ROCKY MOUNTAIN
JOURNAL OF MATHEMATICS
Volume 41, Number 6, 2011

SHARP GLOBAL BOUNDS FOR JENSEN'S INEQUALITY

SLAVKO SIMIC

ABSTRACT. In this article we find the form of a sharp global upper bound for Jensen's inequality. Thereby, previous results on this topic are essentially improved. We also give some applications in Analysis and Information Theory.

1. Introduction. The form of Jensen functional $J_f(\mathbf{p}, \mathbf{x})$ is given by (cf. [**4**]),

$$J_f(\mathbf{p}, \mathbf{x}) := \sum p_i f(x_i) - f\Big(\sum p_i x_i\Big),$$

where, in this case, $f(\cdot)$ stands for a convex function defined on the domain D_f, $D_f \subseteq \mathbf{R}$, $\mathbf{x} := \{x_i\}$ is a finite sequence of numbers from D_f and $\mathbf{p} := \{p_i\}$, $\sum p_i = 1$ denotes a positive weight sequence associated with $\mathbf{x}$.

Throughout the paper we assume that all terms of the sequence $\mathbf{x}$ belong to some closed interval I, i.e., that for some fixed $a, b : x_i \in [a, b] := I \subseteq D_f$, $i = 1, 2, \dots$.

Therefore, the global bounds for $J_f(\mathbf{p}, \mathbf{x})$ will depend only on f and I (cf. [**8**]). For instance, the famous Jensen's inequality asserts that for $\mathbf{x} \in I$,

$$0 \le J_f(\mathbf{p}, \mathbf{x}).$$

One can see that the lower bound zero is of global nature since it does not depend on $\mathbf{p}$ or $\mathbf{x}$ but only on f and the interval I whereupon f is convex. It is also obvious that zero is the best possible global lower bound for the Jensen functional.

In the same sense, an upper global bound was given by Dragomir in [**3**].

2010 AMS *Mathematics subject classification.* Primary 26B25, 26D20.
Keywords and phrases. Jensen's discrete inequality, global bounds, Lagrange means, generalized $A - G - H$ inequality.
Received by the editors on March 11, 2009, and in revised form on March 31, 2009.
DOI:10.1216/RMJ-2011-41-6-2021

Theorem A. *If f is a differentiable convex mapping on I, then we have*

$$J_f(\mathbf{p},\mathbf{x}) \le \frac{1}{4}(b-a)(f'(b)-f'(a)) := R_f(a,b). \tag{1}$$

There are a number of papers where the inequality (1) is utilized in applications concerning some parts of Analysis, Numerical Analysis, Information Theory, etc., (cf. [**1, 2, 3, 5, 9**]).

Our article [**9**] contained an upper global bound without differentiability restriction on f. Namely, we proved the following

Theorem B. *If $\mathbf{p}$, $\mathbf{x}$ are defined as above, we have that*

$$J_f(\mathbf{p},\mathbf{x}) \le f(a)+f(b)-2f\left(\frac{a+b}{2}\right) := S_f(a,b), \tag{2}$$

for any f that is convex over $I := [a,b]$.

Moreover, in [**8**] we introduced the characteristic $C(f)$ of a convex function f, i.e., a constant depending only on f, defined by

$$C(f) := \sup_{p;a,b\in D_f}\left[\frac{pf(a)+qf(b)-f(pa+qb)}{f(a)+f(b)-2f(a+b)/2}\right].$$

For example,

$$C(x^2)=1/2; \quad C(-\sqrt{x})=(\sqrt{2}+1)/4;$$
$$C(x\log x)=(e\log 2)^{-1}; \quad C(-\log x)=1.$$

Since $1/2 \le C(f) \le 1$, the inequality (2) is improved to

$$J_f(\mathbf{p},\mathbf{x}) \le C(f)S_f(a,b) := \widehat{S}_f(a,b). \tag{3}$$

In this article we shall give another global bound $T_f(a,b)$ for Jensen's inequality, which is better than both of the aforementioned bounds $R_f(a,b)$ and $\widehat{S}_f(a,b)$ and is, in fact, best possible.

As an application, we determine $T_f(a,b)$ in the case of the generalized $A-G$ inequality as a combination of some well-known classical means. As a consequence, new global upper bounds for $A-G$ and $G-H$ inequalities are established. Sharp bounds for Shannon's entropy and Kullback-Leibler divergence, probability measures which are of importance in Information Theory, are also given.

2. Results. Our main result is contained in the following

Theorem C. *Let f, $\mathbf{p}, \mathbf{x}$ be defined as above and $p, q > 0$, $p+q=1$. Then*

$$\begin{aligned} J_f(\mathbf{p},\mathbf{x}) &:= \sum p_i f(x_i) - f\Big(\sum p_i x_i\Big) \\ &\le \max_p [pf(a) + qf(b) - f(pa+qb)] \\ &:= T_f(a,b). \end{aligned} \tag{4}$$

This upper bound is very precise. For example,

$$T_{x^2}(a,b) = \max_p (pa^2 + qb^2 - (pa+qb)^2) = \max_p (pq(b-a)^2) = \frac{1}{4}(b-a)^2,$$

and we obtain at once the well-known pre-Grüss inequality

$$\sum p_i x_i^2 - \Big(\sum p_i x_i\Big)^2 \le \frac{1}{4}(b-a)^2,$$

with the constant $1/4$ as best possible.

Remark 1. It is easy to see that, for fixed $a, b \in D_f$, $a \ne b$, the function $g(p) := pf(a) + (1-p)f(b) - f(pa + (1-p)b)$ is concave for $0 \le p \le 1$ with $g(0) = g(1) = 0$. Hence, there exists the unique positive $\max_p g(p) = g(p_0) = T_f(a,b)$ attained at the unique $p_0 = p_0(a,b) \in (0,1)$.

An interesting question suggested by the referee is: how does the result from Theorem C look if, instead of I, one considers a convex subset of a vector space.

The generalization follows immediately if one considers a convex polytope in R^n, but this is not the topic of the actual article. Similar generalizations are left to the readers.

The next theorem approves that the inequality (4) is stronger than (1) or (3).

Theorem D. *Let D_f be the domain of a convex function f and $I := [a, b] \subset D_f$. Then the inequalities*

(i) $T_f(a, b) \le R_f(a, b)$;

(ii) $T_f(a, b) \le \widehat{S}_f(a, b)$,

hold for each $I \subset D_f$.

We also show that the bound $T_f(a, b)$ is sharp by the following

Theorem E. *For an arbitrary convex function f and $I = [a, b] \in D_f$, there exist a sequence $\mathbf{x}_0 \in I$ and an associated weight sequence $\mathbf{p}_0$, such that*

$$J_f(\mathbf{p}_0, \mathbf{x}_0) = T_f(a, b).$$

The explicit form of $T_f(a, b)$ is given by the next assertion.

Theorem F. *For a differentiable convex mapping f, we have that*

$$T_f(a, b) = \frac{f(b) - f(a)}{b - a}\, \Theta_f(a, b) + \frac{bf(a) - af(b)}{b - a} - f(\Theta_f(a, b)), \tag{5}$$

where $\Theta_f(a, b)$ is the Lagrange mean value of numbers a, b, defined by

$$\Theta_f(a, b) := (f')^{-1}\left(\frac{f(b) - f(a)}{b - a}\right).$$

We shall give some applications of the above results in Analysis and Information Theory.

3. Applications.

3.1. Applications in analysis. The following well-known $A-G$ inequality [**7**] asserts that

$$A(\mathbf{p},\mathbf{x}) \geq G(\mathbf{p},\mathbf{x}),$$

where

$$A(\mathbf{p},\mathbf{x}) := \sum p_i x_i; \qquad G(\mathbf{p},\mathbf{x}) := \prod x_i^{p_i},\ \mathbf{x} \in \mathbf{R}^+,$$

are the generalized arithmetic and geometric means, respectively.

As an illustration we determine a new converse of $A-G$ inequality.

Theorem G. *For $x_i \in [a,b]$, $0 < a < b$, $i = 1, 2, \dots,$ we have*

$$1 \leq \frac{A(\mathbf{p},\mathbf{x})}{G(\mathbf{p},\mathbf{x})} \leq L(a,b)I(a,b)/G^2(a,b) := \Lambda_1(a,b), \tag{6}$$

where

$$G(a,b) := \sqrt{ab};\ L(a,b) := \frac{b-a}{\log b - \log a};\ I(a,b) := (b^b/a^a)^{1/(b-a)}/e,$$

are the geometric, logarithmic and identric means, respectively.

As a consequence, we also get a converse of the $G-H$ inequality

$$1 \leq \frac{G(\mathbf{p},\mathbf{x})}{H(\mathbf{p},\mathbf{x})} \leq \Lambda_1(a,b), \tag{7}$$

where the generalized harmonic mean $H(\cdot,\cdot)$ of positive numbers x_1, x_2, $\dots$ is defined by

$$H(\mathbf{p},\mathbf{x}) := \left(\sum \frac{p_i}{x_i}\right)^{-1}.$$

Analogously, we obtain

Theorem H. *For $x_i \in [a,b]$, $0<a<b$, $i=1,2,\dots,$ we have*

$$0 \le A(\mathbf{p},\mathbf{x}) - G(\mathbf{p},\mathbf{x}) \le 2(A(a,b)-L(a,b)) - L(a,b)\log\frac{I(a,b)}{L(a,b)}$$
$$:= \Lambda_2(a,b);$$
$$0 \le \sum p_i x_i \log x_i - \Big(\sum p_i x_i\Big)\log\Big(\sum p_i x_i\Big) \le I(a,b) - \frac{G^2(a,b)}{L(a,b)}$$
$$:= \Lambda_3(a,b).$$

Remark 2. It is well known that, for $0<a<b$,

$$\min\{a,b\} < G(a,b) < L(a,b) < I(a,b) < A(a,b) < \max\{a,b\}.$$

Those inequalities can be used for a simplification of the expressions Λ_i, $i=1,2,3$.

3.2. Applications in information theory. Define probability distributions P and Q of a discrete random variable X by

$$P(X=i)=p_i>0, \qquad Q(X=i)=q_i>0, \quad i=1,2,\dots,r;$$
$$\sum p_i = \sum q_i = 1.$$

Among the other quantities, of utmost importance in Information Theory are the Kullback-Leibler divergence $D_{KL}(P||Q)$ and Shannon's entropy $H(X)$, defined to be

$$D_{KL}(P||Q) := \sum p_i \log\frac{p_i}{q_i};$$
$$H(X) := \sum_1^r p_i \log\frac{1}{p_i}.$$

The distribution P represents here data and observations, while Q typically represents a theoretical model or an approximation of P. Both divergences are always non-negative. For example, Gibbs's inequality states that $D_{KL}(P||Q) \ge 0$ and $D_{KL}(P||Q)=0$ if and only if $P=Q$.

Applying the above results we obtain the following estimates.

Theorem I. (i) *Denoting* $m := \min(q_i/p_i)$; $M := \max(q_i/p_i)$, $i = 1, 2, \ldots$, *we have*

$$0 \leq D_{KL}(P||Q) \leq \log \Lambda_1(m, M). \tag{8}$$

(ii) *Denoting* $\mu := \min\{p_i\}$; $\nu := \max\{p_i\}$, $i = 1, 2, \ldots$, *we have*

$$0 \leq \log r - H(X) \leq \log \Lambda_1(\mu, \nu). \tag{9}$$

Since G, L, I are homogenous means of order one, it follows from (6) that $\Lambda_1(a, b)$ is a homogenous function of order zero, i.e., $\Lambda_1(ta, tb) = \Lambda_1(a, b)$, $t > 0$.

For example, if $p_i \in (a_r, 10a_r)$, $a_r > 0$, we have

$$\log \Lambda_1(a_r, 10a_r) = \log \Lambda_1(1, 10) \approx 0,619.$$

Hence,

$$0 \leq \log r - H(X) < 0,62$$

for each $r \in \mathbf{N}$.

4. Proofs.

Proof of Theorem C. Since $x_i \in [a, b]$, there is a sequence $\{\lambda_i\}$, $\lambda_i \in [0, 1]$, such that $x_i = \lambda_i a + (1 - \lambda_i)b$, $i = 1, 2, \ldots$. Hence,

$$\begin{aligned}
&\sum p_i f(x_i) - f\Big(\sum p_i x_i\Big) \\
&\qquad = \sum p_i f(\lambda_i a + (1 - \lambda_i)b) - f\Big(\sum p_i(\lambda_i a + (1 - \lambda_i)b)\Big) \\
&\qquad \leq \sum p_i(\lambda_i f(a) + (1 - \lambda_i)f(b)) \\
&\qquad\quad - f\Big(a \sum p_i \lambda_i + b \sum p_i(1 - \lambda_i)\Big) \\
&\qquad = f(a)\Big(\sum p_i \lambda_i\Big) + f(b)\Big(1 - \sum p_i \lambda_i\Big) \\
&\qquad\quad - f\Big(a\Big(\sum p_i \lambda_i\Big) + b\Big(1 - \sum p_i \lambda_i\Big)\Big).
\end{aligned}$$

Denoting $\sum p_i\lambda_i := p$; $1-\sum p_i\lambda_i := q$, we have that $0\le p,q\le 1$; $p+q=1$. Consequently,

$$\begin{aligned} J_f(\mathbf{p},\mathbf{x}) &:= \sum p_i f(x_i) - f\Big(\sum p_i x_i\Big) \\ &\le pf(a)+qf(b)-f(pa+qb) \\ &\le \max_p[pf(a)+qf(b)-f(pa+qb)] \\ &:= T_f(a,b), \end{aligned}$$

and the proof of Theorem C is done.

Proof of Theorem D. (i) Since f is convex (and differentiable, in this case), we have for all $x,t\in I$:

$$f(x)\ge f(t)+(x-t)f'(t).$$

In particular,

$$f(pa+qb)\ge f(a)+q(b-a)f'(a); \qquad f(pa+qb)\ge f(b)+p(a-b)f'(b).$$

Therefore,

$$\begin{aligned} &pf(a)+qf(b)-f(pa+qb) \\ &\qquad = p(f(a)-f(pa+qb))+q(f(b)-f(pa+qb)) \\ &\qquad \le p(q(a-b)f'(a))+q(p(b-a)f'(b)) \\ &\qquad = pq(b-a)(f'(b)-f'(a)). \end{aligned}$$

Hence,

$$\begin{aligned} T_f(a,b) &:= \max_p[pf(a)+qf(b)-f(pa+qb)] \\ &\le \max_p[pq(b-a)(f'(b)-f'(a))] \\ &= \frac{1}{4}(b-a)(f'(b)-f'(a)) := R_f(a,b). \end{aligned}$$

(ii) Because

$$\begin{aligned} \frac{pf(a)+qf(b)-f(pa+qb)}{S_f(a,b)} &\le \sup_{p;a,b\in D_f}\left[\frac{pf(a)+qf(b)-f(pa+qb)}{f(a)+f(b)-2f(a+b)/2}\right] \\ &:= C_f(a,b), \end{aligned}$$

we get at once that

$$T_f(a,b) := \max_p [pf(a) + qf(b) - f(pa+qb)] \le C_f S_f(a,b) := \widehat{S}(a,b).$$

Proof of Theorem E. For fixed $[a,b] \subseteq D_f$, a way of constructing the sequences $\mathbf{x}_0$ and $\mathbf{p}_0$ can be the following: choose $x_1 = a$; $x_2 = x_3 = \cdots = b$, and let $p_1 = p_0$ where the number p_0 is defined in Remark 1. Hence, by the same remark, we obtain that

$$J_f(\mathbf{p}_0, \mathbf{x}_0) = T_f(a,b).$$

Proof of Theorem F. Applying the standard technique for finding extremals of a function of one variable, the desired result follows. Details are left to the reader.

Proof of Theorem G. By Theorem C, applied with $f(x) = -\log x$, we obtain

$$0 \le \log \frac{A(\widetilde{p},\widetilde{x})}{G(\widetilde{p},\widetilde{x})} \le T_{-\log x}(a,b) = \max_p [\log(pa+qb) - p\log a - q\log b].$$

By the standard argument, it is easy to find that the unique maximum is attained at point p_0 given by

$$p_0 = \frac{b}{b-a} - \frac{1}{\log b - \log a} = \frac{b - L(a,b)}{b-a}.$$

Since $0 < a < b$, we get $0 < p_0 < 1$ and, after some calculation, it follows that

$$0 \le \log \frac{A(\widetilde{p},\widetilde{x})}{G(\widetilde{p},\widetilde{x})} \le \log\left(\frac{b-a}{\log b - \log a}\right) - \log(ab) + \frac{b\log b - a\log a}{b-a} - 1.$$

Exponentiating, we obtain the first assertion from Theorem G.

By a change of variable $x_i \to (1/x_i)$, $i = 1, 2, \ldots$, the proof of the second proposition easily follows from the previous one since then

$$A\left(\mathbf{p}, \frac{1}{\mathbf{x}}\right) = \frac{1}{H(\mathbf{p}, \mathbf{x})}; \qquad G\left(\mathbf{p}, \frac{1}{\mathbf{x}}\right) = \frac{1}{G(\mathbf{p}, \mathbf{x})},$$

and

$$\Lambda_1(1/b, 1/a) = \Lambda_1(a, b).$$

Proof of Theorem H. Applying Theorems C and F with $f(t) = e^t$, $f(t) = t \log t$, the desired results follow.

Proof of Theorem I. A variant of inequality (6) asserts that

$$0 \le \log\left(\sum p_i x_i\right) - \sum p_i \log x_i \le \log \Lambda_1(a, b).$$

Putting there $x_i = q_i/p_i$, $i = 1, 2, \ldots$ with $a = m := \min\{x_i\}$; $b = M := \max\{x_i\}$, $i = 1, 2, \ldots$, and taking into account that $\sum q_i = 1$, the assertion from part (i) follows.

Taking the distribution Q to be uniform, the last result follows from the previous one by noting that

$$\Lambda_1\left(\frac{1}{r\nu}, \frac{1}{r\mu}\right) = \Lambda_1\left(\frac{1}{\nu}, \frac{1}{\mu}\right) = \Lambda_1(\mu, \nu).$$

REFERENCES

1. P.R. Beesack and J. Pecaric, *On Jensen's inequality for convex functions*, J. Math. Anal. Appl. **110** (1985), 536–552.

2. I. Budimir, S.S. Dragomir and J. Pecaric, *Further reverse results for Jensen's discrete inequality and applications in information theory*, J. Inequal. Pure Appl. Math. **2** (2001), Art. 5.

3. S.S. Dragomir, *A converse result for Jensen's discrete inequality via Grüss inequality and applications in information theory*, Anal. Univ. Oradea. Fasc. Math. **7** (1999–2000), 178–189.

4. ———, *Bounds for the normalized Jensen functional*, Bull. Aust. Math. Soc. **74** (2006), 471–476.

5. S.S. Dragomir and C.J. Goh, *Some bounds on entropy measures in information theory*, Appl. Math. Lett. **10** (1997), 23–28.

6. G.H. Hardy, J.E. Littlewood and G. Pölya, *Inequalities*, Cambridge University Press, Cambridge, 1978.

7. D.S. Mitrinovic, *Analytic inequalities*, Springer, New York, 1970.

8. S. Simic, *On a global upper bound for Jensen's inequality*, J. Math. Anal. Appl. **343** (2008), 414–419 (DOI 10.1016/j. jmaa. 2008.01.060).

9. ———, *Jensen's inequality and new entropy bounds*, Appl. Math. Letter. **22** (2009), 1262–1265.

Kneza Mihaila 36, 11000 Belgrade, Serbia
Email address: ssimic@turing.mi.sanu.ac.rs