EasyChair Preprint

№ 12439

A Unifying Theory for the Reliability of Stochastic Programming Solutions Using Compromise Decisions

Shuotao Diao and Suvrajeet Sen

March 10, 2024

A Unifying Theory for the Reliability of Stochastic Programming Solutions using Compromise Decisions

Shuotao Diao[1] and Suvrajeet Sen[2]

[1] Department of Industrial Engineering and Management Sciences, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA
[2] Daniel J. Epstein Department of Industrial & Systems Engineering, University of Southern California, 3715 McClintock Ave, Los Angeles, CA 90089, USA
shuotao.diao@northwestern.edu
s.sen@usc.edu

Abstract. This paper studies the reliability of stochastic programming solutions using compromise decisions. A compromise decision is obtained by minimizing the aggregation of objective function approximations across the replications while regularizing the candidate decisions of all the replications, which we refer to as Compromise Decision problem. Rademacher average of families of functions is used to bound the sample complexity of the compromise decisions.

Keywords: Stochastic Programming · Sample Average Approximation · Rademacher Average

1 Introduction

Monte Carlo sampling is known to support applications in which uncertainty may be simulated (e.g., simulated annealing [4], simulation-optimization [6]) with relative ease. However, the introduction of sampling introduces estimation errors, and in the case of optimization, additional errors in decision-optimization can be introduced due to unreliable objective function (especially gradient/subgradient) estimates.

Importance sampling is one popular approach to reduce the variance of estimate by Monte Carlo sampling. It has been widely used to reduce stochastic gradient [13, 31] and portfolio credit risk [9]. Kozimík and Morton [16] also propose an importance sampling methodology to reduce the variance of the upper bound estimate of optimal cost from the Stochastic Dual Dynamic Programming (SDDP) algorithms in the risk-averse setting. Carson and Maria summarized that the key success of importance sampling lies in the appropriate change of probability measure for rare event simulation.

Additionally, other popular variance reduction methods include linear control random variables method ([25]), in which a correlated random variable with mean zero is added to the objective function, and common random numbers variance reduction approach ([15]). For a review of modern variance reduction technique in stochastic optimization, we refer the readers to the overview by Homem-de-Mello and Bayraksan [17].

When memory is no longer a bottleneck of the computation ([30]), performing several replications of sampling procedures (possibly in parallel) becomes an alternative to improve objective function estimate and/or optimal solution estimate. Sen and Liu [24] propose a closed-loop methodology, which they refer to as Compromise Decision approach, to aggregate the information of replication of both objective function estimates and decision estimates together. Their computational results of a two-stage SONET Switched Network problem show that compromise decision has a relatively lower validated objective in the minimization problem. Xu and Sen later extends the Compromise Decision approach to solve two-stage stochastic linear programs [24] and multi-stage stochastic linear programs [29]. While the computations reported in these papers have been extremely encouraging, a common theoretical understanding of these procedures have not yet emerged. This paper is intended to present such a theory which is based on a small set of principles which can be used to explain the computational success reported in the above papers.

Indeed, we aim to study the finite-sample complexity of the Compromise Decision approach to solve the following generic stochastic program. Formal definitions will be introduced in the later sections.

$$\min_{x \in X} \; f(x) \triangleq \mathbb{E}_{\tilde{\xi}}[F(x, \tilde{\xi})], \tag{1}$$

where $\tilde{\xi} : \Omega \mapsto \Xi \subset \mathbb{R}^d$ is a random variable defined on a probability space $(\Omega, \Sigma_\Omega, \mathbb{P})$, $X \subset \mathbb{R}^p$ is the feasible region of x, and $F : X \times \Xi \mapsto \mathbb{R}$ is a Carathéodorian function (i.e., continuous in X and measurable for almost every $\xi \in \Xi$).

Given m ($m \geq 2$) replications, we let ξ_i^n, $\hat{f}_n(x; \xi_i^n)$ and $\hat{x}(\xi_i^n)$ denote the sample set with size n, objective function estimate, and decision estimate in the i^{th} replication of using Monte Carlo sampling to solve the problem in (1), respectively. The Compromise Decision problem is formulated as follows:

$$\min_{x \in X} \; \frac{1}{m} \sum_{i=1}^{m} \hat{f}_n(x; \xi_i^n) + \frac{\rho}{2} \| x - \frac{1}{m} \sum_{i=1}^{m} \hat{x}_n(\xi_i^n) \|^2 \tag{2}$$

The addition of the quadratic regularizer to the objective function has been widely used in Stochastic Programming algorithms such as the proximal point method [22], mirror descent method [19], regularized SD [11], and regularized SDDP [1, 10]. The Compromise Decision problem consists of value function approximation aggregation and penalty to the distance to the average decision estimates. Let $\bar{x}_N(\xi^N) = \frac{1}{m} \sum_{i=1}^{m} \hat{x}(\xi_i^n)$ and let $x_N^c(\xi^N)$ denote the optimal solution of (2). It is obvious that if $\bar{x}_N(\xi^N)$ and $x_N^c(\xi^N)$ agree, then both are optimal to (2). Such observation has been transformed into a stopping rule for compromise SD ([24]).

Motivated by the numerical evidence of successful use of the Compromise Decision problem, this paper aims to provide a common mathematical basis for those successes. In particular, we shall address the following concerns for the Compromise Decision problem:

1. What is the sample complexity of the margin of error on the objective function involving replications?
2. How the penalty coefficient ρ interacts with the function estimate and decision estimate?
3. What is the sample complexity of ϵ-optimal solution of (2)? In particular, what is the large deviation bound of the distance between the compromise decision and optimal solution set of (1)?

This paper is organized as follows. In section 2, we review Rademacher complexity and its use in bounding the sample complexity of the objective function estimate and its variance. In section 3, we present the Compromise Decision problem and derive its sample complexity. In section 4, we present the sample complexity analysis of the compromise decision when a Benders' type decomposition algorithm is used to solve each replication of the approximation problem.

1.1 Notations

Let $\tilde{\xi} : \Omega \mapsto \Xi \subset \mathbb{R}^d$ denote a random vector defined on the probability space $(\Omega, \Sigma_\Omega, \mathbb{P})$. We let ξ denote one realization of $\tilde{\xi}$. Let $\tilde{\xi}_1, \tilde{\xi}_2, \ldots, \tilde{\xi}_n$ denote independent and identically distributed (i.i.d.) copies of $\tilde{\xi}$. For $i = 1, 2, \ldots, n$, we let ξ_i denote the realization of $\tilde{\xi}_i$ and let $\xi^n \triangleq \{\xi_1, \xi_2, \ldots, \xi_n\}$ denote the set of realizations of n i.i.d. copies of $\tilde{\xi}$. Let $X \subset \mathbb{R}^p$ be a compact set of decisions, and let $F : X \times \Xi \mapsto \mathbb{R}$ be a Carathéodorain function. We let $\Pr(\cdot)$ denote the probability of an event, and let $\|\cdot\|$ denote the Euclidean norm.

Without further specification, we let n denote the sample size of each replication and let m denote the number of replications. We let ξ_i^n denote the sample set with size n in the i^{th} replication. We let $N = mn$ denote the total sample size and let $\xi^N = \cup_{i=1}^m \xi_i^n$ denote the mega sample set. Without loss of generality, we consider the case in which the sample size for each replication is the same, but it is straightforward to extend the analysis to the heterogeneous sample sizes case.

As for the Rademacher average, we Let $\tilde{\sigma}_1, \tilde{\sigma}_2, \ldots, \tilde{\sigma}_n$ be i.i.d. random variables with $\tilde{\sigma}_i$ for $i = 1, 2, \ldots, n$ being equally likely to be 1 or -1. That is, $\Pr(\tilde{\sigma} = 1) = \frac{1}{2}$ and $\Pr(\tilde{\sigma} = -1) = \frac{1}{2}$. Furthermore, we require that $\tilde{\sigma}_i$ are independent of $\tilde{\xi}$.

To study ϵ-optimality of a solution, we shall define the following metric to measure the distance between two sets (see [7, 8] for more details)

$$\Delta(A, B) \triangleq \sup_{a \in A} \inf_{b \in B} \|a - b\|. \tag{3}$$

In other words, $\Delta(A, B)$ is the largest distance from a point of set A to set B. The following theorem sets up the Lipschitzian behavior of the ϵ-solution set in terms of the metric defined in (3).

Theorem 1. *[7] Assume that X is a nonempty compact convex set and $f : X \mapsto \mathbb{R}$ is a lower semicontinuous convex function. Let the following definitions hold: $D_X = \max_{x,x' \in X} \|x - x'\|$; $\theta^* = \min_{x \in X} f(x)$; $\epsilon' > \epsilon > 0$, $X_\epsilon = \{x \in X : f(x) \leq \theta^* + \epsilon\}$, and $X_{\epsilon'} = \{x \in X : f(x) \leq \theta^* + \epsilon'\}$. Then the following holds:*

$$\Delta(X_{\epsilon'}, X_\epsilon) \leq \frac{\epsilon' - \epsilon}{\epsilon} D_X.$$

Proof. See [7, Theorem 3.11].

The results of Theorem 1 relate the perturbation of the objective function (e.g., estimation error from sampling) to the ϵ-optimal solution set.

2 Background on Sample Complexity of Objective Function Point Estimate and Rademacher Average.

As mentioned by [3], "Rademacher complexity is commonly used to describe the data-dependent complexity of a function class". One key advantage of (empirical) Rademacher average (or complexity) is that it can be measured from a finite sample set (see [3] for i.i.d. cases, and see [18] for non-i.i.d. cases). As a result, it can be used to estimate the finite-sample error of a function class. Radecamher average has been widely used in neural networks ([2]), support vector machine ([27]), and decision trees ([12]).

We begin this section by reviewing the notion of Rademacher average and its use in bounding a sample average approximation of the objective function and a sample variance of the random cost function.

The common finite-sample approximation of (1) is known as the sample average approximation (SAA), and is written as follows.

$$\min_{x \in X} f_n(x; \xi^n) \triangleq \frac{1}{n} \sum_{i=1}^{n} F(x, \xi_i). \tag{4}$$

Since we will deal with multiple replications of sampling, we write the sample set explicitly in the argument of the $f_n(x; \cdot)$ to distinguish different sample sets. A similar writing style will apply sample variance, estimated solutions and compromise decisions.

The variance of the random cost function, $F(x, \tilde{\xi})$, parameterized by the decision x, is defined as

$$\text{Var}[F(x, \tilde{\xi})] \triangleq \mathbb{E}_{\tilde{\xi}} \left[\Big(F(x, \tilde{\xi}) - f(x) \Big)^2 \right].$$

The unbiased estimate of the variance of $F(x, \tilde{\xi})$ is formulated as

$$s_n^2(x; \xi^n) \triangleq \frac{1}{n-1} \sum_{i=1}^{n} [F(x, \xi_i) - \hat{f}_n(x)]^2. \tag{5}$$

Define a compound function $H : X \times Y \times \Xi \mapsto \mathbb{R}$ with $H(x, y, \xi) = (F(x, \xi) - y)^2$. As suggested in [8], the variance of $F(x, \tilde{\xi})$ can be written as a compound function as follows:

$$\mathrm{Var}[F(x, \tilde{\xi})] = \mathbb{E}_{\tilde{\xi}}[H(x, \mathbb{E}_{\tilde{\xi}}[F(x, \tilde{\xi})], \tilde{\xi})] = \mathbb{E}_{\tilde{\xi}}[(F(x, \tilde{\xi}) - \mathbb{E}_{\tilde{\xi}}[F(x, \tilde{\xi})])^2].$$

Furthermore, the sample variance, $s_n^2(x)$, can be rewritten as a compound function as follows.

$$s_n^2(x; \xi^n) = \frac{1}{n-1} \sum_{i=1}^{n} H(x, \frac{1}{n} F(x, \xi_i), \xi_i). \tag{6}$$

Throughout the paper, we make the following assumptions:

A1. $X \subset \mathbb{R}^p$ is a nonempty compact convex set contained in a cube whose edge-length is D.
A2. $F(x, \xi)$ is Hölder continuous in x with constant L_F and $\gamma \in (0, 1]$.
A3. There exists $M_F \in (0, \infty)$ such that $\sup_{x \in X, \xi \in \Xi} |F(x, \xi)| < M_F$. Let $Y \subset \mathbb{R}$ be $Y \triangleq [-M_F, M_F]$.

We note that the assumption of boundedness of the feasible region and the objective function is common in the Stochastic Programming literature [26, 19, 8]. Also, Hölder continuity condition of the objective function is a generalization of its Lipschitzian counterpart. The introduction of the soundness parameters and Hölder continuity-related parameters are later used to bound the Rademacher average of the random cost function and its variance.

The Rademacher average of a function class is defined as follows.

Definition 1 ([8]). *Let $\tilde{\sigma}_1, \tilde{\sigma}_2, \ldots, \tilde{\sigma}_n$ be i.i.d. random variables with $\tilde{\sigma}_i$ for $i = 1, 2, \ldots, n$ being equally likely to be 1 or -1. For a set of points $(\xi_1, \ldots, \xi_n) = \xi^n$ in Ξ and a sequence of functions $\{F(\cdot, \xi_i) : X \mapsto \mathbb{R}\}$, the Rademacher average of a function class is defined by:*

$$R_n(F, \xi^n) \triangleq \mathbb{E}_{\tilde{\sigma}} \left[\sup_{x \in X} \left| \frac{1}{n} \sum_{i=1}^{n} \tilde{\sigma}_i F(x, \xi_i) \right| \right] \tag{7}$$

The upper bound of Rademacher average of the random cost function, $F(x, \xi)$, is given in the following lemma.

Lemma 1 ([8]). *Suppose that assumptions A1 - A3 hold. Then*

$$R_n(F, \xi^n) \leq (L_F D^\gamma d^{\frac{\gamma}{2}} + M_F \sqrt{2(\log 2 + \frac{d}{2\gamma} \log n)})/\sqrt{n}, \tag{8}$$

furthermore, for $\lambda \in (0, \frac{1}{2})$, we have

$$R_n(F, \xi^n) \leq \frac{N_F}{n^\lambda} \tag{9}$$

where $N_F = L_F D^\gamma d^{\frac{\gamma}{2}} + M_F \sqrt{2(\log 2)} + \frac{M_F d^{1/2}}{\sqrt{\gamma(1-2\lambda)e}}$.

The upper bound of the Rademcaher average of the compound function $H(x, y, \xi)$ is given below.

Lemma 2. *Let $\tilde{\sigma}_1, \tilde{\sigma}_2, \ldots, \tilde{\sigma}_n$ be i.i.d. random variables with $\tilde{\sigma}_i$ for $i = 1, 2, \ldots, n$ being equally likely to be 1 or -1. Let the Rademacher average of the set of sequences of H be*

$$R_n(H, \xi^n) = \mathbb{E}_{\tilde{\sigma}} \sup_{x \in X, y \in Y} \left| \frac{1}{n} \sum_{i=1}^{n} \tilde{\sigma}_i H(x, y, \xi_i) \right|.$$

Suppose that assumptions A1 - A3 hold. Let $L_H = 4M_F\sqrt{L_F^2 + 1}$ and $M_H = 4M_F^2$. For any $\lambda \in (0, \frac{1}{2})$, we have the following result:

$$R_n(H, \xi^n) \leq \frac{N_H}{n^\lambda}, \tag{10}$$

where $N_H = L_H D(d+1)^{\frac{1}{2}} + M_H\sqrt{2(\log 2)} + \frac{M_F(d+1)^{1/2}}{\sqrt{(1-2\lambda)e}}$.

Since N_H does not depend on ξ, we observe that

$$R_n(F, \varXi) = \sup_{\xi_1 \in \varXi_1, \ldots, \xi_n \in \varXi_n} R_n(F, \xi^n) \leq \sup_{\xi_1 \in \varXi_1, \ldots, \xi_n \in \varXi_n} \frac{N_H}{n^\lambda} = \frac{N_H}{n^\lambda}.$$

Denote

$$\delta_n^f(\xi^n) = \sup_{x \in X} \left| \frac{1}{n} \sum_{i=1}^{n} F(x, \xi_i) - \mathbb{E}_{\tilde{\xi}}[F(x, \tilde{\xi})] \right| \tag{11}$$

One key property of Rademacher average is that we can use a symmetric argument to bound the estimated error defined in (11). For more details about the symmetric argument, please see [5]. We summarize the bound of $\delta_n^f(\xi^n)$ derived by Ermoliev and Norkin [8] in the following theorem.

Theorem 2. *Suppose that assumptions A1 - A3 hold. Then the following holds:*

1. $$\mathbb{E}[\delta_n^f(\tilde{\xi}^n)] \leq 2R_n(F, \varXi) \leq \frac{2N_F}{n^\lambda},$$
2. $$\Pr\left\{ n^\lambda \delta_n^f(\tilde{\xi}^n) \geq 2N_F + t \right\} \leq \exp\left(-\frac{t^2}{2M_F^2} \right).$$

Similarly, we derive the sample complexity of the sample variance in the next lemma.

Lemma 3. *Denote*

$$\delta_n^h(\xi^n) = \sup_{x \in X} \left| \frac{1}{n} \sum_{i=1}^{n} H(x, \frac{1}{n} \sum_{j=1}^{n} F(x, \xi_j), \xi_i) - \mathbb{E}_{\tilde{\xi}}[H(x, \mathbb{E}_{\tilde{\xi}}[F(x, \tilde{\xi})], \tilde{\xi})] \right|$$

and

$$\hat{\delta}_n(\xi^n) = \sup_{x\in X, y\in Y} \left| \frac{1}{n}\sum_{i=1}^{n} H(x,y,\xi_i) - \mathbb{E}_{\tilde{\xi}}[H(x,y,\tilde{\xi})] \right|$$

Suppose that assumptions A1 - A3 hold. Then the following holds:

1.
$$\delta_n^h(\xi^n) \leq 4M_F\delta_n^f(\xi^n) + \hat{\delta}_n(\xi^n)$$

2.
$$\begin{aligned}\mathbb{E}[\delta_n^h(\tilde{\xi^n})] &\leq 8M_F R_n(F,\Xi) + 2R_n(H,\Xi)\\ &\leq \frac{8M_F N_H + 2N_F}{n^\lambda},\end{aligned}$$

where $\lambda \in (0,\frac{1}{2})$*,* $N_F = L_F D^\gamma d^{\frac{\gamma}{2}} + M_F\sqrt{2(\log 2)} + \frac{M_F d^{1/2}}{\sqrt{\gamma(1-2\lambda)e}}$*, and* $N_H = L_H D(d+1)^{\frac{1}{2}} + M_H\sqrt{2(\log 2)} + \frac{M_F(d+1)^{1/2}}{\sqrt{(1-2\lambda)e}}$.

2.1 Sample Complexity of the Margin of Error

We let $\xi_i^n = \{\xi_{(i-1)n+j}\}_{j=1}^n$ denote the i^{th} sample set with size n. The SAA objective function of the i^{th} replication is

$$f_n(x;\xi_i^n) \triangleq \frac{1}{n}\sum_{j=1}^{n} F(x,\xi_{(i-1)n+j}),\ i=1,2,\ldots,m.$$

$s_n^2(x;\xi_i^n)$ is sample variance of $F(x,\xi)$ in the i^{th} replication:

$$s_n^2(x;\xi_i^n) = \frac{1}{n-1}\sum_{j=1}^{n}[F(x,\xi_{(i-1)n+j}) - f_{n,i}(x)]^2. \tag{12}$$

Let $Z_{1-\frac{\alpha}{2}} = \Phi^{-1}\left(1-\frac{\alpha}{2}\right)$. With m replications, the margin of error or the $1-\alpha$ confidence interval of $f(x)$ is written as:

$$\frac{1}{m}\sqrt{\frac{\sum_{i=1}^{m} s_n^2(x;\xi_i^n)}{n}}Z_{1-\frac{\alpha}{2}} \tag{13}$$

By using Jensen's inequality, the upper bound of the expected margin of error (half-width of the $(1-\alpha)$ confidence interval) is

$$\begin{aligned}\mathbb{E}\left[\frac{1}{m}\sqrt{\frac{\sum_{i=1}^{m} s_n^2(x;\tilde{\xi}_i^n)}{n}}Z_{1-\frac{\alpha}{2}}\right] &\leq \sqrt{\frac{\sigma^2(x)}{mn} + \frac{8M_F N_H + 2N_F}{mn^{(1+\lambda)}} + \frac{4M_F^2}{mn(n-1)}}Z_{1-\frac{\alpha}{2}}\\ &\leq O((mn)^{-\frac{1}{2}}),\end{aligned} \tag{14}$$

where $\lambda \in (0,\frac{1}{2})$. The upper bound given in (14) not only agrees with the common sense, $(O((mn)^{-\frac{1}{2}}))$, but also indicates that the bias terms $(\frac{8M_F N_H+2N_F}{mn^{(1+\lambda)}} + \frac{4M_F^2}{mn(n-1)})$ diminishes faster than the unbiased term $\frac{\sigma^2(x)}{mn}$.

3 Sample Complexity of Compromise Decision Problem

In this section, we shall formulate the Compromise Decision problem and then study the sample complexity of the associated compromise decision. We start with a basic formulation where we can solve each replication of the SAA problem to optimal. In particular, we let $x_n(\xi_i^n)$ denote the optimal solution of the i^{th} replication:

$$x_n(\xi_i^n) \in \arg\min_{x \in X} f_n(x; \xi_i^n),\ i = 1, 2, \dots, m.$$

Given a regularizer $\rho \in (0, \infty)$, the compromise stochastic program can be formulated as follows:

$$\min_{x \in X} \frac{1}{m} \sum_{i=1}^{m} f_n(x; \xi_i^n) + \frac{\rho}{2} \|x - \frac{1}{m} \sum_{i=1}^{m} x_n(\xi_i^n)\|^2. \tag{15}$$

Let $\theta_{N,\rho}(\xi^N)$ and θ^* denote the optimal values of Compromise Decision problem in (15) and true problem in (1), respectively, we shall show that $\mathbb{E}[|\theta_{N,\rho}(\tilde{\xi}^N) - \theta^*|] \leq \frac{4m-2}{mn^\lambda} N_F$, for some problem specific constant N_F.

In practice, we often obtain an ϵ-optimal solution with $\epsilon > 0$. Hence, the more tractable compromise stochastic program is formulated below:

$$\min_{x \in X} \frac{1}{m} \sum_{i=1}^{m} f_n(x; \xi_i^n) + \frac{\rho}{2} \|x - \frac{1}{m} \sum_{i=1}^{m} x_{n,\epsilon}(\xi_i^n)\|^2 \tag{16}$$

where $x_{n,\epsilon}(\xi_i^n)$ is the ϵ-optimal solution of the i^{th} replication of the SAA problem (i.e., $f_n(x_{n,\epsilon}(\xi_i^n); \xi_i^n) \leq \text{minimum}\{f_n(x; \xi_i^n) | x \in X\} + \epsilon$).

We make one extra assumption on the convexity of the objective function below:

A4. $F(x, \xi)$ is convex in $x \in X$ for every $\xi \in \Xi$.

Also, recall that the estimation error of each replication is defined as follows:

$$\delta_n^f(\xi_i^n) = \sup_{x \in X} |f_n(x; \xi_i^n) - f(x)|,\ i = 1, 2, \dots, m.$$

With the symmetric argument (see [5, 8]), we have $\mathbb{E}[\delta_n(\tilde{\xi}_i^n)] \leq 2R_n(F, \Xi)$. In the next theorem, we use this relation to derive the sample complexity of the optimal cost of (15).

Theorem 3. *Suppose that assumptions A1 - A4 hold. Let*

$$\theta_{N,\rho}(\xi^N) = \min_{x \in X} \frac{1}{m} \sum_{i=1}^{m} f_n(x; \xi_i^n) + \frac{\rho}{2} \|x - \frac{1}{m} \sum_{i=1}^{m} x_n(\xi_i^n)\|^2.$$

Then the following holds:

$$\mathbb{E}[|\theta_{N,\rho}(\tilde{\xi}^N) - \theta^*|] \leq \frac{4m-2}{m} R_n(F, \Xi) \leq \frac{4m-2}{mn^\lambda} N_F,$$

where $R_n(F, \Xi)$ is the Rademacher average associated with F and sample size n, and $\lambda \in (0, \frac{1}{2})$, $N_F = L_F D^\gamma d^{\frac{\gamma}{2}} + M_F \sqrt{2(\log 2)} + \frac{M_F d^{1/2}}{\sqrt{\gamma(1-2\lambda)e}}$.

We further define the ϵ-optimal solution set of the compromise problem.

$$\hat{X}_{N,\rho,\epsilon}(\xi^N) = \{x \in X : \frac{1}{m}\sum_{i=1}^{m} f_n(x;\xi_i^n) + \frac{\rho}{2}\|x - \frac{1}{m}\sum_{i=1}^{m} x_{n,\epsilon}(\xi_i^n)\|^2 \leq \theta_{N,\rho,\epsilon} + \epsilon\}, \tag{17}$$

We close the section by giving the finite-sample complexity of the compromise decisions defined in (17).

Theorem 4. *Suppose that assumptions A1 - A4 hold. Let* $\lambda \in (0, \frac{1}{2})$ *and* $N_F = L_F D^\gamma d^{\frac{\gamma}{2}} + M_F\sqrt{2(\log 2)} + \frac{M_F d^{1/2}}{\sqrt{\gamma(1-2\lambda)e}}$. *Then the following hold:*

1.

$$\Delta(\hat{X}_{N,\rho,\epsilon}(\xi^N), X_\epsilon^*) \leq \frac{\frac{2}{m}\sum_{j=1}^{m}\delta_n(\xi_j^n) + \frac{1}{m^2}\sum_{\substack{i,j=1 \\ i\neq j}}^{m}(\delta_n(\xi_i^n) + \delta_n(\xi_j^n))}{\epsilon} D_X + 2\sqrt{\epsilon/\rho}. \tag{18}$$

2.

$$\mathbb{E}[\Delta(\hat{X}_{N,\rho,\epsilon}(\tilde{\xi}^N), X_\epsilon^*)] \leq \frac{D_X N_F}{\epsilon}\frac{8m-4}{mn^\lambda} + 2\sqrt{\epsilon/\rho}.$$

3. Further let $\rho = n$ *and* $C = \frac{\epsilon^{\frac{3}{2}}}{D_X} + \frac{(3m-2)2N_F}{m}$, *then we have*

$$\Pr\left\{\frac{\epsilon n^\lambda}{2D_X}\Delta(\hat{X}_{N,\rho,\epsilon}(\tilde{\xi}^N), X_\epsilon^*) \geq C + t\right\} \leq m\exp\left\{-\frac{m^2t^2}{2M_F(3m-2)^2}\right\}.$$

4 Algorithms for Compromise Decision Problems

In this section, we study the scenario where a proper algorithm is applied to solve each replication of SAA problem. We will provide a framework to merge the computational results from each replication to create a Compromise Decision problem.

Let $\hat{x}_n(\xi^n)$ and $\hat{f}_n(x;\xi^n)$ denote the estimated solution and approximated objective function of $f_n(x;\xi^n)$ output by the algorithm. When the algorithm terminates, we require that the following conditions hold:

C1. There exists $\epsilon_1 \in (0,\infty)$ such that $\hat{x}_n(\xi) \in \arg\min_{x\in X} \hat{f}_n(x;\xi^n)$ and

$$f_n(\hat{x}_n(\xi^n);\xi^n) - \hat{f}_n(\hat{x}_n(\xi^n);\xi^n) \leq \epsilon_1.$$

C2. Given $\epsilon_2 \in (0,\infty)$. $\hat{f}_n(x;\xi^n)$ is a convex piecewise linear approximation (i.e., $\hat{f}_n(x;\xi^n) = \max_{\ell\in L}\{\alpha_\ell + \langle\beta_\ell, x\rangle\}$) of $f_n(x;\xi^n)$ with possibly some errors such that $\hat{f}_n(x;\xi^n) \leq f_n(x;\xi^n) + \epsilon_2$ for all $x \in X$.

In Condition C1, $f_n(\hat{x}_n(\xi^n);\xi^n)$ is the upper bound estimate of the optimal cost and $\hat{f}_n(\hat{x}_n(\xi^n);\xi^n)$ is the lower bound estimate of the optimal cost. Condition C2 ensures that it outputs an inexact outer approximation of the SAA of the objective function. Conditions C1 and C2 altogether ensure that an algorithm yields an $(\epsilon_1 + \epsilon_2)$-optimal solution of the SAA problem. For instance, Kelley's Cutting Plane Methods ([14]) and other Benders' type decomposition-based algorithms ([20, 21, 23, 28]) will satisfy the conditions above.

Here, we discuss a pre-processing step for building Compromise Decision problem. For $i = 1, 2, \dots, m$, we let $\hat{x}_n(\xi_i^n)$ and $\hat{f}_n(\xi_i^n)$ denote the estimated solution of $\min_{x \in X} f_n(x; \xi_i^n)$ and surrogate function of $f_n(x; \xi_i^n)$. In the pre-processing step, we augment $\hat{f}_n(x; \xi_i^n)$ by letting

$$\check{f}_n(x;\xi_i^n) = \max\{\hat{f}_n(x;\xi_i^n), \max_{j=1\dots n}\{f_n(\hat{x}_n(\xi_j^n);\xi_i^n) + \langle f'_{n,\epsilon_2}(\hat{x}_n(\xi_j^n);\xi_i^n), x - \hat{x}_n(\xi_j^n)\rangle\}\},$$

where $f'_{n,\epsilon_2;\xi_i^n}(\hat{x}_n^j)$ is the ϵ_2-subgradient of $f_n(\cdot;\xi_i^n)$ at x.

Now we can set up a Compromise Decision problem below:

$$\min_{x \in X} \frac{1}{m} \sum_{i=1}^{m} \check{f}_n(x;\xi_i^n) + \frac{\rho}{2}\|x - \frac{1}{m}\sum_{j=1}^{m} \hat{x}_n(\xi_j^n)\|^2. \tag{19}$$

Let $\epsilon < \epsilon_1 + \epsilon_2$. We define the ϵ-optimal solution set of problem (19) below:

$$\hat{\theta}_{N,\rho,\epsilon}(\xi^N) = \min_{x \in X} \frac{1}{m} \sum_{i=1}^{m} \check{f}_n(x;\xi_i^n) + \frac{\rho}{2}\|x - \frac{1}{m}\sum_{i=1}^{m} \hat{x}_n(\xi_i^n)\|^2,$$

$$\check{X}_{N,\rho,\epsilon}(\xi^N) = \{x \in X : \frac{1}{m}\sum_{i=1}^{m} f_n(x;\xi_i^n) + \frac{\rho}{2}\|x - \frac{1}{m}\sum_{i=1}^{m}\hat{x}_n(\xi_i^n)\|^2 \le \hat{\theta}_{N,\rho,\epsilon}(\xi^N) + \epsilon\}.$$

Finally, we derive the finite-sample complexity of the compromise decisions in (19) below.

Theorem 5. *Suppose that assumptions A1 - A4 and conditions C1 - C2 hold. Let* $\lambda \in (0, \frac{1}{2})$ *and* $N_F = L_F D^\gamma d^{\frac{\gamma}{2}} + M_F\sqrt{2(\log 2)} + \frac{M_F d^{1/2}}{\sqrt{\gamma(1-2\lambda)e}}$. *Then the following holds:*

1.
$$\mathbb{E}[\varDelta(\check{X}_{N,\rho,\epsilon}(\tilde{\xi}^N), X_\epsilon^*)] \le \frac{D_X N_F}{\epsilon}\frac{8m-4}{mn^\lambda} + \frac{\epsilon_1 + \epsilon_2 - \epsilon}{\epsilon}2D_X + \frac{2m(m-1)\epsilon_2}{m^2\epsilon}D_X + 2\sqrt{\epsilon/\rho}.$$

2. *Furthermore, pick* $\epsilon = \epsilon_1$, $\epsilon_2 = \frac{1}{\sqrt{n}}$, $\rho = n$, *and* $C = \frac{(2m-1)}{m} + \frac{\epsilon^{\frac{3}{2}}}{D_X} + \frac{(3m-2)2N_F}{m}$, *then*
$$\Pr\left\{\frac{\epsilon n^\lambda}{2D_X}\varDelta(\check{X}_{N,\rho,\epsilon}(\tilde{\xi}^N), X_\epsilon^*) \ge C + t\right\} \le m\exp\left\{-\frac{m^2t^2}{2M_F(3m-2)^2}\right\}.$$

References

1. Asamov, T., Powell, W.B.: Regularized decomposition of high-dimensional multistage stochastic programs with markov uncertainty. SIAM Journal on Optimization **28**(1), 575–595 (2018)

2. Bartlett, P.L., Foster, D.J., Telgarsky, M.J.: Spectrally-normalized margin bounds for neural networks. Advances in neural information processing systems **30** (2017)
3. Bartlett, P.L., Mendelson, S.: Rademacher and gaussian complexities: Risk bounds and structural results. Journal of Machine Learning Research **3**(Nov), 463–482 (2002)
4. Bertsimas, D., Tsitsiklis, J.: Simulated annealing. Statistical science **8**(1), 10–15 (1993)
5. Boucheron, S., Bousquet, O., Lugosi, G.: Theory of classification: A survey of some recent advances. ESAIM: probability and statistics **9**, 323–375 (2005)
6. Carson, Y., Maria, A.: Simulation optimization: methods and applications. In: Proceedings of the 29th conference on Winter simulation. pp. 118–126 (1997)
7. Ermoliev, Y.M., Norkin, V.I.: Normalized convergence in stochastic optimization. Annals of Operations Research **30**(1), 187–198 (1991)
8. Ermoliev, Y.M., Norkin, V.I.: Sample average approximation method for compound stochastic optimization problems. SIAM Journal on Optimization **23**(4), 2231–2263 (2013)
9. Glasserman, P., Li, J.: Importance sampling for portfolio credit risk. Management science **51**(11), 1643–1656 (2005)
10. Guigues, V., Lejeune, M.A., Tekaya, W.: Regularized stochastic dual dynamic programming for convex nonlinear optimization problems. Optimization and Engineering **21**, 1133–1165 (2020)
11. Higle, J.L., Sen, S.: Finite master programs in regularized stochastic decomposition. Mathematical Programming **67**(1), 143–168 (1994)
12. Kääriäinen, M., Elomaa, T.: Rademacher penalization over decision tree prunings. In: European Conference on Machine Learning. pp. 193–204. Springer (2003)
13. Kawai, R.: Optimizing adaptive importance sampling by stochastic approximation. SIAM Journal on Scientific Computing **40**(4), A2774–A2800 (2018)
14. Kelley, Jr, J.E.: The cutting-plane method for solving convex programs. Journal of the society for Industrial and Applied Mathematics **8**(4), 703–712 (1960)
15. Kleinman, N.L., Spall, J.C., Naiman, D.Q.: Simulation-based optimization with stochastic approximation using common random numbers. Management Science **45**(11), 1570–1578 (1999)
16. Kozmík, V., Morton, D.P.: Evaluating policies in risk-averse multi-stage stochastic programming. Mathematical Programming **152**, 275–300 (2015)
17. Homem-de Mello, T., Bayraksan, G.: Stochastic constraints and variance reduction techniques. In: Handbook of simulation optimization, pp. 245–276. Springer (2014)
18. Mohri, M., Rostamizadeh, A.: Rademacher complexity bounds for non-iid processes. Advances in Neural Information Processing Systems **21** (2008)
19. Nemirovski, A., Juditsky, A., Lan, G., Shapiro, A.: Robust stochastic approximation approach to stochastic programming. SIAM Journal on optimization **19**(4), 1574–1609 (2009)
20. Oliveira, W., Sagastizábal, C., Scheimberg, S.: Inexact bundle methods for two-stage stochastic programming. SIAM Journal on Optimization **21**(2), 517–544 (2011)
21. Philpott, A.B., Guan, Z.: On the convergence of stochastic dual dynamic programming and related methods. Operations Research Letters **36**(4), 450–455 (2008)
22. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM journal on control and optimization **14**(5), 877–898 (1976)
23. Ruszczyński, A.: A regularized decomposition method for minimizing a sum of polyhedral functions. Mathematical programming **35**, 309–333 (1986)

24. Sen, S., Liu, Y.: Mitigating uncertainty via compromise decisions in two-stage stochastic linear programming: Variance reduction. Operations Research **64**(6), 1422–1437 (2016)
25. Shapiro, A.: Monte carlo sampling methods. Handbooks in operations research and management science **10**, 353–425 (2003)
26. Shapiro, A., Nemirovski, A.: On complexity of stochastic programming problems. Continuous optimization: Current trends and modern applications pp. 111–146 (2005)
27. Sun, S.: Multi-view laplacian support vector machines. In: Advanced Data Mining and Applications: 7th International Conference, ADMA 2011, Beijing, China, December 17-19, 2011, Proceedings, Part II 7. pp. 209–222. Springer (2011)
28. Van Slyke, R.M., Wets, R.: L-shaped linear programs with applications to optimal control and stochastic programming. SIAM journal on applied mathematics **17**(4), 638–663 (1969)
29. Xu, J., Sen, S.: Compromise policy for multi-stage stochastic linear programming: Variance and bias reduction. Computers & Operations Research **153**, 106132 (2023)
30. Zhang, H., Chen, G., Ooi, B.C., Tan, K.L., Zhang, M.: In-memory big data management and processing: A survey. IEEE Transactions on Knowledge and Data Engineering **27**(7), 1920–1948 (2015)
31. Zhao, P., Zhang, T.: Stochastic optimization with importance sampling for regularized loss minimization. In: international conference on machine learning. pp. 1–9. PMLR (2015)