

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

Steftcho P. Dokov and David P. Morton  
Graduate Program in Operations Research and Industrial  
Engineering  
The University of Texas at Austin  
Austin, Texas 78712

`steftcho@mail.utexas.edu` and `morton@mail.utexas.edu`

February 21, 2001

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

STEFCHO P. DOKOV and DAVID P. MORTON

*Graduate Program in Operations Research and Industrial Engineering  
The University of Texas at Austin  
Austin, Texas 78712*

February 21, 2001

**Abstract.** 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 hyper-rectangle. 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.

## 1 Introduction

In this paper we introduce a new 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 hyper-rectangle, and the bounds use limited moment information, requiring only the first and second moments.

Jensen's (1906) 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 (Edmundson 1956, Madansky 1959) is a first-order upper bound on the expectation of a convex function. Madansky (1960) 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 1997, Chpt. 9).

Generalized moment problems (GMPs) (Karr 1983, Kemperman 1968, Kreĭn and Nudel'man 1977) provide a unifying framework for developing bounds that use limited moment information from the underlying distribution. 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 (1966, 1976). Generalizations of the *EM* bound have been developed by Dupačová (1976), Frauendorfer (1988) and Gassmann and Ziemba (1986). For more on using GMPs to find upper bounds for stochastic programs see Birge and Wets (1987) and Kall (1988). 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 regarding the underlying distribution, bounds based on limited moment information can be 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 1986, Birge and Wets 1986, Frauendorfer 1992, Huang, Ziemba and Ben-Tal 1977). The bounds that we develop in this paper can also be applied in this fashion.

One way of tightening first-order bounds is to apply them in a conditional manner as indicated above. Alternatively, when further information is available (e.g., the second moment) the bound can be tightened by using this knowledge. Dupačová (1966) incorporates a variance constraint in a GMP for finding an upper bound. Birge and Dulá (1991), Dulá (1992), Dulá and Murthy (1992), and Kall (1991) use second moment information in computing upper bounds.

There is significantly less work on second-order lower bounds. Edirisinghe (1996) 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 (1996) and is extended to multi-stage stochastic programs in Edirisinghe (1999). (Also see Frauendorfer (1996) for bounds for multi-stage 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 can be more expensive to compute.

We note that bounds based on convexity can be applied to convex-concave saddle functions; see Edirisinghe (1996), Edirisinghe and Ziemba (1994*a*, 1994*b*), and Frauendorfer (1992). In addition to bounds based on distributional approximations there are bounds based on functional approximations. See, for example, Birge and Wallace (1988), Birge and Wets (1986), Birge and Wets (1989), Morton and Wood (1999), Powell and Frantzeskakis (1994), and Wal-

lace (1987).

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 Section 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 Section 5, we generalize the results to convex functions of random vectors. The multivariate bounds are applied to two stochastic programs in Section 6 and the paper is summarized in Section 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

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

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

$$\int_a^b u dP = m_1, \quad (2)$$

$$\int_a^b u^2 dP = m_2. \quad (3)$$

Here,  $\mathcal{F}$  is the Borel field of  $[a, b]$ ,  $f: [a, b] \rightarrow \mathcal{R}$  is a convex function, and  $\xi$  is a nondegenerate random variable defined on  $([a, b], \mathcal{F}, P)$ . Karr (1983) 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 (1977, Chpt. 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 (1991) 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.

If  $x_1$ ,  $x_2$ , and  $x_3$  are distinct then, without loss of generality, we can assume  $a \leq x_1 < x_2 < x_3 \leq b$ . The vectors

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

are linearly independent (they form a  $3 \times 3$  Vandermonde matrix). In this case, (1) can be written in the equivalent form

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

Solving for the  $p_i$ 's in terms of  $x_i$ 's,  $i = 1, 2, 3$ , and the given moments  $m_1$  and  $m_2$  leads to the following optimization problem in three variables.

$$\begin{aligned} \inf \quad & \left\{ \frac{m_2 + x_2 x_3 - m_1(x_2 + x_3)}{(x_2 - x_1)(x_3 - x_1)} f(x_1) + \frac{m_1(x_1 + x_3) - x_1 x_3 - m_2}{(x_2 - x_1)(x_3 - x_2)} f(x_2) \right. \\ & \left. + \frac{m_2 + x_1 x_2 - m_1(x_1 + x_2)}{(x_3 - x_1)(x_3 - x_2)} f(x_3) \right\} \quad (5) \\ \text{s.t.} \quad & m_2 + x_2 x_3 - m_1(x_2 + x_3) \geq 0 \\ & m_1(x_1 + x_3) - x_1 x_3 - m_2 \geq 0 \\ & m_2 + x_1 x_2 - m_1(x_1 + x_2) \geq 0 \\ & a \leq x_1 < x_2 < x_3 \leq b. \end{aligned}$$

Solving for  $p_1$ ,  $p_2$ , and  $p_3$  in this manner requires that the three vectors in (4) be linearly independent. This would not be the case if two of the points coincide. We note that the optimal objective value to (5) remains the same if the strict inequalities in the ordering constraints are replaced with inclusive inequalities (and the ‘‘inf’’ in the objective is replaced with a ‘‘min’’). If, for example,  $x_1 = x_2$ , combining the first two terms in the objective function shows that it remains well-behaved. In what follows we use  $X$  to denote the feasible region of (5) with inclusive inequalities on the ordering constraints.

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 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  $f$  and instead will bound its optimal objective value from below to derive a class of closed-form lower bounds.

### 3 A Class of Second-Order Lower Bounds

The next theorem states the main result of the paper for univariate functions.

**Theorem 1** *Let  $f : [a, b] \rightarrow \mathcal{R}$  be a convex function, and let  $\xi$  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*

$$E^P f(\xi) \geq L(y, z) = \min\{L_1(y), L'_1(y), L_2(z), L'_2(z)\} \quad \forall y \in [B, b], z \in [a, A], \quad (6)$$

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,

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

The feasible region can be re-expressed

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

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

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

In order to derive the lower bound we effectively branch on  $x_2 \leq m_1$  and  $x_2 \geq m_1$ . Specifically, in the first case we find a lower bound on a relaxation of  $X \cap \{x : x_2 \leq m_1\}$ .

And, in the second case we find a lower bound on a relaxation of  $X \cap \{x : x_2 \geq m_1\}$ . Finally, we conclude that the smaller of these two local lower bounds is a global lower bound on  $\min_X F(x)$ .

**Case 1**,  $x_2 \leq m_1$ .

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

$$\frac{f(x_3) - f(x_2)}{x_3 - x_2} \leq \frac{f(x_3) - f(m_1)}{x_3 - m_1} \quad \text{and} \quad \frac{f(x_2) - f(x_1)}{x_2 - x_1} \geq \frac{f(\bar{A}) - f(x_1)}{\bar{A} - x_1} \quad (7)$$

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,

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

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

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

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

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

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

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

and

$$f(x_3) \geq \frac{x_3 - m_1}{B - m_1} f(B) - \frac{x_3 - B}{B - m_1} f(m_1) = \frac{m_1 - a}{\sigma^2} \left[ (x_3 - m_1) f(B) - (x_3 - B) f(m_1) \right]. \quad (10)$$

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

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

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

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

Applying a similar analysis it can be shown that

$$\min_{y \leq x_3 \leq b} g(x_3) \geq \frac{b - m_1}{b - a} f(A) + \frac{m_1 - a}{b - a} \left( \frac{y - B}{y - m_1} f(m_1) + \frac{B - m_1}{y - m_1} f(y) \right) = L'_1(y).$$

Summarizing our results for Case 1, we have shown

$$E^P f(\xi) \geq \min\{L_1(y), L'_1(y)\} \quad \forall y \in [B, b] \text{ when } x_2 \leq m_1.$$

**Case 2**,  $x_2 \geq m_1$ .

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

$$E^P f(\xi) \geq \min\{L_2(z), L'_2(z)\} \quad \forall z \in [a, A] \text{ when } x_2 \geq m_1.$$

Combining the results from Cases 1 and 2 gives the desired result.  $\square$

Theorem 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 tradeoff between the quality of the bound and the effort required to compute it. In Section 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  $f$  in order 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 2 and 3.

**Corollary 2** *Under the hypotheses of Theorem 1,*

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

Moreover,

$$L(y, z) \geq EB, \quad \forall y \in [B, b], z \in [a, A].$$



**Proof:** Let  $z = A$  and  $y = B$ . It is straightforward to show  $L_1(B) \geq L'_1(B)$  and  $L_2(A) \geq L'_2(A)$  by using the fact that  $f$  is convex. The inequality (12) then follows from  $L'_1(B) = L'_2(A) = EB$ .

In order 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  $L_1(y) = \frac{y-m_1}{y-a}f(A_y) + \frac{m_1-a}{y-a}f(B)$ .  $L_1(y') \geq L_1(y'')$  is equivalent to

$$\begin{aligned} \frac{y''-m_1}{y''-a}f(A_{y''}) + \frac{m_1-a}{y''-a}f(B) &\leq \frac{y'-m_1}{y'-a}f(A_{y'}) + \frac{m_1-a}{y'-a}f(B) && \iff \\ f(A_{y''}) &\leq \frac{(y''-a)(y'-m_1)}{(y'-a)(y''-m_1)}f(A_{y'}) + \frac{(m_1-a)(y''-y')}{(y'-a)(y''-m_1)}f(B) && \iff \\ f(A_{y''}) &\leq \frac{B-A_{y''}}{B-A_{y'}}f(A_{y'}) + \frac{A_{y''}-A_{y'}}{B-A_{y'}}f(B), \end{aligned}$$

which holds since  $f$  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  $L'_2(A) = EB$ .  $\square$

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 the price to pay for this savings is that the bound is the weakest from our class. Edirisinghe (1996) 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 hyper-rectangle, and the  $EB$  bound in (12) is the special case of his bound in one dimension. We will return to the monotonicity results established in the proof of Corollary 2 for  $L_1$ ,  $L'_1$ ,  $L_2$ , and  $L'_2$  in Section 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 require 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** *Under the hypotheses of Theorem 1,*

$$\min\{L'_1(B_z), L'_2(z)\} \leq E^P f(\xi) \quad \forall z \in [a, A].$$

**Proof:** Let  $z \in [a, A]$ ,  $y = B_z$ , and note  $B_z \in [B, b]$ . The desired result then follows immediately from (6),  $L_1(B_z) \geq L'_2(z)$  and  $L_2(z) \geq L'_1(B_z)$ . 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

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

Since  $f$  is convex,

$$f(m_1) \leq \frac{m_1 - z}{B - z} f(B) + \frac{B - m_1}{B - z} f(z). \quad (13)$$

The inequality  $L_1(B_z) \geq L'_2(z)$  follows by replacing  $f(m_1)$  in the definition of  $L'_2(z)$  with the right-hand side of (13).  $\square$

The expressions in Theorem 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  $\xi$ . In a similar way,  $L'_1(y)$  is a convex combination of  $f(A)$  and an expression which 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 (1996) 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 in the proof of Corollary 2, 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 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 support of the approximating distribution either has two points or three points, depending on which of the four expectations is smallest. The first moment of all four distributions is  $m_1$ , the mean of the underlying random variable  $\xi$ . 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 Section 2, and hence we cannot guarantee the bound will be sharp.

While  $L(y, z)$  can be strengthened by optimizing over  $y$  and  $z$  (see Section 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 note, however, that a variant of our approach could

lead to a stronger bound. In particular, a valid lower bound would be achieved by solving the univariate optimization problem (9), as well as its analog for Case 2, and taking the smaller of these two optimal values. This possible approach is complicated by the fact these two univariate optimization problems are nonconvex, and a global optimum is required to ensure a valid bound.

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  $\sigma^2$ , of the underlying random variable. While this is clearly less demanding than assuming the distribution of  $\xi$  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  $\sigma$ . There are three important properties: First,  $L(y, z)$  is nondecreasing in  $\sigma$  provided  $m_1$ ,  $a$ , and  $b$  are fixed. Second,  $L(y, z)$  decreases as  $a$  decreases or  $b$  increases provided  $m_1$  and  $\sigma$  are fixed as well as  $b$  or  $a$ , respectively. Third, the  $EB = L(a, b)$  bound from Corollary 2 is a convex function of  $m_1$  provided  $a$ ,  $b$ , and  $m_2$  are fixed. The first two properties can be seen from the geometric interpretation of  $L(y, z)$  illustrated in Figures 1 and 2 and can be easily verified analytically. The third property follows immediately from the definitions of  $EB$  and convexity. The first and third properties provide a way to find a valid second-order lower bound when instead of knowing  $m_1$  and  $\sigma$  precisely we have a ranges:  $m' \leq m_1 \leq m''$  and  $\sigma' \leq \sigma \leq \sigma''$ . Solving the convex minimization problem  $\min_{m' \leq m_1 \leq m''} EB$  yields a valid lower bound for all  $m_1 \in [m', m'']$  while  $L(y, z)|_{\sigma=\sigma'}$  is a valid lower bound 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  $a \rightarrow -\infty$  and  $b \rightarrow \infty$ , 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,  $L(y, z) \leq E^P f(\xi)$ , 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

$$L^* = \max_{y \in [B, b], z \in [a, A]} L(y, z) = \min \left\{ \max_{y \in [B, b]} \min\{L_1(y), L'_1(y)\}, \max_{z \in [a, A]} \min\{L_2(z), L'_2(z)\} \right\}. \quad (14)$$

Consider the two univariate maximization problems defined on the right-hand side of (14), and recall from the proof of Corollary 2 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 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 since

$$L_1(y) \geq L'_1(y) \forall y \in [B, y^*] \text{ and } L_1(y) \leq L'_1(y) \forall y \in [y^*, b]$$

and

$$L_2(z) \leq L'_2(z) \forall z \in [a, z^*] \text{ and } L_2(z) \geq L'_2(z) \forall z \in [z^*, A].$$

The best lower bound from the subclass of bounds defined in Corollary 3 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 1** Let  $\xi$  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

$$\begin{aligned} L_1(y) &= \frac{y-4}{y} f(A_y) + \frac{4}{y} f(5), \\ L'_1(y) &= \frac{1}{3} f(2) + \frac{2}{3} \left( \frac{y-5}{y-4} f(4) + \frac{1}{y-4} f(y) \right), \\ L_2(z) &= \frac{2}{6-z} f(2) + \frac{4-z}{6-z} f(B_z), \\ L'_2(z) &= \frac{1}{3} \left( \frac{2}{4-z} f(z) + \frac{2-z}{4-z} f(4) \right) + \frac{2}{3} f(5), \end{aligned}$$

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 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, 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, Edirisinghe (1996),  $L^*$ , 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  $f$  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 (1977, Chpt. 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).  $\square$

## 5 Multivariate Generalizations

In this section we generalize the bound of Theorem 1 to a convex function of a random vector with independent components. We first adapt our notation to the multivariate setting. Let  $\xi = (\xi_1, \dots, \xi_d)^T$  be a random vector with independent components. The support of  $\xi$  is contained in  $\prod_{i=1}^d [a_i, b_i]$ , the joint distribution of  $\xi$  is denoted  $P$ ,  $E\xi_i = \mu_i$ , and  $\text{var } \xi_i = \sigma_i^2$ . 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  $f: \prod_{i=1}^d [a_i, b_i] \rightarrow \mathcal{R}$  be a convex function. The lower bound on  $E^P f(\xi)$  we define below is denoted  $L(y, z)$  and is parameterized by  $y = (y_1, \dots, y_d)^T$  and  $z = (z_1, \dots, z_d)^T$ .

For the univariate case described in Section 3, the four functions defined in Theorem 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  $i = 1, \dots, d$  with the following probability mass functions:

$$\begin{aligned}
Q_{i1} : \quad Pr(A_{iy_i}) &= \frac{B_i - \mu_i}{B_i - A_{iy_i}}, \quad Pr(B_i) = \frac{\mu_i - A_{iy_i}}{B_i - A_{iy_i}}, \\
Q_{i1'} : \quad Pr(A_i) &= \frac{B_i - \mu_i}{B_i - A_i}, \quad Pr(\mu_i) = \frac{(\mu_i - A_i)(y_i - B_i)}{(B_i - A_i)(y_i - \mu_i)}, \quad Pr(y_i) = \frac{(\mu_i - A_i)(B_i - \mu_i)}{(B_i - A_i)(y_i - \mu_i)}, \\
Q_{i2} : \quad Pr(A_i) &= \frac{B_{iz_i} - \mu_i}{B_{iz_i} - A_i}, \quad Pr(B_{iz_i}) = \frac{\mu_i - A_i}{B_{iz_i} - A_i}, \\
Q_{i2'} : \quad Pr(B_i) &= \frac{\mu_i - A_i}{B_i - A_i}, \quad Pr(z_i) = \frac{(B_i - \mu_i)(\mu_i - A_i)}{(B_i - A_i)(\mu_i - z_i)}, \quad Pr(\mu_i) = \frac{(B_i - \mu_i)(A_i - z_i)}{(B_i - A_i)(\mu_i - z_i)}.
\end{aligned} \tag{15}$$

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 4** *Let  $f : \prod_{i=1}^d [a_i, b_i] \rightarrow \mathcal{R}$  be a convex function. Let  $\xi$  be a random vector with independent components, support contained in  $\prod_{i=1}^d [a_i, b_i]$ ,  $E\xi_i = \mu_i$ , and  $\text{var } \xi_i = \sigma_i^2 > 0$ ,  $i = 1, \dots, d$ . Then*

$$E^P f(\xi) \geq L(y, z) = \min \left\{ E^{Q_{1j_1}} E^{Q_{2j_2}} \dots E^{Q_{dj_d}} f(\xi_1, \dots, \xi_d) : j_1, \dots, j_d = 1, 1', 2, 2' \right\} \tag{16}$$

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 (15),  $y \in \prod_{i=1}^d [B_i, b_i]$  and  $z \in \prod_{i=1}^d [a_i, A_i]$ .

**Proof:** We give an outline of the proof which uses the same ideas as the proof of Theorem 1. Let the marginal distributions of  $\xi$  be denoted  $P_i$ ,  $i = 1, \dots, d$ . Then

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

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 1, with obvious generalizations of the notation to the  $d^{\text{th}}$  component under consideration, we have

$$E^P f(\xi) \geq \begin{cases} E^{P_1} \dots E^{P_{d-1}} E^{Q_{d1}(y_d)} f(\xi_1, \dots, \xi_{d-1}, \xi_d) & \text{for } x_{d2} \leq \mu_d, x_{d3} \in [B_d, y_d] \\
E^{P_1} \dots E^{P_{d-1}} E^{Q_{d1'}(y_d)} f(\xi_1, \dots, \xi_{d-1}, \xi_d) & \text{for } x_{d2} \leq \mu_d, x_{d3} \in [y_d, b_d] \\
E^{P_1} \dots E^{P_{d-1}} E^{Q_{d2}(z_d)} f(\xi_1, \dots, \xi_{d-1}, \xi_d) & \text{for } x_{d2} \geq \mu_d, x_{d1} \in [z_d, A_d] \\
E^{P_1} \dots E^{P_{d-1}} E^{Q_{d2'}(z_d)} f(\xi_1, \dots, \xi_{d-1}, \xi_d) & \text{for } x_{d2} \geq \mu_d, x_{d1} \in [a_d, z_d], \end{cases} \tag{17}$$

where  $Q_{d1}$ ,  $Q_{d1'}$ ,  $Q_{d2}$ , and  $Q_{d2'}$  are defined in (15). Next we consider the  $d-1^{\text{st}}$  component and for each of the four cases in (17) we have four subcases corresponding to: (i)  $x_{d-1,2} \leq \mu_{d-1}$ ,

$x_{d-1,3} \in [B_{d-1}, y_{d-1}]$ , (ii)  $x_{d-1,2} \leq \mu_{d-1}$ ,  $x_{d-1,3} \in [y_{d-1}, b_{d-1}]$ , (iii)  $x_{d-1,2} \geq \mu_{d-1}$ ,  $x_{d-1,1} \in [z_{d-1}, A_{d-1}]$ , and (iv)  $x_{d-1,2} \geq \mu_{d-1}$ ,  $x_{d-1,1} \in [a_{d-1}, z_{d-1}]$ . 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  $\xi$ , resulting in a total of  $4^d$  cases. In order to achieve a valid lower bound we take the minimum over all of these cases, as specified in (16).  $\square$

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 (1996) to random vectors with rectangular support (although those bounds require only  $2^d$  function evaluations).

Obvious analogs of Corollaries 2 and 3 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 (1996, Theorem 5). Analogous to Corollary 2, the multivariate version of  $L(y, z)$  is at least as strong as Edirisinghe's bound on a rectangular domain.

Clearly, the dimension of  $\xi$  must be modest in order for  $L(y, z)$  to be computable. In addition, Theorem 4 requires the components of  $\xi$  to be independent. One commonly used probabilistic modeling technique ameliorates both of these limitations. Often a random vector of dimension  $n$ , say  $\eta$ , with dependent components is approximated via of a deterministic linear transformation,  $\eta = H\xi$ , of another random vector,  $\xi$ . Here,  $H \in \mathcal{R}^{n \times d}$ , and the random vector  $\xi$  typically consists of a small number  $d$  (relative to  $n$ ) of independent factors that “explain” the randomness in  $\eta$ . So in our setting, if we wish to bound  $Eh(\eta)$  where  $h: \mathcal{R}^n \rightarrow \mathcal{R}$  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  $\max \{L(y, z) : y \in \prod_{i=1}^d [B_i, b_i], z \in \prod_{i=1}^d [a_i, A_i]\}$  may be hard. The difficulty is that  $L(y, z)$  is nondifferentiable and is not concave (or even quasiconcave). Nondifferentiability arises even when  $f$  is smooth because of the min operator in (16).

## 6 Computational Results

In this section we empirically analyze the behavior of the multivariate lower bound of Section 5 on two stochastic programming test problems from the literature. These problems are two-stage stochastic programs with recourse that take the form

$$w^* = \min_{x \in X} cx + Ef(x, \xi), \quad (18)$$

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 given  $x$  and  $\xi$ , i.e.,

$$f(x, \xi) = \min_{y \geq 0} \quad qy \tag{19}$$

$$\text{s.t.} \quad Wy = Tx + h.$$

In general,  $\xi$  is the vector of random elements from  $q, W, T$ , and  $h$ . We assume that  $f(x, \xi)$  is finite, w.p.1, for all  $x \in X$ . When the randomness only appears in the right-hand side of (19), i.e., only in  $T$  and  $h$ ,  $f(x, \cdot)$  is convex for all  $x \in X$ .

Two-stage stochastic programs have received significant attention in the operations research literature because of their ability to adequately model many stochastic systems that require optimization. See, for example, the textbooks by Birge and Louveaux (1997), Kall and Wallace (1994) and Prékopa (1995).

The two test problems we use are called CEP1 (Higle and Sen 1994) and PGP2 (Louveaux and Smeers 1988, Higle and Sen 1996b) and are small two-stage stochastic programs with recourse. Both models are capacity-expansion planning models (CEP1 in manufacturing and PGP2 in electric power), and both have a demand vector with three independent and discretely distributed components (in the vector  $\xi = h$ ). Thus,  $f(x, \cdot)$  is a convex function to which we can apply the bounds developed in Theorem 4. The total number of realizations of  $\xi$  is 216 for CEP1 and 576 for PGP2 so we can compute  $w^*$  and  $Ef(x, \xi)$ , for fixed values of  $x$ , exactly and assess the quality of the lower bounds. CEP1 and PGP2 are valuable for analysis purposes because of their complementary nature. In particular, for CEP1 the variance of  $f(x, \xi)$  is relatively large, but the optimization in (18) with respect to  $x$  is relatively easy. On the other hand, for PGP2 the variance of  $f(x, \xi)$  is relatively small but  $Ef(x, \xi)$  is fairly flat in the neighborhood of  $x^*$  and so exact optimization is more difficult. (See Frauendorfer (1992), Higle and Sen (1996a) and Higle (1998).)

Table 2 summarizes our computational results for CEP1 and PGP2. The table displays the lower bounds of Jensen (*JB*) and Edirisinghe (1996) (*EB*) as well as our lower bound (*LB*). The first column indicates the relevant expression. For example, the first row under each problem is labeled  $\min_{x \in X} cx + Ef(x, \xi)$ . The final column of the table for this row is simply the optimal solution value,  $w^*$ . The Jensen bound (*JB*) for this row is  $\min_{x \in X} cx + f(x, E\xi)$  while the *EB* and *LB* values similarly represent the minimum value (with respect to  $x$ ) of the bounding function. There are two contributions to the lower bounds in the first row: (i) the lower bounding approximation (which holds for fixed  $x$ ) and (ii) the minimization with respect to  $x$ . The second row, labeled  $cx^* + Ef(x^*, \xi)$ , eliminates the contribution due to the minimization so that the bound can be examined with respect to only the former reason. The  $x^*$  used throughout is that found in the optimization of the “true” problem. The values for the



final column of the first two rows are identical by definition. Finally, the row labeled  $Ef(x^*, \xi)$  removes the constant term  $cx^*$  because it has nothing to do with the error introduced by the lower-bounding approximation.

For each lower bound, the table lists the numerical value and its percentage of the true value. The fourth and fifth rows under each test problem are labeled “Unif( $x^*$ )” and “Tr\_N( $x^*$ ).” The probability distributions used for these model instances are the continuous uniform and truncated normal. As before, the components of  $\xi$  are assumed to be independent. The supports for each uniform and truncated normal distribution are chosen so that the mean and variance (to four significant digits) are identical to that of the discrete distribution. The numerical values listed are the associated bounds on  $Ef(x^*, \xi)$ , where again the value of  $x^*$  is that from the discretely distributed problem. For these continuous distributions we are unable to compute  $Ef(x^*, \xi)$  and so we estimate it using Monte Carlo sampling. For these rows, the “True value” column contains the unbiased point estimate as well as a 95% confidence interval derived using a sample size of 5000. Theorem 4 provides a family of lower bounds parameterized by the vectors  $y$  and  $z$ . The value listed under column “ $LB$ ” is that obtained by optimizing  $L(y, z)$  over a grid of points in the parameters  $y$  and  $z$ . We note that achieving 100% of the “True value” is likely to be impossible for any bound based on limited information since the “true” distribution we have used almost certainly does not solve a GMP.

Three key factors affect the quality of the second-order bounds: (i) the size of the set containing  $\xi$ ’s support (ii)  $\xi$ ’s variability, and (iii) the shape of the convex function  $f$ . In CEP1, the widths of the support intervals for the continuous uniform 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 lower bounds  $EB$  and  $LB$  are slightly weaker in the first case and significantly weaker in the second case (see the discussion at the end of Section 3). The variability of  $\xi$  for PGP2 is significantly less than for CEP1. The Jensen bound for PGP2 is very strong and the second-order bounds provide only marginal improvement over  $JB$ . Of course, even if the variances of  $\xi$  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 PGP2. These functions aren’t differentiable so we cannot examine metrics such as those given by the Hessian. However, the nature of both problems is that a steep penalty must be 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 PGP2 is 0.00088. The observation from the literature that the optimization in CEP1 is “easy” is consistent with the fact that optimizing the  $EB$  and  $LB$  bounds gave the same  $x^*$  as the true distribution despite the fact that the associated distributional approximations are rather crude. This doesn’t occur for the Jensen approximation in CEP1 or for any of the approximations for

PGP2. For CEP1 the magnitude of improvement of  $LB$  over the  $EB$  bound ranges from 6% to 8%. For PGP2 (where the Jensen bound is already very tight) the improvement is minimal, less than 0.06%.

## 7 Summary

We have developed a class of second-order lower bounds on the expectation of a convex function. We assume 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 (1996). 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 Section 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 1 which uses two cases,  $x_2 \leq m_1$  and  $x_2 \geq m_1$ . 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 grant DMI-9702217.

## References

- Birge, J.R., F. Louveaux 1997. *Introduction to Stochastic Programming*. Springer-Verlag, New York.
- Birge, J.R., J.H. Dulá 1991. Bounding separable recourse functions with limited distribution information. *Annals of Operations Research* **30** 277–298.
- Birge, J.R., R.J.-B. Wets 1986. Designing approximation schemes for stochastic optimization problems, in particular, for stochastic programs with recourse. *Mathematical Programming Study* **27** 54–102.
- Birge, J.R., R.J.-B. Wets 1987. Computing bounds for stochastic programming problems by means of a generalized moment problem. *Mathematics of Operations Research* **12** 149–162.
- Birge, J.R., R.J.-B. Wets 1989. Sublinear upper bounds for stochastic programs with recourse. *Mathematical Programming* **43** 131–149.
- Birge, J.R., S.W. Wallace 1986. Refining bounds for stochastic linear programs with linearly transformed independent random variables. *Operations Research Letters* **5** 73–77.
- Birge, J.R., S.W. Wallace 1988. A separable piecewise linear upper bound for stochastic linear programs. *SIAM Journal on Control and Optimization* **26** 725–739.
- Dulá, J.H. 1992. An upper bound on the expectation of simplicial functions of multivariate random variables. *Mathematical Programming* **55** 69–80.
- Dulá, J.H., R.V. Murthy 1992. A Tchebysheff-type bound on the expectation of sublinear polyhedral functions. *Operations Research* **40** 914–922.
- Dupačová, J. 1976. Minimax stochastic programs with nonconvex nonseparable penalty functions. in A.Prékopa, ed., *Progress in Operations Research*. Mathematica Societatis János Bolyai, Eger, Hungary. pp. 303–316.
- Dupačová, J. (as Žáčková) 1966. On minimax solutions of stochastic linear programming problems. *Časopis pro Pěstování Matematiky* **91** 423–429.
- Edirisinghe, N.C.P. 1996. New second-order bounds on the expectation of saddle functions with applications to stochastic linear programming. *Operations Research* **44** 909–922.
- Edirisinghe, N.C.P. 1999. Bound-based approximations in multistage stochastic programming: Nonanticipativity aggregation. *Annals of Operations Research* **85** 103–127.
- Edirisinghe, N.C.P., G-M. You 1996. Second-order scenario approximation and refinement in optimization under uncertainty. *Annals of Operations Research* **64** 143–178.
- Edirisinghe, N.C.P., W.T. Ziemba 1994a. Bounding the expectation of a saddle function with application to stochastic programming. *Mathematics of Operations Research* **19** 314–340.
- Edirisinghe, N.C.P., W.T. Ziemba 1994b. Bounds for two-stage stochastic programs with fixed recourse. *Mathematics of Operations Research* **19** 292–313.
- Edmundson, H.P. (1956). Bounds on the expectation of a convex function of a random variable. Technical report. The Rand Corporation Paper 982. Santa Monica, California.

- Frauendorfer, K. 1988. Solving SLP recourse problems with arbitrary multivariate distributions – the dependent case. *Mathematics of Operations Research* **13** 377–394.
- Frauendorfer, K. 1992. *Stochastic Two-Stage Programming*. Lecture Notes in Economics and Mathematical Systems 392. Springer-Verlag, Berlin.
- Frauendorfer, K. 1996. Barycentric scenario trees in convex multistage stochastic programming. *Mathematical Programming* **75** 277–293.
- Gassmann, H., W.T. Ziemba 1986. A tight upper bound for the expectation of a convex function of a multivariate random variable. *Mathematical Programming Study* **27** 39–53.
- Higle, J.L. 1998. Variance reduction and objective function evaluation in stochastic linear programs. *INFORMS Journal on Computing* **10**.
- Higle, J.L., S. Sen 1994. Finite master programs in regularized stochastic decomposition. *Mathematical Programming* **67** 143–168.
- Higle, J.L., S. Sen 1996a. Duality and statistical tests of optimality for two stage stochastic programs. *Mathematical Programming* **75** 257–275.
- Higle, J.L., S. Sen 1996b. *Stochastic decomposition: a statistical method for large scale stochastic linear programming*. Kluwer Academic Publishers, Dordrecht.
- 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. *Operations Research* **25** 315–325.
- Jensen, J.L. 1906. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica* **30** 175–193.
- Kall, P. 1988. Stochastic programming with recourse: upper bounds and moment problems — a review. in J. G. et al., ed., *Advances in Mathematical Optimization*. Akademie-Verlag, Berlin. pp. 86–103.
- Kall, P. 1991. An upper bound for slp using first and total second moments. *Analys of Operations Research* **30** 267–276.
- Kall, P., S.W. Wallace 1994. *Stochastic Programming*. John Wiley and Sons, Chichester.
- Karr, A.F. 1983. Extreme points of certain sets of probability measures, with applications. *Mathematics of Operations Research* **8** 74–85.
- Kemperman, J. 1968. The general moment problem, a geometric approach. *Annals of Mathematical Statistics* **39** 93–122.
- Kreĭn, M.G., A.A. Nudel'man 1977. *The Markov Moment Problem and Extremal Problems*. American Mathematical Society, Providence, Rhode Island (Translations of Mathematical Monographs, 50).
- Louveaux, F.V., Y. Smeers 1988. Optimal investments for electricity generation: a stochastic model and a test-problem. in Y. Ermoliev, R.-B. Wets, eds, *Numerical Techniques for Stochastic Optimization*. Springer-Verlag, Berlin. pp. 445–453.

- Madansky, A. 1959. Bounds on the expectation of a convex function of a multivariate random variable. *Annals of Mathematical Statistics* **30** 743–746.
- Madansky, A. 1960. Inequalities for stochastic linear programming problems. *Management Science* **6** 197–204.
- Morton, D.P., R.K. Wood 1999. Restricted-recourse bounds for stochastic linear programming. *Operations Research* **47** 943–956.
- Powell, W.B., L.F. Frantzeskakis 1994. Restricted recourse strategies for dynamic networks with random arc capacities. *Transportation Science* **28** 3–23.
- Prékopa, A. 1995. *Stochastic Programming*. Kluwer Academic Publishers, Dordrecht.
- Wallace, S.W. 1987. A piecewise linear upper bound on the network recourse function. *Mathematical Programming* **38** 133–146.

| $f(\xi)$ | $JB$  |            | $EB$  |            | $L^*$  |            |                | $GMP$ |
|----------|-------|------------|-------|------------|--------|------------|----------------|-------|
|          | value | % of $GMP$ | value | % of $GMP$ | value  | % of $GMP$ | $y^*$ or $z^*$ | value |
| $\xi^2$  | 16    | 80         | 18    | 90         | 18.4   | 92         | 5.6458         | 20    |
| $\xi^3$  | 64    | 64         | 86    | 86         | 91.1   | 91         | 5.5308         | 100   |
| $\xi^4$  | 256   | 51         | 422   | 84         | 452.9  | 91         | 0.5274         | 500   |
| $\xi^5$  | 1024  | 41         | 2094  | 84         | 2237.0 | 89         | 0.36285        | 2500  |

Table 1: The lower bounds of Jensen ( $JB$ ), Edirisinghe (1996) ( $EB$ ),  $L^*$ , 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 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  $\xi^2$  and  $\xi^3$  the value of  $y^*$  is given (since  $L_1(y^*) < L_2(z^*)$ ) and for  $\xi^4$  and  $\xi^5$  the value of  $z^*$  is given (since  $L_2(z^*) < L_1(y^*)$ ).

| Value to be bounded              | <i>JB</i> |           | <i>EB</i> |           | <i>LB</i> |           | True value     |
|----------------------------------|-----------|-----------|-----------|-----------|-----------|-----------|----------------|
|                                  | value     | % of true | value     | % of true | value     | % of true |                |
| CEP1                             |           |           |           |           |           |           |                |
| $\min_{x \in X} cx + Ef(x, \xi)$ | 90200     | 25        | 263161    | 74        | 282911    | 80        | 355160         |
| $cx^* + Ef(x^*, \xi)$            | 113717    | 32        | 263161    | 74        | 282911    | 80        | 355160         |
| $Ef(x^*, \xi)$                   | 97050     | 28        | 246494    | 72        | 266244    | 79        | 338493         |
| Unif( $x^*$ )                    | 97050     | 29        | 212999    | 63        | 238966    | 71        | 338354 ± 11716 |
| Tr_N( $x^*$ )                    | 97050     | 29        | 123993    | 36        | 142523    | 42        | 340351 ± 11980 |
| 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     | 444.55    | 99.38     | 447.32         |
| $Ef(x^*, \xi)$                   | 277.01    | 98.64     | 277.86    | 98.95     | 278.05    | 98.95     | 280.82         |
| Unif( $x^*$ )                    | 277.01    | 98.51     | 278.22    | 99.12     | 278.78    | 99.14     | 281.20 ± 1.9   |
| Tr_N( $x^*$ )                    | 277.01    | 97.12     | 277.69    | 97.36     | 277.86    | 97.42     | 285.23 ± 2.9   |

Table 2: The table displays the lower bounds of Jensen (*JB*), Edirisinghe (1996) (*EB*), this paper’s lower bound (*LB*), and optimal solution values for test problems CEP1 and PGP2. The models and data for these problems are taken from Hige and Sen (1996b). CEP1 and PGP2 are also solved under continuous uniform and truncated normal distributions. The support of these distributions is such that they have same mean and variance (to four significant digits) as in the discrete case. For both 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.

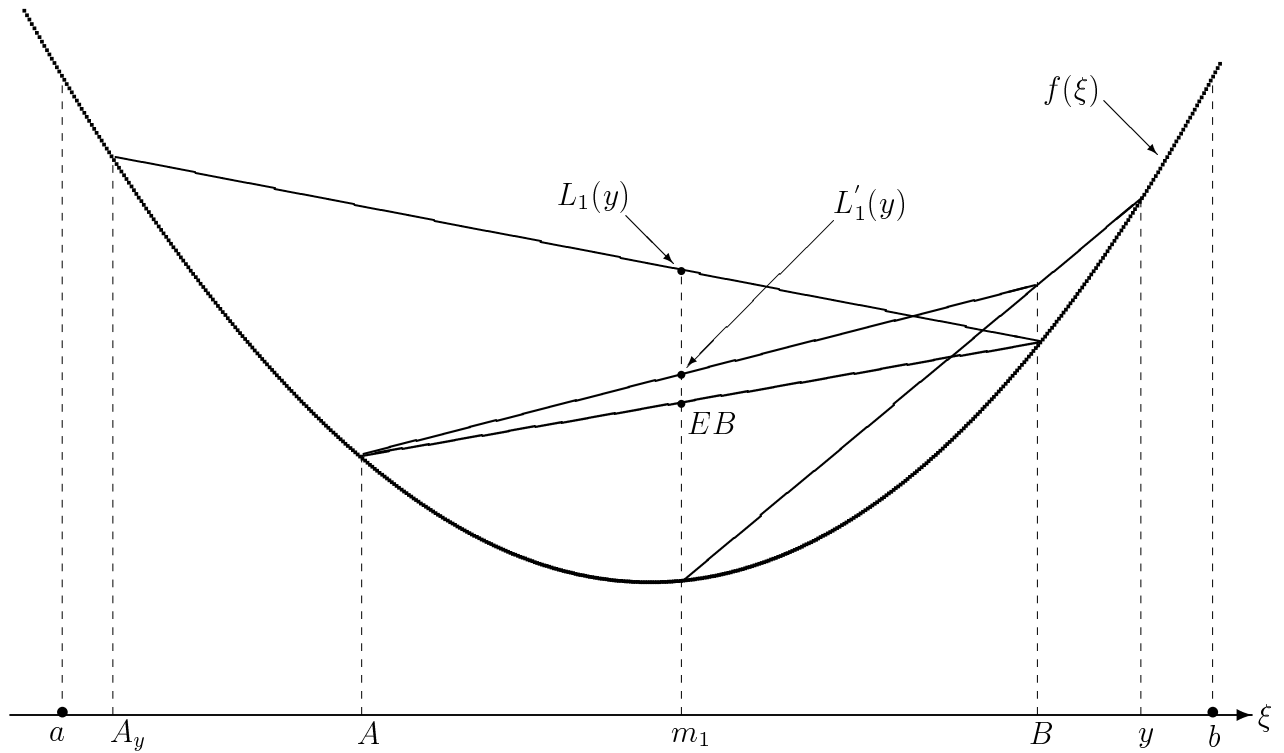


Figure 1: This figure illustrates  $L_1(y)$  and  $L'_1(y)$ , two of the four expressions that define the class of second-order lower bounds on  $E^P f(\xi)$  from Theorem 1. The parameter  $y$  is one of two degrees of freedom associated with the bound  $L(y, z)$ .  $L_1(y)$  is a convex combination 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 (1996) second-order lower bound (see  $EB$  in Corollary 2) is also shown.



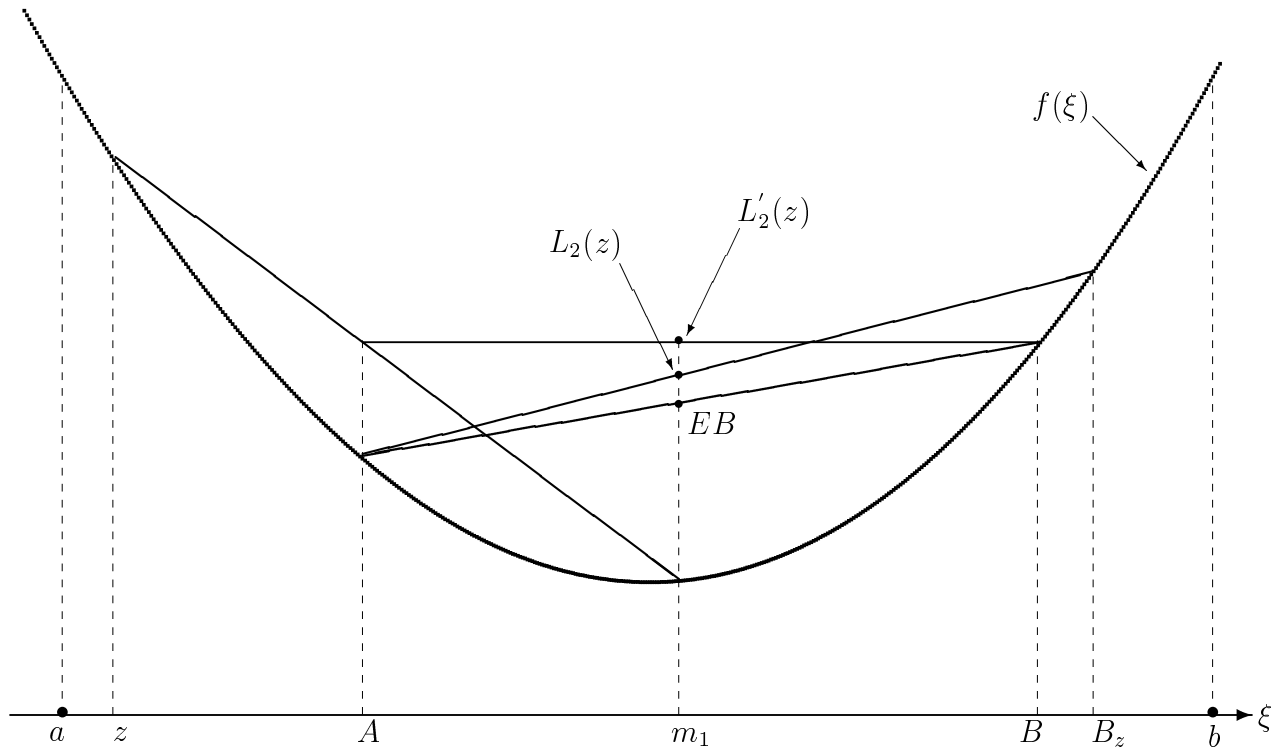


Figure 2: The two expressions associated with the second degree of freedom (i.e., parameter  $z$ ) that define the lower bound in Theorem 1 are shown here.  $L_2(z)$  is a convex combination of  $f(A)$  and  $f(B_z)$  while  $L'_2(z)$  is a convex combination of  $f(B)$  and a term which is, in turn, a convex combination of  $f(z)$  and  $f(m_1)$ .

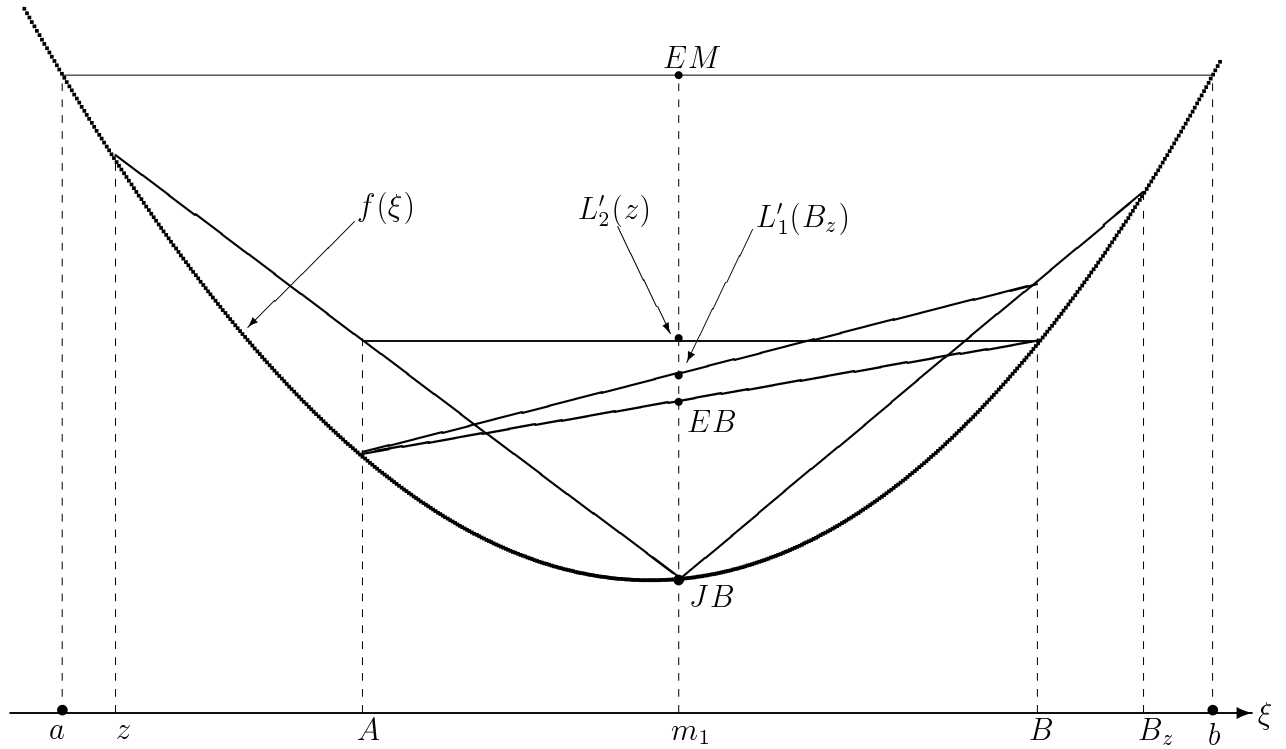


Figure 3: This figure provides a geometric interpretation of the two expressions,  $L'_1(B_z)$  and  $L'_1(z)$ , that define Corollary 3's class of second-order lower bounds. The classical first-order bounds of Jensen ( $JB$ ) and Edmundson-Madansky ( $EM$ ), as well as Edirisinghe's (1996) second-order lower bound ( $EB$ ), are also illustrated in the figure.

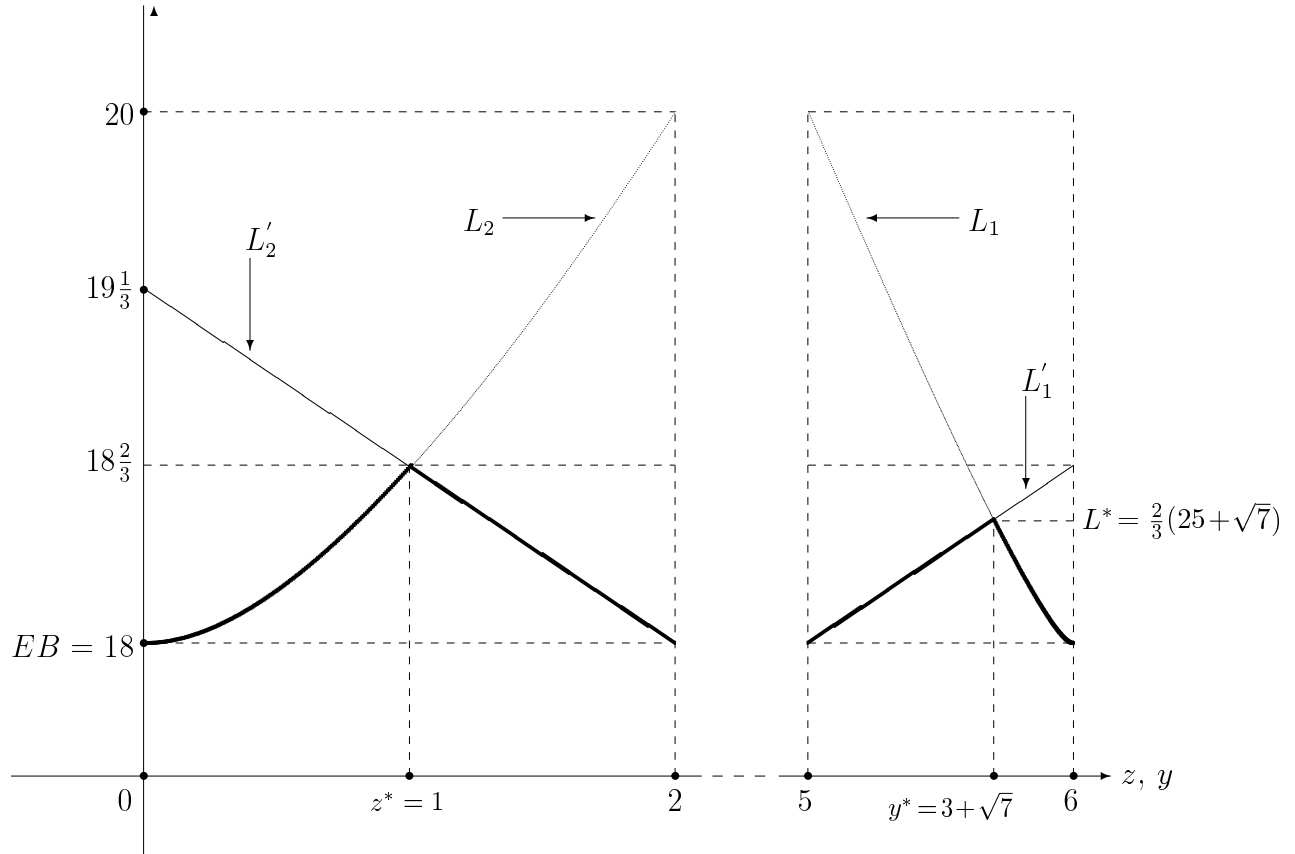


Figure 4: For Example 1 considered in Section 4, this figure shows  $L^*$ , the strongest lower bound from the class of bounds introduced in Theorem 1. The optimal values  $y^*$  and  $z^*$  (see (14)) are also indicated. The shadow lines are the two functions  $\min\{L_1(y), L_1'(y)\}$  and  $\min\{L_2(z), L_2'(z)\}$ .

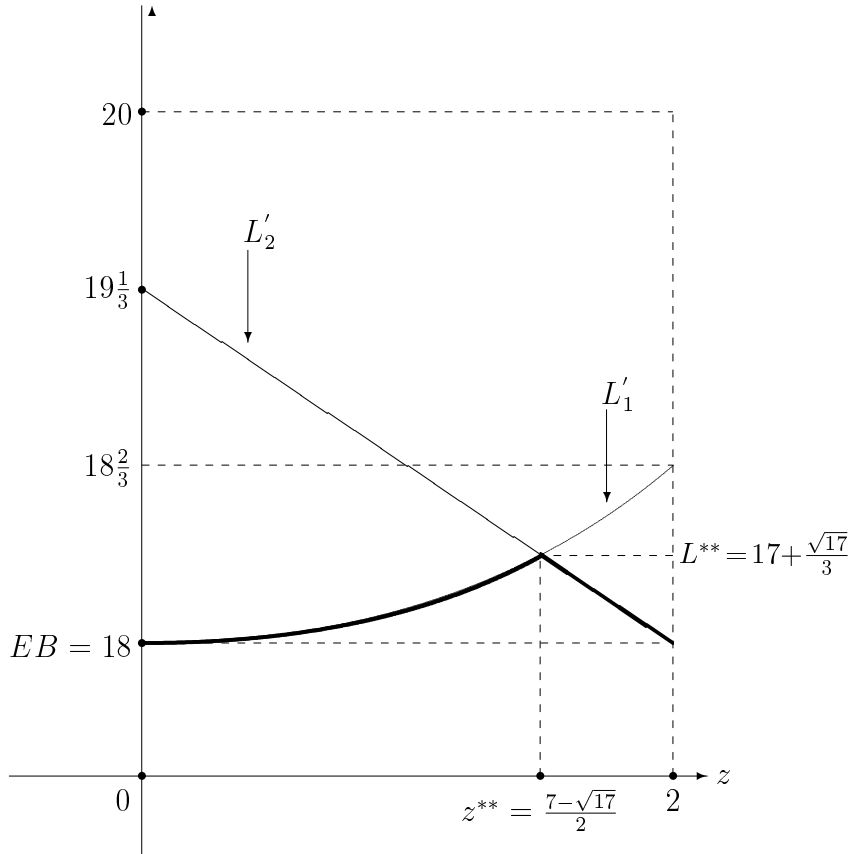


Figure 5: For Example 1, the shadow line is the class of lower bounds  $\min\{L'_1(B_z), L'_2(z)\}$  from Corollary 3. The best lower bound in this class,  $L^{**} = L'_1(B_{z^{**}}) = L'_2(z^{**})$ , is also shown.