1. BASIC INEQUALITIES

The arsenal of an analyst is stocked with inequalities. In this chapter we present briefly some of the simplest and most useful of these. It is an indication of the size of the subject that, although our aims are very modest, this chapter is rather long.

Perhaps the most basic inequality in analysis concerns the arithmetic and geometric means; it is sometimes called the AM-GM inequality. The *arithmetic mean* of a sequence $a = (a_1, \dots, a_n)$ of n reals is

$$A(a) = \frac{1}{n} \sum_{i=1}^{n} a_i;$$

if each a_i is non-negative then the *geometric mean* is

$$G(a) = \left(\prod_{i=1}^{n} a_i \right)^{1/n},$$

where the non-negative nth root is taken.

Theorem 1. The geometric mean of n non-negative reals does not exceed their arithmetic mean: if $a = (a_1, \dots, a_n)$ then

$$G(a) \leqslant A(a). \tag{1}$$

Equality holds iff $a_1 = \cdots = a_n$.

Proof. This inequality has many simple proofs; the witty proof we shall present was given by Augustin-Louis Cauchy in his Cours d'Analyse (1821). (See Exercise 1 for another proof.) Let us note first that the theorem holds for $n = 2$. Indeed,

$$(a_1 - a_2)^2 = a_1^2 - 2a_1a_2 + a_2^2 \geqslant 0;$$

so

$$(a_1+a_2)^2 \geqslant 4a_1a_2,$$

with equality iff $a_1 = a_2$.

Suppose now that the theorem holds for $n = m$. We shall show that it holds for $n = 2m$. Let $a_1, \ldots, a_m, b_1, \ldots, b_m$ be non-negative reals. Then

$$\begin{aligned}(a_1\ldots a_m b_1\ldots b_m)^{1/2m} &= \{(a_1\ldots a_m)^{1/m}(b_1\ldots b_m)^{1/m}\}^{1/2}\\ &\leqslant \tfrac{1}{2}\{(a_1\ldots a_m)^{1/m}+(b_1\ldots b_m)^{1/m}\}\\ &\leqslant \frac{1}{2}\left(\frac{a_1+\ldots+a_m}{m}+\frac{b_1+\ldots+b_m}{m}\right)\\ &= \frac{a_1+\ldots+a_m+b_1+\ldots+b_m}{2m}.\end{aligned}$$

If equality holds then, by the induction hypothesis, we have $a_1 = \cdots = a_m = b_1 = \cdots = b_m$. This implies that the theorem holds whenever n is a power of 2.

Finally, suppose n is an arbitrary integer. Let

$$n < 2^k = N \qquad \text{and} \qquad a = \frac{1}{n}\sum_{i=1}^{n} a_i.$$

Set $a_{n+1} = \cdots = a_N = a$. Then

$$\prod_{i=1}^{N} a_i = a^{N-n}\prod_{i=1}^{n} a_i \leqslant \left(\frac{1}{N}\sum_{i=1}^{N} a_i\right)^N = a^N;$$

so

$$\prod_{i=1}^{n} a_i \leqslant a^n,$$

with equality iff $a_1 = \cdots = a_N$, in other words iff $a_1 = \cdots = a_n$. □

In 1906 Jensen obtained some considerable extensions of the AM-GM inequality. These extensions were based on the theory of convex functions, founded by Jensen himself.

A subset D of a real vector space is *convex* if every convex linear combination of a pair of points of D is in D, i.e. if $x, y \in D$ and $0 < t < 1$ imply that $tx+(1-t)y \in D$. Note that if D is convex, $x_1, \ldots, x_n \in D$, $t_1, \ldots, t_n > 0$ and $\sum_{i=1}^n t_i = 1$ then $\sum_{i=1}^n t_i x_i \in D$.

Indeed, assuming that a convex linear combination of $n-1$ points of D is in D, we find that

$$x_2' = \sum_{i=2}^{n} \frac{t_i}{1-t_1} x_i \in D$$

and so

$$\sum_{i=1}^{n} t_i x_i = t_1 x_1 + (1-t_1) x_2' \in D.$$

Given a convex subset D of a real vector space, a function $f: D \to \mathbb{R}$ is said to be *convex* if

$$f(tx+(1-t)y) \leqslant tf(x)+(1-t)f(y) \tag{2}$$

whenever $x, y \in D$ and $0 < t < 1$. We call f *strictly convex* if it is convex and, moreover, $f(tx+(1-t)y) = tf(x)+(1-t)f(y)$ and $0 < t < 1$ imply that $x = y$. Thus f is strictly convex if strict inequality holds in (2) whenever $x \neq y$ and $0 < t < 1$. A function f is *concave* if $-f$ is convex and it is *strictly concave* if $-f$ is strictly convex. Clearly, f is convex iff the set $\{(x,y) \in D\times\mathbb{R} : y \geqslant f(x)\}$ is convex.

Furthermore, a function $f: D \to \mathbb{R}$ is convex (concave, ...) iff its restriction to every interval $[a,b] = \{ta+(1-t)b : 0 \leqslant t \leqslant 1\}$ in D is convex (concave, ...). Rolle's theorem implies that if $f: (a,b) \to \mathbb{R}$ is differentiable then f is convex iff f' is increasing and f is concave iff f' is decreasing. In particular, if f is twice differentiable and $f'' \geqslant 0$ then f is convex, while if $f'' \leqslant 0$ then f is concave. Also, if $f'' > 0$ then f is strictly convex and if $f'' < 0$ then f is strictly concave.

The following simple result is often called Jensen's theorem; in spite of its straightforward proof, the result has a great many applications.

Theorem 2. Let $f: D \to \mathbb{R}$ be a concave function. Then

$$\sum_{i=1}^{n} t_i f(x_i) \leqslant f\left(\sum_{i=1}^{n} t_i x_i\right) \tag{3}$$

whenever $x_1, \ldots, x_n \in D$, $t_1, \ldots, t_n \in (0,1)$ and $\sum_{i=1}^{n} t_i = 1$. Furthermore, if f is strictly concave then equality holds in (3) iff $x_1 = \cdots = x_n$.

Proof. Let us apply induction on n. As for $n = 1$ there is nothing to prove and for $n = 2$ the assertions are immediate from the definitions, let us assume that $n \geqslant 3$ and the assertions hold for smaller values of n.

Suppose first that f is concave, and let

$$x_1,\dots,x_n \in D, \qquad t_1,\dots,t_n \in (0,1) \qquad \text{with } \sum_{i=1}^n t_i = 1.$$

For $i = 2,\dots,n$, set $t_i' = t_i/(1-t_1)$, so that $\sum_{i=2}^n t_i' = 1$. Then, by applying the induction hypothesis twice, first for $n-1$ and then for 2, we find that

$$\begin{aligned}\sum_{i=1}^n t_i f(x_i) &= t_1 f(x_1) + (1-t_1)\sum_{i=2}^n t_i' f(x_i)\\ &\leqslant t_1 f(x_1) + (1-t_1) f\left(\sum_{i=2}^n t_i' x_i\right)\\ &\leqslant f\left(t_1x_1 + (1-t_1)\sum_{i=2}^n t_i' x_i\right)\\ &= f\left(\sum_{i=1}^n t_i x_i\right).\end{aligned}$$

If f is strictly concave, $n \geqslant 3$ and not all x_i are equal then we may assume that not all of $x_2,\dots,x_n$ are equal. But then

$$(1-t_1)\sum_{i=2}^n t_i' f(x_i) < (1-t_1) f\left(\sum_{i=2}^n t_i' x_i\right);$$

so the inequality in (3) is strict. □

It is very easy to recover the AM-GM inequality from Jensen's theorem: $\log x$ is a strictly concave function from $(0,\infty)$ to $\mathbb{R}$, so for $a_1,\dots,a_n > 0$ we have

$$\frac{1}{n}\sum_{i=1}^n \log a_i \leqslant \log \sum_{i=1}^n \frac{a_i}{n},$$

which is equivalent to (1). In fact, if $t_1,\dots,t_n > 0$ and $\sum_{i=1}^n t_i = 1$ then

$$\sum_{i=1}^n t_i \log x_i \leqslant \log \sum_{i=1}^n t_i x_i, \tag{4}$$

with equality iff $x_1 = \cdots = x_n$, giving the following extension of Theorem 1.

Theorem 3. Let $a_1,\dots,a_n \geqslant 0$ and $p_1,\dots,p_n > 0$ with $\sum_{i=1}^n p_i = 1$. Then

$$\prod_{i=1}^{n} a_i^{p_i} \leqslant \sum_{i=1}^{n} p_i a_i, \tag{5}$$

with equality iff $a_1 = \cdots = a_n$.

Proof. The assertion is trivial if some a_i is 0; if each a_i is positive, the assertion follows from (4). □

The two sides of (5) can be viewed as two different means of the sequence $a_1, \ldots, a_n$: the left-hand side is a generalized geometric mean and the right-hand side is a generalized arithmetic mean, with the various terms and factors taken with different weights. In fact, it is rather natural to define a further extension of these notions.

Let us fix $p_1, \ldots, p_n > 0$ with $\sum_{i=1}^{n} p_i = 1$: the p_i will play the role of *weights* or *probabilities*. Given a continuous and strictly monotonic function $\varphi : (0, \infty) \to \mathbb{R}$, the *$\varphi$-mean* of a sequence $a = (a_1, \ldots, a_n)$ $(a_i > 0)$ is defined as

$$M_\varphi(a) = \varphi^{-1}\left(\sum_{i=1}^{n} p_i \varphi(a_i)\right).$$

Note that M_φ need not be rearrangement invariant: for a permutation π the φ-mean of a sequence $a_1, \ldots, a_n$ need not equal the φ-mean of the sequence $a_{\pi(1)}, \ldots, a_{\pi(n)}$. Of course, if $p_1 = \cdots = p_n = 1/n$ then every φ-mean is rearrangement invariant.

It is clear that

$$\min_{1 \leqslant i \leqslant n} a_i \leqslant M_\varphi(a) \leqslant \max_{1 \leqslant i \leqslant n} a_i.$$

In particular, the mean of a constant sequence $(a_0, \ldots, a_0)$ is precisely a_0.

For which pairs φ and ψ are the means M_φ and M_ψ comparable? More precisely, for which pairs φ and ψ is it true that $M_\varphi(a) \leqslant M_\psi(a)$ for every sequence $a = (a_1, \ldots, a_n)$ $(a_i > 0)$? It may seem a little surprising that Jensen's theorem enables us to give an exact answer to these questions (see Exercise 31).

Theorem 4. Let $p_1, \ldots, p_n > 0$ be fixed weights with $\sum_{i=1}^{n} p_i = 1$ and let $\varphi, \psi : (0, \infty) \to \mathbb{R}$ be continuous and strictly monotone functions, such that $\varphi\psi^{-1}$ is concave if φ is increasing and convex if φ is decreasing. Then

$$M_\varphi(a) \leqslant M_\psi(a)$$

for every sequence $a = (a_1, \ldots, a_n)$ $(a_i > 0)$. If $\varphi\psi^{-1}$ is strictly concave (respectively, strictly convex) then equality holds iff $a_1 = \cdots = a_n$.

Proof. Suppose that φ is increasing and $\varphi\psi^{-1}$ is concave. Set $b_i = \psi(a_i)$ and note that, by Jensen's theorem,

$$\begin{aligned} M_\varphi(a) = \varphi^{-1}\left(\sum_{i=1}^n p_i\varphi(a_i)\right) &= \varphi^{-1}\left(\sum_{i=1}^n p_i(\varphi\psi^{-1})(b_i)\right) \\ &\leqslant \varphi^{-1}\left((\varphi\psi^{-1})\left\{\sum_{i=1}^n p_i b_i\right\}\right) \\ &= \psi^{-1}\left(\sum_{i=1}^n p_i\psi(a_i)\right) = M_\psi(a). \end{aligned}$$

If $\varphi\psi^{-1}$ is strictly concave and not all a_i are equal then the inequality above is strict since not all b_i are equal.

The case when φ is decreasing and $\varphi\psi^{-1}$ is convex is proved analogously. □

When studying the various means of positive sequences, it is convenient to use the convention that a stands for a sequence $(a_1, \ldots, a_n)$, b for a sequence $(b_1, \ldots, b_n)$ and so on; furthermore,

$$a^{-1} = \frac{1}{a} = (a_1^{-1}, \ldots, a_n^{-1}), \qquad a+x = (a_1+x, \ldots, a_n+x) \qquad (x \in \mathbb{R}^+),$$

$$ab = (a_1b_1, \ldots, a_nb_n), \qquad abc = (a_1b_1c_1, \ldots, a_nb_nc_n),$$

and so on.

If $\varphi(t) = t^r$ $(-\infty < r < \infty,\ r \neq 0)$ then one usually writes M_r for M_φ. For $r > 0$ we define the mean M_r for all non-negative sequences: if $a = (a_1, \ldots, a_n)$ $(a_i \geqslant 0)$ then

$$M_r(a) = \left(\sum_{i=1}^n p_i a_i^r\right)^{1/r}.$$

Note that if $p_1 = \cdots = p_n = 1/n$ then M_1 is the usual *arithmetic mean* A, M_2 is the *quadratic mean* and M_{-1} is the *harmonic mean*. As an immediate consequence of Theorem 4, we shall see that M_r is a continuous monotone increasing function of r.

In fact, $M_r(a)$ has a natural extension from $(-\infty, 0) \cup (0, \infty)$ to the whole of the extended real line $[-\infty, \infty]$ such that $M_r(a)$ is a continuous monotone increasing function. To be precise, put

$$M_\infty(a) = \max_{1\leqslant i\leqslant n} a_i, \qquad M_{-\infty}(a) = \min_{1\leqslant i\leqslant n} a_i, \qquad M_0(a) = \prod_{i=1}^n a_i^{p_i}.$$

Thus $M_0(a)$ is the weighted geometric mean of the a_i. It is easily checked that we have $M_r(a) = \{M_{-r}(a^{-1})\}^{-1}$ for all r $(-\infty \leqslant r \leqslant \infty)$.

Theorem 5. Let $a = (a_1, \ldots, a_n)$ be a sequence of positive numbers, not all equal. Then $M_r(a)$ is a continuous and strictly increasing function of r on the extended real line $-\infty \leqslant r \leqslant \infty$.

Proof. It is clear that $M_r(a)$ is continuous on $(-\infty, 0)\cup(0,\infty)$. To show that it is strictly increasing on this set, let us fix r and s, with $-\infty < r < s < \infty$, $r \neq 0$ and $s \neq 0$. If $0 < r$ then t^r is an increasing function of $t > 0$, and $t^{r/s}$ is a concave function, and if $r < 0$ then t^r is decreasing and $t^{r/s}$ is convex. Hence, by Theorem 4, we have $M_r(a) < M_s(a)$.

Let us write $A(a)$ and $G(a)$ for the *weighted* arithmetic and geometric means of $a = (a_1, \ldots, a_n)$, i.e. set

$$A(a) = M_1(a) = \sum_{i=1}^n p_i a_i \qquad \text{and} \qquad G(a) = M_0(a) = \prod_{i=1}^n a_i^{p_i}.$$

To complete the proof of the theorem, all we have to do is to show that

$$M_\infty(a) = \lim_{r\to\infty} M_r(a), \qquad M_{-\infty}(a) = \lim_{r\to-\infty} M_r(a), \qquad G(a) = \lim_{r\to 0} M_r(a).$$

The proofs of the first two assertions are straightforward. Indeed, let $1 \leqslant m \leqslant n$ be such that $a_m = M_\infty(a)$. Then for $r > 0$ we have

$$M_r(a) \geqslant (p_m a_m^r)^{1/r} = p_m^{1/r} a_m;$$

so $\liminf_{r\to\infty} M_r(a) \geqslant a_m = M_\infty(a)$. Since $M_r(a) \leqslant M_\infty(a)$ for every r, we have $\lim_{r\to\infty} M_r(a) = M_\infty(a)$, as required. Also,

$$M_{-\infty}(a) = \{M_\infty(a^{-1})\}^{-1} = \{\lim_{r\to\infty} M_r(a^{-1})\}^{-1} = \lim_{r\to-\infty} M_r(a).$$

The final assertion, $G(a) = \lim_{r\to 0} M_r(a)$, requires a little care. In keeping with our conventions, for $-\infty < r < \infty$ $(r \neq 0)$ let us write $a^r = (a_1^r, \ldots, a_n^r)$. Then, clearly,

$$M_r(a) = A(a^r)^{1/r}.$$

Also, it is immediate that

$$\lim_{r\to 0}\frac{1}{r}(a_i^r - 1) = \frac{\partial}{\partial r}e^{r\log a_i}\Big|_{r=0} = \log a_i$$

and so

$$\lim_{r\to 0}\frac{1}{r}\{A(a^r) - 1\} = \log G(a). \tag{6}$$

Since

$$\log t \leqslant t - 1$$

for every $t > 0$, if $r > 0$ then

$$\log G(a) = \frac{1}{r}\log G(a^r) \leqslant \frac{1}{r}\log A(a^r) \leqslant \frac{1}{r}\{A(a^r) - 1\}.$$

Letting $r \to 0$, we see from (6) that the right-hand side tends to $\log G(a)$ and so

$$\lim_{r\to 0+}\log M_r(a) = \lim_{r\to 0+}\frac{1}{r}\log A(a^r) = \log G(a),$$

implying

$$\lim_{r\to 0+} M_r(a) = G(a).$$

Finally,

$$\lim_{r\to 0-} M_r(a) = \lim_{r\to 0+}\{M_r(a^{-1})\}^{-1} = G(a^{-1})^{-1} = G(a). \qquad \square$$

The most frequently used inequalities in functional analysis are due to Hölder, Minkowski, Cauchy and Schwarz. Recall that a *hermitian form* on a complex vector space V is a function $\varphi : V\times V \to \mathbb{C}$ such that $\varphi(\lambda x + \mu y, z) = \lambda\varphi(x, z) + \mu\varphi(y, z)$ and $\varphi(y, x) = \overline{\varphi(x, y)}$ for all $x, y, z \in V$ and $\lambda, \mu \in \mathbb{C}$. (Thus $\varphi(x, \lambda y + \mu z) = \bar{\lambda}\varphi(x, y) + \bar{\mu}\varphi(x, z)$.) A hermitian form φ is said to be *positive* if $\varphi(x, x)$ is a positive real number for all $x \in V$ ($x \neq 0$).

Let $\varphi(\cdot,\cdot)$ be a positive hermitian form on a complex vector space V. Then, given $x, y \in V$, the value

$$\varphi(\lambda x + y, \lambda x + y) = |\lambda|^2\varphi(x, x) + 2\,\mathrm{Re}(\lambda\varphi(x, y)) + \varphi(y, y)$$

is real and non-negative for all $\lambda \in \mathbb{C}$. For $x \neq 0$, setting $\lambda = -\overline{\varphi(x, y)}/\varphi(x, x)$, we find that

$$|\varphi(x, y)|^2 \leqslant \varphi(x, x)\varphi(y, y)$$

and the same inequality holds, trivially, for $x = 0$ as well. This is the

Cauchy–Schwarz inequality. In particular, as

$$\varphi(x,y) = \sum_{i=1}^{n} x_i \bar{y}_i$$

is a positive hermitian form on $\mathbb{C}^n$,

$$\left|\sum_{i=1}^{n} x_i \bar{y}_i\right| \leqslant \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2} \left(\sum_{i=1}^{n} |y_i|^2\right)^{1/2},$$

and so

$$\left|\sum_{i=1}^{n} x_i y_i\right| \leqslant \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2} \left(\sum_{i=1}^{n} |\bar{y}_i|^2\right)^{1/2} = \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2} \left(\sum_{i=1}^{n} |y_i|^2\right)^{1/2}. \tag{7}$$

Our next aim is to prove an extension of (7), namely *Hölder's inequality*.

Theorem 6. Suppose

$$p, q > 1 \quad \text{and} \quad \frac{1}{p} + \frac{1}{q} = 1.$$

Then for complex numbers $a_1, \ldots, a_n, b_1, \ldots, b_n$ we have

$$\left|\sum_{k=1}^{n} a_k b_k\right| \leqslant \left(\sum_{k=1}^{n} |a_k|^p\right)^{1/p} \left(\sum_{k=1}^{n} |b_k|^q\right)^{1/q} \tag{8}$$

with equality iff all a_k are 0 or $|b_k|^q = t|a_k|^p$ and $a_k b_k = e^{i\theta}|a_k b_k|$ for all k and some t and θ.

Proof. Given non-negative reals a and b, set $x_1 = a^p$, $x_2 = b^q$, $p_1 = 1/p$ and $p_2 = 1/q$. Then, by Theorem 3,

$$ab = x_1^{p_1} x_2^{p_2} \leqslant p_1 x_1 + p_2 x_2 = \frac{a^p}{p} + \frac{b^q}{q}, \tag{9}$$

with equality iff $a^p = b^q$.

Hölder's inequality is a short step away from here. Indeed, if

$$\left(\sum_{k=1}^{n} |a_k|\right)\left(\sum_{k=1}^{n} |b_k|\right) \neq 0$$

then by homogeneity we may assume that

$$\sum_{k=1}^{n} |a_k|^p = \sum_{k=1}^{n} |b_k|^q = 1.$$

But then, by (9),

$$\left|\sum_{k=1}^{n} a_k b_k\right| \leq \sum_{k=1}^{n} |a_k b_k| \leq \sum_{k=1}^{n}\left(\frac{|a_k|^p}{p}+\frac{|b_k|^q}{q}\right) = \frac{1}{p}+\frac{1}{q} = 1 .$$

Furthermore, if equality holds then

$$|a_k|^p = |b_k|^q \qquad \text{and} \qquad \left|\sum_{k=1}^{n} a_k b_k\right| = \sum_{k=1}^{n} |a_k b_k|,$$

implying $a_k b_k = e^{i\theta}|a_k b_k|$. Conversely, it is immediate that under these conditions we have equality in (8). □

Note that if M_r denotes the rth mean with weights $p_i = n^{-1}$ $(i = 1,\ldots,n)$ and for $a = (a_1,\ldots,a_n)$ and $b = (b_1,\ldots,b_n)$ we put $ab = (a_1b_1,\ldots,a_nb_n)$, $|a| = (|a_1|,\ldots,|a_n|)$ and $|b| = (|b_1|,\ldots,|b_n|)$, then Hölder's inequality states that if $p^{-1}+q^{-1} = 1$ with $p,q > 1$, then

$$M_1(|ab|) \leq M_p(|a|)M_q(|b|).$$

A minor change in the second half of the proof implies that (8) can be extended to an inequality concerning the means M_1, M_p and M_q with arbitrary weights (see Exercise 8).

The numbers p and q appearing in Hölder's inequality are said to be *conjugate exponents* (or *conjugate indices*). It is worth remembering that the condition $p^{-1}+q^{-1} = 1$ is the same as

$$(p-1)(q-1) = 1, \qquad (p-1)q = p \qquad \text{or} \qquad (q-1)p = q.$$

Note that 2 is the only exponent which is its own conjugate. As we remarked earlier, the special case $p = q = 2$ of Hölder's inequality is called the Cauchy–Schwarz inequality.

In fact, one calls 1 and ∞ *conjugate exponents* as well. Hölder's inequality is essentially trivial for the pair $(1,\infty)$:

$$M_1(|ab|) \leq M_1(|a|)M_\infty(|b|),$$

with equality iff there is a θ such that $|b_k| = M_\infty(|b|)$ and $a_k b_k = e^{i\theta}|a_k b_k|$ whenever $a_k \neq 0$.

The next result, *Minkowski's inequality*, is also of fundamental importance: in chapter 2 we shall use it to define the classical l_p spaces.

Theorem 7. Suppose $1 \leq p < \infty$ and $a_1,\ldots,a_n,b_1,\ldots,b_n$ are complex numbers. Then

$$\left(\sum_{k=1}^{n} |a_k+b_k|^p\right)^{1/p} \leq \left(\sum_{k=1}^{n} |a_k|^p\right)^{1/p} + \left(\sum_{k=1}^{n} |b_k|^p\right)^{1/p} \tag{10}$$