Inequalities

Sums of squares

Over the complex numbers, every polynomial has at least one root by the fundamental theorem of algebra. Over the reals, however, it is possible to define polynomials that are always greater than (or equal to) zero. These are known as *positive (semi)definite* functions. One such example is $x^2 + y^2 \geq 0$, which is true for all $x, y \in \mathbb{R}$. In general, as squares of real numbers are non-negative, sums of squares are also non-negative. This is the most basic useful inequality.

- If $x_1, x_2, \ldots, x_n \in \mathbb{R}$ and $\alpha_1, \alpha_2, \ldots, \alpha_n > 0$, then $\alpha_1 x_1{}^2 + \alpha_2 x_2{}^2 + \ldots + \alpha_n x_n{}^2 \geq 0$, with equality if and only if $x_1{}^2 = x_2{}^2 = \ldots = x_n{}^2 = 0$. **[Sum of squares inequality]**

Artin proved Hilbert's seventeenth problem, namely that every positive semidefinite polynomial (and, by extension, rational function) can be expressed as the sum of squares of rational functions. Charles Delzell later developed an algorithm to do so. Hence, it is *theoretically* possible to prove any inequality involving rational functions simply by reducing it to the sum of squares inequality. However, this approach is similar in its impracticality to building an automobile using Stone Age tools. Certainly, it is impossible in the 270 minutes allocated in the International Mathematical Olympiad. Nevertheless, we can still tackle *some* basic inequalities in this way, especially if they are expressible as the sums of squares of polynomials.

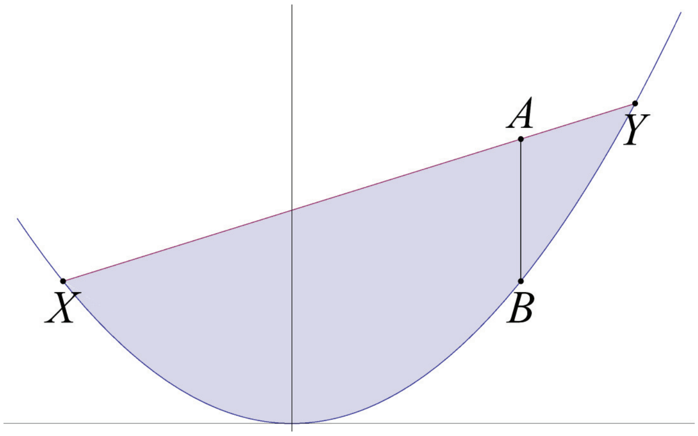
1. Prove that $x^2 + y^2 + z^2 \geq x\,y + y\,z + z\,x$.

Jensen's inequality

According to Ross Atkins, "Jensen's inequality is greater than or equal to all other inequalities". This strongly indicates that it is advisable to assimilate it into one's problem-solving repertoire. It is geometrically very obvious, namely that the barycentre of a convex figure is located inside it. This makes it all the more remarkable that so many useful inequalities, such as the power means inequality, are trivialised by Jensen's inequality.

- A continuous function f is *convex* over an interval (a, b) if, for all $x_1, x_2 \in (a, b)$ and $\alpha_1, \alpha_2 \in [0, 1]$ such that $\alpha_1 + \alpha_2 = 1$, we have $f(\alpha_1 x_1 + \alpha_2 x_2) \leq \alpha_1 f(x_1) + \alpha_2 f(x_2)$. If the reverse inequality holds instead, the function is *concave*. **[Definition of convexity]**

This is most easily represented with the aid of a diagram:



For any two points X and Y on the curve of a convex function, any point A on the line segment $X\,Y$ lies above the curve. The Australian IMO team leader, Ivan Guo, created a mnemonic for remembering the shapes of generic convex and concave functions:

- Ivan: "Con**cave** looks like a **cave**, and con**vex** looks like a **vex**."
- Someone else: "What's a vex?"
- Ivan: "An upside-down cave."

2. Let f be a convex function over the interval (a, b). Let $\{x_1, x_2, \ldots, x_n\} \subset (a, b)$ and $\{\alpha_1, \alpha_2, \ldots, \alpha_n\} \subset [0, 1]$ such that $\alpha_1 + \alpha_2 + \ldots + \alpha_n = 1$. Show that $f(\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_n x_n) \leq \alpha_1 f(x_1) + \alpha_2 f(x_2) + \ldots + \alpha_n f(x_n)$. **[Weighted Jensen's inequality]**

Observe that the $n = 2$ case of the weighted Jensen inequality is just the definition of convexity. It is often quoted as the slightly less general (but asymptotically equivalent) theorem where $\alpha_1 = \alpha_2 = \ldots = \alpha_n = \frac{1}{n}$.

- Let f be a convex function over the interval (a, b), and let $\{x_1, x_2, \ldots, x_n\} \subset (a, b)$. Then $f\left(\frac{1}{n}(x_1 + x_2 + \ldots + x_n)\right) \leq \frac{1}{n}(f(x_1) + f(x_2) + \ldots + f(x_n))$. **[Jensen's inequality]**

3. If $\{x_1, x_2, \ldots, x_n\}$ are all positive, show that $\frac{1}{n}(x_1 + x_2 + \ldots + x_n) \geq \sqrt[n]{x_1 x_2 \ldots x_n}$. **[AM-GM inequality]**

4. If a and b are two non-zero real numbers such that $a \geq b$, show that
$\sqrt[a]{\frac{1}{n}(x_1{}^a + x_2{}^a + \ldots + x_n{}^a)} \geq \sqrt[b]{\frac{1}{n}\left(x_1{}^b + x_2{}^b + \ldots + x_n{}^b\right)}$. **[Power means inequality]**

The *arithmetic mean*, *quadratic mean* and *harmonic mean* arise when a is 1, 2 and -1, respectively. The geometric mean is the limit as $a \to 0$.

Muirhead's inequality

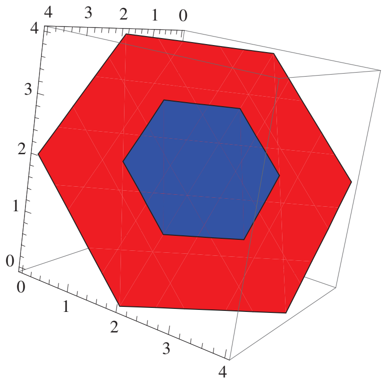
Muirhead's inequality is a powerful generalisation of the AM-GM inequality. Before we can define it, however, it is necessary to introduce the idea of *majorisation*.

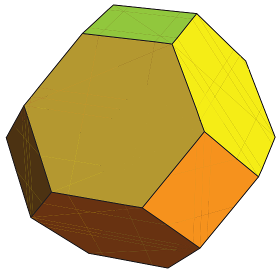
- Let $a_1 + a_2 + \ldots + a_n = 1$ and $b_1 + b_2 + \ldots + b_n = 1$, and all $a_i \in [0, 1]$ and $b_i \in [0, 1]$. Assume further that the sequences are ordered such that $a_1 \geq a_2 \geq \ldots \geq a_n$ and $b_1 \geq b_2 \geq \ldots \geq b_n$. Then $\{a_i\}$ *majorises* $\{b_i\}$ if and only if $a_1 + a_2 + \ldots + a_k \geq b_1 + b_2 + \ldots + b_k$ for all $k \in [1, n]$. **[Definition of majorisation]**

The sequence (4, 0, 0, 0), for example, majorises (1, 1, 1, 1), as they are sorted into descending order and the following inequalities hold:

- $4 \geq 1$;
- $4 + 0 \geq 1 + 1$;
- $4 + 0 + 0 \geq 1 + 1 + 1$;
- $4 + 0 + 0 + 0 = 1 + 1 + 1 + 1$.

Occasionally, the notation $(4, 0, 0, 0) \succcurlyeq (1, 1, 1, 1)$ is used to denote this relationship. Majorisation may appear at first to be a contrived relation, although it has several equivalent and more enlightening formulations. We interpret $\underline{a} = (a_1, a_2, \ldots, a_n)$ as a vector in $\mathbb{R}^n$, and consider the set of $n!$ (not necessarily distinct) vectors obtained by permuting the elements of the vector $\underline{a}$. They all lie in the $(n - 1)$-dimensional plane with equation $x_1 + x_2 + \ldots + x_n = a_1 + a_2 + \ldots + a_n$, and form the vertices of a *permutation polytope*. In general (when all elements are distinct), the three-variable case is a hexagon, whereas the four-variable case is a truncated octahedron.





The red and blue hexagons correspond to the sets {4, 2, 0} and {3, 2, 1}, respectively. The condition that the red hexagon contains the blue hexagon is equivalent to $\{4, 2, 0\} \succcurlyeq \{3, 2, 1\}$, which in turn is equivalent to the *Birkhoff-von Neumann theorem*: (3, 2, 1) can be expressed as a weighted average of permutations of (4, 2, 0). More subtly, this also implies that, for all $x, y, z \geq 0$, the polynomial $x^4 y^2 z^0 + z^4 y^2 x^0 + y^4 z^2 x^0 + x^4 z^2 y^0 + z^4 x^2 y^0 + y^4 x^2 z^0$ is greater than or equal to $x^3 y^2 z^1 + z^3 y^2 x^1 + y^3 z^2 x^1 + x^3 z^2 y^1 + z^3 x^2 y^1 + y^3 x^2 z^1$; a fact known as *Muirhead's inequality*.

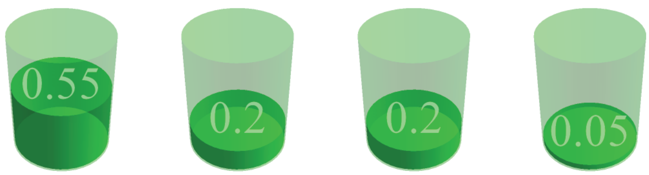
- Let $\{x_1, x_2, \ldots, x_n\}$, $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ and $\{\beta_1, \beta_2, \ldots, \beta_n\}$ be sequences of non-negative real numbers. If $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ majorises $\{\beta_1, \beta_2, \ldots, \beta_n\}$, then $\sum_{\text{sym}} (x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_n^{\alpha_n}) \geq \sum_{\text{sym}} \left(x_1^{\beta_1} x_2^{\beta_2} \ldots x_n^{\beta_n}\right)$. The sigmas denote symmetric sums, *i.e.* sums over all $n!$ permutations of $\{x_1, x_2, \ldots, x_n\}$. **[Muirhead's inequality]**

It is discussed in https://nrich.maths.org/discus/messages/67613/Muirhead-69859.pdf. Geoff Smith described how Muirhead's inequality is not well known amongst members of the IMO jury; occasionally certain inequalities, which were highly amenable to attack by this method, appeared on the IMO as a result of this.

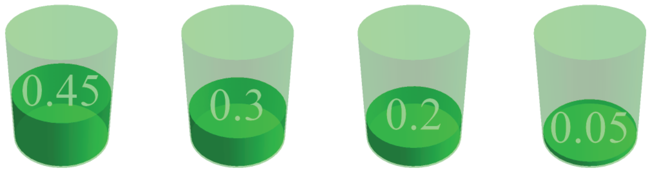
5. Prove that, for all positive real numbers x, y and z, we have
$2x^3 + 2y^3 + 2z^3 \geq x^2 y + y^2 x + y^2 z + z^2 y + z^2 x + x^2 z$.

Majorisation as fluid transfer

We have already defined majorisation in terms of decreasing sequences and permutation polytopes. A third interpretation involves containers of fluid. In each configuration below, the total volume of fluid is 1 unit; we will assume this without loss of generality to simplify things.



Suppose we have a sequence of containers of fluid, such that if container X is immediately to the left of container Y, then X contains at least as much fluid as Y. We are allowed to siphon fluid from X to Y as long as this weak inequality is maintained. From the configuration above, we can siphon up to 0.175 units of fluid from the first container to the second one without breaking the weak inequality. In the diagram below, we have transferred 0.1 units.



This is known as a *valid q-move*, where $q = 0.1$ is the amount of fluid transferred. We can continue in this manner. The *fluid transfer lemma* states that we can get from an initial sequence S_0 to (arbitrarily close to) a target sequence S_ω by applying valid q-moves if and only if S_0 majorises S_ω. A more formal definition follows:

- Suppose S_0 and $S_\omega = \{b_1, b_2, \ldots, b_n\}$ are two weakly decreasing sequences of non-negative real numbers, each with unit sum and length n. Let $\epsilon > 0$ be a small real number. Define a *valid q-move* to be an operation $\{a_1, a_2, \ldots, a_k, a_{k+1}, \ldots, a_n\} \to \{a_1, a_2, \ldots, a_k - q, a_{k+1} + q, \ldots, a_n\}$ such that the sequence remains strictly decreasing and still majorises S_ω. Then there exists some $\delta \ll \epsilon$ such that there exists a finite sequence of N valid δ-moves $S_0 \to S_1 \to \ldots \to S_N$ such that each term of S_N differs by the corresponding term of S_ω by at most ϵ if and only if S_0 majorises S_ω. **[Fluid transfer lemma]**

Proof:

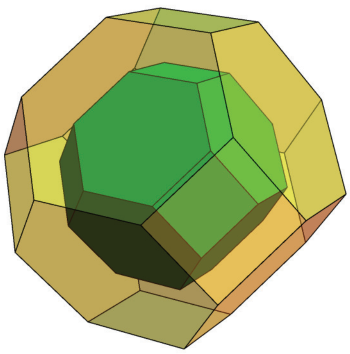
The 'only if' part is much easier, as it is evident that S_0 majorises S_1, which in turn majorises S_2. By induction, S_0 majorises S_N. If S_0 does not majorise S_ω, then one of the weak inequalities must be broken by an amount h. If we let $\epsilon < \frac{h}{n}$, then S_N must be sufficiently close to S_ω to also break one of those inequalities. Hence, S_0 does not majorise S_N, so we have a contradiction.

For the 'if' part, note that there are a finite number of attainable configurations for a given S_0 and δ, and the process cannot cycle, so must eventually terminate. Suppose we perform valid δ-moves arbitrarily until we reach a position S_N where no further valid δ-moves are possible.

By definition, for each pair of adjacent elements (a_i, a_{i+1}) in S_N, it must be the case that either:

- $a_i - a_{i+1} < 2\,\delta$ (in which case applying a δ-move would break the weakly decreasing criterion);
- or $(a_1 + a_2 + \ldots + a_i) - (b_1 + b_2 + \ldots + b_i) < \delta$ (in which case applying a δ-move would break the majorisation criterion).

If the first case applies to all pairs of adjacent elements, we have $a_1 - a_n < 2\,(n-1)\,\delta$. So, each element must be within $2\,(n-1)\,\delta$ of the mean, $\frac{1}{n}$. As S_N majorises S_ω, the same must be true of S_ω. Hence, corresponding elements can differ by no more than $4\,(n-1)\,\delta$, which we can make smaller than ϵ by letting δ be sufficiently small. This leaves the alternative case where there exists some i such that $0 < (a_1 + a_2 + \ldots + a_i) - (b_1 + b_2 + \ldots + b_i) < \delta$. In that case, we can split the problem into two separate problems: one involving the first i elements of the sequences, and the other involving the last $n - i$. (We need not worry that the sum of the first i elements of S_N is slightly greater than that of S_ω, as we can make the difference arbitrarily small. It is not important that things are exact, as long as the largest accumulative error is smaller than ϵ.) By inducting on the number of elements, we prove the fluid transfer lemma.



Returning to the geometric interpretation, this means that we can incrementally move the vertices of the larger polytope inwards (varying two coordinates of any vertex at any one time whilst preserving the full symmetry group) until it becomes arbitrarily close to 'suffocating' the smaller polytope. This is rather intuitive, and implies the Birkhoff-von Neumann theorem.

Energy minimisation lemma

A corollary of this lemma is the *energy minimisation lemma*. The proof relies on concepts from real analysis such as continuity and convergence, which are taught in most undergraduate maths degrees (such as the Cambridge Mathematical Tripos).

- Suppose we have a continuous function $E : \mathbb{R}^n \to \mathbb{R}$, known as the *energy function*. Suppose that applying a valid q-move to $S = \{a_1, a_2, \ldots, a_n\}$ cannot increase the value of $E(S)$. If we have two sequences S_0 and S_ω, such that S_0 majorises S_ω, then $E(S_0) \geq E(S_\omega)$. **[Energy minimisation lemma]**

Effectively, we associate an 'energy function' with the configuration of containers, such that the energy either remains constant or decreases whenever a valid q-move is applied. The energy minimisation lemma states that $E(S_0) \geq E(S_\omega)$ if S_0 majorises S_ω.

Proof:

Due to the fluid transfer lemma, we can apply valid q-moves to S_n to result in a new sequence (S_{n+1}) where each term differs from S_ω by at most $\epsilon = e^{-n}$. Starting from S_0, we produce an infinite sequence of sequences $\{S_0, S_1, \ldots\}$ where each term is an increasingly close approximation to S_ω. More specifically, this sequence of sequences *converges* to S_ω. As E is a continuous function, this means that $\{E(S_0), E(S_1), \ldots\}$ must converge to $E(S_\omega)$. Also, as valid q-moves cannot increase the value of $E(S)$, we have $E(S_0) \geq E(S_1) \geq \ldots$; by the monotone convergence theorem, this means $E(S_\omega)$ is the infimum of these terms, and therefore no larger than any of them. The result then follows.

Generalised Muirhead inequality

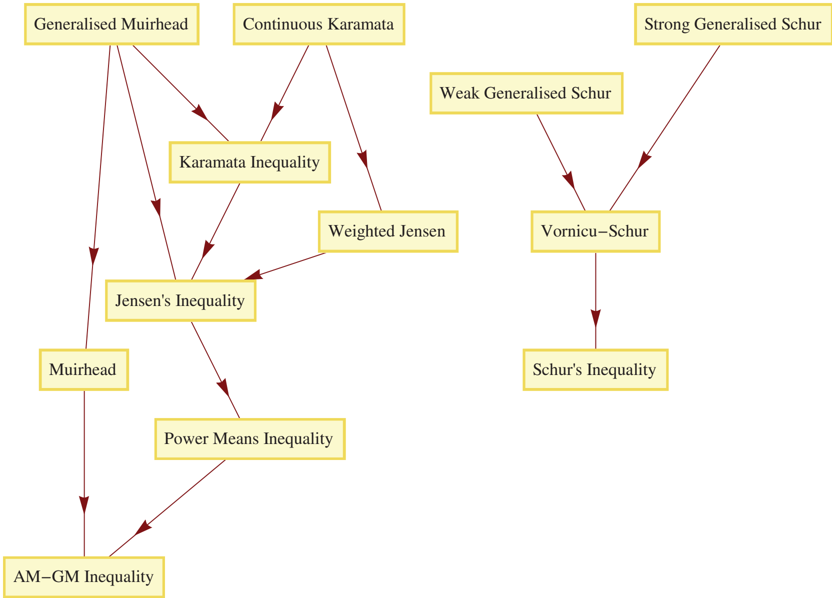
Using the lemmas developed above, it is straightforward to prove the generalised Muirhead inequality.

6. Let $f : (a, b) \to \mathbb{R}$ be a convex continuous function. Let $\alpha_1 \geq \beta_1 \geq \beta_2 \geq \alpha_2 \geq 0$, such that $\alpha_1 + \alpha_2 = \beta_1 + \beta_2 = 1$, and let $\{x_1, x_2\} \subset (a, b)$. Prove that $f(\alpha_1 x_1 + \alpha_2 x_2) + f(\alpha_1 x_2 + \alpha_2 x_1) \geq f(\beta_1 x_1 + \beta_2 x_2) + f(\beta_1 x_2 + \beta_2 x_1)$. **[Generalised Muirhead's inequality for 2 variables]**

7. Let $f : (a, b) \to \mathbb{R}$ be a convex continuous function, and let $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ be a weakly decreasing sequence of non-negative reals with unit sum. Let $\{x_1, x_2, \ldots, x_n\} \subset (a, b)$. Define the function $g(\alpha_1, \alpha_2, \ldots, \alpha_n, x_1, x_2, \ldots, x_n) = \sum_{\text{sym}} f(\alpha_1 x_{\sigma(1)} + \alpha_2 x_{\sigma(2)} + \ldots + \alpha_n x_{\sigma(n)})$, where the sum is taken over all $n!$ permutations σ of $\{1, 2, \ldots, n\}$. Prove that applying a valid q-move to $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ cannot cause g to increase.

8. Let f be a convex continuous function over (a, b) and $\{x_1, x_2, \ldots, x_n\} \subset (a, b)$. Let $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ and $\{\beta_1, \beta_2, \ldots, \beta_n\}$ be weakly decreasing sequences of non-negative reals, each with unit sum. Further, the former sequence majorises the latter. Prove that $\sum_{\text{sym}} f(\alpha_1 x_{\sigma(1)} + \alpha_2 x_{\sigma(2)} + \ldots + \alpha_n x_{\sigma(n)}) \geq \sum_{\text{sym}} f(\beta_1 x_{\sigma(1)} + \beta_2 x_{\sigma(2)} + \ldots + \beta_n x_{\sigma(n)})$, where the sums are taken over all $n!$ permutations σ of $\{1, 2, \ldots, n\}$. **[Generalised Muirhead's inequality]**

We can derive the ordinary Muirhead's inequality by letting $f(x) = e^x$. Similarly, Jensen's inequality follows from using the sequences $\{1, 0, \ldots, 0\} \succcurlyeq \left\{\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right\}$. This idea of inequalities generalising other inequalities gives a hierarchy:

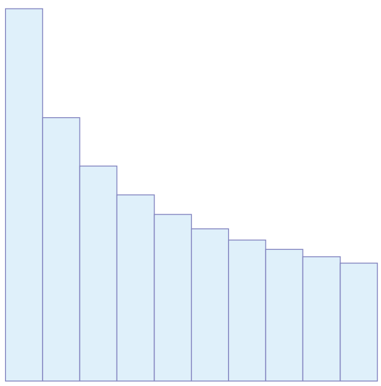


Karamata inequality

Consider the generalised Muirhead inequality. If we let $\{x_1, x_2, \ldots, x_n\} = \{1, 0, 0, \ldots, 0\}$, then we obtain the *Karamata inequality* as a special case.

- Suppose $\{a_i\}$ majorises $\{b_i\}$, and f is a convex function. Then $f(a_1) + \ldots + f(a_n) \geq f(b_1) + \ldots + f(b_n)$. **[Karamata inequality]**

This can be extended in another direction. We can assume without loss of generality that $\sum a_i = \sum b_i = 1$. Effectively, we can consider two new functions, $g'(x) = n\, a_{\lceil x n \rceil}$ and $h'(x) = n\, b_{\lceil x n \rceil}$, which are defined on the open interval $(0, 1)$. As $\{a_i\}$ majorises $\{b_i\}$ and the sequences are sorted in descending order, we have that g' and h' are weakly decreasing and $\int_0^k g'(x)\, dx \geq \int_0^k h'(x)\, dx$ for all $0 < k < 1$. We represent these integrals by $g(x)$ and $h(x)$, respectively. It is clear that $g(0) = h(0) = 0$ and $g(1) = h(1) = 1$.



The graph of $g'(x)$ is a collection of n rectangles of decreasing height. Integrating this to obtain $g(x)$ results in a concave line formed from n straight line segments of decreasing gradient. If we take the limit as n tends towards infinity, the sequences in Karamata's inequality are replaced with arbitrary non-negative decreasing functions, g' and h'.

- Suppose g and h are increasing concave functions with domain $[0, 1]$ such that $g(0) = h(0) = 0$, $g(1) = h(1) = 1$ and $g(k) \geq h(k)$ for all $k \in [0, 1]$. The derivatives of $g(x)$ and $h(x)$ with respect to x are denoted $g'(x)$ and $h'(x)$, respectively. Let f be an arbitrary convex function. Then $\int_0^1 f(g'(x))\, dx \geq \int_0^1 f(h'(x))\, dx$. **[Continuous Karamata inequality]**

Schur's inequality

A useful inequality that can be proved using sums of squares is *Schur's inequality*. Unlike the previous inequalities, which generalise to arbitrarily many variables, this has just three terms.

9. Suppose $a \geq b \geq c$ and $x + z \geq y \geq 0$. Show that $x^2(a-b)(a-c) + y^2(b-c)(b-a) + z^2(c-a)(c-b) \geq 0$. **[Strong 6-variable Schur]**

It is often quoted as the much weaker result shown below.

10. Show also that $x(a-b)(a-c) + y(b-c)(b-a) + z(c-a)(c-b) \geq 0$. **[Weak 6-variable Schur]**

This can be used, with a little work, to form a very powerful inequality.

11. Let $f : \mathbb{R} \to \mathbb{R}^+$ be a function expressible as the sum of non-negative monotonic functions. Let $g : \mathbb{R} \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$ be odd and increasing. Show that
$f(a)\, g(h(a-b)\, h(a-c)) + f(b)\, g(h(b-c)\, h(b-a)) + f(c)\, g(h(c-a)\, h(c-b)) \geq 0$. **[Weak generalised Schur]**

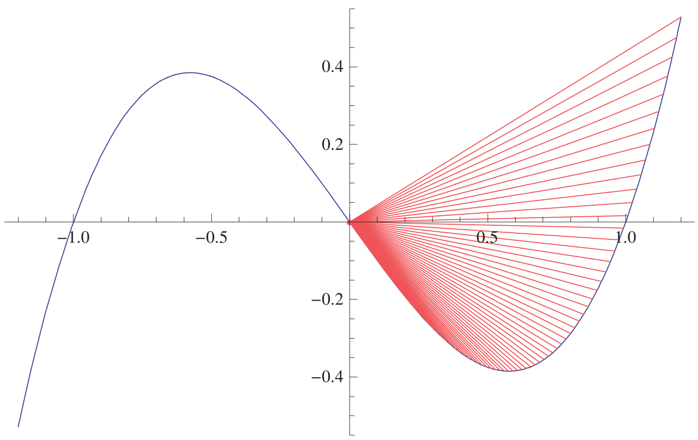
When $h(w) = w^k$ and $g(w) = w$, this is known as the *Vornicu-Schur inequality*. With the additional constraints of $k = 1$ and $f(w) = w^p$, this is simply Schur's inequality.

- If $a, b, c \in \mathbb{R}^+$, then $a^p(a-b)(a-c) + b^p(b-c)(b-a) + c^p(c-a)(c-b) \geq 0$. **[Schur's inequality]**

It is popularly believed that a suitable combination of Muirhead and Schur can conquer any inequality. This is obviously an exaggeration, since neither can prove (for instance) Jensen's inequality. Nevertheless, most symmetric inequalities in three variables submit to such an attack.

12. Prove that $x^6 + y^6 + z^6 + 3\,x^2\,y^2\,z^2 \geq 2\,x^3\,y^3 + 2\,y^3\,z^3 + 2\,z^3\,x^3$.

Nevertheless, we can go further. The strong 6-variable Schur inequality can also be generalised in a similar way to its weaker counterpart. We define a function f to be *positive-illuminable* if $f(\alpha x) \leq \alpha f(x)$ for all $0 \leq \alpha \leq 1$ and $x \geq 0$. Informally, this means that a light source placed infinitesimally above the origin will be able to illuminate every point on the curve $y = f(x)$, $x \geq 0$ from above. This is demonstrated in the following diagram, where no rays emitted from the origin intersect the curve twice. Positive-illuminability is a weaker condition than convexity.



We are now in a position to state and prove the stronger generalised form of Schur's inequality.

13. Let $f : \mathbb{R} \to \mathbb{R}^+$ be a function expressible as the sum of non-negative monotonic functions. Let $g : \mathbb{R} \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$ be odd, increasing and positive-illuminable. Show that $f(a)^2\, g(h(a-b)\, h(a-c)) + f(b)^2\, g(h(b-c)\, h(b-a)) + f(c)^2\, g(h(c-a)\, h(c-b)) \geq 0$. **[Strong generalised Schur]**

Calculus

Although ideas of limiting processes and integration can be traced back to Archimedes, our modern understanding of calculus was developed much later. It was conceived independently, and almost simultaneously, by Sir Isaac Newton and Gottfried Leibniz. As Newton only considered differentiation with respect to time, we currently use Leibniz's (much clearer) notation instead.

In the explorations of various general inequalities, terms such as 'increasing', 'convex' and 'positive-illuminable' appeared. It is possible to express each of these concepts in the environment of calculus. We will represent the first derivative of a function $f(x)$ with $f'(x)$. The second derivative, $f''(x)$, is also of interest.

- A differentiable function f is increasing on an interval I if and only if $f'(x) \geq 0$ for all $x \in I$.

This is intuitive. The derivative measures the rate of increase of a function, which we require to be non-negative. Convex functions have an increasing gradient, so we require the second derivative to be positive.

- A differentiable function f is convex on an interval I if and only if $f''(x) \geq 0$ for all $x \in I$.

The properties 'decreasing' and 'concave' are similarly defined, but with the '$\geq$' operator reversed in direction.

14. Prove that $e^{2x} + e^{2y} \geq 2\, e^{x+y}$ for all $x, y \in \mathbb{R}$.

So far, we have considered calculus in one variable. Nevertheless, it is possible to delve into the realms of *multivariate calculus*. The main approach is to consider the *partial derivative* of a function with respect to a variable. To do this, we allow one variable to vary and force the others to remain constant. For example, $z = y^2 + 2xy$ has the partial derivatives $\frac{\partial z}{\partial x} = 2y$ and $\frac{\partial z}{\partial y} = 2y + 2x$.

If we want to show that the value of a function $z = f(x, y)$ increases as we move parallel to the x-axis, we need to show that $\frac{\partial z}{\partial x}$ is always non-negative. To investigate how it changes as we move parallel to the vector (3, 2), we are interested in $3\frac{\partial z}{\partial x} + 2\frac{\partial z}{\partial y}$.

15. Let x, y and z be positive real numbers. Prove that $4(x+y+z)^3 > 27\left(x^2 y + y^2 z + z^2 x\right)$. **[BMO2 2010, Question 4]**

Warning: A stationary point is a point where all partial derivatives are zero. Be careful, however, as this could be a point of inflection or saddle point instead of a minimum or maximum. Also, calculus does not guarantee that a particular extremum is global; for example, $x^3 - 3x$ has a *local minimum* at $x = 1$, but still takes on arbitrarily low values. You should bear this in mind when attempting to tackle a problem using calculus, especially Lagrange multipliers. If you want to use calculus to locate an extremum of a function, it is invariably a good idea to sketch a graph of the function first. Unfortunately, your two-dimensional paper and three-dimensional imagination are insufficient when there are many variables.

Lagrange multipliers

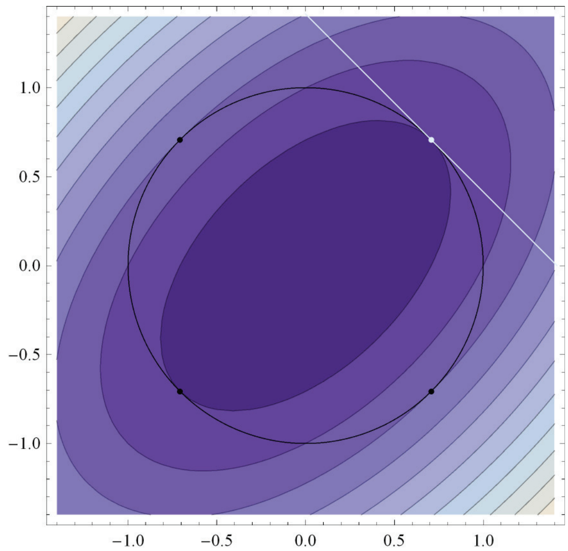
Suppose we have some additional constraints on the variables in an inequality. For example, we encountered a problem where we had to minimise $x^2 + y^2 + z^2$ subject to the constraint that $x^3 + y^3 + z^3 - 3xyz = 1$. One way of incorporating the side constraint is to *homogenise* the inequality. In that example, it would involve making all terms in $x^2 + y^2 + z^2$ of degree zero. In this case, it 'reduces' to the following problem, which is really quite horrible:

- Find the minimum value of $\frac{x^2+y^2+z^2}{(x^3+y^3+z^3-3\,x\,y\,z)^{\frac{2}{3}}}$, where $x,\ y,\ z \in \mathbb{R}$.

If we could guess that the minimum value is 1 (which is by no means obvious), then it is equivalent to proving that $\left(x^2 + y^2 + z^2\right)^3 \geq \left(x^3 + y^3 + z^3 - 3\,x\,y\,z\right)^2$. One could attempt to bash this degree-6 polynomial inequality with any combination of Muirhead, Schur and the $u\,v\,w$ method (as we shall do shortly), but it lacks a certain elegance.

A method that is more amenable to incorporating side constraints into problems is the use of *Lagrange multipliers*, which enable the application of calculus. If we want to minimise the value of f (which is a function of some variables) subject to the algebraic constraint $g = 0$ (where g is a function of those variables), then we introduce a new variable, λ. We consider the function $\Lambda = f + \lambda\,g$, and minimise it by locating its stationary points. We'll start with a simple non-trivial example in two variables:

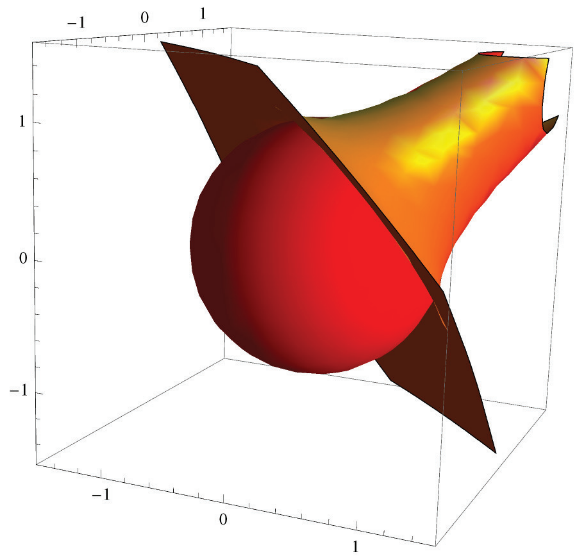
- Find the minimum and maximum of $f = x^2 + y^2 - x\,y$, subject to the constraint $x^2 + y^2 = 1$.



The contours of f are ellipses of the form $f = x^2 + y^2 - x\,y = k$, and we want to find the ones that touch the circle $g = x^2 + y^2 - 1 = 0$. Let $\Lambda = f + \lambda\,g$. Consider a point of tangency, such as that highlighted in the diagram above. We imagine setting a new orthogonal coordinate system centred at this point, with an axis normal to the common tangent. Call this coordinate ω. The partial derivatives $\frac{\partial f}{\partial \omega}$ and $\frac{\partial g}{\partial \omega}$ are both non-zero, whereas the partial derivatives with respect to the other axes are all zero. Hence, if we let $\lambda = -\frac{\frac{\partial f}{\partial \omega}}{\frac{\partial g}{\partial \omega}}$, the partial derivatives of Λ with respect to all of the (new) axes are zero, so the partial derivatives are all zero. In other words, any extremal point of f on the curve $g = 0$ is also a stationary point of Λ. This method only works if $\frac{\partial g}{\partial \omega}$ is non-zero at the extremal points, so it is important to verify this before proceeding with the method of Lagrange multipliers. In this example, g is quadratic and only stationary at the origin, so we can safely apply the method.

- Find the stationary points of $\Lambda = x^2 + y^2 - x\,y + \lambda\left(x^2 + y^2 - 1\right)$.

Equating $\frac{\partial \Lambda}{\partial x} = 0$ and $\frac{\partial \Lambda}{\partial y} = 0$, we have the equations $2\,x - y + 2\,\lambda\,x = 0$ and $2\,y - x + 2\,\lambda\,y = 0$, which simplify to $2\,(\lambda + 1) = \frac{y}{x} = \frac{x}{y}$. Hence, $y^2 = x^2$ and thus $x = \pm\,y$, from whence we obtain all four tangency points: $\left(\pm\frac{1}{\sqrt{2}},\ \pm\frac{1}{\sqrt{2}}\right)$. They correspond to the maximum value $f = 3$ and minimum value $f = 1$.



We shall now contemplate the original problem. As shown above, let $f = x^2 + y^2 + z^2$ and $g = x^3 + y^3 + z^3 - 3\,x\,y\,z - 1$. This simplification is more appetising than the previous attempt at homogenising the problem.

- Find the stationary points of $\Lambda = x^2 + y^2 + z^2 + \lambda\left(x^3 + y^3 + z^3 - 3\,x\,y\,z - 1\right)$.

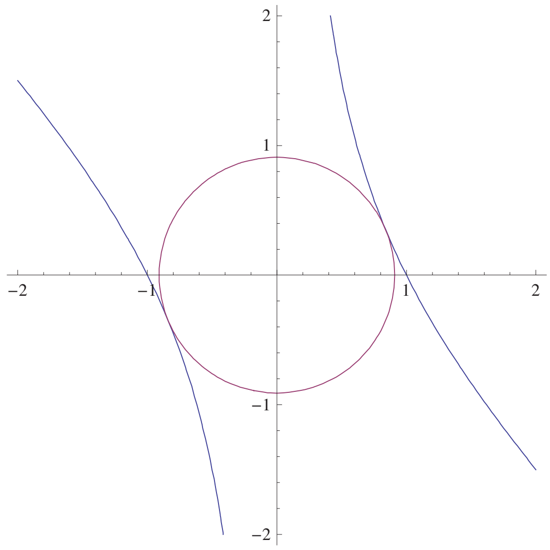
Differentiating it with respect to x gives the partial derivative $\frac{\partial\Lambda}{\partial x} = 2\,x + 3\,\lambda\,x^2 - 3\,\lambda\,y\,z$, which we wish to equate to zero. Similarly, by differentiating with respect to y and z, we obtain two more equations. (The final equation, $\frac{\partial\Lambda}{\partial\lambda} = 0$, is precisely the original side constraint, $x^3 + y^3 + z^3 - 3\,x\,y\,z = 1$.)

More interestingly, we can multiply $2\,x + 3\,\lambda\,x^2 - 3\,\lambda\,y\,z = 0$ by x to result in the cubic equation $3\,\lambda\,x^3 + 2\,x^2 = 3\,\lambda\,x\,y\,z$. Hence, x, y and z are all solutions of the equation $3\,\lambda\,x^3 + 2\,x^2 = k$, where $k = 3\,\lambda\,x\,y\,z$. Either x, y, z are the three distinct roots, or two of them are equal. In the former case, we have $x\,y + y\,z + z\,x = 0$ by Vieta's formulas, resulting in the equation $x^3 + y^3 + z^3 - 3\,x\,y\,z = (x + y + z)^3 = \left(x^2 + y^2 + z^2\right)^{\frac{3}{2}}$, and thus $x^2 + y^2 + z^2 = 1$ and we are done. In the other case, we can assume without loss of generality that $y = z$ and thus eliminate a variable.

- Find the stationary points of $\Lambda = x^2 + 2\,y^2 + \lambda\left(x^3 + 2\,y^3 - 3\,x\,y^2 - 1\right)$.

We obtain $\frac{\partial\Lambda}{\partial y} = 4\,y + 6\,\lambda\,y^2 - 6\,\lambda\,x\,y = 0$, which has solutions $y = 0$ and $\lambda(x - y) = \frac{2}{3}$. The former case clearly results in $(x,\ y,\ z) = (1,\ 0,\ 0)$, again giving a minimum of $x^2 + y^2 + z^2 = 1$. The other solution is more intricate. By considering the other partial derivative, $\frac{\partial\Lambda}{\partial x} = 2\,x + 3\,\lambda\,x^2 - 3\,\lambda\,y^2 = 2\,x + 3\,\lambda(x + y)\,(x - y) = 0$, we get $2\,x + 2\,(x + y) = 0$. This gives $2\,x = -y$, which can be substituted back into the original equation to give $-27\,x^3 = 1$, or $(x,\ y,\ z) = \left(-\frac{1}{3},\ \frac{2}{3},\ \frac{2}{3}\right)$. This also attains the value of $x^2 + y^2 + z^2 = 1$. We now just need to choose the minimum value of $x^2 + y^2 + z^2$, which is 1.

16. Find the distance from the closest points on the hyperbola $x\,y + x^2 = 1$ to the origin $O = (0,\ 0)$.



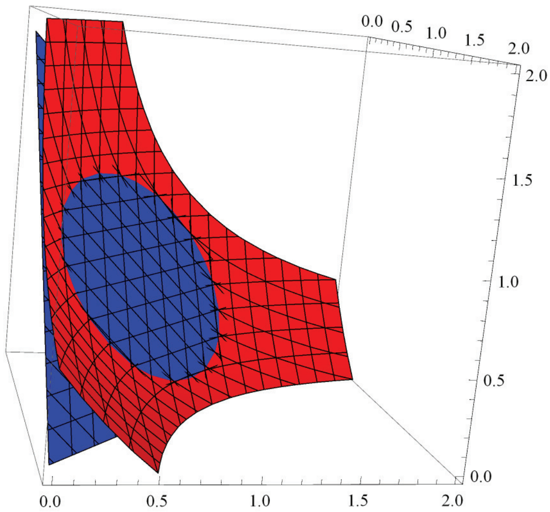
The *u v w* method

Symmetric polynomial inequalities in three positive real variables have frequently appeared in olympiads. The '$u\,v\,w$ method' uses the idea of expressing these as polynomials in the ESPs.

- $3\,u = x + y + z$
- $3\,v^2 = x\,y + y\,z + z\,x$
- $w^3 = x\,y\,z$

17. If $x,\ y,\ z \in \mathbb{R}$, prove that $\left(x^2 + y^2 + z^2\right)^3 \geq \left(x^3 + y^3 + z^3 - 3\,x\,y\,z\right)^2$. When does equality occur?

The full power of the $u\,v\,w$ method is realised when we require that $x,\ y,\ z \geq 0$. By the AM-GM inequality, $u \geq v \geq w$ with equality if and only if $x = y = z$. This leads to an approach for tackling all three-variable symmetric polynomial inequalities of reasonably low degree.



The blue plane $x + y + z = 3\,u$ intersects the red two-sheeted hyperboloid $x\,y + y\,z + z\,x = 3\,v^2$ in a conic. It is the intersection of the blue plane with the sphere with equation $x^2 + y^2 + z^2 = 9\,u^2 - 6\,v^2$, so is a circle. If we fix u and v, we can 'move' around the circumference of the circle and examine how w varies. This can be accomplished by the method of Lagrange multipliers.

18. Show that the stationary points of $\Lambda = x\,y\,z + \lambda\left(x^2 + y^2 + z^2 + 6\,v^2 - 9\,u^2\right) + \mu(x + y + z - 3\,u)$ occur only where two of the variables are equal.

If the circle intersects the planes $x = 0$, $y = 0$ and $z = 0$, however, we must also account for the 'boundary case' where one of the variables is zero.

- If we want to find the maximum or minimum values of w^3 for some fixed u and v^2, it suffices to only check the cases where $x = 0$ or $y = z$.

Now let's suppose we are trying to prove a symmetric polynomial inequality where the degree of the greatest term is 8. It can be expressed as the inequality $F\,w^6 + 2\,G\,w^3 + H \geq 0$, where F, G, H are functions of u and v^2. This is a quadratic in w^3, so its extreme values occur when either w^3 is minimised, maximised, or reaches the stationary point. By differentiating the above expression with respect to w^3, this occurs when $F\,w + G = 0$, *i.e.* $w = -\frac{G}{F}$.

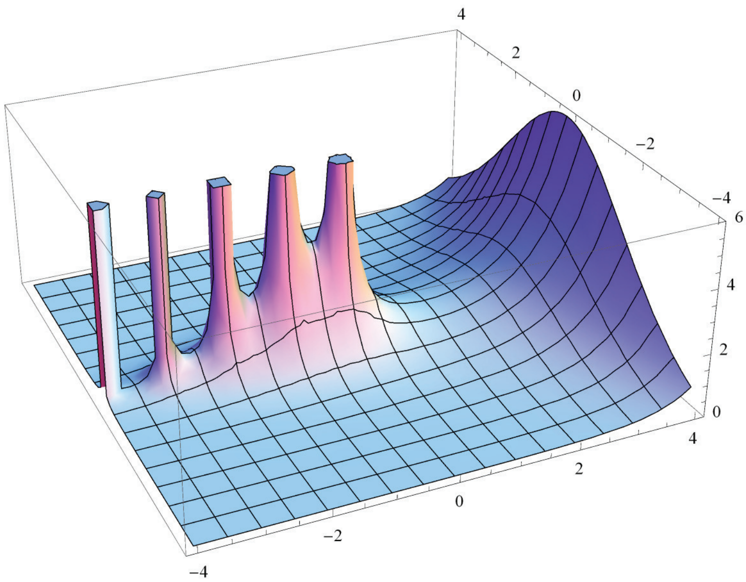
- To prove the inequality $F\,w^6 + 2\,G\,w^3 + H \geq 0$ (which is an arbitrary symmetric polynomial of degree $d \leq 8$ in three variables), where F, G, H are polynomials in u and v^2, it suffices only to check that it holds under each of the following three cases:
  - One variable is zero (without loss of generality, $x = 0$);
  - Two variables are equal (without loss of generality, $y = z$);
  - $F\,w + G = 0$. (only relevant where $d \geq 6$). **[Generalised Tejs' corollary]**

$F\,w + G = 0$ is a degree-$(d - 3)$ symmetric polynomial equation, where d is the degree of the inequality. In some problems, you may be sufficiently fortunate to find that equality can never occur, for instance if $F\,w + G > 0$ in all cases. Since this is a degree-5 inequality, it can be itself verified using Tejs' corollary.

Gamma function

The function $f(x) = 2^x$ can be defined on the positive integers by the product $\underbrace{2\times2\times\ldots\times2}_{n\text{ times}}$. If we want to extend this function to the reals and complex numbers, we can do so by using the recurrence $f(x+1) = 2\,f(x)$. This gives an uncountably infinite number of possible contenders. If we insist that the function is continuous, differentiable and is 'logarithmically convex', then there is only one possible function: $f(x) = \exp(x\log(2))$, where $\exp(x) = e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots$ and $\log(x)$ is its inverse.

Euler did the same for the factorial function. The Gamma function is defined by $\Gamma(x) = (x-1)!$ for $x \in \mathbb{N}$, and more generally over the positive complex numbers with positive real part by the convergent integral $\Gamma(x) = \int_0^\infty e^{-t}\,t^{x-1}\,dt$. For complex numbers with negative real part, we can extrapolate using the recurrence $\Gamma(x+1) = x\,\Gamma(x)$. For example, it is known that $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$, so $\Gamma\left(-\frac{1}{2}\right) = -2\sqrt{\pi}$ and $\Gamma\left(-\frac{3}{2}\right) = \frac{4}{3}\sqrt{\pi}$.



A plot of $|\Gamma(z)|$ is shown above for complex values of z. Observe that for non-positive integers, the function is undefined.

From the integral definition of the Gamma function, it is straightforward to establish this identity:

- $\frac{1}{A^x} = \frac{1}{\Gamma(x)} \int_0^\infty e^{-A t}\, t^{x-1}\, dt$. **[Identity involving the Gamma function]**

If we want to show that $\frac{a}{A^x} + \frac{b}{B^x} + \frac{c}{C^x} \geq 0$, we can convert it to the equivalent inequality $\frac{1}{\Gamma(x)} \int_0^\infty \left(a\, e^{-A t} + b\, e^{-B t} + c\, e^{-C t}\right) t^{x-1}\, dt \geq 0$. If x is positive and $a\, e^{-A t} + b\, e^{-B t} + c\, e^{-C t} \geq 0$, the integrand and integral are therefore also non-negative.

19. Prove that $\sum_{i=1}^{n} \sum_{j=1}^{n} \frac{a_i a_j}{(p_i+p_j)^c} \geq 0$, where $c,\ p_1,\ p_2,\ \ldots,\ p_n > 0$ and $a_1,\ a_2,\ \ldots,\ a_n \in \mathbb{R}$. **[KöMaL, Problem A493, November 2009]**

An interesting fact concerning the Gamma function is that the volume of a n-dimensional hypersphere of radius r is given by $\frac{\pi^{\frac{n}{2}} r^n}{\Gamma\left(\frac{n}{2}+1\right)}$. One can verify easily that this agrees with known formulae for the line segment, circle and sphere.

20. The E_8 lattice consists of points in $\mathbb{R}^8$ such that the coordinates are either all integers or all half-integers, and the sum of the coordinates is an even integer. Suppose we place (hyper)spheres of radius r, centred at each point in E_8. What is the maximum value of r such that the spheres are disjoint, and what is the density of the resulting sphere packing?

Solutions

1. This follows from the non-negativity of $(x-y)^2+(y-z)^2+(z-x)^2$.

2. By induction on the number of variables, the barycentre $B=(\alpha_1 x_1+\ldots+\alpha_n x_n,\ \alpha_1 f(x_1)+\ldots+\alpha_n f(x_n))$ must lie in the convex hull of the points $\{P_i=(x_i,\ f(x_i))\}$. As every point on the perimeter of the convex hull lies above the curve by the definition of convexity, so too must every point in the interior of the convex hull, including the barycentre.

3. This is the special case of Jensen's inequality where $f(x)=e^x$.

4. We can replace a and b with $\frac{a}{b}$ and 1, respectively, without altering anything, and thus assume without loss of generality that $b=1$. Applying Jensen's inequality to $f(x)=x^a$ gives the desired result.

5. (3, 0, 0) majorises (2, 1, 0), so this follows from Muirhead's inequality.

6. Without loss of generality, re-define the interval so that $\alpha_1=0$, $\alpha_2=1$. We then need to prove that $f(x_2)+f(x_1)\geq f(\beta_1 x_1+\beta_2 x_2)+f(\beta_1 x_2+\beta_2 x_1)$. By the definition of convexity, we have $f(\beta_1 x_1+\beta_2 x_2)\leq\beta_1 f(x_1)+\beta_2 f(x_2)$ and $f(\beta_1 x_2+\beta_2 x_1)\leq\beta_1 f(x_2)+\beta_2 f(x_1)$. Adding these together yields the desired inequality.

7. Suppose we apply a q-move to α_i and α_{i+1}. Consider each sum of the form $f(\alpha_1 x_{\sigma(1)}+\ldots+\alpha_i x_{\sigma(i)}+\alpha_{i+1} x_{\sigma(i+1)}+\ldots+\alpha_n x_{\sigma(n)})+$. Note that this is equal to

$$f(\alpha_1 x_{\sigma(1)}+\ldots+\alpha_i x_{\sigma(i+1)}+\alpha_{i+1} x_{\sigma(i)}+\ldots+\alpha_n x_{\sigma(n)})$$

$g(\alpha_i x_{\sigma(i)}+\alpha_{i+1} x_{\sigma(i+1)})+g(\alpha_i x_{\sigma(i+1)}+\alpha_{i+1} x_{\sigma(i)})$ for some convex function $g(w)=f(w+k)$. The previous theorem tells us that this cannot increase when α_i and α_{i+1} are replaced with β_i and β_{i+1}. Apply this principle to all $\frac{n!}{2}$ pairs of terms.

8. This is a corollary of the energy minimisation lemma and the previous question.

9. Let $a-b=d$ and $b-c=e$. Then the inequality becomes $x^2\,d(d+e)-y^2\,d\,e+z^2\,e(d+e)\geq 0$. Rearranging, we obtain the equivalent $(x\,d-z\,e)^2+\left((x+z)^2-y^2\right)d\,e\geq 0$. This is clearly true if $(x+z)^2\geq y^2$.

10. $x^2+z^2\geq y^2$ is a weaker condition than $x+z\geq y$, so the result follows from the previous question.

11. Assume without loss of generality that $a\geq b\geq c$, and let $d=a-b$ and $e=b-c$. For any non-negative monotonic function f, we have $f(a)+f(c)\geq f(b)$; hence, this must be true of any sum of non-negative monotonic functions. The problem reduces to showing that $f(a)\,g(h(d)\,h(d+e))+f(c)\,g(h(e)\,h(d+e))\geq f(b)\,g(h(e)\,h(d))$. As g and h are increasing, we have $f(a)\,g(h(d)\,h(d+e))+f(c)\,g(h(e)\,h(d+e))\geq(f(a)+f(c))\,g(h(e)\,h(d))$, which in turn must be greater than $f(b)\,g(h(e)\,h(d))$, as $f(a)+f(c)\geq f(b)$.

12. Obviously, the worst-case scenario is when all variables are positive. Expanding the Schur inequality $x^2\left(x^2-y^2\right)\left(y^2-z^2\right)+y^2\left(y^2-z^2\right)\left(y^2-x^2\right)+z^2\left(z^2-x^2\right)\left(z^2-y^2\right)\geq 0$ gives the variant $x^6+y^6+z^6+3\,x^2\,y^2\,z^2\geq x^4\,y^2+x^2\,y^4+y^4\,z^2+y^2\,z^4+z^4\,x^2+z^2\,x^4$. The other inequality, $x^4\,y^2+x^2\,y^4+y^4\,z^2+y^2\,z^4+z^4\,x^2+z^2\,x^4\geq x^3\,y^3+y^3\,z^3+z^3\,x^3$, is a simple application of Muirhead.

13. Again, assume without loss of generality that $a\geq b\geq c$, and let $d=a-b$ and $e=b-c$. As h is positive-illuminable, $h(m+n)\geq h(m)+h(n)$ for all $m,n\in\mathbb{R}^+$. As g and h are increasing and odd, we have $f(a)^2\,g(h(d)\,h(d+e))+f(c)^2\,g(h(e)\,h(d+e))\geq f(a)^2\,g\left(h(d)^2+h(d)\,h(e)\right)+f(c)^2\,g\left(h(e)^2+h(d)\,h(e)\right)$. Hence, we can reduce this, effectively, to the case where h is the identity function. As g is positive-illuminable, we can express $g(w) = G(w)^2\, w$ for all $w \in \mathbb{R}^+$, where G is an increasing function. Now, we let $x = f(a)\, G(h(a-b)\, h(a-c))$ and define y and z similarly. We have $x + z \geq y$ by the same argument as in the proof of the weak generalised Schur inequality. The result then follows from the strong 6-variable Schur.

14. We let $f(x) = e^{2x}$. Differentiating this twice gives $4\, e^{2x}$, which is positive-definite. Hence, f is convex and we can apply Jensen's inequality to show that $\frac{1}{2}\,(f(x) + f(y)) \geq f\left(\frac{x+y}{2}\right)$.

15. Let $w = f(x,\ y,\ z) = 4\,(x+y+z)^3 - 27\left(x^2\, y + y^2\, z + z^2\, x\right)$. Differentiate with respect to x to give the partial derivative $\frac{\partial w}{\partial x} = 12\,(x+y+z)^2 - 27\left(2\,x\,y + z^2\right)$. The cyclic sum is $\frac{\partial w}{\partial x} + \frac{\partial w}{\partial y} + \frac{\partial w}{\partial z} = 36\,(x+y+z)^2 - 27\left(x^2 + y^2 + z^2 + 2\,x\,y + 2\,y\,z + 2\,z\,x\right) = 9\,(x+y+z)^2$. This is obviously positive, so the function increases as we move parallel to the vector $(1,\ 1,\ 1)$. Hence, we have $f(x+h,\ y+h,\ z+h) > f(x,\ y,\ z)$ for all $h > 0$. Assume without loss of generality that $x \leq y$ and $x \leq z$. Using the previous statement, $f(x,\ y,\ z) > f(0,\ y-x,\ y-z)$. To prove the strict inequality in general, therefore, we need only prove the weak inequality when one of the variables is zero. We have reduced the problem to showing that $4\,(y+z)^3 \geq 27\, y^2\, z$. This is evident from the factorisation $4\,(y+z)^3 - 27\, y^2\, z = (4\,y + z)\,(y - 2\,z)^2$.

16. We wish to minimise $x^2 + y^2$ subject to the constraint $x\,y + x^2 = 1$. We use Lagrange multipliers to obtain $\Lambda = x^2 + y^2 - \lambda\left(x\,y + x^2 - 1\right)$. We equate each of its partial derivatives, $2\,x - \lambda\,y - 2\,\lambda\,x$ and $2\,y - \lambda\,x$, to zero. The latter gives us the value of λ, namely $\frac{2\,y}{x}$, so we can substitute it into the other equation and obtain $2\,x - 2\,\frac{y^2}{x} - 4\,y = 0$. We can multiply throughout by $\frac{1}{2}\,x$ to give the quadratic $x^2 - y^2 - 2\,x\,y = 0$, or $\left(\frac{x}{y}\right)^2 - 2\left(\frac{x}{y}\right) - 1 = 0$. The Babylonian formula gives us $\frac{x}{y} = 1 \pm \sqrt{2}$. It is sensible to draw a graph of the hyperbola to confirm that the root we are looking for is actually $\frac{x}{y} = 1 + \sqrt{2}$. Hence, $x = \left(1 + \sqrt{2}\right) y$ and $y = \left(\sqrt{2} - 1\right) x$. Substituting this into the equation of the hyperbola gives $x^2 = \frac{1}{\sqrt{2}}$. Similarly, we have $y^2 = \left(\sqrt{2} - 1\right)^2 x^2 = \left(3 - 2\sqrt{2}\right) x^2$, giving $x^2 + y^2 = \frac{\left(4 - 2\sqrt{2}\right)}{\sqrt{2}} = 2\sqrt{2} - 2$. The distance is the square-root of that, namely $\sqrt{2\sqrt{2} - 2}$.

17. The inequality $\left(x^2 + y^2 + z^2\right)^3 \geq \left(x^3 + y^3 + z^3 - 3\,x\,y\,z\right)^2$ can be expressed in the $u\,v\,w$ notation as $\left(9\,u^2 - 6\,v^2\right)^3 \geq \left((3\,u)\left(9\,u^2 - 9\,v^2\right)\right)^2$. We can divide throughout by 3^6, giving the equivalent inequality $\left(u^2 - \frac{2}{3}\,v^2\right)^3 \geq \left(u^3 - v^2\,u\right)^2$, which expands to $u^6 - 2\,u^4\,v^2 + \frac{4}{3}\,u^2\,v^4 - \frac{8}{27}\,v^6 \geq u^6 - 2\,u^4\,v^2 + u^2\,v^4$. Observe that the first two terms on each side of the equation cancel, so we wish to prove that $\frac{4}{3}\,u^2\,v^4 - \frac{8}{27}\,v^6 \geq u^2\,v^4$, or $\frac{1}{3}\,u^2\,v^4 \geq \frac{8}{27}\,v^6$. We can neatly divide by $\frac{1}{27}\,v^4$ (which must be positive, since $9\,v^4 = (x\,y + y\,z + z\,x)^2$), giving $9\,u^2 \geq 8\,v^2$. As u^2 is necessarily positive and greater in magnitude than v^2, this is true. Equality occurs when $v^4 = 0$, *i.e.* $x\,y + y\,z + z\,x = 0$.

18. Firstly, obtain the partial derivative $\frac{\partial \Lambda}{\partial x} = y\,z + 2\,\lambda\,x + \mu = 0$. We can multiply throughout by x to get $2\,\lambda\,x^2 + \mu\,x - w = 0$, where $w = x\,y\,z$. As $x,\ y,\ z$ are all roots of this quadratic equation, which only has two roots, at least two must be identical.

19. We note that this is equivalent to $\frac{1}{\Gamma(c)} \int_0^\infty t^{c-1} \sum_{i=1}^n \sum_{j=1}^n a_i\, a_j\, e^{-(p_i + p_j)\,t}\, dt \geq 0$. Observe that this simplifies to $\frac{1}{\Gamma(c)} \int_0^\infty t^{c-1} \left(\sum_{i=1}^n a_i\, a_j\, e^{-p_i\,t}\right)^2 dt \geq 0$, which is necessarily true.

20. The points in the lattice (regarded as vectors) clearly form a group under addition, so we need only calculate the minimum distance from the zero vector to another vector $\underline{a}$. If the coordinates of $\underline{a}$ are all half-integers, the minimum norm (squared length) of $\underline{a}$ is $8\left(\frac{1}{2}\right)^2 = 2$. Similarly, the closest integer point in the lattice is $(1, 1, 0, 0, 0, 0, 0, 0)$, with a norm of 2. Hence, the maximum value of r is $\frac{1}{2}\sqrt{2}$, so the volume of each sphere is $\frac{\pi^4 r^8}{4!} = \frac{\pi^4}{384}$. We now need to determine the number of lattice points per unit volume. The points in $\mathbb{Z}^8$ and $\left(\mathbb{Z} + \frac{1}{2}\right)^8$ each have one point per unit volume, so $\mathbb{Z}^8 \cup \left(\mathbb{Z} + \frac{1}{2}\right)^8$ has two points per unit volume. E_8 comprises half of these points (those with an even sum of coordinates), so it has one point per unit volume. Hence, the sphere packing has a density of $\frac{\pi^4}{384}$.