The AM-GM inequality from different viewpoints

Elem. Math. 72 (2017) 24 – 34
0013-6018/17/010024-11
DOI 10.4171/EM/320

Elemente der Mathematik

The AM-GM inequality from different viewpoints

Darko Veljan

Darko Veljan is professor (now retired) of mathematics at the University of Zagreb, where he received his B.Sc. and M.Sc. He obtained his Ph.D. at Cornell University. He is the author of several university textbooks. His interests include topology, geometry, combinatorics, mathematical education and history of mathematics.

1 Introduction

The famous Russian mathematician Andrei N. Kolmogorov (1903–1987) once said: "Every serious proof in mathematics eventually boils down to proving an inequality".

One of the most common and useful basic "folklore" inequalities is the *arithmetic mean-geometric mean inequality*, for short the AM-GM *inequality*: $A \geq G$, where $A = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the *arithmetic mean* (average, commonly denoted by $\overline{x}$) and $G = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}$ the *geometric mean* of real numbers $x_1, x_2, \ldots, x_n \geq 0$. Equality occurs if and only if $x_1 = \cdots = x_n$. The rth *power mean* $M_r(x)$ of a vector $x = (x_1, \ldots, x_n) \in \mathbf{R}_+^n$ (all $x_i \geq 0$) is defined by

$$M_r := M_r(x) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^r\right)^{1/r} \qquad \text{for all } r \in \mathbf{R} \cup \{\pm\infty\}.$$

Die Ungleichung vom arithmetischen und geometrischen Mittel gehört zu den grundlegendsten Abschätzungen in der Mathematik. Für zwei Variablen war sie bereits Euklid bekannt, ein Beweis für beliebig viele Variablen findet sich erstmals 1729 in einer Arbeit des schottischen Mathematikers Colin Maclaurin. Auch Cauchy widmet sich in seinem Werk *Analyse algébrique* von 1821 dieser Ungleichung. So sind im Laufe der Geschichte zahlreiche algebraische, geometrische, topologische und kombinatorische Beweise zusammengekommen, welche oftmals anschauliche geometrische oder auch physikalische Interpretationen zulassen. Die Anwendungen und Verallgemeinerungen sind unübersehbar und allgegenwärtig im mathematischen Tagesgeschäft. Der Autor der vorliegenden Arbeit gibt einen Überblick, der bis hin zum arithmetisch-geometrischen Mittel reicht und die Betrachtung neuer gemischter Mittel anregt.

M_1 is the arithmetic mean A, $M_0 (= \lim_{r \to 0} M_r)$ is the geometric mean G, while M_{-1} is the *harmonic mean*, M_2 the *quadratic mean*, $M_{-\infty} = \min\{x_i\}$, $M_\infty (= \lim_{r\to\infty} M_r) = \max\{x_i\}$ etc.

The weighted version is given by

$$M_r(x) = \left(\sum_{i=1}^{n} w_i x_i^r \right)^{1/r},$$

where

$$w = (w_1, \ldots, w_n),\ w_1, \ldots, w_n \geq 0 \quad \text{and} \quad \sum_{i=1}^{n} w_i = 1.$$

There are two important inequalities for (weighted) power means. The first is the *increasing property* (*or monotonicity*): $p \leq q \Rightarrow M_p(x) \leq M_q(x)$ and the second is the *product property*: $M_r(x)M_r(y) \leq M_r(xy)$ for all r, where $xy = (x_1y_1, \ldots, x_ny_n)$ is the (component-wise) product of vectors x and y. In the generic case $p = 0, q = 1$ the increasing property is just the AM-GM inequality, while the case $r = 1$ of the product property is the *Chebyshev inequality* (from 1860): if $x_1 \leq \cdots \leq x_n$ and $y_1 \leq \cdots \leq y_n$ then $A(x)A(y) \leq A(xy)$.

The AM-GM inequality for two numbers was probably known to Pythagoras (about 500 B.C.) and for sure to Euclid (about 300 B.C.). The general AM-GM inequality for any n was probably known to Fermat, Descartes, maybe Galileo and others around 1630, but definitely to Newton about 1705. The first rigorous proof appeared about 1725 by Maclaurin.

Two classical books on inequalities are [1] and [2]. In modern theory, general means are defined quite abstractly in terms of metric (or topological) space with some natural properties (see, e.g., [3]). The *mean* of any list of points (data) in any set of points can be thought of as the point (or more points) "closest" to the list in a given, prescribed sense. For example, the *Fréchet mean* (introduced about 1938) of points $x_1, \ldots, x_N$ on a Riemannian manifold (M, d) is a point $p \in M$ (if exists) such that $\sum_i d^2(p, x_i)$ has minimal value.

2 Standard and less standard proofs

The most common textbook proofs of the AM-GM inequality are by induction or by Jensen's functional inequality $f\left(\frac{x+y}{2}\right) \leq \frac{1}{2}(f(x) + f(y))$ which verbally can be phrased as "the value at the average is not greater than the average of the values". It is just the convexity of the function f. (In fact, Jensen in his paper from 1906 used concavity of the function ln on positive reals.)

The following induction proof of the AM-GM inequality is well known since 1970; it is short and instructive. Here it is. For $n = 1$ it is trivial. Suppose it holds for $n - 1$ and let $x_1, \ldots, x_n \geq 0$ are given. Let A and G be their arithmetic and geometric means, respectively. We may assume that $x_1 \leq x_2 \leq \cdots \leq x_n$. Then clearly $x_1 \leq A \leq x_n$. By induction on $n - 1$ numbers $x_2, x_3, \ldots, x_{n-1}, x_1 + x_n - A$ we have

$$\left(\frac{x_2 + x_3 + \cdots + x_{n-1} + (x_1 + x_n - A)}{n-1} \right)^{n-1} \geq x_2 x_3 \cdots x_{n-1}(x_1 + x_n - A).$$

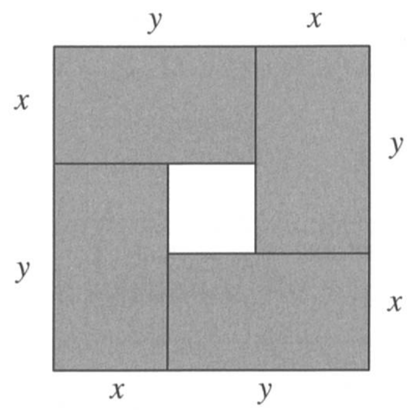

Fig. 1 The square on $x+y$ contains four rectangles with x and y, so for areas we have: $(x+y)^2 \geq 4xy \Rightarrow \frac{x+y}{2} \geq \sqrt{xy}$.

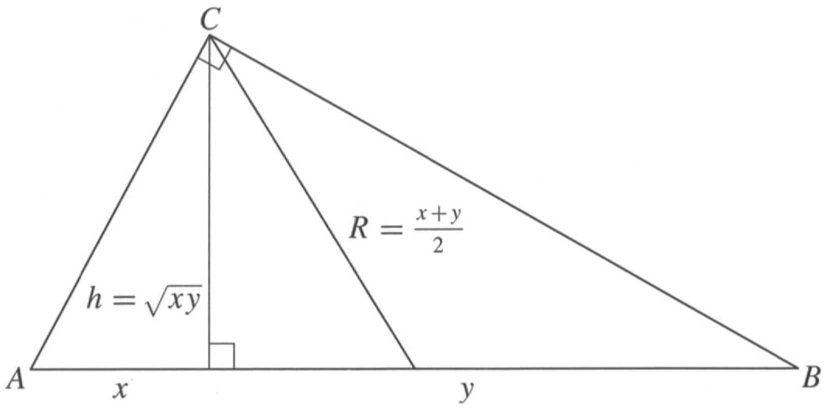

Fig. 2 In the right triangle ABC, the circumradius is $R = \frac{x+y}{2}$ and the height is $h = \sqrt{xy}$; $R \geq h \Rightarrow \frac{x+y}{2} \geq \sqrt{xy}$.

Since $x_1 + x_2 + \cdots + x_n = nA$, it follows that $A^{n-1} \geq x_2 x_3 \cdots x_{n-1}(x_1 + x_n - A)$. From $A - x_1 \geq 0$ and $x_n - A \geq 0$, we get $(A - x_1)(x_n - A) \geq 0$, hence $A(x_1 + x_n - A) \geq x_1 x_n$. By multiplying the above inequality by A we obtain

$$A^n \geq x_2 x_3 \cdots x_{n-1}[A(x_1 + x_n - A)] \geq x_2 x_3 \cdots x_{n-1} x_1 x_n = G^n.$$

Therefore, $A \geq G$. The equality case is clear. A much older induction proof on k where $n = 2^k$ was given by Cauchy around 1821.

The case $n = 2$ as we said was known from the ancient times. The algebraic proof is:

$$(x+y)^2 - 4xy = (x-y)^2 \geq 0,$$

hence $x^2 + y^2 \geq 2xy$. Geometric "visual" proofs are in Figures 1–4.

For $n = 3$ there are also some "quick" algebraic proofs. Here are a few. Consider $x^3 + y^3 + z^3 - 3xyz$ and express it in terms of the elementary symmetric functions (e_1, e_2, e_3).

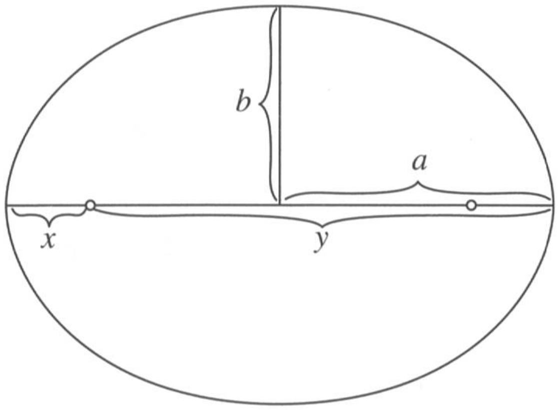

Fig. 3 "Astronomy proof". In the ellipse: $\frac{x+y}{2} = a \geq b = \sqrt{xy}$, since major semi-axes $\geq$ minor semi-axes.

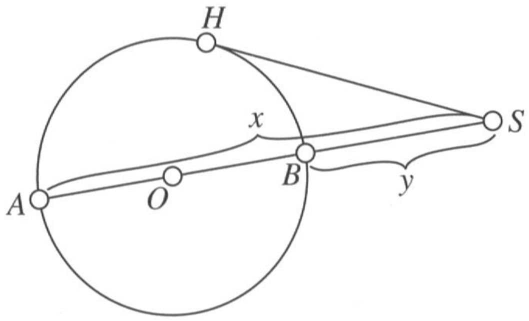

Fig. 4 "Satellite proof". $SA = x$, $SB = y$; $\frac{x+y}{2} = SO \geq SH = \sqrt{xy}$, distance to the Earth's center $\geq$ distance to the horizon.

We obtain by standard methods

$$\begin{aligned} x^3 + y^3 + z^3 - 3xyz &= e_1^3 - 3e_1e_2 = e_1(e_1^2 - 3e_2) \\ &= (x+y+z)[(x+y+z)^2 - 3(xy+yz+zx)] \\ &= (x+y+z)(x^2+y^2+z^2 - xy - yz - zx) \\ &= \frac{1}{2}(x+y+z)[(x-y)^2 + (y-z)^2 + (z-x)^2] \geq 0, \end{aligned}$$

because x, y and z are nonnegative. Hence, $x^3+y^3+z^3 \geq 3xyz$. The following polynomial identity also implies the AM-GM inequality in three variables $x, y, z \geq 0$:

$$(x+y+z)^3 - 27xyz = \frac{1}{2}\big[(x+y+7z)(x-y)^2 + (y+z+7x)(y-z)^2 + (z+x+7y)(z-x)^2\big].$$

In four variables:

$$(x+y+z+w)^4 - 4^4 xyzw = \frac{1}{3}\sum\big((x^2+y^2+11z^2+11w^2+14xy+58zw)(x-y)^2\big),$$

where Σ means the symmetric sum. And in general, as it can be shown, the difference $\sum_{i=1}^n x_i^n - \prod_{i=1}^n (nx_i)$ is of the form $\sum_{i<j} P_{ij}(x_i - x_j)^2$, where P_{ij} are homogeneous polynomials with positive coefficients and hence the AM-GM inequality.

In the next "quick" proof the convexity of the exponential function $e^x = \exp(x)$ is used. We have

$$\frac{1}{3}(x+y+z) = \frac{1}{3}(\exp\ln x + \exp\ln y + \exp\ln z) \geq \exp\frac{1}{3}(\ln x + \ln y + \ln z) = \sqrt[3]{xyz}.$$

Of course, it works for all n, not only for $n = 3$. A similar "quick" proof is to apply Jensen's inequality to the function $f(x) = x\ln x$. The classical (high-school) proof of Pólya (from around 1925) used convexity of e^x and the fact that $e^x \geq x + 1$, but this follows by noticing that the line $y = x + 1$ is the tangent line to the curve $y = e^x$ at $x = 0$. Substitute $\frac{x_i}{A} - 1$, $i = 1, \ldots, n$ and multiply. (Pólya once said that he dreamed this proof and that was his best dream ever.)

The *rearrangement inequality* is the following fact on inner products: $(x, y^\sigma) \leq (x, y)$, for all vectors $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n) \in \mathbf{R}^n$ with $x_1 \leq \cdots \leq x_n$ and $y_1 \leq \cdots \leq y_n$ and all permutations $\sigma \in S_n$ where $(x, y) = x_1y_1 + \cdots + x_ny_n$ and $(x, y^\sigma) = x_1y_{\sigma(1)} + \cdots + x_ny_{\sigma(n)}$. It is not hard to show that this also implies the AM-GM inequality. And the rearrangement inequality can also (standardly) be proved by induction on the number $n-i$, of fixed points of σ. The induction bases is the trivial case $i = 0$.

Newton's classical proof is as follows. Let e_k be the kth elementary symmetric function of $x_1, \ldots, x_n \geq 0$ and $E_k = e_k/\binom{n}{k}$, $E_0 := 1$. Then the *Newton inequality* says that $E_0, E_1, \ldots, E_n$ is a log-concave sequence, i.e., $E_{k-1}E_{k+1} \leq E_k^2$, for all $k = 1, \ldots, n$ with equality if and only if $x_1 = \cdots = x_n$. Now from

$$\prod_{i=1}^k (E_{i-1}E_{i+1})^i \leq \prod_{i=1}^k E_i^{2i}$$

it follows that $E_{k+1}^k \leq E_k^{k+1}$ or $E_k^{\frac{1}{k}} \geq E_{k+1}^{\frac{1}{k+1}}$. Hence (*Newton's lemma*) $E_1 \geq E_2^{\frac{1}{2}} \geq \cdots \geq E_n^{\frac{1}{n}}$ and the AM-GM inequality (and its refinements) follows. The above log-concavity of E_k's is a consequence of the general fact that if a real polynomial $P(x) = \sum_{i=0}^n a_ix^i$ has only real zeroes then a_k (and moreover, $a_k/\binom{n}{k}$), $k = 0, 1, \ldots, n$ is a log-concave sequence. (It seems the first rigorous proof of this fact was given by Sylvester about 1865.) The proof is by using Rolle's theorem (from 1691). Namely, if $P(x)$ has only real zeroes, then so does $Q(x) = D^kP(x)$, where $D = \frac{d}{dx}$ is the derivative. Then $Q_1(x) = x^{n-k}Q(x^{-1})$ also has only real zeroes and so does $R(x) = D^{n-k-2}Q_1(x)$. But $R(x)$ is a quadratic polynomial, so its discriminant is nonnegative. A little calculation shows that this is just the claim.

A quick topological argument is as follows. Let $M = \max\{x_1x_2\cdots x_n : x_1, \ldots, x_n \geq 0, \sum x_i = S\}$. M exists since the (continuous) product is defined on a compact set (simplex). M occurs when all x_i's are mutually equal (and so equal to $S/n := A$), because otherwise if two factors differ and the sum remains the same, the product decreases. Thus $M \leq A^n$, the AM-GM inequality.

We end this repertoire of proofs by remarking only that the increasing property for weighted means $M_r(x)$ is standardly proved by showing that the partial derivative $\frac{\partial}{\partial r}M_r \geq 0$. And this follows from Jensen's inequality for the function $f(x) = x^{\frac{q}{p}}, q > p > 0$, by checking that $f''(x) \geq 0$. And similarly the product property for $M_r(x)$.

3 Some interpretations, applications and generalizations

Let us first give a geometric interpretation of the AM-GM inequality. Consider an n-dimensional box (brick, rectangular parallelepiped) $\mathcal{B}$ whose side lengths from one corner are $x_1, \dots, x_n$. Then the AM-GM inequality is equivalent to $2^{n-1}(x_1 + \cdots + x_n) \geq n2^{n-1}\sqrt[n]{x_1 \dots x_n}$. The left-hand side is the total length of all edges of the box, i.e., it is the perimeter per($\mathcal{B}$) of $\mathcal{B}$. The right-hand side is the perimeter of the cube $\mathcal{C}$ with side length $\sqrt[n]{x_1 \dots x_n}$ and having the same volume $x_1 \dots x_n$ as $\mathcal{B}$. So the AM-GM inequality (vol($\mathcal{B}$) = vol($\mathcal{C}$) $\Rightarrow$ per($\mathcal{B}$) $\geq$ per($\mathcal{C}$)) is a kind of isoperimetric inequality: the cube has the minimal perimeter among all boxes of the given volume. (Is there any clear-short geometric argument for this?) Another way to think of the AM-GM inequality $(x_1 + \cdots + x_n)^n \geq (nx_1)(nx_2)\dots(nx_n)$ is that the cube of edge length $(x_1 + \cdots + x_n)$ has greater volume than any box with side lengths $nx_1, \dots, nx_n$ at one corner.

There is a whole variety of applications of the AM-GM inequality. Let us recall just a few simple ones from geometry. Euler noticed in 1765 that the circumradius R is at least as double as the inradius r of any triangle. Here is a short proof of this fact. Let S be the area of a triangle with side lengths a, b and c and perimeter $2s$. Recall,

$$S = \frac{abc}{4R} = rs = \sqrt{s(s-a)(s-b)(s-c)}.$$

Then $R \geq 2r$ is equivalent to

$$abc \geq 8(s-a)(s-b)(s-c),$$

or by putting $x = s - a$, $y = s - b$, $z = s - c$, to

$$(x+y)(y+z)(z+x) \geq 8xyz.$$

But this follows by multiplying three simple AM-GM inequalities $\frac{x+y}{2} \geq \sqrt{xy}$ etc. Equality holds only for an equilateral triangle. By using the three variables AM-GM inequality we get $(s-a)(s-b)(s-c) \leq \left(\frac{s}{3}\right)^3$, and hence

$$S = [s(s-a)(s-b)(s-c)]^{\frac{1}{2}} \leq \frac{s^2}{3\sqrt{3}},$$

the *isoperimetric property* for triangles with equality again only for an equilateral triangle. By using the AM-GM inequalities, the hyperbolic version of Euler's inequality (for triangles with circumcircle) is $\tanh(R) \geq 2\tanh(r)$, and similarly in the spherical case ([4]).

Euler's inequality holds in general for any Euclidean n-dimensional simplex: $R \geq nr$, with equality only for the regular simplex. A slick proof (given by L. Fejés-Tóth in 1965), that does not make use of the AM-GM inequality is as follows. Let $\Delta = \Delta(v_0, v_1, \ldots, v_n)$ be an n-simplex and $R = R(\Delta)$ its circumradius. The centroid c_i of the facet opposite to v_i is given (as a vector) by $c_i = \frac{1}{n}(v_0 + \cdots + v_{i-1} + v_{i+1} + \cdots + v_n)$. It is easy to check that the simplices Δ and $\Delta(c_0, c_1, \ldots, c_n)$ are similar with ratio n. Hence the distance $d(c_i, c_j) = \frac{1}{n}d(v_i, v_j)$ for all i, j. This similarity implies $R(\Delta) = nR(\Delta(c_0, c_1, \ldots, c_n))$. A ball of radius less than that of the inscribed ball can not meet every facet of Δ. Therefore $R(\Delta(c_0, c_1, \ldots, c_n)) \geq r$. Hence, $R = nR(\Delta(c_0, c_1, \ldots, c_n)) \geq nr$.

The 2-variable *Cauchy–Schwarz inequality* $(a^2+b^2)(c^2+d^2) \geq (ac+bd)^2$ by expanding both sides reduces to $a^2d^2 + b^2c^2 \geq 2abcd$ and this is again the 2-variable AM-GM inequality (it can also be deduced from *Fermat's two square theorem* $(a^2+b^2)(c^2+d^2) = (ac+bd)^2 + (ad-bc)^2$). But the general Cauchy–Schwarz inequality $|(x, y)| \leq \|x\| \, \|y\|$ simply follows from two geometric facts:

$$(x, y) = \|x\| \, \|y\| \cos \angle(x, y) \quad \text{and} \quad |\cos \angle(x, y)| \leq 1$$

for all angles $\angle(x, y)$. Or algebraically from Lagrange's identity

$$\|x\|^2 \, \|y\|^2 = (x, y)^2 + \sum_{i<j}(x_i y_j - x_j y_i)^2$$

(it could also be called the *Pythagoras–Fermat–Lagrange identity*, see more on this topic in [5]). Or analytically, by nonnegativity of the quadratic function $f(t) = \sum_{i=1}^{n}(x_i t + y_i)^2$.

A notorious application of the AM-GM inequality is in proving the *general isoperimetric inequality*: if V is the volume and S the surface area of a convex body $K \subseteq \mathbf{R}^n (S = \mathrm{vol}_{n-1}(\partial K), V = \mathrm{vol}_n(K))$ then $S^n \geq n^n \omega_n V^{n-1}$ with equality if and only if K is an n-ball (here $\omega_n = \pi^{n/2}/\Gamma(n/2+1)$ is the volume of the unit n-ball; Γ is the gamma function, $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$). A standard proof (by approximation) reduces it to the *Brunn–Minkowski* inequality

$$[\mathrm{vol}(X+Y)]^{1/n} \geq [\mathrm{vol}(X)]^{1/n} + [\mathrm{vol}(Y)]^{1/n}$$

for all nonempty compact $X, Y \subseteq \mathbf{R}^n$, and which for boxes with edges $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ at one of the corners reduces to

$$\prod_{i=1}^{n}(x_i + y_i)^{1/n} \geq \prod_{i=1}^{n} x_i^{1/n} + \prod_{i=1}^{n} y_i^{1/n},$$

and this is by the AM-GM inequality equivalent to

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

(It is a special case of the *Aleksandrov–Fenchel inequality* for mixed volumes.) For n-simplices Δ, the *isoperimetric ratio* $S(\Delta)^n/V(\Delta)^{n-1}$ attains the minimum if and only if Δ is a regular simplex. There are also various discrete analogues of isoperimetric inequalities.

Here is a nice application in algebra. In 1967 Motzkin first found a real polynomial $f = f(X,Y) = X^4Y^2 + X^2Y^4 + 1 - 3X^2Y^2$ which is nonnegative (by using the AM-GM inequality), and yet it can not be a sum of squares of real polynomials. Indeed, suppose $f = \sum f_i^2$, for some $f_i \in \mathbf{R}[X,Y]$, $i = 1,\dots,n$. Clearly, each f_i has degree ≤ 3, and so each f_i is a linear combination of $1, X, Y, X^2, XY, Y^2, X^3, X^2Y, XY^2, Y^3$. But X^3 does not appear in some f_i, because otherwise X^6 would appear in f with a positive coefficient. Similarly, Y^3 and then also X^2 and Y^2 and X and Y do not appear. Hence, each f_i is of the form

$$f_i = a_i + b_iXY + c_iX^2Y + d_iXY^2.$$

But then $\sum b_i^2 = -3$, a contradiction. However, every nonnegative real polynomial is a sum of squares of rational functions as Artin showed in 1927 (answering affirmatively to the 17th Hilbert problem from 1900). Similar examples exist in more variables and their positivity follows from the AM-GM inequality.

Now some generalizations of AM-GM. For any vector $a = (a_1,\dots,a_n) \in \mathbf{R}^n$, define the $[a]$-mean of $x_1,\dots,x_n \geq 0$ by

$$[a] = \frac{1}{n!}\sum_{\sigma\in S_n} x_{\sigma_1}^{a_1}\dots x_{\sigma_n}^{a_n}.$$

For example, if $a = (1,0,\dots,0)$, $[a]$ is the arithmetic mean of $x_1,\dots,x_n$ and if $a = \left(\frac{1}{n},\dots,\frac{1}{n}\right)$, then $[a]$ is the geometric mean. In general, $[a]^{1/(a_1+\dots+a_n)}$ is the *Muirhead mean* of $x_1,\dots,x_n$.

Muirhead's inequality (from 1916) says that $[a] \leq [b]$ for all $x_1,\dots,x_n \geq 0$ if and only if there is a doubly stochastic $n\times n$ matrix P such that $a = Pb$. An $n\times n$ real matrix is *doubly stochastic* if all numbers are nonnegative and the sum of every row and every column is equal to 1. In fact, a doubly stochastic matrix is a weighted average of *permutation matrices* (in any row and any column only one unit, the rest are zeroes); this is the *Birkhoff–von Neumann theorem.* Assuming $a_1 \geq \dots \geq a_n$ and $b_1 \geq \dots \geq b_n$, then $[a] \leq [b]$ is equivalent to the fact that b *majorizes* a, i.e.,

$$a_1 \leq b_1,\ a_1 + a_2 \leq b_1 + b_2, \dots, a_1 + \dots + a_n = b_1 + \dots + b_n.$$

The AM-GM is a special case of Muirhead's inequality (and in fact, they are equivalent).

Also Hölder's inequality seems more general, but it is also equivalent to the AM-GM inequality. And there are many other important inequalities equivalent to the AM-GM inequality.

The *generalized f-mean* for a continuous injective function $f\colon I \to \mathbf{R}$ on an interval $I \subseteq \mathbf{R}^+$ is defined by

$$M_f(x_1,\dots,x_n) := f^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} f(x_i)\right).$$

If $I = \mathbf{R}^+$ and $f(x) = x^r$ then the f-mean is the rth power mean $M_r(x)$. Additional assumptions on f yield generalizations of the power mean increasing property (and in particular of the AM-GM inequality).

Direct applications of the AM-GM are also in numerical analysis, in optimization theory, financial mathematics, probability theory and statistics, information theory, mathematical physics, and many other areas.

4 Combinatorial proof

Back to our main AM-GM topic, we give now a combinatorial proof. Let $x_1, \dots, x_n$ be positive integers and $nA = \sum_{i=1}^n x_i$. The AM-GM inequality is equivalent to

$$(nA)^n \geq (nx_1)(nx_2)\dots(nx_n).$$

Let $X_1, \dots, X_n$ and Y be finite disjoint sets, $|X_i| = nx_i, i = 1, \dots, n$ and $|Y| = nA$. Let us find an injection $f\colon \prod_{i=1}^n X_i \to Y^n = Y \times Y \times \dots \times Y$. In case of two sets S and T with $|S| = a < b = |T|$ and $t_0 \in T$, we can define an injection $f\colon S \times T \hookrightarrow (S \cup \{t_0\}) \times (T \setminus \{t_0\})$ by $f(s,t) = (s,t)$ if $t \neq t_0$ and $f(s,t_0) = (t_0, g(s))$, where $g\colon S \to T \setminus \{t_0\}$ is any injection (which exists because $a \leq b-1$; $f = f_{t_0,g}$). This is in fact a combinatorial proof of the inequality $ab \leq (a+1)(b-1)$. In general, if all x_i are equal we have equality; otherwise there exist i and j such that $x_i < A$ and $x_j > A$. Choose an element $z_1 \in X_j$, add it to X_i, and define a new partition of $X = \cup_{k=1}^n X_k$ by $X = \cup_{k=1}^n X_k^{(1)}$ where $X_k^{(1)} = X_k, k \neq i, j$ and $X_i^{(1)} = X_i \cup \{z_1\}, X_j^{(1)} = X_j \setminus \{z_1\}$. Let $f_{z_1,g_1} : X_i \times X_j \hookrightarrow (X_i \cup \{z_1\}) \times \left(X_j \setminus \{z_1\}\right)$ and $f_1\colon \prod_{k=1}^n X_k \to \prod_{k=1}^n X_k^{(1)}$, the corresponding injection. (Recall the number of injections of $N \hookrightarrow X$ where $n = |N| \leq |X| = x$, is $x^{\underline{n}} := x(x-1)(x-2)\dots(x-n+1)$.) Again, if all $|X_k^{(1)}|$ are equal we are done, otherwise form a new partition of $X = \cup_{k=1}^n X_k^{(2)}$ and define an injection $f_2\colon \prod_{k=1}^n X_k^{(1)} \to \prod_{k=1}^n X_k^{(2)}$ and continue this in the same way until we reach equality, i.e., there exists $m \in \mathbf{N}$ such that $|X_k^{(m)}| = |Y|$ for all $1 \leq k \leq n$, and a bijection $h\colon \prod_{k=1}^n X_k^{(m)} \to Y^n$. Then $f := h \circ f_m \circ \dots \circ f_1\colon \prod_{k=1}^n X_k \to Y^n$ is an injection. This proves the AM-GM inequality for all nonnegative integers.

If $x_1, \dots, x_n \geq 0$ are any real numbers then by the above combinatorial reasons we know all 2^n AM-GM inequalities for all combinations of $\lfloor\quad\rfloor$ and $\lceil\quad\rceil$ (lower and upper integer parts, or "floors" and "ceilings") applied to all $x_1, \dots, x_n$ and then by convexity and continuity arguments it holds for them, too. The following is the moral of the above proof. When a partition of a finite set in n blocks has equal sized blocks, then the number of ways to pick just one point from each block is the largest.

5 Physical interpretation

Now a bit of physics (inspired by [6]). Consider n bodies or solids (e.g., boxes or bricks) with the same heat capacity $C > 0$. Suppose the ith box has the temperature $x_i > 0$, $i = 1, \dots, n$. Imagine now that we put all the bricks together. Then the temperatures tend to distribute so that they are equally distributed at the end of the experiment. This is a consequence of the *first law of thermodynamics* (the law of conservation of energy): temperatures tend to differ as little as possible until they eventually become equally distributed (with the same probability everywhere when the equilibrium is achieved).

At the end of the experiment the total entropy of the system did not decrease. This is a consequence of the *second law of thermodynamics*: the total entropy of a physical system increases (rather, does not decrease) until the system reaches its limit (the popular phrase is "the entropy of the universe tends to a maximum"). The entropy S measures the number of ways the thermodynamic system may be rearranged, i.e., it measures unpredictability of a system, or it is a "measure of disorder". By the "heating formula" (Boltzmann) the entropy change is given by $\Delta S = C\ln(T/T_0)$. Here T_0 is the initial temperature, and T the final temperature. The starting temperatures T_0 are $x_1, x_2, \ldots, x_n$, and the boxes (of the same heat capacity $C > 0$) will in a continuous manner by the end of the experiment have temperature equal to the mean value $A = A(x_1, \ldots, x_n)$. The total entropy did not decrease, so $\sum \Delta S = \sum_{i=1}^{n} C \ln \frac{A}{x_i} \geq 0$, and this implies the AM-GM inequality.

6 The arithmetic-geometric mixed mean and final remarks

The arithmetic mean $M_1 = A$ and the geometric mean $M_0 = G$ of two numbers $x, y > 0$ give rise to the new mixed (or composite) *arithmetic-geometric mean* (AGM for short), denoted by $M_{0,1}(x, y) = GA(x, y)$. It is defined as the common limit of the bounded decreasing sequence $(x_n)_{n\geq 0}$ and the bounded increasing sequence $(y_n)_{n\geq 0}$ given by $x_0 = x$, $y_0 = y$ and $x_{n+1} = \frac{1}{2}(x_n + y_n) = M_1(x_n, y_n)$, $y_{n+1} = \sqrt{x_n y_n} = M_0(x_n, y_n)$. The convergence is rather fast since $|x_{n+1} - y_{n+1}| < \frac{1}{2}|x_n - y_n|$. As Gauss noted in 1818 (and independently Abel in 1827), the value

$$M_{0,1}(x, y) = \frac{\pi}{2I(x, y)}$$

where

$$I(x, y) = \int_0^{\pi/2} \frac{d\varphi}{\sqrt{(x\cos\varphi)^2 + (y\sin\varphi)^2}} = I\left(\frac{x+y}{2}, \sqrt{xy}\right),$$

and the AGM can not be expressed any simpler than in terms of complete elliptic integrals. The basic Pythagorean inequality $G(x, y) \leq GA(x, y) \leq A(x, y)$ (or $M_0 \leq M_{0,1} \leq M_1$) is a natural refinement of the AM-GM inequality in two variables. (What is an eloquent meaning of $GA(x, y)$ on Figures 2–4?)

Interesting recent research on AGM are papers [7] and [8]. Let us mention only that the mixed mean $M_{p,q} = M_{p,q}(x, y)$ for parameters $p \leq q$ can also be defined in a similar manner as $M_{0,1}$ and then recursively general means with more parameters and more variables. Inequality like $M_p \leq M_{p,q} \leq M_q$ generalizes the Pythagorean inequality and refines the power mean increasing property. More generally, we can consider a mixed (f, g)-mean for functions f and g and moreover multi-functional mixed means of more variables.

Another type of "mixed-means" was introduced in [9], where it is proved that

$$\begin{aligned} &M_1(M_0(x_1), M_0(x_1, x_2), \ldots, M_0(x_1, \ldots, x_n)) \\ &\quad \leq M_0(M_1(x_1), M_1(x_1, x_2), \ldots, M_1(x_1, \ldots, x_n)). \end{aligned}$$

(Needless to say, $M_0 \leq M_1$, the ordinary AM-GM, is used in the proof.)

In conclusion, we might say that many facets of the AM-GM inequality in elementary algebra, analysis, topology, geometry, combinatorics, physics, modern mixed mean theory etc. exemplarily show that fundamental principles are profound, unifying and amalgamated throughout mathematics and suggest further research and applications.

References

[1] G. Hardy, J. Littlewood, G. Pólya, *Inequalities*, 2*nd* ed., Cambridge University Press, Cambridge, 1952.

[2] J.M. Steele, *Cauchy–Schwarz Master Class*, MAA, Cambridge University Press, Cambridge, 2004.

[3] M. Hajja, Some elementary aspects of means, *Int. J. Math and. Math Sc.* ID68960 (2013), pp. 1–9.

[4] D. Svrtan, D. Veljan, Non euclidean versions of some classical triangle inequalities, *Forum Geom.* **12**, (2012) pp. 197–209.

[5] A.M. Sommariva, The generating identity of Cauchy–Schwarz–Bunyakovsky inequality, *Elem. Math.* **63**, (2008) pp. 1–5.

[6] M. Gromov, In search for a structure, Part 1: On Entropy, preprint, online (2013), pp. 1–27.

[7] S. Adlaj, An eloquent formula for the perimeter of an ellipse, *Notices AMS* **59** (8) (2012), pp. 1094–1099. (see also S. Adlaj, An arithmetic-geometric mean of a third kind!, retrieved online, 29 December 2015).

[8] J. Spandaw, D. van Straten, Hyperellyptic integrals and generalized arithmetic-geometric mean, *Ramanujan J.* **28** (2012), pp. 61–78.

[9] K. Kedlaya, Proof of mixed arithmetic-mean geometric-mean inequality, *Amer. Math. Monthly* **101** (1994), pp. 355–357.

Darko Veljan
Department of Mathematics
University of Zagreb
Bijenicka 30
10000 Zagreb, Croatia
e-mail: darko.veljan@gmail.com