Upper Bounds on the Expected Value of a Convex Function using Gradient and Conjugate Function Information

John R. Birge
Dept. of Industrial and Operations Engineering
The University of Michigan
Ann Arbor, MI 48109 USA

Marc Teboulle
Dept. of Mathematics, Statistics and Computing Science
Dalhousie University
Halifax, NS Canada
B3H 3J5

Technical Report 87-19

August 1987

Upper Bounds on the Expected Value of a Convex Function using Gradient and Conjugate Function Information.

John R. Birge*

Department of Industrial and Operation Engineering

The University of Michigan

Ann Arbor, MI 48109

USA

Marc Teboulle

Department of Mathematics, Statistics and Computing Science

Dalhousie University

Halifax, NS Canada

B3H 3J5

*Supported in part by Office of Naval Research Grant N00014-86-K-0628 and by Dalhousie University. This work was conducted while the first author was visiting the Department of Mathematics, Statistics and Computing Science, Dalhousie University.

Abstract: New upper bounds are given for the expected value of a convex function. The bounds employ subgradient information and the conjugate function. We derive the bounds and compare them with previous bounds with different information requirements.

Keywords: Bounds, Convex Functions, Stochastic Programs, Utility Functions

1 INTRODUCTION

Evaluating the expectation of a convex function is a central requirement in utility theory (see, for example, Fishburn [1970]) and stochastic programming (see, for example, Dempster [1980]). In general, these problems involve optimizing the expectation of some function of certain random variables and decision parameters. We assume that this function is convex and that certain properties of the convex function and the underlying probability measure are known. We show that new upper bounds on this expectation are available when the information includes subgradient and conjugate function information. This result is especially useful when the original integral is not easily computable as we show below.

The most basic bound on expectations of convex functions is Jensen's lower bound (Jensen [1906] which requires knowing only the finite means of the random variables. Madansky [1959], following Edmundson [1956], gave an upper bound based on the theory of moment spaces. This bound again requires finite mean value information and a bounded n-dimensional rectangular domain of stochastically independent random variables. Ben-Tal and Hochman [1972] extended and refined the Edmundson-Madansky bound by including information of the expected value of the absolute difference between the random variable and its mean.Gassmann and Ziemba [1986] provide a weaker bound that does not require independence (as in Dupacova [1974]) or n-dimensional bounded regions. Frauerdorfer [1986] provides the extension of the Edmundson-Madansky bound with dependencies and knowledge of the joint expectations of the random variables.

The general process of obtaining these bounds as solutions of moment problems is described in Birge and Wets [1986]. The solution of linear approximations is given in Birge and Wets [1987]. Explicit solution procedures also appear in Ermoliev, Gaivoranski and Nedeva [1987] and Cipra [1986]. They are also used in Dula [1986] to provide bounds for the expectation of convex functions with additional properties given first and second moment information.

Our results differ from the above results in our not requiring explicit moment information but instead information regarding the conjugate function and the expectation of the gradient and the inner product of the gradient and the random vector. We first give a one-dimenstional result in Section 2. Section 3 provides an extension in n-dimensions. Section 4 compares our bound with previous bounds in $\mathbb{R}$ and Section 5 provides the comparison in $\mathbb{R}^n$. Section 6 describes possible refinements, and Section 7 gives conclusions.

2. AN UPPER BOUND IN $\mathbb{R}$

Let (Ω, Σ, F) be a probability measure space and let X: $\Omega \rightarrow (a,b)$ be a random variable, where $-\infty \leqslant a < b \leqslant +\infty$, with distribution F.

Let ϕ: $(a,b) \rightarrow \mathbb{R}$ be a convex differentiable function. We denote expectation with respect to F by E and throughout this section we assume that $E\phi(X)$, $E\phi'(X)$ and $EX\phi'(X)$ exist and are finite.

<u>Theorem 2.1</u> Let $\phi: (a,b) \rightarrow \mathbb{R}$ be a convex differentiable increasing function and assume that $E\phi'(X) > o$. Then,

$$E\phi(X) \leqslant \phi\left[\frac{EX\phi'(X)}{E\phi'(X)}\right] := D \tag{2.1}$$

<u>Proof:</u> For any convex differentiable function ϕ on (a,b), the following inequality holds:

$$\phi(s) - \phi(t) \geqslant (s-t)\ \phi'(t) \quad \text{for all } s,t \in (a,b) \tag{2.2}$$

Set $t=X$ [clearly in (a,b)], $s = \dfrac{EX\phi'(X)}{E\phi'(X)}$.

Since ϕ is increasing and $E\phi'(X) > o$, $s \in (a,b)$. Substituting s,t in (2.2) we obtain

$$\phi(X) \leqslant \phi\left[\frac{EX\phi'(X)}{E\phi'(X)}\right] - \left[\frac{EX\phi'(X)}{E\phi'(X)} - X\right] \phi'(X) \tag{2.3}$$

Take expectation on both sides of (2.3) with respect to F and observe that in the right hand side of (2.3) (with $E\phi'(X) > o$):

$$\frac{EX\phi'(X)}{E\phi'(X)}\, E\phi'(X) - EX\phi'(X) = 0.$$

The result follows. □

Remarks 2.1

(1) If ϕ is strictly increasing and concave then inequality (2.1) is reversed.

(2) If ϕ is strictly decreasing, then assuming $E\phi'(X)<o$, inequality (2.1) is still valid.

(3) The differentiablity assumption on ϕ can be relaxed. For if ϕ is convex its left and right derivatives $\phi'_-(x)$ and $\phi'_+(x)$ exist, and are finite and non-decreasing. Moreover the subdifferential of ϕ is given by, $\partial\phi(x) = \{z\in\mathbb{R} : \phi'_-(x)\leqslant z\leqslant\phi'_+(x)\}$ (see e.g Rockafellar [1970], pp. 228-229).

Theorem 2.1 remains valid if we substitute any $z \in \partial\,\phi(x) = [\phi'_-(x), \phi'_+(x)]$ for $\phi'(x)$.

□

Jensen's inequality for a convex function ϕ provides us with a lower bound for $E\phi(X)$:

$$\phi(E(X)) \leqslant E\phi(X) \tag{2.4}$$

Combining inequality (2.4) with Theorem 2.1 allow us to derive a re-arrangment type inequality

Corollary 2.1 Under the assumptions of Theorem 2.1, we have

$$EX\phi'(X) \geqslant E(X)\, E\phi'(X) \tag{2.5}$$

Proof Simply follows from (2.1), combined with (2.4), and using that ϕ is increasing and $E\phi'(X) > o$. □

More generally, let g: (a,b)→$\mathbb{R}$ be a given increasing function. Since ϕ is convex, ϕ' is increasing and so $f(t):=\phi'(g(t))$ is increasing. Then inequality (2.5) implies

$$Eg(X)f(X) \geqslant Eg(X)\,Ef(X) \qquad (2.6)$$

Inequalities (2.5) or (2.6) can be used to obtain bounds on system reliability. For general results on rearrangement inequalities and applications see Karlin and Rinotl [1981] and the references therein.

3. AN UPPER BOUND IN $\mathbb{R}^n$

In this section we present a natural extension in $\mathbb{R}^n$ of the upper bound derived in Theorem 2.1.

Let X be a random vector on the probability space (Ω, Σ, F) with distribution function F and let $S \subset \mathbb{R}^n$ be the support of X.

Assume that S is convex and let ϕ: S→$\mathbb{R}$ be a convex differentiable function. The gradient of ϕ at x is denoted by $\nabla\phi(x)$. The conjugate convex function of ϕ is defined by

$$\phi^*(y) = \sup_x (x^t y - \phi(x))$$

In the sequel we assume that $E\phi(X)$, $EX^T\nabla\phi(X)$ and $E\nabla\phi(X)$ exist and are finite.

<u>Theorem 3.1</u> $E\phi(X) \leqslant EX^T\nabla\phi(X) - \phi^*(E\nabla\phi(X))$ (3.1)

<u>Proof</u> Since ϕ: S→$\mathbb{R}$ is convex and differentiable, the gradient inequality holds, i.e., $\phi(\alpha) - \phi(\beta) \geqslant (\alpha-\beta)^T\nabla\phi(\beta)$ for all $\alpha, \beta \in S$. (3.2)

Setting $\beta = X$ in (3.2) and taking expectation with respect to F in inequality (3.2) implies

$$E\phi(X) \leqslant EX^T \nabla\phi(X) + \phi(\alpha) - \alpha^T E\nabla\phi(X) \text{ for all } \alpha \in S.$$

Hence,

$$E\phi(X) \leqslant EX^T \nabla\phi(X) + \inf_\alpha (\phi(\alpha) - \alpha^T E\nabla\phi(X)). \tag{3.3}$$

Note that:

$$\inf_\alpha (\phi(\alpha) - \alpha^T E\nabla\phi(X)) = -\sup_\alpha (\alpha^T E\nabla\phi(X) - \phi(\alpha)) = -\phi^* (E\nabla\phi(X)).$$

Inequality (3.1) follows immediately from (3.3). □

<u>Remark (3.1)</u> An alternative proof of Theorem 3.1 may be derived using the following useful relation: (see Rockafellar [1970], p. 257)

$$\phi^* (\nabla\phi(z)) = z^T \nabla\phi(z) - \phi(z) \tag{3.4}$$

Setting z=X and taking expectation in (3.4) we obtain

$$E\phi(X) = EX^T \nabla\phi(X) - E\phi^* (\nabla\phi(X) \tag{3.5}$$

But since ϕ is convex so is ϕ^* and hence by Jensen's inequality

$$\phi^*(E\nabla\phi(X)) \leqslant E\phi^*(\nabla\phi(X)) \tag{3.6}$$

Then (3.5) combined with (3.6) implied (3.1). This proof will be useful to refine the upper bound; see Section 6. □

<u>Remarks 3.2</u>

(1) If ϕ is concave Inequality 3.1 is reversed.

(2) As mentioned in Remarks 2.1 (3), Theorem 3.1 remains valid if instead of $\nabla\phi(.)$ we substitute any $z \in \partial \phi$. □

The one dimensional version of Theorem 3.1 (n=1, S=(a,b)) provides us with the upper bound $E\phi(X) \leqslant EX\phi'(X) - \phi^*(E\phi'(X)) := C$. The next result shows that the bound C is better than the upper bound D derived in Theorem 2.1.

<u>Theorem 3.2</u> Under assumptions of Theorem 2.1, we have

$$E\phi(X) \leqslant C \leqslant D \tag{3.7}$$

<u>Proof</u> For any convex functio ϕ and any $\alpha \in \text{dom}\ \phi$, $\beta \in \text{dom}\ \phi^*$ the inequality

$$\phi(\alpha) + \phi^*(\beta) \geqslant \alpha\beta \tag{3.8}$$

holds. Substituting in (3.8) $\alpha = \dfrac{EX\phi'(X)}{E\phi'(X)} \in (a,b) = \text{dom}\ \phi$ and $\beta = E\phi'(X) \in \text{range}\ \phi' \subset \text{dom}\ \phi^*$, the result (3.7) follows. □

4. COMPARISONS OF BOUNDS IN $\mathbb{R}$

Throughout this section $\phi: \mathbb{R} \to \mathbb{R}$ is a given convex differentiable convex function and X is random variable with distribution F and density f with support (a,b). We compare the upper bounds

$$C = EX\phi'(X) - \phi^*(E\phi'(X)) \text{ and } D = \phi\left(\frac{EX\phi'(X)}{E\phi'(X)}\right)$$ with the following

well known upper bounds:

<u>Edmundson-Madansky (1956)</u>

$$EM := \frac{(b-\bar{x})\,\phi(a) + (\bar{x}-b)\,\phi(b)}{b-a} \tag{4.1}$$

where $\bar{x} := E(X) < \infty$, [a,b] is a finite interval.

Ben-Tal-Hochman (1972)

$$BH := \frac{d}{2}\left[\frac{\phi(a)}{\bar{x}-a} + \frac{\phi(b)}{b-\bar{x}}\right] + \phi(\bar{x})\left[1 - \frac{d}{2}\frac{(b-a)}{(b-\bar{x})(\bar{x}-a)}\right] \tag{4.2}$$

where $\bar{x} = E(X) < \infty$, $[a,b]$ is a finite interval, and

$$d := E|X - \bar{x}| = 2\int_{\bar{x}}^{b} (x-\bar{x})\, dF(x) = 2\int_{a}^{\bar{x}} (\bar{x}-x)\, dF(x)$$ is the

expected absolute deviation about the mean. Using this additional information on the random variable X, it is shown that BH gets closer to $E\phi(X) := \bar{\phi}$ than EM i.e, $\bar{\phi} \leq BH \leq EM$

Remark 4.1 The upper bound BH can be obtained for an infinite interval (a,b), $-\infty \leq a < b \leq +\infty$, under additionals assumptions on ϕ; see Ben-Tal and Hochman (1972). □

Example 4.1 $\phi(X) = x^2$, $x > 0$, $X \sim U(0,1)$. Then $\bar{x} = 1/2$, $d = 1/4$, $\bar{\phi} = 1/3$. Using (4.1) and (4.2) we obtain $EM = 1/2$ and $BH = 3/8$. Here $\phi^*(y) = 1/4\, y^2$ and then we compute $C = 5/12$, $D = 4/9$. Jensen's inequality yeilds the lower bound $J = 1/4$ and $J \leq \bar{\phi} \leq BH \leq C \leq D \leq EM$. □

The next example illustrates a situation where $EX\phi'(X)$ and $E\phi'(X)$ (and hence C and D) are easy to compute while $E\phi(X)$ requires the evaluation of a complicated integral. Moreover, in this example, the upper bounds BH and EM are shown to be trivial, i.e., $BH = EM \equiv +\infty$.

Example 4.2 Let $\phi(x) = -\mathrm{Ln}\,(1-x^2)$ $-1 \leq x \leq 1$, and let X be a random variable with density $f(x) = 3/2(1-x^2)$ for $0 \leq x \leq 1$.

$$\text{Then, } \bar{\phi} = -\frac{3}{2}\int_0^1 (1-x^2)\,\mathrm{Ln}\,(1-x^2)\,dx \tag{4.3}$$

To compute the integral (4.3) we use the following known integrals (see e.g. Gradshteyn and Ryzhik (1980) pp. 557-558):

$$\int_0^1 x^{\lambda-1}\,\mathrm{Ln}\,(1+x)\,dx = 1/\lambda\,(\mathrm{Ln}2 - \beta(\lambda+1)) \quad (\lambda > -1)$$

$$\int_0^1 x^{\lambda-1}\,\mathrm{Ln}\,(1-x)\,dx = 1/\lambda\,(\psi(1) - \psi\,(\lambda+1)) \quad (\lambda > -1)$$

where $\psi(1) = -\gamma$, $\gamma = 0.577215...$Euler's constant and the functions ψ, β satisfy

$$\psi(x+1) = \psi(x) + \frac{1}{x},\ \psi\left(\frac{x+1}{2}\right) - \psi\left(\frac{x}{2}\right) = 2\beta(x),\ x > 0 \quad \text{(see e.g. (12) p. 945).}$$

After some algebraic manipulations one obtains $\bar{\phi} = 5/3 - 2\mathrm{Ln}2 \approx 0.280372$.

The conjugate function is given by

$$\phi^*(y) = \sqrt{1+y^2} - 1 + \mathrm{Ln}\left(1 - \left(\frac{\sqrt{1+y^2}-1}{y}\right)^2\right).$$

$$EX\phi'(X) = 3\int_0^1 x^2 dx = 1,\ E\phi'(X) = 3\int_0^1 x\,dx = 3/2.$$

Hence, $\phi'(E\phi'(X)) = \phi^*(3/2) \approx 0.4653$ and thus $C \simeq 0.53468$ and $D = \phi\,(2/3) = 9/5 \simeq 0.58778$.

Jensen's inequality yields the lower bound $J = \phi(\bar{x}) = \phi(3/8) \simeq 0.1515$. A rough estimate of $\bar{\phi}$ could be obtained by averaging the upper bound C and the lower bound J to give 0.34309.

Finally we note that since here $\phi(1) = +\infty$, $EM = BH \equiv +\infty$. (It can be verified that $d \neq 0$.) $\square$

In Theorem 3.2 we show that $C \leqslant D$, and Ben-Tal and Hochman prove that $BH \leqslant EM$. Examples 4.1 and 4.2 illustrate situations where $C \leqslant D \leqslant EM$. In fact we tested many other examples and found $C \leqslant EM$. This inequality is not, however, always valid as illustrated in our next example.

Example 4.3 Let $\phi(X) = 1 - (1 - (x-1)^2)^{1/2}$, $0 \leqslant X \leqslant 2$ and

$$f(X) = \frac{2}{4-\pi}\left(1 - (1-(x-1)^2)^{1/2}\right), \quad 0 \leqslant X \leqslant 2.$$

Then, $\bar{x} = 1$, $EM = 1$, $BH \approx 0.7766$, $J = \phi(\bar{x}) = 0$ and $\bar{\phi} = E\phi(X) \simeq 0.447$. The conjugate ϕ^* is

$$\phi^*(y) = y\left(1 + \frac{y}{\sqrt{1+y^2}}\right) + \frac{1}{\sqrt{1+y^2}} - 1, \qquad (4.3)$$

$$EX\phi'(X) = \frac{3\pi - 4}{3(4-\pi)} \simeq 2.1065, \text{ and } E\phi'(X) = 0.$$

Hence $\phi^*(E\phi'(X)) = \phi^*(0) = 0$ and then $C \simeq 2.1065 > EM = 1$. □

Note that the bound D is not computable here since the assumption of Theorem 2.1 $E\phi'(X) > 0$ is violated. The example not only demonstrates that EM is better than C, but also that the bound C may be a "bad" upper bound. However, we show in Section 6 that the bound C can be considerably improved to be even sharper than BH; see Theorem 6.1 and Example 6.2.

We have already mentioned in the introduction, that the computation of the upper bounds EM and BH requires a <u>finite mean</u> $\bar{x}$. This is not the case for C. Further BH requires the value of d which may be difficult to compute. In the last example of this section we consider the case when $\bar{x}$ and $\bar{d}$ fail to exist and therefore the upper bounds EM and BH are not available.

Example 4.4 Let X be a random variable with density

$$f(x) = \frac{2}{\pi(1 + x^2)} \qquad 0 < x < +\infty \text{ and let } \phi(x) = -2\sqrt{x}.$$

$$\text{Then } \bar{\phi} = \frac{-4}{\pi}\int_0^{\infty} \frac{\sqrt{x}}{1 + x^2}\, dx = -2\sqrt{2} \simeq -2.8284, \text{ and } EX\phi'(X) = -\sqrt{2},$$

$EX\phi'(X) = -\sqrt{2}$. Hence $C = -\sqrt{2} - \phi^*(-\sqrt{2}) = \frac{3\sqrt{2}}{2} \simeq -2.1213$. □

5 COMPARISONS OF BOUNDS IN $\mathbb{R}^n$

Gassmann and Ziemba (1986) extend an idea of Edmundson and Madansky (1956) to derive an upper bound on the expected value of a convex function of a random vector. The bound is given as the solution of the following linear program: (see Gassmann, Ziemba (1986),Theorem 1)

$$GZ = \max_{\lambda} \left(\sum_{i=1}^{m} \phi(v_i)\lambda_i : \sum_{i=i}^{m} \lambda_i = 1, \sum_{i=1}^{m} \lambda_i v_i = \bar{x}, \lambda_i \geq 0 \right) \qquad (5.1)$$

where ϕ: $S \to \mathbb{R}$ is a convex function, $S \subset \mathbb{R}^n$ is convex, $(v_1, \ldots, V_m)$ are the extreme points of a bounded convex polyhedron containing S, and $\bar{x} = (EX_1, \ldots, EX_n)^T$ is the finite mean of the random vector X. We compare the bound GZ with the upper bound derived in Theorem 3.1: $C = EX^T \nabla\phi(x) - \phi^*(E\nabla\phi(x))$.

Example 5.1 (Taken from Gassman-Ziemba [1986] p. 42.)

$$\phi(x_1,x_2) = e^{x_1}, \quad f(x_1,x_2) = \begin{cases} 3(\sqrt{1-x_1^2-x_2^2})/2\pi & \text{if } x_1^2+x_2^2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Then $S = \{(x_1, x_2) : x_1^2 + x_2^2 \leqslant 1\}$ and

$$\bar{\phi} = Ee^{X_1} = \frac{3}{2\pi} \int_{-1}^{1} \int_{-\sqrt{1-x_1^2}}^{\sqrt{1-x_1^2}} e^{x_1} \sqrt{1 - x_1^2 - x_2^2}\, dx_1\, dx_2 = \frac{3}{e} \simeq 1.0364 \qquad (5.2)$$

Using (5.1) the best upper bound derived in [11] is shown to be $GZ \simeq 1.54308$. We now compute the bound C. The conjugate of $\phi(x_1, x_2) = e^{x_1}$ is $\phi^*(y_1, y_2) = y_1 \,Ln\, y_1 - y_1$ and $E\nabla\phi(X) = (Ee^{X_1}, 0)^T =$

$(\frac{3}{e}, 0)^T$ (using 5.2), then $\phi^*(E\nabla\phi(X)) = \phi^*(\frac{3}{e}, 0) = (\frac{3}{e}(Ln\, 3-2) \simeq -0.9948$.

Now,

$$EX\nabla\phi(X) = E(X_1 e^{X_1}) = \frac{3}{2\pi} \int_{-1}^{1} \int_{-\sqrt{1-x_1^2}}^{\sqrt{1-x_1^2}} x_1 e^{x_1} \sqrt{1-x_1^2 - x_2^2}\, dx_2\, dx_1 =$$

$\frac{3e^2-21}{2e} \simeq 0.2146$ and then $C \simeq 1.20948 < GZ \simeq 1.54308$.

Note that Jensen's inequality yields the lower bound $J = \phi(\bar{x}) = \phi(0,0) = 1$ and thus an estimate of $\bar{\phi}$ could be obtained by

$\frac{J + C}{2} \simeq 1.10474$ giving an error of about 7%. $\square$

For a random vector $X = (X_1, \ldots X_n)^T$ with <u>independent</u> components X_i the Edmundson-Madansky and Ben-Tal-Hochman upper Bounds are available where S is an n-dimensional rectangle of the form $S = \prod_{i=1}^{n} [a_i, b_i]$. They are given by the following expressions [1]:

$$EM = \sum_{i_1=0}^{1} \ldots\ldots \sum_{i_n=0}^{1} \Pi_{k=1}^{n} \gamma_{ik}^{k} \phi(c_{i_1}^{1}, \ldots, c_{i_n}^{n}) \text{ where } c_0^i = a_i \text{ , } c_1^i = b_i, \gamma_0^k = \frac{b_k - \bar{x}_k}{b_k - a_k}$$

$$\text{and } \gamma_1^k = \frac{\bar{x}_k - a_k}{b_k - a_k} \tag{5.3}$$

$$BH = \sum_{\Delta_3} (\Pi_{k=1}^{m} P_{\delta_k}^{k}) \phi (a_{\delta_1}^{1}, \ldots\ldots, a_{\delta_n}^{n}) \tag{5.4}$$

where δ_k is 1 or 2 or 3, Δ_3 is the set of 3^n n-dimensional vectors whose components are all 1's and/or 2's and/or 3's,

$$P_1^k = \frac{d_k}{2(\bar{x}_k - a_k)}, \; P_2^k = \frac{d_k}{2(b_k - \bar{x}_k)}$$

$$P_3^k = 1 - P_1^k - P_2^k, \; a_1^k = a_k, \; a_2^k = b_k, \quad a_3^k = \bar{x}_k.$$

As mentioned in [11], for the independent case the upper bound GZ is in fact worse than EM (and therefore BH).

Example 5.2.

$$\text{Let } \phi(x_1, x_2) = \frac{5x_1^2}{2} + \frac{x_2^2}{2} + x_1 x_2 - x_1 - x_2,$$

$$f(x_1, x_2) = \begin{cases} 1 \text{ if } 0 \leqslant x_1, x_2 \leqslant 1 \\ 0 \quad \text{otherwise} \end{cases}$$

Then, $\bar{\phi} = E\phi(x_1, x_2) = \frac{1}{4}$. Using (5.3) and (5.4) we obtain respectively $EM = \frac{1}{4}(\phi(0,0) + \phi(0,1) + \phi(1,0) + \phi(1,1)) = \frac{3}{4}$

and $BH = \frac{1}{16}(\phi(0,0) + \phi(0,1) + \phi(1,0) + \phi(1,1)) + \frac{1}{8}(\phi(0, \frac{1}{2}) + \phi(\frac{1}{2}, 0) + \phi(1, \frac{1}{2}) + \phi(\frac{1}{2}, 1)) + \frac{1}{4}\phi(\frac{1}{2}, \frac{1}{2}) = \frac{3}{8}$. From (5.1) we have:

$$GZ = \max\left(-\frac{\lambda_2}{2} + \frac{3\lambda_3}{2} + 2\lambda_4 : \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 1, \lambda_3 + \lambda_4 = \frac{1}{2},\right.$$

$$\left.\lambda_2 + \lambda_4 = \frac{1}{2}\right) = \max\left(1 - \lambda_2 : 0 \leqslant \lambda_2 \leqslant \frac{1}{2}\right) = 1$$

Now we compute the conjugate function at the point $x^* = (x_1^*, x_2^*)^T$

$$\phi^*(x_1^*, x_2^*) = \frac{1}{2}(x^*+e)^T A(x^*+e),\ A = \frac{1}{4}\begin{pmatrix}1 & -1\\ -1 & 5\end{pmatrix},\ e = (1,1)^T.$$

and $E\nabla\phi(x_1, x_2) = (2,0)^T$, $Ex^T\nabla\phi(x) = \frac{3}{2}$. Then $\phi^*(E\nabla\phi(x)) = \phi^*(2,0) = 1$. This yields the upper bound $C = \frac{1}{2}$, and we have $\bar{\phi} < BH < C < EM < GZ$.

Finally Jensen's inequality yields the lower bound $J = \phi(\bar{x}) = \phi(\frac{1}{2}, \frac{1}{2}) = 0$. (Note incidently that $\frac{J + C}{2} = \bar{\phi} = \frac{1}{4}$!)

□

The bound C is in fact sharp (i.e., $C = \bar{\phi}$) for many examples. For example, consider a piecewise linear function of the form:

$$\phi(x) = \sup_{1 \leq i \leq k} (x-t)^T \pi^i \qquad (5.5)$$

where t is fixed. This is the form of the recourse function in stochastic linear programming (see Wets (1966)). In this case,

$$\phi^*(v) = \inf \sum_{i=1}^{k} \lambda^i \phi_i^* (v_i) : \sum \lambda_i = 1, \sum \lambda_i v_i = \mathbf{v}, \lambda_i \geqslant 0\} \qquad (5.6)$$

from Rockafellar (1970), Theorem 16.5, where

$$\phi_i^* (v^i) = \sup (v^T x - (x-t)^T \pi^i) \qquad (5.7)$$

$$= \begin{cases} \pi^i t \ \text{ if } v^i = \pi^i \\ + \infty \text{ otherwise} \end{cases}$$

From (5.7),

$$\phi^*(v) = \begin{cases} vt \text{ if } v \in \text{co } (\pi^i \,|\, i = 1, \ldots, k) \\ + \infty \text{ otherwise.} \end{cases} \qquad (5.8)$$

Let $\phi(x) = (x-t)^T \pi(x)$ ($\pi(x)$ not necessarily unique) and let

$\nabla\phi(x) = \pi(x)$, then $E(\phi(x) = E((x-t)^T \pi(x))$

$$= E(x^T \nabla\phi(x)) - E(t^T \pi(x))$$

$$= E(x^T \nabla\phi(x)) - \phi^*(E(\nabla\phi(X))),$$

where we note that $E(\nabla\phi(x)) \in co(\pi^i \mid i=1, \ldots, k)$. Hence, $\phi = \bar{C}$. (5.9)

6. REFINING THE UPPER BOUNDS IN $\mathbb{R}$ AND $\mathbb{R}^n$

The upper bound C derived in Theorem 3.1 can be naturally sharpened in the one dimensional case, when X is a continuous random variable with density function f(x).

Theorem 6.1

Let ϕ: $\mathbb{R} \to \mathbb{R}$ be a convex differentialbe function and X a random variable with support (a,b) and density f(x). Then,

$$E\phi(X) \leq \frac{1}{2} (K(a,b) - \int_a^b x\phi(x) f'(x)\, dx - \phi^* (E\phi(X)) := \tilde{C}. \qquad (6.1)$$

where $K(a,b) := b\phi(b)f(b) - a\phi(a)f(a)$ (6.2)

Moreover $\tilde{C}$ is sharper than C, i.e., $\tilde{C} \leq C$ (6.3)

Proof:

From Theorem 3.1 in $\mathbb{R}$, $E\phi(x) \leq C = EX\,\phi'(X) - \phi^*(E\phi'(X))$ (6.4)

Integrating by part $EX\,\phi'(X) = \int_a^b x\,\phi'(x)\, f(x)\, dx$ we obtain from (6.4):

$$E\phi(x) \leq (x\phi(x)\, f(x))_a^b - \int_a^b \phi(x)\, f(x)\, dx - \int_a^b x\phi(x) f'(x) dx - \phi^*(E\phi'(X))$$

$$= K(a,b) - E\phi(X) - \int_a^b x\phi(x)\, f'(x)\, dx - \phi^*(E\phi'(X)) \qquad (6.5)$$

and then (6.1) follows. To show that $\tilde{C} \leqslant C$, observe that

$\tilde{C} = \frac{1}{2}(EX\phi'(X) + E\phi(X) - \phi^*(E\phi'(X)) = \frac{1}{2}(C + E\phi(X))$ and that $E\phi(X) \leqslant C$, implying (6.3) □

Example 6.1

We reconsider Example 4.3 given in Section 4.

$K(a,b) = K(0,2) = \frac{8}{4-\pi}$ and $\int_0^2 x\phi(x)f'(x)\,dx = \frac{3\pi-4}{3(4-\pi)}$. This yields the upper bound $\tilde{C} = \frac{16-3\pi}{6(4-\pi)} \simeq 1.2766 < C \simeq 2.1065$. Note that we still have $EM=1 < \tilde{C}$; but see Example 6.2. □

It is interesting to note that when $X \sim U(a,b)$ then the upper bound $\tilde{C}$ is better than EM. For if $X \sim U(a,b)$ then $EM = \frac{\phi(a) + \phi(b)}{2}$

and $\tilde{C} = \frac{1}{2}(b\,\zeta + \phi(a) - \phi^*(\zeta))$ where $\zeta := \frac{\phi(b)-\phi(a)}{b-a} \in (\phi'(a), \phi'(b))$

(since ϕ is convex).

Hence, from the inequality $\phi(b) + \phi^*(\zeta) \geqslant b\,\zeta$, it follows immediately that $\tilde{C} \leqslant EM$.

Following Remark 3.1 of Section 3, we can further sharpen the upper bound C by using a lower bound established by Ben-Tal and Hochman (1972) which is better than Jensen's lower bound. Let $g: \mathbb{R} \to \mathbb{R}$ be a convex function, and Y a random variable with support (a,b) then:

$$Eg(Y) \geq \beta g(\bar{y} + \frac{d}{2\beta}) + (1-\beta)\, g(\bar{y} - \frac{d}{2(1-\beta)}) := L_g(\beta,d) \qquad (6.6)$$

and $L_g(\beta,d) \geq g(E(Y))$, where $\bar{y} = E(Y) < \infty$, $d = E|Y - \bar{y}|$ and $\beta := Pr(Y \geq \bar{y})$.

From (3.5) we have $E\phi(X) = EX\phi'(X) - E\phi^*(\phi'(X))$ (6.7)

Applying (6.6) with $g := \phi^*$ and $Y := \phi'(X)$ one obtain: (we denote $\bar{\phi}' := E\phi'(X)$)

$$E\phi^*(\phi'(X)) \geq \beta\phi^*(\bar{\phi}' + \frac{d}{2\beta}) + (1-\beta)\,\phi^*(\bar{\phi}' - \frac{d}{2(1-\beta)}) \geq (\phi^*(E\phi'(X))) \qquad (6.8)$$

and hence from (6.7) this implies

$$E\phi(X) \leq EX\,\phi'(X) - L_{\phi^*}(d,\beta) := C^R \leq C. \qquad (6.9)$$

where $\beta = Pr(\phi'(X) \geq \bar{\phi}')$ and $d = E|\phi'(X) - \bar{\phi}'|$.

Furthermore, if X is a random variable with density f(x), even sharper bounds are possible. Following the proof of Theorem 6.1 together with (6.9) we obtain

$$\tilde{C}_1^R := \frac{1}{2}(K(a,b) - \int_a^b x\phi(x)f'(x)dx - L_{\phi^*}(d,\beta)) \qquad (6.10)$$

Clearly, $\tilde{C}_1^R \leq C_1^R$ and $\tilde{C}_1^R \leq \tilde{C}_1$. A natural question is whether $\tilde{C}_1^R$ or C_1^R is better than BH? It seems difficult (if not impossible) to prove such a result in general. However in our worst examples (4.3 and 6.1) this appears to be true as demonstrated below.

Example 6.2

We have already computed in Example 4.3 and 6.1 $\bar{\phi} \simeq 0.447$
$BH \simeq 0.7766$, $EX\phi'(X) \simeq 2.107$, $K(0,2) = \frac{8}{4-\pi} \int_0^2 x\phi(x)f'(x)dx = \frac{3\pi-4}{3(4-\pi)}$

For the random variable $\phi'(X)$ we compute $\beta = \frac{1}{2}$ and $d = \frac{2}{4-\pi}$.

Then using the conjugate ϕ^* given in (4.3) we obtain

$L_{\phi^*}(d, \beta) = \frac{1}{2}(\phi^*(d) + \phi^*(-d)) \simeq 1.5354$. Using (6.9) and (6.10)

it follows that $C_1^R \simeq 0.5711$ and $\tilde{C}_1^R = 0.5088$. Thus,

$\bar{\phi} < \tilde{C}_1^R < C_1^R < BH < EM < \tilde{C} < C$. □

We now turn to the problem of refining the bound C in $\mathbb{R}^n$. For a random vector $X = (x_1, \ldots, X_n)$ with independent components X_i the Ben-Tal Hochman lower bound is available when S is an n dimensional rectangle

of the form $S = \prod_{i=1}^{n} (a_i, b_i)$; and is given by:

$$L_h := \sum_{\Delta_2} \left(\prod_{k-1}^{n} \beta_{\delta_1}^{k} \right) h(a_{\delta_1}^{1}, \ldots, a_{\delta_n}^{n}) \tag{6.11}$$

where $h: \mathbb{R}^n \to \mathbb{R}$ is a convex function, δ_k is 1 or 2, Δ_2 is the

set of 2^n n-dimensional vectors whose components are all

1's or/and 2's, $\beta_1^k = \beta_2^k = 1 - \beta_k$, $a_1^k = \bar{x}_k + \frac{d_k}{2\beta_k}$, $a_2^k = \bar{x}_k - \frac{d_k}{2(1-\beta_k)}$

and d_i, β_i, $\bar{x}_i$ denote the corresponding parameters of X_i.

To apply (6.11) as in (6.7) and (6.8) requires showing that the random vector

$\nabla \phi(x_1,\ldots,x_n)$ has independent components, and computing the corresponding parameters β_k and d_k. This fairly complicated task can be avoided by characterizing the class of functions for which the function $\psi(\cdot):=\phi^*(\nabla\phi(\cdot))$ is convex. Moreover this allows us to express the new upper bound explicitly in terms of the problem's data (without requiring knowledge of ϕ^*); see Theorem 6.2. The following result gives a sufficient condition for ψ to be convex for a large class of functions arising in applications.

Lemma 6.1

Let $\phi : S \subset \mathbb{R}^n_+ \to \mathbb{R}$ be a given twice continuously differentiable function. If $g: \mathbb{R}^n_+ \to \mathbb{R}^n$, $g(z) := \nabla \phi(z)$ is convex, then $\psi(z) := \phi^*(\nabla\phi(z))$ is convex.

Proof

ψ is a convex function if and only if it satisfies the gradient inequality, i.e., for any x and y in S

$$\psi(x) - \psi(y) \geq (x-y)^T \psi(y). \tag{6.12}$$

By definition, $\psi(y) = \phi^*(\nabla \phi(y))$ and thus $\nabla \psi(y) = \nabla^2 \phi(y) \nabla\phi^*(\nabla\phi(y))$

where $\nabla^2 \phi(y)$ denotes the Hessian of ϕ. Using $\nabla \phi^* = (\nabla \phi)^{-1}$,

it follows that $\nabla\psi(y) = \nabla^2 \phi(y).y$. Inequality (6.12) then

becomes:

$$\phi^* (\nabla \phi(x)) - \phi^* (\nabla \phi(y)) \geqslant (x-y)^T \nabla^2 \phi(y).y \qquad (6.13)$$

Now, since ϕ^* is convex, applying the gradient inequality to ϕ^* and using $\nabla \phi^* = (\nabla\phi)^{-1}$, we obtain

$$\phi^* (\nabla \phi(x)) - \phi^* (\nabla \phi(y)) \geqslant (\nabla \phi(x) - \nabla\phi(y))^T y.$$

Thus to prove (6.13) it is sufficient to show

$$(\nabla\phi(X)-\nabla\phi(y))^T y \geqslant (x-y)^T\nabla^2 \phi(y)\, y \qquad (6.14)$$

Since $g(x) := \nabla\phi(x)$ is assumed convex, we have

$$\nabla\phi(x) - \nabla\phi(y) \geqslant (x-y) \quad \nabla^2 \phi(y). \qquad (6.15)$$

Multiplying (6.15) by $y \geqslant 0$ (recall that $S \subset \mathbb{R}^n_+$) yields (6.14) .

□

We can now derive a refined upper bound for a random vector X with independent components X_i. We make the following assumptions:

(I) $S = \prod_{i=1}^{n} (a_i, b_i) \subset \mathbb{R}^n_+$

(II) $\nabla\phi(\cdot)$ is convex

<u>Theorem 6.2</u>

Suppose (I) and (II) hold. Then,

$$E\phi(X) \leqslant EX^T\nabla\phi(X) - L_\psi := C^n \text{ where} \qquad (6.16)$$

$$L_\psi = \sum_{\Delta_2} \left(\prod_{k-1}^{n} \beta^k_{\delta_k}\right) (a^1_{\delta_1},\dots,a^n_{\delta_n})^T\nabla\phi\, (a^1_{\delta_1},\dots, a^n_{\delta_n}) - \phi\, (a^1_{\delta_1},\dots, a^n_{\delta_n})) \qquad (6.17)$$

Proof

From (6.7) (in $\mathbb{R}^n$) we have

$$E\phi(X) = EX^T\nabla\phi(X) - E\phi^*(\nabla\phi(X))$$

Under Assumptions (I) and (II), Lemma 6.1 is applicable and thus

$E\psi(X) = E\phi^*(\nabla\phi(X)) \geq L_\psi$ as defined in (6.11). Moreover,

using $\psi(z) = z^T\nabla\phi(z) - \phi(z)$, the expression (6.17) follows. □

Example 6.3

We reconsider Example 5.2 where we already computed

$\bar{\phi} = 0.25$, EM = 0.75, C = 0.50, BH = 0.375, GZ = 1, $EX^T\nabla\phi(X) = 1.5$ and $d_1 = d_2 = \frac{1}{4}$, $\bar{x}_1 = \bar{x}_2 = \frac{1}{2}$. The assumptions (I) and (II) are clearly satisfied and thus Theorem 6.2 is applicable.

We compute $\beta_1 = \beta_2 = \frac{1}{2}$ and thus using (6.17)

$$L_\psi = \frac{1}{4}\left(\psi(\tfrac{3}{4},\tfrac{3}{4}) + \psi(\tfrac{3}{4},\tfrac{1}{4}) + \psi(\tfrac{1}{4},\tfrac{3}{4}) + \psi(\tfrac{1}{4},\tfrac{1}{4})\right)$$

with $\psi(z_1,z_2) = (z_1,z_2)^T\,\nabla\phi(z_1,z_2) - \phi(z_1,z_2) = \frac{5}{2}z_1^2 + \frac{1}{2}z_2^2 + z_1 z_2$

Then $L_\psi = 1.1875$ which yields the upper bound $C^R = 0.3125 <$ BH=0.375.

7. CONCLUSIONS

We have given new upper bounds for the expectation of convex function using gradient and convex conjugate function information. We have shown that these bounds and their extensions can be better than previous bounds in several examples. We also demonstrated how our bounds are especially useful when the original integral is complicated but has a gradient that can be easily integrated or when the information required for other bounds (e.g., moments) is not available. The new bounds are then applicable in a variety of applications with these characteristics.

References

1. Ben-Tal, A. and Hochman, E. (1972) More bounds on the expectation of a convex function of a random variable. J. Appl. Prob. 9, 803-812.

2. Birge, J.R. and Wets, R.J.B. (1986) Designing approximation schemes for stochastic optimization problems, in particular for stochastic programs with recourse. Math. Prog. Stud. 27, 54-102.

3. Birge, J.R. and Wets, R.J.B. (1987) Computing bounds for stochastic programming problems by means of a generalized moment problem. Math. of Operations Res. forthcoming.

4. Cipra, T. (1986) Moment problem with given covariance structure in stochastic programming. cosopis pro Pestovani Matematiky, forthcoming.

5. Dempster, M.A.H. (1980) Introduction to stochastic programming in Stochastic Programming, M.A.H. Dempster, ed., Academic Press, London.

6. Dula, J.H. (1986) Bounds on the expectation of convex function. Ph.D. Dissertation, The University of Michigan, Ann Arbor, Michigan.

7. Edmundson, H.P. (1956) Bounds on the expectation of a convex function of a random variable. The Rand Corporation Paper 982, Santa Monica, California, USA.

8. Ermolier, Y., Gaivoranoki, A., and Nedeva, C. (1987) Stochastic optimization problems with partially known distribution functions. SIAM J. Control and Optimization, forthcoming.

9. Fishburn, P.C. (1970) Utility Theory for Decision Making, R.E. Krieger, Huntington, New York.

10. Frauendorfer, K. (1986) Solving S.L.P. recourse problems with arbitrary multivariate distributions-the dependent case. Institut fuer Operations Research, Universitate Zuerich, working paper.

11. Gassmann, H. and W. Ziemba, (1986) A tight upper bound for the expectation of a convex function of a multivariate random variable. Math. Prog. Stud. 27, 39-53.

12. Gradshteyn I.S. and I.M. Ryzhik. (1980) Table of Integrals, Series and Products. Corrected and Enlarged Edition. Academic Press, New York.

13. Karlin, S. and Y. Renotl, (1981) Univariate and multivariate total positivity, generalized convexity and related inequalities in Generalized concavity in optimization and economics, Edited by S. Schaible and Ziemba W.T., Academic Press, New York.

14. Madansky, A. (1959) Bounds on the expectation of a convex function of a multivariate random variable. <u>Annals Math. Stat.</u> 30, 743-746.

15. Rockafellar, R.T. (1970) <u>Convex Analysis</u>, Princeton University Press, Princeton, New Jersey, USA.

16. Wets, R.J.B. (1974) Stochastic programs with fixed recourse: the equivalent deterministic program. <u>SIAM Review</u> 16, 309-339.