Elementary Inequalities

Proofs with characterization of the equality sign are given for the following elementary inequalities: Jensen's Inequality, Generalized Young's Inequality, AM-GM Inequality, Hölder's Inequality, Cauchy-Schwarz Inequality, Minkowski's Inequality, Power Means Inequalities, Newton's Inequalities, and Maclaurin's Inequalities.

1.1 Jensen's Inequality

A function f defined on an interval I is convex if for $x, y \in I$ and $\lambda \in [0,1]$,

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

It is strictly convex if strict inequality holds in this condition whenever $x \neq y$ and $\lambda \in (0,1)$.

Jensen's Inequality. *Let f be a convex function on the interval I. Then*

$$f(\lambda_1 x_1 + \cdots + \lambda_n x_n) \leq \lambda_1 f(x_1) + \cdots + \lambda_n f(x_n) \ ,$$

where

$$x_1, x_2, \cdots, x_n \in I \ , \quad \lambda_1, \lambda_2, \cdots, \lambda_n \in [0,1] \ ,$$

When f is strictly convex, let

$$I_1 = \{k : \ \lambda_k \in (0,1]\} \ , \qquad \text{and } I_2 = \{k : \ \lambda_k = 0\}.$$

The equality sign in this inequality holds if and only if all $x_k, \ k \in I_1$, are equal.

We point out that the linear combination $\sum_k \lambda_k x_k$ belongs to the same interval. WLOG letting $x_1 \leq x_2 \leq \cdots \leq x_n$,

$$x_1 = \Big(\sum_k \lambda_k\Big) x_1 \leq \sum_k \lambda_k x_k \leq \Big(\sum_k \lambda_k\Big) x_n = x_n \ .$$

Proof. We prove Jensen's Inequality by an inductive argument on the number of points. When $n = 2$, the inequality follows from the definition of convexity. Assuming that it is true for $n-1$ many points, we show its validity for n many points. Let $\lambda_1, \cdots, \lambda_n \in (0,1), \sum_k \lambda_k = 1$, and

$$y = \sum_{k=1}^{n-1} \frac{\lambda_k}{1-\lambda_n} x_k \in [x_1, x_{n-1}] \ .$$

Using first the definition of convexity and then the induction hypothesis,

$$\begin{aligned}
f(\lambda_1 x_1 + \cdots + \lambda_n x_n) &= f((1-\lambda_n)y + \lambda_n x_n) \\
&\le (1-\lambda_n) f(y) + \lambda_n f(x_n) \\
&= (1-\lambda_n) f\left(\sum_{k=1}^{n-1} \frac{\lambda_k}{1-\lambda_n} x_k \right) + \lambda_n f(x_n) \\
&\le (1-\lambda_n) \sum_{k=1}^{n-1} \frac{\lambda_k}{1-\lambda_n} f(x_k) + \lambda_n f(x_n) \\
&= \sum_{k=1}^{n} \lambda_k f(x_k) \ .
\end{aligned}$$

The case when $\lambda_k = 1$ for some k is trivial. On the other hand, when some λ_k is 0, the inequality reduces to one with fewer λ_k's, and its validity comes from the induction hypothesis.

When f is strictly convex and $\lambda_k \in (0,1)$ for all k, it follows straightly from definition that the strict inequality sign in Jensen's inequality holds when $n = 2, x_1 \neq x_2$. In general, let us assume that the strictly inequality sign holds when $x_1, \cdots, x_{n-1}$ are distinct and prove it when $x_1, \cdots, x_n$ are not all equal. For, when all $x_1, \cdots, x_n$ are distinct, the second $\le$ in the above inequalities becomes $<$ due to the induction hypothesis and hence the strict inequality holds for n. When some x_k's are equal, we can group the expression $\sum_{k=1}^n \lambda_k x_k$ into $\sum_{k=1}^m \mu_k y_k$ where all y_k's are distinct and m is less than n. In this case the desired result comes from the induction hypothesis.

The case of equality becomes trivial when some λ_k equals to 1. When $\lambda_k = 0$ for some k, the inequality is the same as we remove all terms containing $\lambda_k,\ k \in I_2$, from both sides. The rest $\lambda_k, k \in I_1$, are in $(0,1)$, so the desired conclusion follows as before.

$\square$

Jensen's inequality asserts there is an inequality associated to every convex function. As an example, we have

Generalized Young's Inequality. *For $a_k \in (0,\infty)$ and $\lambda_k \in (0,1)$ with $\sum_{k=1}^n \lambda_k = 1$,*

$$a_1 a_2 \cdots a_n \le \frac{a_1^{p_1}}{p_1} + \frac{a_2^{p_2}}{p_2} + \cdots + \frac{a_n^{p_n}}{p_n} \ .$$

Moreover, the equality sign in this inequality holds if and only if all $a_k^{p_k}, k = 1, \cdots, n$, *are equal.*

Proof. The function $z \mapsto e^z$ is strictly convex on $(-\infty, \infty)$ (use $(e^z)'' > 0$). Its associated Jensen's Inequality takes the form

$$e^{\sum_k \lambda_k x_k} \leq \sum_k \lambda_k e^{x_k} \ .$$

The Generalized Young's Inequality follows by setting $a_k = e^{\lambda_k x_k}$ and $p_k = 1/\lambda_k$. $\square$

By taking $x_k = a_k^{p_k}$ and $p_k = n$ for all k in the Generalized Young's Inequality, we recover the inequality on arithmetic and geometric means.

AM-GM Inequality. *For* $x_k, k = 1, \cdots, n, \in (0, \infty)$,

$$(x_1 x_2 \cdots x_n)^{1/n} \leq \frac{x_1 + x_2 + \cdots + x_n}{n} \ .$$

Moreover, equality sign in this inequality holds if and only if all x_k*'s are equal.*

Jensen's Inequality concerning convex functions is a parent inequality. In the next section we use it to prove Hölder's Inequality.

1.2 Hölder's Inequality

For $\mathbf{a} = (a_1, \cdots, a_n) \in \mathbb{R}^n$ and $p \geq 1$, define

$$\|\mathbf{a}\|_p = \left(\sum_{k=1}^n |a_k|^p \right)^{1/p} \ .$$

Hölder's Inequality. *For* $p \geq 1$ *and* $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$ *with* $a_k, b_k \geq 0$,

$$\sum_{k=1}^n a_k b_k \leq \|\boldsymbol{a}\|_p \ \|\boldsymbol{b}\|_q \ , \quad \frac{1}{p} + \frac{1}{q} = 1 \ .$$

Moreover, the equality sign in this inequality holds if and only if either (a) one of $\boldsymbol{a}, \boldsymbol{b}$ *is the zero vector or (b) both* $\boldsymbol{a}$ *and* $\boldsymbol{b}$ *are non-zero vectors and* $b_k^q = c a_k^p$, $k = 1, \cdots, n$, *for some positive number* c.

The number q is called conjugate to p if $1/p+1/q=1$. Note that $q>1$ when $p>1$.

We will present two proofs for this basic inequality.

First Proof When $\mathbf{a}$ or $\mathbf{b}$ is a zero vector, the inequality becomes equality and the assertion is trivially satisfied. It suffices to consider the case where $\mathbf{a}$ or $\mathbf{b}$ is a non-zero vector. WLOG we assume $\mathbf{a} \neq (0,\cdots,0)$ in the following proof.

Apply Young's Inequality of two variables to each pair of a_k, b_k and play the "ε-trick":

$$a_k b_k = (\varepsilon a_k)(\varepsilon^{-1} b_k) \le \frac{(\varepsilon a_k)^p}{p} + \frac{(\varepsilon^{-1} b_k)^q}{q} \ , \quad \varepsilon > 0 \ . \tag{1.1}$$

Summing up over k

$$\sum_{k=1}^n a_k b_k \le \frac{\varepsilon^p}{p}\|\mathbf{a}\|_p^p + \frac{\varepsilon^{-q}}{q}\|\mathbf{b}\|_q^q \ .$$

Now we choose ε by the relation

$$\varepsilon^p\|\mathbf{a}\|_p^p = \varepsilon^{-q}\|\mathbf{b}\|_q^q \ ,$$

that is,

$$\varepsilon = \left(\frac{\|\mathbf{b}\|_q^q}{\|\mathbf{a}\|_p^p}\right)^{1/(p+q)} \ .$$

With this ε we find

$$\varepsilon^p\|\mathbf{a}\|_p^p = \varepsilon^{-q}\|\mathbf{b}\|_q^q = \|\mathbf{a}\|_p \ \|\mathbf{b}\|_q \ ,$$

after some straightforward manipulations. Therefore,

$$\sum_{k=1}^n a_k b_k \le \frac{\|\mathbf{a}\|_p \ \|\mathbf{b}\|_q}{p} + \frac{\|\mathbf{a}\|_p \ \|\mathbf{b}\|_q}{q} = \|\mathbf{a}\|_p \ \|\mathbf{b}\|_q \ ,$$

done.

To settle the equality sign, first it is easy to check directly that equality holds if $\mathbf{b}$ is a scalar multiple of $\mathbf{a}$. On the other hand, observe that in (1.1), for each k, equality sign holds if and only if either $a_k=b_k=0$ or $a_k>0$ and $\varepsilon^p a_k = \varepsilon^{-q} b_k^q$, that is $b_k = \varepsilon^{p+q} a_k$. Hölder's Inequality is obtained by summing up all these inequalities. Therefore, letting $J_1=\{k: \ a_k=b_k=0\}$ and $J_2=\{k: a_k, b_k>0\}$, we know that

$$b_k^q = c a_k^p \ , \quad c = \varepsilon^{p+q} > 0 \ ,$$

whenever $k \in J_2$. But, this relation is also valid for $k \in J_1$. We conclude that the equality sign holds in Hölder's Inequality when $\mathbf{a}$ is a non-zero vector implies that $b_k^q = ca_k^p,\ k = 1, \cdots, n$, for some $c > 0$.

Second Proof In this proof we assume $a_k, b_k > 0$. The reader will have no difficulty to extend it to $a_k, b_k \geq 0$. Moreover, I leave the proof of the equality case to you. It is not hard but tedious.

The Jensen's Inequality associated to the strictly convex function $f(x) = x^p,\ x \in (0, \infty),\ p > 1$, is

$$\Big(\sum_{k=1}^n \lambda_k x_k \Big)^p \leq \sum_{k=1}^n \lambda_k x_k^p \ .$$

Choosing

$$a_k = \lambda_k^{1/p} x_j \ ,$$

and writing $c_k = \lambda_k^{1/q}$, the inequality becomes

$$\sum_{k=1}^n a_k c_k \leq \Big(\sum_{k=1}^n a_k^p \Big)^{1/p} \ ,$$

whenever

$$\sum_{k=1}^n c_k^q = 1 \ .$$

Now, given $\mathbf{b} = (b_1, \cdots, b_n),\ b_k > 0$, the numbers

$$c_k = \frac{b_k}{\|\mathbf{b}\|_q}$$

satisfy

$$\sum_k c_k^q = 1 \ .$$

Therefore,

$$\sum_k a_k \frac{b_k}{\|\mathbf{b}\|_q} \leq \|\mathbf{a}\|_p \ ,$$

and Hölder's Inequality follows.

Recall that the Euclidean product (dot product) in $\mathbb{R}^n$ is given by

$$\mathbf{a} \cdot \mathbf{b} = \sum_{k=1}^n a_k b_k \quad .$$

A slightly more general form of Hölder's Inequality is

Hölder's Inequality. *For $p \geq 1$ and $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$,*

$$|\boldsymbol{a} \cdot \boldsymbol{b}| \leq \|\boldsymbol{a}\|_p \ \|\boldsymbol{b}\|_q \ , \quad \frac{1}{p} + \frac{1}{q} = 1 \ .$$

Moreover, the equality sign in this inequality holds if and only if either (a) one of $\boldsymbol{a}$ or $\boldsymbol{b}$ is the zero vector or (b) both $\boldsymbol{a}$ and $\boldsymbol{b}$ are non-zero, their non-zero components are of the same sign, and $|b_k|^q = c|a_k|^p$, $k = 1, \cdots, n$, for some positive number c.

Proof. Applying the first Hölder's Inequality to $|a_k|, |b_k|$ and using the triangle inequality,

$$\left|\sum_k a_k b_k\right| \leq \sum_k |a_k||b_k| \leq \|\mathbf{a}\|_p \ \|\mathbf{b}\|_q \ .$$

To establish the equality case simply observe that when p_j's are non-zero numbers, $|p_1 + \cdots + p_m| = |p_1| + \cdots + |p_m|$ if and only if all p_j's are of the same sign.

□

1.3 Minkowski's Inequality

Minkowski's Inequality. *For $\boldsymbol{a}$, $\boldsymbol{b} \in \mathbb{R}^n$ and $p \geq 1$,*

$$\|\boldsymbol{a} + \boldsymbol{b}\|_p \leq \|\boldsymbol{a}\|_p + \|\boldsymbol{b}\|_p \ .$$

Moreover, the equality sign in this inequality holds if and only if either (a) $\boldsymbol{a}$ or $\boldsymbol{b}$ is a zero vector or (b) $\boldsymbol{b} = c\boldsymbol{a}$ for some $c \geq 0$.

So equality holds if and only if the two non-zero vectors point to the same direction or one of them is null. The case $p = 1$ is just the familiar triangle inequality. The case $p = 2$ follows readily from Cauchy-Schwarz Inequality. In the following proof we assume $p > 1$.

Proof. By Hölder's Inequality

$$
\begin{aligned}
&\|\mathbf{a}+\mathbf{b}\|_p^p \\
&= \sum_k |a_k+b_k|^{p-1}|a_k+b_k| \\
&\le \sum_k |a_k+b_k|^{p-1}(|a_k|+|b_k|) \\
&= \sum_k |a_k+b_k|^{p-1}|a_k| + \sum_k |a_k+b_k|^{p-1}|b_k| \\
&\le \left(\sum_k |a_k+b_k|^{(p-1)q}\right)^{1/q}\left(\sum_k |a_k|^p\right)^{1/p} + \left(\sum_k |a_k+b_k|^{(p-1)q}\right)^{1/q}\left(\sum_k |b_k|^p\right)^{1/p} \\
&= \|\mathbf{a}+\mathbf{b}\|_p^{p/q}(\|\mathbf{a}\|_p+\|\mathbf{b}\|_p)\ ,
\end{aligned}
$$

where in the last step Hölder's Inequality and $(p-1)q=p$ have been used. Now Minkowski's Inequality follows by absorbing the first factor to the left.

When one of $\mathbf{a}$ and $\mathbf{b}$ is a zero vector, the equality sign holds. Let us consider the case where both vectors are non-zero. From what we have just done equality in Minkowski's Inequality holds if and only if the two inequality signs in the above derivations are equality. When the second inequality sign becomes equality, $|a_k+b_k|^p = c_1|a_k|^p$, $|a_k+b_k|^p = c_2|b_k|^p$ for all k and $c_1, c_2 > 0$, which implies that $|b_k| = c|a_k|, k \ge 1$, for some $c > 0$. But then the first inequality becomes equality implies that both a_k and b_k must be of the same sign.

□

1.4 Cauchy-Schwarz Inequality

Cauchy-Schwarz Inequality. *For* $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$,

$$|\mathbf{a}\cdot\mathbf{b}| \le \|\mathbf{a}\|_2\ \|\mathbf{b}\|_2\ ,$$

and equality sign in this inequality holds if and only if $\boldsymbol{a}$ *and* $\boldsymbol{b}$ *are linearly dependent.*

First Proof. It is the special case of Hölder's Inequality ($p=2$).

Second Proof. It suffices to assume all a_k, b_k's are non-negative and $\mathbf{a} \neq \mathbf{0}$. Consider the function

$$
\begin{aligned}
\varphi(t) &= \sum_k (a_k t - b_k)^2 \\
&= At^2 - Bt + C\ ,
\end{aligned}
$$

where
$$A = \|\mathbf{a}\|_2^2, \quad B = 2\sum_k a_k b_k \ , \quad C = \|\mathbf{b}\|_2^2 \ .$$
This function is always non-negative. Therefore, its discriminant $\Delta = B^2 - 4AC$ must be non-positive everywhere. This is precisely the Cauchy-Schwarz Inequality. Moreover, $\varphi(t) = 0$ has a real root t_0 if and only if $a_k t_0 = b_k$ for all k.

Third Proof. It is contained in the Lagrange Identity:
$$\Big(\sum_{k=1}^n a_k b_k\Big)^2 = \sum_{k=1}^n a_k^2 \sum_{j=1}^n b_j^2 - \frac{1}{2}\sum_{j,k=1}^n \big(a_j b_k - a_k b_j\big)^2 \ .$$
To prove this identity we play with the indices:
$$\begin{aligned}
\Big(\sum_j a_j^2\Big)\Big(\sum_k b_k^2\Big) - \Big(\sum_k a_k b_k\Big)^2 &= \Big(\sum_j a_j^2\Big)\Big(\sum_k b_k^2\Big) - \Big(\sum_k a_k b_k \sum_j a_j b_j\Big) \\
&= \frac{1}{2}\sum_{j,k}(a_j^2 b_k^2 + a_k^2 b_j^2) - \Big(\sum_k a_k b_k\Big)\Big(\sum_j a_j b_j\Big) \\
&= \frac{1}{2}\sum_{j,k}\big(a_j^2 b_k^2 - 2a_j b_k a_k b_j + a_k^2 b_j^2\big) \\
&= \frac{1}{2}\sum_{j,k}(a_j b_k - a_k b_j)^2 \ .
\end{aligned}$$

I leave it to you to prove that $a_j b_k - a_k b_j = 0$ for all j, k implies $\mathbf{b} = c\mathbf{a}$ for some $c \in \mathbb{R}$.

1.5 The Power Means Inequalities

In mathematics, a mean is a real-valued function F on $\mathbb{R}^n$ or its subset which satisfies (a) $F(a, a, \cdots, a) = a$, and (b) $\min_k a_k \le F(\mathbf{a}) \le \max_k a_k$.

Let $\lambda_k,\ k = 1, \cdots, n$, satisfy $\sum_k \lambda_k = 1,\ \lambda_k \in [0, 1]$. For each $t \in \mathbb{R},\ t > 0,\ \mathbf{a} = (a_1, \cdots, a_n),\ a_k \ge 0$, define the **t-th power mean** by
$$M_t(\mathbf{a}) = \Big(\sum_{k=1}^n \lambda_k a_k^t\Big)^{1/t} \ .$$
When $a_k > 0$ for all k, $M_t(\mathbf{a})$ is well-defined for $t \in (-\infty, 0)$.

The power mean becomes the generalized arithmetic mean at $t = 1$ and the generalized harmonic mean at $t = -1$. As we will see, it becomes the generalized geometric mean at the limit $t = 0$. Therefore, it inserts $M_t(\mathbf{a}),\ t \in (0,1)$, between the arithmetic and geometric means.

Power Means Inequality. *For* $k = 1, \cdots, n$, *let* $a_k \in (0,\infty)$ *be distinct and* $\lambda_k \in (0,1)$ *satisfying* $\sum_{k=1}^n \lambda_k = 1$. Regarding $M_t(\mathbf{a})$ as a function in t, we have

(a) $M_t(\mathbf{a})$ *is strictly increasing on* $(-\infty, 0)$ and $(0,\infty)$,

(b)
$$\lim_{t\to 0} M_t(\mathbf{a}) = a_1^{\lambda_1} a_2^{\lambda_2} \cdots a_n^{\lambda_n} \ , \qquad \textit{and}$$

(c)
$$\lim_{t\to\infty} M_t(\mathbf{a}) = \max_k a_k \ , \quad \lim_{t\to-\infty} M_t(\mathbf{a}) = \min_k a_k \ .$$

Thus, unless all a_k's are equal, $M_t(\mathbf{a})$ is a strictly increasing, continuous function on $\mathbb{R}$ after defining $M_0(\mathbf{a})$ as in (b).

Proof. Step 1. We first show that for $0 < s < t$, $M_s(\mathbf{a}) \le M_t(\mathbf{a})$. Since $\mathbf{a}$ is fixed throughout the proof we will simply use M_t to stand for $M_t(\mathbf{a})$ below.

We take $p = t/s$ and $q = t/(t-s)$ and apply Hölder's Inequality:

$$\begin{aligned}
\sum_k \lambda_k a_k^s &= \sum_k \lambda_k^{s/t} a_k^s \lambda_k^{(t-s)/t} \\
&\le \left(\sum_k \left(\lambda_k^{s/t} a_k^s\right)^{t/s} \right)^{s/t} \left(\sum_k \left(\lambda_k^{(t-s)/t}\right)^{t/(t-s)} \right)^{(t-s)/t} \\
&= \left(\sum_k \lambda_k a_k^t \right)^{s/t} ,
\end{aligned}$$

which implies $M_s \le M_t$ for $0 < s < t$. In case of equality, we have $\lambda_k a_k^t = c\lambda_k, c > 0$, for all k, whence all a_k's are the same. But this is impossible due to our assumption.

Step 2. We claim:

$$\lim_{t\to 0^+} M_t(\mathbf{a}) = a_1^{\lambda_1} a_2^{\lambda_2} \cdots a_n^{\lambda_n} \ .$$

For,

$$\begin{aligned}\log M_t &= \frac{1}{t}\log\Big(\sum_k \lambda_k e^{t\log a_k}\Big)\\ &= \frac{1}{t}\log\Big(\sum_k \lambda_k\big(1+t\log a_k+\circ(t)\big)\Big)\\ &= \frac{1}{t}\log\Big(1+t\sum_k \lambda_k\log a_k+\circ(t)\Big)\\ &= \frac{1}{t}\Big(t\sum_k \lambda_k\log a_k+\circ(t)\Big)\\ &\to \sum_k \lambda_k \log a_k\ , \qquad \text{as } t\to 0^+\ ,\end{aligned}$$

after using the expansions $e^z = 1+z+\circ(z)$ and $\log(1+z) = z+\circ(z)$ at $z=0$. Using the continuity of the exponential function, we get

$$\begin{aligned}\lim_{t\to 0^+} M_t &= e^{\lim_{t\to 0^+}\log M_t}\\ &= e^{\sum_k \lambda_k\log a_k}\\ &= a_1^{\lambda_1}\cdots a_n^{\lambda_n}\ .\end{aligned}$$

Step 3. We claim that for $t<s<0$, $M_t\le M_s$. For, letting $b_k = 1/a_k$, $k = 1,\cdots,n$, we have

$$M_t(\mathbf{a}) = \frac{1}{M_{-t}(\mathbf{b})}\ .$$

Therefore, using Step 1 for $0<-s<-t$,

$$\begin{aligned}M_t(\mathbf{a}) &= \frac{1}{M_{-t}(\mathbf{b})}\\ &\le \frac{1}{M_{-s}(\mathbf{b})}\\ &= M_s(\mathbf{a})\ ,\end{aligned}$$

and the claim follows. Note that this implies

$$\lim_{t\to 0^-} M_t(\mathbf{a}) = a_1^{\lambda_1}\cdots a_n^{\lambda_n}\ .$$

Step 4. We claim $\lim_{t\to\infty} M_t(\mathbf{a}) = \max_k a_k$. Let us assume $a_1<a_2<\cdots<a_n$. We have

$$a_n \ge \Big(\sum_k \lambda_k a_k^t\Big)^{1/t} = a_n\left(\sum_{k=1}^{n-1}\lambda_k\frac{a_k^t}{a_n^t}+\lambda_n\right)^{1/t} \ge a_n\lambda_n^{1/t}\ .$$

The desired conclusion follows from the Sandwich rule.

Finally, arguing as in Step 3 we get $\lim_{t\to-\infty} M_t(\mathbf{a}) = a_1$. The proof of the Power Means Inequalities is completed.

□

Some remarks are in order.

(a) It is reasonable to assume all λ_k's are positive and less than one so that all a_k's are involved in $M_t(\mathbf{a})$.

(b) If some a_k's are equal, we may group them together. For instance, when $a_{n-1} = a_n$, we set $a'_1 = a_1, \cdots, a'_{n-1} = a_{n-1} + a_n$ and $\mu_1 = \lambda_1, \cdots, \mu_{n-1} = \lambda_{n-1} + \lambda_n$ in obvious notation. Then $\sum_{k=1}^{n-1} \mu_k = 1$ and $M_t(\mathbf{a'}) = M_t(\mathbf{a})$. Therefore, it is without loss of general to assume all a_k's are distinct.

(c) Under $0 \le a_1 < a_2 < \cdots < a_n$ and $\lambda_k \in (0,1),\ k = 1, \cdots, n$, an examination of the proof in Step 1 shows that $M_t(\mathbf{a})$ is strictly increasing for $t \in (0, \infty)$.

1.6 Newton's Inequalities

Let $S_k(\mathbf{a}),\ k = 0, 1, \cdots, n$ be the k-th elementary symmetric function of $\mathbf{a} = (a_1, \cdots, a_n)$. For instance,

$$S_0(\mathbf{a}) = 1,$$

$$S_1(\mathbf{a}) = \sum_k a_k = a_1 + a_2 + \cdots + a_n \ ,$$

$$S_2(\mathbf{a}) = \sum_{1\le j<k\le n} a_j a_k \ ,$$

$$S_n(\mathbf{a}) = a_1 \cdots a_n \ .$$

The normalized elementary symmetric functions σ_k's are given by dividing S_k by the number of elements in the summation. In general,

$$\sigma_k(\mathbf{a}) = \frac{S_k(\mathbf{a})}{\binom{n}{k}} \ .$$

Newton's Inequalities. *For* $\boldsymbol{a} = (a_1, \cdots, a_n),\ a_k > 0,$

$$\sigma_{k-1}(\boldsymbol{a})\sigma_{k+1}(\boldsymbol{a}) \le \sigma_k^2(\boldsymbol{a}), \quad k = 1, 2, \cdots, n-1 \ .$$

Equality sign holds in one of these inequalities if and only if all a_k's are equal.

The proof of these inequalities is based on the following amazing property.

Proposition. *For $\boldsymbol{a} = (a_1, \cdots, a_n)$, $n \geq 2$, there is some $\boldsymbol{b} = (b_1, \cdots, b_{n-1})$ such that*

$$\sigma_k(\boldsymbol{b}) = \sigma_k(\boldsymbol{a}) \ , \quad k = 0, 1, \cdots, n-1 \ .$$

Proof. Define the polynomial of degree n by

$$\begin{aligned} P(x) &= (x-a_1)(x-a_2)\cdots(x-a_n) \\ &= x^n - S_1(\mathbf{a})x^{n-1} + S_2(\mathbf{a})x^{n-2} + \cdots + (-1)^{n-1}S_{n-1}(\mathbf{a})x + (-1)^n S_n(\mathbf{a}) \ . \end{aligned}$$

P has n many real roots (counting multiplicity). By Rolle's Theorem its derivative has $n-1$ real roots (counting multiplicity). Denote them by $b_1, \cdots, b_n$. We have

$$\begin{aligned} \frac{P'(x)}{n} &= (x-b_1)(x-b_2)\cdots(x-b_{n-1}) \\ &= x^{n-1} - S_1(\mathbf{b})x^{n-2} + S_2(\mathbf{b})x^{n-3} + \cdots + (-1)^{n-1}S_{n-1}(\mathbf{b}) \ . \end{aligned}$$

On the other hand, we have

$$P'(x) = nx^{n-1} - (n-1)S_1(\mathbf{a})x^{n-2} + \cdots + (-1)^{n-1}S_{n-1}(\mathbf{a}) \ .$$

By comparing the coefficients of these two polynomials,

$$\frac{n-k}{n} S_k(\mathbf{a}) = S_k(\mathbf{b}), \quad k = 0, 1, \cdots, n-1 \ ,$$

and the proposition follows after dividing both sides by $\binom{n-1}{k}$.

□

Let us look at the first several Newton's Inequalities (we omit the variable $\mathbf{a}$):

- $$n = 2, \quad \sigma_0\sigma_2 \leq \sigma_1^2 \ ,$$
- $$n = 3, \quad \sigma_0\sigma_2 \leq \sigma_1^2 \ , \quad \sigma_1\sigma_3 \leq \sigma_2^2 \ ,$$
- $$n = 4, \quad \sigma_0\sigma_2 \leq \sigma_1^2 \ , \quad \sigma_1\sigma_3 \leq \sigma_2^2 \ , \quad \sigma_2\sigma_4 \leq \sigma_3^2 \ ,$$

- $$n=5, \quad \sigma_0\sigma_2 \le \sigma_1^2\ , \quad \sigma_1\sigma_3 \le \sigma_2^2\ , \quad \sigma_2\sigma_4 \le \sigma_3^2\ , \quad \sigma_3\sigma_5 \le \sigma_4^2\ .$$

The last member in these inequalities is

$$\sigma_{n-2}(\mathbf{a})\sigma_n(\mathbf{a}) \le \sigma_{n-1}^2(\mathbf{a})\ .$$

It is turned into

$$\sigma_0(\mathbf{c})\sigma_2(\mathbf{c}) \le \sigma_1(\mathbf{c})^2\ , \quad \mathbf{c}=(c_1,\cdots,c_n),\ c_k = 1/a_k\ ,$$

after dividing both sides by $(a_1\cdots a_n)^2$. It has become the first inequality for the variable $\mathbf{c}$.

Now we can prove Newton's Inequalities. When $n=2$, the inequality is nothing but the AM-GM Inequality

$$a_1a_2 \le \left(\frac{a_1+a_2}{2}\right)^2\ .$$

Next, for $n=3$, there are two inequalities. Our proposition shows that the first inequality follows from the inequality in the previous case, that is, the inequality above. Moreover, the second inequality can be reduced to the first one after replacing $\mathbf{a}$ by $\mathbf{c}$. The strategy works for all n. After the $(n-1)$-th case has been established, the first $(n-1)$ inequalities in the n-th case are valid by applying the proposition. The n-th inequality can be reduced to the first inequality for $\mathbf{c}$, $c_k=1/a_k$.

To study the equality case, we may apply induction on n, the number of variables, in the statement:

$\sigma_{k-1}(\mathbf{a})\sigma_{k+1}(\mathbf{a}) = \sigma_k^2(\mathbf{a})$ for some $k\in\{1,\cdots,n-1\}$ implies that all a_k's are equal.

Clearly it is valid at $n=2$. In general, assuming that it holds at $n-1$ we consider the case of n variables. If on the contrary $\sigma_{k-1}(\mathbf{a})\sigma_{k+1}(\mathbf{a}) = \big(\sigma_k(\mathbf{a})\big)^2$ for some $k\in\{1,\cdots,n-1\}$ where $\mathbf{a}$ satisfies $a_j<a_{j+1}$ for some j. By Rolle's Theorem the polynomial $P'(x)$ defined above admits a root in (a_j,a_{j+1}). Together with other real roots, $P'(x)$ has at least two distinct roots. When $k\le n-2$, using the proposition above we have $\sigma_{k-1}(\mathbf{b})\sigma_{k+1}(\mathbf{b}) = \sigma_k^2(\mathbf{b})$, contradicting the induction hypothesis. On the other hand, the case $k=n-1$ can be reduced to the case $k=1$ for the variable $\mathbf{c}$. We have completed the proof of the equality case for Newton's Inequalities.

The following inequality can be deduced from Newton's Inequalities.

Maclaurin's Inequalities. *For* $\mathbf{a} = (a_1, \cdots, a_n), a_k > 0, k = 1, \cdots, n,$

$$\big(\sigma_{k+1}(\mathbf{a})\big)^{1/(k+1)} \le \big(\sigma_k(\mathbf{a})\big)^{1/k}\ , \quad k = 1, \cdots, n-1\ .$$

Moreover, equality sign holds for some k *if and only if* $a_1 = \cdots = a_n$.

These inequalities insert $n-2$ many terms between the GM and AM.

Proposition. *Let* $c_0, c_1, \cdots, c_n \in \mathbb{R}$, *satisfying*

$$c_k \le \frac{1}{2}(c_{k-1} + c_{k+1})\ , \quad k = 1, \cdots, n-1\ .$$

Then

$$\frac{c_k - c_0}{k} \le \frac{c_{k+1} - c_0}{k+1}\ , \quad 1 \le k \le n-1\ . \tag{1.2}$$

Proof. When $k = 1$, (1.2) is the same as $2c_1 \le c_0 + c_2$. Assuming this inequality holds at $k-1$, that is,

$$kc_{k-1} \le (k-1)c_k + c_0\ ,$$

we are going to establish it at k. Indeed, we have

$$2c_k \le c_{k-1} + c_{k+1} \le \frac{k-1}{k}c_k + \frac{1}{k}c_0 + c_{k+1}\ ,$$

which implies

$$(k+1)c_k \le kc_{k+1} + c_0\ ,$$

that is, (1.2) holds.

□

Now we prove Maclaurin's Inequalities. Taking logarithm in Newton's Inequalities yields

$$\log \sigma_k(\mathbf{a}) \ge \frac{1}{2}\big(\log \sigma_{k-1}(\mathbf{a}) + \log \sigma_{k+1}(\mathbf{a})\big)\ . \tag{1.3}$$

Applying the proposition to

$$c_k = -\log \sigma_k(\mathbf{a})\ , \quad k = 0, 1, \cdots, n,$$

yields

$$\frac{-\log \sigma_k(\mathbf{a})}{k} = \frac{c_k - c_0}{k} \le \frac{c_{k+1} - c_0}{k+1} = \frac{-\log \sigma_{k+1}(\mathbf{a})}{k+1}\ ,$$

and Maclaurin's Inequalities follow. The equality sign in Maclaurin's Inequality holds at some k means equality in (1.3). By the characterization of the equality sign in Newton's Inequalities we conclude that all $a_1 = \cdots = a_n$.

Recommended Reading: J.M. Steele, The Cauchy-Schwarz Master Class, *An Introduction to the Art of Mathematical Inequalities*, MAA Problem Book Series, Cambridge University Press, 2008.