldsymbol{x} \in \frac{\mu}{\lambda} A$. It follows easily from the definition of g that

$$\frac{1}{\lambda} g(\lambda \boldsymbol{x}) = \frac{1}{\lambda} \inf \{\mu \geq 0 : \lambda \boldsymbol{x} \in \mu A\} = \inf \left\{ \frac{\mu}{\lambda} : \boldsymbol{x} \in \frac{\mu}{\lambda} A \right\} = g(\boldsymbol{x}).$$

  Trivially, $g(0\boldsymbol{x}) = 0 g(\boldsymbol{x})$. Thus g is positively homogeneous.

  Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and let $\lambda, \mu \geq 0$ with $\lambda + \mu = 1$. Then, for each $\varepsilon > 0$, $\boldsymbol{x} \in (g(\boldsymbol{x}) + \varepsilon) A$, $\boldsymbol{y} \in (g(\boldsymbol{y}) + \varepsilon) A$. So $\lambda \boldsymbol{x} + \mu \boldsymbol{y} \in (\lambda g(\boldsymbol{x}) + \mu g(\boldsymbol{y}) + \varepsilon) A$. Since $\varepsilon > 0$ is arbitrary, $g(\lambda \boldsymbol{x} + \mu \boldsymbol{y}) \leq \lambda g(\boldsymbol{x}) + \mu g(\boldsymbol{y})$. This shows that g is convex.

Example

- We find the gauge function g of the n-cube A defined by the equation

$$A = \{(x_1, \ldots, x_n) : |x_1|, \ldots, |x_n| \le 1\}.$$

Let $\boldsymbol{u} = (u_1, \ldots, u_n)$. Then, for $\lambda \ge 0$,

$$\lambda A = \{(x_1, \ldots, x_n) : |x_1|, \ldots, |x_n| \le \lambda\}.$$

So $\boldsymbol{u} \in \lambda A$ if and only if $\max\{|u_1|, \ldots, |u_n|\} \le \lambda$. Thus,

$$g(\boldsymbol{u}) = \max\{|u_1|, \ldots, |u_n|\}.$$

Nonnegative Positively Homogeneous Convex Functions

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a non-negative positively homogeneous convex function. Then the set A defined by the equation

$$A = \{\boldsymbol{x} \in \mathbb{R}^n : f(\boldsymbol{x}) \le 1\}$$

is closed, convex, contains the origin in its interior and has gauge function f.

- The function f is continuous by a previous theorem. Thus A is closed and contains the open set $\{\boldsymbol{x} \in \mathbb{R}^n : f(\boldsymbol{x}) < 1\}$, which contains the origin. The set A is convex, being the level set of a convex function. Hence A is a closed convex set containing the origin in its interior.

Nonnegative Positively Homogeneous Convex Functions II

- Denote by g the gauge function of A.

  Then, as proved earlier, $A = \{\boldsymbol{x} \in \mathbb{R}^n : g(\boldsymbol{x}) \le 1\}$.

  Suppose that $\boldsymbol{u} \in \mathbb{R}^n$ satisfies $g(\boldsymbol{u}) > 0$.

  Since g is positively homogeneous, $g\left(\frac{\boldsymbol{u}}{g(\boldsymbol{u})}\right) = 1$. Hence $\frac{\boldsymbol{u}}{g(\boldsymbol{u})} \in A$.

  Since f is positively homogeneous and $\frac{\boldsymbol{u}}{g(\boldsymbol{u})} \in A$, $f\left(\frac{\boldsymbol{u}}{g(\boldsymbol{u})}\right) = \frac{f(\boldsymbol{u})}{g(\boldsymbol{u})} \le 1$.

  This shows that $f(\boldsymbol{u}) \le g(\boldsymbol{u})$.

  If $g(\boldsymbol{u}) = 0$, then, for all $\lambda > 0$, $\lambda \boldsymbol{u} \in A$.

  So $0 \le f(\lambda \boldsymbol{u}) = \lambda f(\boldsymbol{u}) \le 1$. It follows that $f(\boldsymbol{u}) = 0$.

  Thus $f(\boldsymbol{u}) \le g(\boldsymbol{u})$ for all $\boldsymbol{u} \in \mathbb{R}^n$.

  By a similar argument, $g(\boldsymbol{u}) \le f(\boldsymbol{u})$ for all $\boldsymbol{u} \in \mathbb{R}^n$.

  Hence $f = g$ and f is the gauge function of A.

Example

- We have already seen that the support function of the regular n-crosspolytope

$$\{(x_1,\ldots,x_n) : |x_1| + \cdots + |x_n| \le 1\}$$

and the gauge function of its dual, the n-cube

$$\{(x_1,\ldots,x_n) : |x_1|,\ldots,|x_n| \le 1\}$$

are the same, namely the function $f : \mathbb{R}^n \to \mathbb{R}$ defined by the equation

$$f(\boldsymbol{u}) = \max\{|u_1|,\ldots,|u_n|\}, \text{ for } \boldsymbol{u} = (u_1,\ldots,u_n) \in \mathbb{R}^n.$$

Duality: Support and Gauge Functions

Theorem

Suppose that g, h are the gauge and support functions, respectively, of a compact convex set A in $\mathbb{R}^n$ which has the origin as an interior point. Then the gauge and support functions of the dual A^* of A are h, g, respectively.

- If $\boldsymbol{u} \in A^*$, then $\boldsymbol{u} \cdot \boldsymbol{a} \leq 1$ for all $\boldsymbol{a}$ in A, whence $h(\boldsymbol{u}) \leq 1$.

  Conversely, if $h(\boldsymbol{u}) \leq 1$, then $\boldsymbol{u} \cdot \boldsymbol{a} \leq 1$ for all $\boldsymbol{a}$ in A, and so $\boldsymbol{u} \in A^*$. Thus,

  $$A^* = \{\boldsymbol{x} \in \mathbb{R}^n : h(\boldsymbol{x}) \leq 1\}.$$

  Since A contains the origin, h is non-negative.

  Thus h is a non-negative, positively homogeneous convex function.

  Hence, by the preceding theorem, h is the gauge function of A^*.

  By what we have just proved, the support function of A^* is the gauge function of $A^{**} = A$, viz. g.

Subsection 7

The Convex Programming Problem

The Convex Programming Problem

- Throughout this section $f, g_1, \ldots, g_m$ will denote convex functions defined on $\mathbb{R}^n$.
- The **convex programming problem** is to minimize $f(\boldsymbol{x})$ subject to the constraints $\boldsymbol{x} \geq \boldsymbol{0}$, $g_1(\boldsymbol{x}) \leq 0, \ldots, g_m(\boldsymbol{x}) \leq 0$.
- The **feasible set** for the problem is the convex set X defined by the equation
$$X = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x} \geq \boldsymbol{0}, g_1(\boldsymbol{x}) \leq 0, \ldots, g_m(\boldsymbol{x}) \leq 0\}.$$
- Thus the convex programming problem is to find $\boldsymbol{x}_0 \in X$ such that $f(\boldsymbol{x}_0) \leq f(\boldsymbol{x})$ for all $\boldsymbol{x} \in X$.

Existence of Coefficients

Theorem

Let $f_1, \ldots, f_k$ be convex functions defined on a nonempty convex set Y in $\mathbb{R}^n$. Suppose that there exists no $\boldsymbol{y}$ in Y such that $f_1(\boldsymbol{y}) < 0$, ..., $f_k(\boldsymbol{y}) < 0$. Then there exist $a_1, \ldots, a_k \geq 0$, not all zero, such that

$$a_1 f_1(\boldsymbol{y}) + \cdots + a_k f_k(\boldsymbol{y}) \geq 0, \text{ for all } \boldsymbol{y} \in Y.$$

- Define a set C in $\mathbb{R}^k$ by the equation

$$C = \{(z_1, \ldots, z_k) : \text{there is } \boldsymbol{y} \in Y \text{ such that } f_i(\boldsymbol{y}) < z_i \text{ for } i = 1, \ldots, k\}.$$

Let $\boldsymbol{u} = (u_1, \ldots, u_k)$, $\boldsymbol{v} = (v_1, \ldots, v_k) \in C$. Let $\lambda, \mu \geq 0$ with $\lambda + \mu = 1$. Then there exist $\boldsymbol{a}, \boldsymbol{b} \in Y$ such that, for $i = 1, \ldots, k$, $f_i(\boldsymbol{a}) < u_i$ and $f_i(\boldsymbol{b}) < v_i$. The convexity of $f_1, \ldots, f_k$ shows that, for $i = 1, \ldots, k$,

$$f_i(\lambda \boldsymbol{a} + \mu \boldsymbol{b}) \leq \lambda f_i(\boldsymbol{a}) + \mu f_i(\boldsymbol{b}) < \lambda u_i + \mu v_i.$$

Hence, since $\lambda \boldsymbol{a} + \mu \boldsymbol{b} \in Y$, $\lambda \boldsymbol{u} + \mu \boldsymbol{v} \in C$. Thus C is convex.

Existence of Coefficients (Cont'd)

- By hypothesis, C does not contain the origin of $\mathbb{R}^k$.

  So the origin and C can be separated by a hyperplane.

  Thus, there exist scalars $a_1, \ldots, a_k$, not all zero, such that, for all $\boldsymbol{y} \in Y$ and all $\lambda_1, \ldots, \lambda_k > 0$,

  $$a_1(f_1(\boldsymbol{y}) + \lambda_1) + \cdots + a_k(f_k(\boldsymbol{y}) + \lambda_k) \geq 0.$$

  Letting $\lambda_1 \to \infty$, whilst keeping $\lambda_2, \ldots, \lambda_k$ fixed in, we deduce that $a_1 \geq 0$. Similarly, $a_2 \geq 0$, ..., $a_k \geq 0$.

  Letting $\lambda_1 \to 0^+, \ldots, \lambda_k \to 0^+$, we deduce that, for all $\boldsymbol{y}$ in Y,

  $$a_1 f_1(\boldsymbol{y}) + \cdots + a_k f_k(\boldsymbol{y}) \geq 0.$$

Lagrangian Function and Saddle-Point Problem

- The **Lagrangian function** associated with the convex programming problem is the function F of the $m+n$ variables $x_1,\ldots,x_n$, $y_1,\ldots,y_m$ defined by the equation

$$F(\boldsymbol{x},\boldsymbol{y}) = f(\boldsymbol{x}) + y_1 g_1(\boldsymbol{x}) + \cdots + y_m g_m(\boldsymbol{x}),$$

  where $\boldsymbol{x} = (x_1,\ldots,x_n)$, $\boldsymbol{y} = (y_1,\ldots,y_m)$.
- The **saddle-point problem** is to determine a **saddle point** of F, that is, a point $(\boldsymbol{x}_0,\boldsymbol{y}_0)$ of $\mathbb{R}^{m+n}$ such that $\boldsymbol{x}_0 \geq \mathbf{0}$, $\boldsymbol{y}_0 \geq \mathbf{0}$ and

$$F(\boldsymbol{x}_0,\boldsymbol{y}) \leq F(\boldsymbol{x}_0,\boldsymbol{y}_0) \leq F(\boldsymbol{x},\boldsymbol{y}_0),$$

  for all $\boldsymbol{x} \geq \mathbf{0}$, $\boldsymbol{y} \geq \mathbf{0}$.

Saddle-Points and Convex Programming Problem

Theorem

Let $(\boldsymbol{x}_0, \boldsymbol{y}_0)$ be a saddle point of the Lagrangian function F. Then $\boldsymbol{x}_0$ is a solution to the convex programming problem and $F(\boldsymbol{x}_0, \boldsymbol{y}_0) = f(\boldsymbol{x}_0)$.

- Let $\boldsymbol{x}_0 = (x_1^0, \ldots, x_n^0) \geq \mathbf{0}$ and $\boldsymbol{y}_0 = (y_1^0, \ldots, y_m^0) \geq \mathbf{0}$. For all $\boldsymbol{y} = (y_1, \ldots, y_m) \geq \mathbf{0}$, $F(\boldsymbol{x}_0, \boldsymbol{y}_0) \geq F(\boldsymbol{x}_0, \boldsymbol{y})$. So

$$y_1^0 g_1(\boldsymbol{x}_0) + \cdots + y_m^0 g_m(\boldsymbol{x}_0) \geq y_1 g_1(\boldsymbol{x}_0) + \cdots + y_m g_m(\boldsymbol{x}_0).$$

  By fixing $y_2, \ldots, y_m$ and letting $y_1 \to \infty$, we deduce that $g_1(\boldsymbol{x}_0) \leq 0$.

  Similarly, $g_2(\boldsymbol{x}_0) \leq 0$, ..., $g_m(\boldsymbol{x}_0) \leq 0$.

  Thus $\boldsymbol{x}_0$ is a point of the feasible set X of the convex programming problem.

Saddle-Points and Convex Programming Problem (Cont'd)

- Putting $\boldsymbol{y} = \mathbf{0}$ in the saddle-point inequality $F(\boldsymbol{x}_0, \boldsymbol{y}_0) \geq F(\boldsymbol{x}_0, \boldsymbol{y})$ and using the fact that $\boldsymbol{x}_0 \in X$, we deduce that

$$f(\boldsymbol{x}_0) \leq f(\boldsymbol{x}_0) + y_1^0 g_1(\boldsymbol{x}_0) + \cdots + y_m^0 g_m(\boldsymbol{x}_0).$$

Therefore, since $\boldsymbol{y}_0 \geq \mathbf{0}$ and $g_i(\boldsymbol{x}_0) \leq 0$,

$$0 \leq y_1^0 g_1(\boldsymbol{x}_0) + \cdots + y_m^0 g_m(\boldsymbol{x}_0) \leq 0.$$

Hence

$$y_1^0 g_1(\boldsymbol{x}_0) + \cdots + y_m^0 g_m(\boldsymbol{x}_0) = 0 \text{ and } F(\boldsymbol{x}_0, \boldsymbol{y}_0) = f(\boldsymbol{x}_0).$$

Since $F(\boldsymbol{x}_0, \boldsymbol{y}_0) \leq F(\boldsymbol{x}, \boldsymbol{y}_0)$ for all $\boldsymbol{x} \geq \mathbf{0}$, we deduce that, for all $\boldsymbol{x} \in X$,

$$f(\boldsymbol{x}_0) \leq f(\boldsymbol{x}) + y_1^0 g_1(\boldsymbol{x}) + \cdots + y_m^0 g_m(\boldsymbol{x}) \leq f(\boldsymbol{x}).$$

This shows that $\boldsymbol{x}_0$ is a solution to the convex programming problem.

A Partial Converse

- It is not true that, given any solution $\boldsymbol{x}_0$ of the convex programming problem, there is always a $\boldsymbol{y}_0$ such that $(\boldsymbol{x}_0, \boldsymbol{y}_0)$ is a saddle point of the Lagrangian function F.

Theorem

Suppose that $\boldsymbol{x}_0$ is a solution of the convex programming problem. Suppose also that there exists $\boldsymbol{x}^* \geq 0$ such that $g_1(\boldsymbol{x}^*) < 0$, ..., $g_m(\boldsymbol{x}^*) < 0$. Then there exists $\boldsymbol{y}_0 \in \mathbb{R}^m$ for which $(\boldsymbol{x}_0, \boldsymbol{y}_0)$ is a saddle point of the Lagrangian function F.

- Suppose that $\boldsymbol{x}$ belongs to the nonnegative orthant Y of $\mathbb{R}^n$. Then not all of the following inequalities can hold: $g_1(\boldsymbol{x}) < 0$, ..., $g_m(\boldsymbol{x}) < 0$, $f(\boldsymbol{x}) - f(\boldsymbol{x}_0) < 0$. Thus, by a previous theorem, there exist $a_1, \ldots, a_m, a_0 \geq 0$, not all zero, such that

$$a_1 g_1(\boldsymbol{x}) + \cdots + a_m g_m(\boldsymbol{x}) + a_0(f(\boldsymbol{x}) - f(\boldsymbol{x}_0)) \geq 0$$

whenever $\boldsymbol{x} \in Y$, i.e., $\boldsymbol{x} \geq \boldsymbol{0}$.

A Partial Converse (Cont'd)

- If $a_0 = 0$, then
$$0 > a_1 g_1(\boldsymbol{x}^*) + \cdots + a_m g_m(\boldsymbol{x}^*) \geq 0,$$
which is impossible. Thus $a_0 > 0$. For $i = 1, \ldots, m$, let $y_i^0 = \frac{a_i}{a_0}$ and let $\boldsymbol{y}_0 = (y_1^0, \ldots, y_m^0) \geq \boldsymbol{0}$. Then, for any $\boldsymbol{x} \geq \boldsymbol{0}$, we deduce from the displayed inequality that
$$f(\boldsymbol{x}_0) \leq f(\boldsymbol{x}) + y_1^0 g_1(\boldsymbol{x}) + \cdots + y_m^0 g_m(\boldsymbol{x}) = F(\boldsymbol{x}, \boldsymbol{y}_0).$$
Hence
$$f(\boldsymbol{x}_0) \leq f(\boldsymbol{x}_0) + y_1^0 g_1(\boldsymbol{x}_0) + \cdots + y_m^0 g_m(\boldsymbol{x}_0) \leq f(\boldsymbol{x}_0).$$
So $y_1^0 g_1(\boldsymbol{x}_0) + \cdots + y_m^0 g_m(\boldsymbol{x}_0) = 0$. Thus, for all $\boldsymbol{x} \geq \boldsymbol{0}$, $F(\boldsymbol{x}_0, \boldsymbol{y}_0) = f(\boldsymbol{x}_0) \leq F(\boldsymbol{x}, \boldsymbol{y}_0)$. For $\boldsymbol{y} = (y_1, \ldots, y_m) \geq \boldsymbol{0}$,
$$F(\boldsymbol{x}_0, \boldsymbol{y}_0) = f(\boldsymbol{x}_0) \geq f(\boldsymbol{x}_0) + y_1 g_1(\boldsymbol{x}_0) + \cdots + y_m g_m(\boldsymbol{x}_0) = F(\boldsymbol{x}_0, \boldsymbol{y}).$$
This shows that $(\boldsymbol{x}_0, \boldsymbol{y}_0)$ is a saddle point of F.

Kuhn-Tucker Conditions

Theorem (Kuhn-Tucker Conditions)

Suppose that the convex functions $f, g_1, \ldots, g_m : \mathbb{R}^n \to \mathbb{R}$ are differentiable. Then $(\boldsymbol{x}_0, \boldsymbol{y}_0)$, where $\boldsymbol{x}_0 = (x_1^0, \ldots, x_n^0)$ and $\boldsymbol{y}_0 = (y_1^0, \ldots, y_m^0)$, is a saddle point of the Lagrangian function F if and only if

$$\begin{array}{l} \boldsymbol{x}_0 \geq \boldsymbol{0}, \\ \frac{\partial F}{\partial x_j}(\boldsymbol{x}_0, \boldsymbol{y}_0) = \frac{\partial f}{\partial x_j}(\boldsymbol{x}_0) + \sum_{i=1}^m y_i^0 \frac{\partial g_i}{\partial x_j}(\boldsymbol{x}_0) \geq 0, \\ \frac{\partial F}{\partial x_j}(\boldsymbol{x}_0, \boldsymbol{y}_0) = 0, \text{ if } x_j^0 > 0, \end{array}$$

and

$$\begin{array}{l} \boldsymbol{y}_0 \geq \boldsymbol{0}, \\ \frac{\partial F}{\partial y_j}(\boldsymbol{x}_0, \boldsymbol{y}_0) = g_j(\boldsymbol{x}_0) \leq 0, \\ \frac{\partial F}{\partial y_j}(\boldsymbol{x}_0, \boldsymbol{y}_0) = 0, \text{ if } y_j^0 > 0. \end{array}$$

Proof

- Suppose first that $(\boldsymbol{x}_0, \boldsymbol{y}_0)$ is a saddle point of F.

  Then certainly the first conditions of each triple are satisfied.

  For each $j = 1, \ldots, n$,

$$F(\boldsymbol{x}_0 + \lambda \boldsymbol{e}_j, \boldsymbol{y}_0) \geq f(\boldsymbol{x}_0, \boldsymbol{y}_0), \text{ if } \lambda \geq -x_j^0.$$

  It now follows, by elementary calculus, that

$$\frac{\partial F}{\partial x_j}(\boldsymbol{x}_0, \boldsymbol{y}_0) \geq 0 \text{ and } \frac{\partial F}{\partial x_j}(\boldsymbol{x}_0, \boldsymbol{y}_0) = 0, \text{ if } x_j^0 > 0.$$

  Thus, the last two conditions of the first triple are satisfied.

  By a previous theorem, the remaining conditions are also satisfied.

Proof (Converse)

- Suppose next that the six Kuhn-Tucker conditions are satisfied.

  The function $F(\boldsymbol{x},\boldsymbol{y}_0)$ of $\boldsymbol{x}$, for fixed $\boldsymbol{y}_0$, is convex and differentiable, because $f,g_1,\ldots,g_m$ are, and $\boldsymbol{y}_0\geq\mathbf{0}$. Thus $F(\boldsymbol{x},\boldsymbol{y}_0)$ has unique support at $\boldsymbol{x}_0$. Hence, for all $\boldsymbol{x}=(x_1,\ldots,x_n)\geq\mathbf{0}$,

$$\begin{aligned}F(\boldsymbol{x},\boldsymbol{y}_0) &\geq F(\boldsymbol{x}_0,\boldsymbol{y}_0)+(x_1-x_1^0)\frac{\partial F}{\partial x_1}(\boldsymbol{x}_0,\boldsymbol{y}_0)+\cdots+(x_n-x_n^0)\frac{\partial F}{\partial x_n}(\boldsymbol{x}_0,\boldsymbol{y}_0)\\ &= F(\boldsymbol{x}_0,\boldsymbol{y}_0)+x_1\frac{\partial F}{\partial x_1}(\boldsymbol{x}_0,\boldsymbol{y}_0)+\cdots+x_n\frac{\partial F}{\partial x_n}(\boldsymbol{x}_0,\boldsymbol{y}_0)\\ &\geq F(\boldsymbol{x}_0,\boldsymbol{y}_0).\end{aligned}$$

  The first set of conditions was used here.

  Finally for $\boldsymbol{y}=(y_1,\ldots,y_m)\geq\mathbf{0}$, we have

$$\begin{aligned}F(\boldsymbol{x}_0,\boldsymbol{y}) &= F(\boldsymbol{x}_0,\boldsymbol{y}_0)+(y_1-y_1^0)g_1(\boldsymbol{x}_0)+\cdots+(y_m-y_m^0)g_m(\boldsymbol{x}_0)\\ &= F(\boldsymbol{x}_0,\boldsymbol{y}_0)+y_1g_1(\boldsymbol{x}_0)+\cdots+y_mg_m(\boldsymbol{x}_0)\\ &\leq F(\boldsymbol{x}_0,\boldsymbol{y}_0).\end{aligned}$$

  Here we have used the second set of conditions.

  We have thus shown that $(\boldsymbol{x}_0,\boldsymbol{y}_0)$ is a saddle point of F.

Example

- Solve the convex programming problem:

$$\begin{array}{ll} \text{minimize} & -6x_1 + 2x_1^2 - 2x_1x_2 + 2x_2^2 \\ \text{subject to} & x_1 + x_2 \le 2,\ x_1 \ge 0,\ x_2 \ge 0. \end{array}$$

Write $f(x_1, x_2) = -6x_1 + 2x_1^2 - 2x_1x_2 + 2x_2^2$ and $g(x_1, x_2) = x_1 + x_2 - 2$. The Lagrangian function F is defined by the equation

$$F(\mathbf{x}, \mathbf{y}) = -6x_1 + 2x_1^2 - 2x_1x_2 + 2x_2^2 + y_1(x_1 + x_2 - 2).$$

The Kuhn-Tucker conditions give the following equations and inequalities:

$$\begin{array}{rr} x_1(-6 + 4x_1 - 2x_2 + y_1) = 0, & -6 + 4x_1 - 2x_2 + y_1 \ge 0, \\ x_2(-2x_1 + 4x_2 + y_1) = 0, & -2x_1 + 4x_2 + y_1 \ge 0, \\ y_1(x_1 + x_2 - 2) = 0, & x_1 + x_2 - 2 \le 0, \\ & x_1 \ge 0,\ x_2 \ge 0,\ y_1 \ge 0. \end{array}$$

Example (Cont'd)

- The three equations have the following six solutions:

	x_1	x_2	y_1
(i)	0	0	0
(ii)	0	2	-8
(iii)	$\frac{3}{2}$	0	0
(iv)	2	0	-2
(v)	2	1	0
(vi)	$\frac{3}{2}$	$\frac{1}{2}$	$1.$

  Of these solutions only (vi) satisfies all the remaining inequalities.
  Hence f has minimal value $-\frac{11}{2}$ at $(\frac{3}{2}, \frac{1}{2})$.

Subsection 8

Matrix Inequalities

A Problem Involving Quadratic Forms

- Associated with each real symmetric square matrix $\boldsymbol{A}$ of order n, there is a quadratic function $q : \mathbb{R}^n \to \mathbb{R}$ defined for each $\boldsymbol{x}$ in $\mathbb{R}^n$ by the equation
$$q(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x} = (\boldsymbol{A}\boldsymbol{x}) \cdot \boldsymbol{x}.$$
- Let $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ be an orthonormal sequence in $\mathbb{R}^n$ consisting of eigenvectors of $\boldsymbol{A}$ corresponding to the eigenvalues $\lambda_1, \ldots, \lambda_n$ of $\boldsymbol{A}$.
- Then, for $i = 1, \ldots, n$, $\boldsymbol{u}_i^T \boldsymbol{A}\boldsymbol{u}_i = (\boldsymbol{A}\boldsymbol{u}_i) \cdot \boldsymbol{u}_i = (\lambda_i \boldsymbol{u}_i) \cdot \boldsymbol{u}_i = \lambda_i$.
- Hence $(q(\boldsymbol{u}_1), \ldots, q(\boldsymbol{u}_n)) = (\lambda_1, \ldots, \lambda_n)$.
- We consider the following problem:

  If $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ is any orthonormal sequence in $\mathbb{R}^n$, how are the points $\boldsymbol{u} = (q(\boldsymbol{u}_1), \ldots, q(\boldsymbol{u}_n))$ and $\boldsymbol{v} = (q(\boldsymbol{v}_1), \ldots, q(\boldsymbol{v}_n))$ related to one another?

Answering the Problem

- Express each $\boldsymbol{v}_i$, for $i = 1, \ldots, n$, as a linear combination of $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$, thus:

$$\boldsymbol{v}_i = (\boldsymbol{v}_i \cdot \boldsymbol{u}_1)\boldsymbol{u}_1 + \cdots + (\boldsymbol{v}_i \cdot \boldsymbol{u}_n)\boldsymbol{u}_n.$$

- Hence

$$\begin{aligned} q(\boldsymbol{v}_i) &= ((\boldsymbol{v}_i \cdot \boldsymbol{u}_1)\boldsymbol{A}\boldsymbol{u}_1 + \cdots + (\boldsymbol{v}_i \cdot \boldsymbol{u}_n)\boldsymbol{A}\boldsymbol{u}_n) \cdot \\ &\qquad ((\boldsymbol{v}_i \cdot \boldsymbol{u}_1)\boldsymbol{u}_1 + \cdots + (\boldsymbol{v}_i \cdot \boldsymbol{u}_n)\boldsymbol{u}_n) \\ &= (\lambda_1(\boldsymbol{v}_i \cdot \boldsymbol{u}_1)\boldsymbol{u}_1 + \cdots + \lambda_n(\boldsymbol{v}_i \cdot \boldsymbol{u}_n)\boldsymbol{u}_n) \cdot \\ &\qquad ((\boldsymbol{v}_i \cdot \boldsymbol{u}_1)\boldsymbol{u}_1 + \cdots + (\boldsymbol{v}_i \cdot \boldsymbol{u}_n)\boldsymbol{u}_n) \\ &= (\boldsymbol{v}_i \cdot \boldsymbol{u}_1)^2\lambda_1 + \cdots + (\boldsymbol{v}_i \cdot \boldsymbol{u}_n)^2\lambda_n \\ &= (\boldsymbol{v}_i \cdot \boldsymbol{u}_1)^2 q(\boldsymbol{u}_1) + \cdots + (\boldsymbol{v}_i \cdot \boldsymbol{u}_n)^2 q(\boldsymbol{u}_n). \end{aligned}$$

- Thus $\boldsymbol{v} = \boldsymbol{S}\boldsymbol{u}$, where $\boldsymbol{S}$ is the square matrix of order n whose (i,j)th element is $(\boldsymbol{v}_i \cdot \boldsymbol{u}_j)^2$.

Double Stochasticity of the Matrix $\boldsymbol{S}$

- The matric $\boldsymbol{S}$ is a square matrix all of whose elements are non-negative real numbers.
- Squaring both sides of equation $\boldsymbol{v}_i = (\boldsymbol{v}_i \cdot \boldsymbol{u}_1)\boldsymbol{u}_1 + \cdots + (\boldsymbol{v}_i \cdot \boldsymbol{u}_n)\boldsymbol{u}_n$, and using the orthonormality of the sequences $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ and $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$, we deduce that, for $i = 1, \ldots, n$,

$$(\boldsymbol{v}_i \cdot \boldsymbol{u}_1)^2 + \cdots + (\boldsymbol{v}_i \cdot \boldsymbol{u}_n)^2 = \|\boldsymbol{v}_i\|^2 = 1.$$

- Similarly, for $j = 1, \ldots, n$,

$$(\boldsymbol{u}_j \cdot \boldsymbol{v}_1)^2 + \cdots + (\boldsymbol{u}_j \cdot \boldsymbol{v}_n)^2 = \|\boldsymbol{u}_j\|^2 = 1.$$

- Thus $\boldsymbol{S}$ is a square matrix of order n whose elements are non-negative real numbers, and the sum of the elements in each of its rows and columns is equal to 1.
- Such a matrix is called a **doubly stochastic matrix**.
- The set of all doubly stochastic $n \times n$ matrices will be denoted by Ω_n.

Permutation Matrices

- The simplest example of a doubly stochastic matrix is a **permutation matrix**, which is a square matrix with precisely one 1 in each row and column, all of its other elements being zero.
- Equivalently, a permutation matrix is one that can be obtained by permuting the rows of an identity matrix.
- Clearly every convex combination (in the obvious sense) of permutation matrices is a doubly stochastic matrix.
- The converse of this result, namely that every doubly stochastic matrix is a convex combination of permutation matrices, is also true and it is known as Birkhoff's Theorem.
- This theorem, which will be proven here, is perhaps the most fundamental result in the whole study of doubly stochastic matrices.

$n \times n$ Matrices and $\mathbb{R}^{n^2}$

- In a natural way we may regard each real $n \times n$ matrix $\boldsymbol{A} = [a_{ij}]$ as a point $\boldsymbol{a} = (a_{ij})$ of $\mathbb{R}^{n^2}$, the n^2 elements of $\boldsymbol{A}$ corresponding in some prescribed way to the n^2 coordinates of $\boldsymbol{a}$.
- To be definite, we set up the correspondence

$$\boldsymbol{A} = [a_{ij}] \leftrightarrow (a_{11}, \ldots, a_{1n}, a_{21}, \ldots, a_{2n}, \ldots, a_{n1}, \ldots, a_{nn}) = \boldsymbol{a}.$$

- This correspondence is a bijection between the set of all real $n \times n$ matrices and the set of points in $\mathbb{R}^{n^2}$.
- It preserves linear combinations, and so we can usefully identify the matrix $\boldsymbol{A}$ with the point $\boldsymbol{a}$.
- Under this identification, we may think of the set Ω_n of doubly stochastic $n \times n$ matrices as a set in $\mathbb{R}^{n^2}$ and refer to some of its members as being permutation matrices.

Lemma on Non-Singular 0-1-Block Matrices

Lemma

Let $\boldsymbol{B}$ be a non-singular square matrix of order n that can be partitioned in the form $\begin{bmatrix} \boldsymbol{P} \\ \boldsymbol{Q} \end{bmatrix}$, where $\boldsymbol{P}$ and $\boldsymbol{Q}$ are matrices of 0's and 1's, such that no column of either $\boldsymbol{P}$ or $\boldsymbol{Q}$ contains more than one 1. Then $\det \boldsymbol{B} = \pm 1$.

- We argue by induction on n. The case $n = 2$ is trivial.

  Suppose that $n \geq 3$ and that the assertion is true for square matrices of order $n-1$. Let $\boldsymbol{B}$ be as in the statement of the lemma.

  At least one column of $\boldsymbol{B}$ contains precisely one 1.

  Otherwise the rows of $\boldsymbol{P}$ could be added to the negatives of the rows of $\boldsymbol{Q}$ to produce a zero row, contradicting the non-singularity of $\boldsymbol{B}$.

  Expanding $\det \boldsymbol{B}$ by a column with precisely one 1, $\det \boldsymbol{B} = \pm \det \boldsymbol{C}$.

  But $\boldsymbol{C}$ is a square matrix of order $n-1$ of the form in the lemma.

  Hence, $\det \boldsymbol{B} = \pm 1$, since $\det \boldsymbol{C} = \pm 1$ by the induction hypothesis.

Birkhoff's Theorem

Theorem

The set Ω_n is a polytope in $\mathbb{R}^{n^2}$ whose extreme points are the permutation matrices in Ω_n. Every doubly stochastic matrix is a convex combination of permutation matrices.

- The set Ω_n is polyhedral, since it consists of those points (x_{ij}) in $\mathbb{R}^{n^2}$ satisfying the relations:

$$\begin{array}{rcll} x_{ij} & \geq & 0, & i,j=1,\ldots,n; \\ \sum_{j=1}^n x_{ij} & = & 1, & i=1,\ldots,n; \\ \sum_{i=1}^n x_{ij} & = & 1, & j=1,\ldots,n-1. \end{array}$$

  Note that the equality $x_{1n}+\cdots+x_{nn}=1$ follows from the $2n-1$ equations in the last two lines.
  The relations of the first two lines show that, if $(x_{ij})\in\Omega_n$, then $0\leq x_{ij}\leq 1$. Hence Ω_n is a bounded polyhedral set, i.e., a polytope.

Birkhoff's Theorem (Cont'd)

- That each permutation matrix in Ω_n is one of its extreme points follows easily from the definitions of extreme point and permutation matrix. The non-trivial part of the proof is to show that each extreme point of Ω_n is a permutation matrix.
  Let (a_{ij}) be an extreme point of Ω_n. Then, by a previous theorem, (a_{ij}) is a nonnegative basic solution for the system of the $2n-1$ equations in the last two lines above, i.e., of $\boldsymbol{Ax} = \boldsymbol{b}$, where

$$\boldsymbol{A} = \begin{bmatrix} 111\ldots11 & 000\ldots00 & \cdots & 000\ldots00 \\ 000\ldots00 & 111\ldots11 & \cdots & 000\ldots00 \\ \vdots & \vdots & & \vdots \\ 000\ldots00 & 000\ldots00 & \cdots & 111\ldots11 \\ 100\ldots00 & 100\ldots00 & \cdots & 100\ldots00 \\ 010\ldots00 & 010\ldots00 & \cdots & 010\ldots00 \\ \vdots & \vdots & & \vdots \\ 000\ldots10 & 000\ldots10 & \cdots & 000\ldots10 \end{bmatrix} \text{ and } \boldsymbol{b} = (1,\ldots,1) \in \mathbb{R}^{2n-1}.$$

Birkhoff's Theorem (Cont'd)

- At least $n^2-(2n-1)=(n-1)^2$ of the a_{ij} must be zero. The others, $a_1,\ldots,a_{2n-1}$, say, satisfy a system of linear equations of the form

$$\boldsymbol{B}(a_1,\ldots,a_{2n-1})=\boldsymbol{b},$$

  where $\boldsymbol{B}$ is a non-singular $(2n-1)\times(2n-1)$ submatrix of $\boldsymbol{A}$.

  The matrix $\boldsymbol{B}$ satisfies the conditions of the lemma. So $\det\boldsymbol{B}=\pm1$.

  Thus the elements of $\boldsymbol{B}^{-1}$, and hence of $(a_1,\ldots,a_{2n-1})$, are integers.

  It follows that the doubly stochastic matrix (a_{ij}) has only integer elements. So it must be a permutation matrix.

  We complete the proof by noting that a polytope is the convex hull of its extreme points.

The $\boldsymbol{\lambda}$-Set of a Real Symmetric Matrix

- Suppose now that $\boldsymbol{\lambda}$ is an n-tuple of the (necessarily real) eigenvalues, in some order, of a real symmetric $n \times n$ matrix $\boldsymbol{A}$.
- The set $\Lambda_{\boldsymbol{A}}$ of all such n-tuples $\boldsymbol{\lambda}$ is called the **$\boldsymbol{\lambda}$-set** of $\boldsymbol{A}$.
- Clearly $\Lambda_{\boldsymbol{A}}$ is a finite set containing at most $n!$ points.

Theorem

Let $f : X \to \mathbb{R}$ be a convex function which is defined on a convex set X in $\mathbb{R}^n$ containing the $\boldsymbol{\lambda}$-set $\Lambda_{\boldsymbol{A}}$ of a real symmetric $n \times n$ matrix $\boldsymbol{A}$. Let $(\lambda_1, \ldots, \lambda_n)$ be a point of $\Lambda_{\boldsymbol{A}}$ where f assumes its maximum on $\Lambda_{\boldsymbol{A}}$. Then, for any orthonormal sequence $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ in $\mathbb{R}^n$,

$$f(\boldsymbol{v}_1^T \boldsymbol{A} \boldsymbol{v}_1, \ldots, \boldsymbol{v}_n^T \boldsymbol{A} \boldsymbol{v}_n) \leq f(\lambda_1, \ldots, \lambda_n).$$

Proof of the Theorem

- We show:
  - First that the point $\boldsymbol{v} = (\boldsymbol{v}_1^T \boldsymbol{A} \boldsymbol{v}_1, \ldots, \boldsymbol{v}_n^T \boldsymbol{A} \boldsymbol{v}_n)$ lies in X;
  - Then that $f(\boldsymbol{v}) \le f(\boldsymbol{\lambda})$, where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$.

  Let $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ be an orthonormal sequence of eigenvectors of $\boldsymbol{A}$ corresponding to the eigenvalues $\lambda_1, \ldots, \lambda_n$. Then, as we proved at the beginning of this section, there is a matrix $\boldsymbol{S}$ of Ω_n such that $\boldsymbol{v} = \boldsymbol{S}\boldsymbol{\lambda}$. By Birkhoff's Theorem, there exist $\mu_1, \ldots, \mu_m \ge 0$ with $\mu_1 + \cdots + \mu_m = 1$ such that $\boldsymbol{S} = \mu_1 \boldsymbol{P}_1 + \cdots + \mu_m \boldsymbol{P}_m$, where $\boldsymbol{P}_1, \ldots, \boldsymbol{P}_m$ are the permutation matrices in Ω_n. Hence

$$\boldsymbol{v} = \boldsymbol{S}\boldsymbol{\lambda} = \mu_1(\boldsymbol{P}_1\boldsymbol{\lambda}) + \cdots + \mu_m(\boldsymbol{P}_m\boldsymbol{\lambda}) \in \operatorname{conv}\Lambda_{\boldsymbol{A}} \subseteq X.$$

  The convexity of f shows that

$$f(\boldsymbol{v}) \le \mu_1 f(\boldsymbol{P}_1\boldsymbol{\lambda}) + \cdots + \mu_m f(\boldsymbol{P}_m\boldsymbol{\lambda}) \le \mu_1 f(\boldsymbol{\lambda}) + \cdots + \mu_m f(\boldsymbol{\lambda}) = f(\boldsymbol{\lambda}).$$

Nonnegative Semidefinite Matrices

Theorem

Let $\boldsymbol{A}$ be a non-negative semidefinite $n \times n$ matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$. Then, for any orthonormal sequence $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ in $\mathbb{R}^n$,

$$\det \boldsymbol{A} = \lambda_1 \cdots \lambda_n \leq \prod_{j=1}^{n} \boldsymbol{v}_j^T \boldsymbol{A} \boldsymbol{v}_j.$$

- Since $\boldsymbol{A}$ is non-negative semidefinite, $\lambda_1, \ldots, \lambda_n \geq 0$. The function $f : X \to \mathbb{R}$ defined on the non-negative orthant X of $\mathbb{R}^n$ by the equation

$$f(x_1, \ldots, x_n) = -(x_1 \cdots x_n)^{1/n}, \text{ for } x_1, \ldots, x_n \geq 0,$$

is easily seen to be convex from a previous corollary. The $\boldsymbol{\lambda}$-set of $\boldsymbol{A}$ is clearly contained in X. The preceding theorem shows that

$$-\left(\prod_{j=1}^{n} \boldsymbol{v}_j^T \boldsymbol{A} \boldsymbol{v}_j\right)^{1/n} \leq -(\lambda_1 \cdots \lambda_n)^{1/n} = -(\det \boldsymbol{A})^{1/n}.$$

Hadamard's Determinant Inequality

Theorem (Hadamard's Determinant Inequality)

Let $\boldsymbol{A} = [a_{ij}]$ be a real $n \times n$ matrix. Then

$$(\det \boldsymbol{A})^2 \le (a_{11}^2 + \cdots + a_{n1}^2) \cdots (a_{1n}^2 + \cdots + a_{nn}^2).$$

If $\boldsymbol{A}$ is nonnegative semidefinite, then $\det \boldsymbol{A} \le a_{11} \cdots a_{nn}$.

- Let $\boldsymbol{B} = [b_{ij}]$ denote the nonnegative semidefinite matrix $\boldsymbol{A}^T \boldsymbol{A}$. Applying the preceding theorem to $\boldsymbol{B}$, and using the orthonormal sequence $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ of elementary vectors, we deduce that

$$(\det \boldsymbol{A})^2 = \det \boldsymbol{B} \le \prod_{j=1}^{n} \boldsymbol{e}_j^T \boldsymbol{B} \boldsymbol{e}_j = b_{11} \cdots b_{1n}.$$

Hence $(\det \boldsymbol{A})^2 \le (a_{11}^2 + \cdots + a_{n1}^2) \cdots (a_{1n}^2 + \cdots + a_{nn}^2)$.
When $\boldsymbol{A}$ is itself non-negative semidefinite, we apply the preceding theorem to $\boldsymbol{A}$ and the sequence $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ to get $\det \boldsymbol{A} \le a_{11} \cdots a_{nn}$.

Minkowski's Determinant Inequality

Theorem (Minkowski's Determinant Inequality)

Let $\boldsymbol{A}, \boldsymbol{B}$ be nonnegative semidefinite $n \times n$ matrices. Then

$$(\det(\boldsymbol{A}+\boldsymbol{B}))^{1/n} \geq (\det \boldsymbol{A})^{1/n} + (\det \boldsymbol{B})^{1/n}.$$

- Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ be an orthonormal sequence of eigenvectors of the non-negative semidefinite matrix $\boldsymbol{A}+\boldsymbol{B}$ corresponding to eigenvalues $\lambda_1, \ldots, \lambda_n$. Then, using previous proven inequalities,

$$\begin{aligned}
(\det(\boldsymbol{A}+\boldsymbol{B}))^{1/n} &= (\lambda_1 \cdots \lambda_n)^{1/n} \\
&= (\prod_{j=1}^n \boldsymbol{v}_j^T (\boldsymbol{A}+\boldsymbol{B}) \boldsymbol{v}_j)^{1/n} \\
&= (\prod_{j=1}^n (\boldsymbol{v}_j^T \boldsymbol{A} \boldsymbol{v}_j + \boldsymbol{v}_j^T \boldsymbol{B} \boldsymbol{v}_j))^{1/n} \\
&\geq (\prod_{j=1}^n \boldsymbol{v}_j^T \boldsymbol{A} \boldsymbol{v}_j)^{1/n} + (\prod_{j=1}^n \boldsymbol{v}_j^T \boldsymbol{B} \boldsymbol{v}_j)^{1/n} \\
&\geq (\det \boldsymbol{A})^{1/n} + (\det \boldsymbol{B})^{1/n}.
\end{aligned}$$

Diagonals of a Square Matrix

- A **diagonal** of a real $n \times n$ matrix $\boldsymbol{A} = [a_{ij}]$ is a finite sequence $a_{1\sigma(1)}, \ldots, a_{n\sigma(n)}$ of elements of $\boldsymbol{A}$, where $\sigma(1), \ldots, \sigma(n)$ is a permutation of $1, \ldots, n$.
- To form such a diagonal:
  - We first choose any element d_1 in the first row of $\boldsymbol{A}$.
  - Next we choose any element d_2 in the second row of $\boldsymbol{A}$ not lying in the same column as d_1.
  - Then we choose any element d_3 in the third row of $\boldsymbol{A}$ not lying in the same column as either d_1 or d_2.
  - Continuing in this way, we produce a diagonal $d_1, \ldots, d_n$ of $\boldsymbol{A}$.
- Clearly $\boldsymbol{A}$ has at most $n!$ different diagonals.
- The diagonal $a_{11}, \ldots, a_{nn}$ is called the **leading diagonal** of $\boldsymbol{A}$.

Positive Diagonals and Doubly Stochastic Matrices

- A diagonal $d_1, \ldots, d_n$ of $\boldsymbol{A}$ is said to be **positive** if $d_1, \cdots, d_n > 0$.
- It is a non-trivial fact that a doubly stochastic matrix always has a positive diagonal.

  Indeed, by Birkhoff's Theorem, each doubly stochastic matrix $\boldsymbol{A}$ in Ω_n can be expressed in the form

$$\boldsymbol{A} = \lambda_1 \boldsymbol{P}_1 + \cdots + \lambda_m \boldsymbol{P}_m,$$

  where $\boldsymbol{P}_1, \ldots, \boldsymbol{P}_m$ are permutation matrices and $\lambda_1, \ldots, \lambda_m > 0$ with $\lambda_1 + \cdots + \lambda_m = 1$.

  For each $i = 1, \ldots, n$, let $\boldsymbol{P}_1$, have a 1 in its ith row and $\sigma(i)$th column. Then $a_{1\sigma(1)}, \ldots, a_{n\sigma(n)}$ is a positive diagonal of $\boldsymbol{A}$.

A Corollary to Birkhoff's Theorem

Theorem

Let $\boldsymbol{C} = [c_{ij}]$ be a real $n \times n$ matrix. Then there exists a diagonal $c_{1\sigma(1)}, \ldots, c_{n\sigma(n)}$ of $\boldsymbol{C}$ such that

$$c_{1\sigma(1)} + \cdots + c_{n\sigma(n)} \leq \sum_{i,j=1}^{n} c_{ij} s_{ij},$$

for every doubly stochastic $n \times n$ matrix $\boldsymbol{S} = [s_{ij}]$.

- Define a function $f : \Omega_n \to \mathbb{R}$ by the equation

$$f(\boldsymbol{S}) = \sum_{i,j=1}^{n} c_{ij} s_{ij},$$

for each doubly stochastic matrix $\boldsymbol{S} = [s_{ij}]$ in Ω_n. Let $\boldsymbol{P}_1, \ldots, \boldsymbol{P}_m$ be the permutation matrices in Ω_n. Choose one of these matrices, $\boldsymbol{P} = [p_{ij}]$, say, for which $f(\boldsymbol{P}) = \min\{f(\boldsymbol{P}_1), \ldots, f(\boldsymbol{P}_m)\}$. Suppose that the 1 in the ith row of $\boldsymbol{P}$ lies in its $\sigma(i)$th column.

A Corollary to Birkhoff's Theorem (Cont'd)

- By Birkhoff's Theorem, each doubly stochastic matrix $\boldsymbol{S} = [s_{ij}]$ in Ω_n can be written in the form $\boldsymbol{S} = \lambda_1 \boldsymbol{P}_1 + \cdots + \lambda_m \boldsymbol{P}_m$, for some $\lambda_1, \ldots, \lambda_m \geq 0$ with $\lambda_1 + \cdots + \lambda_m = 1$. Thus,

$$f(\boldsymbol{S}) = \lambda_1 f(\boldsymbol{P}_1) + \cdots + \lambda_m f(\boldsymbol{P}_m) \geq f(\boldsymbol{P}).$$

Finally,

$$\begin{aligned} c_{1\sigma(1)} + \cdots + c_{n\sigma(n)} &= \sum_{i,j=1}^{n} c_{ij} p_{ij} \\ &= f(\boldsymbol{P}) \\ &\leq f(\boldsymbol{S}) \\ &= \sum_{i,j=1}^{n} c_{ij} s_{ij}. \end{aligned}$$

Doubly Stochastic Matrices and Average Size of a Diagonal

Theorem

Each doubly stochastic $n \times n$ matrix has a positive diagonal whose harmonic mean is at least $\frac{1}{n}$.

- Let $\boldsymbol{A} = [a_{ij}]$ be an $n \times n$ doubly stochastic matrix. Define an $n \times n$ matrix $[c_{ij}]$ by the equations

$$c_{ij} = \begin{cases} \frac{1}{a_{ij}}, & \text{for } a_{ij} > 0 \\ n^2 + 1, & \text{for} a_{ij} = 0. \end{cases}$$

By the preceding theorem, some diagonal $c_{1\sigma(1)}, \ldots, c_{n\sigma(n)}$ of $[c_{ij}]$ satisfies the inequalities

$$c_{1\sigma(1)} + \cdots + c_{n\sigma(n)} \leq \sum_{i,j=1}^{n} c_{ij} a_{ij} \leq n^2.$$

Now all the terms on the left-hand side are positive, and so no term can be equal to $n^2 + 1$.

Doubly Stochastic Matrices and Size of a Diagonal (Cont'd)

- This implies that, for $i = 1, \ldots, n$, $a_{i\sigma(i)} > 0$ and $c_{i\sigma(i)} = \frac{1}{a_{i\sigma(i)}}$. Thus, from the inequality, we get

$$\frac{1}{a_{1\sigma(1)}} + \cdots + \frac{1}{a_{n\sigma(n)}} \leq n^2.$$

Consequently, the harmonic mean

$$\left(\frac{1}{n}\left(\frac{1}{a_{1\sigma(1)}} + \cdots + \frac{1}{a_{n\sigma(n)}}\right)\right)^{-1}$$

of the diagonal $a_{1\sigma(1)}, \ldots, a_{n\sigma(n)}$ is at least $\frac{1}{n}$.

A Consequence

Corollary

Each doubly stochastic $n \times n$ matrix $[a_{ij}]$ has a positive diagonal $a_{1\sigma(1)}, \ldots, a_{n\sigma(n)}$ satisfying the inequalities

$$a_{1\sigma(1)} + \cdots + a_{n\sigma(n)} \geq 1 \quad \text{and} \quad a_{1\sigma(1)} \cdots a_{n\sigma(n)} \geq n^{-n}.$$

- By the theorem,

$$\frac{1}{n} \leq \frac{n}{\frac{1}{a_{1\sigma(1)}} + \cdots + \frac{1}{a_{n\sigma(n)}}}.$$

But the harmonic arithmetic and geometric means satisfy

$$\frac{n}{\frac{1}{a_{1\sigma(1)}} + \cdots + \frac{1}{a_{n\sigma(n)}}} \leq \sqrt[n]{a_{1\sigma(1)} \cdots a_{n\sigma(n)}} \leq \frac{a_{1\sigma(1)} + \cdots + a_{n\sigma(n)}}{n}.$$

Therefore, $a_{1\sigma(1)} + \cdots + a_{n\sigma(n)} \geq 1$ and $a_{1\sigma(1)} \cdots a_{n\sigma(n)} \geq n^{-n}$.