On a Perron-Frobenius type result for non-negative impulse response functions

Bernard Hanzon*
Edgeworth Centre for Financial Mathematics
School of Mathematical Sciences
University College Cork, Ireland
b.hanzon@ucc.ie

Finbarr Holland
School of Mathematical Sciences
University College Cork, Ireland
f.holland@ucc.ie

20 Mar 2010

*Supported by Science Foundation Ireland under grant numbers RFP2007-MATF802 and 07/MI/008

Abstract

Keywords: Forward rate of interest curves, matrix-exponential distributions, matrix-geometric distributions, non-negative systems, phase-type distributions, exponential-polynomial-trigonometric functions, state-space systems, linear system realization theory, impulse response functions, GARCH models

1 Introduction

The class of impulse response functions of linear dynamical systems with constant coefficients, both in discrete and continuous time, plays an important role in systems and control theory. In various rather different areas of mathematical modelling, including financial mathematics and probability theory the same class of functions appears. In financial mathematics for instance as forward rate curves in interest rate models (such as the Nelson-Siegel [17] and Svensson [18] curves; see e.g. [12], [13],[14], [16], [15]), and in probability theory as probability density functions (such as the phase-type, matrix geometric and matrix exponential probability density functions; for phase-type distributions see for instance [7], [4],[5], [6]; for matrix geometric and matrix exponential distributions see for instance [20], [9]; see also [19] for an application in chemistry). When such functions are used as probability density functions then they obviously have to be non-negative, and they have to sum or integrate to one. In interest rate theory one also would like these functions to be non-negative (to avoid negative interest rates). In systems and control theory the same non-negative restriction appears when so-called externally non-negative systems are considered (see for instance [2]). In quite a number of these cases there is an interesting link with what is called "positive (or non-negative) state-space realization" in systems theory; in probability theory there is an analogue of a positive or non-negative realization in the form of a Markov chain model underlying the probability density function (which then has the interpretation of the probability density function of a first exit time). However it is well-known (see [2]) that the class of systems with non-negative finite dimensional realization is strictly smaller than the class of systems with non-negative response function. For instance the subclass of non-negative trigonometric polynomials (to be precise: non-negative functions of the form $f : [0, \infty) \to \mathbf{R}$, $f(x) = \gamma + \sum_{j=1}^{J} \alpha_j \cos(\omega_j x + \theta_j)$ with $J \geq 1$, real γ, real and positive $\alpha_j > 0, j = 1, 2, \ldots, J$, real positive and distinct $\omega_1 > \omega_2 > \ldots > \omega_J > 0$ and real θ_j, $j = 1, 2, \ldots, J$,) does not allow for a ("continuous time") non-negative realization. For any non-negative impulse response function for which a non-negative realization exists [2] show that the set of poles of the transfer function (which coincides with the spectrum of the dynamic matrix in a minimal realization) must contain a "non-negative real dominant pole" corresponding to the Perron-Frobenius eigenvalue of some non-negative realization. (A similar result has been shown by [4] in the context of phase-type distributions). This "non-negative real dominant pole" is a pole which has modulus at least as large as the modulus of the other poles (hence "dominant") and is real; in the discrete time case it is also non-negative. A natural question to ask is whether such a dominant real pole exists as well in the case of non-negative impulse response functions which do not allow non-negative realization. In the literature it has been stated at several places that this must indeed be the case, such as in [8], and [9]. However, in the first case no proof has been provided, in the second case the proof provided is incorrect (it uses an incorrect argument regarding application of the fact that the rational complex numbers lie dense in the set of complex numbers). The aim of this paper is to first give a short and correct proof for the discrete time case and then derive a proof for the continuous time case from that. The proof is based on the classical Pringsheim theorem from complex function theory. Also some consequences and generalizations of the result will be discussed. We will also show that the existence of a dominant non-negative ("Perron-Frobenius") eigenvalue for any non-negative matrix follows as a corollary to our result. Here we touch on work by Friedland [3] who has given a proof of the Perron-Frobenius theorem based on Pringsheim's theorem.

2 The discrete time case.

As should be clear from the introduction the mathematical result presented here can be applied to different areas of mathematical modelling. These different applications each require a different terminology. Here we choose to use the terminology that corresponds to the impulse response function interpretation. In this section we consider the discrete-time case. In the probability theory application one would speak of the discrete probability case (compare e.g. [20]).

Consider a sequence of non-negative real numbers $\{h_k\}_{k=1}^{\infty}$ for which any of the following three equivalent conditions hold:

(i) the corresponding transfer function $\sum_{k=1}^{\infty} h_k w^{-k}$ is, for $|w| > K$ for some sufficiently large real K, a rational function $G(w) = \frac{p(w)}{q(w)}$ where w is a complex variable, $p(w), q(w)$ are co-prime polynomials with real coefficients and the degree of $p(w)$ is less than the degree n (say) of $q(w)$; we will assume q to be monic without loss of generality

(ii) there exists a minimal realization (A, b, c), where A is an $n \times n$ matrix, c an $1 \times n$ row vector and b and $n \times 1$ column vector, such that $h_k = cA^{k-1}b,\ k = 1, 2, \dots$ (as is well-known, if a complex minimal realization of a real sequence $\{h_k\}_{k=1}^{\infty}$ exists then a real minimal realization of the same order exists as well)

(iii) the Hankel matrix $(H_{ij})_{i=1,2,\dots,N,j=1,2,\dots,N}$ with $H_{ij} = h_{i+j-1},\ i = 1, 2, \dots, N, j = 1, 2, \dots, N$ has finite rank n for all $N \geq n$.

The equivalence of the three conditions is well-known and well-documented in the literature on linear systems theory (see e.g. [11], the equivalence of (i) and (iii) is known as Kronecker's theorem; the equivalence of (iii) and (ii) follows from the Ho-Kalman realization algorithm). We will speak of the sequence $\{h_k\}_{k=1}^{\infty}$ as the impulse response function of the finite dimensional state space system

$$\xi_k = A\xi_{k-1} + b\nu_k, \qquad \eta_k = c\xi_k, \;\; k = 1, 2, 3, \ldots \tag{1}$$

It is also well-known that if (i),(ii),(iii) hold, then $\frac{p(w)}{q(w)} = c(wI - A)^{-1}b$ and the spectrum (collection of eigenvalues) $\sigma(A)$ is equal to the collection of zeros of q which is the collection of poles of the transfer function $G(w)$ of the system (see e.g. [11]). Obviously these poles (eigenvalues) can both be real or complex numbers.

We can now formulate our first main theorem.

Theorem 2.1 *Let $\{h_k\}_{k=1}^{\infty}$ be a non-negative sequence, but not identically zero, that has the interpretation of an impulse response function of a minimal discrete time finite dimensional state space system (A, b, c). Then $\sigma(A)$ contains a real non-negative element λ_M equal to $\lambda_M = \max_{\lambda\in\sigma(A)} |\lambda|$.*

Proof: Consider the function $f(z) = c(I - zA)^{-1}b$. This is equal to $\frac{1}{z}\frac{p(\frac{1}{z})}{q(\frac{1}{z})}$ for every $z \in \mathbf{C}$ for which both functions are defined. It follows that the points of singularity of $f(z)$ (in the terminology used for meromorphic functions, which includes rational functions) are equal to the reciprocals of the non-zero elements in the spectrum $\sigma(A)$, or in other words the reciprocals of the non-zero poles of the system. Because $f(0) = cb$ is well-defined and f is rational, f has a power series expansion around $z = 0$. It is equal to

$$f(z) = \sum_{k=0}^{\infty} cA^k bz^k = \sum_{k=0}^{\infty} h_{k+1}z^k, \;\; |z| < \rho,$$

where $\rho > 0$ is the radius of convergence of the power series. Note that because f is a rational function, the only singularities of f in $\mathbf{C}$ are the poles of f which are the reciprocals of the non-zero poles of the system. Of course on the disc $D(0, \rho) = \{z|z \in \mathbf{C}, |z| < \rho\}$ the function f is well-defined and has no singularities. Note that the coefficients h_{k+1} of the power series of f around $z = 0$ are non-negative. It now follows from the classical Pringsheim theorem (see e.g. [21], p.215) that if the radius of convergence ρ is finite, then $\rho \in \mathbf{R}_{>0}$ is itself a point of singularity of the function, hence a pole. Let $\lambda_M := \frac{1}{\rho}$ if ρ is finite and $\lambda_M := 0$ if $\rho = \infty$. It follows easily, both if $\rho < \infty$ and if $\rho = \infty$, that λ_M is a pole of the system, so $\lambda_M \in \sigma(A)$; and that $|\lambda| \leq |\lambda_M|$ for each pole λ of the system (so for each $\lambda \in \sigma(A)$.) $\square$

3 The continuous time case

Consider the class of functions $g : [0, \infty) \to \mathbf{R}$ that can be represented (or "realized") as $g(x) = c_1 e^{A_1 x} b_1$, where A_1 is a real $n_1 \times n_1$ matrix, b_1 an

$n_1 \times 1$ real column vector and c_1 an $1 \times n_1$ real row vector. As is well-known g is in this class if and only if it can be written as

$$g(x) = \mathrm{Re}(\sum_{k=1}^{K} p_k(x) e^{\lambda_k x})$$

where Re denotes the operator that gives the real part of a complex number, K is an integer, and p_k is a complex valued polynomial and λ_k a complex number, for each $k = 1, 2, \dots, K$. Note that this includes the trigonometric polynomials, as one can choose λ_k to be purely imaginary. (For instance, choosing $K = 1,\ p_1(x) \equiv 1,\ \lambda_1 = i$ one obtains $g(x) = cos(x)$). Obviously the class contains the class of polynomial functions (including the constant functions), and the class of real exponential functions (i.e. functions of the form e^{rx} for some real number r.) Furthermore this class of functions is a ring, as with any pair $g1, g2$ that it contains, it also contains the product $g_1 g_2$ as well as the sum $g_1 + g_2$. The functions from this class play an important role in the mathematical sciences. The class goes under a variety of different names in the literature, such as quasi-exponential functions, matrix exponential functions, EPT (exponential-polynomial-trigonometric) functions. In the systems and control literature they have been known as "Bohl functions" in the past but unfortunately without any apparent good reason and this terminology is now fading out. We think it would be appropriate to call this class the class of d'Alembert functions as the class emerged historically for the first time in the studies of d'Alembert as the class of solutions of linear differential equations with constant coefficients. Here we will use the more conventional name of EPT class. We have the following well-known characterization: the function $g : [0, \infty) \to \mathbf{R}$ is EPT if and only if any of the following conditions hold:

(i) The integral $\int_{\tau=0}^{\infty} g(\tau) e^{-s\tau} d\tau$ converges for all s in (at least) some non-empty open subset of the set of complex numbers and the ("continuous time") transfer function

$$G(s) := \int_{\tau=0}^{\infty} g(\tau) e^{-s\tau} d\tau$$

is a rational function $G(s) = \frac{p(s)}{q(s)}$ with $p,\ q$ polynomials, where we suppose p and q have been taken to be co-prime and q monic. This rational function will then necessarily be strictly proper, i.e. the degree of p is less than the degree n, say, of q.

(ii) There exists a minimal realization (A, b, c), where A is $n \times n$, b is $n \times 1$ and c is $1 \times n$, such that $g(s) = ce^{As}b$. (Just as in the discrete-time case if a complex minimal realization of order n exists, then also a real minimal realization of order n exists).

(iii) Let $h_k = g^{(k-1)}(0),\ k = 1, 2, \ldots$ denote the sequence of values at zero of the higher order derivatives of g and consider the corresponding $N \times N$ Hankel matrix $H = (H_{ij})_{i=1,2,\ldots,N, j=1,2,\ldots,N}$ with entries $H_{ij} = h_{i+j-1},\ i = 1, 2, \ldots, N, j = 1, 2, \ldots, N$. There exist a non-negative integer n such that the Hankel matrix H has rank n for all $N \geq n$.

As is well-known and easy to show the transfer function of $g(x) = ce^{Ax}b$ is equal to $G(s) = c(sI-A)^{-1}b$. As before, the poles of the transfer function are equal to the set of eigenvalues $\sigma(A)$ of A if (A, b, c) is a minimal realization. (cf. [11]). We can now formulate the second main theorem:

Theorem 3.1 *Consider an EPT function $g : [0, \infty) \to \mathbf{R}$ with a minimal realization (A, b, c) such that $g(x) = ce^{Ax}b$. Suppose that g is non-negative, i.e. $g(x) \geq 0$, for all $x \geq 0$, but that g is not identically zero. Then $\lambda_M := max_{\lambda \in \sigma(A)} Re(\lambda)$ is an element of the spectrum of A.*

Remark: Note that λ_M can be positive, zero or negative. Under the conditions stated, the theorem asserts that the spectrum of $A - \lambda_M I$ lies in the closed left half plane of $\mathbf{C}$ and contains 0.
Proof: Use will be made of the corresponding result in the discrete time case (see previous section). Consider the following integrals:

$$\int_{x=0}^{\infty} x^k e^{-rx} ce^{Ax} b dx = k! c[(rI - A)^{-1}]^{k+1} b$$

for $k = 0, 1, 2, \ldots$ and for $r \in \mathbf{R}$ larger than λ_M where λ_M (as defined in the theorem) is the smallest real number for which $\sigma(A - \lambda_M I)$ lies in the closed left half plane. Note that it follows immediately that if $g(x)$ is non-negative, then for each $r > \lambda_M$ the sequence $c[(rI - A)^{-1}]^k b,\ k = 1, 2, \ldots$ is non-negative (because the integrand is non-negative, and integrable, in that case). Let μ denote the largest real element in $\sigma(A)$. We want to show that μ coincides with λ_M. We will do that by refuting the possibility that $\mu < \lambda_M$. Therefore assume $\mu < \lambda_M$. It follows that there exists a complex number $\lambda_M + iy,\ y \neq 0$ in the spectrum $\sigma(A)$ of A. It is not hard to show that for all $r > r_0 := \frac{1}{2}(\lambda_M + \mu + \frac{y^2}{\lambda_M - \mu})$ one has $\frac{1}{r-\mu} < |\frac{1}{r-(\lambda_M+iy)}|$. Therefore for $r > r_0$ the discrete time system with (minimal!) realization $((rI - A)^{-1}, b, c)$ has non-negative impulse response sequence, however with largest real pole $\frac{1}{r-\mu}$ smaller than the modulus of the complex pole $\frac{1}{r-(\lambda_M+iy)}$. This contradicts theorem (2.1). As $\lambda_M \geq \mu$ by definition, it follows that equality $\lambda_M = \mu$ must hold. □

Remark: This theorem can also be derived from a general result in probability theory for probability density functions that have analytic characteristic functions, which can be found in [25].

4 Some consequences and generalizations

Here we list a number of consequences and generalizations of the main results.

(i) If one replaces the requirement that the impulse response be non-negative by the less stringent requirement that the impulse response function be non-negative eventually, then the conclusion remains unchanged. In the case of discrete time systems this follows from the fact that the Pringsheim theorem also holds if the coefficients in the power series involved are non-negative eventually, i.e. from a certain point in the coefficient sequence onwards. For the continuous time case one can argue as follows: if the EPT function satisfies $g(x) \geq 0$ for all $x \geq x_0$ for some $x_0 > 0$, then one can consider the shifted version $k(x) = g(x + x_0)$. Using a realization (A, b, c) of $g(x)$, it follows that $k(x) = ce^{Ax_0}e^{Ax}b$, so (A, b, ce^{Ax_0}) is a realization for h. Because e^{Ax_0} is a non-singular matrix (with inverse e^{-Ax_0}) (A, b, c) is minimal if and only if (A, b, ce^{Ax_0}) is minimal (the observability matrices of both realizations have the same rank, while the reachability matrices are identical). As h satisfies the conditions of the theorem in the previous section, it follows that $\sigma(A)$ must have a dominant real pole.

(ii) If one replaces, in the discrete time case, the requirement that the elements of the sequence $\{h_k\}$ have to be non-negative, by the requirement that they can be complex but have to have non-negative real parts, while for all non-zero elements, the absolute value of the complex arguments have to be uniformly bounded by a number θ less than $\frac{\pi}{2}$, (i.e. there exist real numbers $r_k,\ \nu_k$ such that $h_k = e^{r_k + i\nu_k}$, with $\nu_k < \theta$, for all k for which $h_k \neq 0$) then the conclusion is still that the system has a dominant real positive pole. Cf [24].

(iii) It may be of interest to note that using the theorem we can derive that any *non-negative matrix* must have a dominant real non-negative eigenvalue. This can be seen as follows. Let P denote an $n \times n$ matrix with only non-negative entries. Then the sequence $h_k = \mathrm{Tr}(P^{k-1}),\ k = 1, 2, \ldots$ is obviously non-negative. The transfer function of this sequence is

$$\sum_{k=1}^{\infty} h_k z^{-k} = \mathrm{Tr} \sum_{k=1}^{\infty} (Pz^{-1})^{k-1}.z^{-1} = \mathrm{Tr}(zI - P)^{-1} = \sum_{j=1}^{m} \frac{\mu_j}{z - \lambda_j},$$

where $\lambda_j,\ j = 1, 2, \ldots, m$ denote the distinct eigenvalues of P and $\mu_j \geq 1$ the algebraic multiplicity of eigenvalue $\lambda_j,\ j = 1, 2, \ldots, m$. So $\sum_{j=1}^{m} \mu_j = n$. It follows that the set of m poles of the transfer function of the sequence coincides with the set of m distinct eigenvalues of P.

Theorem (2.1) now implies that P must have a dominant real positive eigenvalue, which is a classical result of Perron-Frobenius. (Note that ideas similar to this remark can be found in [3]).

(iv) Theorem (2.1) can also be applied to the well-known GARCH models in time-series econometrics. An essential part of such models is described by a linear dynamical models which has non-negative inputs, namely the square of the previous innovations of the time series, and non-negative outputs, namely the variance of the present innovations of the time series (see e.g. [22]). Therefore such linear dynamical models which form the key to describing (univariate) GARCH models and which are in discrete-time, must have a non-negative impulse response sequence. This puts restrictions on the admissible parameter values in GARCH models and Theorem (2.1) implies that a GARCH model must have a dominant real non-negative pole! In the paper [1] the set of admissible parameter values for GARCH models is investigated and the results (cf. Theorems 1 and 2 in that paper) imply the existence of a dominant real non-negative pole (in our terminology–note that [1] works with "roots" which are the reciprocals of our "poles") in case the order of the system is 1 or 2.

(v) Obviously the necessary condition for non-negative impulse response functions of finite dimensional linear dynamical systems that is presented here is not (and far from) sufficient. In separate work we intend to present further investigations into necessary and into sufficient conditions. How to analyze non-negativity of such functions on a finite interval is treated in [23].

References

[1] D. Nelson, C.Q. Cao, *Inequality constraints in the univariate GARCH model*, Journal of Business and Economic Statistics, vol.10, no.2, April 1992, pp. 229-235.

[2] B.D.O. Anderson, M. Deistler, L. Farina, L. Benvenuto, *Nonnegative realizations of a linear system with nonnegative impulse response*, IEEE Transactions on Circuits and Systems, vol. 43, pp. 134-142, 1996.

[3] S. Friedland, *On an inverse problem for nonnegative and eventually nonnegative matrices*, Israel Journal of Mathematics, vol. 29, no.1, 1978.

[4] C.A. O'Cinneide, *Characterization of phase type distributions,* Stochastic Models, vol. 6, issue 1, 1990, pp. 1-57.

[5] R.S. Maier, C.A. O'Cinneide, *A closure characterization of phase-type distributions,* J. Appl. Prob. vol. 29, pp. 92-103, 1992.

[6] S. Asmussen, *Matrix-analytic models and their analysis,*, Scandinavian Journal of Statistics, vol. 27, issue 2, pp. 193-226.

[7] M. Neuts, *Matrix Geometric Solutions in Stochastic Models: An Algorithmic Approach,* Johns Hopkins Univ. Press, Baltimore, MD, 1981.

[8] L. Farina, *Positive systems in the state space approach: main issues and recent results,* Proc MTNS 2002, Notre Dame, 2002;

`http://www.nd.ed/~mtns/papers/14900_1.pdf`

[9] M.W. Fackrell, *Characterization of matrix-exponential distributions*, Ph.D. thesis, School of Applied Mathematics, Univ Adelaide, 2003 (available on-line).

[10] P.C. Bell, J-C. Delrenne, R.M. Jungers, V.D. Blondel, *The Continuous Skolem-Pisot Problem: On the Complexity of Reachability for Linear Ordinary Differential Equations*, arXiv: 0809.2189v2, 23 April 2009.

[11] T. Kailath, *Linear Systems,* Prentice Hall, Englewood Cliffs, 1980.

[12] D. Filipovic, *A note on the Nelson-Siegel family,* Mathematical Finance, vol. 9, no. 4, pp. 349-359.

[13] D. Filipovic, *Exponential-polynomial families and the term structure of interest rates,* Bernoulli, vol. 6, no.6, pp. 1081-1107, 2000.

[14] E. Sharef, D. Filipovic, *Conditions for consistent exponential-polynomial forward rate processes with multiple non-trivial factors,* International Journal of Theoretical and Applied Finance, vol. 7, nr. 6, pp. 685-700, 2004.

[15] T. Bjoerk, B.J. Christensen, *Interest rate dynamics and consistent forward rate curves,* Mathematical Finance, vol. 9, nr. 4, pp. 323-348, 1999.

[16] T. Bjoerk, A. Gombani, *Minimal realizations of interest rate models,* Finance and Stochastics, vol. 3, nr. 4, pp. 413-432, 1999.

[17] C.R. Nelson, A.F. Siegel, *Parsimonious modeling of yield curves,* Journal of Business, vol. 60, nr. 4, pp. 473-489, 1987.

[18] L. Svensson, *Estimating and interpreting forward interest rates: Sweden 1992-4,* Discussion paper, Centre for Economic Policy Research (1051), 1994.

[19] X. Zhang, K. De Cock, M.F. Bugallo, P.M. Djuric, *A general method for the computation of probabilities in systems of first-order chemical reactions,* Journal of Chemical Physics, vol. 122, nr. 10, 2005.

[20] B. Hanzon, R.J. Ober, *State space calculations for discrete probability densities,* Linear Algebra and its Applications, vol. 350, pp. 67-87, 2002.

[21] E.C. Titchmarsh, *The Theory of Functions, 2nd ed,* Oxford University Press, 1939.

[22] C. Gourieroux, *ARCH Models and Financial Applications*, Springer, 1997.

[23] B. Hanzon, F. Holland, *Non-negativity of exponential polynomials and EPT functions (Extended summary)*, accepted for presentation at the Bachelier Finance World Congress, Toronto, July 2010.

[24] B. Hanzon, F. Holland, *The long-term behaviour of Markov sequences*, to be submitted.

[25] E. Lukacs, O. Szasz, *On analytic characteristic functions,* Pacific J. Math., vol. II, no. 4, pp. 615-625, 1952.