UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Aspects of optimization with increasing concave stochastic order constraints

Permalink
https://escholarship.org/uc/item/7864q7b9

Author
Haskell, William Benjamin

Publication Date
2011

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Aspects of optimization with increasing concave stochastic order constraints

By

William B. Haskell

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering – Industrial Engineering and Operations Research

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Zuo-Jun Max Shen, Chair
Professor J. George Shanthikumar
Professor Rhonda Righter
Professor Venkat Anantharam

Spring 2011

Abstract

Aspects of optimization with increasing concave stochastic order constraints

by

William B. Haskell

Doctor of Philosophy in Industrial Engineering and Operations Research

University of California, Berkeley

Professor Zuo-Jun Max Shen, Chair

This dissertation applies convex optimization techniques to a class of stochastic optimization problems. This class of problems is distinguished by a multivariate increasing concave stochastic order constraint on a random-variable-valued mapping. The analysis is divided into four chapters to address four different aspects of optimization with increasing concave stochastic order constraints.

In Chapter 3, we describe this class of stochastic optimization problems in detail. This class of problems does not satisfy the Slater condition, and we must overcome this technical difficulty to apply convex optimization. We introduce a perturbation of the original problem for this purpose and we discuss semi-infinite programming and nonlinear programming relaxations of this perturbation. It is shown that increasing concave functions act as the Lagrange multipliers of increasing concave stochastic order constraints. We conclude this chapter with a discussion about a transformation of these problems via coupling on finite probability spaces.

In Chapter 4, we consider sample average approximation for increasing concave stochastic order constrained programs. We verify consistency of sample average approximation. We also study sample average approximation for the semi-infinite programming and nonlinear programming relaxations presented in Chapter 3. Again, we verify consistency of sample average approximation for these relaxations. The solution method in this chapter is based on the coupling transformation from Chapter 3. This transformation necessarily requires a large-scale implementation. We comment on aggregation, column generation, and row generation techniques as a possible large-scale approach.

In Chapter 5 we introduce a robust version of the increasing concave stochastic order. Specifically, we consider model misspecification for the underlying probability distribution. Robust versions of the semi-infinite programming and

nonlinear programming relaxations from Chapter 3 are also presented and analyzed. The case of polyhedral uncertainty is emphasized for finite probability spaces. There is a large-scale implementation issue here as well, and aggregation is suggested again as a possible solution approach.

In Chapter 6 we consider multi-period stochastic optimization problems with increasing concave stochastic order constraints. We put a multivariate increasing concave stochastic order constraint on a vector of performance measures across all time periods. We derive optimality conditions similar to those in Chapter 3. We also show that there is a companion auxiliary control problem for this class of multi-period problems that can be solved with dynamic programming. We comment on using duality decomposition to solve instances of this problem on finite probability spaces, such as those obtained via conditional sampling.

Contents

CHAPTER 1

Introduction

1.1. Motivation

1.1.1. Optimization under uncertainty. Decision making with incomplete information is called optimization under uncertainty. Many approaches for modeling uncertainty and incorporating uncertainty into optimization problems have been taken. The minimax approach has been studied extensively (see [**5**]). This paradigm finds the decision with the best worse-case scenario. Info-gap decision theory (see [**3**]) finds the decision which meets a pre-specified goal under the greatest range of uncertainty. Stochastic optimization is a third popular paradigm for dealing with uncertainty, and is also perhaps the oldest.

In stochastic optimization, a probability distribution is assigned to the uncertainty. A probability distribution captures the important idea of some events being more likely than others. A probabilistic model of uncertainty also introduces the idea of risk. The idea of risk is intrinsic to stochastic optimization, and many methods for modeling and managing risk are available.

1.1.2. Risk management. This dissertation is built around using multivariate increasing concave stochastic order constraints to manage risk. The family of increasing concave functions is closely related to the idea of risk-aversion and this relationship is well understood in the univariate case. This intuition can be extended to the multivariate case. In particular, there are many problems where a risk-averse decision maker would like to manage multiple random prospects simultaneously. A multivariate increasing concave function is a natural model for the utility of a risk-averse decision maker in this situation, since there may be inter-dependencies among the random prospects.

The multivariate increasing concave stochastic order is built from the spectrum of all such utility functions. Instead of trying to maximize the expectation of a given utility function, we choose a benchmark random variable to make pairwise comparisons. The expectation of the utility of a random-variable-valued mapping is compared to the expectation of the utility of the benchmark for all increasing concave functions. This methodology restricts the set of admissible random variables to those that are acceptable in some sense, rather than those that are maximal with respect to a scalar objective. The multivariate increasing concave stochastic order fits naturally into convex optimization and its powerful set of analytical tools.

There are many applications for this class of problems, particularly in network management and design problems. For example, communication, electric, and transportation networks all have an inherent idea of flow and a corresponding capacity for flow. In real networks, there is variation in the capacity for flow and the demand for capacity utilization. Real networks also face the possibility of catastrophic disruption. We can model capacity, demand, and disruption stochastically. The increasing concave stochastic order then naturally constrains the vector of unmet demand across the network.

1.1.3. Benchmark. We now make some intuitive remarks about the value and implementation of stochastic order constraints. Any time a stochastic order constraint is used in optimization, a benchmark must be chosen. The idea of a benchmark is very intuitive. Throughout this dissertation, a benchmark refers to an exogenously chosen random variable that is desirable in some sense. We emphasize that the benchmark is itself a random variable. The overall strategy of stochastic order constrained optimization is to set a benchmark, and then choose from the alternatives that are larger than the benchmark in a specific stochastic order.

1.1.4. Effective comparison. There is intuitive justification for the use of stochastic order constraints. Stochastic order constraints compare random variables to similar objects, other random variables. Approaches like chance constraints and risk measures turn information about random variables into scalars, and then optimize over the scalars. There are two inherent difficulties with this approach. First, a choice needs to be made about how to scalarize risk. There are no general guidelines for this choice. Second, information about random variables is lost through any scalarization. Stochastic order constraints bypass these deficiencies by using more of the distributional information. Stochastic orders can also be thought of as comparing a continuum of risk measures between two random variables.

1.1.5. Convex integral stochastic order constraints. We will derive optimality conditions and a corresponding duality theory by using convex optimization techniques. This procedure will show that the Lagrange multipliers of increasing concave stochastic order constraints are increasing concave functions. We will emphasize this fact throughout this dissertation. This observation makes the interpretation of the optimality conditions and duality theory more transparent.

One can imagine constructing the Lagrange multiplier of an integral stochastic order constraint by taking sums and non-negative scalar multiples of functions from the defining class. Naturally, any finite sum of non-negative multiples of increasing concave functions is an increasing concave function. An analogous result holds for all of the other integral stochastic orders mentioned in this dissertation. However, the defining classes of these other integral stochastic orders are smaller, so the domain of the corresponding Lagrange multipliers is smaller.

1.2. Contribution

1.2.1. Scope. This dissertation has a wide scope. First, we consider multivariate increasing concave order constraints directly. The paper [**1**] is the only other work that uses multivariate increasing concave order constraints directly, rather than through an approximation. However, this paper does not emphasize the increasing concave order and only affine random-variable-valued mappings are studied.

Second, we allow a general underlying probability space via sampling. The recent paper [**16**] pioneers the study of sample average approximation for stochastic order constrained optimization. This paper focuses on a specific relaxation of the increasing concave order that is built from the univariate increasing concave order and linear combinations of the components of random vectors. This dissertation extends the results of [**16**] to the increasing concave stochastic order. With a firm theoretical foundation for sample average approximation, we can handle general underlying probability spaces in applications.

1.2.2. Optimality conditions and duality theory. Three main problem classes are presented in this dissertation. Optimality conditions are developed for each of these problem classes. All three sets of optimality conditions follow from the necessary and sufficient conditions for optimality in Banach spaces (see [**4**, Chapter 3]). We contribute a transformation of these conditions that shows increasing concave functions are the Lagrange multipliers of increasing concave stochastic order constraints. This transformation is applied in Chapters 3 and 6. Because the increasing concave functions are the Lagrange multipliers of increasing concave stochastic order constraints, the duals of the problems in Chapters 3 and 6 can be defined in terms of these functions.

1.2.3. Sample average approximation. Most probability spaces encountered in practice have an enormous number of scenarios. Sample average approximation is a common methodology for estimating an underlying probability distribution. We verify that sample average approximation is consistent for the three problem classes of Chapter 3.

We also set an important convention for sampling for stochastic order constrained problems. We suppose that sampling for the random-variable-valued mapping and sampling for the benchmark are conducted separately. This assumption generalizes earlier work without much difficulty. Finally, we make some initial observations about a possible large-scale implementation scheme for sample average approximation. As will be shown, sample average approximation is only part of a practical implementation strategy.

1.2.4. Robust optimization. It is possible to combine increasing concave stochastic order constraints with robust optimization. We consider uncertainty about the underlying probability distribution. We study the properties of this problem class and provide optimality conditions.

We emphasize this problem class for finite probability spaces. The large-scale implementation strategies for sample average approximation are extended to handle these robust optimization problems as well.

1.2.5. Multi-period optimization. We are able to extend the increasing concave stochastic order for use in multi-period optimization. We define an increasing concave stochastic order constraint on a vector of performance measures across time in a finite horizon control problem. We derive conditions for an optimal policy. Increasing concave functions appear again as the Lagrange multipliers of the increasing concave stochastic constraint in the multi-period case. This discovery lets us define an auxiliary control problem that can be solved with dynamic programming methods.

We also make some recommendations about solving this problem class for finite probability spaces. We can apply the coupling transformation used in Chapter 3 to the multi-period problem and use duality decomposition.

CHAPTER 2

Literature review

2.1. Stochastic optimization

This section is devoted to reviewing fundamental material from analysis and probability that will be used throughout this dissertation.

2.1.1. Analysis. Let $\mathbb{R}^N$ be the space of real $N-$dimensional vectors, and let $\mathbb{R}^{M\times N}$ be the space of real $M \times N-$dimensional matrices. Let e_n be the n^{th} unit vector in $\mathbb{R}^N$ and $\underline{1}$ be the vector of all ones.

For $x = (x_1, \ldots, x_N) \in \mathbb{R}^N$, define the norm

$$\|x\|_p \triangleq \left(\sum_{n=1}^{N} |x_n|^p \right)^{1/p}$$

for $1 \le p < \infty$ and the supremum norm

$$\|x\|_\infty \triangleq \sup_{n=1,\ldots,N} \{|x_n|\} .$$

Let

$$B_x(\epsilon) \triangleq \left\{ y \in \mathbb{R}^N : \|x - y\|_2 < \epsilon \right\} \subset \mathbb{R}^N$$

be the open ball of radius $\epsilon > 0$ (with respect to the $\|\cdot\|_2$ norm) centered at $x \in \mathbb{R}^N$.

We define $\mathcal{C}\left(\mathbb{R}^M, \mathbb{R}^N\right)$ to be the space of all continuous functions from $\mathbb{R}^M$ to $\mathbb{R}^N$, and use the shorthand $\mathcal{C}\left(\mathbb{R}^M\right)$ for $\mathcal{C}\left(\mathbb{R}^M, \mathbb{R}\right)$. The norm on $\mathcal{C}\left(\mathbb{R}^M\right)$ is

$$\|f\|_{\mathcal{C}(\mathbb{R}^M)} = \sup_{z \in \mathbb{R}^M} |f(z)|.$$

For a general compact set $Z \subset \mathbb{R}^M$, we let $\mathcal{C}\left(Z; \mathbb{R}^N\right)$ denote the space of continuous functions $f : Z \to \mathbb{R}^N$. Similarly, $\mathcal{C}(Z)$ indicates $\mathcal{C}(Z; \mathbb{R})$. Let $\mathcal{C}_+(Z)$ denote the space of nonnegative continuous functions $f : Z \to \mathbb{R}$. The norm on $\mathcal{C}(Z)$ is

$$\|f\|_{\mathcal{C}(Z)} = \sup_{z \in Z} |f(z)|.$$

Let $\mathcal{M}(Z)$ denote the space of finite regular Borel measures on Z, and define the duality pairing between $f \in \mathcal{C}(Z)$ and $\Lambda \in \mathcal{M}(Z)$ as

$$\langle \Lambda, f \rangle = \int_Z f(z)\, d\Lambda(z) .$$

Define the dual cone to $\mathcal{C}_+(Z)$ as:

$$\mathcal{M}_+(Z) \triangleq \{\Lambda \in \mathcal{M}(Z) : \langle \Lambda, f \rangle \ge 0, \forall f \in \mathcal{C}_+(Z)\} .$$

A set $Z \subset \mathbb{R}^M$ is convex if $\alpha\, z_1 + (1-\alpha)\, z_2 \in Z$ for all $z_1, z_2 \in Z$ and $0 \leq \alpha \leq 1$. Suppose Z is convex, then $f \in \mathcal{C}(Z)$ is convex if

$$f(\alpha\, z_1 + (1-\alpha)\, z_2) \leq \alpha\, f(z_1) + (1-\alpha)\, f(z_2)$$

and $f \in \mathcal{C}(Z)$ is concave if

$$f(\alpha\, z_1 + (1-\alpha)\, z_2) \geq \alpha\, f(z_1) + (1-\alpha)\, f(z_2)$$

for all $z_1, z_2 \in Z$ and $0 \leq \alpha \leq 1$.

Let $\leq_{\mathbb{R}^N}$ denote the usual component-wise order on $\mathbb{R}^N$. We say $f \in \mathcal{C}\left(Z, \mathbb{R}^N\right)$ is convex if

$$f(\alpha\, z_1 + (1-\alpha)\, z_2) \leq_{\mathbb{R}^N} \alpha\, f(z_1) + (1-\alpha)\, f(z_2)$$

and $f \in \mathcal{C}(Z)$ is concave if

$$f(\alpha\, z_1 + (1-\alpha)\, z_2) \geq_{\mathbb{R}^N} \alpha\, f(z_1) + (1-\alpha)\, f(z_2)$$

for all $z_1, z_2 \in Z$ and $0 \leq \alpha \leq 1$.

For a convex function $f \in \mathcal{C}(Z)$, the subdifferential of f at $z_0 \in Z$ is

$$\partial f(z_0) \triangleq \left\{ s \in \mathbb{R}^M : f(z) \geq f(z_0) + \langle s, z - z_0 \rangle, \forall z \in Z_0 \right\}.$$

We define the subdifferential of a concave function $f \in \mathcal{C}(Z)$ to be

$$\partial f(z_0) \triangleq -\partial\left[-f(z_0)\right].$$

For a set $A \subset \mathbb{R}^M$, we use $\operatorname{int} A$ to denote the interior and $\operatorname{cl} A$ to denote the closure of A, respectively. Define $\operatorname{conv} A$ to be the set of all finite convex combinations $\sum_{i=1}^{I} \alpha_i z_i$ for $z_i \in A$ and $\alpha_i \geq 0$ for all $i = 1, \ldots, I$, and $\sum_{i=1}^{I} \alpha_i = 1$. Define $\operatorname{cone} A$ to be the set of all finite conic combinations $\sum_{i=1}^{I} \alpha_i z_i$ for $z_i \in A$ and $\alpha_i \geq 0$ for all $i = 1, \ldots, I$, and observe that $\operatorname{cone} A$ is a convex set.

The diameter of a set $A \subset \mathbb{R}^M$ is

$$\mathbb{D}(A) = \sup_{x,\, y \in A} \|x - y\|.$$

Using the definitions from [**25**, Chapter 7], for sets A and B in $\mathbb{R}^M$ the deviation of the set A from the set B is

$$\mathbb{D}(A, B) = \sup_{x \in A} \inf_{y \in B} \|x - y\|$$

and the Hausdorff distance between A and B is

$$\mathbb{H}(A, B) = \min\left\{\mathbb{D}(A, B), \mathbb{D}(B, A)\right\}.$$

2.1.2. Probability. We denote a probability space as $(\Omega, \mathcal{F}, P)$ where Ω is a sample space, $\mathcal{F}$ is a $\sigma-$algebra on Ω, and P is a probability measure on $(\Omega, \mathcal{F})$. Elements $\omega \in \Omega$ are often referred to as scenarios or sample paths. We define $\mathcal{L}_p^1(\Omega, \mathcal{F}, P) \triangleq \mathcal{L}_p(\Omega, \mathcal{F}, P; \mathbb{R})$ for $1 \leq p < \infty$ to be the space of all $\mathcal{F}-$measurable mappings $X : \Omega \to \mathbb{R}$ with $\|X\|_p = \left(\mathbb{E}\left[|X|^p\right]\right)^{1/p} < \infty$. The set

$$\mathcal{L}_\infty^1(\Omega, \mathcal{F}, P) \triangleq \mathcal{L}_\infty(\Omega, \mathcal{F}, P; \mathbb{R})$$

is the space of all essentially bounded measurable mappings $X : \Omega \to \mathbb{R}$. The norm on $\mathcal{L}_\infty^1(\Omega, \mathcal{F}, P)$ is

$$\|X\|_\infty = \inf\{\alpha \geq 0 : P(|X(\omega)| > \alpha) = 0\}.$$

The symbol

$$\mathcal{L}_0^1(\Omega, \mathcal{F}, P) \triangleq \mathcal{L}_0(\Omega, \mathcal{F}, P; \mathbb{R})$$

represents the space of all measurable mappings $X : \Omega \to \mathbb{R}$. The value of a random variable $X \in \mathcal{L}_0^1(\Omega, \mathcal{F}, P)$ at $\omega \in \Omega$ is denoted $X(\omega)$.

The moment generating function of a random variable $X \in \mathcal{L}_0^1(\Omega, \mathcal{F}, P)$ is

$$M_X(s) = \mathbb{E}\left[e^{s\,X}\right].$$

For $N \geq 2$, we define $\mathcal{L}_p^N(\Omega, \mathcal{F}, P) \triangleq \mathcal{L}_p\left(\Omega, \mathcal{F}, P; \mathbb{R}^N\right)$ for $1 \leq p < \infty$ to be the space of all $\mathcal{F}-$measurable mappings $\underline{\mathrm{X}} : \Omega \to \mathbb{R}^N$ with $\left(\mathbb{E}\left[(\|\underline{\mathrm{X}}(\omega)\|_2)^p\right]\right)^{1/p} < \infty$. Vector-valued random variables are written as $\underline{\mathrm{X}} = (X_1, \ldots, X_N)$. The set

$$\mathcal{L}_\infty^N(\Omega, \mathcal{F}, P) \triangleq \mathcal{L}_\infty\left(\Omega, \mathcal{F}, P; \mathbb{R}^N\right)$$

is the space of all essentially bounded measurable mappings $\underline{\mathrm{X}} : \Omega \to \mathbb{R}^N$. The norm on $\mathcal{L}_\infty^N(\Omega, \mathcal{F}, P)$ is

$$\|\underline{\mathrm{X}}\|_\infty = \inf\{\alpha \geq 0 : P(\|\underline{\mathrm{X}}(\omega)\|_2 > \alpha) = 0\}.$$

The symbol

$$\mathcal{L}_0^N(\Omega, \mathcal{F}, P) = \mathcal{L}_0\left(\Omega, \mathcal{F}, P; \mathbb{R}^N\right)$$

represents the space of all measurable mappings $\underline{\mathrm{X}} : \Omega \to \mathbb{R}^N$. The value of a vector-valued random variable $\underline{\mathrm{X}} \in \mathcal{L}_0^N(\Omega, \mathcal{F}, P)$ at $\omega \in \Omega$ is denoted $\underline{\mathrm{X}}(\omega)$.

For vector-valued random variables $\underline{\mathrm{X}}$ and $\underline{\mathrm{Y}}$, the notation $\underline{\mathrm{X}} = \underline{\mathrm{Y}}$ means $\underline{\mathrm{X}}(\omega) = \underline{\mathrm{Y}}(\omega)$ for $P-$almost all $\omega \in \Omega$ and the notation $\underline{\mathrm{X}} \leq \underline{\mathrm{Y}}$ means $\underline{\mathrm{X}}(\omega) \leq \underline{\mathrm{Y}}(\omega)$ for $P-$almost all $\omega \in \Omega$.

Each random variable $\underline{\mathrm{X}} \in \mathcal{L}_p^N$ induces a measure on $\left(\mathbb{R}^N, \mathcal{B}^N\right)$ where $\mathcal{B}^N$ is the Borel $\sigma-$algebra on $\mathbb{R}^N$. For the mapping $\underline{\mathrm{X}}^{-1} : \left(\mathbb{R}^N, \mathcal{B}^N\right) \to (\Omega, \mathcal{F})$, define the measure $P \circ \underline{\mathrm{X}}^{-1}(A) = P\left(\underline{\mathrm{X}}^{-1}(A)\right)$ for $A \in \mathcal{B}^N$. We denote this measure as $P \circ \underline{\mathrm{X}}^{-1} : \left(\mathbb{R}^N, \mathcal{B}^N\right) \to \mathbb{R}$ where $\circ$ denotes composition. The support of the measure $P \circ \underline{\mathrm{X}}^{-1}$ is:

$$\operatorname{supp}\left(P \circ \underline{\mathrm{X}}^{-1}\right) = \left\{x \in \mathbb{R}^N : P \circ \underline{\mathrm{X}}^{-1}(B_x(\epsilon)) > 0,\ \forall \epsilon > 0\right\}.$$

2.1.3. Random-variable-valued mappings. Stochastic optimization is built around the idea of random-variable-valued mappings. Let $Z_0 \subset \mathbb{R}^M$ be the compact convex decision set of a stochastic optimization problem for the rest of this dissertation.

We introduce the random-variable-valued mapping $\underline{\mathrm{G}}(z) : Z_0 \to \mathcal{L}_1^N(\Omega, \mathcal{F}, P)$ to represent a random variable that is conditioned by a decision $z \in Z_0$ within an optimization problem. The value of the random variable $\underline{\mathrm{G}}(z)$ at $\omega \in \Omega$ is denoted $[\underline{\mathrm{G}}(z)](\omega)$. The random-variable-valued mapping $\underline{\mathrm{G}}(z)$ is continuous and concave if $[\underline{\mathrm{G}}(z)](\omega)$ is continuous and concave in $z \in Z_0$ for $P-$almost all $\omega \in \Omega$ (this definition is conceived in [**8**]).

Risk measures, mappings from the space of random variables to the scalars, have received significant attention as a means to manage risk in $\underline{\mathrm{G}}(z)$ ([**23, 22**]). A general stochastic optimization problem with a risk measure looks like

$$\max\left\{f(z)+\rho\left(\underline{\mathrm{G}}(z)\right)\,:\,z\in Z_0\right\}, \tag{1}$$

where $f(z)\in\mathcal{C}(Z_0)$ is concave and

$$\rho:\,\mathcal{L}_1^N(\Omega,\mathcal{F},P)\to\mathbb{R}$$

is increasing and concave. For increasing, we mean $\rho(\underline{\mathrm{X}}_1)\le\rho(\underline{\mathrm{X}}_2)$ for $\underline{\mathrm{X}}_1,\underline{\mathrm{X}}_2\in\mathcal{L}_1^N(\Omega,\mathcal{F},P)$ if $\underline{\mathrm{X}}_1(\omega)\le\underline{\mathrm{X}}_2(\omega)$ for $P-$almost all $\omega\in\Omega$. Concavity is defined similarly, for $\underline{\mathrm{X}}_1,\underline{\mathrm{X}}_2\in\mathcal{L}_1^N(\Omega,\mathcal{F},P)$ and $0\le\alpha\le1$,

$$\rho\left(\alpha\,\underline{\mathrm{X}}_1+(1-\alpha)\,\underline{\mathrm{X}}_2\right)\le\alpha\,\rho(\underline{\mathrm{X}}_1)+(1-\alpha)\,\rho(\underline{\mathrm{X}}_2).$$

Problem (1) is a convex optimization problem under these assumptions.

2.2. Stochastic order constrained optimization

In this section we will present an extension of risk measures that motivates this dissertation.

2.2.1. Family of convex integral stochastic order constraints. Stochastic orders, including many of the ones we study in this dissertation, are examined at length in [**24**]. We should comment that the increasing concave order is just one of a huge family of multivariate stochastic orders. However, the choice of stochastic orders that can be used directly as constraints in convex optimization problems is limited when compared to the whole family of stochastic orders.

All of the stochastic orders discussed in this dissertation are integral stochastic orders (see [**17**]).

Definition 2.2.1. Let $\mathfrak{F}$ be a collection of measurable functions $f:\mathbb{R}^N\to\mathbb{R}$. For $\underline{\mathrm{X}},\underline{\mathrm{Y}}\in\mathcal{L}_1^N(\Omega,\mathcal{F},P)$, the relation $\underline{\mathrm{X}}\ge_{\mathfrak{F}}\underline{\mathrm{Y}}$ on $\mathcal{L}_1^N(\Omega,\mathcal{F},P)$ is defined as

$$\int_\Omega f\left(\underline{\mathrm{X}}(\omega)\right)P(d\omega)\ge\int_\Omega f\left(\underline{\mathrm{Y}}(\omega)\right)P(d\omega),\qquad\forall f\in\mathfrak{F}.$$

Alternatively, integral stochastic orders can be defined on probability measures.

Definition 2.2.2. Let $\mathfrak{F}$ be a collection of measurable functions $f:\mathbb{R}^N\to\mathbb{R}$. For probability measures P and Q on $\left(\mathbb{R}^N,\,\mathcal{B}^N\right)$, the relation $P\ge_{\mathfrak{F}}Q$ is defined as

$$\int_{\mathbb{R}^N}f(x)\,P(dx)\ge\int_{\mathbb{R}^N}f(x)\,Q(dx),\qquad\forall f\in\mathfrak{F}.$$

Introduce a fixed benchmark random variable $\underline{\mathrm{Y}}\in\mathcal{L}_1^N(\Omega,\mathcal{F},P)$. Integral stochastic order constrained optimization problems are of the form:

$$\max_{z\in Z_0}\quad f(z) \tag{2}$$

$$\text{s.t.}\quad \underline{\mathrm{G}}(z)\ge_{\mathfrak{F}}\underline{\mathrm{Y}}. \tag{3}$$

Now we review the choices of $\mathfrak{F}$ that have been studied so far in problem $(2)-(3)$. We define $\mathcal{U}^1\subset\mathcal{C}(\mathbb{R})$ to be the set of all increasing concave functions $u:\mathbb{R}\to\mathbb{R}$ such that

$$\sup\left\{|s|\,:\,s\in\partial\,u(x)\,,\,x\in\mathbb{R}\right\}\le1.$$

The choice of the constant 1 is arbitrary, we just need the entire set $\mathcal{U}^1$ to be equicontinuous. There is no loss of generality here because the inequality

$$\mathbb{E}\left[u\left(\underline{\mathrm{X}}\right)\right] \leq \mathbb{E}\left[u\left(\underline{\mathrm{Y}}\right)\right]$$

is equivalent to $\mathbb{E}\left[\alpha\, u\left(\underline{\mathrm{X}}\right)\right] \leq \mathbb{E}\left[\alpha\, u\left(\underline{\mathrm{Y}}\right)\right]$ for all $\alpha > 0$.

Definition 2.2.3. For random variables $X,\, Y \in \mathcal{L}_1^1\left(\Omega, \mathcal{F}, P\right)$, if

$$\mathbb{E}\left[u\left(X\right)\right] \geq \mathbb{E}\left[u\left(Y\right)\right], \qquad \forall u \in \mathcal{U}^1,$$

then X is larger than Y in the univariate increasing concave order.

The field of stochastic order constrained optimization is originally established in **[7]** with univariate increasing concave stochastic order constraints. This work is extended in **[8]** which introduces the idea of random-variable-valued mappings into stochastic order constrained programs.

Using the techniques from **[11]**, we can incorporate uncertainty about the underlying probability distribution P. Let P_0 be a fixed probability measure on $(\Omega, \mathcal{F})$. The topological dual space of $\mathcal{L}_1^1\left(\Omega, \mathcal{F}, P_0\right)$ is $\mathcal{L}_\infty^1\left(\Omega, \mathcal{F}, P_0\right)$. We adopt the following convention from **[11]**. Any measure Q that is absolutely continuous with respect to P_0 with Radon-Nikodym derivative $\frac{dQ}{dP_0}$ in $\mathcal{L}_\infty^1\left(\Omega, \mathcal{F}, P_0\right)$ can be considered as element of $\mathcal{L}_\infty^1\left(\Omega, \mathcal{F}, P_0\right)$. For $X \in \mathcal{L}_1^1\left(\Omega, \mathcal{F}, P_0\right)$ and $Q \in \mathcal{L}_\infty^1\left(\Omega, \mathcal{F}, P_0\right)$ we define

$$\begin{aligned} \mathbb{E}_Q\left[X\right] &= \langle Q, X\rangle \\ &= \int_\Omega X\left(\omega\right) Q\left(d\omega\right) \\ &= \int_\Omega X\left(\omega\right) \frac{dQ}{dP_0}\left(\omega\right) P_0\left(d\omega\right). \end{aligned}$$

We let $\mathcal{Q}$ represent a set of probability measures and we assume that $\mathcal{Q}$ is convex, closed, and bounded in $\mathcal{L}_\infty^1\left(\Omega, \mathcal{F}, P_0\right)$. Set

$$B = \sup_{P\in\mathcal{Q}} \|\frac{dP}{dP_0}\|_\infty < \infty.$$

Definition 2.2.4. For random variables $X,\, Y \in \mathcal{L}_1^1\left(\Omega, \mathcal{F}, P_0\right)$, if

$$\mathbb{E}_P\left[u\left(X\right)\right] \geq \mathbb{E}_P\left[u\left(Y\right)\right], \qquad \forall u \in \mathcal{U}^1, \qquad \forall P \in \mathcal{Q},$$

then X is larger than Y in the robust increasing concave order with respect to $\mathcal{Q}$.

Multivariate stochastic orders have also been embedded into problem (2) − (3). The following three stochastic orders share the same essence. Each defines a multivariate stochastic order by applying the univariate increasing concave order to weighted sums of the components of random vectors. In the following definitions, we will use $\geq_{icv}$ as shorthand for the univariate increasing concave stochastic order. This choice is also an abuse of notation, because after this subsection we will always use $\geq_{icv}$ to denote the multivariate increasing concave stochastic order.

In **[10]**, a multivariate extension of definition 2.2.3 is proposed and embedded into stochastic optimization problems.

Definition 2.2.5. For random vectors $\underline{X}$, $\underline{Y} \in \mathcal{L}_1^N(\Omega, \mathcal{F}, P)$, if

$$\langle \alpha, \underline{X} \rangle \geq_{icv} \langle \alpha, \underline{Y} \rangle, \quad \forall \alpha \in \mathbb{R}_+^N,$$

then $\underline{X}$ is larger than $\underline{Y}$ in the multivariate linear increasing concave order.

In **[6]** a more general version of the preceding definition is proposed.

Definition 2.2.6. For random vectors $\underline{X}$, $\underline{Y} \in \mathcal{L}_1^N(\Omega, \mathcal{F}, P)$, for a polyhedral convex set $V \subset \mathbb{R}^N$ if

$$\langle v, \underline{X} \rangle \geq_{icv} \langle v, \underline{Y} \rangle, \quad \forall v \in V,$$

then $\underline{X}$ is larger than $\underline{Y}$ in the multivariate polyhedral linear increasing concave order.

This work is further extended in **[16]**, where the linear weights are allowed to range over general convex sets in $\mathbb{R}^N$.

Definition 2.2.7. For random vectors $\underline{X}$, $\underline{Y} \in \mathcal{L}_1^N(\Omega, \mathcal{F}, P)$, for a convex set $C \subset \mathbb{R}^N$ if

$$\langle c, \underline{X} \rangle \geq_{icv} \langle c, \underline{Y} \rangle, \quad \forall c \in C,$$

then $\underline{X}$ is larger than $\underline{Y}$ in the multivariate convex linear increasing concave order.

2.2.2. Increasing concave stochastic order constraints. We now present the central integral stochastic order of this dissertation. Define $\mathcal{U}^N \subset \mathcal{C}\left(\mathbb{R}^N\right)$ to be the set of all increasing concave functions $u : \mathbb{R}^N \to \mathbb{R}$ such that

$$\sup\left\{\|s\|_2 : s \in \partial_x u(x), x \in \mathbb{R}^N\right\} \leq 1.$$

Recall that on $\mathbb{R}^N$ all norms are equivalent. We have

$$C_1 \|s\|_{p_1}\| \leq \|s\|_{p_2} \leq C_2 \|s\|_{p_1}$$

for constants $0 < C_1, C_2 < \infty$ for any $1 \leq p_1, p_2 \leq \infty$. Thus, if

$$\sup\left\{\|s\|_p : s \in \partial_x u(x), x \in \mathbb{R}^N\right\} \leq 1,$$

for any $1 \leq p \leq \infty$, it follows

$$\sup\left\{\|s\|_q : s \in \partial_x u(x), x \in \mathbb{R}^N\right\} \leq C < \infty,$$

for any $1 \leq q \leq \infty$ for some $C < \infty$.

We define the multivariate increasing concave order as follows:

Definition 2.2.8. For random vectors $\underline{X}$, $\underline{Y} \in \mathcal{L}_1^N(\Omega, \mathcal{F}, P)$, if

$$\mathbb{E}\left[u\left(\underline{X}\right)\right] \geq \mathbb{E}\left[u\left(\underline{Y}\right)\right], \quad \forall u \in \mathcal{U}^N,$$

then $\underline{X}$ is larger than $\underline{Y}$ in the increasing concave order (denoted $\underline{X} \geq_{icv} \underline{Y}$).

The relation $\underline{X} \geq_{icv} \underline{Y}$ can be defined in terms of coupling. Recall that for random variables $\underline{X}, \underline{Y} \in \mathcal{L}_1^N(\Omega, \mathcal{F}, P)$, $\underline{X} =_d \underline{Y}$ denotes equality in distribution. The relation $\underline{X} =_d \underline{Y}$ is defined as $P\left(\{\underline{X} \leq \eta\}\right) = P\left(\{\underline{Y} \leq \eta\}\right)$ for all $\eta \in \mathbb{R}^N$.

Theorem 2.2.9. [**24**, Theorem 7.A.2] *Two random vectors* $\underline{X}$, $\underline{Y} \in \mathcal{L}_1^N(\Omega, \mathcal{F}, P)$ *satisfy* $\underline{X} \geq_{icv} \underline{Y}$ *if, and only if, there exist two random vectors* $\underline{X}'$, $\underline{Y}' \in \mathcal{L}_1^N(\Omega, \mathcal{F}, P)$, *such that*

$$\underline{X}' =_d \underline{X}, \tag{4}$$

$$\underline{Y}' =_d \underline{Y}, \tag{5}$$

$$\underline{X}' \geq \mathbb{E}\left[\underline{Y}' \mid \underline{X}'\right], \qquad P-\textit{almost surely.} \tag{6}$$

The following corollary is a powerful application of theorem 2.2.9 for finite probability spaces.

Corollary 2.2.10. [**1**, Corollary 2] *Let* $\Omega_1 = (\omega_{11}, \ldots, \omega_{1J})$ *and* $\Omega_2 = (\omega_{21}, \ldots, \omega_{2K})$ *and let* $\mathcal{F}_1$ *and* $\mathcal{F}_2$ *make all atoms on* Ω_1 *and* Ω_2 *measurable, respectively. Then for* $\underline{X} \in \mathcal{L}_1^N(\Omega_1, \mathcal{F}_1, P_1)$ *and* $\underline{Y} \in \mathcal{L}_1^N(\Omega_2, \mathcal{F}_2, P_2)$*, the constraint* $\underline{X} \geq_{icv} \underline{Y}$ *is equivalent to the existence of* $\pi_{jk} \geq 0$ *for* $j = 1, \ldots, J$ *and* $k = 1, \ldots, K$ *such that*

$$\underline{X}(\omega_{1j}) \geq \frac{\sum_{k=1}^{K} \pi_{jk} \underline{Y}(\omega_{2k})}{P_1(\{\omega_{1j}\})}, \quad j = 1, \ldots, J, \tag{7}$$

$$\sum_{k=1}^{K} \pi_{jk} = P_1(\{\omega_{1j}\}), \quad j = 1, \ldots, J, \tag{8}$$

$$\sum_{j=1}^{J} \pi_{jk} = P_2(\{\omega_{2k}\}), \quad k = 1, \ldots, K, \tag{9}$$

$$\pi \geq 0. \tag{10}$$

2.2.3. Relaxations of increasing concave stochastic order constraints. We now review work from [**14**] that serves as a prelude to this dissertation. In [**14**], a family of relaxations of the increasing concave order is presented based on semi-infinite programming and nonlinear programming.

Choose a compact convex set $W \subset \mathbb{R}^N$ with a nonempty interior. Define

$$\|u\|_W = \sup_{x \in W} |u(x)|$$

and

$$\|\partial u\|_W = \sup\{\|s\|_2 : s \in \partial u(x), x \in W\},$$

for all $u \in \mathcal{U}^N$.

Define $\mathcal{U}^N(W)$ to be the set of all $u \in \mathcal{U}^N$ restricted to the domain W such that:

- $u(w_0) = 0$ where w_0 is the least upper bound of W.
- $\|\partial u\|_W \leq 1$ for all $u \in \mathcal{U}^N(W)$.

Since W is a convex and compact set in $\mathbb{R}^N$, its least upper bound exists and is finite. Under these two assumptions $\mathcal{U}^N(W)$ is uniformly bounded and equicontinuous. It is also compact in $\mathcal{C}(W)$.

For a compact set $\Xi \subset \mathbb{R}^A$ with $A < \infty$, we define $U_\Xi \subset \mathcal{U}^N(W)$ to be a family of increasing concave functions parametrized by $\xi \in \Xi$. The dependence of

U_Ξ on W is left implicit from now on for cleaner notation. It is understood that all functions in U_Ξ are drawn from $\mathcal{U}^N(W)$ and are restricted to the domain W. All $u \in U_\Xi$ are written as $u(\cdot;\xi) : \mathbb{R}^N \to \mathbb{R}$ to indicate the dependence on the parameters.

The function-valued mapping $\boldsymbol{u} : \Xi \to U_\Xi$ maps the parameter space Ξ to the function space U_Ξ. In particular, $[\boldsymbol{u}(\xi)](x) = u(x;\xi)$. We assume the following technical conditions are satisfied by U_Ξ and the associated mapping $\boldsymbol{u}$.

- $\boldsymbol{u}$ is a bijection.
- $\boldsymbol{u}$ and $\boldsymbol{u}^{-1}$ are continuous.
- U_Ξ is equicontinuous and uniformly bounded.

The mapping $\boldsymbol{u}$ is a homeomorphism under these assumptions. Further, $u(x;\xi)$ is continuous in ξ for all $x \in W$.

Consider the problem

$$\text{(SIP)} \qquad \max_{z \in Z_0} \quad f(z)$$
$$\text{(11)} \qquad \text{s.t.} \quad \mathbb{E}[u(\underline{\mathrm{G}}(z))] \geq \mathbb{E}[u(\underline{\mathrm{Y}})], \qquad \forall u \in U_\Xi.$$

The notation (SIP) stands for semi-infinite program, since problem (SIP) is equivalent to:

$$\begin{aligned} \max_{z \in Z_0} \quad & f(z) \\ \text{s.t.} \quad & \mathbb{E}[u(\underline{\mathrm{G}}(z);\xi)] \geq \mathbb{E}[u(\underline{\mathrm{Y}};\xi)], \qquad \forall \xi \in \Xi, \end{aligned}$$

using the definitions of U_Ξ and $\boldsymbol{u}$. We define the feasible region, set of optimal solutions, and optimal value of problem (SIP) as

$$\begin{aligned} Z(U_\Xi) &\triangleq \{z \in Z_0 : \mathbb{E}[u(\underline{\mathrm{G}}(z))] \geq \mathbb{E}[u(\underline{\mathrm{Y}})], \forall u \in U_\Xi\}, \\ S(U_\Xi) &\triangleq \arg\max\{f(z) : z \in Z(U_\Xi)\}, \\ \nu(U_\Xi) &\triangleq \max\{f(z) : z \in Z(U_\Xi)\}, \end{aligned}$$

to emphasize the dependence on U_Ξ. The constraint (11) is an integral stochastic order, but it is a relaxation of the increasing concave stochastic order since $U_\Xi \subset \mathcal{U}^N(W)$.

Optimality conditions for problem (SIP) follow from the usual optimality conditions for convex optimization. We introduce the Lagrangian

$$L(z,u) = f(z) + \mathbb{E}[u(\underline{\mathrm{G}}(z)) - u(\underline{\mathrm{Y}})],$$

where utility functions $u \in \operatorname{cl}\operatorname{cone}\mathcal{U}^N(W)$ act as Lagrange multipliers for the $\geq_{icv}$ constraint. We use the following Slater condition for problem (SIP).

Assumption 2.2.11. *There exists $\tilde{z} \in Z_0$ such that*

$$\mathbb{E}[u(\underline{\mathrm{G}}(\tilde{z}))] > \mathbb{E}[u(\underline{\mathrm{Y}})], \qquad \forall u \in U_\Xi.$$

Assumption 2.2.11 depends on the choice of $\underline{\mathrm{G}}(z)$, if $\underline{\mathrm{G}}(z) \not\geq_{icv} \underline{\mathrm{Y}}$ for all $z \in Z_0$ then the Slater condition cannot be satisfied. However, even if $\underline{\mathrm{G}}(z) \geq_{icv} \underline{\mathrm{Y}}$ for all $z \in Z_0$, assumption 2.2.11 still depends on the choice of U_Ξ. If we want to include functions $u \in \mathcal{U}^N(W)$ with arbitrarily small $\|\partial u\|_W$ in U_Ξ, then we must also include the zero function $u = 0$ by compactness. If U_Ξ contains the zero function, then assumption 2.2.11 cannot be satisfied.

There are intuitive sufficient conditions for problem (SIP) to satisfy assumption 2.2.11. In [**14**], the existence of a family $\{U_{\Xi_k}\}_{k=0}^{\infty}$ is assumed such that $U_{\Xi_k} \subset U_{\Xi_{k+1}}$ for all $k \geq 0$, $\bigcup_{k=0}^{\infty} U_{\Xi_k} = \mathcal{U}^N(W)$, and problem (SIP) over U_{Ξ_k} satisfies the Slater condition for all $k \geq 0$. There are actually many such sequences. For example, we can construct such a sequence out of increasing polyhedral concave functions. Increasing polyhedral concave functions on $\mathbb{R}^N$ with $L < \infty$ faces look like

$$f_{a,b}(x) = \min_{l \in 1,\ldots,L} \{\langle a_l, x\rangle + b_l\},$$

where $a_l \in \mathbb{R}^N_+$ and $b_l \in \mathbb{R}_+$. We restrict to $\|a_l\|_2 \leq 1$ for all $l = 1,\ldots,L$. To ensure that the Slater condition holds, we can take a sequence $\epsilon_k \downarrow 0$ and $L(k) \to \infty$ and define

$$U_{\Xi_k} = \left\{ f_{a,b} : \begin{array}{cc} \epsilon_k \leq \|a_l\|_2 \leq 1, & l = 1,\ldots,L(k), \\ |b_l| \leq 1, & l = 1,\ldots,L(k) \end{array} \right\}.$$

The resulting problems (SIP) over U_{Ξ_k} can all satisfy the Slater condition because U_{Ξ_k} consists of only strictly increasing functions. Also, as $\epsilon_k \to 0$ and $L(k) \to \infty$, it is known that U_{Ξ_k} approximates $\mathcal{U}^N(W)$ arbitrarily closely.

The following two theorems are found in [**14**] and follow from semi-infinite programming optimality conditions (see [**4**, Section 5.4]) and an application of Fubini's theorem.

Theorem 2.2.12. *Suppose assumption 2.2.11 holds. If $\hat{z}$ solves problem (SIP), then there exists $\hat{u} \in \mathrm{cl\,cone}\, U_\Xi$ such that*

$$L(\hat{z}, \hat{u}) = \max\{L(z, \hat{u}) : z \in Z_0\}, \tag{12}$$

$$\mathbb{E}[\hat{u}(\underline{\mathrm{G}}(\hat{z})) - \hat{u}(\underline{\mathrm{Y}})] = 0, \tag{13}$$

The dual functional for problem (SIP) is

$$\mathrm{d}(u) \triangleq \max\{L(z,u) : z \in Z_0\}$$

and the corresponding dual to problem (SIP) is

$$\min\{\mathrm{d}(u) : u \in \mathrm{cl\,cone}\, U_\Xi\}. \tag{SIP$_D$}$$

Strong duality holds between problem (SIP) and problem (SIP$_D$).

Theorem 2.2.13. *Suppose assumption 2.2.11 holds.*

(a) If $\hat{z}$ solves problem (SIP), then there is a solution to problem (SIP$_D$) and the optimal values are equal.

(b) If $\hat{u}$ is an optimal solution to problem (SIP$_D$), then any $\hat{z}$ satisfying the optimality conditions (12) – (13) *corresponding to $\hat{u}$ is optimal to problem (SIP).*

We can consider the same development for finitely many constraint functions. Let

$$\{u_1,\ldots,u_I\}$$

represent a finite set of functions $u_i \in \mathcal{U}^N(W)$ for $i = 1, \ldots, I$. We introduce the problem:

$$\text{(NLP)} \qquad \max_{z \in Z_0} \quad f(z)$$
$$(14) \qquad \text{s.t.} \quad \mathbb{E}\left[u_i\left(\underline{\mathrm{G}}(z)\right)\right] \geq \mathbb{E}\left[u_i\left(\underline{\mathrm{Y}}\right)\right], \qquad i = 1, \ldots, I.$$

The notation (NLP) stands for nonlinear program. We define the feasible region, set of optimal solutions, and optimal value of problem (NLP) as

$$\begin{aligned} Z(\{u_1, \ldots, u_I\}) &\triangleq \{z \in Z_0 : (14)\}, \\ S(\{u_1, \ldots, u_I\}) &\triangleq \arg\max\{f(z) : z \in Z(\{u_1, \ldots, u_I\})\}, \\ \nu(\{u_1, \ldots, u_I\}) &\triangleq \max\{f(z) : z \in Z(\{u_1, \ldots, u_I\})\}. \end{aligned}$$

Optimality conditions and duality theory for problem (NLP) follow directly from the nonlinear programming optimality conditions. The following theorem shows that there is an instance of problem (NLP) with $\{u_1, \ldots, u_I\} \subset U_\Xi$ that gives an optimal solution to problem (SIP) over U_Ξ.

Theorem 2.2.14. *Suppose assumption 2.2.11 holds. If $\hat{z}$ solves problem (SIP) over U_Ξ , then there exists $\hat{u} \in cl\, cone\, U_\Xi$ of the form*

$$\hat{u} = \sum_{m=1}^{M} \lambda_m u(\cdot; \xi_m)$$

for $\lambda_m \geq 0$ and $\xi_m \in \Xi$ for $m = 1, \ldots, M$, such that $(12) - (13)$ hold. It follows that $\hat{z} \in S(\{u_1, \ldots, u_M\})$.

CHAPTER 3

Optimization with increasing concave stochastic order constraints

3.1. Optimization problem

This chapter develops increasing concave stochastic order constrained programs in detail. First, we present a perturbation of increasing concave stochastic order constraints that lets us obtain optimality conditions and duality results. Second, this chapter motivates the idea that the Lagrange multipliers of increasing concave stochastic order constraints are increasing concave functions.

We will work with two distinct probability spaces $(\Omega_1, \mathcal{F}_1, P_1)$ and $(\Omega_2, \mathcal{F}_2, P_2)$, one for the random-variable-valued mapping $\underline{G}(z)$ and the other for the benchmark $\underline{Y}$. There is no loss of generality in assuming two separate probability spaces because the relation $\underline{X} \geq_{icv} \underline{Y}$ only depends on the marginal distributions of $\underline{X}$ and $\underline{Y}$. Further, it is easier to describe our upcoming coupling and sample average approximation schemes when the random-variable-valued mapping and the benchmark exist on two separate probability spaces. We let

$$(\Omega, \mathcal{F}, P) = (\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2, P_1 \times P_2)$$

be the probability space determined by the cross product of

$$(\Omega_1, \mathcal{F}_1, P_1) \text{ and } (\Omega_2, \mathcal{F}_2, P_2).$$

In particular, P is chosen so that the sigma-fields $\mathcal{F}_1$ and $\mathcal{F}_2$ are independent. This distinction is important because later we will construct measures on $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2)$ such that $\mathcal{F}_1$ and $\mathcal{F}_2$ are not independent.

We formally define the random-variable-valued mapping

$$\underline{G}(z) : Z_0 \to \mathcal{L}_1^N(\Omega_1, \mathcal{F}_1, P_1).$$

This mapping can be written explicitly as $[\underline{G}(z)](\omega) : Z_0 \times \Omega_1 \to \mathbb{R}^N$. Recall that $[\underline{G}(z)](\omega)$ is continuous and concave in $z \in Z_0$ for $P-$almost all $\omega \in \Omega$. We now make some common technical assumptions on $\underline{G}(z)$ as described in **[25, 16]**. The following assumption holds throughout Chapters 3, 4, and 5.

Assumption 3.1.1. *(a)* $[\underline{G}(z)](\omega)$ *is Lipschitz continuous on* Z_0 *for* P_1-*almost all* $\omega \in \Omega_1$. *There exists* $\Pi \in \mathcal{L}_1^1(\Omega_1, \mathcal{F}_1, P_1)$ *such that* $\| [\underline{G}(z_1)](\omega) - [\underline{G}(z_2)](\omega) \|_2 \leq \Pi(\omega) \|z_1 - z_2\|_2$ *for* P_1-*almost all* $\omega \in \Omega_1$. *Set* $\pi \triangleq \mathbb{E}[\Pi]$.

(b) The MGF of Π, *denoted* $M_\Pi(s)$, *is finite for* s *in a neighborhood of zero.*

We will use $\geq_{icv}$ to define a constraint on the random vector $\underline{G}(z)$. To complete the $\geq_{icv}$ constraint, we introduce a benchmark random vector

$$\underline{Y} = (Y_1, \ldots, Y_N) \in \mathcal{L}_1^N(\Omega_2, \mathcal{F}_2, P_2).$$

If $\operatorname{supp}\left(P_2 \circ \underline{\mathrm{Y}}^{-1}\right)$ is contained in W, then

$$\mathbb{E}\left[u\left(\underline{\mathrm{Y}}\right)\right] = \mathbb{E}\left[u\left(\underline{\mathrm{Y}}\right) 1\left\{\underline{\mathrm{Y}} \in W\right\}\right] = \mathbb{E}\left[u\left(\underline{\mathrm{Y}}\right) 1\left\{\underline{\mathrm{Y}} \in \operatorname{supp}\left(P_2 \circ \underline{\mathrm{Y}}^{-1}\right)\right\}\right].$$

Thus, to characterize $\underline{\mathrm{G}}(z) \geq_{icv} \underline{\mathrm{Y}}$ we only need to pay attention to functions u defined on W when W contains both $\operatorname{supp}\left(P_1 \circ \left[\underline{\mathrm{G}}(z)\right]^{-1}\right)$ and $\operatorname{supp}\left(P_2 \circ \underline{\mathrm{Y}}^{-1}\right)$. The expression $\mathbb{E}\left[u\left(\underline{\mathrm{G}}(z)\right) - u\left(\underline{\mathrm{Y}}\right)\right]$ is not altered by changing u outside of W. In line with this reasoning, we assume W contains $\operatorname{supp}\left(P_2 \circ \underline{\mathrm{Y}}^{-1}\right)$ and $\operatorname{supp}\left(P_1 \circ \left[\underline{\mathrm{G}}(z)\right]^{-1}\right)$ for all $z \in Z_0$. For clarity, $\left[\underline{\mathrm{G}}(z)\right]^{-1} : \mathbb{R}^N \to \Omega$ is the inverse of the random variable $\underline{\mathrm{G}}(z) \in \mathcal{L}_1^N\left(\Omega_1, \mathcal{F}_1, P_1\right)$, not the inverse of the mapping $\underline{\mathrm{G}}(z) : Z_0 \to \mathcal{L}_1^N\left(\Omega_1, \mathcal{F}_1, P_1\right)$.

The main optimization problem in this dissertation is:

$$\text{(MP)} \qquad \max_{z \in Z_0} \quad f(z)$$

$$\text{(1)} \qquad \text{s.t.} \quad \mathbb{E}\left[u\left(\underline{\mathrm{G}}(z)\right)\right] \geq \mathbb{E}\left[u\left(\underline{\mathrm{Y}}\right)\right], \qquad \forall u \in \mathcal{U}^N(W).$$

The notation (MP) stands for main problem. Problem (MP) is a semi-infinite programming problem, where the semi-infinite constraints are indexed by the compact set $\mathcal{U}^N(W)$ (see [**4**, Section 5.4]). We denote the feasible region, set of optimal solutions, and optimal value of problem (MP) as:

$$\begin{aligned} Z &\triangleq \left\{z \in Z_0 : (1)\right\}, \\ S &\triangleq \arg\max\left\{f(z) : z \in Z\right\}, \\ \nu &\triangleq \max\left\{f(z) : z \in Z\right\}. \end{aligned}$$

Under our assumptions about $\operatorname{supp}\left(P_1 \circ \left[\underline{\mathrm{G}}(z)\right]^{-1}\right)$ and $\operatorname{supp}\left(P_2 \circ \underline{\mathrm{Y}}^{-1}\right)$, if

$$\mathbb{E}\left[u\left(\underline{\mathrm{G}}(z)\right)\right] \geq \mathbb{E}\left[u\left(\underline{\mathrm{Y}}\right)\right], \quad \forall u \in \mathcal{U}^N(W),$$

then

$$\underline{\mathrm{G}}(z) \geq_{icv} \underline{\mathrm{Y}}.$$

3.1.1. Constraint perturbation. Problem (MP) does not satisfy the Slater condition because $u = 0$ is included in $\mathcal{U}^N(W)$. In [**14**], the relaxed problem (SIP) was designed to satisfy the Slater condition by construction. We take a different approach in this dissertation and perturb the constraints. For notational convenience define the function

$$g(z, u) \triangleq \mathbb{E}\left[u\left(\underline{\mathrm{G}}(z)\right)\right] - \mathbb{E}\left[u\left(\underline{\mathrm{Y}}\right)\right].$$

Introduce the perturbed problem:

$$(\epsilon - \text{MP}) \qquad \max_{z \in Z_0} \quad f(z)$$

$$\text{(2)} \qquad \text{s.t.} \quad g(z, u) \geq \epsilon, \qquad \forall u \in \mathcal{U}^N(W),$$

and its feasible region, set of optimal solutions, and optimal value:

$$
\begin{aligned}
Z^{\epsilon} &\triangleq \{z \in Z_0 : (2)\}, \\
S^{\epsilon} &\triangleq \arg\max \{f(z) : z \in Z^{\epsilon}\}, \\
\nu^{\epsilon} &\triangleq \max \{f(z) : z \in Z^{\epsilon}\}.
\end{aligned}
$$

Problem $(\epsilon - \text{MP})$ is able to satisfy the Slater condition for $\epsilon < 0$, cannot satisfy the Slater condition for $\epsilon = 0$, and is infeasible for $\epsilon > 0$ because $\mathcal{U}^N(W)$ includes the zero function.

Proposition 3.1.2. *Problem (ϵ-MP) is convex for $\epsilon \leq 0$.*

Proof. The objective $f(z)$ is concave by assumption.

Suppose Z^{ϵ} is nonempty. The mapping $[G_n(z)](\omega)$ is concave in z for P_1-almost all $\omega \in \Omega_1$ for $n = 1, \ldots, N$. Further, all $u \in \mathcal{U}^N(W)$ are component-wise increasing. Thus

$$u([G_1(z)](\omega), \ldots, [G_N(z)](\omega))$$

is concave in $z \in Z_0$ for P_1-almost all $\omega \in \Omega$ because it is the composition of the increasing concave function u with the component-wise concave function $([G_1(z)](\omega), \ldots, [G_N(z)](\omega))$. The function

$$\mathbb{E}[u(\underline{\mathrm{G}}(z))] = \int_{\Omega} u([\underline{\mathrm{G}}(z)](\omega)) P_1(d\omega)$$

is concave in z as the integral of the concave functions $u([\underline{\mathrm{G}}(z)](\omega))$ with respect to P_1. Each set $A_u^{\epsilon} = \{z : \mathbb{E}[u(\underline{\mathrm{G}}(z))] \geq \mathbb{E}[u(\underline{\mathrm{Y}})] + \epsilon\}$ is convex, and the intersection of the convex sets $\left\{\bigcap_{u \in \mathcal{U}^N(W)} A_u^{\epsilon}\right\} \bigcap Z_0$ is convex. Problem $(\epsilon-\text{MP})$ thus has a convex feasible region and is a convex optimization problem. □

We define a similar perturbation of problem (SIP):

$$
\begin{aligned}
&(\epsilon\text{-SIP}) \qquad \max_{z \in Z_0} \quad f(z) \\
&(3) \qquad\qquad \text{s.t.} \quad g(z, u) \geq \epsilon, \qquad \forall u \in U_{\Xi}.
\end{aligned}
$$

Denote the feasible region, set of optimal solutions, and optimal value of problem $(\epsilon-\text{SIP})$ as:

$$
\begin{aligned}
Z^{\epsilon}(U_{\Xi}) &\triangleq \{z \in Z_0 : (3)\}, \\
S^{\epsilon}(U_{\Xi}) &\triangleq \arg\max \{f(z) : z \in Z^{\epsilon}(U_{\Xi})\}, \\
\nu^{\epsilon}(U_{\Xi}) &\triangleq \max \{f(z) : z \in Z^{\epsilon}(U_{\Xi})\}.
\end{aligned}
$$

The perturbation of problem (NLP) is defined similarly:

$$
\begin{aligned}
&(\epsilon\text{-NLP}) \qquad \max_{z \in Z_0} \quad f(z) \\
&(4) \qquad\qquad \text{s.t.} \quad g(z, u_i) \geq \epsilon, \qquad i = 1, \ldots, I,
\end{aligned}
$$

and its feasible region, set of optimal solutions, and optimal value are:

$$\begin{aligned} Z^{\epsilon}(\{u_1,\ldots,u_I\}) &\triangleq \{z\in Z_0 : (4)\},\\ S^{\epsilon}(\{u_1,\ldots,u_I\}) &\triangleq \arg\max\{f(z) : z\in Z^{\epsilon}(\{u_1,\ldots,u_I\})\},\\ \nu^{\epsilon}(\{u_1,\ldots,u_I\}) &\triangleq \max\{f(z) : z\in Z^{\epsilon}(\{u_1,\ldots,u_I\})\}.\end{aligned}$$

3.1.2. Analytical properties. Define the new random variable:

$$\Phi\left(\left[\underline{\mathrm{G}}(z)\right](\omega),\underline{\mathrm{Y}}(\omega)\right)\triangleq\|\left[\underline{\mathrm{G}}(z)\right](\omega)\|_2+\|\underline{\mathrm{Y}}(\omega)\|_2.$$

The function $\Phi(\cdot)$ is significant because it dominates the random variable $u(\underline{\mathrm{G}}(z))-u(\underline{\mathrm{Y}})$ for all $u\in\mathcal{U}^N(W)$.

Assumption 3.1.3. *For all $z\in Z_0$, the MGF of $\Phi(\underline{G}(z),\underline{Y})$, denoted*

$$M_{\Phi\left(\underline{G}(z),\underline{Y}\right)}(s),$$

is finite in a neighborhood of zero.

Also define the function

$$\phi(\underline{\mathrm{G}}(z),\underline{\mathrm{Y}})\triangleq\mathbb{E}\left[\Phi\left(\left[\underline{\mathrm{G}}(z)\right](\omega),\underline{\mathrm{Y}}(\omega)\right)\right].$$

Proposition 3.1.4. *(a) For all $u\in\mathcal{U}^N(W)$,*

$$|u([\underline{G}(z)](\omega))-u(\underline{Y}(\omega))|\le\Phi([\underline{G}(z)](\omega),\underline{Y}(\omega))$$

for $P-$almost all $\omega\in\Omega$.

(b) $\phi(\cdot)$ is uniformly bounded on Z_0.

(c) The MGF of $u(\underline{G}(z))-u(\underline{Y})$ is finite in a neighborhood of zero for all $(z,u)\in Z_0\times\mathcal{U}^N(W)$.

Proof. (a) Recall the subdifferential mean value inequality [**2**, Theorem 2.1] to see that

$$|u([\underline{\mathrm{G}}(z)](\omega))-u(\underline{\mathrm{Y}}(\omega))|\le\|\partial u\|_W\|\left[\underline{\mathrm{G}}(z)\right](\omega)-\underline{\mathrm{Y}}(\omega)\|_2$$

for all $u\in\mathcal{U}^N(W)$, since all such u are subdifferentiable. Compute

$$\begin{aligned} &\sup_{u\in\mathcal{U}^N(W)}|u([\underline{\mathrm{G}}(z)](\omega))-u(\underline{\mathrm{Y}}(\omega))|\\ \le\;&\sup_{u\in\mathcal{U}^N(W)}\|\partial u\|_W\|\left[\underline{\mathrm{G}}(z)\right](\omega)-\underline{\mathrm{Y}}(\omega)\|_2\\ \le\;&\left(\sup_{u\in\mathcal{U}^N(W)}\|\partial u(\cdot)\|_W\right)(\|\left[\underline{\mathrm{G}}(z)\right](\omega)\|_2+\|\underline{\mathrm{Y}}(\omega)\|_2)\\ =\;&\Phi([\underline{\mathrm{G}}(z)](\omega),\underline{\mathrm{Y}}),\end{aligned}$$

using the assumption that

$$\sup_{u\in\mathcal{U}^N(W)}\|\partial u(\cdot)\|_W\le 1.$$

(b) Fix $z_0\in Z_0$. Under assumption 3.1.1, for all $z\in Z_0$ we have

$$\begin{aligned}\| [\underline{\mathrm{G}}(z)](\omega) \|_2 &\leq \| [\underline{\mathrm{G}}(z_0)](\omega) \|_2 + \Pi(\omega) \|z - z_0\|_2 \\ &\leq \| [\underline{\mathrm{G}}(z_0)](\omega) \|_2 + \Pi(\omega) \max_{z, z_0 \in Z_0} \|z - z_0\|_2 \\ &< \infty,\end{aligned}$$

using the fact that

$$\mathbb{D}(Z_0) = \max_{z, z_0 \in Z_0} \|z - z_0\|_2 < \infty$$

by compactness of Z_0. Then

$$\begin{aligned}\Phi([\underline{\mathrm{G}}(z)](\omega), \underline{\mathrm{Y}}(\omega)) &\leq \| [\underline{\mathrm{G}}(z_0)](\omega) \|_2 + \Pi(\omega) \mathbb{D}(Z_0) + \|\underline{\mathrm{Y}}(\omega)\|_2 \\ &= \Phi([\underline{\mathrm{G}}(z_0)](\omega), \underline{\mathrm{Y}}(\omega)) + \Pi(\omega) \mathbb{D}(Z_0).\end{aligned}$$

Set $[\underline{\mathrm{G}}(z)](\omega_1, \omega_2) = [\underline{\mathrm{G}}(z)](\omega_1)$, $\Pi(\omega_1, \omega_2) = \Pi(\omega_1)$, and $\underline{\mathrm{Y}}(\omega_1, \omega_2) = \underline{\mathrm{Y}}(\omega_1)$ so that $\underline{\mathrm{G}}(z)$, Π, and $\underline{\mathrm{Y}}$ are all defined on $(\Omega, \mathcal{F}, P)$. We have $\Phi(\underline{\mathrm{G}}(z), \underline{\mathrm{Y}}) \in \mathcal{L}_1^1(\Omega, \mathcal{F}, P)$ for all $z \in Z_0$ since $\underline{\mathrm{G}}(z) \in \mathcal{L}_1^N(\Omega, \mathcal{F}, P)$ for all $z \in Z_0$, $\Pi \in \mathcal{L}_1^1(\Omega, \mathcal{F}, P)$, and $\underline{\mathrm{Y}} \in \mathcal{L}_1^N(\Omega, \mathcal{F}, P)$. Then

$$\begin{aligned}0 &\leq \phi(\underline{\mathrm{G}}(z), \underline{\mathrm{Y}}) \\ &\leq \mathbb{E}[\Phi([\underline{\mathrm{G}}(z_0)], \underline{\mathrm{Y}})] + \mathbb{D}(Z_0)\mathbb{E}[\Pi] \\ &< \infty\end{aligned}$$

for all $z \in Z_0$.

(c) Compute

$$\begin{aligned}\mathbb{E}\left[e^{s\left(u\left(\underline{\mathrm{G}}(z)\right) - u\left(\underline{\mathrm{Y}}\right)\right)}\right] &\leq \mathbb{E}\left[e^{\left|s\left(u\left(\underline{\mathrm{G}}(z)\right) - u\left(\underline{\mathrm{Y}}\right)\right)\right|}\right] \\ &\leq \mathbb{E}\left[e^{|s|\,\left|u\left(\underline{\mathrm{G}}(z)\right) - u\left(\underline{\mathrm{Y}}\right)\right|}\right] \\ &\leq \mathbb{E}\left[e^{|s|\,\Phi\left(\underline{\mathrm{G}}(z), \underline{\mathrm{Y}}\right)}\right] \\ &\leq M_{\Phi\left(\underline{\mathrm{G}}(z), \underline{\mathrm{Y}}\right)}(|s|).\end{aligned}$$

The MGF $M_{\Phi\left(\underline{\mathrm{G}}(z), \underline{\mathrm{Y}}\right)}(|s|)$ is finite in a neighborhood of zero by assumption. □

It follows that

$$\max_{z \in Z_0} \phi(z) < \infty.$$

The function $g(z, u)$ is also bounded.

Proposition 3.1.5. *Suppose that assumption 3.1.1 holds. The function $g(z, u)$ is bounded over $z \in Z_0$ for $u \in \mathcal{U}^N(W)$.*

Proof. Observe $|g(z, u)| \leq \phi(z)$ for all $(z, u) \in Z_0 \times \mathcal{U}^N(W)$. The previous proposition shows that $\phi(z)$ is bounded. □

Now we address some continuity properties.

Proposition 3.1.6. *(a) The random variable $u(\underline{G}(z)) - u(\underline{Y})$ is Lipschitz continuous in $u \in \mathcal{U}^N(W)$ for all $z \in Z_0$.*

(b) Under assumption 3.1.1, $u(\underline{G}(z)) - u(\underline{Y})$ is Lipschitz continuous in $Z_0 \times \mathcal{U}^N(W)$.

PROOF. (a) Fix $z \in Z_0$ and choose $u_1, u_2 \in \mathcal{U}^N(W)$. Then

$$\begin{aligned} & |u_1([\underline{G}(z)](\omega)) - u_1(\underline{Y}(\omega)) - u_2([\underline{G}(z)](\omega)) + u_2(\underline{Y}(\omega))| \\ \leq\ & |u_1([\underline{G}(z)](\omega)) - u_2([\underline{G}(z)](\omega))| + |u_1(\underline{Y}(\omega)) - u_2(\underline{Y}(\omega))| \\ \leq\ & \|[\underline{G}(z)](\omega)\|_2 \|u_1 - u_2\|_W + \|\underline{Y}(\omega)\|_2 \|u_1 - u_2\|_W \\ \leq\ & \Phi([\underline{G}(z)](\omega), \underline{Y}(\omega)) \|u_1 - u_2\|_W. \end{aligned}$$

Thus

$$\begin{aligned} & \|u_1(\underline{G}(z)) - u_1(\underline{Y}) - u_2(\underline{G}(z)) + u_2(\underline{Y})\|_1 \\ \leq\ & \mathbb{E}[\Phi([\underline{G}(z)](\omega), \underline{Y}(\omega))] \|u_1 - u_2\|_W \\ =\ & \phi(z) \|u_1 - u_2\|_W \end{aligned}$$

and $\phi(z)$ has already been shown to be uniformly bounded in $z \in Z_0$.

(b) For $z_1, z_2 \in Z_0$ and $u_1, u_2 \in \mathcal{U}^N(W)$ compute

$$\begin{aligned} & |u_1([\underline{G}(z_1)](\omega)) - u_2([\underline{G}(z_2)](\omega))| \\ \leq\ & |u_1([\underline{G}(z_1)](\omega)) - u_2([\underline{G}(z_1)](\omega))| + |u_2([\underline{G}(z_1)](\omega)) - u_2([\underline{G}(z_2)](\omega))| \\ \leq\ & \|[\underline{G}(z_1)](\omega)\|_2 \|u_1 - u_2\|_W + \|\partial u_2\|_W \Pi(\omega) \|z_1 - z_2\|_2 \\ \leq\ & \|[\underline{G}(z_1)](\omega)\|_2 \|u_1 - u_2\|_W + \Pi(\omega) \|z_1 - z_2\|_2. \end{aligned}$$

Combine with the previous part to obtain

$$\begin{aligned} & |u_1([\underline{G}(z_1)](\omega)) - u_1(\underline{Y}(\omega)) + u_2([\underline{G}(z_2)](\omega)) + u_2(\underline{Y}(\omega))| \\ \leq\ & \Phi([\underline{G}(z_1)](\omega), \underline{Y}(\omega)) \|u_1 - u_2\|_W + \|\partial u_2\|_W \Pi(\omega) \|z_1 - z_2\|_2 \\ \leq\ & \Phi([\underline{G}(z_1)](\omega), \underline{Y}(\omega)) \|u_1 - u_2\|_W + \Pi(\omega) \|z_1 - z_2\|_2. \end{aligned}$$

Thus

$$\begin{aligned} & \|u_1(\underline{G}(z_1)) - u_1(\underline{Y}) + u_2(\underline{G}(z_2)) + u_2(\underline{Y})\|_1 \\ \leq\ & \phi(z_1) \|u_1 - u_2\|_W + \pi \|z_1 - z_2\|_2 \\ \leq\ & (\phi(z_1) + \pi)(\|u_1 - u_2\|_W + \|z_1 - z_2\|_2), \end{aligned}$$

and again $\phi(z)$ is uniformly bounded. □

Continuity of the function $g(z, u)$ is established next.

Proposition 3.1.7. *The function $g(z, u)$ is Lipschitz continuous on $Z_0 \times \mathcal{U}^N(W)$.*

PROOF. For any given $(z_1, u_1), (z_2, u_2) \in Z_0 \times \mathcal{U}^N(W)$, we have

$$\begin{aligned}
|g(z_1,u_1)-g(z_2,u_2)| &= |g(z_1,u_1)-g(z_1,u_2)+g(z_1,u_2)-g(z_2,u_2)| \\
&\le |g(z_1,u_1)-g(z_1,u_2)|+|g(z_1,u_2)-g(z_2,u_2)| \\
&= |\mathbb{E}[u_1(\underline{\mathrm{G}}(z_1))-u_1(\underline{\mathrm{Y}})-u_2(\underline{\mathrm{G}}(z_1))+u_2(\underline{\mathrm{Y}})]| \\
&\quad +|\mathbb{E}[u_2(\underline{\mathrm{G}}(z_1))-u_2(\underline{\mathrm{G}}(z_2))]|.
\end{aligned}$$

By the earlier propositions,

$$\begin{aligned}
& |\mathbb{E}[u_1(\underline{\mathrm{G}}(z_1))-u_1(\underline{\mathrm{Y}})-u_2(\underline{\mathrm{G}}(z_1))+u_2(\underline{\mathrm{Y}})]| \\
\le\ & \mathbb{E}[|u_1(\underline{\mathrm{G}}(z_1))-u_1(\underline{\mathrm{Y}})-u_2(\underline{\mathrm{G}}(z_1))+u_2(\underline{\mathrm{Y}})|] \\
\le\ & \mathbb{E}[\Phi([\underline{\mathrm{G}}(z_1)](\omega),\underline{\mathrm{Y}}(\omega))\,\|u_1-u_2\|_W] \\
\le\ & \max_{z\in Z_0}\phi(z)\ \|u_1-u_2\|_W,
\end{aligned}$$

and

$$\begin{aligned}
& |\mathbb{E}[u_2(\underline{\mathrm{G}}(z_1))-u_2(\underline{\mathrm{G}}(z_2))]| \\
\le\ & \mathbb{E}[|u_2(\underline{\mathrm{G}}(z_1))-u_2(\underline{\mathrm{G}}(z_2))|] \\
\le\ & \mathbb{E}[\|\partial u_2\|_w\|\underline{\mathrm{G}}(z_1)-\underline{\mathrm{G}}(z_2)\|_2] \\
\le\ & \pi\,\|z_1-z_2\|_2.
\end{aligned}$$

In summary,

$$|g(z_1,u_1)-g(z_2,u_2)| \le \left(\max_{z\in Z_0}\phi(z)+\pi\right)(\|u_1-u_2\|_W+\|z_1-z_2\|_2).$$

□

Introduce the functions

$$\begin{aligned}
\psi(z) &\triangleq \inf_{u\in\mathcal{U}^N(W)}\{g(z;\,u)\} \\
\psi(z;\,U_\Xi) &\triangleq \inf_{u\in U_\Xi}\{g(z;\,u)\}.
\end{aligned}$$

In the following development, any finite subset $\{u_1,\ldots,u_I\}\subset\mathcal{U}^N(W)$ can be used in place of U_Ξ. Obviously $\psi(z)$ and $\psi(z;\,U_\Xi)$ are concave and increasing. Further, $\psi(z;\,U_\Xi)\ge\psi(z)$ for $z\in Z_0$ and

$$\begin{aligned}
Z^\epsilon &\equiv \{z\in Z_0 : \psi(z)\ge\epsilon\} \\
Z^\epsilon(U_\Xi) &\equiv \{z\in Z_0 : \psi(z;\,U_\Xi)\ge\epsilon\}.
\end{aligned}$$

Next we establish boundedness and continuity properties of $\psi(z)$ and $\psi(z;\,U_\Xi)$.

Proposition 3.1.8. *(a) $\psi(z)$ is bounded on Z_0.*
(b) $\psi(z)$ is Lipschitz continuous on Z_0.
(c) $\psi(z;\,U_\Xi)$ is bounded on Z_0.
(d) $\psi(z;\,U_\Xi)$ is Lipschitz continuous on Z_0.

PROOF. (a) Compute:

$$\begin{aligned} |\psi(z)| &\leq \sup_{u\in\mathcal{U}^N(W)} |g(z,u)| \\ &= \sup_{u\in\mathcal{U}^N(W)} |\mathbb{E}[u(\underline{G}(z))] - \mathbb{E}[u(\underline{Y})]| \\ &\leq \phi(z). \end{aligned}$$

The proof for (c) is identical.

(b) First observe

$$|\psi(z_1) - \psi(z_2)| \leq \sup_{u\in\mathcal{U}^N(W)} |g(z_1,u) - g(z_2,u)|.$$

To verify this statement, compute:

$$\begin{aligned} \psi(z_2) &= \inf_{u\in\mathcal{U}^N(W)} g(z_2,u) \\ &\leq \inf_{u\in\mathcal{U}^N(W)} \{g(z_1,u) + |g(z_1,u) - g(z_2,u)|\} \\ &\leq \psi(z_1) + \sup_{u\in\mathcal{U}^N(W)} |g(z_1,u) - g(z_2,u)|. \end{aligned}$$

By the same reasoning,

$$\begin{aligned} \psi(z_1) &= \inf_{u\in\mathcal{U}^N(W)} g(z_1,u) \\ &\leq \inf_{u\in\mathcal{U}^N(W)} \{g(z_2,u) + |g(z_2,u) - g(z_1,u)|\} \\ &\leq \psi(z_2) + \sup_{u\in\mathcal{U}^N(W)} |g(z_1,u) - g(z_2,u)|. \end{aligned}$$

Now compute:

$$\begin{aligned} |\psi(z_1) - \psi(z_2)| &\leq \sup_{u\in\mathcal{U}^N(W)} |g(z_1,u) - g(z_2,u)| \\ &= \sup_{u\in\mathcal{U}^N(W)} |\mathbb{E}[u(\underline{G}(z_1)) - u(\underline{G}(z_2))]| \\ &\leq \mathbb{E}[\|\underline{G}(z_1) - \underline{G}(z_2)\|_2] \\ &\leq \mathbb{E}[\Pi(\omega)\, \|z_1 - z_2\|_2] \\ &= \pi\, \|z_1 - z_2\|_2. \end{aligned}$$

The proof for (d) is identical. □

As the set U_Ξ is made larger and larger (in set containment), U_Ξ becomes a progressively better approximation of $\mathcal{U}^N(W)$. In this subsection, we use the notation $U_{\Xi_k} \uparrow \mathcal{U}^N(W)$ to denote an increasing sequence $\{U_{\Xi_k}\}_{k=0}^{\infty}$ with $U_{\Xi_k} \subset U_{\Xi_{k+1}}$ for all $k \geq 0$ and $\bigcup_{k=0}^{\infty} U_{\Xi_k} = \mathcal{U}^N(W)$.

Proposition 3.1.9. *For any $U_{\Xi_k} \uparrow \mathcal{U}^N(W)$, $\psi(z; U_{\Xi_k})$ converges to $\psi(z)$ (from above) uniformly on Z_0.*

Proof. Certainly $\psi(z; U_{\Xi_k}) \to \psi(z)$ for any fixed $z \in Z_0$ as $k \to \infty$. Since Z_0 is compact and $\psi(z; U_\Xi)$ and $\psi(z)$ are continuous, this convergence is uniform. $\square$

We reference an important fact that will be used to establish stability.

Proposition 3.1.10. [**16**, Proposition 3.1] *Suppose $Z^\epsilon \equiv \{z \in Z_0 : \psi(z) \geq \epsilon\}$ satisfies the Slater condition. Then $\mathbb{D}(Z^\epsilon, Z^{\epsilon-\gamma}) \to 0$ as $\gamma \downarrow 0$.*

We must assume that the Slater condition is satisfied in the following lemma. If $\epsilon = 0$, then the set $\{z \in Z_0 : \psi(z) > 0\}$ is empty.

Proposition 3.1.11. *Suppose Z^ϵ satisfies the Slater condition (i.e. $\epsilon < 0$).*
(a) $\mathbb{H}(Z^\epsilon(U_{\Xi_k}), Z^\epsilon) \to 0$ as $U_{\Xi_k} \uparrow \mathcal{U}^N(W)$.
(b) $\mathbb{D}(S^\epsilon(U_{\Xi_k}), S^\epsilon) \to 0$ as $U_{\Xi_k} \uparrow \mathcal{U}^N(W)$.
(c) $\nu^\epsilon(U_{\Xi_k}) \to \nu^\epsilon$ as $U_{\Xi_k} \uparrow \mathcal{U}^N(W)$.

Proof. (a) Since $\psi(z; U_\Xi)$ converges to $\psi(z)$ uniformly, it epi-converges and hypo-converges to $\psi(z)$ as well by [**19**, Proposition 7.15]. Then $Z^\epsilon(U_{\Xi_k})$ is upper semiconvergent to Z^ϵ and $\mathbb{D}(Z^\epsilon(U_{\Xi_k}), Z^\epsilon) \to 0$ as $k \to \infty$ by [**26**, Theorem 3.1]. Since Z^ϵ satisfies the Slater condition, we have $Z^\epsilon \subseteq \operatorname{cl}\operatorname{int} Z^\epsilon$ by the previous proposition. Thus $Z^\epsilon(U_{\Xi_k})$ is lower semiconvergent to Z^ϵ and $\mathbb{D}(Z^\epsilon, Z^\epsilon(U_{\Xi_k})) \to 0$ as $k \to \infty$ by [**26**, Theorem 3.5].

(b) Apply [**26**, Theorem 4.1]. If Z^ϵ is empty, then $Z^\epsilon(U_{\Xi_k})$ must be empty for all large k by part (a).

(c) Follows from part (b). $\square$

3.2. Optimality conditions

3.2.1. Optimality conditions for problem (ϵ−MP). The space $\mathcal{U}^N(W)$ is a compact metric space in the topology induced by the supremum norm $\|\cdot\|_W$ on $\mathcal{C}(W)$. Thus, $\mathcal{C}(W)$ is also a locally compact Hausdorff space. We define $\mathcal{C}(\mathcal{U}^N(W))$ to be the space of continuous functions on $\mathcal{U}^N(W)$, all such functions necessarily have compact support because $\mathcal{U}^N(W)$ is compact.

To represent the constraints (2) as a single functional constraint in $\mathcal{C}(\mathcal{U}^N(W))$, define the operator $\boldsymbol{g} : Z_0 \to \mathcal{C}(\mathcal{U}^N(W))$ via

$$[\boldsymbol{g}(z)](u) = g(z, u).$$

We have already shown that $\boldsymbol{g}(z)$ is an element of $\mathcal{C}(\mathcal{U}^N(W))$ by the Lipschitz continuity of $g(z, u)$ in $(z, u) \in Z_0 \times \mathcal{U}^N(W)$. Constraints (2) are then equivalent to

$$\boldsymbol{g}(z) - \epsilon \in \mathcal{C}_+(\mathcal{U}^N(W)),$$

where $\mathcal{C}_+(\mathcal{U}^N(W))$ is the set of non-negative continuous functions on $\mathcal{U}^N(W)$.

Let $\mathcal{M}(\mathcal{U}^N(W))$ be the space of finite signed regular Borel measures on

$$(\mathcal{U}^N(W), \boldsymbol{\mathcal{B}}),$$

where $\boldsymbol{\mathcal{B}}$ is the sigma-algebra generated by the open sets of $\mathcal{C}(W)$. We denote a measure in $\mathcal{M}(\mathcal{U}^N(W))$ as Λ. Define the Lagrangian

$$L(z, \Lambda) = f(z) + \langle \Lambda, \boldsymbol{g}(z) \rangle,$$

where

$$\langle \Lambda, \boldsymbol{g}(z) \rangle = \int_{\mathcal{U}^N(W)} [\boldsymbol{g}(z)](u)\, d\Lambda(u)$$

is the duality pairing between $\mathcal{C}\left(\mathcal{U}^N(W)\right)$ and $\mathcal{M}\left(\mathcal{U}^N(W)\right)$.

The Slater condition for problem ($\epsilon-$MP) is defined next.

Assumption 3.2.1. *There exists $\tilde{z} \in Z_0$ such that*

$$g(\tilde{z}, u) > \epsilon, \quad \forall u \in \mathcal{U}^N(W).$$

The dual to the cone $\mathcal{C}_+\left(\mathcal{U}^N(W)\right)$ is

$$\begin{aligned} & \mathcal{M}_+\left(\mathcal{U}^N(W)\right) \\ \triangleq \quad & \left\{\Lambda \in \mathcal{M}\left(\mathcal{U}^N(W)\right) \,:\, \langle \Lambda, \boldsymbol{f} \rangle \geq 0, \forall \boldsymbol{f} \in \mathcal{C}_+\left(\mathcal{U}^N(W)\right)\right\}. \end{aligned}$$

In the following theorem, for $u \in \mathcal{U}^N(W)$, δ_u is the Dirac delta function at u on $\mathcal{U}^N(W)$.

Theorem 3.2.2. *Suppose assumption 3.2.1 holds.*

(a) If $\hat{z} \in Z_0$ solves problem (ϵ − MP), then there exists $\hat{\Lambda} \in \mathcal{M}_+\left(\mathcal{U}^N(W)\right)$ such that

$$L\left(\hat{z}, \hat{\Lambda}\right) = \max\left\{L\left(z, \hat{\Lambda}\right) \,:\, z \in Z_0\right\}, \tag{5}$$

$$\langle \hat{\Lambda}, \boldsymbol{g}(\hat{z}) \rangle = \epsilon\, \hat{\Lambda}\left(\mathcal{U}^N(W)\right). \tag{6}$$

(b) If $\hat{z} \in Z_0$ solves problem (ϵ − MP), then there exists $\Lambda \in \mathcal{M}_+\left(\mathcal{U}^N(W)\right)$ satisfying (5) − (6) *such that $\Lambda = \sum_{m=1}^{M} \lambda_m \delta_{u_m}$ for $\lambda_m \geq 0$ and $u_m \in \mathcal{U}^N(W)$ for all $m = 1, \ldots, M$.*

Proof. (a) The operator $\boldsymbol{g}(\cdot)$ is continuous. It is also concave: for $z_1, z_2 \in Z_0$ and $0 \leq \alpha \leq 1$ we have

$$\begin{aligned} & [\boldsymbol{g}(\alpha\, z_1 + (1-\alpha)\, z_2)](u) \\ = \quad & g(\alpha\, z_1 + (1-\alpha)\, z_2, u) \\ = \quad & \mathbb{E}\left[u\left(\underline{\mathrm{G}}(\alpha\, z_1 + (1-\alpha)\, z_2)\right) - u(\underline{\mathrm{Y}})\right] \\ \geq \quad & \mathbb{E}\left[u\left(\alpha\, \underline{\mathrm{G}}(z_1) + (1-\alpha)\, \underline{\mathrm{G}}(z_2)\right) - u(\underline{\mathrm{Y}})\right] \\ \geq \quad & \mathbb{E}\left[\alpha\, u\left(\underline{\mathrm{G}}(z_1)\right) + (1-\alpha)\, u\left(\underline{\mathrm{G}}(z_2)\right) - u(\underline{\mathrm{Y}})\right] \\ = \quad & \alpha\, [\boldsymbol{g}(z_1)](u) + (1-\alpha)\, [\boldsymbol{g}(z_2)](u), \end{aligned}$$

for all $u \in \mathcal{U}^N(W)$. Problem ($\epsilon-$MP) can be written as

$$\begin{aligned} \max_{z \in Z_0} \quad & f(z) \\ \text{s.t.} \quad & \boldsymbol{g}(z) - \epsilon \in \mathcal{C}_+\left(\mathcal{U}^N(W)\right). \end{aligned}$$

Under assumption 3.2.1, the above problem satisfies the Slater condition. The Lagrangian for this problem is

$$\Upsilon(z, \Lambda) = f(z) + \langle \Lambda, \boldsymbol{g}(z) - \epsilon \rangle.$$

We can apply [**4**, Theorem 3.4] to obtain the existence of some $\hat{\Lambda} \in \mathcal{M}_+\left(\mathcal{U}^N(W)\right)$ such that

$$\begin{aligned} \Upsilon\left(\hat{z}, \hat{\Lambda}\right) &= \max\left\{\Upsilon\left(z, \hat{\Lambda}\right) : z \in Z_0\right\}, \\ \langle \hat{\Lambda}, \boldsymbol{g}(\hat{z}) - \epsilon \rangle &= 0. \end{aligned}$$

(b) Apply [**4**, Proposition 5.104]. □

Notice that the perturbation ϵ introduces the term $\epsilon\, \hat{\Lambda}\left(\mathcal{U}^N(W)\right)$ into the complementary slackness conditions. In fact, since $\hat{\Lambda} \in \mathcal{M}_+\left(\mathcal{U}^N(W)\right)$ it is a non-negative measure and $\hat{\Lambda}\left(\mathcal{U}^N(W)\right) = \|\hat{\Lambda}\|$, where $\|\hat{\Lambda}\|$ is the total variation norm of the measure $\hat{\Lambda}$.

Given an optimal dual multiplier $\hat{\Lambda}$, we ask if we can transform the expression $\langle \hat{\Lambda}, \boldsymbol{g}(z) \rangle$ using Fubini's theorem to gain additional information.

Theorem 3.2.3. *Define* $u_\Lambda(\cdot) = \int_{\mathcal{U}^N(W)} u(\cdot)\, d\Lambda(u)$ for $\Lambda \in \mathcal{M}_+\left(\mathcal{U}^N(W)\right)$. *Then for* $\underline{X}, \underline{Y} \in \mathcal{L}_\infty^N(\Omega, \mathcal{F}, P)$ *we have*

$$\int_{\mathcal{U}^N(W)} \mathbb{E}\left[u(\underline{X}) - u(\underline{Y})\right] d\Lambda(u) = \mathbb{E}\left[u_\Lambda(\underline{X}) - u_\Lambda(\underline{Y})\right].$$

Proof. The function $u_\Lambda(\cdot) = \int_{\mathcal{U}^N(W)} u(\cdot)\, d\Lambda(u)$ is increasing and concave because each $u(\cdot)$ is increasing and concave, the measure Λ is non-negative, and integration preserves monotonicity and concavity. Thus, u_Λ is a well defined increasing concave function.

Integration on Ω is with respect to the measure space $(\Omega, \mathcal{F}, P)$ and integration on $\mathcal{U}^N(W)$ is with respect to the measure space $\left(\mathcal{U}^N(W), \boldsymbol{\mathcal{B}}, \Lambda\right)$. Product integration is defined on the measure space

$$\left(\Omega \times \mathcal{U}^N(W), \mathcal{F} \times \boldsymbol{\mathcal{B}}, P \times \Lambda\right)$$

formed by the cross-product. First notice

$$|u(\underline{\mathrm{X}}(\omega)) - u(\underline{\mathrm{Y}}(\omega))| \le \|\partial u\|_W \|\underline{\mathrm{X}}(\omega) - \underline{\mathrm{Y}}(\omega)\|_2,$$

by [**2**, Theorem 2.1]. Further, $\|\partial u\|_W \le 1$ for all $u \in \mathcal{U}^N(W)$ by assumption. Then

$$\begin{aligned} & \left|\int_{\Omega \times \mathcal{U}^N(W)} \left[u(\underline{\mathrm{X}}(\omega)) - u(\underline{\mathrm{Y}}(\omega))\right] d(P \times \Lambda)\right| \\ \le & \int_{\Omega \times \mathcal{U}^N(W)} \|\underline{\mathrm{X}}(\omega) - \underline{\mathrm{Y}}(\omega)\|_2 d(P \times \Lambda). \end{aligned}$$

By assumption $\underline{\mathrm{X}}, \underline{\mathrm{Y}} \in \mathcal{L}_\infty^N(\Omega, \mathcal{F}, P)$, so both $\|\underline{\mathrm{X}}(\omega)\|_2$ and $\|\underline{\mathrm{Y}}(\omega)\|_2$ are bounded by some $C > 0$ for $P-$almost all $\omega \in \Omega$. Thus

$$\int_{\Omega \times \mathcal{U}^N(W)} \|\underline{\mathrm{X}}(\omega) - \underline{\mathrm{Y}}(\omega)\|_2\, d(P \times \Lambda) \le 2C \int_{\Omega \times \mathcal{U}^N(W)} d(P \times \Lambda)$$

and

$$\int_{\Omega \times \mathcal{U}^N(W)} d(P \times \Lambda) = P(\Omega)\, \Lambda\left(\mathcal{U}^N(W)\right) < \infty$$

since $P(\Omega) = 1$ and $\Lambda \in \mathcal{M}\left(\mathcal{U}^N(W)\right)$ is a finite measure by definition.

It follows that the function $u(\underline{\mathrm{X}}(\omega)) - u(\underline{\mathrm{Y}}(\omega))$ is integrable on

$$\left(\Omega \times \mathcal{U}^N(W), \mathcal{F} \times \boldsymbol{\mathcal{B}}, P \times \Lambda\right).$$

By assumption, $(\Omega, \mathcal{F}, P)$ and $\left(\mathcal{U}^N(W), \boldsymbol{\mathcal{B}}, \Lambda\right)$ are finite measure spaces. Fubini's theorem [**13**, Theorem 2.37] gives

$$\begin{aligned}
& \int_{\Omega \times \mathcal{U}^N(W)} \left(u(\underline{\mathrm{X}}(\omega)) - u(\underline{\mathrm{Y}}(\omega))\right) d(P \times \Lambda) \\
= & \int_{\mathcal{U}^N(W)} \left[\int_\Omega u(\underline{\mathrm{X}}(\omega)) - u(\underline{\mathrm{Y}}(\omega))\, P(d\omega)\right] d\Lambda(u) \\
= & \int_\Omega \left[\int_{\mathcal{U}^N(W)} u(\underline{\mathrm{X}}(\omega)) - u(\underline{\mathrm{Y}}(\omega))\, d\Lambda(u)\right] P(d\omega).
\end{aligned}$$

Thus,

$$\begin{aligned}
& \int_{\mathcal{U}^N(W)} \mathbb{E}\left[u(\underline{\mathrm{X}}) - u(\underline{\mathrm{Y}})\right] d\Lambda(u) \\
= & \ \mathbb{E}\left[\int_{\mathcal{U}^N(W)} u(\underline{\mathrm{X}})\, d\Lambda(u) - \int_{\mathcal{U}^N(W)} u(\underline{\mathrm{Y}})\, d\Lambda(u)\right] \\
= & \ \mathbb{E}\left[u_\Lambda(\underline{\mathrm{X}}) - u_\Lambda(\underline{\mathrm{Y}})\right].
\end{aligned}$$

□

The preceding theorem shows that

$$\int_{\mathcal{U}^N(W)} \mathbb{E}\left[u(\underline{\mathrm{G}}(z)) - u(\underline{\mathrm{Y}})\right] d\Lambda(u) = \mathbb{E}\left[u_\Lambda(\underline{\mathrm{G}}(z)) - u_\Lambda(\underline{\mathrm{Y}})\right],$$

for any $z \in Z_0$ using the fact that $\underline{\mathrm{G}}(z) \in \mathcal{L}_\infty^N(\Omega, \mathcal{F}, P)$ for any $z \in Z_0$ and $\underline{\mathrm{Y}} \in \mathcal{L}_\infty^N(\Omega, \mathcal{F}, P)$.

In the next theorem, we show that all functions in $\mathrm{cl\,conv}\,\mathcal{U}^N(W)$ can be described in terms of measures on $\left(\mathcal{U}^N(W), \boldsymbol{\mathcal{B}}\right)$.

Theorem 3.2.4. *The sets* $cl\, conv \mathcal{U}^N(W)$ *and*

$$\left\{\int_{\mathcal{U}^N(W)} u(\cdot)\ d\Lambda(u)\ : \Lambda \in \mathcal{M}_+\left(\mathcal{U}^N(W)\right),\ \|\Lambda\| = 1\right\}$$

are equal.

PROOF. The set $\mathrm{conv}\,\mathcal{U}^N(W)$ is convex by definition. Any finite convex combination of elements of $\mathcal{U}^N(W)$ is of the form $\sum_{i=1}^I \lambda_i u_i(\cdot)$ for where $\sum_{i=1}^I \lambda_i = 1$ and $\lambda \geq 0$. Since W has nonempty interior, the relative interiors of the domains of the functions $u_i(\cdot)$ have a nonempty intersection. Then [**20**, Theorem 23.8] states that

$$\partial\left[\sum_{i=1}^I \lambda_i u_i(\cdot)\right] = \sum_{i=1}^I \lambda_i \partial\, u_i(\cdot).$$

Thus

$$\| \sum_{i=1}^{I} \lambda_i u_i (\cdot) \|_W \leq \sum_{i=1}^{I} \lambda_i \| \partial u_i (\cdot) \|_W \leq 1.$$

It follows that all u in $\operatorname{conv}\mathcal{U}^N(W)$ have uniformly bounded subdifferentials, so $\operatorname{conv}\mathcal{U}^N(W)$ is equicontinuous. By construction $\mathcal{U}^N(W)$ is uniformly bounded, so $\operatorname{conv}\mathcal{U}^N(W)$ is uniformly bounded as well. In summary, $\operatorname{conv}\mathcal{U}^N(W)$ is equicontinuous and uniformly bounded (and thus point-wise bounded) on W. Then the closure of $\operatorname{conv}\mathcal{U}^N(W)$, $\operatorname{cl}\operatorname{conv}\mathcal{U}^N(W)$, is compact by Ascoli's Theorem [**21**, Theorem A5].

By Milman's Theorem [**21**, Theorem 3.25], every extreme point of $\operatorname{cl}\operatorname{conv}\mathcal{U}^N(W)$ lies in $\mathcal{U}^N(W)$. Choquet's Theorem [**18**, Chapter 3, Theorem (Choquet)] then shows that every $u \in \operatorname{cl}\operatorname{conv}\mathcal{U}^N(W)$ is of the form

$$\left\{ \int_{\mathcal{U}^N(W)} u(\cdot)\ d\Lambda(u)\ :\ \Lambda \in \mathcal{M}_+\left(\mathcal{U}^N(W)\right),\ \|\Lambda\| = 1 \right\}.$$

We are able to apply to apply Choquet's Theorem because $\operatorname{cl}\operatorname{conv}\mathcal{U}^N(W) \subset \mathcal{C}(W)$, and $\mathcal{C}(W)$ is a locally convex topological vector space. In particular, for $\delta > 0$ any neighborhood

$$B_\delta(0) = \{ f \in \mathcal{C}(W)\ :\ \|f\| < \delta \}$$

is a convex set and the set of neighborhoods of this form is a convex local base at zero. □

We can scale the previous theorem.

Corollary 3.2.5. *The sets* $\operatorname{cl}\operatorname{cone}\mathcal{U}^N(W)$ *and*

$$\left\{ \int_{\mathcal{U}^N(W)} u(\cdot)\ d\Lambda(u)\ :\ \Lambda \in \mathcal{M}_+\left(\mathcal{U}^N(W)\right) \right\}$$

are equal.

Proof. By definition

$$\operatorname{cone}\mathcal{U}^N(W) = \bigcup_{\alpha \geq 0} \left\{ \alpha \operatorname{conv}\mathcal{U}^N(W) \right\}$$

and

$$\operatorname{cl}\operatorname{cone}\left(\mathcal{U}^N(W)\right) = \bigcup_{\alpha \geq 0} \left\{ \alpha \operatorname{cl}\operatorname{conv}\left(\mathcal{U}^N(W)\right) \right\}.$$

For any $\alpha \geq 0$,

$$\begin{aligned}
& \alpha \left\{ \int_{\mathcal{U}^N(W)} u(\cdot)\, d\Lambda(u)\ :\ \Lambda \in \mathcal{M}_+\left(\mathcal{U}^N(W)\right),\ \|\Lambda\| = 1 \right\} \\
= & \left\{ \int_{\mathcal{U}^N(W)} u(\cdot)\, d(\alpha\Lambda)(u)\ :\ \Lambda \in \mathcal{M}_+\left(\mathcal{U}^N(W)\right),\ \|\Lambda\| = 1 \right\} \\
= & \left\{ \int_{\mathcal{U}^N(W)} u(\cdot)\, d\Lambda(u)\ :\ \Lambda \in \mathcal{M}_+\left(\mathcal{U}^N(W)\right),\ \|\Lambda\| = \alpha \right\}
\end{aligned}$$

where the first inequality uses linearity of the integral

$$\alpha \int_{\mathcal{U}^N(W)} u(\cdot)\, d\Lambda(u) = \int_{\mathcal{U}^N(W)} u(\cdot)\, d(\alpha\Lambda)(u).$$

By definition of the total variation norm, $\|\alpha\Lambda\| = \alpha\|\Lambda\|$. We use the previous theorem to show

$$\begin{aligned}
& \operatorname{cl}\operatorname{cone}\mathcal{U}^N(W) \\
= & \bigcup_{\alpha\geq 0} \alpha\left\{\operatorname{cl}\operatorname{conv}\mathcal{U}^N(W)\right\} \\
= & \bigcup_{\alpha\geq 0} \left\{\alpha\left\{\int_{\mathcal{U}^N(W)} u(\cdot)\, d\Lambda(u) \;:\; \Lambda \in \mathcal{M}_+\left(\mathcal{U}^N(W)\right),\ \|\Lambda\| = 1\right\}\right\} \\
= & \bigcup_{\alpha\geq 0} \left\{\int_{\mathcal{U}^N(W)} u(\cdot)\, d\Lambda(u) \;:\; \Lambda \in \mathcal{M}_+\left(\mathcal{U}^N(W)\right),\ \|\Lambda\| = \alpha\right\} \\
= & \left\{\int_{\mathcal{U}^N(W)} u(\cdot)\, d\Lambda(u) \;:\; \Lambda \in \mathcal{M}_+\left(\mathcal{U}^N(W)\right)\right\}.
\end{aligned}$$

□

In summary, the vector-valued integral $\int_{\mathcal{U}^N(W)} u(\cdot)\, d\Lambda(u)$ lies in $\operatorname{cl}\operatorname{cone}\mathcal{U}^N(W)$ for any $\Lambda \in \mathcal{M}_+\left(\mathcal{U}^N(W)\right)$.

3.2.2. Optimality conditions for problem ($\epsilon-$SIP). For this subsection the operator $\boldsymbol{g}(\cdot)$ will be understood as $\boldsymbol{g} : Z_0 \to \mathcal{C}(\Xi)$,

$$[\boldsymbol{g}(z)](\xi) = g(z, u(\cdot;\xi)).$$

For $\Lambda \in \mathcal{M}(\Xi)$, the Lagrangian for problem ($\epsilon -$ SIP) is

$$L(z,\Lambda) = f(z) + \int_{\Xi} [\boldsymbol{g}(z)](u(\cdot;\xi))\, d\Lambda(\xi).$$

The Slater condition for problem ($\epsilon -$ SIP) is defined next.

Assumption 3.2.6. *There exists $\tilde{z} \in Z_0$ such that*

$$g(\tilde{z}, u) > \epsilon, \qquad \forall u \in U_{\Xi}.$$

For $\xi \in \Xi$, δ_ξ indicates the Dirac delta function at ξ on Ξ.

Theorem 3.2.7. *Suppose assumption 3.2.6 holds.*

(a) If $\hat{z}$ is optimal for problem ($\epsilon-$SIP) then there exists $\hat{\Lambda} \in \mathcal{M}_+(\Xi)$ such that

$$L\left(\hat{z},\hat{\Lambda}\right) = \max\left\{L\left(z,\hat{\Lambda}\right) \;:\; z \in Z_0\right\}, \tag{7}$$

$$\langle \hat{\Lambda}, \boldsymbol{g}(\hat{z})\rangle = \epsilon\,\hat{\Lambda}(\Xi). \tag{8}$$

(b) There exists $\Lambda \in \mathcal{M}_+(\Xi)$ satisfying (7) – (8) such that $\Lambda = \sum_{m=1}^{M} \lambda_m \delta_{\xi_m}$ for $\lambda_m \geq 0$ and $\xi_m \in \Xi$ for all $m = 1, \ldots, M$.

Given an optimal dual multiplier $\hat{\Lambda}$, we can again transform the expression $\langle \Lambda, \boldsymbol{g}(z) \rangle$ using Fubini's theorem. The next three results are in line with the previous subsection and have already been established in [**14**].

Corollary 3.2.8. *Define* $u_\Lambda(\cdot) = \int_\Xi u(\cdot;\xi)\, d\Lambda(\xi)$. *Then for* $\underline{X}, \underline{Y} \in \mathcal{L}_\infty^N(\Omega, \mathcal{F}, P)$,

$$\int_\Xi \mathbb{E}\left[u\left(\underline{X};\xi\right) - u\left(\underline{Y};\xi\right)\right] d\Lambda(\xi) = \mathbb{E}\left[u_\Lambda\left(\underline{X}\right) - u_\Lambda\left(\underline{Y}\right)\right].$$

Next we characterize $\operatorname{cl}\operatorname{conv} U_\Xi$.

Corollary 3.2.9. *The sets* $\operatorname{cl}\operatorname{conv} U_\Xi$ *and*

$$\left\{ \int_\Xi u(\cdot;\xi)\; d\Lambda(u) \;:\; \Lambda \in \mathcal{M}_+(\Xi),\, \|\Lambda\| = 1 \right\}$$

are equal.

We can scale the previous theorem.

Corollary 3.2.10. *The sets* $\operatorname{cl}\operatorname{cone} U_\Xi$ *and*

$$\left\{ \int_\Xi u(\cdot;\xi)\; d\Lambda(\xi) \;:\; \Lambda \in \mathcal{M}_+(\Xi) \right\}$$

are equal.

3.2.3. Optimality conditions for problem ($\epsilon-$NLP). In this subsection the operator $\boldsymbol{g}(\cdot)$ will be understood as a vector-valued mapping $\boldsymbol{g} : Z_0 \to \mathbb{R}^I$,

$$\boldsymbol{g}(z) = \begin{pmatrix} \mathbb{E}\left[u_1\left(\underline{\mathrm{G}}(z)\right) - u_1\left(\underline{\mathrm{Y}}\right)\right] \\ \vdots \\ \mathbb{E}\left[u_I\left(\underline{\mathrm{G}}(z)\right) - u_I\left(\underline{\mathrm{Y}}\right)\right] \end{pmatrix}.$$

For $\lambda \in \mathbb{R}^I$, define the Lagrangian for problem ($\epsilon-$NLP) to be

$$L(z,\lambda) = f(z) + \langle \lambda, \boldsymbol{g}(z) \rangle,$$

where

$$\langle \lambda, \boldsymbol{g}(z) \rangle = \sum_{i=1}^{I} \lambda_i \mathbb{E}\left[u_i\left(\underline{\mathrm{G}}(z)\right) - u_i\left(\underline{\mathrm{Y}}\right)\right]$$

is the usual inner product on $\mathbb{R}^I$.

Assumption 3.2.11. *There exists* $\tilde{z} \in Z_0$ *such that*

$$g(\tilde{z}, u_i) > \epsilon, \quad \forall i = 1, \ldots, I.$$

The next theorem follows from the usual nonlinear programming optimality conditions.

Theorem 3.2.12. *Suppose assumption 3.2.11 holds. If* $\hat{z}$ *is optimal for problem ($\epsilon-$NLP), then there exists* $\hat{\lambda} \in \mathbb{R}^I$ *such that*

$$L\left(\hat{z},\hat{\lambda}\right) = \max\left\{L\left(z,\hat{\lambda}\right) : z\in Z_0\right\}, \tag{9}$$

$$\langle\hat{\lambda},\boldsymbol{g}\left(\hat{z}\right)\rangle = \epsilon\sum_{i=1}^{I}\lambda_i. \tag{10}$$

Since we are only dealing with finite sums in this subsection, we immediately obtain the equality

$$\sum_{i=1}^{I}\lambda_i g\left(z,u_i\right)=\mathbb{E}\left[u_\lambda\left(\underline{\mathrm{G}}\left(z\right)\right)-u_\lambda\left(\underline{\mathrm{Y}}\right)\right],$$

for $u_\lambda\left(\cdot\right)=\sum_{i=1}^{I}\lambda_i u_i\left(\cdot\right)$.

3.3. Duality

3.3.1. Duality for problem (ϵ−MP). The dual functional for problem (ϵ−MP) is

$$\mathrm{d}\left(\Lambda\right)=\max\left\{L\left(z,\Lambda\right) : z\in Z_0\right\}.$$

The corresponding dual problem is

$$(\epsilon-\mathrm{MP}_D)\qquad \min\left\{\mathrm{d}\left(\Lambda\right)-\epsilon\,\Lambda\left(\mathcal{U}^N\left(W\right)\right) : \Lambda\in\mathcal{M}_+\left(\mathcal{U}^N\left(W\right)\right)\right\}.$$

By the optimality conditions and weak duality, strong duality holds between problem (ϵ−MP) and problem ($\epsilon-\mathrm{MP}_D$).

Theorem 3.3.1. *Suppose assumption 3.2.1 holds.*

(a) If problem (ϵ−MP) has an optimal solution, then problem ($\epsilon-MP_D$) has an optimal solution and the optimal values are equal.

(b) If problem ($\epsilon-MP_D$) has an optimal solution $\hat{\Lambda}$ then any $\hat{z}$ satisfying the optimality conditions (5) − (6) *with respect to $\hat{\Lambda}$ is optimal to problem (ϵ−MP).*

(c) Problem ($\epsilon-MP_D$) has an optimal solution with support on a finite set $\{u_1,\ldots,u_M\}\subset\mathcal{U}^N\left(W\right)$.

Proof. (a) Clearly,

$$f\left(z\right)\le L\left(z,\Lambda\right)-\epsilon\,\Lambda\left(\mathcal{U}^N\left(W\right)\right)$$

for any $\Lambda\in\mathcal{M}_+\left(\mathcal{U}^N\left(W\right)\right)$ and any $z\in Z_0$ satisfying (2), and thus

$$f\left(z\right)\le \mathrm{d}\left(\Lambda\right)-\epsilon\,\Lambda\left(\mathcal{U}^N\left(W\right)\right).$$

For $\hat{z}$ and the corresponding $\hat{\Lambda}$ satisfying the optimality conditions (5) − (6) we have

$$f\left(\hat{z}\right)=\mathrm{d}\left(\hat{\Lambda}\right)-\epsilon\,\hat{\Lambda}\left(\mathcal{U}^N\left(W\right)\right)$$

so $\hat{\Lambda}$ is optimal for problem (ϵ−MP_D).

(b) By weak duality, if $\hat{\Lambda}$ is an optimal solution of problem ($\epsilon-\mathrm{MP}_D$) then

$$f\left(z\right)\le \mathrm{d}\left(\hat{\Lambda}\right)-\epsilon\,\hat{\Lambda}\left(\mathcal{U}^N\left(W\right)\right).$$

However, if $\hat{z}$ satisfies the optimality conditions (5) – (6) with respect to $\hat{\Lambda}$, then

$$f(\hat{z}) = \mathrm{d}\left(\hat{\Lambda}\right) - \epsilon \, \hat{\Lambda} \left(\mathcal{U}^N(W)\right)$$

by complementary slackness and $\hat{z}$ is necessarily an optimal solution.

(c) This statement follows from the existence of an optimal multiplier $\Lambda = \sum_{m=1}^{M} \lambda_m \delta_{u_m}$ from the optimality conditions. □

Now define the alternative Lagrangian

$$L(z, u) = f(z) + \mathbb{E}\left[u\left(\underline{\mathrm{G}}(z)\right) - u\left(\underline{\mathrm{Y}}\right)\right],$$

and the alternative dual functional

$$\mathrm{d}(u) = \max\left\{L(z, u) \, : \, z \in Z_0\right\}.$$

As in [**8, 14**], problem (ϵ−MP$_D$) can be transformed into an optimization problem over the domain $\mathrm{cl}\,\mathrm{cone}\,\mathcal{U}^N(W)$. This transformation is of interest to us because the $\geq_{icv}$ constraint is defined in terms of $\mathcal{U}^N(W)$. We relate measures in $\mathcal{M}_+\left(\mathcal{U}^N(W)\right)$ with functions in $\mathrm{cl}\,\mathrm{cone}\,\mathcal{U}^N(W)$ via the functional equality:

$$u_\Lambda(\cdot) = \int_{\mathcal{U}^N(W)} u(\cdot)\, d\Lambda(u).$$

For fixed $u \in \mathrm{cl}\,\mathrm{cone}\,\mathcal{U}^N(W)$, we define

$$M(u) = \left\{\Lambda \in \mathcal{M}_+\left(\mathcal{U}^N(W)\right) \, : \, u(\cdot) = \int_{\mathcal{U}^N(W)} u(\cdot)\, d\Lambda(u)\right\}$$

to be the set of all measures in $\mathcal{M}_+\left(\mathcal{U}^N(W)\right)$ that induce the function u.

Lemma 3.3.2. *(a) $M(u)$ is convex for all $u \in \mathrm{cl}\,\mathrm{cone}\,\mathcal{U}^N(W)$.*

(b) $\alpha\, M(u_1) + (1-\alpha)\, M(u_2) \subset M(\alpha\, u_1 + (1-\alpha)\, u_2)$ for $0 \leq \alpha \leq 1$.

Proof. (a) For $\Lambda_1, \Lambda_2 \in M(u)$ and $0 \leq \alpha \leq 1$, we have

$$u(\cdot) = \int_{\mathcal{U}^N(W)} u(\cdot)\, d\Lambda_1(u)$$

and

$$u(\cdot) = \int_{\mathcal{U}^N(W)} u(\cdot)\, d\Lambda_2(u),$$

and thus

$$u(\cdot) = \int_{\mathcal{U}^N(W)} u(\cdot)\, d\left[\alpha\, \Lambda_1 + (1-\alpha)\, \Lambda_2\right](u).$$

By the choice of $0 \leq \alpha \leq 1$, $\alpha\, \Lambda_1 + (1-\alpha)\, \Lambda_2 \in \mathcal{M}_+\left(\mathcal{U}^N(W)\right)$.

(b) Choose $u_1, u_2 \in \mathrm{cl}\,\mathrm{cone}\,\mathcal{U}^N(W)$ and $0 \leq \alpha \leq 1$. Then for $\Lambda_1 \in M(u_1)$ and $\Lambda_2 \in M(u_2)$, $\alpha\, \Lambda_1 + (1-\alpha)\, \Lambda_2 \in M(\alpha\, u_1 + (1-\alpha)\, u_2)$ since

$$\begin{aligned}&\int_{\mathcal{U}^N(W)} u(\cdot)\, d\left[\alpha\,\Lambda_1 + (1-\alpha)\,\Lambda_2\right](u)\\ =\quad & \alpha \int_{\mathcal{U}^N(W)} u(\cdot)\, d\Lambda_1(u) + (1-\alpha)\int_{\mathcal{U}^N(W)} u(\cdot)\, d\Lambda_2(u)\end{aligned}$$

by linearity of the integral. Thus

$$\alpha\, M(u_1) + (1-\alpha)\, M(u_2) \subset M(\alpha\, u_1 + (1-\alpha)\, u_2)\,.$$

□

Define the new operator

$$\mathrm{PX}(u) = \arg\inf\left\{\|\Lambda\| \,:\, \Lambda \in M(u)\right\}.$$

We use the notation $\mathrm{PX}(u)$ because this operator is analogous to the projection of an element in $\mathrm{cl\,cone}\,\mathcal{U}^N(W)$ onto $\mathcal{M}_+\left(\mathcal{U}^N(W)\right)$.

Lemma 3.3.3. *The infimum is attained in $PX(u)$.*

Proof. For any $\tilde{u} \in \mathrm{cl\,cone}\,\mathcal{U}^N(W)$ with $0 \le \|\partial\tilde{u}\|_W \le 1$, we have $\tilde{u} \in \mathrm{cl\,conv}\,\mathcal{U}^N(W)$. For $\tilde{u} \in \mathrm{cl\,cone}\,\mathcal{U}^N(W)$ with $1 < \|\partial\tilde{u}\|_W$, then

$$\frac{\tilde{u}}{\|\partial\tilde{u}\|_W} \in \mathrm{cl\,conv}\,\mathcal{U}^N(W)$$

since

$$\|\partial\left[\frac{\tilde{u}}{\|\partial\tilde{u}\|_W}\right]\|_W = 1.$$

Thus, for the measure

$$\|\partial\tilde{u}\|_W\,\delta_{\tilde{u}/\|\partial\tilde{u}\|_W}$$

we have

$$\tilde{u}(\cdot) = \int_{\mathcal{U}^N(W)} u(\cdot)\, d\left[\|\partial\tilde{u}\|_W\,\delta_{\tilde{u}/\|\partial\tilde{u}\|_W}\right](u)\,.$$

It follows that

$$\|\partial\tilde{u}\|_W\,\delta_{\tilde{u}/\|\partial\tilde{u}\|_W} \in \mathrm{PX}(\tilde{u})\,.$$

If $0 < \lambda < \|\partial\tilde{u}\|_W$, then $\tilde{u}/\lambda \notin \mathcal{U}^N(W)$ since

$$\|\partial\left[\tilde{u}/\lambda\right]\|_W = \|\partial\tilde{u}\|_W/\lambda > 1.$$

If $\|\partial\tilde{u}\|_W < \lambda$, then $\|\mathrm{PX}(\tilde{u})\| < \lambda$ since $\tilde{u}/\lambda \in \mathcal{U}^N(W)$ and $\tilde{u}/\|\partial\tilde{u}\|_W \ge u/\lambda$ on W. □

The next theorem presents two equivalent forms for problem ($\epsilon-\mathrm{MP}_D$).

Theorem 3.3.4. *Problem ($\epsilon-MP_D$) has two equivalent forms:*

$$\min\left\{\mathrm{d}(\Lambda) - \epsilon\,\Lambda\left(\mathcal{U}^N(W)\right) \,:\, \mathcal{M}_+\left(\mathcal{U}^N(W)\right)\right\},$$

and

$$\min\left\{\mathrm{d}(u) - \epsilon\,\|\mathrm{PX}(u)\| \,:\, u \in \mathrm{cl\,cone}\,\mathcal{U}^N(W)\right\}.$$

PROOF. By the equivalence $u_\Lambda(\cdot) = \int_{\mathcal{U}^N(W)} u(\cdot)\, d\Lambda(u)$, it follows that $\mathrm{d}(u_\Lambda) = \mathrm{d}(\Lambda)$.

The functional $\|\mathrm{PX}(u)\|$ is convex in u. For $u_1, u_2 \in \mathrm{cl\,cone}\,\mathcal{U}^N(W)$ and $0 \le \alpha \le 1$, $\alpha\, u_1 + (1-\alpha)\, u_2 \in \mathrm{cl\,cone}\,\mathcal{U}^N(W)$ and

$$\|\mathrm{PX}(\alpha\, u_1 + (1-\alpha)\, u_2)\| \le \|\Lambda\|$$

for all $\Lambda \in M(\alpha\, u_1 + (1-\alpha)\, u_2)$. For any representation $\Lambda_1, \Lambda_2 \in \mathcal{M}_+\left(\mathcal{U}^N(W)\right)$ with $\Lambda = \alpha\,\Lambda_1 + (1-\alpha)\,\Lambda_2$, we have

$$\begin{aligned} \|\Lambda\| &= \|\alpha\,\Lambda_1 + (1-\alpha)\,\Lambda_2\| \\ &\le \alpha\,\|\Lambda_1\| + (1-\alpha)\,\|\Lambda_2\|, \end{aligned}$$

by the triangle inequality.

In particular, restrict to $\Lambda_1 \in M(u_1)$ and $\Lambda_2 \in M(u_2)$. Since $\alpha\, M(u_1) + (1-\alpha)\, M(u_2) \subset M(\alpha\, u_1 + (1-\alpha)\, u_2)$,

$$\|\Lambda\| \le \alpha\,\|\Lambda_1\| + (1-\alpha)\,\|\Lambda_2\|$$

for all $\Lambda_1 \in M(u_1)$ and $\Lambda_2 \in M(u_2)$. Take the minimum over $\Lambda_1 \in M(u_1)$ and $\Lambda_2 \in M(u_2)$ to conclude

$$\begin{aligned} & \|\mathrm{PX}(\alpha\, u_1 + (1-\alpha)\, u_2)\| \\ \le\ & \alpha\,\|\mathrm{PX}(u_1)\| + (1-\alpha)\,\|\mathrm{PX}(u_2)\|. \end{aligned}$$

Finally,

$$\|\mathrm{PX}(u)\| = \Lambda\left(\mathcal{U}^N(W)\right)$$

when $\mathrm{PX}(u) = \Lambda$. □

Using the preceding result, we can recast the optimality conditions for problem ($\epsilon-$MP) in terms of u.

Theorem 3.3.5. *Suppose assumption 3.2.1 holds. If $\hat{z} \in Z_0$ solves problem ($\epsilon -$ MP), then there exists $\hat{u} \in \mathrm{cl\,cone}\,\mathcal{U}^N(W)$ such that*

$$\begin{aligned} L(\hat{z}, \hat{u}) &= \max\{L(z, \hat{u}) \,:\, z \in Z_0\}, \\ \mathbb{E}[\hat{u}(\underline{G}(\hat{z})) - \hat{u}(\underline{Y})] &= \epsilon\,\|PX(\hat{u})\|. \end{aligned}$$

3.3.2. Duality for problem ($\epsilon-$SIP). Define the dual functional

$$\mathrm{d}(\Lambda) = \max\{L(z, \Lambda) \,:\, z \in Z_0\}$$

for problem ($\epsilon-$SIP). The corresponding dual problem is

$$(\epsilon-\mathrm{SIP}_D) \qquad \min\{\mathrm{d}(\Lambda) - \epsilon\,\Lambda(\Xi) \,:\, \Lambda \in \mathcal{M}_+(\Xi)\}.$$

By the optimality conditions and weak duality, strong duality holds between problem ($\epsilon-$SIP) and problem ($\epsilon - \mathrm{SIP}_D$).

Theorem 3.3.6. *Suppose assumption 3.2.6 holds.*

(a) If problem ($\epsilon-$SIP) has an optimal solution, then problem ($\epsilon -$ SIP$_D$) has an optimal solution and the optimal values are equal.

(b) If problem ($\epsilon -$ SIP$_D$) has an optimal solution $\hat{\Lambda}$ then any $\hat{z}$ satisfying the optimality conditions $(7)-(8)$ *with respect to $\hat{\Lambda}$ is an optimal solution for problem ($\epsilon-$SIP).*

(c) Problem ($\epsilon -$ SIP$_D$) has an optimal solution with support on a finite set $\{\xi_1, \ldots, \xi_M\} \subset \Xi$.

Using the same argument as in the previous section, we can identify utility functions in $\operatorname{cl}\operatorname{cone} U_\Xi$ and measures in $\mathcal{M}_+(\Xi)$ via the functional equality

$$u_\Lambda(\cdot) = \int_\Xi u(\cdot;\xi)\, d\Lambda(\xi).$$

For fixed $u \in \operatorname{cl}\operatorname{cone} U_\Xi$, we define

$$\begin{aligned} & M(u;\operatorname{cl}\operatorname{conv} U_\Xi) \\ = \quad & \left\{\Lambda \in \mathcal{M}_+(\Xi)\,:\, u(\cdot) = \int_\Xi u(\cdot;\xi)\, d\Lambda(\xi)\right\} \end{aligned}$$

to be the set of all measures that induce the function u. Define

$$\begin{aligned} & \mathrm{PX}(u;\operatorname{cl}\operatorname{conv} U_\Xi) \\ = \quad & \arg\min\{\|\Lambda\|\,:\,\Lambda \in M(u;\operatorname{cl}\operatorname{conv} U_\Xi)\} \end{aligned}$$

to be the element of $M(u)$ with minimum norm.

Theorem 3.3.7. *Problem ($\epsilon-$SIP$_D$) has two equivalent forms:*

$$\min\{\mathrm{d}(\Lambda) - \epsilon\,\Lambda(\Xi)\,:\,\Lambda \in \mathcal{M}_+(\Xi)\},$$

and

$$\min\{\mathrm{d}(u) - \epsilon\,\|\mathrm{PX}(u;\operatorname{cl}\operatorname{conv} U_\Xi)\|\,:\,u \in \operatorname{cl}\operatorname{cone} U_\Xi\}.$$

Next we recast the optimality conditions for problem ($\epsilon-$SIP).

Theorem 3.3.8. *Suppose assumption 3.2.1 holds. If $\hat{z} \in Z_0$ solves problem ($\epsilon -$ SIP), then there exists $\hat{u} \in cl\,cone\,U_\Xi$ such that*

$$\begin{aligned} L(\hat{z},\hat{u}) &= \max\{L(z,\hat{u})\,:\,z \in Z_0\}, \\ \mathbb{E}[\hat{u}(\underline{G}(\hat{z})) - \hat{u}(\underline{Y})] &= \epsilon\,\|PX(\hat{u};\,cl\,conv\,U_\Xi)\|. \end{aligned}$$

3.3.3. Duality for problem ($\epsilon-$NLP). The dual functional for problem ($\epsilon-$NLP) is defined as

$$\mathrm{d}(\lambda) = \max\{L(z,\lambda)\,:\,z \in Z_0\}$$

and the corresponding dual problem is

$$(\epsilon-\mathrm{NLP}_D) \qquad \min\left\{d(\lambda) - \epsilon\sum_{i=1}^{I}\lambda_i\,:\,\lambda \geq 0\right\}.$$

By the optimality conditions and weak duality, strong duality holds between problem ($\epsilon-$NLP) and problem ($\epsilon-\text{NLP}_D$).

Theorem 3.3.9. *Suppose assumption 3.2.11 holds.*

(a) If problem ($\epsilon-$NLP) has an optimal solution, then problem ($\epsilon-NLP_D$) has an optimal solution and the optimal values are equal.

(b) If problem ($\epsilon-NLP_D$) has an optimal solution $\hat{\lambda}$ then any $\hat{z}$ satisfying the optimality conditions (9) $-$ (10) *with respect to $\hat{\lambda}$ is optimal to problem ($\epsilon-$NLP).*

We will now transform problem ($\epsilon-\text{NLP}_D$) so that its domain is

$$\text{cone}\,\{u_1,\ldots,u_I\}\,.$$

We will use the equivalence $\mathrm{d}(\lambda)=\mathrm{d}(u_\lambda)$ for $u_\lambda(\cdot)=\sum_{i=1}^{I}\lambda_i u_i(\cdot)$. In this subsection, we can suppose that $\{u_1,\ldots,u_I\}$ is a linearly independent set without loss of generality. For any $u\in\text{cone}\,\{u_1,\ldots,u_I\}$, there is a unique λ such that $u(\cdot)=\sum_{i=1}^{I}\lambda_i u_i(\cdot)$.

Let

$$\text{PX}\left(u;\,\text{conv}\,\{u_1,\ldots,u_I\}\right)=\left\{\lambda\in\mathbb{R}_+^I\,:\,u(\cdot)=\sum_{i=1}^{I}\lambda_i u_i(\cdot)\right\}.$$

The set $\text{PX}\left(u;\,\text{conv}\,\{u_1,\ldots,u_I\}\right)$ is a singleton under the linear independence assumption.

Theorem 3.3.10. *Problem ($\epsilon-NLP_D$) has two equivalent forms:*

$$\min\left\{\mathrm{d}(\lambda)-\epsilon\sum_{i=1}^{I}\lambda_i\,:\,\lambda\geq 0\right\},$$

and

$$\min\left\{\mathrm{d}(u)-\epsilon\left\|\text{PX}\left(u;\,\text{conv}\,\{u_1,\ldots,u_I\}\right)\right\|_1\,:\,\lambda\geq 0\right\}.$$

Next we recast the optimality conditions.

Theorem 3.3.11. *Suppose assumption 3.2.1 holds. If $\hat{z}\in Z_0$ solves problem ($\epsilon-$NLP), then there exists $\hat{u}\in$ cone $\{u_1,\ldots,u_I\}$ such that*

$$\begin{aligned} L(\hat{z},\hat{u}) &= \max\left\{L(z,\hat{u})\,:\,z\in Z_0\right\},\\ \mathbb{E}\left[\hat{u}\left(\underline{G}(\hat{z})\right)-\hat{u}\left(\underline{Y}\right)\right] &= \epsilon\left\|PX\left(\hat{u};\,conv\,\{u_1,\ldots,u_I\}\right)\right\|_1. \end{aligned}$$

3.4. Finite probability spaces

On finite probability spaces, theorem 2.2.9 furnishes another form of problem (MP) as discussed in [**1**]. Suppose $\Omega_1=\{\omega_{11},\ldots,\omega_{1J}\}$ and $\Omega_2=\{\omega_{21},\ldots,\omega_{2K}\}$ are both finite sample spaces and the collection $\mathcal{F}_i$ makes all scenarios on Ω_i measurable for $i=1,2$. We define sets $\mathcal{J}=\{1,\ldots,J\}$ and $\mathcal{K}=\{1,\ldots,K\}$ to index the samples. Our coupled problem is:

$$
\begin{array}{llll}
\text{(CP)} & \max\limits_{z \in Z_0} & f(z) & \\
(11) & \text{s.t.} & \left[\underline{G}(z)\right](\omega_{1j}) \geq \dfrac{\sum_{k=1}^{K} \pi_{jk} \underline{Y}(\omega_{2k})}{P_1(\{\omega_{1j}\})}, & \forall j \in \mathcal{J}, \\
(12) & & \displaystyle\sum_{k=1}^{K} \pi_{jk} = P_1(\{\omega_{1j}\}), & \forall j \in \mathcal{J}, \\
(13) & & \displaystyle\sum_{j=1}^{J} \pi_{jk} = P_2(\{\omega_{2k}\}), & \forall k \in \mathcal{K}.
\end{array}
$$

Problem (CP) is a standard convex programming problem with finitely many decision variables and finitely many constraints. Its perturbation is

$$
\begin{array}{llll}
(\epsilon\text{−CP}) & \max\limits_{z \in Z_0} & f(z) & \\
(14) & \text{s.t.} & \left[\underline{G}(z)\right](\omega_{1j}) & \forall j \in \mathcal{J}, \\
 & & \geq \dfrac{\sum_{k=1}^{K} \pi_{jk} \left(\underline{Y}(\omega_{2k}) + \epsilon\, \underline{1}\right)}{P_1(\{\omega_{1j}\})}, & \\
(15) & & \displaystyle\sum_{k=1}^{K} \pi_{jk} = P_1(\{\omega_{1j}\}), & \forall j \in \mathcal{J}, \\
(16) & & \displaystyle\sum_{j=1}^{J} \pi_{jk} = P_2(\{\omega_{2k}\}), & \forall k \in \mathcal{K}.
\end{array}
$$

The inequality

$$\left[\underline{G}(z)\right](\omega_{1j}) \geq \frac{\sum_{k=1}^{K} \pi_{jk} \left(\underline{Y}(\omega_{2k}) + \epsilon\, \underline{1}\right)}{P_1(\{\omega_{1j}\})}$$

is equivalent to

$$\left[\underline{G}(z)\right](\omega_{1j}) \geq \frac{\sum_{k=1}^{K} \pi_{jk} \left(\underline{Y}(\omega_{2k})\right)}{P_1(\{\omega_{1j}\})} + \epsilon\, \underline{1}$$

using the fact that $\sum_{k=1}^{K} \pi_{jk} = P_1(\{\omega_{1j}\})$. The constraints (14) are then equivalent to

$$\left[\underline{G}(z)\right](\omega_{1j}) \geq \frac{\sum_{k=1}^{K} \pi_{jk}\, \underline{Y}(\omega_{2k})}{P_1(\{\omega_{1j}\})} + \epsilon\, \underline{1}, \quad \forall j \in \mathcal{J}.$$

The following proposition describes the relationship between problem $(\epsilon - \text{MP})$ and problem $(\epsilon - \text{CP})$.

Proposition 3.4.1. *For $\epsilon < 0$, problem ($\sqrt{N}\,\epsilon$−MP) is a relaxation of problem (ϵ−CP) on $\Omega_1 = \{\omega_{11}, \ldots, \omega_{1J}\}$ and $\Omega_2 = \{\omega_{21}, \ldots, \omega_{2K}\}$.*

Proof. Problem $(\epsilon - \text{CP})$ is equivalent to

$$\begin{aligned} \max_{z \in Z_0} \quad & f(z) \\ \text{s.t.} \quad & \sum_{j=1}^{J} P_1(\{\omega_{1j}\})\, u([\underline{\mathrm{G}}(z)](\omega_{1j})) && \forall u \in \mathcal{U}^N(W), \\ & \geq \sum_{k=1}^{K} P_2(\{\omega_{2k}\})\, u(\underline{\mathrm{Y}}(\omega_{2k}) + \epsilon\, \underline{1}), \end{aligned}$$

by construction. In this setting, problem $(\epsilon - \mathrm{MP})$ is

$$\begin{aligned} \max_{z \in Z_0} \quad & f(z) \\ \text{s.t.} \quad & \sum_{j=1}^{J} P_1(\{\omega_{1j}\})\, u([\underline{\mathrm{G}}(z)](\omega_{1j})) && \forall u \in \mathcal{U}^N(W), \\ & \geq \sum_{k=1}^{K} P_2(\{\omega_{2k}\})\, u(\underline{\mathrm{Y}}(\omega_{2k})) + \epsilon. \end{aligned}$$

For any $u \in \mathcal{U}^N(W)$ we have

$$|\mathbb{E}[u(\underline{\mathrm{Y}} + \epsilon\, \underline{1})] - \mathbb{E}[u(\underline{\mathrm{Y}})]| \leq \|\epsilon\, \underline{1}\|_2 = \sqrt{N}\,\epsilon,$$

since $\|\partial u\|_W \leq 1$. Rearrange to obtain

$$\mathbb{E}[u(\underline{\mathrm{Y}} + \epsilon\, \underline{1})] \geq \mathbb{E}[u(\underline{\mathrm{Y}})] + \sqrt{N}\,\epsilon.$$

For any $z \in Z_0$ such that $(14) - (16)$ we have

$$\begin{aligned} & \sum_{j=1}^{J} P_1(\{\omega_{1j}\})\, u([\underline{\mathrm{G}}(z)](\omega_{1j})) \quad \forall u \in \mathcal{U}^N(W), \\ \geq & \sum_{k=1}^{K} P_2(\{\omega_{2k}\})\, u(\underline{\mathrm{Y}}(\omega_{2k}) + \epsilon\, \underline{1}), \end{aligned}$$

and thus

$$\begin{aligned} & \sum_{j=1}^{J} P_1(\{\omega_{1j}\})\, u([\underline{\mathrm{G}}(z)](\omega_{1j})) \quad \forall u \in \mathcal{U}^N(W), \\ \geq & \sum_{k=1}^{K} P_2(\{\omega_{2k}\})\, u(\underline{\mathrm{Y}}(\omega_{2k})) + \sqrt{N}\,\epsilon. \end{aligned}$$

□

3.4.1. Optimality conditions for problem ($\epsilon-$CP). We now derive optimality conditions for problem ($\epsilon-$CP). Introduce multipliers $\{\lambda_j\}_{j\in\mathcal{J}} \subset \mathbb{R}^N$ corresponding to the constraints (14). The partial Lagrangian for problem ($\epsilon-$CP) is

$$L(z,\pi,\lambda) = f(z) + \sum_{j=1}^{J}\langle\lambda_j, [\underline{\mathrm{G}}(z)](\omega_{1j}) - \frac{\sum_{k=1}^{K}\pi_{jk}\underline{\mathrm{Y}}(\omega_{2k})}{P_1(\{\omega_{1j}\})} - \epsilon\,\underline{1}\rangle.$$

Since problem ($\epsilon-$CP) is a standard convex programming problem with finitely many constraints, we can appeal to the usual Slater condition.

Assumption 3.4.2. *There exists $\tilde{z} \in Z_0$ and $\tilde{\pi} \geq 0$ such that*

$$\begin{aligned}[\underline{G}(\tilde{z})](\omega_{1j}) > \frac{\sum_{k=1}^{K}\tilde{\pi}_{jk}\underline{Y}(\omega_{2k})}{P_1(\{\omega_{1j}\})} + \epsilon\,\underline{1}, &\quad \forall j \in \mathcal{J},\\ \sum_{k=1}^{K}\tilde{\pi}_{jk} = P_1(\{\omega_{1j}\}), &\quad \forall j \in \mathcal{J},\\ \sum_{j=1}^{J}\tilde{\pi}_{jk} = P_2(\{\omega_{2k}\}), &\quad \forall k \in \mathcal{K}.\end{aligned}$$

The usual optimality conditions for nonlinear programming follow.

Theorem 3.4.3. *Suppose assumption 3.4.2 holds. If $(\hat{z},\hat{\pi})$ solves problem ($\epsilon-$CP), then there exists $\hat{\lambda} \in \mathbb{R}_+^{JN}$ such that*

$$L\left(\hat{z},\hat{\pi},\hat{\lambda}\right) = \max\left\{L\left(z,\pi,\hat{\lambda}\right) : \begin{array}{c}\sum_{k=1}^{K}\pi_{jk} = P_1(\{\omega_{1j}\}), \quad \forall j \in \mathcal{J},\\ \sum_{j=1}^{J}\pi_{jk} = P_2(\{\omega_{2k}\}), \quad \forall k \in \mathcal{K},\\ z \in Z_0,\ \pi \geq 0\end{array}\right\}, \tag{17}$$

$$\sum_{j=1}^{J}\langle\hat{\lambda}_j,\ [\underline{G}(\hat{z})](\omega_{1j}) - \frac{\sum_{k=1}^{K}\hat{\pi}_{jk}\underline{Y}(\omega_{2k})}{P_1(\{\omega_{1j}\})}\rangle = \epsilon\sum_{j=1}^{J}\langle\underline{1},\hat{\lambda}_j\rangle, \tag{18}$$

$$\hat{\lambda} \geq 0. \tag{19}$$

3.4.2. Duality and decomposition for problem ($\epsilon-$CP). The dual functional for problem ($\epsilon-$CP) is

$$\mathrm{d}(\lambda) = \max\left\{L(z,\pi,\lambda) : \begin{array}{c}\sum_{k=1}^{K}\pi_{jk} = P_1(\{\omega_{1j}\}), \quad \forall j \in \mathcal{J},\\ \sum_{j=1}^{J}\pi_{jk} = P_2(\{\omega_{2k}\}), \quad \forall k \in \mathcal{K},\\ z \in Z_0,\ \pi \geq 0\end{array}\right\}.$$

The dual to problem ($\epsilon-$CP) is

$$\min\left\{\mathrm{d}(\lambda) - \epsilon\sum_{j=1}^{J}\langle\underline{1},\lambda_j\rangle : \lambda \geq 0\right\}. \tag{$\epsilon-\mathrm{CP}_D$}$$

We can show that strong duality holds between these two problems.

Theorem 3.4.4. *Suppose assumption 3.4.2 holds.*

(a) If problem ($\epsilon-$CP) has an optimal solution, then problem ($\epsilon-CP_D$) has an optimal solution and the optimal values are equal.

(b) If problem ($\epsilon-CP_D$) has an optimal solution $\hat{\lambda}$, then any $(\hat{z}, \hat{\pi})$ that satisfies $(17)-(19)$ *corresponding to $\hat{\lambda}$ is optimal to problem ($\epsilon-$CP).*

For a candidate dual solution $\lambda \geq 0$, we solve $\mathrm{d}(\lambda)$ to obtain a candidate primal solution $(\hat{z}(\lambda), \hat{\pi}(\lambda))$. Next we use $(\hat{z}(\lambda), \hat{\pi}(\lambda))$ to compute a subgradient of $\mathrm{d}(\lambda)$. We can update λ by moving in the direction of this subgradient. The following result holds by construction of the Lagrangian $L(z, \pi, \lambda)$.

Proposition 3.4.5. *For $\tilde{\lambda} \geq 0$,*

$$\left(\left[\underline{G}\left(\hat{z}\left(\tilde{\lambda}\right)\right)\right](\omega_{1j}) - \frac{\sum_{k=1}^{K} \hat{\pi}_{jk}\left(\tilde{\lambda}\right) \underline{Y}(\omega_{2k})}{P_1(\{\omega_{1j}\})}\right)_{j \in \mathcal{J}} \in \mathbb{R}^{J\,N}$$

is a subgradient of $d(\lambda)$ at $\lambda = \tilde{\lambda}$.

The dual functional $\mathrm{d}(\lambda)$ has decomposable structure by inspection.

Proposition 3.4.6. *The dual functional $d(\lambda)$ decomposes into*

$$\max\left\{f(z) + \sum_{j=1}^{J} \langle \lambda_j, [\underline{G}(z)](\omega_j)\rangle \,:\, z \in Z_0\right\}$$

and

$$\max\left\{-\sum_{j=1}^{J}\sum_{k=1}^{K} \frac{\pi_{jk}\langle \lambda_j, \underline{Y}(\omega_{2k})\rangle}{P_1(\omega_{1j})} \,:\, \begin{array}{c} \sum_{k=1}^{K} \pi_{jk} = P_1(\{\omega_{1j}\}), \quad \forall j \in \mathcal{J}, \\ \sum_{j=1}^{J} \pi_{jk} = P_2(\{\omega_{2k}\}), \quad \forall k \in \mathcal{K}, \\ z \in Z_0,\ \pi \geq 0 \end{array}\right\}.$$

The first layer of $\mathrm{d}(\lambda)$ is a standard convex optimization problem. The second is a linear program.

CHAPTER 4

Sample average approximation of a class of multivariate integral stochastic order constrained programs

4.1. Sample average approximation

In this chapter we study sample average approximation for problem ($\epsilon-$MP). The structure of this chapter is based on the development in [**16**]. There are two important extensions of [**16**] in this chapter. First, we explore SAA for problem ($\epsilon-$MP) and problem ($\epsilon-$SIP), both of which generalize the problem class in [**16**]. Second, we sample for $\underline{\mathrm{G}}(z)$ and $\underline{\mathrm{Y}}$ separately, whereas it is assumed in [**16**] that samples are taken from the jointly distributed random vector $(\underline{\mathrm{G}}(z), Y)$. In practice, the benchmark usually does not need to be sampled at all because it is designed by the decision maker. Further, collecting unnecessary samples for either the random-variable-valued mapping or the benchmark will lead to a significant computational burden as this chapter shows.

This chapter establishes the significant fact that sample average approximation is effective for problem ($\epsilon-$MP) and problem ($\epsilon-$SIP). If it were not, then there would be no hope of solving this class of problems in practice.

4.1.1. Sample average approximation for problem ($\epsilon-$MP). We sample from $(\Omega_1, \mathcal{F}_1, P_1)$ for the random-variable-valued mapping $\underline{\mathrm{G}}(z)$ and we sample from $(\Omega_2, \mathcal{F}_2, P_2)$ for the benchmark $\underline{\mathrm{Y}}$. In some applications, the benchmark does not need to be sampled at all. In that case K is fixed.

We add structure to $(\Omega_1, \mathcal{F}_1)$ and (Ω_2, F_2) to continue.

Definition 4.1.1. [**12**, Section 1.4] A measurable space $(\Omega, \mathcal{F})$ is nice if there is an injective map $\phi : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$, where $\mathcal{B}$ is the Borel $\sigma-$algebra on $\mathbb{R}$, such that both ϕ and ϕ^{-1} are measurable.

A nice measurable space $(\Omega, \mathcal{F})$ can be identified with $\mathbb{R}$ in this manner; that is, a sample point $\omega \in \Omega$ uniquely corresponds to a point in $\mathbb{R}$. We will assume $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ are nice for the rest of this chapter.

Define

$$\mathbb{R}^{\mathbb{N}} = \{\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots) : \omega_i \in \mathbb{R}\} = \{\text{functions } \boldsymbol{\omega} : \mathbb{N} \to \mathbb{R}\},$$

where $\mathbb{N} = \{1, 2, \ldots\}$ is the set of natural numbers. We let $\mathcal{B}^{\mathbb{N}}$ be the $\sigma-$field on $\mathbb{R}^{\mathbb{N}}$ generated by finite dimensional sets of the form $\{\omega : \omega_i \in B_i,\, 1 \leq i \leq I\}$ for $B_i \in \mathcal{B}$. The Kolmogorov extension theorem (see [**12**, Chapter 1, Theorem 4.11]) establishes probability measures $\mathcal{P}_1$ and $\mathcal{P}_2$ on $\left(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}}\right)$ such that

$$\mathcal{P}_1(\boldsymbol{\omega} : \omega_i \in (a_i, b_i],\, 1 \leq i \leq I) = \times_{i=1}^{I} P_1(\omega \in (a_i, b_i])$$

and

$$\mathcal{P}_2\left(\boldsymbol{\omega} : \omega_i \in (a_i, b_i],\ 1 \leq i \leq I\right) = \times_{i=1}^{I} P_2\left(\omega \in (a_i, b_i]\right),$$

for all rectangles $\times_{i=1}^{I}(a_i, b_i]$ with $a_i < b_i$. We use $\left(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}}, \mathcal{P}_1\right)$ to denote the space of random samples from $(\Omega_1, \mathcal{F}_1, P_1)$ and we use $\left(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}}, \mathcal{P}_2\right)$ to denote the space of random samples from $(\Omega_2, \mathcal{F}_2, P_2)$, For convenience, we also introduce

$$\left(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}}, \mathcal{P}\right) = \left(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}}, \mathcal{P}_1\right) \times \left(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}}, \mathcal{P}_2\right)$$

to denote the joint probability space of random samples. The choice of $\mathcal{P} = \mathcal{P}_1 \times \mathcal{P}_2$ indicates that samples for the random-variable-valued mapping and samples for the benchmark are independent.

Let δ_ω be the point-mass at $\omega \in \Omega$, let

$$\hat{P}_{1J}\left(\boldsymbol{\omega}\right) = \frac{1}{J}\sum_{j=1}^{J} \delta_{\omega_{1j}}$$

be the empirical distribution from the first J sample points from $\left(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}}, \mathcal{P}_1\right)$, and let

$$\hat{P}_{2K}\left(\boldsymbol{\omega}\right) = \frac{1}{K}\sum_{k=1}^{K} \delta_{\omega_{2k}}$$

be the empirical distribution from the first K sample points from $\left(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}}, \mathcal{P}_2\right)$. Equal probability is assigned to all points in $\hat{P}_{1J}\left(\boldsymbol{\omega}\right)$ and $\hat{P}_{2K}\left(\boldsymbol{\omega}\right)$. By the Glivenko-Cantelli theorem, $\hat{P}_{1J}\left(\boldsymbol{\omega}\right) \to P_1$ in the topology of weak convergence. That is, for $\mathcal{P}_1-$almost all $\boldsymbol{\omega} \in \mathbb{R}^{\mathbb{N}}$ we have

$$\int_\Omega f\left(\omega\right) \hat{P}_{1J}\left(d\omega\right) \to \int_\Omega f\left(\omega\right) P_1\left(d\omega\right)$$

as $J \to \infty$ for all bounded functions $f \in \mathcal{C}\left(\mathbb{R}\right)$. In the above integration, we are technically abusing notation by writing $\hat{P}_{1J}\left(d\omega\right)$ because $\hat{P}_{1J}$ is a random variable that depends on $\boldsymbol{\omega} \in \mathbb{R}^N$. When we write $\hat{P}_{1J}\left(d\omega\right)$, it will be understood that we are integrating over Ω with respect to the measure $\hat{P}_{1J}$ and the dependence of $\hat{P}_{1J}$ on $\boldsymbol{\omega} \in \mathbb{R}^{\mathbb{N}}$ is suppressed.

The estimate of $g\left(z, u\right)$ based on J sample points from $(\Omega_1, \mathcal{F}_1, P_1)$ and K sample points from $(\Omega_2, \mathcal{F}_2, P_2)$ is

$$g_{JK}\left(z, u\right) \triangleq \frac{1}{J}\sum_{j=1}^{J} u\left(\left[\underline{\mathrm{G}}\left(z\right)\right]\left(\omega_{1j}\right)\right) - \frac{1}{K}\sum_{k=1}^{K} u\left(\underline{\mathrm{Y}}\left(\omega_{2k}\right)\right).$$

The sample average approximation of problem ($\epsilon-$MP) based on J samples from $(\Omega_1, \mathcal{F}_1, P_1)$ and K samples from $(\Omega_2, \mathcal{F}_2, P_2)$ is

$$(\epsilon - \text{SMP}) \qquad \max_{z \in Z_0} \quad f\left(z\right)$$

$$(1) \qquad \text{s.t.} \quad g_{JK}\left(z, u\right) \geq \epsilon, \qquad \forall u \in \mathcal{U}^N\left(W\right),$$

and its feasible region, set of optimal solutions, and optimal value are:

$$
\begin{aligned}
Z^{\epsilon}_{JK} &\triangleq \{z \in Z_0 : (1)\},\\
S^{\epsilon}_{JK} &\triangleq \arg\max \{f(z) : z \in Z^{\epsilon}_{JK}\},\\
\nu^{\epsilon}_{JK} &\triangleq \max \{f(z) : z \in Z^{\epsilon}_{JK}\}.
\end{aligned}
$$

The corresponding perturbed sample average approximation of problem ($\epsilon-$CP) is

$$
\begin{aligned}
&(\epsilon\text{-SCP}) && \max_{z \in Z_0,\, \pi \geq 0} && f(z) && \\
&(2) && \text{s.t.} && \frac{1}{J}\left[\left[\underline{\mathrm{G}}(z)\right](\omega_{1j})\right] \geq \sum_{k=1}^{K} \pi_{jk}\left(\underline{\mathrm{Y}}(\omega_{2k}) + \epsilon\, \underline{1}\right), && \forall j \in \mathcal{J},\\
&(3) && && \sum_{k=1}^{K} \pi_{jk} = \frac{1}{J}, && \forall j \in \mathcal{J},\\
&(4) && && \sum_{j=1}^{J} \pi_{jk} = \frac{1}{K}, && \forall k \in \mathcal{K}.
\end{aligned}
$$

Define the feasible region, set of optimal solutions, and optimal value for problem ($\epsilon-$SCP):

$$
\begin{aligned}
Z^{\epsilon}_{CP,JK} &\triangleq \{(z,\pi) \in Z_0 : (2)-(4)\},\\
S^{\epsilon}_{CP,JK} &\triangleq \arg\max \left\{f(z) : (z,\pi) \in Z^{\epsilon}_{CP,JK}\right\},\\
\nu^{\epsilon}_{CP,JK} &\triangleq \max \left\{f(z) : (z,\pi) \in Z^{\epsilon}_{CP,JK}\right\}.
\end{aligned}
$$

It is easier to study the properties of sample average approximation for problem ($\epsilon-$SMP), but it is easier to solve problem ($\epsilon-$SCP).

In Chapter 3, we perturbed problem (MP) so that it could satisfy the Slater condition. This same perturbation serves a new purpose in this chapter. For $\epsilon = 0$, there is no guarantee that problem ($\epsilon-$SMP) is feasible even when J and K are large and the original problem ($\epsilon-$MP) is feasible. For example, suppose the constraint

$$\mathbb{E}\left[u\left(\left[\underline{\mathrm{G}}(z)\right](\omega_{1j})\right)\right] - \mathbb{E}\left[u\left(\underline{\mathrm{Y}}\right)\right] \geq 0$$

is binding for some nontrivial (i.e. non-constant) $u \in \mathcal{U}^N(W)$ at all $z \in Z_0$ (with a possibly different u for each $z \in Z_0$). We can choose $\tilde{\epsilon} > 0$ and J^* and K^* so that

$$\left|\int u\left(\left[\underline{\mathrm{G}}(z)\right](\omega)\right) P(d\omega) - \int u\left(\left[\underline{\mathrm{G}}(z)\right](\omega)\right) \hat{P}_{1J}(d\omega)\right| < \tilde{\epsilon}$$

and

$$\left|\int u\left(\underline{\mathrm{Y}}(\omega)\right) P(d\omega) - \int u\left(\underline{\mathrm{Y}}(\omega)\right) \hat{P}_{2K}(d\omega)\right| < \tilde{\epsilon}$$

for all $J \geq J^*$ and $K \geq K^*$ for each of these z and u. However, these bounds do not guarantee that the inequality

$$\int u\left(\left[\underline{\mathrm{G}}(z)\right](\omega)\right) \hat{P}_{1J}(d\omega) - \int u\left(\underline{\mathrm{Y}}(\omega)\right) \hat{P}_{2K}(d\omega) \geq 0$$

will hold when originally $\mathbb{E}\left[u\left(\underline{\mathrm{G}}\left(z,\omega\right)\right)\right]-\mathbb{E}\left[u\left(\underline{\mathrm{Y}}\right)\right]=0$, no matter how small $\tilde{\epsilon}$ becomes. The perturbations in problems $(\epsilon-\mathrm{SMP})$ and $(\epsilon-\mathrm{SCP})$ provide the necessary slack in the stochastic order constraint to overcome this feasibility issue.

We estimate problem ($\epsilon-$SIP) with

$$(\epsilon-\mathrm{SSIP})\qquad \max_{z\in Z_0}\quad f\left(z\right)$$

$$(5)\qquad \text{s.t.}\quad g_{JK}\left(z,u\right)\geq\epsilon,\qquad \forall u\in U_{\Xi}.$$

The feasible region, set of optimal solutions, and optimal value of problem ($\epsilon-$SSIP) are:

$$\begin{aligned} Z_{JK}^{\epsilon}\left(U_{\Xi}\right) &\triangleq \left\{z\in Z_0\,:\,(5)\right\},\\ S_{JK}^{\epsilon}\left(U_{\Xi}\right) &\triangleq \arg\max\left\{f\left(z\right)\,:\,z\in Z_{JK}^{\epsilon}\left(U_{\Xi}\right)\right\},\\ \nu_{JK}^{\epsilon}\left(U_{\Xi}\right) &\triangleq \max\left\{f\left(z\right)\,:\,z\in Z_{JK}^{\epsilon}\left(U_{\Xi}\right)\right\}.\end{aligned}$$

The sample average approximation of problem ($\epsilon-$NLP) is

$$(\epsilon\text{-SNLP})\qquad \max_{z\in Z_0}\quad f\left(z\right)$$

$$(6)\qquad \text{s.t.}\quad g_{JK}\left(z,u_i\right)\geq\epsilon,\qquad i=1,\ldots,I,$$

with corresponding feasible region, set of optimal solutions, and optimal value:

$$\begin{aligned} Z_{JK}^{\epsilon}\left(\{u_1,\ldots,u_I\}\right) &\triangleq \left\{z\in Z_0\,:\,(6)\right\},\\ S_{JK}^{\epsilon}\left(\{u_1,\ldots,u_I\}\right) &\triangleq \arg\max\left\{f\left(z\right)\,:\,z\in Z^{\epsilon}\left(\{u_1,\ldots,u_I\}\right)\right\},\\ \nu_{JK}^{\epsilon}\left(\{u_1,\ldots,u_I\}\right) &\triangleq \max\left\{f\left(z\right)\,:\,z\in Z^{\epsilon}\left(\{u_1,\ldots,u_I\}\right)\right\}.\end{aligned}$$

4.1.2. Analytical properties. Set $\pi_J\triangleq\frac{1}{J}\sum_{j=1}^{J}\Pi\left(\omega_{1j}\right)$, which only depends on the sample from $(\Omega_1,\mathcal{F}_1,P_1)$. Also define

$$\phi_{JK}\left(\underline{\mathrm{G}}\left(z\right),\underline{\mathrm{Y}}\right)\triangleq\frac{1}{J}\sum_{j=1}^{J}\|\left[\underline{\mathrm{G}}\left(z\right)\right]\left(\omega_{1j}\right)\|_2+\frac{1}{K}\sum_{k=1}^{K}\|\underline{\mathrm{Y}}\left(\omega_{2k}\right)\|_2.$$

Proposition 4.1.2. *Under assumption 3.1.1, $\phi_{JK}(\cdot)$ is uniformly bounded on Z_0 for $\mathcal{P}-$almost all $\boldsymbol{\omega}\in\mathbb{R}^{\mathbb{N}}$.*

Proof. It has already been shown that the inequality

$$\begin{aligned}\Phi\left(\left[\underline{\mathrm{G}}\left(z\right)\right]\left(\omega\right),\underline{\mathrm{Y}}\left(\omega\right)\right) &\leq \|\left[\underline{\mathrm{G}}\left(z_0\right)\right]\left(\omega\right)\|_2+\Pi\left(\omega\right)\mathbb{D}\left(Z_0\right)+\|\underline{\mathrm{Y}}\left(\omega\right)\|_2\\ &= \Phi\left(\left[\underline{\mathrm{G}}\left(z_0\right)\right]\left(\omega\right),\underline{\mathrm{Y}}\left(\omega\right)\right)+\mathbb{D}\left(Z_0\right)\Pi\left(\omega\right)\end{aligned}$$

holds for any fixed $z_0\in Z_0$.

Let $\hat{P}_{JK}=\hat{P}_{1J}\times\hat{P}_{2K}$ be the probability measure on $(\Omega,\mathcal{F})$ determined by the first J samples from $(\Omega_1,\mathcal{F}_1,P_1)$ and the first K samples from $(\Omega_2,\mathcal{F}_2,P_2)$. We have assumed that $\underline{\mathrm{G}}\left(z\right)$ for all $z\in Z_0$, Π, and $\underline{\mathrm{Y}}$ are all integrable with respect to P. It follows that $\underline{\mathrm{G}}\left(z\right)$ for all $z\in Z_0$, Π, and $\underline{\mathrm{Y}}$ are all integrable with respect to

$\hat{P}_{JK}$. The random variable $\Phi\left(\left[\underline{\mathrm{G}}\left(z_0\right)\right], \underline{\mathrm{Y}}\right)$ is then integrable with respect to $\hat{P}_{JK}$ and

$$\begin{aligned} & \phi_{JK}\left(\underline{\mathrm{G}}\left(z\right), \underline{\mathrm{Y}}\right) \\ \leq \quad & \mathbb{E}_{\hat{P}_{JK}}\left[\Phi\left(\left[\underline{\mathrm{G}}\left(z_0\right)\right], \underline{\mathrm{Y}}\right)\right] + \mathbb{D}\left(Z_0\right) \mathbb{E}_{\hat{P}_{JK}}\left[\Pi\right] \\ < \quad & \infty \end{aligned}$$

for all $z \in Z_0$. □

The next proposition shows that the estimate $g_{JK}\left(z, u\right)$ is a bounded function.

Proposition 4.1.3. *Suppose that assumption 3.1.1 holds, $g_{JK}\left(z, u\right)$ is bounded for $u \in \mathcal{U}^N\left(W\right)$ for $\mathcal{P}-$almost all $\boldsymbol{\omega} \in \mathbb{R}^{\mathbb{N}}$.*

Proof. We have $\left|g_{JK}\left(z, u\right)\right| \leq \phi_{JK}\left(z\right)$ and that $\phi_{JK}\left(z\right)$ is uniformly bounded in $z \in Z_0$ for $\mathcal{P}-$almost all $\boldsymbol{\omega} \in \mathbb{R}^{\mathbb{N}}$. □

Now we establish continuity of $g_{JK}\left(z, u\right)$.

Proposition 4.1.4. *$g_{JK}\left(z, u\right)$ is Lipschitz continuous on $Z_0 \times \mathcal{U}^N\left(W\right)$ for $\mathcal{P}-$almost all $\boldsymbol{\omega} \in \mathbb{R}^{\mathbb{N}}$.*

Proof. Recall the definition of $\hat{P}_{JK}$ from earlier. For any given $\left(z_1, u_1\right), \left(z_2, u_2\right) \in Z_0 \times \mathcal{U}^N\left(W\right)$, we have

$$\begin{aligned} & \left|g_{JK}\left(z_1, u_1\right) - g_{JK}\left(z_2, u_2\right)\right| \\ = \quad & \left|g_{JK}\left(z_1, u_1\right) - g_{JK}\left(z_1, u_2\right) + g_{JK}\left(z_1, u_2\right) - g_{JK}\left(z_2, u_2\right)\right| \\ \leq \quad & \left|g_{JK}\left(z_1, u_1\right) - g_{JK}\left(z_1, u_2\right)\right| + \left|g_{JK}\left(z_1, u_2\right) - g_{JK}\left(z_2, u_2\right)\right| \\ = \quad & \left|\mathbb{E}_{\hat{P}_{JK}}\left[u_1\left(\underline{\mathrm{G}}\left(z_1\right)\right) - u_1\left(\underline{\mathrm{Y}}\right) - u_2\left(\underline{\mathrm{G}}\left(z_1\right)\right) + u_2\left(\underline{\mathrm{Y}}\right)\right]\right| \\ & + \left|\mathbb{E}_{\hat{P}_{JK}}\left[u_2\left(\underline{\mathrm{G}}\left(z_1\right)\right) - u_2\left(\underline{\mathrm{G}}\left(z_2\right)\right)\right]\right|. \end{aligned}$$

By the earlier propositions,

$$\begin{aligned} & \left|\mathbb{E}_{\hat{P}_{JK}}\left[u_1\left(\underline{\mathrm{G}}\left(z_1\right)\right) - u_1\left(\underline{\mathrm{Y}}\right) - u_2\left(\underline{\mathrm{G}}\left(z_1\right)\right) + u_2\left(\underline{\mathrm{Y}}\right)\right]\right| \\ \leq \quad & \mathbb{E}_{\hat{P}_{JK}}\left[\left|u_1\left(\underline{\mathrm{G}}\left(z_1\right)\right) - u_1\left(\underline{\mathrm{Y}}\right) - u_2\left(\underline{\mathrm{G}}\left(z_1\right)\right) + u_2\left(\underline{\mathrm{Y}}\right)\right|\right] \\ \leq \quad & \mathbb{E}_{\hat{P}_{JK}}\left[\Phi\left(\left[\underline{\mathrm{G}}\left(z_1\right)\right]\left(\omega\right), \underline{\mathrm{Y}}\left(\omega\right)\right)\left\|u_1 - u_2\right\|_W\right] \\ \leq \quad & \Gamma_{JK}\left\|u_1 - u_2\right\|_W, \end{aligned}$$

where

$$\Gamma_{JK} = \max_{z \in Z_0} \phi_{JK}\left(z\right) < \infty.$$

Further,

$$\begin{aligned} & \left|\mathbb{E}_{\hat{P}_{JK}}\left[u_1\left(\underline{\mathrm{G}}\left(z_1\right)\right) - u_2\left(\underline{\mathrm{G}}\left(z_2\right)\right)\right]\right| \\ \leq \quad & \mathbb{E}_{\hat{P}_{JK}}\left[\left|u_2\left(\underline{\mathrm{G}}\left(z_1\right)\right) - u_2\left(\underline{\mathrm{G}}\left(z_2\right)\right)\right|\right] \\ \leq \quad & \mathbb{E}_{\hat{P}_{JK}}\left[\Pi\left(\omega\right)\left\|\underline{\mathrm{G}}\left(z_1\right) - \underline{\mathrm{G}}\left(z_2\right)\right\|_2\right] \\ \leq \quad & \pi_J\left\|z_1 - z_2\right\|_2, \end{aligned}$$

where

$$\pi_J = \frac{1}{J} \sum_{j=1}^{J} \Pi\left(\omega_{1j}\right) < \infty.$$

In summary,

$$\left| g_{JK}\left(z_1, u_1\right) - g_{JK}\left(z_2, u_2\right) \right| \leq \left(\Gamma_{JK} + \pi_J\right) \left(\|u_1 - u_2\|_W + \|z_1 - z_2\|_2\right).$$

□

4.2. Convergence analysis of sample average approximation

In this section, we verify consistency of SAA for problem ($\epsilon-$MP) and problem ($\epsilon-$SIP).

4.2.1. Convergence analysis for problem ($\epsilon-$MP). The next proposition verifies that $g_{JK}(\cdot)$ converges to $g(\cdot)$ uniformly on $Z_0 \times \mathcal{U}^N(W)$.

Proposition 4.2.1. *For $\mathcal{P}-$almost all $\boldsymbol{\omega} \in \mathbb{R}^{\mathbb{N}}$,*

$$\sup\left\{ \left| g(z,u) - g_{JK}(z,u) \right| : \ (z,u) \in Z_0 \times \mathcal{U}^N(W) \right\} \to 0$$

as $J, K \to \infty$.

Proof. For fixed $(z,u) \in Z_0 \times \mathcal{U}^N(W)$, $g_{JK}(z,u) \to g(z,u)$ as $J, K \to \infty$ by the Glivenko-Cantelli theorem. The space $Z_0 \times \mathcal{U}^N(W)$ is compact, and $g(z,u)$ and all $g_{JK}(z,u)$ are continuous, so this convergence is uniform.

To formalize this argument, choose $\epsilon > 0$. For any fixed $(z,u) \in Z_0 \times \mathcal{U}^N(W)$ it is true that $g_{JK}(z,u) \to g(z,u)$ as $J, K \to \infty$ for $\mathcal{P}-$almost all $\boldsymbol{\omega} \in \mathbb{R}^{\mathbb{N}}$. In particular, there are J^* and K^* (that depend on (z,u) and $\boldsymbol{\omega}$) such that

$$\left| g_{JK}(z,u) - g(z,u) \right| < \epsilon/3$$

for $J \geq J^*(z,u)$ and $K \geq K^*(z,u)$. It is also true that $g_{JK}(z,u)$ are Lipschitz continuous on $Z_0 \times \mathcal{U}^N(W)$ for $\mathcal{P}-$almost all $\boldsymbol{\omega} \in \mathbb{R}^{\mathbb{N}}$. Thus, there is a $\delta_1 > 0$ (that depends on $\boldsymbol{\omega}$, but not on (z,u) by Lipschitz continuity) such that

$$\|(z,u) - (z',u')\| < \delta_1$$

where

$$\|(z,u) - (z',u')\| = \|z - z'\|_2 + \|u - u'\|_W$$

implies

$$\left| g_{JK}(z,u) - g_{JK}(z',u') \right| < \epsilon/3.$$

Similarly, there is a $\delta_2 > 0$ such that

$$\|(z,u) - (z',u')\| < \delta_2$$

implies

$$\left| g(z,u) - g(z',u') \right| < \epsilon/3.$$

Let $\delta = \min\{\delta_1, \delta_2\}$ and let the set

$$\{B_\delta(z,u)\}_{(z,u)\in Z_0\times\mathcal{U}^N(W)}$$

be an open cover of $Z_0\times\mathcal{U}^N(W)$ determined by the open balls of radius δ at each (z,u). By compactness, there is a finite subcover

$$\{B_\delta(z_i,u_i)\}_{i=1}^I$$

of $Z_0\times\mathcal{U}^N(W)$ determined by $\{(z_1,u_1),\ldots,(z_I,u_I)\}$. Let

$$(J^*,K^*)\geq\max\{(J^*(z_i,u_i),K^*(z_i,u_i))\}_{i=1}^I,$$

where this maximization is with respect to the usual order on $\mathbb{R}^2$. Suppose that $(J,K)\geq(J^*,K^*)$. For any $(z,u)\in Z_0\times\mathcal{U}^N(W)$, let (z_{i^*},u_{i^*}) be chosen such that $(z,u)\in B_\delta(z_{i^*},u_{i^*})$. It follows that

$$\begin{aligned}
& |g_{JK}(z,u)-g(z,u)| \\
\leq\; & |g_{JK}(z,u)-g_{JK}(z_{i^*},u_{i^*})+g_{JK}(z_{i^*},u_{i^*}) \\
& -g(z_{i^*},u_{i^*})+g(z_{i^*},u_{i^*})-g(z,u)| \\
\leq\; & |g_{JK}(z,u)-g_{JK}(z_{i^*},u_{i^*})|+|g_{JK}(z_{i^*},u_{i^*})-g(z_{i^*},u_{i^*})| \\
& +|g(z_{i^*},u_{i^*})-g(z,u)| \\
<\; & \epsilon/3+\epsilon/3+\epsilon/3=\epsilon.
\end{aligned}$$

In the preceding inequality, we use that $|g_{JK}(z,u)-g_{JK}(z_{i^*},u_{i^*})|<\epsilon/3$ and $|g(z_{i^*},u_{i^*})-g(z,u)|<\epsilon/3$ by Lipschitz continuity, and $|g_{JK}(z_{i^*},u_{i^*})-g(z_{i^*},u_{i^*})|<\epsilon/3$ by choice of (J^*,K^*). □

Introduce the function

$$\psi_{JK}(z)\triangleq\inf\left\{g_{JK}(z,u)\;:\;u\in\mathcal{U}^N(W)\right\}$$

and note the equivalence

$$Z^\epsilon_{JK}\equiv\{z\in Z_0\,:\,\psi_{JK}(z)\geq\epsilon\}.$$

Lemma 4.2.2. *$\psi_{JK}(\cdot)$ is Lipschitz continuous on Z_0 for $\mathcal{P}-$almost all $\boldsymbol{\omega}\in\mathbb{R}^{\mathbb{N}}$.*

Proof. For $z_1,z_2\in Z_0$, we have

$$\begin{aligned}
& |\psi_{JK}(z_1)-\psi_{JK}(z_2)| \\
\leq\; & \sup_{u\in\mathcal{U}^N(W)}\left|\frac{1}{J}\sum_{j=1}^J\left[u\left([\underline{G}(z_1)](\omega_{1j})\right)-u\left([\underline{G}(z_2)](\omega_{1j})\right)\right]\right| \\
\leq\; & \frac{1}{J}\sum_{j=1}^J\Pi(\omega_{1j})\,\|\,[\underline{G}(z_1)](\omega_{1j})-[\underline{G}(z_2)](\omega_{1j})\,\|_2 \\
\leq\; & \pi_J\,\|z_1-z_2\|_2.
\end{aligned}$$

□

Next we verify uniform convergence.

Lemma 4.2.3. *$\psi_{JK}(\cdot)$ converges to $\psi(\cdot)$ uniformly on Z_0 for $\mathcal{P}-$almost all $\boldsymbol{\omega} \in \mathbb{R}^{\mathbb{N}}$.*

Proof. Compute

$$\begin{aligned} & |\psi_{JK}(z) - \psi(z)| \\ = \quad & |\inf\left\{g_{JK}(z,u) \,:\, u \in \mathcal{U}^N(W)\right\} \\ & - \inf\left\{g(z,u) \,:\, u \in \mathcal{U}^N(W)\right\}| \\ \leq \quad & \sup\left\{|g_{JK}(z,u) - g(z,u)| \,:\, u \in \mathcal{U}^N(W)\right\} \to 0, \end{aligned}$$

as $J, K \to \infty$. □

The main convergence result is next.

Theorem 4.2.4. *Suppose $Z^\epsilon \equiv \{z \in Z_0 \,:\, \psi(z) \geq \epsilon\}$ satisfies the Slater condition.*
(a) $\mathbb{H}(Z^\epsilon_{JK}, Z^\epsilon) \to 0$ as $J, K \to \infty$ for $\mathcal{P}-$almost all $\boldsymbol{\omega} \in \mathbb{R}^{\mathbb{N}}$.
(b) $\mathbb{D}(S^\epsilon_{JK}, S^\epsilon) \to 0$ as $J, K \to \infty$ for $\mathcal{P}-$almost all $\boldsymbol{\omega} \in \mathbb{R}^{\mathbb{N}}$.
(c) $\nu^\epsilon_{JK} \to \nu^\epsilon$ as $J, K \to \infty$ for $\mathcal{P}-$almost all $\boldsymbol{\omega} \in \mathbb{R}^{\mathbb{N}}$.

Proof. (a) $\psi_{JK}(\cdot)$ converges to $\psi(\cdot)$ uniformly on Z_0. Thus, by [**19**, Proposition 7.15] $\psi_{JK}(\cdot)$ epi-converges and hypo-converges to $\psi(\cdot)$ on Z_0. By [**26**, Theorem 3.1], the set Z^ϵ_{JK} upper semi-converges to Z^ϵ for $\mathcal{P}-$almost all $\boldsymbol{\omega} \in \mathbb{R}^{\mathbb{N}}$. It follows that $\mathbb{D}(Z^\epsilon_{JK}, Z^\epsilon) \to 0$. Since Z^ϵ satisfies the Slater condition, it follows from [**26**, Theorem 3.5] that $\mathbb{D}(Z^\epsilon, Z^\epsilon_{JK}) \to 0$ as $J, K \to \infty$ for $\mathcal{P}-$almost all $\boldsymbol{\omega} \in \mathbb{R}^{\mathbb{N}}$.

(b) By the preceding reasoning, Z^ϵ_{JK} semi-converges to Z^ϵ for $\mathcal{P}-$almost all $\boldsymbol{\omega} \in \mathbb{R}^{\mathbb{N}}$. By [**26**, Theorem 4.1], $\mathbb{D}(S^\epsilon_{JK}, S^\epsilon) \to 0$ as $J, K \to \infty$ for $\mathcal{P}-$almost all $\boldsymbol{\omega} \in \mathbb{R}^{\mathbb{N}}$. If S^ϵ is empty, then S^ϵ_{JK} is empty for all sufficiently large J and K.

(c) Follows from part (b). □

4.2.2. Convergence analysis for problem ($\epsilon-$SIP). Using the fact that U_Ξ is a compact subset of $\mathcal{U}^N(W)$, we obtain the following corollary of the convergence result in the previous subsection.

Corollary 4.2.5. *For $\mathcal{P}-$almost all $\boldsymbol{\omega} \in \mathbb{R}^{\mathbb{N}}$,*

$$\sup\{|g(z,u) - g_{JK}(z,u)| \,:\, (z,u) \in Z_0 \times U_\Xi\} \to 0$$

as $J, K \to \infty$.

Define

$$\psi_{JK}(z; U_\Xi) \triangleq \inf_{u \in U_\Xi} \{g_{JK}(z; u)\}$$

and notice

$$Z^\epsilon_{JK}(U_\Xi) \equiv \{z \in Z_0 \,:\, \psi_{JK}(z; U_\Xi) \geq \epsilon\}.$$

Lemma 4.2.6. *$\psi_{JK}(\cdot; U_\Xi)$ is Lipschitz continuous on Z_0 for $\mathcal{P}-$almost all $\boldsymbol{\omega} \in \mathbb{R}^{\mathbb{N}}$.*

Uniform convergence of $\psi_{JK}(\cdot; U_\Xi)$ to $\psi(\cdot; U_\Xi)$ is established next.

Lemma 4.2.7. *$\psi_{JK}(\cdot; U_\Xi)$ converges to $\psi(\cdot; U_\Xi)$ uniformly on Z_0 for $\mathcal{P}-$almost all $\boldsymbol{\omega} \in \mathbb{R}^{\mathbb{N}}$.*

The main convergence result for problem ($\epsilon-$SIP) now follows.

Corollary 4.2.8. *Suppose problem ($\epsilon -$ SIP) satisfies the Slater condition.*
(a) $\mathbb{H}\left(Z^{\epsilon}_{JK}\left(U_{\Xi}\right), Z^{\epsilon}\left(U_{\Xi}\right)\right) \to 0$ *as* $J, K \to \infty$ *for* $\mathcal{P}-$*almost all* $\boldsymbol{\omega} \in \mathbb{R}^{\mathbb{N}}$.
(b) $\mathbb{D}\left(S^{\epsilon}_{JK}\left(U_{\Xi}\right), S^{\epsilon}\left(U_{\Xi}\right)\right) \to 0$ *as* $J, K \to \infty$ *for* $\mathcal{P}-$*almost all* $\boldsymbol{\omega} \in \mathbb{R}^{\mathbb{N}}$.
(c) $\nu^{\epsilon}_{JK}\left(U_{\Xi}\right) \to \nu^{\epsilon}\left(U_{\Xi}\right)$ *as* $J, K \to \infty$ *for* $\mathcal{P}-$*almost all* $\boldsymbol{\omega} \in \mathbb{R}^{\mathbb{N}}$.

We know that there exists a sequence $\{U_{\Xi_k}\}_{k=0}^{\infty}$ with $U_{\Xi_k} \subset U_{\Xi_{k+1}}$ and

$$\bigcup_{k=0}^{\infty} U_{\Xi_k} = \mathcal{U}^N\left(W\right).$$

Since this sequence is countable, we obtain the following corollary.

Corollary 4.2.9. *Suppose problem ($\epsilon-$SIP) over* U_{Ξ_k} *satisfies the Slater condition for all* $k \geq 0$. *Then* $\mathbb{H}\left(Z_{JK}\left(U_{\Xi_k}\right), Z_{JK}\left(U_{\Xi_k}\right)\right) \to 0$, $\mathbb{D}\left(S_{JK}\left(U_{\Xi_k}\right), S\left(U_{\Xi_k}\right)\right) \to 0$, *and* $\nu^{\epsilon}_{JK}\left(U_{\Xi_k}\right) \to \nu^{\epsilon}\left(U_{\Xi_k}\right)$ *as* $J, K \to \infty$ *for* $\mathcal{P}-$*almost all* $\boldsymbol{\omega} \in \mathbb{R}^{\mathbb{N}}$.

Proof. For each fixed k, this conclusion does not hold on a set of $\mathcal{P}-$measure zero. The countable union of sets with measure zero has measure zero. □

4.3. Upper bounds

In this section we adapt the methodology in [**16**, Section 6] to produce upper bounds for problem ($\epsilon-$SMP) and problem ($\epsilon-$SSIP). It is possible to use duality to construct an upper bound for ν and $\nu\left(U_{\Xi}\right)$. The Lagrange multiplier of constraint (1) is $u \in \mathrm{cl\,cone}\,\mathcal{U}^N\left(W\right)$. By weak duality,

$$\nu \leq \max_{z \in Z_0} L\left(z, u\right) - \epsilon \left\|\mathrm{PX}\left(u; \mathrm{cl\,conv}\,\mathcal{U}^N\left(W\right)\right)\right\|.$$

Analogously, the Lagrange multiplier of constraint (5) is $u \in \mathrm{cl\,cone}\,U_{\Xi}$. By weak duality,

$$\nu\left(U_{\Xi}\right) \leq \max_{z \in Z_0} L\left(z, u\right) - \epsilon \left\|\mathrm{PX}\left(u; \mathrm{cl\,conv}\,U_{\Xi}\right)\right\|.$$

Let

$$L_{JK}\left(z, u\right) \triangleq f\left(z\right) + \frac{1}{J}\sum_{j=1}^{J} u\left(\underline{\mathrm{G}}\left(z, \omega_{1j}\right)\right) - \frac{1}{K}\sum_{k=1}^{K} u\left(\underline{\mathrm{Y}}\left(\omega_{2k}\right)\right)$$

be the sample average approximation of $L\left(z, u\right)$. At any time, our best guess for an upper bound for ν is

$$\min\left\{\max_{z \in Z_0} L_{JK}\left(z, u\right) - \epsilon \|\mathrm{PX}\left(u; \mathrm{cl\,conv}\,\mathcal{U}^N\left(W\right)\right) \,:\, u \in \mathrm{cl\,cone}\,\mathcal{U}^N\left(W\right)\right\}$$

and for $\nu\left(U_{\Xi}\right)$ is

$$\min\left\{\max_{z \in Z_0} L_{JK}\left(z, u\right) - \epsilon \left\|\mathrm{PX}\left(u; \mathrm{cl\,conv}\,U_{\Xi}\right)\right\| \,:\, u \in \mathrm{cl\,cone}\,U_{\Xi}\right\}.$$

4.4. Aggregation, column generation, and row generation

Problem ($\epsilon-$SCP) is difficult to solve directly for large J and K. In this section we discuss a large scale implementation strategy for this problem that combines variable and constraint aggregation with column and row generation.

The Lagrangian of problem ($\epsilon-$SCP) is

$$\begin{aligned} L(z,\pi,\lambda,\delta,\gamma) \quad = \quad & f(z) + \sum_{j=1}^{J} \langle \lambda_j, \frac{1}{J}\left[\underline{G}(z)\right](\omega_{1j}) - \sum_{k=1}^{K} \pi_{jk}\left(\underline{Y}(\omega_{2k}) + \epsilon\, \underline{1}\right)\rangle \\ & + \sum_{j=1}^{J} \delta_j \left(\sum_{k=1}^{K} \pi_{jk} - \frac{1}{J}\right) + \sum_{k=1}^{K} \gamma_k \left(\sum_{j=1}^{J} \pi_{jk} - \frac{1}{K}\right), \end{aligned}$$

and problem ($\epsilon-$SCP) can be written as

$$\max\left\{\min\left\{L(z,\pi,\lambda,\delta,\gamma)\ :\ \lambda \geq 0\right\}\ :\ z \in Z_0,\ \pi \geq 0\right\}.$$

The dual of problem ($\epsilon-$SCP) is then

$$\min\left\{\max\left\{L(z,\pi,\lambda,\delta,\gamma)\ :\ z \in Z_0,\ \pi \geq 0\right\}\ :\ \lambda \geq 0\right\}.$$

For the dual functional

$$\mathrm{d}(\lambda) = \max\left\{ f(z) + \sum_{j=1}^{J} \langle \lambda_j, \frac{1}{J}\left[\underline{G}(z)\right](\omega_{1j})\rangle\ :\ z \in Z_0 \right\},$$

we explicitly obtain the dual of problem ($\epsilon-$SCP):

$$\begin{aligned} (\epsilon-\mathrm{SCP}_D) \qquad \min \quad & \mathrm{d}(\lambda) - \frac{1}{J}\sum_{j=1}^{J} \delta_j - \frac{1}{K}\sum_{k=1}^{K} \gamma_k \\ \text{s.t.} \quad & -\langle \lambda_j, \underline{Y}(\omega_{2k}) + \epsilon\, \underline{1}\rangle + \delta_j + \gamma_k \leq 0, \qquad \forall j \in \mathcal{J},\ \forall k \in \mathcal{K}. \end{aligned}$$

4.4.1. Aggregation. We are really only interested in the decision variables z in problem ($\epsilon-$SCP), rather than (z,π). With this understanding, we would be willing to accept a possibly sub-optimal but feasible solution to problem ($\epsilon-$SCP) in return for a gain in computational tractability. Problem ($\epsilon-$SCP) can be made more tractable via aggregation.

There is a natural aggregation scheme for problem ($\epsilon-$SCP). Choose a partition of $\mathcal{J}$, denoted $\{J_{i_1}\}_{i_1=1}^{I_1}$, where $\bigcup_{i_1=1}^{I_1} J_{i_1} = \mathcal{J}$ and $J_{i_1} \bigcap J_{i_2} = \emptyset$ for $i_1 \neq i_2$. We can aggregate all of the scenarios in each set J_i into a single scenario with probability mass $\sum_{j\in J_{i_1}} 1/J = |J_{i_1}|/J$. Similarly, choose a partition of $\mathcal{K}$, denoted $\{K_{i_2}\}_{i_2=1}^{I_2}$, where $\bigcup_{i_2=1}^{I_2} K_i = \mathcal{K}$ and $K_{i_1} \bigcap K_{i_2} = \emptyset$ for $i_1 \neq i_2$. We can aggregate all of the scenarios in each set K_i into a single scenario with probability mass $\sum_{k\in K_{i_2}} 1/K = |K_{i_2}|/K$. Let

$$\hat{\pi}(J_{i_1}, K_{i_2})\ :\ \{J_{i_1}\}_{i_1=1}^{I_1} \times \{K_{i_2}\}_{i_2=1}^{I_2} \rightarrow \mathbb{R}_+$$

be the joint probability measure on the cross product of the modified sample spaces $\{J_{i_1}\}_{i_1=1}^{I_1}$ and $\{K_{i_2}\}_{i_2=1}^{I_2}$. We recover the original joint probability measure

$\{\pi_{jk}\}_{j\in\mathcal{J},\,k\in\mathcal{K}}$ from the aggregated measure $\hat{\pi}\left(J_{i_1},K_{i_2}\right)$ via the following convention. Set

$$\pi_{jk}=\frac{\hat{\pi}\left(J_{i_1},K_{i_2}\right)}{\left|J_{i_1}\right|\left|K_{i_2}\right|}$$

for $J_{i_1}\ni j$ and $K_{i_2}\ni k$.

We define (abusing notation by using J_{i_1} and K_{i_2} to indicate scenarios)

$$\left[\underline{\mathrm{G}}\left(z\right)\right]\left(J_{i_1}\right)\triangleq\frac{1}{\left|J_{i_1}\right|}\sum_{j\in J_{i_1}}\left[\underline{\mathrm{G}}\left(z\right)\right]\left(\omega_{1j}\right)$$

and

$$\underline{\mathrm{Y}}\left(K_{i_2}\right)\triangleq\frac{1}{\left|K_{i_2}\right|}\sum_{k\in K_{i_2}}\underline{\mathrm{Y}}\left(\omega_{2k}\right).$$

In this new setup, the constraints (2) are recast as

$$\frac{\left|J_{i_1}\right|}{\left|J\right|}\left[\underline{\mathrm{G}}\left(z\right)\right]\left(J_{i_1}\right)\geq\sum_{i_2=1}^{I_2}\hat{\pi}\left(J_{i_1},K_{i_2}\right)\left(\underline{\mathrm{Y}}\left(K_{i_2}\right)+\epsilon\,\underline{1}\right),\qquad i_1=1,\ldots,I_1,$$

or

$$\frac{1}{J}\left[\sum_{j\in J_{i_1}}\left[\underline{\mathrm{G}}\left(z\right)\right]\left(\omega_{1j}\right)\right]\geq\sum_{i_2=1}^{I_2}\hat{\pi}\left(J_{i_1},K_{i_2}\right)\left[\frac{1}{\left|K_{i_2}\right|}\sum_{k\in K_{i_2}}\left(\underline{\mathrm{Y}}\left(\omega_{2k}\right)+\epsilon\,\underline{1}\right)\right],\quad i_1=1,\ldots,I_1,$$

since

$$\frac{\left|J_{i_1}\right|}{J}\left[\underline{\mathrm{G}}\left(z\right)\right]\left(J_{i_1}\right)=\frac{1}{J}\left[\sum_{j\in J_{i_1}}\left[\underline{\mathrm{G}}\left(z\right)\right]\left(\omega_{1j}\right)\right].$$

The constraints (3) become

$$\sum_{i_2=1}^{I_2}\hat{\pi}\left(J_{i_1},K_{i_2}\right)=\frac{\left|J_{i_1}\right|}{J},\qquad i_1=1,\ldots,I_1.$$

Finally, the constraints (4) become

$$\sum_{i_1=1}^{I_1}\hat{\pi}\left(J_{i_1},K_{i_2}\right)=\frac{\left|K_{i_2}\right|}{K},\qquad i_2=1,\ldots,I_2.$$

The aggregate form of any instance of problem ($\epsilon-$SCP) looks exactly like the general form of problem ($\epsilon-$SCP), with appropriate modifications based on this discussion. We obtain the problem

$$(\epsilon\text{-AggSCP}) \quad \max_{z\in Z_0,\,\hat{\pi}\geq 0} \quad f(z)$$

$$\text{(7)} \qquad \text{s.t.} \quad \frac{|J_{i_1}|}{|J|}\left[\underline{\mathrm{G}}(z)\right](J_{i_1}) \geq \sum_{i_2=1}^{I_2} \hat{\pi}(J_{i_1}, K_{i_2})\left(\underline{\mathrm{Y}}(K_{i_2}) + \epsilon\,\underline{1}\right),$$

$$i_1 = 1, \ldots, I_1,$$

$$\text{(8)} \qquad \sum_{i_2=1}^{I_2} \hat{\pi}(J_{i_1}, K_{i_2}) = \frac{|J_{i_1}|}{J},$$

$$i_1 = 1, \ldots, I_1,$$

$$\text{(9)} \qquad \sum_{i_1=1}^{I_1} \hat{\pi}(J_{i_1}, K_{i_2}) = \frac{|K_{i_2}|}{K},$$

$$i_2 = 1, \ldots, I_2,$$

and we denote its feasible region, set of optimal solutions, and optimal value as:

$$\begin{aligned}
& Z^{\epsilon}_{JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1}, \{K_{i_2}\}_{i_2=1}^{I_2}\right) \\
\triangleq\ & \{z \in Z_0,\, \hat{\pi} \geq 0 : \ (7) - (9)\}, \\
& S^{\epsilon}_{JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1}, \{K_{i_2}\}_{i_2=1}^{I_2}\right) \\
\triangleq\ & \arg\max\left\{f(z) \ : \ (z,\hat{\pi}) \in Z^{\epsilon}_{JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1}, \{K_{i_2}\}_{i_2=1}^{I_2}\right)\right\}, \\
& \nu^{\epsilon}_{JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1}, \{K_{i_2}\}_{i_2=1}^{I_2}\right) \\
\triangleq\ & \max\left\{f(z) \ : \ (z,\hat{\pi}) \in Z^{\epsilon}_{JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1}, \{K_{i_2}\}_{i_2=1}^{I_2}\right)\right\}.
\end{aligned}$$

4.4.2. Row generation. We can use the transformation

$$\pi_{jk} = \frac{\hat{\pi}(J_{i_1}, K_{i_2})}{|J_{i_1}|\,|K_{i_2}|}$$

for $J_{i_1} \ni j$ and $K_{i_2} \ni k$ to generate the full set of constraints of problem ($\epsilon-$SCP). The constraints (2) expand to become:

$$\frac{1}{J}\underline{\mathrm{G}}(z, \omega_{1j}) \geq \sum_{i_2=1}^{I_2}\sum_{k\in K_{i_2}}\left(\frac{\hat{\pi}(J_{i_1}, K_{i_2})}{|J_{i_1}|\,|K_{i_2}|}\right)\left(\underline{\mathrm{Y}}(\omega_{2k}) + \epsilon\,\underline{1}\right), \quad \forall j \in \mathcal{J},$$

for $J_{i_1} \ni j$. Equivalently, we have

$$\frac{|J_{i_1}|}{J}\underline{\mathrm{G}}(z, \omega_{1j}) \geq \sum_{i_2=1}^{I_2} \hat{\pi}(J_{i_1}, K_{i_2})\left(\sum_{k\in K_{i_2}} \frac{\underline{\mathrm{Y}}(\omega_{2k})}{|K_{i_2}|} + \epsilon\,\underline{1}\right), \quad \forall j \in \mathcal{J}.$$

The equality in distribution constraints $(3)-(4)$ are automatically satisfied in this situation. Suppose that $\hat{\pi}$ satisfies $(8)-(9)$, then

$$\begin{aligned}\sum_{k=1}^{K} \pi_{jk} &= \sum_{i_2=1}^{I_2} \sum_{k \in K_{i_2}} \pi_{jk} \\ &= \sum_{i_2=1}^{I_2} \sum_{k \in K_{i_2}} \frac{\hat{\pi}\left(J_{i_1}, K_{i_2}\right)}{\left|J_{i_1}\right| \left|K_{i_2}\right|} \\ &= \frac{1}{\left|J_{i_1}\right|} \sum_{i_2=1}^{I_2} \hat{\pi}\left(J_{i_1}, K_{i_2}\right) \\ &= \frac{1}{\left|J_{i_1}\right|} \left(\frac{\left|J_{i_1}\right|}{J}\right) = 1/J,\end{aligned}$$

for all $j \in \mathcal{J}$. A similar calculation establishes that

$$\begin{aligned}\sum_{j=1}^{J} \pi_{jk} &= \sum_{i_1=1}^{I_1} \sum_{j \in J_{i_1}} \pi_{jk} \\ &= \frac{1}{\left|K_{i_2}\right|} \sum_{i_1=1}^{I_1} \hat{\pi}\left(J_{i_1}, K_{i_2}\right) \\ &= \frac{1}{\left|K_{i_2}\right|} \left(\frac{\left|K_{i_2}\right|}{K}\right) = 1/K,\end{aligned}$$

for all $k \in \mathcal{K}$.

We now obtain the problem

(ϵ-RowSCP)

$$\max_{z \in Z_0,\, \hat{\pi} \geq 0} \quad f\left(z\right)$$

(10) s.t. $$\frac{\left|J_{i_1}\right|}{J} \underline{\mathrm{G}}\left(z, \omega_{1j}\right) \geq \sum_{i_2=1}^{I_2} \frac{\hat{\pi}\left(J_{i_1}, K_{i_2}\right)}{\left|J_{i_1}\right|} \left(\frac{1}{\left|K_{i_2}\right|} \sum_{k \in K_{i_2}} \underline{\mathrm{Y}}\left(\omega_{2k}\right) + \epsilon\, \underline{1}\right),$$

$\forall j \in \mathcal{J}$,

(11) $$\sum_{i_2=1}^{I_2} \hat{\pi}\left(J_{i_1}, K_{i_2}\right) = \frac{\left|J_{i_1}\right|}{J},$$

$i_1 = 1, \ldots, I_1$,

(12) $$\sum_{i_1=1}^{I_1} \hat{\pi}\left(J_{i_1}, K_{i_2}\right) = \frac{\left|K_{i_2}\right|}{K},$$

$i_2 = 1, \ldots, I_2$,

and we denote its feasible region, set of optimal solutions, and optimal value as:

$$
\begin{aligned}
& Z^{\epsilon}_{Row,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1}, \{K_{i_2}\}_{i_2=1}^{I_2}\right) \\
\triangleq \quad & \{z \in Z_0,\, \hat{\pi} \geq 0 :\ (10) - (12)\}, \\
& S^{\epsilon}_{Row,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1}, \{K_{i_2}\}_{i_2=1}^{I_2}\right) \\
\triangleq \quad & \arg\max_{(z,\hat{\pi})}\left\{f(z)\ :\ (z,\hat{\pi}) \in Z^{\epsilon}_{Row,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1}, \{K_{i_2}\}_{i_2=1}^{I_2}\right)\right\}, \\
\triangleq \quad & \nu^{\epsilon}_{Row,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1}, \{K_{i_2}\}_{i_2=1}^{I_2}\right) \\
& \max\left\{f(z)\ :\ (z,\hat{\pi}) \in Z^{\epsilon}_{Row,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1}, \{K_{i_2}\}_{i_2=1}^{I_2}\right)\right\}.
\end{aligned}
$$

Problem ($\epsilon-$RowSCP) has many constraints and few variables, and it can be solved via row (cut) generation techniques.

Problem ($\epsilon-$RowSCP) is clearly a restricted version of problem ($\epsilon-$SCP). Thus

$$Z^{\epsilon}_{Row,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1}, \{K_{i_2}\}_{i_2=1}^{I_2}\right) \subset Z^{\epsilon}_{CP,JK},$$

when measures $\hat{\pi}$ on the partitioned space in problem ($\epsilon-$RowSCP) are identified with measures π on $\Omega_1 \times \Omega_2$. Necessarily

$$\nu^{\epsilon}_{Row,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1}, \{K_{i_2}\}_{i_2=1}^{I_2}\right) \leq \nu^{\epsilon}_{CP,JK}.$$

Problem ($\epsilon-$RowSCP) is also clearly a restricted version of problem ($\epsilon-$AggSCP) because it is a disaggregation of the constraints in problem ($\epsilon-$AggSCP). Then

$$Z^{\epsilon}_{Row,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1}, \{K_{i_2}\}_{i_2=1}^{I_2}\right) \subset Z^{\epsilon}_{Agg,JK},$$

when measures $\hat{\pi}$ on the partitioned space in problem ($\epsilon-$RowSCP) are identified with measures π on $\Omega_1 \times \Omega_2$. Necessarily

$$\nu^{\epsilon}_{Row,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1}, \{K_{i_2}\}_{i_2=1}^{I_2}\right) \leq \nu^{\epsilon}_{Agg,JK}.$$

4.4.3. Column generation. We can also use the partitions $\{J_{i_1}\}_{i_1=1}^{I_1}$ and $\{K_{i_2}\}_{i_2=1}^{I_2}$ to generate the full set of variables for problem ($\epsilon-$SCP). We will take the average of the constraints over each partition to keep the problem size manageable. The coupling constraints (2) expand to become:

$$\frac{|J_{i_1}|}{J}\underline{\mathrm{G}}(z, J_{i_1}) \geq \sum_{j \in J_{i_1}} \sum_{k=1}^{K} \pi_{jk}\left(\underline{\mathrm{Y}}(\omega_{2k}) + \epsilon\, \underline{1}\right), \quad i_1 = 1, \ldots, I_1.$$

The equality in distribution constraints $(3) - (4)$ expand to become

$$\sum_{j \in J_{i_1}} \sum_{k=1}^{K} \pi_{jk} = \frac{|J_{i_1}|}{J}, \quad i_1 = 1, \ldots, I_1,$$

and

$$\sum_{k\in K_{i_2}}\sum_{j=1}^{J}\pi_{jk}=\frac{|K_{i_2}|}{K},\quad i_2=1,\ldots,I_2.$$

We obtain the optimization problem

(ϵ-ColSCP)

$$\max_{z\in Z_0,\,\hat{\pi}\geq 0} \quad f(z)$$

$$\text{(13)}\quad \text{s.t.}\quad \frac{|J_{i_1}|}{J}\left[\underline{\mathrm{G}}(z)\right](J_{i_1})\geq\sum_{k=1}^{K}\sum_{j\in J_{i_1}}\pi_{jk}\left(\underline{\mathrm{Y}}(\omega_{2k})+\epsilon\,\underline{1}\right),\quad i_1=1,\ldots,I_1,$$

$$\text{(14)}\quad \sum_{j\in J_{i_1}}\sum_{k=1}^{K}\pi_{jk}=\frac{|J_{i_1}|}{J},\quad i_1=1,\ldots,I_1,$$

$$\text{(15)}\quad \sum_{k\in K_{i_2}}\sum_{j=1}^{J}\pi_{jk}=\frac{|K_{i_2}|}{K},\quad i_2=1,\ldots,I_2,$$

and we denote its feasible region, set of optimal solutions, and optimal value as:

$$\begin{aligned}
&Z^{\epsilon}_{Col,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1},\{K_{i_2}\}_{i_2=1}^{I_2}\right)\\
\triangleq\ &\{z\in Z_0,\,\pi\geq 0:\ (13)-(15)\},\\
&S^{\epsilon}_{Col,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1},\{K_{i_2}\}_{i_2=1}^{I_2}\right)\\
\triangleq\ &\arg\max_{(z,\hat{\pi})}\left\{f(z)\ :\ (z,\pi)\in Z^{\epsilon}_{Col,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1},\{K_{i_2}\}_{i_2=1}^{I_2}\right)\right\},\\
&\nu^{\epsilon}_{Col,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1},\{K_{i_2}\}_{i_2=1}^{I_2}\right)\\
\triangleq\ &\max\left\{f(z)\ :\ (z,\pi)\in Z^{\epsilon}_{Col,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1},\{K_{i_2}\}_{i_2=1}^{I_2}\right)\right\}.
\end{aligned}$$

Problem ($\epsilon-$ColSCP) has few constraints and many variables, and it can be solved via column generation (pricing) techniques. By construction, problem ($\epsilon-$ColSCP) is a relaxation of problem ($\epsilon-$SCP). We see

$$Z^{\epsilon}_{Col,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1},\{K_{i_2}\}_{i_2=1}^{I_2}\right)\supset Z^{\epsilon}_{CP,JK}$$

and

$$\nu^{\epsilon}_{Col,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1},\{K_{i_2}\}_{i_2=1}^{I_2}\right)\geq\nu^{\epsilon}_{CP,JK}.$$

Problem ($\epsilon-$ColSCP) is also relaxation of problem ($\epsilon-$AggSCP)

$$Z^{\epsilon}_{Col,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1},\{K_{i_2}\}_{i_2=1}^{I_2}\right)\supset Z^{\epsilon}_{Agg,JK}$$

and

$$\nu^{\epsilon}_{Col,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1},\{K_{i_2}\}_{i_2=1}^{I_2}\right)\geq\nu^{\epsilon}_{Agg,JK}.$$

CHAPTER 5

Robust optimization with a class of multivariate integral stochastic order constraints

5.1. Robust optimization problem

In this chapter we combine the approach from **[11]** with our present development. In **[11]**, uncertainty about the underlying probability distribution in a stochastic order constrained program is studied for the first time. This paper considers underlying uncertainty for univariate increasing concave stochastic order constraints. In this chapter we seek to make a similar contribution for multivariate increasing concave stochastic order constraints. This chapter is especially important for applications, since there is usually uncertainty about the underlying probability distribution in practice.

We start by described our model of uncertainty for P. We formalize a base probability space $(\Omega, \mathcal{F}, P_0)$ to be used throughout this chapter. The probability measure P_0 is fixed and known to formalize the concept of a probability uncertainty set. We are given a set of probability measures $\mathcal{Q} \subset \mathcal{L}_\infty^1(\Omega, \mathcal{F}, P_0)$. We assume $\mathcal{Q}$ is convex and closed, and

$$B = \sup_{P \in \mathcal{Q}} \| \frac{dP}{dP_0} \|_\infty < \infty.$$

We adopt the following convention. Any measure Q that is absolutely continues with respect to P_0 with Radon-Nikodym derivative dQ/dP_0 in $\mathcal{L}_\infty^1(\Omega, \mathcal{F}, P_0)$ can be considered as element of $\mathcal{L}_\infty^1(\Omega, \mathcal{F}, P_0)$. We propose the following robust version of $\geq_{icv}$. For $\underline{\mathrm{X}}, \underline{\mathrm{Y}} \in \mathcal{L}_1^N(\Omega, \mathcal{F}, P_0)$, if

$$\mathbb{E}_P[u(\underline{\mathrm{X}})] \geq \mathbb{E}_P[u(\underline{\mathrm{Y}})], \quad \forall u \in \mathcal{U}^N(W), \quad \forall P \in \mathcal{Q},$$

then $\underline{\mathrm{X}}$ is larger than $\underline{\mathrm{Y}}$ in the robust increasing concave order with respect to $\mathcal{Q}$.

We will use this order to define a constraint on $\underline{\mathrm{G}}(z)$,

$$\mathbb{E}_P[u(\underline{\mathrm{G}}(z))] \geq \mathbb{E}_P[u(\underline{\mathrm{Y}})], \quad \forall u \in \mathcal{U}^N(W), \quad \forall P \in \mathcal{Q}.$$

We present a robust version of problem (MP):

$$\text{(RobMP)} \qquad \max_{z \in Z_0} \quad f(z)$$

$$\text{(1)} \qquad \text{s.t.} \quad \mathbb{E}_P[u(\underline{\mathrm{G}}(z))] \geq \mathbb{E}_P[u(\underline{\mathrm{Y}})], \quad \forall u \in \mathcal{U}^N(W), \forall P \in \mathcal{Q}.$$

Like problem (MP), problem (RobMP) cannot satisfy the Slater condition because $u = 0$ is included in $\mathcal{U}^N(W)$. We perturb problem (RobMP) to obtain:

$$
\begin{array}{lll}
(\epsilon\mathrm{-RobMP}) & \max_{z \in Z_0} & f(z) \\
(2) & \text{s.t.} & \mathbb{E}_P\left[u\left(\underline{\mathrm{G}}(z)\right)\right] \geq \mathbb{E}_P\left[u\left(\underline{\mathrm{Y}}\right)\right] + \epsilon, \\
 & & \forall u \in \mathcal{U}^N(W),\ \forall P \in \mathcal{Q}.
\end{array}
$$

We define its feasible region, set of solutions, and optimal value:

$$
\begin{aligned}
Z^{\mathcal{Q},\epsilon} &\triangleq \{z \in Z_0 : (2)\}, \\
S^{\mathcal{Q},\epsilon} &\triangleq \arg\max\left\{f(z) : z \in Z^{\mathcal{Q},\epsilon}\right\}, \\
\nu^{\mathcal{Q},\epsilon} &\triangleq \max\left\{f(z) : z \in Z^{\mathcal{Q},\epsilon}\right\}.
\end{aligned}
$$

We will use semi-infinite programming and nonlinear programming to construct relaxations of problem ($\epsilon-$RobMP), as was done for problem ($\epsilon-$MP). These relaxations will be more computationally tractable but can be studied with the same methods. When we use U_Ξ to approximate $\mathcal{U}^N(W)$ we obtain the robust version of problem ($\epsilon-$SIP):

$$
\begin{array}{lll}
(\epsilon\mathrm{-RobSIP}) & \max_{z \in Z_0} & f(z) \\
(3) & \text{s.t.} & \mathbb{E}_P\left[u\left(\underline{\mathrm{G}}(z)\right)\right] \geq \mathbb{E}_P\left[u\left(\underline{\mathrm{Y}}\right)\right] + \epsilon, \\
 & & \forall u \in U_\Xi,\ \forall P \in \mathcal{Q}.
\end{array}
$$

We adopt the following notation for the feasible region, set of optimal solutions, and optimal value of problem ($\epsilon-$RobSIP)

$$
\begin{aligned}
Z^{\mathcal{Q},\epsilon}(U_\Xi) &\triangleq \{z \in Z_0 : (3)\}, \\
S^{\mathcal{Q},\epsilon}(U_\Xi) &\triangleq \arg\max\left\{f(z) : z \in Z^{\mathcal{Q},\epsilon}(U_\Xi)\right\}, \\
\nu^{\mathcal{Q},\epsilon}(U_\Xi) &\triangleq \max\left\{f(z) : z \in Z^{\mathcal{Q},\epsilon}(U_\Xi)\right\}.
\end{aligned}
$$

When $\mathcal{U}^N(W)$ is approximated with a finite set $\{u_1, \ldots, u_I\}$, the robust version of problem ($\epsilon-$NLP) emerges:

$$
\begin{array}{lll}
(\epsilon\mathrm{-RobNLP}) & \max_{z \in Z_0} & f(z) \\
(4) & \text{s.t.} & \mathbb{E}_P\left[u_i\left(\underline{\mathrm{G}}(z)\right)\right] \geq \mathbb{E}_P\left[u_i\left(\underline{\mathrm{Y}}\right)\right] + \epsilon, \\
 & & i = 1, \ldots, I,\ \forall P \in \mathcal{Q}.
\end{array}
$$

Its feasible region, set of solutions, and optimal value are:

$$
\begin{aligned}
Z^{\mathcal{Q},\epsilon}(\{u_1, \ldots, u_I\}) &\triangleq \{z \in Z_0 : (4)\}, \\
S^{\mathcal{Q},\epsilon}(\{u_1, \ldots, u_I\}) &\triangleq \arg\max\left\{f(z) : z \in Z^{\mathcal{Q},\epsilon}(\{u_1, \ldots, u_I\})\right\}, \\
\nu^{\mathcal{Q},\epsilon}(\{u_1, \ldots, u_I\}) &\triangleq \max\left\{f(z) : z \in Z^{\mathcal{Q},\epsilon}(\{u_1, \ldots, u_I\})\right\}.
\end{aligned}
$$

Now define the support functional $\sigma : \mathcal{L}_1^1(\Omega, \mathcal{F}, P_0) \to \mathbb{R}$ of the set $\mathcal{Q}$

$$
\sigma(V) = \inf_{P \in \mathcal{Q}} \mathbb{E}_P[V].
$$

Proposition 5.1.1. [**11**, Proposition 2] *Suppose $\mathcal{Q}$ is convex, closed, and bounded.*
(a) $\sigma(\cdot)$ is convex.
(b) $\sigma(\cdot)$ is Lipschitz continuous on $\mathcal{L}_1^1(\Omega, \mathcal{F}, P_0)$ with modulus B.

For a fixed benchmark $\underline{\mathrm{Y}}$, we define the functional

$$\rho_u(\underline{\mathrm{X}}) = \sigma[u(\underline{\mathrm{X}}) - u(\underline{\mathrm{Y}})]$$

on $\mathcal{L}_1^N(\Omega, \mathcal{F}, P_0)$. The functional $\rho_u(\underline{\mathrm{X}})$ aids the study of the robust constraints in this chapter.

Notice $u(\underline{\mathrm{X}}) - u(\underline{\mathrm{Y}})$ is a univariate random variable. Under our assumption that $u \in \mathcal{U}^N(W)$ and $\underline{\mathrm{X}}$, $\underline{\mathrm{Y}} \in \mathcal{L}_1^N(\Omega, \mathcal{F}, P_0)$, we have $u(\underline{\mathrm{X}}) - u(\underline{\mathrm{Y}}) \in \mathcal{L}_1^1(\Omega, \mathcal{F}, P_0)$. The following proposition is the multi-dimensional analog of [**11**, Proposition 3].

Proposition 5.1.2. *(a) For every $u \in \mathcal{U}^N(W)$ the functional $\rho_u(\cdot)$ is concave.*
(b) $\rho_u(\cdot)$ is increasing.
(c) $\rho_u(\cdot)$ is Lipschitz continuous with modulus B.

Proof. (a) The functional $V \to \inf_{P \in \mathcal{Q}} \mathbb{E}_P(V)$ is concave. Further, each $\mathbb{E}_P(V)$ is increasing so $V \to \inf_{P \in \mathcal{Q}} \mathbb{E}_P(V)$ is increasing.

The function $\underline{\mathrm{X}} \to u(\underline{\mathrm{X}})$ is concave. Since $\rho_u(\cdot)$ is the composition of an increasing concave function with a concave function, it is concave.

(b) $V \to \inf_{P \in \mathcal{Q}} \mathbb{E}_P(V)$ and $\underline{\mathrm{X}} \to u(\underline{\mathrm{X}})$ are both increasing, thus $\rho_u(\cdot)$ is increasing as the composition of increasing functions.

(c) $\sigma(\cdot)$ is Lipschitz continuous with modulus B. Any mapping $\underline{\mathrm{X}} \to u(\underline{\mathrm{X}})$ has Lipschitz constant bounded by 1. Thus, the composition of these mappings is Lipschitz continuous with modulus B. □

Define the functions

$$\begin{aligned} \psi^{\mathcal{Q}}(z) &\triangleq \inf_{u \in \mathcal{U}^N(W)} \{\rho_u(\underline{\mathrm{G}}(z))\}, \\ \psi^{\mathcal{Q}}(z; U_\Xi) &\triangleq \inf_{u \in U_\Xi} \{\rho_u(\underline{\mathrm{G}}(z))\}. \end{aligned}$$

Proposition 5.1.3. *(a) $\psi^{\mathcal{Q}}(z)$ is increasing, concave on Z_0, and Lipschitz continuous.*
(b) $\psi^{\mathcal{Q}}(z; U_\Xi)$ is increasing, concave on Z_0, and Lipschitz continuous.

Proof. (a) $\psi^{\mathcal{Q}}(z)$ is increasing since the infimum of increasing functions is an increasing function. $\psi^{\mathcal{Q}}(z)$ is concave as the infimum of concave functions.

For $z_1, z_2 \in Z_0$ we have

$$\begin{aligned} \psi^{\mathcal{Q}}(z_1) &= \inf_{u \in \mathcal{U}^N(W)} \rho_u(\underline{\mathrm{G}}(z_1)) \\ &\leq \inf_{u \in \mathcal{U}^N(W)} \{\rho_u(\underline{\mathrm{G}}(z_2)) + |\rho_u(\underline{\mathrm{G}}(z_1)) - \rho_u(\underline{\mathrm{G}}(z_2))|\} \\ &\leq \psi^{\mathcal{Q}}(z_2) + \sup_{u \in \mathcal{U}^N(W)} |\rho_u(\underline{\mathrm{G}}(z_1)) - \rho_u(\underline{\mathrm{G}}(z_2))|. \end{aligned}$$

Analogously,

$$\begin{aligned}\psi^{\mathcal{Q}}(z_2) &= \inf_{u\in\mathcal{U}^N(W)} \rho_u\left(\underline{\mathrm{G}}(z_2)\right) \\ &\leq \inf_{u\in\mathcal{U}^N(W)} \left\{\rho_u\left(\underline{\mathrm{G}}(z_1)\right) + |\rho_u\left(\underline{\mathrm{G}}(z_2)\right) - \rho_u\left(\underline{\mathrm{G}}(z_1)\right)|\right\} \\ &\leq \psi^{\mathcal{Q}}(z_1) + \sup_{u\in\mathcal{U}^N(W)} |\rho_u\left(\underline{\mathrm{G}}(z_1)\right) - \rho_u\left(\underline{\mathrm{G}}(z_2)\right)|.\end{aligned}$$

Now compute:

$$\begin{aligned}|\psi^{\mathcal{Q}}(z_1) - \psi^{\mathcal{Q}}(z_2)| &\leq \sup_{u\in\mathcal{U}^N(W)} |\rho_u\left(\underline{\mathrm{G}}(z_1)\right) - \rho_u\left(\underline{\mathrm{G}}(z_2)\right)| \\ &\leq B\,\|\underline{\mathrm{G}}(z_1) - \underline{\mathrm{G}}(z_2)\|_2 \\ &\leq B\,\pi\,\|z_1 - z_2\|_2.\end{aligned}$$

(b) Similar to part (a). □

We next present a convergence result about the quality of the approximation.

Proposition 5.1.4. *Suppose* $Z^\epsilon \equiv \left\{z \in Z_0 : \psi^{\mathcal{Q}}(z) \geq \epsilon\right\}$ *satisfies the Slater condition. As* $U_{\Xi_k} \uparrow \mathcal{U}^N(W)$*:*
(a) $\mathbb{H}\left(Z^{\mathcal{Q},\epsilon}(U_{\Xi_k}), Z^{\mathcal{Q},\epsilon}\right) \to 0$ *as* $k \to \infty$.
(b) $\mathbb{D}\left(S^{\mathcal{Q},\epsilon}(U_{\Xi_k}), S^{\mathcal{Q},\epsilon}\right) \to 0$ *as* $k \to \infty$.
(c) $\nu^{\mathcal{Q},\epsilon}(U_{\Xi_k}) \to \nu^{\mathcal{Q},\epsilon}$ *as* $k \to \infty$.

Proof. (a) The functions $\psi^{\mathcal{Q}}(z; U_{\Xi_k})$ converge to $\psi^{\mathcal{Q}}(z)$ uniformly on the compact set Z_0. Thus, $\psi^{\mathcal{Q}}(z; U_{\Xi_k})$ epi-converges and hypo-converges to $\psi^{\mathcal{Q}}(z)$ as well. Then $\mathbb{D}\left(Z^{\mathcal{Q},\epsilon}(U_{\Xi_k}), Z^{\mathcal{Q},\epsilon}\right) \to 0$ as $k \to \infty$. Since $Z^{\mathcal{Q},\epsilon}$ satisfies the Slater condition, $Z^{\mathcal{Q},\epsilon} = \mathrm{cl}\,\mathrm{int}\, Z^{\mathcal{Q},\epsilon}$ and $\mathbb{D}\left(Z^{\mathcal{Q},\epsilon}, Z^{\mathcal{Q},\epsilon}(U_{\Xi_k})\right) \to 0$.
(b) Apply [**26**, Theorem 4.1].
(c) Follows from part (b). □

5.2. Optimality conditions

5.2.1. Optimality conditions for problem ($\epsilon-$RobMP). Introduce the operator $\boldsymbol{\rho} : \mathcal{L}_1^N(\Omega, \mathcal{F}, P_0) \to \mathcal{C}\left(\mathcal{U}^N(W)\right)$, defined as

$$[\boldsymbol{\rho}(\underline{\mathrm{X}})](u) = \rho_u(\underline{\mathrm{X}}).$$

For a multiplier $\Lambda \in \mathcal{M}\left(\mathcal{U}^N(W)\right)$, the Lagrangian for problem ($\epsilon-$RobMP) is

$$L(z, \Lambda) = f(z) + \int_{\mathcal{U}^N(W)} [\boldsymbol{\rho}(\underline{\mathrm{G}}(z))](u)\, d\Lambda(u).$$

The Slater condition is

Assumption 5.2.1. *There exists* $\tilde{z} \in Z_0$ *such that*

$$\rho_u\left(\underline{G}(\tilde{z})\right) > \epsilon, \quad \forall u \in \mathcal{U}^N(W).$$

Standard optimality conditions are presented in the next theorem.

Theorem 5.2.2. *Suppose assumption 5.2.1 is satisfied.*

(a) If $\hat{z}$ is optimal for problem ($\epsilon-$RobMP), then there exists a measure $\hat{\Lambda} \in \mathcal{M}_+\left(\mathcal{U}^N(W)\right)$ such that

$$L\left(\hat{z},\hat{\Lambda}\right) = \max\left\{L\left(z,\hat{\Lambda}\right) \,:\, z \in Z_0\right\}, \tag{5}$$

$$\int_{\mathcal{U}^N(W)} \left[\boldsymbol{\rho}\left(\underline{G}(\hat{z})\right)\right](u)\, d\hat{\Lambda}(u) = \epsilon\,\hat{\Lambda}\left(\mathcal{U}^N(W)\right). \tag{6}$$

(b) There exists a measure $\Lambda \in \mathcal{M}_+\left(\mathcal{U}^N(W)\right)$ satisfying (5) $-$ (6) such that $\Lambda = \sum_{m=1}^{M} \lambda_m \delta_{u_m}$ for $\lambda_m \geq 0$ and $u_m \in \mathcal{U}^N(W)$ for all $m = 1,\ldots,M$.

Proof. (a) The operator $\boldsymbol{\rho}\left(\underline{G}(\cdot)\right) \,:\, Z_0 \to \mathcal{C}\left(\mathcal{U}^N(W)\right)$ is continuous and concave. First notice that for fixed $z \in Z_0$,

$$\inf_{P\in\mathcal{Q}} \mathbb{E}_P\left[u\left(\underline{G}(z)\right) - u\left(\underline{Y}\right)\right]$$

is continuous in $u \in \mathcal{U}^N(W)$. To see this fact, compute:

$$\begin{aligned}
& \Big|\inf_{P\in\mathcal{Q}} \mathbb{E}_P\left[u_1\left(\underline{G}(z)\right) - u_1\left(\underline{Y}\right)\right] \\
& \quad - \inf_{P\in\mathcal{Q}} \mathbb{E}_P\left[u_2\left(\underline{G}(z)\right) - u_2\left(\underline{Y}\right)\right]\Big| \\
\leq \quad & \sup_{P\in\mathcal{Q}} \Big|\mathbb{E}_P\left[u_1\left(\underline{G}(z)\right) - u_1\left(\underline{Y}\right)\right] \\
& \quad - \mathbb{E}_P\left[u_2\left(\underline{G}(z)\right) - u_2\left(\underline{Y}\right)\right]\Big|.
\end{aligned}$$

We have that

$$\left|\mathbb{E}_P\left[u_1\left(\underline{G}(z)\right) - u_1\left(\underline{Y}\right)\right] - \mathbb{E}_P\left[u_2\left(\underline{G}(z)\right) - u_2\left(\underline{Y}\right)\right]\right| \to 0$$

as $\|u_1 - u_2\|_W \to 0$ for any $P \in \mathcal{Q}$. By a compactness argument (using compactness of $\mathcal{Q}$), we obtain that

$$\sup_{P\in\mathcal{Q}} \left|\mathbb{E}_P\left[u_1\left(\underline{G}(z)\right) - u_1\left(\underline{Y}\right)\right] - \mathbb{E}_P\left[u_2\left(\underline{G}(z)\right) - u_2\left(\underline{Y}\right)\right]\right| \to 0$$

as $\|u_1 - u_2\|_W \to 0$.

Now recall that

$$\rho_u\left(\underline{G}(\cdot)\right) \,:\, Z_0 \to \mathbb{R}$$

is continuous in $z \in Z_0$ for each $u \in \mathcal{U}^N(W)$. It follows that

$$\sup\left\{\left|\rho_u\left(\underline{G}(z_1)\right) - \rho_u\left(\underline{G}(z_2)\right)\right| \,:\, u \in \mathcal{U}^N(W)\right\} \to 0$$

as $\|z_1 - z_2\|_2 \to 0$ by a compactness argument (using compactness of $\mathcal{U}^N(W)$).

Problem ($\epsilon-$RobMP) can be rewritten as

$$\begin{aligned}
\max_{z\in Z_0} \quad & f(z) \\
\text{s.t.} \quad & \boldsymbol{\rho}\left(\underline{G}(z)\right) - \epsilon \in \mathcal{C}_+\left(\mathcal{U}^N(W)\right).
\end{aligned}$$

The Lagrangian for this problem is

$$\Upsilon(z,\Lambda) = f(z) + \langle \Lambda, \boldsymbol{\rho}(\underline{\mathrm{G}}(z)) - \epsilon \rangle$$

Under assumption 5.2.1, the Slater condition is satisfied for problem ($\epsilon-$RobMP). The necessary and sufficient conditions for optimality in Banach spaces, [**4**, Theorem 3.4], then give the desired result:

$$\begin{aligned} \Upsilon\left(\hat{z},\hat{\Lambda}\right) &= \max\left\{\Upsilon\left(z,\hat{\Lambda}\right) \,:\, z \in Z_0\right\}, \\ \int_{\mathcal{U}^N(W)} \left[\boldsymbol{\rho}(\underline{\mathrm{G}}(\hat{z})) - \epsilon\right](u)\, d\hat{\Lambda}(u) &= 0. \end{aligned}$$

(b) Apply [**4**, Proposition 5.104]. □

5.2.2. Optimality conditions for problem ($\epsilon-$RobSIP). The operator

$$\boldsymbol{\rho} : \mathcal{L}_1^N(\Omega,\mathcal{F},P_0) \to \mathcal{C}(\Xi)$$

is understood in this subsection as

$$[\boldsymbol{\rho}(\underline{\mathrm{X}})](\xi) = \rho_{u(\cdot;\xi)}(\underline{\mathrm{X}}).$$

For $\Lambda \in \mathcal{M}_+(\Xi)$, we introduce the Lagrangian

$$L(z,\Lambda) = f(z) + \int_\Xi \rho_{u(\cdot;\xi)}(\underline{\mathrm{G}}(z))\, d\Lambda(\xi).$$

The Slater condition for problem ($\epsilon-$RobSIP) is

Assumption 5.2.3. *There exists $\tilde{z} \in Z_0$ such that*

$$\rho_{u(\cdot;\xi)}(\underline{G}(\tilde{z})) > \epsilon, \quad \forall \xi \in \Xi.$$

Optimality conditions for problem ($\epsilon-$RobSIP) are summarized in the next theorem.

Theorem 5.2.4. *Suppose assumption 5.2.3 is satisfied.*

(a) If $\hat{z}$ is optimal for problem ($\epsilon-$RobSIP), then there exists $\hat{\Lambda} \in \mathcal{M}_+(\Xi)$ such that

$$L\left(\hat{z},\hat{\Lambda}\right) = \max\left\{L\left(z,\hat{\Lambda}\right) \,:\, z \in Z_0\right\}, \tag{7}$$

$$\int_\Xi \rho_{u(\cdot;\xi)}(\underline{G}(\hat{z}))\, d\hat{\Lambda}(\xi) = \epsilon\,\hat{\Lambda}(\Xi). \tag{8}$$

(b) There exists a measure $\Lambda \in \mathcal{M}_+(\Xi)$ satisfying $(7)-(8)$ such that $\Lambda = \sum_{m=1}^M \lambda_m \delta_{\xi_m}$ for $\lambda_m \geq 0$ and $\xi_m \in \Xi$ for all $m = 1,\ldots,M$.

5.2.3. Optimality conditions for problem ($\epsilon-$RobNLP). For this subsection, we set the Lagrangian to be

$$L(z,\lambda) = f(z) + \sum_{i=1}^I \lambda_i \rho_{u_i}(\underline{\mathrm{G}}(z)).$$

The Slater condition for this problem is the usual nonlinear programming Slater condition.

Assumption 5.2.5. *There exists* $\tilde{z} \in Z_0$ *such that*

$$\inf_{P \in \mathcal{Q}} \{g(\tilde{z}, u_i)\} > \epsilon, \quad i = 1, \ldots, I.$$

Optimality conditions for problem ($\epsilon-$RobNLP) are next.

Theorem 5.2.6. *Suppose assumption 5.2.5 is satisfied. If* $\hat{z}$ *is optimal for problem* ($\epsilon-$*RobNLP*)*, then there exists* $\hat{\lambda} \in \mathbb{R}^I_+$ *such that*

$$L\left(\hat{z}, \hat{\lambda}\right) = \max\left\{L\left(z, \hat{\lambda}\right) : z \in Z_0\right\}, \tag{9}$$

$$\sum_{i=1}^{I} \hat{\lambda}_i \rho_{u_i}\left(\underline{G}(\hat{z})\right) = \epsilon \sum_{i=1}^{I} \hat{\lambda}_i. \tag{10}$$

The following result is based on weak duality and it streamlines implementation of the robust $\geq_{icv}$ constraints on finite probability spaces.

Proposition 5.2.7. *Suppose that* $\Omega = \{\omega_1, \ldots, \omega_J\}$ *is finite,* $\mathcal{F}$ *makes all atoms measurable, and* $\mathcal{Q}$ *is polyhedral,*

$$\mathcal{Q} = \left\{p \in \mathbb{R}^J_+ : \begin{array}{c} A\,p \geq b, \\ \sum_{j=1}^J p_j = 1 \end{array}\right\},$$

for some $A \in \mathbb{R}^{J_0 \times J}$ *and* $b \in \mathbb{R}^{J_0}$*. Then for* $\underline{X} \in \mathcal{L}^N_1(\Omega, \mathcal{F}, P)$*, the constraints*

$$\inf_{P \in \mathcal{Q}} \mathbb{E}_P\left[u\left(\underline{X}\right) - u\left(\underline{Y}\right)\right] \geq 0, \quad i = 1, \ldots, I$$

are equivalent to

$$-\langle \mu(i), b\rangle - \gamma(i) \geq 0, \quad i = 1, \ldots, I,$$
$$u_i\left(\underline{X}(\omega_j)\right) - u_i\left(\underline{Y}(\omega_j)\right) + \left(\mu(i)^T A\right)_j + \gamma(i) \geq 0, \quad j \in \mathcal{J}, \quad i = 1, \ldots, I.$$

Proof. In this case, $\sigma(X)$ is the optimization problem:

$$\min_{p \in \mathbb{R}^I_+} \quad \sum_{j=1}^{J} p_j X(\omega_j) \tag{11}$$

$$\text{s.t.} \quad A\,p \geq b, \tag{12}$$

$$\sum_{j=1}^{J} p_j = 1. \tag{13}$$

The Lagrangian of problem (11) – (13) is

$$\Upsilon(p, \mu, \gamma) = \sum_{j=1}^{J} p_j X(\omega_j) + \langle \mu, A\,p - b\rangle + \gamma\left(\sum_{j=1}^{J} p_j - 1\right).$$

We can rewrite problem (11) – (13) as

$$\min_{p}\left\{\max_{\mu,\gamma}\left\{\Upsilon\left(p,\mu,\gamma\right)\,:\,\mu\leq0\right\}\,:\,p\geq0\right\}$$

so the dual to problem $(11)-(13)$ is

$$\max_{\mu,\gamma}\left\{\min_{p}\left\{\Upsilon\left(p,\mu,\gamma\right)\,:\,p\geq0\right\}\,:\,\mu\leq0\right\}.$$

Rearrange $\Upsilon\left(p,\mu,\gamma\right)$ to become

$$\sum_{j=1}^{J}p_j\left[X\left(\omega_j\right)+\left(\mu^TA\right)_j+\gamma\right]-\langle\mu,b\rangle-\gamma.$$

The dual of problem $(11)-(13)$ is then

$$\max_{\mu\leq0}\quad-\langle\mu,b\rangle-\gamma \tag{14}$$

$$\text{s.t.}\quad X\left(\omega_j\right)+\left(\mu^TA\right)_j+\gamma\geq0,\qquad\forall j\in\mathcal{J}. \tag{15}$$

By weak duality, we can replace the constraints

$$\inf_{P\in\mathcal{Q}}\mathbb{E}_P\left[u\left(X\right)-u\left(Y\right)\right]\geq0,\quad i=1,\ldots,I,$$

with the linear inequalities

$$-\langle\mu\left(i\right),b\rangle-\gamma\left(i\right)\geq0,\quad i=1,\ldots,I,$$
$$u_i\left(X\left(\omega_j\right)\right)-u_i\left(Y\left(\omega_j\right)\right)+\left(\mu\left(i\right)^TA\right)_j+\gamma\left(i\right)\geq0,\quad\forall j\in\mathcal{J},\qquad i=1,\ldots,I.$$

□

5.3. Duality

5.3.1. Duality for problem ($\epsilon-$RobMP). Introduce the dual functional

$$\mathrm{d}\left(\Lambda\right)=\max\left\{L\left(z,\Lambda\right)\,:\,z\in Z_0\right\}$$

and the corresponding dual problem,

$$(\epsilon-\text{RobMP}_D)\qquad\min\left\{\mathrm{d}\left(\Lambda\right)-\epsilon\left\|\Lambda\right\|\,:\,\Lambda\in\mathcal{M}_+\left(\mathcal{U}^N\left(W\right)\right)\right\}.$$

Strong duality holds between problem ($\epsilon-$RobMP) and ($\epsilon-\text{RobMP}_D$) by the usual arguments.

Theorem 5.3.1. *Suppose assumption 5.2.1 holds.*

(a) If problem ($\epsilon-$RobMP) has an optimal solution, then problem ($\epsilon-\text{RobMP}_D$) has an optimal solution and the optimal values are equal.

(b) If problem ($\epsilon-\text{RobMP}_D$) has an optimal solution $\hat{\Lambda}$ then any $\hat{z}$ satisfying the optimality conditions $(5)-(6)$ with respect to $\hat{\Lambda}$ is optimal to problem ($\epsilon-$RobMP).

(c) Problem ($\epsilon-\text{RobMP}_D$) has an optimal solution with support on a finite set $\{u_1,\ldots,u_M\}\subset\mathcal{U}^N\left(W\right)$.

5.3.2. Duality for problem ($\epsilon-$RobSIP). The dual functional for problem ($\epsilon-$RobSIP) is

$$\mathrm{d}(\Lambda) = \max\{L(z,\Lambda) : z \in Z_0\}$$

and the dual problem is

$$(\epsilon-\mathrm{RobSIP}_D) \qquad \min\{\mathrm{d}(\Lambda) - \epsilon\|\Lambda\| : \Lambda \in \mathcal{M}_+(\Xi)\}.$$

The expected strong duality result is next.

Theorem 5.3.2. *Suppose assumption 5.2.3 holds.*

(a) Suppose problem ($\epsilon-$RobSIP) has an optimal solution. Then problem ($\epsilon-$RobSIP$_D$) has an optimal solution and the optimal values are equal.

(b) If problem ($\epsilon-$RobSIP$_D$) has an optimal solution $\hat{\Lambda}$ then any $\hat{z}$ satisfying the optimality conditions $(7)-(8)$ *with respect to $\hat{\Lambda}$ is optimal to problem ($\epsilon-$RobSIP).*

(c) Problem ($\epsilon-$RobMP$_D$) has an optimal solution with support on a finite set $\{\xi_1, \ldots, \xi_M\} \subset \Xi$.

5.3.3. Duality for problem ($\epsilon-$RobNLP). The dual functional for problem ($\epsilon-$RobNLP) is

$$\mathrm{d}(\lambda) = \max\{L(z,\lambda) : z \in Z_0\}$$

and the dual problem is

$$(\epsilon-\mathrm{RobSIP}_D) \qquad \min\left\{\mathrm{d}(\lambda) - \epsilon\sum_{i=1}^{I}\lambda_i : \lambda \in \mathbb{R}_+^I\right\}.$$

Strong duality is established in the following theorem.

Theorem 5.3.3. *Suppose assumption 5.2.5 holds.*

(a) If problem ($\epsilon-$RobNLP) has an optimal solution, then problem ($\epsilon-$RobNLP$_D$) has an optimal solution and the optimal values are equal.

(b) If problem ($\epsilon-$RobNLP$_D$) has an optimal solution $\hat{\lambda}$ then any $\hat{z}$ satisfying the optimality conditions $(9)-(10)$ *with respect to $\hat{\lambda}$ is optimal to problem ($\epsilon-$RobNLP).*

5.4. Aggregation, column generation, and row generation

We can define a robust version of problem (CP) as follows. Introduce functions $\pi_{jk}(P) \in \mathcal{C}(\mathcal{Q})$. In this section, $\mathcal{Q}$ is an uncertainty set on the joint distribution of $(\underline{\mathrm{G}}(z), \underline{\mathrm{Y}})$ on $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2)$. Consider the problem

$$
\begin{array}{llll}
\text{(RobCP)} & \max\limits_{z \in Z_0,\, \pi(P) \geq 0} & f(z) & \\
(16) & \text{s.t.} & [\underline{\mathrm{G}}(z)](\omega_{1j}) \geq \dfrac{\sum_{k=1}^{K} \pi_{jk}(P)\, \underline{\mathrm{Y}}(\omega_{2k})}{P(\{\omega_{1j}\})}, & \\
 & & \forall j \in \mathcal{J},\, \forall P \in \mathcal{Q}, & \\
(17) & & \displaystyle\sum_{k=1}^{K} \pi_{jk}(P) = P(\{\omega_{1j}\}), & \\
 & & \forall j \in \mathcal{J},\, \forall P \in \mathcal{Q}, & \\
(18) & & \displaystyle\sum_{j=1}^{J} \pi_{jk}(P) = P(\{\omega_{2k}\}), & \\
 & & \forall k \in \mathcal{K},\, \forall P \in \mathcal{Q}. &
\end{array}
$$

Problem (RobCP) has infinitely many constraints and decision variables in the infinite-dimensional space $\mathcal{C}(\mathcal{Q})$. We point out that if $(\hat{z}, \hat{\pi}(P))$ is a solution to problem (RobCP), then $\underline{\mathrm{G}}(\hat{z}) \geq_{icv}^{P} \underline{\mathrm{Y}}$ for all $P \in \mathcal{Q}$. The perturbation of problem (RobCP) is

$$
\begin{array}{llll}
(\epsilon\text{−RobCP}) & \max\limits_{z \in Z_0,\, \pi(P) \geq 0} & f(z) & \\
(19) & \text{s.t.} & [\underline{\mathrm{G}}(z)](\omega_{1j}) \geq \dfrac{\sum_{k=1}^{K} \pi_{jk}(P)\, (\underline{\mathrm{Y}}(\omega_{2k}) + \epsilon\, \underline{1})}{P(\{\omega_{1j}\})}, & \\
 & & \forall j \in \mathcal{J},\, \forall P \in \mathcal{Q}, & \\
(20) & & \displaystyle\sum_{k=1}^{K} \pi_{jk}(P) = P(\{\omega_{1j}\}), & \\
 & & \forall j \in \mathcal{J},\, \forall P \in \mathcal{Q}, & \\
(21) & & \displaystyle\sum_{j=1}^{J} \pi_{jk}(P) = P(\{\omega_{2k}\}), & \\
 & & \forall k \in \mathcal{K},\, \forall P \in \mathcal{Q}. &
\end{array}
$$

We describe the connection between problem (ϵ−RobMP) and problem (ϵ−RobCP) in the next proposition.

Proposition 5.4.1. *For $\epsilon < 0$, problem ($\sqrt{N}\,\epsilon$−RobMP) is a relaxation of problem (ϵ−RobCP) on $\Omega_1 = \{\omega_{11}, \ldots, \omega_{1J}\}$ and $\Omega_2 = \{\omega_{21}, \ldots, \omega_{2K}\}$.*

Proof. Problem (ϵ − RobCP) is equivalent to

$$\max_{z \in Z_0} \quad f(z)$$
$$\text{s.t.} \quad \sum_{j=1}^{J} P_1(\{\omega_{1j}\})\, u([\underline{G}(z)](\omega_{1j})) \qquad \forall u \in \mathcal{U}^N(W), \forall P \in \mathcal{Q},$$
$$\geq \sum_{k=1}^{K} P_2(\{\omega_{2k}\})\, u(\underline{Y}(\omega_{2k}) + \epsilon\, \underline{1}),$$

by construction. In this setting, problem ($\epsilon -$ RobMP) is

$$\max_{z \in Z_0} \quad f(z)$$
$$\text{s.t.} \quad \sum_{j=1}^{J} P_1(\{\omega_{1j}\})\, u([\underline{G}(z)](\omega_{1j})) \qquad \forall u \in \mathcal{U}^N(W), \forall P \in \mathcal{Q},$$
$$\geq \sum_{k=1}^{K} P_2(\{\omega_{2k}\})\, u(\underline{Y}(\omega_{2k})) + \epsilon.$$

For any $u \in \mathcal{U}^N(W)$ we have

$$|\mathbb{E}[u(\underline{Y} + \epsilon\, \underline{1})] - \mathbb{E}[u(\underline{Y})]| \leq \|\epsilon\, \underline{1}\|_2 = \sqrt{N}\, \epsilon,$$

since $\|\partial u\|_W \leq 1$. Rearrange to obtain

$$\mathbb{E}[u(\underline{Y} + \epsilon\, \underline{1})] \geq \mathbb{E}[u(\underline{Y})] + \sqrt{N}\, \epsilon.$$

For any $z \in Z_0$ such that $(19) - (21)$ we have

$$\sum_{j=1}^{J} P_1(\{\omega_{1j}\})\, u([\underline{G}(z)](\omega_{1j})) \qquad \forall u \in \mathcal{U}^N(W), \qquad \forall P \in \mathcal{Q},$$
$$\geq \sum_{k=1}^{K} P_2(\{\omega_{2k}\})\, u(\underline{Y}(\omega_{2k}) + \epsilon\, \underline{1}),$$

and thus

$$\sum_{j=1}^{J} P_1(\{\omega_{1j}\})\, u([\underline{G}(z)](\omega_{1j})) \qquad \forall u \in \mathcal{U}^N(W), \qquad \forall P \in \mathcal{Q},$$
$$\geq \sum_{k=1}^{K} P_2(\{\omega_{2k}\})\, u(\underline{Y}(\omega_{2k})) + \sqrt{N}\, \epsilon.$$

□

Now suppose that the system $(16) - (18)$ is satisfied for all $P \in \text{bd}\{\mathcal{Q}\}$ where $\text{bd}\{\mathcal{Q}\}$ denotes the boundary of $\mathcal{Q}$ in $\mathbb{R}^J$. By convexity of $\mathcal{Q}$, the system $(16) - (18)$ is satisfied for all $P \in \mathcal{Q}$. To see this fact, set $P_\lambda = \sum_{l=1}^{L} \lambda_l P_l$ with all $P_l \in \text{bd}\{\mathcal{Q}\}$

and $\sum_{l=1}^{L} \lambda_l = 1$ and $\lambda \geq 0$. Then define $\pi_{jk}(P_\lambda) = \sum_{l=1}^{L} \lambda_l \pi_{jk}(P_l)$. We see immediately that

$$\begin{aligned} P_\lambda(\{\omega_{1j}\})\left[\underline{\mathrm{G}}(z)\right](\omega_{1j}) &= \sum_{l=1}^{L} \lambda_l P_l(\{\omega_{1j}\})\left[\underline{\mathrm{G}}(z)\right](\omega_{1j}) \\ &\geq \sum_{l=1}^{L} \lambda_l \left(\sum_{k=1}^{K} \pi_{jk}(P_l)\,\underline{\mathrm{Y}}(\omega_{2k})\right) \\ &= \sum_{k=1}^{K} \left(\sum_{l=1}^{L} \lambda_l \pi_{jk}(P_l)\right)\underline{\mathrm{Y}}(\omega_{2k}) = \sum_{k=1}^{K} \pi_{jk}(P_\lambda)\,\underline{\mathrm{Y}}(\omega_{2k}). \end{aligned}$$

Further,

$$\begin{aligned} \sum_{k=1}^{K} \pi_{jk}(P_\lambda) &= \sum_{k=1}^{K} \left(\sum_{l=1}^{L} \lambda_l \pi_{jk}(P_l)\right) \\ &= \sum_{l=1}^{L} \lambda_l \left(\sum_{k=1}^{K} \pi_{jk}(P_l)\right) \\ &= \sum_{l=1}^{L} \lambda_l P_l(\{\omega_{1j}\}) = P_\lambda(\{\omega_{1j}\}), \end{aligned}$$

and

$$\begin{aligned} \sum_{j=1}^{J} \pi_{jk}(P_\lambda) &= \sum_{j=1}^{J} \left(\sum_{l=1}^{L} \lambda_l \pi_{jk}(P_l)\right) \\ &= \sum_{l=1}^{L} \lambda_l \left(\sum_{j=1}^{J} \pi_{jk}(P_l)\right) \\ &= \sum_{l=1}^{L} \lambda_l P(\{\omega_{2k}\}) = P(\{\omega_{2k}\}), \end{aligned}$$

so that $\pi_{ij}(P_\lambda)$ as defined is a solution to (16) – (18) for $P_\lambda = \sum_{l=1}^{L} \lambda_l P_l$.

When $\mathcal{Q}$ is polyhedral, then we can check $\underline{\mathrm{G}}(z) \geq_{icv}^{\mathcal{Q}} \underline{\mathrm{Y}}$ by checking all the extreme points of $\mathcal{Q}$. Suppose $\{P_1, \ldots, P_T\}$ are the extreme points of $\mathcal{Q}$ indexed by $\mathcal{T} = \{1, \ldots, T\}$. Then we can define the problem

($\epsilon-$PolyRobCP)

$$\max_{z\in Z_0,\,\pi(P)\geq 0} \quad f\left(z\right)$$

$$\text{s.t.} \quad \left[\underline{\mathrm{G}}\left(z\right)\right]\left(\omega_{1j}\right) \geq \frac{\sum_{k=1}^{K}\pi_{jk}\left(P_t\right)\left(\underline{\mathrm{Y}}\left(\omega_{2k}\right)+\epsilon\,\underline{1}\right)}{P_t\left(\left\{\omega_{1j}\right\}\right)}, \tag{22}$$

$$j\in\mathcal{J},\, t\in\mathcal{T},$$

$$\sum_{k=1}^{K}\pi_{jk}\left(P_t\right) = P_t\left(\left\{\omega_{1j}\right\}\right), \tag{23}$$

$$j\in\mathcal{J},\, t\in\mathcal{T},$$

$$\sum_{j=1}^{J}\pi_{jk}\left(P_t\right) = P_t\left(\left\{\omega_{2k}\right\}\right), \tag{24}$$

$$k\in\mathcal{K},\, t\in\mathcal{T}.$$

We introduce the Lagrangian for problem ($\epsilon-$PolyRobCP),

$$\begin{aligned}
& L\left(z,\pi,\lambda,\delta,\gamma\right)\\
= \quad & f\left(z\right)\\
& +\sum_{t=1}^{T}\sum_{j=1}^{J}\langle\lambda_{tj},\left[\underline{\mathrm{G}}\left(z\right)\right]\left(\omega_{1j}\right)-\frac{\sum_{k=1}^{K}\pi_{jk}\left(P_t\right)\left(\underline{\mathrm{Y}}\left(\omega_{2k}\right)+\epsilon\,\underline{1}\right)}{P_t\left(\left\{\omega_{1j}\right\}\right)}\rangle\\
& +\sum_{t=1}^{T}\sum_{j=1}^{J}\delta_{tj}\left(\sum_{k=1}^{K}\pi_{jk}\left(P_t\right)-P_t\left(\left\{\omega_{1j}\right\}\right)\right)\\
& +\sum_{t=1}^{T}\sum_{k=1}^{K}\gamma_{tk}\left(\sum_{j=1}^{J}\pi_{jk}\left(P_t\right)-P_t\left(\left\{\omega_{2k}\right\}\right)\right).
\end{aligned}$$

Assumption 5.4.2. *There exists $\tilde{z}\in Z_0$ and $\tilde{\pi}\geq 0$ such that*

$$\begin{aligned}
\left[\underline{G}\left(\tilde{z}\right)\right]\left(\omega_{1j}\right) > \frac{\sum_{k=1}^{K}\tilde{\pi}_{jk}\left(P_t\right)\left(\underline{Y}\left(\omega_{2k}\right)+\epsilon\,\underline{1}\right)}{P_t\left(\left\{\omega_{1j}\right\}\right)}, &\quad \forall j\in\mathcal{J},\,\forall t\in\mathcal{T},\\
\sum_{k=1}^{K}\tilde{\pi}_{jk}\left(P_t\right) = P_t\left(\left\{\omega_{1j}\right\}\right), &\quad \forall j\in\mathcal{J},\,\forall t\in\mathcal{T},\\
\sum_{j=1}^{J}\tilde{\pi}_{jk}\left(P_t\right) = P_t\left(\left\{\omega_{2k}\right\}\right), &\quad \forall k\in\mathcal{K},\,\forall t\in\mathcal{T}.
\end{aligned}$$

We obtain the following optimality conditions.

Theorem 5.4.3. *Suppose assumption 5.4.2 holds. If $(\hat{z},\hat{\pi})$ is optimal for problem ($\epsilon-$PolyRobCP), then there exists $\hat{\lambda}\geq 0$ and $\left(\hat{\delta},\hat{\gamma}\right)$ such that*

$$L\left(\hat{z},\hat{\pi},\hat{\lambda},\hat{\delta},\hat{\gamma}\right) = \max\left\{L\left(z,\pi,\hat{\lambda},\hat{\delta},\hat{\gamma}\right)\ :\ (23)-(24)\right\},$$

$$0 = \sum_{t=1}^{T}\sum_{j=1}^{J}\langle\hat{\lambda}_{tj},\left[\underline{G}\left(\hat{z}\right)\right]\left(\omega_{1j}\right)-\frac{\sum_{k=1}^{K}\hat{\pi}_{jk}\left(P_t\right)\left(\underline{Y}\left(\omega_{2k}\right)+\epsilon\,\underline{1}\right)}{P_t\left(\{\omega_{1j}\}\right)}\rangle,$$

$$0 = \sum_{t=1}^{T}\sum_{j=1}^{J}\hat{\delta}_{tj}\left(\sum_{k=1}^{K}\hat{\pi}_{jk}\left(P_t\right)-P_t\left(\{\omega_{1j}\}\right)\right),$$

$$0 = \sum_{t=1}^{T}\sum_{k=1}^{K}\hat{\gamma}_{tk}\left(\sum_{j=1}^{J}\hat{\pi}_{jk}\left(P_t\right)-P_t\left(\{\omega_{2k}\}\right)\right).$$

Now we consider the dual of the previous problem. For the dual functional

$$\mathrm{d}\left(\lambda\right)=\max\left\{f\left(z\right)+\sum_{t=1}^{T}\sum_{j=1}^{J}\langle\lambda_{tj},\left[\underline{\mathrm{G}}\left(z\right)\right]\left(\omega_{1j}\right)\rangle\ :\ z\in Z_0\right\},$$

the corresponding dual problem is

$$(\epsilon-\mathrm{PolyRobCP}_D)\quad \min\quad \mathrm{d}\left(\lambda\right)-\sum_{t=1}^{T}\sum_{j=1}^{J}P_t\left(\{\omega_{1j}\}\right)\delta_{tj}-\sum_{t=1}^{T}\sum_{k=1}^{K}P_t\left(\{\omega_{2k}\}\right)\gamma_{tk}$$
$$\text{s.t.}\quad -\langle\lambda_{tj},\frac{\underline{\mathrm{Y}}\left(\omega_{2k}\right)+\epsilon\,\underline{1}}{P_t\left(\{\omega_{1j}\}\right)}\rangle+\delta_{tj}+\gamma_{tk}\le 0,$$
$$\forall t\in\mathcal{T},\ \forall j\in\mathcal{J},\ \forall k\in\mathcal{K}.$$

Under assumption 5.4.2 strong duality holds between problem ($\epsilon-$PolyRobCP) and problem ($\epsilon-$PolyRobCP$_D$). We are justified in solving the dual and recovering the solution to the primal.

Theorem 5.4.4. *Suppose assumption 5.4.2 holds.*

(a) If $(\hat{z},\hat{\pi})$ is optimal for problem ($\epsilon-$PolyRobCP), then problem ($\epsilon-$PolyRobCP$_D$) has an optimal solution and the optimal values are equal.

(b) If problem ($\epsilon-$PolyRobCP$_D$) has an optimal solution then problem ($\epsilon-$PolyRobCP) has an optimal solution.

5.4.1. Aggregation. Again, we are really only interested in the decision variables z in problem (PolyRobCP), but not (z,π) in its totality. We use this observation to justify the expediency of finding a feasible but possibly sub-optimal solution to problem (PolyRobCP).

There is a natural aggregation scheme for problem (PolyRobCP) as in Chapter 4. Let

$$\hat{\pi}\left(J_{i_1},K_{i_2};\,P_t\right)\ :\ \{J_{i_1}\}_{i_1=1}^{I_1}\times\{K_{i_2}\}_{i_2=1}^{I_2}\to\mathbb{R}_+$$

be the joint probability measure on the cross product of the modified sample spaces $\{J_{i_1}\}_{i_1=1}^{I_1}$ and $\{K_{i_2}\}_{i_2=1}^{I_2}$ corresponding to the extremal probability distribution P_t. Even though we are dealing with multiple extremal probability measures $\{P_t\}_{t\in\mathcal{T}}$

and could introduce a separate partition of the state space for each P_t, we use a single partition of the state space. We recover the original joint probability measure

$$\{\pi_{jk}(P_t)\}_{j\in\mathcal{J},\,k\in\mathcal{K}}$$

from the aggregated measure $\hat{\pi}(J_{i_1}, K_{i_2};\, P_t)$ via the following convention. Set

$$\pi_{jk}(P_t) = \frac{P_t(\{\omega_{1j}\})\, P_t(\{\omega_{2j}\})\, \hat{\pi}(J_{i_1}, K_{i_2};\, P_t)}{\left(\sum_{j\in J_{i_1}} P_t(\{\omega_{1j}\})\right)\left(\sum_{k\in K_{i_2}} P_t(\{\omega_{2k}\})\right)}$$

for $J_{i_1} \ni j$ and $K_{i_2} \ni k$.

We adopt the notational conventions

$$P_t(J_{i_1}) \triangleq \sum_{j\in J_{i_1}} P_t(\{\omega_{1j}\})$$

and

$$P_t(K_{i_2}) \triangleq \sum_{k\in K_{i_2}} P_t(\{\omega_{2k}\}).$$

We define (abusing notation by using J_{i_1} and K_{i_2} to indicate scenarios)

$$[\underline{\mathrm{G}}(z)](J_{i_1}) = \frac{\sum_{j\in J_{i_1}} P_t(\{\omega_{1j}\})\,[\underline{\mathrm{G}}(z)](\omega_{1j})}{\sum_{j\in J_{i_1}} P_t(\{\omega_{1j}\})}$$

and

$$\underline{\mathrm{Y}}(K_{i_2}) = \frac{\sum_{k\in K_{i_2}} P_t(\{\omega_{2k}\})\,\underline{\mathrm{Y}}(\omega_{2k})}{\sum_{k\in K_{i_2}} P_t(\{\omega_{2k}\})}.$$

In this new setup, the constraints (22) are recast as

$$P_t(J_{i_1})\,[\underline{\mathrm{G}}(z)](J_{i_1}) \geq \sum_{i_2=1}^{I_2} \hat{\pi}(J_{i_1}, K_{i_2};\, P_t)\left(\underline{\mathrm{Y}}(K_{i_2}) + \epsilon\,\underline{1}\right), \qquad \forall t\in\mathcal{T},\ i_1 = 1,\ldots,I_1.$$

The constraints (23) become

$$\sum_{i_2=1}^{I_2} \hat{\pi}(J_{i_1}, K_{i_2};\, P_t) = P_t(J_{i_1}), \qquad \forall t\in\mathcal{T},\ i_1 = 1,\ldots,I_1.$$

Finally, the constraints (24) become

$$\sum_{i_1=1}^{I_1} \hat{\pi}(J_{i_1}, K_{i_2};\, P_t) = P_t(K_{i_2}), \qquad \forall t\in\mathcal{T},\ i_2 = 1,\ldots,I_2.$$

The aggregate version of problem ($\epsilon-$PolyRobCP) looks exactly like the general form of problem ($\epsilon-$PolyRobCP) with appropriate notational modifications based on this discussion:

($\epsilon-$AggPolyRobCP)

$$\max_{z\in Z_0,\,\pi(P)\geq 0} \quad f(z)$$

$$\text{s.t.} \quad [\underline{G}(z)](J_{i_1}) \geq \frac{\sum_{i_2=1}^{I_2} \hat{\pi}(J_{i_1}, K_{i_2};\, P_t)\left(\underline{Y}(K_{i_2}) + \epsilon\,\underline{1}\right)}{P_t(J_{i_1})}, \tag{25}$$

$$\forall t\in\mathcal{T},\; i_1 = 1,\ldots,I_1,$$

$$\sum_{i_2=1}^{I_2} \hat{\pi}(J_{i_1}, K_{i_2};\, P_t) = P_t(J_{i_1}), \tag{26}$$

$$\forall t\in\mathcal{T},\; i_1 = 1,\ldots,I_1,$$

$$\sum_{i_1=1}^{I_1} \hat{\pi}(J_{i_1}, K_{i_2};\, P_t) = P_t(K_{i_2}) \tag{27}$$

$$\forall t\in\mathcal{T},\; i_2 = 1,\ldots,I_2,$$

and the corresponding feasible region, set of optimal solutions, and optimal value are:

$$\begin{aligned}
& Z_{JK}^{\mathcal{Q},\epsilon}\left(\{J_{i_1}\}_{i_1=1}^{I_1}, \{K_{i_2}\}_{i_2=1}^{I_2}\right) \\
\triangleq\ & \{z\in Z_0,\, \hat{\pi}\geq 0 :\ (25)-(27)\}, \\
& S_{JK}^{\mathcal{Q},\epsilon}\left(\{J_{i_1}\}_{i_1=1}^{I_1}, \{K_{i_2}\}_{i_2=1}^{I_2}\right) \\
\triangleq\ & \arg\max_{(z,\hat{\pi})}\left\{f(z)\ :\ (z,\hat{\pi})\in Z_{JK}^{\mathcal{Q},\epsilon}\left(\{J_{i_1}\}_{i_1=1}^{I_1}, \{K_{i_2}\}_{i_2=1}^{I_2}\right)\right\}, \\
& \nu_{JK}^{\mathcal{Q},\epsilon}\left(\{J_{i_1}\}_{i_1=1}^{I_1}, \{K_{i_2}\}_{i_2=1}^{I_2}\right) \\
\triangleq\ & \max\left\{f(z)\ :\ (z,\hat{\pi})\in Z_{JK}^{\mathcal{Q},\epsilon}\left(\{J_{i_1}\}_{i_1=1}^{I_1}, \{K_{i_2}\}_{i_2=1}^{I_2}\right)\right\}.
\end{aligned}$$

5.4.2. Row generation. We can use the transformation

$$\pi_{jk}(P_t) = \frac{P_t(\{\omega_{1j}\})\,P_t(\{\omega_{2k}\})\,\hat{\pi}(J_{i_1}, K_{i_2};\, P_t)}{\left(\sum_{j\in J_{i_1}} P_t(\{\omega_{1j}\})\right)\left(\sum_{k\in K_{i_2}} P_t(\{\omega_{2k}\})\right)}$$

for $J_{i_1} \ni j$ and $K_{i_2} \ni k$ to generate constraints for problem ($\epsilon-$PolyRobCP). The coupling constraints (22) expand to become:

$$\begin{aligned}
& P_t(\omega_{1j})\,[\underline{G}(z)](\omega_{1j}) \\
\geq & \sum_{i_2=1}^{I_2}\sum_{k\in K_{i_2}}\left(\frac{P_t(\{\omega_{1j}\})\,P_t(\{\omega_{2k}\})\,\hat{\pi}(J_{i_1}, K_{i_2};\, P_t)}{\left(\sum_{j\in J_{i_1}} P_t(\{\omega_{1j}\})\right)\left(\sum_{k\in K_{i_2}} P_t(\{\omega_{2k}\})\right)}\right)\left(\underline{Y}(\omega_{2k}) + \epsilon\,\underline{1}\right), \\
& \forall j\in\mathcal{J}.
\end{aligned}$$

These constraints can be further transformed to give

$$\begin{aligned}
&\left(\sum_{j\in J_{i_1}} P_t\left(\{\omega_{1j}\}\right)\right)\left[\underline{\mathrm{G}}\left(z\right)\right]\left(\omega_{1j}\right) \\
&\geq \sum_{i_2=1}^{I_2} \hat{\pi}\left(J_{i_1}, K_{i_2};\, P_t\right)\left(\sum_{k\in K_{i_2}} \frac{P_t\left(\{\omega_{2k}\}\right)\underline{\mathrm{Y}}\left(\omega_{2k}\right)}{\sum_{k\in K_{i_2}} P_t\left(\{\omega_{2k}\}\right)} + \epsilon\,\underline{1}\right), \\
&\qquad \forall j \in \mathcal{J}.
\end{aligned}$$

The equality in distribution constraints (23) – (24) are automatically satisfied in this situation. Suppose that $\hat{\pi}$ satisfies (23) – (24), then

$$\begin{aligned}
\sum_{k=1}^{K} \pi_{jk}\left(P_t\right) &= \sum_{i_2=1}^{I_2} \sum_{k\in K_{i_2}} \pi_{jk}\left(P_t\right) \\
&= \sum_{i_2=1}^{I_2} \sum_{k\in K_{i_2}} \frac{P_t\left(\{\omega_{1j}\}\right) P_t\left(\{\omega_{2k}\}\right) \hat{\pi}\left(J_{i_1}, K_{i_2};\, P_t\right)}{\left(\sum_{j\in J_{i_1}} P_t\left(\{\omega_{1j}\}\right)\right)\left(\sum_{k\in K_{i_2}} P_t\left(\{\omega_{2k}\}\right)\right)} \\
&= \sum_{i_2=1}^{I_2} \frac{P_t\left(\{\omega_{1j}\}\right) \hat{\pi}\left(J_{i_1}, K_{i_2};\, P_t\right)}{\left(\sum_{j\in J_{i_1}} P_t\left(\{\omega_{1j}\}\right)\right)} \\
&= P_t\left(\omega_{1j}\right),
\end{aligned}$$

for all $t \in \mathcal{T}$ and $j \in \mathcal{J}$. A similar calculation establishes that

$$\begin{aligned}
\sum_{j=1}^{J} \pi_{jk}\left(P_t\right) &= \sum_{i_1=1}^{I_1} \sum_{j\in J_{i_1}} \pi_{jk}\left(P_t\right) \\
&= \sum_{i_1=1}^{I_1} \frac{P_t\left(\{\omega_{2k}\}\right) \hat{\pi}\left(J_{i_1}, K_{i_2};\, P_t\right)}{\left(\sum_{k\in K_{i_2}} P_t\left(\{\omega_{2k}\}\right)\right)} \\
&= P_t\left(\{\omega_{2k}\}\right),
\end{aligned}$$

for all $t \in \mathcal{T}$ and $k \in \mathcal{K}$.

We now obtain the problem

(ϵ-RowPolyRobCP)

$$
\begin{aligned}
&\max_{z\in Z_0,\,\hat{\pi}\geq 0} && f(z) \\
&\text{s.t.} && \left(\sum_{j\in J_{i_1}} P_t\left(\{\omega_{1j}\}\right)\right)\left[\underline{\mathrm{G}}(z)\right](\omega_{1j}) \qquad (28)\\
& && \geq \sum_{i_2=1}^{I_2}\hat{\pi}\left(J_{i_1},K_{i_2};\,P_t\right)\left(\sum_{k\in K_{i_2}}\frac{P_t\left(\{\omega_{2k}\}\right)\underline{\mathrm{Y}}(\omega_{2k})}{\sum_{k\in K_{i_2}}P_t\left(\{\omega_{2k}\}\right)}+\epsilon\,\underline{1}\right),\\
& && \forall t\in\mathcal{T},\,\forall j\in\mathcal{J},\\
& && \sum_{i_2=1}^{I_2}\hat{\pi}\left(J_{i_1},K_{i_2};\,P_t\right)=P_t\left(J_{i_1}\right), \qquad (29)\\
& && \forall t\in\mathcal{T},\,i_1=1,\ldots,I_1,\\
& && \sum_{i_1=1}^{I_1}\hat{\pi}\left(J_{i_1},K_{i_2}\right)=\frac{|K_{i_2}|}{K}, \qquad (30)\\
& && \forall t\in\mathcal{T},\,i_2=1,\ldots,I_2,
\end{aligned}
$$

and we denote its feasible region, set of optimal solutions, and optimal value as:

$$
\begin{aligned}
& Z^{\mathcal{Q},\epsilon}_{Row,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1},\{K_{i_2}\}_{i_2=1}^{I_2}\right)\\
\triangleq\; & \left\{z\in Z_0,\,\hat{\pi}\geq 0:\;(28)-(30)\right\},\\
& S^{\mathcal{Q},\epsilon}_{Row,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1},\{K_{i_2}\}_{i_2=1}^{I_2}\right)\\
\triangleq\; & \arg\max_{(z,\hat{\pi})}\left\{f(z)\;:\;(z,\hat{\pi})\in Z^{\mathcal{Q},\epsilon}_{Row,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1},\{K_{i_2}\}_{i_2=1}^{I_2}\right)\right\},\\
& \nu^{\mathcal{Q},\epsilon}_{Row,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1},\{K_{i_2}\}_{i_2=1}^{I_2}\right)\\
\triangleq\; & \max\left\{f(z)\;:\;(z,\hat{\pi})\in Z^{\mathcal{Q},\epsilon}_{Row,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1},\{K_{i_2}\}_{i_2=1}^{I_2}\right)\right\}.
\end{aligned}
$$

Problem ($\epsilon-$RowPolyRobCP) has many constraints and few variables, and it can be solved via row (cut) generation techniques. Since problem ($\epsilon-$RowPolyRobCP) is a restriction of problem ($\epsilon-$RowPolyRobCP), we see immediately that

$$Z^{\mathcal{Q},\epsilon}_{Row,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1},\{K_{i_2}\}_{i_2=1}^{I_2}\right)\subset Z^{\mathcal{Q},\epsilon}_{Row,JK}$$

and

$$\nu^{\mathcal{Q},\epsilon}_{Row,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1},\{K_{i_2}\}_{i_2=1}^{I_2}\right)\leq \nu^{\mathcal{Q},\epsilon}_{Row,JK}.$$

By the same reasoning,

$$Z^{\mathcal{Q},\epsilon}_{Row,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1},\{K_{i_2}\}_{i_2=1}^{I_2}\right)\subset Z^{\mathcal{Q},\epsilon}_{Agg,JK}$$

and

$$\nu^{\mathcal{Q},\epsilon}_{Row,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1},\{K_{i_2}\}_{i_2=1}^{I_2}\right)\le\nu^{\mathcal{Q},\epsilon}_{Agg,JK}.$$

5.4.3. Column generation. We can also use the partitions $\{J_{i_1}\}_{i_1=1}^{I_1}$ and $\{K_{i_2}\}_{i_2=1}^{I_2}$ to generate the full set of variables for problem ($\epsilon-$PolyRobCP). The coupling constraints (22) expand to become:

$$\begin{aligned}&\sum_{j\in J_{i_1}}P_t(\omega_{1j})\left[\underline{G}(z)\right](\omega_{1j})\\ \geq\quad&\sum_{j\in J_{i_1}}\sum_{k=1}^{K}\pi_{jk}(P_t)\left(\underline{Y}(\omega_{2k})+\epsilon\,\underline{1}\right),\\ &\forall t\in\mathcal{T},\,i_1=1,\ldots,I_1.\end{aligned}$$

The equality in distribution constraints $(23)-(24)$ expand to become

$$\sum_{j\in J_{i_1}}\sum_{k=1}^{K}\pi_{jk}(P_t)=\sum_{j\in J_{I_1}}P_t(\omega_{1j}),\quad\forall t\in\mathcal{T},\,i_1=1,\ldots,I_1,$$

and

$$\sum_{k\in K_{i_2}}\sum_{j=1}^{J}\pi_{jk}(P_t)=\sum_{k\in K_{i_2}}P_t(\omega_{2k}),\quad\forall t\in\mathcal{T},\,i_2=1,\ldots,I_2.$$

We obtain the optimization problem

$$\begin{aligned}
&(\epsilon-\text{ColPolyRobCP}) && \max_{z\in Z_0,\,\hat{\pi}\geq 0} && f(z)\\
&(31) && \text{s.t.} && \sum_{j\in J_{i_1}}P_t(\omega_{1j})\left[\underline{G}(z)\right](\omega_{1j})\\
&&&&& \geq\sum_{j\in J_{i_1}}\sum_{k=1}^{K}\pi_{jk}(P_t)\left(\underline{Y}(\omega_{2k})+\epsilon\,\underline{1}\right),\\
&&&&& \forall t\in\mathcal{T},\,i_1=1,\ldots,I_1,\\
&(32) &&&& \sum_{j\in J_{i_1}}\sum_{k=1}^{K}\pi_{jk}(P_t)=\sum_{j\in J_{I_1}}P_t(\omega_{1j}),\\
&&&&& \forall t\in\mathcal{T},\,i_1=1,\ldots,I_1,\\
&(33) &&&& \sum_{k\in K_{i_2}}\sum_{j=1}^{J}\pi_{jk}(P_t)=\sum_{k\in K_{i_2}}P_t(\omega_{2k}),\\
&&&&& \forall t\in\mathcal{T},\,i_2=1,\ldots,I_2,
\end{aligned}$$

and we denote its feasible region, set of optimal solutions, and optimal value as:

$$\begin{aligned}
& Z^{\mathcal{Q},\epsilon}_{Col,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1}, \{K_{i_2}\}_{i_2=1}^{I_2}\right) \\
\triangleq \quad & \{z \in Z_0,\, \pi \geq 0 :\ (31) - (33)\}, \\
& S^{\mathcal{Q},\epsilon}_{Col,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1}, \{K_{i_2}\}_{i_2=1}^{I_2}\right) \\
\triangleq \quad & \arg\max_{(z,\hat{\pi})}\left\{f(z)\ :\ (z,\pi) \in Z^{\mathcal{Q},\epsilon}_{Col,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1}, \{K_{i_2}\}_{i_2=1}^{I_2}\right)\right\}, \\
& \nu^{\mathcal{Q},\epsilon}_{Col,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1}, \{K_{i_2}\}_{i_2=1}^{I_2}\right) \\
\triangleq \quad & \max\left\{f(z)\ :\ (z,\pi) \in Z^{\mathcal{Q},\epsilon}_{Col,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1}, \{K_{i_2}\}_{i_2=1}^{I_2}\right)\right\}.
\end{aligned}$$

Problem ($\epsilon-$ColPolyRobCP) has few constraints and many variables, and it can be solved via column generation (pricing) techniques. Since problem ($\epsilon-$ColPolyRobCP) is a relaxation of problem ($\epsilon-$ColPolyRobCP), we see immediately that

$$Z^{\mathcal{Q},\epsilon}_{Col,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1}, \{K_{i_2}\}_{i_2=1}^{I_2}\right) \supset Z^{\mathcal{Q},\epsilon}_{Col,JK}$$

and

$$\nu^{\mathcal{Q},\epsilon}_{Col,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1}, \{K_{i_2}\}_{i_2=1}^{I_2}\right) \geq \nu^{\mathcal{Q},\epsilon}_{Col,JK}.$$

By the same reasoning,

$$Z^{\mathcal{Q},\epsilon}_{Col,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1}, \{K_{i_2}\}_{i_2=1}^{I_2}\right) \supset Z^{\mathcal{Q},\epsilon}_{Agg,JK}$$

and

$$\nu^{\mathcal{Q},\epsilon}_{Col,JK}\left(\{J_{i_1}\}_{i_1=1}^{I_1}, \{K_{i_2}\}_{i_2=1}^{I_2}\right) \geq \nu^{\mathcal{Q},\epsilon}_{Agg,JK}.$$

CHAPTER 6

Multi-period optimization with increasing concave stochastic order constraints

6.1. Multi-period optimization problem

In [**9**], a novel class of multi-period stochastic programs is developed. This class of problems is characterized by a stochastic order constraint on the vector of system performance measures across all time periods. We will continue this work in this chapter and study a different type of stochastic order constraint. Specifically, we use the multivariate increasing concave stochastic order in line with the theme of this dissertation. This choice of stochastic order constraint leads to easily implementable sampling methods.

6.1.1. System dynamic and performance measures. Our time horizon is $t = 0, \ldots, T$. We introduce a filtration $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_{T+1}$ on $(\Omega, \mathcal{F}, P)$ where $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_{T+1} = \mathcal{F}$. The $\sigma-$field $\mathcal{F}_t$ represents the information available to the decision maker at the beginning of period t. We have a discrete time dynamical system

$$s_{t+1} = A_t s_t + B_t v_t + e_t, \quad t = 0, \ldots, T-1,$$

where $s_t \in \mathcal{L}_1^{N_s}(\Omega, \mathcal{F}_t, P)$ is the state vector at time t, $v_t \in \mathcal{L}_1^{N_v}(\Omega, \mathcal{F}_t, P)$ is the control vector at time t, and $e_t \in \mathcal{L}_1^{N_s}(\Omega, \mathcal{F}_t, P)$ is the disturbance at time t. The initial state s_0 is given, and the random matrices satisfy $A_t \in \mathcal{L}_\infty^{N_s \times N_s}(\Omega, \mathcal{F}_t, P)$ and $B_t \in \mathcal{L}_\infty^{N_s \times N_v}(\Omega, \mathcal{F}_t, P)$ where $\mathcal{L}_\infty^{M \times N}(\Omega, \mathcal{F}, P)$ is the space of all essentially bounded measurable mappings $X : \Omega \to \mathbb{R}^{M \times N}$.

This system has performance measures

$$G_t : \mathcal{L}_1^{N_s}(\Omega, \mathcal{F}_t, P) \times \mathcal{L}_1^{N_v}(\Omega, \mathcal{F}_t, P) \to \mathcal{L}_1(\Omega, \mathcal{F}_t, P)$$

for $t = 0, \ldots, T-1$, and

$$G_T : \mathcal{L}_1^{N_s}(\Omega, \mathcal{F}_T, P) \to \mathcal{L}_1(\Omega, \mathcal{F}_T, P),$$

denoted as

$$[G_t(s_t, v_t)](\omega), \quad t = 0, \ldots, T-1,$$
$$[G_T(s_T, v_T)](\omega).$$

We use the shorthand $s = (s_1, \ldots, s_T)$, $v = (v_0, \ldots, v_{T-1})$, and

$$\underline{G}(s, v) = (G_0(s_0, v_0), \ldots, G_{T-1}(s_{T-1}, v_{T-1}), G_T(s_T)).$$

We assume $[\underline{G}(s, v)](\omega)$ is continuous and concave in (s, v) for $P-$almost all $\omega \in \Omega$.

Define the space of state trajectories $(s_1, \ldots, s_T)$ to be

$$\mathcal{S} = \mathcal{L}_1^{n_v}(\Omega, \mathcal{F}_1, P) \times \ldots \times \mathcal{L}_1^{n_v}(\Omega, \mathcal{F}_T, P),$$

and the space of controls $(v_0, \ldots, v_{T-1})$ to be

$$\mathcal{V} = \mathcal{L}_1^{n_v}(\Omega, \mathcal{F}_0, P) \times \ldots \times \mathcal{L}_1^{n_v}(\Omega, \mathcal{F}_{T-1}, P).$$

The sets $\mathcal{S}$ and $\mathcal{V}$ implicitly account for non-anticipativity.

We next modify assumption 3.1.1 to apply to the dynamic case.

Assumption 6.1.1. *(a) $\underline{G}(s, v)$ is Lipschitz continuous on $\mathcal{L}_1^{2T}(\Omega, \mathcal{F}, P)$ for $P-$almost all $\omega \in \Omega$. There exists $\Pi \in \mathcal{L}_1^1(\Omega, \mathcal{F}, P)$ such that*

$$\| [\underline{G}(s_1, v_1)](\omega) - [\underline{G}(s_2, v_2)](\omega) \|_2 \quad \leq \quad \Pi(\omega) \| (s_1, v_1) - (s_2, v_2) \|_2$$

for $P-$almost all $\omega \in \Omega$. Set $\pi \triangleq \mathbb{E}[\Pi]$.

(b) The MGF of Π, denoted $M_\Pi(s)$, is finite for s in a neighborhood of zero.

6.1.2. Dynamic stochastic order constraint. We introduce the random vector

$$\underline{Y} = (Y_0, \ldots, Y_T) \in \mathcal{L}_1^{T+1}(\Omega, \mathcal{F}, P)$$

to act as a benchmark for $\underline{G}(s, v)$. For our main problem, we will enforce an increasing concave stochastic order constraint on $\underline{G}(s, v)$,

$$\underline{G}(s, v) \geq_{icv} \underline{Y},$$

defined as

$$\mathbb{E}[u(\underline{G}(s, v))] \geq \mathbb{E}[u(\underline{Y})], \quad \forall u \in \mathcal{U}^N(W).$$

We will assume that $\underline{G}(s, v)$ and $\underline{Y}$ have support contained in W, so that $\mathcal{U}^N(W)$ is sufficient to characterize $\underline{G}(s, v) \geq_{icv} \underline{Y}$.

We have additional constraints on admissible controls, $v_t \in V_t$ for $P-$almost all $\omega \in \Omega$ where $V_t \subset \mathbb{R}^{N_v}$ is a closed convex set for $t = 0, \ldots, T-1$. Our main problem is

$$\text{(DP)} \quad \max_{(s,v) \in \mathcal{S} \times \mathcal{V}} \quad \mathbb{E}\left[\sum_{t=0}^{T-1} G_t(s_t, v_t) + G_T(s_T)\right]$$

$$\text{(1)} \quad \text{s.t.} \quad s_{t+1} = A_t s_t + B_t v_t + e_t, \; t = 0, \ldots, T-1,$$

$$\text{(2)} \quad \mathbb{E}[u(\underline{G}(s, v))] \geq \mathbb{E}[u(\underline{Y})] \quad \forall u \in \mathcal{U}^N(W),$$

$$\text{(3)} \quad v_t \in V_t \text{ a.s.}, \quad t = 0, \ldots, T-1.$$

The notation DP stands for dynamic problem. Also, the expression $v_t \in V_t$ a.s. means $v_t(\omega) \in V_t$ for $P-$almost all $\omega \in \Omega$.

Because of the issue with the Slater condition that we have encountered throughout this dissertation, we introduce the perturbed problem

$$
\begin{aligned}
&(\epsilon\text{-DP}) && \max_{(s,v)\in\mathcal{S}\times\mathcal{V}} && \mathbb{E}\left[\sum_{t=0}^{T-1} G_t\left(s_t, v_t\right) + G_T\left(s_T\right)\right] \\
&(4) && \text{s.t.} && s_{t+1} = A_t s_t + B_t v_t + e_t,\ t = 0, \ldots, T-1, \\
&(5) && && \mathbb{E}\left[u\left(\underline{\mathrm{G}}\left(s, v\right)\right)\right] \geq \mathbb{E}\left[u\left(\underline{\mathrm{Y}}\right)\right] + \epsilon, \qquad \forall u \in \mathcal{U}^N\left(W\right), \\
&(6) && && v_t \in V_t \text{ a.s.}, \qquad t = 0, \ldots, T-1.
\end{aligned}
$$

Proposition 6.1.2. *Problem ($\epsilon-$DP) is a convex optimization problem for all $t = 0, 1, \ldots, T$.*

Proof. The objective function $\mathbb{E}\left[\sum_{t=0}^{T-1} G_t\left(s_t, v_t\right) + G_T\left(s_T\right)\right]$ is concave in (s, v) by construction. The set

$$E = \left\{ (s, v)\ :\ \begin{array}{cc} s_{t+1} = A_t s_t + B_t v_t + e_t, & t = 0, \ldots, T-1, \\ v_t \in V_t \text{ a.s.}, & t = 0, \ldots, T-1 \end{array} \right\}$$

is clearly convex since the system dynamic is linear and all V_t are convex. Further each set

$$A_u^{\epsilon} = \left\{ (s, v) \in \mathcal{S} \times \mathcal{V} : \mathbb{E}\left[u\left(\underline{\mathrm{G}}\left(s, v\right)\right)\right] \geq \mathbb{E}\left[u\left(\underline{\mathrm{Y}}\right)\right] + \epsilon \right\}$$

is convex since $\mathbb{E}\left[u\left(\underline{\mathrm{G}}\left(\cdot\right)\right)\right]$ is concave, and thus

$$\bigcap_{u\in\mathcal{U}^N(W)} A_u^{\epsilon}$$

is convex. □

We also introduce the SIP relaxation of problem ($\epsilon-$DP):

$$
\begin{aligned}
&(\epsilon\text{-DSIP}) && \max_{(s,v)\in\mathcal{S}\times\mathcal{V}} && \mathbb{E}\left[\sum_{t=0}^{T-1} G_t\left(s_t, v_t\right) + G_T\left(s_T\right)\right] \\
&(7) && \text{s.t.} && s_{t+1} = A_t s_t + B_t v_t + e_t, \qquad t = 0, \ldots, T-1, \\
&(8) && && \mathbb{E}\left[u\left(\underline{\mathrm{G}}\left(s, v\right)\right)\right] \geq \mathbb{E}\left[u\left(\underline{\mathrm{Y}}\right)\right] + \epsilon, \qquad \forall u \in U_{\Xi}, \\
&(9) && && v_t \in V_t \text{ a.s.}, \qquad t = 0, \ldots, T-1.
\end{aligned}
$$

The notation DSIP stands for dynamic problem with semi-infinite relaxations.

6.1.3. Nonlinear system dynamic. We can generalize the linear system dynamic (1) to allow for a nonlinear system dynamic

$$s_{t+1} = A_t\left(s_t, v_t\right), \quad t = 0, \ldots, T-1,$$

where $A_t : \mathbb{R}^{N_s} \times \mathbb{R}^{N_v} \to \mathbb{R}^{n_s}$ is a random nonlinear operator with $A_t\left(\omega\right) \in \mathcal{C}\left(\mathbb{R}^{N_s} \times \mathbb{R}^{N_v}; \mathbb{R}^{N_s}\right)$ for $P-$almost all $\omega \in \Omega$, and $A_t \in \mathcal{F}_{t+1}$ for all $t = 0, \ldots, T-1$.

We can consider

$$\max_{(s,v)\in\mathcal{S}\times\mathcal{V}} \quad \mathbb{E}\left[\sum_{t=0}^{T-1} G_t\left(s_t, v_t\right) + G_T\left(s_T\right)\right] \tag{10}$$

$$\text{s.t.} \quad A_t\left(s_t, v_t\right) \geq s_{t+1}, \qquad t = 0, \ldots, T-1, \tag{11}$$

$$\mathbb{E}\left[u\left(\underline{\mathrm{G}}\left(s, v\right)\right)\right] \geq \mathbb{E}\left[u\left(\underline{\mathrm{Y}}\right)\right] + \epsilon, \qquad \forall u \in \mathcal{U}^N\left(W\right), \tag{12}$$

$$v_t \in V_t \text{ a.s.}, \qquad t = 0, \ldots, T-1. \tag{13}$$

Proposition 6.1.3. *Suppose all $G_t\left(s_t, v_t\right)$ are increasing in s_t for $t = 0, \ldots, T-1$ and $G_T\left(s_T\right)$ is increasing in s_T. Also suppose that all A_t are concave for $t = 0, \ldots, T-1$.*

(a) Problem (10) − (13) is a convex optimization problem.

(b) At an optimal solution $\left(\hat{s}, \hat{v}\right)$ of problem (10) − (13), $A_t\left(\hat{s}_t, \hat{v}_t\right) = \hat{s}_{t+1}$ for all $t = 0, \ldots, T-1$.

Proof. (a) The operators A_t are all concave in $\left(s_t, v_t\right)$, so the modified constraints $A_t\left(s_t, v_t\right) \geq s_{t+1}$ for $t = 0, \ldots, T-1$ still induce convex feasible regions in $\left(s, v\right)$.

(b) The objective $\mathbb{E}\left[\sum_{t=0}^{T-1} G_t\left(s_t, v_t\right) + G_T\left(s_T\right)\right]$ is increasing in s by assumption, and if $\underline{\mathrm{G}}\left(s, v\right) \geq_{icv} \underline{\mathrm{Y}}$ then $\underline{\mathrm{G}}\left(s', v\right) \geq_{icv} \underline{\mathrm{Y}}$ for any $s' \geq s$ by the same monotonicity assumption. Since both $s_{t+1} \in \mathcal{F}_{t+1}$ and $A_t\left(s_t, v_t\right) \in \mathcal{F}_{t+1}$, we must have $A_t\left(s_t, v_t\right) = s_{t+1}$ at optimality. □

We will restrict to linear system dynamics for the rest of this chapter, the extension to a concave nonlinear system dynamic is usually immediate.

6.2. Optimality conditions

6.2.1. Optimality conditions for problem ($\epsilon-$DP). We now derive optimality conditions for problem ($\epsilon-$DP). We introduce a multiplier $u \in \mathrm{cl\,cone}\,\mathcal{U}^N\left(W\right)$ corresponding to the increasing concave stochastic order constraint. The Lagrangian $L : \mathcal{S} \times \mathcal{V} \times \mathrm{cl\,cone}\,\mathcal{U}^N\left(W\right) \to \mathbb{R}$ for problem ($\epsilon-$DP) is:

$$L\left(s, v, u\right) \quad = \quad \mathbb{E}\left[\sum_{t=0}^{T-1} G_t\left(s_t, v_t\right) + G_T\left(s_T\right) + u\left(\underline{\mathrm{G}}\left(s, v\right)\right)\right].$$

The Slater condition for problem ($\epsilon-$DP) follows.

Assumption 6.2.1. *There exists $\left(\tilde{s}, \tilde{v}\right) \in \mathcal{S} \times \mathcal{V}$ such that (4) and (6) hold, and*

$$\mathbb{E}\left[u\left(\underline{G}\left(\tilde{s}, \tilde{v}\right)\right)\right] > \mathbb{E}\left[u\left(\underline{Y}\right)\right] + \epsilon \qquad \forall u \in \mathcal{U}^N\left(W\right).$$

We obtain the following optimality conditions for problem problem ($\epsilon-$DP) .

Theorem 6.2.2. *Suppose assumption 6.2.1 holds. If $\left(\hat{s}, \hat{v}\right)$ is an optimal solution of problem ($\epsilon-$DP), then there exists $\hat{u} \in$ cl cone$\mathcal{U}^N\left(W\right)$ such that*

$$L\left(\hat{s}, \hat{v}, \hat{u}\right) \quad = \quad \max_{(s,v)\in\mathcal{S}\times\mathcal{V}} \left\{L\left(s, v, \hat{u}\right) \,:\, (4), (6)\right\},$$

$$\mathbb{E}\left[\hat{u}\left(\underline{G}\left(\hat{s}, \hat{v}\right)\right)\right] \quad = \quad \epsilon \left\| PX\left(\hat{u};\ cl\,conv\mathcal{U}^N\left(W\right)\right) \right\|.$$

PROOF. Define

$$E=\left\{(s,v)\ :\ \begin{array}{cc} s_{t+1}=A_t s_t+B_t v_t+e_t, & t=0,\ldots,T-1, \\ v_t\in V_t \text{ a.s.}, & t=0,\ldots,T-1 \end{array}\right\}.$$

Also define the operator $\boldsymbol{g}(s,v):\mathcal{L}_1^{2T}(\Omega,\mathcal{F},P)\to\mathcal{C}\left(\mathcal{U}^N(W)\right)$ by $[\boldsymbol{g}(s,v)](u)=\mathbb{E}\left[u\left(\underline{\mathrm{G}}(s,v)\right)\right]-\mathbb{E}\left[u\left(\underline{\mathrm{Y}}\right)\right]$, and rewrite problem ($\epsilon-$DP) as

$$\begin{array}{ll} \max\limits_{(s,v)\in\mathcal{S}\times\mathcal{V}} & \mathbb{E}\left[\sum_{t=0}^{T-1}G_t(s_t,v_t)+G_T(s_T)\right] \\ \text{s.t.} & \boldsymbol{g}(s,v)-\epsilon\in\mathcal{C}_+\left(\mathcal{U}^N(W)\right), \\ & (s,v)\in E. \end{array}$$

The mapping $\boldsymbol{g}(\cdot,\cdot)$ is a continuous operator on $\mathcal{S}\times\mathcal{V}$. Using equicontinuity of $\mathcal{U}^N(W)$ and Lipschitz continuity of $\underline{\mathrm{G}}(s,v)$,

$$\begin{array}{ll} & \|\boldsymbol{g}(s_1,v_1)-\boldsymbol{g}(s_2,v_2)\|_{\mathcal{C}(\mathcal{U}^N(W))} \\ \leq & \max\limits_{u\in\mathcal{U}^N(W)}\mathbb{E}\left[\left|u\left(\underline{\mathrm{G}}(s_1,v_1)\right)-u\left(\underline{\mathrm{G}}(s_2,v_2)\right)\right|\right] \\ \leq & \mathbb{E}\left[\|\underline{\mathrm{G}}(s_1,v_1)-\underline{\mathrm{G}}(s_2,v_2)\|_2\right] \\ \leq & \pi\|(s_1,v_1)-(s_2,v_2)\|_2. \end{array}$$

Further, $\boldsymbol{g}(\cdot,\cdot)$ is concave. For (s_1,v_1), (s_2,v_2), $0\leq\alpha\leq 1$, and $u\in\mathcal{U}^N(W)$ we have

$$\begin{array}{ll} & [\boldsymbol{g}(\alpha\, s_1+(1-\alpha)\,s_2,\alpha\, v_1+(1-\alpha)\,v_2)](u) \\ = & \mathbb{E}\left[u\left(\underline{\mathrm{G}}(\alpha\, s_1+(1-\alpha)\,s_1,\alpha\, v_1+(1-\alpha)\,v_2)\right)-u\left(\underline{\mathrm{Y}}\right)\right] \\ \geq & \mathbb{E}\left[u\left(\alpha\,\underline{\mathrm{G}}(s_1,v_1)+(1-\alpha)\,\underline{\mathrm{G}}(s_2,v_2)\right)-u\left(\underline{\mathrm{Y}}\right)\right] \\ \geq & \mathbb{E}\left[\alpha\, u\left(\underline{\mathrm{G}}(s_1,v_1)\right)+(1-\alpha)\,u\left(\underline{\mathrm{G}}(s_2,v_2)\right)\right] \\ & -\,\mathbb{E}\left[\alpha\, u\left(\underline{\mathrm{Y}}\right)+(1-\alpha)\,u\left(\underline{\mathrm{Y}}\right)\right] \\ = & \alpha\,[\boldsymbol{g}(s_1,v_1)](u)+(1-\alpha)\,[\boldsymbol{g}(s_2,v_2)](u), \end{array}$$

using the fact that u is increasing.

For multipliers $\Lambda\in\mathcal{M}_+\left(\mathcal{U}^N(W)\right)$, the Lagrangian of this problem is

$$\Upsilon(s,v,\Lambda)=\mathbb{E}\left[\sum_{t=0}^{T-1}G_t(s_t,v_t)+G_T(s_T)\right]+\int_{\mathcal{U}^N(W)}[\boldsymbol{g}(s,v)](u)\,d\Lambda(u).$$

Under assumption 6.2.1, the Slater condition is satisfied and we obtain the optimality conditions

$$\begin{array}{lcr} \Psi\left(\hat{s},\hat{v},\hat{\Lambda}\right) & = & \max\left\{\Psi\left(s,v,\hat{\Lambda}\right)\ :\ (s,v)\in\mathcal{S}\times\mathcal{V}\right\}, \\ \langle\Lambda,\boldsymbol{g}(\hat{s},\hat{v})\rangle & = & \epsilon\,\Lambda\left(\mathcal{U}^N(W)\right). \end{array}$$

The results in Chapter 3 allow these conditions to be transformed into the desired ones in terms of increasing concave functions. □

We obtain the following optimality conditions for problem ($\epsilon-$DSIP) using the same argument as for problem ($\epsilon-$DP).

Corollary 6.2.3. *Suppose assumption 6.2.1 holds. If $(\hat{s}, \hat{v})$ is an optimal solution of problem ($\epsilon-$DSIP), then there exists $\hat{u} \in cl\, cone U_{\Xi}$ such that*

$$\begin{aligned} L(\hat{s}, \hat{v}, \hat{u}) &= \max_{(s,v)\in\mathcal{S}\times\mathcal{V}} \{L(s, v, \hat{u}) \;:\; (7), (9)\}, \\ \mathbb{E}\left[\hat{u}\left(\underline{G}(\hat{s}, \hat{v})\right) - \hat{u}\left(\underline{Y}\right)\right] &= \epsilon \left\| PX(\hat{u};\; cl\, conv U_{\Xi}) \right\|. \end{aligned}$$

6.2.2. Auxiliary control problem. The preceding optimality conditions induce an auxiliary control problem. This problem is explicitly:

$$\begin{aligned} \max_{(s,v)\in\mathcal{S}\times\mathcal{V}} \quad & \mathbb{E}\left[\sum_{t=0}^{T-1} G_t(s_t, v_t) + G_T(s_T) + u\left(\underline{\mathrm{G}}(s, v)\right)\right] & \\ \text{s.t.} \quad & s_{t+1} = A_t s_t + B_t v_t + e_t, & t = 0, \ldots, T-1, \\ & v_t \in V_t \text{ a.s.}, & t = 0, \ldots, T-1. \end{aligned}$$

The above problem can be solved with dynamic programming, unlike the original forms of problems ($\epsilon-$DP) and ($\epsilon-$DSIP). In problems ($\epsilon-$DP) and ($\epsilon-$DSIP), the stochastic order constraint cannot be checked from a cost-to-go perspective because it requires knowledge of activity on all other sample paths $\omega \in \Omega$.

Define

$$\mathcal{S}_t^T = \times_{\upsilon=t}^{T} \mathcal{L}_1^1(\Omega, \mathcal{F}_\upsilon, P)$$

and

$$\mathcal{V}_t^{T-1} = \times_{\upsilon=t}^{T-1} \mathcal{L}_1^1(\Omega, \mathcal{F}_\upsilon, P).$$

At time $t = 0, \ldots, T-1$, given history $\mathcal{F}_t$ the cost-to-go problem is:

$$\begin{aligned} (\mathrm{DP}_t(\mathcal{F}_t)) \quad \max_{(s,v)\in\mathcal{S}_{t+1}^T\times\mathcal{V}_t^{T-1}} \quad & \mathbb{E}\left[\sum_{\upsilon=t}^{T-1} G_\upsilon(s_\upsilon, v_\upsilon) + G_T(s_T) + u\left(\underline{\mathrm{G}}(s, v)\right) \;\middle|\; \mathcal{F}_t\right] \\ \text{s.t.} \quad & s_{\upsilon+1} = A_\upsilon s_\upsilon + B_\upsilon v_\upsilon + e_\upsilon, \\ & \upsilon = t, \ldots, T-1, \\ & v_\upsilon \in V_\upsilon \text{ a.s.}, \\ & \upsilon = t, \ldots, T-1. \end{aligned}$$

In the function $\mathbb{E}\left[u\left(\underline{\mathrm{G}}(s, v)\right) \mid \mathcal{F}_t\right]$, the first $t-1$ arguments

$$\left(G_0(s_0, v_0), \ldots, G_{t-1}(s_{t-1}, v_{t-1})\right)$$

are constant since $(s_0, \ldots, s_{t-1}, v_0, \ldots, v_{t-1}) \in \mathcal{F}_t$.

Proposition 6.2.4. *Problem ($DP_t(\mathcal{F}_t)$) is a convex optimization problem for $\mathcal{F}_t$ for all $t = 0, 1, \ldots, T$*

Proof. The function $\mathbb{E}\left[\sum_{\upsilon=t}^{T-1} G_\upsilon(s_\upsilon, v_\upsilon) + G_T(s_T)\right]$ is concave by construction. The function $\mathbb{E}\left[u\left(\underline{\mathrm{G}}(s, v)\right)\right]$ is concave in $(s_t, \ldots, s_T, v_t, \ldots, v_{t-1})$ for any fixed $(s_0, \ldots, s_{t-1}, v_0, \ldots, v_{t-1})$ since u is concave. Additionally, the set

$$E = \left\{ (s, v) \in \mathcal{S}_{t+1}^{T} \times \mathcal{V}_{t}^{T-1} : \begin{array}{ll} s_{t+1} = A_t s_t + B_t v_t + e_t, & t = v, \ldots, T-1, \\ v_t \in V_t \text{ a.s.}, & t = v, \ldots, T-1 \end{array} \right\}$$

is convex. □

6.3. Duality

In this section, we will focus on dynamic programming duality for problem ($\epsilon-$DP) and for our auxiliary control problem. Define the dual functional

$$\mathrm{d}(u) = \max_{(s,v) \in \mathcal{S} \times \mathcal{V}} \{ L(s, v, u) \, : \, (4), (6) \}.$$

Notice that the dual functional $\mathrm{d}(u)$ is equivalent to the auxiliary control problem in the previous section, and it does not depend on ϵ. Using the results from Chapter 3, the dual to problem ($\epsilon-$DP) is

$$(\epsilon\text{−DP}_D) \qquad \min \left\{ \mathrm{d}(u) - \epsilon \left\| \mathrm{PX}\left(u; \operatorname{cl}\operatorname{conv} \mathcal{U}^N(W)\right) \right\| \, : \, u \in \operatorname{cl}\operatorname{cone} \mathcal{U}^N(W) \right\}.$$

Similarly, the dual to problem ($\epsilon-$DSIP) is

$$(\epsilon\text{−DSIP}_D) \qquad \min \left\{ \mathrm{d}(u) - \epsilon \left\| \mathrm{PX}\left(u; \operatorname{cl}\operatorname{conv}(U_\Xi)\right) \right\| \, : \, u \in \operatorname{cl}\operatorname{cone} U_\Xi \right\}.$$

The non-anticipativity constraints will be modeled explicitly in this section. The usual mean-vector constraints are:

$$v_t - \mathbb{E}\left[v_t \mid \mathcal{F}_t \right] = 0 \text{ a.s.}, \quad t = 0, \ldots, T-1.$$

However, we will use state-vector constraints (see **[15]**) for greater ease. On sample path $\omega \in \Omega$, the history of observations at time t is denoted by the history operator

$$H_t \omega = \left(A_0(\omega), B_0(\omega), e_0(\omega), \ldots, A_{t-1}(\omega), B_{t-1}(\omega), e_{t-1}(\omega) \right),$$

for $t = 1, \ldots, T$. At time $t = 0$ there are no observations. The inverse of the history operator H_t is denoted as

$$H_t^{-1} \omega = \{ \tilde{\omega} \in \Omega \, : \, H_t \tilde{\omega} = H_t \omega \}.$$

The set $H_t^{-1} \omega \subset \Omega$ is the set of all scenarios $\tilde{\omega} \in \Omega$ that share the same history as scenario ω up to time t. We let $H_t \Omega$ denote the set of all possible histories of observations up to time t. Introduce random variables $\bar{v}_t : H_t \Omega \to \mathbb{R}^{N_v}$ for $t = 1, \ldots, T-1$ (no state vector is needed for time $t = 0$, since the initial state is assumed to be given). Each $\bar{v}_t$ is automatically $\mathcal{F}_t-$measurable by construction. The non-anticipativity constraints can then be expressed as

$$v_t(\omega) - \bar{v}_t(H_t \omega) = 0 \text{ a.s.}, \quad t = 1, \ldots, T-1.$$

Make the non-anticipativity constraints explicit in $\mathrm{d}(u)$ to obtain

$$\begin{array}{lll} \max\limits_{s,v} & \mathbb{E}\left[\sum\limits_{t=0}^{T-1} G_t\left(s_t, v_t\right) + G_T\left(s_T\right) + u\left(\underline{\mathrm{G}}\left(s, v\right)\right)\right] & \\ \text{s.t.} & s_{t+1} = A_t s_t + B_t v_t + e_t, & t = 0, \ldots, T-1, \\ & s_t\left(\omega\right) - \overline{s}_t\left(H_t \omega\right) = 0 \text{ a.s.}, & t = 0, \ldots, T, \\ & v_t\left(\omega\right) - \overline{v}_t\left(H_t \omega\right) = 0 \text{ a.s.}, & t = 0, \ldots, T-1, \\ & v_t \in V_t \text{ a.s.}, & t = 0, \ldots, T-1. \end{array}$$

Define the random set

$$E\left(\omega\right) = \left\{ \left(s, v\right) \; : \; \begin{array}{c} s_{t+1} = A_t\left(\omega\right) s_t + B_t\left(\omega\right) v_t + e_t\left(\omega\right), \\ t = 0, \ldots, T-1, \\ v_t \in V_t\left(\omega\right), \\ t = 0, \ldots, T-1 \end{array} \right\}$$

to represent the system dynamic and control constraints on a particular sample path. The set $E\left(\omega\right)$ as defined satisfies relatively complete recourse: it can be written as

$$E\left(\omega\right) = E_0\left(\omega\right) \times \cdots \times E_T\left(\omega\right)$$

where each $E_t\left(\omega\right)$ is $\mathcal{F}_t-$measurable, and the constraint

$$\left(s\left(\omega\right), v\left(\omega\right)\right) \in E\left(\omega\right)$$

can be written as

$$v_t\left(\omega\right) \in E_t\left(\omega\right)$$

for $t = 0, \ldots, T-1$ and

$$s_t\left(\omega\right) \in E_t\left(\omega\right)$$

for $t = 0, \ldots, T$.

Introduce the function

$$\phi\left(s\left(\omega\right), v\left(\omega\right), \omega\right) = \begin{cases} \left[\sum_{t=0}^{T-1} G_t\left(s_t, v_t\right) + G_T\left(s_T\right) + u\left(\underline{\mathrm{G}}\left(s, v\right)\right)\right]\left(\omega\right), \\ \left(s, v\right) \in E\left(\omega\right), \\ -\infty, \\ \text{otherwise.} \end{cases}$$

The function $\phi\left(\cdot, \omega\right)$ is concave on its effective domain:

$$\left\{\left(s, v\right) \; : \; \phi\left(s, v, \omega\right) > -\infty\right\}$$

by construction. Further, the effective domain of $\phi\left(\cdot; \omega\right)$ is closed and convex. The function $\phi\left(\cdot, \omega\right)$ is a normal integrand in the sense that its hypograph is closed and measurable (see [**19**, Example 14.32]).

The functional $\mathrm{d}\left(u\right)$ is then equivalent to the optimization problem:

$$\begin{array}{lll} \max\limits_{s,v} & \mathbb{E}\left[\phi\left(s\left(\omega\right),v\left(\omega\right),\omega\right)\right] & \\ \text{s.t.} & s_t\left(\omega\right)-\overline{s}_t\left(H_t\omega\right)=0 \text{ a.s.,} & t=0,\ldots,T, \\ & v_t\left(\omega\right)-\overline{v}_t\left(H_t\omega\right)=0 \text{ a.s.,} & t=0,\ldots,T-1. \end{array}$$

Even though s_0 is given and $v_0(\omega) - \overline{v}_0(H_0\omega) = 0$ is a tautology under our assumptions, we still introduce the time index $t = 0$ above for synchronicity with [**15**]. We will introduce multipliers $\Theta_s = (\Theta_{s0}, \ldots, \Theta_{sT}) \in \mathcal{L}_1^{(T+1)N_s}(\Omega, \mathcal{F}, P)$ and $\Theta_v = \left(\Theta_{v0}, \ldots, \Theta_{v(T-1)}\right) \in \mathcal{L}_1^{T\,N_v}(\Omega, \mathcal{F}, P)$ for the non-anticipativity constraints. Let $\Theta = (\Theta_s, \Theta_v)$. We introduce the conjugate function

$$\begin{aligned} & \phi^*\left(\Theta\left(\omega\right),\omega\right) \\ = & \sup_{s(\omega),v(\omega)}\left\{\phi\left(s\left(\omega\right),v\left(\omega\right),\omega\right)-\left\langle\Theta_s\left(\omega\right),s\left(\omega\right)\right\rangle-\left\langle\Theta_v\left(\omega\right),v\left(\omega\right)\right\rangle\right\}. \end{aligned}$$

The following result establishes duality from the perspective of the non-anticipativity constraints. Let $\Theta_t = (\Theta_{st}, \Theta_{vt})$ for $t = 0, \ldots, T-1$ and $\Theta_T = \Theta_{ST}$.

Theorem 6.3.1. [**15**, Theorem 1] *(a) The dual of $d(u)$ is*

$$\begin{array}{lll} \min\limits_{\Theta} & -\mathbb{E}\left[\phi^*\left(\Theta\left(\omega\right),\omega\right)\right] & \\ s.t. & \mathbb{E}\left[\Theta_t\left(\tilde{\omega}\right)\mid\tilde{\omega}\in H_t^{-1}\left(H_t\omega\right)\right]=0 \textit{ a.s.,} & t=0,\ldots,T. \end{array}$$

(b) Suppose $(\hat{s}, \hat{v})$ is optimal for $d(u)$ and that $\partial\phi(\hat{s}(\omega), \hat{v}(\omega), \omega)$ is non-empty for $P-$almost all $\omega \in \Omega$. Then there exists $\hat{\Theta}(\omega) \in \partial\,\phi(\hat{s}(\omega), \hat{v}(\omega), \omega)$ for $P-$almost all $\omega \in \Omega$ such that

$$\mathbb{E}\left[\Theta_t\left(\tilde{\omega}\right)\ \mid\tilde{\omega}\in H_t^{-1}\left(H_t\omega\right)\right]=0 \text{ a.s., } t=0,\ldots,T.$$

(c) $-\mathbb{E}\left[\phi^*\left(\hat{\Theta}(\omega), \omega\right)\right]$ is the optimal value of $d(u)$ and its dual.

6.4. Finite probability spaces

In this section, we will consider problem (DP) on a finite probability space. The dependence on $\epsilon > 0$ is dropped in this section for notational succinctness. We will work with the coupling transformation in this section so we do not have to rely on the ϵ perturbation to satisfy the Slater condition. For this section, we suppose that the performance measure $\underline{\mathrm{G}}(s, v)$ (and the corresponding discrete time dynamic system that it evaluates) and $\underline{\mathrm{Y}}$ are constructed on separate probability spaces $(\Omega_1, \mathcal{F}_1, P_1)$ and $(\Omega_2, \mathcal{F}_2, P_2)$.

The decision variables (s, v) are now finite dimensional $\{(s(\omega_j), v(\omega_j))\}_{j\in\mathcal{J}}$. We denote $s_t(\omega_j)$ and $v_t(\omega_j)$ as the state and control on scenario ω_j at time t. We obtain the problem:

(DCP)

$$\max \quad \sum_{j=1}^{J} P_1\left(\{\omega_{1j}\}\right)\left[\sum_{t=0}^{T-1}\left[G_t\left(s_t,v_t\right)\right]\left(\omega_{1j}\right)+\left[G_T\left(s_T\right)\right]\left(\omega_{1j}\right)\right]$$

$$(14) \quad \text{s.t.} \quad \left[\underline{\mathrm{G}}\left(s,v\right)\right]\left(\omega_{1j}\right) \geq \frac{\sum_{k=1}^{K}\pi_{jk}\underline{\mathrm{Y}}\left(\omega_{2k}\right)}{P_1\left(\{\omega_{1j}\}\right)}, \qquad \forall j \in \mathcal{J},$$

$$(15) \quad \sum_{k=1}^{K}\pi_{jk} = P_1\left(\{\omega_{1j}\}\right), \qquad \forall j \in \mathcal{J},$$

$$(16) \quad \sum_{j=1}^{J}\pi_{jk} = P_2\left(\{\omega_{2k}\}\right), \qquad \forall k \in \mathcal{K}.$$

$$(17) \quad s_{t+1}\left(\omega_{1j}\right) = A_t\left(\omega_{1j}\right)s_t\left(\omega_{1j}\right) + B_t\left(\omega_{1j}\right)v_t\left(\omega_{1j}\right) + e_t\left(\omega_{1j}\right), \quad t=0,\ldots,T-1,\ \forall j\in\mathcal{J},$$

$$(18) \quad v_t\left(\omega_{1j}\right) - \bar{v}_t\left(H_t\left(\omega_{1j}\right)\right) = 0, \quad t=0,\ldots,T-1, \forall j\in\mathcal{J},$$

$$(19) \quad v_t\left(\omega_{1j}\right) \in V_t\left(\omega_{1j}\right), \quad t=0,\ldots,T-1, \forall j\in\mathcal{J}.$$

There is no need to introduce state vector constraints for s, since (17) automatically ensures that $s_t \in \mathcal{F}_t$ will hold for all $t = 1, \ldots, T$.

We will decompose around the constraints (14) and (18). The Lagrangian for problem (DCP) is:

$$\begin{aligned} L\left(s,v,\bar{v},\pi,\lambda,\Theta\right) \quad = \quad & \sum_{j=1}^{J} P_1\left(\{\omega_{1j}\}\right)\left[\sum_{t=0}^{T-1}\left[G_t\left(s_t,v_t\right)\right]\left(\omega_{1j}\right)+\left[G_T\left(s_T\right)\right]\left(\omega_{1j}\right)\right] \\ & + \sum_{j=1}^{J}\langle\lambda_j, \left[\underline{\mathrm{G}}\left(s,v\right)\right]\left(\omega_{1j}\right) - \frac{\sum_{k=1}^{K}\pi_{jk}\underline{\mathrm{Y}}\left(\omega_{2k}\right)}{P_1\left(\{\omega_{1j}\}\right)}\rangle \\ & + \sum_{j=1}^{J} P\left(\{\omega_{1j}\}\right)\sum_{t=1}^{T-1}\langle\Theta_t\left(\omega_{1j}\right), v_t\left(\omega_{1j}\right) - \bar{v}_t\left(H_t\left(\omega_{1j}\right)\right)\rangle. \end{aligned}$$

The Slater condition for problem (DCP) follows.

Assumption 6.4.1. *There exists* $(\tilde{s},\tilde{v},\tilde{\bar{v}},\tilde{\pi})$ *satisfying* (15) – (17) *and* (19) *such that*

$$\left[\underline{G}\left(\tilde{s},\tilde{v}\right)\right]\left(\omega_{1j}\right) > \frac{\sum_{k=1}^{K}\tilde{\pi}_{jk}\underline{Y}(\omega_{2k})}{P_1\left(\{\omega_{1j}\}\right)}, \quad \forall j \in \mathcal{J}.$$

Optimality conditions for problem (DCP) are next and follow from the nonlinear programming optimality conditions.

Theorem 6.4.2. *Suppose assumption 6.4.1 holds. If* $\left(\hat{s},\hat{v},\hat{\bar{v}},\hat{\pi}\right)$ *is optimal for problem (DCP) then there exists* $\hat{\lambda} \in \mathbb{R}^{J(T+1)}$ *and* $\hat{\Theta} \in \mathcal{L}_{\infty}^{T-1}\left(\Omega_1,\mathcal{F}_1,P_1\right)$ *such that*

(20)

$$L\left(\hat{s},\hat{v},\hat{\bar{v}},\hat{\pi},\hat{\lambda},\hat{\Theta}\right) = \max\left\{L\left(s,v,\bar{v},\hat{\pi},\hat{\lambda},\hat{\Theta}\right) \,:\, (15)-(17)\,,\,(19)\right\}$$

$$0 = \sum_{j=1}^{J}\langle\hat{\lambda}_j,\left[\underline{G}\left(\hat{s},\hat{v}\right)\right]\left(\omega_{1j}\right)-\frac{\sum_{k=1}^{K}\hat{\pi}_{jk}\underline{Y}\left(\omega_{2k}\right)}{P_1\left(\left\{\omega_{1j}\right\}\right)}\rangle, \quad (21)$$

$$0 = \sum_{j=1}^{J}P_1\left(\left\{\omega_{1j}\right\}\right)\sum_{t=1}^{T-1}\langle\hat{\Theta}_t\left(\omega_{1j}\right),\hat{v}_t\left(\omega_{1j}\right)-\hat{\bar{v}}_t\left(H_t\left(\omega_{1j}\right)\right)\rangle. \quad (22)$$

The dual functional for problem (DCP) is:

$$\mathrm{d}\left(\lambda,\Theta\right)=\max\left\{L\left(s,v,\bar{v},\pi,\lambda,\Theta\right) \,:\, (15)-(17)\,,\,(19)\right\}.$$

The dual to problem (DCP) is:

$$\min_{\lambda,\Theta}\left\{\mathrm{d}\left(\lambda,\Theta\right) \,:\, \lambda\geq 0\right\}. \quad (\mathrm{DCP}_D)$$

Strong duality holds between problem (DCP) and problem (DCP_D).

Theorem 6.4.3. *Suppose assumption 6.4.1 holds.*

(a) If problem (DCP) has an optimal solution then problem (DCP_D) has an optimal solution and the optimal values are equal.

(b) If problem (DCP_D) has an optimal solution $\left(\hat{\lambda},\hat{\Theta}\right)$, then any feasible solution to the optimality conditions (20) − (22).

The following decomposition result is by inspection.

Proposition 6.4.4. *The dual functional $d\left(\lambda,\Theta\right)$ decomposes to J deterministic layers*

$$\begin{aligned}
\max \quad & \sum_{t=0}^{T-1}\left[G_t\left(s_t,v_t\right)\right]\left(\omega_{1j}\right)+\left[G_T\left(s_T\right)\right]\left(\omega_{1j}\right)\\
& +\langle\lambda_j,\left[\underline{G}\left(s,v\right)\right]\left(\omega_{1j}\right)\rangle+\sum_{t=1}^{T-1}\langle\Theta_t\left(\omega_{1j}\right),v_t\left(\omega_{1j}\right)\rangle\\
\textit{s.t.} \quad & s_{t+1}\left(\omega_{1j}\right)=A_t\left(\omega_{1j}\right)s_t\left(\omega_{1j}\right)+B_t\left(\omega_{1j}\right)v_t\left(\omega_{1j}\right)+e_t\left(\omega_{1j}\right),\\
& t=0,\ldots,T-1,\\
& v_t\left(\omega_{1j}\right)\in V_t,\\
& t=0,\ldots,T-1,
\end{aligned}$$

and a non-anticipativity layer

$$\max\left\{-\sum_{j=1}^{J}P_1\left(\left\{\omega_{1j}\right\}\right)\sum_{t=1}^{T-1}\langle\Theta_t\left(\omega_{1j}\right),\bar{v}_t\left(H_t\left(\omega_{1j}\right)\right)\rangle\right\}.$$

The non-anticipativity layer induces linear constraints on problem (DCP_D):

$$\sum_{\tilde{\omega}_{1k} \in H_t^{-1}(\omega_{1j})} P_1\left(\{\tilde{\omega}_{1k}\}\right) \Theta_t\left(\tilde{\omega}_{1k}\right) = 0, \qquad t = 0, \ldots, T-1, \qquad \forall j \in \mathcal{J}.$$

CHAPTER 7

Conclusion

7.1. Remarks

This section highlights the uniting themes and summarizes the impact of this dissertation.

7.1.1. Uniting themes. We have emphasized three themes throughout this dissertation. The first is the necessary perturbation for problem (MP) that overcomes the inherent difficulty with the Slater condition. The second theme is the interpretation of the optimality conditions in terms of increasing concave functions. The third is the importance of sample average approximation for problem (MP).

In problem (SIP), it is possible to satisfy the Slater condition without a perturbation by suitably choosing U_Ξ as discussed in [**14**]. However, to satisfy the Slater condition for problem (MP) and to satisfy the Slater condition for problem (SIP) for general choices of U_Ξ, we are forced to introduce a perturbation. This perturbation turns out to be essential in Chapters 3, 5, and 6 to derive optimality conditions. The central difficulty with problem (MP), problem (RobMP), and problem (DP) is that the zero function is contained in $\mathcal{U}^N(W)$. We cannot exclude just the zero function because then we lose compactness of the constraint index set. In [**14**], the zero function is excluded along with a set of functions so that the resulting subset U_Ξ of $\mathcal{U}^N(W)$ is still compact. The choice of the constant perturbation function ϵ for all of the constraints can be adjusted. The perturbation can in principle be any continuous function on $\mathcal{C}\left(\mathcal{U}^N(W)\right)$ that is strictly positive at $u = 0$.

The optimality conditions and duality results in this dissertation are in line with earlier work in [**7, 8, 14**]. Increasing concave functions show up as the Lagrange multipliers of the increasing concave stochastic order constraints. Due to the perturbation function ϵ, there is a slight modification to the optimality conditions when compared to [**7, 8, 14**]. A new term appears that penalizes increasing concave functions with large rates of change.

The importance of sample average approximation for the problem class presented in this dissertation cannot be overestimated. Most probability distributions in practice are too large to deal with exhaustively. Further, it is necessary to use SAA to apply the transformation for problem (MP) developed in [**1**].

7.1.2. Impact. This dissertation makes several fundamental contributions. In Chapter 3, we introduce a perturbation of problem (MP) and derive conditions optimality conditions and duality results. It is shown that the increasing concave functions play the role of Lagrange multipliers for the perturbed problems with a suitable modification. In Chapter 4, we take up the issue of sample average approximation for problem (MP). We establish consistency of sample average approximation. This chapter argues for sampling from the random-variable-valued mapping and the benchmark separately. In Chapter 5 we consider the related issue of robustness against uncertainty in the underlying probability distribution. Our development in this chapter parallels [**11**]. Finally, in Chapter 6 we show that increasing concave stochastic order constraints can be used in multi-period optimization.

7.2. Future research

This section discusses future directions for this work.

7.2.1. Estimation. There is an implicit connection between Chapter 4 and Chapter 5. In Chapter 4 we used sample average approximation to estimate P based on sample data. Statistical estimators are associated with a degree of confidence, and an implied confidence interval. Instead of just using a point estimate of P in optimization, we would like to use all of the information from the confidence interval. Confidence intervals are closely related to the uncertainty sets $\mathcal{Q}$ that were considered in Chapter 5. The robust problem (RobMP) can possibly help mitigate the uncertainty in estimation. We would like to quantify this mitigation.

7.2.2. Learning. We want to combine the development in this dissertation with statistical learning strategies. Many real world systems are feedback based and must be able to incorporate new data in real time. We distinguish between learning and estimation in this way because estimation implies a single round of estimation and optimization, while learning implies alternating rounds of estimation and optimization in perpetuity.

Benchmark design also falls under the heading of learning. Instead of using a static benchmark that does not depend on data, the decision maker can select and update his benchmark over time. For many problems, it is necessary to redesign the benchmark online for realistic goal setting.

7.2.3. Network applications. There are many applications for this class of stochastic optimization problems. Our future work will emphasize design and management problems in transportation and supply networks, communication networks, and energy networks. Our objective for this array of problems will be two-fold. First, we want to design resilient networks that recover from disruption quickly. Second, we want to manage networks to ensure consistent reliability given the presence of stochastic noise. Increasing concave stochastic order constraints are an effective tool for both of these purposes and allow these goals to be addressed simultaneously.

7.2.4. Large-scale implementation. Chapter 4 and Chapter 5 mention the issue of large-scale implementation for problems (SCP) and (RobCP). This issue is also implied in Chapter 6 for Problem (DCP). Large-scale implementation for these problems is challenging in two respects. First, these problems are all nonlinear programming problems while column and row generation techniques have mainly been developed for linear programs. Second, we cannot use constraint and variable sampling as in other large-scale approaches. In semi-infinite programming, for example, it is common to sample a finite set of the constraints and only enforce these constraints. Problem (NLP) can be constructed from problem (MP) and problem (SIP) in this way. However, in problems (SCP) and (RobCP) we do not want to discard any distributional information because these problems are designed to use all available distributional information about $\underline{\mathrm{G}}(z)$ and $\underline{\mathrm{Y}}$. Problem (SCP) in particular would become somewhat redundant if we used constraint sampling after the fact, this problem was constructed by random sampling in the first place.

We turn to aggregation since we must use all of the information in problems (SCP) and (RobCP). To solve the aggregate versions of problems ($\epsilon-$SCP) and ($\epsilon-$RobCP), we must determine how to construct the partitions described in Chapters 4 and 5, and how to update them as new problem data become available.

Bibliography

[1] Benjamin Armbruster and James Luedtke. Models and formulations for multivariate dominance constrained stochastic programs. *Submitted for publication*, 2010.

[2] D. Azagra and R. Deville. Subdifferential rolle's and mean value inequality theorems. *Bull. Austral. Math. Soc.*, 56:319–329, 1997.

[3] Yakov Ben-Haim. *Info-Gap Decision Theory: Decisions Under Severe Uncertainty*. Elsevier Ltd., 2nd edition, 2006.

[4] J. Frederic Bonnans and Alexander Shapiro. *Perturbation Analysis of Optimization Problems*. Springer-Verlag New York, Inc., 2000.

[5] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[6] Tito Homem de Mello and Sanjay Mehrotra. A cutting surface method for uncertain linear programs with polyhedral stochastic dominance constraints. *SIAM Journal of Optimization*, 2009.

[7] Darinka Dentcheva and Andrzej Ruszczyński. Optimization with stochastic dominance constraints. *Society of Industrial and Applied Mathematics Journal of Optimization*, 14(2):548–566, 2003.

[8] Darinka Dentcheva and Andrzej Ruszczyński. Optimality and duality theory for stochastic optimization problems with nonlinear dominance constraints. *Mathematical Programming*, 99:329–350, 2004.

[9] Darinka Dentcheva and Andrzej Ruszczyński. Stochastic dynamic optimization with discounted stochastic dominance constraints. *Society of Industrial and Applied Mathematics Journal of Control and Optimization*, 47(5):2540–2556, 2008.

[10] Darinka Dentcheva and Andrzej Ruszczyński. Optimization with multivariate stochastic dominance constraints. *Mathematical Programming*, 117:111–127, 2009.

[11] Darinka Dentcheva and Andrzej Ruszczyński. Robust stochastic dominance and its application to risk-averse optimization. *Mathematical Programming*, 125:85–100, 2010.

[12] Rick Durrett. *Probability: Theory and Examples*. Thomson Brooks/Cole, 2005.

[13] Gerald B. Folland. *Real Analysis: Modern Techniques and Their Applications*. John Wiley and Sons, Inc., 1999.

[14] William B. Haskell, Zuo-Jun Max Shen, and J. George Shanthikumar. Optimization with a class of multivariate integral stochastic order constraints. *Submitted to Mathematical Programming*, 2011.

[15] Julia L. Higle and Suvrajeet Sen. Multistage stochastic convex programs: Duality and its implications. *Annals of Operations Research*, 142:129–146, 2006.

[16] Jian Hu, Tito Homem de Mello, and Sanjay Mehrotra. Sample average approximation of stochastic dominance constrained programs. *Accepted for publication in Mathematical Programming*, 2010.

[17] Alfred Muller. Stochastic orders generated by integrals: A unified study. *Advances in Applied Probability*, 29(2):414–428, 1997.

[18] Robert R. Phelps. *Lectures on Choquet's Theorem*. D. Van Nostrand Company, Inc., 1966.

[19] R. T. Rockafellar and R.J-B. Wets. *Variational Analysis*. Springer-Verlag, Berlin, 1998.

[20] R. Tyrrell Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[21] Walter Rudin. *Functional Analysis*. McGraw-Hill, Inc., 1991.

[22] Andrzej Ruszczyński and Alexander Shapiro. Conditional risk mappings. *Mathematics of Operations Research*, 31(3):544–561, 2006.

[23] Andrzej Ruszczyński and Alexander Shapiro. Optimization of convex risk functions. *Mathematics of Operations Research*, 31(3):433–452, 2006.

[24] Moshe Shaked and J. George Shanthikumar. *Stochastic Orders*. Springer, 2007.

[25] Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński. *Lectures on Stochastic Programming Modeling and Theory*. Society for Industrial and Applied Mathematics and the Mathematical Programming Society, 2009.

[26] Silvia Vogel. A stochastic approach to stability in stochastic programming. *Journal of Computational and Applied Mathematics*, 56:65–96, 1994.