

## Joint densities of hitting times for finite state Markov processes

Tomasz R. BIELECKI<sup>1</sup>, Monique JEANBLANC<sup>2</sup>, Ali Devin SEZER<sup>2,3,\*</sup>

<sup>1</sup>Department of Applied Mathematics, Illinois Institute of Technology, Chicago, IL, USA

<sup>2</sup>Laboratory of Mathematics and Modelization of Évry (LaMME), University of Évry Val d'Essonne  
(UMR CNRS 8071), Évry, France

<sup>3</sup>Institute of Applied Mathematics, Middle East Technical University, Ankara, Turkey

Received: 10.08.2016

Accepted/Published Online: 01.06.2017

Final Version: 24.03.2018

**Abstract:** For a finite state Markov process  $X$  and a finite collection  $\{\Gamma_k, k \in K\}$  of subsets of its state space, let  $\tau_k$  be the first time the process visits the set  $\Gamma_k$ . In general,  $X$  may enter some of the  $\Gamma_k$  at the same time and therefore the vector  $\tau := (\tau_k, k \in K)$  may put nonzero mass over lower dimensional regions of  $\mathbb{R}_+^{|K|}$ ; these regions are of the form  $R_s = \{\mathbf{t} \in \mathbb{R}_+^{|K|} : t_i = t_j, i, j \in s(1)\} \cap \bigcap_{l=2}^{|s|} \{\mathbf{t} : t_m < t_i = t_j, i, j \in s(l), m \in s(l-1)\}$  where  $s$  is any ordered partition of the set  $K$  and  $s(j)$  denotes the  $j^{\text{th}}$  subset of  $K$  in the partition  $s$ . When  $|s| < |K|$ , the density of the law of  $\tau$  over these regions is said to be “singular” because it is with respect to the  $|s|$ -dimensional Lebesgue measure over the region  $R_s$ . We derive explicit/recursive and simple to compute formulas for these singular densities and their corresponding tail probabilities over all  $R_s$  as  $s$  ranges over ordered partitions of  $K$ . We give a numerical example and indicate the relevance of our results to credit risk modeling.

**Key words:** Finite state Markov processes, simultaneous hitting times, densities of singular parts, multiple first hitting times, generalized multivariate phase-type distributions, credit risk modeling

### 1. Introduction

One of the basic random variables associated with a Markov process  $X$  is its first hitting time to a given subset of its state space. In the present work, we will confine ourselves to finite state Markov processes. If  $X$  has an absorbing state and all of the states can communicate with it, the distribution of the first hitting time to the absorbing state is said to be a *phase-type distribution*. Phase-type distributions, which go back to Erlang [13], are used to model a wide range of phenomena in, e.g., reliability theory, communications systems, insurance, and finance. The literature on these distributions is immense; see, e.g., [1–3, 16, 23].

To the best of our knowledge, Assaf et al. [5] were the first to study multivariate (multidimensional) phase-type distributions. Their setup for the two-dimensional case is as follows: take two proper subsets  $\Gamma_1$  and  $\Gamma_2$  of  $E$ , the state space of  $X$ , and assume that with positive probability the process enters their intersection; let  $\tau_k$  be the first time the process hits  $\Gamma_k$ . The joint law of  $(\tau_1, \tau_2)$  is called a two-dimensional phase-type distribution. Higher dimensional versions are defined similarly: for a finite collection of subsets  $\{\Gamma_k, k \in K\}$

\*Correspondence: devin.sezer@gmail.com

The research of TR Bielecki was supported by NSF Grant DMS-0604789 and NSF Grant DMS-0908099.

The research of Monique Jeanblanc was supported by Chair Markets in Transition (Fédération Bancaire Française) and Labex ANR 11-LABX-0019.

The research of Ali Devin Sezer was supported by RBUCE-UP Marie Curie European Project.

of the state space  $E$ , the distribution of the random vector  $\boldsymbol{\tau} := (\tau_k, k \in K)$  is a  $|K|$ -dimensional phase-type distribution, where  $|K|$  denotes the number of elements in  $K$ . In general, the underlying process can hit some of the  $\Gamma_k$  simultaneously and this implies that multidimensional phase-type distributions can put nonzero mass on certain lower dimensional regions of  $\mathbb{R}_+^{|K|}$ ; e.g., for  $K = \{1, 2, 3\}$  these regions are  $\{\mathbf{t} : t_1 = t_2 < t_3\}$ ,  $\{\mathbf{t} : t_1 = t_2 = t_3\}$ ,  $\{\mathbf{t} : t_2 < t_1 = t_3\}$ , etc. In general, each ordered partition  $s$  of  $K$  defines an  $|s|$ -dimensional subset  $R_s$  of  $\mathbb{R}_+^{|K|}$  ( $|s|$  denotes the number of subsets of  $K$  that appear in the partition  $s$ ; the precise definitions are given in Section 2.2) over which the law of  $\boldsymbol{\tau}$  may put nonzero probability mass. The law of  $\boldsymbol{\tau}$ , when restricted to one of these lower dimensional regions, turns out to have a density with respect to the  $|s|$ -dimensional Lebesgue measure of that region; these densities are called “singular” (or “the singular part(s) of the density of  $\boldsymbol{\tau}$ ”) because of the lower dimensionality of  $R_s$ . The focus of the present paper is on these singular densities of  $\boldsymbol{\tau}$ ; our goal is to find simple formulas for them and for the tail probabilities associated with them. To the best of our knowledge, the only paper currently available that develops density or tail probability formulas for the singular parts is [5], which focuses on the case of  $|K| = 2$  and  $\Gamma_1, \Gamma_2$  absorbing. The only currently available density formula for  $|K| > 2$  was also derived in [5] and covers only the absolutely continuous part of the density (i.e. the density over the  $|K|$ -dimensional region  $\{\mathbf{t} \in \mathbb{R}_+^{|K|} : t_i \neq t_j, \text{ for } i \neq j\}$ ) in the case when  $\Gamma_k, k \in K$ , are assumed absorbing; display (45) in Section 4.3 gives this density formula from [5] (which is stated without a proof in [5]; [14] provides a proof for it). Over the last three decades, this formula has found use in a range of application areas, e.g., [20, modeling of plant growth], [9, insurance], and [14, credit risk].

The main contributions of the present paper are Theorem 3.1, which gives an explicit formula for the singular density of the random vector  $\boldsymbol{\tau}$  over each  $R_s \subset \mathbb{R}_+^{|K|}$  as  $s$  ranges over all partitions of  $K$ , covering all possible singular and nonsingular parts, and Theorem 3.2, which gives a recursive formula for the tail probabilities of  $\boldsymbol{\tau}$  using the density formulas. We make no assumptions on whether  $\{\Gamma_k, k \in K\}$  are absorbing and Theorem 3.1 gives a general formula for the joint density of a collection of first hitting times for any finite state Markov process  $X$ . The density formula when  $\Gamma_k$  are absorbing follows as a special case (Proposition 4.1).

One common method of computing a density is to compute the corresponding tail probability and then to differentiate it to get the density. This is the method used in [5, 14]. As discussed below and in Section 3.4, “singular” tail probabilities of  $\boldsymbol{\tau}$  (i.e. tail probabilities where some components of  $\boldsymbol{\tau}$  are equal) turn out to be more complex than the corresponding densities. If one tries to compute singular tail probabilities (for example, using the methodology of taboo probabilities as presented by Syski in [23]), one runs into difficult calculations even when  $|K|$  is small. For this reason, rather than focusing on the tail probabilities of  $\boldsymbol{\tau}$ , we directly compute the singular densities of  $\boldsymbol{\tau}$  over each  $R_s \subset \mathbb{R}_+^{|K|}$  using the following idea: for each  $\mathbf{t} \in R_s \subset \mathbb{R}_+^{|K|}$ , the event  $\{\boldsymbol{\tau} = \mathbf{t}\}$  corresponds to the limit of a specific set of trajectories of the Markov process whose (vanishing) probability can be written in terms of the exponentials of submatrices of the intensity matrix. These sets of trajectories are of the following form:  $X$  stays in a waiting set  $W_1$  until the first jump time  $\bar{t}_1$  and jumps exactly at that time into  $T_1 \subset W_2$ , then stays in the set  $W_2$  until time  $\bar{t}_2$  and jumps exactly then into  $T_2$  and so on, until all of the pairs  $(W_n, T_n)$ ,  $n \leq |s|$ , are exhausted. The waiting and the target sets  $(W_n, T_n)$ , and the jump times  $\bar{t}_n$ ,  $n = 1, 2, 3, \dots, |s|$ , are all determined by  $\mathbf{t}$ . Section 3.1 gives the one-step version of this argument in the computation of the density of a single  $\tau_k$ , given as Proposition 3.1. The argument extends to multiple hitting times in Section 3.2 and the multidimensional density is given as formula (20) in Theorem

**3.1.** Formula (20) gives the density of all singular parts of the distribution of  $\tau$ ; it involves the multiplication of exponentials of appropriate submatrices of the intensity matrix  $\lambda$  determined by the sequence of waiting and target sets. The well-known density formula (45) from [5] for the absolutely continuous part involves only exponentials of the full intensity matrix.

The proof of Theorem 3.1 uses induction on  $|K|$  and starts from the following fact:  $f$  is the density of  $\tau$  if and only if  $\mathbb{E}[g(\tau)] = \int f(x)g(x)dx$  for all bounded continuous functions  $g$ . Other elements of the proof are the strong Markov property of the underlying process at its jump times, properties of the conditional expectation, and the distribution of the process conditioned on events that occur while the process goes through the waiting and target sets mentioned above.

Tail events are events of the form  $\{\tau \in \prod_{k \in K}(t_k, \infty)\}$ . In general, to compute the probability of a tail event, one must decompose it into pieces of the form  $\{\tau \in R_s \cap \prod_{k \in K}(t_k, \infty)\}$  and this makes it difficult to derive closed-form simple expressions for tail probabilities. An alternative and simple way to express them is given in Theorem 3.2 of Section 3.4 as a recursive formula involving a single-dimensional integral over a one-dimensional version of the density formula.

In Section 4, we derive alternative expressions for the density and the tail probability formulas for absorbing  $\{\Gamma_k\}$  and indicate the connections between our results and the density formulas in [5].

The calculation of the density formula (20) involves only  $|s| \leq |K|$  number of matrix multiplications and exponentiations. For moderate sizes of  $|E|$  these computations can be carried out on a standard laptop (the precise speed will depend on the matrix multiplication and matrix exponentiation algorithms used in the implementation; for example, for  $|E| = 1500$ , a matrix exponentiation of a typical matrix of size  $1500 \times 1500$  takes approximately 10 s on our laptop using Octave [11]; for  $|s| = 10$  the density formula requires 10 of these operations). We give several numerical examples with  $|E| = 27$  in Section 5.

Kulkarni [19] generalized multivariate phase-type distributions to a class of distributions called MPH\* and derived a theoretical representation of members of this class (see [19, Equation (35)]), based on an infinite sum characterization of the full occupation measure (the time spent in each state until absorption to an absorbing state) of the process (given in [19, Equation (34)]). Another recent generalization of multivariate phase-type distributions was given by Bladt and Nielsen in [7], in which a distribution on  $\mathbb{R}_+^k$  is called “multivariate matrix-exponential” (MVME) if its Laplace transform is the ratio of two multivariate polynomials; as explained in the same work, these can be seen as generalizations of multivariate phase-type distributions.

From an applied perspective, our primary motivation in deriving the results in the present paper has been to model default times of companies/obligors with first hitting times of a finite state Markov process *where multiple defaults are allowed to happen at the same time*; with the results of the present paper this is now possible in greater generality (for the case of two obligors one could use the results in [5]). The conclusion explains this application starting from the credit risk model of [14] and the numerical example of Section 5. In addition to credit risk, our results may find applications in reliability theory (see, e.g., [5]), counterparty risk (see, e.g., [10]), and insurance (see, e.g., [6]). From a theoretical perspective, the first motivation of the paper has been the solution of a problem whose two-dimensional version was solved in [5], i.e. find simple expressions for the density of  $\tau$ ; surprisingly, prior to the results of the present paper, such expressions were not available in the literature. See the Conclusion (Section 6) for further theoretical motivation for the present study. Another potential application of multivariate phase-type distributions with singularities was identified in [21] in the modeling of simultaneous shocks to a system. With the results of the present paper such models

become implementable; in particular, in the fitting of such models, the singular densities we have derived can be used to run maximum likelihood estimation.

**2. Definitions**

Let  $E$  be a finite set and  $X$  an  $E$ -valued continuous time process defined over a measurable space  $(\Omega, \mathcal{F})$ . Let  $\mathcal{F}_t := \sigma\{X_s, s \leq t\}$  be the filtration generated by  $X$  and for any stopping time  $\tau$  and let  $\mathcal{F}_\tau$  denote the  $\sigma$ -algebra of events determined prior to time  $\tau$  [17, Definition 2.12]. Let  $P_i, i \in E$ , be a family of measures on  $(\Omega, \mathcal{F})$  such that  $X$  is a time-homogeneous Markov chain with intensity matrix  $\lambda$  and such that  $P_i(X_0 = i) = 1$ . The jump rate of the process from state  $i$  is  $-\lambda(i, i) = \sum_{j \neq i} \lambda(i, j)$ . For any probability measure  $\alpha$  on  $E$ , define  $P_\alpha := \sum_{i \in E} \alpha(i)P_i$ , where  $\alpha(i)$  denotes the measure of  $\{i\}, i \in E$ , under  $\alpha$ . Under  $P_\alpha$ , the process  $X$  is a time-homogeneous Markov process with intensity matrix  $\lambda$  and initial distribution  $\alpha$ . For  $\alpha = \delta_i, i \in E$ , where  $\delta_i$  denotes the Dirac measure on  $E$  putting unit mass on  $i$ , we will simply write  $P_i$  instead of  $P_{\delta_i}$ . For a square matrix  $A$ , define its exponential  $e^A := \sum_{j=0}^\infty \frac{1}{j!} A^j$ . The distribution of  $X_t$  on  $E$  under  $P_\alpha$  at time  $t$  is  $\alpha e^{t\lambda}$ .

**Remark 2.1** In what follows we will assume  $-\lambda(i, i) > 0$  for all  $i \in E$ . The cases where  $\lambda(i, i) = 0$  is allowed can be treated by straightforward modifications of the arguments below. Section 4 treats the case when the sets  $\Gamma_k$  are assumed absorbing.

For a finite collection  $\{\Gamma_k \subset E, k \in K\}$  of subsets of  $E$ , define the hitting times

$$\tau_k := \inf\{u \in (0, \infty) : X_u \in \Gamma_k\}.$$

The index set  $K$  can be any finite set, but we will always take it to be a finite subset of the integers. In the next section, we derive formulas for the (conditional) joint density and tail probabilities of the stopping times  $\{\tau_k, k \in K\}$ . To ease notation, unless otherwise noted, we will assume throughout that  $E - \bigcup_{k \in K} \Gamma_k$  is not empty and that the initial distribution  $\alpha$  puts its full mass on this set. See Remark 3.2 and Section 3.3 for comments on how one removes this assumption.

For a set  $a \subset E$ ,  $a^c$  means its complement with respect to  $E$  and  $|a|$  means the number of elements in it. For two subsets  $a, b \subset E$ , we define  $\lambda(a, b)$  as the matrix with elements

$$\begin{cases} \lambda(i, j) \text{ if } i \in a, j \in b, \\ 0, \text{ otherwise.} \end{cases} \tag{1}$$

For  $a \subset E$ , we will write  $\lambda(a)$  for  $\lambda(a, a)$ , so that in particular  $\lambda = \lambda(E)$ .

Throughout we will need to refer to zero matrices and vectors of various dimensions; we write all as 0 and the dimension will always be clear from context.

**2.1. Restriction and extension of vectors and  $\tau$  as a random function**

For any nonempty finite set  $a$ , let  $\mathbb{R}^a$  be the set of functions from  $a$  to  $\mathbb{R}$ . The set  $\mathbb{R}^a$  is the same as  $\mathbb{R}^{|a|}$ , except for the way we index the components of their elements. For two sets  $a \subset b$  and  $y \in \mathbb{R}^b$  we denote the restriction of  $y$  to  $a$  by  $y|_a \in \mathbb{R}^a$ :

$$y|_a(i) := y(i) \text{ for } i \in a. \tag{2}$$

The same notation continues to make sense if we replace the set  $a$  by a set of the form  $b \times c$ , and therefore it can be used to write submatrices of a matrix. Thus, for  $\mathbf{M} \in \mathbb{R}^{E \times E}$  and nonempty  $b, c \subset E$ ,

$$\mathbf{M}|_{b \times c} \tag{3}$$

denotes the submatrix of  $\mathbf{M}$  consisting of its components  $M(i, j)$  with  $(i, j) \in b \times c$ . For  $b = c$  we write  $\mathbf{M}|_b$ .

For  $x \in \mathbb{R}^a$ , and  $a \subset b$ , denote by  $x|_b \in \mathbb{R}^b$  the following extension of  $x$  to  $b$ :

$$x|_b(i) = \begin{cases} x(i) & \text{for } i \in a, \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

The random vector  $\boldsymbol{\tau} = (\tau_k, k \in K)$  can also be thought of as a random function on  $K$ , and we will often do so. Thus, for  $A \subset K$ , we may write  $\boldsymbol{\tau}|_A$  to denote  $(\tau_k, k \in A)$ . The advantage of the notation  $\boldsymbol{\tau}|_A$  is that we are able to index its components with elements of  $A$  rather than with the integers  $\{1, 2, 3, \dots, |A|\}$ ; this is useful when stating the recursive formulas and proofs below.

### 2.2. Subpartitions of $K$

The key feature of the distribution of  $\boldsymbol{\tau}$ , already mentioned in the Introduction, is that it may put nonzero mass on lower dimensional subsets of  $\mathbb{R}_+^{|K|}$ . This happens, for example, when  $X$  hits with positive probability  $\bigcap_{k \in A} \Gamma_k$  before  $\bigcup_{k \in A} \Gamma_k - \bigcap_{k \in A} \Gamma_k$  for some  $A \subset K$  with  $|A| > 1$ . As this example suggests, one can divide  $\mathbb{R}_+^{|K|}$  into a number of regions and associate with each an intersection of events of the form “ $X$  hits  $a$  before  $b$ ” for appropriate subsets  $a, b \subset E$ . To write down the various regions and the corresponding events we will use subpartitions of  $K$ , which we introduce now.

Recall that  $K$  is the set of indices of the stopping times  $\{\tau_k\}$  or, equivalently, of the sets  $\{\Gamma_k\}$ . We call an ordered sequence of disjoint nonempty subsets of  $K$  an ordered *subpartition* of  $K$ . If the union of all elements of a subpartition is equal to  $K$ , then we call it a partition. For example,  $(\{1, 2\}, \{3\}, \{4\})$  [ $(\{1, 2\}, \{4\})$ ] is a [sub]partition of  $\{1, 2, 3, 4\}$ . Denote by  $|s|$  the number of components in the subpartition and by  $s(n)$  its  $n^{\text{th}}$  component,  $n \in \{1, 2, 3, \dots, |s|\}$ . In which order the sets appear in the partition matters. For example,  $(\{3\}, \{4\}, \{1, 2\})$  is different from the previous partition. In the combinatorics literature this is often called an “ordered partition”; see, e.g., [22]. Only ordered subpartitions appear in the present work and, to be brief, we always assume that every subpartition has a definite order and we drop the adjective “ordered.” With a slight abuse of notation we will write  $s(n_1, n_2)$  to denote the  $n_2$ -th element of the  $n_1$ -th set in the subpartition.

Two subpartitions  $s_1$  and  $s_2$  are said to be disjoint if  $\cup_n s_1(n)$  and  $\cup_n s_2(n)$  are disjoint subsets of  $K$ . For a given disjoint pair of subpartitions  $s_1$  and  $s_2$ , let  $s_1 \cup s_2$  be their concatenation, e.g.,  $(\{1, 2\}, \{3\}) \cup (\{4, 6\}) = (\{1, 2\}, \{3\}, \{4, 6\})$ .

For a subpartition  $s$ , let us denote by  $Ls$  its left shift:

$$Ls = L(s(1), s(2), \dots, s(|s|)) := (s(2), s(3), \dots, s(|s|)).$$

Let  $L^m$  denote the left shift applied  $m$  times. Similarly for  $\mathbf{t} \in \mathbb{R}^n$ ,  $n > 1$ , let  $L\mathbf{t} \in \mathbb{R}^{n-1}$  be its left shift. For  $\mathbf{t} \in \mathbb{R}^n$  and  $r \in \mathbb{R}$  let  $\mathbf{t} - r$  denote  $(t_1 - r, t_2 - r, \dots, t_n - r)$ .

Given a subpartition  $s$  and an integer  $n$  such that  $0 < n \leq |s|$ , let  $s - s(n)$  be the subpartition that is the same as  $s$  but without  $s(n)$ , e.g.,  $(\{1, 2\}, \{3\}, \{4, 7\}) - \{3\} = (\{1, 2\}, \{4, 7\})$ . Given a nonempty set  $A$

satisfying  $A \subset K - \bigcup_{n=1}^{|s|} s(n)$ , let  $s + A$  denote the subpartition that contains all the sets in  $s$  and  $A$ , e.g.,  $(\{1, 2\}, \{3\}) + \{4, 7\} = (\{1, 2\}, \{3\}, \{4, 7\})$ .

Define

$$S(s) := \bigcup_{n=1}^{|s|} \bigcup_{k \in s(n)} \Gamma_k, \tag{5}$$

i.e.  $S(s)$  is the set of all states of  $X$  contained in the subpartition  $s$ . We will denote the empty subpartition  $\emptyset$ ,  $S(\emptyset) = \emptyset$  by definition (5). For a partition  $s$ , define  $R_s \subset \mathbb{R}_+^{|K|}$  as

$$R_s := \bigcap_{n=1}^{|s|} \left\{ \mathbf{t} \in \mathbb{R}_+^{|K|} : t_{k_1} = t_{k_2}, k_1, k_2 \in s(n) \right\} \cap \left\{ \mathbf{t} \in \mathbb{R}_+^{|K|} : t_{s(1,1)} < t_{s(2,1)} < \dots < t_{s(|s|,1)} \right\}; \tag{6}$$

this and the expression used for  $R_s$  in the abstract define the same set; the one in the abstract takes less space while (6) is easier to read.

**Example 2.1** For  $K = \{1, 2, 3, 4, 5, 6\}$ , and  $s = (\{1, 4\}, \{2\}, \{3, 5, 6\})$ , we have  $|s| = 3$ ,  $s(1) = \{1, 4\}$ ,  $s(2) = \{2\}$ ,  $s(3) = \{3, 5, 6\}$ ,  $s(1, 1) = 1$ ,  $s(2, 1) = 2$ ,  $s(3, 1) = 3$ , and

$$R_s = \{ \mathbf{t} : t_1 = t_4 < t_2 < t_3 = t_5 = t_6 \}.$$

Let  $\mathcal{S}$  be the set of all partitions of  $K$ . The sets  $R_s, s \in \mathcal{S}$ , are disjoint and their union is  $\mathbb{R}_+^{|K|}$ . Our main result, Theorem 3.1 below, shows that for each  $s \in \mathcal{S}$ , the distribution of  $\tau$  restricted to  $R_s$  is absolutely continuous with respect to the  $|s|$ -dimensional Lebesgue measure on  $R_s$  and gives a formula for the corresponding density.

Let  $\mathbf{I}$  be the identity matrix  $\mathbf{I} \in \mathbb{R}^{|E| \times |E|}$ . For  $a \subset E$ , we replace its rows whose indices appear in  $a^c$  with the 0 vector and call the resulting matrix  $\mathbf{I}_a$ , e.g.,  $\mathbf{I}_E$  is  $\mathbf{I}$  itself and  $\mathbf{I}_\emptyset$  is the zero matrix. Basic linear algebra implies that matrix  $\mathbf{I}_a$  has the following actions on matrices:

**Lemma 2.1** Let  $n$  be a positive integer. For any  $\mathbf{M} \in \mathbb{R}^{|E| \times n}$ , the left multiplication by  $\mathbf{I}_a$ , i.e.  $\mathbf{I}_a \mathbf{M}$ , acts on the rows of  $\mathbf{M}$ , and  $\mathbf{I}_a \mathbf{M}$  is the same as  $\mathbf{M}$  except that its rows whose indices are in  $a^c$  are replaced by 0 (a zero row vector of dimension  $n$ ), i.e. if  $\mathbf{r}_i$  is the  $i^{th}$  row of  $\mathbf{M}$  then the  $i^{th}$  row of  $\mathbf{I}_a \mathbf{M}$  is  $\mathbf{r}_i$  if  $i \in a$  and 0 otherwise. Similarly, right multiplication by  $\mathbf{I}_a$  acts on the columns of a matrix  $\mathbf{M} \in \mathbb{R}^{n \times |E|}$ , and  $\mathbf{M} \mathbf{I}_a$  is the same as  $\mathbf{M}$  except that now the columns with indices in  $a^c$  are set to zero.

It follows from (1) and Lemma 2.1 that  $\lambda(a, b) = \mathbf{I}_a \lambda \mathbf{I}_b$ . The operation of setting some of the columns of the identity matrix to zero commutes with set operations, i.e. one has  $\mathbf{I}_{a \cap b} = \mathbf{I}_a \mathbf{I}_b$ ,  $\mathbf{I}_{a \cup b} = \mathbf{I}_a + \mathbf{I}_b - \mathbf{I}_a \mathbf{I}_b$ ,  $\mathbf{I}_{a^c} = \mathbf{I} - \mathbf{I}_a$ . Using this and Lemma 2.1 one can write formulas involving  $\lambda(\cdot, \cdot)$  in a number of ways. For example,  $\lambda(a^c, a)$  can be written as  $\mathbf{I}_{a^c} \lambda \mathbf{I}_a = (\mathbf{I} - \mathbf{I}_a) \lambda \mathbf{I}_a = \lambda \mathbf{I}_a - \mathbf{I}_a \lambda \mathbf{I}_a$ , and  $\lambda(a, b \cap c)$  can be written as  $\mathbf{I}_a \lambda \mathbf{I}_{b \cap c} = \mathbf{I}_a \lambda \mathbf{I}_b \mathbf{I}_c = \mathbf{I}_a \lambda \mathbf{I}_c \mathbf{I}_b$ .

### 3. The density of first hitting times

We start by deriving the density of a single hitting time over sets of sample paths that avoid a given subset of the state space until hitting occurs.

**3.1. Density of one hitting time**

For any set  $d \subset E$ , any index  $j \in E$ , any probability measure  $\alpha \in \mathcal{P}(E)$ , and any real number  $u \in \mathbb{R}_+$  we define

$$\begin{aligned}
 p_{\alpha,d}^u(j) &:= P_\alpha(X_u = j, X_v \notin d, \forall v \leq u), \\
 p_d^u(i, j) &:= p_{\delta_i,d}^u(j) = P_i(X_u = j, X_v \notin d, \forall v \leq u), \quad p^u(i, j) := P_i(X_u = j).
 \end{aligned}
 \tag{7}$$

Let the symbol  $\mathbf{p}_{\alpha,d}^u$  denote the row vector with components  $p_{\alpha,d}^u(j)$ ;  $\mathbf{p}_d^u$  and let  $\mathbf{p}^u$  denote the  $|E| \times |E|$  matrices with elements  $p_d^u(i, j)$  and  $p^u(i, j)$ , respectively. Note that  $\mathbf{p}_{\alpha,d}^u = \alpha \mathbf{p}_d^u$ . It follows from the definition of  $p^u(i, j)$  that

$$\lim_{u \rightarrow 0} p^u(i, j)/u = \lambda(i, j), \quad i \neq j.
 \tag{8}$$

**Lemma 3.1** *Let  $\alpha$  be an initial distribution on  $E$  with  $\alpha|_d = 0$ . Then*

$$\mathbf{p}_{\alpha,d}^u = \alpha e^{u\lambda(d^c)}.
 \tag{9}$$

**Proof** The following proof parallels the proof of [2, Theorem 3.4, page 48]. Let  $\nu_1$  be the first jump time of  $X$ ; by definition of the process  $X$  and the probability measure  $P_i$ ,  $X_0 = i$  under  $P_i$  and  $\nu_1$  is exponentially distributed with rate  $-\lambda(i, i) > 0^*$  and  $X_{\nu_1}$  is independent of  $\nu_1$  with distribution  $P_i(X_{\nu_1} = l) = -\frac{\lambda(i,l)}{\lambda(i,i)}$ ; let us fix  $(i, j) \in d^c \times d^c$  and break the computation of  $p_d^u(i, j)$  into two cases according to whether  $\nu_1 < u$  or  $\nu_1 > u$ ,

$$\begin{aligned}
 p_d^u(i, j) &= P_i(X_u = j, X_v \notin d, \forall v \leq u, \nu_1 > u) + P_i(X_u = j, X_v \notin d, \forall v \leq u, \nu_1 < u) \\
 &= P_i(\nu_1 > u)\delta_i(j) + \mathbb{E}_i[PX_u = j, X_v \notin d, \forall v \leq u, \nu_1 < u | \mathcal{F}_{\nu_1}].
 \end{aligned}$$

The strong Markov property of  $X$  (applied at time  $\nu_1$ ) implies

$$p_d^u(i, j) = P_i(\nu_1 > u)\delta_i(j) + \mathbb{E}_i[1_{\{X_{\nu_1} \notin d\}} 1_{\{\nu_1 < u\}} p_d^{u-\nu_1}(X_{\nu_1}, j)].$$

The joint distribution of  $(\nu_1, X_{\nu_1})$  pointed out above now gives

$$\begin{aligned}
 p_d^u(i, j) &= e^{\lambda(i,i)u}\delta_i(j) + \int_0^u \lambda(i, i)e^{\lambda(i,i)v} \left( \sum_{l \in d^c - \{i\}} \frac{\lambda(i, l)}{\lambda(i, i)} p_d^{u-v}(l, j) \right) dv \\
 &= e^{\lambda(i,i)u} \left( \delta_i(j) + \int_0^u \lambda(i, i)e^{\lambda(i,i)(v-u)} \left( \sum_{l \in d^c - \{i\}} \frac{\lambda(i, l)}{\lambda(i, i)} p_d^{u-v}(l, j) \right) dv \right) \\
 &= e^{\lambda(i,i)u} \left( \delta_i(j) + \int_0^u \lambda(i, i)e^{-\lambda(i,i)s} \left( \sum_{l \in d^c - \{i\}} \frac{\lambda(i, l)}{\lambda(i, i)} p_d^s(l, j) \right) ds \right).
 \end{aligned}
 \tag{10}$$

Differentiating the last equality with respect to  $u$  gives, for  $i, j \in d^c$ ,

$$\frac{dp_d^u(i, j)}{du} = \left( \sum_{l \in d^c} \lambda(i, l)p_d^u(l, j) \right).$$

---

\*See Remark 2.1.

Let us rewrite the last display in matrix form:

$$\frac{dp_d^u}{du} = (\boldsymbol{\lambda}|_{d^c}) p_d^u. \tag{11}$$

Setting  $u = 0$  in (10) gives the initial condition

$$p_d^0 = \mathbf{I}|_{d^c}. \tag{12}$$

The unique solution of the constant coefficient ODE (11) with initial condition (12) is  $p_d^u = \mathbf{p}_d^u|_{d^c} = e^{u\boldsymbol{\lambda}|_{d^c}} = e^{u\boldsymbol{\lambda}(d^c)}|_{d^c}$ . Equality (9) follows from this and  $\alpha|_d = 0$ .  $\square$

**Remark 3.1** Probabilities that concern sample paths that stay away from a given set are called “taboo probabilities” in [23, Section 1.2]. [23, Equation (F), page 28] is equivalent to (11).

The next corollary is a natural extension of the tail probability formulas in [5]:

**Corollary 3.1** For  $a \subset E$ ,  $\tau_a := \inf\{u > 0 : X_u \in a\}$ , and an initial distribution with  $\alpha|_a = 0$ ,

$$P_\alpha(\tau_a > u) = \alpha e^{u\boldsymbol{\lambda}(a^c)} \mathbf{1}, \tag{13}$$

where  $\mathbf{1}$  denotes the  $|E|$ -dimensional column vector with all components equal to 1.

**Proof**  $P_\alpha(\tau_a > u) = \sum_{j \in a^c} P_\alpha(X_u = j, X_v \notin a, \forall v \leq u) = \alpha e^{u\boldsymbol{\lambda}(a^c)} \mathbf{1}$ , where the last equality is implied by (9).  $\square$

**Remark 3.2** One must modify (13) to  $P_\alpha(\tau_a > u) = \alpha \mathbf{I}_{a^c} e^{u\boldsymbol{\lambda}(a^c)} \mathbf{1}$ ,  $P_\alpha(\tau_a = 0) = \alpha \mathbf{I}_a \mathbf{1}$  if one does not assume  $\alpha|_a = 0$ .

Once  $P_\alpha(\tau_a > u)$  is known, one can differentiate it to compute the density of  $\tau_a$ . This is the main method of derivation and proof in the prior literature on phase-type distributions that we are aware of. We have already noted that this approach is problematic when the distribution of  $\tau$  has singular parts. Instead, we will directly focus on the computation of the densities. The next proposition does so for the case of a single hitting time  $\tau_a$ . The proposition allows also to specify a subset  $b \subset E$  that the process is required to stay away from before the hitting time; this generalization is useful in extending the theorem to multiple hitting times (see the next subsection).

**Proposition 3.1** Let  $a, b \subset E$ ,  $a \cap b = \emptyset$  be given. Define  $\tau_a := \inf\{u > 0 : X_u \in a\}$  and set  $d = a \cup b$ . Then

$$\frac{d}{du} [P_\alpha(\tau_a \in (0, u], X_v \in d^c, \forall v < \tau_a)] = \alpha e^{u\boldsymbol{\lambda}(d^c)} \boldsymbol{\lambda}(d^c, a) \mathbf{1}, \tag{14}$$

where  $\alpha$  is the initial distribution of  $X$  with  $\alpha|_d = 0$ .

In other words, the density of  $\tau_a$  on the set  $\{X_v \in d^c, \forall v < \tau_a\}$  is given by the right side of (14).



**Proof** [Proof of Proposition 3.1] Define

$$\hat{p}(l, j, h) := P_l(X_{\tau_a} = j, \tau_a < h) \tag{15}$$

for  $l \notin a$ , and  $j \in a$ . Let, as before,  $\nu_1$  denote the first jump time of the process  $X$ . Then

$$\begin{aligned} \hat{p}(l, j, h) &= P_l(X_{\tau_a} = j, \tau_a < h, \tau_a = \nu_1) + P_l(X_{\tau_a} = j, \tau_a < h, \tau_a > \nu_1) \\ &= P_l(X_{\nu_1} = j, \tau_{\nu_1} < h) + P_l(X_{\tau_a} = j, \tau_a < h, \tau_a > \nu_1). \end{aligned}$$

The explicit joint distribution of  $(\nu_1, X_{\nu_1})$  and the strong Markov property of  $X$  give

$$\hat{p}(l, j, h) = \int_0^h \lambda(l, l) e^{\lambda(l, l)s} \left( \frac{\lambda(l, j)}{\lambda(l, l)} + \sum_{k \notin a} \frac{\lambda(l, k)}{\lambda(l, l)} \hat{p}(k, j, h - s) \right) ds.$$

This and the fundamental theorem of calculus imply

$$\frac{d\hat{p}(l, j, h)}{dh}(0) = \lambda(l, j). \tag{16}$$

The Markov property of  $X$  (applied at time  $u$ ) gives

$$\begin{aligned} P_i(X_{\tau_a} = j, \tau_a \in (u, u + h], X_v \in d^c, \forall v \leq u) &= \mathbb{E}_i [P(X_{\tau_a} = j, \tau_a \in (u, u + h], X_v \in d^c, \forall v \leq u | \mathcal{F}_u)] \\ &= \mathbb{E}_i [1_{\{X_v \in d^c, v \leq u\}} P(X_{\tau_a} = j, \tau_a \in (u, u + h] | \mathcal{F}_u)] \\ &= \mathbb{E}_i [1_{\{X_v \in d^c, v \leq u\}} P_{X_u}(X_{\tau_a} = j, \tau_a \leq h)] \\ &= \sum_{l \in d^c} p_d^u(i, l) \hat{p}(l, j, h). \end{aligned} \tag{17}$$

Then

$$P_\alpha(\tau_a \in (u, u + h], X_v \in d^c, \forall v < \tau_a) = \sum_{i \in E} \sum_{l \in d^c} \sum_{j \in a} \alpha(i) p_d^u(i, l) \hat{p}(l, j, h).$$

The derivative with respect to  $u$  appearing in (14) equals the derivative of the last display with respect to  $h$  evaluated at 0, which equals

$$\frac{d}{dh} [P_\alpha(\tau_a \in (u, u + h], X_v \in d^c, \forall v < \tau_a)](0) = \sum_{i \in E} \sum_{l \in d^c} \sum_{j \in a} \alpha(i) p_d^u(i, l) \frac{d}{dh} [\hat{p}(l, j, h)](0),$$

which, by (16), equals

$$\sum_{i \in E} \sum_{l \in d^c} \sum_{j \in a} \alpha(i) p_d^u(i, l) \lambda(l, j).$$

By the previous lemma,  $p_d^u(i, l)$  equals the  $(i, l)^{th}$  component of  $e^{u\lambda(d^c)}$ ; substituting this in the last display gives

$$\frac{d}{dh} [P_\alpha(\tau_a \in (u, u + h], X_v \in d^c, \forall v < \tau_a)](0) = \alpha e^{u\lambda(d^c)} \lambda(d^c, a) \mathbf{1},$$

where we have also used  $\alpha(i) = 0$  for  $i \in d$ . This proves (14). □

Setting  $b = \emptyset$  in Proposition 3.1 we get the density of  $\tau_a$ . The following result will be needed in the proof of Theorem 3.1.

**Proposition 3.2** Let  $a, b \subset E$ ,  $a \cap b = \emptyset$ . Define  $\tau_a := \inf\{u : X_u \in a\}$  and  $d = a \cup b$ . Let  $\alpha$  be a probability measure on  $E$ , with  $\alpha|_d = 0$ . Set  $\alpha_1 := \alpha e^{\tau_a \lambda(d^c)} \lambda(d^c, a)$  and  $\mathcal{V} := \{X_v \notin b, \forall v \leq \tau_a\}$ . Then

$$P_\alpha(X_{\tau_a} = j | (\tau_a, 1_{\mathcal{V}})) = (\alpha_1(j) / \alpha_1 \mathbf{1}) \text{ on } \mathcal{V}, \tag{18}$$

where  $1_{\mathcal{V}}$  is the indicator function of the event  $\mathcal{V}$ .

Note that  $\mathcal{V}$  is the event that  $X$  does not visit the set  $b$  before time  $\tau_a$ .

**Proof** The equality (17) implies

$$P_\alpha(X_{\tau_a} = j, \tau_a \in (u, u + h], X_v \in d^c, \forall v \leq u) = \sum_{i \in E} \sum_{l \in d^c} \alpha(i) p_d^u(i, l) \hat{p}(l, j, h).$$

Differentiating the last display with respect to  $h$  and evaluating it at  $h = 0$  gives the density of  $\tau_a$  over the event  $\mathcal{V} \cap \{X_{\tau_a} = j\}$ :

$$f_{\tau_a, j}(u) := \sum_{i \in E} \sum_{l \in d^c} \alpha(i) p_d^u(i, l) \lambda(l, j) = \alpha e^{u \lambda(d^c)} \lambda(d^c, j),$$

where we have used (16) and (9). Summing the last densities over  $j$  gives the density of  $\tau_a$  over the event  $\mathcal{V}$  (already computed in the previous proposition):

$$f_{\tau_a}(u) := \alpha e^{u \lambda(d^c)} \lambda(d^c, a) \mathbf{1}.$$

Then, for any Borel set  $B \subset \mathbb{R}$ ,

$$\begin{aligned} P_\alpha(X_{\tau_a} = j, \tau_a \in B, \mathcal{V}) &= \int_B f_{\tau_a, j}(u) du = \int_B \frac{f_{\tau_a, j}(u)}{f_{\tau_a}(u)} f_{\tau_a}(u) du \\ &= \mathbb{E}_\alpha \left[ 1_B(\tau_a) 1_{\mathcal{V}} \frac{\alpha e^{\tau_a \lambda(d^c)} \lambda(d^c, j)}{\alpha_1 \mathbf{1}} \right] \\ &= \mathbb{E}_\alpha \left[ 1_B(\tau_a) 1_{\mathcal{V}} \frac{\alpha_1(j)}{\alpha_1 \mathbf{1}} \right]; \end{aligned}$$

this and the definition of the conditional expectation imply (18). □

### 3.2. The multidimensional density

One can extend Proposition 3.1 to a representation of the distribution of  $\tau$  using the subpartition notation of Section 2.2 as follows. For a partition  $s$  of  $K$ , an integer  $n \in \{1, 2, \dots, |s|\}$ , and a vector  $\mathbf{t} \in R_s \subset \mathbb{R}_+^{|K|}$ , define the real numbers and the sets

$$\bar{t}_n := t_{s(n,1)}, \quad \bar{t}_0 := 0, \quad W_n := [S(L^{n-1}s)]^c, T_n := \left[ \bigcap_{k \in s(n)} \Gamma_k \right] \cap W_{n+1}, \tag{19}$$

where  $W$  stands for “waiting” and  $T$  for “target” and  $S(\cdot)$  is as defined in (5). In particular,  $W_1 = [S(L^0s)]^c = [S(s)]^c = [\bigcup_{k \in K} \Gamma_k]^c = E - \bigcup_{k \in K} \Gamma_k$  and  $W_{|s|+1} = [S(L^{|s|}s)]^c = E$  (which follows from  $S(\emptyset) = \emptyset$ ).

**Example 3.1** Suppose  $E = \{a, b, c, d, e, f, g, h, i\}$ ,  $\Gamma_1 = \{a, b\}$ ,  $\Gamma_2 = \{b, c, i\}$ ,  $\Gamma_3 = \{c, d, e\}$ , and  $\Gamma_4 = \{a, e, f\}$ ; consider  $\mathbf{t} = (3.1, 2.7, 1.2, 3.1) \in R_s \subset \mathbb{R}^4$  with  $s = (\{3\}, \{2\}, \{1, 4\})$ . For  $\boldsymbol{\tau} = \mathbf{t}$ , the process enters first  $\Gamma_3$  at time 1.2, then  $\Gamma_2$  at time 2.7 and finally  $\Gamma_1$  and  $\Gamma_4$  at time  $t = 3.1$ . We have  $\bar{t}_1 = 1.2$ ,  $\bar{t}_2 = 2.7$ ,  $\bar{t}_3 = 3.1$  with  $W_1 = \{g, h\}$ ,  $T_1 = \{d\}$ ,  $W_2 = \{d, g, h\}$ ,  $T_2 = \{c, i\}$ ,  $W_3 = \{c, d, g, h, i\}$ , and  $T_3 = \{a\}$ .

The idea of the density formula and its proof is the  $|s|$  step version of the one in Proposition 3.1: in order for  $\boldsymbol{\tau} = \mathbf{t} \in \mathbb{R}_+^{|K|}$ ,  $X$  has to stay in set  $W_1$  until time  $\bar{t}_1$  and jump exactly at that time into  $T_1 \subset W_2$ , then stay in set  $W_2$  until time  $\bar{t}_2$  and jump exactly then into  $T_2$  and so on until all of the pairs  $(W_n, T_n)$ ,  $n \leq |s|$ , are exhausted.

The definitions above depend on the collection  $\{\Gamma_k, k \in K\}$ . We will express this dependence explicitly in the following theorem by including the index set  $K$  as a variable of the density function  $f$ . This will be useful in the proof of the theorem in the next subsection where we comment on the case when  $\alpha$  may put nonzero mass on  $\cup_{k \in K} \Gamma_k$  and in Proposition 3.3. For a sequence  $\mathbf{M}_1, \mathbf{M}_2, \dots, \mathbf{M}_n$  of square matrices of the same size  $\prod_{m=1}^n \mathbf{M}_m$  will mean  $\mathbf{M}_1 \mathbf{M}_2 \cdots \mathbf{M}_n$ .

**Theorem 3.1** For any partition  $s \in \mathcal{S}$ , the distribution of  $\boldsymbol{\tau}$  on the set  $R_s$  has density

$$f_s(\alpha, \mathbf{t}, K) = \alpha \left( \prod_{n=1}^{|s|} e^{\lambda(W_n)(\bar{t}_n - \bar{t}_{n-1})} \lambda(W_n, T_n) \right) \mathbf{1}, \quad \mathbf{t} \in R_s \tag{20}$$

with respect to the  $|s|$ -dimensional Lebesgue measure on  $R_s$ , i.e. for any bounded measurable function  $g : \mathbb{R}^{|K|} \rightarrow \mathbb{R}$

$$\mathbb{E}[1_{R_s}(\boldsymbol{\tau})g(\boldsymbol{\tau})] = \int_{R_s} g(\mathbf{t})f_s(\alpha, \mathbf{t}, K)d_s \mathbf{t} \tag{21}$$

holds, where  $d_s \mathbf{t}$  denotes the  $|s|$ -dimensional Lebesgue measure on  $R_s$ .

**Proof** The proof will use induction on  $|K|$ . For  $|K| = 1$ , (20) is the same as (14) with  $b = \emptyset$ . Let  $\kappa > 1$  and suppose that (20) holds for all  $K$  with  $|K| \leq \kappa - 1$ ; we will now argue that (20) also holds for all  $K$  with  $|K| = \kappa$ . Fix a set  $K$  with  $|K| = \kappa$  and a partition  $s$  of  $K$ ; we will show that (21) holds, which will prove that the distribution of  $\boldsymbol{\tau}$  restricted to  $R_s$  has the density (20).

Define the stopping time  $\vartheta := \min\{\tau_k, k \in K\}$ , i.e.  $\vartheta$  is the first time  $X$  enters the set  $\cup_{k \in K} \Gamma_k$ . In the rest of the proof we will assume  $P_\alpha(\vartheta < \infty) = 1$ ; the treatment of the case  $P_\alpha(\vartheta = \infty) > 0$  needs no new ideas and the following argument can be extended to handle it by adding several case by case comments. On the set  $\{\boldsymbol{\tau} \in R_s\}$  the following conditions hold:

$$X_\vartheta \in T_1 \text{ and } X_u \in W_1, \forall u < \vartheta.$$

These imply that the equality  $\vartheta = \tau_{s(1,1)}$  holds on the same set. Therefore,

$$\{\boldsymbol{\tau} \in R_s\} \subset \mathcal{W}_1 := \{X_u \in W_1, \forall u < \vartheta\} \cap \{X_\vartheta \in T_1\}. \tag{22}$$

Proposition 3.1 implies that if  $\lambda(W_1, T_1)$  is zero, then  $\mathcal{W}_1$  has probability zero. Thus, (22) implies that if  $\lambda(W_1, T_1)$  is zero then  $P_\alpha(\boldsymbol{\tau} \in R_s) = 0$  and, indeed,  $f_s(\alpha, \mathbf{t}, K) = 0$  is the density of  $\boldsymbol{\tau}$  on  $R_s$ . From here on, we will treat the case when  $\lambda(W_1, T_1)$  is nonzero.

Next, define the process  $\widehat{X}$  by  $\widehat{X}_u := X_{u+\vartheta}$ ,  $u \geq 0$ , and  $\widehat{\tau} = (\widehat{\tau}_k, k \in S(Ls))$  where  $\widehat{\tau}_k := \inf\{u : \widehat{X}_u \in \Gamma_k\}$ ;  $\widehat{X}$  is the trajectory of  $X$  after time  $\vartheta$ . The strong Markov property of  $X$  implies that  $\widehat{X}$  is a Markov process with intensity matrix  $\lambda$  and starting from  $\widehat{X}_0 = X_\vartheta$ . This and (22) imply

$$\widehat{\tau} = \tau|_{Ls} - \vartheta, \tag{23}$$

where  $\tau|_{Ls}$  is defined in accordance with (2). Finally, the definition of  $\widehat{\tau}$  and that of  $\mathcal{W}_1$  imply

$$\{\tau \in R_s\} = \mathcal{W}_1 \cap \{\widehat{\tau} \in R_{Ls}\}. \tag{24}$$

In words, this equality says that for  $\tau$  to be partitioned according to  $s$ , among all  $\{\Gamma_k\}$ ,  $X$  must visit  $T_1$  first (which ensures that all  $\Gamma_k$ , for  $k \in s(1)$ , are visited at the same time and before the rest of the  $\Gamma_j$ ,  $j \notin s(1)$ ) and after this visit the rest of the hitting times must be arranged according to the partition  $Ls$ .

Denote by  $\mathbb{1}$  the function that maps all elements of  $K$  to 1. Define  $\widehat{g} : \mathbb{R}_+ \times \mathbb{R}_+^{S(Ls)} \rightarrow \mathbb{R}$  as

$$\widehat{g}(u, \widehat{\mathbf{t}}) := g\left(u\mathbb{1} + \widehat{\mathbf{t}}|^{S(s)}\right),$$

where we used the vector extension notation of (4). Equalities (23) and (24) imply

$$\begin{aligned} \mathbb{E}[1_{R_s}(\tau)g(\tau)] &= \mathbb{E}[1_{\mathcal{W}_1}1_{R_{Ls}}(\widehat{\tau})\widehat{g}(\vartheta, \widehat{\tau})] \\ &= \mathbb{E}[\mathbb{E}[1_{\mathcal{W}_1}1_{R_{Ls}}(\widehat{\tau})\widehat{g}(\vartheta, \widehat{\tau})|\mathcal{F}_\vartheta]] \\ &= \mathbb{E}[1_{\mathcal{W}_1}\mathbb{E}[1_{R_{Ls}}(\widehat{\tau})\widehat{g}(\vartheta, \widehat{\tau})|\mathcal{F}_\vartheta]], \end{aligned} \tag{25}$$

where, for the last equality, we used the fact that the set  $\mathcal{W}_1$  is  $\mathcal{F}_\vartheta$  measurable. Property 5A in [8, page 98] implies

$$\mathbb{E}[1_{R_{Ls}}(\widehat{\tau})\widehat{g}(\vartheta, \widehat{\tau})|\mathcal{F}_\vartheta] = h(\vartheta)$$

where

$$h(u) := \mathbb{E}[1_{R_{Ls}}(\widehat{\tau})\widehat{g}(u, \widehat{\tau})|\mathcal{F}_u]. \tag{26}$$

The strong Markov property of  $X$  and the definition of  $\widehat{X}$  imply

$$h(u) = \mathbb{E}[1_{R_{Ls}}(\widehat{\tau})\widehat{g}(u, \widehat{\tau})|X_u] = \mathbb{E}[1_{R_{Ls}}(\widehat{\tau})\widehat{g}(u, \widehat{\tau})|\widehat{X}_0].$$

The random variable  $\widehat{X}_0$  takes values in a finite set and therefore one can compute the conditional expectation  $\mathbb{E}[1_{R_{Ls}}(\widehat{\tau})\widehat{g}(u, \widehat{\tau})|\widehat{X}_0]$  by conditioning on each of these values separately. Since  $\widehat{X}$  is a Markov process with initial value  $\widehat{X}_0$  and intensity matrix  $\lambda$ , one can invoke the induction hypothesis for the set  $K - s(1)$  to conclude that, on the set  $\{\widehat{X}_0 = j\}$ ,

$$h(u) = \mathbb{E}[1_{R_{Ls}}(\widehat{\tau})\widehat{g}(u, \widehat{\tau})|\widehat{X}_0 = j] = \int_{R_{Ls}} f_{Ls}(\delta_j, \widehat{\mathbf{t}}, K - s(1))g(u, \widehat{\mathbf{t}})d_{Ls}\widehat{\mathbf{t}} \tag{27}$$

where  $f_{Ls}$  is given as in (20) with  $s$  changed to  $Ls$  and  $K$  changed to  $K - s(1)$ . Once we substitute (27) in (25) we get an expectation involving only three random variables:  $\vartheta$ ,  $1_A$ , and  $\widehat{X}_0 = X_\vartheta$ , where

$A = \{X_u \in W_1, u < \vartheta\}$ . Proposition 3.1 implies that the density of  $\vartheta$  on the set  $A$  is  $\alpha e^{\lambda(W_1)\bar{t}_1} \lambda(W_1, T_1) \mathbf{1}$ , and Proposition 3.2 implies that the law of  $\widehat{X}_0$  conditioned on  $\vartheta$  and  $1_{W_1}$  is

$$\frac{\alpha e^{\lambda(W_1)\vartheta} \lambda(W_1, T_1)}{\alpha e^{\lambda(W_1)\vartheta} \lambda(W_1, T_1) \mathbf{1}}. \tag{28}$$

Then, starting from (25),

$$\begin{aligned} \mathbb{E}_\alpha[1_{R_s}(\boldsymbol{\tau})g(\boldsymbol{\tau})] &= \mathbb{E}_\alpha[1_{W_1}\mathbb{E}_\alpha[1_{R_{L_s}}(\widehat{\boldsymbol{\tau}})\widehat{g}(\vartheta, \widehat{\boldsymbol{\tau}})|\mathcal{F}_\vartheta]] = \mathbb{E}_\alpha[1_{W_1}h(\vartheta)] \\ &= \mathbb{E}_\alpha \left[ 1_{W_1} \int_{R_{L_s}} f_{L_s}(\delta_{X_\vartheta}, \widehat{\mathbf{t}}, K - s(1)) \widehat{g}(\vartheta, \widehat{\mathbf{t}}) d_{L_s} \widehat{\mathbf{t}} \right] \\ &= \mathbb{E}_\alpha \left[ 1_{W_1} \int_{R_{L_s}} \left( \delta_{X_\vartheta} \prod_{n=2}^{|\mathbf{s}|} e^{\lambda(W_n)(\bar{t}_n - \bar{t}_{n-1})} \lambda(W_n, T_n) \mathbf{1} \right) \widehat{g}(\vartheta, \widehat{\mathbf{t}}) d_{L_s} \widehat{\mathbf{t}} \right] \\ &= \int_0^\infty \int_{R_{L_s}} \sum_{j \in E} \left[ \alpha e^{\lambda(W_1)\bar{t}_1} \lambda(W_1, T_1) \right] (j) \\ &\quad \times \left( \delta_j \prod_{n=2}^{|\mathbf{s}|} e^{\lambda(W_n)(\bar{t}_n - \bar{t}_{n-1})} \lambda(W_n, T_n) \mathbf{1} \right) \widehat{g}(\bar{t}_1, \widehat{\mathbf{t}}) d_{L_s} \widehat{\mathbf{t}} d\bar{t}_1 \\ &= \int_0^\infty \int_{R_{L_s}} \left( \alpha \prod_{n=1}^{|\mathbf{s}|} e^{\lambda(W_n)(\bar{t}_n - \bar{t}_{n-1})} \lambda(W_n, T_n) \mathbf{1} \right) \widehat{g}(\bar{t}_1, \widehat{\mathbf{t}}) d_{L_s} \widehat{\mathbf{t}} d\bar{t}_1 \\ &= \int_{R_s} \left( \alpha \prod_{n=1}^{|\mathbf{s}|} e^{\lambda(W_n)(\bar{t}_n - \bar{t}_{n-1})} \lambda(W_n, T_n) \mathbf{1} \right) g(\mathbf{t}) d_s \mathbf{t}, \end{aligned} \tag{29}$$

where in going from line (29) to the one following it, we use the conditional distribution (28) and the density of  $\vartheta$ . The last line above finishes the proof of the induction step and completes the proof of the theorem.  $\square$

The following corollary follows from (21) and the disjoint decomposition  $\mathbb{R}_+^{|\mathbf{K}|} = \bigcup_{s \in \mathcal{S}} R_s$ .

**Corollary 3.2** For any bounded and measurable  $g : \mathbb{R}_+^{|\mathbf{K}|} \rightarrow \mathbb{R}$  we have

$$\mathbb{E}_\alpha[g(\boldsymbol{\tau})] = \sum_{s \in \mathcal{S}} \int_{R_s} g(\mathbf{t}) f_s(\alpha, \mathbf{t}, K) d_s \mathbf{t}.$$

In what follows, to ease exposition, we will sometimes refer to  $f$  as the “density” of  $\boldsymbol{\tau}$  without explicitly mentioning the reference measures  $d_s$ ,  $s \in \mathcal{S}$ .

**Remark 3.3** The first  $\kappa > 0$  jump times of a standard Poisson process with rate  $\lambda \in (0, \infty)$  have the joint density  $\prod_{n=1}^\kappa e^{\lambda(t_n - t_{n-1})} \lambda$ ,  $0 = t_0 < t_1 < t_2 < \dots < t_\kappa$ . Similarly, the first  $\kappa > 0$  jump times of a Markov arrival process with intensity matrix  $C + D$  (where  $C$  [ $D$ ] is the matrix of transition intensities with [without] arrivals) have joint density  $\boldsymbol{\alpha} \left( \prod_{n=1}^\kappa e^{D(t_n - t_{n-1})} C \right) \mathbf{1}$ , see [4] or [2, page 304]. The density (20) can also be interpreted as a generalization of these formulas.

**3.3. When  $\alpha$  puts positive mass on  $\cup_k \Gamma_k$**

If  $\alpha$  puts positive mass on  $\gamma := \cup_{k \in K} \Gamma_k$ , one best describes the law of  $\tau$  proceeding as follows. Define  $\bar{\alpha}' := 1 - \sum_{i \in \gamma} \alpha(i) = \sum_{i \in \gamma^c} \alpha(i)$ , and, for  $\bar{\alpha}' > 0$ :

$$\alpha' := \frac{1}{\bar{\alpha}'} (\alpha - \sum_{i \in \gamma} \alpha(i) \delta_i) = \frac{1}{\bar{\alpha}'} \sum_{i \in \gamma^c} \alpha(i) \delta_i;$$

note that  $\bar{\alpha}'$  is the measure of  $\gamma^c$  under  $\alpha$  and,  $\alpha'$ , whenever defined, is just the measure  $\alpha$  conditioned on  $\gamma^c$ .

First consider the case when  $\bar{\alpha}' > 0$ . The foregoing definitions imply

$$P_\alpha(\tau \in U) = \bar{\alpha}' P_{\alpha'}(\tau \in U) + \sum_{i \in \gamma} \alpha(i) P_i(\tau \in U) \tag{30}$$

for any Borel set  $U \subset \mathbb{R}_+^{|K|}$ . By its definition  $\alpha'$  puts no mass on  $\gamma = \cup_{k \in K} \Gamma_k$  and therefore Theorem 3.1 is applicable and  $f(\alpha', \cdot, K)$  is the density of the probability measure  $P_{\alpha'}(\tau \in \cdot)$ . For the second summand of (30), it is enough to compute each  $P_i(\tau \in U)$  separately. Define  $K_i := \{k : i \in \Gamma_k\}$ ,  $U_i := \{\mathbf{t} : \mathbf{t} \in U, t_k = 0, k \in K_i\}$ ,  $\bar{U}_i := \{\mathbf{t}|_{K_i^c} : \mathbf{t} \in U_i\}$ . Now remember that  $i \in \gamma$ ; thus, if  $i \in \Gamma_k$  then  $\tau_k = 0$  under  $P_i$ , and therefore  $P_i(\tau \in U) = P_i(\tau \in U_i)$ . For  $\tau \in U_i$ , the stopping times  $\tau|_{K_i}$  are all 0. Thus, to compute  $P_i(\tau \in U_i)$  it suffices to compute  $P_i(\tau|_{K_i^c} \in \bar{U}_i)$ , but by definition  $i \notin \cup_{k \in K_i^c} \Gamma_k$  and once again Theorem 3.1 is applicable and gives the density of  $\tau|_{K_i^c}$  under  $P_i$  as  $f(\delta_i, \cdot, K_i^c)$ .

If  $\bar{\alpha}' = 0$ , then

$$P_\alpha(\tau \in U) = \sum_{i \in \gamma} \alpha(i) P_i(\tau \in U)$$

and the computation of  $P_i(\tau \in U)$  goes as above.

**3.4. Tail probabilities of  $\tau$**

Probabilities of tail events have representations as integrals of densities given in Theorem 3.1 over appropriate subsets of  $\mathbb{R}_+^{|K|}$ . In the present subsection, we derive a simple recursive representation of these integrals that uses a version of the density formula and the ideas used in its derivation.

By tail probabilities we mean probabilities of sets of the form  $\{\tau_2 = \tau_4 > t_1, \tau_3 > t_2, \tau_1 = \tau_5 > t_2, \tau_3 \neq \tau_2, \tau_1 \neq \tau_3\}$ , or more generally

$$\left( \bigcap_{n=1}^{|s|} \bigcap_{k_1, k_2 \in s(n)} \{\tau_{k_1} = \tau_{k_2}\} \cap \{\tau_{s(n,1)} > t_n\} \right) \cap \bigcap_{n_1 \neq n_2, n_1, n_2 \leq |s|} \{\tau_{s(n_1,1)} \neq \tau_{s(n_2,1)}\}, \tag{31}$$

where  $s$  is a partition of  $K$ , and  $\mathbf{t} \in \mathbb{R}_+^{|s|}$  is such that  $t_n < t_{n+1}$ . In (31) all equality and inequality conditions are explicitly specified. One can write standard tail events in terms of these, e.g.,  $\{\tau_1 > t_1\} \cap \{\tau_2 > t_2\}$  is the same as the disjoint union

$$(\{\tau_1 > t_1, \tau_2 > t_2\} \cap \{\tau_1 \neq \tau_2\}) \cup \{\tau_1 = \tau_2 > \max(t_1, t_2)\}.$$

Both of these sets are of the form (31). Thus, it is enough to be able to compute probabilities of events of the form (31).

**Remark 3.4** From here on, to keep the notation short, we will assume that, over tail events, unless explicitly stated with an equality condition, all stopping times appearing in them are distinct from each other (therefore, when writing formulas, we will omit the last intersection in (31)).

A tail event of the form (31) consists of a sequence of constraints of the form

$$\{\tau_{s(n,1)} = \tau_{s(n,2)} = \dots = \tau_{s(n,|s(n)|)} > t_n\}.$$

There are two types of subconstraints involved here: that entrances to all  $\Gamma_k$ ,  $k \in s(n)$ , happen at the same time and that this event occurs after time  $t_n$ . Keeping track of these constraints as they evolve in time requires that we introduce yet another class of events that generalize (31). For two disjoint subpartitions  $s_1$  and  $s_2$  of  $K$  and an element  $\mathbf{t} \in \mathbb{R}_+^{|s_1|}$  such that  $t_{|s_1|} > t_{|s_1|-1} > \dots > t_2 > t_1$  (by convention,  $\mathbf{t} = 0$  if  $|s_1| = 0$ ) define

$$\begin{aligned} \mathcal{T}(s_1, s_2, \mathbf{t}) := & \left( \bigcap_{n=1}^{|s_1|} \bigcap_{k_1, k_2 \in s_1(n)} \{\tau_{k_1} = \tau_{k_2}\} \cap \{\tau_{s_1(n,1)} > t_n\} \right) \cap \\ & \bigcap_{m=1}^{|s_2|} \bigcap_{\ell_1, \ell_2 \in s_2(m)} \{\tau_{\ell_1} = \tau_{\ell_2}\}. \end{aligned} \tag{32}$$

In view of Remark 3.4, setting  $s_1 = s$  and  $s_2 = \emptyset$  reduces (32) to (31). The indices in  $s_1$  appear both in equality constraints and time constraints while indices in  $s_2$  appear only in equality constraints.

**Remark 3.5** The definition (32) implies that if a component of  $s_2$  has only a single element, that component has no influence on  $\mathcal{T}(s_1, s_2, \mathbf{t})$ . For example,  $\mathcal{T}(s_1, (\{1\}, \{2, 3\}), \mathbf{t})$  is the same as  $\mathcal{T}(s_1, (\{2, 3\}), \mathbf{t})$ .

To express  $P_\alpha(\mathcal{T}(s_1, s_2, \mathbf{t}))$ , we will define a collection of functions  $p_i$ ,  $i \in E$ , of  $s_1$ ,  $s_2$ , and  $\mathbf{t}$ . We will denote by  $\mathbf{p}$  the column vector with components  $p_i, i \in E$ .

For  $s_1 = \emptyset$  and  $i \in E$ , define  $p_i(\emptyset, s_2, 0) := P_i(\mathcal{T}(\emptyset, s_2, 0))$ . If  $s_2$  is empty or if it consists of components with single elements, then the definitions of  $\mathbf{p}$  and  $\mathcal{T}$  and Remark 3.5 imply

$$\mathbf{p}(\emptyset, s_2, 0) = \mathbf{1}. \tag{33}$$

For a given disjoint pair of subpartitions  $s_1$  and  $s_2$ , define

$$T_n(s_1, s_2) := \bigcap_{k \in s_2(n)} \Gamma_k - S(s_1 \cup s_2 - s_2(n)), \quad T(s_1, s_2) := \bigcup_{n=1}^{|s_2|} T_n(s_1, s_2).$$

If  $s_1 \neq \emptyset$ , define  $\mathbf{p}$  recursively as

$$\begin{aligned} \mathbf{p}(s_1, s_2, \mathbf{t}) := & \int_0^{t_1} e^{u\lambda(W)} \boldsymbol{\lambda}(W, T(s_1, s_2)) \left( \sum_{n=1}^{|s_2|} \mathbf{I}_{T_n(s_1, s_2)} \mathbf{p}(s_1, s_2 - s_2(n), \mathbf{t} - u) \right) du \\ & + e^{t_1\lambda(W)} \mathbf{p}(Ls_1, s_2 + s_1(1), L\mathbf{t} - t_1), \end{aligned} \tag{34}$$

where  $W = [S(s_1 \cup s_2)]^c$ .

**Theorem 3.2** *Suppose  $E - S(s_1 \cup s_2)$  is not empty and that  $\alpha$  is a probability measure on  $E$  that puts all of its mass on this set. Then*

$$P_\alpha(\mathcal{T}(s_1, s_2, \mathbf{t})) = \alpha \mathbf{p}(s_1, s_2, \mathbf{t}).$$

We omit the proof, which is parallel to that of Theorem 3.1 and proceeds by induction. Theorem 3.2 holds for all finite state Markov processes and does not require that any of the  $\{\Gamma_k\}$  be absorbing. The evaluations of  $\mathbf{p}$  on the right side of the recursion (34) will have smaller subpartitions in its arguments; then, in a finite number of steps, these recursions will lead to an evaluation of  $\mathbf{p}$  with  $s_1 = \emptyset$ .

Note that (34) reduces to

$$\mathbf{p}(s_1, \emptyset, \mathbf{t}) = e^{\lambda(S(s_1)^c)t_1} \mathbf{p}(Ls_1, s_1(1), Lt - t_1), \tag{35}$$

if  $s_2 = \emptyset$ .

When  $s_1$  has no equality constraints and  $s_2 = \emptyset$ , one can invoke (35)  $|s_1|$  times along with Remark 3.5 and (33) and get:

**Corollary 3.3** *Let  $\alpha$  be as in Theorem 3.2. If  $|s_1| > 0$  equals the dimension of  $\mathbf{t}$  (in particular, there are no equality constraints), then*

$$P_\alpha(\mathcal{T}(s_1, \emptyset, \mathbf{t})) = \alpha \mathbf{p}(s_1, \emptyset, \mathbf{t}) = \alpha \left( \prod_{n=1}^{|s_1|} e^{\lambda(W_n)(t_n - t_{n-1})} \right) \mathbf{1} \tag{36}$$

where  $W_n = [S(L^{n-1}(s_1))]^c$ .

The formula (36) is a generalization of [5, equation (7)] to general finite state Markov processes.

### 3.5. Conditional formulas

Let us next derive formulas for the conditional density of  $\tau$  given  $\mathcal{F}_{u_0}$  where  $u_0 > 0$  is a fixed deterministic point in time. Introduce the set valued process

$$V_u := \{k \in K, \tau_k < u\}.$$

The set  $K$  is finite and then so is its power set  $2^K$ ; thus,  $V_u$  takes values in a finite set. The set  $V_u$  is the collection of  $\Gamma_k$  that  $X$  has visited up to time  $u$ . We will denote the complement of  $V_u$  in  $K$  by  $V_u^c$ . The times  $\tau|_{V_{u_0}}$  are known by time  $u_0$  and hence they are constant given  $\mathcal{F}_{u_0}$ . Thus, to write the conditional density of  $\tau$  given  $\mathcal{F}_{u_0}$  it suffices to write down the regular conditional density of  $\tau|_{V_{u_0}^c}$ , i.e. of the hitting times to the  $\Gamma_k$  that have not been visited by time  $u_0$ . From here on the idea is the same as in the proof of Theorem 3.1. Define  $\widehat{X}_u := X_{u+u_0}$  and for  $k \in V_{u_0}^c$

$$\widehat{\tau}_k := \inf\{u \in (0, \infty) : \widehat{X}_u \in \Gamma_k\}.$$

The definitions of  $\widehat{X}$  and  $\widehat{\tau}$  imply

$$\widehat{\tau} = \tau|_{V_{u_0}^c} - u_0. \tag{37}$$



The random variable  $\widehat{X}_0 = X_{u_0}$  is a constant given  $\mathcal{F}_{u_0}$ . Thus, the process  $\widehat{X}$  has exactly the same distribution as  $X$  with initial point  $X_{u_0}$  and Theorem 3.1 applies and gives the density of  $\widehat{\tau}$ , which is, by (37), the regular conditional distribution of  $\tau|_{V_{u_0}^c} - u_0$ . Therefore, for any bounded measurable  $g : \mathbb{R}^{V_{u_0}^c} \rightarrow \mathbb{R}$  and a partition  $s'$  of  $V_{u_0}^c$ ,

$$\mathbb{E} \left[ g \left( \tau|_{V_{u_0}^c} \right) 1_{R_{s'}} \left( \tau|_{V_{u_0}^c} \right) | \mathcal{F}_{u_0} \right] = \int_{R_{s'}} g(u_0 + u) f(\delta_{X_{u_0}}, u, V_{u_0}^c) d_{s'} u.$$

We record this as:

**Proposition 3.3** *The regular conditional density of  $\tau|_{V_{u_0}^c} - u_0$  given  $\mathcal{F}_{u_0}$  is  $f(\delta_{X_{u_0}}, \mathbf{t}, V_{u_0}^c)$ .*

#### 4. Absorbing $\{\Gamma_k\}$ and connections with earlier results

The next subsection shows how the formulas in the previous sections can be simplified when  $\Gamma_k$  are absorbing; this assumption is made in [5, 14]. The subsection following it shows how formulas from [5] can be interpreted as special cases of the formulas derived in the present work.

##### 4.1. Density formula for absorbing $\{\Gamma_k\}$

A nonempty subset  $d \subset E$  is said to be *absorbing* if  $\lambda(i, j) = 0$  for all  $i \in d$  and  $j \in d^c$ , i.e. if  $\boldsymbol{\lambda}(d, d^c) = 0$ . Proposition 4.1 gives an alternative expression for the density formula (20) under the assumption that all  $\{\Gamma_k, k \in K\}$  are absorbing. Its proof follows from the next lemma:

**Lemma 4.1** *Let  $\alpha$  be a probability measure on  $E$  and  $d \subset E$ . If  $d$  is absorbing and  $\alpha|_d = 0$ , then*

$$\mathbf{p}_{\alpha, d}^u = \alpha e^{\boldsymbol{\lambda} u} \mathbf{I}_{d^c}, \tag{38}$$

where  $\mathbf{p}_{\alpha, d}^u$  is the row vector whose  $j^{\text{th}}$  component is the probability  $p_{\alpha, d}^u(j) = P_\alpha(X_u = j, X_v \notin d, \forall v \leq u)$  (see (7)).

**Proof** We already know from Lemma 3.1 that  $\mathbf{p}_{\alpha, d}^u = \alpha e^{\boldsymbol{\lambda}(d^c)u}$ ; we would like to show that one has the alternative formula  $\mathbf{p}_{\alpha, d}^u = \alpha e^{\boldsymbol{\lambda} u} \mathbf{I}_{d^c}$  when  $d$  is absorbing and  $\alpha|_d = 0$ . The distribution of  $X$  at time  $u$  is  $\alpha e^{\boldsymbol{\lambda} u}$ , i.e.  $P_\alpha(X_u = j) = \alpha e^{\boldsymbol{\lambda} u}(j)$  for all  $j \in E$ . The fact that  $d$  is absorbing implies that if  $X_{u_0} \in d$  then  $X_u \in d$  for all  $u \geq u_0$ , Therefore, for  $j \in d^c$ ,  $P_\alpha(X_u = j) = P_\alpha(X_u = j, X_v \notin d, \forall v < u)$ , i.e.

$$(\mathbf{p}_{\alpha, d}^u)|_{d^c} = (\alpha e^{\boldsymbol{\lambda} u} \mathbf{I}_{d^c})|_{d^c}. \tag{39}$$

The equality  $\mathbf{p}_{\alpha, d}^u = \alpha e^{\boldsymbol{\lambda}(d^c)u}$  and  $\alpha|_d = 0$  imply  $\mathbf{p}_{\alpha, d}^u|_d = 0$ ; The definition of  $\mathbf{I}_{d^c}$  implies  $(\alpha e^{\boldsymbol{\lambda} u} \mathbf{I}_{d^c})|_d = 0$ . These and (39) imply (38). □

The previous lemma implies that, when all  $\Gamma_k$  are absorbing, one can replace the  $\boldsymbol{\lambda}(W_n)$  appearing in the density formula (20) with  $\boldsymbol{\lambda}$ ; this observation gives the following:

**Proposition 4.1** *Assume that all  $\Gamma_k$ ,  $k \in K$ , are absorbing and let  $\alpha$  be such that  $\alpha|_{W_1^c} = 0$ . Then, for any  $s \in \mathcal{S}$  and  $\mathbf{t} \in R_s$ , the density  $f_s$  given in Theorem 3.1 takes the form*

$$f_s(\alpha, \mathbf{t}, K) = \alpha \left( \prod_{n=1}^{|\mathbf{s}|} e^{\lambda(\bar{t}_n - \bar{t}_{n-1})} \boldsymbol{\lambda}(W_n, T_n) \right) \mathbf{1}. \tag{40}$$

**4.2. Tail probabilities for absorbing  $\{\Gamma_k\}$**

When  $\{\Gamma_k, k \in K\}$  are absorbing, in view of (38), one can write the tail probability that appears in Theorem 3.2 as

$$\begin{aligned} P_\alpha(\mathcal{T}(s_1, s_2, \mathbf{t})) = & \\ & \alpha \int_0^{t_1} e^{\lambda u} \boldsymbol{\lambda}(W, T(s_1, s_2)) \left( \sum_{n=1}^{|\mathbf{s}_2|} \mathbf{I}_{T_n(s_1, s_2)} \mathbf{p}(s_1, s_2 - s_2(n), \mathbf{t} - u) \right) du \\ & + \alpha e^{\lambda t_1} \mathbf{I}_W \mathbf{p}(Ls_1, s_2 + s_1(1), L\mathbf{t} - t_1) \end{aligned}$$

and, in particular,

$$P_\alpha(\mathcal{T}(s_1, \emptyset, \mathbf{t})) = \alpha e^{\lambda t_1} \mathbf{I}_{S(s_1)^c} \mathbf{p}(s_1 - s_1(1), s_1(1), L\mathbf{t} - t_1). \tag{41}$$

**4.3. Connections with earlier results**

This subsection relates the phase-type density/tail probability formulas from [5] to the formulas derived in the present work. In [5], the authors assume that  $E$  has a single absorbing state called  $\Delta$  and they denote by  $A$  what in our paper is denoted by  $\boldsymbol{\lambda}|_{\{\Delta\}^c}$ . Moreover, [5] uses the letter  $\alpha$  to denote the initial distribution of  $X$ , but on the set  $\widehat{E} := E - \{\Delta\}$ , rather than on the set  $E$  as it is done here; in particular, [5] implicitly assumes  $P(X_0 = \Delta) = 0$ . We will use the symbol  $\widehat{\alpha}$  to denote the ‘ $\alpha$  of [5]’. The relation between  $\alpha$  and  $\widehat{\alpha}$  is  $\alpha|_{\{\Delta\}^c} = \widehat{\alpha}$ .

As far as the singular densities/tail probabilities of  $\boldsymbol{\tau}$ , [5] treats only the case of  $|K| = 2$ . Using the notation of that paper, we are given two sets  $\Gamma_1, \Gamma_2 \subset E$  with  $\Gamma_1 \cap \Gamma_2 = \{\Delta\}$ , and  $T_k$  is the first hitting time to  $\Gamma_k$ . The formula [5, Equation (5), page 692] says

$$P_\alpha(T_1 = T_2 > u) = \widehat{\alpha} e^{Au} A^{-1} (Ag_1g_2 - [A, g_1] - [A, g_2])\mathbf{e}, \tag{42}$$

where  $g_k = \mathbf{I}_{\Gamma_k|_{\{\Delta\}^c}}$ ,  $\mathbf{e}$  is the vector with dimension  $|\widehat{E}| = |E| - 1$  with all components equal to 1, and for two matrices  $B$  and  $C$ ,  $[B, C] := BC - CB$ . The absorbing property of  $\Gamma_1$  and  $\Gamma_2$  implies that the matrix inside the parentheses on the right side of (42) equals  $g'A$ , where  $g' = \mathbf{I}_{(\Gamma_1 \cup \Gamma_2)^c|_{\widehat{E}}}$ , i.e. the same matrix as  $A$  except that the rows whose indices appear in  $\Gamma_1 \cup \Gamma_2$  are replaced with 0. Thus,  $(Ag_1g_2 - [A, g_1] - [A, g_2])\mathbf{e}$  is another way to take the  $\Delta$  column of  $\boldsymbol{\lambda}$  and replace its components whose indices appear in  $\Gamma_1 \cup \Gamma_2$  with 0. Denote this vector by  $C_\Delta$ . Then the right side of (42) is

$$\alpha|_{\widehat{E}} (e^{\lambda u}|_{\widehat{E}}) A^{-1} C_\Delta. \tag{43}$$

In the present work, the same probability is expressed by a special case of (41); for the present case one sets  $K = \{1, 2\}$ ,  $s_1 = (\{1, 2\})$ ; for these values, (35) and conditioning on the initial state gives

$$P_\alpha(\tau_1 = \tau_2 > u) = \alpha e^{\lambda u} \mathbf{I}_w \mathbf{p}(\emptyset, (\{1, 2\}), 0), \tag{44}$$

where  $w = (\Gamma_1 \cup \Gamma_2)^c$ . One sees that this is equivalent to (43) as follows. On the right side of the last equality  $s_1 = \emptyset$  and we have no time constraints (the inequality constraints related to  $\mathbf{t}$ ) and thus  $\mathbf{p}(\emptyset, (\{1, 2\}), 0)$  is the probability of the event  $\{\tau_1 = \tau_2\}$ ; the expression following the matrix exponential in (43) represents this probability. Finally, the absorbing property of the underlying chain and  $X_0 \neq \Delta$  imply that we can ignore the restriction to  $\hat{E}$  in (43).

The second density formula from [5] is for the absolutely continuous part of the distribution of  $\tau$ ; [14] makes use of this formula in the following context. The process  $X$  of [14] is a Markov jump process (with absorbing boundary) taking values in  $\mathbb{Z}_2^m := \{0, 1\}^m$  (the  $m$ -fold Cartesian product), with jumps in  $\{-e_k, k = 1, 2, 3, \dots, m\}$ , where  $e_k$  is the unit vector with  $k^{th}$  coordinate equal to 1. In [14], the absorbing sets are denoted as  $\Delta_i$ , see the display after [14, (2.3)], and they correspond to the  $\Gamma_k = \{z \in \mathbb{Z}_2^m : z_k = 0\}$  of our notation. The key property of the setup in [14] is this: take any collection  $\{\Gamma_{k_1}, \Gamma_{k_2}, \dots, \Gamma_{k_n}\}$  with  $n > 1$ ; because the only increments of  $X$  are the  $\{-e_k\}$ , the process cannot enter the sets in the collection at the same time. Thus, in this formulation,  $X$  must hit the  $\{\Gamma_k\}$  at separate times and the distribution of  $\tau$  has no singular part, i.e.  $P(\tau \in R_s) = 0$  for  $|s| < m$ , and one needs only the density of  $\tau$  with respect to the full Lebesgue measure in  $\mathbb{R}^m$  (the ‘‘absolutely continuous part’’); thus, for the purposes of [14], the density of the absolutely continuous part of the distribution of  $\tau$  is sufficient and a formula for this is already available in [5] and is given in [14, display (3.1.1)] as follows:

$$f(\mathbf{t}) = (-1)^m \alpha \left( \prod_{n=1}^{m-1} e^{\lambda(\bar{t}_n - \bar{t}_{n-1})} (\lambda G_{k_n} - G_{k_n} \lambda) \right) e^{\lambda(\bar{t}_m - \bar{t}_{m-1})} \lambda G_{k_m} \mathbf{1}, \tag{45}$$

for  $\mathbf{t} \in R_s$  with  $|s| = m$ ; here  $G_k = I_{\Gamma_k^c}$  and  $k_n$  is the index for which  $t_{k_n} = \bar{t}_n$  ([14] uses the letter  $Q$  for the rate matrix  $\lambda$ ). We have derived the full density formula (40) in the absorbing case in Section 4 describing the density of  $\tau$  over its all possible parts (singular and nonsingular). Arguments similar to those given for the two-dimensional formula can be used to show that (40) reduces to (45) when all components of  $\mathbf{t}$  are distinct.

### 5. Numerical example

The state space of our numerical example is  $E = \mathbb{Z}_3^3$ , where  $\mathbb{Z}_3 = \{0, 1, 2\}$ ; the state space has 27 elements. For  $z \in \mathbb{Z}_3^3$  and  $k \in K = \{1, 2, 3\}$  let  $z_k$  denote the  $k^{th}$  component of  $z$ . For the collection  $\{\Gamma_k\}$  take

$$\Gamma_k = \{z \in E : z_k = 0\}.$$

As before,  $\tau_k$  is the first time the process  $X$  hits the set  $\Gamma_k$ . The initial distribution  $\alpha$  will be the uniform distribution over the set

$$E - \bigcup_{k \in K} \Gamma_k = \left\{ z \in E : \min_{k \in K} z_k > 0 \right\}.$$

We will compute the density of  $\tau = (\tau_1, \tau_2, \tau_3)$  over the sets  $R_{s_1}, R_{s_2} \subset \mathbb{R}_+^3$  defined by the partitions  $s_1 = (\{2, 3\}, \{1\})$  and  $s_2 = (\{1, 2, 3\})$ ; the first corresponds to the event  $\{\tau \in R_{s_1}\} = \{\tau_2 = \tau_3 < \tau_1\}$  and the second to  $\{\tau \in R_{s_2}\} = \{\tau_1 = \tau_2 = \tau_3\}$ .

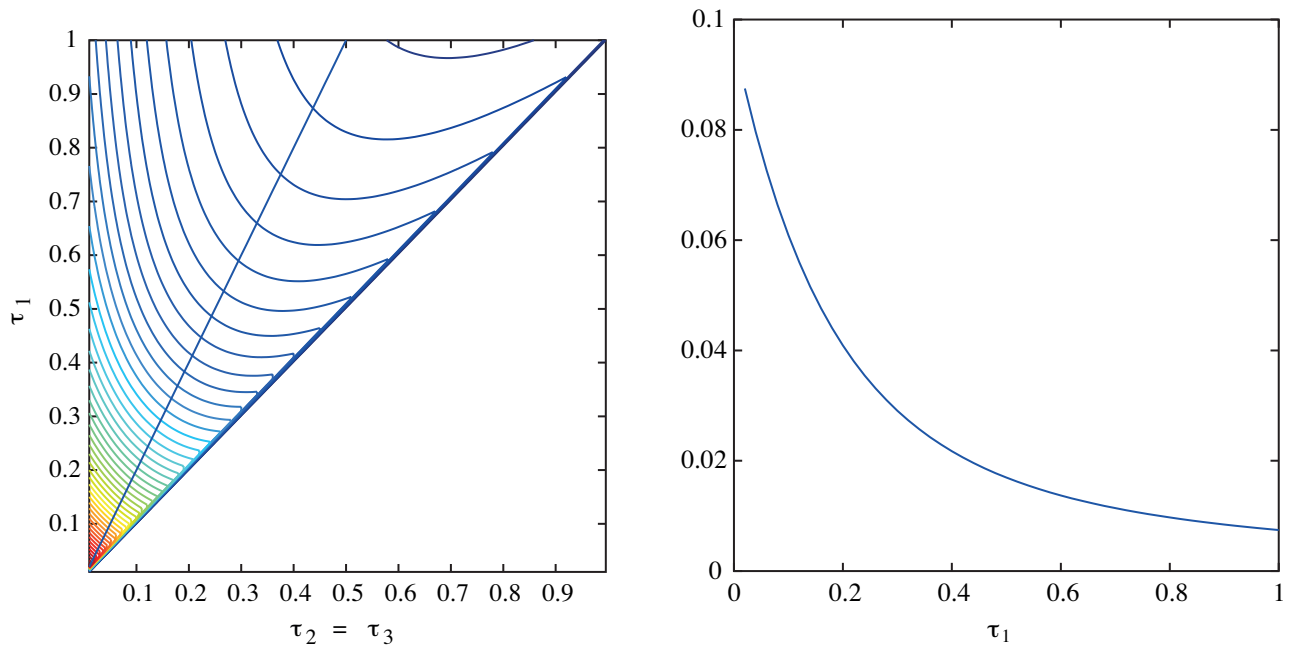
The dynamics of  $X$  on  $\mathbb{Z}_3^3$  for our numerical example will be that of a constrained random walk with the following increments:

$$\pm e_k, \pm(e_1 + e_2), \pm(e_1 + e_2 + e_3), k \in K, \tag{46}$$

where  $e_1 := (1, 0, 0)$ ,  $e_2 := (0, 1, 0)$ , and  $e_3 := (0, 0, 1)$ ; the  $\{\Gamma_k\}$  are assumed to be absorbing, i.e. if  $X_{u_0} \in \Gamma_k$  any increment involving  $\pm e_k$  can no longer be an increment of  $X$  for  $u > u_0$ . The sets  $B_k := \{z : z_k = 2\}$  are “reflecting” in the sense that if  $X_t \in B_k$  for some  $t$ , increments involving  $+e_k$  cannot be the first increment of  $X$  in the time interval  $[t, \infty)$ . We assume the following jump rates for the increments listed in (46):

$$2, 1, 2, 1, 3, 1, 0.5, 0.5, 0.2, 0.2;$$

e.g., if  $X_0 = (1, 1, 1)$  and  $\sigma_1$  denotes the first jump time of  $X$ ,  $\sigma_1$  is exponentially distributed with rate  $s$  where  $s$  is the sum of the rates in the above display and  $P(X_{\sigma_1} = X_0 + e_1 + e_2 + e_3) = 0.2/s$ . These rates and the aforementioned dynamics give a  $27 \times 27$   $\lambda$  matrix. The level sets  $f(\alpha, \cdot, K)|_{R_{s_1}}$  are depicted in Figure 1 and the graph of  $f(\alpha, \cdot, K)|_{R_{s_2}}$  is depicted in Figure 2.



**Figure 1.** The level curves of the density  $f$  for  $\tau_2 = \tau_3 < \tau_1$ . On the right: the values of  $f$  over the line segment connecting  $(0,0)$  to  $(0.5,1)$ .

For the parameter values of this numerical example,  $P_\alpha(\cap_{k \neq k'} \tau_k \neq \tau_{k'}) = 0.899$  and thus the singular parts account for around 10% of the distribution of  $\tau$ .

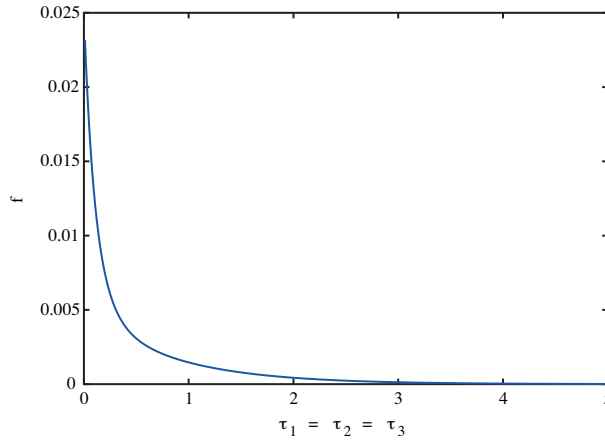


Figure 2. The density  $f$  for  $\tau_1 = \tau_2 = \tau_3$ .

### 6. Conclusion

Our primary motivation in deriving the formulas in the present paper has been their potential applications to credit risk modeling, especially to model default times of companies/obligors with first hitting times of a finite state Markov process *where multiple defaults are allowed to happen at the same time*; given the results of our paper this is now possible in a reasonably general framework (for the case of two obligors one could use the results in [5]). Let us explain a bit further this application starting from the credit risk model of [14] and the numerical example of Section 5.

With the results in the present work one can extend the modeling approach of [14] in two directions. Remember that the underlying process in [14] can only move by increments of  $\{-e_k\}$ , i.e. the model assumes that the obligors can default only one at a time. However, for highly correlated obligors it may make sense to allow simultaneous defaults, i.e. allow increments of the form  $-\sum_n e_{k_n}$ . Once multiple defaults are allowed, the default times will have nonzero singular parts and the formulas in the present work can be used to compute them, as is done in the numerical example of Section 5. Secondly, the default sets  $\{\Gamma_k\}$  no longer have to be assumed to be absorbing. Thus, with the formulas derived in the present work, one can treat models that allow recovery from default.

As  $|E|$  increases, (20) and other formulas derived in the present paper will take too long a time to compute (the same holds for earlier density formulas in the prior literature) and it can be of interest to derive asymptotic approximations for these densities.

A second theoretical contribution of the present work is to the line of research that originated in [12] and continued in [18] and [15]. In [12] the following problem was studied: given a filtration  $\mathbb{G} = \{\mathcal{G}_u, u \in \mathbb{R}_+\}$  and a multivariate random time  $\tau = (\tau_1, \dots, \tau_m)$  study the conditional law, say  $\mu_u^{\mathbb{G}}$ , of  $\tau$  given  $\mathcal{G}_u$ , in the case that where  $P(\tau_i = \tau_j) = 0$  for  $i \neq j, i, j = 1, 2, \dots, m$ . Thus, a (random) measure was sought so that

$$P(\tau \in B | \mathcal{G}_u) = \int_B \mu_u^{\mathbb{G}}(d\mathbf{t}), \tag{47}$$

for any Borel subset  $B$  of  $\mathbb{R}_+^m$ . If the measure  $\mu_u^{\mathbb{G}}$  is represented as

$$\mu_u^{\mathbb{G}}(d\mathbf{t}) = \varphi_u^{\mathbb{G}}(\mathbf{t})\nu(d\mathbf{t}), \tag{48}$$

where  $\nu$  is a (possibly random) measure on  $\mathbb{R}_+^m$ , then  $\varphi_u^{\mathbb{G}}$  is called the density of  $P(\cdot|\mathcal{G}_u)$  with respect to  $\nu$ , and the process  $\varphi^{\mathbb{G}}$  is called the conditional density process. This study was extended in [15], for  $n = 2$ , to the case where joint defaults were allowed, that is  $P(\tau_1 = \tau_2) > 0$ .

Now, let  $\mathbb{F} = \{\mathcal{F}_u, u \in \mathbb{R}_+\}$  be the filtration generated by  $X$ . The Markov property of  $X$  implies that the conditional density of  $\tau$  given  $\mathcal{F}_u$  directly follows from the density formula (20), as we show in Proposition 3.3. Thus, our results generalize (47) and (48) to the case of arbitrary  $m \geq 1$  where the restriction  $P(\tau_i = \tau_j) = 0$  for  $i \neq j$ ,  $i, j = 1, 2, \dots, m$  is no longer required. This generalization covers only the Markovian case, i.e. when  $\mathbb{G} = \mathbb{F}$  and when  $\tau$  is defined as first hitting times of the process  $X$ . Still, it allows one to study and model probabilities related to simultaneous multivariate trigger events, such as simultaneous defaults in a large pool of obligors.

### References

- [1] Aldous D, Shepp L. The least variable phase type distribution is Erlang. *Stoch Models* 1987; 3: 467-473.
- [2] Asmussen S. *Applied Probability and Queues*. New York, NY, USA: Springer, 2003.
- [3] Asmussen S, Albrecher H. *Ruin Probabilities*. Singapore: World Scientific, 2010.
- [4] Asmussen S, Bladt M. Point processes with finite-dimensional conditional probabilities. *Stoch Proc Appl* 1999; 82: 127-142.
- [5] Assaf D, Langberg NA, Savits TH, Shaked M. Multivariate phase-type distributions. *Oper Res* 1984; 32: 688-702.
- [6] Barges M, Loisel S, Venel X. On finite-time ruin probabilities with dependence between reinsurance cycles and the claim arrival process. *Scand Actuar J* 2013; 3: 163-185.
- [7] Bladt M, Nielsen BF. Multivariate matrix-exponential distributions. *Stoch Models* 2010; 26: 1-26.
- [8] Borovkov AA. *Probability Theory*. New York, NY, USA: Springer, 2009.
- [9] Cai J, Li H. Conditional tail expectations for multivariate phase-type distributions. *J Appl Probab* 2005; 42: 810-825.
- [10] Crépey S, Bielecki TR, Brigo D. *Counterparty Risk and Funding: A Tale of Two Puzzles*. Boca Raton, FL, USA: CRC Press, 2014.
- [11] Eaton JW. *GNU Octave Manual*. London, UK: Network Theory Limited, 2002.
- [12] El Karoui N, Jeanblanc M, Jiao Y. Density approach in modeling successive defaults. *SIAM J Fin Math* 2015; 6: 1-21.
- [13] Erlang AK. Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges. *Elektroteknikeren* 1917; 13: 5-13.
- [14] Herbertsson A. Modelling default contagion using multivariate phase-type distributions. *Review of Derivatives Research* 2011; 14: 1-36.
- [15] Jiao Y, Li S. Generalized density approach in progressive enlargement of filtrations. *Electron J Probab* 2015; 20: 1-21.
- [16] Johnson MA, Taaffe MR. Matching moments to phase distributions: mixtures of Erlang distributions of common order. *Stoch Models* 1989; 5: 711-743.
- [17] Karatzas I, Shreve SE. *Brownian Motion and Stochastic Calculus*. 2nd ed. New York, NY, USA: Springer-Verlag, 1991.
- [18] Kchia Y, Larsson M. Credit contagion and risk management with multiple non-ordered defaults. [arxiv.org/abs/1104.5272](https://arxiv.org/abs/1104.5272), 2013.
- [19] Kulkarni VG. A new class of multivariate phase type distributions. *Oper Res* 1989; 37: 151-158.

- [20] Loi C, Cournde PH. Generating functions of stochastic L-systems and application to models of plant development. In: Discrete Mathematics and Theoretical Computer Science 2008, DMTCS Proceedings Vol. AI, pp. 325-338.
- [21] O'Conneide CA, Raftery AE. A continuous multivariate exponential distribution that is multivariate phase type. Stat Prob Lett 1989; 7: 323-325.
- [22] Stanley RP. Enumerative Combinatorics. Cambridge, UK: Cambridge University Press, 2011.
- [23] Syski R. Passage Times for Markov Chains. Amsterdam, the Netherlands: IOS Press, 1992.