ELSEVIER

Linear Algebra and its Applications 350 (2002) 67–87

LINEAR ALGEBRA AND ITS APPLICATIONS

www.elsevier.com/locate/laa

State space calculations for discrete probability densities ☆

Bernard Hanzon [a], Raimund J. Ober [b],*

[a] *Department of Econometrics and Operations Research, Vrije Universiteit, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands*

[b] *Center for Engineering Mathematics EC33, University of Texas at Dallas, Richardson, TX 75083, USA*

Received 1 August 2001; accepted 18 December 2001

Submitted by P. Fuhrmann

Abstract

State space methods have proved to be powerful theoretical and computational tools in a number of areas of applications, in particular filtering and control theory. In this paper we advocate the use of state space methods for the study of discrete probability densities on the set {0, 1, 2, ...}. The fundamental approach is to consider the class $\mathscr{D}$ of all discrete probability densities that can be represented as the impulse response/convolution kernel of a finite dimensional discrete time state space system. We show that all standard operations such as the calculation of moments, convolution, scaling, translation, product, etc. can be carried out using system representations. © 2002 Elsevier Science Inc. All rights reserved.

Keywords: Discrete probability; Systems theory; State space methods; Realization theory; Linear algebra

1. Introduction

State space methods have proved to be theoretically and computationally powerful tools in various areas of applications, including filtering, signal processing and in particular control theory. The strength of the approach in the realm of theory lies in the fact that often very general situations can be treated with linear algebra tools. From a computational point of view the large array of methods from numerical linear algebra can be employed.

☆ The work by the second author was also supported in part by NSF grant DMS 9803186.

* Corresponding author.

E-mail addresses: ober@utdallas.edu (R.J. Ober).

PII: S0024-3795(02)00268-9

In this paper we show that state space methods can be used to deal with discrete probability densities. In particular, we show that essentially all standard probability theoretic calculations can be performed in a state space setting. This is done with the expectation that state space methods might become a tool equally important in probability theory as they are in other applications. In another publication [4] the authors have studied the question how state space methods can be used to investigate a continuous time filtering problem that involves rational probability densities.

In this paper we are going to consider state space methods for the calculation with discrete probability densities. In particular, we are going to analyze the class $\mathscr{D}$ of discrete time probability densities that admit a finite dimensional state space description. More precisely, we consider discrete probability densities, i.e. sequences p_k, $k = 0, 1, 2, \ldots,$ with $p_k \geqslant 0$, $k = 0, 1, 2, \ldots,$ and $\sum_{k=0}^{\infty} p_k = 1$, that admit a finite dimensional state space realization (A, b, c, d) such that

$$p_0 = d, \quad p_k = cA^{k-1}b, \ \ k = 1, 2, \ldots, \tag{1}$$

where the *system* (A, b, c, d) is such that A is an $n \times n$ matrix, b is an $n \times 1$ vector, c is an $1 \times n$ vector and d is a scalar. The system (A, b, c, d) is called a *minimal* realization of the probability density $(p_k)_{k\geqslant 0}$ if n is the smallest dimension for which a system exists such that (1) is satisfied. Therefore the set $\mathscr{D}$ consists of all sequences $(p_k)_{k\geqslant 0}$ such that

(1) $p_k \geqslant 0, k = 0, 1, 2, \ldots,$
(2) there exists a system (A, b, c, d) such that $p_0 = d$, $p_k = cA^{k-1}b, k = 1, 2, \ldots,$
(3) $\sum_{k=0}^{\infty} p_k = 1.$

For later technical reasons we also define the set $\mathscr{U}$ that consists of all 'un-normalized' probability densities and the zero sequence, i.e. all sequences that satisfy conditions (1) and (2) while (3) is replaced by $\sum_{k=0}^{\infty} p_k < \infty$.

In the following lemma we are going to quote a fundamental stability result that asserts that for a minimal realization of a probability density in $\mathscr{D}$ all eigenvalues of A are in the open unit disk.

Lemma 1.1. *Let (A, b, c, d) be a minimal realization of a discrete-time probability density in $\mathscr{D}$. Then A is discrete-time asymptotically stable, i.e.*

$$|\lambda(A)| < 1$$

for all eigenvalues $\lambda(A)$ of A.

Proof. Let $(p_k)_{k\geqslant 0} \in \mathscr{D}$ and let (A, b, c, d) be a minimal realization of $(p_k)_{k\geqslant 0}$. Since $\sum_{k=0}^{\infty} p_k = d + \sum_{k=1}^{\infty} |cA^{k-1}b| < \infty$, the result is a standard system theoretic stability test (see, e.g. [1, p. 208]). □

An important tool in system theory is the transfer function which we will show is very closely related to the generating function of the probability density. Let $(p_k)_{k\geqslant 0}$

be a probability density in $\mathscr{D}$ with state space realization (A, b, c, d). The *transfer function* of the (discrete-time) system (A, b, c, d) is given by

$$\begin{aligned}
G(z) &= \sum_{k=0}^{\infty} z^{-k} p_k \\
&= d + \sum_{k=1}^{\infty} z^{-k} c A^{k-1} b \\
&= d + z^{-1} c \left(\sum_{k=1}^{\infty} z^{-(k-1)} A^{k-1} \right) b \\
&= d + z^{-1} c \left(\sum_{k=0}^{\infty} (z^{-1} A)^{-k} \right) b \\
&= d + z^{-1} c \left(I - (z^{-1} A) \right)^{-1} b \\
&= c \, (zI - A)^{-1} \, b + d.
\end{aligned}$$

Note that this is a rational function in the independent variable z. Standard results in system theory show that two (not necessarily minimal) realizations of a probability density function in $\mathscr{D}$ have the same transfer function (see, e.g. [5]). As a consequence Lemma 1.1 implies that the poles of the transfer function of a probability density in $\mathscr{D}$ are in the open unit disk.

As mentioned earlier, the transfer function of the realization of a discrete-time probability density $(p_k)_{k\geqslant 0}$ in $\mathscr{D}$ is closely related to the generating function of the density. If X is a random variable with probability density $(p_k)_{k\geqslant 0}$ then the generating function is given by

$$E(s^X) = \sum_{k=0}^{\infty} s^k p_k = \sum_{k=1}^{\infty} s^k c A^{k-1} b + d = G(s^{-1}) = c \left(s^{-1} I - A \right)^{-1} b + d.$$

Hence the generating function can be obtained from the transfer function by a simple inversion of the independent variable. In this context it is important to mention that Kronecker's theorem [5] implies that each probability density with a rational transfer function of McMillan degree n has a minimal state space realization of dimension n.

We now discuss a number of examples.

Example 1 (*Densities with finite support*). Let $(p_k)_{k\geqslant 0}$ be a discrete probability density with finite support, i.e. there exists n such that $p_n \neq 0,\ p_k = 0,\ k > n$. A realization for such a density is given by

$$A = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

$$c = \begin{pmatrix} p_1 & p_2 & \cdots & p_{n-1} & p_n \end{pmatrix}, \quad d = p_0,$$

which implies that the density is in $\mathscr{D}$. The transfer function is given by

$$\begin{aligned} G(z) &= p_0 + z^{-1} p_1 + z^{-2} p_2 + \cdots + z^{-n} p_n \\ &= \frac{z^n p_0 + z^{n-1} p_1 + \cdots + z^1 p_{n-1} + p_n}{z^n}. \end{aligned}$$

Note that if $p_n \neq 0$, the numerator and denominator of this rational function are coprime. In this case we also therefore have that the realization (A, b, c, d) is minimal.

Example 2 (*Geometric distribution*). For $\lambda > 1$,

$$p_k = \frac{\lambda - 1}{\lambda^{k+1}}, \quad k = 0, 1, 2, \ldots,$$

is the *geometric distribution* with *parameter* λ. A minimal realization for this density is given by the one-dimensional system (A, b, c, d) with

$$A = \left(\frac{1}{\lambda}\right), \quad b = (1),$$

$$c = \left(\frac{\lambda - 1}{\lambda^2}\right), \quad d = \frac{\lambda - 1}{\lambda},$$

which implies that the geometric density is in $\mathscr{D}$. The transfer function of this realization is

$$G(z) = \frac{\lambda - 1}{\lambda^2 (z - \lambda^{-1})} + \frac{\lambda - 1}{\lambda} = \frac{(\lambda - 1)z}{\lambda z - 1}.$$

Example 3 (*Poisson density*). Consider the Poisson density, i.e. let $p_k = \mathrm{e}^{-\lambda} \lambda^k / k!$ for $k = 0, 1, 2, \ldots$ Then we can see that no finite dimensional system (A, b, c, d) exists that realizes the density, i.e. such that (1) is satisfied. It is well known that the generating function of a random variable X with Poisson density with parameter λ is given by $Es^X = \mathrm{e}^{\lambda(s-1)}$. We know that the generating function of a random variable with density in $\mathscr{D}$ is rational. This therefore shows that the Poisson density is not in $\mathscr{D}$. However, in many applications where detectors have a finite range a Poisson density has to be 'cut off' to properly model the actual detection process. This results in a density that has finite support and therefore falls in our class. The fact that the

Poisson density can be approximated appropriately by one with finite support also follows from the well-known Poisson limit theorem [2].

The material in this paper is organized as follows. In the subsequent section we show how moments of densities in $\mathscr{D}$ can be calculated using state space techniques. In a further section we present state space formulae for standard calculations with probability densities such as scaling, shifting, convolution, etc. This is followed by examples that demonstrate the use of the methods presented here.

We use the following notation: The symbol $\mathscr{C}$ stands for the field of complex numbers. The symbol $\mathscr{D}$ (Section 1) denotes the set of discrete probability densities that admit state space realizations. Similarly, the symbol $\mathscr{U}$ stands for the set that contains the 'un-normalized' probability densities that admit a state space realization (see, Section 1) and the zero sequence.

2. Moments

In this section we are going to study the calculation of moments of random variables with discrete densities in $\mathscr{D}$ from the point of view of state space systems. The nth moment $E(X^n)$ $(n \geqslant 1)$ of a random variable X with discrete probability density $(p_k)_{k\geqslant 0}$ is, if it exists, given by

$$E(X^n) = \sum_{k=0}^{\infty} k^n p_k.$$

Assume that the discrete probability density $(p_k)_{k\geqslant 0}$ is in $\mathscr{D}$ and has a state space realization (A, b, c, d). First consider the Laurent expansion for the transfer function $G(z) = c(zI - A)^{-1}b + d$ given by

$$G(z) = \sum_{k=0}^{\infty} z^{-k} p_k = d + \sum_{k=1}^{\infty} z^{-k} cA^k b.$$

Lemma 1.1 implies that G is analytic in an open neighborhood of infinity that includes the unit circle. Hence, in particular $G(z)$ exists for all $z \in \mathscr{C}$ with $|z| = 1$. We can therefore also define the following differentiation operator combined with a shift by

$$\Delta G(z) := -z\frac{\mathrm{d}}{\mathrm{d}z}G(z) := -z\sum_{k=1}^{\infty}(-k)z^{-k-1}p_k = \sum_{k=1}^{\infty} kz^{-k}p_k = \sum_{k=0}^{\infty} z^{-k}\tilde{p}_k,$$

where we have set $\tilde{p}_0 := 0$, $\tilde{p}_k := kp_k$, $k \geqslant 1$. Note that the above mentioned analyticity of G guarantees that the sum converges in an open neighborhood of infinity that includes the unit circle. We therefore immediately have the first part of the following lemma which is closely related to standard results on the generating function of a random variable with discrete probability density.

Lemma 2.1.
(1) *Let X be a random variable with discrete probability density* $(p_k)_{k\geqslant 0}$ *in* $\mathscr{D}$. *Then*

$$EX = \Delta G(1).$$

(2) *Let* (A, b, c, d) *be a stable realization of the probability density. Then*

$$EX = c(I - A)^{-2}b.$$

Proof.
(1) Follows from the above arguments since

$$EX = \sum_{k=0}^{\infty} kp_k = \Delta G(1).$$

(2) We have

$$\begin{aligned}\Delta G(z) &= -z\frac{\mathrm{d}}{\mathrm{d}z}(c(zI - A)^{-1}b + d)\\ &= -zc\left(-(zI - A)^{-1}\left(\frac{\mathrm{d}}{\mathrm{d}z}(zI - A)\right)(zI - A)^{-1}\right)b\\ &= zc\left((zI - A)^{-1}(zI - A)^{-1}\right)b.\end{aligned}$$

Evaluating this expression at $z = 1$ we obtain the result. Here we use the fact that $I - A$ is invertible because A is assumed have all its eigenvalues in the open unit disk. □

In order to calculate higher order moments we need to collect a few more details.

Lemma 2.2. *Let* (A, b, c, d) *be a system and let*

$$r_0 := d, \quad r_k := cA^{k-1}b, \;\; k = 1, 2, \ldots$$

Assume that
(1) $r_k \geqslant 0$ *for all* $k = 0, 1, 2, \ldots$
(2) *All eigenvalues of A have modulus strictly less than* 1.
Then $(r_k)_{k\geqslant 0} \in \mathscr{U}$.

This is the consequence of a standard stability result (see e.g. [1, p. 208]).

Proposition 2.1. *Let X be a random variable with discrete probability density* $(p_k)_{k\geqslant 0}$ *in* $\mathscr{D}$ *and let G be its transfer function.*
(1) *Let* $(\tilde{p}_k)_{k\geqslant 0}$ *be the Laurent coefficients of* ΔG, *i.e.* $\tilde{p}_0 := 0$, $\tilde{p}_k := kp_k$, $k \geqslant 1$. *Then* $(\tilde{p}_k)_{k\geqslant 0} \in \mathscr{U}$.
(2) *For* $n \geqslant 1$,

$$(\Delta^n G)(z) = \sum_{k=1}^{\infty} k^n z^{-k} p_k.$$

(3) *For* $n \geqslant 1$, *the moment* EX^n *exists and is given by*

$$EX^n = (\Delta^n G)(1).$$

(4) *Let* (A, b, c, d) *be an asymptotically stable state space representation of* $\Delta^n G$, $n = 0, 1, 2, \ldots$, *i.e.*

$$\Delta^n G(z) = c(zI - A)^{-1}b + d,$$

(*here* $\Delta^0 G = G$). *Then*

$$\Delta^{n+1} G(z) = zc(zI - A)^{-2}b = c_d(zI - A_d)^{-1}b_d,$$

where

$$A_d = \begin{pmatrix} A & 0 \\ I & A \end{pmatrix}, \quad b_d = \begin{pmatrix} b \\ 0 \end{pmatrix}, \quad c_d = \begin{pmatrix} c & cA \end{pmatrix}.$$

Proof. (1) Clearly $\tilde{p}_k \geqslant 0$ for $k = 0, 1, 2, \ldots$ That this sequence has a rational transfer function was shown in the proof of Lemma 2.1, part (2). Therefore a state space realization of $(\tilde{p}_k)_{k\geqslant 0}$ exists. From Lemma 2.1, part (2), it follows that $\sum_{k=0}^{\infty} \tilde{p}_k < \infty$. Therefore $(\tilde{p}_k)_{k\geqslant 0}$ is an element of the set $\mathscr{U}$, by definition.

(2) and (3) By Lemma 2.1 the result is correct for $n = 1$ and by (1) the Laurent coefficients of ΔG are in $\mathscr{U}$ and are given by $\tilde{p}_0 = 0$, $\tilde{p}_k = kp_k$. Now assume that the result is correct for $n = n_0 \geqslant 1$ and that the Laurent coefficients $(\tilde{p}_k)_{k\geqslant 0}$ of $(\Delta^{n_0} G)(z)$ are in $\mathscr{U}$ and are given by $\tilde{p}_0 = 0$, $\tilde{p}_k = k^{n_0} p_k$. Then

$$(\Delta^{n_0+1} G)(z) = \Delta\left(\sum_{k=1}^{\infty} z^{-k}\tilde{p}_k\right) = \sum_{k=1}^{\infty} kz^{-k}\tilde{p}_k = \sum_{k=1}^{\infty} z^{-k}k^{n_0+1}p_k.$$

By (1) the Laurent coefficients of $(\Delta^{n_0+1} G)(z) = \Delta\,(\Delta^{n_0} G)\,(z)$ are in $\mathscr{U}$. Hence, the induction is complete. We therefore have for all $n \geqslant 1$ that EX^n exists and is given by

$$(\Delta^n G)(1) = \sum_{k=0} k^n p_k = EX^n.$$

(4) For $n = 0$ this result follows from the proof of Lemma 2.1 and a simple computation. For $n \geqslant 1$ the proof is analogous. □

Recursively applying part (4) of the previous result will provide state space representations of $\Delta^n G$ for $n \geqslant 1$. These state space representations can then be used with part (1) of the result to obtain a state space formulation for the nth moment. As an example we are now going to provide a state space description for the second moment.

Example. Let X be a random variable whose discrete probability density is in $\mathscr{D}$ and has an asymptotically stable state space realization (A, b, c, d). Then by Proposition 2.1 part (4) ΔG has a state space realization (A_1, b_1, c_1, d_1) given by

$$A_1 = \begin{pmatrix} A & 0 \\ I & A \end{pmatrix}, \quad b_1 = \begin{pmatrix} b \\ 0 \end{pmatrix}, \quad c_1 = \begin{pmatrix} c & cA \end{pmatrix}, \quad d_1 = 0.$$

Note that this realization is again asymptotically stable, in fact the set of eigenvalues of A_1 is equal to the set of eigenvalues of A, which forms a subset of the open unit disk. By Proposition 2.1 part (4) the first moment *EX* is then

$$EX = c_1(I - A_1)^{-1}b_1.$$

To obtain an asymptotically stable state space representation (A_2, b_2, c_2, d_2) of $\Delta^2 G$ we apply Proposition 2.1 to the state space representation (A_1, b_1, c_1, d_1) of ΔG and obtain

$$A_2 = \left(\begin{array}{cc|cc} A & 0 & 0 & 0 \\ I & A & 0 & 0 \\ \hline I & 0 & A & 0 \\ 0 & I & I & A \end{array}\right), \quad b_2 = \left(\begin{array}{c} b \\ 0 \\ \hline 0 \\ 0 \end{array}\right),$$

$$c_2 = \left(\begin{array}{cc|cc} c & cA & 2cA & cA^2 \end{array}\right), \quad d_2 = 0.$$

Now by Proposition 2.1 part (3) we have that

$$EX^2 = c_2(I - A_2)^{-1}b_2.$$

This essentially establishes the required result. It is, however, worthwhile pointing out that the state space representation (A_2, b_2, c_2, d_2) is far from minimal, i.e. another realization of $\Delta^2 G$ exists with lower state space dimension. To see this we perform a state space transformation with the non-singular matrix

$$T = \begin{pmatrix} I & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & 0 & I \\ 0 & -I & I & 0 \end{pmatrix},$$

i.e. consider the system

$$(\tilde{A}_2, \tilde{b}_2, \tilde{c}_2, \tilde{d}_2) := (TA_2T^{-1}, Tb_2, c_2T^{-1}, d_2),$$

which has the same transfer function $\Delta^2 G$ as (A_2, b_2, c_2, d_2). Now,

$$\tilde{A}_2 = \begin{pmatrix} I & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & 0 & I \\ 0 & -I & I & 0 \end{pmatrix} \begin{pmatrix} A & 0 & 0 & 0 \\ I & A & 0 & 0 \\ 0 & 0 & A & 0 \\ 0 & I & I & A \end{pmatrix} \begin{pmatrix} I & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & I & 0 & I \\ 0 & 0 & I & 0 \end{pmatrix}$$

$$= \left(\begin{array}{ccc|c} A & 0 & 0 & 0 \\ I & A & 0 & 0 \\ 0 & 2I & A & I \\ \hline 0 & 0 & 0 & A \end{array}\right),$$

$$\tilde{b}_2 = \begin{pmatrix} I & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & 0 & I \\ 0 & -I & I & 0 \end{pmatrix} \begin{pmatrix} b \\ 0 \\ 0 \\ 0 \end{pmatrix} = \left(\begin{array}{c} b \\ 0 \\ 0 \\ \hline 0 \end{array}\right),$$

$$\tilde{c}_2 = \begin{pmatrix} c & cA & 2cA & cA^2 \end{pmatrix} \begin{pmatrix} I & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & I & 0 & I \\ 0 & 0 & I & 0 \end{pmatrix} = \left(c \quad 3cA \quad cA^2 \;\middle|\; 2cA \right),$$

$$\tilde{d}_2 = 0.$$

As can easily be verified (see also the Kalman decomposition in e.g. [1,5]) we have that the subsystem $(\hat{A}_2, \hat{b}_2, \hat{c}_2, \hat{d}_2)$

$$\hat{A}_2 = \begin{pmatrix} A & 0 & 0 \\ I & A & 0 \\ 0 & 2I & A \end{pmatrix}, \quad \hat{b}_2 = \begin{pmatrix} b \\ 0 \\ 0 \end{pmatrix},$$

$$\hat{c}_2 = \begin{pmatrix} c & 3cA & cA^2 \end{pmatrix}, \quad \tilde{d}_2 = 0,$$

has the same transfer function ΔG^2 as the higher dimensional full system (A_2, b_2, c_2, d_2). Furthermore this realization is again asymptotically stable. Hence

$$EX^2 = \hat{c}_2(I - \hat{A}_2)^{-1}\hat{b}_2.$$

If the nth moment is to be calculated, a realization of $\Delta^n G(z)$ can be computed by repeated application of the realization formula for the image of Δ when applied to a transfer function. If the realization for $\Delta^n G(z)$ that is found in this way is (A_n, b_n, c_n, d_n) then the nth moment is given by $E(X^n) = c_n(I - A_n)^{-1}b_n + d_n$. There is an alternative way to calculate these moments that circumvents the realization of $\Delta^n G(z)$ and is given in terms of the realization (A, b, c, d) of $G(z)$. We have the following proposition.

Proposition 2.2. *Let X be a random variable with discrete probability density* $(p_k)_{k\geqslant 0}$ *in* $\mathscr{D}$. *Let* (A, b, c, d) *be an asymptotically stable realization of the probability density and G its transfer function. Define a sequence of polynomials*

$$q(\alpha; n) = q_0(n) + q_1(n)\alpha + \cdots + q_{n-1}(n)\alpha^{n-1}, \quad n = 1, 2, \ldots,$$

recursively as follows:

$$q(\alpha; 1) = 1,$$

$$q(\alpha; n+1) = (n\alpha + 1)q(\alpha; n) + \alpha(\alpha - 1)q'(\alpha; n), \quad n = 1, 2, \ldots,$$

where $q'(\alpha; n)$ *denotes the derivative of* $q(\alpha; n)$ *with respect to* α. *Then the* nth *moment of the density with transfer function G and asymptotically stable realization* (A, b, c, d) *is given by*

$$E(X^n) = cq(A;n)(I-A)^{-n-1}b = \sum_{j=0}^{n-1} q_j(n)cA^j(I-A)^{-n-1}b, \quad n=1,2,\ldots$$

Proof. The proof will be by induction on $n=1,2,\ldots$ Use will be made of the homogeneous polynomials $\tilde{q}(z,\alpha;n)$, $n=1,2,\ldots$, that are related to the polynomials $q(\alpha;n)$, $n=1,2,\ldots$, by the formula

$$\tilde{q}(z,\alpha;n) = z^{n-1}q\left(\frac{\alpha}{z};n\right), n=1,2,\ldots$$

For each $n=1,2,\ldots$ the degree of homogeneity of $\tilde{q}(z,\alpha;n)$ is $n-1$, which is also the degree of $q(\alpha;n)$. This can be shown easily by induction from the definition of $q(\alpha;n)$, $n=1,2,\ldots$ Therefore Euler's theorem on homogeneous functions can be applied to give the following equality:

$$\frac{\partial\tilde{q}(z,\alpha;n)}{\partial z} = (n-1)\frac{\tilde{q}(z,\alpha;n)}{z} - \frac{\alpha}{z}\frac{\partial\tilde{q}(z,\alpha;n)}{\partial\alpha}, \quad n=1,2,\ldots$$

Note further that $\tilde{q}(1,\alpha;n)=q(\alpha;n)$, $n=1,2,\ldots$, and

$$\frac{\partial\tilde{q}(1,\alpha;n)}{\partial\alpha} = q'(\alpha;n),\ n=1,2,\ldots$$

What will be shown is that the following equality holds:

$$\Delta^n G(z) = z^n c\tilde{q}(z,A;n)(zI-A)^{-n-1}b, \quad n=1,2,\ldots$$

From this the formula in the proposition follows by substituting $z=1$. The inverse $(I-A)^{-1}$ exists because (A,b,c,d) is an asymptotically stable realization. For $n=1$ this says that $\Delta G(z)=zc(zI-A)^{-2}b$, which was shown to hold in Lemma 2.1. Now suppose that the formula is correct for $n=n_0$. By applying Δ to the formula for $\Delta^{n_0}G(z)$ we obtain the following equality:

$$\begin{aligned}\Delta^{n_0+1}G(z) &= \Delta\big(zc\tilde{q}(z,A;n_0)(zI-A)^{-(n_0+1)}b\big)\\ &= -zc\tilde{q}(z,A;n_0)(zI-A)^{-(n_0+1)}\\ &\quad -z^2c\frac{\partial\tilde{q}(z,A;n_0)}{\partial z}(zI-A)^{-(n_0+1)}\\ &\quad -z^2c\tilde{q}(z,A;n_0)(-n_0-1)(zI-A)^{-(n_0+1)}b.\end{aligned}$$

The three terms appearing in this formula can be rewritten as follows. The first one can be rewritten as

$$zc\big(-\tilde{q}(z,A;n_0)(zI-A)\big)(zI-A)^{-n_0-2}b.$$

Using Euler's formula presented above, the second term can be rewritten as

$$zc\left(-(n_0-1)\tilde{q}(z,A;n_0)(zI-A) + A\frac{\partial\tilde{q}}{\partial\alpha}(z,A;n_0)(zI-A)\right)\times(zI-A)^{-n_0-2}b.$$

The third term can be rewritten as

$$zc\,((n_0+1)z\tilde{q}(z,A;n_0))\,(zI-A)^{-n_0-2}b.$$

Adding up these terms we obtain

$$\Delta^{n_0+1}G(z) = zc\left((n_0A+zI)\tilde{q}(z,A;n_0) + A(zI-A)\frac{\partial\tilde{q}(z,A;n_0)}{\partial\alpha}\right) \times (zI-A)^{-n_0-2}b,$$

which can easily be seen to be equal to

$$c\tilde{q}(z,a;n_0+1)(zI-A)^{-n_0-2}b$$

as was to be shown. □

3. Operations on probability densities

In this section we are going to discuss state-space formulations of operations on discrete probability density functions in $\mathscr{D}$.

3.1. Convolution of two probability densities

If X_1, X_2 are two independent random variables with probability densities ρ_1, ρ_2, then it is well known that $X = X_1 + X_2$ has probability density $\rho = \rho_1 * \rho_2$, where $*$ denotes the convolution [2]. The following proposition shows that if $\rho_1, \rho_2 \in \mathscr{D}$ then ρ is also in $\mathscr{D}$. An explicit realization is available in terms of the realization of ρ_1 and ρ_2.

Proposition 3.1. *Let ρ_i, $i = 1, 2$, be two discrete probability densities in $\mathscr{D}$, with state space realizations (A_i, b_i, c_i, d_i) and transfer functions G_i, $i = 1, 2$.*

Then the convolution ρ, i.e. $\rho := \rho_1 * \rho_2$ is in $\mathscr{D}$ and has a realization (A_c, b_c, c_c, d_c) and transfer function G given by

$$A_c = \begin{pmatrix} A_1 & b_1c_2 \\ 0 & A_2 \end{pmatrix}, \quad b_c = \begin{pmatrix} b_1d_2 \\ b_2 \end{pmatrix},$$

$$c_p = \begin{pmatrix} c_1 & d_1c_2 \end{pmatrix}, \quad d_c = d_1d_2,$$

$$G_c(z) = G_1(z)G_2(z).$$

If the realizations (A_i, b_i, c_i, d_i), $i = 1, 2$, are asymptotically stable, then the realization (A_c, b_c, c_c, d_c) is asymptotically stable as well.

Proof. That $G_c = G_1G_2$ follows since the z-transform transforms convolution into a pointwise product of the transfer functions. It is easily verified that (A_c, b_c, c_c, d_c)

is a realization of G_c. From the block upper triangular structure of A_c it follows that $\det(zI - A_c) = \det(zI - A_1)\det(zI - A_2)$. Therefore any eigenvalue of A_c is either an eigenvalue of A_1 or an eigenvalue of A_2 or both. So if (A_1, b_1, c_1, d_1) and (A_2, b_2, c_2, d_2) are asymptotically stable realizations then the realization (A_c, b_c, c_c, d_c) is also asymptotically stable. That ρ is a probability density is a standard result, which can also be seen from the fact that $G_c(1) = G_1(1)G_2(1)$. Hence ρ is in $\mathscr{D}$. □

3.2. Translation and scaling of a probability density

In the next straightforward lemma the effect of translation and scaling of a random variable on the state space realization of the density is considered.

Proposition 3.2. *Let X be a random variable with probability density ρ in $\mathscr{D}$. Let (A, b, c, d) be a realization of ρ with transfer function G.*

(1) *Let $x_0 > 0$ be an integer. Then the random variable $X + x_0$ has a probability density $q(x) = \rho(x - x_0)$ which has a realization (A_t, b_t, c_t, d_t) with transfer function G_t given by*

$$A_t = \left(\begin{array}{ccccc|c} 0 & 0 & \cdots & 0 & 0 & c \\ 1 & 0 & \cdots & 0 & 0 & 0\cdots0 \\ 0 & 1 & \cdots & 0 & 0 & \vdots \\ \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0\cdots0 \\ \hline 0 & 0 & \cdots & 0 & 0 & A \end{array}\right), \quad b_t = \begin{pmatrix} d \\ 0 \\ \vdots \\ 0 \\ \hline b \end{pmatrix},$$

$$c_t = \left(\begin{array}{ccccc|c} 0 & 0 & \cdots & 0 & 1 & 0\cdots0 \end{array}\right), \quad d_t = 0,$$

and

$$G_t(z) = z^{-x_0}G(z), \quad z \in \mathscr{C}.$$

The size of the first block diagonal subblock of A_t is $x_0 \times x_0$. If (A, b, c, d) is an asymptotically stable realization then so is (A_t, b_t, c_t, d_t).

(2) *Let X be a random variable with probability density ρ is in $\mathscr{D}$. Let (A, b, c, d) be an n-dimensional realization of ρ with transfer function G.*
Let $a \geqslant 1$ be an integer. Then the random variable aX has a probability density ρ_a which is in $\mathscr{D}$ and has a realization (A_s, b_s, c_s, d_s) given by

$$A_s = \begin{pmatrix} 0 & 0 & \cdots & 0 & A \\ I_n & 0 & \cdots & 0 & 0 \\ 0 & I_n & \ddots & 0 & 0 \\ \ddots & \ddots & & \ddots & \vdots \\ 0 & 0 & \cdots & I_n & 0 \end{pmatrix}, \quad b_s = \begin{pmatrix} b \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

$$c_s = \begin{pmatrix} 0 & 0 & \cdots & 0 & c \end{pmatrix}, \quad d_s = d.$$

The system matrices are presented in block partitioned form with A_s being an $a \times a$ block matrix whose blocks are of size $n \times n$. Its transfer function G_s is given by $G_s(z) := G_s(z^a)$, $z \in \mathscr{C}$. If (A, b, c, d) is an asymptotically stable realization, then so is (A_s, b_s, c_s, d_s).

Proof. (1) First note that x_0 can be considered to be a random variable with probability density function $p_k = 0$ for all $k \neq x_0$ and $p_{x_0} = 1$. Note that $(p_k)_{k\geqslant 0} \in \mathscr{D}$. By Example 1 in Section 1 we have that the x_0-dimensional system (A_1, b_1, c_1, d_1) is a minimal realization of $(p_k)_{k\geqslant 0}$, where

$$A_1 = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}, \quad b_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

$$c_1 = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 \end{pmatrix}, \quad d_1 = 0.$$

By the convolution formulae of Proposition 3.1 we have the desired result. Because A_1 is a nilpotent matrix, its only eigenvalue is zero and hence (A_1, b_1, c_1, d_1) is asymptotically stable. If (A, b, c, d) is asymptotically stable, it then follows from Proposition 3.1 that (A_t, b_t, c_t, d_t) is asymptotically stable as well.

(2) If the random variable X has probability density $\rho = (\rho(k))_{k\geqslant 0}$ then aX with $a > 1$, an integer, has probability density $\rho_a = (\rho_a(k))_{k\geqslant 0}$, with $\rho_a(k) = \rho(n)$ for each k such that $k = an$ for some positive integer n and $\rho_a(k) = 0$ otherwise. This implies that the transfer function of ρ_a is given by

$$G_a(z) = \sum_{k=0}^{\infty} z^{-k}\rho_a(k) = \sum_{k=0}^{\infty} z^{-ak}\rho(k) = \sum_{k=0}^{\infty} (z^a)^{-k}\rho(k) = G(z^a),$$

where G is the transfer function of ρ.

To verify that (A_s, b_s, c_s, d_s) is indeed a realization of ρ_a first note that

$$\begin{aligned} A_s b_s &= \begin{pmatrix} 0 & b^{\mathrm{T}} & 0 & \cdots & 0 \end{pmatrix}^{\mathrm{T}} \\ A_s^2 b_s &= \begin{pmatrix} 0 & 0 & b^{\mathrm{T}} & 0 & \cdots & 0 \end{pmatrix}^{\mathrm{T}} \\ &\vdots \\ A_s^{a-1} b_s &= \begin{pmatrix} 0 & 0 & 0 & 0 & \cdots & b^{\mathrm{T}} \end{pmatrix}^{\mathrm{T}} \\ A_s^a b_s &= \begin{pmatrix} b^{\mathrm{T}} A^{\mathrm{T}} & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}^{\mathrm{T}}. \end{aligned}$$

Continuing inductively, this shows that

$$\begin{aligned}&d_s = d = \rho_a(0) = \rho(0), \quad c_s b_s = 0 = \rho_a(1),\\ &c_s A_s b_s = 0 = \rho_a(2), \ldots,\\ &c_s A_s^{a-2} b_s = 0 = \rho_a(a-1), \quad c_s A_s^{a-1} b_s = \rho(1) = \rho_a(a),\\ &c_s A_s^{a} b_s = 0 = \rho_a(a+1), \ldots\end{aligned}$$

Recursively, this shows that (A_s, b_s, c_s, d_s) is a realization of ρ_a. Hence ρ_a is in $\mathscr{D}$. Now suppose (A, b, c, d) is an asymptotically stable realization. Then the eigenvalues of A all have modulus less than one. Now suppose λ is an eigenvalue of A_s and let $\xi = (\xi_1^{\mathrm{T}}, \ldots, \xi_a^{\mathrm{T}})^{\mathrm{T}}$ be an associated eigenvector, partioned into a subvectors of size n. From the eigenvalue equation $A_s\xi = \lambda\xi$ it follows that $A\xi_a = \lambda\xi_1$ and $\xi_i = \lambda\xi_{i+1},\ i = 1, \ldots, a-1$. Therefore $A\xi_a = \lambda^a\xi_a$, which implies that λ^a is an eigenvalue of the matrix A. Hence $|\lambda^a| < 1$, which implies $|\lambda| < 1$, because a is positive. Therefore the realization (A_s, b_s, c_s, d_s) is asymptotically stable. □

3.3. Product of two probability densities

In conditional probability calculations it is regularly necessary to calculate the product of two probability densities. The following result shows that multiplication of probability densities leads to a sequence in $\mathscr{U}$. If this sequence is non-zero, a normalization results in a probability density in $\mathscr{D}$. For an example of such calculations see Section 3.6.

Proposition 3.3. *Let ρ_i, $i = 1, 2$, be two discrete probability densities in $\mathscr{D}$, with state space realizations (A_i, b_i, c_i, d_i) and transfer functions G_i, $i = 1, 2$.*

Consider the pointwise product $\tilde{\rho}$, i.e. $\tilde{\rho}(k) := \rho_1(k)\rho_2(k),\ k = 0, 1, 2, \ldots$ Assume that $\tilde{\rho}$ is not identically zero, i.e. that $\sum_{k=0}^{\infty} \tilde{\rho}(k) > 0$. Then ρ is in $\mathscr{D}$, where

$$\rho(k) := \frac{1}{\sum_{k=0}^{\infty} \tilde{\rho}(k)} \tilde{\rho}(k), \quad k = 0, 1, 2, \ldots$$

The system $(\tilde{A}_p, \tilde{b}_p, \tilde{c}_p, \tilde{d}_p)$ given by

$$\tilde{A}_p = (A_1 \otimes A_2), \quad \tilde{b}_p = (b_1 \otimes b_2),$$

$$\tilde{c}_p = (c_1 \otimes c_2), \quad \tilde{d}_p = d_1 d_2$$

(here $\otimes$ denotes the Kronecker product) is a realization of $\tilde{\rho}$, i.e.

$$\tilde{\rho}(0) = \tilde{d}_p, \quad \tilde{\rho}(k) = \tilde{c}_p \tilde{A}_p^{k-1} \tilde{b}_p, \ \ k = 1, 2, \ldots$$

If (A_1, b_1, c_1, d_1), (A_2, b_2, c_2, d_2) are asymptotically stable realizations, then so is $(\tilde{A}_p, \tilde{b}_p, \tilde{c}_p, \tilde{d}_p)$. If $(\tilde{A}_p, \tilde{b}_p, \tilde{c}_p, \tilde{d}_p)$ is an asymptotically stable realization, then the system (A_p, b_p, c_p, d_p) given by

$$A_p := \tilde{A}_p, \quad b_p := \tilde{b}_p,$$

$$c_p := c\tilde{c}_p, \quad d_p := c\tilde{d}_p$$

with $c = 1/(\tilde{c}_p(I - \tilde{A}_p)^{-1}\tilde{b}_p + \tilde{d}_p)$, *is an asymptotically stable realization of* ρ.

Proof. We first verify that the system $(\tilde{A}_p, \tilde{b}_p, \tilde{c}_p, \tilde{d}_p)$ is a realization of $\tilde{\rho}$. For $k = 1, 2, \ldots,$ we have

$$\begin{aligned}\tilde{\rho}(k) &= \rho_1(k)\rho_2(k)\\ &= \big(\tilde{c}_1\tilde{A}_1^{k-1}\tilde{b}_1\big)\big(\tilde{c}_2\tilde{A}_2^{k-1}\tilde{b}_2\big)\\ &= \big(\tilde{c}_1\tilde{A}_1^{k-1}\tilde{b}_1\big) \otimes \big(\tilde{c}_2\tilde{A}_2^{k-1}\tilde{b}_2\big)\\ &= \big(\tilde{c}_1 \otimes \tilde{c}_2\big)\big(\tilde{A}_1^{k-1}\tilde{b}_1 \otimes \tilde{A}_2^{k-1}\tilde{b}_2\big)\\ &= \big(\tilde{c}_1 \otimes \tilde{c}_2\big)\big(\tilde{A}_1^{k-1} \otimes \tilde{A}_2^{k-1}\big)\big(\tilde{b}_1 \otimes \tilde{b}_2\big)\\ &= \big(\tilde{c}_1 \otimes \tilde{c}_2\big)\big(\tilde{A}_1 \otimes \tilde{A}_2\big)^{k-1}\big(\tilde{b}_1 \otimes \tilde{b}_2\big),\end{aligned}$$

and for $k = 0$,

$$\tilde{\rho}(0) = \rho_1(0)\rho_2(0) = d_1 d_2.$$

Each eigenvalue η of $\tilde{A}_p$ can be written as a product $\eta = \lambda\mu$, where λ and μ are eigenvalues of A_1 and A_2, respectively. Therefore, if (A_1, b_1, c_1, d_1) and (A_2, b_2, c_2, d_2), are asymptotically stable realizations, then $|\lambda| < 1$ and $|\mu| < 1$, hence $|\eta| = |\lambda||\mu| < 1$ and so $(\tilde{A}_p, \tilde{b}_p, \tilde{c}_p, \tilde{d}_p)$ is an asymptotically stable realization. Here we have used standard results on Kronecker products (see, e.g. [6]). This shows $(A_p, b_p, c_p, d_p) = (A_1 \otimes A_2, b_1 \otimes b_2, c \otimes c_2, d_1 d_2)$ is a realization of $\tilde{\rho} = \rho_1\rho_2$. Note that

$$\sum_{k=0}^{\infty} \tilde{p}_k = \tilde{d}_k + \sum_{k=1}^{\infty} \tilde{c}_p\tilde{A}_p^{k-1}\tilde{b}_p = \tilde{G}(1),$$

where $\tilde{G}$ is the transfer function of the system $(\tilde{A}_p, \tilde{b}_p, \tilde{c}_p, \tilde{d}_p)$. Note that by assumption $\sum_{k=0}^{\infty} \tilde{p}_k = \tilde{G}(1) \neq 0$.

Therefore $\rho = (1/\tilde{G}(1))\tilde{\rho}$ is a discrete probability density. If $(\tilde{A}_p, \tilde{b}_p, \tilde{c}_p, \tilde{d}_p)$ is an asymptotically stable realization of $\tilde{\rho}$ then $\tilde{G}(1) = \tilde{c}_p(I - \tilde{A}_p)^{-1}\tilde{b}_p + \tilde{d}_p$ and it easily follows that (A_p, b_p, c_p, d_p) is a realization of ρ with $G(1) = 1$. This shows that ρ is in $\mathscr{D}$. □

3.4. Mixing of random variables

One way to construct new random variables is by mixing of random variables. The next proposition shows that if X_1, X_2 have probability densities in $\mathscr{D}$ then a mixing of these random variables leads to a new random variable Y which has a probability density in $\mathscr{D}$.

Proposition 3.4. *Let X_1, X_2 be random variables with discrete probability densities ρ_1, ρ_2 in $\mathscr{D}$. Let $0 < p < 1$ and define a random variable Y with probability density ρ_Y by*

$$Y = \begin{cases} X_1 & \text{with probability } p, \\ X_2 & \text{with probability } 1-p. \end{cases}$$

Then:

(1) *The probability density ρ_Y of Y is in $\mathscr{D}$.*

(2) *If (A_i, b_i, c_i, d_i) is a realization of ρ_i, $i = 1, 2$, then (A_Y, b_Y, c_Y, d_Y) given by*

$$A_Y = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}, \quad b_Y = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix},$$

$$c_Y = \begin{pmatrix} pc_1 & (1-p)c_2 \end{pmatrix}, \quad d_Y = pd_1 + (1-p)d_2,$$

is a realization of ρ_Y. If (A_1, b_1, c_1, d_1) and (A_2, b_2, c_2, d_2), are asymptotically stable realizations, then (A_Y, b_Y, c_Y, d_Y) is an asymptotically stable realization.

(3) *If G_i is the transfer function of ρ_i, $i = 1, 2$, then the transfer function G_Y of ρ_Y is given by*

$$G_Y = pG_1 + (1-p)G_2.$$

Proof. (2) The proof is a straightforward verification. The set of eigenvalues of A_Y is the union of the sets of eigenvalues of A_1 and A_2. Therefore if (A_i, b_i, c_i, d_i), $i = 1, 2$, are asymptotically stable realizations, then all these eigenvalues have modulus less than one, hence (A_Y, b_Y, c_Y, d_Y) is an asymptotically stable realization.

(1) Since ρ_Y has a realization by (2) we have that ρ_Y is in $\mathscr{D}$.

(3) Follows immediately from (2). □

3.5. Realizations of discrete probability densities

An important consequence of the operations that were presented earlier is that the class of densities $\mathscr{D}$ is closed under these operations, i.e. that all the respective calculations can be performed without leaving the class. In particular, this means that all the calculations can be done using state space methods.

An interesting question in the same context is which discrete-time systems (A, b, c, d) give rise to discrete probability densities in $\mathscr{D}$. This issue is addressed in the following proposition.

Proposition 3.5. *Let (A, b, c, d) be an n-dimensional discrete-time system, i.e. let A be an $n \times n$ matrix, b an $n \times 1$ vector, c a $1 \times n$ vector and d a scalar. Let $(p_k)_{k \geqslant 0}$ be defined by*

$$p_0 := d, \quad p_k := cA^{k-1}b, \ \ k = 1, 2, 3, \ldots$$

(1) *If* (A, b, c, d) *is an* internally positive system, *i.e. all entries of* A, b, c *and* d *are non-negative, then* $p_k \geqslant 0$, $k = 0, 1, 2, \ldots$

(2) *Assume that* (A, b, c, d) *is minimal. We have that* $\sum_{k=0}^{\infty} |p_k| < \infty$ *if and only if* (A, b, c, d) *is asymptotically stable, i.e. if and only if all eigenvalues of A are strictly less than* 1.

(3) *Assume that* (A, b, c, d) *is asymptotically stable and that* $p_k \geqslant 0$, $k = 0, 1, 2, \ldots$ *Then* $(p_k)_{k\geqslant 0}$ *is in* $\mathscr{D}$ *if and only if* $c(I - A)^{-1}b + d = 1$.

(4) *Assume that* (A, b, c, d) *is asymptotically stable, that* $p_k \geqslant 0$, $k = 0, 1, 2, \ldots,$ *and that* $c(I - A)^{-1}b + d > 0$. *Then the system* $(A, b, \tilde{c}, \tilde{d})$ *with*

$$\tilde{c} := \frac{1}{c(I - A)^{-1}b + d}c,$$

$$\tilde{d} := \frac{1}{c(I - A)^{-1}b + d}d$$

is such that $(\tilde{p}_k)_{k\geqslant 0}$ *is in* $\mathscr{D}$, *where*

$$\tilde{p}_0 := \tilde{d}, \quad \tilde{p}_k := \tilde{c}A^{k-1}b, \ \ k = 1, 2, 3, \ldots$$

Proof. (1) This is elementary.

(2) This is a restatement of a fundamental stability result for finite dimensional discrete time systems (see, e.g. [1, p. 208]).

(3) This follows since $c(I - A)^{-1}b + d = \sum_{k=0}^{\infty} p_k$, if the system is asymptotically stable.

(4) The system $(A, b, \tilde{c}, \tilde{d})$ is the result of a simple rescaling. Hence the result follows from (3). □

An important consequence of our results above is that the defined state space operations preserve internal positivity of the systems involved.

We will not discuss the inherently difficult question to characterize all systems whose impulse response, i.e. in our case whose associated probability density, is positive. We refer the interested reader to the literature (see, e.g. [3] and the references therein). While this is unquestionably an important theoretical issue, it is not crucial to apply the results presented here. Our results only require that a realization exists and that it can be calculated. These results are classical and are discussed e.g. in [5] or [1].

3.6. Applications

We are now going to briefly discuss how the above presented formulae can be used in practical examples. The emphasis here is not to discuss implementation issues in great detail but rather to illustrate the above presented techniques. To this end we consider the following example.

Example. Suppose an observation $Y = hX + V$ is made of a random variable X, the signal, scaled with a positive integer factor h, with some additive measurement noise V. Assume that X and V are independent random variables both with probability density in the class $\mathscr{D}$. The task is to a find an expression for the conditional probability for X given a measured value y of Y, i.e. $\rho_{X|Y}(X = x|Y = y)$.

Let (A_X, b_X, c_X, d_X) denote a realization of the probability density ρ_X of X and (A_V, b_V, c_V, d_V) a realization of the probability density ρ_V of V. Let $S_Y = \{y \in \{0, 1, 2, \ldots\} \mid Pr(Y = y) > 0\}$ denote the support of Y.

The probability density for X given Y is given by the Bayesian formula [2]

$$\rho_{X|Y}(X = x|Y = y) = N_{X|Y}^{-1}\rho_{Y|X}(Y = y|X = x)\rho(X = x),$$

where $N_{X|Y}$ is the scaling factor designed to guarantee that the right-hand side defines a probability density in x. Its precise form is not relevant here. If $X = x$ is given then observing $Y = y \in S_Y$ is equivalent to observing $V = y - hx$. Therefore the first factor on the right-hand side of the Bayesian formula is equal to

$$\begin{aligned}\rho_{Y|X}(Y = y|X = x) &= \rho_V(V = y - hx)\\ &= \begin{cases} c_V A_V^{y-hx-1} b_V & \text{if } y - hx > 0,\\ d_V & \text{if } y - hx = 0.\end{cases}\end{aligned}$$

Consider the sequence of non-negative numbers

$$p(x) := \begin{cases} c_V A_V^{y-hx-1} b_V & \text{for } x = 0, 1, 2, 3, \ldots, n(y) - 1,\\ r & \text{for } x = n(y),\\ 0 & \text{for } x = n(y) + 1, \ldots \end{cases}$$

with

$$r := \begin{cases} c_V A_V^{y-hx-1} b_V & \text{if } n(y) < y/h,\\ d_V & \text{if } n(y) = y/h,\end{cases}$$

where $n(y) = [y/h]$ is the largest integer less than or equal to y/h. All elements of $p := (p(x))_{x \geqslant 0}$ are non-negative. Since by assumption $Pr(Y = y) > 0$, standard arguments show that not all elements of p can be zero. Hence there exists a constant $K > 0$ such that $\tilde{p} := (1/K)p$ is a probability density of finite support and hence in $\mathscr{D}$. It therefore has a realization $(A, b, \tilde{c}, \tilde{d})$ given by

$$A = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0\\ 1 & 0 & \cdots & 0 & 0\\ 0 & 1 & \cdots & 0 & 0\\ \vdots & \vdots & & \vdots & \vdots\\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}, \quad b = \begin{pmatrix} 1\\ 0\\ \vdots\\ 0 \end{pmatrix},$$

$$c = \begin{pmatrix} p(1) & p(2) & \cdots & p(n-1) & p(n) \end{pmatrix}, \quad d = p(0),$$

and $\tilde{c} = (1/K)c$, $\tilde{d} = (1/K)d$.

Using the above realization, the desired state space realization $(A_{X|Y}, b_{X|Y}, c_{X|Y}, d_{X|Y})$ of the probability density $\rho_{X|Y}(X = x|Y = y)$ can be obtained by applying the formula for the realization of the product of two probability densities from $\mathscr{D}$. It is given by

$$A_{X|Y} = A \otimes A_X,$$

$$b_{X|Y} = b \otimes b_X,$$

$$c_{X|Y} = N^{-1} c \otimes c_X,$$

$$d_{X|Y} = N^{-1} d d_X,$$

where $N = (c \otimes c_X)(I - A \otimes A_X)^{-1}(b \otimes b_X) + dd_X$. (Note that the previous normalization constants are all absorbed in the present normalization constant N.)

It is worthwhile pointing out that a filtering problem can also be easily treated using our framework. Consider the following linear dynamical model with a scalar state space:

$$X_{t+1} = f_t X_t + W_t,$$

$$Y_t = h_t X_t + V_t, \quad t = 0, 1, 2, \ldots,$$

where $X_t, Y_t, V_t, W_t, t = 0, 1, 2, \ldots$, are random variables, $f_t, t = 0, 1, 2, \ldots$, and $h_t, t = 0, 1, 2, \ldots$, are sequences of non-negative integers, X_0 and $V_t, W_t, t=0, 1, 2, \ldots$, stochastically independent with probability densities in $\mathscr{D}$. It follows from our results on scaling and addition of random variables that X_t, $t = 1, 2, \ldots$, and Y_t, $t = 0, 1, 2, \ldots$, also have probability densities in $\mathscr{D}$. The *filtering problem* is to find for each value of $t = 0, 1, 2, \ldots$ the conditional probability density $\rho_{X_t|Y_0,\ldots,Y_t}(X_t = x|Y_0, \ldots, Y_t)$ and the corresponding conditional mean of X_t given the observations $y_0, y_1, \ldots, y_t$. The solution to this problem is given by a recursive set of equations, describing the calculations to be performed at each time step. For brevity we present here only an outline of the steps to be taken. This will show that all conditional probability densities are in our class $\mathscr{D}$ and that they can be calculated using the operations presented above. The fact that all conditional densities occurring in the filtering problem are in the class $\mathscr{D}$ of probability density functions considered is an important advantage of this class, not shared by many other such classes. For linear filtering problems two other such classes are the class of Gaussian probability density functions (the corresponding filter is the well-known Kalman filter) and the class of rational probability density functions (cf. [4]).

Suppose now that the conditional probability density of X_{t_0} given $Y_0 = y_0, \ldots, Y_{t_0} = y_{t_0}$ has been calculated for some non-negative integer t_0 and that it has an associated realization $(A_{t_0|t_0}, b_{t_0|t_0}, c_{t_0|t_0}, d_{t_0|t_0})$. The filtering equations are given by a prediction step and update step.

The prediction step consists of calculating the probability density function

$$p(x) = \rho_{X_{t_0+1}|Y_0,\ldots,Y_{t_0}}(X_{t_0+1} = x|Y_0 = y_0, \ldots, Y_{t_0} = y_{t_0})$$

of $X_{t_0+1} = f_{t_0}X_{t_0} + W_{t_0}$, from the probability density function $q(x) = \rho_{X_{t_0}|Y_0,\ldots,Y_{t_0}}(X_{t_0} = x|Y_0 = y_0, \ldots, Y_{t_0} = y_{t_0})$, the scalar factor f_{t_0} and the probability density function $r(w) = \rho_{W_{t_0}}(W_{t_0} = w)$, where $q, r \in \mathscr{D}$. Because $q \in \mathscr{D}$ and f_{t_0} is a scalar constant the probability density function $s(x) = \rho_{f_{t_0}X_{t_0}|Y_0,\ldots,Y_{t_0}}(f_{t_0}X_{t_0} = x|Y_0 = y_0, \ldots, Y_{t_0} = y_{t_0})$ is again in $\mathscr{D}$ and its state space realization can be calculated from the state space realization of q using the formulas for the scaling operation given in Proposition 3.2, part (2). The probability density function $p(x)$ can be found by using the fact that X_{t_0+1} is the sum of $f_{t_0}X_{t_0}$ and W_{t_0}, which are stochastically independent. Because $s, r \in \mathscr{D}$, their convolution $p = s * r$ is also in $\mathscr{D}$ and its state space realization can be calculated from the state space realizations of s and r, using the formulas for the convolution operation given in Proposition 3.1.

The update step consists of calculating the probability density function

$$u(x) = \rho_{X_{t_0+1}|Y_0,\ldots,Y_{t_0+1}}(X_{t_0+1} = x|Y_0 = y_0, \ldots, Y_{t_0+1} = y_{t_0+1})$$

from the probability density function $p(x)$ of X_{t_0+1} given $Y_0 = y_0, \ldots, Y_{t_0} = y_{t_0}$, and the number y_{t_0+1}. How one can obtain u from p and y_{t_0+1} was described in detail in the previous example, with $Y = Y_{t_0+1}$, $y = y_{t_0+1}$, $h = h_{t_0+1}$, $X = X_{t_0+1}$, $V = V_{t_0+1}$. It follows that $u \in \mathscr{D}$ and a state space realization of u is obtained from the state space realization of p and the number y_{t_0+1}.

4. Concluding remarks

The class of probability density functions on $\{0, 1, 2, \ldots\}$ for which the generating function is a rational function has been investigated. It is shown that using a state space representation of such probability density functions, the moments can be calculated and elementary operations on probability density functions can be performed, using only linear algebra techniques. It is shown as an application that this allows us to solve a linear filtering problem in which the disturbances have distributions of the type investigated. This study should be regarded as a first step in applying state space formulas for probabilistic calculations involving this class of densities. We have not addressed numerical issues here. It should be pointed out, however, that in other areas of applications, computations based on state space realizations have often proved to be numerically better than those based on other approaches. This is due to the fact that the inherent non-uniqueness of a state space realization allows the user to pick a realization that has good numerical properties. The formulae show that the state space dimensions of the systems that realize the probability densities can be expected to be rather large in specific cases, and that the realizations may be non-minimal. That last issue can be resolved by applying a standard minimal realization algorithm to the state space realization (see, e.g. [5]). A further step in this direction to avoid very large state space dimensions would be to apply model reduction techniques to produce a lower order system that approximates the original with sufficient accuracy.

References

[1] F.M. Callier, C.A. Desoer, Linear System Theory, Springer, Berlin, 1991.
[2] K.L. Chung, A Course in Probability Revised, Harcourt, New York, 2000.
[3] L. Farina, S. Rinaldi, Linear Positive Systems, Wiley, New York, 2000.
[4] B. Hanzon, R.J. Ober, A state-space calculus for rational probability density functions and applications to non-Gaussian filtering, SIAM J. Control Optim. 40 (2001) 724–740.
[5] T. Kailath, Linear Systems, Prentice-Hall, Englewood Cliffs, NJ, 1980.
[6] P. Lancaster, M. Tismenetsky, The Theory of Matrices, second ed., Academic Press, New York, 1985.