STOCHASTIC OPTIMIZATION

ADDIS ABABA UNIVERSITY
COLLEGE OF COMPUTATIONAL AND NATURAL SCIENCE
DEPARTMENT OF MATHEMATICS

A project Submitted in partial fulfillment of the requirement of the degree of master of science in mathematics

By: Solomon Zerfu
Stream: Optimization
Advisor: Semu Mitiku(PhD)

February, 2014

ADDIS ABABA UNIVERSITY
DEPARTMENT OF MATHEMATICS

The undersigned hereby certify that they have read and recommend to the department of mathematics for acceptance of this project entitled "**Stochastic Optimization**" by **Solomon Zerfu** in partial fulfillment of the requirements for the degree of Master of Science in mathematics.

Advisor: Dr. Semu Mitiku Kassa

Signature:_____

Date_____
Examiner 1: Dr. Birhanu Guta

Signature:_____

Date_____
Examiner 2: Dr. Tadesse Abdi

Signature:_____

Date_____

ADDIS ABABA UNIVERSITY
DEPARTMENT OF MATHEMATICS

By:SOLOMON ZERFU

Title:STOCHASTIC OPTIMIZATION

Department: Mathematics

Degree: M.Sc.

Convocation: February

Year: 2014

Solomon Zerfu:
Signature:_____

Date_____

Acknowledgment

First I would like to thank my advisor, **Dr. Semu Mitiku**, to whom I am very grateful for all his support throughout this entire process. He has shown great patience and always has been helpful, giving many hours of his time and hard work in assisting me in finishing this project. Without his support, expertise and guidance, this project's completion would not have been possible. I am also very thankful to my friends Ashenafi Teklay Woldemariam and Abebe Taye for their comment and supporting different material. I would also like to express my appreciation to **my families** specially Tize, Baby and Bireh for their continuous encouragement.
I would like to extend my thanks to the Mathematics Department of Addis Ababa University for giving me the opportunity to do my graduate work and also for supplying me with the necessary facilities and materials for the study. My special thanks also goes to Dr. Berhanu Bekele head of Mathematics department and Dr. Tilahun Abebawu for their help and encouragement.
Last, but not least, I would like to forward my grateful thanks to **my friends** who comments and helping me in all activities.

Contents

Abstract

Stochastic optimization is a leading approach to model optimization problems in which there is uncertainty in the input data, whether from measurement noise or an inability to know the future. This paper focuses on types of Stochastic optimization such as Stochastic optimization problems with recourse and Chance constrained optimization problems as well as how to change one Stochastic optimization problems to deterministic equivalent form.

Keywords: Probability, Random Variable, Expected value, Measure, Convex, Stochastic Optimization, Recourse, Chance Constrained.

Chapter 1

Preliminary

Definitions of Various Terms. In this section we will define and explain the various terms which are used in the definition of probability.

1.1 Probability

Definition 1.1. ***Trial and Event***. Consider an experiment which, though repeated under essentially identical conditions, does not give unique results but may result in any one of the several possible outcomes. *[20] The experiment is known as a trial and the outcomes are known as events or casts.* For example:

- throwing of a die is a trial and getting l(or 2 or 3, ..., or 6) is an event.
- tossing of a coin is a trial and getting head (H) or tail (T) is an event.
- drawing two cards from a pack of well-shuffled cards is a trial and getting a king and a queen are events. [21]

Generally an event is the occurrence or nonoccurrence of a phenomena. *[12]* ***Exhaustive Events***. *The total number of possible outcomes in any trial is known as exhaustive events or exhaustive cases.* For example :

- in tossing of a coin there are two exhaustive cases, *viz.*, head and tail (the possibility of the coin standing on an edge being ignored).
- in throwing of two dice, the exhaustive number of cases is $6^2 = 36$, since any of the 6 numbers 1 to 6 on the first die can be associated with any of the six numbers on the other die. [21]

Favourable Events or Cases. *The number of cases favourable to an event in a trial is the number of outcomes which entail the happening of the event.* For example,

- in drawing a card from a pack of cards the number of cases favourable to drawing of an ace is 4, for drawing a spade is 13 and for drawing a red card is 26.
- in throwing of two dice, the number of cases favourable to getting the sum 5 is 4 i.e.,$(1,4)(4,1)(2,3)(3,2)$.

Mutually Exclusive Events.
Events are said to be mutually exclusive or incompatible if the happening of anyone of them precludes the happening of all the others (i.e., if no two or more of them can happen simultaneously in the same trial). For example:

- in throwing a die all the 6 faces numbered 1 to 6 are mutually exclusive since if anyone of these faces comes, the possibility of others in the same trial is ruled out.

Equally Likely Events. *Outcomes of a trial are said to be equally likely if taking into consideration all the relevant evidence, there is no reason to expect one in preference to the others.* For example:

- in throwing an unbiased die, all the six faces are equally likely to come. [21]

Independent events. *Several events are said to be independent if the happening (or non-happening) of an event is not affected by the supplementary knowledge concerning the occurrence of any number of the remaining events.* For example:

- in tossing an unbiased coin the event of getting head in the first toss is independent of getting a head, in the second, third, and subsequent throws. [21]

Definition 1.2. *[21]If a trial results in n exhaustive, mutually exclusive and equally likely cases and m of them are favourable to the happening of an event E, then the probability 'p' of happening of E is given by*

$$p = P(E) = \frac{Favourable\ number\ of\ cases}{Exhaustive\ number\ of\ cases} = \frac{m}{n} \tag{1.1}$$

Remark 1.1. *1. Probability 'p' of the happening of an event is also known as the probability of success and the Probability 'q' of the non- happening of the event as the probability of failure. [6]*

2. *If $P(E) = 1$, E is called a certain event and if $P(E) = 0$, E is called an impossible event. [21]*

Definition 1.3. *(Von Mises) [21]If a trial is repeated, a number of times under essentially homogeneous and identical conditions; then the limiting value of the ratio of the number of times the event-happens to the number of trials, as the number of trials become indefinitely large is called the probability of happening of the event. (It is assumed that the limit is finite and unique). Symbolically, if in n trials an event E happens m times, then the probability 'p' of the happening of E is given by*

$$p = P(E) = \lim_{n\to\infty} \frac{m}{n} \tag{1.2}$$

1.2 Random Variable

A function that assigns numbers to outcomes is called a *random variable.* [21] The purpose of such functions in practice is to define a new sample space whose outcomes speak more directly to the objectives of the experiment. [20] For a mathematical and rigorous definition of the random variable, let us consider the probability space, the triplet $(\Omega, \mathcal{A}, P)$, where Ω is the sample space, viz., space of outcomes, $\mathcal{A}$ is the σ-field of subsets in Ω, and P is a probability function on $\mathcal{A}$. [6]

Definition 1.4. *[21]A random variable (r.v.) is a function $X(\omega)$ with domain Ω and range $(-\infty, \infty)$ such that for every real number a,*
the event $[\omega : X(\omega) < a] \in \mathcal{A}$.

Discrete Random Variable.
If a random variable takes at most a countable number of values, it is called a *discrete random variable.* In other words, a real valued function defined on a discrete sample space is called a *discrete random variable.* [21]
Probability Mass Function (and probability distribution of a discrete random variable).
Suppose X is a one-dimensional discrete random variable taking at most a countably infinite number of values $x_1, x_2, ...$ with each possible outcome x_i, we associate a number $P_i = P(X = x_i) = p(x_i)$ called the probability of x_i. The numbers $p(x_i); \quad i = 1, 2, \ldots$ must satisfy the following conditions:
(i) $p(x_i) \geq 0 \quad \forall i$,
(ii) $\sum_{i=1}^{\infty} p(x_i) = 1$.
This function p is called the probability mass function of the random variable X and the set $(x_i, p(x_i))$ is called the *probability distribution* (p.d.) of the

random variable X. [21]
Continuous Random Variable.
A random variable X is said to be continuous if, it can take all possible values between certain limits. [21]

1.3 Expected Values, Variance and Standard Deviation

The expected value of a random variable X denoted $E(X)$, is the long-run average value of the random variable over many repeated trials or occurrences. The expected value of X is also called the expectation of X or the mean of X. [21]
The variance of random variable X is the expected value of the square of the deviation of X from its mean:

$$\begin{aligned} Var(X) &= E(X - E(X))^2 \\ &= \sum_x (x - \mu X)^2 f(x), \text{if} \quad x \quad \text{is discrete.} \\ &= \int_x (x - \mu X)^2 f(x) dx, \text{if} \quad x \quad \text{is continuous.} \end{aligned}$$

where $f(x)$ is probability density function.
The Standard deviation is the square root of the variance. [21]

1.4 Convex Set, Function and Concave

Definition 1.5. *[18]Let S be a subset of a real space.*
(a) The set S is said to be convex *if for all $x, y \in S$*

$$\lambda x + (1 - \lambda) y \in S \quad \forall \lambda \in [0, 1].$$

(b) Let set S be non empty and convex. A functional $f : S \longrightarrow \Re$ is said to be convex *if for all $x, y \in S$*

$$f(\lambda x + (1 - \lambda) y) \leq \lambda f(x) + (1 - \lambda) f(y) \quad \forall \lambda \in [0, 1].$$

(c) Let set S be non empty and convex. A functional $f : S \longrightarrow \Re$ is said to be concave *if the functional $-f$ is convex.*
(d) Let set S be non empty and convex. If every $\alpha \in \Re$ the set $S_\alpha := \{x \in S \mid f(x) \leq \alpha\}$ is convex, then functional f is quasiconvex.

1.5 Lipschitz Continuous

Definition 1.6. *Let S be a non empty subset of a real normed space $(X, \| \ . \ \|)$, let $f : S \longrightarrow \Re$ be a given function, let $x_0 \in S$. Then f is said to be* Lipschitz continuous *at $x_0 \in S$ if there is a constant $\mathcal{K} \geq 0$ and some $\epsilon \geq 0$ such that*

$$| \ f(x) - f(y) \ | \geq \mathcal{K} \ \| \ x - y \ \| \ \forall x, y \in S \cap B(x_0, \epsilon)$$

where $B(x_0, \epsilon) = \{x \in X \ | \| \ x - x_0 \ \| \ \epsilon\}$.

Definition 1.7. *[5]Uncertainty is the unknown future events that cannot be predicted quantitatively within useful limits.*

Uncertainty can be classified as epistemic and aleatoric uncertainties. Epistemic uncertainties are called systemic uncertainties. These are internal (endogenous) uncertainties that reveal themselves, for instance, due to imprecision in the coefficients or constants in the mathematical model of a process.
Aleatoric uncertainties are uncertainties that are caused due external (exogenous) influences.

1.6 Measurable Set, Function and Fatou Lemma

Definition 1.8. *We say a subset S of $\Re$ is Lebesgue measurable if for every subset of A of $\Re$.*

$$\mu A = \mu(A \cap S) + \mu(A \cap S^c).$$

Definition 1.9. *Let $f : X \longrightarrow \Re$ be a function. If $f^{-1}(\mathcal{O})$ is a measurable set for every open subset $\mathcal{O}$ of $\Re$, then f is called a measurable function.*

Definition 1.10. *A non negative measurable function f on Ω is said to be integrable over Ω if*

$$\int_\Omega f < \infty.$$

Lemma 1.1. *(Fatous Lemma) Let $f_n : \Omega \longrightarrow [0, \infty)$, measurable for each $n = 1, 2, ...$*
Suppose $f_n \to f$ a.e on Ω. Then

$$\int_\Omega f \leq \liminf_{n \to \infty} \int_\Omega f_n.$$

1.7 Constrained and Unconstrained Optimization

The online dictionary defines the word optimization as: *"the design and operation of a system or process to make it as good as possible in some defined sense"*. Thus, one may think of optimization as the art or science of determining the best solution to certain mathematically defined problems. An objective function defined by a set of independent decision variables is used to determine the goodness of a solution. Optimization is a mathematical discipline that concerns the finding of minima and maxima of functions, subject to so-called constraints. Optimization problems can be categorized into constrained optimization problems and unconstrained optimization problems.

Definition 1.11. *[4]Constrained optimization problem can be defined as a regular constraint satisfaction problem in which constraints are weighted and the goal is to find a solution maximizing the weight of satisfied constraints. A general constrained optimization problem may be written as follows:*

$$\begin{array}{ll} & max f(x) \\ subject\ to & g_i(x) = c_i \quad for \quad i = 1, \ldots, n \quad Equality\ constraints. \\ & h_j(x) \leq d_j \quad for \quad j = 1, \ldots, m \quad Inequality\ constraints. \end{array} \tag{1.3}$$

where x is a vector residing in an n-dimensional space, $f(x)$ is a scalar valued objective function, $g_i(x) = c_i$ for $i = 1, \ldots, n$ and $h_j(x) \leq d_j$ for $j = 1, \ldots, m$ are constraint functions that need to be satisfied [18].

Definition 1.12. *[4]Unconstrained minimization is the problem of finding a vector x that is a local minimum to a scalar function $f(x)$:*

$$min_x f(x) \tag{1.4}$$

The term unconstrained means that no restriction is placed on the range of x.

Chapter 2

Stochastic Optimization

2.1 Introduction

Definition 2.1. *[14]Stochastic Optimization is the process of maximizing or minimizing the value of a mathematical or statistical function when one or more of the input parameters is subject to randomness.*

The word stochastic means involving chance or probability.
Stochastic processes always involve probability, such as trying to predict the water level in a reservoir at a certain time based on random distribution of rainfall and water usage or estimating the number of dropped connections in a communications network based on random variable traffic but constant bandwidth [8].
In contrast, deterministic processes never involve probability; outcomes occur (or fail to occur) based on predictable and exact input values. Deterministic methods provide a theoretical guarantee of locating the global minimum, or at least a local minimum whose objective function value differs by at worst ϵ from the global one for a given $\epsilon > 0$. Stochastic methods only offer a guarantee in probability. On the other hand, stochastic methods are usually faster in locating a global optimum than deterministic ones. Advanced Stochastic techniques use stochastic methods to search for the location of local minima and then employ deterministic methods to solve a local minimization problem. Stochastic optimization lends itself to real-life situations because many phenomena in the physical world involve uncertainty, impression or randomness [11].
Stochastic Optimization plays a significant role in the analysis, design, and operation of modern system [3].
Stochastic Optimization algorithms have broad application to problems in statistics (e.g., design of experiments and response surface modeling), science,

engineering, and business. Algorithms that employ some form of stochastic optimization have become widely available [11].
Specific applications include:

business (making short- and long-term investment decisions in order to increase profit),

aerospace engineering (running computer simulations to refine the design of a missile or aircraft),

medicine (designing laboratory experiments to extract the maximum information about the efficacy of a new drug),

traffic engineering (setting the timing for the signals in a traffic network). [8]

Stochastic optimization is decision making under risk. This means that some of the model coefficients are random variables with known or estimated distributions whose realizations are revealed after some or all of the decisions have been made [23]. An optimization problem that involves uncertain or random parameters is known as a non-deterministic optimization problem. In general, optimization problems with random variables are stochastic optimization problems and Optimization problems that involve uncertain parameters are known as robust optimization problems [5].

2.2 Expected Value Function

To study about properties of stochastic optimization let we say some thing about properties of the expected value. Let $(\Omega, \mathbb{A}, \mathbb{P})$ be a complete probability space; $f : \Re^n \times \Omega \longrightarrow \Re$ and, for any given $x, f(x, .)$ is an (Lebesgue) integrable function. In general, the objective functions of a stochastic optimization problem is of the form $E[f(x, \xi)]$ [16].
Expected Value:

$$\varphi = E[f(x, \xi)] = \int f(x, \xi)\phi(\xi)d\xi.$$

Theorem 2.1. *(Convexity of the Expected Value Function)*
If the function $f(x, \xi)$ convex a.s., then the function $\varphi = E[f(x, \xi)]$ is also convex.

In a minimization problems, the objective function should be at least lower semi-continuous.

Definition 2.2. *(Lower Semi-Continuous)*
[16]Let $f : \Re^n \longrightarrow (-\infty, +\infty)$ and $x_0 \in Dom(f)$. Then f is lower semi-continuous (l.s.c.) at x_0 iff

$$\liminf_{x \to x_0} f(x) \geq f(x_0).$$

Theorem 2.2. *(Continuity of the Expected Value Function)*
If the function $f(.,\xi)$ is lower semi-continuous at x_0 for almost all ξ, and suppose the following two conditions are satisfied.

1. *$E[\inf_{x \in \cup(x_0)} f(x,\xi)] > -\infty$, where $\cup(x_0)$ is some neighborhood of x_0.*
2. *$f(x,\xi)$ is convex on some neighborhood $\cup(x_0)$ for almost all ξ, then $E[f(.,\xi)]$ is lower semi-continuous at x_0 [16].*

Proof. Assume that condition (1) holds true. Using Fatou's Lemma, Let $x_n \to x_0$ and define $z_n = f(x_n,\xi)$ and $y := \inf_{x \in \cup(x_0)} f(x,\xi)$ Hence,

$$\begin{aligned}
\varphi(x_0) &= E[f(x_0,\xi)] \\
&= \int_{\Re^m} f(x_0,\xi)\phi(\xi)d\xi \\
&\underbrace{\leq}_{l.s.c} \int_{\Re^m} \liminf_{n\to\infty} f(x_n,\xi)\phi(\xi)d\xi \\
&\underbrace{\leq}_{Fatou} \liminf_{n\to\infty} \int_{\Re^m} f(x_n,\xi)\phi(\xi)d\xi \\
&= \liminf_{n\to\infty} \varphi(x_n).
\end{aligned}$$

Suppose now condition (2) hold true. A convex function is always continuous over its domain of definition. Then by convexity Theorem, the function $E[f(.,\xi)]$ is convex on an open neighborhood $\tilde{\cup}(x_0) \subset \cup(x_0)$. As a result $E[f(.,\xi)]$ is continuous on $\tilde{\cup}(x_0)$; hence, continuous at x_0. □

Theorem 2.3. *(Lebesgue Dominated Convergence Theorem)*
Let $\{g_n(\xi)\}$ be a sequence of measurable functions with $g_n : \Omega \longrightarrow \Re$. If

1. *$g_n(\xi) \longrightarrow g(\xi)$ a.e., and*
2. *there is an integrable function $\mathcal{K} : \Omega \longrightarrow \Re_+$ such that $\mid g_n(\xi) \mid \leq \mathcal{K}(\xi)$.*

then g is integrable and

$$\begin{aligned}
\lim_{n\to\infty} \int g_n(\xi)d\mu(\xi) &= \int g d\mu\xi \\
&= \int \lim_{n\to\infty} g_n(\xi)d\mu(\xi),
\end{aligned}$$

where $\mu(\xi)$ is the probability measure associated with ξ (i.e., $d\mu(\xi) = \phi(\xi)d\xi$).

Theorem 2.4. *(Differentiability of the Expected Value Function)*
Let $x_0 \in \Re^n$. If
(i) for each $x \in \cup(x_0)$ in a neighborhood of x_0 the function $f(x,.)$ is measurable,
(ii) $E[\mid f(x_0,\xi) \mid] < \infty$,
(iii) there exists a measurable function $\mathcal{K} : \Omega \longrightarrow \Re_+$ (i.e., a positive valued function) such that $E[\mathcal{K}(\xi)] =: \mathcal{K} < \infty$ and $\mid f(x_1,\xi) - f(x_2,\xi) \mid \leq \mathcal{K}(\parallel x_1 - x_2 \parallel)$ for almost every $\xi \in \Omega$, and
(iv) $f(.,\xi)$ is differentiable at x_0 for almost every $\xi \in \Omega$; then the expected value function $\varphi(x) = E[f(x,\xi)]$ is Lipschitz continuous in the neighborhood $\cup(x_0)$, differentiable at x_0 and $\nabla\varphi(x_0) = \nabla E[f(x_0,\xi)] = E[\nabla_x f(x_0,\xi)]$.

[5]

Proof. (a) From (iii) we have for any x_1 and x_2 in a neighborhood of x_0, that

$$\mid \varphi(x_1) - \varphi(x_2) \mid \leq \int_\Omega \mid f(x_1,\xi) - f(x_2,\xi) \mid \leq \mathcal{K} \parallel x_1 - x_2 \parallel .$$

$$\mid \varphi(x_1) - \varphi(x_2) \mid \leq \mathcal{K} \parallel x_1 - x_2 \parallel .$$

Consequently, $\varphi(.)$ is locally Lipschitz continuous.
(b) By (ii) $\varphi(x_0) = E[f(x_0,\xi)] < \infty$. This implies $\varphi(.)$ well defined, finite valued and Lipschitz continuous in $\cup(x_0)$. Let $\{x_n\}$ be a sequence such that $x_n \longrightarrow x_0$. Then $\exists N : x_n \in \cup(x_0), \forall n \geq N$. Using $e_k \in \Re^n$ as the k^{th} unit vector

$$\underbrace{\frac{\partial f(x_0,\xi)}{dx_k}}_{=:g(\xi)} = \lim_{t_n \to 0} \underbrace{\frac{f(x_0 + t_n e_k,\xi) - f(x_0,\xi)}{t_n}}_{g_n(\xi)}$$

Hence,$g_n(\xi) \longrightarrow g(\xi)$ on Ω and

$$\mid g_n(\xi) \mid = \mid t_n^{-1}[f(x_0 + t_n e_k,\xi) - f(x_0,\xi)] \mid \leq \mathcal{K}(\xi)$$

Then by using Lebesgue Dominated Convergence Theorem, we get

$$\begin{aligned}
\frac{\partial\varphi(x_0)}{\partial x_k} &= [\frac{\partial}{\partial x_k}\int_\Omega f(x,\xi)d\mu(\xi)]_{x=x_0} \\
&= \lim_{t_n \to 0}\int_\Omega \frac{f(x_0 + t_n e_k,\xi) - f(x_0,\xi)}{t_n} d\mu(\xi) \\
&= \int_\Omega \lim_{t_n \to 0} \frac{f(x_0 + t_n e_k,\xi) - f(x_0,\xi)}{t_n} d\mu(\xi) \\
&= \int_\Omega \frac{\partial f(x_0,\xi)}{dx_k}.
\end{aligned}$$

Additionally, each partial derivative $\frac{\partial \varphi(x)}{\partial x_k}$ is continuous in $\cup(x_0)$. Therefore, $\varphi(x) = E[f(x,\xi)]$ is differentiable at x_0 and $\nabla\varphi(x_0) = \nabla E[f(x_0,\xi)] = E[\nabla_x f(x_0,\xi)]$. □

2.3 Types of Stochastic Optimization Problems

There are two major class of Stochastic Optimization problems. These are:

1. Stochastic Optimization Problems with Recourse
2. Chance Constrained Optimization Problems

2.3.1 Stochastic Optimization Problems with Recourse

A stochastic optimization problem with recourse takes the form:

$$\begin{aligned} (SOPR) \quad & \min_x\{f_1(x) + E[Q(x,\xi)]\} \\ & \text{subject to:} \\ & g_j(x) \leq 0, j = 1,\ldots,m_1, \end{aligned} \tag{2.1}$$

where

$$\begin{aligned} & Q(x,\xi) = \inf_y f_2(x,\xi,y) \\ & \text{subject to:} \\ & h_i(x) \leq 0, i = 1,\ldots,m_2, \end{aligned} \tag{2.2}$$

where a decision x should be made irrespective of what will occur in the future first-stage decision. But after the first-stage decision has been made, the outcomes of the first-stage decision could be affected by random events. And $Q(x,\xi)$ is called the *recourse* which is cost of compensation, second-stage value function. The objective $\min_x\{f_1(x) + E[Q(x,\xi)]\}$ minimize the performance function for the first-stage decision plus the cost of compensation [16]. To generalize, a two-stage stochastic(recourse) linear program can be formulated as follows:

$$\begin{aligned} & \min_x\{C^T(x) + E[Q(x,\xi)]\} \\ & \text{subject to:} \\ & Ax = b, \\ & x \geq 0, \end{aligned} \tag{2.3}$$

$$
\begin{aligned}
&Q(x,\xi) = \min_x\{q^T(\omega)y(\omega)\} \\
&\text{subject to:} \\
&T(\omega)x + Wy(\omega) = h(\omega), \\
&y(\omega) \geq 0,
\end{aligned}
\tag{2.4}
$$

where (2.3) represents the first-stage problem, while (2.4) represents the second-stage problem.
- $x \in \Re^{n_1}$ the deterministic first-stage decision vector,
- the first-stage matrix $A \in \Re^{m_1 \times n_1}$ and first-stage vectors $c \in \Re^{n_1}$, $b \in \Re^{m_1}$;
- $y(\omega) \in \Re^{n_2}$ the random second-stage decision vector;
- for a given realization of the random event $\omega \in \Omega$, the second-stage matrix $T(\omega) \in \Re^{m_2 \times n_1}$ and the second-stage vectors $q(\omega) \in \Re^{n_2}$, $h(\omega) \in \Re^{m_2}$ are fixed;
- the second-stage matrix $W \in \Re^{m_2 \times n_2}$ is known as the recourse matrix.
- $\xi^T = (q(\omega), T(\omega), W(\omega), h(\omega))$ the random vector.

Specifically, $\xi(\omega) = \begin{pmatrix} q(\omega) \\ T_1(\omega) \\ \vdots \\ T_{n_1}(\omega) \\ W_1(\omega) \\ \vdots \\ W_{n_2}(\omega) \\ h(\omega) \end{pmatrix}$

Here $W_i(\omega)$ and $T_j(\omega)$ represents i-th and j-th columns of matrices $W(\omega)$ and $T(\omega)$ respectively. If the matrix W is a fixed deterministic matrix, then the Stochastic Optimization Problems with Recourse is called a two-stage linear stochastic optimization with fixed recourse. Otherwise, if W is random matrix, then ξ should also contain the columns of W [16].

2.3.2 Chance Constrained Optimization Problems

The chance constrained approach states constraints as probabilities.
(i) Single (or Separate) Chance Constraints.
Given probability levels $\alpha_1, ..., \alpha_m \in [0, 1]$

$$pr\{g_i(x, \xi) \leq 0\} \geq \alpha_i, i = 1, ..., m,$$

(ii) Joint Chance Constraints.
Given probability levels $\alpha \in [0, 1]$

$$pr\{g_i(x, \xi) \leq 0, i = 1, ..., m\} \geq \alpha.$$

Then we have the general form of Chance Constrained Optimization Problems as follows:

$$\begin{array}{ll} (CCOPT) & min_x\{\gamma_1 E[f(x,\xi)] + \gamma_2 Var[f(x,\xi)]\} \\ & s.t. \\ & pr\{g_i(x,\xi) \leq 0\} \geq \alpha_i, i = 1, ..., m, \\ & x \in X. \end{array} \tag{2.5}$$

is a CCOPT under the single chance constraints $Pr\{g_i(x,\xi) \leq 0\} \geq \alpha_i, i = 1, ..., m$ and

$$\begin{array}{ll} (CCOPT) & min_x\{\gamma_1 E[f(x,\xi)] + \gamma_2 Var[f(x,\xi)]\} \\ & s.t. \\ & pr\{g_i(x,\xi) \leq 0, i = 1, ..., m, \} \geq \alpha \\ & x \in X. \end{array} \tag{2.6}$$

is a CCOPT with joint-chance constraints $Pr\{g_i(x,\xi) \leq 0, i = 1, ..., m\} \geq \alpha$. The values of $\alpha_1, \ldots, \alpha_m \in [0,1]$ and $\alpha \in [0,1]$ are usually pre-given (user-defined). The value α_i is the reliability level of satisfaction of the constraint $Pr\{g_i(x,\xi) \leq 0\} \geq \alpha_i$.
Now our Objective is to determine x that provides a compromised minimum between the expected value $E[f(x,\xi)]$ and variance $Var[f(x,\xi)]$ of f satisfying the constraints

$$Pr\{g_i(x,\xi) \leq 0\} \geq \alpha_i, i = 1, \ldots, m.$$

The parameters $\gamma_1, \gamma_2 \geq 0$ are weighing factors. By choosing a larger value for either γ_1 or γ_2, we can decide which one ($E[f(x,\xi)]$ or $Var[f(x,\xi)]$) we would like to minimize the most. Traditionally, we have $\gamma_1 = 1$ and $\gamma_2 = 0$ so that the objective function consists of only $E[f(x,\xi)]$. If $f(x,\xi) = f(x)$, then we have $E[f(x,\xi)] = f(x)$ and $Var[f(x,\xi)] = Var[f(x)] = 0$.
In general, joint-chance constrained optimization problems [16] pose more difficulties than single chance constrained problems. Here we restrict ourselves to single chance constraints. The concise notation $\{g(x,\xi) \leq 0\}$ represents

$$\{g(x,\xi) \leq 0\} = \{\xi \in \Re^p \mid g(x,\xi) \leq 0\}.$$

Hence, for each fixed $x \in \Re^n$, the set

$$B(x) := \{\xi \in \Re^p \mid g(x,\xi) \leq 0\}$$

is a random set (note that $B(.)$ is a set-valued map).
If $g(x,.) : \Re^p \longrightarrow \Re$ is a measurable function, then $B(x)$ is a random measurable set. Therefore,

$$Pr\{g(x,\xi) \leq 0\} = Pr(B(x))$$

which is the probability measure of the random set $B(x)$. Consequently, x is a feasible point of the chance constraint if and only $Pr(B(x)) \geq \alpha$, i.e., if the probability measure of $B(x)$ is at least equal to α. The given value of α specifies by how much probability or reliability level that a given x should satisfy $g(x, \xi) \leq 0$. In many practical applications (customer demand, product quality, etc), constraints are required to be satisfied with a higher reliability level, i.e., above average. Hence, $0.5 \leq \alpha \leq 1$. Commonly, $\alpha = 0.95, \alpha = 0.98, \alpha = 0.99$, etc [10].
Note that:
if $\alpha = 0$, then every $x \in \Re^n$ satisfies $Pr\{g(x, \xi) \leq 0\} \geq \alpha$, i.e.,

$$\{x \in \Re^n \mid Pr\{g(x, \xi) \leq 0\} \geq \alpha\} = \Re^n,$$

if $\alpha = 1$, the constraint $Pr\{g(x, \xi) \leq 0\} \geq \alpha$ should be satisfied with a 100% guarantee. In real-world applications 100% guarantees are hard to provide. Hence, for $\alpha = 1$, the constraint $Pr\{g(x, \xi) \leq 0\} \geq \alpha$ is said to be conservative. Observe that

$$Pr\{g(x, \xi) \leq 0\} + Pr\{g(x, \xi) > 0\} = 1.$$

Hence, the probability constraint $Pr\{g(x, \xi) \leq 0\} \geq \alpha$ with reliability level α is equivalent to

$$Pr\{g(x, \xi) > 0\} \leq 1 - \alpha.$$

Which represents a risk-constraint, implying that there is a risk probability of $1 - \alpha$ violating the constraint $g(x, \xi) \leq 0$ [10].
Let $\alpha_1, \alpha_2 \in [0, 1]$ be such that $\alpha_1 \leq \alpha_2$. Then,

$$\{x \in \Re^n \mid Pr\{g(x, \xi) \leq 0\} \geq \alpha_2\} \subset \{x \in \Re^n \mid Pr\{g(x, \xi) \leq 0\} \geq \alpha_1\}.$$

As α increases in $[0, 1]$ the set

$$B(\alpha) = \{x \in \Re^n \mid Pr\{g(x, \xi) \leq 0\} \geq \alpha\}$$

shrinks. Since the probability value $Pr\{g(x, \xi) \leq 0\}$ depends on x, this can be indicated by the functional relation as follows:

$$p(x) := Pr\{g(x, \xi) \leq 0\}$$

For the single-chance constrained CCOPT we have

$$p_i(x) := Pr\{g_i(x, \xi) \leq 0\}, i = 1, ..., m.$$

Consequently, the feasible set of CCOPT can be written as

$$B := \{x \in \Re^n \mid p_i(x) \geq \alpha_i, i = 1, ..., m\}$$

for a fixed $\alpha = (\alpha_1, ..., \alpha_m)$ [5]. Hence, the knowledge of continuity, convexity and differentiability of the function $p(x)$ is of paramount importance. Note that, the $B(\alpha)$ above can be also written as

$$B(\alpha) = \{x \in \Re^n \mid p(x) \geq \alpha\}.$$

Remark 2.1. *Given a fixed $\alpha \in [0, 1]$ and a probability constraint*

$$p(x) := Pr\{g(x, \xi) \leq 0\} \geq \alpha,$$

in general, it is very hard to determine the set of all x's that satisfy this inequality. In fact,

$$\begin{aligned} Pr\{g(x, \xi) \leq 0\} &= \int_{g(x,\xi)\leq 0} \phi(\xi) d\xi \\ &= \int_{\Re^p} \chi_{(-\infty, g(x,\xi)]} \phi(\xi) d\xi. \end{aligned}$$

where χ represents the indicator function

$$\chi_{(-\infty,0)}(\varsigma) = \begin{cases} 1 & \text{if } \varsigma \leq 0 \\ 0 & \text{if } \varsigma > 0 \end{cases}$$

Therefore, the value of $p(x)$ is obtained by evaluating a multidimensional integral and $p(x)$ can be discontinuous.

Remark 2.2. *Suppose $\xi = a^T = (a_1, a_2, ..., a_n)$ be normally distributed (not necessarily independent) random variables with mean μ and covariance matrix $\sum$[5]. Then the feasible set*

$$B(\alpha) = \{x \in \Re^n \mid p(x) \geq \alpha\},$$

with $p(x) = pr\{a^T x \leq b\} \geq \alpha$ and $b \in \Re$, can be exactly represented by

$$B(\alpha) = \{x \in \Re^n \mid \mu^T x + \Phi^{-1}(\alpha)\sqrt{x^T \sum x} \leq b\},$$

Properties of Chance Constrained Optimization Problems

1. Convexity

Proposition 2.1. *Suppose $\xi = a^T = (a_1, a_2, ..., a_n)$ be normally distributed (not necessarily independent) random variables with mean μ and covariance matrix $\sum$. Then the feasible set*

$$B(\alpha) = \{x \in \Re^n \mid p(x) \geq \alpha\},$$

with $p(x) = pr\{a^T x \leq b\} \geq \alpha$ and $b \in \Re$, is convex for any $\alpha \in [0.5, 1]$ [15].

Proof. Since $p(x) \geq \alpha$ is equivalent to $\mu^T x + \Phi^{-1}(\alpha)\sqrt{x^T \sum x} \leq b$, the feasible set $B(\alpha)$ is convex if the function $g(x) := \mu^T x + \Phi^{-1}(\alpha)\sqrt{x^T \sum x}$ is convex. But g is a convex w.r.t. x if $\Phi^{-1}(\alpha) \geq 0$ holds true because of the assumption $\alpha \in [0.5, 1]$. Hence the proof. □

Definition 2.3. *(Quasi-concave Probability Measure)*
[15]A probability measure $Pr(.)$ is said to be quasi-concave if

$$Pr(\lambda S_1 + (1-\lambda)S_2) \geq min(Pr(S_1), Pr(S_2)$$

for all convex measurable sets S_1, S_2 and all $\lambda \in [0, 1]$.

Theorem 2.5. *(Wets, 1989)*
If $g : \Re^n \times \Re^p \longrightarrow \Re$ is a (jointly) convex function w.r.t. (x, ξ) and the probability measure $Pr(.)$ is quasi-concave, then the feasible set

$$B(\alpha) = \{x \in \Re^n \mid Pr\{g(x, \xi) \leq 0\} \geq \alpha\}$$

is a convex set for all $\lambda \in [0, 1]$ [16].

Proof. Given α, let $x_1, x_2 \in B(\alpha)$ and $\lambda \in [0, 1]$. Then we want to show that: $\lambda x_1 + (1-\lambda)x_2 \in B(\alpha)$.
Define the set $S(x) = \{\xi \in \Re^p \mid g(x, \xi) \leq 0\}$. Then the sets $S(x_1)$ and $S(x_2)$ are convex and measurable and $Pr(S(x_1)) \geq \alpha$ and $S(x_2) \geq \alpha$. For any $\xi_1 \in S(x_1)$ and $\xi_2 \in S(x_2)$ we have

$$g(\lambda(x_1, \xi_1) + (1-\lambda)(x_2, \xi_2)) \leq \lambda g(x_1, \xi_1) + (1-\lambda)g(x_2, \xi_2) \leq 0.$$

Set $x_\lambda := \lambda x_1 + (1-\lambda)x_2$ and $\xi_\lambda := \lambda\xi_1 + (1-\lambda)\xi_2$. It follows that $\xi_\lambda \in S(x_\lambda)$ and for all $\xi_1 \in S(x_1)$ and $\xi_2 \in S(x_2)$ we have

$$\lambda\xi_1 + (1-\lambda)\xi_2 \in S(\lambda x_1 + (1-\lambda)x_2).$$

$$\Rightarrow \lambda S(x_1) + (1-\lambda)S(x_2) \in S(\lambda x_1 + (1-\lambda)x_2).$$

$$\begin{aligned} \Rightarrow Pr(S(x_\lambda)) &\geq Pr(\lambda S(x_1) + (1-\lambda)S(x_2)) \\ &\geq min\{Pr(S(x_1)), Pr(S(x_2))\} \\ &\geq \alpha. \end{aligned}$$

Hence, the feasible set $B(.)$ is convex. □

Corollary 2.1. *If $g : \Re^n \times \Re^p \longrightarrow \Re$ is a (jointly) convex function w.r.t. (x, ξ) and the probability measure $Pr(.)$ is quasi-concave, then the probability function*

$$p(x) = Pr\{g(x, \xi) \leq 0\}$$

is a quasi-concave.

Definition 2.4. *(Log-concave Measure, Prekopa 1971)*
[16]A probability measure $Pr(.)$ is said to be log-concave if

$$Pr(\lambda S_1 + (1-\lambda)S_2) \geq [Pr(S_1]^{\lambda}.[Pr(S_2)]^{1-\lambda}$$

for all convex measurable sets S_1, S_2 and all $\lambda \in [0,1]$.

Proposition 2.2. *A log-concave probability measure on $\mathcal{F}$ is quasi-concave.*

Proof. Let $S_1, S_2 \in \mathcal{F}$ be convex sets such that $Pr(S_i) \geq 0, i = 1,2$ By assumption, for any $\lambda \in (0,1)$ we have

$$Pr(\lambda S_1 + (1-\lambda)S_2) \geq Pr^{\lambda}(S_1).Pr^{1-\lambda}(S_2),$$

By the monotonicity of the logarithm, it follows that

$$\begin{aligned} \ln[Pr(\lambda S_1 + (1-\lambda)S_2)] &\geq \lambda \ln[Pr(S_1)] + (1-\lambda)\ln[Pr(S_2)] \\ &\geq min\{\ln[Pr(S_1)], \ln[Pr(S_2)]\} \end{aligned}$$

and hence

$$Pr(\lambda S_1 + (1-\lambda)S_2) \geq min[Pr(S_1), Pr(S_2)].$$

□

Note that: For $0 \leq a, b \leq 1$ we have

$$a^{\lambda}b^{1-\lambda} \geq [min(a,b)]^{\lambda}.[min(a,b)]^{1-\lambda} = min(a,b)$$

Proposition 2.3. *Suppose $Pr(.)$ be a probability measure associated with a density function $\phi(\xi)$ of a random variable $\xi \in \Re^p$ such that*

$$Pr(A) = \int_A \phi(\xi) d\xi$$

for any measurable set A. If ϕ is a log-concave function, then $Pr(.)$ is a log-concave probability measure.

Theorem 2.6. *(Borell 1975)*
Suppose $Pr(.)$ be a probability measure associated with a density function $\phi(\xi)$ of a continuous random variable $\xi \in \Re^p$ such that

$$Pr(A) = \int_A \phi(\xi) d\xi$$

If $\phi^{-\frac{1}{p}}$ is convex on $\Re^p$, then the probability measure $Pr(.)$ is a quasi-convex [5].

2. Continuity

Proposition 2.4. *If $g : \Re^n \times \Re^p \longrightarrow \Re$ is an upper semi-continuous, then $p(x)$ is upper semi-continuous and the feasible set*

$$B(\alpha) = \{x \in \Re^p \mid g(x, \xi) \leq 0\} \geq \alpha$$

is a closed set [15].

Proposition 2.5. *If $g : \Re^n \times \Re^p \longrightarrow \Re$ is continuous, and for each $x \in \Re^n$,then $p(x)$ is upper semi-continuous and the feasible set*

$$Pr\{\xi \in \Re^p \mid g_j(x, \xi) = 0\} = 0, j = 1, ..., m$$

then probability function $p(x) = \{g_j(x, \xi)\} \leq 0, j = 1, ..., m$ is continuous; hence, $B(\alpha) = \{x \in \Re^p \mid p(x) \geq \alpha\}$ is a closed set.

3. Differentiability

Assumption 1: For each $x \in \Re^n, \nabla g(x, \xi) \neq 0$ for each ξ on the boundary of the random set $S(x) := \{\xi \in \Omega \mid g(x, \xi) \leq 0\}$. Obviously the set $S_0(x) := \{\xi \in \Omega \mid g(x, \xi) = 0\}$ is a subset of the boundary $\partial S(x)$ of the random set $S(x) = \{\xi \in \Omega \mid g(x, \xi) \leq 0\}$ [5].

Theorem 2.7. *(Uryasev 1995)*
given $x \in \Re^n$. If Assumption 1 is satisfied and
(i) the function $g : \Re^n \times \Re^p \longrightarrow \Re$ has continuous partial derivatives $\nabla_x g(x, \xi)$ and $\nabla_\xi g(x, \xi)$,
(ii) the random set $S(.)$ is bounded in a neighborhood of x.
(iii) there is a continuous matrix function $H : \Re^n \times \Re^m \longrightarrow \Re^{n \times m}$ such that

$$H(x, \xi)\nabla_\xi g(x, \xi) + \nabla_\xi g(x, \xi) = 0$$

(iv) the matrix function $H(x, \xi)$ has continuous partial derivatives $\nabla_\xi H(x, \xi)$, then
(a) the probability function is differentiable and
(b) its gradient is given by

$$\nabla_\xi p(x) = \int_{S(x)} div\xi(H(x, \xi))\phi(\xi)d\xi - \int_{\partial S(x)} (\frac{\nabla_x g(x, \xi) + H(x, \xi)}{\parallel \nabla_\xi g(x, \xi) \parallel})$$

[5].

2.4 Optimality Condition

Definition 2.5. *[1] Let $p \in (0,1)$. A point $v \in \Re^m$ is called a p- efficient point of the probability distribution function F if $f(v) \geq p$ and there is no $z \leq v, z \neq v$ such that $F(z) \geq p$.*

Definition 2.6. *[1] The p- level set of the distribution function $F_Z(z) = Pr\{Z \leq z\}$ of Z is defined as follows:*

$$Z_p = \{z \in \Re^m : F_Z(z) \geq p\} \tag{2.7}$$

Assume $c : \Re^n \longrightarrow \Re$ is convex function.The mapping $g : \Re^n \longrightarrow \Re^m$ has concave components $g_i : \Re^n \longrightarrow \Re$. The set $X \in \Re^n$ is convex and closed, the vector Z takes values in $\Re^m$ for Z_p defined in (2.7), we have

$$\begin{aligned} &\min_{x,z} c(x) \\ &\text{s.t.} \\ &g(x) \geq z, \\ &x \in X, \\ &z \in Z_p \text{ [1].} \end{aligned} \tag{2.8}$$

Associating a lagrange multiplier $u \in \Re^m_+$ with the constraint $g(x) \geq z$, we obtain

$$L(x, z, u) = c(x) + u^T(z - g(x)).$$

The dual functional has the form

$$\Psi(u) = \inf_{(x,z)\in X\times Z_p} L(x, z, u) = h(u) + d(u), \tag{2.9}$$

where

$$h(u) = \inf\{c(x) - u^T g(x) : x \in X\}, \tag{2.10}$$

$$d(u) = \inf\{u^T z : z \in Z^p\}. \tag{2.11}$$

For any $u \in \Re^m_+$ the value of $\Psi(u)$ is a lower bound on the optimal value c^* of the original problem [1]. The best Lagrangian lower bound will be given by the optimal value Ψ^* of the problem:

$$\sup_{u\geq 0} \Psi(u) \tag{2.12}$$

which is called the dual problem to problem (2.8). For $u \ngeq 0$ one has $d(u) = -\infty$, because the set Z_p contains a transformation of $\Re^m_+$. The function $d(.)$ is concave. Note that $d(u) = -S_{Z_p}(u)$, where $S_{Z_p}(.)$ is the support function

of the set Z_p, the we have $d(u) = \inf\{u^T z : z \in ConvZ_p\}$ [1]. Observe that the normal cone to the positive orthant at a point $u \geq 0$ is the following:

$$V(u) = \{v \in \Re^m : u^T v = d(u) \quad \text{and} \quad v \quad \text{is a p-efficient.} \tag{2.13}$$

we define the set

$$N_{\Re^m_+}(u) = \{d \in \Re^m_- : d_i = 0 \quad \text{if} \quad u_i > 0, i = 1, ..., m\}, \tag{2.14}$$

Let us consider the convex hull problem:

$$\begin{aligned} &\min_{x,z} c(x) \\ &\text{s.t.} \\ &g(x) \geq z, \\ &x \in X, \\ &z \in ConvZ_p \end{aligned} \tag{2.15}$$

we impose the following constraint qualification condition:

$$\text{There exist points} \quad x^0 \in X \quad \text{and} \quad z^0 \in ConvZ_p \quad \text{such that} \quad g(x^0) > z^0. \tag{2.16}$$

If this constraint qualification condition is satisfied, then the duality theory implies that there exists $\hat{u} \geq 0$ at which the minimum attained and $\Psi^* = \Psi(\hat{u})$ is optimal value of (2.15) [1].

Definition 2.7. *A distribution function F is called α-concave on the set $A \subset \Re^s$ with $\alpha \in [-\infty, \infty]$ if*

$$F(z) \geq m_\alpha(F(x), F(y), \lambda)$$

for all $z, x, y \in A$, and $\lambda \in (0, 1)$ such that

$$z \geq \lambda x + (1 - \lambda)y.$$

Theorem 2.8. *Assume that the constraint qualification in*(2.16) *is satisfied, the probability distribution of the vector Z is α-concave for some $\alpha \in [-\infty, \infty]$, and the set X is compact. If a point $(\hat{x}, \hat{z})$ is an optimal solution of* (2.8)*, then there exists a vector $\hat{u} \geq 0$, which is an optimal solution of* (2.12) *and the optimal values of both problems are equal. If $\hat{u}$ is an optimal solution of* (2.12)*, then there exist a point $\hat{x}$ such that $(\hat{x}, \hat{g(x)})$ is a solution of* (2.12) *and the optimal values of both problems are equal [1].*

Proof. By the assumption of α-concavity, problems (2.8) and (2.15) are the same. If $\hat{u}$ is an optimal solution of (2.12), we obtain the existence of points $\hat{x} \in X(\hat{u}), v^1, ..., v^{m+1} \in V(u)$ and scalars $\beta_1, ..., \beta_{m+1} \geq 0$ with$\sum_{j=1}^{m+1} \beta_j = 1$ such that the optimality conditions are satisfied. Setting $\hat{z} = g(\hat{x})$ we show that $\hat{x}, \hat{z}$ is an optimal solution of (2.8) and that the optimal values of both problems are equal. First we observe that this point is feasible. We choose $y \in -N_{\Re^m_+}(\hat{u})$ such that

$$y = g(\hat{x}) - \sum_{j=1}^{m+1} \beta_j v^j.$$

From the definitions of $X(\hat{u}), V(\hat{u})$, and the normal cone, we obtain,

$$h(\hat{u}) = c(\hat{x}) - \hat{u}^T g(\hat{x}) = c(\hat{x}) - \hat{u}^T (\sum_{j=1}^{m+1} \beta_j v^j + y)$$

$$= c(\hat{x}) - \sum_{j=1}^{m+1} \beta_j d(\hat{u}) - \hat{u}^T y = c(\hat{x}) - d(\hat{u}).$$

Thus

$$c(\hat{x}) = h(\hat{u}) + d(\hat{u}) = \Psi^* \geq c^*,$$

which proves that $(\hat{x}, \hat{z})$ is an optimal solution of (2.8) and $\Psi^* = c^*$. If $(\hat{x}, \hat{z})$ is an optimal solution of (2.8), then there is a vector $\hat{u} \geq 0$ such that $\hat{u}_i(\hat{z}_i - g_i(\hat{x})) = 0$ and

$$0 \in \partial c(\hat{x}) + \partial \hat{u}^T g(\hat{x}) - \hat{z} + N_{X \times Z}(\hat{x}, \hat{z}).$$

This means that

$$0 \in \partial c(\hat{x}) \partial u^T g(\hat{x}) + N_X(\hat{x}) \tag{2.17}$$

and

$$0 \in \hat{u} + N_Z(\hat{z}) \tag{2.18}$$

The first inclusion (2.17) is optimality condition for problem (2.10) and thus $x \in X(\hat{u})$, the inclusion (2.11) is equivalent to $\hat{z} \in \partial \chi^*_{Z_p}(\hat{u})$, then we get that

$$\hat{z} \in \partial d(\hat{u}) = ConvV(\hat{u}) - N_{\Re^m_+}(\hat{u})$$

Thus, there exists point $v^1, ..., v^{m+1} \in V(u)$ and scalars $\beta_1, ..., \beta_{m+1} \geq 0$ with$\sum_{j=1}^{m+1} \beta_j = 1$ such that

$$\hat{z} - \sum_{j=1}^{m+1} \beta_j v^j \in -N_{\Re^m_+}(\hat{u})$$

Using the complementarity condition $\hat{u}_i(\hat{z}_i - g_i(\hat{x})) = 0$ we conclude that the optimality condition holds and $\hat{u}$ is an optimal solution of problem (2.12) $\square$

In the linear probabilistic optimization problem, let the functions $c : \Re^n \times \Re^s \longrightarrow \Re$ and $g : \Re^n \times \Re^s \longrightarrow \Re^m$ are continuous and the set $X \subset \Re^n$ is a closed convex set. $g(x) = B^T x$, where B is an $m \times n$ matrix, and $c(x) = c^T x$ with $c \in \Re^n$[1]. The problem read as follows:

$$\begin{array}{ll} & \min_x c^T x \\ (CCOPT) & \text{s.t.} \\ & Pr\{Bx \geq \xi\} \geq p, \\ & Ax \geq b, \\ & x \geq 0. \end{array} \tag{2.19}$$

Here B is an $s \times n$ matrix and $b \in \Re^s, p \in (0, 1)$.

Definition 2.8. *[1]Problem* (2.19) *satisfies the dual feasibility condition if*

$$\Lambda = \{(u, w) \in \Re_+^{s+m} : A^T w + B^T u \leq c\} \neq \emptyset.$$

Theorem 2.9. *Assume that the feasible set of* (2.19) *is non empty and that* ξ *has a discrete distribution on* ξ^n*. Then* (2.19) *has an optimal solution iff it satisfies the LQ condition defined in definition* (2.8)*[1].*

Proof. If (2.19) has an optimal solution, then for some $j \in \varepsilon$ the linear optimization problem.

$$\begin{array}{l} \min_x c^T x \\ \text{s.t.} \\ Pr\{Bx \geq v^j\} \geq p, \\ Ax \geq b, \\ x \geq 0. \end{array} \tag{2.20}$$

has an optimal solution. By duality, its dual problem

$$\begin{array}{l} \min_{u,w} u^T v^j + b^T w \\ \text{s.t.} \\ B^T u + A^T w \leq c, \\ u, w \geq 0, \end{array} \tag{2.21}$$

has an optimal solution and the optimal values of both programs are equal. Thus, the dual feasibility condition in definition 2.8 must be satisfied. On

the other hand, if the dual feasibility condition is satisfied, all dual programs (2.21) for $j \in \varepsilon$ have nonempty feasible sets, so the objective values of all primal problems (2.20) are bounded from below. Since at least one of them has a nonempty feasible set by assumption, an optimal solution must exist. □

Chapter 3

Methods in Stochastic Optimization

3.1 Introduction

To solve the optimization problem

$$\begin{aligned} & min_x E[f(x,\xi)] \\ & s.t. \\ & Pr\{g_i(x,\xi) \leq 0\} \geq \alpha_i, i = 1, ..., m, \\ & x \in X. \end{aligned} \tag{3.1}$$

defining

$$F(x) := E[f(x,\xi)] = \int_\Omega f(x,\xi)\phi(\xi)d\xi$$

and

$$\begin{aligned} G(x) &:= Pr\{g(x,\xi) \leq 0\} \\ &= \int_{g(x,\xi)\leq 0} \phi(\xi)d\xi \\ &= \int_\Omega \chi_{(-\infty,0]}(g(x,\xi))\phi(\xi)d\xi \\ &= E[\chi_{(-\infty,0]}(g(x,\xi))] \\ & x \in X. \end{aligned}$$

Then, problem (3.1) is equivalently written as

$$\begin{array}{ll} & min_x F(x) \\ (NLP) & s.t. \\ & G(x) - \alpha \geq 0 \\ & x \geq 0. \end{array} \tag{3.2}$$

which is NLP [5]. Then how we solve this problem? The most difficult task in solving (3.1) is the evaluation of the values of the chance constraint

$$G(x) := Pr\{g(x, \xi) \leq 0\}$$

for a given x. Note that $z = g(x, \xi)$ is a random variable, since ξ is random. If $z = g(x, \xi)$ is non-linear w.r.t.ξ it is difficult to determine the distribution of the random variable z from that of ξ [5].

3.2 Some Methods of solving Stochastic Optimization Problems

The major idea we deal here is how we can solve one optimization problem with stochastic data and how we can change the stochastic problems to an equivalent deterministic form and solving it. There are different methods used for solving stochastic optimization problems. Below we state some of them.

3.2.1 Back-mapping (Projection) Method

The back-mapping (constraint transformation) was proposed by Wendt et al. [9] for the transformation of the chance constraints from the space of output variables into the space of uncertain input variables whose joint distribution is known. The transformation of chance constraints is performed based on the assumption of the existence of strict monotonic relations between a chance constrained variable $x_i, i \in I$, and some uncertain variable ξ_j. In higher dimensions (experimentally or analytically) one studies the equation $z = g(x, \xi)$ and among $\xi_1, \xi_2, ..., \xi_m$ find a ξ_j which has a strict monotonic relation with z, so that $z = \varphi_x(\xi_j)$[5]. Here,
$\varphi_x(\xi_j)$ is either strictly increasing as $\xi_j \uparrow z$ or strictly decreasing as $\xi_j \downarrow z$.

Thus $\xi_j = \varphi_x^{-1}(z)$.

$$\begin{aligned}\xi_j \uparrow z \Rightarrow Pr\{g(x,\xi) \leq 0\} & \\ &= Pr\{\xi_j \leq \varphi_x^{-1}(0)\} \\ &= \int ... \int (\int_{-\infty}^{\varphi_x^{-1}(0)} \phi(\xi_j)) d\tilde{\xi}, \\ & x \in X.\end{aligned}$$

and

$$\begin{aligned}\xi_j \downarrow z \Rightarrow Pr\{g(x,\xi) \leq 0\} & \\ &= Pr\{\xi_j \geq \varphi_x^{-1}(0)\} \\ &= \int ... \int (\int_{\varphi_x^{-1}(0)}^{+\infty} \phi(\xi_j)) d\tilde{\xi}, \\ & x \in X.\end{aligned}$$

where $\tilde{\xi} = (\xi_1, ..., \xi_{j-1}, \xi_{n+1}, ..., \xi_m)$. Now, for $\xi_j \uparrow z$, (3.1) (CCOPT) is equivalent to

$$\begin{array}{ll} & \min_x E[f(x,\xi)] \\ (CCOPT) & \text{s.t} \\ & Pr(x) = Pr\{\xi_j \leq \varphi_x^{-1}(0)\} \geq \alpha, \\ & x \in X. \end{array}$$

$$\begin{aligned} Pr(x) &= \int ... \int (\int_{-\infty}^{\varphi_x^{-1}(0)} \phi(\xi)) d\xi \\ \nabla Pr(x) &= \int ... \int (\int_{-\infty}^{\varphi_x^{-1}(0)} \nabla_x \varphi_x^{-1}(0) \phi(\xi)) d\xi \end{aligned}$$

and $\xi_j \downarrow z$, (3.1) (CCOPT) is equivalent to

$$\begin{array}{ll} & \min_x E[f(x,\xi)] \\ (CCOPT) & \text{s.t} \\ & Pr(x) = Pr\{\xi_j \geq \varphi_x^{-1}(0)\} \leq \alpha, \\ & x \in X. \end{array}$$

$$\begin{aligned} Pr(x) &= \int ... \int (\int_{\varphi_x^{-1}(0)}^{+\infty} \phi(\xi)) d\xi \\ \nabla Pr(x) &= \int ... \int (\int_{\varphi_x^{-1}(0)}^{+\infty} \nabla_x \varphi_x^{-1}(0) \phi(\xi)) d\xi \end{aligned}$$

This method has advantage if monotonic relations ($z = g(x, \xi)$) are easy to find and provides direct representation of chance constraints. But it has disadvantage since monotonic relation may not exist and difficult to verify. [5]

3.2.2 Sample Average Approximation (SAA)

The main idea of Sample Average Approximation (SAA) approach for solving stochastic programs is as follows. We define the indicator χ:

$$\chi_{(0,+\infty]}(g(x,\xi)) = \begin{cases} 0 & \text{if } g(x,\xi) > 0 \\ 1 & \text{if } g(x,\xi) \leq 0 \end{cases}$$

and determine the samples $\{\xi_1, \xi_2, ..., \xi_N\} \subset \Omega$ then, replace the chance constrains with

$$Pr_N(x) = \frac{1}{N} \sum_{k=1}^{N} \chi_{(-\infty,0]}(g(x,\xi_k)) \geq \alpha$$

($Pr_N(x)$ = Relative-frequency count for the satisfaction of $g(x, \xi) \leq 0$) Sampling techniques have been integrated with decomposition algorithms to successfully solve stochastic linear programs of enormous sizes to great precision.

Example 3.1. *Let take the classical maximum likelihood method of estimation. That is, let $g(x, \xi)$ be a family of probability density functions (pdf), parameterized by the parameter vector $y \in Y \subset \Re^m$, and let $\xi_1, \xi_2, ..., \xi_N$ be an i.i.d. random sample with a probability distribution P. Define*

$$\hat{f}_N(y) := -N^{-1} \sum_{i=1}^{N} \ln g(\xi_i, y)$$

By the law of large numbers we have that, for any fixed value of y, $\hat{f}_N(y)$ converges to

$$f(y) := -E_p\{\ln g(\xi_i, y)\} = -\int \ln g(x, y) p dx,$$

with probability one, as $N \longrightarrow \infty$. This leads to the "true" and "approximating" optimization problems of minimizing $f(y)$ and $\hat{f}_N(y)$, respectively, over the parameter set Y.
In particular, suppose that the distribution P is given by a pdf $g(x, y_0), y_0 \in Y$, then y_0 is an unconstrained minimizer of $f(y)$, and hence is an optimal solution of the "true" problem. Indeed, by using concavity of the logarithm

function, we obtain

$$f(y_0) - f(y) = \int \ln[\frac{g(x,y)}{g(x,y_0)}]g(x,y_0)dx$$
$$\leq \int [\frac{g(x,y)}{g(x,y_0)} - 1]g(x,y_0)dx.$$

[2]

SAA has advantages by avoiding computation of multidimensional integrals and preserving convexity structures. But its disadvantage is that it leads to a non-smooth optimization and feasibility of the obtained solution to the (3.1) is guaranteed only when $N \to \infty$[5].

3.2.3 Robust Optimization Technique

The origins of robust optimization date back to the establishment of modern decision theory in the 1950s and the use of worst case analysis and Wald's maximin model as a tool for the treatment of severe uncertainty. It became a discipline of its own in the 1970s with parallel developments in several scientific and technological fields. Robust optimization considers the (worst-case) problem

$$\begin{array}{ll} & \min_x E[f(x,\xi)] \\ (RO) & \text{s.t} \\ & g(x,\xi) \leq 0, \xi \in \Omega, \\ & x \in X. \end{array} \tag{3.3}$$

where $g(x,\xi) \leq 0$ is required to be satisfied for as many realizations of ξ from Ω as possible[5]. By generating independent identically distributed random samples $\xi_1, ..., \xi_N$ form Ω we solve the problem

$$\begin{array}{ll} & \min_x \frac{1}{N} \sum_{i=1}^{N} f(x,\xi_i) \\ (NLP)_{RO} & \text{s.t} \\ & g(x,\xi_i) \leq 0, i = 1,2,...,N \\ & x \in X. \end{array} \tag{3.4}$$

Theorem 3.1. *Suppose $\alpha \in (0,1)$ and $f(.,\xi)$ is convex w.r.t. $x \in \Re^n$. If the number of random samples $\xi_1, ..., \xi_N$*

$$N \geq \frac{2n}{1-\alpha} \ln(\frac{1}{1-\alpha}) + (\frac{2}{1-\alpha}) \ln(\frac{1}{\alpha}) + 2n,$$

then the optimal solution obtained from (3.4) *is an optimal solution of* (3.3) *with reliability* α.

This method has advantages in preserving convexity, simple to implement and solve, and there is no need to compute integrals. Its disadvantage is that solution of (3.4) may not feasible to the (3.1) and for a higher reliability level α, very large number of scenarios $\xi_1, ..., \xi_N$ are required.

3.3 Solution Technique for Linear Chance Constrained Optimization Problems

To solve one stochastic optimization problem and see how the stochastic problem changed to equivalent deterministic problem and solved let we limit ourselves to linear chance constrained optimization problem. Linear Optimization problems with single chance constraints [5] has the form:

$$\begin{aligned} (LCCOPT) \quad & min_{x\in\Re^n} E[c^T x] \\ & \text{s.t.} \\ & Pr\{a_i^T x \leq b_i\} \geq \alpha_i, i = 1, ..., m \\ & x \geq 0, \end{aligned} \tag{3.5}$$

where:
the decision variable $x = (x_1, ..., x_n)^T \in \Re^n$ is deterministic,
the matrix $A = \begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_m^T \end{bmatrix}$ is a random matrix and $c \in \Re^n$ is a random vector.
We consider different cases.
(a) The matrix $A = (a_{ij})$ is random, b and c are deterministic.
Assumption 2 : The elements a_{ij} of the matrix A are independently normally distributed with mean μ_{ij} and standard deviation σ_{ij} , i.e., $a_{ij} \sim \mathcal{N}(\mu_{ij}, \sigma_{ij})$.
Define $d_i := a_i^T x = \sum_{j=1}^n a_{ij} x_j$.
Where d_i is a linear combination of normally distributed random variables. For each $i \in \{1, 2, ..., n\}, d_i$ is normally distributed, with mean: $\mu d_i = \sum_{j=1}^n \mu_{ij} x_j$. and variance (standard deviation): $\sigma^2 d_i = \sum_{j=1}^n \sigma_{ij} x_j^2$.

Hence,

$$\begin{aligned} Pr\{a_i^T x \leq b_i\} \geq \alpha_i &\equiv Pr\{\sum_{j=1}^{n} a_{ij}x_j \leq b_i\} \geq \alpha_i \\ &\equiv Pr\{d_i \leq b_i\} \geq \alpha_i, i = 1, ..., m. \\ &\equiv Pr\{\frac{d_i - \mu d_i}{\sigma d_i} \leq \frac{b_i - \mu d_i}{\sigma d_i}\} \geq \alpha_i, i = 1, ..., m. \end{aligned}$$

The random variable $s_i := \frac{d_i - \mu d_i}{\sigma d_i}$ has a standard normal distribution,i.e., $s_i \sim \mathcal{N}(0, 1), i = 1, ..., m.$

$$pr\{\frac{d_i - \mu d_i}{\sigma d_i} \leq \frac{b_i - \mu d_i}{\sigma d_i}\} = \Phi(\frac{b_i - \mu d_i}{\sigma d_i})$$

where $\Phi(.)$ is the cumulative standard normal distribution function of s_i. Thus, we have

$$\begin{aligned} &\Phi(\frac{b_i - \mu d_i}{\sigma d_i}) \geq \alpha_i, i = 1, ..., m. \\ \Leftrightarrow &\frac{b_i - \mu d_i}{\sigma d_i} \geq \Phi^{-1}(\alpha_i) \\ \Leftrightarrow &b_i - \mu d_i \geq \sigma d_i \Phi^{-1}(\alpha_i), i = 1, ..., m. \end{aligned}$$

where Φ^{-1} is the inverse normal distribution function.

$$\begin{aligned} &b_i - \mu d_i \geq \sigma d_i \Phi^{-1}(\alpha_i) \\ \Leftrightarrow &\sum_{j=1}^{m} \mu_j x_j + \Phi^{-1}(\alpha_j)\sqrt{\sum_{j=1}^{m} \sigma_{ij}^2 x_j^2} - b_i \leq 0. \end{aligned}$$

Therefore, an equivalent representation of (LCCOPT) is

$$\begin{aligned} &min_{x \in \Re^n} c^T x \\ &\text{s.t.} \\ &\sum_{j=1}^{m} \mu_j x_j + \Phi^{-1}(\alpha_j)\sqrt{\sum_{j=1}^{m} \sigma_{ij}^2 x_j^2} - b_i \leq 0, i = 1, ..., m, \\ &x \geq 0, \end{aligned} \tag{3.6}$$

We call this equation as deterministic nonlinear optimization problem.
(b) The vector b is random, the matrix $A = (a_{ij})$ is and the vector c are

deterministic.
Assumption 3 : The components of the vector $b^T = (b_1, b_2, ..., b_m)$ independently normally distributed with mean μ_i and variance $\sigma_i, i = 1, 2, ..., m$. Hence,

$$\begin{aligned} Pr\{a_i^T x \leq b_i\} &= Pr\{\sum_{j=1}^{n} a_{ij} x_j \leq b_i\} \\ &= Pr\{\frac{\sum_{j=1}^{n} a_{ij}^T x_j - \mu_i b_i}{\sigma b_i} \leq \frac{b_i - \mu b_i}{\sigma b_i}\} \geq \alpha_i \\ &\Leftrightarrow Pr\{\frac{b_i - \mu b_i}{\sigma b_i} \leq \frac{\sum_{j=1}^{n} a_{ij}^T x_j - \mu_i b_i}{\sigma b_i}\} \geq 1 - \alpha_i \\ &\Leftrightarrow \Phi(\frac{\sum_{j=1}^{n} a_{ij}^T x_j - \mu_i b_i}{\sigma b_i}) \leq 1 - \alpha_i \\ &\Leftrightarrow \sum_{j=1}^{n} a_{ij}^T x_j - \mu_i b_i \leq \Phi^{-1}(1 - \alpha_i)\sigma b_i. \end{aligned}$$

As a result we obtain

$$\begin{aligned} &min_{x \in \Re^n} c^T x \\ &\text{s.t.} \\ &a_i^T x \leq \mu_i b_i + \Phi^{-1}(1 - \alpha_i)\sigma b_i, i = 1, ..., m, \\ &x \geq 0, \end{aligned} \tag{3.7}$$

which is a deterministic linear optimization problem [16].

3.3.1 Examples

Find the minimum solution of the following problems:

Example 3.2.

$$\begin{aligned} &min_{x \in \Re^1} 2x^2 + 4 \\ &s.t. \\ &Pr\{\xi x \leq 0\} \geq 0.9 \\ &x \geq 0, \end{aligned}$$

where ξ is a random matrix normally distributed with a mean $\mu = 2$ and variance $\sigma^2 = 0.01$

Using the procedure stated in case (a), an equivalent deterministic form of the given stochastic problem is:

$$
\begin{aligned}
& min_{x\in\Re^1} 2x^2 + 4 \\
& s.t. \\
& \mu x + \Phi^{-1}(0.9)\sqrt{\sigma^2 x^2} - 0 \leq 0 \\
& x \geq 0,
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow \quad & min_{x\in\Re^1} 2x^2 + 4 \\
& s.t. \\
& 2x + \Phi^{-1}(0.9)\sqrt{0.01x^2} \leq 0 \\
& x \geq 0,
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow \quad & min_{x\in\Re^1} 2x^2 + 4 \\
& s.t. \\
& 2x + 1.2816 \times 0.1x \leq 0 \\
& x \geq 0,
\end{aligned}
$$

Then solving the deterministic problem,

$$
\begin{aligned}
& min_{x\in\Re^1} 2x^2 + 4 \\
& s.t. \\
& 2.12816x \leq 0 \\
& x \geq 0,
\end{aligned}
$$

gives the solution $x = 0$ which minimize $f(x)$.

Example 3.3.

$$
\begin{aligned}
& min_{x\in\Re^1} 5x - 20 \\
& s.t. \\
& Pr\{-2x \leq \xi\} \geq 0.9 \\
& x \geq 0,
\end{aligned}
$$

where the vector $b = \xi$ is a random vector normally distributed with mean $\mu = 5$ and variance $\sigma^2 = 0.09$
Then since b is a random vector by the procedure stated in case (b), we have,

$$
\begin{aligned}
& min_{x\in\Re^1} 5x - 20 \\
& s.t. \\
& ax \leq \mu b + \Phi^{-1}(1-\alpha)\sigma b \\
& x \geq 0,
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow \quad & min_{x\in\Re^1} 5x - 20 \\
& s.t. \\
& -2x \leq 5 + \Phi^{-1}(0.1)(0.3) \\
& x \geq 0, \\
\Rightarrow \quad & min_{x\in\Re^1} 5x - 20 \\
& s.t. \\
& 2x + 4.61552 \geq 0 \\
& x \geq 0,
\end{aligned}
$$

is an equivalent linear deterministic and lastly solving this, we get $x = 0$ *which minimizes the given function* $f(x) = 5x - 20$.

Bibliography

[1] Alexander Shapiro, D.Dentcheva and A.Ruszczyriski, *Lectures on Stochastic Programming Modelling and Theory*, Saim and mps, USA, 2009.

[2] Alexander Shapiro, *Statistical Inference of Stochastic Optimization Problem*, Annals of Statistics, USA, 1989, pp.2-3.

[3] A.S.Poznyak and K.Najim,*Learning Automata and Stochastic Optimization*, Springer,London 1997.

[4] Dennis, J.E, and Schnabel, R.B., *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, 1983.

[5] Dr. Abebe Geletu, *Lecture Notes on Stochastic Optimization*, AAU, summer 2013.

[6] Fwu-Rang Chang, *Stochastic Optimization in Continuous Time*, Cambridge University Press, Cambridge, 2004.

[7] Geng Deng, *Simulation-Based Optimization*, University of Wisconsin, Madison, 2007.

[8] James C.Spall, *Introduction to Stochastic Search and Optimization*, Hoboken, New Jersey, 2003.

[9] Jitka Dupacova, Jan Hurt and Josef Stepan, *Stochastic Modeling In Economics and Finance*, Kluwer Academic Publisher, New York, 2003.

[10] John R. Birge and François Louveaux, *Introduction to Stochastic Programming*, Springer, New York, London, 2011.

[11] J.Gentle, W.Härdle, and Y.Mori, *Handbook of Computational Statistics*, Springer, Heidelberg, 2004.

[12] Kaddour Najim, Enso Ikonen and Ait-Kadi Daoud, *Stochastic Processes-Estimation, Optimization and Analysis*, Kagon Page Science, London and Sterling, VA, 2004.

[13] Karl-George Steffens, George A.Anastassion, (Series in Operations Research and Financial Engineering)*Introduction to Stochastic Programming*, Springer, Verlag New York, 1997.

[14] Kurt Marti, *Stochastic Optimization Methods*, Springer, Berlin-Heidelberg,2005.

[15] Peter Kall, *Stochastic Linear Programming*, Springer, Berlin, 1976.

[16] Peter Kall and Stein W. Wallace, *Stochastic Programming*, John Wiley and Sons, New York, First, 2003.

[17] Minkyu Choi, Alisherov F. A, Seung-Hwan Jeon and Sattarova F. Y., *Review on Transfer and Processing of the Continuous Information*, International Journal of Smart Home Vol.3, No.4, October, pp.37-40, 2009.

[18] Mokhtar S. Bazaraa, *Nonlinear Programming Theory and Algorithms*, John Wiley and Sons, Inc., Hobken, New Jersey, 2006.

[19] M.Thoma, *Lecture Notes In Control and Information Sciences*, CUP, Cambridge, 2002.

[20] Richard J.Larsen and Morris L.Marx, *An Introduction To Mathematical Statistics and Its Application*, Prentice Hall, Boston, 2006.

[21] S.C.Gupta, *Fundamental of Mathematical Statistics*, Sultan Chand and Sons, New Delhi, 2000.

[22] Singiresu S. Rao, *Engineering Optimization Theory and Practice*, John Wiley and Sons, Inc., Hobken, New Jersey, 2009.

[23] Stein W. Wallace and William T. Ziemba, *Applications of Stochastic Programming*, Siam, Newyork, 2005.