STOCHASTIC DOMINANCE-CONSTRAINED MARKOV DECISION PROCESSES

WILLIAM B. HASKELL† AND RAHUL JAIN‡

Abstract. We are interested in risk constraints for infinite horizon discrete time Markov decision processes (MDPs). Starting with average reward MDPs, we show that increasing concave stochastic dominance constraints on the empirical distribution of reward lead to linear constraints on occupation measures. An optimal policy for the resulting class of dominance-constrained MDPs is obtained by solving a linear program. We compute the dual of this linear program to obtain average dynamic programming optimality equations that reflect the dominance constraint. In particular, a new pricing term appears in the optimality equations corresponding to the dominance constraint. We show that many types of stochastic orders can be used in place of the increasing concave stochastic order. We also carry out a parallel development for discounted reward MDPs with stochastic dominance constraints. A portfolio optimization example is used to motivate the paper.

Key words. Markov decision processes, stochastic dominance constraints, infinite-dimensional linear programs, strong duality

AMS subject classifications. 90C40, 91B30, 90C34, 49N15

DOI. 10.1137/120874679

1. Introduction. Markov decision processes (MDPs) are a natural and powerful framework for stochastic control problems. In the present paper, we take up the issue of risk constraints in MDPs. Convex analytic methods for MDPs have been successful at handling many types of constraints. Our specific goal is to find and study MDPs with risk constraints that are amenable to a convex analytic formulation. It turns out that stochastic dominance constraints are natural risk constraints for MDPs.

Convex analytic methods are well studied for Markov decision processes. The linear programming approach for MDPs is pioneered in [34] and further developed in [32]. An early survey is found in [3]. The main idea is that some MDPs can be written as convex optimization problems in terms of appropriate occupation measures. The works [7, 23, 8, 27] discuss a rigorous theory of convex optimization for MDPs with general Borel state and action spaces. Another survey on convex analytic methods for MDPs is found in [38]. Detailed monographs on Markov decision processes are found in [28, 29, 39]. Constrained MDPs can naturally be embedded in this framework. Constrained discounted MDPs are explored in [20, 21]. The book [1] is a substantial monograph on constrained MDPs. Constrained discounted MDPs in Borel spaces are analyzed in [24], and constrained average cost MDPs in Borel spaces are developed in [25]. Infinite-dimensional linear programming plays a fundamental role in both [24, 25], and the theory of infinite-dimensional linear programming is developed in [2]. The special case of constraints on expected utility in discounted MDPs is considered in [31]. MDPs with expected constraints and pathwise constraints, also called hard
constraints, are considered in [36] using convex analytic methods. An inventory system
is detailed to motivate the theoretical results.

Policies in MDPs induce stochastic processes, and Markov policies induce Markov
chains. Typically, policies are evaluated with respect to some measure of expected
reward, such as long-run average reward or discounted reward. The variation/spread/
dispersion of policies is also critical to their evaluation. Given two policies with equal
expected performance, we would prefer the one with smaller variation in some sense.
Consider a discounted portfolio optimization problem, for example. The expected
costs of an investment policy is a key performance measure; the downside
variation of an investment policy is also a key performance measure. When rewards
and costs are involved, the variation of a policy can also be called its risk.

Risk management for MDPs has been considered from many perspectives in the
literature. The paper [22] includes penalties for the variance of rewards in MDPs. The
optimal policy is obtained by solving a nonlinear programming problem in occupation
measures. In [42], the mean-variance trade-off in MDPs is further explored in a Pareto-
optimality sense. The conditional value-at-risk of the total cost in finite horizon MDPs
is constrained in [6]. It is argued that convex analytic methods do not apply to this
problem type and an offline iterative algorithm is employed to solve for the optimal
policy. Ruszczyński [40] develops Markov risk measures for finite horizon and infinite
horizon discounted MDPs. Dynamic programming equations are derived that reflect
the risk aversion, and policy iteration is shown to solve the infinite horizon problem.

Our notion of risk-constrained MDPs differs from this literature survey. We are
interested in the empirical distribution of reward, rather than in its expectation, vari-
ance, or other summary statistics. Our approach is based on stochastic orders, which
are partial orders on the space of random variables. See [37, 41] for extensive mono-
graphs on stochastic orders. Dentcheva and Ruszczyński [11, 12] use the increasing
concave stochastic order to define stochastic dominance constraints in single-stage
stochastic optimization. The increasing concave stochastic order is notable for its
connection to risk-averse decision makers, i.e., it captures the preferences of all risk-
averse decision makers. A benchmark random variable is introduced, and a concave
random variable-valued mapping is constrained to dominate the benchmark in the
increasing concave stochastic order. It is shown that increasing concave functions are
the Lagrange multipliers of the dominance constraints. The dual problem is a search
over a certain class of increasing concave functions, interpreted as utility functions,
and strong duality is established. Stochastic dominance constraints are applied to
finite horizon stochastic programming problems with linear system dynamics in [14].
Specifically, a stochastic dominance constraint is placed on a vector of state and ac-
tion dependent reward functions across the finite planning horizon. The Lagrange
multipliers of this dynamic stochastic dominance constraint are again determined to
be increasing concave functions, and strong duality holds. In contrast, we place a
stochastic dominance constraint on the empirical distribution of reward in infinite
horizon MDPs. We argue that this type of constraint comprehensively accounts for
the variation in policies in MDPs.

We make two main contributions in this paper. First, we show how to formu-
late stochastic dominance constraints for long-run average reward maximizing MDPs.
More immediately, we show that stochastic dominance-constrained MDPs can be
solved via linear programming over occupation measures. Our model is more gen-
eral than [14] because it allows for an arbitrary transition kernel and is also infinite
horizon. Also, our model is more computationally tractable than the stochastic pro-
gramming model in [14] because it leads to linear programs (LPs). Second, we apply
infinite-dimensional linear programming duality to gain more insight: the resulting
duals are similar to the linear programming form of the average reward dynamic pro-
gramming optimality equations. However, new decision variables corresponding to the
stochastic dominance constraint appear in an intuitive way. The new decision vari-
ables are increasing concave functions that price rewards. This observation parallels
the results in [11, 12, 15] and is natural because our stochastic dominance constraints
are defined in terms of increasing concave functions. The upcoming dual problems are
themselves LPs, unlike the dual problems in [11, 12, 15], which are general (nonlinear)
infinite-dimensional convex optimization problems.

This paper is organized as follows. In section 2, we consider stochastic dominance
constraints for long-run average reward maximizing MDPs. In section 3 we formulate
this problem as a static optimization problem, in fact, a linear programming problem,
in a space of occupation measures. Section 4 develops the dual for this problem
using infinite-dimensional linear programming duality and reveals the form of the
Lagrange multipliers. In section 5, we discuss a number of immediate variations and
extensions, especially the drastically simpler development on finite state and action
spaces. Section 6 develops a portfolio optimization problem, and the paper concludes
in section 7.

2. MDPs and stochastic dominance. The first subsection presents a general
model for average reward MDPs, and the second explains how to apply stochastic
dominance constraints.

2.1. Average reward MDPs. A typical representation of a discrete time MDP
is the 5-tuple

\[(S, A, \{A(s) : s \in S\}, Q, r)\].

The state space \(S\) and the action space \(A\) are Borel spaces, measurable subsets of
complete and separable metric spaces, with corresponding Borel \(\sigma\)-algebras \(B(S)\) and
\(B(A)\). We define \(\mathcal{P}(S)\) to be the space of probability measures over \(S\) with respect to
\(B(S)\), and we define \(\mathcal{P}(A)\) analogously. For each state \(s \in S\), the set \(A(s) \subset A\) is a
measurable set in \(B(A)\) and indicates the set of feasible actions available in state \(s\).

The set of feasible state-action pairs is written

\[K = \{(s, a) \in S \times A : a \in A(s)\} ,\]

and \(K\) is assumed to be closed in \(S \times A\). The transition law \(Q\) governs the system
evolution. Explicitly, \(Q(B|s, a)\) for \(B \in \mathcal{B}(S)\) is the probability of visiting the set \(B\)
given the state-action pair \((s, a)\). Finally, \(r : K \to \mathbb{R}\) is a measurable reward function
that depends on state-action pairs.

We now describe two classes of policies for MDPs. Let \(H_t\) be the set of histories at
time \(t\), \(H_0 = S\), \(H_1 = K \times S\), and \(H_t = K^t \times S\) for all \(t \geq 2\). A specific history \(h_t \in H_t\) records the state-action pairs visited at times \(0, 1, \ldots, t - 1\) and the current state \(s_t\).
Define \(\Pi\) to be the set of all history-dependent randomized policies: collections of
mappings \(\pi_t : H_t \to \mathcal{P}(A)\) for all \(t \geq 0\). Given a history \(h_t \in H_t\) and a set \(B \in \mathcal{B}(A)\),
\(\pi_t(B| h_t)\) is the probability of selecting an action in \(B\). Define \(\Phi\) to be the class of
stationary randomized Markov policies: mappings \(\phi : S \to \mathcal{P}(A)\) which only depend
on history through the current state. For a given state \(s \in S\) and a set \(B \in \mathcal{B}(A)\),
\(\phi(B| s)\) is the probability of choosing an action in \(B\). The class \(\Phi\) will be viewed as
a subset of \(\Pi\). We explicitly assume that both \(\Pi\) and \(\Phi\) only include feasible policies
that respect the constraints \(K\).
The state and action at time $t$ are denoted $s_t$ and $a_t$, respectively. With respect to an initial distribution $\nu \in \mathcal{P}(S)$, any policy $\pi \in \Pi$ determines a probability measure $P^\pi_\nu$ and stochastic process $\{(s_t, a_t), t \geq 0\}$ defined on a measurable space $(\Omega, \mathcal{F})$. The expectation operator with respect to $P^\pi_\nu$ is denoted $\mathbb{E}^\pi_\nu[\cdot]$. Consider the long-run expected average reward

$$ R(\pi, \nu) = \liminf_{T \to \infty} \frac{1}{T} \mathbb{E}^\pi_\nu \left[ \sum_{t=0}^{T-1} r(s_t, a_t) \right]. $$

The classic long-run expected average reward maximization problem is

$$ \sup_{(\pi, \nu) \in \Pi \times \mathcal{P}(S)} R(\pi, \nu). $$

It is known that a stationary policy in $\Phi$ is optimal for problem (2.1) under suitable conditions. (This result is found in [39] for finite and countable state spaces and in [28, 29] for general Borel state and action spaces.)

### 2.2. Stochastic dominance.

Now we will motivate and formalize stochastic dominance constraints for problem (2.1). To begin, let $z : K \to \mathbb{R}$ be another measurable reward function, possibly different from $r$. A risk-averse decision maker with an increasing concave utility function $u : \mathbb{R} \to \mathbb{R}$ would be interested in maximizing his long-run average expected utility

$$ \liminf_{T \to \infty} \frac{1}{T} \mathbb{E}^\pi_\nu \left[ \sum_{t=0}^{T-1} u(z(s_t, a_t)) \right]. $$

However, it is difficult to choose one utility function to represent a risk-averse decision maker without considerable information. We will use the increasing concave order to express a continuum of risk preferences in MDPs.

**Definition 2.1.** For random variables $X, Y \in \mathbb{R}$, $X$ dominates $Y$ in the increasing concave stochastic order, written $X \geq_{icv} Y$, if $\mathbb{E}[u(X)] \geq \mathbb{E}[u(Y)]$ for all increasing concave functions $u : \mathbb{R} \to \mathbb{R}$ such that both expectations exist.

Let $C(\mathbb{R})$ be the set of all continuous functions $f : \mathbb{R} \to \mathbb{R}$. Let $\mathcal{U}(\mathbb{R}) \subset C(\mathbb{R})$ be the set of all increasing concave functions $u : \mathbb{R} \to \mathbb{R}$ such that

$$ \lim_{x \to \infty} u(x) = 0 $$

and

$$ u(x) = u(x_0) + \kappa(x - x_0) $$

for all $x \leq x_0$ for some $\kappa > 0$ and $x_0 \in \mathbb{R}$. (The choices of $\kappa$ and $x_0$ differ among $u$.)

The second condition just means that all $u \in \mathcal{U}(\mathbb{R})$ become linear as $x \to -\infty$. By construction, functions $u \in \mathcal{U}(\mathbb{R})$ are bounded from above by zero. We will use the set $\mathcal{U}(\mathbb{R})$ to characterize $X \geq_{icv} Y$.

Now define $(x)_- \equiv \min\{x, 0\}$. We note that any function in $\mathcal{U}(\mathbb{R})$ can be written in terms of the family $\{(x - \eta)_- : \eta \in \mathbb{R}\}$. To understand this result, choose $u \in \mathcal{U}(\mathbb{R})$ and a finite set of points $\{x_1, \ldots, x_j\}$. By concavity, there exist $a_i \in \mathbb{R}$ such that $a_i(x - x_i) + u(x_i) \geq u(x)$ for all $x \in \mathbb{R}$ and for all $i = 1, \ldots, j$. Each linear function $a_i(x - x_i) + u(x_i)$ is a global overestimator of $u$. The piecewise linear increasing concave function

$$ \min_{i=1,\ldots,j} \{a_i(x - x_i) + u(x_i)\} $
is also a global overestimator of \( u \), and certainly

\[
  u(x) \leq \min_{i=1,\ldots,j} \{ a_i(x-x_i) + u(x_i) \} \leq a_i(x-x_i) + u(x_i)
\]

for all \( i = 1,\ldots,j \) and \( x \in \mathbb{R} \). As the number of sample points \( j \) increases, the polyhedral concave function \( \min_{i=1,\ldots,j} \{ a_i(x-x_i) + u(x_i) \} \) becomes a better approximation of \( u \). We realize that the function \( \min_{i=1,\ldots,j} \{ a_i(x-x_i) + u(x_i) \} \) is equal to a finite sum of nonnegative scalar multiples of functions from \( \{(x-\eta)_- : \eta \in \mathbb{R} \} \). It follows that the relation \( X \geq_{icv} Y \) is equivalent to \( \mathbb{E}[(X-\eta)_-] \geq \mathbb{E}[(Y-\eta)_-] \) for all \( \eta \in \mathbb{R} \). This result is related to integral stochastic orders; a detailed proof of this result can be found in [37], for example. When the support of \( Y \) is contained in a compact interval \([a, b]\), the condition \( \mathbb{E}[(X-\eta)_-] \geq \mathbb{E}[(Y-\eta)_-] \) for all \( \eta \in [a, b] \) is sufficient for \( X \geq_{icv} Y \) (see [12]).

From now on, let \( Y \) be a fixed reference random variable on \( \mathbb{R} \) to benchmark the empirical distribution of reward \( z \). We assume that \( Y \) has support in a compact interval \([a, b]\) throughout the rest of this paper. Define

\[
  Z_\eta(\pi, \nu) \triangleq \liminf_{T \to \infty} \frac{1}{T} \mathbb{E}_\nu^\pi \left[ \sum_{t=0}^{T-1} (z(s_t, a_t) - \eta)_- \right]
\]

to be the long-run expected average shortfall in \( z \) at level \( \eta \). We propose the class of stochastic dominance-constrained MDPs:

\[
\begin{align*}
  (2.2) \quad & \sup_{(\pi, \nu) \in \Pi \times \mathcal{P}(S)} R(\pi, \nu) \\
  & \text{s.t.} \quad Z_\eta(\pi, \nu) \geq \mathbb{E}[(Y-\eta)_-] \quad \forall \eta \in [a, b].
\end{align*}
\]

For emphasis, we index \( \eta \) over the compact set \([a, b]\) in (2.3). Allowing \( \eta \) to range over all \( \mathbb{R} \) would lead to major technical difficulties, as first observed in [11, 12].

Constraint (2.3) is a continuum of constraints on the long-run expected average shortfall of the policy \( \pi \) for all \( \eta \in [a, b] \). We will approach problem (2.2)–(2.3) by casting it in the space of long-run average occupation measures.

We can denote the feasible region of problem (2.2)–(2.3) succinctly as

\[
  \Delta \triangleq \{ (\pi, \nu) \in \Pi \times \mathcal{P}(S) : R(\pi, \nu) > -\infty \text{ and } Z_\eta(\pi, \nu) \geq \mathbb{E}[(Y-\eta)_-] \quad \forall \eta \in [a, b] \},
\]

allowing problem (2.2)–(2.3) to be written as

\[
  \rho^* \triangleq \sup \{ R(\pi, \nu) : (\pi, \nu) \in \Delta \},
\]

where \( \rho^* \) is the optimal value.

\textbf{Remark 2.2.} We focus on the average reward case in this paper. The extension to the average cost case is immediate. Let \( c : K \to \mathbb{R} \) be a measurable cost function. The long-run expected average cost is

\[
  C(\pi, \nu) = \limsup_{T \to \infty} \frac{1}{T} \mathbb{E}_\nu^\pi \left[ \sum_{t=0}^{T-1} c(s_t, a_t) \right].
\]

Similarly, let \( z : K \to \mathbb{R} \) be another measurable cost function that possibly differs from \( c \). Since \( z \) represents costs, we want the empirical distribution of \( z \) to be “small” in a stochastic sense. For costs, it is logical to use the increasing convex order rather
than the increasing concave order. For random variables $X, Y \in \mathbb{R}$, $X$ dominates $Y$ in the increasing convex stochastic order, written $X \geq_{icx} Y$, if $E[f(X)] \geq E[f(Y)]$ for all increasing convex functions $f : \mathbb{R} \to \mathbb{R}$ such that both expectations exist. Define $(x)_+ \triangleq \max\{x, 0\}$, and recall that the relation $X \geq_{icx} Y$ is equivalent to $E[(X - \eta)_+] \geq E[(Y - \eta)_+]$ for all $\eta \in \mathbb{R}$. When the support of $Y$ is contained in an interval $[a, b]$, the relation $X \geq_{icx} Y$ is equivalent to $E[(X - \eta)_+] \geq E[(Y - \eta)_+]$ for all $\eta \in [a, b]$.

Momentarily, let $Y$ be a benchmark random variable that we require to dominate the empirical distribution of $\mathbf{z}$. Define

$$3_\eta (\pi, \nu) \triangleq \limsup_{T \to \infty} \frac{1}{T} \mathbb{E}_\nu \left[ \sum_{t=0}^{T-1} (\mathbf{z}_t(s_t, a_t) - \eta)_+ \right]$$

for all $\eta \in [a, b]$. We obtain the cost minimization problem

$$\inf_{(\pi, \nu) \in \Pi \times \mathcal{P}(S)} C(\pi, \nu) \quad \text{s.t.} \quad 3_\eta (\pi, \nu) \leq E[(Y - \eta)_+] \quad \forall \eta \in [a, b].$$

The upcoming results of this paper all have immediate analogues for the average cost case.

3. A linear programming formulation. This section develops problem (2.2)–(2.3) as an infinite-dimensional LP. First, we discuss occupation measures on the set $K$. Occupation measures on $K$ can be interpreted as the long-run average expected number of visits of a stochastic process $\{(s_t, a_t), t \geq 0\}$ to each state-action pair. Next, we argue that a stationary policy in $\Phi$ is optimal for problem (2.2)–(2.3). It will follow that the functions $R(\phi, \nu)$ and $Z_\eta(\phi, \nu)$ can be written as linear functions of the occupation measure corresponding to $\phi$ and $\nu$. These linear functions give us the desired LP.

To proceed, we recall several well-known results in convex analytic methods for MDPs. We will use $\mu$ to denote probability measures on $K$ and the set of all probability measures on $K$ is denoted $\mathcal{P}(K)$. Probability measures on $K$ can be equivalently viewed as probability measures on all of $S \times A$ with all mass concentrated on $K$, $\mu(K) = 1$. For any $\mu \in \mathcal{P}(K)$, the marginal of $\mu$ on $S$ is the probability measure $\hat{\mu} \in \mathcal{P}(S)$ defined by $\hat{\mu}(B) = \mu(B \times A)$ for all $B \in \mathcal{B}(S)$.

The following fact is ubiquitous in the literature on convex analytic methods for MDPs (see [17], for example). If $\mu$ is a probability measure on $K$, then there exists a stationary randomized Markov policy $\phi \in \Phi$ such that $\mu$ can be disintegrated as $\mu = \hat{\mu} \cdot \phi$, where $\hat{\mu}$ is the marginal of $\mu$. Specifically, $\mu = \hat{\mu} \cdot \phi$ is defined by

$$\mu(B \times C) = \int_B \phi(C|s) \hat{\mu}(ds)$$

for all $B \in \mathcal{B}(S)$ and $C \in \mathcal{B}(A)$.

We can integrate measurable functions $f : K \to \mathbb{R}$ with respect to measures $\mu \in \mathcal{P}(K)$. Define

$$\langle \mu, f \rangle \triangleq \int_K f(s,a) \mu(d(s,a))$$

as the integral of $f$ over state-action pairs $(s,a) \in K$ with respect to $\mu$. We are assuming the integral $\langle \mu, f \rangle$ is well defined throughout this paper. Then

$$\langle \mu, r \rangle = \int_K r(s,a) \mu(d(s,a))$$
is the expected reward with respect to the probability measure \( \mu \) and
\[
\langle \mu, (z - \eta)_- \rangle = \int_K (z(s, a) - \eta)_- \mu(d(s, a))
\]
is the expected shortfall in \( z \) at level \( \eta \) with respect to the probability measure \( \mu \). Notice that \( r \) and \( (z - \eta)_- \) are both nonpositive, so the preceding two integrals are well defined.

We need to restrict to a certain class of probability measures. For notational convenience, define \( r(s, \phi) \triangleq \int_A r(s, a) \phi(da|s) \) and \( Q(\cdot|s, \phi) \triangleq \int_A Q(\cdot|s, a) \phi(da|s) \).

**Definition 3.1** (see [25, Definition 3.4]). A probability measure \( \mu = \hat{\mu} \cdot \phi \) is called stable if
\[
\langle \mu, r \rangle = \int r(s, a) \mu(d(s, a)) > -\infty
\]
and the marginal \( \hat{\mu} \) is invariant with respect to \( Q(\cdot|\cdot, \phi) \), i.e., \( \hat{\mu}(B) = \int_S Q(B|s, \phi) \hat{\mu}(ds) \) for all \( B \in \mathcal{B}(S) \).

When \( \mu \) is stable, the long-run expected average cost \( R(\phi, \hat{\mu}) \) is
\[
R(\phi, \hat{\mu}) = \liminf_{T \to \infty} \frac{1}{T} \mathbb{E}^0_\mu \left[ \sum_{t=0}^{T-1} r(s_t, a_t) \right] = \langle \mu, r \rangle
\]
by the individual ergodic theorem [43, p. 388, Theorem 6]. Then for stable \( \mu = \hat{\mu} \cdot \phi \in \mathcal{P}(K) \), it follows that
\[
R(\phi, \hat{\mu}) = \langle \mu, r \rangle = \int_S r(s, \phi) \hat{\mu}(ds).
\]

Similarly, for stable \( \mu = \hat{\mu} \cdot \phi \), it is true that
\[
Z_\eta(\phi, \hat{\mu}) = \langle \mu, (z - \eta)_- \rangle = \int_S \left[ \int_A (z(s, a) - \eta)_- \phi(da|s) \right] \hat{\mu}(ds)
\]
for all \( \eta \in [a, b] \).

There is a connection between problem (2.2)–(2.3) and stable policies. Let \( I_\Gamma \) be the indicator function of a set \( \Gamma \) in \( \mathcal{B}(K) \), where \( \mathcal{B}(K) \) is the Borel \( \sigma \)-algebra on \( K \), and define the \( T \)-stage expected occupation measure \( \mu^\pi_{\nu,T} \) on \( K \) via
\[
\mu^\pi_{\nu,T}(\Gamma) = \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}^\pi_\nu \left[ I_\Gamma(s_t, a_t) \right] = \frac{1}{T} \sum_{t=0}^{T-1} P^\pi_\nu \{ (s_t, a_t) \in \Gamma \}
\]
for all \( \Gamma \in \mathcal{B}(K) \). Then,
\[
R(\pi, \nu) = \liminf_{T \to \infty} \frac{1}{T} \mathbb{E}^\pi_\nu \left[ \sum_{t=0}^{T-1} r(s_t, a_t) \right] = \liminf_{T \to \infty} \langle \mu^\pi_{\nu,T}, r \rangle
\]
and
\[
Z_\eta(\pi, \nu) = \liminf_{T \to \infty} \frac{1}{T} \mathbb{E}^\pi_\nu \left[ \sum_{t=0}^{T-1} (z(s_t, a_t) - \eta)_- \right] = \liminf_{T \to \infty} \langle \mu^\pi_{\nu,T}, (z - \eta)_- \rangle
\]
for all \( \eta \in [a, b] \).
With this notation, we can now interpret constraint (2.3) as an explicit stochastic dominance constraint on a well-defined random variable. We introduce the measurable mapping \( Z(\pi, \nu) : K \to \mathbb{R} \) for all \((\pi, \nu) \in \Pi \times \mathcal{P}(S)\) via the convention
\[
[Z(\pi, \nu)](s, a) \triangleq z(s, a),
\]
for all \((s, a) \in K\). The definition of the mapping above does not depend on \((\pi, \nu)\), however, the underlying probability distribution on \(K\) does depend on \((\pi, \nu)\) - in fact, it is the long-run average occupation measure \(\mu^\pi_\nu\) defined by
\[
\mu^\pi_\nu(\Gamma) \triangleq \liminf_{T \to \infty} \mu^\pi_\nu,T(\Gamma)
\]
for all \(\Gamma \in \mathcal{B}(K)\). Then \(Z(\pi, \nu)\) can be formally defined on the probability space \((K, \mathcal{B}(K), \mu^\pi_\nu)\), and we obtain the explicit stochastic dominance constraint
\[
Z(\pi, \nu) \geq icvY.
\]

To continue, we introduce some technical assumptions for the rest of the paper. Let \(\mathcal{C}_b(K)\) be the space of continuous and bounded functions on \(K\). The transition law \(Q\) is defined to be weakly continuous when
\[
\int_S h(\xi)Q(d\xi|\cdot) \in C_b(K)
\]
for all \(h \in C_b(S)\).

**Assumption 3.2.**

(a) Problem (2.2)–(2.3) is consistent, i.e., the set \(\Delta\) is nonempty.

(b) The reward function \(r\) is nonpositive, and for any \(\epsilon \geq 0\) the set \(\{(s, a) \in S \times A : r(s, a) \geq -\epsilon\}\) is compact.

(c) The function \(z\) is bounded and upper semicontinuous on \(S \times A\).

(d) The transition law \(Q\) is weakly continuous.

A function \(f\) on \(K\) is called a moment if there exists a nondecreasing sequence of compact sets \(K_n \uparrow K\) such that
\[
\lim_{n \to \infty} \inf_{(s, a) \notin K_n} f(s, a) = \infty;
\]
see [28, Definition E.7]. When \(K\) is compact, then any function on \(K\) is a moment. Assumption 3.2(b) implies that \(-r\) is a moment. By construction, all the functions \((z(s, a) - \eta) - \) are bounded above by zero on \(S \times A\) for all \(\eta \in [a, b]\).

The next lemma reduces the search for optimal policies to stable policies. We define
\[
\Delta_\phi \triangleq \{\mu \in \mathcal{P}(K) : \mu\text{ is stable, }\mu = \bar{\mu} \cdot \phi \text{ and } (\phi, \bar{\mu}) \in \Delta\}
\]
to be the set of all stable probability measures \(\mu\) that are feasible for problem (2.2)–(2.3).

**Lemma 3.3.** Suppose Assumption 3.2 holds. For each feasible pair \((\pi, \nu) \in \Delta\), there exists a stable probability measure \(\mu = \bar{\mu} \cdot \phi\) such that \((\phi, \bar{\mu}) \in \Delta\) and \(R(\pi, \nu) \leq R(\phi, \bar{\mu}) = \langle \mu, r \rangle\).

**Proof.** For any \((\pi, \nu) \in \Delta\), there exists a stable policy \(\mu = \nu \cdot \phi\) such that
\[
R(\pi, \nu) \leq R(\phi, \bar{\mu}) = \langle \mu, r \rangle
\]
by [28, Lemma 5.7.10]. By the same reasoning,
\[
\mathbb{E}[(Y - \eta) -] \leq Z_\eta(\pi, \nu) \leq Z_\eta(\phi, \bar{\mu}) = \langle \mu, (z - \eta) - \rangle
\]
for all \(\eta \in [a, b]\) so that \(\mu = \bar{\mu} \cdot \phi\) is feasible. \(\square\)
Problem (2.2)–(2.3) is solvable if there exists a pair $(\pi^*, \nu^*) \in \Delta$ with $R(\pi^*, \nu^*) = \rho^*$, i.e., the optimal value is attained. When an optimization problem is solvable, we can replace “sup” with “max.” We use the preceding lemma to show that problem (2.2)–(2.3) is solvable.

**Theorem 3.4.** Problem (2.2)–(2.3) is solvable.

**Proof.** By Lemma 3.3,

\[
\rho^* = \sup \{ \langle \mu, r \rangle : \mu \in \Delta_s \}.
\]

Now apply the proof of [28, Theorem 5.7.9]. Let $\{\epsilon_n\}$ be a sequence with $\epsilon_n \downarrow 0$ and $\epsilon_n \leq 1$. For any $\epsilon_n$, there is a pair $(\pi^n, \nu^n) \in \Delta$ with $R(\pi^n, \nu^n) \geq \rho^* - \epsilon_n$ by the definition of $\rho^*$. Again, by Lemma 3.3, for each $(\pi^n, \nu^n) \in \Delta$ there is a pair $(\phi^n, \hat{\mu}^n) \in \Delta$ such that $\mu^n = \hat{\mu}^n \cdot \phi^n$ is stable and $R(\pi^n, \nu^n) \leq R(\phi^n, \hat{\mu}^n) = \langle \mu^n, r \rangle$.

By construction, $\langle \mu^n, r \rangle \geq \rho^* - \epsilon_n$ and $\epsilon_n \in (0, 1)$ for all $n$, so $\inf_n \langle \mu^n, r \rangle \geq \rho^* - 1$. It follows that $\sup_n \langle \mu^n, -r \rangle \leq 1 - \rho^*$. Since $-r$ is a moment, the preceding inequality along with [28, Proposition E.8] and [28, Proposition E.6] imply that there exists a subsequence of measures $\{\mu^{n_i}\}$ converging weakly to a measure $\mu$ on $K$. Now

\[
\rho^* \leq \limsup_{i \to \infty} \langle \mu^{n_i}, r \rangle
\]

holds since $\langle \mu^n, r \rangle \geq \rho^* - \epsilon_n$ for all $n$ and $\epsilon_n \downarrow 0$. By [28, Proposition E.2],

\[
\limsup_{i \to \infty} \langle \mu^{n_i}, r \rangle \leq \langle \mu, r \rangle,
\]

so we obtain

\[
\rho^* \leq \langle \mu, r \rangle.
\]

Since $\langle \mu, r \rangle \leq \rho^*$ must hold by definition of $\rho^*$, the preceding inequality shows that $\langle \mu, r \rangle = \rho^*$, i.e., $\mu$ attains the optimal value $\rho^*$ and is stable. By a similar argument,

\[
E[(Y - \eta)_-] \leq \limsup_{i \to \infty} \langle \mu^{n_i}, (z - \eta)_- \rangle \leq \langle \mu, (z - \eta)_- \rangle
\]

since each $\langle \mu^{n_i}, (z - \eta)_- \rangle \geq E[(Y - \eta)_-]$ for all $i$ and all $\eta \in [a, b]$. Thus, $\mu$ is feasible.

Let $\mu^*$ be the optimal stable measure just guaranteed, and disintegrate to obtain $\mu^* = \hat{\mu}^* \cdot \phi^*$. The pair $(\phi^*, \hat{\mu}^*)$ is then optimal for problem (2.2)–(2.3) since

\[
R(\phi^*, \hat{\mu}^*) = \langle \mu^*, r \rangle = \rho^*
\]

and

\[
Z_\eta(\phi^*, \hat{\mu}^*) = \langle \mu^*, (z - \eta)_- \rangle \geq E[(Y - \eta)_-]
\]

for all $\eta \in [a, b]$. \qed

From the preceding theorem, we can now write “max” instead of “sup” in the objective of problem (2.2)–(2.3),

\[
\rho^* \triangleq \max \{ R(\pi, \nu) : (\pi, \nu) \in \Delta_s \}.
\]

We are now ready to formalize problem (2.2)–(2.3) as an LP. Introduce the weight function

\[
w(s, a) = 1 - r(s, a)
\]
on $K$. Under our assumption that $r$ is nonpositive, $w$ is bounded from below by one. The space of signed Borel measures on $K$ is denoted $\mathcal{M}(K)$. With the preceding weight function, define $\mathcal{M}_w(K)$ to be the space of signed measures $\mu$ on $K$ such that

$$\|\mu\|_{\mathcal{M}_w(K)} \triangleq \int_K w(s,a) |\mu(d(s,a))| < \infty.$$ 

If $\|\mu\|_{\mathcal{M}_w(K)} < \infty$, then certainly

$$\langle \mu, r \rangle = \int_K r(s,a) \mu(d(s,a)) > -\infty$$

since $1 - r = w$.

We can identify stable policies as elements of $\mathcal{M}_w(K)$. If $\mu$ is a stable probability measure, then it is an element of $\mathcal{M}_w(K)$ since

$$\int_K w(s,a) |\mu(d(s,a))| = \int_K (1 - r(s,a)) \mu(d(s,a)) = \mu(K) - \langle \mu, r \rangle < \infty.$$ 

Also define the weight function

$$\hat{w}(s) = 1 - \sup_{a \in A(s)} r(s,a)$$

on $S$ which is bounded from below by one as well. The space $\mathcal{M}_w(S)$ is defined analogously with $\hat{w}$ and $S$ in place of $w$ and $S \times A$.

The topological dual of $\mathcal{M}_w(K)$ is $\mathcal{F}_w(K)$, the vector space of measurable functions $h : K \to \mathbb{R}$ such that

$$\|h\|_{\mathcal{F}_w(K)} \triangleq \sup_{(s,a) \in K} \frac{|h(s,a)|}{w(s,a)} < \infty.$$ 

Certainly, $r \in \mathcal{F}_w(K)$ by definition of $w$ since

$$\|r\|_{\mathcal{F}_w(K)} = \sup_{(s,a) \in K} \frac{|r(s,a)|}{w(s,a)} = \sup_{(s,a) \in K} \frac{|r(s,a)|}{1 + |r(s,a)|} \leq 1.$$ 

Every element $h \in \mathcal{F}_w(K)$ induces a continuous linear functional on $\mathcal{M}_w(K)$ defined by

$$\langle \mu, h \rangle \triangleq \int_K h(s,a) \mu((d(s,a))).$$ 

The two spaces $(\mathcal{M}_w(K), \mathcal{F}_w(K))$ are called a dual pair, and the duality pairing is the bilinear form $\langle u, h \rangle : \mathcal{M}_w(K) \times \mathcal{F}_w(K) \to \mathbb{R}$ just defined. The topological dual of $\mathcal{M}_w(S)$ is $\mathcal{F}_w(S)$, which is defined analogously with $S$ and $\hat{w}$ in place of $K$ and $w$.

We now make some additional technical assumptions.

**Assumption 3.5.**

(a) The function $(z - \eta)_-$ is an element of $\mathcal{F}_w(K)$ for all $\eta \in [a,b]$.

(b) The function $\int_S \hat{w}(\xi)Q(d\xi | s,a) : S \times A \to \mathbb{R}$ is an element of $\mathcal{F}_w(K)$.

Notice that Assumption 3.5(a) is satisfied if $z \in \mathcal{F}_w(K)$. To see this fact, reason that

$$\| (z - \eta)_- \|_{\mathcal{F}_w(K)} \leq \| z - \eta \|_{\mathcal{F}_w(K)} \leq \| z \|_{\mathcal{F}_w(K)} + \| \eta \|_{\mathcal{F}_w(K)},$$
where the first inequality follows from $|(z - \eta)_-| \leq |z - \eta|$. The constant function $f(x) = \eta$ on $K$ is in $\mathcal{F}_w(K)$ since

$$||\eta||_{\mathcal{F}_w(K)} = \sup_{(s,a) \in K} \frac{|\eta|}{w(s,a)} \leq |\eta|.$$ 

The linear mapping $L_0 : \mathcal{M}_w(K) \rightarrow \mathcal{M}_w(S)$ defined by

$$(3.1) \quad [L_0 \mu](B) \triangleq \hat{\mu}(B) - \int_K Q(B|s,a) \mu(d(s,a)) \quad \forall B \in \mathcal{B}(S),$$

is used to verify that $\mu$ is an invariant probability measure on $K$ with respect to $Q$. The mapping (3.1) appears in all work on convex analytic methods for long-run average reward/cost MDPs. When $L_0 \mu(B) = 0$, it means that the long-run proportion of time in state $B$ is equal to the rate at which the system transitions to state $B$ from all state-action pairs $(s,a) \in K$.

**Lemma 3.6.** The condition $\mu \in \Delta_\alpha$ is equivalent to $\langle \mu, r \rangle > -\infty$ and

- $L_0 \mu = 0$,
- $\langle \mu, 1 \rangle = 1$,
- $\langle \mu, (z - \eta)_- \rangle \geq \mathbb{E}[(Y - \eta)_-] \quad \forall \eta \in [a,b], \quad \mu \geq 0$.

**Proof.** The linear constraints $\langle \mu, 1 \rangle = \int_K \mu(d(s,a)) = 1$ and $\mu \geq 0$ just ensure that $\mu$ is a probability measure on $K$. The condition $L_0 \mu = 0$ is equivalent to invariance of $\mu$ with respect to $Q$. For stable $\mu = \hat{\mu} \cdot \phi$, $R(\phi, \hat{\mu}) = \langle \mu, r \rangle > -\infty$ and $Z_{\eta}(\phi, \hat{\mu}) = \langle \mu, (z - \eta)_- \rangle$. Since $Z_{\eta}(\phi, \hat{\mu}) \geq \mathbb{E}[(Y - \eta)_-]$ for all $\eta \in [a,b]$, the conclusion follows.

Next we continue with the representation of the dominance constraints (2.3). We would like to express the constraints $\langle \mu, (z - \eta)_- \rangle \geq \mathbb{E}[(Y - \eta)_-]$ for all $\eta \in [a,b]$ through a single linear operator.

**Lemma 3.7.** For any $\mu \in \mathcal{P}(K)$, $\langle \mu, (z - \eta)_- \rangle$ is uniformly continuous in $\eta$ on $[a,b]$.

**Proof.** Write $\langle \mu, (z - \eta)_- \rangle = \int_K (z(s,a) - \eta)_- \mu(d(s,a))$. Certainly, each function $(z(s,a) - \eta)_-$ is continuous in $\eta$ for fixed $s \times a$. Choose $\epsilon > 0$ and $|\eta' - \eta| < \epsilon$. Then

$$|(z(s,a) - \eta')_- - (z(s,a) - \eta)_-| \leq |z(s,a) - \eta' - z(s,a) + \eta| \leq \epsilon$$

by definition of $(x)_-$. It follows that

$$\left| \int_K (z(s,a) - \eta')_- \mu(d(s,a)) - \int_K (z(s,a) - \eta)_- \mu(d(s,a)) \right| \leq \left| \int_K \epsilon \mu(d(s,a)) \right| = \epsilon,$$

since $\mu$ is a probability measure.
The preceding lemma allows us to write the dominance constraints (2.3) as a linear operator in the space of continuous functions. Recall that we have assumed \([a, b]\) to be a compact set. Let \(C([a, b])\) be the space of continuous functions on \([a, b]\) in the supremum norm, 
\[
\|f\|_{C([a, b])} = \sup_{a \leq x \leq b} |f(x)|
\]
for \(f \in C([a, b])\). The topological dual of \(C([a, b])\) is \(M([a, b])\), the space of finite signed Borel measures on \([a, b]\). Every measure \(\Lambda \in M([a, b])\) induces a continuous linear functional on \(C([a, b])\) through the bilinear form 
\[
\langle \Lambda, f \rangle = \int_{a}^{b} f(\eta) \Lambda(d\eta).
\]
Define the linear operator \(L_1 : M_w(K) \rightarrow C([a, b])\) by
\[
[L_1 \mu](\eta) = \langle \mu, (z - \eta)_- \rangle \quad \forall \eta \in [a, b].
\]
Also define the continuous function \(y \in C([a, b])\), where 
\[
y(\eta) = E[(Y - \eta)_-] \quad \text{is the shortfall in } Y \text{ at level } \eta \text{ for all } \eta \in [a, b].
\]
The dominance constraints are then equivalent to 
\[
[L_1 \mu](\eta) \geq y(\eta) \quad \forall \eta \in [a, b],
\]
which can be written as the single inequality 
\[
L_1 \mu \geq y \quad \text{in } C([a, b]).
\]
The linear programming form of problem (2.2)–(2.3) is
\[
\begin{align*}
\text{(3.3)} & \quad \max \quad \langle \mu, r \rangle \\
\text{(3.4)} & \quad \text{s.t.} \quad L_0 \mu = 0, \\
\text{(3.5)} & \quad \langle \mu, 1 \rangle = 1, \\
\text{(3.6)} & \quad L_1 \mu \geq y, \\
\text{(3.7)} & \quad \mu \in M_w(K), \quad \mu \geq 0.
\end{align*}
\]
Since \(\rho^* = \max\{R(\pi, \nu) : (\pi, \nu) \in \Delta_s\} \), and stable probability measures on \(K\) can be identified as elements of \(M_w(K)\), problem (3.3)–(3.7) is equivalent to problem (2.2)–(2.3).

4. Computing the dual. In this section we compute the dual of the infinite linear programming problem (3.3)–(3.7). The development in [2] is behind our duality development and the duality theory for linear programming for MDPs on Borel spaces in general.

We will introduce Lagrange multipliers for constraints (3.4), (3.5), and (3.6); each Lagrange multiplier is drawn from the appropriate topological dual space. We introduce Lagrange multipliers \(h \in \mathcal{F}_w(S)\) for constraint (3.4). The constraint \(\langle \mu, 1 \rangle = 1\) is an equality in \(\mathbb{R}\), so we introduce Lagrange multipliers \(\beta \in \mathbb{R}\) for constraint (3.5). Finally, we introduce Lagrange multipliers \(\Lambda \in M([a, b])\) for constraints (3.6). The Lagrange multipliers \((h, \beta, \Lambda) \in \mathcal{F}_w(S) \times \mathbb{R} \times M([a, b])\) will be the decision variables in the upcoming dual to problem (3.3)–(3.7).

To proceed with duality, we compute the adjoints of \(L_0\) and \(L_1\). The adjoint is analogous to the transpose for linear operators in Euclidean spaces.

**Lemma 4.1.** (a) The adjoint of \(L_0\) is \(L_0^* : \mathcal{F}_w(S) \rightarrow \mathcal{F}_w(K)\), where
\[
[L_0^* h](s, a) \triangleq h(s) - \int_S h(\xi) Q(d\xi | s, a)
\]
for all \((s, a) \in K\).
The adjoint of $L_1$ is $L_1^\ast : \mathcal{M}([a,b]) \to \mathcal{F}_w(K)$, where

$$[L_1^\ast \Lambda](s,a) = \int_a^b (z(s,a) - \eta)_- \Lambda(d(s,a)).$$

**Proof.** (a) This result is well known; see [28, 29].

(b) Write

$$\langle \Lambda, L_1 \mu \rangle = \int_a^b \left[ \int_K (z(s,a) - \eta)_- \mu(d(s,a)) \right] \Lambda(d\eta).$$

When $z$ is bounded on $K$, then

$$\left| \int_K (z(s,a) - \eta)_- (\mu \times \Lambda)(d((s,a) \times \eta)) \right|$$

$$= \left| \int_K \frac{(z(s,a) - \eta)_-}{w(s,a)} w(s,a)(\mu \times \Lambda)(d((s,a) \times \eta)) \right|$$

$$\leq \|z - \eta\| \|\mathcal{F}_w(K)\| \|\mu\| \|\mathcal{M}_w(K)\| \|\Lambda\| \mathcal{M}([a,b])$$

$$< \infty,$$

since $\|\mu\| \mathcal{M}(K) = 1$ and $\|\Lambda\| \mathcal{M}([a,b]) < \infty$. The Fubini theorem applies to justify interchanging the order of integration,

$$\langle \Lambda, L_1 x \rangle = \int_a^b \left[ \int_K (z(s,a) - \eta)_- \mu(d(s,a)) \right] \Lambda(d\eta)$$

$$= \int_K \int_a^b (z(s,a) - \eta)_- \Lambda(d\eta) \mu(d(s,a))$$

$$= \int_K \langle \Lambda, (z(s,a) - \eta)_- \mu(d(s,a)) \rangle,$$

revealing $L_1^\ast : \mathcal{M}([a,b]) \to \mathcal{F}_w(K)$. □

We compute the dual to problem (3.3)–(3.7) in the next theorem. Our proof is based on an interchange of max and inf; see [33, 5] for more about duality in topological vector spaces. Also see [2] for duality in infinite-dimensional linear programming.

**Theorem 4.2.** The dual to problem (3.3)–(3.7) is

$$\inf_{\beta \geq 0} \beta - \langle \Lambda, y \rangle$$

s.t. $r + L_0^\ast h - \beta 1 + L_1^\ast \Lambda \leq 0$,

$$(h, \beta, \Lambda) \in \mathcal{F}_w(S) \times \mathbb{R} \times \mathcal{M}([a,b]), \Lambda \geq 0.$$

**Proof.** The Lagrangian for problem (3.3)–(3.7) is

$$\vartheta(\mu, h, \beta, \Lambda) \triangleq \langle \mu, r \rangle + \langle h, L_0 \mu \rangle + \beta(\langle \mu, 1 \rangle - 1) + \langle \Lambda, L_1 \mu - y \rangle,$$

allowing problem (3.3)–(3.7) to be expressed as

$$\max_{\mu \in \mathcal{M}_w(K)} \left\{ \inf_{(h, \beta, \Lambda) \in \mathcal{F}_w(S) \times \mathbb{R} \times \mathcal{M}([a,b])} \{ \vartheta(\mu, h, \beta, \Lambda) : \Lambda \geq 0 \} : \mu \geq 0 \right\}. $$
We rearrange the Lagrangian to obtain
\[
\vartheta (\mu, h, \beta, \Lambda) = \langle \mu, r \rangle + \langle h, L_0 \mu \rangle + \beta (\langle \mu, 1 \rangle - 1) + \langle \Lambda, L_1 \mu - y \rangle \\
= \langle \mu, r \rangle + \langle L_0^* h, \mu \rangle + \langle \mu, 1 \rangle - \beta + \langle L_1^* \Lambda, \mu \rangle - \langle \Lambda, y \rangle \\
= \langle \mu, r + L_0^* h + \beta 1 + L_1^* \Lambda \rangle - \beta - \langle \Lambda, y \rangle.
\]

The dual to problem (3.3)–(3.7) is then
\[
\inf_{(h, \beta, \Lambda) \in F_w(S) \times \mathbb{R} \times \mathcal{M}([a, b])} \left\{ \max_{\mu \in \mathcal{M}_w(K)} \{ \vartheta (\mu, h, \beta, \Lambda) : \mu \geq 0 \} : \Lambda \geq 0 \right\}.
\]

Since \(\mu \geq 0\), the constraint \(r + L_0^* h + \beta 1 + L_1^* \Lambda \leq 0\) is implied. Since \(\beta\) is unrestricted, take \(\beta = -\beta\) to get the desired form. \(\Box\)

We write problem (4.1)–(4.3) with the “inf” objective rather than the “min” objective because we must verify that the optimal value is attained. The dual problem (4.1)–(4.3) is explicitly

\[
(4.4) \quad \inf \beta - \int_a^b \mathbb{E} \left[ (Y - \eta)_- \right] \Lambda (d\eta)
\]

\[
\text{s.t.} \quad r(s, a) + \int_a^b (z(s, a) - \eta)_- \Lambda (d\eta) \leq \beta + h(s) - \int_S h(\xi) Q(d\xi | s, a)
\]

(4.5) \quad \forall (s, a) \in K,

(4.6) \quad (h, \beta, \Lambda) \in F_w(S) \times \mathbb{R} \times \mathcal{M}([a, b]), \Lambda \geq 0.

Since \(r \leq 0\), problem (4.4)–(4.6) is readily seen to be consistent by choosing \(h = 0\), \(\beta = 0\), and \(\Lambda = 0\).

Problem (4.4)–(4.6) has another, more intuitive form. In [11, 12, 14], it is recognized that the Lagrange multipliers of stochastic dominance constraints are utility functions. This result is true in our case as well. Using the family \(\{(x - \eta)_- : \eta \in [a, b]\}\), any measure \(\Lambda \in \mathcal{M}([a, b])\) with \(\Lambda \geq 0\) induces an increasing concave function in \(\mathcal{C}([a, b])\) defined by

\[
u(x) = \int_a^b (x - \eta)_- \Lambda (d\eta)
\]

for all \(x \in \mathbb{R}\). In fact, the above definition of \(\nu\) gives a function in \(\mathcal{C}(\mathbb{R})\) as well. Define

\[
\mathcal{U}([a, b]) = \left\{ \nu(x) = \int_a^b (x - \eta)_- \Lambda (d\eta) \text{ for } \Lambda \in \mathcal{M}([a, b]), \Lambda \geq 0 \right\}
\]

to be the set of utility functions generated by the family \(\{(x - \eta)_- : \eta \in [a, b]\}\). The set \(\mathcal{U}([a, b]) \subset \mathcal{U}(\mathbb{R})\) is the set of all utility functions that can be constructed by limits of sums of scalar multiples of functions in \(\{(x - \eta)_- : \eta \in [a, b]\}\).

**COROLLARY 4.3.** Problem (4.4)–(4.6) is equivalent to

\[
(4.7) \quad \inf \beta - \mathbb{E} [\nu(Y)]
\]

(4.8) \quad \text{s.t.} \quad r(s, a) + u(z(s, a)) \leq \beta + h(s) - \int_S h(\xi) Q(d\xi | s, a) \quad \forall (s, a) \in K,

(4.9) \quad (h, \beta, u) \in F_w(S) \times \mathbb{R} \times \mathcal{U}([a, b]).
Proof. Notice that the function
\[ u(x) = \int_a^b (x - \eta) \Lambda(d\eta) \]
is an increasing concave function in \( x \) for any \( \Lambda \in \mathcal{M}([a, b]) \) with \( \Lambda \geq 0 \). By using this definition of \( u \), we see that for each state-action pair \((s, a)\),
\[ \langle \Lambda, (z(s, a) - \eta) \rangle = \int_a^b (z(s, a) - \eta) \Lambda(d\eta) = u(z(s, a)). \]
Further, we can apply the Fubini theorem again to obtain
\[ \langle \Lambda, y \rangle = \int_a^b \mathbb{E}[(Y - \eta) \Lambda(d\eta)] = \mathbb{E}[\int_a^b (Y - \eta) \Lambda(d\eta)] = \mathbb{E}[u(Y)]. \]

Next we verify that there is no duality gap between the primal problem (3.3)–(3.7) and its dual (4.1)–(4.3). All three dual problems (4.1)–(4.3), (4.4)–(4.6), and (4.7)–(4.9) are equivalent so the upcoming results apply to all of them.

The following result states that the optimal values of problems (3.3)–(3.7) and (4.1)–(4.3) are equal. Afterward, we will show that the optimal value of problem (4.1)–(4.3) is attained, establishing strong duality.

**Theorem 4.4.** Suppose Assumptions 3.2 and 3.5 hold. The optimal values of problems (3.3)–(3.7) and (4.1)–(4.3) are equal,
\[ \rho^* = \max \{ R(\pi, \nu) : (\pi, \nu) \in \Delta \} \]
\[ = \inf \{ \beta - \langle \Lambda, y \rangle : (\Lambda, y) \in \mathcal{F}_w(S) \times \mathbb{R} \times \mathcal{M}([a, b]), \Lambda \geq 0 \}. \]

**Proof.** Apply [29, Theorem 12.3.4], which in turn follows from [2, Theorem 3.9]. Introduce slack variables \( \alpha \in \mathcal{C}([a, b]) \) for the dominance constraints \( L_1 \mu \geq y \). We must show that the set
\[ H \triangleq \{(L_0 \mu, \langle \mu, 1 \rangle, L_1 x - \alpha, \langle \mu, r \rangle - \zeta) : \mu \geq 0, \alpha \geq 0, \zeta \geq 0 \} \]
is weakly closed (closed in the weak topology). Let \((D, \leq)\) be a directed (partially ordered) set, and consider a net
\[ \{(\mu_\kappa, \alpha_\kappa, \zeta_\kappa) : \kappa \in D\}, \]
where \( \mu_\kappa \geq 0, \alpha_\kappa \geq 0, \) and \( \zeta_\kappa \geq 0 \) in \( \mathcal{M}_w(K) \times \mathbb{R} \times \mathcal{C}([a, b]) \) such that
\[ (L_0 \mu_\kappa, \langle \mu_\kappa, 1 \rangle, L_1 \mu_\kappa - \alpha_\kappa, \langle \mu_\kappa, r \rangle - \zeta_\kappa) \]
has weak limit \( (\nu^*, \gamma^*, f^*, \rho^*) \in \mathcal{M}_w(S) \times \mathbb{R} \times \mathcal{C}([a, b]) \times \mathbb{R} \). Specifically,
\[ \langle \mu_\kappa, 1 \rangle \rightarrow \gamma^* \]
and
\[ \langle \mu_\kappa, r \rangle - \zeta_\kappa \rightarrow \rho^*, \]
since weak convergence on \( \mathbb{R} \) is equivalent to the usual notion of convergence,
\[ (L_0 \mu_\kappa, g) \rightarrow (\nu^*, g) \]
for all $g \in \mathcal{F}_{\hat{w}}(S)$ and
\[
\langle L_1 \mu_\kappa - \alpha_\kappa, \Lambda \rangle \to \langle f^*, \Lambda \rangle
\]
for all $\Lambda \in \mathcal{M}([a, b])$. We must show that $(\nu^*, \gamma^*, f^*, \rho^*) \in H$ under these conditions, i.e., that there exist $x \geq 0$, $\alpha \geq 0$, and $\zeta \geq 0$ such that
\[
\nu^* = L_0 \mu, \quad \gamma^* = \langle \mu, 1 \rangle, \quad f^* = L_1 \mu - \alpha, \quad \rho^* = \langle \mu, r \rangle - \zeta.
\]
The fact that there exist $\mu \geq 0$ and $\zeta \geq 0$ such that
\[
\nu^* = L_0 \mu, \quad \gamma^* = \langle \mu, 1 \rangle, \quad \rho^* = \langle \mu, r \rangle - \zeta
\]
is already established in [29, Theorem 12.3.4] and applies to our setting without modification.

It remains to verify that there exists $\alpha \in C([a, b])$ with $\alpha \geq 0$ and $f^* = L_1 \mu - \alpha$. Choose $\Lambda = \delta_\eta$ for the Dirac delta function at $\eta \in [a, b]$ to see that
\[
[L_1 \mu_\kappa](\eta) - \alpha_\kappa(\eta) \to f^*(\eta)
\]
for all $\eta \in [a, b]$, establishing pointwise convergence. Pointwise convergence on a compact set implies uniform convergence, so in fact
\[
L_1 \mu_\kappa - \alpha_\kappa \to f^*
\]
in the supremum norm topology on $C([a, b])$. Since $L_1 \mu_\kappa \in C([a, b])$ and $f^* \in C([a, b])$, it follows that $L_1 \mu_\kappa - f^* \in C([a, b])$ for any $\kappa$. Define $\alpha_\kappa = L_1 \mu_\kappa - f^*$ and $\alpha = L_1 \mu - f^*$, and notice that $\alpha \geq 0$ necessarily.

The next theorem shows that the dual problem (4.1)–(4.3) is solvable, i.e., there exists $(h^*, \beta^*, \Lambda^*)$ satisfying $r + L_0 h^* - \beta^* 1 + L_1 \Lambda^* \leq 0$ that attains the optimal value
\[
\beta^* - \langle \Lambda^*, y \rangle = \rho^*.
\]
When problem (4.1)–(4.3) is solvable, we are justified in saying that strong duality holds: the optimal values of both problems (3.3)–(3.7) and (4.1)–(4.3) are equal and both problems attain their optimal value.

To continue, we make an assumption about the existence of a particular minimizing sequence in problem (4.1)–(4.3).

Assumption 4.5. There exists a minimizing sequence $(h^n, \beta^n, \Lambda^n)$ in problem (4.1)–(4.3) such that
(a) $\{\beta^n\}$ is bounded in $\mathbb{R}$,
(b) $\{h^n\}$ is bounded in $\mathcal{F}_{\hat{w}}(S)$, and
(c) $\{\Lambda^n\}$ is bounded in the weak* topology on $\mathcal{M}([a, b])$.

We acknowledge that Assumption 4.5 is a strong requirement and difficult to check in practice. In [28, 29, 24, 25], similar assumptions are made about the existence of a suitably bounded minimizing sequence to ensure strong duality for infinite LPs. An analogue of Assumption 4.5 is even required for the unconstrained case.

However, we have already shown there is no duality gap under much milder assumptions that are easy to check, namely, Assumptions 3.2 and 3.5. The absence of a duality gap means that we can get arbitrarily good certificates of optimality in a primal-dual approximation scheme. Strong duality is not required for an approximation scheme to work.
We establish strong duality next under the preceding assumption. To reiterate, strong duality holds when the optimal values of problems (3.3)–(3.7) and (4.1)–(4.3) are equal and both problems are solvable.

**Theorem 4.6.** Suppose Assumption 4.5 holds. Strong duality holds between problem (3.3)–(3.7) and problem (4.1)–(4.3).

**Proof.** Let \((h^n, \beta^n, \Lambda^n) \in \mathcal{F}_w(S) \times \mathbb{R} \times \mathcal{M}([a, b])\) for \(n \geq 0\) be a minimizing sequence of triples given in Assumption 4.5:

\[
\begin{align*}
    r(s, a) + \int_a^b (z(s, a) - \eta) \Lambda^n(d\eta) \\
    \leq \beta^n + h^n(s) - \int_S h^n(\xi) Q(d\xi | s, a) \\
    \forall (s, a) \in K
\end{align*}
\]

for all \(n \geq 0\) and

\[
\begin{align*}
    \beta^n - \int_a^b \mathbb{E}[(Y - \eta)] \Lambda^n(d\eta) \downarrow \rho^*.
\end{align*}
\]

Since the sequence \(\{\beta^n\}\) is bounded, it has a convergent subsequence with \(\lim_{n \to \infty} \beta^n = \beta^*\).

Now \(\{\Lambda^n\}\) is bounded in \(\mathcal{M}([a, b])\) in the weak* topology induced by \(\mathcal{C}([a, b])\) by assumption. Since \(\{\Lambda^n\}\) is bounded, the sequence can be scaled to lie in the closed unit ball of \(\mathcal{M}([a, b])\) in the weak* topology. Since \(\mathcal{C}([a, b])\) is separable (there exists a countable dense set, i.e., the polynomials with rational coefficients), the weak* topology on \(\mathcal{M}([a, b])\) is metrizable. By the Banach–Alaoglu theorem, it follows that \(\{\Lambda^n\}\) has a subsequence that converges to some \(\Lambda^*\) in the weak* topology, i.e.,

\[
\langle \Lambda^n, f \rangle \to \langle \Lambda^*, f \rangle
\]

for all \(f \in \mathcal{C}([a, b])\). In particular, since \(\mathbb{E}[(Y - \eta)]\) and \((z(s, a) - \eta)\) are continuous functions on \([a, b]\) for all \((s, a) \in K\), it follows that

\[
\lim_{n \to \infty} \int_a^b \mathbb{E}[(Y - \eta)] \Lambda^n(d\eta) = \int_a^b \mathbb{E}[(Y - \eta)] \Lambda^*(d\eta)
\]

and

\[
\lim_{n \to \infty} \int_a^b (z(s, a) - \eta) \Lambda^n(d\eta) = \int_a^b (z(s, a) - \eta) \Lambda^*(d\eta).
\]

Finally, since \(\{h^n\}\) is bounded in \(\mathcal{F}_w(S)\) we can define

\[
h^*(s) \triangleq \liminf_{m \to \infty} h^m(s)
\]

for all \(s \in S\). Then the function \(h^*(s)\) is bounded in \(\mathcal{F}_w(S)\), and

\[
\liminf_{n \to \infty} \int_S h^n(\xi) Q(d\xi | s, a) \geq \int_S h^*(\xi) Q(d\xi | s, a)
\]

by Fatou’s lemma. Taking the limit, it follows that \((h^*, \beta^*, \Lambda^*)\) is an optimal solution to the dual problem. □

The role of the utility function \(u\) in problem (4.7)–(4.9) is fairly intuitive. The function \(u\) serves as a pricing variable for the performance function \(z\), and the total
reward is treated as if it were \( r(\cdot) + u(\cdot) \). Problem (4.7)–(4.9) leads to a new version of the optimality equations for average reward based on infinite-dimensional linear programming complementary slackness.

**Theorem 4.7.** Let \( \mu^* = \hat{\mu}^* \cdot \phi^* \) be an optimal solution to problem (3.3)–(3.7) and \( (h^*, \beta^*, u^*) \) be an optimal solution to problem (4.1)–(4.3). Then

\[
\langle \mu^*, u^* (z) \rangle = \mathbb{E} [u^* (Y)]
\]

and

\[
\beta^* + h^*(s) = \sup_{a \in A(s)} \left\{ r(s, a) + u^*(z(s, a)) + \int_S h^*(\xi) Q(d\xi | s, a) \right\}
\]

for \( \hat{\mu}^* \)-almost all \( s \in S \).

**Proof.** There is a corresponding optimal solution \( (h^*, \beta^*, \Lambda^*) \) to problem (4.1)–(4.3). Complementary slackness between problems (3.3)–(3.7) and (4.1)–(4.3) gives \( \langle \Lambda^*, L_1 \mu^* - y \rangle = 0 \), where \( (h^*, \beta^*, u^*) \) is a corresponding optimal solution of problem (4.1)–(4.3). Then

\[
\langle \Lambda^*, L_1 \mu^* \rangle = \langle L_1^* \Lambda^*, \mu^* \rangle = \langle \mu^*, u^* (z) \rangle
\]

and \( \langle \Lambda^*, y \rangle = \mathbb{E}[u^*(Y)] \).

Complementary slackness also gives

\[
\langle r + L_0^* h^* - \beta^* 1 + L_1^* \Lambda^*, \mu^* \rangle = 0,
\]

which yields the second statement since \( \mu^* \geq 0 \) and \( r + L_0^* h^* - \beta^* 1 + L_1^* \Lambda^* \leq 0 \).

**5. Variations and extensions.**

**5.1. Multivariate integral stochastic orders.** We extend our repertoire in this subsection to include some additional stochastic orders. Integral stochastic orders (see [37]) refer to stochastic orders that are defined in terms of families of functions. The increasing concave stochastic order is an example of an integral stochastic order, because it is defined in terms of the family of increasing concave functions. We now consider multivariate integral stochastic orders. So far, we have considered a \( z : K \to \mathbb{R} \) that is a scalar-valued function. In practice there are usually many system performance measures of interest, so it is logical to consider vector valued \( z : K \to \mathbb{R}^n \) as well. For example, \( z(s, a) \) may represent the service rate to \( n \) customers in a wireless network. The empirical distribution \( \lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} z(s_t, a_t) \) is now a vector-valued random variable on \( \mathbb{R}^n \).

Recall the multivariate increasing concave stochastic order. For random vectors \( X, Y \in \mathbb{R}^n \), \( X \) dominates \( Y \) in the increasing concave stochastic order, written \( X \succeq_{icv} Y \), if \( \mathbb{E}[u(X)] \geq \mathbb{E}[u(Y)] \) for all increasing concave functions \( u : \mathbb{R}^n \to \mathbb{R} \) such that both expectations exist. Unlike univariate \( \succeq_{icv} \), there is no parametrized family of functions (like \( (x - \eta)_- \)) that generates all the multivariate increasing concave functions. This result rests on the fact that the set of extreme points of the increasing concave functions on \( \mathbb{R}^n \) to \( \mathbb{R} \) is dense for \( n \geq 2 \); see [30, 9].

As in [15], we can relax the condition \( X \succeq_{icv} Y \) by constructing a tractable parametrized family of increasing concave functions. Let \( u(\cdot : \xi) : \mathbb{R}^n \to \mathbb{R} \) represent a family of increasing concave functions parametrized by \( \xi \in \Xi \subset \mathbb{R}^p \), where \( \Xi \) is compact. Then, the family of functions \( \{u(\cdot : \xi)\}_{\xi \in \Xi} \) is a subset of all increasing
concave functions and leads to a relaxation of $\geq_{icv}$. We say $X$ dominates $Y$ with respect to the integral stochastic order generated by \{u(\cdot; \xi)\}_{\xi \in \Xi}$ if $\mathbb{E}[u(X; \xi)] \geq \mathbb{E}[u(Y; \xi)]$ for all $\xi \in \Xi$. Define

$$Z_\xi(\pi, \nu) = \lim \inf_{T \to \infty} \frac{1}{T} \mathbb{E}^\pi_{\nu} \left[ \sum_{t=0}^{T-1} u(z(s_t, a_t); \xi) \right]$$

for all $\xi \in \Xi$. For convenience, we assume $u(x; \xi)$ is continuous in $\xi \in \Xi$ for any $x \in \mathbb{R}^n$.

We propose the multivariate dominance-constrained MDP

\begin{align}
&\sup_{(\pi, \nu) \in \Pi \times \mathcal{P}(S)} R(\pi, \nu) \\
&\text{s.t. } Z_\xi(\pi, \nu) \geq \mathbb{E}[u(Y; \xi)] \quad \forall \xi \in \Xi
\end{align}

using \{u(\cdot; \xi)\}_{\xi \in \Xi}.

By the same reasoning as earlier,

$$Z_\xi(\phi, \hat{\mu}) = \langle \mu, u(z(s, a); \xi) \rangle = \int_S \int_A u(z(s, a); \xi) \phi(da|s) \hat{\mu}(ds)$$

for all $\xi \in \Xi$ when $\mu = \hat{\mu} \cdot \phi \in \Delta_s$.

**Lemma 5.1.** For any $\mu \in \mathcal{P}(K)$, $\langle \mu, u(z; \xi) \rangle$ is continuous in $\xi$.

**Proof.** Write $\langle \mu, u(z; \xi) \rangle = \int_K u(z(s, a); \xi) \mu(ds, a)$. Certainly each function $u(z(s, a); \xi)$ is continuous in $\xi$ for any fixed $s \times a$. Since $\mu$ is finite, it follows that the integral of $u(z(s, a); \xi)$ with respect to $\mu$ is continuous in $\xi$. 

Let $C(\Xi)$ be the space of continuous functions on $\Xi$ in the supremum norm,

$$\|f\|_{C(\Xi)} \triangleq \sup_{x \in \Xi} |f(\xi)|.$$  

We will express the dominance constraints (5.2) as a linear operator in $C(\Xi)$. This operator depends on the parametrization $u(\cdot; \xi)$. The preceding lemma justifies defining $L_1 : \mathcal{M}(S \times A) \to C(\Xi)$ by

\begin{align}
[L_1 x](\xi) &\triangleq (x, u(z; \xi)) \quad \forall \xi \in \Xi.
\end{align}

Also define the continuous function $y \in C(\Xi)$ by $y(\xi) = \mathbb{E}[u(Y; \xi)]$ for all $\xi \in \Xi$ to represent the benchmark.

The steady-state version of problem (5.1)–(5.2) is the modified LP:

\begin{align}
\max &\quad \langle \mu, r \rangle \\
\text{s.t.} &\quad L_0 \mu = 0, \\
\langle \mu, 1 \rangle &\quad = 1, \\
L_1 \mu &\quad \geq y, \\
\mu &\in \mathcal{M}(K), \mu \geq 0.
\end{align}

Problem (5.4)–(5.8) is almost the same as problem (3.3)–(3.7), except that now $L_1 \mu$ is an element in $C(\Xi)$ to reflect the multivariate dominance constraint.

We now compute the adjoint of $L_1$, which depends on the choice of family \{u(\cdot; \xi) : \xi \in \Xi\}. The parametrization $u(\cdot; \xi)$ will appear explicitly in this computation.
Lemma 5.2. The adjoint of $L_1$ is $L_1^*: \mathcal{M}(\Xi) \to \mathcal{F}_w(K)$, where

$$[L_1^* \Lambda](s, a) \triangleq \int_{\Xi} u(z(s, a); \xi) \Lambda(d\xi).$$

Proof. Write

$$\langle \Lambda, L_1 \mu \rangle = \int_{\Xi} \langle \mu, u(z; \xi) \rangle \Lambda(d\xi) = \int_{K} \left[ \int_{\Xi} u(z(s, a); \xi) \mu(d(s, a)) \right] \Lambda(d\xi).$$

When $z$ is bounded on $S \times A$, then

$$\left| \int_{\Xi} u(z; \xi)(\mu \times \Lambda)(d((s, a) \times \xi)) \right| \leq \|u(z; \xi)\|_{\mathcal{F}_w(K)} \|\mu\|_{\mathcal{M}_w(K)} \|\Lambda\|_{\mathcal{M}([a, b])} < \infty.$$ 

The Fubini theorem applies to justify interchange of the order of integration,

$$\langle \Lambda, L_1 \mu \rangle = \int_{K} \left[ \int_{\Xi} u(z(s, a); \xi) \Lambda(d\xi) \right] \mu(d(s, a)) = \int_{K} (\Lambda, u(z(s, a); \xi)) \mu(d(s, a)).$$

The dual to problem (5.4)–(5.8) looks identical to problem (4.1)–(4.3) and is now explicitly

$$\inf \beta - \int_{\Xi} \mathbb{E}[u(Y; \xi)] \Lambda(d\xi) \quad \text{(5.9)}$$

subject to

$$r(s, a) + \int_{\Xi} u(z(s, a); \xi) \Lambda(d\xi) \leq \beta + h(s) - \int_{S} h(\xi) Q(d\xi|s, a) \quad \forall (s, a) \in K,$$

$$\text{(5.10)}$$

$$\text{(5.11)} \quad (h, \beta, \mu) \in \mathcal{F}_w(S) \times \mathbb{R} \times \mathcal{M}(\Xi), \Lambda \geq 0.$$

Define

$$U(\Xi) = \left\{ u(x) = \int_{\Xi} u(x; \xi) \Lambda(d\xi) \quad \text{for} \quad \Lambda \in \mathcal{M}(\Xi), \Lambda \geq 0 \right\}$$

to be the closure of the cone of functions generated by $\{u(x; \xi) : \xi \in \Xi\}$. In this case $U(\Xi)$ is a family of functions in $\mathcal{C}(\mathbb{R}^n)$, the space of continuous functions $f: \mathbb{R}^n \to \mathbb{R}$.

We see immediately that problem (5.9)–(5.11) is equivalent to

$$\inf \beta - \mathbb{E}[u(Y)] \quad \text{(5.12)}$$

subject to

$$r(s, a) + u(z(s, a)) \leq \beta + h(s) - \int_{S} h(\xi) Q(d\xi|s, a) \quad \forall (s, a) \in K,$$

$$\text{(5.13)}$$

$$\text{(5.14)} \quad (h, \beta, u) \in \mathcal{F}_w(S) \times \mathbb{R} \times U(\Xi).$$

The variables $u \in U(\Xi)$ in problem (5.12)–(5.14) are now pricing variables for the vector $z$. Assumption 4.5(c) can be modified so that each $\Lambda^n \in \mathcal{M}(\Xi)$. Then, under this modification of Assumption 4.5, strong duality holds between problem (5.4)–(5.8) and problem (5.12)–(5.14).
Theorem 5.3. The optimal values of problems (5.4)--(5.8) and (5.9)--(5.11) are equal. Further, the dual problem (5.9)--(5.11) is solvable and strong duality holds between problems (5.4)--(5.8) and (5.9)--(5.11).

5.2. Discounted reward. We briefly sketch the development for discounted reward; it is mostly similar. Discounted cost MDPs in Borel spaces with finitely many constraints are considered in [24]. In this subsection, we do not treat the initial distribution $\nu$ as a decision variable since the discounted reward is highly dependent on the initial distribution.

Introduce the discount factor $\delta \in (0, 1)$ and consider the long-run expected discounted reward

$$R(\pi, \nu) = \mathbb{E}_\nu^\pi \left[ \sum_{t=0}^{\infty} \delta^t r(s_t, a_t) \right].$$

We are interested in the distribution of discounted reward $z$,

$$\sum_{t=0}^{\infty} \delta^t (s_t, a_t).$$

Define

$$Z_\eta(\pi, \nu) \triangleq \mathbb{E}_\nu^\pi \left[ \sum_{t=0}^{\infty} \delta^t (z(s_t, a_t) - \eta) \right].$$

We propose the dominance-constrained MDP:

$$\begin{align*}
& (5.15) \sup_{\pi \in \Pi} R(\pi, \nu) \\
& (5.16) \text{s.t. } Z_\eta(\pi, \nu) \geq \mathbb{E} [(Y - \eta)_{-}] \quad \forall \eta \in [a, b].
\end{align*}$$

For emphasis, the initial distribution $\nu \in \mathcal{P}(S)$ is fixed in problem (5.15)--(5.16).

We work with the $\delta$-discounted expected occupation measure

$$\mu^\pi(\Gamma) \triangleq \sum_{t=0}^{\infty} \delta^t P^\pi_{\nu}(s_t, a_t) \in \Gamma$$

for all $\Gamma \in \mathcal{B}(S \times A)$. Now let

$$\begin{align*}
& (5.17) \quad [L_0 \mu](B) \triangleq \hat{\mu}(B) - \delta \int_{S \times A} Q(B | s, a) \mu(d(s, a)) \quad \forall B \in \mathcal{B}(S)
\end{align*}$$

and

$$\begin{align*}
& (5.18) \quad [L_1 \mu](\eta) \triangleq \langle \mu, (z - \eta)_{-} \rangle \quad \forall \eta \in [a, b].
\end{align*}$$

Also continue to define $y \in \mathcal{C}([a, b])$ by $y(\eta) = \mathbb{E}[(Y - \eta)_{-}]$ for all $\eta \in [a, b]$. Problem (5.15)--(5.16) is then equivalent to the LP

$$\begin{align*}
& (5.19) \quad \max \quad \langle \mu, r \rangle \\
& (5.20) \quad \text{s.t. } L_0 \mu = \nu, \\
& (5.21) \quad L_1 \mu \geq y, \\
& (5.22) \quad \mu \in \mathcal{M}(K), \mu \geq 0.
\end{align*}$$
Introduce Lagrange multipliers $h \in \mathcal{F}_w(S)$ for constraint $L_0 \mu = \nu$ and multipliers $\Lambda \in \mathcal{M}([a, b])$ for constraint $L_1 \mu \geq y$; the Lagrangian is then
\[
\vartheta(h, \nu, \Lambda) = \langle \mu, r \rangle + \langle h, L_0 \mu - \nu \rangle + \langle \Lambda, L_1 \mu - y \rangle.
\]

The adjoint of $L_0$ is $L_0^* : \mathcal{F}_w(S) \to \mathcal{F}_w(S \times A)$ defined by
\[
[L_0^* h](s, a) \triangleq h(s) - \delta \int_S h(\xi) Q(d\xi | s, a).
\]

The adjoint of $L_1$ is still $L_1^* : \mathcal{M}([a, b]) \to \mathcal{F}_w(S \times A)$, where
\[
[L_1^* \Lambda](s, a) \triangleq \int_a^b (z(s, a) - \eta) \Lambda(d\eta).
\]

The form of the dual follows.

**Theorem 5.4.** The dual to problem (5.19)–(5.22) is
\[
\begin{align*}
\text{(5.23)} & \quad \min \langle h, \nu \rangle - \langle \Lambda, y \rangle \\
\text{(5.24)} & \quad \text{s.t.} \quad r + L_0^* h + L_1^* \Lambda \geq 0, \\
\text{(5.25)} & \quad h \in \mathcal{F}_w(K), \ \Lambda \in \mathcal{M}([a, b]), \ \Lambda \geq 0.
\end{align*}
\]

The optimal values of problems (5.19)–(5.22) and (5.23)–(5.25) are equal, and problem (5.23)–(5.25) is solvable.

This dual is explicitly
\[
\begin{align*}
\text{(5.26)} & \quad \min \langle h, \nu \rangle - \mathbb{E}[u(Y)] \\
\text{(5.27)} & \quad \text{s.t.} \quad r(s, a) + u(z(s, a)) \leq h(s) - \delta \int_S h(\xi) Q(d\xi | s, a) \quad \forall (s, a) \in K, \\
\text{(5.28)} & \quad h \in \mathcal{F}_w(K), \ u \in \mathcal{U}([a, b]).
\end{align*}
\]

Problem (5.26)–(5.28) leads to a modified set of optimality equations for the infinite horizon discounted reward case, namely,
\[
h(s) = \max_{a \in A(s)} \left\{ r(s, a) + u(z(s, a)) + \delta \int_S h(\xi) Q(d\xi | s, a) \right\}
\]
for all $s \in S$.

**5.3. Approximate linear programming.** Various approaches have been put forward for solving infinite-dimensional LPs with sequences of finite-dimensional LPs, such as in [26, 35]. Approximate linear programming (ALP) has been put forward as an approach to the curse of dimensionality, and it can be applied to our present setting. The average reward LP (3.3)–(3.7) and the discounted reward LP (5.19)–(5.22) generally have uncountably many variables and constraints.

ALP for average cost dynamic programming is developed in [10]. Previous work on ALP for dynamic programming has focused on approximating the cost-to-go function $h$ rather than the steady-state occupation measure $\mu$. It is more intuitive to design basis functions for the cost-to-go function than the occupation measure. For problem (3.3)–(3.7), we approximate the cost-to-go function $h \in \mathcal{F}_w(S)$ with the basis functions
\{\phi_1, \ldots, \phi_m \} \subset \mathcal{F}_w (S)$. We approximate the pricing variable $u \in \mathcal{U}([a, b])$ with basis functions $\{u_1, \ldots, u_n \} \subset \mathcal{U}([a, b])$. The resulting approximate LP is

\begin{align}
\min & \quad \beta - \mathbb{E} \left[ \sum_{i=1}^{n} \alpha_i u_i (Y) \right] \\
\text{s.t.} & \quad r (s, a) + \sum_{i=1}^{n} \alpha_i u_i (z (s, a)) \leq \beta + \sum_{j=1}^{m} \gamma_j h_j (s) - \int_{S} \left[ \sum_{j=1}^{m} \gamma_j h_j \right] (\xi) Q (d\xi | s, a) \\
& \quad \forall (s, a) \in K, \tag{5.30}
\end{align}

\begin{align}
(\gamma, \beta, \alpha) \in \mathbb{R}^{m} \times \mathbb{R} \times \mathbb{R}^{n}. 
\end{align}

We are justified in writing minimization instead of infimum in problem (5.29)–(5.31) because there are only finitely many decision variables. ALP has been studied extensively for the linear programming representation of the optimality equations for discounted infinite horizon dynamic programming (see [18, 19, 16]). The discounted approximate LP is

\begin{align}
\min & \quad \langle h, \nu \rangle - \mathbb{E} \left[ \sum_{i=1}^{n} \alpha_i u_i (Y) \right] \\
\text{s.t.} & \quad r (s, a) + \sum_{i=1}^{n} \alpha_i u_i (z (s, a)) \leq \sum_{j=1}^{m} \gamma_j h_j (s) - \delta \int_{S} \left[ \sum_{j=1}^{m} \gamma_j h_j \right] (\xi) Q (d\xi | s, a) \\
& \quad \forall (s, a) \in K, \tag{5.33}
\end{align}

\begin{align}
(\gamma, \alpha) \in \mathbb{R}^{m} \times \mathbb{R}^{n}. 
\end{align}

Both problems (5.29)–(5.31) and (5.32)–(5.34) are restrictions of the corresponding problems (4.7)–(4.9) and (5.26)–(5.28).

Problems (5.29)–(5.31) and (5.32)–(5.34) have a manageable number of decision variables but an intractable number of constraints. Constraint sampling has been a prominent tool in ALP, and we cite a relevant result now. Let

\begin{align}
(\gamma, r, \kappa) \geq 0 \quad \forall \gamma \in \mathcal{L}
\end{align}

be a set of linear inequalities in the variables $r \in \mathbb{R}^{k}$ indexed by an arbitrary set $\mathcal{L}$. Let $\psi$ be a probability distribution on $\mathcal{L}$. We would like to take independent and identically (i.i.d.) samples from $\mathcal{L}$ to construct a set $\mathcal{W} \subseteq \mathcal{L}$ with

\begin{align}
\sup \left\{ \psi \left( \{ y : (\gamma_y, r) + \kappa_y \geq 0 \} \right) \leq \epsilon \right. \\
\left. \forall r \in \mathcal{W} \right\} \leq \epsilon.
\end{align}

**Theorem 5.5 (see [19, Theorem 2.1]).** For any $\delta \in (0, 1)$ and $\epsilon \in (0, 1)$, and

\begin{align}
m \geq \frac{4}{\epsilon} \left( k \ln \frac{12}{\epsilon} + \ln \frac{2}{\delta} \right),
\end{align}

where $k$ is the number of basis functions.
a set \( \mathcal{W} \) of \( m \) i.i.d. samples drawn from \( \mathcal{L} \) according to distribution \( \psi \), satisfies

\[
\sup_{\{r, \langle z, r \rangle + \kappa z \geq 0 \forall z \in \mathcal{W}\}} \psi \left( \{y : \langle \gamma_y, r \rangle + \kappa_y < 0\} \right) \leq \epsilon
\]

with probability at least \( 1 - \delta \).

Thus, we can sample state-action pairs from any distribution \( \psi \) on \( K \) to obtain tractable relaxations of problems (5.29)–(5.31) and (5.32)–(5.34) with probabilistic feasibility guarantees. Note that the number of samples required is \( O\left(\frac{1}{\epsilon \ln \frac{1}{\epsilon}, \ln \frac{1}{\delta}}\right) \).

5.4. Finite state and action spaces. The development for finite state and action spaces is much simpler. In particular, the usual linear programming duality theory applies immediately to establish strong duality. For this section, let \( x \) denote an occupation measure on \( K \) to emphasize that it is finite-dimensional. Also suppose the benchmark \( Y \) has finite support \( \text{supp} \ Y \subset \mathbb{R} \), so that constraint (2.3) is equivalent to

\[
E_x \left[ (z(s,a) - \eta)_- \right] \geq E \left[ (Y - \eta)_- \right] \quad \forall \eta \in \text{supp} \ Y
\]

by [11, Proposition 3.2]. Each expectation

\[
E_x \left[ (z(s,a) - \eta)_- \right] = \sum_{(s,a) \in K} x(s,a)(z(s,a) - \eta)_-
\]

is a linear function of \( x \).

For finite state and action spaces, the steady-state version of problem (2.2)–(2.3) is

\[
\begin{align*}
\text{max} & \quad \sum_{(s,a) \in K} r(s,a)x(s,a) \\
\text{s.t.} & \quad \sum_{a \in A(s)} x(j,a) - \sum_{s \in S} \sum_{a \in A(s)} P(j|s,a)x(s,a) = 0 \quad \forall j \in S, \\
& \quad \sum_{(s,a) \in K} x(s,a) = 1, \\
& \quad E_x \left[ (z(s,a) - \eta)_- \right] \geq E \left[ (Y - \eta)_- \right] \quad \forall \eta \in \text{supp} \ Y, \\
& \quad x \geq 0.
\end{align*}
\]

Duality for problem (3.3)–(3.7) is immediate from linear programming duality. As discussed in [39, Chapter 8], the dual of the linear programming problem without the dominance constraints is

\[
\begin{align*}
\text{min} & \quad g \\
\text{s.t.} & \quad g + h(s) - \sum_{j \in S} P(j|s,a)h(j) \geq r(s,a) \quad \forall (s,a) \in K, \\
& \quad g \in \mathbb{R}, \ h \in \mathbb{R}^{[S]}.
\end{align*}
\]

The vector \( h \) is interpreted as the average cost-to-go function. To proceed with the dual for problem (3.3)–(3.7), let \( \lambda \in \mathbb{R}^{[\text{supp} \ Y]} \) with \( \lambda \geq 0 \) and consider the piecewise linear increasing concave function

\[
u(\xi) = \sum_{\eta \in \text{supp} \ Y} \lambda(\eta)(\xi - \eta)_-
\]
with breakpoints at $\eta \in \operatorname{supp} Y$. The above function $u(\xi)$ can be interpreted as a utility function for a risk-averse decision maker. We define

$$\mathcal{U}(\operatorname{supp} Y) = \text{cl cone} \left\{ (x - \eta)_- : \eta \in \operatorname{supp} Y \right\}$$

$$= \left\{ u(x) = \sum_{\eta \in \operatorname{supp} Y} \lambda(\eta) (x - \eta)_- \text{ for } \lambda \in \mathbb{R}^{\operatorname{supp} Y}, \lambda \geq 0 \right\}$$

to be the set of all such functions. Since $\operatorname{supp} Y$ is assumed to be finite, $\mathcal{U}(\operatorname{supp} Y)$ is a finite-dimensional set.

**Theorem 5.6.** The dual to problem (5.37)–(5.41) is

$$\min g - \mathbb{E}[u(Y)]$$

(s.t. $r(s, a) + u(z(s, a)) \leq g + h(s) - \sum_{j \in S} P(j | s, a) h(j) \quad \forall (s, a) \in K,$

$$g \in \mathbb{R}, h \in \mathbb{R}^{|S|}, u \in \mathcal{U}(\operatorname{supp} Y).$$

Strong duality holds between problem (5.37)–(5.41) and problem (5.42)–(5.44).

**Proof.** Introduce the Lagrangian

$$L(x, g, h, \lambda) \triangleq \sum_{(s, a) \in K} r(s, a) x(s, a) + g \left[ \sum_{(s, a) \in K} x(s, a) - 1 \right]$$

$$+ \sum_{j \in S} h(j) \left[ \sum_{a \in A(s)} x(j, a) - \sum_{s \in S} \sum_{a \in A(s)} P(j | s, a) x(s, a) \right]$$

$$+ \sum_{\eta \in \operatorname{supp} Y} \lambda(\eta) \left( \left[ \sum_{(s, a) \in K} x(s, a) (z(s, a) - \eta)_- - \mathbb{E}[(Y - \eta)_-] \right) \right).$$

Define the increasing concave function

$$u(\xi) = \sum_{\eta \in \operatorname{supp} Y} \lambda(\eta) (\xi - \eta)_- ;$$

then

$$\sum_{\eta \in \operatorname{supp} Y} \lambda(\eta) \left[ \left[ \sum_{(s, a) \in K} x(s, a) (z(s, a) - \eta)_- - \mathbb{E}[(Y - \eta)_-] \right) \right]$$

$$= \sum_{(s, a) \in K} x(s, a) (u(z(s, a))) - \mathbb{E}[u(Y)]$$

by interchanging finite sums. So, the Lagrangian could also be written as

$$L(x, g, h, u) = \sum_{(s, a) \in K} r(s, a) x(s, a) + g \left[ \sum_{(s, a) \in K} x(s, a) - 1 \right]$$

$$+ \sum_{j \in S} h(j) \left[ \sum_{a \in A(s)} x(j, a) - \sum_{s \in S} \sum_{a \in A(s)} P(j | s, a) x(s, a) \right]$$

$$+ \sum_{(s, a) \in K} x(s, a) u(z(s, a)) - \mathbb{E}[u(Y)]$$
for \( u \in U(\text{supp} \ Y) \). The dual to problem (5.41)–(5.40) is defined as
\[
\min_{g \in \mathbb{R}, h \in \mathbb{R}^{|S|}, u \in U(\text{supp} \ Y)} \left\{ \max_{x \geq 0} L(x, g, h, u) \right\}.
\]
Rearranging the Lagrangian gives
\[
L(x, g, h, u) = \sum_{(s, a) \in K} x(s, a) \left[ r(s, a) + g + h(s) - \sum_{j \in S} P(j | s, a) h(j) + u(z(s, a)) \right] - g - \mathbb{E}[u(Y)],
\]
so that the dual to problem (5.41)–(5.40) is
\[
\min -g - \mathbb{E}[u(Y)]
\]
\[
\text{s.t. } r(s, a) + g + h(s) - \sum_{j \in S} P(j | s, a) h(j) + u(z(s, a)) \leq 0 \quad \forall (s, a) \in K,
\]
\[
\lambda \in \mathbb{R}, \ h \in \mathbb{R}^{|S|}, \ u \in U(\text{supp} \ Y).
\]
Since \( g \) and \( h \) are unrestricted, take \( g = -g \) and \( h = -h \) to get the desired result. \( \blacksquare \)

We used linear programming duality in the preceding proof for illustration. Alternatively, we could have just applied our general strong duality result from earlier. It is immediate that problem (5.43)–(5.44) is the finite-dimensional version of problem (4.7)–(4.9).

There is no difficulty with the Slater condition for problems (5.40)–(5.41) and (5.43)–(5.44) as there is in [11, 12]. In [11, 12], the decision variable in a stochastic program is a random variable, so stochastic dominance constraints are nonlinear. In our case, the decision variable \( x \) is in the space of measures and the dominance constraints are linear. Linear programming duality does not depend on the Slater condition.

The development for the discounted case is similar. In terms of discounted occupation measures \( x \), problem (5.15)–(5.16) is
\[
\max \sum_{(s, a) \in K} r(s, a) x(s, a).
\]
\[
\text{s.t. } \sum_{a \in A(s)} x(j, a) - \sum_{s \in S} \sum_{a \in A(s)} \gamma P(j | s, a) x(s, a) = \alpha(j) \quad \forall j \in S,
\]
\[
\mathbb{E}_x \left[ (z(s, a) - \eta)_- \right] \geq \mathbb{E} \left[ (Y - \eta)_- \right] \quad \forall \eta \in \text{supp} \ Y,
\]
\[
x \geq 0.
\]
We compute the dual to problem (5.45)–(5.48) in the next theorem using the space of utility functions \( U(\text{supp} \ Y) \) from earlier.

**Theorem 5.7.** The dual to problem (5.45)–(5.48) is
\[
\min \sum_{j \in S} \alpha(j) v(j) - \mathbb{E}[u(Y)]
\]
\[
\text{s.t. } v(s) - \sum_{s \in S} \sum_{a \in A(s)} \gamma P(j | s, a) v(j) \geq r(s, a) + u(z(s, a)) \quad \forall (s, a) \in K,
\]
\[
v \in \mathbb{R}^{|S|}, \ u \in U(\text{supp} \ Y).
\]
**Strong duality holds between problem (5.45)–(5.48) and problem (5.49)–(5.51).**
6. Portfolio optimization. We use an infinite horizon discounted portfolio optimization problem to illustrate our ideas in this section. See [4] for a recent monograph on applications of MDPs to finance. A single period portfolio optimization with stochastic dominance constraints is analyzed in [13]. Specifically, the model in [13] puts a stochastic dominance constraint on the return rate of a portfolio allocation. We use this model as our motivation for the dynamic setting and put a stochastic dominance constraint on the discounted infinite horizon return rate.

Suppose there are $n$ assets whose random return rates evolve according to a discrete time Markov chain. We can include a riskless asset with a constant return rate in this set. The asset returns at time $t$ are

$$\rho_t = (\rho_t(1), \ldots, \rho_t(n)) \in \mathbb{R}^n_+,$$

where $\rho_t(i)$ is the return rate of asset $i$ at time $t$. The portfolio at time $t$ is captured by

$$x_t = (x_t(1), \ldots, x_t(n)) \in \mathbb{R}^n_+,$$

where $x_t(i)$ is the proportion of the portfolio invested in asset $i$ at time $t$. We require $\sum_{i=1}^{n} x_t(i) = 1$ and $x_t \geq 0$ for all $t \geq 0$; there is no shorting. We define the state space of this MDP to be $S = \mathbb{R}^n_+ \times \mathbb{R}^n_+$, where $(\rho_{t-1}, x_t)$ is the state at time $t$.

At each time $t \geq 0$, the investor observes the current prices of the assets and then updates portfolio positions subject to transaction costs before new prices are realized. Let $a_t \subset \mathbb{R}^n$ be the buying and selling decisions at time $t$, where $a_t(i)$ is the total change in the number of shares held of asset $i$. Buy and sell decisions are defined on $A = \mathbb{R}^n$. Define

$$A(x) \triangleq \left\{ a \in \mathbb{R}^n : x(i) + a(i) \geq 0 \forall i = 1, \ldots, n, \sum_{i=1}^{n} a(i) = 0 \right\}$$

to be the set of feasible buying and selling decisions given current holdings $x_t$. Here

$$K \triangleq \{(\rho, x, a) \in S \times A : a \in A(x)\}.$$ 

We note that $K$ is closed since the functions $x(i) + a(i)$ and $\sum_{i=1}^{n} a(i)$ are linear in $(x, a)$ and do not depend on $\rho$. The set $K$ is also compact since we are dealing with allocations (and thus $x$ and $a$ must be bounded).

We use $Q$ to represent the controlled transition kernel. Holdings evolve deterministically, so the equalities

$$x_{t+1}(i) = x_t(i) + a_t(i), \quad i = 1, \ldots, n, \quad t \geq 0,$$

are satisfied almost surely with respect to $P^\pi_\nu$ for any $\pi \in \Pi$. Since return rates $\{r_t, t \geq 0\}$ are exogenous, we can use $\tilde{Q}$ to represent the transition kernel for $\{\rho_t, t \geq 0\}$ alone. When $\tilde{Q}$ is weakly continuous, then $Q$ is weakly continuous by linearity of the evolution of $\{x_t, t \geq 0\}$.

The transaction costs $c : A \to \mathbb{R}$ are defined to be

$$c(a_t) \triangleq \sum_{i=1}^{n} a_t(i)^2.$$
Notice $c$ is a moment on $K$ and it is nonnegative $c \geq 0$. The overall return rate between time $t - 1$ and $t$ (evaluated at time $t$) is

$$z(\rho_{t-1}, x_t) \triangleq \langle \rho_{t-1}, x_t \rangle.$$ 

It follows that $z$ is continuous on $K$. We can also suppose that $z$ is uniformly bounded on $K$, since the relative change in wealth cannot realistically exceed a certain order of magnitude.

We want to minimize discounted transaction costs

$$C(\pi, \nu) \triangleq E^\pi_\nu \left[ \sum_{t \geq 0} \delta^t c(a_t) \right]$$

subject to a stochastic dominance constraint on the discounted return rate. Define

$$Z_\eta(\pi, \nu) \triangleq E^\pi_\nu \left[ \sum_{t=0}^{\infty} \delta^t (z(\rho_{t-1}, x_t) - \eta)_- \right]$$

to be the expected discounted shortfall in relative returns at level $\eta$. We introduce a benchmark $Y$ for the discounted return rate, and we suppose the support of $Y$ is bounded within $[a, b]$. In this example, the benchmark can be taken as any market index. Our resulting portfolio optimization problem is then

\begin{align*}
\max_{\pi \in \Pi} & -C(\pi, \nu) \\
\text{s.t.} & \quad Z_\eta(\pi, \nu) \geq E \left[ (Y - \eta)_- \right], \quad \eta \in [a, b].
\end{align*}

An optimal policy for problem (6.1)–(6.2) tracks a market index $Y$ in a stochastic dominance sense with minimal transaction costs.

We introduce the weight functions

$$w(s, a) = 1 + c(a)$$

and

$$\hat{w}(s) = 1 + \sup_{a \in A(s)} c(a),$$

which are both bounded below by one. The function $z$ is an element of $\mathcal{F}_w(K)$ since

$$\sup_{(\rho, x, a) \in K} \frac{|z(\rho, x)|}{1 + c(a)} \leq \sup_{(\rho, x, a) \in K} |z(\rho, x)| < \infty$$

under the assumption that $z$ is bounded on $K$. Also, $\hat{w}$ is uniformly bounded on $S$ so $\int_S \hat{w}(\xi)Q(d\xi \mid s, a) : S \times A \to \mathbb{R}$ is automatically an element of $\mathcal{F}_w(K)$.

Based on this discussion, Assumptions 3.2 and 3.5 are satisfied. Using the definitions from subsection 5.2, the primal LP for problem (6.1)–(6.2) is

\begin{align*}
\max & -\langle \mu, c \rangle \\
\text{s.t.} & \quad L_0 \mu = \nu, \\
& \quad L_1 \mu \geq y, \\
& \quad \mu \in \mathcal{M}(K), \, \mu \geq 0.
\end{align*}
The dual LP for problem (6.1)–(6.2) is
\[
\begin{align*}
\min & \quad \langle h, \nu \rangle - E [u (Y)] \\
\text{s.t.} & \quad - c (a) + u (z (s, a)) \leq h(s) - \delta \int_S h (\xi) Q (d\xi | s, a) \quad \forall (s, a) \in K, \\
& \quad h \in F_w (K), u \in U ([a, b]).
\end{align*}
\]

There is no duality gap between these LPs.

When return rates \{\rho_t, t \geq 0\} and buy/sell decisions \{a_t, t \geq 0\} are discrete, then \{x_t, t \geq 0\} are discrete as well. These quantities are also all bounded. We can then justify applying the development in subsection 5.4 to both the primal and the dual linear programming problem when \(K\) and \(\text{supp} Y\) are approximated with finite sets. Since there is no duality gap, we can get arbitrarily good certificates of optimality by choosing successively finer discretizations.

7. Conclusion. We have shown how to use stochastic dominance constraints in infinite horizon MDPs. Convex analytic methods establish that stochastic dominance-constrained MDPs can be solved via linear programming and have corresponding dual linear programming problems. Conditions are given for strong duality to hold between these two LPs. Utility functions appear in the dual as pricing variables corresponding to the stochastic dominance constraints. This result has intuitive appeal, since our stochastic dominance constraints are defined in terms of utility functions, and parallels earlier results [11, 12, 14]. Our results are shown to be extendable to many types of stochastic dominance constraints, particularly multivariate ones.

There are three main directions for our future work. First, we will consider efficient strategies for computing the optimal policy to stochastic dominance-constrained MDPs. Second, we would like to explore other methods for modeling risk in MDPs using convex analytic methods. Specifically, we are interested in solving MDPs with convex risk measures and chance constraints with “static” optimization problems as we have done here. Third, as suggested by our portfolio optimization example, we will consider online data-driven optimization for the stochastic dominance-constrained MDPs in this paper. The transition probabilities of underlying MDPs are not known in practice and must be learned online.

REFERENCES

A. Muller and D. Stoyan

D. Pucci de Farias and B. Van Roy


