MATHEMATICS OF OPERATIONS RESEARCH
Vol. 30, No. 3, August 2005, pp. 662–677
ISSN 0364-765X | EISSN 1526-5471 | 05 | 3003 | 0662
DOI 10.1287/moor.1040.0136

Second-Order Lower Bounds on the Expectation of a Convex Function

Steftcho P. Dokov, David P. Morton
Graduate Program in Operations Research, The University of Texas at Austin, Austin, Texas 78712,
{steftcho@msn.com, morton@mail.utexas.edu}

We develop a class of lower bounds on the expectation of a convex function. The bounds utilize the first two moments of the underlying random variable, whose support is contained in a bounded interval or hyperrectangle. Our bounds have applications to stochastic programs whose random parameters are known only through limited-moment information. Computational results are presented for two-stage stochastic linear programs.

Key words: stochastic programming approximations; generalized moment problems
MSC2000 subject classification: Primary: 90C15; secondary: 41A17
OR/MS subject classification: Primary: Programming, stochastic
History: Received February 26, 2001; revised June 7, 2003, and August 9, 2004.

1. Introduction. In this paper, we introduce a class of lower bounds on the expected value of a convex function. The underlying random variable's support is contained in a bounded interval or hyperrectangle, and the bounds use limited-moment information, requiring only the first and second moments.

Jensen's [26] inequality provides a first-order (i.e., uses only the first moment) lower bound on the expectation of a convex function. For random variables with bounded support, the Edmundson-Madansky (EM) bound [17, 33] is a first-order upper bound on the expectation of a convex function. Madansky [34] applied these bounds to stochastic linear programming problems, and subsequently there has been significant work on developing approximations to stochastic programs via such bounding schemes (see, e.g., Birge and Louveaux [2, Chapter 9]).

Generalized-moment problems (GMPs) (Karr [29], Kemperman [30], Kreĭn and Nudel'man [31]) provide a unifying framework for developing bounds that use limited-moment information. In a GMP, the objective is to optimize the expected value of a function over probability measures that are constrained to incorporate information (e.g., certain moments) assumed to be known. When only the first moment is known, Jensen's bound minimizes the objective and the EM bound maximizes the objective. Because solutions to GMPs are probability measures, the resulting bounds may be viewed as "distributional approximations" that are optimal when there is incomplete knowledge regarding the underlying distribution. Distributional approximations replace the true (and perhaps unknown) distribution with an approximating distribution that is usually designed to facilitate computation of the associated bound. Typically, these approximations have finite support. This view of bounds for stochastic programs began with the work of Dupačová [11, 10]. Generalizations of the EM bound have been developed by Dupačová [11], Frauendorfer [18], and Gassmann and Ziemba [21]. For more on using GMPs to find upper bounds for stochastic programs, see Birge and Wets [6] and Kall [27]. As we will show, while the bounds we develop do not actually solve GMPs, they are motivated by moment problems and are distributional approximations.

Even when there is complete knowledge of the underlying distribution, bounds based on limited-moment information are useful. Computing the expected value of a function can be difficult, especially when function evaluations are expensive (as is typically the case in stochastic programming). Sequential-approximation algorithms apply bounds to a partition of the random variable's support, requiring conditional mass and moment calculations. These procedures iteratively improve upper- and lower-bound approximations; see

Birge and Wallace [3], Birge and Wets [5], Frauendorfer [19], and Huang et al. [24]. The bounds that we develop in this paper can also be applied in this fashion.

Therefore, one way of tightening first-order bounds is to apply them in a conditional manner. Alternatively, the bound can be tightened by using further information (e.g., the second moment). Dupačová [10] incorporates a variance constraint in a GMP for finding an upper bound. Birge and Dulá [1], Dulá [8], Dulá and Murthy [9], and Kall [28] also use second-moment information in computing upper bounds.

There is significantly less work on second-order lower bounds. Edirisinghe [12] develops a second-order lower bound that is tighter than the first-order Jensen bound. This second-order bound is employed in a sequential approximation procedure for two-stage stochastic programs in Edirisinghe and You [14] and is extended to multistage stochastic programs in Edirisinghe [13]. (Also see Frauendorfer [20] for bounds for multistage problems.) For random vectors with independent components, the class of lower bounds we develop includes Edirisinghe's result as a special case and, in general, provides a tighter lower bound, but is more expensive to compute.

We note that bounds based on convexity can be applied to convex-concave saddle functions; see Edirisinghe [12], Edirisinghe and Ziemba [15, 16], and Frauendorfer [19]. In addition to bounds based on distributional approximations, there are bounds based on functional approximations. See, for example, Birge and Wallace [4], Birge and Wets [5, 7], Morton and Wood [35], Powell and Frantzeskakis [36], and Wallace [37].

This paper is organized as follows. Section 2 provides background on a generalized-moment problem and related results. A new class of second-order lower bounds for univariate convex functions is developed in §3. Section 4 shows that the best lower bound in this class can be determined by finding the roots of two monotone univariate functions. In §5, we extend the results to convex functions of random vectors. The multivariate bounds are applied to three stochastic programs in §6, and the paper is summarized in §7.

2. A generalized-moment problem. To compute valid lower bounds for all distributions with support contained in the finite interval $[a, b]$ and known first and second moments m_1 and m_2, we consider a GMP

$$\inf_{P\in\mathcal{P}} \int_a^b f(u)\, dP = E^P f(\xi), \tag{1}$$

where $\mathcal{P}$ is a set of probability measures on $([a, b], \mathcal{F})$ such that

$$\int_a^b u\, dP = m_1, \tag{2}$$

$$\int_a^b u^2\, dP = m_2. \tag{3}$$

Here, $\mathcal{F}$ is the Borel field of $[a, b]$, $f\colon [a, b] \to \mathcal{R}$ is a convex function, and ξ is a nondegenerate random variable defined on $([a, b], \mathcal{F}, P)$. Karr [29] studies properties of solutions to GMPs like (1) in a more general form. An application of his results tells us that the set of probability measures, $\mathcal{P}$, which are feasible for the infinite-dimensional problem (1) is convex and compact with respect to the weak* topology. As a result, $\mathcal{P}$ can be expressed as the convex hull of its extreme points. Furthermore, an optimal solution of the GMP is obtained at an extreme point of $\mathcal{P}$, and for (1) these extreme points are probability measures that have *at most* three distinct points of support x_1, x_2, and x_3 in $[a, b]$. Sometimes only two points are used in an optimal solution. When f is continuously differentiable and df/du is strictly convex on $[a, b]$, an application of Kreĭn and Nudel'man [31, Chapter 4, Theorem 1.1] implies that a unique two-point support solves (1). One of these points is a and the other point and weights are determined by (2), (3), and $\int_a^b dP = 1$. Birge and Dulá [1] develop a more general condition that is sufficient to ensure the two-point property when the objective in (1) is to maximize $E^P f(\xi)$. Under their condition, the two points may be interior to $[a, b]$ and a line search is required to find the points and their weights. The bounds we derive do not require f to be differentiable and are applicable when a three-point support may arise.

We can assume three distinct points, $a \le x_1 < x_2 < x_3 \le b$, without loss of generality, because a two-point or one-point solution can be obtained via associated weights p_1, p_2, and p_3. The vectors

$$\begin{pmatrix} 1 \\ x_1 \\ x_1^2 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ x_2 \\ x_2^2 \end{pmatrix}, \quad \text{and} \quad \begin{pmatrix} 1 \\ x_3 \\ x_3^2 \end{pmatrix} \tag{4}$$

are linearly independent (they form a 3×3 Vandermonde matrix). Therefore, (1) can be written in the equivalent form

$$\begin{aligned} \inf\ & p_1 f(x_1) + p_2 f(x_2) + p_3 f(x_3) \\ \text{s.t.}\ & p_1 + p_2 + p_3 = 1, \\ & p_1 x_1 + p_2 x_2 + p_3 x_3 = m_1, \\ & p_1 x_1^2 + p_2 x_2^2 + p_3 x_3^2 = m_2, \\ & a \le x_1 < x_2 < x_3 \le b, \\ & p_1, p_2, p_3 \ge 0. \end{aligned}$$

Solving for the p_is in terms of x_is, $i = 1, 2, 3$, and the given moments m_1 and m_2 leads to the following optimization problem in three variables:

$$\begin{aligned} \inf & \left\{ \frac{m_2 + x_2 x_3 - m_1(x_2 + x_3)}{(x_2 - x_1)(x_3 - x_1)} f(x_1) + \frac{m_1(x_1 + x_3) - x_1 x_3 - m_2}{(x_2 - x_1)(x_3 - x_2)} f(x_2) \right. \\ & \qquad \left. + \frac{m_2 + x_1 x_2 - m_1(x_1 + x_2)}{(x_3 - x_1)(x_3 - x_2)} f(x_3) \right\} \\ \text{s.t.}\ & m_2 + x_2 x_3 - m_1(x_2 + x_3) \ge 0, \\ & m_1(x_1 + x_3) - x_1 x_3 - m_2 \ge 0, \\ & m_2 + x_1 x_2 - m_1(x_1 + x_2) \ge 0, \\ & a \le x_1 < x_2 < x_3 \le b. \end{aligned} \tag{5}$$

Solving (5) exactly would yield a *sharp* lower bound on the expectation of a convex function when only the mean and variance of the underlying distribution are known. Saying that a bound is sharp means that it is achieved for some distribution with first and second moments m_1 and m_2. We are unable to solve (5) for an arbitrary convex f, and instead will bound its optimal objective value from below to derive a class of closed-form lower bounds. In what follows, we use X to denote the closure of the feasible region of (5), i.e., with inclusive inequalities on the ordering constraints.

3. A class of second-order lower bounds. The next theorem states the main result of the paper for univariate functions.

THEOREM 3.1. *Let $f\colon [a, b] \to \mathcal{R}$ be a convex function, and let ξ be a nondegenerate random variable with support contained in $[a, b]$ and first and second moments m_1 and m_2, respectively. Let $\sigma^2 = m_2 - m_1^2$, $A_v = m_1 - \sigma^2/(v - m_1)$, $B_v = m_1 + \sigma^2/(m_1 - v)$, $A = A_b$, and $B = B_a$. Then,*

$$E^P f(\xi) \ge L(y, z) = \min\{L_1(y), L_1'(y), L_2(z), L_2'(z)\} \quad \forall\, y \in [B, b], z \in [a, A], \tag{6}$$

where

$$\begin{aligned}
L_1(y) &= \frac{B-m_1}{B-A_y} f(A_y) + \frac{m_1 - A_y}{B-A_y} f(B),\\
L_1'(y) &= \frac{B-m_1}{B-A} f(A) + \frac{m_1-A}{B-A}\left(\frac{y-B}{y-m_1} f(m_1) + \frac{B-m_1}{y-m_1} f(y)\right),\\
L_2(z) &= \frac{B_z - m_1}{B_z - A} f(A) + \frac{m_1 - A}{B_z - A} f(B_z),\\
L_2'(z) &= \frac{B-m_1}{B-A}\left(\frac{m_1-A}{m_1-z} f(z) + \frac{A-z}{m_1-z} f(m_1)\right) + \frac{m_1-A}{B-A} f(B).
\end{aligned}$$

Proof. Let $x = (x_1, x_2, x_3)$ and let $F(x)$ denote the objective function in (5). Then, by algebraic manipulation,

$$\begin{aligned}
F(x) = {}& \frac{x_3 - m_1}{x_3 - x_1} f(x_1) + \frac{m_1 - x_1}{x_3 - x_1} f(x_3)\\
& - \frac{m_1(x_1+x_3) - x_1x_3 - m_2}{x_3 - x_1}\left(\frac{f(x_3)-f(x_2)}{x_3-x_2} - \frac{f(x_2)-f(x_1)}{x_2-x_1}\right).
\end{aligned}$$

The feasible region can be re-expressed as

$$X = \{(x_1, x_2, x_3)\colon a \le x_1 \le \bar{A} \le x_2 \le \bar{B} \le x_3 \le b\},$$

where $\bar{A} \equiv A_{x_3}$ and $\bar{B} \equiv B_{x_1}$. To see this, we need only observe

$$\begin{aligned}
m_2 + x_2x_3 - m_1(x_2+x_3) \ge 0 &\Leftrightarrow x_2 \ge \bar{A},\\
m_1(x_1+x_3) - x_1x_3 - m_2 \ge 0 &\Leftrightarrow x_1 \le \bar{A},\\
m_1(x_1+x_3) - x_1x_3 - m_2 \ge 0 &\Leftrightarrow x_3 \ge \bar{B},\\
m_2 + x_1x_2 - m_1(x_1+x_2) \ge 0 &\Leftrightarrow x_2 \le \bar{B}.
\end{aligned}$$

To derive the lower bound, we effectively branch on $x_2 \le m_1$ and $x_2 \ge m_1$. Specifically, in the first case we find a lower bound on a relaxation of $X \cap \{x\colon x_2 \le m_1\}$. In the second case, we find a lower bound on a relaxation of $X \cap \{x\colon x_2 \ge m_1\}$. Finally, we conclude that the smaller of these two local lower bounds is a global lower bound on $\min_X F(x)$.

Case 1. $x_2 \le m_1$. Let $X_1 = \{(x_1, x_2, x_3)\colon a \le x_1 \le \bar{A} \le x_2 \le m_1 \le B \le x_3 \le b\}$ and note $X_1 \supseteq X \cap \{x\colon x_2 \le m_1\}$ because $\bar{B} = B_{x_1}$ has range $[B, b]$. The inequalities

$$\frac{f(x_3)-f(x_2)}{x_3-x_2} \le \frac{f(x_3)-f(m_1)}{x_3-m_1} \quad \text{and} \quad \frac{f(x_2)-f(x_1)}{x_2-x_1} \ge \frac{f(\bar{A})-f(x_1)}{\bar{A}-x_1} \tag{7}$$

follow from the definition of convexity of f by expressing m_1 as a convex combination of x_2 and x_3 and $\bar{A}$ as a convex combination of x_1 and x_2, respectively. Therefore,

$$\begin{aligned}
\min_{X_1} F(x) \ge {}& \min_{X_1}\left\{\frac{x_3-m_1}{x_3-x_1} f(x_1) + \frac{m_1-x_1}{x_3-x_1} f(x_3)\right.\\
& \left. - \frac{m_1(x_1+x_3) - x_1x_3 - m_2}{x_3-x_1}\left(\frac{f(x_3)-f(m_1)}{x_3-m_1} - \frac{f(\bar{A})-f(x_1)}{\bar{A}-x_1}\right)\right\}\\
= {}& \min_{X_1}\left\{\frac{x_3-m_1}{x_3-x_1} f(\bar{A}) + \frac{\bar{A}-x_1}{x_3-x_1} f(m_1) + \frac{m_1-\bar{A}}{x_3-x_1} f(x_3)\right\}.
\end{aligned} \tag{8}$$

Differentiating the objective function in (8) with respect to x_1 yields

$$\frac{x_3-m_1}{(x_3-x_1)^2} f(\bar{A}) - \frac{x_3-\bar{A}}{(x_3-x_1)^2} f(m_1) + \frac{m_1-\bar{A}}{(x_3-x_1)^2} f(x_3),$$

which is nonnegative. This can be seen by expressing m_1 as a convex combination of $\bar{A}$ and x_3 and using the fact that f is convex. As a result, (8) is bounded below by

$$\min_{B\le x_3\le b} \underbrace{\left\{ \frac{x_3-m_1}{x_3-a} f(\bar{A}) + \frac{\bar{A}-a}{x_3-a} f(m_1) + \frac{m_1-\bar{A}}{x_3-a} f(x_3) \right\}}_{g(x_3)}. \tag{9}$$

Fixing $y\in[B,b]$, we have

$$\min_{B\le x_3\le b} g(x_3) = \min\left\{ \min_{B\le x_3\le y} g(x_3),\ \min_{y\le x_3\le b} g(x_3)\right\}.$$

Under domain $[B, y]$, $\bar{A}=A_{x_3}$ has range $[a, A_y]$. Therefore, for $x_3\in[B,y]$ we have, by convexity,

$$f(\bar{A}) \ge \frac{m_1-\bar{A}}{m_1-A_y} f(A_y) - \frac{A_y-\bar{A}}{m_1-A_y} f(m_1) = \frac{y-m_1}{x_3-m_1} f(A_y) - \frac{y-x_3}{x_3-m_1} f(m_1)$$

and

$$f(x_3) \ge \frac{x_3-m_1}{B-m_1} f(B) - \frac{x_3-B}{B-m_1} f(m_1) = \frac{m_1-a}{\sigma^2}[(x_3-m_1)f(B)-(x_3-B)f(m_1)]. \tag{10}$$

Applying these two inequalities to the first and third terms of the objective in (9) yields

$$g(x_3) \ge \frac{y-m_1}{x_3-a} f(A_y) + \frac{m_1-a}{x_3-a} f(B) - \frac{y-x_3}{x_3-a} f(m_1) \quad \forall\, x_3\in[B,y]. \tag{11}$$

The function on the right-hand side of (11) has a nonpositive derivative with respect to x_3 (this, again, follows from convexity of f) so

$$\min_{B\le x_3\le y} g(x_3) \ge \frac{y-m_1}{y-a} f(A_y) + \frac{m_1-a}{y-a} f(B) = L_1(y).$$

Applying a similar analysis, it can be shown that

$$\min_{y\le x_3\le b} g(x_3) \ge \frac{b-m_1}{b-a} f(A) + \frac{m_1-a}{b-a}\left(\frac{y-B}{y-m_1} f(m_1) + \frac{B-m_1}{y-m_1} f(y)\right) = L_1'(y).$$

Summarizing our results for Case 1, we have shown that

$$E^P f(\xi) \ge \min\{L_1(y), L_1'(y)\} \quad \forall\, y\in[B,b] \text{ when } x_2\le m_1.$$

Case 2. $x_2\ge m_1$. Repeating, by analogy, the steps of Case 1 using $z\in[a,A]$ and B_z, it can be shown that

$$E^P f(\xi) \ge \min\{L_2(z), L_2'(z)\} \quad \forall\, z\in[a,A] \text{ when } x_2\ge m_1.$$

Combining the results from Cases 1 and 2 gives the desired result. □

Theorem 3.1 gives us a family of lower bounds, $L(y,z)$, on $E^P f(\xi)$ that is valid for any distribution with support in $[a,b]$ and known mean and variance. When computing a bound, we have the freedom to choose any $y\in[B,b]$ and any $z\in[a,A]$. In choosing specific values for these parameters, we should consider the trade-off between the quality of the bound and the effort required to compute it. In §4, we show how to compute $\max_{y\in[B,b],\, z\in[a,A]} L(y,z)$, the strongest bound from our class. In some cases, evaluating f is expensive, and arbitrary choices of y and z require seven function evaluations of f to compute $L(y,z)$. However, specific choices of y and z provide a bound requiring fewer function evaluations. Two such results are provided in Corollaries 3.1 and 3.2. Edirisinghe [12] develops a second-order lower bound that may be applied to convex functions of random vectors whose support is contained in a simplex or a hyperrectangle, and the *EB* bound of Corollary 3.1 is the special case of his bound in one dimension.

COROLLARY 3.1. *Under the hypotheses of Theorem 3.1,*

$$EB = \frac{B - m_1}{B - A} f(A) + \frac{m_1 - A}{B - A} f(B) \le E^P f(\xi). \tag{12}$$

Moreover,

$$L(y, z) \ge EB \quad \forall\, y \in [B, b], z \in [a, A]. \tag{13}$$

PROOF. Let $z = A$ and $y = B$. It is straightforward to show $L_1(B) \ge L_1'(B)$ and $L_2(A) \ge L_2'(A)$ by using the fact that f is convex. Inequality (12) then follows from $L_1'(B) = L_2'(A) = EB$.

To prove the second result, we first establish monotonicity properties of the four functions used to define $L(y, z)$. Beginning with $L_1(y)$, let $y' < y''$, where $y', y'' \in [B, b]$ and note that

$$L_1(y) = \frac{y - m_1}{y - a} f(A_y) + \frac{m_1 - a}{y - a} f(B).$$

$L_1(y') \ge L_1(y'')$ is equivalent to

$$\frac{y'' - m_1}{y'' - a} f(A_{y''}) + \frac{m_1 - a}{y'' - a} f(B) \le \frac{y' - m_1}{y' - a} f(A_{y'}) + \frac{m_1 - a}{y' - a} f(B)$$

$$\iff \quad f(A_{y''}) \le \frac{(y'' - a)(y' - m_1)}{(y' - a)(y'' - m_1)} f(A_{y'}) + \frac{(m_1 - a)(y'' - y')}{(y' - a)(y'' - m_1)} f(B)$$

$$\iff \quad f(A_{y''}) \le \frac{B - A_{y''}}{B - A_{y'}} f(A_{y'}) + \frac{A_{y''} - A_{y'}}{B - A_{y'}} f(B),$$

which holds because f is convex. Thus, L_1 decreases from $L_1(B)$ to $L_1(b) = EB$. One can similarly show: L_1' increases from $L_1'(B) = EB$ to $L_1'(b)$; L_2 increases from $L_2(a) = EB$ to $L_2(A)$; and, L_2' decreases from $L_2'(a)$ to $L_2'(A) = EB$. □

Calculating a second-order lower bound via (12) has the advantage of minimizing computational effort, reducing the number of function evaluations from seven to two, but as Inequality (13) shows, the price to pay for this savings is that the bound is the weakest from our class. We will return to the monotonicity results established in the proof of Corollary 3.1 for L_1, L_1', L_2, and L_2' in §4.

Computing $L(y, z)$ for general choices of $y \in [B, b]$ and $z \in [a, A]$ requires seven function evaluations of f at z, A_y, A, m_1, B, y, and B_z. Two degrees of freedom are provided via y and z. Selecting $z \in [a, A]$ and setting $y = B_z$ yields a subclass of bounds with one degree of freedom that requires only five function evaluations (five because $y = B_z$ and $A_y = A_{B_z} = z$). This result is summarized in the next corollary.

COROLLARY 3.2. *Under the hypotheses of Theorem 3.1,*

$$\min\{L_1'(B_z), L_2'(z)\} \le E^P f(\xi) \quad \forall\, z \in [a, A].$$

PROOF. Let $z \in [a, A]$, $y = B_z$, and note that $B_z \in [B, b]$. The desired result then follows immediately from (6), $L_1(B_z) \ge L_2'(z)$, and $L_2(z) \ge L_1'(B_z)$. We sketch the proof for the former inequality; the latter inequality can be verified in analogous fashion. Using the fact that $A_{B_z} = z$ and the definition of L_1, we have

$$L_1(B_z) = \frac{B - m_1}{B - z} f(z) + \frac{m_1 - z}{B - z} f(B).$$

Because f is convex,

$$f(m_1) \le \frac{m_1 - z}{B - z} f(B) + \frac{B - m_1}{B - z} f(z). \tag{14}$$

The inequality $L_1(B_z) \ge L_2'(z)$ follows by replacing $f(m_1)$ in the definition of $L_2'(z)$ with the right-hand side of (14). □

The expressions in Theorem 3.1 of the four functions that determine $L(y, z)$ afford a geometric interpretation illustrated in Figure 1 (for $L_1(y)$ and $L_1'(y)$) and Figure 2 (for $L_2(z)$

and $L_2'(z)$). For example, for any $y \in [B, b]$, $L_1(y)$ is a convex combination of $f(A_y)$ and $f(B)$, and the weights in this convex combination may be viewed as a two-point probability distribution that preserves the first moment, m_1, of the underlying random variable ξ. In a similar way, $L_1'(y)$ is a convex combination of $f(A)$ and an expression that itself is a convex combination of $f(m_1)$ and $f(y)$. The associated three-point distribution again has first moment m_1.

An analogous geometric interpretation holds for $L_2(z)$ and $L_2'(z)$ and is shown in Figure 2. The one-dimensional version of Edirisinghe's [12] lower bound (EB) is also shown in Figures 1 and 2. The monotonicity properties of $L_1(y)$, $L_1'(y)$, $L_2(z)$, and $L_2'(z)$ developed

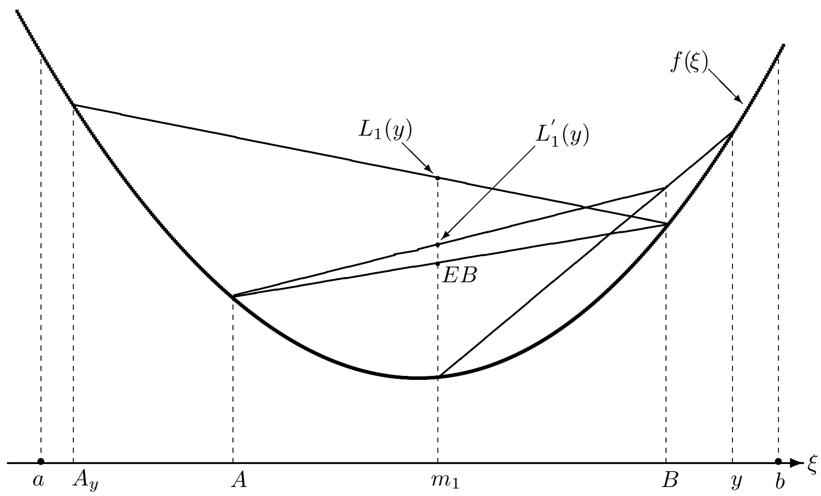

Figure 1. This figure illustrates $L_1(y)$ and $L_1'(y)$, two of the four expressions that define the class of second-order lower bounds on $E^P f(\xi)$ from Theorem 3.1.

Note. The parameter y is one of two degrees of freedom associated with the bound $L(y, z)$. $L_1(y)$ is a convex combination of $f(A_y)$ and $f(B)$ while $L_1'(y)$ is a convex combination of $f(A)$ and a term which is, in turn, a convex combination of $f(m_1)$ and $f(y)$. Edirisinghe's [12] second-order lower bound (see EB in Corollary 3.1) is also shown.

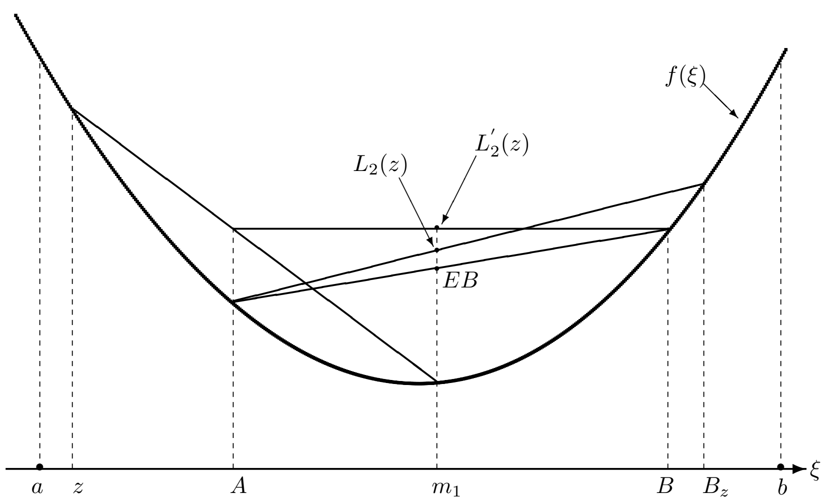

Figure 2. The two expressions associated with the second degree of freedom (i.e., parameter z) that define the lower bound in Theorem 3.1 are shown here.

Note. $L_2(z)$ is a convex combination of $f(A)$ and $f(B_z)$, while $L_2'(z)$ is a convex combination of $f(B)$ and a term which is, in turn, a convex combination of $f(z)$ and $f(m_1)$.

in the proof of Corollary 3.1, and the fact that EB is a limiting value for each of these functions, are easily seen by shifting y and z in the figures and tracking the corresponding changes in A_y and B_z.

Figure 3 shows the lower bound $\min\{L'_1(B_z), L'_2(z)\}$ of Corollary 3.2, which requires five function evaluations. The functions $L'_1(B_z)$ and $L'_2(z)$ are convex combinations of function values of f at A, m_1, B_z and z, m_1, B, respectively. The well-known first-order lower and upper bounds of Jensen (JB) and Edmundson-Madansky (EM) are also indicated in Figure 3.

The bound $L(y, z)$ may be viewed as arising from a distributional approximation. For fixed values of y and z there are four possible approximating distributions, and the expected value of f under these respective distributions is $L_1(y)$, $L'_1(y)$, $L_2(z)$, and $L'_2(z)$. Of these four distributions, we must choose the one that minimizes the expectation of f; see (6). The first moment of all four distributions is m_1, the mean of the underlying random variable ξ. However, the second moments of the approximating distributions are smaller than m_2. So, while yielding a valid lower bound, the approximating distributions do not solve the GMP (1) posed in §2, and hence we cannot guarantee that the bound will be sharp.

While $L(y, z)$ can be strengthened by optimizing over y and z (see §4), one of its attractive features is that it is not necessary to use an optimization algorithm to compute a valid bound. This is in contrast to some approaches that require a GMP be solved to optimality for the bound to be valid.

We conclude this section by briefly analyzing the bound $L(y, z)$ with respect to its defining parameters. Computing the bound requires knowing the endpoints of the interval of support, $[a, b]$, and the mean and variance, m_1 and σ^2, of the underlying random variable. While this is clearly less demanding than assuming that the distribution of ξ is known, even these parameters may not be known precisely. So, here we give qualitative insight for the sensitivity of the bound with respect to $[a, b]$, m_1, and σ. There are three important properties: First, $L(y, z)$ is nondecreasing in σ provided m_1, a, and b are fixed. Second, $L(y, z)$ decreases as a decreases or b increases provided m_1 and σ are fixed as well as b or a, respectively. Third, the $EB = L(b, a)$ bound from Corollary 3.1 is a convex function of m_1 provided a, b, and m_2 are fixed. The first and third properties provide a way to find a valid second-order lower bound when, instead of knowing m_1 and σ precisely, we have ranges: $m' \le m_1 \le m''$ and $\sigma' \le \sigma \le \sigma''$. Solving the convex minimization problem $\min_{m' \le m_1 \le m''} EB$ yields a valid lower bound for all $m_1 \in [m', m'']$, while $L(y, z)|_{\sigma=\sigma'}$ is a valid lower bound

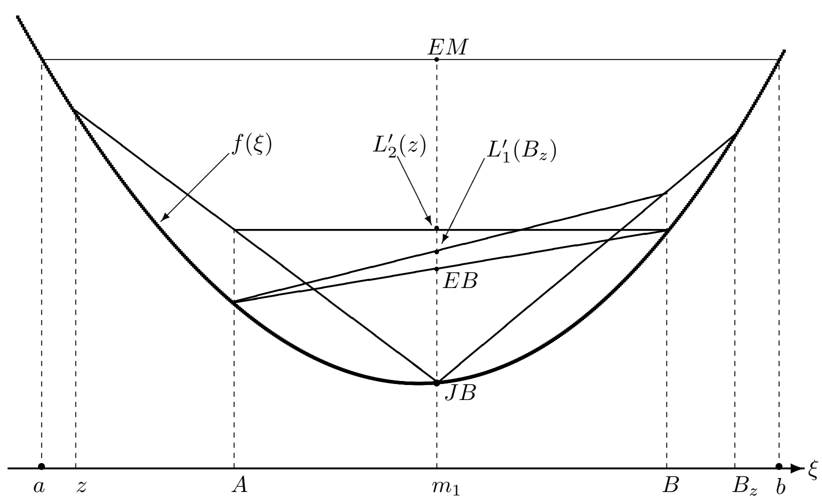

FIGURE 3. This figure provides a geometric interpretation of the two expressions, $L'_1(B_z)$ and $L'_2(z)$, that define Corollary 3.2's class of second-order lower bounds.

Note. The classical first-order bounds of Jensen (JB) and Edmundson-Madansky (EM), as well as Edirisinghe's [12] second-order lower bound (EB), are also illustrated in the figure.

for all $\sigma \in [\sigma', \sigma'']$. (Unfortunately, the convexity property with respect to m_1 does not extend to $L(y,z)$.) Finally, the second property shows that the bound becomes stronger as the interval of support shrinks and weaker as it grows. In a limiting argument, as the endpoints of the interval containing the support satisfy $a \to -\infty$ and $b \to \infty$, the EB and $L(y,z)$ bounds collapse to the Jensen bound.

4. Optimizing the bounds. In the previous section, we found a class of second-order lower bounds on the expectation of a convex function, $L(y,z) \le E^P f(\xi)$, where $L(y,z) = \min\{L_1(y), L_1'(y), L_2(z), L_2'(z)\}$. Because $y \in [B,b]$ and $z \in [a,A]$ are at our disposal, a natural question is: What choice of these parameters gives the strongest lower bound? The answer is given by y^* and z^* that solve

$$L^* = \max_{y\in[B,b],\, z\in[a,A]} L(y,z) = \min\left\{ \max_{y\in[B,b]} \min\{L_1(y), L_1'(y)\},\ \max_{z\in[a,A]} \min\{L_2(z), L_2'(z)\} \right\}. \quad (15)$$

Consider the two univariate maximization problems defined on the right-hand side of (15), and recall from the proof of Corollary 3.1 the following monotonicity results of the four associated functions:

- $L_1(y)$ decreases on $[B,b]$ with $L_1(b) = EB$,
- $L_1'(y)$ increases on $[B,b]$ with $L_1'(B) = EB$,
- $L_2(z)$ increases on $[a,A]$ with $L_2(a) = EB$, and
- $L_2'(z)$ decreases on $[a,A]$ with $L_2'(A) = EB$.

As a result, there exists $y^* \in [B,b]$ such that

$$L_1(y^*) = L_1'(y^*) = \max_{y\in[B,\,b]} \min\{L_1(y), L_1'(y)\},$$

and there exists $z^* \in [a,A]$ such that

$$L_2(z^*) = L_2'(z^*) = \max_{z\in[a,\,A]} \min\{L_2(z), L_2'(z)\}.$$

The best lower bound from the class of bounds introduced in Theorem 3.1 is

$$L^* = \min\{L_1(y^*), L_2(z^*)\}.$$

Values for y^* and z^* (and hence L^*) may be found by running two bisection searches because

$$L_1(y) \ge L_1'(y) \quad \forall\, y \in [B, y^*] \quad \text{and} \quad L_1(y) \le L_1'(y) \quad \forall\, y \in [y^*, b],$$

and

$$L_2(z) \le L_2'(z) \quad \forall\, z \in [a, z^*] \quad \text{and} \quad L_2(z) \ge L_2'(z) \quad \forall\, y \in [z^*, A].$$

The best lower bound from the subclass of bounds defined in Corollary 3.2 is given by $\max_{z\in[a,A]} \min\{L_1'(B_z), L_2'(z)\}$. Now, B_z is an increasing function of z, so as z increases from a to A, L_1' increases from EB, and L_2' decreases to EB. Thus, a single bisection search can be used to compute $z^{**} \in [a,A]$ with

$$L^{**} = L_1'(B_{z^{**}}) = L_2'(z^{**}) = \max_{z\in[a,\,A]} \min\{L_1'(B_z), L_2'(z)\}.$$

EXAMPLE 4.1. Let ξ be a random variable with support $[0,6]$, mean $m_1 = 4$, and variance $\sigma^2 = 4$. Thus, $A = 2$, $B = 5$, $A_y = 4(y-5)/(y-4)$, $B_z = 4(5-z)/(4-z)$, and

$$L_1(y) = \frac{y-4}{y} f(A_y) + \frac{4}{y} f(5),$$
$$L_1'(y) = \frac{1}{3} f(2) + \frac{2}{3}\left(\frac{y-5}{y-4} f(4) + \frac{1}{y-4} f(y)\right),$$
$$L_2(z) = \frac{2}{6-z} f(2) + \frac{4-z}{6-z} f(B_z),$$
$$L_2'(z) = \frac{1}{3}\left(\frac{2}{4-z} f(z) + \frac{2-z}{4-z} f(4)\right) + \frac{2}{3} f(5),$$

where $y \in [5,6]$ and $z \in [0,2]$. With $f(\xi) = \xi^2$, the unique value of y^* such that $L_1(y^*) = L_1'(y^*)$ is the solution of

$$4\left(\frac{4y^* - 15}{y^* - 4}\right) = L_1(y^*) = L_1'(y^*) = \frac{2}{3}(y^* + 22);$$

i.e., $y^* = 3 + \sqrt{7}$. Similarly, the unique value of z^* such that $L_2(z^*) = L_2'(z^*)$ solves

$$8\left(\frac{2z^* - 9}{z^* - 4}\right) = L_2(z^*) = L_2'(z^*) = \frac{2}{3}(29 - z^*),$$

which yields $z^* = 1$. Therefore, the best lower bound in Theorem 3.1's class of bounds is

$$L^* = L(3+\sqrt{7}, 1) = \frac{2}{3}(25 + \sqrt{7}) \approx 18.43.$$

This bound is illustrated in Figure 4.

Figure 5 shows the class of lower bounds developed in Corollary 3.2, and the best bound from this class, $L^{**} = \max_{z\in[0,2]} \min\{L_1'(B_z), L_2'(z)\} = 17 + \sqrt{17}/3 \approx 18.37$, which is achieved at $z^{**} = (7 - \sqrt{17})/2$.

For $f(\xi) = \xi^n$, $n = 2, 3, 4, 5$, Table 1 provides the bounds of Jensen [26], Edirisinghe [12], L^*, and the optimal value of the GMP defined in (1). As the table indicates, the second-order bounds, EB and L^*, are significantly stronger than the first-order Jensen bound as f

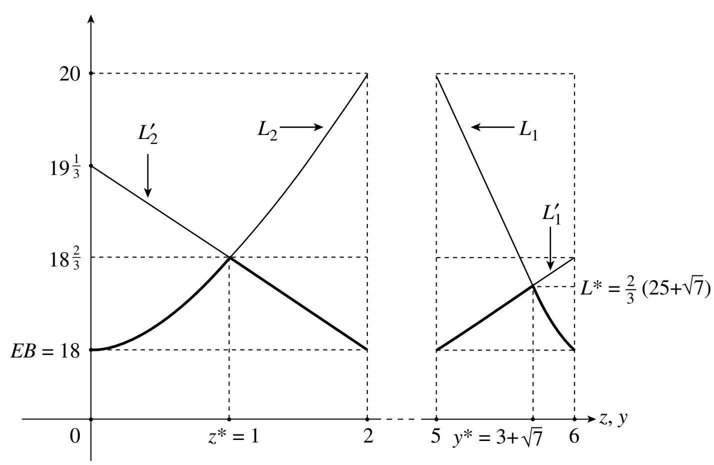

FIGURE 4. For Example 4.1 considered in §4, this figure shows L^*, the strongest lower bound from the class of bounds introduced in Theorem 3.1. The optimal values y^* and z^* (see (15)) are also indicated. The shadow lines are the two functions $\min\{L_1(y), L_1'(y)\}$ and $\min\{L_2(z), L_2'(z)\}$.

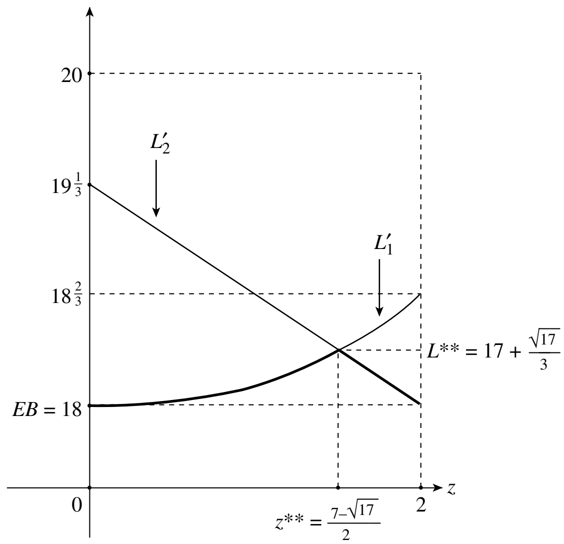

FIGURE 5. For Example 4.1, the shadow line is the class of lower bounds $\min\{L_1'(B_z), L_2'(z)\}$ from Corollary 3.2. The best lower bound in this class, $L^{**} = L_1'(B_{z^{**}}) = L_2'(z^{**})$, is also shown.

becomes "more convex." For $f(\xi) = \xi^n$, $n = 3, 4, 5$, the derivative of f is strictly convex, and so by Kreĭn and Nudel'man [31, Chapter 4, Theorem 1.1], we know that the unique optimal solution of the GMP has two-point support $\{a, B\}$ with weights p and $1-p$ that can be found by solving $pa + (1-p)B = m_1$. For $f(\xi) = \xi^2$, the objective function of the GMP has value $m_2 = 20$ by the second-moment constraint (3).

5. Multivariate extensions. In this section, we extend the bound of Theorem 3.1 to a convex function of a random vector with independent components. We first adapt our notation to the multivariate setting. Let $\xi = (\xi_1, \ldots, \xi_d)^T$ be a random vector with independent components. The support of ξ is contained in $\prod_{i=1}^d [a_i, b_i]$, the joint distribution of ξ is denoted P, $E\xi_i = \mu_i$, and $\operatorname{var} \xi_i = \sigma_i^2$. Define $A_{iv} = \mu_i - \sigma_i^2/(v - \mu_i)$, $B_{iv} = \mu_i + \sigma_i^2/(\mu_i - v)$, $A_i = A_{ib_i}$, and $B_i = B_{ia_i}$. Let $f\colon \prod_{i=1}^d [a_i, b_i] \to \mathscr{R}$ be a convex function. The lower bound on $E^P f(\xi)$ we define below is denoted $L(y, z)$ and is parameterized by $y = (y_1, \ldots, y_d)^T$ and $z = (z_1, \ldots, z_d)^T$.

For the univariate case described in §3, the four functions defined in Theorem 3.1 are simply the expectation of f under certain two-point (L_1 and L_2) and three-point (L_1' and L_2') distributions. With this view, we define distributions Q_{i1}, $Q_{i1'}$, Q_{i2}, and $Q_{i2'}$ for $i = 1, \ldots, d$

TABLE 1. The lower bounds of Jensen (JB), Edirisinghe [12] (EB), L^*, and the solution of the generalized-moment problem (1) (GMP). The rows in the table give the bounds on $E^P f(\xi)$ with $f(\xi) = \xi^n$, $n = 2, 3, 4, 5$, for the class of random variables in Example 4.1, i.e., with mean $m_1 = 4$, second moment $m_2 = 20$, and support $[a, b] = [0, 6]$. Each of the bounds, JB, EB, and L^*, are also shown as a percentage of GMP. For ξ^2 and ξ^3, the value of y^* is given (because $L_1(y^*) < L_2(z^*)$) and for ξ^4 and ξ^5 the value of z^* is given (because $L_2(z^*) < L_1(y^*)$).

	JB		*EB*		L^*			
$f(\xi)$	Value	% of GMP	Value	% of GMP	Value	% of GMP	y^* or z^*	GMP value
ξ^2	16	80	18	90	18.4	92	5.6458	20
ξ^3	64	64	86	86	91.1	91	5.5308	100
ξ^4	256	51	422	84	452.9	91	0.5274	500
ξ^5	1,024	41	2,094	84	2,237.0	89	0.36285	2,500

with the following probability mass functions:

$$\begin{aligned}
&Q_{i1}\colon \Pr(A_{iy_i})=\frac{B_i-\mu_i}{B_i-A_{iy_i}}, \quad \Pr(B_i)=\frac{\mu_i-A_{iy_i}}{B_i-A_{iy_i}},\\
&Q_{i1'}\colon \Pr(A_i)=\frac{B_i-\mu_i}{B_i-A_i}, \quad \Pr(\mu_i)=\frac{(\mu_i-A_i)(y_i-B_i)}{(B_i-A_i)(y_i-\mu_i)}, \quad \Pr(y_i)=\frac{(\mu_i-A_i)(B_i-\mu_i)}{(B_i-A_i)(y_i-\mu_i)},\\
&Q_{i2}\colon \Pr(A_i)=\frac{B_{iz_i}-\mu_i}{B_{iz_i}-A_i}, \quad \Pr(B_{iz_i})=\frac{\mu_i-A_i}{B_{iz_i}-A_i},\\
&Q_{i2'}\colon \Pr(B_i)=\frac{\mu_i-A_i}{B_i-A_i}, \quad \Pr(z_i)=\frac{(B_i-\mu_i)(\mu_i-A_i)}{(B_i-A_i)(\mu_i-z_i)}, \quad \Pr(\mu_i)=\frac{(B_i-\mu_i)(A_i-z_i)}{(B_i-A_i)(\mu_i-z_i)}.
\end{aligned} \tag{16}$$

Note that these distributions are parameterized by y and z; i.e., $Q_{i1}=Q_{i1}(y_i)$, $Q_{i1'}=Q_{i1'}(y_i)$, $Q_{i2}=Q_{i2}(z_i)$, and $Q_{i2'}=Q_{i2'}(z_i)$, $i=1,\dots,d$.

THEOREM 5.1. *Let $f\colon \prod_{i=1}^d[a_i,b_i]\to\mathscr{R}$ be a convex function. Let ξ be a random vector with independent components, support contained in $\prod_{i=1}^d[a_i,b_i]$, $E\xi_i=\mu_i$, and $\operatorname{var}\xi_i=\sigma_i^2>0$, $i=1\dots,d$. Then,*

$$E^P f(\xi)\ge L(y,z)=\min\{E^{Q_{1j_1}}E^{Q_{2j_2}}\cdots E^{Q_{dj_d}}f(\xi_1,\dots,\xi_d)\colon j_1,\dots,j_d=1,1',2,2'\}, \tag{17}$$

where the distributions $Q_{ij_i}=Q_{ij_i}(y_i)$, $j_i=1,1'$, and $Q_{ij_i}=Q_{ij_i}(z_i)$, $j_i=2,2'$, $i=1,\dots,d$, are defined in (16), $y\in\prod_{i=1}^d[B_i,b_i]$ and $z\in\prod_{i=1}^d[a_i,A_i]$.

PROOF. We give an outline of the proof that uses the same ideas as the proof of Theorem 3.1. Let the marginal distributions of ξ be denoted P_i, $i=1,\dots,d$. Then,

$$E^P f(\xi)=\int_{a_1}^{b_1}\cdots\int_{a_d}^{b_d} f(u_1,\dots,u_d)\,dP_d\cdots dP_1.$$

For fixed values of $u_1,\dots,u_{d-1}$, $f(u_1,\dots,u_{d-1},\cdot)$ is a convex univariate function. So, following the proof of Theorem 3.1, with obvious generalizations of the notation to the dth component under consideration, we have

$$E^P f(\xi)\ge\begin{cases}E^{P_1}\cdots E^{P_{d-1}}E^{Q_{d1}(y_d)}f(\xi_1,\dots,\xi_{d-1},\xi_d) & \text{for } x_{d2}\le\mu_d,\ x_{d3}\in[B_d,y_d],\\ E^{P_1}\cdots E^{P_{d-1}}E^{Q_{d1'}(y_d)}f(\xi_1,\dots,\xi_{d-1},\xi_d) & \text{for } x_{d2}\le\mu_d,\ x_{d3}\in[y_d,b_d],\\ E^{P_1}\cdots E^{P_{d-1}}E^{Q_{d2}(z_d)}f(\xi_1,\dots,\xi_{d-1},\xi_d) & \text{for } x_{d2}\ge\mu_d,\ x_{d1}\in[z_d,A_d],\\ E^{P_1}\cdots E^{P_{d-1}}E^{Q_{d2'}(z_d)}f(\xi_1,\dots,\xi_{d-1},\xi_d) & \text{for } x_{d2}\ge\mu_d,\ x_{d1}\in[a_d,z_d],\end{cases} \tag{18}$$

where Q_{d1}, $Q_{d1'}$, Q_{d2}, and $Q_{d2'}$ are defined in (16). Next, we consider the $d-1$st component, and for each of the four cases in (18) we have four subcases corresponding to: (i) $x_{d-1,2}\le\mu_{d-1}$, $x_{d-1,3}\in[B_{d-1},y_{d-1}]$; (ii) $x_{d-1,2}\le\mu_{d-1}$, $x_{d-1,3}\in[y_{d-1},b_{d-1}]$; (iii) $x_{d-1,2}\ge\mu_{d-1}$, $x_{d-1,1}\in[z_{d-1},A_{d-1}]$; and (iv) $x_{d-1,2}\ge\mu_{d-1}$, $x_{d-1,1}\in[a_{d-1},z_{d-1}]$. These four respective subcases lead to expectations with respect to $Q_{d-1,1}(y_{d-1})$, $Q_{d-1,1'}(y_{d-1})$, $Q_{d-1,2}(z_{d-1})$, and $Q_{d-1,2'}(z_{d-1})$. We continue in this fashion until we have applied the result to the first component of ξ, resulting in a total of 4^d cases. To achieve a valid lower bound, we take the minimum over all of these cases, as specified in (17). □

Computing $L(y,z)$ for general values of y and z requires 7^d function evaluations of f. These are performed at each point in $\prod_{i=1}^d\{z_i,A_{iy_i},A_i,\mu_i,B_i,y_i,B_{iz_i}\}$. Thus, the effort required to compute $L(y,z)$ grows exponentially in the dimension d. This is consistent with applications of the bounds of Edmundson-Madansky and Edirisinghe [12] to random vectors with rectangular support (although those bounds require only 2^d function evaluations).

Obvious analogs of Corollaries 3.1 and 3.2 can be developed for the multivariate case, yielding bounds that require fewer function evaluations, namely 2^d and 5^d, respectively. The former bound (requiring 2^d function evaluations) is the bound of Edirisinghe [12, Theorem 5]. zAnalogous to Corollary 3.1, the multivariate version of $L(y,z)$ is at least as strong as Edirisinghe's bound on a rectangular domain.

Clearly, the dimension of ξ must be modest for $L(y,z)$ to be computable. In addition, Theorem 5.1 requires the components of ξ to be independent. One commonly used probabilistic modeling technique ameliorates both of these limitations. Often a random vector of dimension n, say η, with dependent components is approximated via a deterministic linear transformation, $\eta = H\xi$, of another random vector, ξ. Here, $H \in \mathcal{R}^{n\times d}$, and the random vector ξ typically consists of a small number d (relative to n) of independent factors that "explain" the randomness in η. Therefore, in our setting, if we wish to bound $Eh(\eta)$ where h: $\mathcal{R}^n \to \mathcal{R}$ is convex, we apply the above methodology to $f(\xi) = h(H\xi)$.

Even though the functions that form $L(y, z)$ are monotonic in each component of y and z and $L(y, z)$ is unimodal, solving $\max\{L(y, z)\colon\ y \in \prod_{i=1}^{d}[B_i, b_i], z \in \prod_{i=1}^{d}[a_i, A_i]\}$ may be hard. The difficulty is that $L(y, z)$ is nondifferentiable and is not concave (or even quasi-concave). Nondifferentiability arises even when f is smooth because of the min operator in (17).

6. Computational results. This section empirically analyzes our lower bounds of §5 on two-stage stochastic linear programs with recourse that take the form

$$w^* = \min_{x\in X} cx + Ef(x, \xi), \tag{19}$$

where X is a polyhedral set, x is the first-stage decision vector, cx is the first-stage cost, and f is defined as the optimal value of a linear program (LP) given x and ξ; i.e.,

$$\begin{aligned} f(x, \xi) = \min_{y\geq 0}\ & qy \\ \text{s.t.}\ \ & Wy = Tx + h. \end{aligned} \tag{20}$$

We assume that $|f(x, \xi)| < \infty$, w.p.1 $\forall\, x \in X$, and $\xi = (T, h)$ so that $f(x, \cdot)$ is convex $\forall\, x \in X$.

We use three test problems APL1P [25], CEP1 [22], and PGP2 [23, 32]. All three are small models of Type (19)/(20) and involve capacity-expansion planning: CEP1 in manufacturing and APL1P and PGP2 in electric power.

In all three models, the components of ξ are independent and discretely distributed. CEP1 and PGP2 both have three random demands so that ξ consists only of elements of h. APL1P has three random demands and two random power-generator availabilities so that ξ consists of five elements of h and T. The total number of realizations of ξ is 1,280 for APL1P, 216 for CEP1, and 576 for PGP2. Therefore, we can compute exact values and assess the quality of the lower bounds. APL1P has randomness parameterized by the decision vector; i.e., the available capacities are of the form $\xi_i x_i$, $i = 1, 2$, and we discuss associated issues below.

Table 2 summarizes our computational results, displaying the lower bounds of Jensen (JB), Edirisinghe [12] (EB), and our lower bound (LB). The first column indicates the relevant expression. For example, the first row for each problem is "$\min_{x\in X} cx + Ef(x, \xi)$." The final column for this row is the optimal value, w^*, and JB for this row is $\min_{x\in X} cx + f(x, E\xi)$. The EB and LB values similarly represent minimized bounding functions. Two factors contribute to the lower bounds in the first row: the lower-bounding approximation (which holds for fixed x) and the minimization with respect to x. The second row, labeled $cx^* + Ef(x^*, \xi)$, eliminates the latter contribution. The x^* used throughout is that found in the optimization of the discretely distributed "true" problem. The row labeled $Ef(x^*, \xi)$ removes the constant term cx^*. The table lists the numerical value of each lower bound and its percentage of the true value. The multivariate lower bound is parameterized by y and z, and the "LB" values use y_i and z_i values that maximize the geometric mean of the lengths of the eight intervals $[a_i, z_i]$, $[z_i, A_{iy_i}]$, $[A_{iy_i}, A_i]$, $[A_i, \mu_i]$, $[\mu_i, B_i]$, $[B_i, B_{iz_i}]$, $[B_{iz_i}, y_i]$, and $[y_i, b_i]$ (two of which are of constant length), $i = 1, \ldots, d$. This "spreads out" the points of evaluation.

The fourth and fifth rows under each test problem correspond to componentwise independent continuous-uniform and truncated-normal distributions. The supports of these distributions are chosen so that the mean and variance match (to four significant digits) that of

Table 2. The table displays the lower bounds of Jensen (JB), Edirisinghe [12] (EB), this paper (LB), and optimal solution values for test problems CEP1, PGP2, and APL1P. The models and data for CEP1 and PGP2 are taken from Higle and Sen [23] and for APL1P from Infanger [25]. All three models are also solved under continuous-uniform and truncated-normal distributions. The support of these distributions is such that they have the same mean and variance (to four significant digits) as in the discrete case. For all problems, the first rows in the table give the lower bounds on the original model "$\min_{x\in X} cx + Ef(x,\xi)$" where the bounds, in general, are computed with respect to different first-stage decision vectors x. The second rows compare the bounds using a consistent x, namely $x = x^*$, an optimal solution of the "true" discretely distributed problem. The third rows remove the constant term cx^*, while the fourth and fifth rows use the uniform and truncated normal distributions. The reported CPU times are the average of the "Unif(x^*)" and "Tr_N(x^*)" computations, with the "True Value" entries estimated using a sample size of 25,000.

Value to be bounded	JB		EB		LB		True value (TV)
	Value	% of TV	Value	% of TV	Value	% of TV	
CEP1							
$\min_{x\in X} cx + Ef(x,\xi)$	90,250	25	263,161	74	279,162	79	355,160
$cx^* + Ef(x^*,\xi)$	113,717	32	263,161	74	279,162	79	355,160
$Ef(x^*,\xi)$	97,050	29	246,495	73	262,495	78	338,493
Unif(x^*)	97,050	28	212,997	62	238,487	70	$341{,}642 \pm 5{,}237$
Tr_N(x^*)	97,050	29	125,162	37	147,811	44	$335{,}709 \pm 5{,}279$
CPU time in sec.	0.039		0.047		0.27		68.9
PGP2							
$\min_{x\in X} cx + Ef(x,\xi)$	428.51	95.79	428.93	95.89	428.93	95.89	447.32
$cx^* + Ef(x^*,\xi)$	443.51	99.15	444.36	99.34	444.45	99.36	447.32
$Ef(x^*,\xi)$	277.01	98.64	277.86	98.95	277.95	98.98	280.82
Unif(x^*)	277.01	98.74	278.72	99.35	278.78	99.37	280.55 ± 0.85
Tr_N(x^*)	277.01	95.72	277.72	95.97	277.80	96.00	289.38 ± 1.75
CPU time in sec.	0.12		0.17		0.47		111.4
APL1P							
$\min_{x\in X} cx + Ef(x,\xi)$	23,700	96.2	24,103	97.8	24,282	98.5	24,642
$cx^* + Ef(x^*,\xi)$	23,778	96.5	24,262	98.5	24,340	98.8	24,642
$Ef(x^*,\xi)$	12,649	93.6	13,133	97.2	13,211	97.8	13,514
Unif(x^*)	12,649	96.0	12,898	97.9	12,938	98.2	$13{,}179 \pm 38$
Tr_N(x^*)	12,649	96.7	12,840	98.1	12,866	98.3	$13{,}085 \pm 36$
CPU time in sec.	0.08		0.14		38.3		68.0

the original discrete distribution. Here, the table's values bound $Ef(x^*,\xi)$, and the "True Value" column contains Monte Carlo point estimates as well as a 95% confidence intervals on $Ef(x^*,\xi)$ formed using sample sizes of 25,000.

Computing JB for fixed x requires solving one LP of Form (20). Computing EB on hyperrectangular support requires solving 2^d such LPs, where $d = 3$ for CEP1 and PGP2 and $d = 5$ for APL1P. Computing LB requires solving 7^d LPs. The number of function evaluations indicates the associated computational effort. For example, we anticipate the ratio of effort to compute LB to EB to be $(7/2)^d$. The last row under each test problem in Table 2 is the CPU time (sec.) to compute each bound on a 1.7 GHz Pentium 4 machine with 2 GB of memory using GAMS/CPLEX. The "True Value" entries average the times of the uniform and truncated-normal instances (which were very close) under the sample size of 25,000. The effect of going from $d = 3$ for CEP1 and PGP2 to $d = 5$ for APL1P is shown in the longer run time for LB in the latter case.

Three key factors affect the bounds' quality: (i) the size of the set containing ξ's support, (ii) ξ's variability, and (iii) the shape of f. In CEP1, the uniform support intervals are slightly wider than for the discrete distributions, and the truncated-normal supports are significantly wider than those of the discrete distributions. Not surprisingly, the EB and LB bounds are slightly weaker in the first case and significantly weaker in the second case. The variability of ξ for APL1P and PGP2 is significantly less than for CEP1. The Jensen bound for the former two problems is strong and the second-order bounds provide only small improvements. Of course, even if the variances of ξ and $f(x,\xi)$ are large, JB (as well as

EB and *LB*) will be exact if $f(x, \cdot)$ is linear. $f(x^*, \cdot)$ is significantly more "nonlinear" for CEP1 than for APL1P or PGP2. (In all three problems, a steep penalty is paid when demand is not met, or is met by "outsourcing." For CEP1 the probability of having unmet demand is 0.569, while the same probability for APL1P is 0.098 and for PGP2 is 0.00088.) For CEP1 the magnitude of improvement of *LB* over the *EB* bound ranges from 6% to 18%. For PGP2 and APL1P (where the Jensen bound is already fairly tight) the improvement is minimal—less than 0.8%.

As described in §1, it would be possible to tighten these bounds by applying them in a conditional fashion to a partition of ξ's support. The values in Table 2 are not formed in this manner, and instead represent a single application of the bounds.

In APL1P, the random availability elements, $x_i\xi_i$, $i = 1, 2$, are parameterized by x. Therefore, when we form the Jensen bound, we require the mean of, for example, $x_1\xi_1$—i.e., $x_1 E\xi_1$. This linear dependency allows us to minimize the Jensen bound with respect to x to obtain a lower bound on w^*. Fortunately, the same property holds for the *EB* and *LB* bounds (see, e.g., Theorem 3.1), allowing minimization of the approximating problems associated with the *EB* and *LB* bounds via two-stage stochastic linear programs.

7. Summary. We have developed a class of second-order lower bounds on the expectation of a convex function. We assume that the underlying random variable's support is contained in a finite interval. The random variable's distribution is assumed to be known only through its mean and variance. The bounds may be viewed as arising from a distributional approximation that uses either two or three points of support. We described extensions to the multivariate case where the random vector can be expressed as a linear transformation of a vector with independent components. In this case, our bounds are tighter than the second-order lower bound of Edirisinghe [12]. We note that Edirisinghe's bound may be applied to random vectors defined on a simplex and under more general dependency assumptions.

The family of bounds, $L(y, z)$, has two degrees of freedom via the parameters y and z. Finding the best bound from this class simply requires solving two univariate optimization problems using bisection search. Our bounds have application to stochastic programs with limited moment information when the recourse function is convex in its argument corresponding to the random parameters.

There are several ways that $L(y, z)$ might be improved upon. One idea, discussed in §3, would produce a tighter bound, but would require the solution of two nonconvex univariate optimization problems. A second idea concerns the "branching" performed in the proof of Theorem 3.1, which uses two cases, $x_2 \le m_1$ and $x_2 \ge m_1$. Tighter bounds might be available by branching, for example, on four intervals: $[a, A]$, $[A, m_1]$, $[m_1, B]$, and $[B, b]$. Finally, the bounding scheme we apply (e.g., in (7) and (10)) relaxes the objective function of the generalized moment problem using linear approximations. Using a properly constructed quadratic function could lead to stronger bounds (but would likely require additional function evaluations). Of course, we cannot ensure that these approaches will lead to closed-form bounds that have the kind of geometric interpretation of the bounds we have developed.

Acknowledgments. This research was supported by the National Science Foundation through Grants DMI-9702217 and DMI-0217927. The authors are grateful for the suggestions of two anonymous referees.

References

[1] Birge, J. R., J. H. Dulá. 1991. Bounding separable recourse functions with limited distribution information. *Ann. Oper. Res.* **30** 277–298.
[2] Birge, J. R., F. Louveaux. 1997. *Introduction to Stochastic Programming*. Springer-Verlag, New York.
[3] Birge, J. R., S. W. Wallace. 1986. Refining bounds for stochastic linear programs with linearly transformed independent random variables. *Oper. Res. Lett.* **5** 73–77.
[4] Birge, J. R., S. W. Wallace. 1988. A separable piecewise linear upper bound for stochastic linear programs. *SIAM J. Control Optim.* **26** 725–739.
[5] Birge, J. R., R. J.-B. Wets. 1986. Designing approximation schemes for stochastic optimization problems, in particular, for stochastic programs with recourse. *Math. Programming Stud.* **27** 54–102.

[6] Birge, J. R., R. J.-B. Wets. 1987. Computing bounds for stochastic programming problems by means of a generalized moment problem. *Math. Oper. Res.* **12** 149–162.
[7] Birge, J. R., R. J.-B. Wets. 1989. Sublinear upper bounds for stochastic programs with recourse. *Math. Programming* **43** 131–149.
[8] Dulá, J. H. 1992. An upper bound on the expectation of simplicial functions of multivariate random variables. *Math. Programming* **55** 69–80.
[9] Dulá, J. H., R. V. Murthy. 1992. A Tchebysheff-type bound on the expectation of sublinear polyhedral functions. *Oper. Res.* **40** 914–922.
[10] Dupačová, J. (as Žáčková). 1966. On minimax solutions of stochastic linear programming problems. *Časopis pro Pěstování Matematiky* **91** 423–429.
[11] Dupačová, J. 1976. Minimax stochastic programs with nonconvex nonseparable penalty functions. A. Prékopa, ed. *Progress in Operations Research (Proc. Sixth Hungarian Conf., Eger, Hungary, 1974). Colloquia Mathematica Societatis János Bolyai*, No. 12, North-Holland, Amsterdam, The Netherlands, 303–316.
[12] Edirisinghe, N. C. P. 1996. New second-order bounds on the expectation of saddle functions with applications to stochastic linear programming. *Oper. Res.* **44** 909–922.
[13] Edirisinghe, N. C. P. 1999. Bound-based approximations in multistage stochastic programming: nonanticipativity aggregation. *Ann. Oper. Res.* **85** 103–127.
[14] Edirisinghe, N. C. P., G.-M. You. 1996. Second-order scenario approximation and refinement in optimization under uncertainty. *Ann. Oper. Res.* **64** 143–178.
[15] Edirisinghe, N. C. P., W. T. Ziemba. 1994. Bounding the expectation of a saddle function with application to stochastic programming. *Math. Oper. Res.* **19** 314–340.
[16] Edirisinghe, N. C. P., W. T. Ziemba. 1994. Bounds for two-stage stochastic programs with fixed recourse. *Math. Oper. Res.* **19** 292–313.
[17] Edmundson, H. P. 1956. Bounds on the expectation of a convex function of a random variable. Technical report, Paper 982. The Rand Corporation, Santa Monica, CA.
[18] Frauendorfer, K. 1988. Solving SLP recourse problems with arbitrary multivariate distributions—the dependent case. *Math. Oper. Res.* **13** 377–394.
[19] Frauendorfer, K. 1992. Stochastic two-stage programming. *Lecture Notes in Economics and Mathematical Systems*, No. 392. Springer-Verlag, Berlin, Germany.
[20] Frauendorfer, K. 1996. Barycentric scenario trees in convex multistage stochastic programming. *Math. Programming* **75** 277–293.
[21] Gassmann, H., W. T. Ziemba. 1986. A tight upper bound for the expectation of a convex function of a multivariate random variable. *Math. Programming Stud.* **27** 39–53.
[22] Higle, J. L., S. Sen. 1994. Finite master programs in regularized stochastic decomposition. *Math. Programming* **67** 143–168.
[23] Higle, J. L., S. Sen. 1996. *Stochastic Decomposition: a Statistical Method for Large Scale Stochastic Linear Programming*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
[24] Huang, C. C., W .T. Ziemba, A. Ben-Tal. 1977. Bounds on the expectation of a convex function of a random variable: with applications to stochastic programming. *Oper. Res.* **25** 315–325.
[25] Infanger, G. 1994. *Planning Under Uncertainty: Solving Large-Scale Stochastic Linear Programs*. The Scientific Press Series, Boyd and Fraser, Danvers, MA.
[26] Jensen, J. L. 1906. *Sur les fonctions convexes et les inégalités entre les valeurs moyennes*. *Acta Math.* **30** 175–193.
[27] Kall, P. 1988. Stochastic programming with recourse: upper bounds and moment problems—a review. J. Guddat, B. Bank, H. Hollatz, P. Kall, D. Klatte, B. Kummer, K. Lommatzch, K. Tammer, M. Vlach, K. Zimmermann, eds. *Advances in Mathematical Optimization*. Akademie-Verlag, Berlin, Germany, 86–103.
[28] Kall, P. 1991. An upper bound for SLP using first and total second moments. *Ann. Oper. Res.* **30** 267–276.
[29] Karr, A. F. 1983. Extreme points of certain sets of probability measures, with applications. *Math. Oper. Res.* **8** 74–85.
[30] Kemperman, J. 1968. The general moment problem: a geometric approach. *Ann. Math. Statist.* **39** 93–122.
[31] Kreĭn, M. G., A. A. Nudel'man. 1977. *The Markov Moment Problem and Extremal Problems. Translations of Mathematical Monographs*, No. 50. American Mathematical Society, Providence, RI.
[32] Louveaux, F. V., Y. Smeers. 1988. Optimal investments for electricity generation: a stochastic model and a test-problem. Y. Ermoliev, R. J.-B. Wets, eds. *Numerical Techniques for Stochastic Optimization*. Springer-Verlag, Berlin, Germany, 445–453.
[33] Madansky, A. 1959. Bounds on the expectation of a convex function of a multivariate random variable. *Ann. Math. Statist.* **30** 743–746.
[34] Madansky, A. 1960. Inequalities for stochastic linear programming problems. *Management Sci.* **6** 197–204.
[35] Morton, D. P., R. K. Wood. 1999. Restricted-recourse bounds for stochastic linear programming. *Oper. Res.* **47** 943–956.
[36] Powell, W. B., L. F. Frantzeskakis. 1994. Restricted recourse strategies for dynamic networks with random arc capacities. *Transportation Sci.* **28** 3–23.
[37] Wallace, S. W. 1987. A piecewise linear upper bound on the network recourse function. *Math. Programming* **38** 133–146.