A Convex Region in a vector space is a region which, together with any two points in that region, includes all of the straight line segment joining them. For example, interiors of elipses and triangles and parallelograms are convex regions in the plane, but a star or annulus is not.

The Convex Hull of a set $\{ z_1, z_2, z_3, ... \}$ of given points or vectors is the smallest convex region that contains all of them. The convex hull can be shown to be the set of all …

$$\textbf{Positively Weighted Averages}\quad \frac{w_1z_1 + w_2z_2 + w_3z_3 + \dots + w_nz_n}{w_1 + w_2 + w_3 + \dots + w_n}$$

of finite subsets, here $\{ z_1, z_2, z_3, ..., z_n \}$, of those points with positive **Weights** $w_1 > 0$, $w_2 > 0$, $w_3 > 0$, ..., $w_n > 0$. In fact, Caratheodory's theorem says that the convex hull is the union of all simplices whose vertices are chosen from the given point set, and every such simplex can easily be shown to consist of the set of positively weighted averages of its vertices. In the plane simplices are triangles; in 3-space simplices are tetrahedra; in N-space a simplex is the convex hull of N+1 points not lying in a (hyper)plane of dimension less than N. Usually $n > N$.

Another interpretation of the foregoing positively weighted average is as the **Center of Mass** or **Center of Gravity** of a collection of positive masses $w_1, w_2, w_3, ..., w_n$ positioned respectively at the points $z_1, z_2, z_3, ..., z_n$. Note that the center of mass of positive masses lies strictly inside their convex hull when $n > 2$, and strictly between them when $n = 2$.

All the foregoing assertions above about convex regions seem too obvious for plane regions to need explicit proofs here, but proofs for convex regions in spaces of arbitrarily high dimensions may be unobvious. (See texts by R.V. Benson, H.G. Egglestone, S.R. Lay, F.A. Valentine, … .)

A real function $U(x)$ of the vector argument x is called a **Convex Function** when its graph is the lower boundary of a convex region. Consequently every **Secant** (a straight line that crosses a graph at least twice) lies above the graph of $U(x)$ between every two crossings, so there can be only two crossings. Moreover the graph of $U(x)$ lies above its every **Support-Line**, which is a straight line, like a tangent, that touches the graph somewhere without crossing it nor running above it. The letters U and V without serifs are examples of graphs of convex functions; the letter V has no tangent at its sharp bottom vertex, but it has infinitely many support-lines there. A round bowl, a satellite-dish, and an ice-cream cone can also be examples of convex graphs.

A convex function's domain must be convex; do you see why? Every convex function must be continuous in the interior of its domain (this isn't obvious), and can be proved to be differentiable almost everywhere, and one-sidedly differentiable once everywhere inside its domain, but not necessarily twice differentiable anywhere; however, the second derivative must be *nonnegative definite* wherever it exists. For instance, if x is a scalar this means $U''(x) \ge 0$.

What this means for vector arguments x is best explained with the aid of a *Taylor Series*:

$$U(z+h) = U(z) + U'(z)\cdot h + (U''(z) + Ø(z, h))\cdot h\cdot h/2\,, \quad \text{where } Ø(z, h) \to 0 \text{ as } h \to o\,,$$

wherever the second derivative $U''(z)$ exists. Here U'' is a *bilinear* operator: $U''\cdot p\cdot b$ is a real-valued functional linear in each of the vectors p and b separately. If we think of $U(x)$ as a function of the components $\xi_1, \xi_2, \xi_3, \dots$ of x, then $U''(x)\cdot p\cdot b = \sum_j \sum_k (\partial^2U(x)/\partial\xi_j\partial\xi_k)\, \pi_j\, ß_k = p^THb$ where $\pi_1, \pi_2, \pi_3, \dots$ and $ß_1, ß_2, ß_3, \dots$ are the components of p and of b respectively, and H is the Hessian matrix of second partial derivatives of U. A well-known theorem due to H.A. Schwarz asserts that $\partial^2U(x)/\partial\xi_j\partial\xi_k = \partial^2U(x)/\partial\xi_k\partial\xi_j$ provided either side exists and is continuous, as is normally the case; then $H = H^T$ is symmetric and $U''(x)\cdot p\cdot b = U''(x)\cdot b\cdot p$.

The convexity of U implies that $U(x) \ge U(z) + U'(z)\cdot(x–z)$ for every x in the domain of U, which says the graph of $U(x)$ lies above a (hyper)plane tangent at z; substituting this inequality into the Taylor series implies that $(x–z)^TH(x–z) \ge 0$ for all such x, especially those very near z, which implies that H and thus $U''(z)$ is nonnegative definite as asserted earlier. Conversely, if $U''(z)$ is nonnegative definite at every z inside its domain then $U(x)$ is convex, though to prove this we must replace $U'' + Ø$ by a weighted average of U'' on the segment from x to z.

Analogously, a real function $C(x)$ of a vector argument x is called **Concave** when its graph is the upper boundary of a convex region; and then $C''(x)$ must be nonpositive definite wherever $C''(x)$ exists. Only an *Affine-Linear* (or *Non-Homogeneous Linear*) function $c^T x + b$ can be both concave and convex everywhere on its domain.

Suppose now that $x_1, x_2, ..., x_n$ are *distinct* vector arguments in the domain of a convex function $U(x)$, and suppose $w_1, w_2, ..., w_n$ are all positive weights. Then the weighted average

$$x_0 := \frac{w_1x_1 + w_2x_2 + w_3x_3 + \ldots + w_nx_n}{w_1 + w_2 + w_3 + \ldots + w_n}$$

lies in the domain of U also. Now set $y_k := U(x_k)$ for $k = 0, 1, 2, \ldots, n$; **Jensen's Inequality** is this **Theorem**:

$$y_0 \le \ddot{y} := \frac{w_1y_1 + w_2y_2 + w_3y_3 + \ldots + w_ny_n}{w_1 + w_2 + w_3 + \ldots + w_n}$$

Its proof goes roughly as follows: Let $z_k = (x_k, y_k)$ for $k = 0, 1, 2, \ldots, n$; all these points lie on the graph of $U(x)$ which, as the lower boundary of its convex hull, also falls below or on the boundary of the convex hull of the points $z_1, z_2, ..., z_n$. In particular, z_0 lies below or on that last boundary, so z_0 lies directly below or on the positively weighted average

$$(x_0, \ddot{y}) = \frac{w_1z_1 + w_2z_2 + w_3z_3 + \ldots + w_nz_n}{w_1 + w_2 + w_3 + \ldots + w_n} .$$

This confirms Jensen's Inequality. Replacing a convex function $U(x)$ by a concave function $C(x)$ merely reverses the inequality after y_0 .

Jensen's Inequality becomes equality only when $n = 1$ or function U is affine-linear over at least the convex hull of the given arguments x_j ; can you see why? (It takes a while.)

Jensen's Inequality has many applications. An important one is the Inequality among the Arithmetic, Geometric and Harmonic Means:

Given n positive weights $w_1, w_2, ..., w_n$ and n positive numbers $x_1, x_2, ..., x_n$, define

$$A := \frac{w_1x_1 + w_2x_2 + w_3x_3 + \ldots + w_nx_n}{w_1 + w_2 + w_3 + \ldots + w_n} ,$$

$$G := (x_1^{w_1} \cdot x_2^{w_2} \cdot x_3^{w_3} \cdot \ldots \cdot z_n^{w_n})^{1/(w_1 + w_2 + w_3 + \ldots + w_n)} , \quad \text{and}$$

$$H := \frac{w_1 + w_2 + w_3 + \ldots + w_n}{\frac{w_1}{x_1} + \frac{w_2}{x_2} + \frac{w_3}{x_3} + \ldots + \frac{w_n}{x_n}} .$$

Here A is a Weighted Arithmetic Mean of the set $\{ x_1, x_2, ..., x_n \}$ and G a Weighted Geometric Mean with the same set of weights $\{ w_1, w_2, ..., w_n \}$, and H is a Weighted Harmonic Mean with the same weights.

Theorem: $H \le G \le A$, with equality only when $x_1 = x_2 = ... = x_n$.

To prove that $G \le A$ apply Jensen's Inequality to the concave function $\log(x)$ to deduce that $\log(G) \le \log(A)$. To prove $H \le G$ apply the inequality of the arithmetic and geometric means to the reciprocals of $x_1, x_2, ..., x_n$, G and H .

An elementary application of the inequalities among the arithmetic and geometric means yields a proof that $e = 2.718281828459\ldots$ exists as the limit of the increasing sequence $s_n := (1 + 1/n)^n$ and also of the decreasing sequence $S_n := 1/(1 - 1/n)^n$ as integer n approaches $+\infty$. First observe that $s_n < s_{n+1}$ because the geometric mean of $n+1$ numbers $\{ 1 + 1/n, 1 + 1/n, ..., 1 + 1/n, 1 \}$ is less than their arithmetic mean. Therefore s_n really does increase with n. Next observe that $1/S_n < 1/S_{n+1}$ because the $n+1$ numbers $\{ 1 - 1/n, 1 - 1/n, ..., 1 - 1/n, 1 \}$ have a smaller geometric mean than arithmetic mean; S_n really does decrease as n increases. Meanwhile $s_n < S_n$ because $s_n/S_n = (1 - 1/n^2)^n < 1$. This implies that every term in the increasing sequence s_n is less than every term in the decreasing sequence S_n, so both sequences must converge; s_n increases to some limit $\ddot{e}$ while S_n decreases to some limit $\hat{e}$ as $n \to +\infty$, and $\hat{e} \ge \ddot{e}$. Must $\hat{e} = \ddot{e}$? For all $n > 2$ consider

$$0 \le \hat{e} - \hat{e} < S_n - s_n = (1 - (1 - 1/n^2)^n)\cdot S_n = (1 - (1/S_{n\cdot n})^{1/n})\cdot S_n$$
$$< (1 - 1/S_4^{\,1/n})\cdot S_2 \to 0 \quad \text{as } n \to +\infty .$$

So $\hat{e} - \ddot{e} = 0$. Therefore both sequences converge to the same limit, called e, as claimed. (The sequence $\sqrt{(s_n S_n)} \to e$ faster, but still requires n to be huge to achieve appreciable accuracy. Can you see why roundoff may blight the computation of terms in any of the three sequences for very huge n unless it is restricted to powers of 10 on a calculator, or powers of 2 on a computer?)

. . . .

Another important application of Jensen's Inequality is a proof of …

Hölder's Inequality:

Suppose $q/(q-1) = p > 1$ (and so $p/(p-1) = q > 1$ too), and suppose all of $u_1, u_2, u_3, \ldots$ and $v_1, v_2, v_3, \ldots$ are positive. Then

$$\sum_j u_j\cdot v_j \le \sqrt[p]{(\textstyle\sum_j u_j^p)}\cdot \sqrt[q]{(\textstyle\sum_j v_j^q)} ,$$

with equality just when $u_1^p/v_1^q = u_2^p/v_2^q = u_3^p/v_3^q = \ldots$.

To prove this, apply Jensen's Inequality to the convex function x^q with weights $w_j = u_j^p$ and arguments $x_j = u_j\cdot v_j/w_j$. Note that the inequality stays true in the limits as $p \to 1$ or $\to +\infty$.

The special case $p = q = 2$ is called Cauchy's Inequality, and justifies calling the angle $\arccos(\mathbf{a}\bullet\mathbf{b}/(\|\mathbf{a}\|\cdot\|\mathbf{b}\|) = \arccos((\sum_j a_j\cdot b_j)/\sqrt{(\sum_j a_j^2 \cdot \sum_k b_k^2)})$ the unsigned angle between two column vectors $\mathbf{a}$ and $\mathbf{b}$ in Euclidean N-space. In non-Euclidean spaces generally no useful notion of angle need exist, but often the length of row-vectors $\mathbf{a}^T := [a_1, a_2, ..., a_N]$ is defined usefully as $\|\mathbf{a}^T\|_q := \sqrt[q]{(\sum_j |a_j|^q)}$ in conjunction with $\|\mathbf{b}\|_p := \sqrt[p]{(\sum_j |b_j|^p)}$ for column vectors $\mathbf{b} := [b_1, b_2, \ldots, b_N]^T$, and then " $\mathbf{a}^T\mathbf{b} \le \|\mathbf{a}^T\|_q\cdot\|\mathbf{b}\|_p$ " is the mnemonic for Hölder's inequality. These definitions of length $\|\ldots\|_{\ldots}$ make sense because they satisfy …

Minkowski's Triangle Inequality:

Suppose $p > 1$, and all of $u_1, u_2, u_3, ...$ and $v_1, v_2, v_3, ...$ are positive. Then

$$\sqrt[p]{(\textstyle\sum_j (u_j + v_j)^p)} \le \sqrt[p]{(\textstyle\sum_j u_j^p)} + \sqrt[p]{(\textstyle\sum_j v_j^p)} ,$$

with equality just when $u_1/v_1 = u_2/v_2 = u_3/v_3 = \ldots$.

To prove this, apply Jensen's Inequality again, now with $w_j = v_j^p$ and $x_j = (u_j/v_j)^p$ and the *concave* function $(1 + \sqrt[p]{x})^p$. Note again that the inequality remains true in the limits as $p \to 1$ or $\to +\infty$. The mnemonic for Minkowski's inequality is " $\|\mathbf{b} + \mathbf{c}\|_p \le \|\mathbf{b}\|_p + \|\mathbf{c}\|_p$ ".