Jensen’s inequality for nonconvex functions

SANJO ZLOBEC

Abstract. Jensen’s inequality is formulated for convexifiable (generally nonconvex) functions.

Key words: Jensen’s inequality, convexifiable function, arithmetic mean theorem

AMS subject classifications: 26B25, 52A40

Received June 23, 2004 Accepted November 2, 2004

1. Introduction

Jensen’s inequality is 100 years old, e.g., [1, 2, 3]. It says that the value of a convex function at a point, which is a convex combination of finitely many points, is less than or equal to the convex combination of values of the function at these points. Using symbols: If \( f: \mathbb{R}^n \to \mathbb{R} \) is convex then

\[
f\left(\sum_{i=1}^{p} \lambda_i x^i\right) \leq \sum_{i=1}^{p} \lambda_i f(x^i)
\]

for every set of \( p \) points \( x^i, i = 1, \ldots, p \), in the Euclidean space \( \mathbb{R}^n \) and for all real scalars \( \lambda_i \geq 0, i = 1, \ldots, p \), such that \( \sum_{i=1}^{p} \lambda_i = 1 \).

In this note the inequality (1) is extended from convex to convexifiable functions, e.g., [4, 5]. These include all twice continuously differentiable functions, all once continuously differentiable functions with Lipschitz derivative and all analytic functions. As a special case we obtain a new form of the arithmetic mean theorem.

2. Convexifiable functions

If \( f: \mathbb{R}^n \to \mathbb{R} \) is a continuous function in \( n \) variables defined on a convex set \( C \) of \( \mathbb{R}^n \), then the function is said to be convex on \( C \) if

\[
f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)
\]

\(^*\)Research supported in part by a grant from NSERC of Canada.

\(^\dagger\)McGill University, Department of Mathematics and Statistics, Burnside Hall, 805 Sherbrooke Street West, Montreal, Quebec, Canada H3A 2K6, e-mail: zlobec@math.mcgill.ca
for every \( x, y \in C \) and scalar \( 0 \leq \lambda \leq 1 \). Note that this is (1) for \( p = 2 \). Let us recall several recent results.

**Definition 1** [[5]]. Given a continuous \( f : \mathbb{R}^n \to \mathbb{R} \) defined on a convex set \( C \), consider the function \( \varphi : \mathbb{R}^{n+1} \to \mathbb{R} \) defined by \( \varphi(x, \alpha) = f(x) - \frac{1}{2}\alpha x^T x \), where \( x^T \) is the transposed of \( x \). If \( \varphi(x, \alpha) \) is a convex function on \( C \) for some \( \alpha = \alpha^* \), then \( \varphi(x, \alpha) \) is a convexification of \( f \) and \( \alpha^* \) is its convexifier on \( C \). Function \( f \) is convexifiable if it has a convexification.

**Observation 1.** If \( \alpha^* \) is a convexifier of \( f \), then so is every \( \alpha \leq \alpha^* \).

In order to characterize a convexifiable function, the mid-point acceleration function

\[
\Psi(x, y) = \frac{4}{\|x - y\|^2} \left( f(x) + f(y) - 2f\left( \frac{x + y}{2} \right) \right), \quad x, y \in C, x \neq y
\]

was introduced in [5]. There it was shown that a continuous \( f : \mathbb{R}^n \to \mathbb{R} \), defined on a nontrivial convex set \( C \) (i.e., a convex set with at least two distinct points) in \( \mathbb{R}^n \) is convexifiable on \( C \) if, and only if, its mid-point acceleration function \( \Psi \) is bounded from below on \( C \).

For two important classes of functions a convexifier \( \alpha \) can be given explicitly. If \( f \) is twice continuously differentiable then its second derivative at \( x \) is represented by the Hessian matrix \( H(x) = \left( \partial^2 f(x)/\partial x_i \partial x_j \right) \). This is a symmetric matrix with real eigenvalues. Denote its smallest eigenvalue by \( \lambda(x) \) and its “globally” smallest eigenvalue over a compact convex set \( C \) by

\[
\lambda^* = \min_{x \in C} \lambda(x).
\]

**Lemma 1** [[4, 5]]. Given a twice continuously differentiable function \( f : \mathbb{R}^n \to \mathbb{R} \) on a nontrivial compact convex set \( C \) in \( \mathbb{R}^n \). Then \( \alpha = \lambda^* \) is a convexifier.

We say that a continuously differentiable function \( f \) has Lipschitz derivative if

\[
|\|\nabla f(x) - \nabla f(y)\|_2| = L\|x - y\|_2 \text{ for every } x, y \in C \text{ and some constant } L.
\]

Here \( \nabla f(u) \) is the (Fréchet) derivative of \( f \) at \( u \) and \( \|u\| = (u^T u)^{1/2} \) is the Euclidean norm. We represent the derivative at \( x \) as a row n-tuple gradient \( \nabla f(x) = (\partial f(x)/\partial x_i) \).

**Lemma 2** [[5]]. Given a continuously differentiable function \( f : \mathbb{R}^n \to \mathbb{R} \) with Lipschitz derivative and a constant \( L \) on a nontrivial compact convex set \( C \) in \( \mathbb{R}^n \). Then \( \alpha = -L \) is a convexifier.

One can show that every convexifiable scalar function \( f : \mathbb{R} \to \mathbb{R} \) is Lipschitz, i.e., \( |f(s) - f(t)| \leq K|s - t| \) for every \( s \) and \( t \) and some constant \( K \). This means that a scalar non-Lipschitz function is not convexifiable. However, almost all smooth functions of practical interest are convexifiable; e.g., [5].

3. Jensen’s inequality for convexifiable functions

In this section we formulate (1) for convexifiable functions.

**Theorem 1 [Jensen’s inequality for convexifiable functions].** Consider a convexifiable function \( f : \mathbb{R}^n \to \mathbb{R} \) on a bounded nontrivial convex set \( C \) of \( \mathbb{R}^n \) and its convexifier \( \alpha \). Then
Corollary 1 holds, in particular, for analytic functions with convex set \( \alpha \) and its constant \( \lambda \). Given a twice continuously differentiable function \( f \), one obtains

\[
\alpha = \lambda^* \quad \text{for every set of } \alpha, \lambda \in \mathbb{R}.
\]

Hence Jensen’s inequality works for \( \varphi(x, \alpha) \). After substitution one obtains

\[
f(\lambda x) \leq \lambda^*(x) - \frac{\lambda^*(x)^T(x - x^*)}{2}
\]

After more rearranging the more pleasing form (3) follows.

Using the fact that for a convex function \( f \) one can choose the convexifier \( \alpha = 0 \), one recovers (1). For a twice continuously differentiable function one can specify \( \alpha = \lambda^* \) (by Lemma 1) and for a continuously differentiable function with Lipschitz derivative and its constant \( L \), one can specify \( \alpha = -L \) (by Lemma 2). Hence we have, respectively, the following special cases:

**Corollary 1 [Jensen’s inequality for twice continuously differentiable functions].** Given a twice continuously differentiable function \( f : \mathbb{R}^n \to \mathbb{R} \) on a nontrivial compact convex set \( C \) in \( \mathbb{R}^n \). Then

\[
f(\sum_{i=1}^{p} \lambda_i x_i) \leq \sum_{i=1}^{p} \lambda_i f(x_i) - \frac{\lambda^*}{2} \left( \sum_{i,j=1}^{p} \lambda_i \lambda_j \|x_i - x_j\|^2 \right)
\]

for every set of \( p \) points \( x_i, i = 1, \ldots, p \), in \( C \) and all real scalars \( \lambda_i \geq 0, i = 1, \ldots, p \), with \( \sum_{i=1}^{p} \lambda_i = 1 \).

**Observation 2.** If \( f \) in Corollary 1 is strictly convex, then the lowest eigenvalue of the Hessian is \( \lambda^* \geq 0 \) (often \( \lambda^* > 0 \)) and (4) may provide a better bound than (1). Since every analytic function \( f : \mathbb{R} \to \mathbb{R} \) is twice continuously differentiable, Corollary 1 holds, in particular, for analytic functions with \( \lambda^* = \min_{t \in C} f''(t) \).

**Corollary 2 [Jensen’s inequality for once continuously differentiable functions with Lipschitz derivative].** Given a continuously differentiable function \( f : \mathbb{R}^n \to \mathbb{R} \) with Lipschitz derivative and a constant \( L \) on a nontrivial compact convex set \( C \) in \( \mathbb{R}^n \). Then

\[
f(\sum_{i=1}^{p} \lambda_i x_i) \leq \sum_{i=1}^{p} \lambda_i f(x_i) + \frac{L}{2} \left( \sum_{i,j=1}^{p} \lambda_i \lambda_j \|x_i - x_j\|^2 \right)
\]

for every set of \( p \) points \( x_i, i = 1, \ldots, p \), in \( C \) and all real scalars \( \lambda_i \geq 0, i = 1, \ldots, p \), with \( \sum_{i=1}^{p} \lambda_i = 1 \).

**Special Case:** For a scalar function \( f : \mathbb{R} \to \mathbb{R} \) and two scalar points \( a \) and \( b \) Jensen’s inequality is

\[
f(\lambda a + (1 - \lambda)b) \leq \lambda f(a) + (1 - \lambda)f(b), \quad \text{for every } 0 \leq \lambda \leq 1
\]
while for a convexifiable $f$, it is

$$f(\lambda a + (1 - \lambda)b) \leq \lambda f(a) + (1 - \lambda)f(b) - \frac{\alpha}{2}\lambda(1 - \lambda)(a - b)^2$$

for every convexifier $\alpha$ and for every $0 \leq \lambda \leq 1$. We will use this special case to illustrate the basic difference between the two inequalities.

**Illustration 1.** Consider $f(t) = \sin t$ on $0 \leq t \leq 2\pi$. Take $a = 0$ and $b = 2\pi$. Then (1) and its extension yield, respectively

$$\sin(2\pi(1 - \lambda)) \leq 0, \quad 0 \leq \lambda \leq 1$$

(6) and

$$\sin(2\pi(1 - \lambda)) \leq 2\pi^2\lambda(1 - \lambda), \quad 0 \leq \lambda \leq 1.$$  (7)

Inequality (6) is not satisfied on the region where $f(t)$ is not convex, i.e., $1/2 \leq \lambda \leq 1$. On the other hand the new upper bound in (7) holds (see Figure 1).

**Illustration 2.** Consider $f(t) = t^4$ between $a = 1$ and $b = 2$. Then (1) and its extension yield $(2 - \lambda)^4 \leq 16 - 15\lambda$ and $(2 - \lambda)^4 \leq 16 - 9\lambda - 6\lambda^2$, $0 \leq \lambda \leq 1$, respectively. The upper bounds are compared against the original function in Figure 2.
Jensen's inequality for nonconvex functions

Jensen's inequality is closely related to the arithmetic mean theorem for real numbers. The following theorem says that the value of a convex function at the arithmetic mean of p numbers is less than or equal to the arithmetic mean of the values of the function at these numbers.

Theorem 2 [Classic arithmetic mean theorem for convex functions, e.g., [3]]. Consider a convex scalar function \( f : \mathbb{R} \to \mathbb{R} \) on a nontrivial compact interval \([a, b]\). Then

\[
f\left(\frac{1}{p} \sum_{i=1}^{p} t_i\right) \leq \frac{1}{p} \sum_{i=1}^{p} f(t_i) \tag{8}
\]

for every set of \( p \) points \( t_i \in [a, b], i = 1, \ldots, p \).

Specifying \( x^* = t_i, \lambda_i = 1/p, i = 1, \ldots, p \), in (3) one obtains, after rearrangement, the following extension:

Theorem 3 [Arithmetic mean theorem for convexifiable functions]. Consider a convexifiable scalar function \( f : \mathbb{R} \to \mathbb{R} \) on a nontrivial compact interval \([a, b]\) and its convexifier \( \alpha \). Then

\[
f\left(\frac{1}{p} \sum_{i=1}^{p} t_i\right) \leq \frac{1}{p} \sum_{i=1}^{p} f(t_i) - \frac{\alpha}{2} \left(\frac{1}{p} \sum_{i=1}^{p} t_i^2 - \left(\frac{1}{p} \sum_{i=1}^{p} t_i\right)^2\right) \tag{9}
\]

for every set of \( p \) points \( t_i \in [a, b], i = 1, \ldots, p \).

Observation 3. In (9) one can set \( \alpha = 0 \) if \( f \) is convex, \( \alpha = \lambda^* = \min_{t \in [a, b]} f''(t) \) if \( f \) is twice continuously differentiable or \( \alpha = -L \) if \( f \) is Lipschitz continuously differentiable with a constant \( L \). The first special case recovers the classic result.
Observation 4. The term corresponding to the convexifier is positive, provided that at least one $t_i$ is non-zero. Indeed, denote $A = (t_i) \in \mathbb{R}^p$, $E = (1, \ldots, 1)^T \in \mathbb{R}^p$. Then this term is $[(1/p)(A, A) - (1/p)^2(A, E)^2]$. Since $(A, E)^2 \leq \|A\|^2\|E\|^2 = (A, A) \cdot p$ and $p < p^2$, the term is positive. Since for a twice continuously differentiable strictly convex $f$, we know that $\lambda^* = \min_{t \in [a, b]} f''(t) \geq 0$, it follows that (9) typically provides in this case a better estimate than (8).

Special Case: For a scalar function $f : \mathbb{R} \to \mathbb{R}$ and only two points $t_1$ and $t_2$, (3) (and after some rearrangement (9)) yields

$$f\left(\frac{t_1 + t_2}{2}\right) \leq \frac{1}{2}(f(t_1) + f(t_2)) - \alpha \cdot (t_1 - t_2)^2$$

References


