ON INTERIOR-POINT BASED RETROSPECTIVE APPROXIMATION METHODS FOR SOLVING TWO-STAGE STOCHASTIC LINEAR PROGRAMS

Soumyadip Ghosh   Raghu Pasupathy

IBM T.J. Watson Research Center   Virginia Tech
Yorktown Heights, NY 10598, USA   Blacksburg, VA 24061, USA

ABSTRACT

In a recent paper, Gongyun Zhao introduced what appears to be the first interior point formulation for solving two-stage stochastic linear programs for finite support random variables. In this paper, we generalize Gongyun Zhao's formulation by incorporating it into a retrospective approximation framework. What results is an implementable interior-point solution paradigm that can be used to solve general two-stage stochastic linear programs. After discussing some basic properties, we characterize the complexity of the algorithm, leading to guidance on the number of samples that should be generated to construct the sub-problem linear programs, effort expended in solving the sub-problems, and the effort expended in solving the master problem.

1 INTRODUCTION

1.1 Problem Statement

In this paper, we consider problems of the form

$$\min \quad c^T x + \rho(x) \qquad \text{s.t.} \quad Ax = a, \quad x \geq 0, \qquad\qquad (M)$$

where $A \in \mathbb{R}^{c \times q_M}, a \in \mathbb{R}^c$ and $x \in \mathbb{R}^{q_M}$, and $\rho(x) = \mathrm{E}[G(x,\xi)]$ is called the recourse function. The term $G(x,\xi)$ is the optimal value of the problem

$$\min \quad d^T(\xi) y(\xi) \qquad \text{s.t.} \quad w(\xi) y(\xi) = h(\xi) - T(\xi)x, \quad y(\xi) \geq 0. \qquad (S)$$

In the above, ξ is a random vector, and $T(\xi) \in \mathbb{R}^{l \times q_M}, W(\xi) \in \mathbb{R}^{l \times q_S}$ and $y(\xi) \in \mathbb{R}^{q_S}$ are linear coefficients for each realization ξ. The random vector ξ has a distribution law over a support Ξ. No assumptions are made about the nature of the distribution or the support Ξ.

Problems in the general form $(M) - (S)$ are called two-stage stochastic linear programs, and have an extensive body of literature dedicated to them (see e.g., Birge and Louveaux (1997), Shapiro (2004), Chap. 2 in Shapiro et al. (2009) for comprehensive reviews.) This formulation typically represents decision making in two stages, where the first (or *master*) stage concerns decisions x that are to be taken under uncertainty about the eventual outcome ξ in the second stage, and the second stage (or *sub-problem*) decisions $y(\xi)$ act as recourse actions undertaken once the uncertain parameters ξ are realized. Numerous applications in diverse fields like supply-chain (Bassok et al. 1999), energy (Wallace and Fleten 2003), healthcare (Denton et al. 2010) etc. make this a key part of the stochastic optimization literature.

Monte-Carlo sampling based methods form an essential approach to solving such two-stage problems. The prototypical algorithm samples a finite set of n realizations $\{\xi_i, i = 1, \dots, n\}$ of ξ, and then solves the approximate optimization problem

$$\min \quad c^T x + \hat{\rho}(n,x) \qquad \text{s.t.} \quad Ax = a, \quad x \geq 0, \qquad\qquad (\hat{M}_n)$$

which is obtained by replacing the recourse function $\rho(x)$ with its approximation $\hat{\rho}(n,x) = \frac{1}{n}\sum_{i=1}^{n} G(x,\xi_i)$, and where $G(x,\xi_i)$ are obtained from (S). The samples ξ_i are usually i.i.d. copies of the random variable ξ, although Homem-de-Mello (2008) studies the non-i.i.d. setting.

The two-stage linear program literature concentrates on two related goals. One aim is to study the statistical properties of the approximation $(\hat{M}_n)$ to the original problem (M) with the objective of selecting an appropriate sample size. Let η^* be the optimal value for the master problem (M), and $\hat{\eta}^*(n)$ the optimal value of the sample-path problem $(\hat{M}_n)$. Then under fairly general assumptions, it is well known that $\hat{\eta}^*(n) \to \eta^*$ (c.f. King and Rockafellar 1993.) In addition Shapiro et al. (2009)[Chapter 3] and Shapiro (2004) show that a rate-of-convergence result for the optimal objective value can also be obtained under additional conditions. For example, if $P(\|\hat{\rho}(n,x) - \rho(x)\| \leq \varepsilon)$ converges exponentially fast as $n \to \infty$, then the same can be shown to hold for the optimal objective value $P(|\hat{\eta}^*(n) - \eta^*| \leq \varepsilon)$. Homem-de-Mello (2008) remarks that in general the functional CLT that holds for the estimator of the stochastic function can be translated to hold also for optimal objective value, and even for the optimal solution values under more restrictive conditions. Such rates-of-convergence results have also inspired some work on sequential algorithms towards identifying methods that return, with a guaranteed confidence, a solution of a stipulated standard (Bayraksan and Morton 2009).

The second goal is to solve the approximation $(\hat{M}_n)$ for a fixed size n in the most efficient possible manner. Zhao (2001) lays out a nice mosaic of all the different approaches taken to solve the approximation problem $(\hat{M}_n)$ efficiently. The solutions are broadly classified into *direct* methods, where the simplex method or interior point methods are used to directly solve the finite linear program $(\hat{M}_n) - (S)$ written as one large program, and *decomposition* methods, which keep the two problems $(\hat{M}_n)$ and (S) separate and pass relevant information between them iteratively. Decomposition methods are further divided into cutting-plane (CPDM) and derivative based (DDM) decomposition methods. The CPDM paradigm generates a set of cuts from (S) in order to approximate the non-linear and non-smooth $\hat{\rho}(n,x)$ in $(\hat{M}_n)$. The DDM approach evaluates the derivative (or its approximate) of $\hat{\rho}(n,x)$ at the current iterate of the master problem $(\hat{M}_n)$ using the sub-problems (S) and chooses the next iterate for the master problem using any one of the standard non-linear gradient-based descent algorithms.

Decomposition based methods enjoy the clear advantage over direct methods of scaling well with increasing sample size n. The CPDM methods suffer the handicap of not being able to change their samples as the method progresses, as this might invalidate a cut added earlier. Thus, they can only be used to solve $(\hat{M}_n)$ to optimality for a specified sample set. The DDM methods do not suffer this limitation since they use samples only to determine local information (the gradient at the current iterate), but a substantial obstacle exists in estimating the derivative of the non-linear non-smooth function $\rho(n,x)$ (refer Section 2.2 in Shapiro et al. 2009 for an explanation of this property.) Zhao (2001) suggests fixing this limitation of DDMs by solving an interior point version of the two-stage program by introducing barrier-functions for the linear constraints in both stages. Subsequent work along this line (e.g., in Chen and Mehrotra 2007, Mehrotra and Ozevin 2007) has further expanded the use to more efficient interior-point methods in a decomposition scheme which takes advantage of warm-starts to be performed in both stages.

1.2 Solution Context

Interestingly, all of the work discussed thus far appear within the realm of sample average approximation (SAA), where a single approximate problem, "generated" with a large-enough sample size, is solved to identify a solution estimator. We situate our contribution in this paper within an arguably more efficient variation of SAA called retrospective approximation (RA) (Chen and Schmeiser 1994, Pasupathy and Schmeiser 2009, Pasupathy 2010), and within the DDM class of optimization approaches. The broad strategy of RA is easily stated: generate a sequence of approximate problems with increasing sample sizes, and solve each of the generated problems using an appropriate solver to within a specified error tolerance. RA's efficiency comes from the use of "warm starts," i.e., solutions to prior problems are used as

initial guesses to subsequent problems, and from appropriate choice of the sample size and error tolerance sequences.

The RA approach can be efficiently adapted to solving two-stage programs if two conditions are met: a) an appropriate notion of accuracy is developed for any feasible master-stage candidate solution, and b) the effort required in estimating the recourse function $\rho(x)$ using a specified sample-size estimate $\hat{\rho}(n,x)$ is quantified. Usual solver implementations for linear programs are based on the simplex-method, which has few theoretical guarantees on the number of steps required to obtain a solution with a pre-specified quality guarantee. But recent work by Zhao (2001) and Chen and Mehrotra (2007) on using barrier-function based interior-point methods in this context inspires us to assume similarly that the second stage problems (S) are solved using interior point methods. Barrier methods have a guaranteed worst case solution time that scales polynomially with the (inverse of) the precision required, which is measured in terms of duality gap for linear programs. Thus, an ε-close estimate of the approximate recourse function $\hat{\rho}(n,x)$ can be obtained in computational time that is polynomial in n and $1/\varepsilon$.

The algorithm embodiment presented here assumes only that the second stage LPs (S) are solved using interior point methods to evaluate the recourse function at current iterate, and is agnostic of the method used to solve the first stage problem. The recourse function $\rho(x)$ in the master stage problem is a convex function in general, and is strictly so under certain regularity conditions (refer Section 2.2 in Shapiro et al. 2009). It is thus reasonable to make the following assumption, which will be operative for the rest of this paper:

Assumption 1 The master-stage problem (M) attains its optimal value η^* in the strict interior of the feasible set $\{Ax = a, \ x \geq 0\}$ of the master-stage variables.

So, the master-stage problem (M) is locally an unconstrained problem as the iteration count grows. The master-stage solver may utilize any Newtonian convergence scheme to realize a linear or higher reduction in a figure-of-merit, which is usually the objective function value itself, as iterations proceed. For example, the barrier-based methods of Zhao (2001) provide such a polynomial converge guarantee.

We sketch an outline of the RA paradigm for general two-stage linear programs in Section 2. In analyzing this method, we rely heavily on Pasupathy (2010) where optimal parameter choice for classical retrospective optimization is studied. In congruence with the RA approach, we assume that each iteration requires approximate solutions for problems $(\hat{M}_n)$ of a specified sample size n to within a specified accuracy. The accuracy required is specified as an ε^S for the interior point method for the second-stage problems (S) and a ε^M for the master problem solver. The following quantities are important for the analysis and bear defining upfront:

(i) η^*, as before, is the optimal value for the master problem (M).
(ii) $\hat{\eta}^*(n)$ is the optimal value of the sample-path problem $(\hat{M}_n)$.
(iii) $\hat{\eta}^*(n, \varepsilon^S)$ is the optimal value when the subproblems in $(\hat{M}_n)$ are solved to within ε^S.
(iv) $\hat{\hat{\eta}}^*(n, \varepsilon^S, \varepsilon^M)$ is the value obtained when the sample-path problem $(\hat{M}_n)$ is solved to within ε^M, and the sub-problems in $(\hat{M}_n)$ are solved to within ε^S.

Section 2.2 establishes that $\hat{\hat{\eta}}^*(n, \varepsilon^S, \varepsilon^M)$ is a consistent estimator of η^* as n grows and $\varepsilon^S, \varepsilon^M$ falls. In Section 2.3 we address how outer iterations should increase sample size n and simultaneously decrease solution guarantees ε^S and ε^M for the sub- and master-problems respectively in an efficient manner.

1.3 Summary of Contributions

The following are the main contributions of this paper.

1. We outline an interior-point based RA framework for solving two-stage stochastic LPs. Consistent with usual RA frameworks, the problem is solved iteratively. Each iteration involves approximately solving a convex programming problem that approximates the true problem.

2. We characterize the complexity of the problem in terms of the parameters used within the framework. Specifically, we derive an expression for the total computational effort in terms of the sample size (i.e., number of LPs solved when observing the convex objective function constructed during each iteration) and the extent to which the master and the subproblems are solved.

3. We derive the best achievable complexity within the framework, and the corresponding stipulations on the algorithm parameters needed to achieve it.

We emphasize that neither the proposed framework nor the methods of analysis followed in this paper are original contributions. Instead, our contribution lies in the use of the framework and the analysis to a very specific but important class of simulation optimization problems, and in the process aiding implementability.

2   INTERIOR-POINT BASED RA FOR SOLVING TWO-STAGE STOCHASTIC LPS

The proposed interior-point based RA algorithm is conceptually very simple. During the kth iteration, a problem of the type $(\hat{M}_{n_k})$ is independently generated and solved to within error-tolerance ε_k^M, thereby yielding a solution X_k. In the subsequent iteration, X_k is used as the initial solution ("warm start") to solve the problem $(\hat{M}_{n_{k+1}})$ to within error-tolerance ε_{k+1}^M. The process is then repeated until some user-specified termination criterion is satisfied. As we shall see, under certain conditions on the sample-size sequence $\{n_k\}$ and the error-tolerance sequence $\{\varepsilon_k^M\}$, the sequence of generated solutions $\{X_k\}$ converges to the correct solution wp1.

We emphasize here that the generated problem $(\hat{M}_{n_k})$ is convex, and that its solution is assumed to be identified through a "black-box" solver. This is because the objective function in $(\hat{M}_{n_k})$ is unknown until a specific point x is visited by the algorithm being used to solve $(\hat{M}_{n_k})$, at which time n_k LPs are independently generated and solved. Moreover, each of the generated LPs are solved (using an interior-point solver) only to within error-tolerance ε_k^S.

2.1 Algorithm Listing

We now formally list the proposed algorithm along with its components.

Algorithm Components:

(i)   A interior-point procedure for solving the LPs generated as sub-problems (S).
(ii) A procedure for solving convex programming problems of the type $(\hat{M}_n)$
(iii) A rule to compute the sample size sequence $\{n_k\}$.
(iv) A rule to compute the error-tolerance sequences $\{\varepsilon_k^S\}, \{\varepsilon_k^M\}$.

Algorithm Logic:

0.   Initialize the iteration number $k = 1$. Set $n_1, \varepsilon_1^S, \varepsilon_1^M$.
1.   Use component (ii) to solve the sample-path problem $(\hat{M}_{n_k})$ to within error-tolerance ε_k^M. (At each x, n_k LPs of the type (S) are generated independently, and solved to within error-tolerance ε_k^S using component (i).) Obtain a retrospective solution X_k.
5.   Use components (iii) and (iv) to obtain $n_{k+1}, \varepsilon_{k+1}^S$, and ε_{k+1}^M. Set $k \leftarrow k+1$ and go to Step 1.

Two aspects are crucial to the efficient functioning of the proposed algorithm. First, solutions from the previous iterations ought to be used as "warm starts" solutions to the subsequently generated problems, i.e., X_k should be the initial guess to the problem $(\hat{M}_{n_{k+1}})$. As the sample sizes become larger, this will usually yield efficiency gains because subsequent problems will "look similar" to previous problems and thereby have solutions in close proximity. Second, the sequences $\{\varepsilon_k^M\}$ and $\{\varepsilon_k^S\}$ should be chosen carefully. Choosing them too small will result in too much work per iteration compared to the inherent sampling

variability in the sample-path problem ($\hat{M}_{n_k}$). Choosing them too large will result in a large bias in the obtained solutions $\{X_k\}$. We discuss this issue rigorously and provide a recommendation in Section 2.3.

2.2 Convergence

In this section, we present some basic results on convergence and rate of convergence of the (true) objective function values of the solutions generated by the proposed paradigm. Recall that through the framework just proposed, a sequence of sample-path problems $\{\hat{M}_{n_k}\}$ are solved to respective tolerances $\{\varepsilon^M\}$, to obtain a sequence of solutions along with a sequence of corresponding optimal values $\{\hat{\hat{\eta}}^*(n, \varepsilon^S, \varepsilon^M)\}$.

Towards preparing for the main results, we first state a result that provides sufficient conditions that guarantee that the sequence of recourse-function approximations $\{\hat{\rho}_{n_k}(x)\}$ converges to the recourse function $\hat{\rho}(x)$ uniformly (in x) wp1.

Proposition 1 Let $\partial F(x, \xi)$ denote the sub-differential of $F(x, \xi)$ (with respect to x). Suppose there exists a constant $b < \infty$ satisfying $b > \sup_{x \in \mathscr{D}} b(x)$, where

$$b(x) = \sup_{y \in \partial F(x, \xi)} \{\|y\|\} \quad wp1.$$

Let the sequence of positive constants $\{n_k\} \to \infty$ as $k \to \infty$. Then $\hat{\rho}(n_k, x) \to \rho(x)$ as $k \to \infty$ uniformly (in x) wp1.

Proof Sketch. If uniform convergence does not hold, there exists a convergent sequence $\{x_k\}$ and $\varepsilon > 0$ such that $\{x_k\} \to x$ and $|\hat{\rho}(n_k, x_k) - \rho(x_k)| > \varepsilon$ with positive probability. However,

$$|\hat{\rho}(n_k, x_k) - \rho(x_k)| \leq |\hat{\rho}(n_k, x_k) - \hat{\rho}(n_k, x)| + |\hat{\rho}(n_k, x) - \rho(x)| + |\rho(x) - \rho(x_k)|. \tag{1}$$

The first term on the right-hand side of (1) satisfies $|\hat{\rho}(n_k, x_k) - \hat{\rho}(n_k, x)| \leq b\|x_k - x\|$. The second term is small by pointwise convergence and the third term is small due to the continuity of $\rho(\cdot)$. $\qquad\square$

Proposition 1 essentially states that as long as the supremum-norm of the sub-differentials of the sample-paths $F(x, \xi)$ are bounded, the approximations to the recourse function converge in a uniform sense. Such convergence is very useful to demonstrate that the sequence of objective function values $\{\hat{\hat{\eta}}^*(n, \varepsilon^S, \varepsilon^M)\}$ of the solutions returned by the procedure converges to the true optimal value η^* wp1. We state this formally next, along with a sketch of the proof.

Proposition 2 Suppose $\{\varepsilon_k^M\}$ and $\{\varepsilon_k^S\}$ are positive-valued sequences satisfying $\{\varepsilon_k^M\} \to 0$ and $\{\varepsilon_k^S\} \to 0$ as $k \to \infty$ wp1. Also, suppose $\hat{\rho}(n_k, x) \to \rho(x)$ uniformly (in x) wp1 as $k \to \infty$. Then

$$|\hat{\hat{\eta}}^*(n, \varepsilon^S, \varepsilon^M) - \eta^*| \to 0 \text{ wp1 as } k \to \infty.$$

Proof Sketch. Since $\hat{\rho}(n_k, x) \to \rho(x)$ uniformly (in x) wp1, it can be shown that $\eta^*(n_k) \to \eta^*$ as $k \to \infty$ wp1. Now use

$$\begin{aligned}|\hat{\hat{\eta}}^*(n, \varepsilon^S, \varepsilon^M) - \eta^*| &\leq |\hat{\hat{\eta}}^*(n, \varepsilon^S, \varepsilon^M) - \hat{\eta}(n_k, \varepsilon_k^S)| + |\hat{\eta}(n_k, \varepsilon_k^S - \eta^*(n_k)| + |\eta^*(n_k) - \eta^*| \\ &\leq \varepsilon_k^M + \varepsilon_k^S + |\eta^*(n_k) - \eta^*| \end{aligned} \tag{2}$$

to see that the assertion in the proposition holds. $\qquad\square$

Proposition 2 states that as long as the sequence of tolerances for the master and sub-problems are chosen to converge to zero, and the recourse-function approximations converge to the recourse function uniformly, the (true) objective function values will converge to the optimal value. The rate at which such convergence happens is given, as usual, through a central limit theorem, as stated next without proof.

Proposition 3 Suppose $\{\varepsilon_k^M\}$ and $\{\varepsilon_k^S\}$ are positive-valued sequences satisfying $\varepsilon_k^M = o(1/\sqrt{n_k})$ and $\varepsilon_k^S = o(1/\sqrt{n_k})$ as $k \to \infty$. (Given two sequences $\{u_n\}$ and $\{v_n\}$, we say that $u_n = o(v_n)$ if $u_n/v_n \to 0$ as $n \to \infty$ (Serfling 1980)[pp. 8].) Suppose there exists a constant $b < \infty$ satisfying $b > \sup_{x \in \mathscr{D}} b(x)$, where $b(x) = \sup_{y \in \partial F(x,\xi)}\{\|y\|\}$ wp1. Also suppose that $\sup_x \mathrm{E}[(F(x,\xi))^2] < \infty$. Then

$$\sqrt{n_k}(\hat{\eta}^*(n,\varepsilon^S,\varepsilon^M) - \eta^*) \xrightarrow{d} \mathscr{N}(0,\sigma^2(x^*)),$$

where x^* is the unique solution of the problem (M), $\mathscr{N}(\mu,\sigma^2)$ refers to a normal random variables with mean μ and variance σ^2, and $\sigma^2(x^*) = \mathrm{Var}(F(x^*,\xi))$.

2.3 Complexity Analysis and Optimal Parameter Characterization

In this section, we first characterize the complexity of the proposed method as a function of the error-tolerance sequences $\{\varepsilon_k^S\}$, $\{\varepsilon_k^M\}$, and the sequence of sample sizes $\{n_k\}$. This work characterization is then used to provide guidance on the relative rates at which the error-tolerances should tend to zero while the sample-size sequence tends to infinity. Towards facilitating exposition, we introduce the following additional notation:

– $w^S(\varepsilon)$: the number of algorithmic steps involved in solving each sub-problem to within tolerance $\varepsilon > 0$, i.e., until an ε-optimal solution is obtained;
– $w^M(x,\varepsilon)$: the number of steps involved in solving each master problem to within tolerance $\varepsilon > 0$, starting from the initial solution x;
– β: a constant that captures the relative computational cost of the steps in the master and the sub-problems.

(Notice that the notation $w^S(\varepsilon)$ implies independence of the chosen initial point when solving the sub-problems. This is partially motivated by the fact that when implementing the proposed framework using a solver, an initial point is chosen automatically and with no prior information. Also, the the notation $w^S(\varepsilon)$ and $w^M(x,\varepsilon)$ implicitly assumes that, when all other conditions are the same, the number of steps involved in solving any realization is the same. This assumption, while somewhat stringent, can be relaxed by assuming instead that $w^S(\varepsilon)$ and $w^M(x,\varepsilon)$ are upper bounds on the actual number of steps expended by the solver in solving the sub-problem and the master problem respectively.)

Given the above notation, the total amount of work involved during the ith iteration of the proposed method is given as

$$W_i = (\beta + n_i w_i^S(\varepsilon_{n_i}^S))w^M(X_{i-1},\varepsilon_i^M). \tag{3}$$

For a barrier method that uses Newton steps to perform centering on the sub-problems, the number of steps $w^S(\varepsilon_i)$ can be shown to be

$$
\begin{aligned}
w^S(\varepsilon_i) &= \left\lceil \frac{\log(q^S/\varepsilon_i t^0)}{\log v} \right\rceil \\
&\approx \frac{\log(q^S/t^0) + \log(1/\varepsilon_i^S)}{\log v},
\end{aligned} \tag{4}
$$

where q^S is the number of constraints in the sub-problem, t^0 is the initial penalty, and v is the multiplicative constant on the penalty (Boyd 2004)[Chapter 11]. Using (3) and (4), we see that the total work expended after k iterations of the proposed method is

$$W_k \approx \sum_{i=1}^{k}\left(\beta + n_i \frac{\log(q^S/t^0)}{\log v} + \frac{\log(1/\varepsilon_i^S)}{\log v}\right)w^M(X_{i-1},\varepsilon_i^M). \tag{5}$$

We now use the expression developed in (5) to guide choices of the error-tolerance sequences $\{\varepsilon_i^S\}$, $\{\varepsilon_i^M\}$ relative to the sequence of sample sizes $\{n_i\}$. As a measure of efficiency of a particular choice of the sequences, we use the product of the squared error of the optimality gap of the solution returned by the proposed method and the total computational effort expended in arriving at the returned solution. Specifically, given that W_k, expressed through (5), is the total work expended at the end of the kth iteration, and $\|\hat{\eta}^*(n, \varepsilon^S, \varepsilon^M) - \eta^*\|$ is the resulting error, we analyze the random variable sequence $\{T_k\}$ given by

$$
\begin{aligned}
T_k &= W_k \|\hat{\eta}^*(n, \varepsilon^S, \varepsilon^M) - \eta^*\|^2 \\
&= \|\hat{\eta}^*(n, \varepsilon^S, \varepsilon^M) - \eta^*\|^2 \sum_{i=1}^k \left(\beta + n_i \frac{\log(q^S/t^0)}{\log v} + \frac{\log(1/\varepsilon_i^S)}{\log v} \right) w^M(X_{i-1}, \varepsilon_i^M).
\end{aligned}
\tag{6}
$$

(T_k is a random variable with the interpretation of "work $\times$ squared error.") If we assume that the solver that is being used to solve the master problem converges linearly, we can write the following "big Oh" expression for $w^M(X_{i-1}, \varepsilon_i^M)$, after exploiting a central limit theorem on the sequence of solutions $\{X_k\}$.

$$
w^M(X_{i-1}, \varepsilon_i^M) = O_p \left(1 + \frac{1}{\log s} \log \frac{\varepsilon_i^M}{1/\sqrt{n_{i-1}} + 1/\sqrt{n_i} + \varepsilon_{i-1}^M} \right),
\tag{7}
$$

where $s > 0$ is a constant that is related to the solver that is in use.

We now use (6) and (7) to make a statement on the behavior of the sequence of random variables $\{T_k\}$ through Proposition 4.

Proposition 4 Suppose $\{\varepsilon_i^S\}$ and $\{\varepsilon_i^M\}$ are $O_p(n_i^{-q})$ and $O_p(n_i^{-r})$ respectively, where $q, r > 0$.

(a)   If $q < 1/2$ or $r < 1/2$, then $T_k / \log n_k \xrightarrow{p} \infty$.
(b)   If $q \geq 1/2$ and $r \geq 1/2$, then $T_k / \log n_k = O_p(1)$.

Proof Sketch of (a).   We see that if $q < 1/2$, then

$$
w^M(X_{i-1}, \varepsilon_i^M) = O_p \left(1 + \frac{1}{\log s} \log \frac{\varepsilon_i^M}{1/\sqrt{n_{i-1}} + 1/\sqrt{n_i} + \varepsilon_{i-1}^M} \right) = O_p(1).
\tag{8}
$$

Also, if $q < 1/2$ or $r < 1/2$,

$$
\begin{aligned}
n_i \|\hat{\eta}^*(n, \varepsilon^S, \varepsilon^M) - \eta^*\|^2 &= n_i O_p \left((1/\sqrt{n_i} + \varepsilon_i^M + \varepsilon_i^S)^2 \right) \\
&\xrightarrow{p} \infty.
\end{aligned}
\tag{9}
$$

Also,

$$
T_k / \log n_k \geq \frac{1}{v} \frac{\log(1/\varepsilon_k^S)}{\log(n_k)} n_k \|\hat{\eta}^*(n, \varepsilon^S, \varepsilon^M) - \eta^*\|^2 w^M(X_{k-1}, \varepsilon_i^M).
\tag{10}
$$

Now use (8), (9), and (10) to conclude that the assertion in (a) holds.   □

3   CONCLUDING REMARKS

The recent paper Zhao (2001) incorporates interior-point methods within a sample-average approximation framework for solving two-stage stochastic LPs. We believe that this is an important step in the direction of reliably solving large-scale two-stage stochastic LPs — an important class of simulation optimization problems. In this paper, we have attempted to extend Zhao's framework by incorporating it into a refinement of sample-average approximation called retrospective approximation. In the resulting framework, a sequence of convex approximations to the original two-stage stochastic LP are generated and solved using a user-specified interior point method.

The strength of the proposed framework is that it is implementable. This is primarily owing to its generality (the framework does not assume that the support of the underlying random variables is finite), its use of a progressive sequence of approximate problems (instead of using a single approximation involving a potentially enormous sample size), and its explicit incorporation of interior-point methods, which are amongst the favored techniques for solving large-scale convex programs. We emphasize that neither the individual elements in the proposed paradigm nor the type of analysis performed in the paper are new — they have appeared in various forms and in more generality in some of the references that have been cited. However, "putting together" these individual methodological elements within the important specific context of two-stage stochastic has seemed particularly worthwhile to us.

ACKNOWLEDGMENTS

REFERENCES

Bassok, Y., R. Anupindi, and R. Akella. 1999. "Single-Period Multiproduct Inventory Models with Substitution". *Operations Research* 47:632–642.

Bayraksan, G., and D. P. Morton. 2009. "A Sequential Sampling Procedure for Stochastic Programming". *Operations Research* 59 (4): 898–913.

Birge, J. R., and F. Louveaux. 1997. *Introduction to Stochastic Programming*. New York, NY.: Springer.

Boyd, S. 2004. *Convex Optimization*. Cambridge, U.K.: Cambridge University Press.

Chen, H., and B. W. Schmeiser. 1994, December. "Retrospective approximation algorithms for stochastic root finding". In *Proceedings of the 1994 Winter Simulation Conference*, edited by J. D. Tew, S. Manivannan, D. A. Sadowski, and A. F. Seila, WSC '94, 255–261. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Chen, M., and S. Mehrotra. 2007. "Self-concordant Tree and Decomposition Based Interior Point Methods for Stochastic Convex Optimization Problems". *Stochastic Programming E-Print Series* 2007:–.

Denton, B. T., A. J. Miller, H. J. Balasubramanian, and T. R. Huschka. 2010. "Optimal Allocation of Surgery Blocks to Operating Rooms Under Uncertainty". *Operations Research* 58:802–816.

Homem-de-Mello, T. 2008. "On rates of convergence for stochastic optimization problems under non-i.i.d. sampling". *SIAM Journal on Optimization* 19 (2): 524–551.

King, A. J., and R. T. Rockafellar. 1993. "Asymptotic Theory for Solutions in Statistical Estimation and Stochastic Programming". *Mathematics of Operations Research* 18:148–162.

Mehrotra, S., and M. G. Ozevin. 2007. "Convergence Analysis of a Weighted Barrier Decomposition Algorithm for Two Stage Stochastic Programming". Technical report, Dept of IEMS, Northwestern University.

Pasupathy, R. 2010. "On choosing parameters in retrospective-approximation algorithms for stochastic root finding and simulation optimization". *Operations Research* 58:889–901.

Pasupathy, R., and B. W. Schmeiser. 2009, March. "Retrospective-approximation algorithms for the multidimensional stochastic root-finding problem". *ACM Trans. Model. Comput. Simul.* 19:5:1–5:36.

Serfling, R. J. 1980. *Approximation Theorems of Mathematical Statistics*. New York, NY.: John Wiley & Sons, Inc.

Shapiro, A. 2004. "Monte Carlo sampling methods". In *Stochastic Programming*, edited by A. Ruszczynski and Shapiro, Handbooks in Operations Research and Management Science, 353–426. Elsevier.

Shapiro, A., D. Dentcheva, and A. Ruszczynski. 2009. *Lectures on Stochastic Programming: Modeling and Theory*. Philadelphia, PA: MPSSIAM Series on Optimization.

Wallace, S. W., and S.-E. Fleten. 2003. "Stochastic Programming Models in Energy". In *Stochastic Programming*, edited by A. Ruszczynski and A. Shapiro, Volume 10 of *Handbooks in Operations Research and Management Science*, 637 – 677. Elsevier.

Zhao, G. 2001. "A Log-Barrier method with Benders decomposition for solving two-stage stochastic linear programs". *Mathematical Programming Series A* 90:507–536.

AUTHOR BIOGRAPHIES

RAGHU PASUPATHY is an assistant professor in the Industrial and Systems Engineering Department at Virginia Tech. His research interests lie broadly in Monte Carlo methods with a specific focus on simulation optimization and stochastic root finding. He is a member of INFORMS, IIE, and ASA, and serves as an Associate Editor for ACM TOMACS and INFORMS Journal on Computing. His e-mail address is pasupath@vt.edu and his web page is https://filebox.vt.edu/users/pasupath/pasupath.htm.

SOUMYADIP GHOSH is a Research Staff Member of the Mathematical Sciences Division at IBM T.J. Watson Research Center, Yorktown Heights, NY. His simulation research interests include theory and practice in particular on simulation-based optimization techniques and input dependence modeling and risk modeling. His other interests lie in the fields of smarter energy management, supply-chain analysis, and queueing theory based scheduling of large-scale production systems. He can be contacted at ghoshs@us.ibm.com and his web-page is https://researcher.ibm.com/researcher/view.php?person=us-ghoshs.