Approximation schemes for countably-infinite linear programs with moment bounds

Juan Kuntz† Philipp Thomas‡ Guy-Bart Stan§ Mauricio Barahona¶

October 10, 2018

Abstract

We introduce approximation schemes for a type of countably-infinite-dimensional linear programs (CILPs) whose feasible points are unsigned measures and whose optimal values are bounds on the averages of these measures. In particular, we explain how to approximate the program’s optimal value, optimal points, and minimal point (should one exist) by solving finite-dimensional linear programs. We show that the approximations converge to the CILP’s optimal value, optimal points, and minimal point as the size of the finite-dimensional program approaches that of the CILP. Inbuilt in our schemes is a degree of error control: they yield lower and upper bounds on the optimal values and we give a simple bound on the approximation error of the minimal point. To motivate our work, we discuss applications of our schemes taken from the Markov chain literature: stationary distributions, occupation measures, and exit distributions.

1 Introduction

Countably infinite linear programs (CILPs) are linear programs (LPs) with countably many decision variables and constraints. CILPs arise in network flow problems [31, 35, 32], infinite-horizon planning and stochastic programming applications such as production planning, equipment replacement, and capacity expansion [17, 24, 23], semi-infinite linear programs [3, 19], search problems in robotics [7], robust optimisation [11], and, prominently, in optimal control problems tied to Markov chains [30, 21, 2, 10]. In this paper, we consider CILPs that arise in the analysis of Markov chains with unbounded state space. In particular, measures of importance to these chains (e.g., the stationary distributions and occupation measures in Sec. 2.1) are feasible points of the CILPs and the optimal values of these programs are bounds on the averages of these measures.

Due to their infinite size, CILPs cannot be solved directly and one must resort to an approximation scheme. We introduce two such schemes—Schemes A and B—that ignore all but finitely many of the CILP’s decision variables and replace its countably-many constraints with finitely-many ones.

†Department of Mathematics and Department of Bioengineering, Imperial College London, London, SW7 2AZ, UK (juan.kuntz08@imperial.ac.uk).
‡Department of Mathematics, Imperial College London, London, SW7 2AZ, UK (p.thomas@imperial.ac.uk).
§Department of Bioengineering, Imperial College London, London, SW7 2AZ, UK (Co-corresponding author: g.stan@ic.ac.uk).
¶Department of Mathematics, Imperial College London, London, SW7 2AZ, UK (Co-corresponding author: m.barahona@ic.ac.uk).

*The first author was supported by a BBSRC PhD Studentship (BB/F017510/1). PT was supported through a Fellowship of The Royal Commission for the Exhibition of 1851. GBS acknowledges support by an EPSRC Fellowship for Growth (EP/M002187/1) and the EU H2020-FETOPEN-2016-2017 project 766840 (COSY-BIO). MB acknowledges support from EPSRC grant EP/N014529/1 supporting the EPSRC Centre for Mathematics of Precision Healthcare.
Scheme A yields lower and upper bounds on the optimal value of the CILP and approximations of its optimal points. Scheme B consists of iterating scheme A and is applicable to CILPs that possess a minimal point, i.e., a feasible point that is element-wise no greater than every other feasible point. It yields approximations of this minimal point whose error is straightforward to bound in practice. Both schemes are shown to converge: the bounds on the optimal value, the approximations of the optimal points, and those of the minimal point tend to the optimal value, optimal point, and minimal point (respectively) as the size of the approximating LPs approaches that of the original CILP.

To derive our approximation schemes, we append a moment bound as an explicit constraint to our CILP (see (2.4) in Sec. 2). The moment bound is satisfied by the feasible points of the CILP and can readily be computed using either Foster-Lyapunov criteria [28, 29, 16, 25] or mathematical programming [22, 34, 18, 25, 26, 33, 8, 9, 15]. The inclusion of this bound allows us to establish the strong form of convergence of the schemes and to quantify the approximation error. The latter property is our main innovation compared with previous approximation schemes [14, 1, 19, 5, 13, 35, 27, 32, 12].

The outline of the paper is as follows. In Sec. 2, we introduce the class of LPs we consider, motivate our work with two problems taken from the Markov chain literature, and introduce the notation and assumptions used throughout the paper. In Sec. 3, we introduce Scheme A and prove its convergence. In Sec. 4, we prove the convergence of Scheme B and show how to bound its error. In Sec. 5, we explain how to apply Schemes A and B to the motivating problems introduced in Sec. 2. We conclude the paper with a discussion in Sec. 6. The paper has two appendices: the first contains the proof of Theorem 3.2 omitted from the main text to ease the paper’s reading, the second explains how to extend our results beyond Assumption (3.2) made throughout the paper.

2 The linear program and two motivating problems

We consider a measure $\rho$ on a countable set $\mathcal{X}$ that a) satisfies $|\mathcal{X}|$ linear equations
\begin{equation}
\rho H(x) := \sum_{z \in \mathcal{X}} \rho(z) h(z, x) = \phi(x) \quad \forall x \in \mathcal{X},
\end{equation}
where $|\mathcal{X}|$ denotes the cardinality of $\mathcal{X}$ and $H := (h(x, y))_{x, y \in \mathcal{X}}$ is a Metzler matrix; and b) its image
\begin{equation}
\psi(y) := \rho G(y) = \sum_{x \in \mathcal{S}} \rho(x) g(x, y) \quad \forall y \in \mathcal{Y},
\end{equation}
under a non-negative matrix $G := (g(x, y))_{x \in \mathcal{X}, y \in \mathcal{Y}}$ is a probability distribution:
\begin{equation}
\rho(g) = \sum_{x \in \mathcal{X}} g(x) \rho(x) = 1, \quad \text{where} \quad g(x) := \sum_{y \in \mathcal{Y}} g(x, y) \quad \forall x \in \mathcal{X}
\end{equation}
and $\mathcal{Y}$ denotes a second countable set.

If $\mathcal{X}$ is infinite, or finite but large, (2.1)–(2.3) cannot be solved directly. However, a variety of analytical and numerical methods yield bounds on the moments of $\rho$ (or moment bounds for short). That is, a constant $c \geq 0$ such that
\begin{equation}
\rho(w) := \sum_{x \in \mathcal{X}} w(x) \rho(x) \leq c
\end{equation}
where $w$ is a norm-like function meaning a non-negative function on $\mathcal{X}$ with finite sublevel sets:
\begin{equation}
\mathcal{X}_r := \{ x \in \mathcal{X} : w(x) < r \} \quad r > 0.
\end{equation}
The set of (unsigned) measures that satisfy (2.1), (2.3), and (2.4),

\[ \mathcal{L} := \left\{ \rho \in \mathbb{R}^{|\mathcal{X}|} : \begin{cases} \rho H(x) = \phi(x) & \forall x \in \mathcal{X}, \\ \rho(y) = 1, \\ \rho(w) \leq c, \\ \rho \geq 0. \end{cases} \right\}, \tag{2.6} \]

forms the feasible set of a linear program with \(|\mathcal{X}|\) decision variables, \(|\mathcal{X}| + 1\) equality constraints, and \(|\mathcal{X}| + 1\) inequality constraints. In particular, the linear programs we approximate in this paper are

\[ l_f := \inf \{ \rho(f) : \rho \in \mathcal{L} \}, \quad u_f := \sup \{ \rho(f) : \rho \in \mathcal{L} \}. \tag{2.7} \]

where \(l_f\) and \(u_f\) denote the optimal values and \(f\) is any given real-valued function on \(\mathcal{X}\) such that the integrals in (3.8) are well-defined. Any feasible point \(\rho \in \mathcal{L}\) achieving this infimum (resp. supremum):

\[ \rho_*(f) = l_f, \quad (\text{resp. } \rho_*(f) = u_f) \]

is said to be an optimal point. It is straightforward to verify that \(\mathcal{L}\) possesses at most one feasible point \(\rho_m\) that is element-wise no greater than all other feasible points:

\[ \rho_m(x) \leq \rho(x) \quad \forall x \in \mathcal{X}, \quad \rho \in \mathcal{L}. \tag{2.8} \]

If it exists, we refer to \(\rho_m\) as the minimal point of \(\mathcal{L}\). Even though we focus on the LPs in (2.7), the constraints in the definition of \(\mathcal{L}\) may be substantially modified without affecting the results of this paper, see Sec. 6.

### 2.1 Motivating problems

Two problems encountered in the analysis of Markov processes on countably infinite state spaces (Markov chains or chains, for short) encouraged us to develop the approaches presented in this paper. The first is computation of exit distributions and occupation measures associated with the exit times (or first passage times) of the chain. The second is the computation of the stationary distributions of chains.

In what follows, we use \(\{X_n\}_{n \in \mathbb{N}}\) to denote a time-homogeneous discrete-time Markov chain taking values in a countable state space \(\mathcal{S}\). Let \(P := (p(x,y))_{x,y \in \mathcal{S}}\) denote the chain’s one-step matrix and \(\lambda := \{\lambda(x)\}_{x \in \mathcal{S}}\) its initial distribution:

\[ p(x,y) = \mathbb{P}_\lambda \{X_1 = y|\{X_0 = x\}\}, \quad \lambda(x) = \mathbb{P}_\lambda (\{X_0 = x\}), \]

for each \(x, y \in \mathcal{S}\), where \(\mathbb{P}_\lambda\) denotes the underlying probability measure (we append the subscript \(\lambda\) to emphasise that the chain’s starting position is sampled from the distribution \(\lambda\)).

Similarly, we use \(\{X_t\}_{t \geq 0}\) to denote a minimal time-homogeneous continuous-time Markov chain that also takes values in a countable state space \(\mathcal{S}\). Let \(Q := (q(x,y))_{x,y \in \mathcal{S}}\) denote the chain’s rate matrix and \(\lambda := \{\lambda(x)\}_{x \in \mathcal{S}}\) its initial distribution:

\[ q(x,y) = \lim_{t \to 0} \mathbb{P}_\lambda (\{X_t = y, t < T_\infty\}|\{X_0 = x\}), \quad \lambda(x) = \mathbb{P}_\lambda (\{X_0 = x\}), \]

for each \(x, y \in \mathcal{S}\), where \(T_\infty\) denotes the chain’s explosion time and \(\mathbb{P}_\lambda\) denotes the underlying probability measure. We assume that \(Q\) is stable and conservative:

\[ -q(x,x) = \sum_{y \neq x} q(x,y) < \infty \quad \forall x \in \mathcal{S}, \tag{2.9} \]

where the sum is meant to be taken over all states \(y\) in \(\mathcal{S}\) except for \(x\).
Exit distributions and occupation measures

We begin with the discrete-time case and then proceed to the continuous-time one.

**Discrete-time chains**  The exit time $\sigma$ of $\{X_n\}_{n \in \mathbb{N}}$ from a given subset $D$ of the state space (known as the domain) is the first time that the chain lies outside of $D$ (or plus infinity if it never does):

$$\sigma := \inf \{ n \in \mathbb{N} : X_n \in S \setminus D \}.$$

With the exit time $\sigma$, we associate the exit distribution $\mu$ and occupation measure $\nu$ defined by

$$\mu(x) := \mathbb{P}_\lambda(\{X_\sigma = x, \sigma < \infty\}) \quad \forall x \notin D, \quad \nu(x) := \mathbb{E}_\lambda \left[ \sum_{m=0}^{\sigma-1} 1_x(X_m) \right] \quad \forall x \in D.$$

In other words, $\mu(x)$ denotes the probability that the chain exits the domain via state $x$ while $\nu(x)$ denotes the expected number of visits that the chain will make to $x$ before exiting the domain. As shown in [25, Theorem 2.9], if the chain starts inside of the domain with probability one ($\lambda(D) = 1$) and the occupation measure $\nu$ satisfies the moment bound $\nu(w) \leq c$, then the occupation measure is the minimal feasible point of $L$ in (2.6) with

$$\mathcal{X} := D, \quad \mathcal{Y} := S \setminus D, \quad h(x, y) := p(x, y) - 1_x(y) \quad \forall x, y \in D, \quad \phi(x) := -\lambda(x) \quad \forall x \in D, \quad g(x, y) := p(x, y) \quad \forall x \in D, \quad y \notin D,$$

and the exit distribution is given by $\mu = \nu G$. Moments bounds of the sort (2.4) can be obtained analytically using the Foster-Lyapunov criterion in [29, Sec. 14.2.1] or computationally using the semidefinite programming approach discussed in [25, Chap. 4].

**Continuous-time chains**  Just as in the discrete-time case, the exit time $\tau$ of the continuous time chain $\{X_t\}_{t \geq 0}$ from a domain $D$ is the first time that the chain lies outside of $D$:

$$\tau = \inf\{ t \geq 0 : X_t \in S \setminus D \}.$$

With the exit time $\tau$, we associate the exit distribution $\mu$ and occupation measure $\nu$ defined by

$$\mu(x) := \mathbb{P}_\lambda(\{X_\tau = x, \tau < \infty\}) \quad \forall x \notin D, \quad \nu(x) := \mathbb{E}_\lambda \left[ \int_0^{\tau \wedge T} 1_x(X_t) dt \right] \quad \forall x \in D.$$

These tell us where on the boundary of the domain the chain exits ($\mu$) and where inside the domain it spends its time up until the exit time ($\nu$). As shown in [25, Theorem 2.37], if the exit time is almost surely finite, the chain starts inside of the domain ($\lambda(D) = 1$), and the occupation measure $\nu$ satisfies the moment bound $\nu(w) \leq c$, then the occupation measure is the minimal feasible point of $L$ in (2.6) where

$$\mathcal{X} := D, \quad \mathcal{Y} := S \setminus D, \quad h(x, y) := q(x, y) \quad \forall x, y \in D, \quad \phi(x) = -\lambda(x) \quad \forall x \in D, \quad g(x, y) := q(x, y) \quad \forall x \in D, \quad y \notin D,$$

and the exit distribution is given by $\mu = \nu G$. In this case, moment bounds can be obtained using the Foster-Lyapunov criterion [25, Thrm. 2.57] or the computational approaches discussed in [18, 25].

**Stationary distributions**

Once again, we begin with the discrete-time case and then proceed to the continuous-time one.
**Discrete-time chains** A probability measure $\pi$ on $\mathcal{S}$ is said to be a stationary distribution of the chain if sampling its initial position from $\pi$ ensures that the chain remains distributed according to $\pi$ for all time:

$$P_\pi(\{X_n = x\}) = \pi(x) \quad \forall x \in \mathcal{S}, \ n \in \mathbb{N}.$$

Classical theory (e.g., [4, Theorem A.I.3.1]) tells us that the set of stationary distributions that satisfy the moment bound $\pi(w) \leq c$ is $\mathcal{L}$ in (2.6) with

$$\mathcal{X} = \mathcal{Y} := \mathcal{S}, \quad h(x, y) = p(x, y) - 1_x(y) \quad \forall x, y \in \mathcal{S},$$

$$\phi(x) = 0 \quad \forall x \in \mathcal{S}, \quad g(x, y) = 1_x(y) \quad \forall x, y \in \mathcal{S}.$$  (2.12)

In this case, moment bounds can be computed using the Foster-Lyapunov criteria in [16, 29] (and references therein) or the semidefinite programming approach of [22, Sec. 12.4].

**Continuous-time chains** A probability measure $\pi$ on $\mathcal{S}$ is said to be a stationary distribution of the chain if sampling the chain’s starting location from $\pi$ ensures that the chain remains distributed according to $\pi$ for all time:

$$P_\pi(\{X_t = x, t < T_\infty\}) = \pi(x) \quad \forall x \in \mathcal{S}, \ t \geq 0.$$

Assuming that the rate matrix $Q$ is regular, the set of stationary distributions that satisfy the moment bound $\pi(w) \leq c$ is $\mathcal{L}$ in (2.6) with

$$\mathcal{X} = \mathcal{Y} := \mathcal{S}, \quad h(x, y) = q(x, y) \quad \forall x, y \in \mathcal{S},$$

$$\phi(x) = 0 \quad \forall x \in \mathcal{S}, \quad g(x, y) = 1_x(y) \quad \forall x, y \in \mathcal{S},$$  (2.13)

see [25, Thrm. 2.41]. Moment bounds on the stationary distributions of continuous-time chains can be obtained using the Foster-Lyapunov criteria in [28, 16] or the computational approaches in [34, 26, 25, 33, 8, 9, 15]. We extensively study the application of the schemes in this paper to this problem in [26].

### 2.2 Notation and assumptions

Throughout this paper, we adhere to the convention that the supremum (resp. infimum) of the empty set is minus infinity (resp. plus infinity) and we denote the sets of real numbers, natural numbers, and positive integers by $\mathbb{R}$, $\mathbb{N}$, and $\mathbb{Z}_+$. We use $\mathcal{X}$ and $\mathcal{Y}$ to denote two countable sets. We denote the cardinality of $\mathcal{X}$ by $|\mathcal{X}|$ and its power set by $2^\mathcal{X}$. By a “measure $\rho$ on $\mathcal{X}$”, we mean a $\sigma$-finite (possibly signed) measure on $(\mathcal{X}, 2^\mathcal{X})$ and use $\rho \geq 0$ to signify that $\rho$ is unsigned where it may otherwise be ambiguous. We abuse our notation by using $\rho$ to denote both the measure and its density with respect to the counting measure on $(\mathcal{X}, 2^\mathcal{X})$. That is, we write $\rho(x) := \rho(\{x\})$ for each $x \in \mathcal{X}$, so that

$$\rho(A) = \sum_{x \in A} \rho(x), \quad \forall A \subseteq \mathcal{X}.$$  

Given the above, we identify the space of measures, and that of functions, on $\mathcal{X}$ with $\mathbb{R}^{|\mathcal{X}|}$. Clearly, $\rho$ is unsigned if and only if $\rho(x) \geq 0$ for all $x \in \mathcal{X}$. Next, to quantify the error of our approximations, we use the total variation norm:

$$||\rho|| := \sup_{A \subseteq \mathcal{X}} |\rho(A)|.$$  (2.14)

In (2.1)–(2.6), $H := (h(z,x))_{z,x \in \mathcal{X}}$ and $G := (g(x,y))_{x \in \mathcal{X}, y \in \mathcal{Y}}$ denote matrices of real numbers indexed by $\mathcal{X}^2$ and $\mathcal{X} \times \mathcal{Y}$ (respectively) that satisfy Assumption 2.1 below.
Assumption 2.1. (i) $H$ is Metzler and $G$ is nonnegative:

$$h(z, x) \geq 0, \quad \forall x \neq y, \quad g(x, y) \geq 0, \quad \forall x \in \mathcal{X}, \ y \in \mathcal{Y};$$  

the rows of $G$ are summable (i.e., $g(x)$ in (2.3) is finite for all $x \in \mathcal{X}$), and we have at our disposal constants $a_1, a_2, \ldots$ such that

$$\sup_{x \notin \mathcal{X}} \frac{g(x)}{w(x)} \leq a_r \quad \forall r \in \mathbb{Z}_+, \quad \lim_{r \to \infty} a_r = 0.$$  

(ii) For each $x \in \mathcal{X}$, $w(x) > 0$, or $g(x) > 0$, or there exists $x_1, \ldots, x_l \in \mathcal{X}$ such that $g(x_l) > 0$ and

$$h(x, x_1)h(x_1, x_2) \ldots h(x_{l-1}, x_l) > 0.$$

Assumption 2.1(i) ensures that the sums $\rho H(x)$ and $\rho G(y)$ are absolutely convergent for any $x \in \mathcal{X}, y \in \mathcal{Y}$, and feasible point $\rho$ of $\mathcal{L}$. Assumption 2.1(ii) and the constraints $\rho \geq 0$ and $\rho(g) = 1$ in (2.6) guarantee that the entries $\rho(x)$ of the feasible points $\rho$ of $\mathcal{L}$ are bounded: a fact that will be important for the proofs of Secs. 3–4.

In terms of the problems discussed in Sec. 2.1, (2.15) is satisfied because one-step matrices $P$ of discrete-time chains and rate matrices $Q$ of continuous-time chains are Metzler. In the case of the stationary distributions, $G$ is the identity matrix on $\mathbb{R}^{|\mathcal{X}|}$ and so Assumption 2.1(ii) is trivially satisfied. Similarly, our choice of truncations $\mathcal{X}_r$ in (2.5) ensures that (2.16) is satisfied with $a_r := r^{-1}$. Because one-step matrices $P$ of discrete-time chains are row stochastic ($\sum_{y \in \mathcal{X}} p(x, y) = 1$ for all $x$), (2.16) also holds with $a_r := r^{-1}$ for the exit time problem of these chains. In the continuous-time case, the matter is not so simple for the exit time problem unless we pick $w(x)$ to be $(-q(x, x))^d$ for some $d > 1$, in which case (2.9) implies that (2.16) holds with $a_r = r^{1-d}$. However, these bounds $a_r$ for the exit time problem (both in discrete-time and in continuous-time) will be generally conservative: sharper bounds can be obtained on a case-by-case basis. Lastly, even though Assumption 2.1(ii) is not automatically satisfied for the exit time problem, it is a very mild restriction: it asks that the chain must be able to leave the domain from every state $x$ for which $w(x) = 0$. Indeed, if the exit time is almost surely finite, states from which the chain cannot leave the domain are irrelevant to the question as the chain has zero probability of visiting them.

3 Scheme A: Bounding optimal values and approximating optimal points

The starting point of our schemes is replacing the indexing set $\mathcal{X}$ with its truncation $\mathcal{X}_r$ defined in (2.5). We then derive a sequence of computationally tractable sets $\{\mathcal{L}_r\}_{r \in \mathbb{N}}$ that approximate $\mathcal{L}$; computationally tractable meaning that optimising over $\mathcal{L}_r$ consists of solving finite-dimensional linear programs. In particular, we will choose these sets to be outer approximations of $\mathcal{L}$ in the sense that a) if $\rho$ belongs to $\mathcal{L}$, then its restriction $\rho|_{\mathcal{X}_r}$ to the truncation $\mathcal{X}_r$ belongs to $\mathcal{L}_r$, where the restriction is defined by:

$$\rho_{|r}(x) := \begin{cases} \rho(x) & \text{if } x \in \mathcal{X}_r \\ 0 & \text{if } x \notin \mathcal{X}_r \end{cases} \quad \forall x \in \mathcal{X};$$  

and b) $\mathcal{L}_r$ is approximately equal to $\mathcal{L}$ for large enough $r$ (we formalise this in Sec. 3.2).
To obtain these outer approximations, we make one further assumption: each equation in (2.1) only involves finitely many entries of \( \rho \) (this assumption can be circumvented, see App. B). Hence the support of each of the columns of \( H \) is finite:

\[
\text{supp}(h(\cdot, x)) := \{ z \in \mathcal{X} : h(z, x) \neq 0 \} \text{ is finite, for each } x \in \mathcal{X}; \tag{3.2}
\]
similar assumptions can be found in the works [35, 13, 14] discussing approximation schemes for other types of CILPs. In the context of the Markov chains in Sec. 2.1, (3.2) requires that each state is reachable in a single jump from at most finitely many others.

### 3.1 Outer approximations and bounding the optimal values

Let \( \mathcal{E}_r \) denote the set of indices \( x \) in the truncation such that the support of the corresponding column \( h(\cdot, x) \) of \( H \) is contained in the truncation:

\[
\mathcal{E}_r := \{ x \in \mathcal{X}_r : \text{supp}(h(\cdot, x)) \subseteq \mathcal{X}_r \}.
\]

Because, for any \( x \) in \( \mathcal{E}_r \), the equation \( \rho H(x) = \phi(x) \) only involve entries \( \rho(z) \) of \( \rho \) indexed by \( z \)s inside the truncation \( \mathcal{X}_r \),

\[
\rho|_r H(x) = \rho H(x) = \phi(x) \quad \forall x \in \mathcal{E}_r, \quad \rho \in \mathcal{L}. \tag{3.3}
\]

The first step in deriving our outer approximation is to strike all other equations from (2.1) resulting in a set of finitely many equations each involving finitely many unknowns (i.e., \(|\mathcal{E}_r| \leq |\mathcal{X}_r| \) equations and \(|\mathcal{X}_r| \) unknowns).

Because \( w \) is a nonnegative function, it follows immediately from the definition (2.6) of \( \mathcal{L} \) that

\[
\rho|_r(w) \leq \rho(w) \leq c, \quad \rho|_r \geq 0, \quad \forall \rho \in \mathcal{L}. \tag{3.4}
\]

Given (3.3)–(3.4), we only need now to approximate the constraints “\( \rho(g) = 1 \)” with one that only involves \( \rho|_r \). To do so, we use the moment bound (2.4) and the following generalisation of Markov’s inequality that allows us to control the tail \( \{ \rho(x) \}_{x \notin \mathcal{X}_r} \) of \( \rho \):

\[
\rho(1_{\mathcal{X}_c} f) = \sum_{x \notin \mathcal{X}_r} \rho(x) f(x) \leq \left( \sup_{x \notin \mathcal{X}_r} \frac{f(x)}{w(x)} \right) \sum_{x \notin \mathcal{X}_r} w(x) \rho(x) \leq \left( \sup_{x \notin \mathcal{X}_r} \frac{f(x)}{w(x)} \right) \rho(w) \leq c \left( \sup_{x \notin \mathcal{X}_r} \frac{f(x)}{w(x)} \right), \tag{3.5}
\]
as long as \( f \) is a nonnegative function on \( \mathcal{X} \) and \( \rho \) belongs to \( \mathcal{L} \), where \( 1_{\mathcal{X}_c} \) denotes the indicator function of the complement \( \mathcal{X}_c \) of the truncation \( \mathcal{X}_r \). Setting \( f \) to be the row sums \( g \) of \( G \) in (2.3), we find that

\[
1 - ca_r \leq 1 - c \left( \sup_{x \notin \mathcal{X}_r} \frac{g(x)}{w(x)} \right) \leq 1 - \rho(1_{\mathcal{X}_c} g) = \rho(1_{\mathcal{X}_c} g) = \rho|_r(g) \leq \rho(g) = 1, \tag{3.6}
\]
for all \( \rho \) in \( \mathcal{L} \), where the first inequality follows from (2.16), the second from Markov’s inequality (3.5), and the first and last equalities follow from the constraint \( \rho(g) = 1 \) in (2.6).
Putting (3.3)–(3.6) together shows that

\[
\mathcal{L}_r := \left\{ \rho \in \mathbb{R}^{\left| \mathcal{X}_r \right|} : \begin{array}{l}
\rho H(x) = \phi(x) \quad \forall x \in \mathcal{X}_r, \\
1 - c a_r \leq \rho(g) \leq 1, \\
\rho(w) \leq c, \\
\rho \geq 0, \\
\rho(\mathcal{X}_r^c) = 0.
\end{array} \right. \tag{3.7}
\]

is an outer approximation of \( \mathcal{L} \) in (2.6). In contrast with \( \mathcal{L} \) which involves \( \left| \mathcal{X} \right| \) decision variables and a comparable number of constraints, \( \mathcal{L}_r \) only involves \( \left| \mathcal{X}_r \right| \) decision variables, \( \left| \mathcal{E}_r \right| \) equality constraints, and \( \left| \mathcal{X}_r \right| + 3 \) inequality constraints (note that the constraints \( \rho(\mathcal{X}_r^c) = 0 \) and \( \rho \geq 0 \) imply that \( \rho(x) = 0 \) for every \( x \) outside of the truncation \( \mathcal{X}_r \)). Because \( w \) is a norm-like function, \( \left| \mathcal{X}_r \right| \) is a finite number for any positive integer \( r \) and so \( \mathcal{L}_r \) is the feasible set of a finite-dimensional linear program regardless of whether \( \left| \mathcal{X} \right| \) is finite.

Due to the outer approximation property, we obtain bounds on the feasible points of \( \mathcal{L} \) by optimising over \( \mathcal{L}_r \). In particular, for any real-valued function \( f \) on \( \mathcal{X} \), consider the finite-dimensional LPs:

\[
l_f^r := \inf \{ \rho(f) : \rho \in \mathcal{L}_r \}, \quad u_f^r := \sup \{ \rho(f) : \rho \in \mathcal{L}_r \}. \tag{3.8}
\]

Because \( \mathcal{L}_r \) is an outer approximation of \( \mathcal{L} \) and \( \rho(f) = \rho(1_{\mathcal{X}_r} f) + \rho(1_{\mathcal{X}_r^c} f) \), we have that

\[
l_f^r + \rho(1_{\mathcal{X}_r^c} f) \leq \rho(f) \leq u_f^r + \rho(1_{\mathcal{X}_r} f).
\]

For this reason, \( \rho(f) \geq l_f^r \) if \( f \) is non-negative outside of the truncation. If instead it is non-positive, we have that \( \rho(f) \leq u_f^r \). Otherwise, we use Markov’s inequality (3.5) to bound \( \left| \rho(1_{\mathcal{X}_r} f) \right| \):

\[
\left| \rho(1_{\mathcal{X}_r} f) \right| \leq \rho(1_{\mathcal{X}_r} |f|) \leq c \left( \sup_{x \in \mathcal{X}_r} \frac{|f(x)|}{w(x)} \right). \tag{3.9}
\]

In summary, we have the following.

**Lemma 3.1.** Suppose that Assumption 2.1 is satisfied. If \( \mathcal{L} \) is non-empty, then \( \mathcal{L}_r \) is non-empty for each \( r > 0 \). Moreover, if \( f \) is any real-valued function on \( \mathcal{X} \), \( l_f^r \) and \( u_f^r \) are defined as in (3.8), and \( \rho \) belongs to \( \mathcal{L} \), then

\[
l_f^r \leq \rho|_{\mathcal{X}}(f) \leq u_f^r,
\]

where \( \rho|_{\mathcal{X}} \) is the restriction of \( \rho \) to \( \mathcal{X} \) defined in (3.1). Consequently, if \( f \) is \( \rho \)-integrable, we have that

\[
l_f^r - c \left( \sup_{x \notin \mathcal{X}_r} \frac{|f(x)|}{w(x)} \right) \leq \rho(f) \leq u_f^r + c \left( \sup_{x \notin \mathcal{X}_r} \frac{|f(x)|}{w(x)} \right). \tag{3.10}
\]

Furthermore, if \( f \) is non-negative (resp. non-positive) on \( \mathcal{X}_r^c \), then

\[
l_f^r \leq \rho(f) \quad (\text{resp. } \rho(f) \leq u_f^r). \tag{3.11}
\]

**Proof.** Shown in (3.3)–(3.9). □

Putting together the lower and upper bound in (3.10)–(3.11), we quantify the uncertainty in our knowledge of the \( \rho(f) \). Moreover, because the bounds (3.10)–(3.11) hold for all feasible points \( \rho \) of \( \mathcal{L} \) they also hold if we replace the average \( \rho(f) \) with the corresponding optimal values in (2.7). In the next section, we show that the sequences \( \{l_f^r\}_{r \in \mathbb{Z}_+} \) and \( \{u_f^r\}_{r \in \mathbb{Z}_+} \), obtained by computing these bounds for ever larger truncations \( \mathcal{X}_r \) converge to \( l_f \) and \( u_f \), respectively.
3.2 The scheme’s convergence

Our definition (2.5) of the truncations implies that they form an increasing sequence with \( \mathcal{X} \) as its limit: \( \mathcal{X}_1 \subseteq \mathcal{X}_2 \subseteq \ldots, \quad \bigcup_{r=1}^{\infty} \mathcal{X}_r = \mathcal{X}. \) (3.12)

Here, we show that the outer approximations \( \mathcal{L}_r \) converge in a strong manner to \( \mathcal{L} \). The type of convergence we deal with is \( \rho^r \to \rho \) as \( r \to \infty \) if and only if

\[
\lim_{r \to \infty} \rho^r(f) = \rho(f) \quad \forall f \in \mathcal{W},
\] (3.13)

where \( \mathcal{W} \) denotes the vector space of real-valued functions on \( \mathcal{X} \) that eventually grow strictly slower than the norm-like function \( w \) in (2.4):

\[
\mathcal{W} := \left\{ f \in \mathbb{R}^{|\mathcal{X}|} : \lim_{r \to \infty} \sup_{x \not\in \mathcal{X}_r} \frac{|f(x)|}{w(x)} = 0 \right\}. \] (3.14)

If (3.13) holds, we say that \( \rho^r \) converges to \( \rho \) in weak*. In our setting of countable spaces \( \mathcal{X} \), weak* convergence implies convergence in total variation [25, Rem. 5.10]. The next theorem shows that any sequence \( \{\rho^r\}_{r \in \mathbb{Z}^+} \) of feasible points \( \rho^1 \in \mathcal{L}_1, \rho^2 \in \mathcal{L}_2, \ldots \) is weak* sequentially compact and that each of its accumulation points belongs to \( \mathcal{L} \). In other words, each subsequence \( \{\rho^{r_k}\}_{k \in \mathbb{Z}^+} \) of \( \{\rho^r\}_{r \in \mathbb{Z}^+} \) has a weak* converging subsequence \( \{\rho^{r_{i_k}}\}_{i \in \mathbb{Z}^+} \) and its limit belongs to \( \mathcal{L} \).

**Theorem 3.2.** Suppose that \( w \) is a norm-like function, that Assumption 2.1 and (3.2) are satisfied, and that \( \mathcal{L} \) is non-empty. For each \( r > 0 \), \( \mathcal{L}_r \) in (3.7) is non-empty. Moreover, if \( \rho^r \) belongs to \( \mathcal{L}_r \) for each \( r \in \mathbb{Z}^+ \), then \( \{\rho^r\}_{r \in \mathbb{Z}^+} \) is weak* sequentially compact, and all of its accumulation points belong to \( \mathcal{L} \) in (2.6). Consequently, for each \( f \in \mathcal{W} \), the sequences \( \{l^r_f\}_{r \in \mathbb{Z}^+} \) and \( \{u^r_f\}_{r \in \mathbb{Z}^+} \) defined in (3.8) converge and

\[
\lim_{r \to \infty} l^r_f = l_f \quad \lim_{r \to \infty} u^r_f = u_f
\]

with \( l_f \) and \( u_f \) as in (2.7).

**Proof.** See Appendix A. \( \square \)

Theorem 3.2 tells us that, for sufficiently large \( r \), the optimal values \( l^r_f \) and \( u^r_f \) in (3.8) of the finite-dimensional LPs are close to the optimal values \( l_f \) and \( u_f \) of the the infinite-dimensional LP. As we now discuss, the same also holds for the optimal points of \( \mathcal{L}_r \) achieving \( l^r_f \) and \( u^r_f \) and those of \( \mathcal{L} \) achieving \( l_f \) and \( u_f \).

3.3 Approximating the optimal points and their images

We have the following corollary of Theorem 3.2.

**Corollary 3.3.** Suppose that the premise of Theorem 3.2 is satisfied.

(i) Suppose that there exists a unique \( \rho_* \) in \( \mathcal{L} \) such that \( \rho_*(f) = l_f \) (resp. \( \rho_*(f) = u_f \)) for a given \( f \) in \( \mathcal{W} \). If, for each positive integer \( r \), \( \rho^r \) in \( \mathcal{L}_r \) is such that \( \rho^r(f) = l^r_f \) (resp. \( \rho^r(f) = u^r_f \)), then the sequence \( \{\rho^r\}_{r \in \mathbb{Z}^+} \) converges in weak* to \( \rho_* \).

(ii) Suppose that \( \mathcal{L} \) is the singleton \( \{\rho\} \). If \( \rho^r \) belongs \( \mathcal{L}_r \) for each positive integer \( r \), then the sequence \( \{\rho^r\}_{r \in \mathbb{Z}^+} \) converges in weak* to \( \rho \).
Proof. (i) We only show this for \( \{ l^r \}_{r \in \mathbb{Z}_+} \): the proof for the \( \{ u^r \}_{r \in \mathbb{Z}_+} \) is entirely analogous. Theorem 3.2 tells us that \( \{ \rho^r \}_{r \in \mathbb{Z}_+} \) is weak* sequentially compact with accumulation points in \( L \). Pick any accumulation point \( \rho \) of \( \{ \rho^r \}_{r \in \mathbb{Z}_+} \) and let \( \{ \rho^r_j \}_{j \in \mathbb{Z}_+} \) be any subsequence of \( \{ \rho^r \}_{r \in \mathbb{Z}_+} \) converging to \( \rho \). Theorem 3.2 implies that,

\[
\rho(f) = \lim_{j \to \infty} \rho^r_j(f) = l^r_f = l_f = \rho_* (f).
\]

Because \( \rho \) belongs to \( L \), the uniqueness of \( \rho_* \) then implies that \( \rho = \rho_* \). Since this holds for any accumulation point \( \rho \) of \( \{ \rho^r \}_{r \in \mathbb{Z}_+} \) and the sequence is sequentially compact, we have that \( \{ \rho^r \}_{r \in \mathbb{Z}_+} \) converges in weak* to \( \rho_* \). (ii) This follows immediately from (i) by setting \( f := 0 \). \qed

In other words, if \( L \) is a singleton \( \{ \rho \} \), then Corollary 3.3(ii) shows that any feasible point \( \rho^r \) of \( L_r \) is close to \( \rho \) for large enough \( r \). If \( L \) is not a singleton, then Corollary 3.3(i) justifies using Scheme A to approximate optimal points \( \rho_* \) of \( L \) achieving infimum (resp. supremum) \( \rho_* (f) = l_f \) (resp. \( \rho_* (f) = u_f \)), assuming that these are unique.

To approximate the images \( \psi \) through \( G \) of the feasible points \( \rho \) (see (2.2)), we combine Corollary 3.3 with the following lemma.

**Lemma 3.4.** Suppose that Assumption 2.1 (i) is satisfied, that \( w \) is norm-like, and that \( \rho^1 \in L_1, \rho^2 \in L_2, \ldots \) converges in weak* to \( \rho \in L \). If \( \psi^1 := \rho^1 G, \psi^2 := \rho^2 G, \ldots \), then \( \psi^r \) converges in total variation to \( \psi := \rho G \) as \( r \) tends to infinity.

**Proof.** Using the definition of the total variation norm (2.14) we have that

\[
||\psi^r - \psi|| \leq \sum_{y \in \mathcal{Y}} |\psi^r(y) - \psi(y)| \leq \sum_{x \in \mathcal{X}_a} |\rho^r(x) - \rho(x)| g(x) + \sum_{x \notin \mathcal{X}_a} (\rho^r(x) + \rho(x)) g(x),
\]

for any \( r, n \in \mathbb{Z}_+ \). Due to (2.16), the remainder of the proof is analogous to Part (c) of the proof of Theorem 3.2. \qed

Suppose that \( L \) has a unique feasible point \( \rho \). In practice, we are faced with the crucial question: “what \( r \) should we pick to ensure that Scheme A returns accurate approximations of \( \rho \)?” Herein lies the main drawback of Scheme A: we know of no practical way of deciding whether the truncation employed is large enough. To overcome this issue, we introduce Scheme B in the next section. It computes a collection of element-wise lower bounds on the unknown measure \( \rho \) by repeatedly applying Scheme A. Using this collection of lower bounds as an approximation of \( \rho \), it is straightforward to bound its error (defined as the total variation distance between the lower bounds and \( \rho \)). If this error bound proves to be unsatisfactorily large, we increase the truncation, re-compute the bounds, and re-evaluate the error bound. As we will show, the collection of bounds also converges in weak* to \( \rho \) and the error bound converges to zero as the truncation approaches the entire index set \( \mathcal{X} \). For this reason, by repeatedly increasing \( r \), we will always be able to achieve any desired error tolerance.

### 4 Scheme B: Approximating the minimal point

Suppose that \( L \) has a minimal point meaning a feasible point \( \rho_m \) satisfying (2.8). Let

\[
l^r(x) := \begin{cases} 
  l^r_x & \text{if } x \in \mathcal{X}, \\
  0 & \text{if } x \notin \mathcal{X},
\end{cases} \quad \forall x \in \mathcal{X}, \ r \in \mathbb{Z}_+,
\]

(4.1)

where \( l^r_x \) is as in (3.8) with \( f := 1_{\{x\}} \). Lemma 3.1 shows that \( l^r \) bounds all feasible points \( \rho \) of \( L \) from below:

\[
l^r(x) \leq \rho(x), \quad \forall x \in \mathcal{X}, \ r \in \mathbb{Z}_+.
\]

(4.2)
Theorem 3.2 tell us that $l^r$ converges pointwise to $\rho_m$:
\[
\lim_{r \to \infty} l^r(x) = \rho_m(x) \quad \forall x \in \mathcal{X}.
\] (4.3)

These lower bounds actually converge in weak*. Furthermore, in contrast with the approximations $\rho^r$ of Scheme A, bounding the total variation distance between $l^r$ and $\rho_m$ is straightforward:

**Theorem 4.1.** Suppose that premise of Theorem 3.2 is satisfied.

(i) For each positive integer $r$,
\[
||\rho - l^r|| = \rho(\mathcal{X}) - l^r(\mathcal{X}) \leq u^r_{\mathcal{X}} + \frac{c}{r} - l^r(\mathcal{X}) \quad \forall \rho \in \mathcal{L},
\]
where $u^r_{\mathcal{X}}$ is as in (3.8) with $f := 1_{\mathcal{X}}$.

(ii) If $\mathcal{L}$ has a minimal point $\rho_m$ (in the sense of (2.8)), then $l^r$ converges to $\rho_m$ in weak* as $r$ tends to infinity.

(iii) If $\mathcal{L}$ has a unique feasible point $\rho$, then not only does the error $||\rho - l^r||$ tend to zero as $r$ approaches infinity, but so does the error bound $u^r_{\mathcal{X}} + c/r - l^r(\mathcal{X})$.

**Proof.** (i) Because (4.2) shows that $\rho - l^r$ is an unsigned measure for each $\rho \in \mathcal{L}$, this follows immediately from Lemma 3.1 and the fact that the total variation norm of an unsigned measure is its mass. (ii) Theorem 3.2 shows that the limit $l(x) := \lim_{r \to \infty} l^r(x)$ exists for each $x \in \mathcal{X}$. Fatou’s Lemma and (4.2)–(4.3) show that
\[
l(w) \leq \lim_{r \to \infty} l^r(w) \leq \lim_{r \to \infty} \rho_m(w) \leq \rho_m(w) \leq c.
\] (4.4)

Pick any $f \in \mathcal{W}$, fix $r,n \in \mathbb{Z}_+$, and note that
\[
|\rho_m(f) - l^r(f)| \leq \sum_{x \in \mathcal{X}} (\rho_m(x) - l^r(x)) |f(x)| + \sum_{x \not\in \mathcal{X}} (\rho_m(x) + l^r(x)) |f(x)|.
\]

Given (4.3), (4.4), and the above, the remainder of the proof is analogous to Part (c) of the proof of Theorem 3.2. (iii) This follows immediately from (ii) and Theorem 3.2.

To approximate the image $\psi_m := \rho_m G$ through $G$ of the minimal solution $\rho_m$ we use
\[
l^r_\psi(y) := \begin{cases} l^r_{g_y} & \text{if } y \in \mathcal{Y}_r \\ 0 & \text{if } y \not\in \mathcal{Y}_r \end{cases} \quad \forall y \in \mathcal{Y},
\]
where $l^r_{g_y}$ is as in (3.8) with $f(\cdot) = g_y(\cdot) := g(\cdot, y)$ and $\mathcal{Y}_r$ is any finite subset of $\mathcal{Y}$.

**Corollary 4.2.** Suppose that premise of Theorem 3.2 is satisfied and that $\mathcal{Y}_1, \mathcal{Y}_2, \ldots$ is an increasing sequence of finite subsets of $\mathcal{Y}$ such that $\bigcup_{r=1}^{\infty} \mathcal{Y}_r = \mathcal{Y}$.

(i) For each positive integer $r$,
\[
l^r_\psi(y) \leq \rho G(y) \quad \forall y \in \mathcal{Y}, \quad \rho \in \mathcal{L}.
\]

(ii) For each positive integer $r$,
\[
||\rho G - l^r_\psi|| = 1 - l^r(g) \quad \forall \rho \in \mathcal{L}.
\]

11
(iii) If \( L \) has a minimal point \( \rho_m \) (in the sense of (2.8)), then \( \psi_m := \rho_m G = \rho G \) for all \( \rho \in L \) and the lower bounds \( l^r_\psi \) converge in total variation to \( \psi_m \) as \( r \) tends to infinity.

Proof. (i) This follows directly from (4.2) and non-negativity of \( G \). (ii) Part (i) tells us that \( \rho G - l^r_\psi \) is an unsigned measure on \( (\mathcal{Y}, 2^\mathcal{Y}) \) and the result follows. (iii) Inequality (2.8) and non-negativity of \( G \) implies that \( \rho_m G(y) \leq \rho G(y) \) for every \( y \in \mathcal{Y} \) and \( \rho \in L \). Thus,

\[
\sum_{y \in \mathcal{Y}} |\rho G(y) - \rho_m G(y)| = \sum_{y \in \mathcal{Y}} (\rho G(y) - \rho_m G(y)) = \rho(g) - \rho_m(g) = 1 - 1 = 0,
\]

showing that \( \rho G = \rho_m G \) for each \( \rho \in L \). Fix an \( \varepsilon > 0 \). Note that \( \psi_m(\mathcal{Y}) = \rho_m(g) = 1 \) so there exists an \( R \) such that \( \psi_m(\mathcal{Y}_R^c) \leq \varepsilon/4 \). Thus,

\[
||\psi_m - l^r_\psi|| \leq \sum_{y \in \mathcal{Y}_R} |\psi_m(y) - l^r_\psi(y)| + \sum_{y \notin \mathcal{Y}_R} (\psi_m(y) + l^r_\psi(y)) \leq \sum_{y \in \mathcal{Y}_R} (\psi_m(y) - l^r_\psi(y)) + 2\psi_m(\mathcal{Y}_R^c) \leq \sum_{y \in \mathcal{Y}_R} (l_{g_y} - l^r_{g_y}) + \frac{\varepsilon}{2},
\]

where \( l_{g_y} \) is as in (2.7) with \( f := g_y \), the first inequality follows from non-negativity of \( \psi_m \) and \( l^r_\psi \), the second from (i), and the final from the fact that \( \rho G = \rho_m G \) for each \( \rho \in L \). Because the non-negativity of \( G \) and (2.16) implies that \( g_y \) belongs to \( \mathcal{W} \) for each \( y \in \mathcal{Y} \), finiteness of \( \mathcal{Y}_R \) and Theorem 3.2 imply that \( \sum_{y \in \mathcal{Y}_R}(l_{g_y} - l^r_{g_y}) \) is smaller than \( \varepsilon/2 \) for all sufficiently large \( r \) completing the proof.

We refer to the process of computing the lower bounds \( l^r \) (or \( l^r_\psi \)) and evaluating the error using Theorem 4.1 (or its corollary) as “Scheme B”. The error control of Scheme B comes at a price: the scheme is more computationally expensive than Scheme A as it entails solving \( |\mathcal{X}_r| \) (or \( |\mathcal{Y}_r| \)) LPs instead of a single one. Parallelising the computation of the bounds mitigates some of this extra cost. Lastly, replacing \( l^r_\psi \) with \( u^r_\psi \) in (4.1), where \( u^r_\psi \) denotes \( u^r_\psi \) in (3.8) with \( f := 1_x \), we obtain a measure \( u^r \) of upper bounds on the feasible points of \( L \) analogous to \( l^r \) (also accompanied by an error bound). Even though Lemma 3.1 shows that \( u^r \) converges pointwise, no weak* convergence can be recovered, see [25] for details.

5 Motivating problems revisited

Before concluding the paper, we quickly point out how to apply the results of Secs. 3–4 to the problems of Sec. 2.1 that motivate this work.

Exit distributions and occupation measures

With \( H \) and \( G \) as in (2.10) (resp. (2.11) in the continuous-time case), Corollary 3.3(i) (with \( f := 1 \)) and Lemma 3.4 yield compute converging approximations of the occupation measure \( \nu \) and exit distribution \( \mu \). If error bounds are important, then Theorem 4.1 and Corollary 4.2 yield approximations accompanied by such bounds.

Stationary distributions

In the case of a unique stationary distribution \( \pi \) satisfying the moment bound \( \pi(w) \leq c \), Corollary 3.3(ii) yields converging approximations of \( \pi \) as long as \( H \) and \( G \) are as in (2.12) (resp. (2.13) in
exists a \( \pi \) be a singleton. Otherwise (5.1) and the disjointness of the communicating classes imply that there
f with \( L \) of the feasible points \( \rho \) can be removed from (2.6) and/or any equality
\[ \rho_{H} = \phi \] as in ours. In their words, we obtain the approximating finite-dimensional LP
approximating an infinite-dimensional LP by a finite-dimensional one in their schemes is the same
for infinite-dimensional linear programs that are closely related to ours. The underlying idea of approximating
an infinite-dimensional LP by a finite-dimensional one in their schemes is the same
as in ours. In their words, we obtain the approximating finite-dimensional LP \( \mathcal{L}_{r} \) by aggregating the
constraints \( \{ \rho H = \phi(x) \ \forall x \in X \} \) in \( \mathcal{L} \) into \( \{ \rho H(x) = \phi(x) \ \forall x \in X \} \) and relaxing the constraint
\( \rho(g) = 1 \) in \( \mathcal{L} \) into \( 1 - c a_{r} \leq \rho(g) \leq 1 \). Similarly, in the proof of Theorem 3.2 we employed ideas analogous to theirs even if their proofs require more technically involved results (e.g. Prokhorov’s Theorem or functional analytic machinery) due to the more general setting they consider.

The main difference between their approach and ours is that they do not employ an explicit
moment bound in their LPs. Instead, they choose an objective that consists of minimising the
integral of a norm-like function to recover the weak* sequential compactness. Because of their fixed
objective, they only obtain lower bounds on the optimal value. The absence of upper bounds makes it
difficult to bound the approximation error. Similarly, because they cannot bound other averages,
it is not possible to construct measures of bounds with easy-to-quantify-errors along the lines of \( l^{r} \)

6 Concluding remarks

In this paper, we approximate a class of countably infinite linear programs with finite-dimensional
ones and show how to bound their approximation errors. As discussed at the end of Sec. 3, any
desired error tolerance is achieved by iterating our approach with ever increasing truncations. The
convergence results of Secs. 3–4 guarantee that this procedure will terminate.

To simplify the exposition we focused on the particular type of LP in (2.7) motivated by the
problems discussed in Sec. 2.1. However, our approximation schemes can be modified to apply to
other CILPs that are not necessarily tied to Markov chains. Specifically, the constraints \( \rho H = \phi \) can be
removed from (2.6) and/or any equality \( \rho(f) = \alpha \) or inequality \( \rho(f) \leq \alpha \) can be added as long
as \( f \) is a function belonging to \( \mathcal{W} \). The techniques of Sec. 3 carry over if the constraints of the new
CILP either involve only finitely many entries of the feasible points \( \rho \), or are of the type \( \rho(f) \leq \alpha \)
for a nonnegative function \( f \). Otherwise, the techniques of App. B apply. The critical ingredients
required to guarantee convergence of the schemes are a moment bound and the boundedness of the
entries of the feasible points \( \rho \), which in our case this followed from Assumption 2.1(ii) and the
constraints \( \rho(g) = 1 \) and \( \rho \geq 0 \).

Hernández-Lerma and Lasserre [19, 20, 21, 22] introduced a series of approximation schemes
for infinite-dimensional linear programs that are closely related to ours. The underlying idea of approximating
an infinite-dimensional LP by a finite-dimensional one in their schemes is the same
as in ours. In their words, we obtain the approximating finite-dimensional LP \( \mathcal{L}_{r} \) by aggregating the
constraints \( \{ \rho H(x) = \phi(x) \ \forall x \in X \} \) in \( \mathcal{L} \) into \( \{ \rho H(x) = \phi(x) \ \forall x \in X \} \) and relaxing the constraint
\( \rho(g) = 1 \) in \( \mathcal{L} \) into \( 1 - c a_{r} \leq \rho(g) \leq 1 \). Similarly, in the proof of Theorem 3.2 we employed ideas analogous to theirs even if their proofs require more technically involved results (e.g. Prokhorov’s Theorem or functional analytic machinery) due to the more general setting they consider.

The main difference between their approach and ours is that they do not employ an explicit
moment bound in their LPs. Instead, they choose an objective that consists of minimising the
integral of a norm-like function to recover the weak* sequential compactness. Because of their fixed
objective, they only obtain lower bounds on the optimal value. The absence of upper bounds makes it
difficult to bound the approximation error. Similarly, because they cannot bound other averages,
it is not possible to construct measures of bounds with easy-to-quantify-errors along the lines of \( l^{r} \)

\[ \mathcal{L} = \left\{ \sum_{i \in I} \theta_{i} \pi_{i} : \theta_{i} \geq 0 \ \forall i \in I, \ \sum_{i \in I} \theta_{i} = 1 \right\} \] (5.1)

where \( \{ \pi_{i} : i \in I \} \) denotes the set of ergodic distributions satisfying \( \pi_{i}(w) \leq c \). Each of these
distributions \( \pi_{i} \) has support on a closed communicating class \( C_{i} \subseteq S \) of the state space. Because
these classes are disjoint sets, we obtain converging approximations of \( \{ \pi_{i} : i \in I \} \) with \( L \) by using Corollary 3.3(i)
with \( f := 1_{x} \) for any state \( x \) inside the class \( C_{i} \). Moreover, because \( L_{r} \) is an outer approximation
of \( \mathcal{L} \), our lower bounds function as a uniqueness test: if \( l^{r}(x) > 0 \) for some \( r \) and \( x \), then \( \mathcal{L} \) must
be a singleton. Otherwise (5.1) and the disjointness of the communicating classes imply that there
exists a \( \pi \) in \( \mathcal{L} \) such that \( \pi(x) = 0 \) contradicting the lower-bound property of \( l^{r}(x) \). See [26] and
[25, Chap. 6.2] for the details.
and $l^p$ in Sec. 4. Conversely, their approach has the advantage that it does not rely on a moment bound but only assumes that optimal value is finite.

The applications in [19, 20, 21, 22] involve uncountable index sets $X$. To obtain finite-dimensional LPs, Hernández-Lerma and Lasserre must discretise $X$ in one manner or another (they find a dense countable subset of a space that contains the measures they wish to approximate). Because of this discretisation, the restrictions of the measure of interest are not necessarily feasible points of the finite-dimensional LPs. For this reason, their schemes yield only converging approximations of the quantities of interest instead of converging bounds. Since using an explicit moment bound does not overcome this difficulty, we did not attempt to extend the schemes beyond the countable case. For a more detailed comparison, see [25, Sec. 5.5].

Finally, we mention that our method can be adapted to compute marginal probability distributions by redefining $G$, see [26, Sec. IV.B.3] or [25, Sec. 5.1] for details and [26, Sec. 5] for a concrete example. In this case, we generally have to solve far less LPs than in that for the full distribution.

References

[1] E. Altman. Denumerable constrained Markov decision processes and finite approximations. Math. Oper. Res., 19(1):169–191, 1994.

[2] E. Altman. Constrained Markov decision processes. Chapman & Hall/CRC, 1999.

[3] E. J. Anderson and P. Nash. Linear programming in infinite-dimensional spaces: theory and applications. Wiley, 1987.

[4] S. Asmussen. Applied Probability and Queues. Springer-Verlag New York, 2003.

[5] R. Cavazos-Cadena. Finite-state approximations for denumerable state discounted markov decision processes. Appl. Math. Optim., 14(1):1–26, 1986.

[6] K. L. Chung. Markov chains with stationary transition probabilities. Springer, 2nd edition, 1967.

[7] E. D. Demaine, S. P. Fekete, and S. Gal. Online searching with turn cost. Theor. Comput. Sci., 361(2-3):342–355, sep 2006.

[8] G. R. Dowdy and P. I. Barton. Using Semidefinite Programming to Calculate Bounds on Stochastic Chemical Kinetic Systems at Steady State. Computer Aided Chemical Engineering, 40:2239–2244, 2017.

[9] G. R. Dowdy and P. I. Barton. Bounds on stochastic chemical kinetic systems at steady state. J. Chem. Phys., 148(8):084106, 2018.

[10] E. Feinberg and A. Shwartz. Handbook of Markov decision processes: methods and algorithms. Kluwer, Boston, 2002.

[11] A. Ghate. Circumventing the Slater conundrum in countably infinite linear programs. Eur. J. Oper. Res., 246(3):708–720, 2015.

[12] A. Ghate. Inverse optimization in countably infinite linear programs. Oper. Res. Lett., 43(3):231–235, 2015.
[13] A. Ghate, D. Sharma, and R. L. Smith. A shadow simplex method for infinite linear programs. *Oper. Res.*, 58(4-part-1):865–877, 2010.

[14] A. Ghate and R. L. Smith. A linear programming approach to nonstationary infinite-horizon Markov decision processes. *Oper. Res.*, 61(2):413–425, 2013.

[15] K. R. Ghusinga, C. A. Vargas-Garcia, A. Lamperski, and A. Singh. Exact lower and upper bounds on stationary moments in stochastic biochemical systems. *Phys. Biol.*, 14(4):04LT01, 2017.

[16] P. W. Glynn and A. Zeevi. Bounding stationary expectations of Markov processes. In *Markov Processes and Related Topics: A Festschrift for Thomas G. Kurtz*, pages 195–214. Institute of Mathematical Statistics, 2008.

[17] Richard C. Grinold. Finite horizon approximations of infinite horizon linear programs. *Math. Program.*, 12(1):1–17, 1977.

[18] K. Helmes, S. Röhl, and R. H. Stockbridge. Computing moments of the exit time distribution for Markov processes by linear programming. *Oper. Res.*, 49(4):516–530, 2001.

[19] O. Hernández-Lerma and J. B. Lasserre. Approximation schemes for infinite linear programs. *SIAM J. Optim.*, 8(4):973–988, 1998.

[20] O. Hernández-Lerma and J. B. Lasserre. Linear programming approximations for Markov control processes in metric spaces. *Acta Appl. Math.*, 51(2):123–139, 1998.

[21] O. Hernández-Lerma and J. B. Lasserre. *Further topics on discrete-time Markov control processes*. Springer, 1999.

[22] O. Hernández-Lerma and J. B. Lasserre. *Markov chains and invariant probabilities*. Birkhäuser, 2003.

[23] K. Huang. *Multi-stage stochastic programming models in production planning*. PhD thesis, Georgia Institute of Technology, 2005.

[24] P. C. Jones, J. L. Zydiak, and W. J. Hopp. Stationary dual prices and depreciation. *Math. Program.*, 41(1-3):357–366, 1988.

[25] J. Kuntz. *Deterministic approximation schemes with computable errors for the distributions of Markov chains*. PhD thesis, Imperial College London, 2017.

[26] J. Kuntz, P. Thomas, G. B. Stan, and M. Barahona. Rigorous bounds on the stationary distributions of the chemical master equation via mathematical programming. 2017.

[27] I. Lee, M. A. Epelman, H. Edwin Romeijn, and R. L. Smith. Simplex algorithm for countable-state discounted Markov decision processes. *Oper. Res.*, 65(4):1029–1042, 2017.

[28] S. P. Meyn and R. L. Tweedie. Stability of Markovian processes III: Foster-Lyapunov criteria for continuous-time processes. *Adv. Appl. Probab.*, 25(3):518–548, 1993.

[29] S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Cambridge University Press, 2nd edition, 2009.
A Proof of Theorem 3.2

We break down the proof into the following steps:

(a) We use a routine diagonal argument to show that the set \( \{\rho^r\}_{r \in \mathbb{Z}^+} \) is pointwise sequentially compact.

(b) We show that every pointwise accumulation point \( \rho \) of \( \{\rho^r\}_{r \in \mathbb{Z}^+} \) satisfies \( \rho \geq 0 \) and \( \rho(w) \leq c \).

(c) We use (b) to strengthen (a) to “\( \{\rho^r\}_{r \in \mathbb{Z}^+} \) is weak* sequentially compact”.

(d) We argue that the accumulation points \( \rho \) of \( \{\rho^r\}_{r \in \mathbb{Z}^+} \) belong to \( \mathcal{L} \).

(e) We use (d) to establish the desired convergence of \( \{l^r_f\}_{r \in \mathbb{Z}^+} \) and \( \{u^r_f\}_{r \in \mathbb{Z}^+} \).

Let’s begin:

(a) Pick any subsequence \( \{\rho^{r_k}\}_{k \in \mathbb{Z}^+} \) of \( \{\rho^r\}_{r \in \mathbb{Z}^+} \). Enumerate the elements of \( \mathcal{X} \) as \( x_1, x_2, \ldots \). Assumption 2.1(ii) and the constraints \( \rho(\mathcal{X}^c) = 0 \), \( \rho(g) \leq 1 \), and \( \rho \geq 0 \) in (3.7) imply that the sequence \( \{\rho^{r_k}(x_i)\}_{k \in \mathbb{Z}^+} \) is contained in a bounded interval. For this reason, the Bolzano-Weierstrass Theorem tells us that \( \{\rho^{r_k}(x_1)\}_{k \in \mathbb{Z}^+} \) has a converging subsequence \( \{\rho^{r_{k_{j_1}}}(x_1)\}_{j_1 \in \mathbb{Z}^+} \). Repeating the same argument for \( x_2 \) and \( \{\rho^{r_{k_{j_1}}}(x_2)\}_{j_1 \in \mathbb{Z}^+} \), we can find a convergent subsequence \( \{\rho^{r_{k_{j_2}}}(x_2)\}_{j_2 \in \mathbb{Z}^+} \) of \( \{\rho^{r_{k_{j_1}}}(x_2)\}_{j_1 \in \mathbb{Z}^+} \), and so on. Define the subsequence

\[
\rho^{r_{k_i}} := \rho^{r_{k_{j_i}}}, \quad \forall i = 1, 2, \ldots
\]

We have that \( \{\rho^{r_{k_i}}\}_{i \in \mathbb{Z}^+} \) converges pointwise (that is, for each \( x \in \mathcal{X} \), \( \rho^{r_{k_i}}(x) \) converges). Because the subsequence \( \{\rho^{r_k}\}_{k \in \mathbb{Z}^+} \) was arbitrary, we have that \( \{\rho^r\}_{r \in \mathbb{Z}^+} \) is pointwise sequentially compact.

(b) Pick any pointwise converging subsequence \( \{\rho^{r_k}\}_{k \in \mathbb{Z}^+} \) of \( \{\rho^r\}_{r \in \mathbb{Z}^+} \) with limit \( \rho \). Since \( \rho^{r_k} \geq 0 \) for all \( k \in \mathbb{Z}^+ \), the pointwise convergence implies that \( \rho \geq 0 \). Since \( w \) is norm-like, \( \mathcal{X}_n = \{x \in \mathbb{R}^n \mid \rho^r(x) \leq c \} \) is a bounded set, and hence \( \mathcal{X}_n \) is sequentially compact. Therefore, there exists a subsequence \( \{\rho^{r_{k_{j_n}}}\}_{n \in \mathbb{N}} \) of \( \{\rho^{r_{k_i}}\}_{i \in \mathbb{Z}^+} \) that converges to \( \rho \) in the norm topology.
If \( S \) is a finite set for each \( n \in \mathbb{Z}_+ \). The pointwise convergence then tell us that for any fixed \( n \in \mathbb{Z}_+ \),

\[
\sum_{x \in \mathcal{X}_n} \rho(x) w(x) = \lim_{k \to \infty} \sum_{x \in \mathcal{X}_n^k} \rho^k(x) w(x) \leq \lim_{k \to \infty} \rho^k(w) \leq c.
\]

The Monotone Convergence Theorem and the above imply that \( \rho(w) \leq c \).

(c) Pick any pointwise converging subsequence \( \{\rho^k\}_{k \in \mathbb{Z}_+} \) of \( \{\rho^r\}_{r \in \mathbb{Z}_+} \) with limit \( \rho \) and any \( f \in \mathcal{W} \). Fix any \( n \in \mathbb{Z}_+ \), and note that

\[
|\rho^k(f) - \rho(f)| \leq \sum_{x \in \mathcal{X}_n} |\rho^k(x) - \rho(x)| |f(x)| + \sum_{x \notin \mathcal{X}_n} (\rho^k(x) + \rho(x)) |f(x)|.
\]

Markov’s inequality \( (3.9) \) and the moment bounds \( \rho(w) \leq c \) and \( \rho^k(w) \leq c \) tell us that

\[
\sum_{x \notin \mathcal{X}_n} (\rho^k(x) + \rho(x)) |f(x)| \leq 2c \left( \sup_{x \notin \mathcal{X}_n} \frac{|f(x)|}{w(x)} \right).
\]

Fix \( \epsilon > 0 \). Because \( f \) belongs to \( \mathcal{W} \), we can find an \( m \in \mathbb{Z}_+ \) such that \( \sup_{x \notin \mathcal{X}_n} (|f(x)|/w(x)) \leq \frac{\epsilon}{4c} \) and thus

\[
\sum_{x \notin \mathcal{X}_n} (\rho^k(x) + \rho(x)) |f(x)| \leq \frac{\epsilon}{2}.
\]

Because \( w \) is norm-like, \( S_m \) is a finite set and so the pointwise convergence \( \rho^k \) to \( \rho \) implies that there exist a \( K \) such that

\[
\sum_{x \in \mathcal{X}_m^k} |\rho^k(x) - \rho(x)||f(x)| \leq \frac{\epsilon}{2}, \quad \forall k \geq K.
\]

Combining \( (A.1) \)–\( (A.3) \) gives us that

\[
|\rho^k(f) - \rho(f)| \leq \epsilon, \quad \forall k \geq K.
\]

Since the \( \epsilon \) was arbitrary, we have the desired limit

\[
\lim_{k \to \infty} \rho^k(f) = \rho(f).
\]

Because the above holds for every \( f \in \mathcal{W} \), we have that \( \rho^k \) not only converges pointwise to \( \rho \) but also in the weak* topology. Weak* sequential compactness of the sequence follows.

(d) Pick any converging subsequence \( \{\rho^k\}_{k \in \mathbb{Z}_+} \) of \( \{\rho^r\}_{r \geq 0} \) with limit \( \rho \). Assumption 2.1 and \( (3.2) \) imply that \( g \) and \( y \mapsto h(y, x) \), for any \( x \in \mathcal{X} \), belong to \( \mathcal{W} \). Because \( w \) is norm-like, \( (3.2) \) further implies that \( \{E_r\}_{r \in \mathbb{Z}_+} \) is also an increasing sequence of sets that approaches \( \mathcal{X} \):

\[
E_1 \subseteq E_2 \subseteq \ldots, \quad \lim_{r \to \infty} E_r = \bigcup_{r=1}^{\infty} E_r = \mathcal{X}.
\]

Putting these observations together with the weak* convergence of the subsequence we have that

\[
\rho H(x) = \lim_{k \to \infty} \rho^k H(x) = \phi(x) \quad \forall x \in \mathcal{X}, \quad \rho(g) = \lim_{k \to \infty} \rho^k(g) = 1,
\]

where the last equality follows from the constraint \( 1 - c \alpha_r \leq \rho(g) \leq 1 \) in \( (3.7) \). Because we already argued in \( (b) \) that \( \rho \geq 0 \) and \( \rho(w) \leq c \), the above shows that that \( \rho \in \mathcal{L} \).
that their limit points satisfy the equation (a prerequisite for the limit points belonging to guarantees that the weak$^\ast$ to retain the outer approximation property of our finite-dimensional LPs, we require a sequence variables in our finite-dimensional LPs and making further use of the moment bound (2.4). In order the proof of Theorem 3.2 for details).

Theoretically, it ensures that, for each in any sufficiently large truncation: a fact that is key for the implementation of our schemes. b) each$^\ast$ violating (3.2)). The assumption is important for two reasons: a) Practically, it ensures that, for every$^\ast$ in Assumption 2.1, and the fact that$^\ast$ in (3.14). This guarantees that the weak$^\ast$ convergence of the approximating sequences$^\ast$ of Sec. 3 implies that their limit points satisfy the equation (a prerequisite for the limit points belonging to$^\ast$, see the proof of Theorem 3.2 for details).

As we now explain, circumventing this assumption involves doubling the number of decision variables in our finite-dimensional LPs and making further use of the moment bound (2.4). In order to retain the outer approximation property of our finite-dimensional LPs, we require a sequence$^\ast$ of known constants such that

$$\sup_{x \in \mathcal{X}_r} \frac{\sum_{z \in \mathcal{X}_r} h(x, z)}{w(x)} \leq b_r \quad \forall r \in \mathbb{Z}_+, \quad \lim_{r \to \infty} b_r = 0. \quad (B.1)$$

The above implies that$^\ast$ belongs to$^\ast$ for each$^\ast$ and we recover b). To see why, fix any$^\ast$ and$^\ast$ and pick a sufficiently large$^\ast$ such that$^\ast$ belongs$^\ast$ and that$^\ast$ is no greater than$^\ast$. Inequality (B.1), the Metzler property of$^\ast$ in Assumption 2.1, and the fact that the truncations are increasing with$^\ast$ imply that

$$\sup_{x \in \mathcal{X}_r} \frac{h(x, z)}{w(x)} \leq \sup_{x \not\in \mathcal{X}_R} \frac{h(x, z)}{w(x)} \leq \sup_{x \not\in \mathcal{X}_R} \frac{\sum_{z \in \mathcal{X}_r} h(x, z)}{w(x)} \leq b_R \leq \varepsilon \quad \forall r \geq R.$$
Now, to recover a finite dimensional outer approximation of $\mathcal{L}$, pick any index $x$ in our truncation $\mathcal{X}_r$ and consider its associated equation:

$$\rho H(x) = \sum_{z \in \mathcal{X}_r} \rho(z) h(z, x) + \epsilon^r(x) = \phi(x), \quad (B.2)$$

where the measure $\epsilon^r$ is defined by

$$\epsilon^r(x) := \sum_{z \notin \mathcal{X}_r} \rho(z) h(z, x), \quad \forall x \in \mathcal{X}.$$ 

Because $H$ is Metzler and any feasible point $\rho$ of $\mathcal{L}$ is non-negative, we have that

$$\epsilon^r(x) \geq 0, \quad \forall x \in \mathcal{X}_r. \quad (B.3)$$

Tonelli’s Theorem, Markov’s inequality (3.5), and (B.1) imply that

$$\epsilon^r(\mathcal{X}_r) = \sum_{z \in \mathcal{X}_r} \epsilon^r(z) = \sum_{x \notin \mathcal{X}_r} \rho(x) \left( \sum_{z \in \mathcal{X}_r} h(x, z) \right) \leq b_r \left( \sum_{x \notin \mathcal{X}_r} \rho(x) w(x) \right)$$

$$\leq b_r \rho(w) \leq cb_r, \quad \forall \rho \in \mathcal{L}. \quad (B.4)$$

Putting (3.4), (3.6), and (B.2)–(B.4) together, we recover a finite dimensional outer approximation of $\mathcal{L}$: if $\rho$ belongs to $\mathcal{L}$ and $\epsilon|_r$ is the restriction of $\epsilon$ to $\mathcal{X}_r$ (defined analogously to $\rho|_r$ in (3.1)), then the pair $(\rho|_r, \epsilon^r|_r)$ belongs to

$$\tilde{\mathcal{L}}_r := \left\{ (\rho, \epsilon) \in \mathbb{R}^{|\mathcal{X}|} \times \mathbb{R}^{|\mathcal{X}|} : \begin{align*}
\rho H(x) + \epsilon(x) &= \phi(x), \quad \forall x \in \mathcal{X}_r, \\
1 - ca_r &\leq \rho(g) \leq 1, \\
\epsilon(\mathcal{X}_r) &\leq cb_r, \quad \rho(w) \leq c, \\
\rho \geq 0, \quad \epsilon \geq 0, \\
\rho(\mathcal{X}_r^c) + \epsilon(\mathcal{X}_r^c) &= 0. \end{align*} \right\}. $$

For this reason, replacing (3.2) with (B.1), $\mathcal{L}_r$ with $\tilde{\mathcal{L}}_r$, and $l_f^r$ and $u_f^r$ in (3.8) with

$$\tilde{l}_f^r := \inf \{ \rho(f) : (\rho, \epsilon) \in \tilde{\mathcal{L}}_r \}, \quad \tilde{u}_f^r := \inf \{ \rho(f) : (\rho, \epsilon) \in \tilde{\mathcal{L}}_r \},$$

the results of Sections 3–4 hold identically. The only difference is that in contrast with $\mathcal{L}_r$ in (3.7), $\tilde{\mathcal{L}}_r$ involves $|\mathcal{X}_r|$ equalities, $2|\mathcal{X}_r| + 4$ inequalities, and $2|\mathcal{X}_r|$ variables.