Convex functions

This handout contains a fairly broad overview of matters regarding convex functions. It contains a lot of optional material in a series of Remarks: the mandatory part is formulas (1-3), the second (easy, as it follows directly from (3)) bullet of Theorem 1, Theorems 2,3, and the inequalities of Jensen, Cauchy and Cauchy-Schwartz.

Definitions and representations of convexity condition.

`Definition 1:` A function $f:\mathbb{R}^n\to\mathbb{R}$ is called convex if for any pair of non-equal $\boldsymbol{x}_1,\boldsymbol{x}_2$ in the domain of f (which is assumed to be a closed convex set) and any pair of real positive numbers θ_1,θ_2, such that $\theta_1+\theta_2=1$, one has

$$f(\theta_1\boldsymbol{x}_1+\theta_2\boldsymbol{x}_2)\le\theta_1 f(\boldsymbol{x}_1)+\theta_2 f(\boldsymbol{x}_2). \tag{1}$$

If the inequality is always strict, f is called strictly convex; f is called (strictly) concave iff $-f$ is (strictly) convex.

Geometrically, it means that on the graph $\{(\boldsymbol{x},y)\in\mathbb{R}^{n+1}: y=f(\boldsymbol{x})\}$ of f, for any $\boldsymbol{x}$ lying on the line segment, connecting a pair of chosen points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ in the domain of f, the point $(\boldsymbol{x},f(\boldsymbol{x}))$ lies below a chord, connecting the pair of points $(\boldsymbol{x}_1,f(\boldsymbol{x}_1))$ and $(\boldsymbol{x}_2,f(\boldsymbol{x}_2))$, for all the possible choices of the pair $\boldsymbol{x}_1,\boldsymbol{x}_2$ (the height being measured in terms of the y-coordinate). This makes the majority of the convexity issues essentially one-dimensional: for instance $f(\boldsymbol{x})$ is convex if and only if for any chosen pair of points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ in the domain of f, the function $\tilde f(t)=f(t\boldsymbol{x}_1+(1-t)\boldsymbol{x}_2)$ of one variable $t\in\mathbb{R}$ is convex.

Unless specified, let's deal with the one-dimensional case, assuming that f is defined and bounded on some closed interval $[a,b]$. If in (1) one sets $x_2>x_1$ and $x=\theta_1x_1+\theta_2x_2$, then using $\theta_2=1-\theta_1$, (1) can be rewritten as

$$f(x)\le\frac{x_2-x}{x_2-x_1}f(x_1)+\frac{x-x_1}{x_2-x_1}f(x_2)\equiv l(x). \tag{2}$$

Namely, the right hand side is a linear function $l(x)$ of x (given $x_1,x_2,f(x_1),f(x_2)$), the geometric description above thus having been made precise.

Remark – optional: Note that (2) can be further rewritten as

$$(x_2-x)f(x_1)-(x_2-x_1)f(x)+(x-x_1)f(x_2)=\begin{vmatrix}1&1&1\\x_1&x&x_2\\f(x_1)&f(x)&f(x_2)\end{vmatrix}\ge0.$$

The latter determinant is twice the signed area of the triangle with vertices $A_1=(x_1,f(x_1))$, $A=(x,f(x))$, $A_2=(x_2,f(x_2))$, which is positive iff going around the triangle A_1AA_2 along the itinerary $A_1\to A\to A_2\to A_1$ occurs counterclockwise.

One proceeds further by multiplying $f(x)$ in the left hand side of (2) by $1=\dfrac{(x_2-x)+(x-x_1)}{x_2-x_1}$, which results in

$$\frac{f(x)-f(x_1)}{x-x_1}\le\frac{f(x_2)-f(x)}{x_2-x}. \tag{3}$$

Analysis of this representation results in the following Theorem 1.

Namely, assuming differentiability of f, letting $x\to x_1^+$ (x approaches x_1 on the right) one gets $f'(x_1)\le\dfrac{f(x_2)-f(x_1)}{x_2-x_1}$; letting $x\to x_2^-$ (x approaches x_2 on the left), one gets $f'(x_2)\ge\dfrac{f(x_2)-f(x_1)}{x_2-x_1}$. Thus $f'(x_2)\ge f'(x_1)$, which implies that the derivative $f'(x)$ is a non-decreasing function of x, and hence the second derivative $f''(x)$, if it exists, is *non-negative.*

Conversely, if one assumes differentiability of $f(x)$ and the fact that $f'(x)$ is a non-decreasing function of x, for any three points $x_1<x<x_2$ and some $\xi_1\in(x_1,x)$ and $\xi_2\in(x,x_2)$, by the Mean value theorem one has $\dfrac{f(x)-f(x_1)}{x-x_1}=f'(\xi_1)$, $\dfrac{f(x_2)-f(x)}{x_2-x}=f'(\xi_2)$. Then (3) follows, as it has been assumed that $f'(\xi_2)\ge f'(\xi_1)$, so the function f is convex.

Remark –optional: The characterisation (3) can be further dug into as follows. For any strictly increasing sequence $\{x_i\}_{i\ge1}$ of points within the interval (a,b), the ratios $\dfrac{f(x_i)-f(x_{i-1})}{x_i-x_{i-1}}$ for $i\ge2$ are non-decreasing and bounded from above, as long as all $x_i < b - \epsilon$ for any $\epsilon > 0$. Thus, at every point $x \in (a, b)$ there exists the left derivative $f'_-(x) = \lim_{h \to 0_+} \dfrac{f(x-h) - f(x)}{h} < \infty$ (although one can have $\lim_{x \to b_-} f'_-(x) = \infty$, e.g. take a "lower semicircle" $f(x) = -\sqrt{1-x^2}$ on $[-1, 1]$). In the same fashion, there exists the right derivative $f'_+(x) = \lim_{h \to 0_+} \dfrac{f(x+h) - f(x)}{h} > -\infty$ for any $x \in (a, b)$. Then by (3), $f'_+(x) \geq f'_-(x)$. Thus, as $f'_+(x)$ is finite, *any* convex function $f(x)$, defined on a closed interval $[a, b]$ must be *continuous* in its interior (a, b) (but not necessarily at the end points a, b themselves: take again $f(x) = -\sqrt{1-x^2}$, but only for $x \in (-1, 1)$, and define $f(-1) = f(1) = 100$.) Moreover, there can be no more than a countable number of points $x \in (a, b)$, where $f'_-(x) < f'_+(x)$, namely where the first derivative has a jump discontinuity.

`Theorem 1:`

- Any convex function $f(x)$ on $[a, b]$ is continuous on (a, b) and has a finite right derivative $f'_+(x)$ and a left derivative $f'_-(x)$ at each point $x \in (a, b)$. Moreover, for all $x \in (a, b)$, $f'_+(x) \geq f'_-(x)$, the equality occurring and yielding the derivative $f'(x)$ everywhere, except possibly a countable number of points inside (a, b). Wherever it exists, $f'(x)$ is a non-decreasing function of x. If $f(x)$ is strictly convex, $f'(x)$ is strictly increasing. If a convex $f(x)$ has the second derivative at $x \in (a, b)$, then $f'(x) \geq 0$.
- If a differentiable function $f(x)$ on $[a, b]$ has a non-decreasing (increasing) derivative $f'(x)$ everywhere on $[a, b]$, then $f(x)$ is (strictly) convex. If $f(x)$ has a (positive, except maybe a finite number of points, where it's zero) non-negative second derivative $f''(x)$ everywhere on (a, b), then $f(x)$ is (strictly) convex on $[a, b]$.

`Remark:` The second bullet is what one should use to check whether a given function is (strictly) convex. For functions of many variables $f(\boldsymbol{x})$, one should look at the Hessian $D^2 f(\boldsymbol{x})$ and see whether it is non-negative (positive) definite. Indeed, assuming the existence and continuity of $D^2 f(\boldsymbol{x})$, the (strict) convexity will follow if the derivative $\nabla f(\boldsymbol{x}) \cdot \boldsymbol{d}$ in any direction $\boldsymbol{d} \in \mathbb{R}^n$ is non-decreasing (increasing), which would be guaranteed if for all $\boldsymbol{x}$ in the domain of f, $\boldsymbol{d} \cdot D^2 f(\boldsymbol{x}) \boldsymbol{d} \geq 0$ ($\boldsymbol{d} \cdot D^2 f(\boldsymbol{x}) \boldsymbol{d} > 0$) except maybe a finite number of points, where it's zero).

Some easy properties of convex functions:

1. If f is convex and c is a constant, then the function cf is convex.
2. If a pair of functions f and g are convex, then the function $f + g$ is convex.
3. Optional: If a function $f(u)$ is convex and increasing, and a function $u(x)$ is convex, then the composition $f[u(x)]$ is convex.

   `Proof:` $f[u(\theta_1 x_1 + \theta_2 x_2)] \leq f[\theta_1 u(x_1) + \theta_2 u(x_2)] \leq \theta_1 f[u(x_1)] + \theta_2 f[u(x_2)]$.
4. Optional: If $f(x)$ is convex on $[a, b]$ and is not a constant, then it cannot have a local maximum inside (a, b).

   `Proof:` Suppose $x_0 \in (a, b)$ is a local maximum. Then there exists a pair of points $x_1, x_2 \in (a, b)$ such that $x_1 < x_0$, $x_2 > x_0$, thus $x_0 = \theta_1 x_1 + \theta_2 x_2$ for some $\theta_1, \theta_2 > 0$, $\theta_1 + \theta_2 = 1$. Moreover, $f(x_1) \leq f(x_0)$, $f(x_2) \leq f(x_0)$, with at least one of the two inequalities being strict. Multiply the first inequality by θ_1 and the second one by θ_2, then add them, getting $f(x_0) > \theta_1 f(x_1) + \theta_2 f(x_2)$. Contradiction with the convexity of f!
5. Optional: Given a function $f(x)$, convex on $x \in [a, b]$, fix a pair of points $x_1, x_2 \in [a, b]$ with $x_2 > x_1$. Then the defining inequality (1), for the above x_1 and x_2 and for all $x \in (x_1, x_2)$ is either always an equality or is always strict.

   `Proof:` Consider a function $g(x) = f(x) - l(x) = f(x) + (-l(x))$ for $x \in [x_1, x_2]$, where $l(x)$ is a chord, defined by the right hand side of (2). Then $g(x)$ is convex as a sum of two convex functions; in addition $g(x_1) = g(x_2) = 0$. Then, according to the previous item, $g(x)$ is either a constant or it cannot have a local minimum for $x \in (x_1, x_2)$, which corresponds to the two alternatives in question.

Finally, there is an important for the case of several variables characterisation of convex functions.

`Theorem 2:` A function $f(\boldsymbol{x})$ is convex if and only if the sublevel set $F_c \equiv \{\boldsymbol{x} : f(\boldsymbol{x}) \leq c\}$ for any $c \in \mathbb{R}$ is convex.

`Proof -- Necessity:` If f is convex, then one should prove that if $\boldsymbol{x}_1, \boldsymbol{x}_2 \in F_c$ for some c, then their convex combination $\theta_1 \boldsymbol{x}_1 + \theta_2 \boldsymbol{x}_2$ is also in F_c. Indeed, $f(\theta_1 \boldsymbol{x}_1 + \theta_2 \boldsymbol{x}_2) \leq \theta_1 f(\boldsymbol{x}_1) + \theta_2 f(\boldsymbol{x}_2) \leq (\theta_1 + \theta_2) c = c$.

`Proof -- Sufficiency, optional):` Take two points $\boldsymbol{x}_1, \boldsymbol{x}_2$ in the domain of f. First, suppose, $f(x_1) = f(x_2) = c$. Then, as F_c is convex, a convex combination $\boldsymbol{x} = \theta_1\boldsymbol{x}_1 + \theta_2\boldsymbol{x}_2 \in F_c$, and then by definition of F_c, one has $f(\theta_1\boldsymbol{x}_1 + \theta_2\boldsymbol{x}_2) \leq c = \theta_1 f(\boldsymbol{x}_1) + \theta_2 f(\boldsymbol{x}_2)$.

If $f(\boldsymbol{x}_1) \neq f(\boldsymbol{x}_2)$, then instead of f consider a function $g(\boldsymbol{x})$, obtained by subtracting from f a linear function $l(\boldsymbol{x})$, such that $l(\boldsymbol{x}_1) = f(\boldsymbol{x}_1)$ and $l(\boldsymbol{x}_2) = f(\boldsymbol{x}_2)$, analogous to the one, defined by the right hand side of the equation (2). It's easy to check that the sublevel sets of the resulting function g are also going to be convex, and the points $\boldsymbol{x}_1, \boldsymbol{x}_2$ will now belong to the same sublevel set of g, corresponding to $c = 0$. Therefore, as has been shown in the preceding passage, one should have $g(\theta_1\boldsymbol{x}_1 + \theta_2\boldsymbol{x}_2) \leq 0$, which means $f(\theta_1\boldsymbol{x}_1 + \theta_2\boldsymbol{x}_2) \leq \theta_1 f(\boldsymbol{x}_1) + \theta_2 f(\boldsymbol{x}_2)$, by definition of g. But the latter statement means that f is convex.

Convex functions and optimisation

Convex functions are important for (non-linear) optimisation[1], because they make it easy and essentially similar to linear programming. Consider the optimisation problem (OP) Min $f(\boldsymbol{x})$, such that $\boldsymbol{x} \in F$ for some feasible set F. Clearly, a Min problem for f is a Max problem for $-f$.

`Definition 2:` A point $\boldsymbol{x} = \boldsymbol{x}_0 \in F$ is a local (strict) minimizer for the above OP if there exists an open neighbourhood $\mathcal{U}$ of $\boldsymbol{x}_0$ such that $\forall\, \boldsymbol{x} \in \mathcal{U} \cap F,\ f(\boldsymbol{x}) \geq (>)\, f(\boldsymbol{x}_0)$.
A point $\boldsymbol{x} = \boldsymbol{x}_0 \in F$ is a global (strict) minimizer if the above inequalities hold for all $\boldsymbol{x} \in F$.

In the same way one can define a local or global (strict) maximizer and a local or global (strict) *extremizer* which is either a local or global (strict) minimizer or a maximizer.

`Theorem 3:` If the objective function $f(\boldsymbol{x})$ is (strictly) convex and the set F is convex, then a local (strict) minimizer is a global one. If $f(\boldsymbol{x})$ is (strictly) concave and F is convex, then a local (strict) maximizer is a global one.

`Proof:` Prove it for the (strict) minimizer, for a maximizer it should be modified in an obvious way. Suppose the contrary, namely $\boldsymbol{x}_0$ is a local (strict) minimizer, but there exists some $\boldsymbol{x}_1 \in F$, such that $f(\boldsymbol{x}_1) < (\leq)\, f(\boldsymbol{x}_0)$. The line segment $\boldsymbol{x}_0\boldsymbol{x}_1$ is contained in F, as the latter set is convex. Then by convexity of the function f, for any $\boldsymbol{x}$ inside the line segment $\boldsymbol{x}_0\boldsymbol{x}_1$, which will be a convex combination of $\boldsymbol{x}_0$ and $\boldsymbol{x}_1$ with the coefficients θ_0 and θ_1, one has $f(\boldsymbol{x}) \leq \theta_0 f(\boldsymbol{x}_0) + \theta_1 f(\boldsymbol{x}_1) < (\leq)\, f(\boldsymbol{x}_0)$. This contradicts the fact that $\boldsymbol{x}_0$ is a (strict) local minimizer, as $\boldsymbol{x}$ can get arbitrarily close to $\boldsymbol{x}_0$, thus entering any neighbourhood $\mathcal{U}$ of $\boldsymbol{x}_0$ and violating Definition 2.

Combining this result with the preceding theorem on the convexity of sublevel sets of convex functions, suppose the feasible set F for the OP is described in terms of the inequalities $g_i(\boldsymbol{x}) \leq b_i$, for $i = 1, \ldots, m$ and possibly the sign constraints $\boldsymbol{x} \geq 0$. Suppose, all the functions g_i are convex. Then the feasible set F for such an OP is convex, and Theorem 3 applies. Hence, to find the global minimizer (maximizer), one should only find a local one. This can be done using the methods for local extrema.

Generalisation of convexity definition

The purpose is to generalise the formula (1) for a convex function $f(\boldsymbol{x})$ to

$$f\left(\sum_{i=1}^{s} \theta_i \boldsymbol{x}_i\right) \leq \sum_{i=1}^{s} \theta_i f(\boldsymbol{x}_i), \quad \text{where} \ \ \theta_i \geq 0, \forall\, i = 1, \ldots, s \ \ \text{and} \ \ \sum_{i=1}^{s} \theta_i = 1. \tag{4}$$

This is easily done by induction on s as follows. For $s = 2$ it's true. Assume, it is true for some $s - 1 \geq 2$. Then if $\theta_s < 1$ (otherwise, there's nothing to prove if $\theta_s = 1$)

$$f\left(\sum_{i=1}^{s} \theta_i \boldsymbol{x}_i\right) = f\left(\theta_s \boldsymbol{x}_s + (1-\theta_s)\sum_{i=1}^{s-1} \frac{\theta_i}{1-\theta_s}\boldsymbol{x}_i\right) \leq \theta_s f(\boldsymbol{x}_s) + (1-\theta_s) f\left(\sum_{i=1}^{s-1} \frac{\theta_i}{1-\theta_s}\boldsymbol{x}_i\right),$$

by definition (1). To the second term, one can apply the induction assumption, as

$$\sum_{i=1}^{s-1} \frac{\theta_i}{1-\theta_s} = \frac{1}{1-\theta_s}\sum_{i=1}^{s-1} \theta_i = \frac{1-\theta_s}{1-\theta_s} = 1.$$

[1]For linear optimisation, all the functions involved are linear, hence convex.

Hence,

$$(1-\theta_s) f\left(\sum_{i=1}^{s-1} \frac{\theta_i}{1-\theta_s} \boldsymbol{x}_i\right) \leq (1-\theta_s) \sum_{i=1}^{s-1} \frac{\theta_i}{1-\theta_s} f(\boldsymbol{x}_i) = \sum_{i=1}^{s-1} \theta_i f(\boldsymbol{x}_i),$$

which proves (4).

Remark: The formula (4) reads geometrically as follows: given s points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_s \in \mathbb{R}^n$ in the domain of f, for $\boldsymbol{x}$ is in the convex hull of the points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_s \in \mathbb{R}^n$, the point $(\boldsymbol{x}, f(\boldsymbol{x})) \in \mathbb{R}^{n+1}$ on the graph of a convex function $f(\boldsymbol{x})$, lies below the *convex hull* of s points $(\boldsymbol{x}_1, f(\boldsymbol{x}_1)), \ldots (\boldsymbol{x}_s, f(\boldsymbol{x}_s)) \in \mathbb{R}^{n+1}$.

The formula (4) is often rewritten as follows: to dispense with the requirement $\sum \theta_i = 1$ (in the sequel let's skip the range if i in the sums, unless it is necessary), one sets $\theta_i = \dfrac{p_i}{\sum p_i}$ for some collection of positive numbers p_i. Clearly, this guarantees $\sum \theta_i = 1$. Then (4) can be rewritten as *Jensen's inequality*:

$$f\left(\frac{\sum p_i \boldsymbol{x}_i}{\sum p_i}\right) \leq \frac{\sum p_i f(\boldsymbol{x}_i)}{\sum p_i}, \tag{5}$$

for a convex f.

Remark: This inequality is non-trivial if and only if there is a pair of two different points amongst $\boldsymbol{x}_i$'s. Otherwise it is a trivial statement $f(\boldsymbol{x}) = f(\boldsymbol{x})$. In the former non-trivial case, the inequality is strict for a strictly convex f. It gets reversed for a concave f.

Classical inequalities

Jensen's inequality (5) in one dimension enables one to prove a variety of useful inequalities, simply by applying it to some concrete convex or concave functions, such as powers, exponentials and logarithms. The wealth of such inequalities comes form the fact that (5) is true for *any* convex f and *any* arrays of $\boldsymbol{x}$'s and $p's$ (of the same size). Besides, sums can be substituted by integrals.

For instance, let $x_1 = a$, $x_s = b$ and take the rest of x_i's ordered within the interval $[a, b]$ with $x_{i+1} > x_i$. Let $p_i = \Delta x_i = x_{i+1} - x_i$ for $1 < i \leq s-1$, $p_s = 0$, with $\sum p_i = b - a > 0$. Then as the number of terms $s \to \infty$, in the numerator of the right-hand side of (5) one simply has the integral $\displaystyle\int_a^b f(x)dx$, while in the left-hand side one gets $f\left(\dfrac{\sum x_i \Delta x_i}{b-a}\right)$. Thus, as $\sum x_i \Delta x_i \to \displaystyle\int_a^b x dx = \frac{b^2 - a^2}{2}$, one gets

$$f\left(\frac{b+a}{2}\right) \leq \frac{\int_a^b f(x)dx}{b-a}. \tag{6}$$

In the same fashion, suppose, $g(x)$ is an integrable function of x. Then one can use (5) in the same way as above by letting $p_i = \Delta x_i$, but rather having $g(x_i)$ instead of x_i. Then

$$f\left(\frac{\int_a^b g(x)dx}{b-a}\right) \leq \frac{\int_a^b f[g(x)]dx}{b-a}. \tag{7}$$

Furthermore, let $f(x) = \ln x$, $x > 0$. Its second derivative is strictly negative for $x > 0$, so the function is concave, hence the converse of (5) must hold for it, namely

$$\ln\left(\frac{1}{\sum p_i} \sum p_i x_i\right) \geq \frac{\sum p_i \ln x_i}{\sum p_i}.$$

Putting everything under the log sign on both sides yields *Young's inequality*: for all $x_i > 0$,

$$\left(\prod x_i^{p_i}\right)^{\frac{1}{\sum p_i}} \leq \frac{\sum p_i x_i}{\sum p_i}. \tag{8}$$

Letting all $p_i = 1$, using n for the number of terms, yields

$$\sqrt[n]{\prod_{i=1}^{n} x_i} \leq \frac{\sum_{i=1}^{n} x_i}{n}. \tag{9}$$

Namely, the *geometric* mean of a finite collection of positive numbers is always less than the *arithmetic* mean, unless all the numbers are equal, when the two means coincide. This inequality bears the name of Cauchy.

Replacing in (9) x_i with $\frac{1}{x_i}$ yields

$$\sqrt[n]{\prod_{i=1}^{n} x_i} \geq \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}. \tag{10}$$

Namely, the geometric mean of a finite collection of positive numbers is always greater than their *harmonic* mean, unless all the numbers are equal when the two means coincide.

Applying Jensen's inequality to a convex function $f(x) = x^k$, $x > 0$, $k > 1$ gives

$$\left(\sum p_i x_i\right)^k \leq \left(\sum p_i x_i^k\right)\left(\sum p_i\right)^{k-1}. \tag{11}$$

Let $k = 2$. For arrays of numbers a_i, b_i Denote $p_i = b_i^2$, $x_i = \frac{a_i}{b_i}$ and substitute in the above. This yields the well-known Cauchy-Schwartz (et. al.) inequality:

$$\sum a_i b_i \leq \sqrt{\sum a_i^2}\sqrt{\sum b_i^2}. \tag{12}$$

The rest is optional – the inequalities I require you to know are Jensen, Cauchy, and Cauchy-Schwartz:

Remark: If one denotes k' a *conjugate exponent* to k by the rule $k' = \frac{k}{k-1}$ or $\frac{1}{k} + \frac{1}{k'} = 1$ (if $k = 2$, then $k' = 2$) and then in (11) sets $p_i = b_i^{k'}$, $x_i = \frac{a_i}{b_i^{k-1}}$ for some positive numbers a_i, b_i, the previous inequality (11) turns into the Hölder inequality:

$$\sum a_i b_i \leq \left(\sum a_i^k\right)^{\frac{1}{k}}\left(\sum b_i^{k'}\right)^{\frac{1}{k'}}. \tag{13}$$

Both Cauchy-Schwartz and Hölder inequalities are quite useful in a particular case when all $b_i = 1$, which implies, in the case of Cauchy-Schwartz that

$$\left(\sum_{i=1}^{n} a_i\right)^2 \leq n\sum_{i=1}^{n} a_i^2.$$

Remark: The Hölder inequality (13) is often derived from the Young inequality (8) for two numbers. In the latter inequality one first lets $p_1 = p$, $p_2 = 1 - p$ as well as $\alpha = x_1^{p_1}$, $\beta = x_2^{p_2}$, thus reducing it to $\alpha\beta \leq p\alpha^{\frac{1}{p}} + p'\beta^{\frac{1}{p'}}$. Then one lets $k = \frac{1}{p}$, $k' = \frac{1}{p'}$, and they are conjugate exponents. Thus

$$\alpha\beta \leq \frac{1}{k}\alpha^k + \frac{1}{k'}\beta^{k'}.$$

Finally, one takes two arrays a_i, b_i of positive numbers and applies the last inequality to each pair

$$\alpha_i = \frac{a_i}{\left(\sum a_i^k\right)^{\frac{1}{k}}}, \quad \beta_i = \frac{b_i}{\left(\sum b_i^{k'}\right)^{\frac{1}{k'}}}.$$

Summing it for all the pairs α_i, β_i yields simply

$$\frac{\sum a_i b_i}{\left(\sum a_i^k\right)^{\frac{1}{k}}\left(\sum b_i^{k'}\right)^{\frac{1}{k'}}} \leq \frac{1}{k} + \frac{1}{k'} = 1,$$

which is equivalent to (13).

Besides, the Hölder inequality (13) is used to produce another important inequality, bearing the name of Minkowski and indicating that for a vector $\boldsymbol{a} \in \mathbb{R}^n$ with components a_i, the expression

$$\|\boldsymbol{a}\|_k = \left(\sum a_i^k\right)^{\frac{1}{k}},$$

for any $k > 1$ is a norm, equivalent to the Euclidean one, corresponding to the case $k = 2$. The Minkowski inequality is the triangle inequality for the norm $\|\cdot\|_k$, namely

$$\|\boldsymbol{a} + \boldsymbol{b}\|_k \leq \|\boldsymbol{a}\|_k + \|\boldsymbol{b}\|_k \quad \text{or} \quad \left(\sum_{i=1}^n (a_i + b_i)^k\right)^{\frac{1}{k}} \leq \left(\sum_{i=1}^n a_i^k\right)^{\frac{1}{k}} + \left(\sum_{i=1}^n b_i^k\right)^{\frac{1}{k}}. \tag{14}$$

Indeed,

$$\sum_{i=1}^n (a_i + b_i)^k = \sum_{i=1}^n a_i (a_i + b_i)^{k-1} + \sum_{i=1}^n b_i (a_i + b_i)^{k-1}.$$

Applying the Hölder inequality (13) to the first term with a_i for a_i and $(a_i + b_i)^{k-1}$ for b_i one gets

$$\sum_{i=1}^n a_i (a_i + b_i)^{k-1} \leq \left(\sum_{i=1}^n a_i^k\right)^{\frac{1}{k}} \left(\sum_{i=1}^n (a_i + b_i)^k\right)^{\frac{k-1}{k}}.$$

In the same fashion, applying the Hölder inequality (13) to the second term with b_i for a_i and $(a_i + b_i)^{k-1}$ for b_i yields

$$\sum_{i=1}^n b_i (a_i + b_i)^{k-1} \leq \left(\sum_{i=1}^n b_i^k\right)^{\frac{1}{k}} \left(\sum_{i=1}^n (a_i + b_i)^k\right)^{\frac{k-1}{k}}.$$

Then, adding the two together and dividing by $\left(\sum_{i=1}^n (a_i + b_i)^k\right)^{\frac{k-1}{k}}$ yields the desired result (14).