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 : $\mathbb{R}^n \to \mathbb{R}$ is convex then

$$f\Big(\sum_{i=1}^{p} \lambda_i \mathbf{x}^i\Big) \leq \sum_{i=1}^{p} \lambda_i f(\mathbf{x}^i) \tag{1}$$

for every set of p points $\mathbf{x}^i, i = 1, \ldots, p$, 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

$$f\big(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}\big) \leq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}) \tag{2}$$

*Research supported in part by a grant from NSERC of Canada.

†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 $\mathbf{x}, \mathbf{y} \in C$ 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 $\varphi(\mathbf{x}, \alpha) = f(\mathbf{x}) - \frac{1}{2}\alpha \mathbf{x}^T \mathbf{x}$, where $\mathbf{x}^T$ is the transposed of $\mathbf{x}$. If $\varphi(\mathbf{x}, \alpha)$ is a convex function on C for some $\alpha = \alpha^\star$, then $\varphi(\mathbf{x}, \alpha)$ is a* convexification *of f and $\alpha^\star$ is its* convexifier *on C. Function f is* convexifiable *if it has a convexification.*

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

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

$$\Psi(\mathbf{x}, \mathbf{y}) = \frac{4}{\|\mathbf{x} - \mathbf{y}\|^2}\left(f(\mathbf{x}) + f(\mathbf{y}) - 2f\left(\frac{\mathbf{x} + \mathbf{y}}{2}\right)\right), \qquad \mathbf{x}, \mathbf{y} \in C, \mathbf{x} \neq \mathbf{y}$$

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 Ψ is bounded from below on C.

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

$$\lambda^\star = \min_{\mathbf{x} \in C} \lambda(\mathbf{x}).$$

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 $\alpha = \lambda^\star$ is a convexifier.*

We say that a continuously differentiable function f has Lipschitz derivative if $|\ [\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})]\,(\mathbf{x} - \mathbf{y})\ | \leq L\|\mathbf{x} - \mathbf{y}\|$ for every $\mathbf{x}, \mathbf{y} \in C$ and some constant L. Here $\nabla f(\mathbf{u})$ is the (Fréchet) derivative of f at $\mathbf{u}$ and $\|\mathbf{u}\| = (\mathbf{u}^T \mathbf{u})^{1/2}$ is the Euclidean norm. We represent the derivative at $\mathbf{x}$ as a row n-tuple gradient $\nabla f(\mathbf{x}) = \left(\partial f(\mathbf{x})/\partial \mathbf{x}_i\right)$.

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 α. Then*

$$f\Big(\sum_{i=1}^{p}\lambda_i\mathbf{x}^i\Big)\le\sum_{i=1}^{p}\lambda_i f(\mathbf{x}^i)-\frac{\alpha}{2}\Bigg(\sum_{\substack{i,j=1\\ i<j}}^{p}\lambda_i\lambda_j\|\mathbf{x}^i-\mathbf{x}^j\|^2\Bigg) \tag{3}$$

for every set of p points $\mathbf{x}^i, i=1,\dots,p$, in C and all real scalars $\lambda_i\ge 0$, $i=1,\dots,p$, with $\sum_{i=1}^{p}\lambda_i=1$.

Proof. Since f is convexifiable, $\varphi(\mathbf{x},\alpha)=f(\mathbf{x})-\frac{1}{2}\alpha\mathbf{x}^T\mathbf{x}$ is a convex function for every convexifier α. Hence Jensen's inequality works for $\varphi(\mathbf{x},\alpha)$. After substitution one obtains

$$f\Big(\sum_{i=1}^{p}\lambda_i\mathbf{x}^i\Big)\le\sum_{i=1}^{p}\lambda_i f(\mathbf{x}^i)-\frac{\alpha}{2}\Bigg(\sum_{i,j=1}^{p}\lambda_i\lambda_j(\mathbf{x}^i)^T(\mathbf{x}^i-\mathbf{x}^j)\Bigg).$$

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 $\alpha=\lambda^\star$ (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*

$$f\Big(\sum_{i=1}^{p}\lambda_i\mathbf{x}^i\Big)\le\sum_{i=1}^{p}\lambda_i f(\mathbf{x}^i)-\frac{\lambda^\star}{2}\Bigg(\sum_{\substack{i,j=1\\ i<j}}^{p}\lambda_i\lambda_j\|\mathbf{x}^i-\mathbf{x}^j\|^2\Bigg) \tag{4}$$

for every set of p points $\mathbf{x}^i, i=1,\dots,p$, in C and all real scalars $\lambda_i\ge 0$, $i=1,\dots,p$, 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 $\lambda^\star\ge 0$ (often $\lambda^\star>0$) 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 $\lambda^\star=\min_{t\in C}f''(t)$.*

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*

$$f\Big(\sum_{i=1}^{p}\lambda_i\mathbf{x}^i\Big)\le\sum_{i=1}^{p}\lambda_i f(\mathbf{x}^i)+\frac{L}{2}\Bigg(\sum_{\substack{i,j=1\\ i<j}}^{p}\lambda_i\lambda_j\|\mathbf{x}^i-\mathbf{x}^j\|^2\Bigg) \tag{5}$$

for every set of p points $\mathbf{x}^i, i=1,\dots,p$, in C and all real scalars $\lambda_i\ge 0$, $i=1,\dots,p$, 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

$$f\big(\lambda a+(1-\lambda)b\big)\le\lambda f(a)+(1-\lambda)f(b),\qquad\text{for every } 0\le\lambda\le 1$$

while for a convexifiable f, it is

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

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

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

$$\sin(2\pi(1-\lambda)) \le 0, \qquad 0 \le \lambda \le 1 \tag{6}$$

and

$$\sin(2\pi(1-\lambda)) \le 2\pi^2\lambda(1-\lambda), \qquad 0 \le \lambda \le 1. \tag{7}$$

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

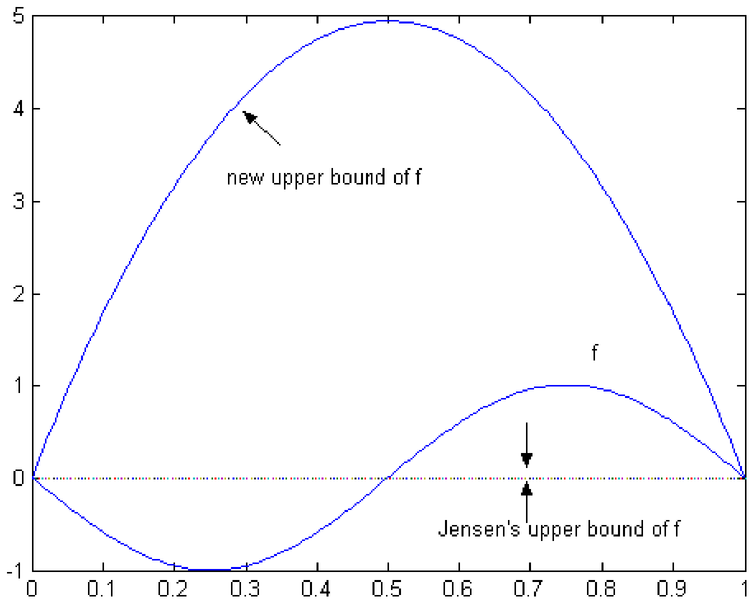

Figure 1. *Jensen's inequality for a convexifiable function*

A situation where the new bound is sharper than the one provided by Jensen's inequality for a convex function is illustrated in the following example.

Illustration 2. *Consider $f(t) = t^4$ between $a = 1$ and $b = 2$. Then (1) and its extension yield $(2-\lambda)^4 \le 16 - 15\lambda$ and $(2-\lambda)^4 \le 16 - 9\lambda - 6\lambda^2$, $0 \le \lambda \le 1$, respectively. The upper bounds are compared against the original function in* Figure 2*.*

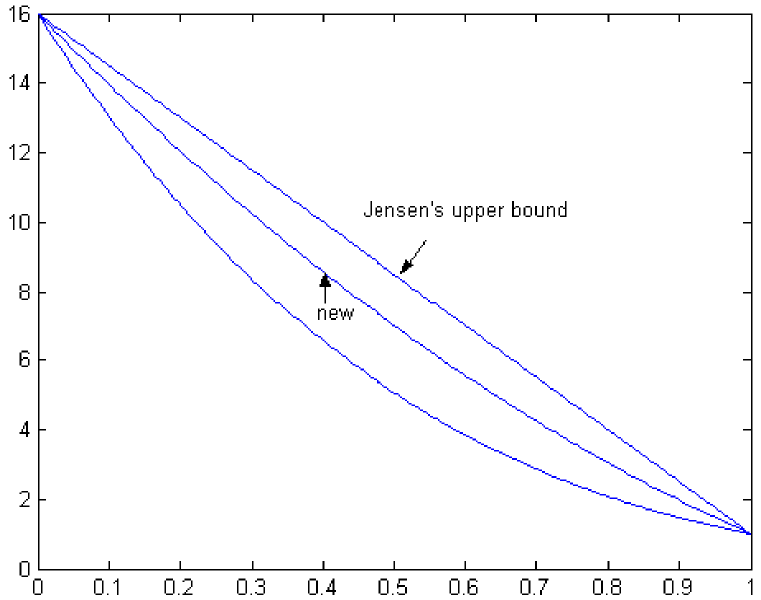

Figure 2. *Improvement for a strictly convex function*

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 $t_i \in [a, b], i = 1, \dots, p$.

Specifying $\mathbf{x}^i = t_i$, $\lambda_i = 1/p$, $i = 1, \dots, p$, 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 α. 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 $t_i \in [a, b], i = 1, \dots, p$.

Observation 3. *In (9) one can set $\alpha = 0$ if f is convex, $\alpha = \lambda^\star = \min_{t\in[a,b]} f''(t)$ 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* $\mathbf{A} = (t_i) \in \mathbb{R}^p$*,* $\mathbf{E} = (1, \dots, 1)^T \in \mathbb{R}^p$*. Then this term is* $[(1/p)(\mathbf{A}, \mathbf{A}) - (1/p)^2(\mathbf{A}, \mathbf{E})^2]$*. Since* $(\mathbf{A}, \mathbf{E})^2 \leq \|\mathbf{A}\|^2\|\mathbf{E}\|^2 = (\mathbf{A}, \mathbf{A}) \cdot p$ *and* $p < p^2$*, the term is positive. Since for a twice continuously differentiable strictly convex* f*, we know that* $\lambda^\star = \min_{t\in[a,b]} f''(t) \geq 0$*, 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

$$f\left(\frac{t_1 + t_2}{2}\right) \leq \frac{1}{2}\big(f(t_1) + f(t_2)\big) - \frac{\alpha}{8} \cdot (t_1 - t_2)^2$$

References

[1] J. L. W. V. JENSEN, *Om konvexe Funktioner og Uligheder mellem Middelvaerdier*, Nyt Tidsskr. Math. **16B**(1905), 49–69.

[2] J. L. W. V. JENSEN, *Sur les fonctions convexes et les inegalites entre les valeur moyennes*, Acta Math. **30**(1906), 175–193.

[3] A. W. ROBERTS, D. E. VARBERG, *Convex Functions*, Academic Press, New York, 1973.

[4] S. ZLOBEC, *Estimating convexifiers in continuous optimization*, Mathematical Communications **8**(2003), 129–137.

[5] S. ZLOBEC, *Characterization of convexifiable functions*, McGill University, June 2004, to be published.