Weierstraß-Institut
für Angewandte Analysis und Stochastik
Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint ISSN 0946 – 8633

A generalization of Lagrange's algebraic identity and connections with Jensen's inequality

Constantin P. Niculescu[1], Holger Stephan[2]

submitted: December 18, 2012

[1] University of Craiova
Department of Mathematics
Craiova 200585, Romania
E-Mail: cniculescu47@yahoo.com

[2] Weierstrass Institute
Mohrenstr. 39
10117 Berlin, Germany
E-Mail: holger.stephan@wias-berlin.de

No. 1756
Berlin 2012

2010 *Mathematics Subject Classification.* 26B25, 26D15.

Key words and phrases. Jensen's inequality, Lagrange's algebraic identity, convex function, slope function, barycenter of a measure .

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 20372-303
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Abstract

We discuss a generalization of Lagrange's algebraic identity that provides valuable insights into the nature of Jensen's inequality and of many other inequalities of convexity.

Joseph Louis Lagrange (1736-1813), who have made significant contributions to the development of analysis, variational calculus, mechanics and astronomy, is also well known for several identities that bear his name. Lagrange's algebraic identity asserts that *for every family of points* $z, x_1, ..., x_n$ *in the Euclidean space* $\mathbb{R}^N$ *and every family of real weights* $p_1, ..., p_n$ *with* $\sum_{k=1}^{n} p_k = 1$, *we have*

$$\sum_{k=1}^{n} p_k \|z - x_k\|^2 = \left\| z - \sum_{k=1}^{n} p_k x_k \right\|^2 + \sum_{1 \le i < j \le n} p_i p_j \|x_i - x_j\|^2. \tag{L}$$

The roots and various ramifications of this identity made the subject of a recent paper [4], where it was noticed that (L) contains as a special case the much more familiar identity,

$$\left(\sum_{i=1}^{n} a_i^2 \right) \left(\sum_{i=1}^{n} b_i^2 \right) = \left(\sum_{i=1}^{n} a_i b_i \right)^2 + \sum_{1 \le i < j \le n} (a_i b_j - a_j b_i)^2,$$

valid for all families a_i, b_i $(i = 1, \dots, n)$ of real numbers.
Lagrange's identity encompasses many other identities like the *parallelogram rule*,

$$2\|x\|^2 + 2\|y\|^2 = \|x+y\|^2 + \|x-y\|^2,$$

and its 2-dimensional analogue, the *Hlawka identity*,

$$\|x\|^2 + \|y\|^2 + \|z\|^2 + \|x+y+z\|^2 = \|x+y\|^2 + \|y+z\|^2 + \|z+x\|^2.$$

At first glance, Lagrange's algebraic identity seems limited to the translates of the square norm function $x \to ||x||^2$, but it can be easily extended to all functions. This will be done below.

1 The generalization of Lagrange's identity

Suppose that f is a real-valued function defined on a subset C of the Euclidean space $\mathbb{R}^N$. Then for every family of points $x_1, ..., x_n \in C$ and every family of nonzero real weights $p_1, ..., p_n$ such that $\sum_{i=1}^{n} p_i = 1$ and $x_G = \sum_{i=1}^{n} p_i x_i \in C \setminus \{x_1, ..., x_n\}$, the following extension of (L) takes place:

$$\sum_{i=1}^{n} p_i f(x_i) = f(x_G) + \sum_{1 \le i < j \le n} p_i p_j \langle s(x_i) - s(x_j), x_i - x_j \rangle, \tag{GL}$$

where

$$s(x) = \frac{f(x) - f(x_G)}{\|x - x_G\|} \cdot \frac{x - x_G}{\|x - x_G\|} \quad \text{for } x \in C \setminus \{x_G\}.$$

In the case of functions of one real variable, $s(x)$ is precisely the slope of the secant line joining the points $(x, f(x))$ and $(x_G, f(x_G))$.
The point x_G is the *equilibrium point* for the weighted family of points $(x_1, p_1), ..., (x_n, p_n)$. Notice that it coincides with the *center of gravity* of that family when all weights $p_1, ..., p_n$ are nonnegative.
Clearly, the presence of a pair (x_i, p_i) in the formula (GL) is effective only when $p_i \neq 0$. That's way we imposed the condition that all the weights are nonzero. The case where one of the

points, say x_n, coincides with x_G is also a case of degeneracy because $(x_1, p_1), ..., (x_n, p_n)$ can be replaced by $\left(x_1, \frac{p_1}{1-p_n}\right), ..., \left(x_{n-1}, \frac{p_{n-1}}{1-p_n}\right)$ with the same equilibrium point x_G and moreover

$$\sum_{i=1}^{n} p_i f(x_i) - f(x_G) = \sum_{i=1}^{n-1} \frac{p_i}{1-p_n} f(x_i) - f(x_G).$$

As shows the argument below, in the case of smooth functions this replacement can be avoided simply by defining the slope function at $x = x_G$ by the formula

$$s(x_G) = \nabla f(x_G).$$

Indeed, the proof of (GL) is based on the formulas giving the expression of these secant lines, more precisely, on the fact that

$$f(x_i) = f(x_G) + \langle s(x_i), x_i - x_G \rangle, \quad \text{for } i \in \{1, ..., n\}.$$

Then

$$\begin{aligned}
\sum_{i=1}^{n} p_i f(x_i) - f(x_G) &= \sum_{i=1}^{n} p_i \left(f(x_i) - f(x_G) \right) \\
&= \sum_{i=1}^{n} p_i \langle s(x_i), x_i - x_G \rangle \\
&= \sum_{i=1}^{n} p_i \left\langle s(x_i), \sum_{j=1}^{n} p_j x_i - \sum_{j=1}^{n} p_j x_j \right\rangle \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} p_i p_j \langle s(x_i), x_i - x_j \rangle \\
&= \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} p_i p_j \langle s(x_i) - s(x_j), x_i - x_j \rangle \\
&= \sum_{1 \le i < j \le n} p_i p_j \langle s(x_i) - s(x_j), x_i - x_j \rangle,
\end{aligned}$$

and the proof of (GL) is done.

We will next prove that the identity (GL) reduces to Lagrange's identity when f is the square norm function on the Euclidean space. Indeed, in this case the vectors $s(x_i)$ are given by the formula

$$\begin{aligned}
s(x_i) &= \frac{\|x_i\|^2 - \|x_G\|^2}{\|x_i - x_G\|} \cdot \frac{x_i - x_G}{\|x_i - x_G\|} \\
&= \frac{\langle x_i + x_G, x_i - x_G \rangle}{\|x_i - x_G\|} \cdot \frac{x_i - x_G}{\|x_i - x_G\|}.
\end{aligned}$$

Then

$$\begin{aligned}
&\langle s(x_i) - s(x_j), x_i - x_j \rangle \\
&\qquad = \langle x_i + x_G, x_i - x_G \rangle - \frac{\langle x_i + x_G, x_i - x_G \rangle}{\|x_i - x_G\|} \cdot \frac{\langle x_i - x_G, x_j - x_G \rangle}{\|x_i - x_G\|} \\
&\qquad - \frac{\langle x_j + x_G, x_j - x_G \rangle}{\|x_j - x_G\|} \cdot \frac{\langle x_j - x_G, x_i - x_G \rangle}{\|x_j - x_G\|} + \langle x_j + x_G, x_j - x_G \rangle \\
&= \|x_i\|^2 + \|x_j\|^2 - 2\|x_G\|^2 - \left(\frac{\|x_i\|^2 - \|x_G\|^2}{\|x_i - x_G\|^2} + \frac{\|x_j\|^2 - \|x_G\|^2}{\|x_j - x_G\|^2} \right) \langle x_i - x_G, x_j - x_G \rangle ,
\end{aligned}$$

so that

$$\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} p_i p_j \langle s(x_i) - s(x_j), x_i - x_j \rangle = \sum_{i=1}^{n} p_i \|x_i\|^2 - \|x_G\|^2 = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} p_i p_j \|x_i - x_j\|^2 .$$

Therefore, for $f(x) = \|x\|^2$, the identity (GL) reduces to a special case of Lagrange's algebraic identity (L), precisely, to the case where $z = 0$. However, due to the translation invariance of the Euclidean metric, this case is actually equivalent to (L) in full generality.

2 A probabilistic interpretation of Lagrange's generalized identity

From the probabilistic point of view, the identity (GL) provides a formula indicating how much the expectation $\mathcal{E}(f;\mu)$ of f, relative to a discrete probability measure $\mu = \sum_{i=1}^{n} p_i \delta_{x_i}$, differs from the value of f at the barycenter b_μ of μ. Recall that

$$b_\mu = \int_C x d\mu(x) = \sum_{i=1}^{n} p_i x_i,$$

that is, b_μ coincides with the center of gravity x_G.
When $f = \|\cdot\|^2$, we noticed above that

$$\begin{aligned}
\mathcal{E}(f;\mu) - f(b_\mu) &= \sum_{i=1}^{n} p_i f(x_i) - f(x_G) \\
&= \sum_{1 \le i < j \le n} p_i p_j \langle s(x_i) - s(x_j), x_i - x_j \rangle
\end{aligned}$$

equals the *variance* of the given family of points with respect to μ,

$$\sigma_\mu^2 = \sum_{1 \le i < j \le n} p_i p_j \|x_i - x_j\|^2 .$$

As we shall show in Example 1 below, one can obtain valuable estimates (from above and from below) of the variance by using appropriate identities (GL).

In the general case, the term

$$SV_{\mu}^{2}(f)=\sum_{1\leq i<j\leq n}p_{i}p_{j}\left\langle s(x_{i})-s(x_{j}),x_{i}-x_{j}\right\rangle$$

represents a measure of how far the slopes of f at $x_1,...,x_n$ are spread out and will be referred to as the *slope variance.* Using the slope variance one can put the identity (GL) in the following probabilistic form

$$\mathcal{E}(f;\mu)=f\left(b_{\mu}\right)+SV_{\mu}^{2}(f).$$

Clearly, the slope variance is linear and can take negative values (unlike the usual variance which is nonnegative). Indeed, in the case of the function $f(x)=\langle Ax,x\rangle$, $x\in\mathbb{R}^N$, attached to a symmetric matrix A with real coefficients, the slope variance corresponding to the discrete probability measure $\mu=\sum_{i=1}^{n}p_i\delta_{x_i}$, is

$$SV_{\mu}^{2}(f)=\sum_{1\leq i<j\leq n}p_{i}p_{j}\left\langle A(x_{i}-x_{j}),x_{i}-x_{j}\right\rangle .$$

Then the condition $SV_{\mu}^{2}(f)\geq 0$ for every μ is equivalent to $\langle Ax,x\rangle\geq 0$ for every $x\in\mathbb{R}^N$.

3 Several examples illustrating (GL)

Example 1. By applying the identity (GL) to the function $f(x)=1/x$ for a family of points $x_1,...,x_n$ contained in the interval $[m,M]\subset(0,\infty)$, and positive weights $p_1,...,p_n$ that sum to 1, we get a formula relating the (weighted) arithmetic mean $A=\sum_{i=1}^{n}p_ix_i$ of these points to their (weighted) harmonic mean $H=\left(\sum_{i=1}^{n}\frac{p_i}{x_i}\right)^{-1}$:

$$\frac{A}{H}-1=\sum_{1\leq i<j\leq n}p_{i}p_{j}\frac{(x_{i}-x_{j})^{2}}{x_{i}x_{j}}.$$

Thus the variance $\sigma_{\mu}^{2}=\sum_{1\leq i<j\leq n}p_ip_j(x_i-x_j)^2$ of the given family (with respect to the discrete probability measure $\mu=\sum_{i=1}^{n}p_i\delta_{x_i}$) verifies the two sided inequality

$$m^{2}\left(\frac{A}{H}-1\right)\leq\sigma_{\mu}^{2}\leq M^{2}\left(\frac{A}{H}-1\right).$$

As a consequence we infer the following estimate for the discrepancy between the harmonic mean and the arithmetic mean of a family of points as above

$$1+\frac{\sigma_{\mu}^{2}}{M^{2}}\leq\frac{A}{H}\leq 1+\frac{\sigma_{\mu}^{2}}{m^{2}}.$$

A better upper bound for σ^2, precisely

$$\sigma_{\mu}^{2}\leq\left(M-A\right)\left(A-m\right),$$

was found by Bhatia and Davis [2], but their result is also a consequence of the identity (GL) when applied to the function $f(x)=x^2$, $x\in[m,M]$, and to families of points with equal weights. Indeed, in this case the identity (GL) becomes

$$\sum_{i=1}^{n}p_{i}\left(M-x_{i}\right)\left(x_{i}-m\right)=\left(M-A\right)\left(A-m\right)-\sigma_{\mu}^{2}.$$

Example 2. We next consider the case of the identity (GL) when applied to the logarithmic function and to a family of positive points $x_1, ..., x_n$ with positive weights $p_1, ..., p_n$ that sum to 1. This case provides the following curious formula relating the geometric mean $G = \prod_{i=1}^n x_i^{p_i}$ to the arithmetic mean A :

$$G = A \cdot \prod_{1 \le i < j \le n} \left(\left(\frac{x_i}{A} \right)^{1/(x_i - A)} \left(\frac{A}{x_j} \right)^{1/(x_j - A)} \right)^{p_i p_j (x_i - x_j)}.$$

Assuming that not all points $x_1, ..., x_n$ are equal each other, we know that $G < A$, so in this case we get the inequality

$$\prod_{1 \le i < j \le n} \left(\left(\frac{x_i}{A} \right)^{1/(x_i - A)} \left(\frac{A}{x_j} \right)^{1/(x_j - A)} \right)^{p_i p_j (x_i - x_j)} < 1.$$

Example 3. Even more sophisticated appears the case of the entropy function $x \ln x$. Using the *identric mean* of two variables,

$$I(s,t) = \begin{cases} \frac{1}{e} \left(\frac{t^t}{s^s} \right)^{1/(t-s)} & \text{if } s \ne t \\ s & \text{if } s = t, \end{cases}$$

we can put the corresponding identity (GL) under the form

$$\prod_{i=1}^n \left(\frac{x_i}{A} \right)^{p_i x_i} = \prod_{1 \le i < j \le n} \left(\frac{I(x_i, A)}{I(x_j, A)} \right)^{p_i p_j (x_i - x_j)}.$$

Example 4. Our last example enriches the list of identities verified by the square norm function in the Euclidean space:

$$\begin{aligned} &6 \left(\|x_1\|^2 + \|x_2\|^2 + \|x_3\|^2 \right) + 2 \|x_1 + x_2 + x_3\|^2 \\ &\qquad = 3 \left(\|x_1 + x_2\|^2 + \|x_2 + x_3\|^2 + \|x_3 + x_1\|^2 \right) + \sum_{1 \le i < j \le 3} \|x_i - x_j\|^2 . \quad (N) \end{aligned}$$

The weighted version of this identity as well as its extension to families of more than 3 points is left to the reader as an exercise.

For the proof of (N), divide both sides by 18 and notice that in this form it can be derived from (GL) as follows:

$$\begin{aligned} &\frac{\|x_1\|^2 + \|x_2\|^2 + \|x_3\|^2}{3} + \left\| \frac{x_1 + x_2 + x_3}{3} \right\|^2 - \frac{2}{3} \left(\left\| \frac{x_1 + x_2}{2} \right\|^2 + \left\| \frac{x_2 + x_3}{2} \right\|^2 + \left\| \frac{x_3 + x_1}{2} \right\|^2 \right) \\ &\qquad = \frac{\|x_1\|^2 + \|x_2\|^2 + \|x_3\|^2}{3} - \left\| \frac{x_1 + x_2 + x_3}{3} \right\|^2 \\ &\qquad - 2 \left(\frac{\left\| \frac{x_1 + x_2}{2} \right\|^2 + \left\| \frac{x_2 + x_3}{2} \right\|^2 + \left\| \frac{x_3 + x_1}{2} \right\|^2}{3} - \left\| \frac{x_1 + x_2 + x_3}{3} \right\|^2 \right) = \frac{1}{18} \sum_{1 \le i < j \le 3} \|x_i - x_j\|^2 . \end{aligned}$$

4 The connection with convexity inequalities

Very close to our generalization of Lagrange's algebraic identity is Jensen's inequality for convex functions. Indeed, if f is a convex function on an interval $[a,b]$ then the slopes

$$s_c(x) = \frac{f(x) - f(c)}{x - c},$$

of the secant lines passing to an arbitrarily fixed point $(c, f(c))$, are nondecreasing on $[a,b]\backslash\{c\}$. See [6], Theorem 1.3.1, p. 20. As a consequence, all products $(s_c(x_i) - s_c(x_j))(x_i - x_j)$ are nonnegative, so from the identity (GL) we infer the discrete form of Jensen's inequality:

$$f(x_G) \leq \sum_{i=1}^{n} p_i f(x_i) \tag{J}$$

for every family of points $x_1, ..., x_n \in [a,b]$ and every family of nonnegative weights $p_1, ..., p_n$ such that $\sum_{i=1}^{n} p_i = 1$. The continuous form of this inequality can be established in the same manner via a suitable extension of (GL) that will be described in the next section.

Jensen's inequality (J) can work even for nonconvex functions provided the barycenters x_G are well placed. Indeed the fact that the slopes from some point x_G are nondecreasing is not characteristic to convex functions. This is discussed in the appendix, where the case of polynomials of 4th degree is presented.

Given a real-valued function f on an interval $[a,b]$ assumed to be bounded from below, we define its convex hull $\text{co}(f)$ as the supremum of all convex functions $h \leq f$. Of course, $f = \text{co}(f)$ when f is convex. The convex hull of a zigzag function is convex polygonal function.

If f meets $\text{co}(f)$ at an interior point x_G, then clearly f verifies Jensen's inequality (J) for every family of points $x_1, ..., x_n \in [a,b]$ and every family of nonnegative weights $p_1, ..., p_n$ such that $\sum_{i=1}^{n} p_i = 1$ and $\sum_{i=1}^{n} p_i x_i = x_G$. It turns out that this condition simply means that f has a support line at x_G that is,

$$f(x) \geq f(x_G) + \lambda(x - x_G),$$

for some $\lambda \in \mathbb{R}$. See [6], Lemma 1.5.1, p. 30.

We didn't exploit here the fact that the identity (GL) actually works for families of real weights rather than of positive weights. The interested reader will find a more systematic approach of Jensen's inequality for nonconvex functions and signed measures in [7].

It is worth mentioning that the identity (GL) touches many other aspects of convexity. For example, from (N) we immediately infer the inequality

$$\frac{\|x_1\|^2 + \|x_2\|^2 + \|x_3\|^2}{3} + \left\| \frac{x_1 + x_2 + x_3}{3} \right\|^2 \geq \frac{2}{3} \left(\left\| \frac{x_1 + x_2}{2} \right\|^2 + \left\| \frac{x_2 + x_3}{2} \right\|^2 + \left\| \frac{x_3 + x_1}{2} \right\|^2 \right),$$

which illustrates the phenomenon of $(2D)$-*convexity* as developed in [1]. This is related to Popoviciu's characterization of convexity in the case of real-valued functions defined on intervals: $f : I \to \mathbb{R}$ is convex if and only if it is continuous at the interior of I and verifies the inequality

$$\frac{f(x_1) + f(x_2) + f(x_3)}{3} + f\left(\frac{x_1 + x_2 + x_3}{3}\right) \geq \frac{2}{3}\left[f\left(\frac{x_1 + x_2}{2}\right) + f\left(\frac{x_2 + x_3}{2}\right) + f\left(\frac{x_3 + x_1}{2}\right)\right]$$

for every $x_1, x_2, x_3 \in I$. See [6], p.12.

5 Further extensions

Our generalization of Lagrange's identity has a continuous analogue in the framework of Borel measures μ supported by a Borel subset C of $\mathbb{R}^N$ that verify the conditions

$$\mu(C) = 1 \text{ and } b_\mu = \int_C x d\mu(x) \in C.$$

Then for every (μ-) integrable function $f : C \to \mathbb{R}$ the following analogue of the identity (GL) holds true:

$$\int_C f(x)d\mu(x) - f(b_\mu) = \frac{1}{2}\int_C \int_C \langle s(x) - s(y), x - y\rangle \, d\mu(x)d\mu(y),$$

The details are practically the same as in the discrete case with the only observation that the slope function $s(x)$ is defined by a formula similar to (S), where the role of x_G is taken by the barycenter b_μ.
The whole discussion above extends verbatim from the case of Euclidean spaces to that of inner product spaces. Surprisingly, an analogue of the identity (GL) can be established in the general setting of normed linear spaces, by using the so called *normalized duality mapping*. This mapping attaches to each element x in a normed linear space E a subset $J(x)$ of the dual space E' defined by

$$J(x) = \left\{x' \in E' : [x', x] = \|x\|^2 = \|x'\|^2\right\}.$$

Here $[x', x] = x'(x)$ represents the duality pairing of functionals in E' with elements x in E.
The fact that $J(x) \neq \emptyset$ is assured by the Hahn-Banach extension theorem. The duality mapping J is single-valued when E is a reflexive, in particular when E is one of the Lebesgue spaces $L^p(\mathbb{R})$ for $1 < p < \infty$. See [11], Theorem 16.1, p. 69. The usefulness of the normalized duality mapping in nonlinear analysis is nicely presented in [3].
Suppose now that f is a real-valued function defined on a subset C of a normed vector space E. Then, exactly as in the Euclidean case, for every family of points $x_1, ..., x_n \in C$ and every family of nonzero real weights $p_1, ..., p_n$ such that $\sum_{i=1}^n p_i = 1$ and $x_G = \sum_{i=1}^n p_i x_i \in C \setminus \{x_1, ..., x_n\}$, we have the identity

$$\sum_{i=1}^n p_i f(x_i) = f(x_G) + \sum_{1 \le i < j \le n} p_i p_j \left[s(x_i) - s(x_j), x_i - x_j\right].$$

Here the slope function is defined by the formula

$$s(x) = \frac{f(x) - f(x_G)}{\|x - x_G\|^2} \cdot \varphi_{x - x_G} \quad \text{for } x \neq x_G,$$

where φ_{x-x_G} is any functional in $J(x - x_G)$.
Last but not least, Lagrange's generalized identity (GL) holds not only for real-valued functions, but also for functions taking values in an arbitrary Banach space.

6 An open problem

While every nonnegative polynomial of a single variable can be expressed as a sum of squares (of polynomials), in the case of several variables this does not works. This led Hilbert to address the following problem:

(Hilbert's seventeenth problem). Given a multivariate polynomial that takes only nonnegative values over the reals, can it be represented as a sum of squares of rational functions?

In 1926, E. Artin solved this problem in the affirmative.
A notable example of nonnegative polynomial that cannot be represented as a sum of squares of polynomials was provided by Motzkin in 1966 (see [5], or [8], p. 46):

$$1 + x^4y^2 + x^2y^4 - 3x^2y^2 = \left(\frac{x^2y(x^2+y^2-2)}{x^2+y^2}\right)^2 + \left(\frac{xy^2(x^2+y^2-2)}{x^2+y^2}\right)^2 + \left(\frac{xy(x^2+y^2-2)}{x^2+y^2}\right)^2 + \left(\frac{x^2-y^2}{x^2+y^2}\right)^2. \quad (M)$$

The fact that

$$1 + x^4y^2 + x^2y^4 - 3x^2y^2 \geq 0 \quad \text{for all } x, y \in \mathbb{R}$$

can be established via the arithmetic mean - geometric mean inequality (which in turns is a consequence of Jensen's inequality).
Can (GL) explain Motzkin's identity or at least is it able to provide an illustration to Hilbert's seventeenth problem? At the moment we lack a clear answer.

Appendix. Jensen's inequality for nonconvex functions

In this Appendix, we derive Jensen's inequality via Lagrange's identity in detail and state an example, how Jensen's inequality can be used in nonconvex situations. On the way to Jensen's inequality we pass the Cauchy-Schwarz inequality and Chebyshev's inequality.
Given an interval $I \subset \mathbb{R}$, a sequences $(a_i)_{i=1}^n$, a set of non negative weights $(p_i)_{i=1}^n$ with $\sum_{i=1}^n p_i = 1$ and a convex function $f : I \longrightarrow \mathbb{R}$, Jensen's famous inequality reads

$$f(a_0) \leq \sum_{i=1}^{n} p_i f(a_i) \ , \tag{A.1}$$

where $a_0 = \sum_{i=1}^n p_i a_i$ is the average of the sequence $(a_i)_{i=1}^n$. Its proof can be found in any textbook on convex analysis (see, e.g., [6]). Actually, the case $n = 2$ is the definition of convexity. Thus, Jensen's inequality holds if and only if f in convex. But this is true, only if we allow the points $(a_i)_{i=1}^n$ to be anywhere on I.
The following proof of Jensen's inequality via Lagrange's identity and Chebyshev's inequality shows that Jensen's inequality is true even for nonconvex function, if the points a_i are chosen in a special manner. It turns out that this inequality is more a consequence of monotonicity than convexity.
Since Jensen's inequality is one of the most important inequalities in mathematics and physics (see, e.g., [10] for the role of Jensen's inequality as origin of the second law of thermodynamics), this can be useful for applying methods of convex analysis in some nonconvex situations.
Starting from Lagrange's identity, there are two ways for the derivation of inequalities. We set

$$L(x,y) = \sum_{k=1}^{n} p_k x_k^2 \cdot \sum_{k=1}^{n} p_k y_k^2 - \left(\sum_{k=1}^{n} p_k x_k y_k \right)^2 - \sum_{1 \leq i < j \leq n} p_i p_j (x_i y_j - x_j y_i)^2 \ .$$

Lagrange's identity reads $L(x, y) \equiv 0$. The last part has a defined sign. This leads to the Cauchy-Schwarz inequality

$$\sum_{k=1}^{n} p_k x_k^2 \cdot \sum_{k=1}^{n} p_k y_k^2 \leq \left(\sum_{k=1}^{n} p_k x_k y_k \right)^2 \ .$$

Now, we define

$$\begin{aligned} C(a,b) &= L\left(\frac{a+b}{2}, 1\right) - L\left(\frac{a+b}{2}, 1\right) = \\ &= \sum_{k=1}^{n} p_k a_k b_k - \left(\sum_{k=1}^{n} p_k a_k \right) \left(\sum_{k=1}^{n} p_k b_k \right) - \sum_{1 \leq i < j \leq n} p_i p_j (a_i - a_j)(b_i - b_j) \ . \end{aligned}$$

From $L(x, 1) \equiv 0$ follows $C(a, b) \equiv 0$. Thus,

$$\sum_{i=1}^{n} p_i a_i b_i - \sum_{i=1}^{n} p_i a_i \cdot \sum_{i=1}^{n} p_i b_i = \sum_{1 \leq i < j \leq n} p_i p_j (a_i - a_j)(b_i - b_j) \ . \tag{A.2}$$

In this identity, the right hand side is nonnegative, if (a_i) and (b_i) are equal-monotone sequences, for example, both are increasing ones. If so, (A.2) implies Chebyshev's inequality

$$\sum_{i=1}^{n} p_i a_i b_i \geq \sum_{i=1}^{n} p_i a_i \cdot \sum_{i=1}^{n} p_i b_i \ .$$

Setting $a_0 = \sum_{i=1}^{n} p_i a_i$, the barycenter or mean value of the sequence (a_i), we obtain from (A.2) the identity

$$\sum_{i=1}^{n} p_i b_i (a_i - a_0) = \sum_{1 \le i < j \le n} p_i p_j (a_i - a_j)(b_i - b_j) \ . \tag{A.3}$$

Given the sequence (a_i), an example for a sequence (b_i) is $b_i = s_{a_0}(a_i)$, a sequence defined by the slope function

$$s_b(x) = \frac{f(x) - f(b)}{x - b}$$

of a function $f : I \subset \mathbb{R} \longrightarrow \mathbb{R}$ in some fixed pivot b. If (a_i) is in increasing order and this slope is a increasing function in x for given $b = a_0$, (a_i) and (b_i) are equal-monotone sequences and we obtain from (A.3)

$$0 \le \sum_{i=1}^{n} p_i \frac{f(a_i) - f(a_0)}{a_i - a_0}(a_i - a_0) = \sum_{i=1}^{n} p_i f(a_i) - \sum_{i=1}^{n} p_i f(a_0) \ .$$

This is precisely Jensen's inequality (A.1), For the derivation we did not use convexity of f. Thus, Jensen's inequality holds not only for convex function as well known, but for any function with monotone slope for a fixed pivot.
$s_b(x)$ is the slope of the secant through the points in b and x. $s_b(x)$ is increasing, if looking from the point $(b, f(b))$ on the graph of $f(x)$, no point $(x, f(x))$ "lies in the shadow" of the graph. This can happen for some points b even if the function is not convex.
If $f(x)$ is convex, $s_b(x)$ is increasing for any point b. In this case, because of

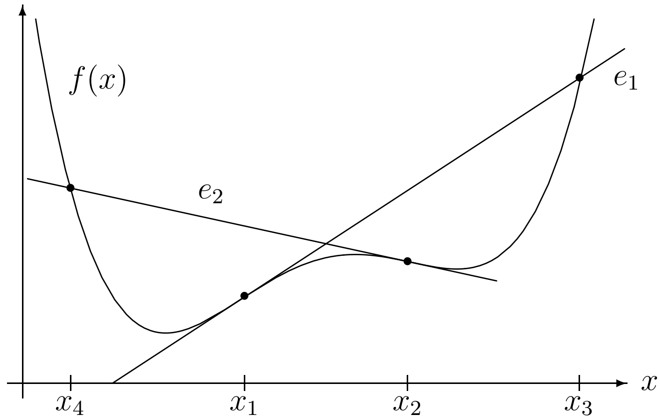
$$\begin{aligned} (a_i - a_j)(b_i - b_j) &= (a_i - a_j)\left(\frac{f(a_i) - f(a_0)}{a_i - a_0} - \frac{f(a_j) - f(a_0)}{a_j - a_0}\right) = \\ &= \frac{f(a_i)}{\frac{a_i - a_0}{a_i - a_j}} + \frac{f(a_j)}{\frac{a_0 - a_j}{a_i - a_j}} - \frac{f\left(\frac{a_0 - a_j}{a_i - a_j} a_i + \frac{a_i - a_0}{a_i - a_j} a_j\right)}{\frac{(a_i - a_0)(a_0 - a_j)}{(a_i - a_j)(a_i - a_j)}} \ge 0 \ , \end{aligned}$$

(a_i) and $b_i = s_{a_0}(a_i)$ are equal-monotone for any mean value a_0, since

$$0 \le \frac{1}{\alpha_2} f(x_1) + \frac{1}{\alpha_1} f(x_2) - \frac{1}{\alpha_1 \alpha_2} f(\alpha_1 x_1 + \alpha_2 x_2)$$

is true for any real numbers α_1 and α_2, if only x_1, x_2 and $\alpha_1 x_1 + \alpha_2 x_2$ are points in the interval of convexity I of f.

An amazing example, published in [9], are polynomials of forth order – typical non-convex functions. We consider such a function $f(x)$ and its turning points with x-coordinates x_1 and x_2. The turning point tangents e_1 and e_2 intersect the graph of $f(x)$ in points with x-coordinates x_3 and x_4. Now, Jensen's inequality holds for any sequence (a_i) with a mean value a_0 satisfying $a_0 \ge x_3$ or $a_0 \le x_4$.



References

[1] M. Bencze, C. P. Niculescu and Florin Popovici, Popoviciu's inequality for functions of several variables. *J. Math. Anal. Appl.*, **365** (2010), Issue 1, 399–409.

[2] R. Bhatia and C. Davis, A Better Bound on the Variance, *The American Mathematical Monthly*, **107** (2000), No. 4, 353-357.

[3] I. Cioranescu, *Geometry of Banach spaces*, *Duality Mappings and Nonlinear Problems*, Kluwer Academic Publishers, Dordrecht, 1990.

[4] M. Gidea and C. P. Niculescu, A Brief Account on Lagrange's Algebraic Identity, *The Mathematical Intelligencer*, **34** (2012), No. 3, 55-61.

[5] T. S. Motzkin, The arithmetic-geometric inequality. In *Inequalities* (O. Shisha, ed.), 483–489, Academic Press, Boston, 1967.

[6] C. P. Niculescu and L.-E. Persson, *Convex Functions and their Applications. A Contemporary Approach*, CMS Books in Mathematics vol. **23**, Springer-Verlag, New York, 2006.

[7] C. P. Niculescu and C. Spiridon, New Jensen-type inequalities. To appear in *Journal of Math. Anal. Appl.*

[8] J. Michael Steele, *The Cauchy-Schwarz master class: an introduction to the art of mathematical inequalities*, Cambridge University Press, 2004.

[9] H. Stephan, Jensen's inequality for non-convex functions, *Recreaţii matematice* (Mathematical Recreations), **IX** (2007), No. 2, 109-110.

[10] H. Stephan, A mathematical framework for general classical systems and time irreversibility as its consequence. *WIAS Berlin 2011*; Preprint n°1629.

[11] J. Willem, *Analyse fonctionelle élémentaire*, Cassini, Paris, 2003.