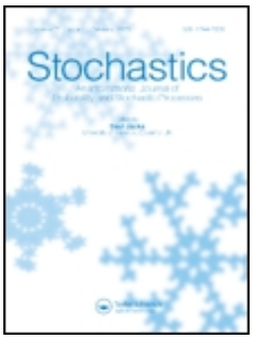

Communications in Statistics - Simulation and Computation

ISSN: 1045-1129 (Print) (Online) Journal homepage: www.tandfonline.com/journals/gssr19

Minimality and reductibility of conditionally poisson systems with finite state space

Peter Spreij

To cite this article: Peter Spreij (1990) Minimality and reductibility of conditionally poisson systems with finite state space, Communications in Statistics - Simulation and Computation, 31:1-4, 55-77, DOI: 10.1080/03610919008833648

To link to this article: https://doi.org/10.1080/03610919008833648

Published online: 02 Nov 2010.

Submit your article to this journal

Article views: 9

View related articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=gssr20

Stochastics and Stochastics Reports, Vol. 31, pp. 55–77
Reprints available directly from the publisher
Photocopying permitted by license only

© 1990 Gordon and Breach, Science Publishers, Inc.
Printed in the United Kingdom

MINIMALITY AND REDUCIBILITY OF CONDITIONALLY POISSON SYSTEMS WITH FINITE STATE SPACE

PETER SPREIJ

Department of Econometrics, Free University, P.O. Box 7161, 1007 MC Amsterdam, The Netherlands

(Received 15 August 1989; in final form 16 November 1989)

In this paper we consider stochastic systems with finite state space and counting process output. In particular we address the question whether a given system has a minimal representation, where roughly speaking minimality means minimality of the size of the state space. We show that minimality is connected to a suitably defined notion of observability. Finally we present an algorithm that enables us, starting from a given representation, to construct a minimal representation for the same system.

KEY WORDS: Stochastic system, counting process system, reducibility, minimality, observability.

INTRODUCTION

In this paper we treat some problems for counting process systems with a finite state space. The main problem we address is the characterization of minimality of a system, which means minimality of the state space. The reason why this topic is important lies partly in identification problems for such systems in the situation where the state process cannot directly be observed. It is known for instance in deterministic linear system theory that a state space, which is too large for explaining the behaviour of the output process, contains unobservable components. This implies among other things that if one wants to perform output-based parameter estimation one will not be able to identify the true parameter values that govern the behaviour of the state process in an unobservable part of the state space. For counting process systems to be treated in the next section a similar reasoning holds. If for instance one wants to identify transition rates of the state process (which turns out to be a Markov process) and if two different states yield the same behaviour of the observed counting process, then one is clearly not able to distinguish whether the state process assumes one of these two values, let alone that one is able to draw reliable conclusions about rates that govern a transition from one of these states to the other one. The lesson of these considerations, as is well known, is that one should always work with minimal representation of a stochastic system.

1. COUNTING PROCESS SYSTEMS

Counting process systems form a subclass of what is known as stochastic systems.

Roughly speaking a stochastic system without input consists of two stochastic processes X and Z where X is called the state process and Z the output process. As in deterministic system theory the state process at time t should summarize all the relevant information about the past of the system in order to describe the future output. Contrary to what can be done in deterministic system theory the state process at time t cannot exactly predict the values of Z_s for $s \geqq t$. It can only describe the probabilistic behaviour of the output process. These notions are made precise in Definition 1.1 that in abstract terms describes what a stochastic system without inputs is. This definition is followed by a more detailed treatment of stochastic systems where the output process is a counting process. First we have to introduce some notation. Let a complete probability space $(\Omega, \mathcal{F}, P)$ be given together with a filtration $\mathbb{F}$. Let X and Z be $\mathbb{F}$-adapted stochastic processes. Then $\mathcal{F}_t^X = \sigma\{X_s, s \leqq t\}$ and $\mathcal{F}_t^Z = \sigma\{Z_s, s \leqq t\}$ are the σ-algebras generated by the past of the processes X and Z. Similarly $\mathcal{F}_t^{X+} = \sigma\{X_s, s \geqq t\}$ contains the information of the future of X after t. We also use the σ-algebra that describes the future increments of the output process Z, $\mathcal{F}_t^{\Delta Z+} = \sigma\{Z_s - Z_t, s \geqq t\}$.

If $\mathcal{F}_1$, $\mathcal{F}_2$ and $\mathcal{G}$ are sigma algebras contained in $\mathcal{F}$, then we say that $\mathcal{F}_1$ and $\mathcal{F}_2$ are conditionally independent given $\mathcal{G}$, if for all integrable $\mathcal{F}_1$-measurable functions X_1 the following relation holds

$$E[X_1 | \mathcal{F}_2 \vee \mathcal{G}] = E[X_1 | \mathcal{G}].$$

We will use the notation $(\mathcal{F}_1, \mathcal{F}_2 | \mathcal{G}) \in CI$.

DEFINITION 1.1 (van Schuppen [5]) A *continuous time stochastic system* is a multiple $(\Omega, \mathcal{F}, P, T, \mathbb{F}, X, Z, \mathcal{X}, \mathcal{Y})$ such that

i) $(\Omega, \mathcal{F}, P)$ is a complete probability space;

ii) $T \subset \mathbb{R}$, T an interval;

iii) $\mathbb{F} = \{\mathcal{F}_t\}_{t \in T}$ filtration on $(\Omega, \mathcal{F}, P)$;

iv) X and Z are $\mathbb{F}$-adapted processes with values in the measurable spaces $\mathcal{X}$ and $\mathcal{Y}$ respectively;

v) $(\mathcal{F}_t^{X+} \vee \mathcal{F}_t^{\Delta Z+}, \mathcal{F}_t | \sigma(X_t)) \in CI$ for all $t \geqq 0$.

Formally speaking each of the components of the multiple in Definition 1.1 is part of the definition. However if no confusion can arise we will often write (X, Z) for a stochastic system. The crucial property in the definition of a stochastic system is (v), which says that given a whole past $\mathcal{F}_t$ it is sufficient to use only X_t for the prediction of the future values of X and the future increments of Z. Observe that 1.1(v) implies that X is a Markov process with respect to the filtration $\mathbb{F}$. Finally it is noticed that usually $\mathcal{F}_t = \mathcal{F}_t^X \vee \mathcal{F}_t^Z$ and $T = (-\infty, \infty)$ or $T = [0, \infty)$.

Clearly the above definition is too abstract for practical purposes. In particular cases one has to specify the distribution of the state and output process. One way to do this is to pose stochastic differential equations that X and Z satisfy. In this

paper we treat stochastic systems where the output process is a counting process and X a finite state process.

DEFINITION 1.2 A *counting process system* is a stochastic system where the output Z is a counting process. We write in this case N for the output process instead of Z. The shorthand notation is then (X, N) for a counting process system with state process X.

We will treat in more detail the class of conditionally Poisson systems.

DEFINITION 1.3 (Brémaud [1]) *Let* $N:\Omega\times[0,\infty)\to\mathbb{N}$ be a counting process, $\mathbb{F}$-adapted with Doob–Meyer decomposition w.r.t. $\mathbb{F}: dN_t=\lambda_t\,dt+dm_t$. Let $\mathscr{F}_\infty^\lambda=\sigma\{\lambda_t, t\geqq 0\}$. $\mathbb{N}$ is called a *conditionally Poisson process*, or a *doubly stochastic Poisson process*, iff for all $t, h\geqq 0$, $u\in\mathbb{R}$

$$E[\exp(iu(N_{t+h}-N_t))\,|\,\mathscr{F}_t\vee\mathscr{F}_\infty^\lambda]=\exp\left((e^{iu}-1)\int_t^{t+h}\lambda_s\,ds\right). \tag{1}$$

So conditioned upon $\mathscr{F}_t\vee\mathscr{F}_\infty^\lambda N_{t+h}-N_t$ has a Poisson distribution with mean $\int_t^{t+h}\lambda_s\,ds$.

PROPOSITION 1.4 *N is a conditionally Poisson process iff m as given in 1.3 is a martingale w.r.t.* $\tilde{\mathbb{F}}=\{\tilde{\mathscr{F}}_t\}_{t\geqq 0}$, *where* $\tilde{\mathscr{F}}_t=\mathscr{F}_t\vee\mathscr{F}_\infty^\lambda$.

Proof If N is conditionally Poisson, then

$$E[m_{t+h}-m_t\,|\,\tilde{\mathscr{F}}_t]=E[N_{t+h}-N_t\,|\,\tilde{\mathscr{F}}_t]-E\left[\int_t^{t+h}\lambda_s\,ds\,\middle|\,\tilde{\mathscr{F}}_t\right]$$

$$=\int_t^{t+h}\lambda_s\,ds-\int_t^{t+h}\lambda_s\,ds=0.$$

Conversely assume that m is a martingale w.r.t. $\tilde{\mathbb{F}}$. Apply the stochastic calculus rule to $\exp(iuN_t)$ to obtain

$$\exp(iuN_{t+h})=\exp(iuN_t)+(e^{iu}-1)\int_t^{t+h}\exp(iuN_{s-})\,dN_s$$

$$=\exp(iuN_t)+(e^{iu}-1)\int_t^{t+h}\exp(iuN_{s-})(\lambda_s\,ds+dm_s).$$

Take conditional expectation w.r.t. $\tilde{\mathscr{F}}_t$ and get

$$E[\exp(iuN_{t+h})\,|\,\tilde{\mathscr{F}}_t]=\exp(iuN_t)+(e^{iu}-1)\int_t^{t+h}E[\exp(iuN_s)\,|\,\tilde{\mathscr{F}}_t]\lambda_s\,ds.$$

Define $g(t+h,t)=E[\exp(iu(N_{t+h}-N_t))|\tilde{\mathcal{F}}_t]$. Then we get

$$g(t+h,t)=1+(e^{iu}-1)\int_t^{t+h} g(s,t)\lambda_s\,ds,$$

from which we find $g(t+h,t)=\exp((e^{iu}-1)\int_t^{t+h}\lambda_s\,ds)$. ∎

Next we present a method for the construction of a counting process system. Let a probability space $(\Omega,\mathcal{F},P_0)$ be given together with a standard Poisson process N and a Markov process X (with state space $\mathcal{X}$) defined on it such that N and X are independent processes. Notice that such a probability space always exists. We assume that X has cadlag paths. Consider the following filtrations: $\mathbb{F}^N,\mathbb{F}^X,\mathbb{F}=\{\mathcal{F}_t^N\vee\mathcal{F}_t^X\}_{t\geq 0}$, $\tilde{\mathbb{F}}=\{\mathcal{F}_t^N\vee\mathcal{F}_\infty^X\}_{t\geq 0}$. The following observation is important. Let $\tilde{m}_t=N_t-t$. By definition $\tilde{m}$ is an $\mathbb{F}^N$-martingale. However because of the independence assumption $\tilde{m}$ is also an $\mathbb{F}$- and $\tilde{\mathbb{F}}$-martingale. Similarly X is also Markov with respect to the filtration $\mathbb{F}$. Let $\lambda:[0,\infty)\times\mathcal{X}\to(0,\infty)$ be a measurable function such that $E_0\int_0^t\lambda(s,X_s)\,ds<\infty$, $\forall t$. Write $\lambda_t=\lambda(t,X_{t-})$. Then $\{\lambda_t\}$ is clearly both $\mathbb{F}$ and $\tilde{\mathbb{F}}$-predictable. Then M defned by $M_t=\int_0^t(\lambda_s-1)\,d\tilde{m}_s$ is an $\tilde{\mathcal{F}}$-martingale and let $\Lambda_t=\mathcal{E}(M_t)$. Then

$$\Lambda_t=\exp\left(\int_0^t\log\lambda_s\,dN_s-\int_0^t(\lambda_s-1)\,ds\right)$$

and Λ is an $\mathbb{F}$- and $\tilde{\mathbb{F}}$-local martingale. We make the following assumption: $E_0\Lambda_t=1$, $\forall t\geq 0$. We can now define a new measure P on $(\Omega,\tilde{\mathcal{F}}_\infty)=(\Omega,\mathcal{F}_\infty)$ as follows. If $A\in\tilde{\mathcal{F}}_t$ then by definition $P(A)=E_0[1_A\Lambda_t]$. The extension to $\tilde{\mathcal{F}}_\infty$ follows by Caratheodory's theorem. Observe that the restriction of P to $\tilde{\mathcal{F}}_t$ is absolutely continuous with respect to the restriction of P to $\tilde{\mathcal{F}}_t$ with Λ_t as Radon–Nikodym derivative and that $\Lambda_t>0$ P_0 a.s. Observe also that the restrictions of P and P_0 to $\mathcal{F}_\infty^X$ coincide.

PROPOSITION 1.5 *Under the new measure P*

i) $m_t=N_t-\int_0^t\lambda_s\,ds$ *defines a martingale with respect to* $\mathbb{F}$ *and* $\tilde{\mathbb{F}}$;

ii) X *is a Markov process with respect to* $\mathbb{F}$.

Proof The first assertion follows from Girsanov's theorem (van Schuppen and Wong [6]). So here we prove only (ii). Let f be a bounded measurable function on $\mathcal{X}$ and $h>0$. Then because Λ_t is the Radon–Nikodym derivative $dP|\tilde{\mathcal{F}}_t/dP_0|\tilde{\mathcal{F}}_t$

$$E[f(X_{t+h})|\mathcal{F}_t]=\frac{E_0[f(X_{t+h})\Lambda_t|\tilde{\mathcal{F}}_t]}{E_0[\Lambda_t|\tilde{\mathcal{F}}_t]}=E_0[f(X_{t+h})|\tilde{\mathcal{F}}_t]=E_0[f(X_{t+h})|\sigma(X_t)].$$

In the second equality we have used the fact Λ_t is $\mathscr{F}_t$-measurable and in the third one that X is $\mathbb{F}$-Markov under P_0. ∎

THEOREM 1.6 *Under the new measure P the pair (X, N) forms a stochastic system.*

Proof From part (i) of Proposition 1.5 and Proposition 1.4 we obtain that N is conditionally Poisson. Notice that we even have

$$E[\exp(iu(N_{t+h}-N_t))\,|\,\mathscr{F}_t] = \exp\left((e^{iu}-1)\int_t^{t+h}\lambda_s\,ds\right).$$

Hence

$$E[\exp(iuN_{t+h}-N_t))\,|\,\mathscr{F}_t \vee \mathscr{X}_t^X] = E[\exp(iu(N_{t+h}-N_t))\,|\,\mathscr{F}_t^{X+}]$$

which shows that

$$(\mathscr{F}_t^{\Delta N+}, \mathscr{F}_t\,|\,\mathscr{F}_t^{X+}) \in CI, \qquad \forall\, t \geqq 0.$$

The fact that X is $\mathbb{F}$-Markov yields

$$(\mathscr{F}_t^{X+}, \mathscr{F}_t\,|\,\sigma(X_t)) \in CI, \qquad \forall\, t \geqq 0.$$

Now we can use the following result which is obvious. Let F_1, F_2, G be σ-algebras. Then $(F_1, F_2\,|\,G) \in CI$ and $(F_1, F_3\,|\,G \vee F_2) \in CI$ is equivalent with $(F_1, F_2 \vee F_3\,|\,G) \in CI$. In our case we take $G = \sigma(X_t)$, $F_2 = \mathscr{F}_t^{X+}$, $F_1 = \mathscr{F}_t$ and $F_3 = \mathscr{F}_t^{\Delta N+}$ and we obtain $(\mathscr{F}_t, \mathscr{F}_t^{X+} \vee \mathscr{F}_t^{\Delta N+}\,|\,\sigma(X_t)) \in CI$. ∎

Thus we have constructed a stochastic system where (as always) X is a Markov process and the output process is a conditional Poisson process. Notice that so far we have used an evolution equation for N whereas for X we only have the Markov property. The next objective is to describe the evolution of X in terms of a stochastic differential equation. Throughout the rest of this chapter the following assumption will be in force.

Assumption 1.7 The state process X takes its values in the finite set $\mathscr{X} = \{x_1, \ldots, x_n\}$, where the x_i are different. Moreover for all i and $t > 0$: $P(X_t = x_i) > 0$.

Define $Y: \Omega \times [0, \infty) \to \{0, 1\}^n$ by its components $Y_{it} := 1_{\{X_t = x_i\}}$. Denote by $\Phi(t, s)$ the matrix of transition probabilities of X. That is for $t \geqq s$, with the notation $z^+ = z^{-1}1_{\{z \neq 0\}}$ and the understanding $0/0 = 0$

$$\Phi_{ij}(t, s) = P(X_t = x_i\,|\,X_s = x_j) = (EY_{js})^+ E(Y_{js}Y_{it}).$$

Then we have the following well-known facts. Semigroup property: $\Phi t, s) = \Phi(t, u)\Phi(u, s)$ for $t \geqq u \geqq s$. Assume that for all $t \geqq 0$ the following limit exists

$$A(t) := \lim_{h \downarrow 0} \frac{1}{h} [\Phi(t+h, t) - I].$$

$A(t)$ will be called the generator of X at time t. So $A(t)$ has nonpositive diagonal elements, the other entries are nonnegative and the column sums are zero.

Proposition 1.8 gives a representation of Markov processes in terms of a linear stochastic differential equation driven by a martingale.

PROPOSITION 1.8 *Let $X: \Omega \times [0, \infty) \to \mathcal{X}$ be a stochastic process, $\mathbb{F}$-adapted and let Y be associated with X as before. Assume that Y satisfies*

$$dY_t = A(t) Y_t \, dt + dm_t^Y. \tag{2}$$

Here $A: [0, \infty) \to \mathbb{R}^{n \times n}$ is a Lebesgue measurable function (deterministic!) and m^Y an $\mathbb{F}$-adapted martingale. Then X and Y are $\mathbb{F}$-Markov processes, with generator $A(t)$. Conversely, if X is $\mathbb{F}$-Markov with generator $A(t)$, then Y satisfies (2).

Proof (Spreij [4]) ∎

Next we give a result on Markov solutions of stochastic differential equations (see also Protter [3] for related problems).

PROPOSITION 1.9 *Let X be the solution of the stochastic differential equation*

$$dX_t = g(t, X_t) \, dt + dm_t^X, X_0 \tag{3}$$

where m^x is an $\mathbb{F}$-martingale and $g: [0, \infty) \times \mathcal{X} \to \mathbb{R}$. Assume that the jump measure μ of X admits a compensator ν (with respect to $\mathbb{F}$ and P) such that $\nu(dt, dy, \omega) = p(t, X_t(\omega), dy) \, dt$. Then X is an $\mathbb{F}$-Markov process.

Proof We show that for the indicator process Y the representation of Proposition 1.8 holds. From (3) we get the stochastic calculus rule for all $k \geqq 0$:

$$\begin{aligned} dX_t^k &= k X_{t-}^{k-1} \, dX_t + \int_{\mathcal{X}} [(X_{t-} + y)^k - X_{t-}^k - k X_{t-}^{k-1} y] \mu(dt, dy) \\ &= k X_t^{k-1} \left(g(t, X_t) + \int_{\mathcal{X}} [(X_t + y)^k - X_t^k - k X_t^{k-1} y] p(t, X_t, dy) \right) dt + d\tilde{m}_t^{(k)}. \end{aligned} \tag{4}$$

Here $d\tilde{m}_t^{(k)}$ summarizes all the martingale terms in (4). In a more compact notation we can write (4) as

$$dX_t^k = g^{(k)}(t, X_t) \, dt + d\tilde{m}_t^{(k)} \tag{5}$$

where $g^{(k)}: [0, \infty) \times \mathcal{X} \to \mathbb{R}$. Now we can write X_t^k as $[x_1^k, \ldots, x_n^k] Y_t$ and $g^{(k)}(t, X_t)$ as $G^{(k)}(t) Y_t$ where $G^{(k)}(t) = [g^{(k)}(t, x_1), \ldots, g^{(k)}(t, x_n)]$. Introduce the following notation. V is the $(n \times n)$ matrix with kth row equal to $[x_1^{k-1}, \ldots, x_n^{k-1}]$ $(k = 1, \ldots, n)$. $G(t)$ is the $(n \times n)$ matrix with kth row $G^{(k-1)}(t)$ $(k = 1, \ldots, n)$. $\tilde{M}_t$ is the martingale with

components $\tilde{m}_t^{(k)}$. If we consider (5) as a system of equations for $k=0,\dots,n-1$ we can summarize it (with $G(t)$ and V as defined above) as

$$V\,dY_t = G(t)Y_t\,dt + d\tilde{M}_t. \tag{6}$$

Observe that V is a Vandermonde matrix, that is nonsingular because all the x_i are different. Let $A(t)=V^{-1}G(t)$ and $M_t^Y = V^{-1}\tilde{M}_t$ then (6) becomes

$$dY_t = A(t)Y_t\,dt + dM_t^Y. \tag{7}$$

Because M is an $\mathbb{F}$-martingale and $A(t)$ is nonrandom, we obtain from (7) by applying Proposition 1.8 that X is $\mathbb{F}$-Markov, with generator $A(t)$. ∎

If we collect the above results we get the following

THEOREM 1.10 *Let the process X and the counting process N satisfy the following equation*

$$dX_t = g(t, X_t)\,dt + dm_t^X, \qquad X_0$$

$$dN_t = \lambda(t, X_t)\,dt + dm_t, \qquad N_0 = 0.$$

Here λ and g are measurable functions from $[0,\infty)\times\mathcal{X}$ to $\mathbb{R}$ and $\mathbb{R}^+$ respectively and m^X and m are $\mathbb{F}$-martingales. Assume moreover that m is a martingale with respect to $\tilde{\mathbb{F}} = \{\mathcal{F}_t^N \vee \mathcal{F}_\infty^X\}$ and that the jump measure μ of X admits a compensator ν of the form $\nu(dt, dy, \omega) = p(t, X_t(\omega), dy)\,dt$. Then the pair (X, N) is a counting process system.

2. MINIMALITY OF CONDITIONALLY POISSON SYSTEMS

In this section we will confine ourselves to stationary systems., This means that the functions A, g and λ in Theorem 1.10 are not explicitly dependent on t. So we use the representations

$$dY_t = AY_t\,dt + dM_t^Y, \qquad Y_0 \tag{8a}$$

$$dN = CY_t\,dt + dm_t, \qquad N_0 = 0. \tag{8b}$$

Here C is a row vector in $\mathbb{R}^n$ with elements $c_i = \lambda(x_i)$.

Equation (8) is called the *forward representation* of the system (X, N). It is also possible to give a *backward representation.* The starting point of this section is the system Eq. (8). The word minimality in the title refers to the minimality of size of the state space $\mathcal{X}$ in a way to be made precise below. The external behaviour of the system (X, N) is the process N. We call (X, N) minimal if we cannot find a system $(\bar{X}, N)$ where $\bar{X}$ has a smaller state space than X. Observe that the external behaviours $(\bar{X}, N)$ and (X, N) are both given by the same process N. For $(\bar{X}, N)$ we use Eq. (8) with Y, A and C replaced with $\bar{Y}$, $\bar{C}$, $\bar{A}$.

DEFINITION 2.1 The forward representation (8) of the system (X, N) is called *strongly reducible* if there exists a set $\bar{\mathcal{X}}$ of lower cardinality than $\mathcal{X}$ and a function $f:\mathcal{X}\to\bar{\mathcal{X}}$ such that with $\bar{X}_t=f(X_t)$, the pair $(\bar{X}, N)$ is a stochastic system with a forward representation of the form (8) and such that $CY_t=\bar{C}\bar{Y}_t$. In this case (X, N) is called *strongly forwardly reducible.* If (X, N) is not strongly forwardly reducible, it is called *strongly forwardly minimal.*

Some remarks are appropriate.

1) If (X, N) is strongly reducible then the "new" state process $\bar{X}$ is again Markov.

2) The adverb strongly in Definition 2.1 can be thought of as opposed to weakly. One may call a system weakly reducible if there exists a counting process system $(\bar{X}, \bar{N})$ on some possibly different probability space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P})$ such that the state space of $\bar{X}$ has strictly smaller cardinality than that of $\bar{X}$ and such that $\bar{N}$ is equal to N in distribution. One can also define strong reducibility for the backward representation of (X, N). We will not treat weak problems and problems for the backward representation. For this reason we will speak of minimality and reducibility throughout this section when we mean strong forward minimality and strong forward reducibility.

The problem that we want to treat is the characterization of minimal counting process systems. In view of Remark 1 above we first focus our attention on functions of a Markov process.

From the equivalence of $\mathbb{F}$-Markov processes and solutions of certain linear stochastic differential equations (Propositions 1.7 and 1.8) it is easy to see when functions of a Markov chain again yield a Markov chain. A similar result also holds for nonstationary chains (Spreij [4]).

To be specific let as before X be a $\mathbb{F}$-Markov chain with state space $\mathcal{X}$. Let H be another set and $f:\mathcal{X}\to H$ a function. Clearly $f(X)$ is again Markov if f is injective. To avoid trivialities let us assume that $H=\{h_1, \ldots, h_m\}$, $m<n$ and that f is onto. Write $Z_t=f(X_t)$. Associate with Z the indicator process W as usual:

$$W:\Omega\times[0,\infty)\to\{0,1\}^m, \quad W_{it}=1_{\{Z_t=h_i\}}.$$

Define $F\in\mathbb{R}^{m\times n}$ by $F_{ij}=1_{\{f(x_j)=h_i\}}$. Notice that $\mathbf{1}_m^T F=\mathbf{1}_n^T$, where $\mathbf{1}_m$ is a column vector with as elements $+1$. Then $W_t=FY_t$. Notice that because f is onto, F has rank m, i.e. it has full row rank. Let $K\in\mathbb{R}^{n\times(n-m)}$ be a fixed matrix such that its columns span $\operatorname{Ker} F$. Let as before A be the matrix of transition intensities of X. We have the following theorem, similar to a result in discrete time (Kemeny and Snell [2, p. 126]).

THEOREM 2.2 *Let X be $\mathbb{F}$-Markov with finite state space $\mathcal{X}$. Let $f:\mathcal{X}\to H$. Then $f(X)$ is again $\mathbb{F}$-Markov iff $FAK=0$ where the columns of K span* $\operatorname{Ker} F$ *and F is related to f as indicated above. If this condition is satisfied, then the matrix B of transition intensities of $f(X)$ is given by $B=FA\hat{F}$, where $\hat{F}$ is any right inverse of F.*

Proof (Spreij [4]) ∎

We will work with a special right inverse of F, the Moore–Penrose inverse which is defined as $F^+=F^T(FF^T)^{-1}$. Because of the prominent role that matrices

F as defined before play, we will refer to these as reduction matrices. Observe that the only invertible transformations of the state space $\mathcal{X}$ are permutations, which correspond to special reduction matrices F, the permutation matrices, that also have the property $F\mathbf{1}=\mathbf{1}$.

PROPOSITION 2.3 *The pair (X, N) is reducible iff there exists a reduction matrix F such that with $\bar{A}=FAF^+$, $\bar{C}=CF^+$ the equalities $FA=\bar{A}F$ and $C=\bar{C}F$ hold. Moreover for the reduced system $(\bar{X}, N)$ the generator of $\bar{X}$ is $\bar{A}$ and the intensity of N is given by $\bar{C}\bar{Y}_t$. In this case one says that F reduces (X, N).*

Proof Obvious in view of Remark 1 after Definition 2.1.

Remark Observe that from purely algebraic considerations $FA=\bar{A}F$ implies that indeed $\bar{A}$ is a generating matrix of some Markov process. Indeed, let $k=k(j)$ be the unique integer such that $F_{kj}=1$. Then

$$\bar{A}_{ik}=\sum_l F_{il}A_{lj}=\sum_{l\neq j} F_{il}A_{lj}+F_{ij}A_{jj}.$$

Now if $i=k$, then $\bar{A}_{ik}=\sum_{l\neq j}F_{il}A_{lj}+A_{jj}\leqq\sum_{l\neq j}A_{lj}+A_{jj}=0$. And if $i\neq k$, then $\bar{A}_{ik}=\sum_{l\neq j}F_{il}A_{lj}\geqq 0$. Furthermore $\mathbf{1}^T\bar{A}=0$. Observe also that $FA=\bar{A}F$ is equivalent with $F\Phi(t)=\bar{\Phi}(t)F$, where $\Phi(t)=\exp(At)$ and $\bar{\Phi}(t)=\exp(\bar{A}t)$.

Since the stochastic nature of the pair (X, N) is determined by the pair (A, C) in view of Eq. (8), we will often speak of minimality or reducibility of (A, C) instead of (X, N).

Observe that the reduction procedure is transitive, which means the following. Suppose F_1 reduces (X, N) into a new system (X_1, N) and suppose that F_2 reduces (X_1, N). Then F_2F_1 reduces the original system (X, N). Indeed if F_1 reduces (X, N) then $F_1A=A_1F_1$ for $A_1=F_1AF_1^+$ and $C=C_1F_1$ for $C_1=CF_1^+$. If then also F_2 reduces (X_1, N), then we can write $F_2A_1=A_2F_2$ and $C_1=C_2F_2$. But then $F_2F_1A=F_2A_1F_1=A_2F_2F_1$ and $C=C_1F_1=C_2F_2F_1$ which is what we have to prove. Notice however that given a reduction matrix F that reduces (X, N) one cannot always decompose F as $F=F_2F_1$, where F_1 reduces (X, N) and F_2 reduces (X_1, N). A simple example is the following. Suppose that X has generator

$$\mathrm{A}=\begin{bmatrix}-2 & 1 & 2\\ 1 & -3 & 2\\ 1 & 2 & -4\end{bmatrix}$$

and N has constant intensity $\lambda=\lambda\mathbf{1}^TY_t$. Then clearly $F=[1\ 1\ 1]$ reduces (X, N) but no reduction matrix $F\in\mathbb{R}^{2\times 3}$ reduces (X, N) as can easily be checked.

DEFINITION 2.4 Let the row vector $C\in\mathbb{R}^n$ be given. Then D is defined to be the diagonal matrix diag(C) which has as the jth diagonal element c_j. For $u\in\mathbb{R}$, $D(u)=(e^{iu}-1)D$.

LEMMA 2.5 *Let F be a reduction matrix, with right inverse F^+ and let K be a*

matrix whose columns span $\operatorname{Ker} F$. *Let* $\bar{C}=CF^+$ *and* $\bar{D}=FDF^+$. *The following statements are equivalent*

i) $C=\bar{C}F$;

ii) $FDK=0$;

iii) $FD=\bar{D}F$.

Proof (i)⇒(ii):

$$(FDK)_{ij}=\sum_k F_{ik}c_kK_{kj}=\sum_{k,l} F_{ik}\bar{c}_lF_{ik}K_{kj}.$$

Because of the special form of the matrix F, there is only one nonzero element in each column. Hence a product $F_{ik}F_{il}$ equals zero if $i\neq l$. Therefore the last summation can be written as

$$\sum_k F_{ik}\bar{c}_iF_{ik}K_{kj}=\bar{c}_i\sum_k F_{ik}^2K_{kj}=\bar{c}_i\sum_k F_{ik}K_{kj}=0.$$

(ii)⇒(iii) $FDK=0$ means that FD is contained in the left kernel of K which is F. Hence there is a matrix L such that $FD=LF$. But then by postmultiplying with F^+ we obtain $L=FDF^+=\bar{D}$.

(iii)⇒(i) $Fd=\bar{D}F$ implies that $\mathbf{1}^TFD=\mathbf{1}^T\bar{D}F$ or $\mathbf{1}^TD=\mathbf{1}^T\bar{D}F$. However $\mathbf{1}^TD=C$ and $\mathbf{1}^T\bar{D}=\bar{C}$. ∎

Remark Assume that $FD=\bar{D}F$ for some reduction matrix F. Then $\bar{D}$ is necessarily diagonal. Indeed we have from this assumption: $F_{ij}c_j=\bar{D}_{ik}F_{kj}+\sum_{l\neq j}\bar{D}_{il}F_{lj}$. Assume that $i\neq k$ and multiply this equation with F_{kj}. Then, since $F_{ij}F_{kj}=0$ for $i\neq k$ we have $0=\bar{D}_{ik}F_{kj}$, and hence $\bar{D}_{ik}\sum_j F_{kj}=0$. Since the summation $\sum_j F_{kj}\geqq 1$ for all k, we have $\bar{D}_{ik}=0$.

LEMMA 2.6 *Let* F *and* K *be as in Lemma 2.5 and let* e_i *be the* ith *basis vector of* $\mathbb{R}^n$. *Let* (X,N) *be a stochastic system as in* (8). *Assume that* $FAK=0$. *Then* F *reduces* (X,N) *if* C *is such that* $Fe_k=Fe_j$ *for some* k *and* j *implies* $c_k=c_j$.

Proof We only have to prove that we can write $C=\bar{C}F$, where $\bar{C}=CF^+$. Observe first that

$$(FF^T)_{ij}=\sum_k F_{ik}F_{jk}=\sum_k F_{ik}\delta_{ij},$$

where δ_{ij} is the Kronecker symbol. In particular $(FF)_{il}^T=\sum_k F_{ik}$. Observe furthermore that for all i, j, k, $c_kF_{ik}F_{ij}=c_jF_{ik}F_{ij}$ because of the assumption on C. Now we calculate

$$(\bar{C}F)_j=(CF^T(FF^T)^{-1}F)_j=\sum_{i,k,l} c_kF_{ik}(FF^T)_{il}^{-1}F_{lj}=\sum_{i,k} c_kF_{ik}F_{ij}(FF^T)_{il}^{-1}$$

$$= c_j \sum_i F_{ij} (FF^T)^{-1}_{ii} \sum_k F_{ik} = c_j \sum_i F_{ij} = c_j.$$

So $\bar{C}F = C$. ∎

Remark From Proposition 2.3 it follows that a necessary condition for reduction of (X, N) (or (A, C)) is that some of the c_i are identical. However this condition is not sufficient, since also the transformed process $f(X)$ has to be Markov. See the example that follows after Proposition 2.3.

However if F reduces a pair (A, C), then, as follows from Lemma 2.6, at the same time it reduces any other pair $(A, \hat{C})$, where $\hat{C} = \tilde{C}F$ for some $\tilde{C}$. Observe that here all the $\tilde{c}_i$ may be different, which is not necessarily the case for the $\bar{c}_i$. This means that if F reduces (A, C), it also reduces any other pair $(A, \hat{C}) = (A, \tilde{C}F)$, if there exists a map g such that $g(\tilde{c}_j) = \bar{c}_i$. Or, equivalently, if there exists a map g such that $g(\hat{c}_i) = c_i$. Indeed this equivalence holds, because $g(\hat{c}_i) = g(\sum_j \tilde{c}_j F_{ji}) = \sum_j g(\tilde{c}_j) F_{ji} = \sum_j \bar{c}_j F_{ji} = c_i$.

To see whether a system (X, N) is reducible one may check whether the criteria of Proposition 2.3 hold for a reduction matrix F. If the state space $\mathcal{X}$ is very large this is of course quite a task. So we are looking for more easily verifiable criteria. It turns out, as can be expected, that a definition of stochastic observability offers an alternative approach to find a possible reduction. Before defining this concept, we have to introduce some notation and we also need some properties that are satisfied by the objects that play a role in the following definition.

DEFINITION 2.7 Let for each integer $k \geqq 1$, U^k be the set of bounded left continuous functions from $\mathbb{R}^+$ to $\mathbb{R}^k$. Write $U = U^1$ and if $u \in U^k$, then $u(t)$ will be written as a row vector. Define for $u \in U$, $v \in U^m$, a reduction matrix $F \in \mathbb{R}^{m \times n}$ and $T \geqq t \geqq 0$

$$g_F^{u,v}(t, T) = E\left[\exp\left(i \int_t^T u(s)\, dN_s + i \int_t^T v(s) F Y_s\, ds \right) \middle| \mathcal{F}_t \right]. \tag{9}$$

Because (X, N) is a stochastic system, we may replace the conditioning σ-algebra in (9) by $\sigma(X_t)$. Hence there exists a deterministic $h_F^{u,v}(t, T) \in \mathbb{C}^{1 \times n}$, such that $g_F^{u,v}(t, T) = h_F^{u,v}(t, T) Y_t$.

The following proposition gives a representation for $h_F^{u,v}(t, T)$ as defined above. We use the following notation throughout the rest of this section. Let x be a row or column vector in $\mathbb{R}^n$. Then $\operatorname{diag}(x)$ is the $n \times n$ diagonal matrix with ith diagonal element equal to x_i. Note that $g_F^{u,v}(t, T)$ can be interpreted as a conditional characteristic function of part of the future behaviour of the system, given its entire past.

PROPOSITION 2.8 *Let $h_F^{u,v}(t, T)$ be as in Definition 2.7. Then it satisfies the integral equation:*

$$h_F^{u,v}(t,T)=\mathbf{1}^T+\int_t^T h_F^{u,v}(s,T)(i\,\mathrm{diag}(v(s)F)+D(u(s)))\Phi(s-t)\,ds. \tag{10}$$

In the points where $h_F^{u,v}(\cdot,T)$ is differentiable, we have

$$\frac{\partial}{\partial t}h_F^{u,v}(t,T)=-h_F^{u,v}(t,T)(i\,\mathrm{diag}(v(t)F)+D(u(t))+A). \tag{11}$$

Proof We suppress in this proof the dependence on u, v and F.

Let $\tilde{g}(T)=\exp(i\int_0^T u(s)\,dN_s+i\int_0^T v(s)FY_s\,ds)$. Then from the stochastic calculus rule we obtain

$$\tilde{g}(T)=1+\int_0^T \tilde{g}(s-)(e^{iu(s)}-1)\,dN_s+\int_0^T \tilde{g}(s)iv(s)FY_s\,ds. \tag{12}$$

Now we take conditional expectations in (12) w.r.t. $\tilde{\mathcal{F}}_t=\mathcal{F}_t\vee\mathcal{F}_\infty^X$. Because CY is also the $\tilde{\mathbb{F}}$-intensity of N, we get

$$\begin{aligned}E[\tilde{g}(T)|\tilde{\mathcal{F}}_t]&=\tilde{g}(t)+E\left[\int_t^T \tilde{g}(s)(e^{iu(s)}-1)CY_s\,ds\Big|\tilde{\mathcal{F}}_t\right]+E\left[\int_t^T \tilde{g}(s)iv(s)FY_s\,ds\Big|\tilde{\mathcal{F}}_t\right]\\&=\tilde{g}(t)+\int_t^T E[\tilde{g}(s)|\mathcal{F}_t](e^{iu(s)}-1)CY_s\,ds+\int_t^T E[\tilde{g}(s)|\tilde{\mathcal{F}}_t]iv(s)FY_s\,ds.\end{aligned} \tag{13}$$

Define $\tilde{g}(t,T)=E[\tilde{g}(t)^{-1}\tilde{g}(T)|\tilde{\mathcal{F}}_t]=\tilde{g}(t)^{-1}E[\tilde{g}(T)|\tilde{\mathcal{F}}_t]$. Then (13) yields

$$\tilde{g}(t,T)=1+\int_t^T \tilde{g}(t,s)((e^{iu(s)}-1)CY_s+iv(s)FY_s)\,ds.$$

So, $\tilde{g}(t,T)=\exp(\int_t^T((e^{iu(s)}-1)C+iv(s)F)Y_s\,ds)$. From this expression we also obtain a "backward" integral equation:

$$\tilde{g}(t,T)=1+\int_t^T \tilde{g}(s,T)((e^{iu(s)}-1)C+iv(s)F)Y_s\,ds. \tag{14}$$

Define now $g(t,T)=E[\tilde{g}(t,T)|\mathcal{F}_t]$, and observe that this is indeed the quantity in Definition 2.7. So we can write $g(t,T)=h(t,T)Y_t$. Then from (14)

$$g(t,T)=1+E\left[\int_t^T E[\tilde{g}(s,T)|\mathcal{F}_s]((e^{iu(s)}-1)C+iv(s)F)Y_s\,ds\Big|\mathcal{F}_t\right]$$

$$= 1 + E\left[\int_t^T g(s, T)((e^{iu(s)} - 1)C + iv(s)F)Y_s\, ds \middle| \mathscr{F}_t\right]$$

$$= 1 + E\left[\int_t^T h(s, T)Y_s((e^{iu(s)} - 1)C + iv(s)F)Y_s\, ds \middle| \mathscr{F}_t\right]$$

$$= 1 + E\left[\int_t^T h(s, T)(D(u(s)) + i\,\mathrm{diag}(v(s)F))Y_s\, ds \middle| \mathscr{F}_t\right]$$

$$= 1 + \int_t^T h(s, T)(D(u(s)) + i\,\mathrm{diag}(v(s)F))\Phi(s-t)\, ds\, Y_t. \tag{15}$$

Or, since $\mathbf{1}^T Y_t = 1$, $g(t, T) = h(t, T)Y_t$ and (15) has to hold for all possible outcomes of Y_t, we get $h(t, T) = \mathbf{1}^T + \int_t^T h(s, T)(D(u(s)) + i\,\mathrm{diag}(v(s)F)\Phi(s-t)\, ds$. This proves (10).

Furthermore, if differentiation w.r.t. t is allowed, (10) yields

$$\frac{\partial h}{\partial t}(t, T) = -h(t, T)(D(u(t)) + i\,\mathrm{diag}(v(t)F))$$

$$-\int_t^T h(s, T)(D(u(s)) + i\,\mathrm{diag}(v(s)F))\Phi(s-t)A\, ds$$

$$= -h(t, T)(D(u(t)) + i\,\mathrm{diag}(v(t)F)) + (\mathbf{1}^T - h(t, T))A$$

$$= -h(t, T)(D(u(t)) + i\,\mathrm{diag}(v(t)F) + A).$$

because $\mathbf{1}^T A = 0$. Hence (11) holds. ∎

In several cases an explicit expression for $h_F^{u,v}(t, T)$ is available. We need the following notation. Let $M_1, \ldots, M_k$ be square matrices of the same order. Then we denote by $\prod_{i=1}^{k} * M_i$ the ordered product $M_k M_{k-1} \ldots M_1$.

COROLLARY 2.9 *Let $t = t_0 < t_1 < \cdots < t_k = T$. Let for $j = 1, \ldots, k$, $u_j \in \mathbb{R}$ and $v_j \in \mathbb{R}^m$ and define $u(s) = \sum_{j=1}^{k} u_j 1_{(t_{j-1}, t_j]}(s)$, $v(s) = \sum_{j=1}^{k} v_j 1_{(t_{j-1}, t_j]}(s)$. Then with this choice of the functions u and v we have*

$$h_F^{u,v}(t, T) = \mathbf{1}^T \prod_{j=1}^{k} * \exp((i\,\mathrm{diag}(v_j F) + D(u_j) + A)(t_j - t_{j-1})). \tag{16}$$

Proof Follows directly from Eq. (11).

The usefulness of the $h_F^{u,v}(t, T)$ is partly the content of the next lemma.

LEMMA 2.10 *Assume that F reduces (X, N). Let $\mathscr{K} = \operatorname{Ker} F$ and K be a matrix whose columns span $\mathscr{K}$. Then $h_F^{u,v}(t, T)K \equiv 0$, and hence there exists a factorization $h_F^{u,v}(t, T) = \bar{h}^{u,v}(t, T)F$.*

Proof Observe first that, always, $F\operatorname{diag}(v(s)F) = \operatorname{diag}(v(s))F$. From the fact that F reduces (A, C), we have $FAK = FDK = 0$. Hence there exist matrices $N_1, N_1(t)$ and N_2 such that $AK = KN_1$, $\Phi(t)K = KN_1(t)$, $DK = KN_2$. Therefore, with suppression of the dependence on u, v and F:

$$h(t, T)K = \mathbf{1}^T K + \int_t^T h(s, T)(i\operatorname{diag}(v(s)F) + D(u(s)))\Phi(s-t)K\,ds$$

$$= 0 + \int_t^T h(s, T)K(iN(s) + (e^{iu(s)} - 1)N_2)N_1(s-t)\,ds. \tag{17}$$

Since $h(t, T)K \equiv 0$ is a solution of (17), and since solutions are unique, the proof is finished. ∎

Apparently, for stepfunctions u and v as in Corollary 2.9, only the differences $t_j - t_{j-1}$ are important. Therefore we introduce functions $h_{F,k}^{u,v}(t)$ as follows. Let $\{t_j\}_{j=1}^{\infty} \subset \mathbb{R}^+$, $\{u_j\}_{j=1}^{\infty} \subset \mathbb{R}$, $\{v_j\}_{j=1}^{\infty} \subset \mathbb{R}^m$ and define $h_{F,k}^{u,v}(t)$ as in (16) with the differences $t_j - t_{j-1}$ replaced by t_j. Let H be the cone $\{(t, T) \in \mathbb{R}^2 : T \geqq t \geqq 0\}$. Clearly for all $u \in U$, $v \in U^m$ and $(t, T) \in H$, $h_F^{u,v}(t, T)$ induces a linear map from $\mathbb{R}^n$ into $\mathbb{C}$. So we can introduce $h_F \in \mathscr{L}(\mathbb{R}^n, \mathbb{C}^{U \times U^m \times H})$ by $h_F(u, v, (t, T)) = h_F^{u,v}(t, T) \in \mathscr{L}(\mathbb{R}^n, \mathbb{C})$. Denote by $\mathscr{K}_F$ the kernel of h_F.

In a similar way we can introduce operators $h_{F,k}$, by considering the functions $h_{F,k}^{u,v}(t)$, and their kernels $\mathscr{K}_{F,k}$. Now we can prove the following.

THEOREM 2.11 *Let $\mathscr{K}_F$ be $\operatorname{Ker} h_F$ and $\mathscr{K}_{F,k} = \operatorname{Ker} h_{F,k}$. Then*

i) $\mathscr{K}_{F,1} \supset \mathscr{K}_{F,2} \supset \cdots$ *and* $\bigcap_{j=1}^{\infty} \mathscr{K}_{F,j} = \mathscr{K}_F$;

ii) *If for some j $\mathscr{K}_{F,j} = \mathscr{K}_{F,j+1}$, then $\mathscr{K}_{F,j}$ is D, A and $\operatorname{diag}(vF)$ invariant, for all $v \in \mathbb{R}^m$ and $\mathscr{K}_F = \mathscr{K}_{F,j}$;*

iii) $\mathscr{K}_F \subset \operatorname{Ker} F$;

iv) *If moreover h_F factorizes as $h_F = \bar{h}F$, then $\mathscr{K}_F = \operatorname{Ker} F$.*

Proof Since we work with fixed F, we suppress the dependence on F. For notational convenience we also suppress dependence on u and v.

i) Let $\bar{t}_j = t_{j+1}$, $j \geqq 1$. Let $v_1 = v$, $u_1 = u$. Then

$$h^{j+1}(t) = h_j(\bar{t})\exp((i\operatorname{diag}(vF) + D(u) + A)t_1). \tag{18}$$

Now $h_{j+1}(t)\mathscr{K}_{j+1} \equiv 0$. So in particular for $t_1 = 0$, we get $h_j(\bar{t})\mathscr{K}_{j+1} \equiv 0$, which shows that $\mathscr{K}_{j+1} \subset \mathscr{K}_j$.

Certainly $\mathscr{K} \subset \mathscr{K}_j$, for all j, so $\mathscr{K} \subset \bigcap_{j=1}^{\infty} \mathscr{K}_j$. But since any $u \in U$ and $v \in U^m$ are pointwise limits of stepfunctions, also the reversed inequality holds.

ii) Assume $\mathscr{K}_j = \mathscr{K}_{j+1}$. Differentiation of (18) with respect to t_1 gives

$$0 \equiv \frac{\partial}{\partial t_1} h_{j+1}(t)\mathscr{K}_j = h_{j+1}(t)(i\,\mathrm{diag}(vF) + D(u) + A)\mathscr{K}_j. \tag{19}$$

Now take in (19) $t_1 = 0$, $u = 0$ and $v = 0$. Then

$$0 \equiv h_j(\bar{t})A\mathscr{K}_j$$

which yields $\mathscr{K}_j$ to be A-invariant. With this information we take in (19) $t_1 = 0$ and $u = 0$ but we allow v to be free. This yields $\mathscr{K}_j$ is also $\mathrm{diag}(vF)$ invariant for all v. Similarly $\mathscr{K}_j$ is also $D(u)$ invariant for all u, hence D invariant. Hence $\mathscr{K} = \bigcap_{l=1}^{\infty} \mathscr{K}_l = \mathscr{K}_j$.

iii) From (ii) we know that $\mathscr{K}$ is $\mathrm{diag}(vF)$ invariant (for all v). Hence $F\,\mathrm{diag}(vF)\mathscr{K} = 0$ or $\mathrm{diag}(v)F\mathscr{K} = 0 \Rightarrow F\mathscr{K} = 0$.

iv) Obvious in view of (iii).

PROPOSITION 2.12 *The following statements are equivalent.*

i) $h_F^{u,v}(t, T) = \bar{h}^{u,v}(t, T)F$;

ii) $\bar{h}^{u,v}$ *satisfies the integral equation*

$$\bar{h}^{u,v}(t, T) = \mathbf{1}^T + \int_t^T \bar{h}^{u,v}(s, T)(i\,\mathrm{diag}(v(s)) + \bar{D}(u(s)))\bar{\Phi}(s-t)\,ds \tag{20}$$

where $FD = \bar{D}F$, $F\Phi(t) = \bar{\Phi}(t)F$.

Proof (i)⇒(ii) From Theorem 2.11, we know that $\mathrm{Ker}\,F = \mathrm{Ker}\,h_F$ is a D, A and $\mathrm{diag}(v(s)F)$ invariant subspace of $\mathbb{R}^n$. So there exist matrices $\bar{D}$ and $\bar{A}$ such that $FD = \bar{D}F$, $FA = \bar{A}F$, and as always we have $F\,\mathrm{diag}(v(s)F) = \mathrm{diag}(v(s))F$. Hence

$$\bar{h}_F^{u,v}(t, T)F = \mathbf{1}^T + \int_t^T \bar{h}^{u,v}(s, T)F(i\,\mathrm{diag}(v(s)F) + D(u(s)))\Phi(s-t)\,ds$$

$$= \mathbf{1}^T + \int_t^T \bar{h}^{u,v}(s, T)(i\,\mathrm{diag}(v(s)) + \bar{D}(u(s)))\bar{\Phi}(s-t)\,dsF.$$

After postmultiplication with F^+, the claim follows.

(ii)⇒(i) Postmultiply (20) by F, then we see that $\bar{h}^{u,v}(t, T)F$ satisfies the same integral equation as $h_F^{u,v}(t, T)$. Because $\bar{h}^{u,v}(T, T)F = \mathbf{1}^T F = \mathbf{1}^T = h_F^{u,v}(t, T)$, the claim follows. ∎

The following proposition, that summarizes some of the preceding results forms

the basis of Definition 2.14 below and makes it understandable if one keeps the interpretation of $g_F^{u,v}(t, T)$ as a conditional characteristic function in mind.

PROPOSITION 2.13 *Let* $\bar{X}_t = f(X_t)$, $\bar{Y}_t = FY_t$, *where the reduction matrix* F *is associated with* f *as usual. There is equivalence between*

i) $E[g_F^{u,v}(t, T) | \sigma(\bar{X}_t)] = g_F^{u,v}(t, T)$ *for all* u, v *and* $t \leqq T$.

ii) $(\bar{X}, N)$ *is a stochastic system and* $g_F^{u,v}(t, T) = E[g_F^{u,v}(t, T) | \mathscr{F}_t^{\bar{X}} \vee \mathscr{F}_t^N]$ *for all* u, v *and* $t \leqq T$.

iii) *There exists a factorization* $h_F^{u,v}(t, T) \equiv \bar{h}^{u,v}(t, T)F$.

Proof (i)⇒(iii) There exists a matrix Q_t such that $E[Y_t | \sigma(\bar{X}_t)] = Q_t \bar{Y}_t$. (i) then implies $h_F^{u,v}(t, T)Y_t = h_F^{u,v}(t, T)Q_t\bar{Y}_t = h_F^{u,v}(t, T)Q_tFY_t$. So take $\bar{h}^{u,v}(t, T) = h_F^{u,v}(t, T)Q_t$.

(iii)⇒(i)

$$E[g_F^{u,v}(t, T) | \sigma(\bar{X}_t)] = E[\bar{h}^{u,v}(t, T)FY_t | \sigma(\bar{X}_t)] = E[\bar{h}^{u,v}(t, T)\bar{Y}_t | \sigma(\bar{X}_t)]$$

$$= \bar{h}^{u,v}(t, T)\bar{Y}_t = h_F^{u,v}(t, T)Y_t.$$

(iii)⇒(ii)

$$E[g_F^{u,v}(t, T) | \mathscr{X}_t^{\bar{X}} \vee \mathscr{F}_t^N] = E[E[g_F^{u,v}(t, T) | \mathscr{F}_t^X \vee \mathscr{F}_t^N] | \mathscr{X}_t^{\bar{X}} \vee \mathscr{F}_t^N]$$

$$= E[h_F^{u,v}(t, T)Y_t | \mathscr{F}_t^{\bar{X}} \vee \mathscr{F}_t^N] = E[\bar{h}_F^{u,v}(t, T)\bar{Y}_t | \mathscr{F}_t^{\bar{X}} \vee \mathscr{F}_t^N]$$

$$= g_F^{u,v}(t, T).$$

This, together with Proposition 2.12 also shows that $(\bar{X}, N)$ is a stochastic system.

(ii)⇒(iii)

$$h_F^{u,v}(t, T)Y_t = E[g_F^{u,v}(t, T) | \mathscr{F}_t^{\bar{X}} \vee \mathscr{F}_t^N] = E[g_F^{u,v}(t, T) | \sigma(\bar{X}_t)] = \bar{h}_F^{u,v}(t, T)\bar{Y}_t$$

for some deterministic $\bar{h}_F^{u,v}(t, T)$ since the last conditional expectation is a function of $\bar{X}_t$. Because $\bar{Y}_t = FY_t$ the result now follows. ∎

DEFINITION 2.14 The nth order system (X, N) is said to be *strongly stochastically observation equivalent* with some mth order system $(m \leqq n)$ if there exists a reduction matrix $F \in \mathbb{R}^{m \times n}$ such that a factorization $h_F = \bar{h}F$ holds. If any such factorization implies that F is a permutation matrix, then (X, N) will be called *strongly stochastically observable.*

Some comments are appropriate. Let (X, N) be described by Eq. (8). If (X, N) is strongly stochastically observation equivalent with some mth order system, then from Propositions 2.12 and 2.13 it follows that this one is described via matrices $\bar{A}$ and $\bar{C}$ by an equation like (8). Therefore we will also say that (A, C) is strongly observation equivalent with $(\bar{A}, \bar{C})$.

The interpretation is as follows. If we condition the distribution of the future of the bivariate stochastic process $(f(X), N)$ on the entire past of (X, N), or

equivalently just on the current state X_t, then this determines $f(X_t)$ only, instead of X_t itself. We also know from Proposition 2.13 that $(f(X), N)$ is again a stochastic system.

Suppose now that (X, N) is strongly stochastically observable and that F is a $m \times n$ reduction matrix $(m < n)$. Then a factorization $h_F = \tilde{h}G$ always exists for another reduction matrix G, which may be the identity (or a permutation matrix). However, because then $\operatorname{Ker} G \subset \operatorname{Ker} h_F \subset \operatorname{Ker} F$ [see Theorem 2.11(iii)], where the last inclusion is strict, it follows that there exists yet another reduction matrix H such that $HG = F$. Hence the conditioning of the distribution of $(f(X), N)$ on X_t determines strictly more than $f(X_t)$. Stated otherwise, $f(X_t)$ is not sufficient to predict the future distribution of $(f(X), N)$. Note also that in this case (X, N) can only be strongly stochastically observation equivalent with another nth order system.

We also mention thaat this definition differs from the current definition of stochastic observability in the literature [van Schuppen (1989, p. 490)] for linear Gaussian systems, where the future evolution of the state processes is disregarded. However the Gaussian analogue of our definition is equivalent with what can be found in the literature. The reason behind our alternative is that we now force the transformed process $f(X)$ to be Markov, which is automatically the case in the linear Gaussian situation. Therefore a slightly different terminology appears to be advisable. The idea behind strong stochastical observation equivalence, is that it should provide us with a link to (strong forward) reducibility. Moreover it should give us information about what reductions of the original systems are possible. This is the content of the next result which, although obvious, brings the concepts minimality and observability together.

THEOREM 2.15 *Let (X, N) be given by Eq. (8). Let F be a reduction matrix and define $\bar{A} = FAF^+$ and $\bar{C} = CF^+$.*

i) *F reduces (A, C) if and only if (A, C) is strongly stochastically observation equivalent with $(\bar{A}, \bar{C})$.*

ii) *(A, C) is (strongly forwardly) minimal if and only if (A, C) is strongly stochastically observable.*

Proof Direct consequence of Definition 2.14, Theorem 2.11 and Proposition 2.12. ∎

At first glance this theorem seems to be not very helpful, if one is looking for possible reduction of (A, C), since Definition 2.14 involves the unknown F that describes the reduction. But it turns out that it is a useful step to the finding of the F (if any) that reduces (A, C). We first introduce some new notation and an auxiliary result. Take in the definition of $h_F^{u,v}(t, T)$ the function v to be identically zero and write instead $h^u(t, T)$. Observe that this quantity does not depend on the specific F any more. By taking u to be a stepfunction we can again, parallel to what we did after Lemma 2.10, construct $h_k^u(t) \in \mathbb{C}^n$ and from these the operators h and h_k. Some of the properties of the $h_F^{u,v}(t, T)$ and $h_{F,k}^{u,v}(t)$ carry over to $h^u(t, T)$ and the $h_k^u(t)$. There are however some differences. The precise result is the following.

PROPOSITION 2.16 *Let* $\mathscr{K} = \operatorname{Ker} h$, $\mathscr{K}_j = \operatorname{Ker} h_j$. *Then*

i) $\mathscr{K}_1 \supset \mathscr{K}_2 \supset \cdots$, *and* $\bigcap_{j=1}^{\infty} \mathscr{K}_j = \mathscr{K}$.

ii) *If for some* j $\mathscr{K}_j = \mathscr{K}_{j+1}$, *then* $\mathscr{K} = \mathscr{K}_j$ *and* $\mathscr{K} = \mathscr{K}_j$ *is* D *and* A *invariant.*

iii) *If* F *reduces* (A, C), *then* $\operatorname{Ker} F \subset \mathscr{K}$, *hence there exists a factorization* $h = \bar{h}F$.

iv) *If a factorization* $h = \bar{h}F$ *exists such that* $\operatorname{Ker} h = \operatorname{Ker} F$ *for a reduction matrix* F, *then* F *reduces* (A, C).

v) *There is equivalence between*

a) $h^u(t, T) \equiv \bar{h}^u(t, T)G$ *and* $\operatorname{Ker} h = \operatorname{Ker} G$ *for some matrix* G; *and*

b) $\bar{h}^u(t, T)$ *satisfies the integral equation*

$$\bar{h}^u(t, T) = \mathbf{1}^T + \int_t^T \bar{h}^u(s, T)\bar{D}(u(s))\bar{\Phi}(s-t)\,ds$$

where $\bar{D}G = GD$, $\bar{D}(u(s)) = (e^{iu(s)} - 1)\bar{D}$ *and* $\bar{\Phi}(t)G = G\Phi(t)$.

Proof (i) and (ii) are proved in the same way as (i), (ii) of Theorem 2.11, (iii) follows in the same way as Lemma 2.10, (iv) follows from (ii) and (v) can be proved as Proposition 2.12. ∎

Remark The most striking difference between h and h_F is the following. Suppose that $\mathscr{K} = \operatorname{Ker} h \neq \{0\}$. Then we have a factorization $h = \bar{h}G$ for some G which is such that $\operatorname{Ker} h = \operatorname{Ker} G$. It may happen that it is impossible to choose G to be a reduction matrix. See Examples 2.23 and 2.24. Notice also that we imposed in (v) of Proposition 2.16 that $\operatorname{Ker} h = \operatorname{Ker} G$, whereas for the analogous statement of Proposition 2.12 the equality $\operatorname{Ker} h_F = \operatorname{Ker} F$ automatically holds. The next proposition implicitly offers a way to compute the $\mathscr{K}_j$ and $\mathscr{K}$.

PROPOSITION 2.17 *There exist a sequence of matrices* W_j, *as indicated in the proof, such that* $\operatorname{Ker} h_j = \operatorname{Ker} W_j$ *for all* $j \geqq 1$.

Proof Let $z = e^{iu} - 1$ and let (with a little abuse of notation) $h_1^z(t) = h_1^u(t)$. Let $W_1(z)$ be the $n \times n$ matrix with jth row equal to $(\partial/\partial t)^j h_1^z(0) = \mathbf{1}^T(zD + A)^{j-1}$ (use Eq. (11) with $v = 0$). By the Caley–Hamilton theorem for $k \geqq n$ one has $(\partial/\partial t)^k h_1^z(0) = \sum_{j=0}^{n-1} \alpha_{kj} \mathbf{1}^T(zD + A)^j$ for some real numbers α_{kj}. Hence $\operatorname{Ker} h_1 = \mathscr{K}$ if and only if $W_1(z)\mathscr{K} = \{0\}$ for all z. Next we form the matrix W_1 in the following way. Each row $\mathbf{1}^T(zD + A)^{j-1}$ of $W_1(z)$ can be written as $\sum_{k=0}^{j-1} z^k \beta_{kj}$, where the β_{kj} are row vectors in $\mathbb{R}^n$. W_1 is now the matrix obtained by stacking all the β_{kj} in a large matrix with n columns. It is evident that $\operatorname{Ker} h_1 = \operatorname{Ker} W_1$. In an analogous way we can also construct matrices $W_j(z)$ via the partial derivatives of $h_j^z(t)$ with respect to the vector t evaluated at $t = 0$. And as above by grouping the equal powers of z that appear in the rows of $W_j(z)$, we obtain a matrix W_j. Hence the $\mathscr{K}_j$ appearing in Proposition 2.16 are the same as the kernels of the matrices W_j. ∎

Some additional properties of h are described by the following lemma.

LEMMA 2.18

i) *For all* $u \in U$ *the function* $h^u(\cdot, T)$ *is left differentiable at* $t = T$ *and* $(\partial/\partial t)h^u(T, T) = -(e^{iu(T)} - 1)C$.

ii) *Let V be the $n \times n$ Vandermonde matrix with jth row equal to $\mathbf{1}^T D^{j-1}$. Then* $\operatorname{Ker} h \subset \operatorname{Ker} V$.

iii) *Assume that there exists a reduction matrix F such that $h^u(t, T) \equiv \bar{h}^u(t, T)F$. Then C can be written as $\bar{C}F$. If moreover all the elements of $\bar{C}$ are different from each other, then* $\mathscr{K} = \operatorname{Ker} h = \operatorname{Ker} F$.

iv) *If all the c_i are different then* $\operatorname{Ker} h = \{0\}$.

Proof

i) Immediately follows from Eq. (11) since $h(T, T) = \mathbf{1}^T$ and $\mathbf{1}^T A = 0$.

ii) From Proposition 2.16 we know that $\operatorname{Ker} h$ is D invariant and since $\mathbf{1}^T \operatorname{Ker} h = h(T, T) \operatorname{Ker} h = \{0\}$ we have $\mathbf{1}^T D^{j-1} \operatorname{Ker} h = \mathbf{1}^T \operatorname{Ker} h = \{0\}$.

iii) We have to prove that $\bar{\mathscr{K}} = \operatorname{Ker} \bar{h} = \{0\}$. Because of (i) and Lemma 2.5 there exists a diagonal matrix $\bar{D}$ such that $FD = \bar{D}F$. Now $\bar{\mathscr{K}} = F\mathscr{K}$. Hence $\bar{D}\bar{\mathscr{K}} = \bar{D}F\mathscr{K} = FD\mathscr{K} \subset F\mathscr{K} = \bar{\mathscr{K}}$. So $\bar{\mathscr{K}}$ is $\bar{D}$ invariant. If $\bar{V}$ is the Vandermonde matrix with jth row equal $\mathbf{1}^T \bar{D}^{j-1}$ then we have as in (ii) $\bar{\mathscr{K}} \subset \operatorname{Ker} \bar{V}$. The latter is zero, since all the elements of $\bar{C}$ are assumed to be different.

iv) Follows from (iii).

The role that the $h^u(t, T)$ play in the finding of a matrix F that reduces (X, N) is revealed by the following theorem.

THEOREM 2.19 *There is equivalence between*

i) *(A, C) is strongly stochastically observation equivalent with $(\bar{A}, \bar{C})$.*

ii) *There exists a reduction matrix F such that $h^u(t, T) = \bar{h}^u(t, T)F$ for all $t \leqq T$ and all $u \in U$ and a similar factorization holds for any other pair $(A, \hat{C})$ where $\hat{C} = \tilde{C}F$. So if $\hat{h}$ is related to $(A, \hat{C})$ as h is to (A, C), then: $\hat{h}^u(t, T) = \tilde{h}^u(t, T)F$ for all $t \leqq T$ and all $u \in U$.*

Proof (i)⇒(ii) From Theorem 2.15 we know that there exists a reduction matrix F such that $FA = \bar{A}F$ and $C = \bar{C}F$. But then in view of the remark after Lemma 2.6 F also reduces any $(A, \hat{C})$ where $\hat{C}$ can be written as $\tilde{C}F$. Hence from Proposition 2.16 we have both the factorization $h^u(t, T) = \bar{h}^u(t, T)F$ and $\hat{h}^u(t, T) = \tilde{h}^u(t, T)F$.

(ii)⇒(i) Since the assumption holds for any $\hat{C} = \tilde{C}F$, we may take all the elements of $\tilde{C}$ to be different. Then from Lemma 2.18(iii) $\operatorname{Ker} \hat{h} = \operatorname{Ker} F$ and from Proposition 2.16 $\operatorname{Ker} F$ is A invariant, so $FA = \bar{A}F$, with $\bar{A} = FAF^+$. By assumption and from Lemma 2.5 $FD = \bar{D}F$. Hence F is a matrix that reduces (A, C). The result now follows from Theorem 2.15(i).

The following result is closely related to Theorem 2.19.

PROPOSITION 2.20 *Let (X, N) satisfy Eq.* (8) *and let $(X, \hat{N})$ be another stochastic system that satisfies an equation like* (8) *with $C = [c_1, \ldots, c_n]$ replaced with $\hat{C} = [\hat{c}_1, \ldots, \hat{c}_n]$ and let $\mathscr{K} = \operatorname{Ker} h$ and $\hat{\mathscr{K}} = \operatorname{Ker} \hat{h}$. Assume that there exists a map g such that $c_i = g(\hat{c}_i)$. Then $\hat{\mathscr{K}} \subset \mathscr{K}$.*

Proof $\hat{\mathscr{K}}$ is a $\hat{D}$ and A invariant subspace of $\mathbb{R}^n$, where $\hat{D}=\operatorname{diag}(\hat{C})$ (Proposition 2.16). We claim that $\hat{\mathscr{K}}$ is also D invariant. If the claim holds, then it immediately follows from Eq. (10) with $v\equiv 0$, that $h^u(t,T)\hat{\mathscr{K}}=\{0\}$, since $\mathbf{1}^T\hat{\mathscr{K}}=\{0\}$. Since $\hat{\mathscr{K}}$ is $\hat{D}$ invariant, it is spanned by some eigenvectors of $\hat{D}$. So let $k\in\hat{\mathscr{K}}$ be such that $\hat{D}k=\hat{c}_i k$, for one of the eigenvalues of $\hat{c}_i$ of $\hat{D}$. Hence for all j we have $\hat{c}_j k_j=\hat{c}_i k_j$ if $k=[k_1,\ldots,k_n]^T$. If $k_j=0$ then certainly $c_j k_j=c_i k_j$. If $k_j\neq 0$ then $\hat{c}_j=\hat{c}_i$, but then also $c_j=c_i$. So again we have $c_j k_j=c_i k_j$. Hence k is an eigenvector of D with eigenvalue $c_i=g(\hat{c}_i)$, which shows that $\hat{\mathscr{K}}$ is also D invariant.

All results so far obtained form the basis of Algorithm 2.21 below, that yields for a stochastic system (X,N) a minimal representation.

ALGORITHM 2.21

1) Compute $\operatorname{Ker} h$ and find a reduction matrix F such that $h^u(t,T)\equiv\bar{h}^u(t,T)F$ and such that $\bar{h}^u(t,T)$ cannot be factorized further by means of some other reduction matrix.
2) If $\operatorname{Ker} h=\operatorname{Ker} F$, then the algorithm produces F as its outcome. Else we go to step 3.
3) Let $\hat{C}=\tilde{C}F$, where all the $\tilde{c}_i$ are different from each other. Form $\hat{h}^u(t,T)$ (which is related to $(A,\hat{C})$ as was $h^u(t,T)$ to (A,C)).
4) Apply step 1 to $\hat{h}^u(t,T)$ in lieu of $h^u(t,T)$.

Before proving that the matrix F produced by the algorithm, induces a minimal pair $(\bar{A},\bar{C})$ defined by $\bar{A}=FAF^+$, $\bar{C}=CF^+$, we discuss the way it works. The finding of F in step 1 is relatively simple. Compute $\operatorname{Ker} h$ by using the matrices W_k of Proposition 2.9 as far as needed. This results in a factorization $h^u(t,T)=\tilde{h}^u(t,T)G$, where G is such that $\operatorname{Ker} G=\operatorname{Ker} h$. Next one inspects the columns of G. If any two of them are identical, then the same holds for the corresponding columns of F, which determines F up to a permutation of its columns. An alternative way is to inspect the elements of the $h^u_k(t)$ for all k as far as needed. If two columns of F are identical then the same holds for the corresponding elements of all the $h^u_k(t)$ and vice versa. If the algorithm stops at step 2, then it follows from Proposition 2.16 that F reduces (A,C). If instead $\operatorname{Ker} h\neq\operatorname{Ker} F$ and step 3 is performed then we known from Lemma 2.18 that some of the elements of $\bar{C}$ (which is such that $C=\bar{C}F$) are identical. Hence it makes sense to construct $\hat{C}$ as prescribed. Then from Proposition 2.20 we obtain that $\operatorname{Ker}\hat{h}\subset\operatorname{Ker} h$ and moreover that this inclusion is strict, since also $\operatorname{Ker}\hat{h}\subset\operatorname{Ker} F$ in view of Lemma 2.18(ii), applied to the Vandermonde matrix with rows $\mathbf{1}^T\hat{D}^{j-1}$, which has kernel equal to $\operatorname{Ker} F$. Hence the algorithm constructs a strictly decreasing sequence of kernels, until it terminates which happens after finitely many iterations.

THEOREM 2.22 *Let F be the final result of Algorithm 2.21. Then F reduces (A,C). Hence there exist $\bar{A},\bar{C}$ with $FA=\bar{A}F$, $C=\bar{C}F$. Moreover $(\bar{A},\bar{C})$ is minimal.*

Proof The resulting F has the property that in the final iteration a factorization of the form $\hat{h}^u(t,T)=\tilde{h}^u(t,T)F$ holds, where $\hat{h}^u(t,T)$ corresponds to some pair $(A,\hat{C})$ and where $\operatorname{Ker}\hat{h}=\operatorname{Ker} F$. So $\operatorname{Ker} F$ is $\hat{D}$ and A invariant (Proposition 2.16)

and also D invariant (see the proof of Proposition 2.20). Hence F reduces (A, C). Now let F_1 be a matrix that gives a maximal reduction of (A, C). So with $\bar{A}=F_1AF_1^+$ and $\bar{C}=CF_1^+$ we have that $(\bar{A}, \bar{C})$ is a minimal pair. F_1 is determined up to a permutation of its columns. Then in step 1 of the algorithm we have a factorization (as follows from Proposition 2.16) $h^u(t, T)=\bar{h}^u(t, T)F_2F_1$, where possibly another reduction matrix is involved. Suppose that step 2 is skipped, otherwise the proof is complete. So we construct $\hat{C}=\tilde{C}F_2F_1$. Then of course (Lemma 2.5) $\operatorname{Ker} F_1$ is $\hat{D}$ invariant and therefore $\hat{h}^u(t, T)$ factorizes as $\tilde{h}^u(t, T)F_3F_1$, with possibly again another reduction matrix F_3, which has the property that $\operatorname{Ker} F_3 \subset \operatorname{Ker} F_2$, because $\operatorname{Ker}(F_3F_1) \subset \operatorname{Ker} \hat{h} \subset \operatorname{Ker} \hat{V}=\operatorname{Ker}(F_2F_1)$, where $\hat{V}$ is the Vandermonde matrix with jth row equal to $\mathbf{1}^T\hat{D}^{j-1}$. (Use also Lemma 2.18.) Hence in each iteration of the algorithm a factorization of functions like $h^u(t, T)$ holds, where the matrix F_1 is always part of the factorization, and where the kernels of the F_2, F_3 etc. are shrinking. Therefore in the final step of the algorithm we have a factorization of the form $\hat{h}=\tilde{h}F_*F_1$. From the first part of the proof we know that F_*F_1 is a matrix that reduces (A, C), but since F_1 gives the minimal reduction F_* has to be a permutation matrix. ∎

In the next two examples, we apply Algorithm 2.21.

Example 2.23 Let X take its values in $\{1, 2, 3, 4, 5\}$ and let

$$A=\begin{bmatrix} -14 & 1 & 1 & 1 & 1 \\ 1 & -14 & 2 & 3 & 1 \\ 9 & 9 & -7 & 7 & 1 \\ 1 & 1 & 2 & -12 & 1 \\ 3 & 3 & 2 & 1 & -4 \end{bmatrix}.$$

Assume that N has the intensity CY, where $C=[1\ 1\ 1\ 1\ 2]$.

The matrix W_1 of Proposition 2.17 now becomes (use $\mathbf{1}^TA=0$):

$$W_1=\begin{bmatrix} \mathbf{1}^T \\ \mathbf{1}^TD \\ \mathbf{1}^TD^2 \\ \mathbf{1}^TDA \\ \mathbf{1}^TD^3 \\ \mathbf{1}^T(D^2A+DAD) \\ \mathbf{1}^TDA^2 \\ \mathbf{1}^TD^4 \\ \mathbf{1}^T(D^3A+D^2AD+DAD^2) \\ \mathbf{1}^T(D^2A^2+DADA+DA^2D) \\ \mathbf{1}^TDA^3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 2 \\ 1 & 1 & 1 & 1 & 4 \\ 3 & 3 & 2 & 1 & -4 \\ 1 & 1 & 1 & 1 & 8 \\ 12 & 12 & 8 & 4 & -20 \\ -32 & -32 & -11 & 10 & 25 \\ 1 & 1 & 1 & 1 & 16 \\ 33 & 33 & 22 & 11 & -68 \\ -172 & -172 & -63 & 46 & 166 \\ 402 & 402 & 51 & -300 & 165 \end{bmatrix}.$$

Now $\operatorname{Ker} h_1=\operatorname{Ker} W_1$ is spanned by $[1\ -1\ 0\ 0\ 0]^T$ and $[0\ 1\ -2\ 1\ 0]^T$. Observe that these two vectors are eigenvectors of both D and A. So $\operatorname{Ker} h_1=\operatorname{Ker} h$. The reduction matrix F in step 1 of Algorithm 2.21 is easily seen to be

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

since the first two columns of W_1 are identical. Clearly $\operatorname{Ker} F \neq \operatorname{Ker} h$. So step 3 of the algorithm applies. Let $\tilde{C}=[1\ 3\ 4\ 2]$, $\hat{C}=[1\ 1\ 3\ 4\ 2]$. Of course one can now construct a matrix $\hat{W}_1$. Then $\operatorname{Ker} \hat{W}_1 \subset \operatorname{Ker} F$ (Lemma 2.18(ii)), which is spanned by $[1\ -1\ 0\ 0\ 0]^T$. Since, as observed above $[1\ -1\ 0\ 0\ 0]^T$ is A and D invariant, we see that $\operatorname{Ker} \hat{W}_1 = \operatorname{Ker} F$ and also, as above, $\operatorname{Ker} \hat{h} = \operatorname{Ker} \hat{W}_1$.

Hence the outcome of the algorithm is

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The next (partially worked) example is, apart from an illustration of Algorithm 2.21, also interesting in the light of the remark that followed Proposition 2.16.

Example 2.24 Change the matrix A in the preceding example into

$$A = \begin{bmatrix} -4 & 1 & 1 & 2 & 1 \\ 1 & -5 & 4 & 2 & 2 \\ 1 & 3 & -8 & 1 & 3 \\ 1 & 0 & 1 & -8 & 4 \\ 1 & 1 & 2 & 3 & -10 \end{bmatrix}$$

but let C be the same. If one again computes the matrix W_1 then it turns out that its kernel $\mathscr{K}_1$ is again spanned by $k_1^T=[1\ -1\ 0\ 0\ 0]^T$ and $k_2^T=[0\ 1\ -2\ 1\ 0]^T$. Let $K=[k_1 k_2]$. A calculation shows that

$$AK = K\begin{bmatrix} -5 & 1 \\ 1 & -10 \end{bmatrix} \quad \text{and} \quad DK = K.$$

Hence $\operatorname{Ker} h$ is spanned by k_1 and k_2, since already $\mathscr{K}_1$ is a D and A invariant subspace. The matrix F in step 1 of Algorithm 2.21 is the same as in the preceding example. Take again $\hat{C}=[1\ 1\ 3\ 4\ 2]$. The matrix $\hat{W}_1$ contains one row equal to $\mathbf{1}^T\hat{D}A=[6\ \ 7\ \ -11\ \ -19\ \ 8]$. Since $\operatorname{Ker} \hat{h}$ belongs to both $\operatorname{Ker} \hat{W}_1$ and $\operatorname{Ker} F$ as explained in the discussion after the description of the algorithm, we see that $\operatorname{Ker} \hat{h}=\{0\}$. The F resulting from the algorithm is therefore the identity matrix (or another permutation matrix). The intriguing feature of this example is, that in spite of the fact that most of the c_i are equal, no reduction is possible.

We close this section with some considerations that indicate ways of future research. Let first $\bar{X}_t = f(X_t)$ and let F be the reduction matrix associated with f.

Assume $f:\{1,\ldots,n\}\to\{1,\ldots,m\}$. Trivially each entry F_{ij} of F has the following interpretation: $F_{ij}=P(\bar{X}_t=i\,|\,X_t=j)$. In both the two examples above we can factorize h as $\bar{h}G$, where

$$G=\begin{bmatrix}1 & 1 & \frac{1}{2} & 0 & 0\\ 0 & 0 & \frac{1}{2} & 1 & 0\\ 0 & 0 & 0 & 0 & 1\end{bmatrix}.$$

Observe that each column of G can be considered as a probability vector. The idea is now to extend the interpretation of the F_{ij} as a conditional probability to the entries of G. This idea allows us to consider so-called probabilistic reductions of the system (X,N) by looking at suitably defined random functions of X_t. This new approach seems to be connected with the behaviour of the solutions of the filtering problem that is defined by the finding of $E[Y_t|\mathcal{F}_t^N]$. Results in this direction will be reported in another publication. We only mention that in the last example the following identity holds: $GA=\bar{A}G$, where

$$\bar{A}=\begin{bmatrix}-2\frac{1}{2} & 4\frac{1}{2} & 4\frac{1}{2}\\ 1\frac{1}{2} & -7\frac{1}{2} & 5\frac{1}{2}\\ 1 & 3 & -10\end{bmatrix},$$

which is indeed the rate matrix of some Markov process, that lives on a state space with three elements. This already indicates that some reduction, of another type than described in this section, should be possible.

References

[1] P. Brémaud, *Point Processes and Queues*, Springer, 1981.
[2] J. G. Kemeny and J. L. Snell, *Finite Markov Chains*, Van Nostrand, 1960.
[3] Ph. Protter, Markov solutions of stochastic differential equations, *ZW* **41** (1977), 39–58.
[4] P. J. C. Spreij, Selfexciting counting process systems with finite state space. To appear in *Stochastic Processes and their Applications* (1990).
[5] J. H. van Schuppen, Stochastic realization problems. In: *Three Decades of Mathematical System Theory* (H. Nijmeijer and J. M. Schumacher, eds.), *Lecture Notes in Control and Information Sciences* **135** (1989), 480–523.
[6] J. H. van Schuppen and E. Wong, Transformation of local martingales under a change of law, *Ann. Prob.* **2** (1974), 879–888.