

# THE EXIT TIME FINITE STATE PROJECTION SCHEME: BOUNDING EXIT DISTRIBUTIONS AND OCCUPATION MEASURES OF CONTINUOUS-TIME MARKOV CHAINS\*

JUAN KUNTZ<sup>†</sup>, PHILIPP THOMAS<sup>‡</sup>, GUY-BART STAN<sup>§</sup>, AND MAURICIO BARAHONA<sup>¶</sup>

**Abstract.** We introduce the exit time finite state projection (ETFSP) scheme, a truncation-based method that yields approximations to the exit distribution and occupation measure associated with the time of exit from a domain (i.e., the time of first passage to the complement of the domain) of time-homogeneous continuous-time Markov chains. We prove that (i) the computed approximations bound the measures from below; (ii) the total variation distances between the approximations and the measures decrease monotonically as states are added to the truncation; and (iii) the scheme converges, in the sense that, as the truncation tends to the entire state space, the total variation distances tend to zero. Furthermore, we give a computable bound on the total variation distance between the exit distribution and its approximation, and we delineate the cases in which the bound is sharp. We also revisit the related finite state projection scheme and give a comprehensive account of its theoretical properties. We demonstrate the use of the ETFSP scheme by applying it to two biological examples: the computation of the first passage time associated with the expression of a gene, and the fixation times of competing species subject to demographic noise.

**Key words.** exit times, first passage times, continuous-time Markov chains, exit time finite state projection, finite state projection, exit distribution, occupation measure

**AMS subject classifications.** 60J27, 60J28, 65C40, 65G20

**DOI.** 10.1137/18M1168261

**1. Introduction.** The time of exit of a continuous-time Markov chain from a domain (or *exit time*, for short) is the time at which the chain leaves the domain for the first time. The exit time is also known as the *first passage time* or, alternatively, the *hitting time* of the complement of the domain. Two measures are associated with an exit event: the *exit distribution*, which describes when and where the chain exits the domain, and the *occupation measure*, which describes which states the chain visits before exiting and at what times they are visited. These two measures can be expressed in terms of the *time-varying law* (i.e., the state space distribution of the chain as a function of time) of an *auxiliary chain* that is identical to the original chain except that every state outside of the domain is turned into an absorbing state [54, 35].

There exists a rich literature on exit times, especially in physics and biomath-

---

\*Submitted to the journal's Methods and Algorithms for Scientific Computing section February 1, 2018; accepted for publication (in revised form) January 23, 2019; published electronically March 12, 2019.

<http://www.siam.org/journals/sisc/41-2/M116826.html>

**Funding:** The first author's work was supported by a BBSRC Ph.D. Studentship (BB/F017510/1). The second author's work was supported by a Fellowship of The Royal Commission for the Exhibition of 1851. The third author's work was supported by an EPSRC Fellowship for Growth (EP/M002187/1) and EU H2020-FETOPEN-2016-2017 project 766840 (COSY-BIO). The fourth author's work was supported by EPSRC grant EP/N014529/1 supporting the EPSRC Centre for Mathematics of Precision Healthcare.

<sup>†</sup>Department of Mathematics and Department of Bioengineering, Imperial College London, London, SW7 2AZ, UK (juan.kuntz08@imperial.ac.uk).

<sup>‡</sup>Department of Mathematics, Imperial College London, London, SW7 2AZ, UK (p.thomas@imperial.ac.uk).

<sup>§</sup>Co-corresponding author. Department of Bioengineering, Imperial College London, London, SW7 2AZ, UK (g.stan@imperial.ac.uk).

<sup>¶</sup>Co-corresponding author. Department of Mathematics, Imperial College London, London, SW7 2AZ, UK (m.barahona@imperial.ac.uk).

ematics [31, 41, 49]. Recently, there has been renewed interest in exit times of continuous-time Markov chains with discrete state space [2, 23, 55, 51], such as those we study in this paper. While the exit problem from a small finite domain is tractable [25, 26], the exit problem from an infinite or large domain can only be solved in special cases [17, 23, 47]. Various approximation schemes have been developed to address this issue [3, 12, 29, 51]. However, most of them do not provide bounds or error estimates on their accuracy.

The popular *finite state projection* (FSP) scheme [45] yields lower bounds on the time-varying law of the continuous-time chain of interest. The scheme chooses a finite truncation of the state space and solves an associated system of linear ordinary differential equations (ODEs) indexed by the states contained in the truncation. Built into the procedure is a computable upper bound on the total variation distance between the lower bounds obtained and the time-varying law. However, the FSP does not provide information about the exit from a domain.

To fill this gap, we introduce the *exit time finite state projection* (ETFSP) scheme (section 1.1), which involves applying an FSP-like scheme to the auxiliary chain with an absorbing complement mentioned above. We show that the scheme yields lower bounds on the exit distribution and the occupation measure associated with an exit time. For the exit distribution, we explain how to compute a bound on the error of its approximation. Theorem 1.1 delineates the theoretical properties of ETFSP showing that (i) the error bound is sharp if and only if the exit event occurs with probability one, and (ii) the scheme converges in total variation to the exit distribution and occupation measure as the truncation approaches the entire state space.

A secondary contribution is Theorem 2.5, which gathers the theoretical properties of the FSP scheme. In particular, we show that the error bound of the FSP scheme is sharp if and only if the chain is nonexplosive, in which case the error bound can indeed be made arbitrarily small by including enough states in the truncation. In the explosive case, the error bound remains nonzero, as observed in [37], and is limited by the probability of explosion.

The final contribution of this paper is a new proof of an old theorem: Theorem 2.6 expresses the exit distribution in terms of the time-varying law of the aforementioned auxiliary chain. Versions of this theorem pepper the literature (e.g., [21, 56, 48, 38, 49, 54]). Our variant relaxes the nonexplosive and deterministic initial condition assumptions in [54] and adds the analogous result for the occupation measure.

*Related literature.* To the best of our knowledge, the ideas underpinning the FSP and ETFSP schemes were first delineated in the 1980s queuing literature (see [27, 38, 39] and references therein) centered around randomization techniques for continuous-time chains with bounded rate matrices. Recently, schemes based on the FSP have been used to bound the cumulative density function of the exit time of stochastic reaction networks, a subclass of the continuous-time chains that we consider here [5, 6, 12]. Given that the ETFSP scheme bounds not only the cumulative density functions but also the corresponding densities, our results imply the convergence of those other schemes as a special case.

**1.1. The ETFSP scheme: Statement of the problem and main result.** We briefly define our problem setting, introduce the ETFSP scheme, and state our main result (Theorem 1.1) detailing the theoretical properties of the scheme.

**Problem definition.** Let  $X := \{X_t : 0 \leq t < T_\infty\}$  be a minimal time-homogeneous continuous-time Markov chain on a probability triplet  $(\Omega, \mathcal{F}, \mathbb{P})$  with countable

state space  $\mathcal{S}$ , stable and conservative rate matrix  $Q := (q(x, y))_{x,y \in \mathcal{S}}$ , explosion time  $T_\infty$ , and initial distribution  $\gamma(x) := \mathbb{P}(\{X_0 = x\}) \forall x \in \mathcal{S}$ .

We single out a subset  $\mathcal{D}$  of the state space  $\mathcal{S}$  and refer to it as the *domain*. The *exit time*  $\tau$  from the domain is the time when the chain first leaves  $\mathcal{D}$ :

$$(1.1) \quad \tau(\omega) := \inf\{t \in [0, T_\infty(\omega)) : X_t(\omega) \notin \mathcal{D}\} \quad \forall \omega \in \Omega,$$

with the convention that the infimum of the empty set is infinity:  $\inf\{\emptyset\} = \infty$ .

The exit distribution  $\mu$  and occupation measure  $\nu$  associated with the exit time are defined as

$$(1.2) \quad \mu([a, b], x) := \mathbb{P}(\{\tau \in [a, b], X_\tau = x\}) \quad \forall 0 \leq a < b < \infty, \forall x \in \mathcal{S},$$

$$(1.3) \quad \nu([a, b], x) := \mathbb{E} \left[ \int_{a \wedge \tau \wedge T_\infty}^{b \wedge \tau \wedge T_\infty} 1_x(X_t) dt \right] \quad \forall 0 \leq a < b < \infty, \forall x \in \mathcal{S},$$

where  $c \wedge d = \min(c, d)$ ,  $c, d \in \mathbb{R}$ , and  $1_x$  denotes the indicator function of state  $x$ :  $1_x(y) = 1$  if  $y = x$  and 0 otherwise.

For each state  $x$ , the measures  $\mu(dt, x)$  and  $\nu(dt, x)$  have densities  $\mu(t, x)$  and  $\nu(t, x)$  with respect to the Lebesgue measure. (We distinguish a measure from its density by writing  $dt$  or  $t$  in its argument.) For small  $h > 0$ , the distribution  $t \mapsto \mu(t, x)$  is a function such that  $\mu(t, x)h$  is the probability that the chain first exits the domain via state  $x$  during the time interval  $[t, t+h]$ . Similarly (and assuming nonexplosivity for the chain),  $\nu(t, x)$  is the average fraction of the interval  $[t, t+h]$  that the chain spends in state  $x$  before exiting the domain. Formally, the relationship between the exit distribution and occupation measure and their densities is

$$(1.4) \quad \mu([a, b], x) = 1_{\mathcal{D}^c}(x) \gamma(x) 1_0(a) + \int_a^b \mu(t, x) dt \quad \forall 0 \leq a < b < \infty, \forall x \in \mathcal{S},$$

$$(1.5) \quad \nu([a, b], x) = \int_a^b \nu(t, x) dt \quad \forall 0 \leq a < b < \infty, \forall x \in \mathcal{S},$$

where  $\mathcal{D}^c$  denotes the complement of the domain and the term  $1_{\mathcal{D}^c}(x) \gamma(x) 1_0(a)$  captures the event that the chain is started outside of the domain.

In this paper, we introduce the *exit time finite state projection* (ETFSP) scheme to approximate the exit distribution and occupation measure in a systematic manner. ETFSP yields approximations of the densities  $\mu(t, x)$  and  $\nu(t, x)$ , and, consequently, of their marginals, including the distribution of the exit time,  $\tau$ , and of the exit location,  $X_\tau$ .

**The exit time finite state projection (ETFSP) scheme.** The numerical scheme consists of the following steps:

1. Choose a finite subset, or *truncation*,  $\mathcal{S}_r$  of the state space  $\mathcal{S}$  and a *final computation time*,  $t_f^r \in [0, \infty)$ .
2. Solve the set of  $|\mathcal{S}_r|$  linear ODEs,

$$(1.6) \quad \dot{\nu}^r(t, x) = \sum_{y \in \mathcal{D}_r} \nu^r(t, y) q(y, x),$$

$$\nu^r(0, x) = \gamma(x) \quad \forall x \in \mathcal{D}_r,$$

$$(1.7) \quad \dot{\mu}^r(t, x) = \sum_{y \in \mathcal{D}_r} \nu^r(t, y) \left( \sum_{z \in \mathcal{D}_r} q(y, z) q(z, x) \right),$$

$$\mu^r(0, x) = \sum_{y \in \mathcal{D}_r} \gamma(y) q(y, x) \quad \forall x \in \mathcal{S}_r \cap \mathcal{D}^c,$$

over the time interval  $[0, t_f^r]$ , where  $\mathcal{D}_r$  denotes the *truncated domain*  $\mathcal{D} \cap \mathcal{S}_r$ .  
 3. Pad  $\nu^r$  and  $\mu^r$  with zeros:

$$(1.8) \quad \begin{aligned} \nu^r(t, x) &:= 0 && \text{if } x \notin \mathcal{D}_r \text{ or } t > t_f^r, \\ \mu^r(t, x) &:= 0 && \text{if } x \notin \mathcal{S}_r \cap \mathcal{D}^c \text{ or } t > t_f^r. \end{aligned}$$

The approximations of the measures  $\mu(dt, x)$  and  $\nu(dt, x)$  are defined as

$$(1.9) \quad \mu^r([a, b], x) := 1_{\mathcal{D}^c \cap \mathcal{S}_r}(x) \gamma(x) 1_0(a) + \int_a^b \mu^r(t, x) dt \quad \forall 0 \leq a < b < \infty, \forall x \in \mathcal{S},$$

$$(1.10) \quad \nu^r([a, b], x) := \int_a^b \nu^r(t, x) dt \quad \forall 0 \leq a < b < \infty, \forall x \in \mathcal{S}.$$

**Theoretical characterization of the ETFSP scheme.** Our main result is Theorem 1.1, which summarizes the theoretical properties of the scheme. (For its proof, see section 3.) We show that  $\mu^r$  and  $\nu^r$  do not just approximate the exit distribution  $\mu$  and occupation measure  $\nu$ , but also bound them from below. We give simple expressions for the mass of the approximations and their errors in terms of the exit time  $\tau$ , the final computation time  $t_f^r$ , and exit time from the truncation  $\mathcal{S}_r$ ,

$$(1.11) \quad \tau_r := \inf\{0 \leq t < T_\infty : X_t \notin \mathcal{S}_r\} \quad \forall r \in \mathbb{N}.$$

To quantify the *approximation errors*  $\|\mu - \mu^r\|$  and  $\|\nu - \nu^r\|$ , we use the total variation norm

$$(1.12) \quad \|\rho\| := \sup\{|\rho(\mathcal{A})| : \mathcal{A} \in \mathcal{G}\}$$

on the measures. We give easy-to-compute bounds for the approximation errors, and we show that the bounds are sharp if the chain exits the domain almost surely (a property that can be verified using Foster-Lyapunov criteria [40, 35]). Last, we prove that the approximation errors and their bounds decrease monotonically as we increase the truncation  $\mathcal{S}_r$ , and that the errors tend to zero as  $\mathcal{S}_r$  tends to  $\mathcal{S}$ .

**THEOREM 1.1** (exit time finite state projection scheme). *Consider a minimal time-homogeneous continuous-time Markov chain with countable state space  $\mathcal{S}$ , stable and conservative rate matrix  $Q := (q(x, y))_{x, y \in \mathcal{S}}$ , explosion time  $T_\infty$ , and initial distribution  $\gamma := (\gamma(x))_{x \in \mathcal{S}}$ . Suppose that the initial distribution satisfies*

$$(1.13) \quad \sum_{x \in \mathcal{D}} \gamma(x) |q(x, x)| < \infty$$

for a given domain  $\mathcal{D} \subseteq \mathcal{S}$ , and let  $\mu$  and  $\nu$  denote, respectively, the exit distribution and occupation measure associated with the exit time  $\tau$  from  $\mathcal{D}$ . Let  $\{\mathcal{S}_r\}_{r \in \mathbb{N}}$  be an increasing sequence of finite sets contained in  $\mathcal{S}$ ,  $\{t_f^r\}_{r \in \mathbb{N}}$  an increasing sequence of nonnegative final computation times, and  $\{\mu^r\}_{r \in \mathbb{N}}$  and  $\{\nu^r\}_{r \in \mathbb{N}}$  the sequences of ETFSP approximations of the exit distribution and occupation measure, respectively, defined by (1.6)–(1.8). Then the following properties hold:

(i) (Increasing sequence of lower bounds.)

$$\begin{aligned} \mu^0(t, x) &\leq \mu^1(t, x) \leq \dots \leq \mu(t, x) && \forall x \in \mathcal{S}, \quad t \in [0, \infty), \\ \nu^0(t, x) &\leq \nu^1(t, x) \leq \dots \leq \nu(t, x) && \forall x \in \mathcal{S}, \quad t \in [0, \infty). \end{aligned}$$

- (ii) (*Mass of the approximations.*) The mass of  $\mu^r$  is the probability that the chain exits the domain no later than exiting the truncation or the final time, i.e.,

$$\mu^r([0, \infty), \mathcal{S}) = \mathbb{P}(\{\tau \leq t_f^r \wedge \tau_r\}) \quad \forall r \in \mathbb{N}.$$

The mass of  $\nu^r$  is

$$\nu^r([0, \infty), \mathcal{S}) = \mathbb{E}[(\tau \wedge t_f^r) \mathbf{1}_{\{\tau \leq \tau_r\}}] \quad \forall r \in \mathbb{N}.$$

- (iii) (*Computable error bounds.*) For any  $r \in \mathbb{N}$ ,

$$(1.14) \quad \begin{aligned} \|\mu - \mu^r\| &= \mathbb{P}(\{\tau < \infty\}) - \mathbb{P}(\{\tau \leq t_f^r \wedge \tau_r\}) \\ &\leq 1 - \left( \gamma(\mathcal{D}^c \cap \mathcal{S}_r) + \sum_{x \in \mathcal{D}^c \cap \mathcal{S}_r} \int_0^{t_f^r} \mu^r(t, x) dt \right) =: \varepsilon_r, \end{aligned}$$

$$(1.15) \quad \begin{aligned} \|\nu - \nu^r\| &= \mathbb{E}[\tau \wedge T_\infty] - \mathbb{E}[(\tau \wedge t_f^r) \mathbf{1}_{\{\tau \leq \tau_r\}}] \\ &\leq \mathbb{E}[\tau] - \sum_{x \in \mathcal{D}_r} \int_0^{t_f^r} \nu^r(t, x) dt =: \varepsilon_r^\nu. \end{aligned}$$

Equality holds for (1.14) if and only if  $\mathbb{P}(\{\tau < \infty\}) = 1$ , i.e., when the chain exits the domain with probability one.

- (iv) (*Monotonicity of the error and of the error bound.*) The approximation errors and their upper bounds are decreasing in  $r$ :

$$(1.16) \quad \|\mu - \mu^r\| \geq \|\mu - \mu^{r+1}\| \quad \text{and} \quad \varepsilon_r \geq \varepsilon_{r+1} \quad \forall r \in \mathbb{N},$$

$$(1.17) \quad \|\nu - \nu^r\| \geq \|\nu - \nu^{r+1}\| \quad \text{and} \quad \varepsilon_r^\nu \geq \varepsilon_{r+1}^\nu \quad \forall r \in \mathbb{N}.$$

- (v) (*Convergence of bounds.*) If  $\cup_r \mathcal{S}_r = \mathcal{S}$  and  $t_f^r \rightarrow \infty$  as  $r \rightarrow \infty$ , the approximation  $\mu^r$  converges in total variation to the exit distribution  $\mu$ :

$$\lim_{r \rightarrow \infty} \|\mu - \mu^r\| = 0.$$

Consequently, it follows from (iii) that

$$\lim_{r \rightarrow \infty} \varepsilon_r = 0 \iff \mathbb{P}(\{\tau < \infty\}) = 1.$$

If  $\mathbb{E}[\tau \wedge T_\infty] < \infty$ , the approximation  $\nu^r$  converges in total variation to the occupation measure  $\nu$ :

$$\lim_{r \rightarrow \infty} \|\nu - \nu^r\| = 0.$$

We refer to the upper bound  $\varepsilon_r$  defined in (1.14) as the *error bound* of the scheme because it bounds the approximation error of  $\mu^r$ . Note that the error bound is easily calculated from  $\mu^r$ ; hence assessing the quality of the approximation requires no extra effort. The bound  $\varepsilon_r^\nu$  for the occupation measure  $\nu^r$  is harder to evaluate because the mean exit time,  $\mathbb{E}[\tau]$ , is unknown in general. However, an upper bound on  $\mathbb{E}[\tau]$  can be obtained through additional computations beyond the scope of this paper (section 5).

Condition (1.13) is a mild technical assumption (e.g., it is satisfied if the chain is initialized deterministically) made to simplify the exposition by ensuring that the density of the exit distribution is finite at time zero:  $\mu(0, x) < \infty \forall x \in \mathcal{S}$ .

**Organization of the paper.** The remainder of the paper is structured as follows. In section 2, we formally define the chain and give several preliminary lemmas required in the subsequent proofs. Specifically, we review the forward equations and provide proofs for theoretical properties of the original FSP (section 2.1), and we give the analytical characterization of the exit distribution and occupation measure and the marginals of these measures (section 2.1). For the sake of readability, we have relegated the technical proofs relevant to section 2 to the supplementary material. Section 3 contains the proof of Theorem 1.1. In section 4, we apply the ETFSP scheme to two biologically motivated examples. We conclude by discussing possible implementations and extensions of the ETFSP scheme in section 5.

**2. Preliminaries.** The starting point in our definition of a continuous-time chain is a stable and conservative rate matrix  $Q := (q(x, y))_{x, y \in \mathcal{S}}$ , that is, a matrix of real numbers indexed by the countable state space  $\mathcal{S}$  satisfying

$$(2.1) \quad q(x, y) \geq 0 \quad \forall x \neq y, \quad q(x, x) = -\sum_{y \neq x} q(x, y) > -\infty \quad \forall x \in \mathcal{S}.$$

Whenever we write “a rate matrix  $Q$ ” in this paper, we mean “a stable and conservative rate matrix  $Q$ .“ We construct our Markov chain  $X$  recursively by running the *Gillespie algorithm* [20, 33, 24] (see Appendix A in the supplementary material). In particular, the algorithm returns the *jump times*  $\{T_n\}_{n \in \mathbb{N}}$  at which transitions occur and the sequence  $Y := \{Y_n\}_{n \in \mathbb{N}}$  of states visited by the chain; both of these are defined on the same probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . The sequence  $Y$  is itself a discrete-time Markov chain known as the *jump chain* (or *embedded chain*), and its one-step matrix is

$$(2.2) \quad \pi(x, y) := \begin{cases} (1_x(y) - 1) q(x, y)/q(x, x) & \text{if } q(x, x) \neq 0, \\ 1_x(y) & \text{otherwise} \end{cases} \quad \forall x, y \in \mathcal{S}.$$

The sample paths  $t \mapsto X_t(\omega)$  of the continuous-time chain  $X$  are defined by

$$(2.3) \quad X_t(\omega) := Y_n(\omega) \quad \forall t \in [T_n(\omega), T_{n+1}(\omega)), \quad \omega \in \Omega.$$

These paths are defined only up until the *explosion time*

$$T_\infty(\omega) := \lim_{n \rightarrow \infty} T_n(\omega) \quad \forall \omega \in \Omega.$$

The limit exists because  $\{T_n(\omega)\}_{n \in \mathbb{N}}$  is an increasing sequence for each  $\omega \in \Omega$ . In other words,  $X_t(\omega)$  is defined only for pairs  $(t, \omega)$  such that  $t < T_\infty(\omega)$ . The reason behind the name “explosion time” given to  $T_\infty$  is that, by this moment in time, the chain has left every finite subset of the state space. In particular, let  $\mathcal{S}_0 \subseteq \mathcal{S}_1 \subseteq \dots$  be an increasing sequence of finite subsets (or *truncations*) of  $\mathcal{S}$  such that  $\cup_r \mathcal{S}_r = \mathcal{S}$ , and let  $\tau_r$  be the time (1.11) that the chain  $X$  first exits  $\mathcal{S}_r$ . That our truncations form an increasing sequence implies that  $\{\tau_r\}_{r \in \mathbb{N}}$  is an increasing sequence of random variables and the limit  $\lim_{r \rightarrow \infty} \tau_r(\omega)$  exists for each  $\omega \in \Omega$ .

LEMMA 2.1 (Lemma 2.18 of [35]). *If  $\{\mathcal{S}_r\}_{r \in \mathbb{N}}$  is an increasing sequence of finite sets such that  $\cup_r \mathcal{S}_r = \mathcal{S}$ , then  $\tau_r$  tends to  $T_\infty$  almost surely.*

The limiting random variable  $\lim_{r \rightarrow \infty} \tau_r$  is the point in time by which the chain has left each of the truncations in the sequence  $\{\mathcal{S}_r\}_{r \in \mathbb{N}}$ . The above tells us that  $\lim_{r \rightarrow \infty} \tau_r$  is (almost surely) equal to  $T_\infty$  regardless of the particular sequence of

finite truncations  $\{\mathcal{S}_r\}_{r \in \mathbb{N}}$  in its definition. For this reason, we interpret  $T_\infty$  as the point in time that the chain leaves the state space, or, in other words, *explodes*.

We now give two technical lemmas we will use throughout the paper. The first delineates the simple relationship between the exit time

$$(2.4) \quad \sigma := \inf\{n \in \mathbb{N} : Y_n \notin \mathcal{D}\}$$

of the jump chain  $Y$  and the exit time  $\tau$  of  $X$  (defined in (1.1)).

LEMMA 2.2 (Lemma 2.27 of [35]). *If  $\tau$  and  $\sigma$  are as in (1.1) and (2.4), then*

$$\tau(\omega) = \begin{cases} T_{\sigma(\omega)}(\omega) & \text{if } \sigma(\omega) < \infty, \\ \infty & \text{if } \sigma(\omega) = \infty \end{cases} \quad \forall \omega \in \Omega.$$

The other lemma allows us to build *auxiliary chains* that will be key in the proofs in this paper.

LEMMA 2.3. *Suppose that a second rate matrix  $\bar{Q}$  coincides with  $Q$  on  $\mathcal{D}$ :*

$$q(x, y) = \bar{q}(x, y) \quad \forall x \in \mathcal{D}, \quad y \in \mathcal{S}.$$

*There exists a chain  $\bar{X} := \{\bar{X}_t\}_{t \geq 0}$  also defined on  $(\Omega, \mathcal{F}, \mathbb{P})$  with rate matrix  $\bar{Q}$ , jump times  $\{\bar{T}_n\}_{n \in \mathbb{N}}$ , jump chain  $\bar{Y} := \{\bar{Y}_n\}_{n \in \mathbb{N}}$ , explosion time  $\bar{T}_\infty$ , and exit times*

$$\bar{\sigma} := \inf\{n \in \mathbb{N} : \bar{Y}_n \notin \mathcal{D}\}, \quad \bar{\tau} := \inf\{t \in [0, \bar{T}_\infty) : \bar{X}_t \notin \mathcal{D}\},$$

*such that  $X$  and  $\bar{X}$  exit the domain at the same time,*

$$(2.5) \quad \sigma(\omega) = \bar{\sigma}(\omega), \quad \tau(\omega) = \bar{\tau}(\omega), \quad \tau(\omega) \wedge T_\infty(\omega) = \bar{\tau}(\omega) \wedge \bar{T}_\infty(\omega) \quad \forall \omega \in \Omega,$$

*and such that  $X$  and  $\bar{X}$  are identical up to (and including) this instant:*

$$(2.6) \quad X_t(\omega) = \bar{X}_t(\omega) \quad \forall (t, \omega) \in [0, \infty) \times \Omega : t \leq \tau(\omega) \wedge T_\infty(\omega).$$

*In particular, until the moment of exit, the jump chain and jump times of both chains are identical:*

$$(2.7) \quad Y_n(\omega) = \bar{Y}_n(\omega), \quad T_n(\omega) = \bar{T}_n(\omega) \quad \forall (n, \omega) \in \mathbb{N} \times \Omega : n \leq \sigma(\omega).$$

*Proof.* See Appendix A in the supplementary material.  $\square$

**2.1. The time-varying law of the chain and the FSP scheme.** The time-varying law of the chain

$$(2.8) \quad p_t(x) := \mathbb{P}(\{X_t = x, t < T_\infty\}) \quad \forall x \in \mathcal{S}$$

satisfies  $|\mathcal{S}|$  linear ODEs known as *Kolmogorov's forward equations* (or *the chemical master equation* or, simply, *the forward equations*).

THEOREM 2.4 (Kolmogorov's forward equations; Corollary 2.21 of [35]). *Suppose that the diagonal of the rate matrix is  $\gamma$ -integrable:*

$$(2.9) \quad \mathbb{E}[|q(X_0, X_0)|] = \sum_{x \in \mathcal{S}} \gamma(x) |q(x, x)| < \infty.$$

*For each  $x \in \mathcal{S}$ ,  $t \mapsto p_t(x)$  is a continuously differentiable function on  $[0, \infty)$ . Furthermore, the time-varying law  $p_t := \{p_t(x)\}_{x \in \mathcal{S}}$  is the minimal nonnegative solution of the equations*

$$(2.10) \quad \dot{p}_t(x) = \sum_{y \in \mathcal{S}} p_t(y) q(y, x), \quad p_0(x) = \gamma(x) \quad \forall x \in \mathcal{S}, \quad t \in [0, \infty).$$

In the above, by “minimal nonnegative solution” we mean that if  $k_t$  is any other nonnegative ( $k_t(x) \geq 0$  for each  $x \in \mathcal{S}$  and  $t \in [0, \infty)$ ) differentiable function satisfying (2.10), then  $k_t(x) \geq p_t(x)$  for each  $x \in \mathcal{S}$  and  $t \geq 0$  (if the chain is explosive, then the equations can have multiple solutions; see [9, 22, 50]). Except for a few special cases, no analytical expressions for this minimal solution are known. If  $\mathcal{S}$  is infinite, or finite but large, direct numerical computation of this solution is not possible either. Instead, we can use the popular finite state projection (FSP) algorithm [45]: a numerical scheme that yields a set of lower bounds  $p_t^r := \{p_t^r(x)\}_{x \in \mathcal{S}}$  on the chain’s time-varying law  $p_t := \{p_t(x)\}_{x \in \mathcal{S}}$ . We identify these bounds with the measure on  $(\mathcal{S}, 2^\mathcal{S})$  defined by  $p_t^r(A) = \sum_{x \in A} p_t^r(x) \forall A \subseteq \mathcal{S}$ , where  $2^\mathcal{S}$  denotes the power set of  $\mathcal{S}$ . The FSP scheme consists of choosing a (finite) truncation  $\mathcal{S}_r$  of the state space  $\mathcal{S}$ , solving numerically the set of  $|\mathcal{S}_r|$  linear ODEs

$$(2.11) \quad p_t^r(x) = \sum_{y \in \mathcal{S}_r} p_t^r(y) q(y, x), \quad p_0^r(x) = \gamma(x) \quad \forall x \in \mathcal{S}_r,$$

over the time interval  $[0, t]$ , and padding  $p_t^r$  with zeros:  $p_t^r(x) := 0 \forall x \notin \mathcal{S}_r$ .

We collect various useful properties of the FSP scheme in Theorem 2.5 below. Most of these properties can be found elsewhere: (i) and (v) are shown in Proposition 2.14 of [1] (however, there is a small mistake therein; see [8]); (iv) and the bound in (iii) are proven in [45]. Although (ii) is mentioned in [16, 46], we have not encountered a proof elsewhere. Similarly, the explicit expression of the error (i.e., the total variation distance between  $p_t$  and its approximation) in (iii) and the necessary and sufficient condition for the bound to be sharp appear to be new.

**THEOREM 2.5** (finite state projection scheme). *Let  $\{\mathcal{S}_r\}_{r \in \mathbb{N}}$  be an increasing sequence of finite sets contained in  $\mathcal{S}$ ,  $\tau_r$  the exit time from the truncation  $\mathcal{S}_r$ , and  $\{p_t^r\}_{r \in \mathbb{N}}$  the sequence of FSP approximations defined by (2.11). Then the following properties hold:*

(i) *(Increasing sequence of lower bounds.)*

$$p_t^0(x) \leq p_t^1(x) \leq \cdots \leq p_t(x) \quad \forall x \in \mathcal{S}, \quad t \geq 0.$$

(ii) *(Mass of the approximation.) The mass of the approximation is the probability that the chain has not yet exited the truncation:*

$$p_t^r(\mathcal{S}) = p_t^r(\mathcal{S}_r) = \mathbb{P}(\{t < \tau_r\}) \quad \forall t \geq 0.$$

(iii) *(Computable error bound.) For any  $r \in \mathbb{N}$ ,*

$$\|p_t - p_t^r\| = \mathbb{P}(\{t < T_\infty\}) - \mathbb{P}(\{t < \tau_r\}) \leq 1 - p_t^r(\mathcal{S}_r) \quad \forall t \geq 0,$$

*and equality holds if and only if  $\mathbb{P}(\{T_\infty = \infty\}) = 1$ , i.e., when the chain is nonexplosive.*

(iv) *(Monotonicity of the error and of the error bound.) The approximation error  $\|p_t - p_t^r\|$  and its upper bound  $1 - p_t^r(\mathcal{S}_r)$  are decreasing in  $r$ ,*

$$(2.12) \quad \|p_t - p_t^r\| \leq \|p_t - p_t^s\|, \quad 1 - p_t^r(\mathcal{S}_r) \leq 1 - p_t^s(\mathcal{S}_s) \quad \forall s \leq r, \forall t \geq 0,$$

*and increasing in  $t$ ,*

$$(2.13) \quad \|p_t - p_t^r\| \geq \|p_u - p_u^r\|, \quad 1 - p_t^r(\mathcal{S}_r) \geq 1 - p_u^r(\mathcal{S}_r) \quad \forall u \leq t, \forall r \in \mathbb{N}.$$

Consequently, the FSP scheme returns not only the approximation  $p_t^r$  of  $p_t$ , but also an approximation  $p_s^r$  of  $p_s$  for each  $s \leq t$  with an error that is bounded uniformly in  $s$ :

$$\sup_{s \in [0, t]} \|p_s - p_s^r\| = \|p_t - p_t^r\| \leq 1 - p_t^r(\mathcal{S}_r) \quad \forall t \geq 0.$$

(v) (*Convergence of bounds.*) If  $\cup_{r \in \mathbb{N}} \mathcal{S}_r = \mathcal{S}$ , then the scheme converges:

$$\lim_{r \rightarrow \infty} \|p_t - p_t^r\| = 0 \quad \forall t \geq 0.$$

*Proof.* See Appendix B in the supplementary material.  $\square$

The ideas behind Theorem 2.5 emerge from the following construction. Consider a second chain  $X^r$  which is identical to  $X$  except that every state outside of the truncation  $\mathcal{S}_r$  is turned into an absorbing state. In particular, let  $X^r$  be the chain of Lemma 2.3 with  $\mathcal{S}_r$  replacing  $\mathcal{D}$ , and  $Q^r := (q^r(x, y))_{x, y \in \mathcal{S}}$  replacing  $\bar{Q}$ , where

$$(2.14) \quad q^r(x, y) := \begin{cases} q(x, y) & \text{if } x \in \mathcal{S}_r, \\ 0 & \text{if } x \notin \mathcal{S}_r. \end{cases}$$

Lemma 2.3 states that the chains  $X$  and  $X^r$  coincide until (and including) the time  $\tau_r$  at which they simultaneously leave the truncation  $\mathcal{S}_r$  for the first time, at which point  $X^r$  becomes trapped in a state outside of the truncation and never returns to  $\mathcal{S}_r$ . In contrast,  $X$  may return to the truncation, and hence the probability  $p_t(x)$  that  $X$  is at any given state  $x$  inside the truncation at time  $t$  is greater than or equal to the probability that  $X^r$  is in the same state at the same time. Since the law of  $X_t^r$  (restricted to  $\mathcal{S}_r$ ) is the solution of (2.11) (Theorem 2.4), we arrive at Theorem 2.5(i).

The probability  $p_t^r(\mathcal{S}_r)$  that  $X^r$  is inside the truncation at time  $t$  is the same as the probability  $\mathbb{P}(\{t < \tau_r\})$  that it has not yet left. Theorem 2.5(ii)–(iii) follows from this fact. If  $X^r$  has not left the truncation by time  $t$ , then it has not left the *larger* truncation  $\mathcal{S}_{r+1}$  by  $t$ . Similarly, if the chain has not left  $\mathcal{S}_r$  by time  $t$ , it has not left by any earlier time  $s \leq t$ . For these reasons, Theorem 2.5(iv) holds.

Due to (iii), proving the convergence of the scheme consists of showing that  $\mathbb{P}(\{t < \tau_r\})$  converges to  $\mathbb{P}(\{t < T_\infty\})$  as  $r$  tends to infinity. Recall that  $\mathbb{P}(\{t < T_\infty\})$  is the probability that the chain has not left the state space by time  $t$ , while  $\mathbb{P}(\{t < \tau_r\})$  is the probability that the chain has not left the truncation  $\mathcal{S}_r$  by time  $t$ . Because the truncations  $\mathcal{S}_r$  approach the complete state space as  $r$  tends to infinity, it must be the case that  $\mathbb{P}(\{t < \tau_r\})$  approaches  $\mathbb{P}(\{t < T_\infty\})$  or, equivalently, that the scheme converges as stated in Theorem 2.5(v).

The FSP *algorithm* as proposed in [45] consists of repeatedly computing  $p_t^r$  while increasing the size of the truncation until the error bound  $1 - p_t^r(\mathcal{S}_r)$  is smaller than some prescribed tolerance. As noted in [37], the algorithm may not terminate, even if the truncations tend to the state space as  $r$  tends to infinity. Theorem 2.5(v) clarifies this issue. Although the *scheme* converges (i.e.,  $p_t^r$  tends to  $p_t$  in total variation as  $r$  tends to infinity or, equivalently,  $p_t^r(\mathcal{S}_r)$  tends to  $\mathbb{P}(\{t < T_\infty\})$ ), this does not imply that the error bound  $1 - p_t^r(\mathcal{S}_r)$  converges to zero. This is only the case if the chain is nonexplosive (i.e.,  $\mathbb{P}(\{T_\infty < \infty\}) = 0$ ). Otherwise,  $\mathbb{P}(\{t < T_\infty\}) > 0 \ \forall t > 0$  (see the proof of Theorem 2.5(iii)) and the algorithm will not terminate if the tolerance is set to be smaller than  $1 - \mathbb{P}(\{t < T_\infty\})$ . In practice, nonexplosivity can be established using a Foster–Lyapunov criterion [7, 42].

We close this section by pointing out that the FSP scheme can also be used to compute converging approximations of the occupation measure associated with a

deterministic time  $t$ , which tells us how long the chain has spent in state  $x$  by time  $t$  (see [35, Corollary 3.2] for details).

**2.2. The exit time and its associated exit distribution and occupation measure.** Let  $\mu$ ,  $\nu$ ,  $\mu^r$ , and  $\nu^r$  be defined as in (1.2)–(1.3) and (1.9)–(1.10). Our convention of  $\inf \emptyset = \infty$  and the exit time's definition in (1.1) implies that it is finite if and only if it is strictly less than the explosion time:  $\tau(\omega) < \infty \iff \tau(\omega) < T_\infty(\omega)$  for any  $\omega \in \Omega$ . Therefore,  $X_\tau$  is defined on  $\{\tau < \infty\}$  and  $\mu$  is well-defined. Technically,  $\mu$ ,  $\nu$ ,  $\mu^r$ , and  $\nu^r$  are unsigned measures on  $([0, \infty) \times \mathcal{S}, \mathcal{X})$ , where  $\mathcal{X}$  is the product sigma algebra of  $2^\mathcal{S}$  and the Borel sigma  $\mathcal{B}([0, \infty))$  on  $[0, \infty)$ . When using (1.2)–(1.3) and (1.9)–(1.10) to define these four measures, we exploit the fact that  $\{[a, b) \times \{x\} : 0 \leq a < b < \infty, x \in \mathcal{S}\}$  is a  $\pi$ -system that generates  $\mathcal{X}$ .

From the definition (1.2) of the exit distribution  $\mu$ , it follows that its mass is the probability that the chain eventually leaves the domain:

$$(2.15) \quad \mu([0, \infty), \mathcal{S}) = \mathbb{P}(\{\tau < \infty, X_\tau \in \mathcal{S}\}) = \mathbb{P}(\{\tau < \infty\}).$$

Similarly, it follows from (1.3) that the mass of the occupation measure is

$$(2.16) \quad \nu([0, \infty), \mathcal{S}) = \mathbb{E} \left[ \int_0^{\tau \wedge T_\infty} \left( \sum_{x \in \mathcal{S}} 1_x(X_t) \right) dt \right] = \mathbb{E}[\tau \wedge T_\infty].$$

If the chain is nonexplosive (i.e.,  $\mathbb{P}(\{T_\infty = \infty\}) = 1$ ), the mass is the mean exit time. For explosive chains (i.e.,  $\mathbb{P}(\{T_\infty = \infty\}) < 1$ ), the same holds as long as the chain cannot explode without first exiting the domain (i.e.,  $\mathbb{P}(\{\tau \leq T_\infty\}) = 1$ ).

In (1.2)–(1.3), we defined the exit distribution  $\mu$  and occupation measure  $\nu$  probabilistically in terms of the chain  $X$ . These measures are characterized analytically in terms of the solutions of the ODEs (2.18) in the following theorem.

**THEOREM 2.6** (analytical characterization of  $\mu$  and  $\nu$ ). *Suppose that (1.13) holds. The exit distribution  $\mu$  and occupation measure  $\nu$  decompose as in (1.4)–(1.5), and their densities  $\nu(t, x)$  and  $\mu(t, x)$  are nonnegative and continuous functions on  $[0, \infty)$  for each  $x \in \mathcal{S}$ . Moreover,*

$$(2.17) \quad \mu(t, x) = 1_{\mathcal{D}^c}(x) \dot{\hat{p}}_t(x), \quad \nu(t, x) = 1_{\mathcal{D}}(x) \hat{p}_t(x) \quad \forall x \in \mathcal{S}, \quad t \in [0, \infty),$$

where  $\hat{p}_t$  is the minimal nonnegative solution (as in Theorem 2.4) of

$$(2.18) \quad \dot{\hat{p}}_t(x) = \sum_{y \in \mathcal{D}} \hat{p}_t(y) q(y, x), \quad \hat{p}_0(x) = \gamma(x) \quad \forall x \in \mathcal{S}, \quad t \in [0, \infty).$$

*Proof.* See Appendix C in the supplementary material.  $\square$

The ideas behind the above theorem are similar to those behind Theorem 2.5. In particular, we consider a second chain  $\hat{X}$  identical to  $X$  except that every state outside of the domain  $\mathcal{D}$  is turned into an absorbing state. That is, let  $\hat{X}$  be the chain of Lemma 2.3 after replacing  $\bar{Q}$  with  $\hat{Q} := (\hat{q}(x, y))_{x, y \in \mathcal{S}}$ , where

$$(2.19) \quad \hat{q}(x, y) := \begin{cases} q(x, y) & \text{if } x \in \mathcal{D}, \\ 0 & \text{if } x \notin \mathcal{D}. \end{cases}$$

The chains  $X$  and  $\hat{X}$  are identical up until (and including) the time at which they both simultaneously exit the domain via the same state. Therefore the probability

$\mu([0, t], x)$  that  $X$  has exited the domain by time  $t$  via state  $x \in \mathcal{D}^c$  is also the probability that  $\hat{X}$  exited via  $x$  by time  $t$ . Because  $\hat{X}$  is trapped in the first state that it enters once leaving the truncation, it follows that  $\mu([0, t], x)$  is the probability that  $\hat{X}$  is in state  $x$  by time  $t$ . The characterization of the exit distribution then follows from Theorem 2.4. The characterization of the occupation measure follows similarly. The key observation is that once  $\hat{X}$  leaves the domain it cannot return, and hence the amount of time that  $\hat{X}$  spends in a state  $x \in \mathcal{D}$  until the moment it exits the domain is the *total* time it will spend in that state.

**The marginals.** In applications, we are often interested in the distribution of the exit time itself, that is, the *time marginal* of the exit distribution

$$(2.20) \quad \mu_T(B) := \mu(B, \mathcal{S}) = \mathbb{P}(\{\tau \in B, X_\tau \in \mathcal{S}\}) = \mathbb{P}(\{\tau \in B\}) \quad \forall B \in \mathcal{B}([0, \infty)).$$

Technically, the above is the distribution of  $\tau$  restricted to  $[0, \infty)$ . However, we recover the complete distribution from  $\mu_T$  as  $\mathbb{P}(\{\tau = \infty\}) = 1 - \mathbb{P}(\{\tau < \infty\}) = 1 - \mu_T([0, \infty))$ . Equations (2.17)–(2.18) imply that  $\mu(t, x)$  is nonnegative, and so combining (1.4) and Tonelli's theorem shows that the time marginal  $\mu_T(dt)$  of the exit distribution also has a density  $\mu_T(t)$  with respect to the Lebesgue measure and that this density is given by  $\mu_T(t) = \sum_{x \in \mathcal{S}} \mu(t, x)$ :

$$(2.21) \quad \mu_T(B) = \gamma(\mathcal{D}^c)\delta_0(B) + \int_B \mu_T(t)dt \quad \forall B \in \mathcal{B}([0, \infty)), \quad x \in \mathcal{S},$$

where  $\delta_0$  denotes the Dirac measure at zero ( $\delta_0(B) = 1$  if  $0 \in B$  and 0 otherwise).

In other cases, we are interested in where on the boundary the exit occurs or where in the domain the chain spends time up until exiting. The *space marginals* of the exit distribution and the occupation measure provide this information:

$$(2.22) \quad \mu_S(x) := \mu([0, \infty), x) = \mathbb{P}(\{X_\tau = x, \tau < \infty\}) \quad \forall x \in \mathcal{S},$$

$$(2.23) \quad \nu_S(x) := \nu([0, \infty), x) = \mathbb{E} \left[ \int_0^{\tau \wedge T_\infty} 1_x(X_t)dt \right] \quad \forall x \in \mathcal{S}.$$

Clearly, one can obtain explicit expressions for  $\mu_T, \mu_S$ , and  $\nu_S$  in terms of  $\hat{p}$ .

**3. Theoretical characterization of the ETFSP scheme: Proof of Theorem 1.1 and bounding the marginal distributions.** We now prove Theorem 1.1, which delineates the theoretical properties of the ETFSP scheme. Before delving into the proof, we discuss briefly some of the intuitive ideas underlying the proof for the exit distribution (the occupation measure is analogous).

Consider the auxiliary chain  $X^r$  introduced above, which is identical to the original chain except that each state outside of the truncation  $\mathcal{S}_r$  is turned into an absorbing state. Once  $X^r$  exits the truncation, it becomes trapped in whichever state it just entered. For this reason, if  $X^r$  has not exited the domain by the time it exits  $\mathcal{S}_r$ , then it will never exit. Theorem 2.6 tells us that  $\mu^r(dt, x)$  is the exit distribution  $\rho^r(dt, x)$  of  $X^r$  restricted to  $[0, t_f^r] \times \mathcal{S}_r$ . Thus, Theorem 1.1(ii) follows from the fact that  $X$  and  $X^r$  are identical up until, and including, the moment that they simultaneously exit the truncation (Lemma 2.3). In contrast with  $X^r$ , the original chain  $X$  may still exit the domain after it leaves the truncation because it does not necessarily get trapped in an absorbing state. During a small interval of time  $[t, t+h]$ , the probability of exiting the domain  $\mathcal{D}$  via state  $x$  is  $\mu(t, x)h$  for  $X$  and  $\rho^r(t, x)h$  for  $X^r$ . Given that,

for any interval size  $h$ , this probability cannot be greater for  $X^r$  than for  $X$ , the lower bound property in Theorem 1.1(i) follows from the continuity of  $\mu(\cdot, x)$  and  $\rho^r(\cdot, x)$  (Theorem 2.6). The remainder of the theorem then follows (i)–(ii) and the fact that  $\tau_r$  is an increasing sequence with limit  $T_\infty$  (Lemma 2.1).

*Proof of Theorem 1.1.* Let  $Y^r := \{Y_n^r\}_{n \in \mathbb{N}}$ ,  $\{T_n^r\}_{n \in \mathbb{N}}$ , and  $T_\infty^r$  be the jump chain, jump times, and explosion time of  $X^r$ .

(i) Theorem 2.6 tells us that  $\nu(t, x) = 1_{\mathcal{D}}(x)\hat{p}_t(x)$ , where  $\hat{p}_t$  is the minimal non-negative solution of (2.18). Theorem 2.4 tells us that  $\hat{p}_t$  is the time-varying law of the auxiliary chain  $\hat{X}$  with rate matrix  $\hat{Q}$  defined in (2.19). Applying the FSP scheme to  $\hat{X}$  instead of  $X$  entails solving

$$(3.1) \quad \dot{k}_t^r(x) = \sum_{y \in \mathcal{S}_r} k_t^r(y)\hat{q}(y, x) = \sum_{y \in \mathcal{D}_r} k_t^r(y)\hat{q}(y, x), \quad k_0^r(x) = \gamma(x) \quad \forall x \in \mathcal{S}_r$$

and setting  $k_t^r(x) = 0 \forall x \notin \mathcal{S}_r$ . Comparing (1.6) and (3.1) we can see that

$$(3.2) \quad \nu^r(t, x) = k_t^r(x) \quad \forall x \in \mathcal{D}_r, \quad t \leq t_f^r.$$

Given (1.8) and the fact that the final times  $t_f^r$  are increasing, the second set of inequalities then follows directly from (2.17) and Theorem 2.5(i). Similarly, (3.2), (1.7), and the finiteness of  $\mathcal{D}_r$  imply that

$$\mu^{r+1}(t, x) - \mu^r(t, x) = \sum_{z \in \mathcal{D}_r} (k_t^{r+1}(z) - k_t^r(z))q(z, x) \geq 0 \quad \forall x \in \mathcal{S}_r \cap \mathcal{D}^c,$$

where the inequality follows from Theorem 2.5(i) and the fact that  $x \neq z$  in the above sum so that  $q(z, x) \geq 0$ . Replacing  $\mu^{r+1}$  by  $\mu$  and  $k^{r+1}$  by  $\hat{p}$  in the above argument and applying (1.8) and (2.17) gives us the other set of inequalities.

(ii) Aside from having to use the fact that the explosion time of the chain  $X^r$  is almost surely infinite (see (B.3) in the supplementary material), the proof of the expression for the mass of  $\nu^r$  is analogous to that for  $\mu^r$ , and so we skip it. Applying Theorem 2.6 to  $X^r$  instead of  $X$  shows that  $\mu^r$ , restricted to  $[0, t_f^r] \times \mathcal{S}_r$ , coincides with the corresponding restriction of the density of the exit distribution associated with the first time that  $X^r$  exits the domain:

$$\tau_{\mathcal{D}}^r := \inf\{0 \leq t < T_\infty^r : X_t^r \notin \mathcal{D}\}.$$

Thus, (1.8) and the definition of the exit distribution (1.2) imply that the mass of  $\mu^r$  is the probability that  $X^r$  exits the domain no later than the final time  $t_f^r$  and via a state inside the truncation:

$$\mu^r([0, \infty), \mathcal{S}) = \mu^r([0, t_f^r], \mathcal{S}) = \mathbb{P}_\lambda(\{\tau_{\mathcal{D}}^r \leq t_f^r, X_{\tau_{\mathcal{D}}^r}^r \in \mathcal{S}_r\}).$$

As we now show, this probability is the same as the probability that the original chain exited the domain no later than the truncation and the final time. The key observation is that whenever  $X^r$  leaves the truncation, it becomes trapped in whichever state it just entered. This implies that if  $X^r$  has not left the domain by the time it exits the truncation, then it never will. Formally, it follows from Lemma 2.2 and (B.2) in the supplementary material that

$$\{\tau_r < \tau_{\mathcal{D}}^r \leq t_f^r\} = \{X_{\tau_{\mathcal{D}}^r}^r = X_{\tau_r}^r, \tau_r < \tau_{\mathcal{D}}^r \leq t_f^r\}$$

(recall that Lemma 2.3 implies that  $X^r$  and  $X$  exit the truncation at the same time  $\tau_r$ ). However, the latter set must be the empty set since  $X_{\tau_r(\omega)}^r(\omega)$  (resp.,  $X_{\tau_D^r(\omega)}(\omega)$ ) lies inside (resp., outside) of the domain in order for  $X^r$  to exit the truncation before it exits the domain ( $\tau_r(\omega) < \tau_D^r(\omega)$ ). Thus,

$$\{\tau_D^r \leq t_f^r, X_{\tau_D^r}^r \in \mathcal{S}_r\} = \{\tau_D^r \leq t_f^r, \tau_D^r \leq \tau_r, X_{\tau_D^r}^r \in \mathcal{S}_r\} = \{\tau_D^r \leq t_f^r \wedge \tau_r\}.$$

Since (2.7) implies  $\{\tau_D^r \leq t_f^r \wedge \tau_r\} = \{\tau \leq t_f^r \wedge \tau_r\}$ , the result follows.

(iii) As (i) shows that  $\mu - \mu^r$  and  $\nu - \nu^r$  are unsigned measures, (1.14)–(1.15) follow from the fact that the total variation norm of a unsigned measure is its mass, (ii), the definitions (1.6)–(1.8) of  $\mu^r$  and  $\nu^r$ , and the expression for the masses of  $\mu$  and  $\nu$  in (2.15)–(2.16).

(iv) This follows directly from (i) and (1.14)–(1.15).

(v) Because Lemmas 2.1 and 2.2 imply that

$$\lim_{r \rightarrow \infty} 1_{\{\tau \leq t_f^r \wedge \tau_r\}} = 1_{\{\tau < \infty\}}, \quad \lim_{r \rightarrow \infty} \tau \wedge t_f^r \wedge \tau_r = \tau \wedge T_\infty \quad \text{almost surely,}$$

the convergence follows from the monotone convergence theorem and (1.14)–(1.15).  $\square$

**Bounding the marginals.** Using the ETFSP scheme we also obtain converging approximations of the marginals  $\mu_T$ ,  $\mu_S$ , and  $\nu_S$  of the exit distribution and occupation measure (see (2.20)–(2.23)). In particular, marginalizing (1.9)–(1.10), we obtain approximations of  $\mu_T$ ,  $\mu_S$ , and  $\nu_S$ :

$$(3.3) \quad \mu_T^r(B) := \mu^r(B, \mathcal{S}) = \gamma(\mathcal{D}^c \cap \mathcal{S}_r) \delta_0(B) + \int_B \mu_T^r(t) dt \quad \forall B \in \mathcal{B}([0, \infty)),$$

$$(3.4) \quad \mu_S^r(x) := \mu^r([0, \infty), x), \quad \nu_S^r(x) := \nu^r([0, \infty), x) \quad \forall x \in \mathcal{S},$$

where  $\mu_T^r(t) := \sum_{x \in \mathcal{S}_r} \mu^r(t, x)$ . The fact that  $\mu^r(dt, x)$  and  $\nu^r(dt, x)$  bound from below the exit distribution and occupation measure (Theorem 1.1(i)) implies that the marginals of the approximations  $\mu_T^r(dt)$ ,  $\mu_S^r(x)$ , and  $\nu_S^r(x)$  bound  $\mu_T(dt)$ ,  $\mu_S(x)$ , and  $\nu_S(x)$  from below. For this reason, the fact that the total variation norm of an unsigned measure is its mass implies that

$$(3.5) \quad \|\mu_T - \mu_T^r\| = \|\mu_S - \mu_S^r\| = \|\mu - \mu^r\|, \quad \|\nu_S - \nu_S^r\| = \|\nu - \nu^r\|.$$

In other words, the errors of the marginal approximations are the same as those of the complete approximations. In full, we have the following corollary of Theorem 1.1.

**COROLLARY 3.1.** *Suppose that the premise of Theorem 1.1 is satisfied. Consider the approximations of the marginals  $\mu_T^r$ ,  $\mu_S^r$ , and  $\nu_S^r$  defined in (3.3)–(3.4).*

(i) *(Increasing sequence of lower bounds.) The approximations form an increasing sequence of lower bounds:*

$$\begin{aligned} \mu_T^0(t) &\leq \mu_T^1(t) \leq \cdots \leq \mu_T(t) & \forall t \in [0, \infty), \\ \mu_S^0(x) &\leq \mu_S^1(x) \leq \cdots \leq \mu_S(x), \quad \nu_S^0(x) \leq \nu_S^1(x) \leq \cdots \leq \nu_S(x) & \forall x \in \mathcal{S}. \end{aligned}$$

(ii) *(Computable error bounds and monotonicity properties.) The equalities and inequalities in (1.14)–(1.17) hold identically if we replace  $\mu, \mu^r$  with  $\mu_T, \mu_T^r$  (or  $\mu_S, \mu_S^r$ ) and  $\nu, \nu^r$  with  $\nu_S, \nu_S^r$ .*

(iii) *(Convergence of bounds.) Suppose that  $\cup_r \mathcal{S}_r = \mathcal{S}$  and that  $t_f^r \rightarrow \infty$  as  $r \rightarrow \infty$ . The approximations of the marginals of the exit distribution converge:*

$$\lim_{r \rightarrow \infty} \|\mu_T - \mu_T^r\| = \lim_{r \rightarrow \infty} \|\mu_S - \mu_S^r\| = 0.$$

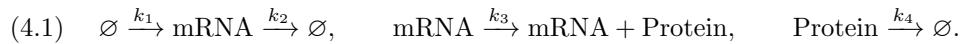
Furthermore, if  $\mathbb{E}[\tau \wedge T_\infty] < \infty$ , then the approximation of the space marginal of the occupation measure converges:

$$\lim_{r \rightarrow \infty} \|\nu_S - \nu_S^r\| = 0.$$

*Proof.* Given (3.5), the corollary follows immediately from Theorem 1.1.  $\square$

**4. Applications.** In this section, we apply the ETFSP scheme to two biological examples from the literature. To simplify the exposition, we assume without loss of generality that the chain starts inside the domain:  $\mathbb{P}(\{X_0 \in \mathcal{D}\}) = \gamma(\mathcal{D}) = 1$ .

**4.1. Threshold model for stochastic gene expression.** Proteins perform essential functions inside living cells. These molecules are expressed from genes through a series of biochemical reactions, and their absolute levels (and the timings in which these are reached) are critical to cell decisions, such as differentiation [12] or lysis in the bacteriophage  $\lambda$  [53]. Let us consider a simple model of gene expression involving the transcription and degradation of mRNA molecules (with rates  $k_1$  and  $k_2$ , respectively), the synthesis of a protein from each mRNA molecule (with rate  $k_3$ ), and the degradation of proteins (with rate  $k_4$ ):



The state of the system is  $x = (m, p)$ , where  $m$  is the number of mRNAs and  $p$  is the number of proteins; hence the state space is  $\mathcal{S} = \mathbb{N}^2$ . The reactions obey mass-action kinetics, and the rate matrix is given by

$$q((m_1, p_1), (m_2, p_2)) = \begin{cases} -k_1 - k_2 m_1 - k_3 m_1 - k_4 p_1 & \text{if } (m_2, p_2) = (m_1, p_1), \\ k_1 & \text{if } (m_2, p_2) = (m_1 + 1, p_1), \\ k_2 m_1 & \text{if } (m_2, p_2) = (m_1 - 1, p_1), \\ k_3 m_1 & \text{if } (m_2, p_2) = (m_1, p_1 + 1), \\ k_4 p_1 & \text{if } (m_2, p_2) = (m_1, p_1 - 1), \\ 0 & \text{otherwise.} \end{cases}$$

We are interested in characterizing the time taken for the protein number  $p$  to attain a critical level  $p_c$ . To this end, we consider the domain

$$\mathcal{D} := \{(m, p) \in \mathbb{N}^2 : p < p_c\},$$

so that the exit time  $\tau$  from the domain  $\mathcal{D}$  corresponds to the first instant at which  $p_c$  proteins accumulate. We compute the lower bounds  $\mu^r(t, (m, p_c))$  of the exit distribution  $\mu(t, (m, p_c))$ , the joint distribution of the exit time, and the number of mRNAs present at exit. For ease of notation, in the rest of this section, we omit the protein number argument (as it is  $p_c$  at time  $\tau$ ), and we write  $(m, p_c)$  as  $m$ .

We use the truncations

$$\mathcal{S}_r = \{(m, p) \in \mathbb{N}^2 : p \leq p_c, m < r\} \quad \forall r \in \mathbb{N},$$

shown in Figure 4.1(a). Figure 4.1(b) shows how the error bound  $\varepsilon_r$  decreases to zero with  $r$  and  $t_f^r$ , whereas Figure 4.1(c) shows the lower bounds  $\mu^r(t, m)$  for various values of the truncation parameter  $r$  and  $t_f^r = 30$  (with  $\varepsilon_r < 10^{-4}$  for  $r = 16$ ).

The exit time correlates negatively with the level of mRNA: the more mRNA molecules present, the higher the expression, and the quicker the protein number

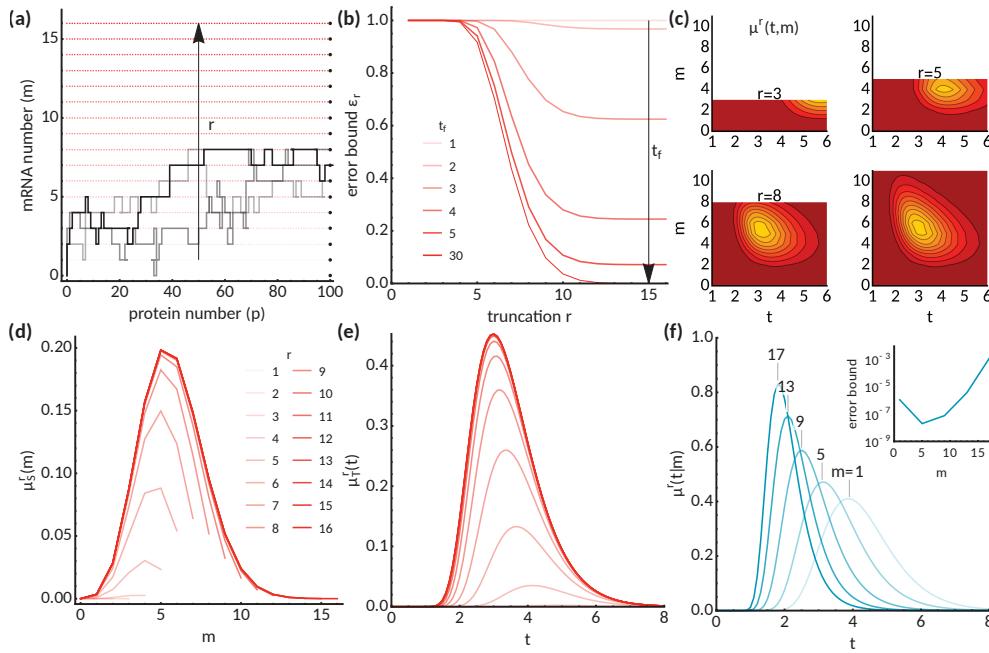


FIG. 4.1. Threshold statistics for stochastic gene expression. (a) Three representative sample paths (light gray, gray, and black lines) of the gene expression model (4.1) exiting at  $p_c = 100$  protein molecules. The red dashed lines indicate the boundaries of the truncations  $S_r$  with increasing  $r$ . (b) The error bound  $\varepsilon_r$  decreases with increasing truncation parameter  $r$  and final computation time  $t_f$ . (c) Lower bounds on the exit distribution for the truncations  $r = 3, 5, 8, 16$  (yellow indicates maximum probability panelwise). (d) The lower bounds on the mRNA-marginal exit distribution are monotonically increasing with  $r$  and become visually indistinguishable for  $r > 12$ . (e) Corresponding bounds on the exit time density. (f) Bounds on the conditional exit time distributions  $\mu_T(t|m)$  with  $r = 20$ ; inset shows error bounds (4.3). Parameters:  $k_1 = 5$ ,  $k_2 = 1$ ,  $k_3 = 10$ ,  $k_4 = 0.1$  and initial condition  $\gamma(x) = \mathbf{1}_0(m)\mathbf{1}_0(p)$  for all panels.

rises. Figures 4.1(d) and (e) show the corresponding lower bounds  $\mu_S^r(m)$  and  $\mu_T^r(t)$  on the space and time marginals  $\mu_S(m)$  and  $\mu_T(t)$ , respectively.

To gain a quantitative understanding of the anticorrelation between the exit time and mRNA numbers, we also compute the density of the exit time conditioned on the number of mRNA present:

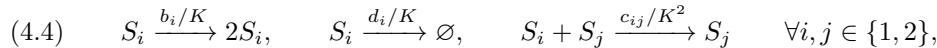
$$(4.2) \quad \mu(t|m) := \frac{\mu(t,m)}{\mu_S(m)} \geq \frac{\mu^r(t,m)}{\mu_S^r(m) + \varepsilon_r} =: \mu^r(t|m).$$

The bound on the right-hand side follows from the fact that  $\mu_S(m)$  is no greater than  $\mu_S^r(m) + \varepsilon_r$  due to the definition of the total variation norm. Integrating both sides of (4.2), we obtain the following bound on the total variation distance between the conditional density and its approximation:

$$(4.3) \quad \|\mu(\cdot|m) - \mu^r(\cdot|m)\| \leq \frac{\varepsilon_r}{\mu_S^r(m) + \varepsilon_r}.$$

Figure 4.1(f) shows this density computed using  $r = 20$  for various values of  $m$ . As expected, the mode of the distribution decreases with increasing mRNA number, but, interestingly, the density also narrows with increasing  $m$ . In the inset of Figure 4.1(f) we verify that the approximation error is small for each  $m$ .

**4.2. Fixation statistics in population dynamics.** The ETFSP framework can be used to provide insights into the fixation (or extinction) statistics of competing populations with small numbers. Common models in ecology and evolution are of the Lotka–Volterra type [10]. Let us consider the population dynamics of two competing species  $S_1$  and  $S_2$ :



with state space  $\mathcal{S} := \mathbb{N}^2$ . The first and second reactions describe the birth and death of individuals with rates  $b_i > 0$  and  $d_i > 0$ , respectively. The third reaction describes intra- and interspecies competition of strength  $c_{ij} > 0$ . The parameter  $K > 0$  is the effective carrying capacity. Let us denote the numbers of individuals by  $x = (x_1, x_2)$ . The rate matrix is given by

$$q(x, y) = \begin{cases} -\sum_{i=1}^2 (w_i^+(x) + w_i^-(x) + w_{i1}(x) + w_{i2}) & \text{if } y = x, \\ w_i^+(x) & \text{if } y = x + e_i \quad \forall i \in \{1, 2\}, \\ w_i^-(x) + w_{i1}(x) + w_{i2} & \text{if } y = x - e_i \quad \forall i \in \{1, 2\}, \\ 0 & \text{otherwise,} \end{cases}$$

where  $e_1 := (1, 0)$ ,  $e_2 := (0, 1)$  and

$$w_i^+(x) := b_i \frac{x_i}{K}, \quad w_i^-(x) := d_i \frac{x_i}{K}, \quad w_{ij}(x) := c_{ij} \frac{x_i x_j}{K^2} \quad \forall i, j \in \{1, 2\}.$$

For simplicity, we fix  $c_{11} = c_{12} = c_{21} = c_{22} = 1$  and  $K = 30$ .

*Deterministic dynamics.* The deterministic dynamics of the populations is modeled with the set of ODEs

$$(4.5) \quad \dot{x}_i = w_i^+(x) - w_i^-(x) - w_{i1}(x) - w_{i2}(x) \quad \forall i = 1, 2.$$

The equilibrium  $(x_1, x_2) = (0, 0)$  representing the extinction of both populations is unstable. Fixation of  $S_1$  occurs when  $S_2$  goes extinct, and vice versa, i.e., when the dynamics approaches one of the two axes. Which of the two species becomes extinct depends on the growth rate difference:

$$(4.6) \quad \Delta\lambda := \lambda_1 - \lambda_2, \quad \text{where } \lambda_i = b_i - d_i \quad \forall i, j \in \{1, 2\}.$$

For  $\Delta\lambda > 0$  (Figure 4.2(a)), there is an unstable fixed point on the  $x_1 = 0$  axis and a stable one on the  $x_2 = 0$  axis; hence the trajectories approach the stable fixed point leading to fixation of  $S_1$ . For  $\Delta\lambda < 0$ , the situation is reversed, resulting in the fixation of  $S_2$  (Figure 4.2(b)). For equal growth rates ( $\Delta\lambda = 0$ ), the dynamics approaches an invariant manifold (a line of fixed points) on which the two species coexist with ratios depending on their initial populations (Figure 4.2(c)).

*Computation of fixation probabilities and times.* In the stochastic setting, both species  $S_1$  and  $S_2$  have nonzero probability of becoming fixed regardless of the value of  $\Delta\lambda$ . To study this phenomenon, we consider the exit time from the domain

$$\mathcal{D} = \{(x_1, x_2) \in \mathbb{N}^2 \mid x_1 > 0, x_2 > 0\},$$

with complement  $\mathcal{D}^c$  that can be decomposed into the disjoint subsets

$$\mathcal{D}_1^c := \{(x_1, 0) \in \mathbb{N}^2 : x_1 > 0\}, \quad \mathcal{D}_2^c := \{(0, x_2) \in \mathbb{N}^2 : x_2 > 0\},$$

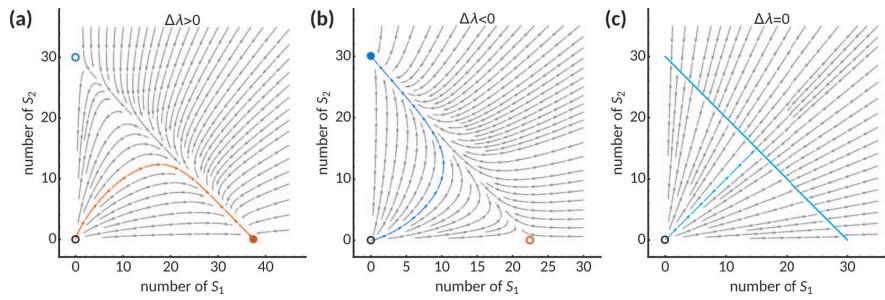


FIG. 4.2. Dynamics of the deterministic fixation model. *Phase portraits of the ODE model of two competing species (4.5) for different values of the growth rate difference  $\Delta\lambda$ .* (a), (b) If  $\Delta\lambda \neq 0$ , the system has a single stable fixed point corresponding to the fixation of the species with the highest growth rate. Filled dots denote stable fixed points; open circles denote unstable fixed points. (c) For equal growth rates ( $\Delta\lambda = 0$ ), the dynamics approaches a line of fixed points, representing coexistence of the two species (neutral case).

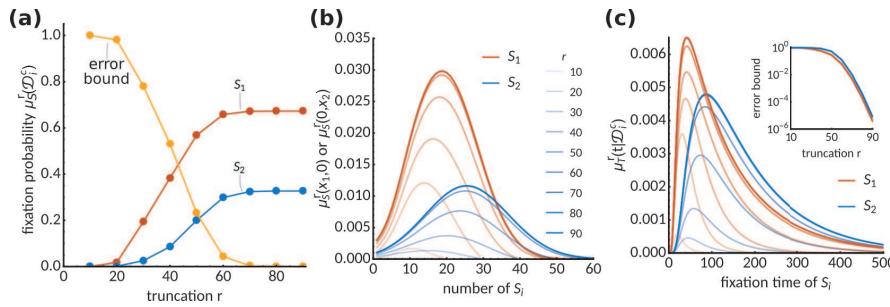


FIG. 4.3. Computation of fixation probabilities and times. (a) The lower bounds on the fixation probabilities converge (red =  $S_1$ , blue =  $S_2$ ) with increasing truncation parameter  $r$ , and the error bound (yellow) approaches zero. (b) The lower bounds on the exit location distributions converge with increasing  $r$  (from light to dark). (c) The lower bounds on the density of fixation times also converge. Parameters:  $b_1 = 2$ ,  $b_2 = 5$ ,  $d_1 = 1$ ,  $d_2 = 4$ ,  $\Delta\lambda = 0$ .

representing, respectively, the fixation of  $S_1$  and of  $S_2$ , and a third subset  $\{(0, 0)\}$  representing the extinction of both species.

We compute lower bounds  $\mu_S^r(\mathcal{D}_1^c)$  and  $\mu_S^r(\mathcal{D}_2^c)$  on the fixation probabilities using the ETFSP scheme and the truncations

$$\mathcal{S}_r = \{(x_1, x_2) \in \mathbb{N}^2 : x_1 + x_2 \leq r\},$$

with final computation time  $t_f^r := 3000$  and initial condition  $\gamma = 1_{(10, 10)}$ .

In Figure 4.3(a), the results for the neutral case  $\Delta\lambda = 0$  with different death rates  $d_1 < d_2$  show that the error bound  $\varepsilon_r$  decreases with  $r$  and can be made arbitrarily small. However, in contrast with the deterministic case,  $S_1$  fixes with higher probability, and the fixation dynamics depends not only on the growth rate difference  $\Delta\lambda$  but also on the difference in death rates  $\Delta d := d_1 - d_2$ . This demographic noise drives the species with the higher death rate ( $S_2$ ) to extinction more frequently [11].

To study this effect, we consider the distribution of  $S_1$  or  $S_2$  upon fixation (Figure 4.3(b)). The probabilities of exiting either through the states  $x_1$  in  $\mathcal{D}_1^c$  or the states  $x_2$  in  $\mathcal{D}_2^c$  are bounded by

$$\mu_S(x_1, 0) \geq \mu_S^r(x_1, 0), \quad \mu_S(0, x_2) \geq \mu_S^r(0, x_2).$$

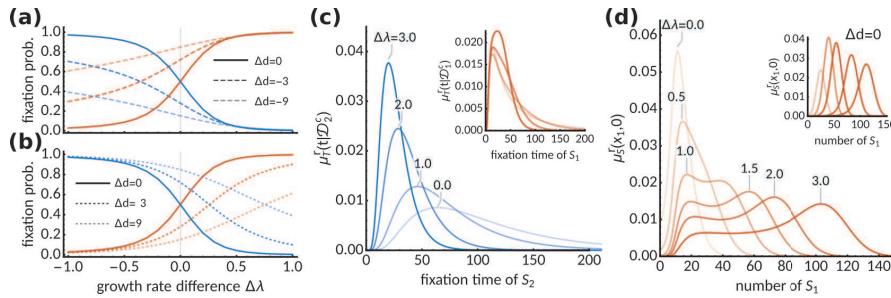


FIG. 4.4. Fixation in the presence of demographic noise. (a),(b) *Fixation probabilities as a function of the growth rate difference  $\Delta\lambda$  for various  $\Delta d$  values with  $d_1 = 1$  in (a) and  $d_2 = 1$  in (b)).* Red lines,  $S_1$ ; blue lines,  $S_2$ . (c) *The density of  $S_2$  fixation times for different  $\Delta\lambda$  values with  $d_1 = 1, d_2 = 10$ . Inset shows the corresponding densities of  $S_1$ .* (d) *Distribution of  $S_1$  individuals at fixation with  $d_1 = 1, d_2 = 10$  and varying  $\Delta\lambda$ . Inset shows  $d_1 = d_2 = 1$  case (no demographic noise).* All computations carried out with  $r = 200$  (20301 states) ensuring that  $\varepsilon_{200} < 10^{-8}$ . Birth rates:  $b_1 = 1 + \Delta\lambda + d_1$  and  $b_2 = 1 + d_2$ .

As shown in Corollary 3.1, the bounds  $\mu_S^r(x_1, 0)$  and  $\mu_S^r(0, x_2)$  increase monotonically in  $r$  and converge—in our numerics, the approximations are visually indistinguishable for  $r > 60$ . Note that the exit location distributions are wide and not clearly peaked around the intersections of  $\mathcal{D}_1^c$  and  $\mathcal{D}_2^c$  with the deterministic manifold. Indeed, the  $S_1$ -exit location distribution peaks at smaller values than deterministically plausible due to higher demographic noise along the direction of  $S_2$  disturbing the dynamics away from the deterministic stable manifold.

To characterize the time at which either fixation occurs, we compute bounds on the fixation time densities. The fixation time of  $S_1$  (resp.,  $S_2$ ) is the exit time conditioned on  $S_1$  (resp.,  $S_2$ ) fixing, and its density is given by

$$(4.7) \quad \mu_T(t|\mathcal{D}_i^c) := \frac{\mu(t, \mathcal{D}_i^c)}{\mu_S(\mathcal{D}_i^c)} \geq \frac{\mu^r(t, \mathcal{D}_i^c)}{\mu_S^r(\mathcal{D}_i^c) + \varepsilon_r} =: \mu_T^r(t|\mathcal{D}_i^c) \quad \forall i = 1, 2.$$

Figure 4.3(c) shows that the bounds on the conditional densities are monotonically increasing, whereas the inset shows that the bound of the approximation error

$$\|\mu_T(\cdot|\mathcal{D}_i^c) - \mu_T^r(\cdot|\mathcal{D}_i^c)\| \leq \frac{\varepsilon_r}{\mu_S^r(\mathcal{D}_i^c) + \varepsilon_r}$$

decreases with  $r$ .

*The effects of demographic noise.* Using ETFSP with a large truncation, we investigate how the behavior of the model depends on the growth rate difference  $\Delta\lambda$  and the death rate difference  $\Delta d$ , a measure of demographic noise. In the absence of demographic noise ( $\Delta d = 0$ ), as in the deterministic case, the fixation of  $S_1$  is favored if  $\Delta\lambda > 0$ , and the converse is true if  $\Delta\lambda < 0$  (Figure 4.4(a)). However, an increase in the demographic noise of  $S_2$  ( $\Delta d < 0$ ) leads to a higher fixation probability of  $S_1$ . If the demographic noise is large enough, the fixation of  $S_1$  becomes favored even if  $\Delta\lambda < 0$ . Conversely, Figure 4.4(b) shows that increasing the demographic noise of  $S_1$  ( $\Delta d > 0$ ) favors fixation of  $S_2$  over  $S_1$  even if  $\Delta\lambda > 0$ .

Next, we focus on the case  $\Delta\lambda > 0, \Delta d < 0$  where the most likely outcome is consistent with the deterministic case (i.e.,  $S_1$  is more likely to fix). Figure 4.4(c) shows the density of fixation times computed using (4.7) for  $S_1$  and  $S_2$  as a function of  $\Delta\lambda$ . The density of fixation times becomes narrower with increasing  $\Delta\lambda$  (inset),

indicating that large growth rate differences attenuate the stochasticity. Despite the fixation of  $S_2$  being less likely than that of  $S_1$ , the time required for this event decreases with  $\Delta\lambda$ : the mode of the conditional distribution  $\mu_T(t|\mathcal{D}_2^c)$  shifts to smaller times.

Figure 4.4(d) shows that the distribution of  $S_1$  individuals at fixation is bimodal for moderate values of  $\Delta\lambda$  and remains broad for larger values. This is the result of strong demographic fluctuations in the direction of  $S_2$  such that fixation of  $S_1$  can occur at small population size. If no demographic noise is present ( $\Delta d = 0$ ), the distributions are unimodal (inset) and considerably narrower regardless of the value of  $\Delta\lambda$ . In summary, demographic noise significantly alters the dynamics of small populations and can even reverse the direction of fixation predicted by deterministic models.

**5. Discussion.** In this paper, we have introduced and characterized the ETFSP scheme, which yields converging approximations of the exit distribution and occupation measure associated with the exit from a domain of continuous-time Markov chains. The ETFSP scheme consists of solving the system of coupled linear ODEs (1.6)–(1.7) and yields approximations of the desired measures. The total variation distance between the exit distribution and its approximation is bounded by one minus the mass of the approximation. Hence the quality of the approximation can be evaluated with no extra effort than that required for its computation.

We have considered minimal chains, i.e., those that do not explode or those that are killed off after exploding. A distinction arises for nonminimal chains, which are reinitialized after exploding [9, 22, 50]. In this case, the FSP and ETFSP still yield monotonically increasing lower bounds on the relevant measures, and the computable error bounds hold identically. However, they do not converge to the measures associated with nonminimal chains but to those associated with minimal chains (Theorems 1.1 and 2.5). The details pertinent to nonminimal chains are left as future work. Although we have not discussed time-inhomogeneous chains relevant in some applications [57, 13], we anticipate that both the ETFSP and FSP schemes apply identically when the rate matrix  $Q$  is replaced with its time-inhomogeneous analogue.

There are several issues worth considering for the application of the ETFSP scheme. Chief among them is the fact that the number of states often grows quickly with the desired accuracy, resulting in large systems of ODEs. Resource-efficient implementations of the FSP scheme have been developed to tackle this issue and can be adapted to the ETFSP setting (see [16] and references therein). To do so, notice that (1.6) is the set of ODEs obtained by applying the FSP scheme to (2.10) with  $\mathcal{D}$  replacing  $\mathcal{S}$ . In other words,  $\nu^r(t, \cdot)$  is  $\gamma_{\mathcal{D}_r} \exp(tQ_{\mathcal{D}_r})$ , where  $\gamma_{\mathcal{D}_r}$  and  $Q_{\mathcal{D}_r}$  are restrictions to the truncated domain  $\mathcal{D}_r$ . The corresponding approximation of the exit distribution is then obtained by rewriting (1.7) as

$$(5.1) \quad \mu^r(t, x) = \sum_{y \in \mathcal{D}_r} \nu^r(t, y) q(y, x) \quad \forall x \in \mathcal{S}_r \cap \mathcal{D}^c.$$

The rapid growth in the number of states can also be mitigated by guiding the truncation choice using simulation-based criteria [46, 52], moment bounds and Markov's inequality to obtain a priori error bounds [35, 36], or other state space exploration techniques (see [15, 16] and references therein). For cases where there are too many important states for ETFSP to handle, Galerkin methods [19, 18, 32] could be adapted to the exit time setting using Theorem 2.6.

Solving (1.6)–(1.7) numerically introduces an additional source of error [44]. A simple way to control this error is to apply randomization techniques [27, 16] on (1.7)

to obtain lower bounds  $\nu^r$ , and using (5.1) to compute lower bounds on  $\mu^r$ . The error bounds in Theorem 1.1(iii) hold if  $\mu^r$  and  $\nu^r$  are replaced with their lower bounds.

Numerically solving the ODEs (1.6) to obtain  $\nu^r$  and performing the matrix-vector multiplication in (5.1) often leads to an accumulation of errors in  $\mu^r$ . We circumvented this issue using an adaptive ODE solver [30] to solve the joint system (1.6)–(1.7), hence ensuring that the errors of both  $\nu^r$  and  $\mu^r$  are taken into account by the solver. A promising alternative here is to apply Krylov methods of the type in [4] to this joint system of ODEs.

Lastly, we did not address how to bound the approximation error of the occupation measure in practice. As shown in Theorem 1.1, the approximation error (1.15) depends on  $\mathbb{E}[\tau \wedge T_\infty]$ , which is bounded from above by the mean exit time  $\mathbb{E}[\tau]$ . For a broad class of chains (those with “rational rate matrices”), the mean exit time can itself be bounded using linear or semidefinite programming approaches [28, 35]. For more general chains, one can employ Foster–Lyapunov criteria [40, 35].

In summary, the ETFSP computes converging approximations of the exit distribution and occupation measure with controlled errors. As demonstrated in section 4, such highly accurate approximations can provide valuable insights into the dynamics of biochemical networks and interacting populations. Although our examples were biological, computing these measures is important to other fields, for instance, quantifying customer waiting times [38, 39], modeling computer-communication and transaction processing systems [34], computing reliability measures of complex systems [14], or in model checking [43].

**Acknowledgments.** We thank the two anonymous referees for their helpful remarks that have significantly improved this manuscript. J.K. gratefully thanks Prof. Sophia Yaliraki for an important stint in her research group, during which the material presented in this manuscript was partially developed.

## REFERENCES

- [1] W. J. ANDERSON, *Continuous-Time Markov Chains: An Applications-Oriented Approach*, Springer-Verlag New York, 1991.
- [2] M. ASSAF AND B. MEERSON, *WKB theory of large deviations in stochastic populations*, J. Phys. A, 50 (2017), 263001.
- [3] B. BARZEL AND O. BIHAM, *Calculation of switching times in the genetic toggle switch and other bistable systems*, Phys. Rev. E, 78 (2008), 041919.
- [4] K. BURRAGE, M. HEGLAND, S. MACNAMARA, AND R. B. SIDJE, *A Krylov-based finite state projection algorithm for solving the chemical master equation arising in the discrete modelling of biological systems*, in Markov Anniversary Meeting: An International Conference to Celebrate the 150th Anniversary of the Birth of A. A. Markov, Boson Books, 2006, pp. 21–38.
- [5] Y. CAO AND J. LIANG, *Adaptively biased sequential importance sampling for rare events in reaction networks with comparison to exact solutions from finite buffer dCME method*, J. Chem. Phys., 139 (2013), 025101.
- [6] Y. CAO, A. TEREBUS, AND J. LIANG, *Accurate chemical master equation solution using multi-finite buffers*, Multiscale Model. Simul., 14 (2016), pp. 923–963, <https://doi.org/10.1137/15M1034180>.
- [7] M. CHEN, *On three classical problems for Markov chains with continuous time parameters*, J. Appl. Probab., 28 (1991), pp. 305–320.
- [8] M. F. CHEN, *A comment on the book “Continuous-Time Markov Chains” by W. J. Anderson*, Chinese J. Appl. Prob. Statist., 12 (1996), pp. 55–59.
- [9] K. L. CHUNG, *Markov Chains with Stationary Transition Probabilities*, 2nd ed., Springer, Berlin, 1967.
- [10] G. W. A. CONSTABLE AND A. J. MCKANE, *Exploiting fast-variables to understand population dynamics and evolution*, J. Stat. Phys., 172 (2018), pp. 3–43.

- [11] G. W. A. CONSTABLE, T. ROGERS, A. J. MCKANE, AND C. E. TARNITA, *Demographic noise can reverse the direction of deterministic selection*, Proc. Natl. Acad. Sci. USA, 113 (2016), pp. E4745–E4754.
- [12] S. H. DANDACH AND M. KHAMMASH, *Analysis of stochastic strategies in bacterial competence: A master equation approach*, PLOS Comput. Biol., 6 (2010), e1000985.
- [13] J. DATTANI AND M. BARAHONA, *Stochastic models of gene transcription with upstream drives: Exact solution and sample path characterization*, J. R. Soc. Interface, 14 (2017), 20160833.
- [14] E. DE SOUZA E SILVA AND H. R. GAIL, *Transient solutions for Markov chains*, in Computational Probability, Internat. Ser. Oper. Res. Management Sci. 24, W. K. Grassmann, ed., Springer US, Boston, MA, 2000, pp. 43–79.
- [15] E. DE SOUZA E SILVA AND P. MEJÍA OCHOA, *State space exploration in Markov models*, SIGMETRICS Perform. Eval. Rev., 20 (1992), pp. 152–166.
- [16] K. N. DINH AND R. B. SIDJE, *Understanding the finite state projection and related methods for solving the chemical master equation*, Phys. Biol., 13 (2016), 035003.
- [17] P. D. DRUMMOND, T. G. VAUGHAN, AND A. J. DRUMMOND, *Extinction times in autocatalytic systems*, J. Phys. Chem. A, 114 (2010), pp. 10481–10491.
- [18] S. ENGBLOM, *Galerkin spectral method applied to the chemical master equation*, Commun. Comput. Phys., 5 (2009), pp. 871–896.
- [19] S. ENGBLOM, *Spectral approximation of solutions to the chemical master equation*, J. Comput. Appl. Math., 229 (2009), pp. 208–221.
- [20] W. FELLER, *On the integro-differential equations of purely discontinuous Markoff processes*, Trans. Amer. Math. Soc., 48 (1940), pp. 488–515.
- [21] W. FELLER, *An Introduction to Probability Theory and Its Applications: Volume 2*, 2nd ed., John Wiley & Sons, New York, 1971.
- [22] D. FREEDMAN, *Markov Chains*, Springer, New York, 1983.
- [23] K. R. GHUSINGA, J. J. DENNEHY, AND A. SINGH, *First-passage time approach to controlling noise in the timing of intracellular events*, Proc. Natl. Acad. Sci. USA, 114 (2017), pp. 693–698.
- [24] D. T. GILLESPIE, *A general method for numerically simulating the stochastic time evolution of coupled chemical reactions*, J. Comput. Phys., 22 (1976), pp. 403–434.
- [25] D. T. GILLESPIE, *Markov Processes: An Introduction for Physical Scientists*, Academic Press, Boston, MA, 1991.
- [26] R. GRIMA AND A. LEIER, *Exact product formation rates for stochastic enzyme kinetics*, J. Phys. Chem. B, 121 (2016), pp. 13–23.
- [27] D. GROSS AND D. R. MILLER, *The randomization technique as a modeling tool and solution procedure for transient Markov processes*, Oper. Res., 32 (1984), pp. 102–123.
- [28] 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.
- [29] R. HINCH AND S. J. CHAPMAN, *Exponentially slow transitions on a Markov chain: The frequency of calcium sparks*, European J. Appl. Math., 16 (2005), pp. 427–446.
- [30] A. C. HINDMARSH, P. N. BROWN, K. E. GRANT, S. L. LEE, R. SERBAN, D. E. SHUMAKER, AND C. S. WOODWARD, *SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers*, ACM Trans. Math. Softw., 31 (2005), pp. 363–396.
- [31] S. IYER-BISWAS AND A. ZILMAN, *First-passage processes in cellular biology*, Adv. Chem. Phys. Vol. 160, (2016), pp. 261–306.
- [32] T. JAHNKE AND T. UDRESCU, *Solving chemical master equations by adaptive wavelet compression*, J. Comput. Phys., 229 (2010), pp. 5724–5741.
- [33] D. G. KENDALL, *An artificial realization of a simple “birth-and-death” process*, J. R. Stat. Soc. Ser. B Stat. Methodol., 12 (1950), pp. 116–119.
- [34] W. J. KNOTTENBELT AND J. T. BRADLEY, *Tackling large state spaces in performance modelling*, in Formal Methods for Performance Evaluation, Lecture Notes in Comput. Sci. 4486, M. Bernardo and J. Hillston, eds., Springer, Berlin, 2007, pp. 318–370.
- [35] J. KUNTZ, *Deterministic Approximation Schemes with Computable Errors for the Distributions of Markov Chains*, Ph.D. thesis, Imperial College London, 2018.
- [36] J. KUNTZ, P. THOMAS, G. B. STAN, AND M. BARAHONA, *Rigorous Bounds on the Stationary Distributions of the Chemical Master Equation via Mathematical Programming*, preprint, <https://arxiv.org/abs/1702.05468>, 2017.
- [37] S. MACNAMARA, R. B. SIDJE, AND K. BURRAGE, *An improved dynamic finite state projection algorithm for the numerical solution of the chemical master equation with applications*, ANZIAM J., 49 (2007), pp. C413–C432.
- [38] B. MELAMED AND M. YADIN, *Numerical computation of sojourn-time distributions in queueing networks*, J. ACM, 31 (1984), pp. 839–854.

- [39] B. MELAMED AND M. YADIN, *Randomization procedures in the computation of cumulative-time distributions over discrete state Markov processes*, Oper. Res., 32 (1984), pp. 926–944.
- [40] M. MENSHKOV AND D. PETRITIS, *Explosion, implosion, and moments of passage times for continuous-time Markov chains: A semimartingale approach*, Stochastic Process. Appl., 124 (2014), pp. 2388–2414.
- [41] R. METZLER, G. OSANIN, AND S. REDNER, *First-Passage Phenomena and Their Applications*, World Scientific, Hackensack, NJ, 2014.
- [42] S. P. MEYN AND R. L. TWEEDIE, *Stability of Markovian processes III: Foster-Lyapunov criteria for continuous-time processes*, Adv. in Appl. Probab., 25 (1993), pp. 518–548.
- [43] D. MILIOS, G. SANGUINETTI, AND D. SCHNOERR, *Probabilistic model checking for continuous-time Markov chains via sequential Bayesian inference*, in International Conference on Quantitative Evaluation of Systems, Springer, Cham, 2018, pp. 289–305.
- [44] C. MOLER AND C. VAN LOAN, *Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later*, SIAM Rev., 45 (2003), pp. 3–49, <https://doi.org/10.1137/S00361445024180>.
- [45] B. MUNSKY AND M. KHAMMASH, *The finite state projection algorithm for the solution of the chemical master equation*, J. Chem. Phys., 124 (2006), 044104.
- [46] B. MUNSKY AND M. KHAMMASH, *A multiple time interval finite state projection algorithm for the solution to the chemical master equation*, J. Comput. Phys., 226 (2007), pp. 818–835.
- [47] B. MUNSKY, I. NEMENMAN, AND G. BEL, *Specificity and completion time distributions of biochemical processes*, J. Chem. Phys., 131 (2009), 235103.
- [48] M. F. NEUTS, *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*, 2nd ed., Dover, New York, 1994.
- [49] S. REDNER, *A Guide to First-Passage Processes*, Cambridge University Press, Cambridge, UK, 2001.
- [50] L. C. G. ROGERS AND D. WILLIAMS, *Diffusions, Markov Processes and Martingales: Volume 1. Foundations*, 2nd ed., Cambridge University Press, Cambridge, UK, 2000.
- [51] D. SCHNOERR, B. CSEKE, R. GRIMA, AND G. SANGUINETTI, *Efficient Low-Order Approximation of First-Passage Time Distributions*, Phys. Rev. Lett., 119 (2017), 210601.
- [52] R. B. SIDJE AND H. D. VO, *Solving the chemical master equation by a fast adaptive finite state projection based on the stochastic simulation algorithm*, Math. Biosci., 269 (2015), pp. 10–16.
- [53] A. SINGH AND J. J. DENNEHY, *Stochastic holin expression can account for lysis time variation in the bacteriophage  $\lambda$* , J. Royal Soc. Interface, 11 (2014), 20140140.
- [54] R. SYSKI, *Passage Times for Markov Chains*, Vol. 1, IOS Press, Amsterdam, 1992.
- [55] A. VALLERIANI, X. LI, AND A. B. KOLOMEISKY, *Unveiling the hidden structure of complex stochastic biochemical networks*, J. Chem. Phys., 140 (2014), 064101.
- [56] N. G. VAN KAMPEN, *Stochastic Processes in Physics and Chemistry*, 3rd ed., Elsevier, Amsterdam, 2007.
- [57] M. VOLIOTIS, P. THOMAS, R. GRIMA, AND C. G. BOWSHER, *Stochastic simulation of biomolecular networks in dynamic environments*, PLoS Comput. Biol., 12 (2016), e1004923.