Stochastic Combinatorial Optimization with Controllable Risk Aversion Level

Anthony Man–Cho So
Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, N. T., Hong Kong
email: manchoso@se.cuhk.edu.hk

Jiawei Zhang
Department of Information, Operations, and Management Sciences, Stern School of Business, New York University, New York, NY 10012, USA
email: jzhang@stern.nyu.edu

Yinyu Ye
Department of Management Science and Engineering and, by courtesy, Electrical Engineering, Stanford University, Stanford, CA 94305, USA
email: yinyu-ye@stanford.edu

Most of the recent work on 2–stage stochastic combinatorial optimization problems have focused on the minimization of the expected cost or the worst–case cost of the solution. Those two objectives can be viewed as two extreme ways of handling risk. In this paper we propose to use an one–parameter family of functionals to interpolate between them. Although such a family has been used in the mathematical finance and stochastic programming literature before, its use in the context of approximation algorithms seems new. We show that under standard assumptions, a broad class of risk–adjusted 2–stage stochastic programs can be efficiently treated by the Sample Average Approximation (SAA) method. In particular, our result shows that it is computationally feasible to incorporate some degree of robustness even when the underlying distribution can only be accessed in a black–box fashion. We also show that when combined with suitable rounding procedures, our result yields new approximation algorithms for many risk–adjusted 2–stage stochastic combinatorial optimization problems under the black–box setting.

Key words: Stochastic Optimization; Risk Measure; Sample Average Approximation (SAA); Approximation Algorithm; Combinatorial Optimization

MSC2000 Subject Classification: Primary: 90C15, 68Q25; Secondary: 90C27, 68W25

OR/MS subject classification: Primary: Programming, Stochastic; Secondary: Analysis of Algorithms

---

1. Introduction A fundamental challenge that faces all decision–makers is the need to cope with an uncertain environment while trying to achieve some predetermined objectives. One certainly does not need to go far to encounter such situations — for example, an office clerk trying to get to work as fast as possible while avoiding possibly congested roads; a customer in the supermarket trying to checkout while avoiding lines that may take a long time, and so on. From a decision–maker's perspective, it is then natural to ask whether one can determine the optimal decision given one's assessment of the uncertain environment. This is a motivating question in the field of stochastic optimization. To keep our discussion focused, we shall consider the class of 2–stage stochastic programs with recourse [1, 5], particularly those that arise in the context of combinatorial optimization problems. Roughly speaking, in the 2–stage recourse model, one commits *irrevocably* to some initial (i.e. first stage) action x based on one's knowledge of the underlying probability distribution. The actions in the second stage cannot be determined in advance, since they depend on the actions of the first stage as well as the uncertain parameters of the problem. However, once those parameters are realized (according to the distribution), a recourse (i.e. second stage) action r can be taken so that, together with the first–stage actions, all the requirements of the problem are satisfied. Naturally, one would seek for the action (x, r) that minimizes the "total cost". However, since the outcome is random, such an objective can have many possible interpretations. In this paper we shall consider the problem of risk minimization. Specifically, let X be the set of permissible actions, and let $(\Omega, \mathscr{B}, \mathbb{P})$ be the underlying probability space. In accordance with the convention in the literature, we assume that the probability distribution is specified via one of the following models:

(a) **Scenario Model**: The set of scenarios $\mathcal{S}$ and their associated probabilities are explicitly given. Hence, under this model, a "polynomial–time" algorithm is allowed to take time polynomial in $|\mathcal{S}|$, the number of scenarios.

(b) **Black–Box Model**: The distribution of the scenarios is given as a black box. An algorithm can use this black box to draw independent samples from the distribution of scenarios.

Note that the scenario model is much more restrictive than the black–box model. For instance, the scenario model is applicable only when the underlying probability distribution is discrete. Now, we are interested in solving problems of the form:

$$\min_{x \in X} \{g(x) \equiv c(x) + \Phi(q(x, \omega))\} \tag{1}$$

where $c : X \to \mathbb{R}_+$ is a (deterministic) cost function, $q : X \times \Omega \to \mathbb{R}_+$ is another cost function that depends both on the decision $x \in X$ and some uncertain parameter $\omega \in \Omega$, and $\Phi : L^2(\Omega, \mathscr{B}, \mathbb{P}) \to \mathbb{R}$ is some *risk measure*. We shall refer to Problem (1) as a *risk–adjusted 2–stage stochastic program with recourse*. Two typical examples of Φ are the *expectation* operator and the *max* operator. The former gives rise to a risk–neutral objective, while the latter gives rise to an extremely risk–averse objective. Both of these risk measures have been studied in recent work on approximation algorithms for stochastic combinatorial optimization problems (see, e.g., [7, 18, 13, 10, 11, 23, 3, 6, 8]). For the case where Φ is the expectation operator, it turns out that under the black–box model, one can obtain a near–optimal solution to Problem (1) with high probability by the so–called Sample Average Approximation (SAA) method [14]. Roughly speaking, the SAA method works as follows. Let $\omega^1, \ldots, \omega^N$ be N i.i.d. samples drawn from the underlying distribution, and consider the sampled problem:

$$\min_{x \in X} \frac{1}{N} \sum_{i=1}^{N} \left(c(x) + q(x, \omega^i) \right) \tag{2}$$

Under some mild assumptions, it has been shown [14, 21] that the optimal value of (2) is a good approximation to that of (1) with high probability, and that the number of samples N can be bounded. Unfortunately, the bound on N depends on the maximum variance V (over all $x \in X$) of the random variables $q(x, \omega)$, which need not be polynomially bounded by the input size. However, in a recent breakthrough, Shmoys and Swamy [23] have been able to circumvent this problem for a large class of 2–stage stochastic *linear* programs. Specifically, by bounding the relative factor by which the second–stage actions are more expensive than the first–stage actions by a parameter λ (called the *inflation factor*), they are able to show that an adaptation of the ellipsoid method will yield an $(1 + \epsilon)$–approximation with the number of samples (i.e. black–box accesses) bounded by a polynomial of the input size, λ and $1/\epsilon$. Subsequently, Charikar et al. [3] have established a similar but more general result using the SAA method. We should mention, however, that both of these results assume that the objective function is linear. Thus, in general, they do not apply to Problem (1).

On another front, motivated by robustness concerns, Dhamdhere et al. [6] have considered the case where Φ is the max operator and developed approximation algorithms for various 2–stage stochastic combinatorial optimization problems with recourse under that setting. Unfortunately, their techniques only work under the more restrictive scenario model[1]. Recently, Feige et al. [8] have shown that such a difficulty can be partially circumvented in the case of covering problems. Specifically, they developed approximation algorithms for a class of 2–stage robust covering problems where the list of scenarios is given implicitly by an upper bound on the number of *active* elements (i.e. those elements that need to be covered) in any scenario. It should be noted, however, that the above robust optimization framework has some limitations. First, it is clear that the approaches used by Dhamdhere et al. [6] and Feige et al. [8] for specifying the underlying probability distribution are less general than the black–box model. Secondly, since the worst–case scenario may occur with an exponentially small probability, it seems unlikely that sampling techniques can be applied to efficiently solve the aforementioned robust optimization problems. Finally, the extreme risk–averse nature of the robust optimization framework may lead to overly conservative solutions. This may not be desirable in some applications.

From the above discussion, a natural question arises whether we can incorporate a certain degree of robustness (possibly with some other risk measures Φ) in the problem while still being able to solve it in polynomial time under the black–box model. If so, can we also develop approximation algorithms for some well–studied combinatorial optimization problems under the new robust setting?

Our Contribution. In this paper we answer both of the above questions in the affirmative. Using techniques from the mathematical finance literature [19, 2], we provide a unified framework for treating

[1] In fact, since only the worst case matters, it is not even necessary to specify any probabilities in their framework.

the aforementioned risk–adjusted stochastic optimization problems. Specifically, we use an one–parameter family of functionals $\{\varphi_\alpha\}_{0\le\alpha<1}$ to capture the degree of risk aversion, and we consider the problem:

$$\min_{x\in X}\{c(x)+\varphi_\alpha(q(x,\omega))\}$$

As we shall see, such a family arises naturally from a change of the underlying probability measure $\mathbb{P}$ and possesses many nice properties. In particular, it includes $\Phi=\mathbb{E}$ as a special case and $\Phi=\max$ as a limiting case. Thus, our framework provides a generalization of previous work. Moreover, our framework works under the most general black–box model, and we show that as long as one does not insist on considering the worst–case scenario, one can use sampling techniques to obtain near–optimal solutions to the problems discussed above efficiently. Our sampling theorem and its analysis can be viewed as a generalization of those by Charikar et al. [3]. Consequently, our result extends the class of problems that can be efficiently treated by the SAA method. Finally, by combining with techniques developed in earlier work [18, 10, 23, 6], we obtain the first approximation algorithms for a large class of 2–stage stochastic combinatorial optimization problems under the risk–adjusted setting.

The rest of the paper is organized as follows. In Section 2 we give some motivation for the risk measure we use and introduce a powerful representation theorem due to Rockafellar and Uryasev [19] (see also Ben–Tal and Teboulle [2]) concerning that measure. In Section 3 we prove the main result of this paper, namely a sampling theorem for a broad class of risk–adjusted 2–stage stochastic programs with recourse. We then use the sampling theorem to design approximation algorithms for several risk–adjusted stochastic combinatorial optimization problems in Section 4. Finally, we close with some concluding remarks and future directions in Section 5.

2. Motivation: Risk Aversion as Change of Probability Measure We begin with the setup and some notation. Let $(\Omega,\mathscr{B},\mathbb{P})$ be a probability space, and let $L^2(\Omega,\mathscr{B},\mathbb{P})$ be the Hilbert space of square–integrable random variables with inner product $\langle\cdot,\cdot\rangle$ given by $\langle U,V\rangle=\int_\Omega UV\,d\mathbb{P}$. We assume that the second–stage cost function q satisfies the following:

(a) $q(x,\cdot)$ is measurable w.r.t. $\mathscr{B}$ for each $x\in X$;
(b) q is continuous w.r.t. x; and
(c) $\mathbb{E}[q(x,\omega)]<\infty$ for each $x\in X$.

To motivate our approach, let us investigate how the following problems capture risk:

$$\min_{x\in X}\{c(x)+\mathbb{E}[q(x,\omega)]\} \tag{3}$$

$$\min_{x\in X}\left\{c(x)+\sup_{\omega\in\Omega}q(x,\omega)\right\} \tag{4}$$

Problem (3) is a standard stochastic optimization problem, in which a first–stage decision $x^*\in X$ is sought so that the sum of the first–stage cost $c(x^*)$ and the expected second–stage cost $\mathbb{E}[q(x^*,\omega)]$ is minimized. In particular, we do not consider any single scenario as particularly important, and hence we simply weigh them by their respective probabilities. On the other hand, Problem (4) is a pessimist's version of the problem, in which one considers the worst–case second–stage cost over all scenarios. Thus, for each $x\in X$, we consider the scenario ω_x that gives the maximum second–stage cost as most important, and we put a weight of 1 on ω_x and 0 on all $\omega\neq\omega_x$, regardless of what their respective probabilities are. These observations suggest the following approach for capturing risk. For each $x\in X$, let $f_x:\Omega\to\mathbb{R}_+$ be a measurable weighing function such that:

$$\int_\Omega f_x(\omega)\,d\mathbb{P}(\omega)=1$$

Now, consider the problem:

$$\min_{x\in X}\{c(x)+\mathbb{E}[f_x(\omega)q(x,\omega)]\} \tag{5}$$

Observe that Problem (5) captures both Problems (3) and (4) as special cases. Indeed, if we set $f_x\equiv 1$, then we recover Problem (3). On the other hand, suppose that Ω is finite, with $\mathbb{P}(\omega)>0$ for all $\omega\in\Omega$. Consider a fixed $x\in X$, and let $\omega'=\arg\max_{\omega\in\Omega}q(x,\omega)$. Then, by setting $f_x(\omega')=\frac{1}{\mathbb{P}(\omega')}$ and $f_x(\omega)=0$ for all $\omega\neq\omega'$, we recover Problem (4).

From the above discussion, we see that one way of addressing risk is by changing the underlying probability measure $\mathbb{P}$ using a weighing function. Indeed, the new probability measure is given by:

$$\mathbb{Q}_x(\omega) \equiv f_x(\omega)\mathbb{P}(\omega) \tag{6}$$

and we may write $\mathbb{E}_{\mathbb{P}}[f_x(\omega)q(x,\omega)] = \mathbb{E}_{\mathbb{Q}_x}[q(x,\omega)]$. Alternatively, we can specify the probability measure $\mathbb{Q}_x$ directly without using weighing functions. As long as the new measure $\mathbb{Q}_x$ is absolutely continuous w.r.t. $\mathbb{P}$ for each $x \in X$ (i.e. $\mathbb{P}(\omega) = 0$ implies that $\mathbb{Q}_x(\omega) = 0$), there will be a corresponding weighing function f_x given precisely by (6). Thus, in this context, we see that f_x is simply the Radon–Nikodym derivative of $\mathbb{Q}_x$ w.r.t. $\mathbb{P}$.

Note that in the above formulation, we are allowed to choose a different weighing function f_x for each $x \in X$. Clearly, there are many possible choices for f_x. However, our goal is to choose the f_x's so that Problem (5) is computationally tractable. Towards that end, let us consider the following strategy. Let $\alpha \in [0,1)$ be a given parameter (the *risk aversion level*), and define:

$$\mathcal{Q} = \left\{ f \in L^2(\Omega, \mathscr{B}, \mathbb{P}) : 0 \le f(\omega) \le \frac{1}{1-\alpha} \text{ for all } \omega \in \Omega,\ \langle f, 1 \rangle = 1 \right\}$$

For each $x \in X$, we take f_x to be the optimal solution to the following optimization problem:

$$f_x = \arg\max_{f \in \mathcal{Q}} \mathbb{E}_{\mathbb{P}}[f(\omega)q(x,\omega)] \tag{7}$$

Note that such a f_x always exists (i.e. the maximum is always attained), since the functional $f \mapsto \langle f, q(x,\cdot)\rangle$ is continuous, and the set $\mathcal{Q}$ is compact (in the weak*–topology) by the Banach–Alaoglu theorem (see p. 120 of [15]). Intuitively, the function f_x boosts the weights of those scenarios ω that have high second–stage costs $q(x,\omega)$ by a factor of at most $(1-\alpha)^{-1}$, and zeroes out the weights of those scenarios that have low second–stage costs. Note also that when $\alpha = 0$, we have $f_x \equiv 1$; and as $\alpha \nearrow 1$, f_x tends to a delta function at the scenario ω that has the highest cost $q(x,\omega)$. Thus, the definition of f_x in (7) captures the intuitive notion of risk as discussed earlier. Now, we define φ_α by:

$$\varphi_\alpha(q(x,\omega)) \equiv \mathbb{E}_{\mathbb{P}}[f_x(\omega)q(x,\omega)]$$

where f_x is given by (7).

At this point, it may seem that we need to perform the non–trivial task of computing f_x for many $x \in X$. However, it turns out that this can be circumvented by the following representation theorem of Rockafellar and Uryasev [19] (see [2] for an interesting historical account of this result). Such a theorem forms the basis of our sampling approach.

Theorem 2.1 *(Rockafellar and Uryasev [19], Ben–Tal and Teboulle [2]) Let $\alpha \in (0,1)$, and for $x \in X$ and $\beta \in \mathbb{R}$, define:*

$$F_\alpha(x,\beta) = \beta + \frac{1}{1-\alpha}\mathbb{E}_{\mathbb{P}}\left[(q(x,\omega)-\beta)^+\right]$$

Then, $F_\alpha(x,\cdot)$ is finite and convex, with $\varphi_\alpha(q(x,\omega)) = \min_\beta F_\alpha(x,\beta)$. In particular, if q is convex w.r.t. x, then φ_α is convex w.r.t. x as well. Indeed, F_α is jointly convex in (x,β).

The power of the above representation theorem lies in the fact that it reduces the risk–adjusted stochastic optimization problem:

$$\min_{x \in X} \{c(x) + \varphi_\alpha(q(x,\omega))\} \tag{8}$$

to the well–studied problem of minimizing the expectation of a certain random function. Thus, it seems plausible that the machinery developed for solving the latter can be applied to Problem (8) as well. Moreover, when c, q are convex w.r.t. x and X is convex, both Problem (8) and its Sample Average Approximation (SAA) are convex optimization problems. This opens up the possibility of using powerful convex optimization techniques [9, 17] to design efficient algorithms for such problems. Before we discuss the algorithmic aspects of Problem (8) and its SAA, however, let us first establish the approximation quality of the SAA approach.

3. Sampling Theorem for Risk–Adjusted Stochastic Optimization Problems In this section we show that for any fixed $\alpha \in [0,1)$, it suffices to have only a polynomial number of samples in order for the Sample Average Approximation (SAA) method [14] to yield a near–optimal solution to Problem (8). Our result and analysis generalize those in [3]. To begin, let X be a finite set, and let us assume that the functions $c : X \to \mathbb{R}$ and $q : X \times \Omega \to \mathbb{R}$ satisfy the following properties:

(a) **(Non–Negativity)** The functions c and q are non–negative for every first–stage action $x \in X$ and every scenario $\omega \in \Omega$.

(b) **(Empty First Stage)** There exists a first–stage action $\phi \in X$ such that $c(\phi) = 0$ and $q(x,\omega) \le q(\phi,\omega)$ for every $x \in X$ and $\omega \in \Omega$.

(c) **(Bounded Inflation Factor)** There exists a $\lambda \ge 1$ such that $q(\phi,\omega) - q(x,\omega) \le \lambda c(x)$ for every $x \in X$ and $\omega \in \Omega$.

We remark that the assumptions above are the same as those in [3] and capture those considered in recent work (see, e.g., [18, 13, 10, 23]). Now, let $g_\alpha(x) = c(x) + \varphi_\alpha(q(x,\omega))$. By Theorem 2.1, we have:

$$\min_{x\in X} g_\alpha(x) = \min_{(x,\beta)\in X\times\mathbb{R}} g'_\alpha(x,\beta)$$

where:

$$g'_\alpha(x,\beta) \equiv c(x) + \mathbb{E}_{\mathbb{P}}\left[q'(x,\beta,\omega)\right] \qquad \text{and} \qquad q'(x,\beta,\omega) \equiv \beta + \frac{1}{1-\alpha}(q(x,\omega)-\beta)^+$$

Let $(x^*,\beta^*) \in X \times [0,\infty)$ be an exact minimizer of g'_α, and set $Z^* = g'_\alpha(x^*,\beta^*)$. It is easy to show that $\beta^* \in [0, Z^*]$. Indeed, observe that:

$$q'(x,\beta,\omega) = \begin{cases} \beta & \text{if } q(x,\omega) \le \beta \\ \dfrac{q(x,\omega)-\alpha\beta}{1-\alpha} & \text{otherwise} \end{cases} \tag{9}$$

Thus, if $\beta > Z^*$, then we have $q'(x,\beta,\omega) > Z^*$ for all $x \in X$ and $\omega \in \Omega$.

We first establish the following lemma.

Lemma 3.1 *Let $\alpha \in [0,1)$, and let c, q and q' be as above.*

(a) Let $\kappa \ge 1$ be fixed. For any $x \in X$, $\omega \in \Omega$ and $\beta \in [0, \kappa Z^*]$, we have:

$$q'(x,\beta,\omega) \le q'(\phi,\beta,\omega) \le \max\left\{q'(\phi,0,\omega), q'(\phi,\kappa Z^*,\omega)\right\}$$

(b) For any $x \in X$, $\omega \in \Omega$ and $\beta \in [0,\infty)$, we have:

$$q'(\phi,\beta,\omega) - q'(x,\beta,\omega) \le \frac{\lambda c(x)}{1-\alpha}$$

Proof. The first inequality in (a) follows directly from the empty first stage assumption. The second inequality follows from equation (9) above. To establish (b), we compute:

$$\begin{aligned} q'(\phi,\beta,\omega) - q'(x,\beta,\omega) &= \frac{1}{1-\alpha}\left[(q(\phi,\omega)-\beta)^+ - (q(x,\omega)-\beta)^+\right] \\ &= \begin{cases} \dfrac{q(\phi,\omega)-q(x,\omega)}{1-\alpha} & \text{if } \beta \le q(x,\omega) \\ \dfrac{q(\phi,\omega)-\beta}{1-\alpha} & \text{if } q(x,\omega) < \beta \le q(\phi,\omega) \\ 0 & \text{if } q(x,\omega) \le q(\phi,\omega) < \beta \end{cases} \end{aligned}$$

The desired result then follows from the bounded inflation factor assumption. □

Before we proceed further, let us first introduce a definition and state the version of the Hoeffding bound that we will be using later.

DEFINITION 3.1 *We say that $x^* \in X$ is an exact (resp. γ–approximate) minimizer of a function f if we have $f(x^*) \le f(x)$ (resp. $f(x^*) \le \gamma f(x)$) for all $x \in X$.*

LEMMA 3.2 (HOEFFDING BOUND; CF. [12]) *Let $V_1, \ldots, V_n$ be independent random variables with $0 \le V_i \le 1$ for $i = 1, \ldots, n$. Set $V = \sum_{i=1}^n V_i$. Then, for any $\epsilon > 0$, we have $P\left(\left| V - \mathbb{E}[V] \right| > \epsilon n \right) \le 2e^{-\epsilon^2 n}$.*

Here is our main sampling theorem.

THEOREM 3.1 *Let $g'_\alpha(x,\beta) = c(x) + \mathbb{E}_{\mathbb{P}}[q'(x,\beta,\omega)]$, where c and q' satisfy the assumptions above, and $\alpha \in [0,1)$ is the risk aversion level. Let $\epsilon \in (0,1/3]$ and $\delta \in (0,1/2)$ be given. Set:*

$$\lambda_\alpha = \frac{\lambda}{1-\alpha}, \quad \eta = \max\left\{1, \frac{\alpha}{1-\alpha}\right\}$$

and define:

$$\hat{g}_\alpha^N(x,\beta) = c(x) + \beta + \frac{1}{N(1-\alpha)} \sum_{i=1}^N \left(q(x,\omega^i) - \beta \right)^+$$

to be the SAA of g'_α, where $\omega^1, \ldots, \omega^N$ are N i.i.d. samples from the underlying distribution, and

$$N = \Theta\left(\frac{\lambda_\alpha^2}{\epsilon^4 (1-\alpha)^2} \log\left(\frac{\eta}{\epsilon} \cdot |X| \cdot \frac{1}{\delta} \right) \right)$$

Let $\kappa \ge 1$ be fixed, and suppose that $(\bar{x}, \bar{\beta}) \in X \times [0, \kappa Z^*]$ is an exact minimizer of $\hat{g}_\alpha^N$ over the domain $X \times [0, \kappa Z^*]$. Then, with probability at least $1 - 2\delta$, the solution $(\bar{x}, \bar{\beta})$ is an $(1 + \Theta(\epsilon\kappa))$–approximate minimizer of g'_α.

REMARKS.

(a) Note that $(\bar{x}, \bar{\beta})$ need not be a global minimizer of $\hat{g}_\alpha^N$ over $X \times [0, \infty)$, since such a global minimizer may have $\beta > \kappa Z^*$. In other words, the optimal solutions to the problems:

$$\min_{(x,\beta) \in X \times [0,\infty)} \hat{g}_\alpha^N(x,\beta) \tag{10}$$

and

$$\min_{(x,\beta) \in [0,\kappa Z^*]} \hat{g}_\alpha^N(x,\beta) \tag{11}$$

could be different. From a practitioner's point of view, it may be easier to solve (10) than (11), because in many applications, it is difficult to estimate Z^* without actually solving the problem. However, it can be shown (see Theorem 3.2) that by repeating the sampling sufficiently many times, we can obtain a sample average approximation $\hat{g}_\alpha^N$ whose exact minimizers $(\bar{x}^*, \bar{\beta}^*)$ over $X \times [0, \infty)$ satisfy $\bar{\beta}^* \le (1+\epsilon) Z^*$ with high probability. Thus, we can still apply the theorem even though we are solving Problem (10).

(b) Note that this theorem does not follow from a direct application of [3, Theorem 1] for two reasons. First, the domain of our optimization problem is $X \times [0, \kappa Z^*]$, which is compact but not finite. However, this can be circumvented by using a suitably chosen grid on $[0, \kappa Z^*]$. A second, and perhaps more serious, problem is that there may not exist a $\beta_0 \in [0, \kappa Z^*]$ such that $q'(x,\beta,\omega) \le q'(\phi, \beta_0, \omega)$ for *all* $x \in X$ and $\omega \in \Omega$. Such an assumption is crucial in the analysis in [3]. On the other hand, we have the weaker statement of Lemma 3.1(a), and that turns out to be sufficient for establishing our theorem.

PROOF. Let (x^*, β^*) be an exact minimizer of g'_α. Then, we have $Z^* = g'_\alpha(x^*, \beta^*)$. Our proof consists of three steps.

<u>**Step 1**</u>: Isolate the high–cost scenarios and bound their total probability mass.

We divide the scenarios into two classes based on a parameter $M > 0$ to be determined later. Specifically, we say that a scenario ω is *high* if $q(\phi, \omega)$ exceeds the threshold M (i.e. $q(\phi, \omega) > M$); otherwise, we say that ω is *low*. Let $p = \mathbb{P}(\omega : \omega \text{ is high})$, and define:

$$\hat{l}_\alpha^N(x,\beta) = \frac{1}{N} \sum_{i:\, \omega^i \; low} q'(x,\beta,\omega^i)$$

$$\hat{h}_\alpha^N(x,\beta) = \frac{1}{N} \sum_{i:\, \omega^i \; high} q'(x,\beta,\omega^i)$$

Then, it is clear that $\hat{g}_\alpha^N(x,\beta) = c(x) + \hat{l}_\alpha^N(x,\beta) + \hat{h}_\alpha^N(x,\beta)$. Similarly, we define:

$$l'_\alpha(x,\beta) = \mathbb{E}_\mathbb{P}\left[q'(x,\beta,\omega)\cdot \mathbf{1}_{\{\omega \text{ is low}\}}\right] = (1-p)\cdot \mathbb{E}_\mathbb{P}\left[q'(x,\beta,\omega)\,|\,\omega \text{ is low}\right]$$
$$h'_\alpha(x,\beta) = \mathbb{E}_\mathbb{P}\left[q'(x,\beta,\omega)\cdot \mathbf{1}_{\{\omega \text{ is high}\}}\right] = p\cdot \mathbb{E}_\mathbb{P}\left[q'(x,\beta,\omega)\,|\,\omega \text{ is high}\right]$$

whence $g'_\alpha(x,\beta) = c(x) + l'_\alpha(x,\beta) + h'_\alpha(x,\beta)$. Now, we can bound p by choosing M appropriately:

Lemma 3.3 *Let $M = \epsilon^{-1}\lambda_\alpha Z^*$. Then, we have $p \le \frac{\epsilon}{\lambda_\alpha(1-\epsilon)}$.*

Proof. We compute:

$$Z^* = g'_\alpha(x^*,\beta^*) \ge h'_\alpha(x^*,\beta^*) = p\cdot \mathbb{E}_\mathbb{P}\left[q'(x^*,\beta^*,\omega)\,|\,\omega \text{ is high}\right] \tag{12}$$

By Lemma 3.1(b), we have $q'(x^*,\beta^*,\omega) \ge q'(\phi,\beta^*,\omega) - \lambda_\alpha c(x^*)$ for all $\omega \in \Omega$. Moreover, by equation (9), we have $q'(\phi,\beta^*,\omega) > M$ whenever ω is a high scenario. Thus, it follows from (12) that:

$$Z^* \ge p(M - \lambda_\alpha c(x^*)) \ge p(M - \lambda_\alpha Z^*)$$

Upon setting $M = \epsilon^{-1}\lambda_\alpha Z^*$, we have $Z^* \ge p\lambda_\alpha Z^*(\epsilon^{-1} - 1)$, which implies the desired result. □

Armed with Lemma 3.3, we can prove the following result:

Lemma 3.4 *Let N_h be the number of high scenarios in the samples $w^1,\dots,w^N$. Then, with probability at least $1-\delta$, we have $N_h/N \le 2\epsilon/\lambda_\alpha$.*

Proof. Define the indicator random variables $X_1,\dots,X_N$ as follows:

$$X_i = \begin{cases} 1 & \text{if } \omega^i \text{ is high} \\ 0 & \text{otherwise} \end{cases}$$

Then, the random variables $X_1,\dots,X_N$ are i.i.d., and we have $N_h = \sum_{i=1}^N X_i$ and $\mathbb{E}_\mathbb{P}[N_h] = pN \le \frac{\epsilon N}{\lambda_\alpha(1-\epsilon)}$. Using Lemma 3.2, we have:

$$\mathbb{P}\left\{N_h - pN > \frac{\epsilon N}{\lambda_\alpha}\left(2 - \frac{1}{1-\epsilon}\right)\right\} \le \exp\left(-\frac{\epsilon^2(1-2\epsilon)^2 N}{\lambda_\alpha^2(1-\epsilon)^2}\right)$$

Then, for $\epsilon \in (0,1/3]$, the above probability is at most δ by our choice of N. Thus, with probability at least $1-\delta$, we have:

$$N_h - pN \le \frac{\epsilon N}{\lambda_\alpha}\left(2 - \frac{1}{1-\epsilon}\right)$$

or equivalently, $N_h/N \le 2\epsilon/\lambda_\alpha + p - \frac{\epsilon}{\lambda_\alpha(1-\epsilon)} \le 2\epsilon/\lambda_\alpha$, as desired. □

Step 2: Establish the quality of the scenario partition.

We claim that each of the following events occurs with probability at least $1-\delta$:

$$A_1 = \left\{\left|l'_\alpha(x,\beta) - \hat{l}_\alpha^N(x,\beta)\right| \le 2\epsilon\kappa Z^* \;\text{ for every } (x,\beta)\in X\times[0,\kappa Z^*]\right\}$$

$$A_2 = \left\{\hat{h}_\alpha^N(\phi,\beta) - \hat{h}_\alpha^N(x,\beta) \le 2\epsilon c(x) \;\text{ for every } (x,\beta)\in X\times[0,\infty)\right\}$$

$$A_3 = \{h'_\alpha(\phi,\beta) - h'_\alpha(x,\beta) \le 2\epsilon c(x) \;\text{ for every } (x,\beta)\in X\times[0,\infty)\}$$

Indeed, observe that by Lemma 3.1(b) and Lemma 3.4, for any $(x,\beta) \in X \times [0,\infty)$, we have:

$$\hat{h}_\alpha^N(\phi,\beta) - \hat{h}_\alpha^N(x,\beta) = \frac{1}{N} \sum_{i:\, \omega^i \, high} \left(q'(\phi,\beta,w^i) - q'(x,\beta,w^i) \right) \le \frac{N_h}{N} \cdot \lambda_\alpha c(x) \le 2\epsilon c(x)$$

with probability at least $1-\delta$. Similarly, for any $(x,\beta) \in X \times [0,\infty)$ and $\epsilon \in (0,1/3]$, we have:

$$h'_\alpha(\phi,\beta) - h'_\alpha(x,\beta) = p \cdot \mathbb{E}_{\mathbb{P}} \left[\left(q'(\phi,\beta,\omega) - q'(x,\beta,\omega) \right) \mid \omega \text{ is high} \right] \le \frac{\epsilon}{\lambda_\alpha(1-\epsilon)} \cdot \lambda_\alpha c(x) \le 2\epsilon c(x)$$

with probability 1. Thus, we conclude that both events A_2 and A_3 occur with the required probability.

To prove that event A_1 occurs with probability at least $1-\delta$, we need the following crucial observation:

Lemma 3.5 *For each $x \in X$ and $\omega \in \Omega$, the function $q'(x,\cdot,\omega)$ is η–Lipschitz (i.e. $\left| q'(x,\beta_1,\omega) - q'(x,\beta_2,\omega) \right| \le \eta \left| \beta_1 - \beta_2 \right|$), where $\eta = \max\left\{1, \frac{\alpha}{1-\alpha}\right\}$.*

Proof. Let $\beta_1, \beta_2 \ge 0$ be such that $\beta_1 < \beta_2$. Then, we have:

$$\left| q'(x,\beta_1,\omega) - q'(x,\beta_2,\omega) \right| \left\{ \begin{array}{lll} = & \beta_2 - \beta_1 & \text{if } q(x,\omega) \le \beta_1 \\ \le & \max\left\{1, \dfrac{\alpha}{1-\alpha}\right\} \cdot (\beta_2 - \beta_1) & \text{if } \beta_1 \le q(x,\omega) < \beta_2 \\ = & \dfrac{\alpha}{1-\alpha}(\beta_2 - \beta_1) & \text{otherwise} \end{array} \right.$$

This completes the proof. □

Now, we use a standard meshing argument to establish the desired result. Define $S = \{i\epsilon\kappa Z^*/\eta : i = 0,1,\ldots,\lceil \eta/\epsilon \rceil\}$, and consider a fixed $(x,\beta) \in X \times S$. Let W be the random variable given by:

$$W = \left\{ \begin{array}{ll} q'(x,\beta,\omega) & \text{if } \omega \text{ is low} \\ 0 & \text{otherwise} \end{array} \right.$$

It follows that $\mathbb{E}_{\mathbb{P}}[W] = l'_\alpha(x,\beta)$. Now, let $W_1,\ldots,W_N$ be N i.i.d. samples of W. Observe that $\frac{1}{N}\sum_{i=1}^N W_i$ has the same distribution as $\hat{l}_\alpha^N(x,\beta)$. By Lemma 3.1(a), we have:

$$W_i \le \max\{q'(\phi,0,\omega), q'(\phi,\kappa Z^*,\omega)\}$$

for $i = 1,2,\ldots,N$. However, for a low scenario ω, we have:

$$q'(\phi,0,\omega) = \frac{q(\phi,\omega)}{1-\alpha} \le \frac{M}{1-\alpha}$$

and

$$q'(\phi,\kappa Z^*,\omega) \le \max\left\{\kappa Z^*, \frac{M - \alpha\kappa Z^*}{1-\alpha}\right\} \le \max\left\{\kappa Z^*, \frac{M}{1-\alpha}\right\}$$

(recall that M is the threshold for determining whether a scenario is high or not). Hence, by setting:

$$M' = \max\left\{\kappa Z^*, \frac{M}{1-\alpha}\right\}, \quad Y_i = W_i/M', \quad Y = \sum_{i=1}^N Y_i$$

we see that $Y_i \in [0,1]$ and $\mathbb{E}_{\mathbb{P}}[Y] = \frac{N}{M'} l'_\alpha(x,\beta)$. Moreover, the random variable Y has the same distribution as $\frac{N}{M'}\hat{l}_\alpha^N(x,\beta)$. It then follows from Lemma 3.2 and our choice of N that:

$$\mathbb{P}\left(\left| Y - \frac{N}{M'} l'_\alpha(x,\beta) \right| > \frac{\epsilon^2(1-\alpha)N}{\lambda_\alpha} \right) \le 2\exp\left(-\frac{\epsilon^4(1-\alpha)^2 N}{\lambda_\alpha^2} \right) \le \frac{\delta}{|X| \cdot |S|}$$

By applying the union bound over all $(x,\beta) \in X \times S$, it follows that:

$$\left| l'_\alpha(x,\beta) - \hat{l}_\alpha^N(x,\beta) \right| \le \frac{\epsilon^2(1-\alpha)M'}{\lambda_\alpha} \le \epsilon\kappa Z^*$$

for all $(x,\beta) \in X \times S$ with probability at least $1-\delta$.

Now, consider an arbitrary pair $(x, \beta) \in X \times [0, \kappa Z^*]$. Then, there exists a $\beta' \in S$ such that $|\beta - \beta'| \leq \epsilon\kappa Z^*/2\eta$. Since $q'(x, \cdot, \omega)$ is η–Lipschitz by Lemma 3.5, we have $|q'(x, \beta, \omega) - q'(x, \beta', \omega)| \leq \epsilon\kappa Z^*/2$. It follows that $\left|l'_\alpha(x, \beta) - l'_\alpha(x, \beta')\right| \leq \epsilon\kappa Z^*/2$ and $\left|\hat{l}^N_\alpha(x, \beta) - \hat{l}^N_\alpha(x, \beta')\right| \leq \epsilon\kappa Z^*/2$. Using the triangle inequality, we conclude that $\left|\hat{l}^N_\alpha(x, \beta) - l'_\alpha(x, \beta)\right| \leq 2\epsilon\kappa Z^*$ whenever $\left|\hat{l}^N_\alpha(x, \beta') - l'_\alpha(x, \beta')\right| \leq \epsilon\kappa Z^*$. This shows that event A_1 occurs with probability at least $1 - \delta$.

Step 3: Establish the approximation guarantee.

With probability at least $1 - 2\delta$, we may assume that all of the events A_1, A_2 and A_3 occur. Then, for any $(x, \beta) \in X \times [0, \kappa Z^*]$, we have:

$$
\begin{aligned}
l'_\alpha(x, \beta) &\leq \hat{l}^N_\alpha(x, \beta) + 2\epsilon\kappa Z^* && \text{(Event } A_1\text{)} \\
h'_\alpha(x, \beta) &\leq h'_\alpha(\phi, \beta) && \text{(Lemma 3.1(a))} \\
0 &\leq \hat{h}^N_\alpha(x, \beta) + 2\epsilon c(x) - \hat{h}^N_\alpha(\phi, \beta) && \text{(Event } A_2\text{)}
\end{aligned}
$$

Upon summing the above inequalities, we obtain:

$$
g'_\alpha(x, \beta) - \hat{g}^N_\alpha(x, \beta) \leq 2\epsilon\kappa Z^* + 2\epsilon c(x) + h'_\alpha(\phi, \beta) - \hat{h}^N_\alpha(\phi, \beta) \tag{13}
$$

Similarly, we have:

$$
\begin{aligned}
\hat{l}^N_\alpha(x, \beta) &\leq l'_\alpha(x, \beta) + 2\epsilon\kappa Z^* && \text{(Event } A_1\text{)} \\
\hat{h}^N_\alpha(x, \beta) &\leq \hat{h}^N_\alpha(\phi, \beta) && \text{(Lemma 3.1(a))} \\
0 &\leq h'_\alpha(x, \beta) + 2\epsilon c(x) - h'_\alpha(\phi, \beta) && \text{(Event } A_3\text{)}
\end{aligned}
$$

from which it follows that:

$$
\hat{g}^N_\alpha(x, \beta) - g'_\alpha(x, \beta) \leq 2\epsilon\kappa Z^* + 2\epsilon c(x) + \hat{h}^N_\alpha(\phi, \beta) - h'_\alpha(\phi, \beta) \tag{14}
$$

Now, let $(\bar{x}, \bar{\beta}) \in X \times [0, \kappa Z^*]$ be an exact minimizer of $\hat{g}^N_\alpha$ over $[0, \kappa Z^*]$. Upon instantiating (x, β) by $(\bar{x}, \bar{\beta})$ in (13) and by (x^*, β^*) in (14) (recall that (x^*, β^*) is an exact minimizer of g'_α) and summing, we have:

$$
\begin{aligned}
& g'_\alpha(\bar{x}, \bar{\beta}) - g'_\alpha(x^*, \beta^*) + \hat{g}^N_\alpha(x^*, \beta^*) - \hat{g}^N_\alpha(\bar{x}, \bar{\beta}) \\
\leq \quad & 4\epsilon\kappa Z^* + 2\epsilon c(\bar{x}) + 2\epsilon c(x^*) + h'_\alpha(\phi, \bar{\beta}) - h'_\alpha(\phi, \beta^*) + \hat{h}^N_\alpha(\phi, \beta^*) - \hat{h}^N_\alpha(\phi, \bar{\beta})
\end{aligned}
$$

Using Lemma 3.3 and Lemma 3.5, we bound:

$$
\left|h'_\alpha(\phi, \bar{\beta}) - h'_\alpha(\phi, \beta^*)\right| \leq p \cdot \eta \left|\bar{\beta} - \beta^*\right| \leq \frac{\epsilon\eta\kappa Z^*}{\lambda_\alpha(1 - \epsilon)} \leq 2\epsilon\kappa Z^*
$$

where the last inequality follows from the facts that $\alpha \in [0, 1)$, $\epsilon \in (0, 1/2]$ and $\lambda \geq 1$. Similarly, together with Lemma 3.4, we have:

$$
\left|\hat{h}^N_\alpha(\phi, \beta^*) - \hat{h}^N_\alpha(\phi, \bar{\beta})\right| \leq \frac{N_h}{N} \cdot \eta \left|\beta^* - \bar{\beta}\right| \leq \frac{2\epsilon\eta\kappa Z^*}{\lambda_\alpha} \leq 2\epsilon\kappa Z^*
$$

Since we have $\hat{g}^N_\alpha(\bar{x}, \bar{\beta}) \leq \hat{g}^N_\alpha(x^*, \beta^*)$, we conclude that:

$$
\begin{aligned}
(1 - 2\epsilon) g'_\alpha(\bar{x}, \bar{\beta}) &\leq g'_\alpha(\bar{x}, \bar{\beta}) - 2\epsilon c(\bar{x}) \\
&\leq g'_\alpha(x^*, \beta^*) + 2\epsilon c(x^*) + 4\epsilon\kappa Z^* + 4\epsilon\kappa Z^* \\
&\leq (1 + 10\epsilon\kappa) Z^*
\end{aligned}
$$

It follows that $g'_\alpha(\bar{x}, \bar{\beta}) \leq (1 + \Theta(\epsilon\kappa)) Z^*$ as desired. This completes the proof of Theorem 3.1. □

The next theorem shows that by repeating the sampling sufficiently many times, we can obtain a SAA $\hat{g}^N_\alpha$ whose exact minimizers $(\bar{x}^*, \bar{\beta}^*)$ over $X \times [0, \infty)$ satisfy $\bar{\beta}^* \leq (1 + \epsilon) Z^*$ with high probability.

THEOREM 3.2 *Let $\alpha \in [0,1)$, $\epsilon \in (0,1/3]$ and $\delta \in (0,1/3)$ be given, and let*

$$k = \Theta\left(\left(1+\frac{1}{\epsilon}\right)\log\frac{1}{\delta}\right), \quad N = \Theta\left(\frac{\lambda_\alpha^2}{\epsilon^4(1-\alpha)^2}\log\left(\frac{\eta}{\epsilon}\cdot|X|\cdot\frac{1}{\delta}\cdot k\right)\right)$$

Consider a collection $\hat{g}_\alpha^{1,N},\ldots,\hat{g}_\alpha^{k,N}$ of independent SAAs of g'_α, where each $\hat{g}_\alpha^{i,N}$ uses N i.i.d. samples of the scenarios. For $i=1,2,\ldots,k$, let $(\bar{x}^i,\bar{\beta}^i)$ be an exact minimizer of $\hat{g}_\alpha^{i,N}$ over $X\times[0,\infty)$. Set $v = \arg\min_i \hat{g}_\alpha^{i,N}(\bar{x}^i,\bar{\beta}^i)$. Then, with probability at least $1-3\delta$, the solution $(\bar{x}^v,\bar{\beta}^v)$ satisfies $\bar{\beta}^v \le (1+\epsilon)Z^*$ and is an $(1+\Theta(\epsilon))$–minimizer of g'_α.

PROOF. We set $\kappa = 1+\epsilon$ in Theorem 3.1. Let us call $\hat{g}_\alpha^{i,N}$ *good* if both events A_1 and A_2 hold. Upon following the computations in Step 2 of Theorem 3.1, we conclude that $\hat{g}_\alpha^{i,N}$ is good with probability at least $1-2\delta/k$. It follows that all the $\hat{g}_\alpha^{i,N}$'s are good with probability at least $1-2\delta$. Now, let (x^*,β^*) be an exact minimizer of g'_α. Since $\mathbb{E}_\mathbb{P}\left[\hat{g}_\alpha^{i,N}(x^*,\beta^*)\right] = g'_\alpha(x^*,\beta^*)$, we have, by Markov's inequality, that:

$$\mathbb{P}\left(\hat{g}_\alpha^{i,N}(x^*,\beta^*) > (1+\epsilon)g'_\alpha(x^*,\beta^*)\right) \le \frac{1}{1+\epsilon} = 1-\frac{1}{1+1/\epsilon}$$

for $i=1,2,\ldots,k$. It follows that:

$$\mathbb{P}\left(\hat{g}_\alpha^{i,N}(x^*,\beta^*) \le (1+\epsilon)g'_\alpha(x^*,\beta^*) \text{ for some } i=1,\ldots,k\right) \ge 1-\left(1-\frac{1}{1+1/\epsilon}\right)^k \ge 1-\delta$$

Thus, with probability at least $1-3\delta$, there exists an index u such that $\hat{g}_\alpha^{u,N}$ is good, and that:

$$\hat{g}_\alpha^{u,N}(\bar{x}^u,\bar{\beta}^u) \le \hat{g}_\alpha^{u,N}(x^*,\beta^*) \le (1+\epsilon)g'_\alpha(x^*,\beta^*) = (1+\epsilon)Z^*$$

It follows that $\bar{\beta}^u \le (1+\epsilon)Z^*$. In particular, we may apply Theorem 3.1 and conclude that $g'_\alpha(\bar{x}^u,\bar{\beta}^u) \le (1+\Theta(\epsilon))Z^*$. This completes the proof. □

Note that in Theorems 3.1 and 3.2, we assume that the problem of minimizing $\hat{g}_\alpha^N$ can be solved exactly. In many cases of interest, however, we can only get an approximate minimizer of $\hat{g}_\alpha^N$. The following theorem shows that we can still guarantee a near–optimal solution in this case.

THEOREM 3.3 *Let $\alpha \in [0,1)$, $\epsilon \in (0,1/3]$ and $\delta \in (0,1/5)$ be given. Let*

$$k = \Theta((1+\epsilon^{-1})\log\delta^{-1}), \quad k' = \Theta((1+\epsilon^{-1})\log(k\delta^{-1}))$$

and set:

$$N = \Theta\left(\frac{\lambda_\alpha^2}{\epsilon^4(1-\alpha)^2}\log\left(\frac{\eta}{\epsilon}\cdot|X|\cdot\frac{1}{\delta}\cdot kk'\right)\right)$$

Consider a collection $\left\{\hat{g}_\alpha^{(i,j),N}\right\}_{i=1,j=1}^{i=k,j=k'}$ of independent SAAs of g'_α, where each $\hat{g}_\alpha^{(i,j),N}$ uses N i.i.d. samples of the scenarios. Then, with probability at least $1-5\delta$, one can find a pair of indices (u,v) such that any γ–approximate minimizer of $\hat{g}_\alpha^{(u,v),N}$ is an $(1+\Theta(\epsilon))\gamma$–minimizer of g'_α.

PROOF. As before, let (x^*,β^*) be an exact minimizer of g'_α. We call $\hat{g}_\alpha^{(i,j),N}$ *good* if events A_1 and A_2 hold for both $\kappa=1$ and $\kappa=(1+\epsilon)\gamma$. Similar to the proof of Theorem 3.2, we conclude that with probability at least $1-4\delta$, we can find indices $v_1,\ldots,v_k$ such that for all $i=1,2,\ldots,k$, we have:

(a) $\hat{g}_\alpha^{(i,v_i),N}$ is good; and

(b) $\hat{g}_\alpha^{(i,v_i),N}(x^*,\beta^*) \le (1+\epsilon)Z^*$.

Note that (b) implies that if $(\bar{x}^i,\bar{\beta}^i)$ is a γ–approximate minimizer of $\hat{g}_\alpha^{(i,v_i),N}$, then we have $\hat{g}_\alpha^{(i,v_i),N}(\bar{x}^i,\bar{\beta}^i) \le (1+\epsilon)\gamma Z^*$. In particular, we have $\bar{\beta}^i \le (1+\epsilon)\gamma Z^*$. Now, by Markov's inequality, we have:

$$\mathbb{P}\left(\hat{h}_\alpha^{(i,v_i),N}(\phi,\beta) \le (1+\epsilon)h'_\alpha(\phi,\beta) \text{ for some } i=1,\ldots,k\right) \ge 1-\left(1-\frac{1}{1+1/\epsilon}\right)^k \ge 1-\delta$$

Thus, with probability at least $1-5\delta$, we have a collection $\left\{\hat{g}_\alpha^{(i,v_i),N}\right\}$ of functions that satisfy (a) and (b) above, and that there exists an index u such that:

$$\hat{h}_\alpha^{(u,v_u),N}(\phi,\beta) \le (1+\epsilon)h'_\alpha(\phi,\beta) \tag{15}$$

For simplicity of notation, let us write $\hat{g}_\alpha^{i,N}$ for $\hat{g}_\alpha^{(i,v_i),N}$. Now, let $v = \arg\min_i \hat{g}_\alpha^{i,N}(\bar{x}^i, \bar{\beta}^i)$, where $(\bar{x}^i, \bar{\beta}^i)$ is a γ–approximate minimizer of $\hat{g}_\alpha^{i,N}$, for $i = 1, 2, \ldots, k$. Note that we have $\bar{\beta}^v \le (1+\epsilon)\gamma Z^*$. Since we have $\hat{g}_\alpha^{v,N}(\bar{x}^v, \bar{\beta}^v) \le \hat{g}_\alpha^{u,N}(\bar{x}^u, \bar{\beta}^u)$, it follows that:

$$\hat{g}_\alpha^{v,N}(\bar{x}^v, \bar{\beta}^v) \le \frac{1}{\gamma}\hat{g}_\alpha^{v,N}(\bar{x}^v, \bar{\beta}^v) + \frac{\gamma-1}{\gamma}\hat{g}_\alpha^{u,N}(\bar{x}^u, \bar{\beta}^u) \le \hat{g}_\alpha^{v,N}(x^*, \beta^*) + (\gamma-1)\hat{g}_\alpha^{u,N}(x^*, \beta^*) \tag{16}$$

Upon instantiating (x, β) by $(\bar{x}^v, \bar{\beta}^v)$ in (13) and by (x^*, β^*) in (14), we obtain:

$$g'_\alpha(\bar{x}^v, \bar{\beta}^v) - \hat{g}_\alpha^{v,N}(\bar{x}^v, \bar{\beta}^v) \le 2\epsilon(1+\epsilon)\gamma Z^* + 2\epsilon c(\bar{x}^v) + h'_\alpha(\phi, \bar{\beta}^v) - \hat{h}_\alpha^{v,N}(\phi, \bar{\beta}^v) \tag{17}$$

$$\hat{g}_\alpha^{v,N}(x^*, \beta^*) - g'_\alpha(x^*, \beta^*) \le 2\epsilon(1+\epsilon)\gamma Z^* + 2\epsilon c(x^*) + \hat{h}_\alpha^{v,N}(\phi, \beta^*) - h'_\alpha(\phi, \beta^*) \tag{18}$$

Upon summing (16), (17) and (18) and following the computations in Step 3 of Theorem 3.1, we obtain:

$$g'_\alpha(\bar{x}^v, \bar{\beta}^v) - g'_\alpha(x^*, \beta^*) \le 8\epsilon(1+\epsilon)\gamma Z^* + 2\epsilon c(\bar{x}^v) + 2\epsilon c(x^*) + (\gamma-1)\hat{g}_\alpha^{u,N}(x^*, \beta^*)$$

Now, upon instantiating (x, β) by (x^*, β^*) in (14) and noting that $\beta^* \le Z^*$, we have:

$$\begin{aligned} \hat{g}_\alpha^{u,N}(x^*, \beta^*) &\le g'_\alpha(x^*, \beta^*) + 2\epsilon Z^* + 2\epsilon c(x^*) + \hat{h}_\alpha^{u,N}(\phi, \beta^*) - h'_\alpha(\phi, \beta^*) \\ &\le Z^* + 2\epsilon Z^* + 2\epsilon c(x^*) + \epsilon(1+2\epsilon)Z^* \qquad (19) \\ &\le (1+6\epsilon)Z^* \end{aligned}$$

where (19) follows from the facts that $\hat{h}_\alpha^{u,N}(\phi, \beta^*) - h'_\alpha(\phi, \beta^*) \le \epsilon h'_\alpha(\phi, \beta^*)$ (see (15)), and that on the almost sure event A_3, we have:

$$h'_\alpha(\phi, \beta^*) \le h'_\alpha(x^*, \beta^*) + 2\epsilon c(x^*) \le (1+2\epsilon)Z^*$$

It follows that:

$$(1-2\epsilon)g'_\alpha(\bar{x}^v, \bar{\beta}^v) \le (1+14\epsilon\gamma)Z^* + (\gamma-1)\hat{g}_\alpha^{u,N}(x^*, \beta^*) \le (1+20\epsilon)\gamma Z^*$$

whence:

$$g'_\alpha(\bar{x}^v, \bar{\beta}^v) \le \frac{1+20\epsilon}{1-2\epsilon}\gamma Z^* = (1+\Theta(\epsilon))\gamma Z^*$$

as desired. □

As we shall see, Theorems 3.2 and 3.3 play an important role in the design of efficient approximation algorithms for various risk–adjusted stochastic combinatorial optimization problems under the black–box model. In particular, they allow us to generalize the recent results in [23, 3, 6] to the risk–adjusted setting.

4. Applications In this section we consider three stochastic combinatorial optimization problems that are special cases of Problem (8) and develop approximation algorithms for them. As we shall see, a key feature that is common to all these algorithms is *locality*, i.e. for each scenario A in the second stage, given a fractional second–stage solution to scenario A, the algorithm will return an integral second–stage solution to scenario A whose value is at most some factor times the value of the fractional solution. Such a feature allows us to take advantage of the following easily–checked properties of φ_α: for random variables $Z_1, Z_2 \in L^2(\Omega, \mathscr{B}, \mathbb{P})$ and any $\alpha \in [0,1)$, we have:

(a) **(Positive Homogeneity)** $\varphi_\alpha(cZ_1) = c\varphi_\alpha(Z_1)$ for any constant $c > 0$.

(b) **(Monotonicity)** If $Z_1 \le Z_2$ a.e., then $\varphi_\alpha(Z_1) \le \varphi_\alpha(Z_2)$.

In the sequel, we assume that the cost functions satisfy the properties in Section 3. In view of Theorems 3.2 and 3.3, we may also assume that for each of the problems under consideration, there is only a polynomial number (say N) of scenarios, each occurring with probability $1/N$.

4.1 Covering Problems Before we define the 2–stage stochastic set cover problem, let us recall the setting of the usual (deterministic) set cover problem. We are given a universe U of elements $e_1, \ldots, e_n$ and a collection $\mathcal{S}$ of subsets of U, say $\mathcal{S} = \{S_1, \ldots, S_m\}$. Each set $S \in \mathcal{S}$ has a non–negative weight w_S, and our goal is to choose a minimum–weight collection of sets from $\mathcal{S}$ so that every element in U

belongs to some set in the collection. As is well–known (see, e.g., [26, Chapter 13]), the following is an LP relaxation of the deterministic set cover problem:

$$\begin{aligned} \text{minimize} \quad & \sum_{S \in \mathcal{S}} w_S x_S \\ \text{subject to} \quad & \sum_{S : e \in S} x_S \ge 1 \quad \forall\, e \in U \\ & x_S \ge 0 \qquad\quad \forall\, S \in \mathcal{S} \end{aligned} \tag{20}$$

In the 2–stage stochastic set cover problem, we are again given a universe U of elements and a collection $\mathcal{S}$ of subsets of U. However, the elements to be covered are not known in advance. Instead, there is a probability distribution over scenarios, where each scenario specifies a subset $A \subset U$ of elements to be covered by the sets in $\mathcal{S}$. For notational convenience, we shall use $A \subset U$ to index a scenario in the sequel. Now, each set $S \in \mathcal{S}$ has a first–stage weight $w_S^I \ge 0$ and a second–stage weight $w_S^{II} \ge 0$. In the first stage, one selects some of these sets and incurs their first–stage weights. Then, a scenario $A \subset U$ is drawn according to the underlying distribution. If the union of the sets chosen in the first stage cannot cover A, then additional sets from $\mathcal{S}$ may be selected, and their second–stage weights will be incurred. At the end, the union of the sets chosen in the first and second stage must cover A. Now, let $\alpha \in [0,1)$ be the risk aversion level. Then, we can formulate the risk–adjusted 2–stage stochastic set cover problem as follows:

$$\text{minimize} \ \sum_{S \in \mathcal{S}} w_S^I x_S + \varphi_\alpha(q(x,A)) \quad \text{subject to } x_S \in \{0,1\} \ \forall\, S \in \mathcal{S} \tag{21}$$

where:

$$\begin{aligned} q(x,A) &= \text{minimize} \ \sum_{S \in \mathcal{S}} w_S^{II} r_{A,S} \quad \text{subject to } r_A \in \mathcal{F}(x,A) \\ \mathcal{F}(x,A) &= \left\{ r_A : \sum_{S : e \in S} r_{A,S} \ge 1 - \sum_{S : e \in S} x_S \ \ \forall\, e \in A, \ r_{A,S} \in \{0,1\} \ \forall\, S \in \mathcal{S} \right\} \end{aligned}$$

If we index the scenarios by $A_1, \ldots, A_N$, then we can write Problem (21) in the following equivalent form:

$$\begin{aligned} \text{minimize} \quad & \sum_{S \in \mathcal{S}} w_S^I x_S + \beta + \frac{1}{N(1-\alpha)} \sum_{i=1}^N \left(\sum_{S \in \mathcal{S}} w_S^{II} r_{A_i,S} - \beta \right)^+ \\ \text{subject to} \quad & \sum_{S : e \in S} x_S + \sum_{S : e \in S} r_{A_i,S} \ge 1 && \forall\, e \in A_i, \ i = 1, \ldots, N \\ & x_S \in \{0,1\} && \forall\, S \in \mathcal{S} \\ & r_{A_i,S} \in \{0,1\} && \forall\, S \in \mathcal{S}, \ i = 1, \ldots, N \\ & \beta \ge 0 \end{aligned}$$

Thus, by relaxing the binary constraints in the straightforward manner and reformulating, we obtain a linear program that can be solved in polynomial time. The following theorem shows that such an LP relaxation can be used to obtain a good approximation to the risk–adjusted 2–stage stochastic set cover problem. It can be viewed as a generalization of a result by Shmoys and Swamy [23, Theorem 2.1] to the risk–adjusted setting.

THEOREM 4.1 *Suppose that we have a procedure that, for any instance of the deterministic set cover problem, produces a feasible solution whose value is at most ρ times the optimal value of the LP relaxation (20). Then, we can convert any feasible solution (x,r) to the LP relaxation of (21) into an integral solution whose value is at most 2ρ times the value of (x,r). In particular, we have a 2ρ–approximation algorithm for the risk–adjusted 2–stage stochastic set cover problem.*

PROOF. Let (x,r) be a feasible solution to the LP relaxation of (21). Observe that for each element $e \in U$, we either have $\sum_{S : e \in S} x_S \ge 1/2$ or $\sum_{S : e \in S} r_{A,S} \ge 1/2$ for every scenario A that contains e. Now, define:

$$E = \left\{ e \in U : \sum_{S : e \in S} x_S \ge \frac{1}{2} \right\}$$

Then, $\{2x_S\}_S$ is a fractional solution to the deterministic set cover problem whose universe is E. By our assumption, we can convert the fractional solution into an integral solution $\{\bar{x}_S\}_S$ whose value is at most $2\rho\sum_{S\in\mathcal{S}} w_S^I x_S$. Similarly, for each scenario A, $\{2r_{A,S}\}_S$ is a fractional solution to the deterministic set cover problem whose universe is $A\backslash E$. Thus, we can convert the fractional solution into an integral solution $\{\bar{r}_{A,S}\}_S$ whose value is at most $2\rho\sum_{S\in\mathcal{S}} w_S^{II} r_{A,S}$. It follows that the value of the integral solution $(\bar{x},\bar{r})$ is given by:

$$\sum_{S\in\mathcal{S}} w_S^I \bar{x}_S + \varphi_\alpha\left(\sum_{S\in\mathcal{S}} w_S^{II}\bar{r}_{A,S}\right) \le 2\rho\sum_{S\in\mathcal{S}} w_S^I x_S + \varphi_\alpha\left(2\rho\sum_{S\in\mathcal{S}} w_S^{II} r_{A,S}\right) \tag{22}$$

$$= 2\rho\left[\sum_{S\in\mathcal{S}} w_S^I x_S + \varphi_\alpha\left(\sum_{S\in\mathcal{S}} w_S^{II} r_{A,S}\right)\right] \tag{23}$$

where (22) (resp. (23)) follows from the monotonicity (resp. positive homogeneity) of φ_α. This completes the proof. □

Together with known results in the literature, Theorem 4.1 immediately yields new approximation algorithms for various risk–adjusted 2–stage stochastic covering problems. For instance,

- using the standard LP rounding algorithm for vertex cover [16] (see also [26, Chapter 14]), we obtain a 4–approximation algorithm for the risk–adjusted 2–stage stochastic vertex cover problem;
- using the greedy set cover algorithm of Chvátal [4] (see also [26, Chapter 13]), we obtain an $O(\log n)$–approximation algorithm for the risk–adjusted 2–stage stochastic set cover problem.

4.2 Facility Location Problem Recall that in the usual (deterministic) facility location problem, we are given a set of facilities F and a set of clients D. Each facility $i\in F$ has an opening cost of $f(i)\ge 0$. Each client in D must be assigned to an opened facility, and the cost of assigning client $j\in D$ to the opened facility $i\in F$ is $c_{ij}\ge 0$. We assume that the assignment costs $\{c_{ij}\}_{i\in F,j\in D}$ satisfy the triangle inequality. The goal is then to open a set of facilities and assign the clients to those facilities so that the total cost (i.e. the sum of opening and assignment costs) is minimized.

In the 2–stage stochastic facility location problem, the set of clients to be assigned is not known in advance. Instead, we have scenarios $A_1,\ldots,A_N$, where each scenario A specifies a subset $D_A\subseteq D$ of clients to be assigned. Moreover, the opening cost of a facility depends on the stage. Specifically, facility $i\in F$ has a first–stage opening cost of $f_0(i)\ge 0$ and a scenario–dependent second–stage opening cost of $f_A(i)\ge 0$. Note that the assignment of clients to opened facilities occurs in the second stage, and hence the assignment costs are relevant only in the second stage. We assume that the assignment costs are the same across the scenarios, and that they satisfy the triangle inequality. Now, our goal is to minimize the total opening and assignment costs w.r.t. the risk measure φ_α, where $\alpha\in[0,1)$ is the risk aversion level. Specifically, we are interested in the following optimization problem:

$$\begin{array}{lll} \text{minimize} & \displaystyle\sum_{i\in F} f_0(i)x_i + \beta + \frac{1}{N(1-\alpha)}\sum_{k=1}^{N}(q(x,A_k)-\beta)^+ & \\ \text{subject to} & x_i\in\{0,1\} & \forall\, i\in F \\ & \beta\ge 0 & \end{array}$$

where:

$$\begin{array}{rclll} q(x,A) & = & \text{minimize} & \displaystyle\sum_{i\in F} f_A(i)r_{A,i} + \sum_{i\in F}\sum_{j\in D_A} c_{ij}y_{A,ij} & \\ & & \text{subject to} & \displaystyle\sum_{i\in F} y_{A,ij}\ge 1 & \forall\, j\in D_A \\ & & & y_{A,ij}\le x_i + r_{A,i} & \forall\, i\in F,\ j\in D_A \\ & & & r_{A,i},\ y_{A,ij}\in\{0,1\} & \forall\, i\in F,\ j\in D_A \end{array}$$

Here, the variable x_i indicates whether the facility $i\in F$ is opened in the first stage, and the variable $r_{A,i}$ indicates whether $i\in F$ is opened in the second stage when scenario A is realized. The variable $y_{A,ij}$ indicates whether client $j\in D$ is assigned to facility $i\in F$ when scenario A is realized.

It is clear that the LP relaxation, which is obtained by replacing the binary constraints with non–negativity constraints, can be solved in polynomial time. Let $(x^*, r^*_{A_1}, \ldots, r^*_{A_N}, y^*_{A_1}, \ldots, y^*_{A_N})$ be an optimal fractional solution to the LP relaxation. Upon applying the rounding algorithm of Shmoys et al. [22], we can get a feasible integral solution $(\bar{x}, \bar{r}_{A_1}, \ldots, \bar{r}_{A_N}, \bar{y}_{A_1}, \ldots, \bar{y}_{A_N})$ such that:

$$\sum_{i \in F} f_0(i)\bar{x}_i \leq \frac{2}{t} \sum_{i \in F} f_0(i)x^*_i$$

$$\sum_{i \in F} f_A(i)\bar{r}_{A,i} \leq \frac{2}{t} \sum_{i \in F} f_A(i)r^*_{A,i} \qquad \forall\, A$$

$$\sum_{i \in F} c_{ij}\bar{y}_{A,ij} \leq \frac{6}{1-t} \sum_{i \in F} c_{ij}y^*_{A,ij} \qquad \forall\, j \in D_A,\, A$$

for some $t \in (0,1)$. By choosing $t = 1/4$, we know that the first–stage cost given by the solution $\bar{x}$ is at most $8\sum_{i \in F} f_0(i)x^*_i$. Moreover, for each scenario A, we have:

$$q(\bar{x}, A) = \sum_{i \in F} f_A(i)\bar{r}_{A,i} + \sum_{i \in F} \sum_{j \in D_A} c_{ij}\bar{y}_{A,ij} \leq 8q(x^*, A)$$

This, together with the monotonicity and positive homogeneity of φ_α, implies the following theorem:

Theorem 4.2 *There exists an 8–approximation algorithm for the risk–adjusted 2–stage stochastic facility location problem.*

4.3 Steiner Tree Problem Let $G = (V, E)$ be an undirected graph with a given root vertex $r \in V$. Recall that in the usual (deterministic) Steiner tree problem, we are given a non–negative weight function $c : E \to \mathbb{R}_+$ on the edges and a set of terminals $S \subset V$ with $r \in S$, and the goal is to find a minimum–weight subgraph of G that spans S. In the 2–stage stochastic Steiner tree problem, the set of terminals is known only in the second stage. Specifically, each scenario A in the second stage specifies a set of terminals $S_A \subset V$. Without loss of generality, we assume that $r \in S_A$. Furthermore, each edge $e \in E$ has a first–stage weight $c_0(e) \geq 0$ and a scenario–dependent second–stage weight $c_A(e) \geq 0$. Now, a solution to the 2–stage stochastic Steiner tree problem consists of a set E_0 of edges to be selected in the first stage, and for each scenario A, a set E_A of edges to be selected in the second stage, so that the subgraph induced by the edges in $E_0 \cup E_A$ spans S_A. Our goal is to minimize the total weight of the subgraph w.r.t. the risk measure φ_α, where $\alpha \in [0,1)$ is the risk aversion level. In other words, we are interested in the following optimization problem:

$$\text{minimize} \sum_{e \in E_0} c_0(e) + \varphi_\alpha\left[\sum_{e \in E_A} c_A(e)\right] \quad \text{subject to } E_0 \cup E_A \text{ spans } S_A \text{ for every scenario } A$$

In the sequel, we shall write $c_0(E') \equiv \sum_{e \in E'} c_0(e)$ and $c_A(E') \equiv \sum_{e \in E'} c_A(e)$ for any $E' \subset E$ and scenario A. Now, using the properties of φ_α, we can prove the following structural lemma, which generalizes the corresponding results of Gupta et al. [11, Lemma 3.1] and Dhamdhere et al. [6, Lemma 4.1] to the risk–adjusted setting.

Lemma 4.1 *The following hold for the risk–adjusted 2–stage stochastic Steiner tree problem.*

(a) *There exists a first–stage solution $\tilde{E}_0 \subset E$ and a subset $\mathcal{A} \subset \{A_1, \ldots, A_N\}$ of scenarios such that $\tilde{E}_0$ is a minimal feasible solution to the scenarios in $\mathcal{A}$.*

(b) *The first–stage solution $\tilde{E}_0 \subset E$ can be extended to a solution to the remaining scenarios in the second stage such that the value of the final solution (w.r.t. φ_α) is at most twice the optimal.*

Proof. Let E^*_0 be the optimal integral first–stage solution, and let E^*_i be the optimal integral second–stage solution when scenario A_i is realized, where $i = 1, \ldots, N$. Using the procedure in Lemma 4.1 of Dhamdhere et al. [6], we can construct a first–stage solution $\tilde{E}_0 \subset E$ such that (i) $c_0(\tilde{E}_0) \leq 2c_0(E^*_0)$, and (ii) $\tilde{E}_0$ is a minimal feasible solution to a subset of scenarios. Moreover, we also obtain second–stage solutions $(\tilde{E}_1, \ldots, \tilde{E}_N)$ such that $c_{A_i}(\tilde{E}_i) \le 2c_{A_i}(E_i^*)$ for $i = 1, \ldots, N$. Thus, using the monotonicity and positive homogeneity of φ_α, we conclude that:

$$c_0(\tilde{E}_0) + \varphi_\alpha\left(c_{A_i}(\tilde{E}_i)\right) \le 2\big[c_0(E_0^*) + \varphi_\alpha(c_{A_i}(E_i^*))\big]$$

as desired. □

As in [11, 6], the upshot of Lemma 4.1 is that it implies the existence of a near–optimal solution $(\tilde{E}_0, \tilde{E}_1, \ldots, \tilde{E}_N)$ to the risk–adjusted 2–stage stochastic Steiner tree problem in which the first–stage solution $\tilde{E}_0$ forms a tree containing the root vertex r. In particular, this implies that in the near–optimal solution, the path from any terminal to the root will consist of a portion of second–stage edges, followed by a portion of first–stage edges. Consequently, we can use the following flow–based integer programming formulation to find that near–optimal solution (cf. [11, 6]):

$$\begin{aligned}
\text{minimize} \quad & \sum_{e\in E} c_0(e)x_e + \beta + \frac{1}{N(1-\alpha)}\sum_{i=1}^{N}\left(\sum_{e\in E} c_{A_i}(e) r_{A_i,e} - \beta\right)^+ \\
\text{subject to} \quad & \sum_{e\in\delta_+(t)} (y_{0,e}(t) + y_{A,e}(t)) \ge 1 && \forall\, t \in S_A,\ \forall\, A && (24)\\
& \sum_{e\in\delta_+(v)} (y_{0,e}(t) + y_{A,e}(t)) = \sum_{e\in\delta_-(v)} (y_{0,e}(t) + y_{A,e}(t)) && \forall\, v \notin \{r,t\},\ \forall\, t\in S_A,\ \forall\, A && (25)\\
& \sum_{e\in\delta_-(v)} y_{0,e}(t) \le \sum_{e\in\delta_+(v)} y_{0,e}(t) && \forall\, v \notin \{r,t\},\ \forall\, t\in S_A,\ \forall\, A && (26)\\
& y_{0,e}(t) \le x_e,\ \ y_{A,e}(t) \le r_{A,e} && \forall\, e\in E,\ \forall\, t\in S_A,\ \forall\, A && (27)\\
& x_e,\ r_{A,e},\ y_{0,e}(t),\ y_{A,e}(t) \in \{0,1\} && \forall\, e\in E,\ \forall\, t\in S_A,\ \forall\, A && (28)\\
& \beta \ge 0
\end{aligned}$$

Here, the variable x_e indicates whether the edge $e \in E$ is selected in the first stage, and the variable $r_{A,e}$ indicates whether $e \in E$ is selected in the second stage when scenario A is realized. The variable $y_{0,e}(t)$ indicates whether the edge $e \in E$ is selected in the first stage as part of terminal t's path to the root. Similarly, for each terminal $t \in S_A$ in scenario A, the variable $y_{A,e}(t)$ indicates whether the edge $e \in E$ is selected in the second stage as part of t's path to the root. The variables $y_{0,e}(t)$ and $y_{A,e}(t)$ are directed in the sense that for $e = (u,v) \in E$, the variable $y_{A,(u,v)}(t)$ denotes t's flow along a second–stage edge in the direction of u to v. Given these directed flow variables, we can define the cut sets $\delta_+(S)$ and $\delta_-(S)$ for any $S \subset V$ as follows:

$$\begin{aligned}
\delta_+(S) &= \{e = (u,v) \in E : u \in S,\ v \notin S\} \\
\delta_-(S) &= \delta_+(V \backslash S)
\end{aligned}$$

We remark that the variables $\{x_e\}_e$ and $\{r_{A,e}\}_{A,e}$ are undirected, and the graph G itself remains undirected.

Now, constraint (24) ensures that there is one unit of flow leaving each terminal $t \in S_A$ in each scenario A, and constraint (25) imposes flow conservation at each non–terminal vertex $v \in V$. As argued earlier, in the near–optimal solution, the path from any terminal to the root will consist of a portion of second–stage edges, followed by a portion of first–stage edges. Thus, at each non–terminal vertex $v \in V$, the net first–stage inflow corresponding to each terminal $t \in S_A$ must be bounded above by the net first–stage outflow. This is enforced by the constraint (26). Finally, constraint (27) ensures that edges with flow are indeed paid for in the objective function.

By Lemma 4.1, the optimal value of the above integer program is at most twice the value of the optimal solution to the risk–adjusted 2–stage stochastic Steiner tree problem. Now, by relaxing the binary constraints (28) in the straightforward manner, we obtain a linear program that can be solved in polynomial time. Let $(x^*, r^*_{A_1}, \ldots, r^*_{A_N})$ be an optimal fractional solution to the LP relaxation. Using the rounding algorithm of Gupta et al. [11], we can get an integral solution $(\bar{E}_0, \bar{E}_1, \ldots, \bar{E}_N)$ with the properties that (i) the subgraph induced by the edges in $\bar{E}_0 \cup \bar{E}_i$ spans S_{A_i} for $i = 1, \ldots, N$, and (ii) we have:

$$c_0(\bar{E}_0) \le 20 \sum_{e \in E} c_0(e) x_e^*$$

$$c_{A_i}(\bar{E}_i) \le 20 \sum_{e \in E} c_{A_i}(e) r_{A_i,e}^* \qquad \forall\, i = 1, \ldots, N$$

Thus, upon using Lemma 4.1 and the monotonicity and positive homogeneity of φ_α, we obtain the following theorem:

Theorem 4.3 *There exists a* 40*–approximation algorithm for the risk–adjusted* 2*–stage stochastic Steiner tree problem.*

5. Conclusion and Future Work In this paper we have motivated the use of a risk measure to capture robustness in stochastic combinatorial optimization problems. By generalizing the sampling theorem in [3], we have shown that the risk–adjusted objective can be efficiently treated by the SAA method. Furthermore, we have exhibited approximation algorithms for various stochastic combinatorial optimization problems under the risk–adjusted setting. Our work opens up several interesting directions for future research. For instance, it would be interesting to develop approximation algorithms for other stochastic combinatorial optimization problems under the risk–adjusted setting. Also, there are other risk measures that can be used to capture robustness (see, e.g., [20]). Can theorems similar to those established in this paper be proven for those risk measures? Finally, it would be worthwhile to study multistage versions of the risk–adjusted stochastic optimization problems considered in this paper. One immediate research problem would be to extend our sampling theorem to the multistage setting (cf. [25]).

Acknowledgment. Part of this work was done when the first author was a doctoral student in the Computer Science Department at Stanford University. Research by the first and third author was supported in part by the Boeing Company. A preliminary version of this paper [24] has appeared in the Proceedings of the 9th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX 2006). The authors would like to thank the anonymous referee for the valuable comments.

References

[1] E. M. L. Beale, *On Minimizing a Convex Function Subject to Linear Inequalities*, Journal of the Royal Statistical Society, Series B (Methodological) **17** (1955), no. 2, 173–184.

[2] Aharon Ben-Tal and Marc Teboulle, *An Old–New Concept of Convex Risk Measures: The Optimized Certainty Equivalent*, Mathematical Finance **17** (2007), no. 3, 449–476.

[3] Moses Charikar, Chandra Chekuri, and Martin Pál, *Sampling Bounds for Stochastic Optimization*, Proceedings of the 9th International Workshop on Randomization and Computation (RANDOM 2005) (Chandra Chekuri, Klaus Jansen, José D. P. Rolim, and Luca Trevisan, eds.), Lecture Notes in Computer Science, vol. 3624, 2005, pp. 257–269.

[4] V. Chvátal, *A Greedy Heuristic for the Set–Covering Problem*, Mathematics of Operations Research **4** (1979), no. 3, 233–235.

[5] George B. Dantzig, *Linear Programming under Uncertainty*, Management Science **1** (1955), no. 3/4, 197–206.

[6] Kedar Dhamdhere, Vineet Goyal, R. Ravi, and Mohit Singh, *How to Pay, Come What May: Approximation Algorithms for Demand–Robust Covering Problems*, Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2005), 2005, pp. 367–378.

[7] Shane Dye, Leen Stougie, and Asgeir Tomasgard, *The Stochastic Single Resource Service–Provision Problem*, Naval Research Logistics **50** (2003), no. 8, 869–887.

[8] Uriel Feige, Kamal Jain, Mohammad Mahdian, and Vahab Mirrokni, *Robust Combinatorial Optimization with Exponential Scenarios*, Proceedings of the 12th International Conference on Integer Programming and Combinatorial Optimization (IPCO XII) (Matteo Fischetti and David P. Williamson, eds.), Lecture Notes in Computer Science, vol. 4513, 2007, pp. 439–453.

[9] Martin Grötschel, László Lovász, and Alexander Schrijver, *Geometric Algorithms and Combinatorial Optimization*, second corrected ed., Algorithms and Combinatorics, vol. 2, Springer–Verlag, Berlin/Heidelberg, 1993.

[10] Anupam Gupta, Martin Pál, R. Ravi, and Amitabh Sinha, *Boosted Sampling: Approximation Algorithms for Stochastic Optimization*, Proceedings of the 36th Annual ACM Symposium on Theory of Computing (STOC 2004), 2004, pp. 417–426.

[11] Anupam Gupta, R. Ravi, and Amitabh Sinha, *LP Rounding Approximation Algorithms for Stochastic Network Design*, Mathematics of Operations Research **32** (2007), no. 2, 345–364.

[12] Wassily Hoeffding, *Probability Inequalities for Sums of Bounded Random Variables*, Journal of the American Statistical Association **58** (1963), no. 301, 13–30.

[13] Nicole Immorlica, David Karger, Maria Minkoff, and Vahab Mirrokni, *On the Costs and Benefits of Procrastination: Approximation Algorithms for Stochastic Combinatorial Optimization Problems*, Proceedings of the 15th Annual ACM–SIAM Symposium on Discrete Algorithms (SODA 2004), 2004, pp. 691–700.

[14] Anton J. Kleywegt, Alexander Shapiro, and Tito Homem-De-Mello, *The Sample Average Approximation Method for Stochastic Discrete Optimization*, SIAM Journal on Optimization **12** (2001), no. 2, 479–502.

[15] Peter D. Lax, *Functional Analysis*, Pure and Applied Mathematics: A Wiley–Interscience Series of Texts, Monographs, and Tracts, John Wiley & Sons, Inc., New York, 2002.

[16] G. L. Nemhauser and L. E. Trotter, Jr., *Properties of Vertex Packing and Independence System Polyhedra*, Mathematical Programming **6** (1974), 48–61.

[17] Yurii Nesterov and Arkadii Nemirovskii, *Interior–Point Polynomial Algorithms in Convex Programming*, SIAM Studies in Applied Mathematics, vol. 13, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 1994.

[18] R. Ravi and Amitabh Sinha, *Hedging Uncertainty: Approximation Algorithms for Stochastic Optimization Problems*, Mathematical Programming, Series A **108** (2006), no. 1, 97–114.

[19] R. Tyrrell Rockafellar and Stanislav Uryasev, *Conditional Value–at–Risk for General Loss Distributions*, Journal of Banking and Finance **26** (2002), no. 7, 1443–1471.

[20] Andrzej Ruszczyński and Alexander Shapiro, *Optimization of Risk Measures*, Probabilistic and Randomized Methods for Design under Uncertainty (Giuseppe Calafiore and Fabrizio Dabbene, eds.), Springer–Verlag, London, 2005, pp. 119–157.

[21] Alexander Shapiro and Arkadi Nemirovski, *On Complexity of Stochastic Programming Problems*, Continuous Optimization: Current Trends and Modern Applications (Vaithilingam Jeyakumar and Alexander Rubinov, eds.), Applied Optimization, vol. 99, Springer Science+Business Media Inc., New York, 2005, pp. 111–146.

[22] David B. Shmoys, Éva Tardos, and Karen Aardal, *Approximation Algorithms for Facility Location Problems*, Proceedings of the 29th Annual ACM Symposium on Theory of Computing (STOC 1997), 1997, pp. 265–274.

[23] David B. Shmoys and Chaitanya Swamy, *An Approximation Scheme for Stochastic Linear Programming and Its Application to Stochastic Integer Programs*, Journal of the ACM **53** (2006), no. 6, 978–1012.

[24] Anthony Man-Cho So, Jiawei Zhang, and Yinyu Ye, *Stochastic Combinatorial Optimization with Controllable Risk Aversion Level*, Proceedings of the 9th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX 2006) (Josep Díaz, Klaus Jansen, José D. P. Rolim, and Uri Zwick, eds.), Lecture Notes in Computer Science, vol. 4110, 2006, pp. 224–235.

[25] Chaitanya Swamy and David B. Shmoys, *Sampling–Based Approximation Algorithms for Multi–Stage Stochastic Optimization*, Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2005), 2005, pp. 357–366.

[26] Vijay V. Vazirani, *Approximation Algorithms*, Springer–Verlag, Berlin/Heidelberg, 2001.