Lagrange's Barycentric Identity From An Analytic Viewpoint

by

Constantin P. Niculescu and Holger Stephan

Abstract

A probabilistic extension of Lagrange's barycentric identity is proved and used to provide new insights into the mechanism of Jensen's inequality. In particular, it is shown that every instance of Jensen's inequality comes from an identity.

Key Words: Jensen's inequality, Lagrange's algebraic identity, convex function, barycenter of a measure.
2010 Mathematics Subject Classification: Primary 26B25;
Secondary 26D15.

1 Introduction

An important mathematical tool that relates the scalar product and the cross product in an Euclidean space is provided by Lagrange's algebraic identity,

$$\|u\|^2 \|v\|^2 = \langle u, v\rangle^2 + \|u \times v\|^2. \qquad (LAI)$$

Lagrange [6], who was interested in the metric geometry of tetrahedra, considered only the case of 3-dimensional space. The general case, and the attached inequality, $|\langle u, v\rangle| \leq \|u\| \|v\|$, were first proved by Cauchy [4] in his celebrated *Cours d'Analyse*.

In a paper from 1783, devoted to the location of barycenter of a mass system, Lagrange [7] proved an identity even more general, which will be referred to as Lagrange's barycentric identity: *For every family of points $z, x_1, ..., x_n$ in the Euclidean space $\mathbb{R}^3$ and every family of positive weights $p_1, ..., p_n$ with $\sum_{k=1}^{n} p_k = 1$, we have*

$$\sum_{k=1}^{n} p_k \|z - x_k\|^2 = \left\| z - \sum_{k=1}^{n} p_k x_k \right\|^2 + \sum_{1 \leq i < j \leq n} p_i p_j \|x_i - x_j\|^2. \qquad (LBI)$$

The roots and various ramifications of the identities (LAI) and (LBI) made the subject of a recent paper [5].

The aim of the present paper is to extend Lagrange's barycentric identity to a probabilistic framework and to provide in this way new insights into the mechanism of Jensen's inequality. In particular, we prove that every instance of Jensen's inequality comes from an identity. See Theorem 3. Thus, the case of Cauchy's inequality (which follows from Lagrange's algebraic identity) is typical for most known inequalities. Even more, each instance of Jensen's inequality can be paired by a converse, a fact that explains the usefulness of these kind of inequalities in the different applications of mathematics.

2 A first generalization of Lagrange's barycentric identity

In what follows, C denotes a subset of the Euclidean space $\mathbb{R}^N$ (or, more generally, of a real Hilbert space), on which there is defined a real measure $\mu = \sum_{i=1}^{n} p_i \delta_{x_i}$ whose weights p_i are all nonzero and sum up to 1. We also assume that the *barycenter* of μ,

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

belongs to $C \backslash \{x_1, ..., x_n\}$.

Theorem 1. *Under the above assumptions on C and μ, every function $f : C \to \mathbb{R}$ verifies the following generalization of Lagrange's barycentric identity:*

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

where

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

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 $(b_\mu, f(b_\mu))$.

Proof: The proof of (GL) is based on the formulas giving the expression of the above mentioned secant lines,

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

Then,

$$\sum_{i=1}^{n} p_i f(x_i) - f(b_\mu) = \sum_{i=1}^{n} p_i (f(x_i) - f(b_\mu)) = \sum_{i=1}^{n} p_i \langle s(x_i), x_i - b_\mu \rangle$$

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

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

$$= \sum_{1 \le i < j \le n} p_i p_j \langle s(x_i) - s(x_j), x_i - x_j \rangle,$$

and the proof of (GL) is done. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

Remark 1. *When f is a continuously differentiable function defined on a convex subset C of $\mathbb{R}^N$, one can state the identity (GL) in terms of gradients. Indeed, according to the fundamental theorem of integral calculus,*

$$f(x_i) - f(b_\mu) = \int_0^1 \langle \nabla f(tx_i + (1-t)b_\mu), x_i - b_\mu \rangle dt,$$

so that by repeating the argument of Theorem 1 we get the identity

$$\sum_{i=1}^n p_i f(x_i) = f(b_\mu)$$
$$+ \sum_{1 \le i < j \le n} p_i p_j \int_0^1 \langle \nabla f(tx_i + (1-t)b_\mu) - \nabla f(tx_j + (1-t)b_\mu), x_i - x_j \rangle dt. \quad (GL_0)$$

Moreover, in this case the restriction $b_\mu \notin \{x_1, ..., x_n\}$ can be discarded by defining $s(b_\mu) = \nabla f(b_\mu)$.

The identity (GL_0) (i.e., the smooth version of the identity (GL)) can be seen as a convexification of the fundamental formula of integral calculus. It is the first term in a sequence of identities obtained from Taylor's formulas of different orders via convexification. See Section 5 below.

The identity (GL) reduces to Lagrange's barycentric identity when f is the square norm function on the Euclidean space. Indeed, this function is continuously differentiable with $\nabla f(x) = 2x$, and we can use the smooth variant (GL_0) of (GL). The resulting identity is precisely Lagrange's barycentric identity (LBI) for $z = 0$. However, due to the translation invariance of the Euclidean metric, this particular case covers (LBI) in full generality.

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

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

Precisely, 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 case of Theorem 1.

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 the 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.

When $f = \|\cdot\|^2$, we noticed above that

$$\mathcal{E}(f;\mu) - f(b_\mu) = \sum_{1 \leq i < j \leq n} p_i p_j \langle s(x_i) - s(x_j), x_i - x_j \rangle,$$

so, in this case $\mathcal{E}(f;\mu) - f(b_\mu)$ equals the *variance* of the given probability measure μ,

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

More generally, from Remark 1 one can infer easily the following result:

Theorem 2. *If f is continuously differentiable, then*

$$|\mathcal{E}(f;\mu) - f(b_\mu)| \leq \|\nabla f\|_{Lip} \, \sigma_\mu^2,$$

where

$$\|\nabla f\|_{Lip} = \sup \left\{ \frac{\|\nabla f(x) - \nabla f(y)\|}{\|x - y\|} : x, y \in C, \ x \neq y \right\}$$

is the Lipschitz constant of ∇f.

More estimates (from above and from below) of the variance are obtained in Example 1 below.

The double sum

$$SV_\mu^2(f) = \sum_{1 \leq i < j \leq n} p_i p_j \langle s(x_i) - s(x_j), x_i - x_j \rangle,$$

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,

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

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

$$SV_\mu^2(f) = \sum_{1 \leq i < j \leq n} p_i p_j \langle A(x_i - x_j), x_i - x_j \rangle.$$

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 Two 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 $A = \sum_{i=1}^{n} p_i x_i$ 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 $\sigma_\mu^2 = \sum_{1 \leq i < j \leq n} p_i p_j (x_i - x_j)^2$ 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

$$\sigma_\mu^2 \leq (M - A)(A - m),$$

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

$$\sum_{i=1}^{n} p_i (M - x_i)(x_i - m) = (M - A)(A - m) - \sigma_\mu^2.$$

Example 2. Our second example enriches the list of identities verified by the square norm function in the Euclidean space:

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

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

$$\frac{\|x_1\|^2 + \|x_2\|^2 + \|x_3\|^2}{3} + \left\|\frac{x_1 + x_2 + x_3}{3}\right\|^2$$

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

$$= \frac{\|x_1\|^2 + \|x_2\|^2 + \|x_3\|^2}{3} - \left\|\frac{x_1 + x_2 + x_3}{3}\right\|^2$$

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

$$= \frac{1}{18}\sum_{1 \le i < j \le 3}\|x_i - x_j\|^2.$$

From (N) one can immediately infer the inequality

$$\frac{\|x_1\|^2 + \|x_2\|^2 + \|x_3\|^2}{3} + \left\|\frac{x_1 + x_2 + x_3}{3}\right\|^2$$

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

which illustrates the phenomenon of $(2D)$-*convexity* as developed in [1]. This is also related to Popoviciu's characterization of convexity. See [8], p. 12.

4   The connection of (GL) with Jensen's inequality

Very close to our generalization of Lagrange's barycentric 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 [8], 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, a fact that allows us to infer from the identity (GL) the discrete form of Jensen's inequality:

$$f(b_\mu) \le \sum_{i=1}^{n} p_i f(x_i) \tag{J}$$

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 integral 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.

Remark 2. *According to Remark* 1, *a function* f *of class* C^1 *verifies Jensen's inequality if its gradient is monotone in the sense that*

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0$$

for all x, y. *An application to matrix analysis can be easily exhibited by considering the space* $\mathrm{M}_n(\mathbb{R})$ *of all* $n \times n$-*dimensional real matrices, when endowed with the Hilbert-Schmidt norm*

$$\|A\|_{HS} = \left(\sum_{i,j=1}^{n} a_{ij}^2 \right)^{1/2} \quad \textit{for } A = (a_{ij})_{i,j=1}^{n}.$$

This is a Hilbert norm associated to the scalar product

$$\langle A, B \rangle = \operatorname{trace} AB,$$

which makes $\mathrm{M}_n(\mathbb{R})$ *isomorphic to the Euclidean space* $\mathbb{R}^{n \times n}$.

We will show that the function $f(A) = \ln \det A$ *is concave on the subset* S_{++}^n *of all positively definite matrices* $A \in \mathrm{M}_n(\mathbb{R})$. *Taking into account that* $\nabla f(A) = A^{-1}$ *(see* [3]*, pp. 641-642), the monotonicity of the gradient of* f *is equivalent to*

$$\operatorname{trace} \left(\left(A^{-1} - B^{-1} \right) (A - B) \right) n \leq 0$$

for every $A, B \in S_{++}^n$. *Since* $\operatorname{trace} XY = \operatorname{trace} YX$, *we have*

$$\operatorname{trace} \left(\left(A^{-1} - B^{-1} \right) (A - B) \right) = -\operatorname{trace} \left(A^{-1} (A - B) B^{-1} (A - B) \right)$$

$$= -\operatorname{trace} \left(A^{-1/2} (A - B) B^{-1} (A - B) A^{-1/2} \right) \leq 0,$$

because $A^{-1/2} (A - B) B^{-1} (A - B) A^{-1/2}$ *equals*

$$\left(A^{-1/2} (A - B) B^{-1/2} \right) \left(A^{-1/2} (A - B) B^{-1/2} \right)^*.$$

Here A^* *denote the transpose of* A. *It is worth noticing that the* log-*concavity of the function* det *is a particular case of the famous Brunn-Minkowski inequality. See* [8]*, Section 3.12.*

Jensen's inequality (J) can work even for nonconvex functions provided the barycenters b_μ are well placed. Indeed, the fact that the slopes from some point b_μ are nondecreasing is not characteristic to convex functions. See the case of polynomials of 4th degree, which is presented in [10].

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

If f meets $\operatorname{co}(f)$ at an interior point b_μ, 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 $\sum_{i=1}^{n} p_i x_i = b_\mu$. It turns out that this condition simply means that f has a support line at b_μ that is,

$$f(x) \geq f(b_\mu) + \lambda(x - b_\mu),$$

for some $\lambda \in \mathbb{R}$. See [8], 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 [9].

5 Higher order extensions of (LBI)

The smooth extension of Lagrange's barycentric identity (as indicated in Remark 1 above) is just the first step in an infinite string of identities resulting from Taylor's formulas with integral remainder via our convexification procedure. Each such identity adds more insight into the mechanism of Jensen's inequality, a fact that will be made clear by considering Taylor's formula of second order.

Given a differentiable function f of class C^2, defined on a convex open subset $U \subset \mathbb{R}^N$, this formula asserts that for every two points z and x in U,

$$f(z) = f(x) + \langle \nabla f(x), z - x \rangle + \int_0^1 (1-t)\langle \nabla^2 f(x + t(z-x))(z-x), z-x \rangle dt. \qquad (T)$$

Here $\nabla f(x)$ and $\nabla^2 f(x)$ represents respectively the gradient and the Hessian matrix of f at x.

Under the presence on U of a discrete Borel measure $\mu = \sum_{i=1}^n p_i \delta_{x_i}$ of total mass 1, whose barycenter $b_\mu = \int_U x d\mu(x) = \sum_{i=1}^n p_i x_i$ belongs to U, we can infer from Taylor's formula the following fact:

$$f(b_\mu) = \sum_{i=1}^n p_i f(b_\mu) = \sum_{i=1}^n p_i f(x_i) + \sum_{i=1}^n p_i \langle \nabla f(x_i), b_\mu - x_i \rangle$$

$$+ \sum_{i=1}^n p_i \int_0^1 (1-t)\langle \nabla^2 f(x_i + t(b_\mu - x_i))(b_\mu - x_i), b_\mu - x_i \rangle dt$$

$$= \sum_{i=1}^n p_i f(x_i) - \sum_{i=1}^n \sum_{j=1}^n p_i p_j \langle \nabla f(x_i), x_i - x_j \rangle$$

$$+ \sum_{i=1}^n p_i \int_0^1 (1-t)\langle \nabla^2 f(x_i + t(b_\mu - x_i))(b_\mu - x_i), b_\mu - x_i \rangle dt.$$

This can be put in a more symmetric way by summing it side by side with the formula obtained by interchanging i and j. The resulting identity is as follows:

Theorem 3. *Under the above assumptions on f and $\mu = \sum_{i=1}^n p_i \delta_{x_i}$,*

$$f(b_\mu) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n p_i p_j \langle \nabla f(x_i) - \nabla f(x_j), x_i - x_j \rangle \qquad (GL_1)$$

$$= \sum_{i=1}^n p_i f(x_i) + \sum_{i=1}^n p_i \int_0^1 (1-t)\langle \nabla^2 f(x_i + t(b_\mu - x_i))(b_\mu - x_i), b_\mu - x_i \rangle dt.$$

Taking into account that the Hessian matrix of a smooth convex function is positive semidefinite, one can infer from the identity (GL_1) the following upper estimate of Jensen's inequality:

Theorem 4. *Suppose that K is a compact convex subset of the Euclidean space $\mathbb{R}^N$, endowed with a Borel probability measure μ, and f is a convex function of class C^1, defined on a neighborhood of K. Then*

$$\frac{1}{2} \int_K \int_K \langle \nabla f(x) - \nabla f(y), x - y \rangle d\mu(x) d\mu(y) \geq \int_K f(x) d\mu(x) - f(b_\mu) \geq 0.$$

It is worth noticing that Theorem 4 provides a converse to each illustration of Jensen's inequality. For example, in the case of the concave function $\ln \det$ (see Remark 2 above), this converse reads as

$$\ln \det \left(\sum_{i=1}^n p_i A_i \right) - \sum_{i=1}^n p_i \ln \det A_i \leq - \sum_{1 \leq i < j \leq n} p_i p_j \operatorname{trace} \left(A_i^{-1} - A_j^{-1} \right) (A_i - A_j)),$$

for every family $A_1, ..., A_n$ of positive definite matrices and every family $p_1, ..., p_n$ of positive numbers that sum to 1.

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] S. BOYD AND L. VANDENBERGHE, *Convex Optimization*, Cambridge University Press, 2008.

[4] A.-L. CAUCHY, *Cours d'Analyse de l'École Royale Polytechnique, I ère partie, Analyse Algébrique*, Paris, 1821. Reprinted by Ed. Jacques Gabay, Paris, 1989.

[5] M. GIDEA AND C. P. NICULESCU, A Brief Account on Lagrange's Algebraic Identity, *The Mathematical Intelligencer*, **34** (2012), No. 3, 55-61.

[6] J. L. LAGRANGE, Solutions analytiques de quelques problémes sur les pyramides triangulaires, *Nouveaux Mémoirs de l'Académie Royale de Berlin*, 1773; see *Oeuvres* de Lagrange, vol. 3, pp. 661-692, Gauthier-Villars, Paris, 1867.

[7] J. L. LAGRANGE, Sur une nouvelle proprieté du centre de gravité, *Nouveaux Mémoirs de l'Académie Royale de Berlin*, 1783; see *Oeuvres* de Lagrange, vol. 5, pp. 535-540, Gauthier-Villars, Paris, 1870.

[8] 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.

[9] C. P. Niculescu and C. Spiridon, New Jensen-type inequalities, *Journal of Math. Anal. Appl.* **401** (2013), No. 1, 343–348.

[10] C. P. Niculescu and H. Stephan, A generalization of Lagrange's algebraic identity and connections with Jensen's inequality, Weierstraß-Institut für Angewandte Analysis und Stochastik, Preprint No. 1756/2012. Available at http://www.wias-berlin.de/publications/wias-publ

Department of Mathematics, University of Craiova
E-mail: `cniculescu47@yahoo.com`

Weierstrass Institut für Angewandte Analysis und Stochastik, Berlin.
E-mail:`stephan@wias-berlin.de`