APPROXIMATIONS OF COUNTABLY-INFINITE LINEAR PROGRAMS OVER BOUNDED MEASURE SPACES[*]

JUAN KUNTZ[†], PHILIPP THOMAS[‡], GUY-BART STAN[§], AND MAURICIO BARAHONA[¶]

Abstract. We study a class of countably-infinite-dimensional linear programs (CILPs) whose feasible sets are bounded subsets of appropriately defined spaces of measures. The optimal value, optimal points, and minimal points of these CILPs can be approximated by solving finite-dimensional linear programs. We show how to construct finite-dimensional programs that lead to approximations with easy-to-evaluate error bounds, and we prove that the errors converge to zero as the size of the finite-dimensional programs approaches that of the original problem. We discuss the use of our methods in the computation of the stationary distributions, occupation measures, and exit distributions of Markov chains.

Key words. Countably-infinite linear programs, outer approximations, moment bounds, Markov ■ chains, stationary distributions, exit distributions, occupation measures

AMS subject classifications. 90C05, 90C90, 90C06, 60J10, 60J22, 60J27, 65C40, 65G20

1. Introduction. A countably infinite linear program (CILP) is a linear program (LP) with countably many decision variables and constraints. CILPs arise in a variety of applications, e.g. network flow problems [42, 46, 43], production planning [47, 25], equipment replacement and capacity expansion [26, 25], semi-infinite linear programs [4, 21], search problems in robotics [7], robust optimisation [11], and, prominently, in optimal control problems tied to Markov chains [41, 23, 3, 10, 17].

Here, we consider a class of CILPs stemming from the analysis of Markov chains with countably infinite state spaces. Certain measures ρ associated with Markov chains (e.g. the stationary distributions and occupation measures) are feasible points of CILPs, and the optimal values of these CILPs bound integrals of the measures of the form $\rho(f) := \int f \, d\rho$, where f denotes a ρ-integrable function. Typically, the CILPs cannot be solved directly, yet their feasible sets often lie inside a bounded space of measures (see Remark 2.4), a fact that can be made explicit by finding moment bounds satisfied by the measures (i.e. bounds on $\rho(w)$ for norm-like functions w, as in (2.5)–(2.6)). Such bounds can be obtained using either Foster-Lyapunov criteria [34, 35, 15, 27, 30] or mathematical programming [24, 45, 19, 27, 29, 44, 8, 9, 14].

In this paper, we derive finite-dimensional LPs that approximate a given CILP by using a moment bound to truncate the (infinite) set of decision variables and replace its (infinite) set of constraints with finitely many. Based on the ensuing finite-dimensional LPs, we introduce two approximation schemes:

- Scheme A yields lower and upper bounds on the optimal value of the CILP

[†]Department of Mathematics and Department of Bioengineering, Imperial College London, London SW7 2AZ, United Kingdom. *Current address:* Department of Statistics, University of Warwick, Coventry, CV4 7AL, United Kingdom (juan.kuntz-nussio@warwick.ac.uk).

[‡]Department of Mathematics, Imperial College London, London SW7 2AZ, United Kingdom (p.thomas@imperial.ac.uk).

[§]Department of Bioengineering, Imperial College London, London SW7 2AZ, United Kingdom (g.stan@imperial.ac.uk).

[¶]Department of Mathematics, Imperial College London, London SW7 2AZ, United Kingdom (m.barahona@imperial.ac.uk).

and approximations of its optimal points;
- Scheme B yields approximations of the minimal point (i.e. a feasible point that is element-wise no greater than every other feasible point), if it exists, with an easily computable bound on the error.

Both schemes are shown to converge as the size of the approximating LP approaches that of the original CILP (i.e. as r in (3.2) approaches infinity). The moment bound also allows us to quantify the error of the approximations.

The idea of approximating infinite-dimensional LPs with finite-dimensional ones is not new: it traces back to [48], if not earlier, and has been extensively explored (see, e.g. [2, 3, 12, 18, 16] for CILPs, and also [41, 40, 6, 39] for similar approximation techniques involving dynamic programming). The main novelty of our work lies in the practical error bounds that accompany our schemes. In this paper, we have three aims: (i) to present the schemes and their error bounds in a general setting, clarifying the ingredients necessary for their design (see Section 7); (ii) to provide the proofs and technical details pertaining to their convergence (see, in particular, Theorems 2.3, 3.5, and 4.1 and Corollaries 3.6 and 4.3); and (iii) to discuss the use of these methods in the exit problem for Markov chains (Sections 2.2 and 6). The technical proofs and theorems presented in this paper provide the theoretical foundation for the applications presented in [29], where we demonstrated without proofs the practical use of the schemes in a particular setting of interest (i.e. the approximation of stationary distributions of stochastic reaction networks).

We remark that our schemes are not reliant on the Lasserre hierarchy [32, 31]. Specifically: the schemes do not use moment and localising matrices to constrain their feasible sets; the problem data need not be polynomial (or more generally, have any algebraic structure); and the convergence proofs do not require *positivstellensätze*. If the problem data is polynomial, then our schemes can use moment bounds obtained with the Lasserre hierarchy to: (i) tackle problems beyond the reach of the Lasserre hierarchy (e.g. to produce bounds for non-polynomial test functions); (ii) recover converging bounds in cases where the hierarchy has no convergence guarantees and its progress stalls (see [29, Section V.B] for an example). In this sense, the schemes proposed here enlarge the already impressive list of applications of the Lasserre hierarchy.

The paper is organised as follows. In Section 2, we introduce the class of CILPs we consider, and we motivate our work with two problems taken from the Markov chain literature. In Section 3, we present approximations of the CILPs' optimal values and points, prove their convergence, and obtain practically useful error bounds for the optimal values. In Section 4, we derive approximations of the CILPs' minimal points and show how to bound their error. In Section 5, we present two numerical procedures (Schemes A and B) to compute these approximations. Afterwards, in Section 6, we describe their application to the motivating problems of Section 2. We conclude the paper with a discussion in Section 7. The paper has three appendices: Appendix A contains the more technical proofs; Appendix B clarifies the relationship between the various notions of convergence we use; Appendix C explains how to dispense of a technical assumption in our results, which is only assumed to simplify the exposition.

2. The linear program and two motivating problems.

2.1. Statement of the problem.
Let $\mathcal{X}, \mathcal{Y}$ denote two countable (possibly infinite) sets. In various settings, we are interested in linear programs where the solutions $\rho = (\rho(x))_{x \in \mathcal{X}}$ lie in ℓ^1, the space of absolutely summable sequences indexed

by $\mathcal{X}$,

$$\ell^1 := \left\{ (\rho(x))_{x \in \mathcal{X}} : \sum_{x \in \mathcal{X}} |\rho(x)| < \infty \right\},$$

and satisfy the following properties:

    (i) ρ is non-negative:

$$(2.1) \qquad\qquad\qquad\qquad \rho(x) \geq 0 \quad \forall x \in \mathcal{X};$$

    (ii) ρ satisfies a system of linear equations

$$(2.2) \qquad\qquad \rho H(x) := \sum_{x' \in \mathcal{X}} \rho(x')h(x', x) = \phi(x) \quad \forall x \in \mathcal{X},$$

        where $H := (h(x', x))_{x, x' \in \mathcal{X}}$ is a given Metzler matrix (c.f. Assumption 2.2(i)) and $\phi := (\phi(x))_{x \in \mathcal{X}}$ a given measure in ℓ^1;

    (iii) the image of ρ under a given non-negative matrix $G := (g(x, y))_{x \in \mathcal{X}, y \in \mathcal{Y}}$,

$$(2.3) \qquad\qquad \psi(y) := \rho G(y) = \sum_{x \in \mathcal{X}} \rho(x)g(x, y) \quad \forall y \in \mathcal{Y},$$

        is a probability distribution; that is, ψ is non-negative and has unit mass:

$$(2.4) \qquad\qquad \rho(g) := \sum_{x \in \mathcal{X}} g(x)\rho(x) = \sum_{y \in \mathcal{Y}} \psi(y) = 1,$$

$$\text{where} \qquad g(x) := \sum_{y \in \mathcal{Y}} g(x, y) \quad \forall x \in \mathcal{X}.$$

        Here, $\rho(g)$ denotes the ρ-integral of g.

If $\mathcal{X}$ and $\mathcal{Y}$ are infinite (or very large), linear programs constrained by (2.1)–(2.4) cannot be solved directly.

    In many important cases, it is possible to find a real-valued function w that fulfills two properties:

    (i) w is *norm-like*:
        • it is non-negative,

$$(2.5) \qquad\qquad\qquad\qquad w(x) \geq 0 \quad \forall x \in \mathcal{X},$$

        • and it has finite sublevel sets,

$$(2.6) \qquad\qquad \mathcal{X}_r := \{x \in \mathcal{X} : w(x) < r\} \quad \forall r \in \mathbb{Z}_+;$$

    (ii) the ρ-integral of w fulfills a so-called *moment bound*:

$$(2.7) \qquad\qquad\qquad \rho(w) = \sum_{x \in \mathcal{X}} w(x)\rho(x) \leq c,$$

        where c is a constant. Bounds of this type can be obtained through various analytical and numerical methods. For instance, the moment bounds for the applications in this paper can be computed using either Foster-Lyapunov criteria [15, 35, 34, 27, 30] or mathematical programming approaches, i.e. by solving linear or semidefinite programs [24, 19, 27, 45, 29, 44, 8, 9, 14].

When such a function w and constant c are found, the solutions ρ belong to a set constrained by linear equalities and inequalities:

$$(2.8) \qquad \mathcal{L} := \left\{ \rho \in \ell^1 : \begin{array}{l} \rho H(x) = \phi(x) \quad \forall x \in \mathcal{X}, \\ \rho(g) = 1, \\ \rho(w) \leq c, \\ \rho(x) \geq 0 \quad \forall x \in \mathcal{X}. \end{array} \right\}.$$

Our aim is to solve linear programs whose feasible set is $\mathcal{L}$. Specifically, our problems of interest can be stated as follows.

PROBLEM 2.1 (CILPs over measures satisfying moment bounds).   *We wish to optimise the ρ-integral of a given real-valued function $f : \mathcal{X} \to \mathbb{R}$ over the set of measures $\mathcal{L}$, i.e. to compute the optimal values l_f, u_f in*

$$(2.9) \qquad l_f := \inf\{\rho(f) : \rho \in \mathcal{L}\},$$

$$(2.10) \qquad u_f := \sup\{\rho(f) : \rho \in \mathcal{L}\}.$$

Should they exist, we are also interested in finding optimal points $\rho_*, \rho^* \in \mathcal{L}$ achieving the optimal values:

$$\rho_*(f) = l_f \quad and \quad \rho^*(f) = u_f,$$

as well as the minimal point $\rho_m \in \mathcal{L}$ (should it exist): the unique feasible point that is element-wise no greater than all feasible points,

$$(2.11) \qquad \rho_m(x) \leq \rho(x) \quad \forall x \in \mathcal{X}, \ \forall \rho \in \mathcal{L}.$$

Here we study approximations of the CILPs (2.9)–(2.10) under the following general assumption.

ASSUMPTION 2.2.   *Consider a CILP over the set $\mathcal{L}$ (2.8) defined by the real matrices $H := (h(x', x))_{x',x \in \mathcal{X}}$ and $G := (g(x, y))_{x \in \mathcal{X}, y \in \mathcal{Y}}$, and the measure $\phi \in \ell^1$. We assume that we have available a norm-like function $w : \mathcal{X} \to [0, \infty)$ such that:*

(i) H is Metzler: $h(x', x) \geq 0, \ \forall x' \neq x \in \mathcal{X}$.
(ii) G is nonnegative: $g(x, y) \geq 0, \ \forall x \in \mathcal{X}, \ \forall y \in \mathcal{Y}$.
(iii) There exist known positive constants $a_1, a_2, \ldots$ such that

$$(2.12) \qquad \sup_{x \notin \mathcal{X}_r} \frac{g(x)}{w(x)} \leq a_r \quad \forall r \in \mathbb{Z}_+, \qquad \lim_{r \to \infty} a_r = 0,$$

where $g(x)$ is x-row sum (2.4) of G and $\mathcal{X}_r$ is the r-sublevel set (2.7) of w.
(iv) For each $x \in \mathcal{X}$, at least one of the following three conditions is fulfilled: either $w(x) > 0$; or $g(x) > 0$; or there exists $x_1, \ldots, x_l \in \mathcal{X}$ such that $h(x, x_1)h(x_1, x_2) \ldots h(x_{l-1}, x_l) > 0$ and $g(x_l) > 0$.
(v) The support of each of the columns of H is finite:

$$\mathrm{supp}\,(h(\cdot, x)) := \{x' \in \mathcal{X} : h(x', x) \neq 0\} \ \text{is finite,} \quad \forall x \in \mathcal{X}.$$

(vi) $\mathcal{L}$ is non-empty.

Assumptions (i, ii) ensure that $\rho H(x)$ and $\rho G(y)$ are absolutely convergent or $+\infty$ for any non-negative $\rho \in \ell^1$. Assumption (iii) allows us to control the feasible points outside of the sublevel sets of w and is key in deriving our finite-dimensional approximating LPs. Assumption (iv) guarantees that the entries $\rho(x)$ are bounded—a

requirement for the proofs of Sections 3–4. The particular statement of Assumption (iv) is motivated by the applications in Section 2.2 for which it is natural, (e.g. for the exit problem it asks that the chain may leave the domain from every state x for which $w(x) = 0$). Assumption (v) is trivially satisfied in many applications [46, 12, 13] (for Markov chains, it requires that each state is reachable in a single jump from at most a finite number of states), and simplifies the derivation of some of our results. It can, however, be circumvented by increasing the size of the approximating LPs (see Appendix C). Lastly, Assumption (vi) ensures that the CILPs (2.9)–(2.10) are non-trivial.

For a broad class of functions f, the CILPs (2.9)–(2.10) are guaranteed to be well-posed and *solvable*, i.e. their optimal values are finite and their optimal points exist.

THEOREM 2.3 (The CILPs are solvable). *Suppose that Assumption 2.2 is satisfied and assume that f belongs to the set $\mathcal{W}$ of real-valued functions on $\mathcal{X}$ that eventually grow strictly slower than the norm-like function w:*

$$(2.13) \qquad \mathcal{W} := \left\{ f : \lim_{r \to \infty} \left(\sup_{x \notin \mathcal{X}_r} \frac{|f(x)|}{w(x)} \right) = 0 \right\}.$$

Then, we have that:
  (i) the sum $\rho(f)$ is absolutely convergent, $\forall \rho \in \mathcal{L}$,
  (ii) the optimal value l_f of (2.9) is finite,
  (iii) there exists at least one optimal point $\rho_* \in \mathcal{L}$ satisfying $\rho_*(f) = l_f$.
Note that by replacing f by $-f$, the above holds identically for optimal value u_f (2.10) and corresponding optimal points ρ^*.

  Proof. See Appendix A.

As explained in the following remark, Assumption 2.2 ensures that the feasible sets of CILPs (2.9)–(2.10) are bounded subsets of an appropriate normed space, a fact that underlies the results of this paper.

REMARK 2.4 (The feasible set is contained in a bounded measure space). *If $\mathcal{N}$ denotes the null set $\{x \in \mathcal{X} : w(x) = 0\}$ of the norm-like function w, then*

$$\tilde{w}(x) := \left\{ \begin{array}{cc} 1 & x \in \mathcal{N} \\ w(x) & x \notin \mathcal{N} \end{array} \right. \qquad \forall x \in \mathcal{X}$$

defines the norm

$$||\rho||_{\tilde{w}} := \sum_{x \in \mathcal{X}} |\rho(x)| \, \tilde{w}(x)$$

on the weighted space $\{\rho \in \ell^1 : ||\rho||_{\tilde{w}} < \infty\}$. For any x in the null set $\mathcal{N}$, Assumption 2.2(i, ii, iv) and the constraints $\rho H = \phi$ and $\rho(g) = 1$ in (2.8) imply that $\rho(x) \leq c_x$ uniformly over ρ in $\mathcal{L}$ for some constant $c_x < \infty$. Because w is norm-like, $\mathcal{N}$ is finite and it follows that $c_{\mathcal{N}} := \sum_{x \in \mathcal{N}} c_x < \infty$. Consequently, the constraints $\rho(w) \leq c$ and $\rho \geq 0$ in (2.8) imply that the feasible set $\mathcal{L}$ lies inside the bounded normed space $(\mathbb{B}_{c+c_{\mathcal{N}}}, ||\cdot||_{\tilde{w}})$, where $\mathbb{B}_{c+c_{\mathcal{N}}} := \{\rho \in \ell^1 : ||\rho||_{\tilde{w}} \leq c + c_{\mathcal{N}}\}$. Therefore we say that (2.9)–(2.10) are CILPs over bounded measure spaces.

Notation. Throughout, we denote the set of positive integers $\{1, 2, 3, \dots\}$ by $\mathbb{Z}_+$ and we use the conventions $\sup \emptyset = -\infty$ and $\inf \emptyset = +\infty$. A countable set $\mathcal{X}$ has cardinality $|\mathcal{X}|$ and power set $2^{\mathcal{X}} := \{A : A \subseteq \mathcal{X}\}$. We abuse notation by using ρ to

denote both a measure on $(\mathcal{X}, 2^{\mathcal{X}})$ and its density ($\rho(x) := \rho(\{x\})$ $\forall x \in \mathcal{X}$) so that

$$(2.14) \qquad \rho(A) = \sum_{x \in A} \rho(x) \quad \forall A \subseteq \mathcal{X}.$$

Given the above, we identify the space of finite, signed measures on $(\mathcal{X}, 2^{\mathcal{X}})$ with ℓ^1.

2.2. Two motivating problems from Markov chain theory: stationary distributions and exit problems. Our work is motivated by the analysis of Markov processes on countably infinite state spaces (a.k.a. Markov chains, or *chains* for short). Specifically, we are interested in computing the stationary distributions of a chain and the exit distributions and occupation measures associated with its exit times. We now show briefly how these two problems can be mapped to Problem 2.1 for both discrete-time and continuous-time chains.

The discrete-time case. Let $(X_n)_{n \in \mathbb{N}}$ denote a time-homogeneous discrete-time Markov chain taking values in a countable state space $\mathcal{S}$, with one-step matrix $P := (p(x,y))_{x,y \in \mathcal{S}}$ and initial distribution $\gamma := (\gamma(x))_{x \in \mathcal{S}}$:

$$p(x,y) = \mathbb{P}_\gamma\left(\{X_1 = y\}|\{X_0 = x\}\right), \quad \gamma(x) = \mathbb{P}_\gamma\left(\{X_0 = x\}\right), \quad \forall x, y \in \mathcal{S},$$

where $\mathbb{P}_\gamma$ denotes the underlying probability measure (and the subscript emphasises that the initial state is sampled from γ).

Stationary distributions. A probability measure π on $\mathcal{S}$ is a *stationary distribution* of the chain if sampling its initial state from π ensures that the chain remains distributed according to π for all time:

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

As the following well-known corollary shows, the stationary distributions (satisfying the moment bound) are the feasible points of a CILP of the type in Problem 2.1. The optimal points are those maximising or minimising an average of interest over the set of stationary distributions and the optimal value is said maximum or minimum.

COROLLARY 2.5 ([5, Theorem A.I.3.1]). *The set of stationary distributions π that satisfy the moment bound $\pi(w) \leq c$ is the set $\mathcal{L}$ (2.8) with*

$$(2.15) \qquad \begin{aligned} &\mathcal{X} := \mathcal{S}, \quad \mathcal{Y} := \mathcal{S}, \\ &h(x', x) := p(x', x) - 1_{x'}(x) \quad \forall x', x \in \mathcal{S}, \\ &\phi(x) := 0 \quad \forall x \in \mathcal{S}, \\ &g(x, y) := 1_x(y) \quad \forall x, y \in \mathcal{S}, \end{aligned}$$

where $1_{x'}$ denotes the indicator function of state x'.

Exit distribution and occupation measure. The *exit time* σ of the chain $(X_n)_{n \in \mathbb{N}}$ from a subset $\mathcal{D} \subseteq \mathcal{S}$ (the *domain*) is the first time that the chain lies outside $\mathcal{D}$:

$$\sigma := \inf\{n \in \mathbb{N} : X_n \notin \mathcal{D}\},$$

with $\sigma := \infty$ if the chain never leaves $\mathcal{D}$. Associated with the exit time, there is an *exit distribution* μ and *occupation measure* ν:

$$\mu(x) := \mathbb{P}_\gamma\left(\{X_\sigma = x, \sigma < \infty\}\right) \quad \forall x \notin \mathcal{D},$$

$$\nu(x) := \mathbb{E}_\gamma\left[\sum_{m=0}^{\sigma-1} 1_x(X_m)\right] \qquad \forall x \in \mathcal{D},$$

i.e. $\mu(x)$ denotes the probability that the chain exits $\mathcal{D}$ via state x, and $\nu(x)$ denotes the expected number of visits that the chain makes to x before exiting $\mathcal{D}$. The occupation measure is the minimal point of a CILP of the type in Problem 2.1 and the exit distribution is its image through an appropiately defined G:

COROLLARY 2.6 ([28, Theorem 11.7]). *If $\sigma < \infty$ almost surely (i.e. chain eventually leaves $\mathcal{D}$ with probability one), $\gamma(\mathcal{D}) = 1$ (i.e. the chain starts inside $\mathcal{D}$), and the occupation measure ν satisfies a moment bound $\nu(w) \leq c$, then ν is the minimal point (2.11) of the set $\mathcal{L}$ in (2.8) with*

$$(2.16) \qquad \begin{aligned} &\mathcal{X} := \mathcal{D}, \quad \mathcal{Y} := \mathcal{D}^c \\ &h(x', x) := p(x', x) - 1_{x'}(x) \quad \forall x', x \in \mathcal{D}, \\ &\phi(x) := -\gamma(x) \quad \forall x \in \mathcal{D}, \\ &g(x, y) := p(x, y) \quad \forall x \in \mathcal{D}, \ \forall y \notin \mathcal{D}, \end{aligned}$$

where $\mathcal{D}^c = \mathcal{S} \backslash \mathcal{D}$ denotes the complement of $\mathcal{D}$ in $\mathcal{S}$. Furthermore, the exit distribution is given by $\mu = \nu G$.

The continuous-time case. Let $(X_t)_{t \geq 0}$ denote a minimal time-homogeneous continuous-time Markov chain (e.g. see [28, Sections 26, 37]) taking values in a countable state space $\mathcal{S}$ with rate matrix $Q := (q(x, y))_{x, y \in \mathcal{S}}$ and initial distribution $\gamma := (\gamma(x))_{x \in \mathcal{S}}$:

$$q(x, y) = \lim_{t \to 0} \frac{\mathbb{P}_\gamma \left(\{X_t = y, t < T_\infty\} | \{X_0 = x\} \right)}{t}, \quad \gamma(x) = \mathbb{P}_\gamma \left(\{X_0 = x\} \right), \ \forall x, y \in \mathcal{S},$$

where T_∞ denotes the explosion time of the chain and $\mathbb{P}_\gamma$ is the underlying probability measure. We assume that Q is stable and conservative, i.e.

$$(2.17) \qquad -q(x, x) = \sum_{\substack{y \in \mathcal{S} \\ y \neq x}} q(x, y) < \infty \qquad \forall x \in \mathcal{S}.$$

Stationary distribution. A probability measure π on $\mathcal{S}$ is a stationary distribution of the continuous-time chain if

$$\mathbb{P}_\pi \left(\{X_t = x, t < T_\infty\} \right) = \pi(x) \quad \forall x \in \mathcal{S}, \ t \geq 0.$$

As for the discrete-time case, the stationary distributions satisfying the moment bound are the feasible points of CILPs of the type in Problem 2.1:

COROLLARY 2.7 ([36, Theorem 1]). *If Q is regular (i.e. $\mathbb{P}_\gamma \left(\{T_\infty = \infty\} \right) = 1$ for all probability distributions γ), the set of stationary distributions that satisfy the moment bound $\pi(w) \leq c$ is the set $\mathcal{L}$ in (2.8) with*

$$(2.18) \qquad \begin{aligned} &\mathcal{X} := \mathcal{S}, \quad \mathcal{Y} := \mathcal{S}, \\ &h(x', x) := q(x', x) \quad \forall x', x \in \mathcal{S}, \\ &\phi(x) := 0 \quad \forall x \in \mathcal{S}, \\ &g(x, y) := 1_x(y) \quad \forall x, y \in \mathcal{S}. \end{aligned}$$

Exit distribution and occupation measure. The exit time τ of $(X_t)_{t \geq 0}$ from a domain $\mathcal{D}$ is the first time that the chain lies outside of $\mathcal{D}$:

$$\tau = \inf\{0 \leq t < T_\infty : X_t \notin \mathcal{D}\},$$

with $\tau := \infty$ if the chain never leaves $\mathcal{D}$. Associated with τ, there is an *exit distribution* μ and *occupation measure* ν:

$$\mu(x) := \mathbb{P}_\gamma (\{X_\tau = x, \tau < \infty\}) \quad \forall x \notin \mathcal{D},$$

$$\nu(x) := \mathbb{E}_\gamma \left[\int_0^{\min\{\tau, T_\infty\}} 1_x(X_t) dt \right] \quad \forall x \in \mathcal{D},$$

which characterise where the chain exits the domain (μ), and where inside the domain the chain spends its time until it exits (ν). Similarly as in the discrete-time case, the occupation measure is the minimal point of a CILP of the type in Problem 2.1 and the exit distribution is its image through an appropriately defined G:

COROLLARY 2.8 ([28, Theorem 36.11]).   *If $\tau < \infty$ almost surely, $\gamma(\mathcal{D}) = 1$, and $\nu(w) \leq c$, the occupation measure is the minimal point (2.11) of the set $\mathcal{L}$ in (2.8) with*

(2.19)
$$\begin{aligned} &\mathcal{X} := \mathcal{D}, \quad \mathcal{Y} := \mathcal{D}^c, \\ &h(x', x) := q(x', x) \quad \forall x', x \in \mathcal{D}, \\ &\phi(x) := -\gamma(x) \quad \forall x \in \mathcal{D}, \\ &g(x, y) := q(x, y) \quad \forall x \in \mathcal{D}, \ \forall y \notin \mathcal{D}, \end{aligned}$$

Furthermore, the exit distribution is given by $\mu = \nu G$.

3. Bounding the optimal values and approximating the optimal points of the CILP. Consider Problem 2.1 under Assumption 2.2. To derive finite-dimensional approximations of CILPs (2.9)–(2.10), we start by truncating the set $\mathcal{X}$ using the sub-level sets $\mathcal{X}_r$ in (2.6) of the norm-like function w. We then define the restriction $\rho_{|r}$ to $\mathcal{X}_r$ of each feasible point ρ in $\mathcal{L}$,

(3.1)
$$\rho_{|r}(x) := \begin{cases} \rho(x) & \text{if } x \in \mathcal{X}_r \\ 0 & \text{if } x \notin \mathcal{X}_r \end{cases} \quad \forall x \in \mathcal{X},$$

and the following set of measures ρ^r:

(3.2)
$$\mathcal{L}_r := \left\{ \rho^r \in \ell^1 : \begin{array}{l} \rho^r H(x) = \phi(x) \quad \forall x \in \mathcal{E}_r, \\ 1 - ca_r \leq \rho^r(g) \leq 1, \\ \rho^r(w) \leq c, \\ \rho^r(x) \geq 0 \quad \forall x \in \mathcal{X}, \\ \rho^r(\mathcal{X}_r^c) = 0. \end{array} \right\} \quad \forall r \in \mathbb{Z}_+.$$

Here $\mathcal{X}_r^c = \mathcal{X} \backslash \mathcal{X}_r$ denotes the complement of $\mathcal{X}_r$ in $\mathcal{X}$; the positive constant a_r is as in (2.12); c denotes the moment bound constant in (2.7); and

(3.3)
$$\mathcal{E}_r := \{x \in \mathcal{X}_r : h(x', x) = 0 \ \forall x' \notin \mathcal{X}_r\}.$$

The set $\mathcal{L}_r$ is an *outer approximation* of $\mathcal{L}$ in the sense that the restriction of every feasible point in $\mathcal{L}$ belongs to $\mathcal{L}_r$.

LEMMA 3.1 (The outer approximation property of $\mathcal{L}_r$).   *Suppose that Assumption 2.2 is satisfied. For any r in $\mathbb{Z}_+$, $\mathcal{L}_r$ is an outer approximation of $\mathcal{L}$:*

$$\rho_{|r} \in \mathcal{L}_r \quad \forall \rho \in \mathcal{L}.$$

Proof. Let ρ be any feasible point in $\mathcal{L}$ and $\rho_{|r}$ be its restriction (3.1) to $\mathcal{X}_r$. If x belongs to $\mathcal{E}_r$, $\rho H(x) = \phi(x)$ only involves entries $\rho(x')$ indexed by x' in $\mathcal{X}_r$, and so

$$(3.4) \qquad \rho_{|r} H(x) = \rho H(x) = \phi(x) \quad \forall x \in \mathcal{E}_r.$$

Because w is a non-negative function, it follows from the definition (2.8) that

$$(3.5) \qquad \rho_{|r}(w) \le \rho(w) \le c, \quad \rho_{|r}(x) \ge 0 \quad \forall x \in \mathcal{X}.$$

Next, consider the following generalisation of Markov's inequality: for any nonnegative function f on $\mathcal{X}$,

$$(3.6) \qquad \sum_{x \notin \mathcal{X}_r} f(x)\rho(x) \le \left(\sup_{x \notin \mathcal{X}_r} \frac{f(x)}{w(x)} \right) \sum_{x \notin \mathcal{X}_r} w(x)\rho(x) \le \left(\sup_{x \notin \mathcal{X}_r} \frac{f(x)}{w(x)} \right) \rho(w)$$
$$\le c \left(\sup_{x \notin \mathcal{X}_r} \frac{f(x)}{w(x)} \right).$$

Setting f to be the row-sum function g in (2.4), we find that

$$(3.7) \qquad 1 - ca_r \le 1 - c \left(\sup_{x \notin \mathcal{X}_r} \frac{g(x)}{w(x)} \right) \le 1 - \sum_{x \notin \mathcal{X}_r} g(x)\rho(x) = \rho_{|r}(g) \le \rho(g) = 1,$$

where the first inequality follows from Assumption 2.2(*iii*), the second from (3.6), and the remainder from $\rho(g) = 1$ in (2.8). Combining (3.4)–(3.7), we have that $\rho_{|r}$ belongs to $\mathcal{L}_r$. □

The outer approximation property of $\mathcal{L}_r$ has an important consequence: the optimal values of the CILPs (2.9)–(2.10) can be bounded by optimising over $\mathcal{L}_r$, i.e. by solving the finite-dimensional LPs

$$(3.8) \qquad l_f^r := \inf\{\rho^r(f) : \rho^r \in \mathcal{L}_r\},$$
$$(3.9) \qquad u_f^r := \sup\{\rho^r(f) : \rho^r \in \mathcal{L}_r\}.$$

Note that these LPs are solvable because they entail optimising a linear function over a compact subset of $\mathbb{R}^{|\mathcal{X}_r|}$. Lemma 3.1 then yields the following bounds:

COROLLARY 3.2 (Bounding the optimal values). *Suppose that Assumption 2.2 is satisfied and let f be any function in $\mathcal{W}$. If l_f and u_f are the optimal values of the CILPs (2.9)–(2.10) and l_f^r and u_f^r are those of (3.8)–(3.9), then we have that*
(3.10)

$$l_f^r - c \left(\sup_{x \notin \mathcal{X}_r} \frac{|f(x)|}{w(x)} \right) \le l_f \le \rho(f) \le u_f \le u_f^r + c \left(\sup_{x \notin \mathcal{X}_r} \frac{|f(x)|}{w(x)} \right), \quad \forall \rho \in \mathcal{L}, \ r \in \mathbb{Z}_+.$$

Under the following additional assumptions, the bounds (3.10) *can be sharpened:*

$$(3.11) \qquad \text{If } f(x) \ge 0, \forall x \notin \mathcal{X}_r, \text{ then } l_f^r \le l_f \le \rho(f) \quad \forall \rho \in \mathcal{L}, \ r \in \mathbb{Z}_+.$$
$$(3.12) \qquad \text{If } f(x) \le 0, \forall x \notin \mathcal{X}_r, \text{ then } \rho(f) \le u_f \le u_f^r \quad \forall \rho \in \mathcal{L}, \ r \in \mathbb{Z}_+.$$

Proof. From Lemma 3.1 and the definitions (3.8)–(3.9), we have

$$(3.13) \qquad l_f^r \le \rho_{|r}(f) \le u_f^r \Rightarrow l_f^r + \sum_{x \notin \mathcal{X}_r} f(x)\rho(x) \le \rho(f) \le u_f^r + \sum_{x \notin \mathcal{X}_r} f(x)\rho(x),$$

for every feasible point ρ in $\mathcal{L}$. The inequalities (3.11)–(3.12) follow immediately because all ρ in $\mathcal{L}$ are non-negative. For (3.10), replace f with $|f|$ in (3.6) to obtain

$$(3.14) \qquad \left| \sum_{x \notin \mathcal{X}_r} f(x)\rho(x) \right| \leq \sum_{x \notin \mathcal{X}_r} |f(x)|\, \rho(x) \leq c \left(\sup_{x \notin \mathcal{X}_r} \frac{|f(x)|}{w(x)} \right). \qquad \square$$

Our definition (2.6) implies that the truncations $\mathcal{X}_r$ form an increasing sequence that approaches the entire index set $\mathcal{X}$ as r tends to ∞:

$$\mathcal{X}_1 \subseteq \mathcal{X}_2 \subseteq \ldots \subseteq \mathcal{X}_r \subseteq \mathcal{X}_{r+1} \subseteq \ldots \quad \text{and} \quad \bigcup_{r=1}^{\infty} \mathcal{X}_r = \mathcal{X}.$$

In the following, we show that the associated outer approximations $\mathcal{L}_r$ converge to $\mathcal{L}$, and that the sequences of lower bounds $(l_f^r)_{r \in \mathbb{Z}_+}$ and upper bounds $(u_f^r)_{r \in \mathbb{Z}_+}$ obtained by solving (3.8)–(3.9) for increasing truncations converge to the optimal values l_f and u_f, respectively. The notions of convergence we consider here are as follows:

DEFINITION 3.3 (Convergence in weak*). *With $\mathcal{W}$ as in (2.13) for any given non-negative w, we say that a sequence $(\rho^r)_{r \in \mathbb{Z}_+} \subseteq \ell^1$ converges to a point $\rho \in \ell^1$ in weak* as $r \to \infty$ if and only if*

$$(3.15) \qquad \lim_{r \to \infty} \rho^r(f) = \rho(f) \quad \forall f \in \mathcal{W}.$$

REMARK 3.4 (Convergence in weak* implies convergence in total variation). *If w is norm-like, weak* convergence of a sequence $(\rho^r)_{r \in \mathbb{Z}_+} \subseteq \ell^1$ to $\rho \in \ell^1$ implies convergence in total variation:*

$$\lim_{r \to \infty} ||\rho - \rho^r|| = 0,$$

where $||\cdot||$ denotes the total variation norm:

$$(3.16) \qquad ||\rho - \rho^r|| := \sup_{A \subseteq \mathcal{X}} |\rho(A) - \rho^r(A)|.$$

See Appendix B for details.

The following theorem formalises the manner in which the outer approximations $\mathcal{L}_r$ converge to the set $\mathcal{L}$ as r tends to infinity.

THEOREM 3.5 (The outer approximations $\mathcal{L}_r$ converge to $\mathcal{L}$). *Suppose that Assumption 2.2 is satisfied. The weak* accumulation points of any given sequence $(\rho^r \in \mathcal{L}_r)_{r \in \mathbb{Z}_+}$ belong to $\mathcal{L}$, and every subsequence of $(\rho^r)_{r \in \mathbb{Z}_+}$ has a weak* convergent subsequence.*

Proof. This proof is similar to that of the weak* sequential compactness of $\mathcal{L}$ in Appendix A.1, hence we only sketch it here. A diagonal argument as in Appendix A.1(a) shows that every subsequence of $(\rho^r)_{r \in \mathbb{Z}_+}$ has a pointwise convergent subsequence. Applying Fatou's lemma as in Appendix A.1(b), we have that the pointwise accumulation points ρ^∞ of $(\rho^r)_{r \in \mathbb{Z}_+}$ are non-negative and satisfy $\rho^\infty(w) \leq c$. Applying the generalisation (3.6) of Markov's inequality as in Appendix A.1(c) shows that $(\rho^r)_{r \in \mathbb{Z}_+}$ has a weak* convergent subsequence. Assumption 2.2(ii, iii, v) implies that g and $x' \mapsto h(x', x)$ belong to $\mathcal{W}$ and

$$\mathcal{E}_1 \subseteq \mathcal{E}_2 \subseteq \ldots, \qquad \lim_{r \to \infty} \mathcal{E}_r = \bigcup_{r=1}^{\infty} \mathcal{E}_r = \mathcal{X},$$

whence it follows that the weak* accumulation points of $(\rho^r)_{r \in \mathbb{Z}_+}$ belong to $\mathcal{L}$. $\square$

Theorem 3.5 has the important consequence that, for sufficiently large r, the optimal values and points of the finite-dimensional LPs (3.8)–(3.9) are close to those of the infinite-dimensional LPs (2.9)–(2.10).

COROLLARY 3.6 (The optimal values and points converge). *Suppose that Assumption 2.2 is satisfied and f belongs to $\mathcal{W}$.*

(i) *(The optimal values) The sequence $(l_f^r)_{r \in \mathbb{Z}_+}$ of optimal values of (3.8) converges to the optimal value l_f of (2.9):*

$$\lim_{r \to \infty} l_f^r = l_f.$$

(ii) *(The optimal points) All weak* accumulation points of any sequence $(\rho_*^r \in \mathcal{L}_r)_{r \in \mathbb{Z}_+}$ of optimal points of (3.8) belong to the set of optimal points of (2.9)*

$$\mathcal{O}_l := \{\rho_* \in \mathcal{L} : \rho_*(f) = l_f\},$$

and every subsequence of $(\rho_*^r)_{r \in \mathbb{Z}_+}$ has a weak* convergent subsequence. In particular, if (2.9) has a unique optimal point ρ_* (i.e. $\mathcal{O}_l = \{\rho_*\}$), then $(\rho_*^r)_{r \in \mathbb{Z}_+}$ converges in weak* to ρ_*.

(iii) *(The unique feasible point case) If (2.9) has a unique feasible point ρ (i.e. $\mathcal{L} = \{\rho\}$), then any sequence $(\rho^r \in \mathcal{L}_r)_{r \in \mathbb{Z}_+}$ of feasible points of (3.8) converges in weak* to ρ.*

Proof. (i, ii) Let ρ^∞ be a weak* accumulation point of $(\rho_*^r)_{r \in \mathbb{Z}_+}$. Given Theorem 3.5, all we need to show is that all $\rho^\infty \in \mathcal{O}_l$. Pick any subsequence $(\rho^{r_k})_{k \in \mathbb{Z}_+}$ converging to ρ^∞ in weak*. Theorem 3.5 implies that ρ^∞ belongs to $\mathcal{L}$, and so $\rho^\infty(f) \geq l_f$. On the other hand, f belongs to $\mathcal{W}$ and (3.10) implies that

$$\rho^\infty(f) = \lim_{k \to \infty} l_f^{r_k} = \lim_{k \to \infty} l_f^{r_k} - \lim_{k \to \infty} c \left(\sup_{x \notin \mathcal{X}_{r_k}} \frac{|f(x)|}{w(x)} \right)$$

$$= \lim_{k \to \infty} \left(l_f^{r_k} - c \left(\sup_{x \notin \mathcal{X}_{r_k}} \frac{|f(x)|}{w(x)} \right) \right) \leq l_f.$$

Hence, $\rho^\infty(f) = l_f$ and $\rho^\infty \in \mathcal{O}_l$ for any given accumulation point ρ^∞ of $(\rho_*^r)_{r \in \mathbb{Z}_+}$.

(iii) follows immediately from (ii) by setting $f := 0$. $\qquad \square$

REMARK 3.7. *Replacing f by $-f$, Corollary 3.6 holds identically for the optimal values (u_f and u_f^r) and optimal points (ρ^* and ρ^{*r}) of (2.10) and (3.9).*

In some applications, including those discussed in Section 6, the images $\psi = \rho G$ (2.3) of the feasible points $\rho \in \mathcal{L}$ are of interest. To approximate these, we use the images $\psi^r = \rho^r G$ of the feasible points of $\mathcal{L}_r$:

PROPOSITION 3.8 (Convergence of the images of the optimal points). *Suppose that Assumption 2.2 is satisfied. If the sequence of points $(\rho^r \in \mathcal{L}_r)_{r \in \mathbb{Z}_+}$ converges in weak* to $\rho \in \mathcal{L}$, then the sequence of images $(\psi^r = \rho^r G)_{r \in \mathbb{Z}_+}$ converges in total variation to $\psi = \rho G$, i.e.*

$$(3.17) \qquad \lim_{r \to \infty} ||\psi^r - \psi|| = 0.$$

where the norm is defined in (3.16)

Proof. From the definition (3.16), we have $\forall r, r' \in \mathbb{Z}_+$ that

$$||\psi^r - \psi|| \leq \sum_{y \in \mathcal{Y}} |\psi^r(y) - \psi(y)| \leq \sum_{x \in \mathcal{X}_{r'}} |\rho^r(x) - \rho(x)|\, g(x) + \sum_{x \notin \mathcal{X}_{r'}} (\rho^r(x) + \rho(x))g(x).$$

Due to Assumption 2.2(*iii*), the rest of the proof is analogous to part (c) in the proof of the weak* sequential compactness of $\mathcal{L}$, see Appendix A.1. □

In the case of a unique feasible point ρ in $\mathcal{L}$, Corollary 3.6(*iii*) and Proposition 3.8 show that any sequence of feasible points $(\rho^r \in \mathcal{L}_r)_{r \in \mathbb{Z}_+}$ of the approximating LPs converges to ρ, and that the images $(\psi^r = \rho^r G)_{r \in \mathbb{Z}_+}$ converge to the image $\psi = \rho G$. Yet ρ^r and ψ^r are uncontrolled approximations of ρ and ψ in the sense that there is no known practical way to compute or bound their errors. To remedy this, we show in the following section how to repeatedly apply the results of this section to obtain collections of element-wise lower bounds on ρ and ψ with easily quantifiable errors.

4. Approximating the minimal point of the CILP. The setup here is identical to that of Section 3, only with the additional assumption that $\mathcal{L}$ has a minimal point (i.e. a feasible point ρ_m satisfying (2.11)). We now focus on approximating ρ_m and its image $\psi_m := \rho_m G$. Given a truncation $\mathcal{X}_r$, let

$$(4.1) \qquad l^r(x) := \begin{cases} l_x^r & \text{if } x \in \mathcal{X}_r \\ 0 & \text{if } x \notin \mathcal{X}_r \end{cases} \qquad \forall x \in \mathcal{X}, \ r \in \mathbb{Z}_+,$$

where l_x^r is as in (3.8) with f being the indicator function 1_x of the state x.

THEOREM 4.1 (Approximation of the minimal point with computable error bounds). *Suppose that Assumption 2.2 is satisfied and that $\mathcal{L}$ has a minimal point ρ_m. Let the approximation $l^r = (l^r(x))_{x \in \mathcal{X}}$ be as in (4.1). Then, the following hold:*

(i) The approximation error has an easy-to-compute bound:

$$(4.2) \qquad ||\rho_m - l^r|| \leq u_{\mathcal{X}_r}^r + \frac{c}{r} - l^r(\mathcal{X}_r) =: \Gamma_r \quad \forall r \in \mathbb{Z}_+,$$

where $u_{\mathcal{X}_r}^r$ is as in (3.9) with f being the indicator function $1_{\mathcal{X}_r}$ of the truncation $\mathcal{X}_r$.

(ii) The approximation l^r converges to ρ_m in weak* as $r \to \infty$.

(iii) In the case of a unique feasible point ρ, the error bound Γ_r converges to zero:

$$\lim_{r \to \infty} \Gamma_r = 0.$$

Proof. (*i*) From Corollary 3.2, it follows that l^r bounds ρ_m from below,

$$(4.3) \qquad l^r(x) \leq \rho_m(x) \quad \forall x \in \mathcal{X}, \ r \in \mathbb{Z}_+.$$

Hence, $\rho_m - l^r$ is an unsigned measure. The total variation norm of an unsigned measure is its mass, hence (4.2) follows immediately by setting $f := 1_{\mathcal{X}_r}$ in (3.10).

(*ii*) Note that Corollary 3.6(*i*) shows that l^r converges pointwise to ρ_m:

$$(4.4) \qquad \lim_{r \to \infty} l^r(x) = \rho_m(x) \quad \forall x \in \mathcal{X}.$$

Using Fatou's Lemma and (4.3)–(4.4) we have

$$(4.5) \qquad l(w) \leq \lim_{r \to \infty} l^r(w) \leq \lim_{r \to \infty} \rho_m(w) \leq \rho_m(w) \leq c.$$

Pick any f in $\mathcal{W}$, fix r, n in $\mathbb{Z}_+$, and note that

$$(4.6) \quad |\rho_m(f) - l^r(f)| \leq \sum_{x \in \mathcal{X}_n} (\rho_m(x) - l^r(x)) |f(x)| + \sum_{x \notin \mathcal{X}_n} (\rho_m(x) + l^r(x)) |f(x)|.$$

Given (4.4)–(4.6), the rest of the proof is analogous to part (c) in the proof of the weak* sequential compactness of $\mathcal{L}$, see Appendix A.1.

(iii) This follows immediately from (i) and Corollary 3.6. $\qquad\square$

REMARK 4.2. *The analogous measure u^r, composed of upper bounds u_x^r obtained from (3.9) with $f := 1_x$, is also an approximation on the feasible points of $\mathcal{L}$, and is also accompanied by an error bound. Although Corollary 3.6 and Remark 3.7 show that u^r converges pointwise, no weak* convergence can be recovered, see [27] for details.*

To approximate the image $\psi_m := \rho_m G$ of the minimal point, we define:

$$(4.7) \qquad l_\psi^r(y) := \begin{cases} l_{g_y}^r & \text{if } y \in \mathcal{Y}_r \\ 0 & \text{if } y \notin \mathcal{Y}_r \end{cases} \qquad \forall y \in \mathcal{Y},$$

where $l_{g_y}^r$ is as in (3.8) with $f(\cdot) = g_y(\cdot) := g(\cdot, y)$, and $\mathcal{Y}_r$ belongs to any increasing sequence of finite subsets that approach $\mathcal{Y}$:

$$\mathcal{Y}_1 \subseteq \mathcal{Y}_2 \subseteq \ldots \subseteq \mathcal{Y}_r \subseteq \mathcal{Y}_{r+1} \subseteq \ldots \quad \text{and} \quad \bigcup_{r=1}^{\infty} \mathcal{Y}_r = \mathcal{Y}.$$

COROLLARY 4.3 (Controlled approximation of the image of the minimal point). *Suppose that Assumption 2.2 is satisfied and $\mathcal{L}$ has a minimal point ρ_m with image $\psi_m = \rho_m G$, and let $l_\psi^r = (l_\psi^r(y))_{y \in \mathcal{Y}}$ be the approximation in (4.7). We have that:*

(i) *The approximation error is given by*

$$\left\| \psi_m - l_\psi^r \right\| = 1 - l_\psi^r(\mathcal{Y}_r).$$

(ii) *The lower bounds l_ψ^r converge in total variation to ψ_m:*

$$\lim_{r \to \infty} \left\| \psi_m - l_\psi^r \right\| = 0.$$

Proof. (i) Since G is non-negative, Corollary 3.2 implies that

$$(4.8) \qquad l_\psi^r(y) \leq \psi_m(y) \quad \forall y \in \mathcal{Y}, \ r \in \mathbb{Z}_+.$$

Hence, $\psi_m - l_\psi^r$ is an unsigned measure and its total variation norm is its mass.

(ii) Because G is non-negative, it is easy to see that

$$\rho G(y) \geq \rho_m G(y) = \psi_m(y) \quad \forall \rho \in \mathcal{L}, \ y \in \mathcal{Y},$$

which implies that

$$l_{g_y} := \inf_{\rho \in \mathcal{L}} \sum_{x \in \mathcal{S}} \rho(x) g(x, y) \geq \psi_m(y) \quad \forall y \in \mathcal{Y}.$$

On the other hand, $\psi_m(y) = \rho_m G(y)$ and ρ_m belongs to $\mathcal{L}$. Hence,

$$l_{g_y} = \psi_m(y) \quad \forall y \in \mathcal{Y}.$$

Choose any $\varepsilon > 0$. Because ψ_m is a probability distribution and $\mathcal{Y}_r$ approaches $\mathcal{Y}$ as $r \to \infty$, there exists an $r' \in \mathbb{Z}_+$ such that $\psi_m(\mathcal{Y}^c_{r'}) \leq \varepsilon/2$. From (4.8), we have that

$$(4.9) \qquad \left|\left|\psi_m - l^r_\psi\right|\right| \leq \sum_{y \in \mathcal{Y}_{r'}} (\psi_m(y) - l^r_\psi(y)) + 2\psi_m(\mathcal{Y}^c_{r'}) \leq \sum_{y \in \mathcal{Y}_{r'}} (l_{g_y} - l^r_{g_y}) + \frac{\varepsilon}{2}.$$

Because $\mathcal{Y}_{r'}$ is finite and, for each y in $\mathcal{Y}$, the function $x \mapsto g(x, y)$ belongs to $\mathcal{W}$ (Assumption 2.2(ii, iii)), Corollary 3.6(i) implies that there exists an R in $\mathbb{Z}_+$ such that

$$\sum_{y \in \mathcal{Y}_{r'}} (l_{g_y} - l^r_{g_y}) \leq \frac{\varepsilon}{2} \quad \forall r \geq R.$$

Combining the above with (4.9), we have that

$$\left|\left|\psi_m - l^r_\psi\right|\right| \leq \varepsilon \quad \forall r \geq R.$$

Because ε is arbitrary, the result follows. $\qquad\square$

5. Approximation schemes. The results in Sections 3 and 4 can be used to establish computational procedures to approximate CILPs using finite-dimensional LPs.

Approximation Scheme A. The results in Section 3 establish a procedure to approximate CILPs (2.9)–(2.10) by solving the finite-dimensional LPs (3.8)–(3.9) defined over outer approximations $\mathcal{L}_r$ associated with truncations $\mathcal{X}_r$. These finite-dimensional LPs have $|\mathcal{X}_r|$ decision variables, $|\mathcal{E}_r| \leq |\mathcal{X}_r|$ equality constraints, and $|\mathcal{X}_r| + 3$ inequality constraints. We refer to this general approach as *Scheme A*, which can be used for different purposes:

- In general, the computed bounds provide converging approximations to the optimal values of the CILP, as shown by Corollary 3.6(i). Specifically, consider the upper and lower bounds in (3.10):

$$(5.1) \qquad \tilde{l}^r_f := l^r_f - c\left(\sup_{x \notin \mathcal{X}_r} \frac{|f(x)|}{w(x)}\right), \quad \tilde{u}^r_f := u^r_f + c\left(\sup_{x \notin \mathcal{X}_r} \frac{|f(x)|}{w(x)}\right), \quad \forall r \in \mathbb{Z}_+.$$

The gap

$$(5.2) \qquad \gamma_r := \tilde{u}^r_f - \tilde{l}^r_f = u^r_f - l^r_f + 2c\left(\sup_{x \notin \mathcal{X}_r} \frac{|f(x)|}{w(x)}\right) \quad \forall r \in \mathbb{Z}_+,$$

bounds the approximation errors:

$$(5.3) \qquad \left|u_f - \tilde{u}^r_f\right| \leq \gamma_r \quad \text{and} \quad \left|l_f - \tilde{l}^r_f\right| \leq \gamma_r \quad \forall r \in \mathbb{Z}_+.$$

- In other applications, the optimal points ρ^* and ρ_* are of interest. As shown in Corollary 3.6(ii), the optimal points ρ^r_* and ρ^{r*} of the LPs (3.8)–(3.9) converge to the set of optimal points ρ_* and ρ^* of the CILP (2.9)–(2.10), respectively.
  If the optimal point ρ_* is unique, then solving the LP (3.8) yields convergent approximations ρ^r_* of ρ_* (Corollary 3.6(ii)), yet with no easy-to-compute quantification of the error. If ρ^* is unique, the same applies to ρ^{r*} from (3.9).
In the case of a unique feasible point ($\mathcal{L} = \{\rho\}$), our results strenghten considerably:

- Using the mid-point of $\tilde{l}_f^r$ and $\tilde{u}_f^r$ to approximate the integral $\rho(f)$, we find that the error is bounded above by

(5.4)
$$\left| \rho(f) - \frac{\tilde{u}_f^r + \tilde{l}_f^r}{2} \right| \leq \frac{\gamma_r}{2} \quad \forall r \in \mathbb{Z}_+,$$

which itself converges to zero:

$$\lim_{r \to \infty} \gamma_r = 0, \quad \text{if } f \in \mathcal{W}.$$

Therefore any desired tolerance can be verified by computing bounds from truncations with large enough r. From the definition (5.2), note that

$$\gamma_r \geq 2c \left(\sup_{x \notin \mathcal{X}_r} \frac{|f(x)|}{w(x)} \right) \quad \forall r \in \mathbb{Z}_+.$$

Hence, the truncation parameter r should be chosen large enough that

$$\sup_{x \notin \mathcal{X}_r} \frac{|f(x)|}{w(x)} \leq \frac{\varepsilon}{c},$$

where ε denotes the desired error tolerance.
- If the feasible point ρ (or its image $\psi = \rho G$) is of interest, any feasible point ρ^r of $\mathcal{L}_r$ (or its image $\psi^r = \rho^r G$) provides a convergent approximations to ρ (or to ψ), for large enough r. However, there is no easy-to-compute quantification of the error of the approximation for ρ (or ψ). To remedy this, we introduce Scheme B, which yields converging approximations of ρ (and ψ) with easy-to-compute error bounds.

Approximation Scheme B. We refer to *Scheme B* as the process of repeatedly applying Scheme A to compute $|\mathcal{X}_r|$ pointwise lower bounds l_x^r (or $|\mathcal{Y}_r|$ lower bounds $l_{g_y}^r$) in order to construct an approximation of the minimal point (or its image) with controlled error given by Theorem 4.1 (or Corollary 4.3).

6. Motivating problems revisited. Schemes A and B are applicable to the motivating problems in Section 2.2. They yield controlled approximations of: (i) the stationary distributions of a chain, and (ii) the exit distributions and occupation measures associated with exit times of the chain. First, note that Assumption $2.2(i, ii)$ is satisfied automatically because both one-step matrices P of discrete-time chains and rate matrices Q of continuous-time ones are Metzler.

Stationary distributions. Let H and G be as in (2.15) for the discrete-time case or (2.18) for the continuous-time one, and suppose that every stationary distribution π satisfies the moment bound $\pi(w) \leq c$. Because G is the identity matrix, Assumption $2.2(iii, iv)$ is trivially satisfied with $a_r := r^{-1}$.

In the case of a unique stationary distribution π, Corollary $3.6(iii)$ yields converging approximations ρ^r of π. Furthermore, Corollary 4.3 yields converging approximations l_ψ^r of π with computable error bounds.

In the non-unique case, there exist several *ergodic distributions* π_i, each with support on a disjoint subset of the state space called a closed communicating class $\mathcal{C}_i$. Because the set of stationary distributions equals that of all convex combinations of the ergodic distributions [28, Theorems 17.10, 43.19], Corollary $3.6(ii)$ and Remark 3.7

show that the optimal points ρ^{*r} of (3.9) (with $f := 1_x$ for any state x inside the class $\mathcal{C}_i$) form converging approximations of π_i.

The schemes developed here are applied extensively to compute the stationary distributions of continuous-time chains in [29].

Exit distributions and occupation measures. Let H and G be as in (2.16) for the discrete-time case or (2.19) in the continuous-time case. For discrete-time chains, Assumption 2.2(iii) holds with $a_r := r^{-1}$ because one-step matrices P are row stochastic. For continuous-time chains, choosing $w(x) := (-q(x, x))^d$ for $d > 1$ means that (2.12) holds with $a_r = r^{1-d}$. Corollary 3.6(ii) (with $f := 1$) and Proposition 3.8 yield converging approximations ρ_*^r and ψ_*^r of the occupation measure ν and exit distribution μ, respectively. If error bounds are important, then Theorem 4.1 and Corollary 4.3 yield the approximations l^r and l_ψ^r accompanied by such bounds.

For the exit problem, Assumption 2.2(iv) is a very mild restriction: it requires that the chain must be able to leave the domain (in one or more jumps) from every state x for which $w(x) = 0$. If the exit time is almost surely finite, states from which the chain cannot leave the domain are irrelevant, as the chain has zero probability of visiting them.

7. Concluding remarks. In this paper, we present results on the approximation of countably infinite linear programs defined over bounded measure spaces. The approximations are linear programs defined over finite-dimensional subspaces of the original infinite-dimensional space (i.e. we go from ℓ^1 to its subspace $\{\rho \in \ell^1 : \rho(x) = 0, \forall x \notin \mathcal{X}_r\}$), and are accompanied by bounds on the approximation error incurred. The convergence properties of our approximations ensure that any desired error tolerance may be achieved by picking a large enough truncation $\mathcal{X}_r$.

We presented two approaches. Scheme A yields error-controlled approximations of the optimal values and uncontrolled approximations of the optimal points and their images. Scheme B yields error-controlled approximations of the minimal point (should it exist) and its image. The error control of Scheme B comes at a computational price: it entails solving a potentially large number of LPs, instead of one or two as in Scheme A. Yet, because each point-wise bound is calculated independently, it is straightforward to parallelise their computation and mitigate this additional computational cost.

In related work, Hernández-Lerma and Lasserre [21, 22, 23, 24] introduced a series of numerical schemes to approximate infinite-dimensional linear programs with finite-dimensional ones. Within their setting and notation, our approximation consists of: (i) *aggregating* the constraints $\{\rho H(x) = \phi(x), \forall x \in \mathcal{X}\}$ into $\{\rho H(x) = \phi(x), \forall x \in \mathcal{X}_r\}$; and (ii) *relaxing* the constraint $\rho(g) = 1$ to $1 - ca_r \leq \rho(g) \leq 1$. Our proof of Theorem 3.5 follows analogous ideas to those behind the convergence proofs in [21, 22, 23, 24], although we require less technical machinery due to our countable setting.

A notable difference of our work is that Hernández-Lerma and Lasserre's approach does not require a known moment bound; however, they assume finiteness of the optimal value, which must be verified in practice using the same type of tools (e.g. Foster-Lyapunov criteria) used to obtain moment bounds. To obtain weak* convergence without an explicit moment bound, Hernández-Lerma and Lasserre fix the objective of their LPs to minimising the integral of a norm-like function. This tacitly guarantees that the optimal points of their approximating LPs satisfy a moment bound. Instead, we directly append a moment bound as a constraint to the LP, thus retaining the ability to choose the objective at will without sacrificing the convergence guarantee. By choosing the objective carefully, we are then able to con-

struct approximations with easy-to-compute errors—something that is not possible with the approach in Refs. [21, 22, 23, 24]. The development of numerical methodologies [32, 31, 38, 20, 33, 49, 37, 1] that facilitate the computation of moment bounds in the two decades since the original work of Hernández-Lerma and Lasserre has made it possible to include such refinements in optimisation frameworks, and has encouraged us to develop the schemes presented in this paper.

The other main difference with the work of Hernández-Lerma and Lasserre is that our results are focused on countable index sets $\mathcal{X}$, whereas the applications in [21, 22, 23, 24] involve uncountable ones. To obtain finite-dimensional LPs, Hernández-Lerma and Lasserre discretise the index set using a dense countable subset of the space that contains the measures they wish to approximate. As a result of this discretisation, the restrictions of the desired measure are not necessarily feasible points of the finite-dimensional LPs. Hence the schemes in [21, 22, 23, 24] yield only converging *approximations* instead of converging *bounds*. For a more detailed comparison, see [27, Section 5.5].

Finally, we remark that although we focused here on a particular type of LPs motivated by the problems discussed in Section 2.2, variations of our approximation schemes can be applied to other CILPs. The critical ingredients required to guarantee convergence of Schemes A and B are: a moment bound, and boundedness of the entries of the feasible points ρ. In particular, any constraint of the form $\rho(f) = \alpha$ or $\rho(f) \leq \alpha$ with f in $\mathcal{W}$ may be added to, or removed from, the definition of the feasible set $\mathcal{L}$. The techniques of Section 3 carry over if the constraints involve functions f that are are non-negative or have finite supports. Otherwise, the techniques of Appendix C apply.

REFERENCES

[1] A. A. Ahmadi and A. Majumdar, *DSOS and SDSOS optimization: more tractable alternatives to sum of squares and semidefinite optimization*, SIAM Journal on Applied Algebra and Geometry, 3 (2019), pp. 193–230, doi:10.1137/18M118935X.

[2] E. Altman, *Denumerable constrained Markov decision processes and finite approximations*, Math. Oper. Res., 19 (1994), pp. 169–191, doi:10.1287/moor.19.1.169.

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

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

[5] S. Asmussen, *Applied Probability and Queues*, Springer-Verlag New York, 2003, doi:10.1007/b97236.

[6] R. Cavazos-Cadena, *Finite-state approximations for denumerable state discounted markov decision processes*, Appl. Math. Opt., 14 (1986), pp. 1–26, doi:10.1109/TAC.2009.2033848.

[7] E. D. Demaine, S. P. Fekete, and S. Gal, *Online searching with turn cost*, Theor. Comput. Sci., 361 (2006), pp. 342–355, doi:10.1016/J.TCS.2006.05.018.

[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 (2017), pp. 2239–2244, doi:10.1016/B978-0-444-63965-3.50375-5.

[9] G. R. Dowdy and P. I. Barton, *Bounds on stochastic chemical kinetic systems at steady state*, J. Chem. Phys., 148 (2018), p. 084106, doi:10.1063/1.5009950.

[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 (2015), pp. 708–720, doi:10.1016/J.EJOR.2015.04.026.

[12] A. Ghate, D. Sharma, and R. L. Smith, *A shadow simplex method for infinite linear programs*, Oper. Res., 58 (2010), pp. 865–877, doi:10.1287/opre.1090.0755.

[13] A. Ghate and R. L. Smith, *A linear programming approach to nonstationary infinite-horizon Markov decision processes*, Oper. Res., 61 (2013), pp. 413–425, doi:10.1287/opre.1120.1121.

[14] 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 (2017), p. 04LT01, doi:10.1088/1478-3975/aa75c6.

[15] P. W. GLYNN AND A. ZEEVI, *Bounding stationary expectations of Markov processes*, in Markov Processes and Related Topics: A Festschrift for Thomas G. Kurtz, Institute of Mathematical Statistics, 2008, pp. 195–214, doi:10.1214/074921708000000381.

[16] R. C. GRINOLD, *Finite horizon approximations of infinite horizon linear programs*, Math. Program., 12 (1977), pp. 1–17, doi:10.1007/BF01593765.

[17] X. GUO AND A. PIUNOVSKIY, *Discounted continuous-time Markov decision processes with constraints: unbounded transition and loss rates*, Math. Oper. Res., 36 (2011), pp. 105–132, doi:10.1287/moor.1100.0477.

[18] X. GUO AND W. ZHANG, *Convergence of controlled models and finite-state approximation for discounted continuous-time Markov decision processes with constraints*, Eur. J. Oper. Res., 16 (2014), pp. 486–496, doi:10.1016/j.ejor.2014.03.037.

[19] K. HELMES, S. RÖHL, AND R. H. STOCKBRIDGE, *Computing moments of the exit time distribution for Markov processes by linear programming*, Oper. Res., 49 (2001), pp. 516–530, doi:10.1287/opre.49.4.516.11221.

[20] D. HENRION, J. B. LASSERRE, AND J. LÖFBERG, *GloptiPoly 3: moments, optimization and semidefinite programming*, Optim. Methods Softw., 24 (2009), pp. 761–779, doi:10.1080/10556780802699201.

[21] O. HERNÁNDEZ-LERMA AND J. B. LASSERRE, *Approximation schemes for infinite linear programs*, SIAM J. Optim., 8 (1998), pp. 973–988, doi:10.1137/S1052623497315768.

[22] O. HERNÁNDEZ-LERMA AND J. B. LASSERRE, *Linear programming approximations for Markov control processes in metric spaces*, Acta Appl. Math., 51 (1998), pp. 123–139, doi:10.1023/A:1005826226226.

[23] O. HERNÁNDEZ-LERMA AND J. B. LASSERRE, *Further topics on discrete-time Markov control processes*, Springer, 1999, doi:10.1007/978-1-4612-0561-6.

[24] O. HERNÁNDEZ-LERMA AND J. B. LASSERRE, *Markov chains and invariant probabilities*, Birkhäuser, 2003, doi:10.1007/978-3-0348-8024-4.

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

[26] P. C. JONES, J. L. ZYDIAK, AND W. J. HOPP, *Stationary dual prices and depreciation*, Math. Program., 41 (1988), pp. 357–366, doi:10.1007/BF01580773.

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

[28] J. KUNTZ, *Markov chains revisited*, 2020, arXiv:2001.02183v2.

[29] J. KUNTZ, P. THOMAS, G.-B. STAN, AND M. BARAHONA, *Bounding the stationary distributions of the chemical master equation via mathematical programming*, The Journal of Chemical Physics, 151 (2019), p. 034109, doi:10.1063/1.5100670.

[30] J. KUNTZ, P. THOMAS, G.-B. STAN, AND M. BARAHONA, *Stationary distributions of continuous-time markov chains: a review of theory and truncation-based approximations*, SIAM Rev., to appear - arXiv preprint arXiv:1909.05794, (2021).

[31] J. B. LASSERRE, *Moments, positive polynomials and their applications*, Imperial College Press, 2009, doi:10.1142/p665.

[32] M. LAURENT, *Sums of squares, moment matrices and optimization over polynomials*, in Emerging applications of algebraic geometry, Springer-Verlag, 2009, pp. 157–270, doi:10.1007/978-0-387-09686-5_7.

[33] J. LÖFBERG, *YALMIP: A toolbox for modeling and optimization in MATLAB*, in IEEE International Symposium on Computer Aided Control System Design, 2004, pp. 284–289, doi:10.1109/CACSD.2004.1393890.

[34] S. P. MEYN AND R. L. TWEEDIE, *Stability of Markovian processes III: Foster-Lyapunov criteria for continuous-time processes*, Adv. Appl. Probab., 25 (1993), pp. 518–548, doi:10.2307/1427522.

[35] S. P. MEYN AND R. L. TWEEDIE, *Markov chains and stochastic stability*, Cambridge University Press, 2nd ed., 2009.

[36] R. G. J. MILLER, *Stationarity Equations in Continuous Time Markov Chains*, Trans. Amer. Math. Soc., 109 (1963), pp. 35–44, doi:10.2307/1993645.

[37] D. PAPP AND S. YILDIZ, *Sum-of-squares optimization without semidefinite programming*, SIAM J. Optim., 29 (2019), pp. 822–851, doi:10.1137/17M1160124.

[38] S. PRAJNA, A. PAPACHRISTODOULOU, AND P. A. PARRILO, *Introducing SOSTOOLS: a general purpose sum of squares programming solver*, in Proceedings of the 41st IEEE Conference on Decision and Control, vol. 1, IEEE, 2002, pp. 741–746, doi:10.1109/CDC.2002.1184594.

[39] T. PRIETO-RUMEAU AND O. HERNÁNDEZ-LERMA, *Discounted continuous-time controlled*

Markov chains: convergence of control models, J. Appl. Probab., 49 (2012), pp. 1072–1090, doi:10.1239/jap/1354716658.

[40] T. PRIETO-RUMEAU AND J. M. LORENZO, *Approximating ergodic average reward continuous-time controlled Markov chains*, IEEE Trans. Autom. Control, 55 (2009), pp. 201–207, doi:10.1109/TAC.2009.2033848.

[41] M. L. PUTERMAN, *Markov decision processes : discrete stochastic dynamic programming*, Wiley, 1994.

[42] H. E. ROMEIJN, D. SHARMA, AND R. L. SMITH, *Extreme point characterizations for infinite network flow problems*, Networks, 48 (2006), pp. 209–222, doi:10.1002/net.20134.

[43] C. T. RYAN, R. L. SMITH, AND M. A. EPELMAN, *A simplex method for uncapacitated pure-supply infinite network flow problems*, SIAM J. Optim., 28 (2018), pp. 2022–2048, doi:10.1137/17M1137553.

[44] Y. SAKURAI AND Y. HORI, *A convex approach to steady state moment analysis for stochastic chemical reactions*, in IEEE 56th Annual Conference on Decision and Control, 2017, pp. 1206–1211, doi:10.1109/CDC.2017.8263820.

[45] E. SCHWERER, *A linear programming approach to the steady-state analysis of Markov processes*, PhD thesis, Stanford University, 1996.

[46] T. C. SHARKEY AND H. E. ROMEIJN, *A simplex algorithm for minimum-cost network-flow problems in infinite networks*, Networks, 52 (2008), pp. 14–31, doi:10.1002/net.20221.

[47] R. L. SMITH AND R. Q. ZHANG, *Infinite Horizon Production Planning in Time-Varying Systems with Convex Production and Inventory Costs*, Manag. Sci., 44 (1998), pp. 1313–1320, doi:10.1287/mnsc.44.9.1313.

[48] A. M. VERSHIK AND V. TEMELT, *Some questions concerning the approximation of the optimal value of infinite-dimensional problems in linear programming*, Sib. Math. J., 9 (1968), pp. 591–601, doi:10.1007/BF02199093.

[49] Y. ZHENG, G. FANTUZZI, AND A. PAPACHRISTODOULOU, *Fast ADMM for sum-of-squares programs using partial orthogonality*, IEEE Trans. Autom. Control, 64 (2018), pp. 3869–3876, doi:10.1109/TAC.2018.2886170.

Appendix A. Proof of Theorem 2.3. First, we show that $\rho(f)$ is absolutely convergent for any ρ in $\mathcal{L}$. Replacing f by $|f|$ in (3.6) we find that

$$\text{(A.1)} \qquad \rho(|f|) = \sum_{x \in \mathcal{X}_1} |f(x)| \, \rho(x) + \sum_{x \notin \mathcal{X}_1} |f(x)| \, \rho(x)$$

$$\leq \sum_{x \in \mathcal{X}_1} |f(x)| \, \rho(x) + c \left(\sup_{x \notin \mathcal{X}_1} \frac{|f(x)|}{w(x)} \right) < \infty$$

where the right-most inequality follows from finiteness of $\mathcal{X}_1$ and our assumption that f belongs in $\mathcal{W}$.

Next, we show that (2.9) is solvable (i.e. Theorem 2.3(ii, iii)). To do so, we prove in Appendix A.1 that the set $\mathcal{L}$ is weak* sequentially compact in the sense that every sequence of points $(\rho^r)_{r \in \mathbb{Z}_+}$ contained in $\mathcal{L}$ has a subsequence $(\rho^{r_k})_{r \in \mathbb{Z}_+}$ that converges in weak* (c.f. Definition 3.3) to a limit ρ^∞ belonging to $\mathcal{L}$. We can then set $(\rho^r)_{r \in \mathbb{Z}_+}$ to be any sequence of feasible points of (2.9) satisfying

$$\lim_{r \to \infty} \rho^r(f) = l_f.$$

Because $\mathcal{L}$ is weak* sequentially compact and f belongs to $\mathcal{W}$, we can find a subsequence $(\rho^{r_k})_{r \in \mathbb{Z}_+}$ of $(\rho^r)_{r \in \mathbb{Z}_+}$ with a limit ρ^∞ belonging to $\mathcal{L}$ that satisfies

$$\rho^\infty(f) = \lim_{k \to \infty} \rho^{r_k}(f) = l_f.$$

Setting $\rho_* := \rho_\infty$ and replacing ρ in (A.1) with ρ_* then shows that (2.9) is solvable.

A.1. $\mathcal{L}$ is weak* sequentially compact. The argument consists of four steps:
(a) Using a standard diagonal argument to find a pointwise convergent subsequence $(\rho^{r_k})_{k \in \mathbb{Z}_+}$ of $(\rho^r)_{r \in \mathbb{Z}_+}$.

(b) Using $\rho^r(w) \leq c$ for all $r \in \mathbb{Z}_+$ and Fatou's lemma to show that the limit ρ^∞ of $(\rho^{r_k})_{k \in \mathbb{Z}_+}$ is non-negative ($\rho^\infty(x) \geq 0 \; \forall x \in \mathcal{X}$) and satisfies $\rho^\infty(w) \leq c$.

(c) Using (a,b) and show that $(\rho^{r_k})_{k \in \mathbb{Z}_+}$ converges in weak* to ρ^∞.

(d) Using (c) to show that ρ^∞ belongs to $\mathcal{L}$.

Let's begin:

(a) Enumerate the elements of $\mathcal{X}$ as $x_1, x_2, \ldots$ Assumption 2.2(i, ii, iv) and the constraints $\rho H = \phi$, $\rho(g) = 1$, and $\rho(w) \leq c$ in (2.8) imply that the sequence $(\rho^r(x_1))_{r \in \mathbb{Z}_+}$ is contained in a bounded interval. For this reason, the Bolzano-Weierstrass Theorem tells us that $(\rho^r(x_1))_{r \in \mathbb{Z}_+}$ has a converging subsequence $(\rho^{r_{k_1}}(x_1))_{k_1 \in \mathbb{Z}_+}$. Repeating the same argument for x_2 and $(\rho^{r_{k_1}})_{k_1 \in \mathbb{Z}_+}$, we obtain a convergent subsequence $(\rho^{r_{k_2}}(x_2))_{k_2 \in \mathbb{Z}_+}$ of $(\rho^{r_{k_1}}(x_2))_{k_1 \in \mathbb{Z}_+}$, and so on. Setting

$$\rho^{r_k} := \rho^{r_{k_k}} \quad \forall k \in \mathbb{Z}_+,$$

we obtain the desired pointwise convergent subsequence.

(b) Given (a), this follows directly from Fatou's lemma.

(c) For any f in $\mathcal{W}$ and r' in $\mathbb{Z}_+$,

$$(A.2) \qquad |\rho^{r_k}(f) - \rho^\infty(f)|$$
$$\leq \sum_{x \in \mathcal{X}_{r'}} |\rho^{r_k}(x) - \rho^\infty(x)| \, |f(x)| + \sum_{x \notin \mathcal{X}_{r'}} (\rho^{r_k}(x) + \rho^\infty(x)) \, |f(x)| \, .$$

The generalisation (3.6) of Markov's inequality (with $|f|$ replacing f) and the moment bounds $\rho^\infty(w) \leq c$ and $\rho^{r_k}(w) \leq c$ tell us that

$$\sum_{x \notin \mathcal{X}_{r'}} (\rho^{r_k}(x) + \rho^\infty(x)) \, |f(x)| \leq 2c \left(\sup_{x \notin \mathcal{X}_{r'}} \frac{|f(x)|}{w(x)} \right) .$$

Fix $\varepsilon > 0$. Because f belongs to $\mathcal{W}$, we can find an r' in $\mathbb{Z}_+$ such that $\sup_{x \notin \mathcal{X}_{r'}} (|f(x)| / w(x)) \leq \frac{\varepsilon}{4c}$. It follows from the above that

$$(A.3) \qquad \sum_{x \notin \mathcal{X}_{r'}} (\rho^{r_k}(x) + \rho^\infty(x)) \, |f(x)| \leq \frac{\varepsilon}{2}.$$

Because $\mathcal{X}_{r'}$ is a finite set (w is norm-like), the pointwise convergence of ρ^{r_k} to ρ^∞ implies that there exist a K such that

$$(A.4) \qquad \sum_{x \in \mathcal{X}_{r'}} |\rho^{r_k}(x) - \rho^\infty(x)| \, |f(x)| \leq \frac{\varepsilon}{2} \quad \forall k \geq K.$$

Combining (A.2)–(A.4) yields

$$|\rho^{r_k}(f) - \rho^\infty(f)| \leq \varepsilon \quad \forall k \geq K.$$

Because the ε was arbitrary, we have the desired limit:

$$\lim_{k \to \infty} \rho^{r_k}(f) = \rho^\infty(f).$$

Since the above holds for every $f \in \mathcal{W}$, we have that ρ^{r_k} not only converges pointwise to ρ but also in weak*.

(d) Assumption $2.2(ii, iii, v)$ implies that g and $x' \mapsto h(x', x)$ (for any $x \in \mathcal{X}$) belong to $\mathcal{W}$. For this reason, the weak* convergence of the subsequence implies that

$$\rho^\infty H(x) = \lim_{k \to \infty} \rho^{r_k} H(x) = \phi(x) \quad \forall x \in \mathcal{X}, \qquad \rho^\infty(g) = \lim_{k \to \infty} \rho^{r_k}(g) = 1.$$

Because we already argued in (b) that $\rho^\infty(x) \geq 0$ for all x in $\mathcal{X}$ and $\rho^\infty(w) \leq c$, the above shows that that ρ^∞ belongs to $\mathcal{L}$.

Appendix B. Weak* convergence implies convergence in total variation.
The argument is as follows: given our identification (2.14) of ℓ^1 with the space of finite signed measures on $(\mathcal{X}, 2^{\mathcal{X}})$, the total variation norm $||\cdot||$ on ℓ^1 is dominated by (in fact, equivalent to) the ℓ^1-norm:

$$||\rho|| \leq \sum_{x \in \mathcal{S}} |\rho(x)| =: ||\rho||_1 \quad \forall \rho \in \ell^1.$$

Because the dual of $(\ell^1, ||\cdot||_1)$ is isomorphic to the space of bounded functions on $\mathcal{X}$, Schur's Theorem tells us that a sequence $(\rho^r)_{r \in \mathbb{Z}_+} \subseteq \ell^1$ converges in the ℓ^1-norm to $\rho \in \ell^1$ if and only if (3.15) holds for all bounded functions f. Because the norm-like assumption implies that all bounded functions belong to $\mathcal{W}$, it follows that weak* convergence implies convergence in ℓ^1 and, consequently, in total variation.

Appendix C. Shedding Assumption 2.2(v). While Assumptions $2.2(i--iv)$ are very mild and widely satisfied in applications, Assumption $2.2(v)$ requiring that every column of H has finitely many non-zero entries is more restrictive. Indeed, it is not exceedingly rare for a discrete-time chain to possess one or more states that are reachable from infinitely many other states in a single step, in which case the one-step matrix P would violate Assumption $2.2(v)$. Although Assumption $2.2(v)$ does simplify the content of this paper, it can be circumvented as follows.

The assumption is important for two reasons: (a) practically, it ensures that, for each x in $\mathcal{X}$, the equation

$$\phi(x) = \rho H(x) = \sum_{x' \in \mathcal{X}} \rho(x') h(x', x) = \sum_{x' \in \mathrm{supp}(h(\cdot, x))} \rho(x') h(x', x)$$

involves only finitely many entries of ρ, a key fact in the implementation of our schemes; and (b) theoretically, it ensures that, for each x in $\mathcal{X}$, the function $x' \mapsto h(x', x)$ belongs to $\mathcal{W}$ in (2.13). For this reason, the weak* convergence of the approximating sequences $(\rho^r)_{r \in \mathbb{Z}_+}$ of Section 3 guarantees that their limit points ρ^∞ satisfy the equations $\rho H = \phi$, a fact necessary when showing that the limit points belonging to $\mathcal{L}$ (see the proof of Theorem 3.5 for details).

This assumption can be avoided using the moment bound (2.7) at the expense of doubling the number of decision variables in our finite-dimensional LPs. To do so, we require a sequence $b_1, b_2, \ldots$ of known constants such that

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

The above implies that $x' \mapsto h(x', x)$ belongs to $\mathcal{W}$ for each $x \in \mathcal{X}$ and we recover (b). To recover finite-dimensional outer approximations of $\mathcal{L}$ of the sort in Section 3, pick any index x in our truncation $\mathcal{X}_r$ and consider its associated equation:

$$(C.2) \qquad \rho H(x) = \sum_{x' \in \mathcal{X}_r} \rho(x') h(x', x) + \epsilon_r(x) = \phi(x),$$

where the measure ϵ_r is defined by

$$\epsilon_r(x) := \begin{cases} \sum_{x' \notin \mathcal{X}_r} \rho(x') h(x', x) & \text{if } x \in \mathcal{X}_r \\ 0 & \text{if } x \notin \mathcal{X}_r \end{cases} \qquad \forall x \in \mathcal{X}.$$

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

(C.3) $$\epsilon_r(x) \geq 0 \quad \forall x \in \mathcal{X}.$$

Tonelli's Theorem, the generalisation (3.6) of Markov's inequality, and (C.1) imply

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

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

Putting (3.5), (3.7), and (C.2)–(C.4) together, we recover a finite dimensional outer approximation of $\mathcal{L}$: if ρ belongs to $\mathcal{L}$, then the pair $(\rho_{|r}, \epsilon_r)$ belongs to

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

For this reason, replacing Assumption 2.2(v) with 'there exists a known sequence $(b_r)_{r \in \mathbb{Z}_+}$ such that (C.1) holds', $\mathcal{L}_r$ with $\tilde{\mathcal{L}}_r$, and l_f^r, u_f^r in (3.8)–(3.9) with

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

the results of Sections 3–4 hold identically. The only difference is that in contrast with $\mathcal{L}_r$ in (3.2), $\tilde{\mathcal{L}}_r$ involves $2|\mathcal{X}_r|$ variables decision variables, $|\mathcal{X}_r|$ equalities, and $2|\mathcal{X}_r| + 4$ inequalities.