FILTERING OF FINITE-STATE TIME-NONHOMOGENEOUS MARKOV PROCESSES, A DIRECT APPROACH

N.V. KRYLOV AND A. ZATEZALO

ABSTRACT. Filtering equation is derived for $P(x_t = x|y_s, s \in [0,t])$ for a continuous-time finite-state two-component time-nonhomogeneous cadlag Markov process $z_t = (x_t, y_t)$. The derivation is based on some new ideas in the filtering theory and does not require any knowledge of stochastic integration.

1. INTRODUCTION

Let $(\Omega, \mathcal{F}, P)$ be a probability space, N and M be integers, $Z = \{1, \dots, N\} \times \{1, \dots, M\}$, and $z_t = (z_t^1, z_t^2)$ be a Z–valued random process defined for $t \geq 0$. We make the following assumption.

Assumption 1.1. The process z_t is a (time–nonhomogeneous) Markov process with cadlag trajectories.

Recall that a function f_t on $[0, \infty)$ is said to be cadlag if it is right continuous on $[0, \infty)$ and has finite left limits on $(0, \infty)$. Quite often we denote

$$x_t = z_t^1, \qquad y_t = z_t^2$$

and we interpret x_t as a signal process and y_t as an observation process. Notice that $x_t \in X = \{1, \dots, N\}$ and $y_t \in Y = \{1, \dots, M\}$ and neither of the processes x_t or y_t is assumed to be Markov alone.

The main goal of the article is to derive the so–called filtering equations for $P(x_t = x|y_s, s \in [0,t])$. This is a very old subject with a vast bibliography which along with historical remarks can be found in [3], [6] and [9]. In the case in which x_t is a time–homogeneous Markov process and y_t is a more general multivariate point process these equations can be found in [3] (see there pages 98, 106, 179). In the general case these equations can be derived on the basis of the results in Chapter 4 of [6]. Therefore in a sense our results are not new. However to the best of authors' knowledge the filtering equations for time–nonhomogeneous case have not been ever derived before. Actually, a case in which x_t is a time–nonhomogeneous Markov process is considered in Sec. 7.3 of [5] (also see [1]), but it seems to us that the result presented there has no sense (the derivation is based on Lemma 2.6 of Chapter 7 of [5] which is false, see Remark 3.1 later). One of motivations of this article is to derive correct equations in the general case of time–nonhomogeneous Markov processes z_t.

Another motivation is related to the ways of deriving filtering equations generally adopted in the literature (cf. [3] and [6]). The point is that the known methods

1991 *Mathematics Subject Classification.* 60G35, 62M20.

Key words and phrases. Finite-state Markov processes, filtering.

The work of the first author was partially supported by the NSF Grant DMS–9625483.

are based on exploiting many fine properties of general discontinuous martingales, particularly a quite sophisticated theory of stochastic integration against them, stochastic integral representation of general discontinuous martingales, the theory of innovation processes, and the method of change of measures. One has to work really hard even to get only acquainted with all needed notions. If the reader is only interested in understanding the derivation of filtering equations in our particular case, he/she is in a very tough situation following these general lines.

It turns out that there exists a new and quite elementary derivation based on what we call "a direct method". The main idea of this method is to use results like Theorems 5.1 and 5.2. In what concerns integration in time variable, the proofs of these theorem, as well as of our main results, do not require anything more sophisticated than Lebesgue integration with respect to Lebesgue measure and integration against piece–wise constant functions. Actually, even Lebesgue integration is not needed if we confine ourselves to the case in which local intensities of our time–nonhomogeneous Markov process are continuous in time variable (for the time–homogeneous case they are independent of time and one need not know Lebesgue integration either). Also, we do not use Doob's optional sampling theorem, "Clark's" representation theorems, right–continuity of filtrations, and σ–algebras like $\mathcal{F}_\tau$ or $\mathcal{F}_{\tau-}$ where τ is a stopping time. However the reader should know the notions of σ–algebra, conditional expectation, martingale, and stopping time. We also use the notion of predictable process. It is worth noting that the most general predictable process in this article is the finite sum of products of deterministic functions of t and left–continuous adapted processes. Moreover, in the case in which the local intensities are left continuous, the only predictable processes in this article will be left continuous and adapted. The reader just getting acquainted with the theory can assume from the beginning that all intensities are left continuous or even continuous and in this case he/she will see all details of a new derivation of filtering equations. In a subsequent article we intend to apply a similar idea to deriving filtering equations for diffusion processes.

The article is organized as follows. In Sec. 2 we recall basic definitions and basic facts from the theory of finite–state time–nonhomogeneous Markov processes. Sec. 3 contains the main results about the filtering equations. Since their proofs in Sec. 5 are quite formal and have nothing to do with considering $P(x_t = x|y_s, s \in [0,t])$ as a function of time, in Sec. 4 we explain heuristically the ideas behind finding the filtering equations. The arguments in Sec. 5 are based on Theorems 5.1 and 5.2 which we have mentioned before and which are proved in Sec. 6 and 7 respectively. The final Sec. 8 contains elementary proofs of some "well–known" (although we could not find exact references) facts about time–nonhomogeneous finite–space Markov processes. These facts provide a useful insight into the behavior of our processes, they are proved on the basis of previous sections and can be used in the heuristic derivation of filtering equations for time–nonhomogeneous processes as in Sec. 4.

2. Finite–state Markov processes

Here we briefly discuss the basics of the theory of finite–state Markov processes. One can find many books where finite–state Markov time–homogeneous processes are treated. However, we could not find in the literature some facts, which we are going to use, concerning the important in practice (cf. [5]) case of time–nonhomogeneous processes. In this section we do not need to assume that z_t has two components. It can be just arbitrary cadlag Markov process with values in some finite set Z.

For $0 \le s \le t < \infty$, and $u, v \in Z$ we define

$$p(s,u,t,v) = P(z_t = v | z_s = u) := \begin{cases} \frac{P(z_t=v, z_s=u)}{P(z_s=u)} & \text{if} \quad P(z_s = u) > 0, \\ P(z_t = v) & \text{if} \quad P(z_s = u) = 0. \end{cases} \tag{2.1}$$

Also, for $t \ge 0$ let $\mathcal{F}^z_t$ be the σ-algebra generated by the sets $\{\omega : z_s = v\}, s \in [0,t], v \in Z$, and by the sets of probability zero. By Assumption 1.1 the process z_t is a Markov process, which is and only is to say that for any function f on Z and any $0 \le t < \infty$ we have (a.s.)

$$\begin{aligned} E\{f(z_t)|\mathcal{F}^z_s\} &= E\{f(z_t)|z_s\} \\ &= \sum_{v\in Z} f(v) P(z_t = v|z_s) = \sum_{v \in Z} p(s, z_s, t, v) f(v). \end{aligned} \tag{2.2}$$

It is also worth noting that owing to right continuity of z_t (and the Dominated Convergence Theorem) the functions $p(s,u,t,v)$ are right continuous in t for any s, u, v.

Our next assumption is about existence of the so–called infinitesimal intensities $\lambda_{uv}(t)$. Roughly speaking, we assume that, for t and $\Delta t \ge 0$,

$$P(z_{t+\Delta t} = v | z_t = u) = \begin{cases} \lambda_{uv}(t)\Delta t + o(\Delta t) & \text{if} \quad u \ne v, \\ 1 + \lambda_{uu}(t)\Delta t + o(\Delta t) & \text{if} \quad u = v \end{cases} \tag{2.3}$$

provided $P(z_t = u) > 0$. In rigorous terms this is formulated in the spirit of Stroock–Varadhan martingale problem as follows.

Assumption 2.1. There exist functions $\lambda_{uv}(t)$ defined for $u, v \in Z$, and $t \ge 0$ such that

$$\int_0^T |\lambda_{uv}(t)|\, dt < \infty \quad \forall T < \infty, u, v \in Z$$

and for any $z \in Z$ the process

$$I_{z_t = z} - \int_0^t \lambda_{z_s z}(s)\, ds \tag{2.4}$$

is an $\mathcal{F}^z_t$–martingale.

The following three remarks intend to give an additional information about z_t. We do not use these remarks in the main part of the article and prove only the first of them for the sake of completeness in Sec. 8.

Remark 2.2. An *equivalent* form of Assumption 2.1 is that the function $p(s,u,t,v)$ satisfies Kolmogorov's equation

$$p(s,u,t,v) = \delta^{uv} + \sum_{z \in Z} \int_s^t p(s,u,r,z) \lambda_{zv}(r)\, dr \tag{2.5}$$

for all $t \ge s$, $v \in Z$, and $u \in Z$ such that $P(z_s = u) \ne 0$, where $\delta^{uu} = 1$ and $\delta^{uv} = 0$ if $u \ne v$. In differential form this equation means that

$$\frac{d}{dt} p(s,u,t,v) = \sum_{z\in Z} p(s,u,t,z)\lambda_{zv}(t).$$

Remark 2.3. Given $\lambda_{uv}(t)$ satisfying

$$\sum_{z\in Z}\lambda_{uz}(t)=0\quad\forall u\in Z,\quad(1-\delta^{uv})\lambda_{uv}(t)\geq 0\quad\forall u,v\in Z,t\geq 0,\tag{2.6}$$

one can solve (2.5) and for any given initial distribution one can construct a Markov cadlag process z_t having these intensities.

Remark 2.4. The σ–algebras $\mathcal{F}^z_t$ are right continuous in t.

One can find a justification of condition (2.6) in the following lemma and in the remark after it.

Lemma 2.5. *For all $z\in Z$ and almost all (ω,t) it holds that*

$$\sum_{v\in Z}\lambda_{z_t v}(t)=0,\qquad I_{z_t\neq z}\lambda_{z_t z}(t)\geq 0.\tag{2.7}$$

We prove this lemma in Sec. 6. The reader which intends to read through the whole article can go to this section for the proof right away.

Remark 2.6. For $u\in Z$ let

$$T(u)=\{t\geq 0:\sum_{z\in Z}\lambda_{uz}(t)=0,\lambda_{uv}(t)\geq 0\text{ for all }v\in Z\setminus\{u\}\ \}.$$

Define

$$\bar\lambda_{uv}(t):=\lambda_{uv}(t)I_{T(u)}(t)\quad t\in[0,\infty),u,v\in Z.$$

We claim then that $\bar\lambda_{uv}(t)$ satisfy

$$\int_0^T|\bar\lambda_{uv}(t)|\,dt<\infty\quad\forall T<\infty,u,v\in Z$$

and for any $z\in Z$ the process

$$I_{z_t=z}-\int_0^t\bar\lambda_{z_s z}(s)\,ds$$

is an $\mathcal{F}^z_t$–martingale. In other words, the functions $\bar\lambda_{uv}(t)$ are intensities of z_t.

Indeed, the first statement is trivial because $|\bar\lambda_{uv}(t)|\leq|\lambda_{uv}(t)|$. To prove the second one observe that Lemma 2.5 says that the inclusion $t\in T(z_t)$ holds for almost all (ω,t). Therefore, $\bar\lambda_{z_t v}(t)=\lambda_{z_t v}(t)$ for almost all (ω,t) and by Fubini's theorem

$$\int_0^t\bar\lambda_{z_s z}(s)\,ds=\int_0^t\lambda_{z_s z}(s)\,ds\quad\forall t\geq 0$$

almost surely, so that one can replace $\lambda_{z_s z}(s)$ with $\bar\lambda_{z_s z}(s)$ in (2.4).

Observe that by definition $\bar\lambda_{uv}(t)$ satisfy (2.6). Since $\bar\lambda_{uv}(t)$ are as good intensities of z_t as the original $\lambda_{uv}(t)$, we may and will always (apart from the proof of Lemma 2.5) assume that (2.6) holds.

3. The filtering equation

It is convenient to number the moments of jumps of z_t. Since it is a cadlag process with finite range, it only has finitely many jumps on every finite time interval. Define

$$\tau(0)=0, \quad \tau(n+1)=\inf\{s\geq\tau(n): z_s\neq z_{\tau(n)}\} \quad n=0,1,...,$$
$$\mathcal{T}=\{\tau(1),\tau(2),...\}, \quad \mathcal{T}(t)=\mathcal{T}\cap(0,t].$$

The finite set $\mathcal{T}(t)$ is the set of all moments of jumps of z_s which occur on the interval of time $(0,t]$. Observe that $\tau(n)$ are stopping times with respect to $\{\mathcal{F}^z_t\}$, that is $\{\omega: t<\tau(n)\}\in\mathcal{F}^z_t$ for any $t\geq 0$. Indeed,

$$\{\tau(n)\leq t\}=\bigcup_{m=n}^{\infty}\bigcup_{0=i_0<i_1<...<i_m\leq m}\{z_{ti_1/m}\neq z_{ti_0/m},...,z_{ti_m/m}\neq z_{ti_{m-1}/m}\}.$$

For $z\in Z$ the process $I_{z_t=z}$ is cadlag and only takes two values 0 or 1. So, this process is piece–wise constant and one knows how to integrate functions against such processes. Let us briefly recall the construction of such integrals. Let $I=\{0=t_0<t_1<t_2<...\}$ be a finite or infinite partition of $[0,\infty)$. Assume that for any $T<\infty$ there are only finitely many k such that $t_k\in[0,T]$. Let f_t be a function such that for any k for which $t_k\in I$ we have $f_t=f_{t_k}$ on $[t_k,t_{k+1})$ with the agreement that $t_{k+1}=\infty$ if t_{k+1} is not defined. For any function g_t and $0\leq t<\infty$ we set

$$\int_{(0,t]}g_s\,df_s:=\int_0^t g_s\,df_s:=\sum_{t_k\leq t}g_{t_k}\Delta f_{t_k}, \tag{3.1}$$

where and in the future we use the notation

$$\Delta f_t=f_t-f_{t-}.$$

Denote h_t the function of t introduced by (3.1). One can easily check that $h_t=h_{t_k}$ on $[t_k,t_{k+1})$ and $\Delta h_{t_k}=g_{t_k}\Delta f_{t_k}$ for any $t_k\in I$. Furthermore, one can integrate against h as well and for any function u_t it holds that

$$\int_0^t u_s\,dh_s=\int_0^t u_s g_s\,df_s$$

if $0\leq t<\infty$. In this sense the definition $dh_t=g_t\,df_t$ is consistent.

Since the jumps of $I_{z_t=z}$ are also jumps of z_t, for instance,

$$\int_0^t g_s\,dI_{z_s=z}=\sum_{s\in\mathcal{T}(t)}g_s\Delta I_{z_s=z},$$

Next, define

$$J_t(z):=\int_0^t I_{z_s=z}dI_{z_s=z}, \quad J_t=\sum_{z\in Z}J_t(z).$$

It is not hard to understand that $J_t(z)$ is the number of all jumps of z_s into the state z which occur on the interval $(0,t]$ and J_t is the number of all jumps of z_s on the interval $(0,t]$.

We will also need similar processes related to the process y_t which takes values in Y and is the second component of z_t. Define

$$\sigma(0)=0, \quad \sigma(n+1)=\inf\{s\geq\sigma(n): y_s\neq y_{\sigma(n)}\} \quad n=0,1,...,$$

$$\mathcal{S} = \{\sigma(1), \sigma(2), ...\}, \quad \mathcal{S}(t) = \mathcal{S} \cap (0, t].$$

$$K_t(y) := \int_0^t I_{y_s=y} dI_{y_s=y}, \quad K_t = \sum_{y \in Y} K_t(y). \tag{3.2}$$

Here $K_t(z)$ is the number of all jumps of y_s into the state y which occur on the interval $(0, t]$ and K_t is the number of all jumps of y_s on the interval $(0, t]$. The random variables $\sigma(n)$ are stopping times with respect to $\{\mathcal{F}_t^y\}$, where $\mathcal{F}_t^y$ is the σ-algebra generated by the sets $\{\omega : y_s = y\}, s \in [0, t], y \in Y$, and by the sets of probability zero.

Remark 3.1. Since for any fixed y the process $K_t(y)$ cannot jump twice unless meantime some other $K_t(\eta)$ has jumped, the processes $K_t(y)$ for different y are not independent. On the other hand, they do not have common jumps. It follows that they are not Poisson and cannot be made Poisson by any change of time or/and probability measure.

Remark 3.2. One can easily check that the processes $K_t(y)$ from (3.2) also admit the following representation in terms of integrals of predictable (moreover, left–continuous) functions

$$K_t(y) := \int_0^t I_{y_{s-} \neq y} dI_{y_s=y}. \tag{3.3}$$

We will also use $dI_{y_t \neq y} = \Delta I_{y_t \neq y} \, dK_t$ or in other words

$$I_{y_t \neq y} = I_{y_0 \neq y} + \int_0^t \Delta I_{y_s \neq y} \, dK_s. \tag{3.4}$$

Actually, (3.4) is obvious since

$$(I_{y_s \neq y} - I_{y_{s-} \neq y})(K_s - K_{s-}) = I_{y_s \neq y} - I_{y_{s-} \neq y}$$

because every time $I_{y_s \neq y}$ has a jump, the process K_t jumps as well.

Remark 3.3. By the above definitions and since $\Delta K_r = 1 \iff r \in \mathcal{S}(t)$, we have

$$\int_0^t g_s \, dK_s = \sum_{s \in \mathcal{S}(t)} g_s.$$

Consider the following initial–value problem

$$\begin{aligned} dq_t(z) =& I_{y_t=z^2} \sum_{\zeta \in Z} q_t(\zeta) \lambda_{\zeta z}(t) \, dt + \big(\sum_{\zeta \in Z} q_{t-}(\zeta) \lambda_{\zeta z}(t) I_{y_t=z^2} - q_{t-}(z) \big) \, dK_t, \\ q_0(z) =& P(z_0 = z | y_0), \end{aligned} \tag{3.5}$$

where $P(z_0 = z|y_0) = P(z_0 = z|y_0 = y)$ for those ω for which $y_0 = y$. It is worth noting that we understand the system (3.5) in the integral form.

Lemma 3.4. *For any ω there exists a unique solution of (3.5) on* $[0, \infty)$.

Proof. One can solve (3.5) consecutively on each time interval $[\sigma(n), \sigma(n+1))$. On $[\sigma(0), \sigma(1))$ system (3.5) is equivalent to

$$q_t(z) = q_0(z) + \int_0^t I_{z^2=y_0} \sum_{\zeta \in Z} q_s(\zeta) \lambda_{\zeta z}(s) ds \qquad t < \sigma(1), \tag{3.6}$$

which is just a system of ordinary linear integral equations. One knows that the solution of such a system exists and is unique even for all t whenever $\lambda_{uv}(t)$ are locally integrable, which is true in our case.

Next, system (3.5) says that

$$q_{\sigma(1)}(z) = q_0(z) + \int_{(0,\sigma(1))} \dots + \int_{[\sigma(1),\sigma(1)]} \dots = q_{\sigma(1)-}(z)$$

$$+\Big(\sum_{\zeta\in Z} q_{\sigma(1)-}(\zeta)\lambda_{\zeta z}(\sigma(1))I_{y_{\sigma(1)}=z^2} - q_{\sigma(1)-}(z)\Big)\Delta K_{\sigma(1)}$$

$$= I_{y_{\sigma(1)}=z^2}\sum_{\zeta\in Z} q_{\sigma(1)-}(\zeta)\lambda_{\zeta z}(\sigma(1)), \tag{3.7}$$

where $q_{\sigma(1)-}(\zeta)$ are known from (3.6) an we have used $\Delta K_{\sigma(1)} = 1$. After having uniquely defined $q_{\sigma(1)}(z)$ we solve

$$q_t(z) = q_{\sigma(1)}(z) + \int_{\sigma(1)}^{t} I_{z^2=y_{\sigma(1)}} \sum_{\zeta\in Z} q_s(\zeta)\lambda_{\zeta z}(s)ds \qquad \sigma(1) \le t < \sigma(2).$$

Then again we get an explicit expression for $q_{\sigma(2)}(z)$ and so on. The lemma is proved.

We now pass to our main results. They will be only proved under the following additional assumption which is imposed only for simplicity of proving. Actually, the results hold without this additional assumption as well.

Assumption 3.5. For all $t \ge 0$, we have

$$\sum_{z\in Z}\lambda_{uz}(t) = 0 \quad \forall u \in Z, \quad \lambda_{uv}(t) > 0 \quad \forall u \ne v, u, v \in Z. \tag{3.8}$$

Remark 3.6. By Remark 2.6 the second relation in (3.8) always holds with $\ge$ in place of $>$. In connection with this it is worth mentioning that one of ways to understand that our main results are true without Assumption 3.5 is to consider instead of z_t a "close" process which one obtains by adding to the intensities at times when they vanish some small constants. If one does this appropriately, the trajectories of the new process will coincide with the trajectories of the old one apart from a set of very small probability.

Also notice that Assumption 3.5 is only used in the very last lines in Sec. 7.

Theorem 3.7. *Let $q_t(z)$ be the function defined in Lemma 3.4. Then $\sum_{z\in Z} q_t(z) > 0$ (a.s.) and*

$$\frac{q_t(z)}{\sum_{\zeta\in Z} q_t(\zeta)} = P(z_t = z|\mathcal{F}_t^y) \qquad (a.s.). \tag{3.9}$$

It turns out that system (3.5) is relatively easy to derive but it contains many "trivial" equations. Indeed obviously, one always has

$$P(z_t = z|\mathcal{F}_t^y) = I_{y_t=z^2}P(z_t = z|\mathcal{F}_t^y) \quad \text{(a.s.)} \tag{3.10}$$

and the same is true for $q_t(z)$ (see. (3.9)). That is why we also give a theorem in terms of $P(x_t = x|\mathcal{F}_t^y)$.

Theorem 3.8. *For $\xi, x \in X$ define*

$$\hat{\lambda}_t(\xi, x) = \lambda_{(\xi, y_{t-})(x, y_t)}(t),$$

and let $\hat{q}_t(x)$ be a unique solution of the system

$$\begin{aligned} d\hat{q}_t(x) =& \sum_{\xi \in X} \hat{q}_t(\xi)\hat{\lambda}_t(\xi, x)\, dt + \big(\sum_{\xi \in X} \hat{q}_{t-}(\xi)\hat{\lambda}_t(\xi, x) - \hat{q}_{t-}(x)\big)\, dK_t, \\ \hat{q}_0(x) =& P(x_0 = x | y_0). \end{aligned} \tag{3.11}$$

Then $\sum_{x \in X} \hat{q}_t(x) > 0$ (a.s.) and

$$P(x_t = x | \mathcal{F}_t^y) = \frac{\hat{q}_t(x)}{\sum_{\xi \in X} \hat{q}_t(\xi)} \qquad (a.s.).$$

4. Heuristic derivation of Theorem 3.7

Since the proofs of Theorems 3.7 and 3.8 are quite formal, we explain how equation (3.5) was found. Take the simplest case of intensities independent of t and such that $\lambda_{uv} > 0$ for $u \neq v$. Then the probabilistic behavior of z_t is very well known. One knows (see, for instance [3] p. 293 or [4], Section 1, Chapter 6) that there are independent sequences of independent random variables $\{\tau_1(z), \tau_2(z), ...\}$, $z \in Z$, such that z_0 is independent of $\{\tau_i(z)\}$ and $P(\tau_i(z) > t) = e^{\lambda_{zz} t}$ for any i, z. Moreover, the process z_t stays at z_0 for $t < \tau_1(z_0) =: \tau(1)$, then at time $\tau_1(z_0)$ jumps at point z independently of $\{\tau_i(z)\}$ with conditional probability given $z_0 = u$ ($\neq z$) equal to $\lambda_{uz}/|\lambda_{zz}|$. After this the process stays at $z_{\tau(1)}$ for $\tau(1) \leq t < \tau_1(z_{\tau(1)})$. Then the process jumps to z independently of $\{\tau_i(z)\}$ with conditional probability given $z_{\tau(1)} = u$ ($\neq z$) equal to $\lambda_{uz}/|\lambda_{zz}|$ and so on.

Denote

$$\bar{p}_t(z) = P(z_t = z | \mathcal{F}_t^y).$$

Intuitively, it is not hard to see that for $0 \leq t + dt < \sigma(1)$ and

$$\bar{q}_t(z) := \bar{p}_t(z) P(t < \sigma(1) | y_0)$$

we have

$$\bar{p}_{t+dt}(z) = P(z_{t+dt} = z | t + dt < \sigma(1), y_0),$$

$$\begin{aligned} \bar{q}_{t+dt}(z) = P(z_{t+dt} = z, t + dt < \sigma(1) | y_0) \\ = \sum_{\zeta \in Z, \zeta^2 = z^2} P(z_{t+dt} = z, z_t = \zeta, t + dt < \sigma(1) | y_0) \end{aligned}$$

with

$$\begin{aligned} & P(z_{t+dt} = z, z_t = \zeta, t + dt < \sigma(1) | y_0 = y) \\ &= P(z_{t+dt} = z, t + dt < \sigma(1), z_t = \zeta, t < \sigma(1) | y_0 = y) \\ &= P(z_{t+dt} = z, t + dt < \sigma(1) | z_t = \zeta, t < \sigma(1), y_0 = y) \\ &\quad \times P(z_t = \zeta | t < \sigma(1), y_0 = y) P(t < \sigma(1) | y_0 = y). \end{aligned}$$

The first factor on the right can be further written as

$$P(z_{t+dt} = z | z_t = \zeta, t < \sigma(1), y_0 = y) = P(z_{t+dt} = z | z_t = \zeta)$$

since, for $\zeta^2 = z^2$, the event $\{(z_{t+dt} = z, t + dt \geq \sigma(1) z_t = \zeta, t < \sigma(1), y_0 = y\}$ implies that there are at least two jumps of y_s and hence of z_s between t and $t + dt$, which happens with probability $O(dt^2)$. Upon remembering (2.3), we conclude

$$\bar{p}_{t+dt}(z) = \sum_{\zeta \in Z, \zeta^2 = z^2} P(z_{t+dt} = z | z_t = \zeta) \bar{q}_t(\zeta)$$

$$= \sum_{\zeta \in Z, \zeta^2 = z^2} \{I_{\zeta = z}(1 + \lambda_{zz}\, dt) + I_{\zeta \neq z} \lambda_{\zeta z}\, dt\} \bar{q}_t(\zeta),$$

$$\bar{q}_{t+dt}(z) = \bar{q}_t(z) + \sum_{\zeta \in Z, \zeta^2 = z^2} \bar{q}_t(\zeta) \lambda_{\zeta z}\, dt.$$

Next (cf. (3.10)) we have $\bar{q}_t(u) \not\equiv 0$ for $t < \sigma(1)$ only if $u^2 = y_t = y_0$ and hence

$$d\bar{q}_t(z) = I_{y_t = z^2} \sum_{\zeta \in Z, \zeta^2 = z^2} \bar{q}_t(\zeta) I_{y_t = \zeta^2} \lambda_{\zeta z}\, dt$$

$$= I_{y_t = z^2} \sum_{\zeta \in Z} \bar{q}_t(\zeta) I_{y_t = \zeta^2} \lambda_{\zeta z}\, dt = I_{y_t = z^2} \sum_{\zeta \in Z} \bar{q}_t(\zeta) \lambda_{\zeta z}\, dt$$

We see that $\bar{q}_t(z)$ satisfies (3.6) for $t < \sigma(1)$.

We also see that, since $\sum_z \bar{q}_t(z) = P(t < \sigma(1) | y_0)$, we have

$$\bar{p}_t(z) = \frac{\bar{q}_t(z)}{\sum_\zeta \bar{q}_t(\zeta)},$$

and, for $t < \sigma(1)$,

$$\Big\{ \sum_\zeta \bar{q}_t(\zeta) \Big\}^{-1} \Big(\sum_z \bar{q}_t(z) \Big)'_t = \sum_{z, \zeta \in Z} I_{y_t = z^2} \bar{p}_t(\zeta) \lambda_{\zeta z} = \bar{\lambda}_t(y_t),$$

where

$$\bar{\lambda}_t(y) := \sum_{\zeta, z \in Z, z^2 = y} \bar{p}_{t-}(\zeta) \lambda_{\zeta z} \tag{4.1}$$

(we write $t-$ instead of t for convenience in the future, at this point this is irrelevant since $\bar{p}_t$ is differentiable for $t < \sigma(1)$). This yields, for $t < \sigma(1)$,

$$\sum_\zeta \bar{q}_t(\zeta) = e^{\int_0^t \bar{\lambda}_s(y_s)\, ds}, \quad \bar{p}_t(z) = \bar{q}_t(z) e^{-\int_0^t \bar{\lambda}_s(y_s)\, ds},$$

$$d\bar{p}_t(z) = I_{y_t = z^2} \sum_{\zeta \in Z} \bar{p}_t(\zeta) \lambda_{\zeta z}\, dt - \bar{p}_t(z) \bar{\lambda}_t(y_t)\, dt, \tag{4.2}$$

Computing jumps of $\bar{p}_t(z)$ is slightly more delicate. Just before $\sigma(1)$ a posteriori distribution of z_t given $\mathcal{F}_t^y$ is $\bar{p}_t(z)$. At moment $\sigma(1)$ the new information comes that the process y_t has a jump and where it has jumped to. Also the above explanation of the nature of jumps allows us to conclude that if $\zeta, z \in Z$ and $\zeta^2 \neq z^2$, then

$$P(\sigma(1) \in (t, t + dt), z_{\sigma(1)} = z | z_t = \zeta, t < \sigma(1), y_0)$$

$$= |\lambda_{\zeta\zeta}|\, dt \lambda_{\zeta z} / |\lambda_{\zeta\zeta}| = \lambda_{\zeta z}\, dt.$$

Also for ω such that $\sigma(1) \in (t, t + dt) =: \delta t$ by denoting $\Phi(t) = dt / P(\sigma(1) \in \delta t | y_0)$, we have

$$P(z_{\sigma(1)} = z, z_{\sigma(1)-} = \zeta | y_0, \sigma(1))$$

$$= P(\sigma(1) \in \delta t, z_{\sigma(1)} = z, z_{\sigma(1)-} = \zeta | y_0) \Phi(t) / dt.$$

Here the first factor on the right equals

$$
\begin{aligned}
&P(\sigma(1)\in\delta t, z_{\sigma(1)}=z, z_t=\zeta, t<\sigma(1)|y_0)\\
&=P(\sigma(1)\in\delta t, z_{\sigma(1)}=z|z_t=\zeta, t<\sigma(1), y_0)P(z_t=\zeta, t<\sigma(1)|y_0)\\
&=\lambda_{\zeta z}P(z_t=\zeta, t<\sigma(1)|y_0)\,dt,
\end{aligned}
$$

where for ω such that $\sigma(1)\in\delta t$

$$
P(z_t=\zeta, t<\sigma(1)|y_0)=\bar q_t(\zeta)=\bar q_{\sigma(1)-}(\zeta)=\bar p_{\sigma(1)-}(\zeta)e^{\int_0^{\sigma(1)}\bar\lambda_s(y_s)\,ds}.
$$

Thus, for ω such that $\sigma(1)\in\delta t$ and $\zeta,z\in Z$ such that $\zeta^2\neq z^2$ we obtain

$$
P(z_{\sigma(1)}=z, z_{\sigma(1)-}=\zeta|y_0,\sigma(1))=\lambda_{\zeta z}\bar p_{\sigma(1)-}(\zeta)\Theta(\sigma(1)-), \tag{4.3}
$$

where $\Theta(t):=\Phi(t)\exp\{\int_0^t\bar\lambda_s(y_s)\,ds\}$.

We are going to use (4.3) on sets $\{y_{\sigma(1)}=z^2\}$, on which $y_{\sigma(1)-}\neq z^2$ so that automatically (see (3.10)) $\bar p_{\sigma(1)-}(\zeta)=I_{\zeta^2\neq z^2}\bar p_{\sigma(1)-}(\zeta)$. Upon summing up in (4.3) with respect to ζ and then z^1, on the set $\{y_{\sigma(1)}=z^2\}$ we find

$$
P(z_{\sigma(1)}=z|y_0,\sigma(1))=\sum_{\zeta\in Z}\bar p_{\sigma(1)-}(\zeta)\lambda_{\zeta z}\Theta(\sigma(1)-),
$$

$$
P(y_{\sigma(1)}=z^2|y_0,\sigma(1))=\bar\lambda_{\sigma(1)}(y_{\sigma(1)})\Theta(\sigma(1)-).
$$

Note in passing that this and obvious inequality $\Theta(t)>0$ yield that

$$
\bar\lambda_{\sigma(1)}(y_{\sigma(1)})\Theta(\sigma(1)-)=\sum_{y\in Y}I_{y_{\sigma(1)}=y}P(y_{\sigma(1)}=y|y_0,\sigma(1))>0.
$$

Finally,

$$
\begin{aligned}
\bar p_{\sigma(1)}(z)&=P(z_{\sigma(1)}=z|y_0,y_{\sigma(1)},\sigma(1))\\
&=I_{y_{\sigma(1)}=z^2}P(z_{\sigma(1)}=z|y_0,y_{\sigma(1)}=z^2,\sigma(1))\\
&=I_{y_{\sigma(1)}=z^2}\frac{P(z_{\sigma(1)}=z|y_0,\sigma(1))}{P(y_{\sigma(1)}=z^2|y_0,\sigma(1))},
\end{aligned}
$$

so that

$$
\bar p_{\sigma(1)}(z)=\frac{1}{\bar\lambda_{\sigma(1)}(y_{\sigma(1)})}I_{y_{\sigma(1)}=z^2}\sum_{\zeta\in Z}\bar p_{\sigma(1)-}(\zeta)\lambda_{\zeta z}. \tag{4.4}
$$

If we multiply $\bar p_t$ by any *continuous* function of t, the equation (4.4) will still hold. Hence, the function $\bar q_t(z)$ defined on $[0,\sigma(1)]$ by

$$
\bar q_t(z)=\begin{cases}\bar p_t(z)e^{\int_0^t\bar\lambda_s(y_s)\,ds} & t<\sigma(1),\\ \bar\lambda_t(y_t)\bar p_t(z)e^{\int_0^t\bar\lambda_s(y_s)\,ds} & t=\sigma(1)\end{cases}
$$

satisfies both (3.6) and (3.7). This is to say that it satisfies (3.5) on $[0,\sigma(1)]$.

After $\sigma(1)$ everything starts over again and step by step one gets that

$$
\bar q_t(z):=\bar p_t(z)e^{\int_0^t\bar\lambda_s(y_s)\,ds}\prod_{s\in\mathcal S(t)}\bar\lambda_s(y_s)=\bar p_t(z)e^{\int_0^t\log\bar\lambda_s(y_s)\,dK_s+\int_0^t\bar\lambda_s(y_s)\,ds}
$$

should satisfy (3.5) on $[0,\infty)$. Multiplying $\bar p_t(z)$ by a function of t introduces a new function which does not sum up to one and therefore is called an *unnormalized* a posteriori distribution of z_t given $\mathcal F^y_t$. The equations for unnormalized distributions are generally called Zakai's equations. Thus, (3.5) is Zakai's equation for our filtering problem. Also notice that continuing with equations (4.2) and (4.4) one gets that $\bar p_t(z)$ itself satisfies the following nonlinear filtering equation:

$$d\bar p_t(z) = I_{y_t=z^2}\sum_{\zeta\in Z}\bar p_t(\zeta)\lambda_{\zeta z}(t)\,dt - \bar\lambda_t(y_t)\bar p_t(z)\,dt \\ + \bar\lambda_t^{-1}(y_t)\sum_{\zeta\in Z}\bar p_{t-}(\zeta)\lambda_{\zeta z}(t)I_{y_t=z^2}\,dK_t - \bar p_{t-}(z)\,dK_t. \quad (4.5)$$

This equation can easily be derived rigorously from Theorem 3.7.

5. Proofs of Theorems 3.7 and 3.8

First, we derive Theorem 3.8 from Theorem 3.7. Notice that existence and uniqueness for (3.11) is proved in the same way as for (3.5). Since

$$P(x_t = x|\mathcal{F}_t^y) = \sum_{z^2\in Y} P(z_t = z|\mathcal{F}_t^y),$$

it follows from (3.9) and from the above mentioned uniqueness that to prove Theorem 3.8 it only remains to check that the function

$$\hat q_t(z^1) := \sum_{z^2\in Y} q_t(z)$$

satisfies (3.11).

Again, it follows from (3.9) (also see (3.10)) that $q_t(z) = q_t(z)I_{z^2=y_t}$ almost surely for any t. Since both parts are right continuous in t, the equality holds for all t at once with probability one. Thus, we also have

$$q_t(z) = q_t(z^1, y_t)I_{z^2=y_t}, \quad \hat q_t(z^1) = q_t(z^1, y_t),$$
$$q_t(z) = \hat q_t(z^1)I_{z^2=y_t}, \quad q_{t-}(\zeta) = \hat q_{t-}(\zeta^1)I_{\zeta^2=y_{t-}}$$

for all t at once with probability one. Owing to this and summing up in (3.5) with respect to z^2, we obtain that $\hat q_t(z^1)$ satisfies (3.11) indeed.

To prove Theorem 3.7 we need two more results.

Theorem 5.1. *Let g be a real–valued function given on Z and $h(s,y)$ be a real–valued function defined for $t \ge 0$ and $y \in Y$. Assume that h is measurable (Lebesgue) in t (actually, we will only need the result for h which are piece–wise constant and left continuous in t). Then the function*

$$v(t,z) := Eg(y_0)I_{z_t=z}e^{i\int_0^t h(s,y_s)\,dK_s} \quad (5.1)$$

satisfies the following system of ordinary differential equations

$$v_t(t,z) = \sum_{\zeta\in Z} v(t,\zeta)\lambda_{\zeta z}(t)\big(I_{\zeta^2=z^2} + e^{ih(t,z^2)}I_{\zeta^2\ne z^2}\big) \quad z \in Z. \quad (5.2)$$

Theorem 5.2. *For $y \in Y$ define*

$$\tilde\lambda_t(y) = \sum_{z\in Z, z^2=y}\lambda_{z_{t-}z}(t). \quad (5.3)$$

Then the following definitions make sense

$$\Lambda_t = e^{-\int_0^t \log\tilde\lambda_s(y_s)\,dK_s - \int_0^t\tilde\lambda_s(y_s)\,ds}, \quad \tilde p_t(z) = q_t(z)\Lambda_t,$$

and $E|\tilde p_t(z)| < \infty$. In addition, for g and h as in Theorem 5.1, the function

$$v(t,z) := Eg(y_0)\tilde p_t(z)e^{i\int_0^t h(s,y_s)\,dK_s}$$

satisfies the same system (5.2).

Remark 5.3. By Remark 2.6

$$I_{y_{t-}\neq y}\tilde\lambda_t(y)\geq 0,\quad \sum_{y\in Y}\tilde\lambda_t(y)=0,$$

$$\tilde\lambda_t(y_{t-})=\sum_{y\in Y}I_{y_{t-}=y}\tilde\lambda_t(y)=-\sum_{y\in Y}I_{y_{t-}\neq y}\tilde\lambda_t(y)\leq 0.$$

The first relation here also implies that $\tilde\lambda_{\sigma(n)}(y_{\sigma(n)})\geq 0$, $n\geq 1$, since by definition of moments of jumps of y_t, we have $y_{\sigma(n)-}\neq y_{\sigma(n)}$ for $n\geq 1$. Therefore, the logarithm of $\tilde\lambda_{\sigma(n)}(y_{\sigma(n)})$ and the integral

$$\int_0^t \log\tilde\lambda_s(y_s)\,dK_s=\sum_{s\in\mathcal{S}(t)}\log\tilde\lambda_s(y_s) \tag{5.4}$$

are well defined even without Assumption 3.5 and may be equal to minus infinity. However, without this assumption Theorem 5.2 is not true in general. Indeed Theorems 5.1 and 5.2 imply

$$Eg(y_0)\sum_{z\in Z}\tilde p_t(z)e^{i\int_0^t h(s,y_s)dK_s}=Eg(y_0)e^{i\int_0^t h(s,y_s)dK_s}. \tag{5.5}$$

Since g and h are arbitrary, it follows from (5.5) (see the proof of Theorem 3.7 below) that

$$1=E\{\Lambda_t|\mathcal{F}_t^y\}\sum_{z\in Z}q_t(z)\qquad\text{(a.s.)}\ . \tag{5.6}$$

But the equality (5.6) is not true in general without Assumption 3.5 as we can see in the following example.

Example. We consider a Markov process $z_t=(x_t,y_t)$, where $M=2$, $N=2$,

$$\lambda_{uv}(t)=\begin{cases}1 & \text{for } u=(1,1) \text{ and } v=(1,2),\\ -1 & \text{for } u=v=(1,1),\\ 0 & \text{otherwise },\end{cases}$$

for $t\in(2n,2n+1]$, $n=0,1,2,\cdots$ with continuity at zero and

$$\lambda_{uv}(t)=\begin{cases}1 & \text{for } u=(2,1) \text{ and } v=(2,2),\\ -1 & \text{for } u=v=(2,1),\\ 0 & \text{otherwise },\end{cases}$$

for $t\in(2n+1,2(n+1)]$, $n=0,1,2,\cdots$.

The initial condition of the model is

$$P(z_0=z)=\begin{cases}\frac12 & \text{for } z=(1,1) \text{ or } z=(2,1),\\ 0 & \text{otherwise.}\end{cases}$$

Calculating transitional probabilities $p(s,u,t,v)$ from (2.5) we see that we have only one jump from one of the states with second coordinate 1 to the corresponding states with second coordinate 2.

For $s, t \in [0,1]$, $s \le t$, we have from the equation (2.5) that

$$p(s,u,t,v) = \begin{cases} e^{s-t} & \text{for } u = v = (1,1), \\ 1 - e^{s-t} & \text{for } u = (1,1) \text{ and } v = (1,2), \\ 1 & \text{for } u = v = (2,1), \\ P(z_t = v) & \text{for } u = (2,2), \\ 0 & \text{otherwise.} \end{cases} \tag{5.7}$$

It is obvious that from (5.7) it follows $P(\sigma(1) \in (0,1)) > 0$.

From (5.3), we have

$$\tilde{\lambda}_{\sigma(1)}(y_{\sigma(1)}) = 1 \tag{5.8}$$

and (a.e.)

$$\tilde{\lambda}_s(y_s) = \sum_{x \in X} \lambda_{z_s(x,y_s)}(s) = \begin{cases} -1 & \text{for } s \le \sigma(1), \\ 0 & \text{for } s > \sigma(1). \end{cases} \tag{5.9}$$

Therefore from (5.8) and (5.9), we have

$$\Lambda_t = e^{\sigma(1) \wedge t}. \tag{5.10}$$

From (5.10), Λ_t is $\mathcal{F}_t^y$-measurable i.e. $E\{\Lambda_t | \mathcal{F}_t^y\} = \Lambda_t$.

Let's now calculate $q_t(z)$, $t \ge 0$, $z \in Z$ from (3.5). First, we want to get initial condition. Since $y_0 = 1$ P-a.s. we have $P(z_0 = (1,1)|y_0) = P(z_0 = (2,1)|y_0) = \frac{1}{2}$ and $P(z_0 = (1,2)|y_0) = P(z_0 = (2,2)|y_0) = 0$ i.e.

$$q_0((1,1)) = q_0((2,1)) = \frac{1}{2}\ , \ q_0((1,2)) = q_0((2,2)) = 0.$$

From (5.7) and discussion after we can assume that $\sigma(1) < 1$ on the set with positive probability and in what follows we consider such set.

Therefore from (3.6) we have that for $t < \sigma(1)$

$$q_t(z) = \begin{cases} \frac{1}{2} e^{-t} & \text{for } z = (1,1) \text{ or } z = (2,1), \\ 0 & \text{otherwise.} \end{cases}$$

For $t = \sigma(1)$, as above we have

$$\begin{aligned} q_{\sigma(1)-}((1,1)) &= q_{\sigma(1)-}((2,1)) = \frac{1}{2} e^{-\sigma(1)}, \\ q_{\sigma(1)-}((1,2)) &= q_{\sigma(1)-}((2,2)) = 0. \end{aligned}$$

By combining this with (3.7) and $y_{\sigma(1)} = 2$ P-a.s., we obtain

$$q_{\sigma(1)}(z) = \begin{cases} \frac{1}{2} e^{-\sigma(1)} & \text{for } z = (1,2), \\ 0 & \text{otherwise.} \end{cases} \tag{5.11}$$

As a consequence of (3.5) and (5.11) we get

$$q_t(z) = \begin{cases} \frac{1}{2} e^{-\sigma(1)} & \text{for } z = (1,2), \\ 0 & \text{otherwise,} \end{cases} \tag{5.12}$$

for $t \in (\sigma(1), 1]$. Finally, from (5.10) and (5.12), we have

$$\Lambda_t \sum_{z \in Z} q_t(z) = \frac{1}{2}$$

on the interval $[\sigma(1), 1]$ which shows that (5.6) doesn't hold on a set of positive probability.

Proof of Theorem 3.7. By Theorems 5.1 and 5.2 and by uniqueness of solutions to (5.2) we have that

$$Eg(y_0)\big(I_{z_t=z}-\tilde{p}_t(z)\big)e^{i\int_0^t h(s,y_s)\,dK_s}=0 \tag{5.13}$$

for any g,h. Observe that

$$\int_0^t h(s,y_s)\,dK_s=\sum_{y\in Y}\int_0^t h(s,y)\,dK_s(y),$$

which easily implies that by the integrals $\int_0^t h(s,y_s)\,dK_s$ one can approximate any linear combination

$$\sum_{i=i}^n \mu^i K_{s^i}(y^i) \qquad s^i\le t, y^i\in Y, \mu^i\in(-\infty,\infty).$$

Next by trigonometric functions one can approximate any bounded continuous function. This and (5.13) implies that

$$Eg(y_0)H(K_{s^1}(y^1),...,K_{s^n}(y^n))\big(I_{z_t=z}-\tilde{p}_t(z)\big)=0,$$

where H is any bounded continuous function, $y^i\in Y$, and $0\le s^i\le ...\le s^n\le t$. General theory or the fact that $y_0, K_s(y)$ only take values in discrete countable sets show that the last equality actually says that

$$E\{I_{z_t=z}|\mathcal{F}_t^K\}=E\{\tilde{p}_t(z)|\mathcal{F}_t^K\} \quad \text{(a.s.)}, \tag{5.14}$$

where $\mathcal{F}_t^K$ is the σ–algebra generated by y_0 and by all random variables $K_s(y)$, $s\le t$, $y\in Y$.

We now claim that $\mathcal{F}_t^K=\mathcal{F}_t^y$. Indeed, the inclusion $\mathcal{F}_t^K\subset\mathcal{F}_t^y$ is obvious. On the other hand, if one knows y_0 and $K_s(y)$ for $s=t$ and all rational $s\le t$ and $y\in Y$, then one knows all moments of jumps $\sigma(1),\sigma(2),...$ of the process y_s which occur on $(0,t]$ because at any jump of y_s one and only one of $K_s(y)$ has a jump. Also knowing y_0 and which of $K_s(y)$ has a jump at $\sigma(1)$, we know where the process got at time $\sigma(1)$ and thus we know the trajectory of y_s on $[0,\sigma(1)]$. We can continue the argument in the same way considering the trajectory after $\sigma(1)$. It is seen, that if one knows y_0 and $K_s(y)$ for $s=t$ and all rational $s\le t$ and $y\in Y$, then one knows y_s for $s\le t$. Thus, $\mathcal{F}_t^y\subset\mathcal{F}_t^K$ and $\mathcal{F}_t^K=\mathcal{F}_t^y$ indeed.

We add to this that $q_t(z)$ is obviously $\mathcal{F}_t^y$–measurable and from (5.14) we conclude

$$P\{z_t=z|\mathcal{F}_t^y\}=q_t(z)E\{\Lambda_t|\mathcal{F}_t^y\} \quad \text{(a.s.)}. \tag{5.15}$$

By summing up these equalities over $z\in Z$ we get

$$1=\sum_{z\in Z}q_t(z)E\{\Lambda_t|\mathcal{F}_t^K\}, \quad \sum_{z\in Z}q_t(z)>0,$$

$$E\{\Lambda_t|\mathcal{F}_t^K\}=\frac{1}{\sum_{z\in Z}q_t(z)} \quad \text{(a.s.)}.$$

After coming back to (5.15) we get (3.9) and this proves Theorem 3.7.

6. Proof of Theorem 5.1

In this section we do not use Assumption 3.5.

Lemma 6.1. *For any bounded $\mathcal{F}^z_t$–predictable process g_t, any bounded $\mathcal{F}^z_t$–stopping time τ, and $z \in Z$ we have*

$$E\int_0^\tau g_t\,dI_{z_t=z} = E\int_0^\tau g_t\lambda_{z_tz}(t)\,dt. \tag{6.1}$$

Proof. First fix $T \in [0,\infty)$ and assume that $\tau \le T$. From equalities like

$$E\int_0^\tau g_t\,dI_{z_t=z} = E\int_0^T g_tI_{t\le\tau}\,dI_{z_t=z},$$

where $I_{t\le\tau}$ is a predictable and even left–continuous $\mathcal{F}^z_t$–adapted process, it follows that we only need to prove (6.1) for nonrandom τ, say $\tau = T$. In this case if, in addition, g is a piece–wise constant function, we get (6.1) from Assumption 2.1 which says that for any $s \le t$ and any bounded $\mathcal{F}^z_s$–measurable ξ we have

$$E\xi(I_{z_t=z} - I_{z_s=z}) = E\xi E\{(I_{z_t=z} - I_{z_s=z})|\mathcal{F}^z_s\}$$
$$= E\xi\int_s^t \lambda_{z_rz}(r)\,dr. \tag{6.2}$$

Next, for any bounded left–continuous $\mathcal{F}^z_t$–adapted process g_t we have $g^n_t \to g_t$, where $g^n_t = g_{k/n}$ if $t \in (k/n,(k+1)/n]$. Also, by (6.2)

$$E\int_0^T g^n_t\,dI_{z_t=z} = E\int_0^T g^n_t\lambda_{z_tz}(t)\,dt. \tag{6.3}$$

Here

$$\int_0^T g^n_t\,dI_{z_t=z} \to \int_0^T g_t\,dI_{z_t=z}, \quad \int_0^T g^n_t\lambda_{z_tz}(t)\,dt \to \int_0^T g_t\lambda_{z_tz}(t)\,dt,$$

where the first relation is obvious and the second one follows from the Dominated Convergence Theorem (remember that $\lambda_{uv}(t)$ are assumed to be integrable over $[0,T]$). By the same theorem we will get (6.1) from (6.3) (for $\tau = T$ and left–continuous functions g) if we prove that

$$E\int_0^T |dI_{z_t=z}| := E\sum_{t\in\mathcal{T}(T)} |\Delta I_{z_t=z}| < \infty. \tag{6.4}$$

Observe that for any integer $n \ge 1$

$$I^2_{z_T=z} - I^2_{z_0=z} = \sum_{k=0}^{n-1}(I^2_{z_{T(k+1)/n}=z} - I^2_{z_{Tk/n}=z}) =$$

$$2\sum_{k=0}^{n-1} I_{z_{Tk/n}=z}(I_{z_{T(k+1)/n}=z} - I_{z_{Tk/n}=z}) + \sum_{k=0}^{n-1}(I_{z_{T(k+1)/n}=z} - I_{z_{Tk/n}=z})^2.$$

Take expectations of these expressions and use (6.2). Then we see that

$$E\sum_{k=0}^{n-1}(I_{z_{T(k+1)/n}=z} - I_{z_{Tk/n}=z})^2 = P(z_T = z) - P(z_0 = z)$$

$$-2E\sum_{k=0}^{n-1} I_{z_{Tk/n}=z}\int_{Tk/n}^{T(k+1)/n}\lambda_{z_t z}(t)\,dt\le 1+2\sum_{u\in Z}\int_0^T|\lambda_{uz}(t)|\,dt.$$

This leads to (6.4) by Fatou's theorem since obviously

$$\sum_{t\in\mathcal{T}(T)}|\Delta I_{z_t=z}|=\sum_{t\in\mathcal{T}(T)}|\Delta I_{z_t=z}|^2=\lim_{n\to\infty}\sum_{k=0}^{n-1}(I_{z_{T(k+1)/n}=z}-I_{z_{Tk/n}=z})^2.$$

Finally, to derive (6.1) in the case of general bounded predictable process g one uses general theory of integration which shows that by virtue of (6.4), equation (6.1) follows from its particular case in which $g=I_{t\le\sigma}$, where σ is any $\mathcal{F}^z_t$–stopping time (the indicator is left continuous). The lemma is proved.

Remark 6.2. In the next corollary and many times in the future we use that if f_t is a piece–wise constant function and h_t is a measurable function such that f_th_t is Lebesgue integrable over $[0,T]$, then

$$\int_0^T f_{t-}h_t\,dt=\int_0^T f_th_t\,dt.$$

This fact is true since $f_{t-}h_t\ne f_th_t$ only at finitely many points.

Corollary 6.3. *For any bounded and predictable g_t and nonnegative, bounded, and predictable h_t and any $T\in[0,\infty)$ we have*

$$E\int_0^T g_t\sum_{z\in Z}\lambda_{z_tz}(t)\,dt=0,\quad E\int_0^T h_tI_{z_t\ne z}\lambda_{z_tz}(t)\,dt\ge 0. \tag{6.5}$$

Indeed, to get the first equality it suffices to sum up in (6.1) with respect to $z\in Z$ and to notice that $\sum_{z\in Z}I_{z_t=z}\equiv 1$. To prove the second one it suffices to plug $g_t=h_tI_{z_{t-}\ne z}$ in (6.1) and notice that obviously $I_{z_{t-}\ne z}\Delta I_{z_t=z}\ge 0$.
Proof of Lemma 2.5. One can replace $\lambda_{z_tz}(t)$ with $\lambda_{z_{t-}z}(t)$ in (6.5). The latter is predictable since it is equal to $\sum_\zeta I_{z_{t-}=\zeta}\lambda_{\zeta z}(t)$ where the indicators are left continuous and $\lambda_{\zeta z}(t)$ are nonrandom. ¿From general theory of integration or after substituting $g_t=\sum_{z\in Z}\lambda_{z_{t-}z}(t)$ into the thus modified first equality in (6.5) one sees that $\sum_{z\in Z}\lambda_{z_{t-}z}(t)=0$ and $\sum_{z\in Z}\lambda_{z_tz}(t)=0$ for almost all (ω,t) indeed. Again, general theory or the substitution of the negative part of $I_{z_{t-}\ne z}\lambda_{z_{t-}z}(t)$ instead of h_t into the second equation in (6.5) proves that $I_{z_t\ne z}\lambda_{z_tz}(t)\ge 0$ for almost all (ω,t). The lemma is proved.

Lemma 6.4. *For any bounded or nonnegative $\mathcal{F}^z_t$–predictable g_t, any bounded $\mathcal{F}^z_t$–stopping time τ, and any $z\in Z$ we have*

$$E\int_0^\tau g_t\,dJ_t(z)=E\int_0^\tau g_tI_{z_t\ne z}\lambda_{z_tz}(t)\,dt, \tag{6.6}$$

$$E\int_0^\tau g_t\,dJ_t=-E\int_0^\tau g_t\lambda_{z_tz_t}(t)\,dt. \tag{6.7}$$

Also (6.6) and (6.7) hold true if g is $\mathcal{F}^z_t$–predictable and

$$E\int_0^\tau|g_t|I_{z_t\ne z}\lambda_{z_tz}(t)\,dt<\infty,\quad E\int_0^\tau|g_t|\,|\lambda_{z_tz_t}(t)|\,dt<\infty$$

respectively.

Proof. The last assertion follows from (6.6) and (6.7) and from the formula $g = g_+ - g_-$. In the first assertion formula (6.7) follows from (6.6) after summing up with respect to z and applying Lemma 2.5. In proving (6.6) as follows from Lemma 2.5 and the Monotone Convergence Theorem, we only need to consider the case of bounded g. In this case, (6.6) follows immediately from Lemma 6.1 since similarly to Remark 3.2 we have $dJ_t(z) = I_{z_{t-}\neq z}\, dI_{z_t=z}$. The lemma is proved.

Corollary 6.5. *For any nonnegative $\mathcal{F}^z_t$–predictable g_t, any $\mathcal{F}^z_t$–stopping time τ, and any $y \in Y$ we have*

$$E\int_0^\tau g_t\, dK_t(y) = E\int_0^\tau g_t I_{y_t\neq y}\tilde{\lambda}_t(y)\, dt,$$

$$E\int_0^\tau g_t\, dK_t = -E\int_0^\tau g_t\tilde{\lambda}_t(y_t)\, dt,$$

where $\tilde{\lambda}_t(y)$ is taken from (5.3). The first and the second equalities also hold true if g_t is an $\mathcal{F}^z_t$–predictable function for which

$$E\int_0^\tau |g_t| I_{y_t\neq y}\tilde{\lambda}_t(y)\, dt < \infty \quad or \quad E\int_0^\tau |g_t|\,|\tilde{\lambda}_t(y_t)|\, dt < \infty$$

respectively.

Indeed, for nonnegative g the second equation follows by summing up the first one over $y \in Y$ and applying Lemma 2.5. To get the first equation from (6.6) one only needs to observe that

$$dK_t(y) = \sum_{z\in Z, z^2=y} I_{y_{t-}\neq y}\, dJ_t(z).$$

For arbitrary g one uses the formula $g_t = (g_t)_+ - (g_t)_-$.

Lemma 6.6. *For any real-valued function $h(t,y)$ defined on $[0,\infty)\times Y$, any ω, and $t\in[0,\infty)$ we have*

$$e^{i\int_0^t h(s,y_s)\, dK_s} = 1 + \int_0^t e^{i\int_0^{r-} h(s,y_s)\, dK_s}(e^{ih(r,y_r)} - 1)\, dK_r. \tag{6.8}$$

Proof. Straightforward calculations give

$$e^{i\int_0^t h(s,y_s)\, dK_s} = 1 + \sum_{r\in\mathcal{S}(t)} (e^{\int_0^r h(s,y_s)\, dK_s} - e^{\int_0^{r-} h(s,y_s)\, dK_s}) =$$
$$= 1 + \sum_{r\in\mathcal{S}(t)} e^{\int_0^{r-} h(s,y_s)\, dK_s}(e^{h(r,y_r)} - 1),$$

(cf. Remark 3.3) which is (6.8). The lemma is proved.

Next result is related to integration against function of locally bounded variation. Actually we will only use it for functions which are representable as sums of Lebesgue integrals and of piece–wise constant functions. For such functions the following lemma is quite elementary.

Lemma 6.7 (see [8] p. 253)**.** *Let a_t and b_t, $t \geq 0$, be right continuous functions of locally bounded variation (that is having bounded variation on any finite interval*

$[0,T]$). *Then* $d(a_tb_t) = a_{t-}\,db_t + b_t\,da_t = a_{t-}\,db_t + b_{t-}\,da_t + \Delta b_t\,da_t$ *in the sense that*

$$\begin{aligned} a_tb_t =& a_0b_0 + \int_0^t a_{s-}db_s + \int_0^t b_s da_s \\ =& a_0b_0 + \int_0^t a_{s-}db_s + \int_0^t b_{s-}da_s + \int_0^t \Delta b_s\,da_s. \end{aligned} \tag{6.9}$$

Proof of Theorem 5.1. By Lemmas 6.7 and 6.6 we have

$$df_t(z) := d\big(I_{z_t=z}e^{i\int_0^t h(s,y_s)\,dK_s}\big) = I_{z_t=z}\,de^{i\int_0^t h(s,y_s)\,dK_s}$$
$$+e^{i\int_0^{t-} h(s,y_s)\,dK_s}\,dI_{z_t=z} = e^{i\int_0^{t-} h(s,y_s)\,dK_s}(e^{ih(t,z^2)}-1)I_{z_t=z}\,dK_t$$
$$+e^{i\int_0^{t-} h(s,y_s)\,dK_s}\,dI_{z_t=z}.$$

Observe that $I_{z_t=z}\,dK_t = I_{y_{t-}\neq z^2}\,dJ_t(z)$ (which we use to convert nonpredictable $I_{z_t=z}$ into predictable $I_{y_{t-}\neq z^2}$) and use Lemmas 6.1 and 6.4. Then

$$v(t,z) = Eg(y_0)f_t(z) =$$
$$Eg(y_0)I_{z_0=z} + \int_0^t e^{ih(s,z^2)}Eg(y_0)e^{i\int_0^s h(r,y_r)\,dK_r}I_{y_s\neq z^2}\lambda_{z_sz}(s)\,ds$$
$$+\int_0^t Eg(y_0)e^{i\int_0^s h(r,y_r)\,dK_r}I_{y_s=z^2}\lambda_{z_sz}(s)\,ds.$$

We finish the proof by noticing that

$$Eg(y_0)e^{i\int_0^s h(r,y_r)\,dK_r}I_{y_s\neq z^2}\lambda_{z_sz}(s) = \sum_{\zeta\in Z}\lambda_{\zeta z}(s)Eg(y_0)e^{i\int_0^s h(r,y_r)\,dK_r}I_{y_s\neq z^2}I_{z_s=\zeta}$$
$$=\sum_{\zeta\in Z}\lambda_{\zeta z}(s)I_{\zeta^2\neq z^2}Eg(y_0)e^{i\int_0^s h(r,y_r)\,dK_r}I_{z_s=\zeta} = \sum_{\zeta\in Z}\lambda_{\zeta z}(s)I_{\zeta^2\neq z^2}v(s,\zeta),$$
$$Eg(y_0)e^{i\int_0^s h(r,y_r)\,dK_r}I_{y_s=z^2}\lambda_{z_sz}(s) = \sum_{\zeta\in Z}\lambda_{\zeta z}(s)Eg(y_0)e^{i\int_0^s h(r,y_r)\,dK_r}I_{y_s=z^2}I_{z_s=\zeta}$$
$$=\sum_{\zeta\in Z}\lambda_{\zeta z}(s)I_{\zeta^2=z^2}Eg(y_0)e^{i\int_0^s h(r,y_r)\,dK_r}I_{z_s=\zeta} = \sum_{\zeta\in Z}\lambda_{\zeta z}(s)I_{\zeta^2=z^2}v(s,\zeta).$$

The theorem is proved.

7. Proof of Theorem 5.2

We start with simple properties of the solution $q_t(z)$ of (3.5). The first one is suggested by (3.9) and the second one by our way of proving Lemma 3.4.

Lemma 7.1. *Almost surely for all* t,z *we have* $q_t(z) = q_t(z)I_{y_t=z^2}$ *and* $q_t(z)\geq 0$.

Proof. By Lemma 6.7, (3.5), and Remark 3.2 we get

$$d(q_t(z)I_{y_t\neq z^2}) = q_{t-}(z)\,dI_{y_t\neq z^2} + I_{y_t\neq z^2}\,dq_t(z)$$
$$= q_{t-}(z)\,(dI_{y_t\neq z^2} - I_{y_t\neq z^2}\,dK_t) = -q_{t-}(z)I_{y_{t-}\neq z^2}\,dK_t.$$

Thus, the function $q_t(z)I_{y_t\neq z^2}$ satisfies the equation $df_t = -f_{t-}\,dK_t$. Also $f_0 = P(z_0=z|y_0)I_{y_0\neq z^2} = 0$ (a.s.). Therefore (cf. the proof of Lemma 3.4), $f_t\equiv 0$ (a.s.), which proves our first assertion.

We also have $q_{t-}(\zeta) = q_{t-}(\zeta) I_{y_{t-}=\zeta^2}$, which we use to prove the second assertion. By using Lemma 6.7 it is easy to prove that the function $\tilde{q}_t(z) := q_t(z) e^{-\int_0^t \lambda_{zz}(s)\,ds}$ satisfies (a.s.)

$$d\tilde{q}_t(z) = I_{y_t=z^2} \sum_{\zeta \in Z \setminus \{z\}} \tilde{q}_t(\zeta) \lambda_{\zeta z}(t)\,dt + [\sum_{\zeta \in Z} \tilde{q}_{t-}(\zeta) \lambda_{\zeta z}(t) I_{y_t=z^2 \neq \zeta^2} - \tilde{q}_{t-}(z)]\,dK_t. \quad (7.1)$$

Therefore, in equation (3.6) corresponding to (7.1) only nonnegative $\lambda_{uv}(t)$ appear, which shows that while constructing the solution of (7.1) on $[0, \sigma(1))$ by the method of successive approximations all approximations will be nonnegative if the first one is say, $q_0(z)$. Therefore, $\tilde{q}_t(z) \geq 0$ on $[0, \sigma(1))$ (a.s.). Furthermore, in equation (3.7) corresponding to (7.1) only nonnegative $\lambda_{uv}(t)$ appear, which shows that $\tilde{q}_{\sigma(1)}(z) \geq 0$ by virtue of $\tilde{q}_{\sigma(1)-}(\zeta) \geq 0$. One can continue in an obvious way getting that $\tilde{q}_t(z) \geq 0$ for all t (a.s.) whence $q_t(z) \geq 0$ for all t (a.s.). The lemma is proved.

So far in this section we did not use Assumption 3.5. We do not use this assumption in the proof of the following lemma either.

Lemma 7.2. *(i) For any t, ω, and y, if $y \neq y_{t-}$, then $\tilde{\lambda}_t(y) \geq 0$. In addition, almost surely*

$$\tilde{\lambda}_{\sigma(n)}(y_{\sigma(n)}) > 0 \quad \forall n \geq 1, \quad 0 < \Lambda_t < \infty \quad \forall t.$$

(ii) Let p_t be a nonnegative cadlag $\mathcal{F}_t^z$-adapted process satisfying the inequality

$$dp_t \leq c_t p_{t-} \big(\sum_{y \in Y} \frac{1}{\tilde{\lambda}_t(y)}\,dK_t(y) + dt \big),$$

where c_t is a nonrandom nonnegative function which is locally integrable on $[0, \infty)$. Then

$$Ep_t \leq Ep_0 e^{(M+1)\int_0^t c_s\,ds}.$$

Proof. (i). The first assertion in (i) follows by our agreement (see Remark 2.6). To prove the second one, notice that by Corollary 6.5 we have

$$E \int_0^T I_{\tilde{\lambda}_t(y)=0}\,dK_t(y) = E \int_0^T I_{\tilde{\lambda}_t(y)=0} \tilde{\lambda}_t(y)\,dt = 0$$

for any $y \in Y$. This means that almost surely for any $y \in Y$

$$\int_0^T I_{\tilde{\lambda}_t(y)=0}\,dK_t(y) = 0 \quad \forall T, \qquad \int_0^\infty I_{\tilde{\lambda}_t(y)=0}\,dK_t(y) = 0,$$

$$I_{\tilde{\lambda}_{\sigma(n)}(y)=0} I_{y_{\sigma(n)}=y} = 0 \quad \forall n \geq 1.$$

Hence almost surely for all $n \geq 1$

$$I_{\tilde{\lambda}_{\sigma(n)}(y_{\sigma(n)})=0} = \sum_{y \in Y} I_{\tilde{\lambda}_{\sigma(n)}(y_{\sigma(n)})=0} I_{y_{\sigma(n)}=y} = 0.$$

This proves (i) since $y_{t-} \neq y_t$ if $t = \sigma(n)$.

(ii). Let $r_t := p_t \exp\{-(M+1)\int_0^t c_s\,ds\}$. We have

$$dr_t = \exp\{-(M+1) \int_0^t c_s\,ds\}\,dp_t - r_{t-}(M+1)c_t\,dt.$$

This yields

$$dr_t \le c_t r_{t-} \sum_{y\in Y} \big(\frac{1}{\tilde\lambda_t(y)}\, dK_t(y) - I_{y_t\neq y}\, dt\big) \tag{7.2}$$

and we only need to prove that $Er_t \le Er_0$. Rewrite equation (7.2) in the integral form:

$$r_t \le r_0 + \int_0^t c_s r_{s-} \sum_{y\in Y} \big(\frac{1}{\tilde\lambda_s(y)}\, dK_s(y) - I_{y_s\neq y} ds\big). \tag{7.3}$$

By (i) for any bounded stopping time τ

$$E\int_0^\tau c_s r_{s-}\,\frac{1}{\tilde\lambda_s(y)}\, dK_s(y) = E\int_0^\tau c_s r_{s-}\,\frac{1}{\tilde\lambda_s^+(y)}\, dK_s(y), \tag{7.4}$$

where

$$\tilde\lambda_s^+(y) = \begin{cases} \tilde\lambda_s(y) & \text{if} \quad \tilde\lambda_s(y) > 0,\\ 1 & \text{otherwise.}\end{cases}$$

Corollary 6.5 implies

$$E\int_0^\tau c_s r_{s-}\,\frac{1}{\tilde\lambda_s(y)}\, dK_s(y) = E\int_0^\tau c_s r_{s-} I_{y_s\neq y}\,\frac{1}{\tilde\lambda_s^+(y)}\tilde\lambda_s(y)\, ds$$
$$\le E\int_0^\tau c_s r_{s-} I_{y_s\neq y}\, ds. \tag{7.5}$$

Now take a number $\alpha \in (0,\infty)$ and put here $t\wedge\tau_\alpha$ in place of τ, where $\tau_\alpha = \inf\{t\ge 0 : \sup_{s\le t} r_s \ge \alpha\}$. Since $\sup_{s\le t} r_s = \sup_{s\in Q_t\cup\{t\}} r_s$ is an $\mathcal{F}_t^z$–adapted process, where Q_t is the set of all rational numbers on $[0,t]$, and for any $u\ge 0$

$$\{u<\tau_\alpha\} = \{\sup_{s\le u} r_s < \alpha\} \in \mathcal{F}_u^z,$$

τ_α and $t\wedge\tau_\alpha$ are $\mathcal{F}_t^z$–stopping time. This combined with (7.5) and the observation that $0\le r_{s-}\le\alpha$ for *all* $s\in(0,\tau_\alpha]$ yields

$$E\int_0^{t\wedge\tau_\alpha} c_s r_{s-}\,\frac{1}{\tilde\lambda_s(y)}\, dK_s(y) \le E\int_0^{t\wedge\tau_\alpha} c_s r_{s-} I_{y_s\neq y}\, ds \le \alpha\int_0^t c_s\, ds.$$

The last integral being finite, this and (7.3) lead us to

$$E\int_0^{t\wedge\tau_\alpha} c_s r_{s-} \sum_{y\in Y} \big(\frac{1}{\tilde\lambda_s(y)}\, dK_s(y) - I_{y_s\neq y} ds\big) \le 0, \quad Er_{t\wedge\tau_\alpha} \le Er_0.$$

We finish the proof upon using Fatou's theorem and noticing that $\tau_\alpha \to\infty$ as $\alpha\to\infty$. The lemma is proved.

Proof of Theorem 5.2. We have

$$d\Lambda_t = e^{-\int_0^t \log\tilde\lambda_s(y_s)\, dK_s}\, de^{-\int_0^t \tilde\lambda_s(y_s)\, ds} + e^{-\int_0^t \tilde\lambda_s(y_s)\, ds}\, de^{-\int_0^t \log\tilde\lambda_s(y_s)\, dK_s}$$
$$= -\Lambda_t\tilde\lambda_t(y_t)\, dt + \Lambda_{t-}(e^{-\log\tilde\lambda_t(y_t)} - 1)\, dK_t = \Lambda_{t-}\{(\tilde\lambda_t^{-1}(y_t) - 1)\, dK_t - \tilde\lambda_t(y_t)\, dt\}.$$

Furthermore,

$$d\tilde p_t(z) = d(q_t(z)\Lambda_t) = q_{t-}(z)\, d\Lambda_t + \Lambda_t\, dq_t(z)$$
$$= q_{t-}(z)\, d\Lambda_t + \Lambda_{t-}\, dq_t(z) + \Delta\Lambda_t\, dq_t(z).$$

By (3.5)

$$\Lambda_{t-}\, dq_t(z)$$

$$= \Lambda_{t-}\{I_{y_t=z^2} \sum_{\zeta\in Z} q_t(\zeta)\lambda_{\zeta z}(t)\, dt + \sum_{\zeta\in Z} q_{t-}(\zeta)\lambda_{\zeta z}(t)\, dK_t(z^2) - q_{t-}(z)\, dK_t\}$$

$$= I_{y_t=z^2} \sum_{\zeta\in Z} \tilde{p}_t(\zeta)\lambda_{\zeta z}(t)\, dt + \sum_{\zeta\in Z} \tilde{p}_{t-}(\zeta)\lambda_{\zeta z}(t)\, dK_t(z^2) - \tilde{p}_{t-}(z)\, dK_t.$$

Use again (3.5) and the fact that $\Delta K_t\, dt = 0$ (which is true since $\Delta K_t \neq 0$ only at countably many points). Also observe that $\Delta K_t\, dK_t = dK_t$ and $\Delta K_t\, dK_t(y) = dK_t(y)$. Then we see that

$$\Delta\Lambda_t\, dq_t(z) = \Lambda_{t-}(\tilde{\lambda}_t^{-1}(y_t) - 1)\{\sum_{\zeta\in Z} q_{t-}(\zeta)\lambda_{\zeta z}(t)\, dK_t(z^2) - q_{t-}(z)\, dK_t\}$$

$$= (\tilde{\lambda}_t^{-1}(y_t) - 1)\{\sum_{\zeta\in Z} \tilde{p}_{t-}(\zeta)\lambda_{\zeta z}(t)\, dK_t(z^2) - \tilde{p}_{t-}(z)\, dK_t\}.$$

Hence

$$\Lambda_{t-}\, dq_t(z) + \Delta\Lambda_t\, dq_t(z)$$

$$= I_{y_t=z^2} \sum_{\zeta\in Z} \tilde{p}_t(\zeta)\lambda_{\zeta z}(t)\, dt + \tilde{\lambda}_t^{-1}(y_t)\{\sum_{\zeta\in Z} \tilde{p}_{t-}(\zeta)\lambda_{\zeta z}(t)\, dK_t(z^2) - \tilde{p}_{t-}(z)\, dK_t\}.$$

This allows us to obtain

$$d\tilde{p}_t(z) = \tilde{p}_{t-}(z)\{(\tilde{\lambda}_t^{-1}(y_t) - 1)\, dK_t - \tilde{\lambda}_t(y_t)\, dt\}$$

$$+ I_{y_t=z^2} \sum_{\zeta\in Z} \tilde{p}_t(\zeta)\lambda_{\zeta z}(t)\, dt + \tilde{\lambda}_t^{-1}(y_t)\{\sum_{\zeta\in Z} \tilde{p}_{t-}(\zeta)\lambda_{\zeta z}(t)\, dK_t(z^2) - \tilde{p}_{t-}(z)\, dK_t\},$$

which after further simplifications yields (cf. (4.5))

$$d\tilde{p}_t(z) = I_{y_t=z^2} \sum_{\zeta\in Z} \tilde{p}_t(\zeta)\lambda_{\zeta z}(t)\, dt - \tilde{\lambda}_t(y_t)\tilde{p}_t(z)\, dt + \tilde{\lambda}_t^{-1}(y_t) \sum_{\zeta\in Z} \tilde{p}_{t-}(\zeta)\lambda_{\zeta z}(t)\, dK_t(z^2) - \tilde{p}_{t-}(z)\, dK_t. \quad (7.6)$$

Observe that in (7.6) $\tilde{\lambda}_t^{-1}(y_t)$ may be replaced by $\tilde{\lambda}_t^{-1}(z^2)$ because $dK_t(z^2) = I_{y_t=z^2} dK_t(z^2)$. Furthermore, since $\tilde{p}_t(z)$ are nonnegative, for $p_t := \sum_z \tilde{p}_t(z)$ we get

$$dp_t \le 2p_t \sum_{\zeta,\eta\in Z} |\lambda_{\zeta\eta}(t)|\, dt + p_{t-} \sum_{\zeta,\eta\in Z} |\lambda_{\zeta\eta}(t)| \sum_{z\in Z} \tilde{\lambda}_t^{-1}(z^2)\, dK_t(z^2).$$

By Lemma 7.2 (ii) this implies that Ep_t and $E\tilde{p}_t(z)$ are bounded functions on $[0, T]$ for any $T < \infty$ and $z \in Z$.

Next as above and also similarly to (7.4) for

$$u_t(z) := \tilde{p}_t(z) e^{i\int_0^t h(s,y_s)\, dK_s}$$

we get

$$u_t(z) = u_0(z) - \int_0^t \tilde{\lambda}_s(y_s) u_s(z)\, ds - \int_0^t u_{s-}(z)\, dK_s + \int_0^t I_{y_s=z^2} \sum_{\zeta\in Z} u_s(\zeta)\lambda_{\zeta z}(s)\, ds + \int_0^t \sum_{\zeta\in Z} u_{s-}(\zeta) e^{ih(s,z^2)} \lambda_{\zeta z}(s)\, \frac{1}{\tilde{\lambda}_s^+(z^2)}\, dK_s(z^2). \quad (7.7)$$

Take expectations of these expressions and apply Corollary 6.5 splitting complex–valued functions into real and imaginary parts. Also notice that

$$E\int_0^t |\tilde{\lambda}_s(y_s)|\,|u_s(z)|\,ds \leq \sum_{\zeta,z\in Z}\int_0^t |\lambda_{\zeta z}(s)|\,ds\,\sup_{s\leq t} E\tilde{p}_s(z) < \infty,$$

$$E\int_0^t \sum_{\zeta\in Z}|u_s(\zeta)|\,|\lambda_{\zeta z}(s)|I_{y_s\neq z^2}\tilde{\lambda}_s(z^2)\,\frac{1}{\tilde{\lambda}^+_s(z^2)}\,ds$$
$$\leq E\int_0^t \sum_{\zeta\in Z}|u_s(\zeta)|\,|\lambda_{\zeta z}(s)|\,ds < \infty.$$

Then from (7.7) for $v(t,z) := Eg(y_0)u_t(z)$ we get

$$v(t,z) = v(0,z) + E\int_0^t g(y_0)I_{y_s=z^2}\sum_{\zeta\in Z}u_s(\zeta)\lambda_{\zeta z}(s)\,ds$$
$$+E\int_0^t g(y_0)\sum_{\zeta\in Z}u_s(\zeta)e^{ih(s,z^2)}\lambda_{\zeta z}(s)I_{y_s\neq z^2}\tilde{\lambda}_s(z^2)\,\frac{1}{\tilde{\lambda}^+_s(z^2)}\,ds.$$

Here is the only place where we use Assumption 3.5 which implies that if $y_s\neq z^2$, then $\tilde{\lambda}_s(z^2)/\tilde{\lambda}^+_s(z^2)=1$. Hence and from the equality $u_s(\zeta)=I_{y_s=\zeta^2}u_s(\zeta)$ it follows that

$$v(t,z) = v(0,z) + E\int_0^t g(y_0)\sum_{\zeta\in Z,\zeta^2=z^2}u_s(\zeta)\lambda_{\zeta z}(s)\,ds$$
$$+E\int_0^t g(y_0)\sum_{\zeta\in Z,\zeta^2\neq z^2}u_s(\zeta)e^{ih(s,z^2)}\lambda_{\zeta z}(s)I_{y_s\neq z^2}\,ds$$
$$= v(0,z) + \int_0^t \sum_{\zeta\in Z,\zeta^2=z^2}v(s,\zeta)\lambda_{\zeta z}(s)\,ds + \int_0^t \sum_{\zeta\in Z,\zeta^2\neq z^2}v(s,\zeta)e^{ih(s,z^2)}\lambda_{\zeta z}(s)\,ds.$$

The theorem is proved.

8. Kolmogorov's equation and infinitesimal structure of z_t

Here we give some additional information concerning the Markov process from Sec. 2 with the state space Z which is just arbitrary finite set.

Proof of Remark 2.2. First let Assumption 2.1 be satisfied. Multiply (2.4) by a bounded $\mathcal{F}^z_s$–measurable function ξ and take expectations. Then we obtain (see (2.2))

$$E\xi p(s,z_s,t,z) = E\xi E\{I_{z_t=z}|\mathcal{F}^z_s\} = E\xi I_{z_t=z} = E\xi I_{z_s=z} + E\xi\int_s^t \lambda_{z_r z}(r)\,dr.$$

Here by Fubini's theorem the last term equals

$$\int_s^t E\xi\lambda_{z_r z}\,dr = \int_s^t E\xi E\{\lambda_{z_r z}|\mathcal{F}^z_s\}\,dr = \int_s^t E\xi\sum_{\zeta\in Z}p(s,z_s,r,\zeta)\lambda_{\zeta z}(r)\,dr.$$

We can again interchange the expectation with the integral since $p(s,\eta,r,\zeta)$ is right continuous in r. Thus,

$$E\xi p(s,z_s,t,z) = E\xi I_{z_s=z} + E\xi\int_s^t\sum_{\zeta\in Z}p(s,z_s,r,\zeta)\lambda_{\zeta z}(r)\,dr. \tag{8.1}$$

For $\xi = I_{z_s=u}$ this yields

$$P(z_s = u)p(s,u,t,z) = P(z_s = u)\delta^{uz} + P(z_s = u)\int_s^t \sum_{\zeta\in Z} p(s,u,r,\zeta)\lambda_{\zeta z}(r)\,dr,$$

which proves that $p(s,u,t,v)$ satisfies Kolmogorov's equation (2.5) indeed.

Now let us start from (2.5). Then

$$p(s,z_s,t,z) = I_{z_s=z} + \int_s^t \sum_{\zeta\in Z} p(s,z_s,r,\zeta)\lambda_{\zeta z}(r)\,dr \quad \text{(a.s.)}$$

and (8.1) holds for all bounded random variables ξ. If ξ is $\mathcal{F}^z_s$–measurable, we can repeat the above argument in the backward direction finally getting that

$$E\xi\big(I_{z_t=z} - I_{z_s=z} - \int_s^t \lambda_{z_r z}(r)\,dr\big) = 0.$$

This by definition means that (2.4) is a martingale and we have proved the assertion of Remark 2.2.

Theorem 8.1. *Assume*

$$\int_t^\infty |\lambda_{z_t z_t}(s)|\,ds = \infty \quad \textit{(a.s.)} \tag{8.2}$$

for any $t < \infty$. *For* $m = 0, 1, ...$ *let*

$$Z_m := (\tau(0), z_{\tau(0)}, \tau(1), z_{\tau(1)}, ..., \tau(m), z_{\tau(m)}),$$

$$\eta_m := -\int_{\tau(m)}^{\tau(m+1)} \lambda_{z_{\tau(m)} z_{\tau(m)}}(t)\,dt.$$

Then for any $m \geq 0$ *the conditional distribution of* η_m *given* Z_m *is standard exponential (with density* e^{-x}*). Furthermore, on each set* $\{z_{\tau(m)} = \zeta\}$ *for all* $z \neq \zeta$ *we have*

$$P(z_{\tau(m+1)} = z | Z_m, \tau(m+1)) = \lambda_{\zeta z}(\tau(m+1))/|\lambda_{\zeta\zeta}(\tau(m+1))| \quad \textit{(a.s.)}. \tag{8.3}$$

Observe that as in Lemma 7.2 (i), for $t \in \mathcal{T}$ and $t > 0$ we have $\lambda_{z_{t-} z_t}(t) > 0$ and $\lambda_{z_{t-} z_{t-}}(t) < 0$ (a.s.). Therefore $|\lambda_{\zeta\zeta}(\tau(m+1))| > 0$ (a.s.) on the set $\{z_{\tau(m+1)-} = \zeta\} = \{z_{\tau(m)} = \zeta\}$. This shows that the right–hand side of (8.3) is well defined. Also notice that on the set $\{z_{\tau(m)} = \zeta\}$ we have $\{z_{\tau(m+1)} \neq \zeta\}$ so that the left hand side of (8.3) is zero if $z = \zeta$.

Lemma 8.2. *Let* $\xi_t(z)$ *be a predictable bounded function given for any* $z \in Z$ *such that* $\xi_t(z_{t-}) > -1$ *for all* $t > 0, \omega$. *Let* τ *and* γ *be finite stopping times such that* $\tau \leq \gamma$. *Let* g *be a nonnegative bounded function of* ω *such that the process* $g_t := gI_{\tau<t}$ *is predictable. Then*

$$Eg\Lambda_\tau \geq Eg\Lambda_\gamma, \tag{8.4}$$

where

$$\Lambda_t = \exp\{\int_0^t \log(1 + \xi_s(z_{s-}))\,dJ_s + \int_0^t \xi_s(z_s)\lambda_{z_s z_s}(s)\,ds\}.$$

Furthermore, if $E\Lambda_\gamma = 1$, *then* $Eg\Lambda_\tau = Eg\Lambda_\gamma$. *Finally, if* γ *is bounded, then* $E\Lambda_\gamma = 1$.

Proof. The restriction $\xi_t(z_{t-}) > -1$ implies that Λ_t is well defined and $\Lambda_t > 0$. Furthermore, by using Lemma 6.7 one easily gets

$$d\Lambda_t = \Lambda_{t-}\xi_t(z_{t-})(dJ_t + \lambda_{z_t z_t}(t)\,dt). \tag{8.5}$$

Hence

$$\begin{aligned} g\Lambda_\gamma = g\Lambda_\tau &+ \int_\tau^\gamma g\Lambda_{t-}\xi_t(z_{t-})dJ_t + \int_\tau^\gamma g\Lambda_{t-}\xi_t(z_{t-})\lambda_{z_t z_t}(t)\,dt \\ &= g\Lambda_\tau + \int_0^\gamma g_t\Lambda_{t-}\xi_t(z_{t-})dJ_t + \int_0^\gamma g_t\Lambda_{t-}\xi_t(z_{t-})\lambda_{z_t z_t}(t)\,dt. \end{aligned} \tag{8.6}$$

As in the proof of Lemma 7.2 (ii) for a constant $\alpha > 0$ define

$$\tau_\alpha = \inf\{t \geq \tau : \sup_{s \leq t} \Lambda_s + \int_\tau^t \sum_{z \in Z} |\lambda_{zz}(s)|\,ds \geq \alpha\}.$$

One can easily check that τ_α is a stopping time. Also $\tau_\alpha \geq \tau$ by definition. Now put $\gamma \wedge \tau_\alpha$ instead of γ in (8.6). Notice that $\Lambda_{s-} \leq \alpha$ for $s \leq \tau_\alpha$ and $g_t = 0$ for $t \leq \tau$. Therefore, for $K := \sup|g| \sup|\xi|$,

$$\int_0^{\gamma\wedge\tau_\alpha} |g_t\Lambda_{t-}\xi_t(z_{t-})|\,|\lambda_{z_t z_t}(t)|\,dt \leq \alpha K \int_\tau^{\tau_\alpha} \sum_{z \in Z} |\lambda_{zz}(t)|\,dt \leq \alpha^2 K.$$

By Lemma 6.4 (see (6.7)) we have $Eg\Lambda_\tau = Eg\Lambda_{\gamma\wedge\tau_\alpha}$ and we get (8.4) by Fatou's theorem.

It follows from (8.4) ($g = 1$, $\tau = 0$) that $E\Lambda_\gamma \leq 1$ for any finite stopping time γ. In particular, $E\Lambda_t \leq 1$. This implies that if γ is bounded by a constant $T < \infty$, then by Fubini's theorem

$$E\int_0^\gamma |g_t\Lambda_t\xi_t(z_{t-})|\,|\lambda_{z_t z_t}(t)|\,dt \leq K \sum_{z \in Z} \int_0^T |\lambda_{zz}(t)|\,dt < \infty$$

and from (8.6) and Lemma 6.4 we see that $Eg\Lambda_\tau = Eg\Lambda_\gamma$. In particular, $E\Lambda_\gamma = 1$.

In general case, for $0 \leq g \leq 1$ from (8.4) it follows that

$$Eg\Lambda_\tau \geq Eg\Lambda_\gamma, \quad E(1-g)\Lambda_\tau \geq E(1-g)\Lambda_\gamma. \tag{8.7}$$

By summing up we get $E\Lambda_\tau \geq E\Lambda_\gamma$. If $E\Lambda_\gamma = 1$, then this is only possible when $E\Lambda_\tau = 1$ and we have equalities instead of the inequalities in (8.7). Hence $Eg\Lambda_\tau = Eg\Lambda_\gamma$ for $0 \leq g \leq 1$ and trivially for all bounded nonnegative g as well. The lemma is proved.

Proof of Theorem 8.1. By Lemma 8.2 we have $E\Lambda_{\tau(n+1)\wedge T} = 1$ for any $n \geq 0$ and $T < \infty$. If $K \geq \xi \geq 0$ where K is a finite constant, then obviously $\Lambda_{T\wedge\tau(n+1)} \leq (1+K)^n$. In this case for $T \to \infty$ by the Dominated Convergence Theorem we get $E\Lambda_{\tau(n+1)} = 1$ and again by Lemma 8.2

$$Eg\Lambda_{\tau(m)} = Eg\Lambda_{\tau(n+1)} \tag{8.8}$$

for any $0 \leq m \leq n+1$ and bounded g such that $gI_{\tau(m)<t}$ is predictable. Since $gI_{\tau(m)<t}$ is left continuous in t we only need $gI_{\tau(m)<t}$ to be $\mathcal{F}_t^z$–measurable for any t. By the way, obviously $Z_m I_{\tau(m)<t}$ can be expressed through the values of z_s at rational $s \leq t$. Therefore this product is $\mathcal{F}_t^z$–measurable for any t. The same is true for $g(Z_m)I_{\tau(m)<t}$, where $g(t_0, z(0), t_1, z(1), ..., t_m, z(m))$ is any bounded measurable function.

Hence, from (8.8) by definition for $K \geq \xi \geq 0$ we have that

$$E\{\Lambda_{\tau(n+1)}|Z_m\} = \Lambda_{\tau(m)} \quad \text{(a.s.)}. \tag{8.9}$$

Take here $n = m$, $\xi_t(z) = 0$ for $t \leq \tau(m)$, $\xi_t(m) = c$ for $t > \tau(m)$, where c is nonrandom nonnegative constant. Then

$$E\{\exp(\log(1+c)I_{\tau(m+1)<\infty} + \int_{\tau(m)}^{\tau(m+1)} c\lambda_{z_{\tau(m)}z_{\tau(m)}}(t)\,dt)|Z_m\} = 1 \quad \text{(a.s.)}. \tag{8.10}$$

Taking expectations in (8.10) and since (8.2) implies that for $c > 0$

$$\begin{aligned} &\exp(\log(1+c)I_{\tau(m+1)<\infty} + \int_{\tau(m)}^{\tau(m+1)} c\lambda_{z_{\tau(m)}z_{\tau(m)}}(t)\,dt) \\ &= I_{\tau(m+1)<\infty}(1+c)e^{c\eta_m} + I_{\tau(m+1)=\infty}\exp(\int_{\tau(m)}^{\infty} \lambda_{z_{\tau(m)}z_{\tau(m)}}(t)\,dt) \\ &= I_{\tau(m+1)<\infty}(1+c)e^{c\eta_m} + I_{\tau(m+1)=\infty}I_{\tau(m)=\infty} \qquad \text{(a.s.)} \end{aligned}$$

we get

$$(1+c)EI_{\tau(m+1)<\infty}e^{-c\eta_m} + EI_{\tau(m+1)=\infty}I_{\tau(m)=\infty} = 1.$$

By letting $c \downarrow 0$ we get $P(\tau(m+1) < \infty) + P(\tau(m) = \infty) = 1$. Since $P(\tau(0) = \infty) = 0$, by induction we get $P(\tau(m+1) < \infty) = 1$ for all m. With this new information (8.10) yields $E\{e^{-c\eta_m}|Z_m\} = 1/(1+c)$ (a.s.). Since by [2] p. 294 Theorem 22.2, a distribution of any positive random variable is uniquely determined by its moment generating function, we get our assertion about the distribution of η_m given Z_m.

In particular, $E\eta_m = 1$ for $m \geq 0$ and

$$\begin{aligned} E\int_0^{\tau(m)} g_t\lambda_{z_t z}(t)I_{z_t\neq z}\,dt &\leq \sup|g|\sum_{z\in Z} E\int_0^{\tau(m)} \lambda_{z_t z}(t)I_{z_t\neq z}\,dt \\ &= -\sup|g|E\int_0^{\tau(m)} \lambda_{z_t z_t}(t)\,dt = m\sup|g| < \infty \end{aligned}$$

if g is a bounded process. If in addition it is predictable, then by Lemma 6.4 equalities (6.6) and (6.7) hold. As above we convince ourselves that for any $z \in Z$

$$E\int_{\tau(m)}^{\tau(m+1)} f(Z_m,t)\,dJ_t(z) = E\int_{\tau(m)}^{\tau(m+1)} f(Z_m,t)I_{z_t\neq z}\lambda_{z_t z}(t)\,dt,$$

$$E\int_{\tau(m)}^{\tau(m+1)} f(Z_m,t)\,dJ_t = -E\int_{\tau(m)}^{\tau(m+1)} f(Z_m,t)\lambda_{z_t z_t}(t)\,dt$$

if the function $f(\cdot,\cdot)$ is nonrandom, nonnegative, and measurable. Hence, for $\zeta \neq z$

$$\begin{aligned} &Ef(Z_m,\tau(m+1))|\lambda_{\zeta\zeta}(\tau(m+1))|I_{z_{\tau(m)}=\zeta}I_{z_{\tau(m+1)}=z} \\ &= E\int_{\tau(m)}^{\tau(m+1)} f(Z_m,t)|\lambda_{\zeta\zeta}(t)|I_{z_{\tau(m)}=\zeta}\,dJ_t(z) \\ &= E\int_{\tau(m)}^{\tau(m+1)} f(Z_m,t)|\lambda_{\zeta\zeta}(t)|I_{z_{\tau(m)}=\zeta}\lambda_{\zeta z}(t)\,dt \\ &= E\int_{\tau(m)}^{\tau(m+1)} f(Z_m,t)I_{z_{\tau(m)}=\zeta}\lambda_{\zeta z}(t)\,dJ_t \\ &= Ef(Z_m,\tau(m+1))I_{z_{\tau(m)}=\zeta}\lambda_{\zeta z}(\tau(m+1)). \end{aligned}$$

Because of the arbitrariness of f we conclude that on the set $\{z_{\tau(m)} = \zeta\}$ we have

$$P(z_{\tau(m+1)} = z | Z_m, \tau(m+1)) |\lambda_{\zeta\zeta}(\tau(m+1))| = \lambda_{\zeta z}(\tau(m+1))$$

and this is (8.3). The theorem is proved.

References

[1] L. Aggoun, R.L. Elliott, *Finite-dimensional models for Hidden Markov chains*, Adv. Appl. Prob. 27, 146-160, 1995.

[2] P. Billingsley, *Probability and measure*, Second Edition, John Willey & Sons, 1986.

[3] P. Brémaud, *Point processes and queues, martingale dynamics*, Springer-Verlag, New York, Heidelberg, Berlin, 1981.

[4] J.L. Doob, *Stochastic processes*, Wiley, New York, 1953.

[5] R.J. Elliott, L. Aggoun, J.B. Moore, *Hidden Markov models*, Springer, New York, 1995.

[6] R. S. Liptser, A. N. Shiryayev, *Theory of martingales*, Kluver, Dordrecht–Boston, 1989.

[7] R. S. Liptser, A. N. Shiryayev, *Statistics of random processes*, Vol. 1, Springer-Verlag, New York , 1977-1978.

[8] R. S. Liptser, A. N. Shiryayev, *Statistics of random processes*, Vol. 2, Springer-Verlag, New York, 1977-1978.

[9] N. Portenko, H. Salehi, A. Skorokhod, *On optimal filtering of multitarget tracking systems based on point processes observations*, Random Oper. and Stoch. Equ., Vol. 5, No. 1, pp. 1-34, 1997.

127 Vincent Hall, University of Minnesota, Minneapolis, MN, 55455
E-mail address: `krylov@math.umn.edu`

127 Vincent Hall, University of Minnesota, Minneapolis, MN, 55455
E-mail address: `zatezalo@math.umn.edu`