BFS Toronto Poster 429: Non-negativity of Exponential Polynomial Trigonometric Functions–a Budan-Fourier sequence approach

Bernard Hanzon and Finbarr Holland
Edgeworth Centre for Financial Mathematics [1] and Department of Mathematics, University College Cork; e-mail: b.hanzon@ucc.ie and f.holland@ucc.ie

Introduction

Consider the class of real EPT functions which contains the real polynomials, the real exponential functions and the "scaled" real trigonometric polynomials (such as $\sin(\nu t), \cos(\nu t), \nu \in \mathbf{R}$ but not $\tan(\nu t)$, ..), and products and sums of such functions (the set of functions is a ring over $\mathbf{R}$).
Such functions appear virtually everywhere in mathematics! To give some examples:
(i) In *financial mathematics* they appear as *forward rate curves*, e.g. the Nelson-Siegel forward rate curves.As these are interest rates we want these to be non-negative!
(ii) In *probability theory* the EPT functions appear as *probability density functions*, in the form of Gamma densities with positive integer shape parameter k for instance.Such density functions must be non-negative and integrable (to one).
(iii) In *systems theory* they appear as impulse response functions of linear systems. In case of *positive systems* the impulse response functions will be positive, and for non-negative systems, non-negative.
There are many more applications. Here we ask ourselves: *How can we analyze (and as a result: guarantee) non-negativity of such a function?*
Here we will consider the question of how to determine the non-negativity of a real EPT function on a given finite closed interval $[0, T]$, $T \in \mathbf{R}$. The question of the tail behavior of general non-negative EPT functions will be treated elsewhere.
As an application of our results we will consider the class of Nelson-Siegel curves as a special case.

Section 1: A Matrix-Vector Representation of EPT Functions

Consider a linear differential equation with real constant coefficients:

$$y^{(n)}(t) + q_1 y^{(n-1)}(t) + q_2 y^{(n-2)}(t) + \ldots + q_n y(t) = 0,$$

The initial conditions are given by:

$$y(0) = b_1, y^{(1)}(0) = b_2, \ldots, y^{(n-1)}(0) = b_n$$

where b_k, $k = 1, 2, \ldots, n$ are real numbers.
The solutions, in general, where described by d'Alembert (in the 18th century).
One (modern) way to obtain these solutions is as follows. Let

$$x_1(t) := y(t), x_2(t) := y^{(1)}(t), x_3(t) := y^{(2)}(t), \ldots, x_n(t) := y^{(n-1)}(t), \quad t \in [0, \infty)$$
$$x_1(0) = y(0) = b_1, x_2(0) = y^{(1)}(0) = b_2, \ldots, x_n(0) = y^{(n-1)}(0) = b_n,$$

and let $x(t) := (x_1(t), x_2(t), \ldots, x_n(t))'$ and $b := (b_1, b_2, \ldots, b_n)'$.
Then our differential equation is:

$$\dot{x}(t) = Ax(t),$$
$$y(t) = x_1(t) = (1, 0, \ldots, 0)x(t);$$
$$A = \begin{bmatrix} 0 & 1 & 0 & \ldots \ldots & 0 \\ 0 & 0 & 1 & 0 \ \ldots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \ldots & \ldots \ldots & 0 & 1 \\ -q_n & -q_{n-1} & \ldots \ldots & \ldots & -q_1 \end{bmatrix},$$

hence A is a so-called *companion matrix*.
The solution can now be obtained as:

$$y(t) = (1, 0, 0, \ldots, 0)e^{At}b = ce^{At}b,$$

where $c = (1, 0, \ldots, 0)$ and $e^{At} := \sum_{k=0}^{\infty} \frac{1}{k!} A^k t^k$
In this approach $x(t)$ is called the state vector.
It is straightforward to show that if T is an $n \times n$ non-singular matrix and

$$\tilde{A} := TAT^{-1}, \ \tilde{b} := Tb, \ \tilde{c} := cT^{-1},$$

then

$$y(t) = \tilde{c}e^{\tilde{A}t}\tilde{b}$$

Actually for any triple (A, b, c), $A : n \times n, b : n \times 1, c : 1 \times n$, $ce^{At}b$ is what we could call a *"d'Alembert function"*: it satisfies a homogeneous linear differential equation with constant coefficients. (This can be seen by application of the theorem of Cayley-Hamilton).
We can always replace any triple $(\hat{A}, \hat{b}, \hat{c})$, $\hat{A} : \hat{n} \times \hat{n}$, $\hat{b} : \hat{n} \times 1$, $\hat{c} : 1 \times \hat{n}$ by a minimal triple (this follows from the so-called *Kalman decomposition*) (A, b, c), $A : n \times n, b : n \times 1, c : 1 \times n$, $n \leq \hat{n}$ such that

$ce^{At}b = \hat{c}e^{\hat{A}t}\hat{b}$, and rank $[b|Ab|A^2b|\ldots|A^nb] = n$, hence (A, b, c) is *reachable*; and rank $\begin{bmatrix} c \\ cA \\ \vdots \\ cA^n \end{bmatrix} = n$

hence (A, b, c) is *observable*.
Consider such a "d'Alembert" function $y(t) = ce^{At}b$, (A, b, c) minimal. Let $\sigma(A)$ denote the set of eigenvalues of A. We will distinguish three cases:
(1) The function $y(t)$ is polynomial. This is the case if and only if $\sigma(A) = \{0\}$. Then A is *nilpotent*, hence $A^n = 0$, and
$e^{At} = I + At + \frac{1}{2}A^2t^2 + \ldots + \frac{1}{(n-1)!}A^{n-1}t^{n-1}$, hence
$y(t) = ce^{At}b = cb + cAbt + \frac{1}{2}cA^2bt^2 + \ldots + \frac{1}{(n-1)!}cA^{n-1}bt^{n-1}$ is polynomial indeed.
(2) If $\sigma(A) \subset \mathbf{R}$, real, then (e.g. by bringing A in Jordan canonical form) one can show
$y(t) = ce^{At}b = \sum_{k=1}^{K} p_k(t)e^{\lambda_k t}$,
where K is the number of distinct (real) eigenvalues of A, and for each $k = 1, 2, \ldots, K$, $p_k(t) \in \mathbf{R}[t]$ is a *real* polynomial and $\lambda_k \in \mathbf{R}$ a real eigenvalue of A, $k = 1, 2, \ldots, K$. Such functions will be called real *exponential polynomials* ("EP" class). The subset of functions for which $\deg(p_k) = 0$, $k = 1, 2, \ldots, K$, will be called the real *exponential sums* class ("E").
(3) If $\sigma(A) \subset \mathbf{C}$ then one can show that there exist disjoint subsets I_1, I_2, I_3 such that
$I_1 \smile I_2 \smile I_3 = \{1, 2, \ldots, K\}$, and such that
$y(t) = ce^{At}b = \sum_{k \in I_1} q_k(t)e^{\lambda_k t} + \sum_{k \in I_2} r_k(t)e^{\theta_k t}\cos(\nu_k t) + \sum_{k \in I_3} s_k(t)e^{\theta_k t}\sin(\nu_k t)$,
where K is the number of distinct eigenvalues ($\in \mathbf{C}$) of A, and
$q_k(t) \in \mathbf{R}[t]$, $k \in I_1$; $r_k(t) \in \mathbf{R}[t]$, $k \in I_2$; $s_k(t) \in \mathbf{R}[t]$, $k \in I_3$ are *real* polynomials and $\lambda_k \in \mathbf{R}$, $k \in I_1$ are the distinct real eigenvalues of A, $k \in I_1$, while $\{\theta_k \pm i\nu_k, \ k \in I_2\} = \{\theta_k \pm i\nu_k, \ k \in I_3\}$ are the distinct complex, non-real eigenvalues of A. A function of this form will be called a real exponential-polynomial-trigonometric function or, in short, a function of the "EPT" class. This coincides with the "d'Alembert" class.
These classes go under *many* different names in the literature such as quasi-exponential, exponential-polynomial, matrix exponential; in control theory for a long time these functions were called Bohl functions, however as the motivation for this terminology is unclear this usage is now apparently fading out.

Section 2: Finding the Minimum of a Polynomial using a Budan-Fourier Sequence

Now consider the problem of determining the minimum of such an EPT function on a finite closed interval $[0, T] \subset \mathbf{R}$.
When the EPT function is identical to the zero function, the minimum is zero and the problem is trivial. Therefore from now on let us consider EPT functions that are not identically equal to zero. The minimum is attained at a point where the derivative $cAe^{At}b$ is zero or at one of the boundary points $0, T$. Actually zeros that correspond to a minimum are "sign-changing zeros", zeros where the sign of the function changes from $+$ to $-$ or vice versa. (Any real, not-identically-zero EPT function has *isolated* zeros, because it is real analytic).
Can we construct an algorithm that produces all sign-changing zeros of an EPT function on a given finite interval?
For the sub-class of polynomials the answer is "Yes"!
This can be shown using a so-called "Fourier" or "Budan-Fourier" sequence, which in this case is just the sequence of higher order derivatives:

$$p(t) = c.e^{At}.b, \ \sigma(A) = \{0\}$$
$$p^{(1)}(t) = c.Ae^{At}.b,$$
$$p^{(2)}(t) = c.A^2.e^{At}.b,$$
$$\vdots$$
$$p^{(n)}(t) = c.A^n.e^{At}.b \equiv 0, (A^n = 0)$$

Let k be the largest integer value for which $p^{(k+1)} \equiv 0$ then $p^{(k)}$ is a non-zero constant and $p^{(k-1)}$ is either strictly increasing (if $p^{(k)}(t) > 0$) or strictly decreasing (if $p^{(k)}(t) < 0$).
If $p^{(k-2)}(0)$ and $p^{(k-2)}(T)$ have the same (non-zero) sign then $p^{(k-2)}(t) \neq 0, \ \forall t \in [0, T]$.
If $p^{(k-2)}(0)$ and $p^{(k-2)}(T)$ have *opposite* (non-zero) signs then there is exactly one sign-changing zero of $p^{(k-2)}(t)$ on $[0, T]$. This zero can be calculated using *bisection*, with *arbitrary precision*!
One way of viewing this is as follows:

Definition

An open interval $(a, b) \subset \mathbf{R}$ will be called *simple* for the function $f : D \to \mathbf{R}, D \subseteq \mathbf{R}_{\geq 0}$, f continuous, if $(a, b) \subseteq D$ and f has at most one sign-changing zero on (a, b).

Remark. Suppose f has simple interval (a, b). For any number $x \in \mathbf{R}$, let Sign$(x) = 1, 0$ or -1 depending on whether $x > 0$, $x = 0$ or $x < 0$ respectively. Then if

$$\lim_{\epsilon \downarrow 0} \text{Sign}\,[f(a + \epsilon)] = \lim_{\epsilon \downarrow 0} \text{Sign}\,[f(b - \epsilon)] \neq 0 \tag{1}$$

then there is *no* sign-changing zero of f in (a, b). If $\lim_{\epsilon \downarrow 0} \text{Sign}\,[f(a + \epsilon)] \times \lim_{\epsilon \downarrow 0} \text{Sign}\,[f(b - \epsilon)] < 0$, then a bisection algorithm gives the unique sign-changing zero $c \in (a, b)$ such that $f(c) = 0$. (Of course the case $f \equiv 0$ on (a, b) can be handled in a straightforward manner.)

Definition

A *grid* $\{a_0, a_1, \ldots, a_N\}$, $a_0 = 0, a_N = T$, $a_0 < a_1 < \ldots < a_N$, is called *simple* for f if each interval (a_{k-1}, a_k), $k = 1, 2, \ldots, N$ is simple for f.

Remark. Given a simple grid the *sign-changing zeros* of f on $[0, T]$ can be found *all* and with *arbitrary precision*, using bisection.
Now note that $\forall k \in \{1, 2, \ldots, n\}$ the sign-changing zeros of $p^{(k)}(t)$, together with the boundary points $0, T$, form a simple grid for $p^{(k-1)}(t)$, where $p^{(0)}(t) \equiv p(t)$ is the original polynomial. Any finite sequence with this property, for which the last element is the zero function, will be called a generalized Budan-Fourier sequence. Therefore the generalized Budan-Fourier sequence gives a guaranteed method to find all the real zeros of the original function p.

Section 3: Simple Grid Properties and a B-F Sequence for EP Functions

(i) If $\{a_0, a_1, \ldots, a_N\}$ is a simple grid for f, then also for $f.g$, where $g(t) \neq 0, \forall t \in (a_0, a_N)$ and g continuous on (a_0, a_N).
(ii) If $\{a_0, a_1, \ldots, a_N\}$ consists of the boundary points together with the sign-changing zeros of a function h, h continuous on (a_0, a_N); then the same grid is obtained if h is replaced by $h.k$ for any continuous function k with $k(t) \neq 0$, $\forall t \in (a_0, a_N)$.
Let $h(t) = ce^{At}b, \sigma(A) \subset \mathbf{R}$, (A, b, c) minimal and real. Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ denote the (real) eigenvalues of $A : n \times n$. Let

$$h^{(0)}(t) := h(t) = c.e^{At}.b$$
$$h^{(1)}(t) := c(A - \lambda_1.I)e^{At}b$$
$$\vdots$$
$$h^{(k)} := c(A - \lambda_k.I)\ldots(A - \lambda_1.I).e^{At}.b$$
$$\vdots$$
$$h^{(n)}(t) := c(A - \lambda_n.I)\ldots(A - \lambda_1.I)e^{At}b \equiv 0 \text{ (Cayley-Hamilton!)}$$

Section 3 continued

Theorem

This is a generalized Budan-Fourier sequence, for any pair of boundary points $\{0, T\}$, $T > 0$.

Proof. Note that for every $k \in \{0, 1, 2, \ldots, n-1\}$, $h^{(k+1)}$ is obtained from $h^{(k)}$ by applying the formula:

$$h^{(k+1)} = e^{\lambda_{k+1}.t}\frac{d}{dt}\left[e^{-\lambda_{k+1}.t}.h^{(k)}(t)\right] \tag{2}$$

hence

$$\frac{d}{dt}\left[e^{-\lambda_{k+1}.t}.h^{(k)}(t)\right] = e^{-\lambda_{k+1}.t}.h^{(k+1)}(t) \tag{3}$$

Using (i),(ii) and the fact that $e^{-\lambda_{k+1}.t} > 0$, $\forall t \in [0, \infty)$, it follows that the sign-changing zeros of $h^{(k+1)}(t)$ (which are also the sign-changing zeros of $e^{-\lambda_{k+1}.t}h^{(k+1)}(t)$), together with the boundary points $\{0, T\}$, form a simple grid for $e^{-\lambda_{k+1}.t}h^{(k)}(t)$, hence for $h^{(k)}(t)$.
As $h^{(n)} \equiv 0$ because $(A - \lambda_n I)(A - \lambda_{n-1} I)\ldots(A - \lambda_1 I) = 0$, it follows that $h^{(n-1)}$ has constant sign (no zeros). □

Section 4: Generalized B-F Sequence for an EPT Function

Now consider a general EPT function $h(t) = c.e^{At}.b$, where $\sigma(A)$ can contain complex elements, but $h(t) \in \mathbf{R}$, $\forall t \in [0, \infty)$. Assuming (A, b, c) minimal, it follows that if $\lambda \in \sigma(A)$ with multiplicity $m(\lambda)$, then $\bar{\lambda} \in \sigma(A)$, with the same multiplicity $m(\hat{\lambda}) = m(\lambda)$. Let us order the eigenvalues (with their algebraic multiplicities) $\lambda_1, \lambda_2, \ldots, \lambda_N$ such that the complex conjugates come in pairs
$\lambda_k = \theta_k + i\nu_k, \lambda_{k+1} = \theta_k - i\nu_k$, with $\nu_k > 0$.

Theorem

A generalized Budan-Fourier sequence with boundary points, for $h(t) = ce^{At}b$, (A, b, c) minimal and real, is obtained as follows:
$h^{(0)}(t) = h(t)$, *with boundary points* $\{0, T\}$;
if $\lambda_k \in \mathbf{R}$:
$h^{(k)} = c(A - \lambda_k I)(A - \lambda_{k-1} I)\ldots(A - \lambda_1 I).e^{At}.b$ *with boundary points* $\{0, T\}$;
if $\lambda_k = \theta_k + i\nu_k, \lambda_{k+1} = \theta_k - i\nu_k$, $\nu_k > 0$ *is a complex conjugate pair then*
$h^{(k)} := Im\left[e^{-\bar{\lambda}_k t}c(A - \lambda_k I)\ldots(A - \lambda_1 I)e^{At}b\right]$ *with extended set of boundary points*
$\{0, \frac{\pi}{\nu_k}, 2\frac{\pi}{\nu_k}, \ldots, \lfloor\frac{T}{\pi/\nu_k}\rfloor\frac{\pi}{\nu_k}, T\}$; *and* $h^{(k+1)} := c(A - \lambda_{k+1} I)(A - \lambda_k I)\ldots(A - \lambda_1 I)e^{At}b$, *with (same) extended set of boundary points* $\{0, \frac{\pi}{\nu_k}, 2\frac{\pi}{\nu_k}, \ldots, \lfloor\frac{T}{\pi/\nu_k}\rfloor\frac{\pi}{\nu_k}, T\}$

Note that $(A - \lambda_{k+1} I)(A - \lambda_k I) = (A - \bar{\lambda}_k)(A - \lambda_k)$ is a real matrix. Notation for the *entier* of a real number x is: $\lfloor x \rfloor = \max\{n \in \mathbf{N}; n \leq x\}$.

Application to Nelson-Siegel forward rate curves

Theorem

A Nelson-Siegel forward rate curve $f_{NS} = z_1 + z_2e^{-\lambda x} + z_3xe^{-\lambda x}$ with $z_1 > 0, \lambda > 0$ is non-negative for all $x \geq 0$ iff $(z_2, z_3) \in A \cup B$ where $A = \{(z_2, z_3)|z_2 \geq -z_1, z_3 \geq -\lambda z_1\}$ and $B = \{(z_2, z_3)|z_3 \leq -\lambda z_1, z_2 \geq \frac{z_3}{\lambda}(1 - \log(\frac{|z_3|}{\lambda z_1}))\}$

Proof (sketch) Consider first the special case $NS(t) = 1 - (b + at)e^{-t}$ then a B-F sequence is $NS_0(t) = NS(t), NS_1(t), NS_2(t)$, with $NS_1(t) = 1 - ae^{-t}, NS_2(t) = 1$. So NS is non-negative if it is non-negative at the boundary points and at all the sign-changing zeros of NS_1. But NS_1 only has sign-changing zeros if its values at $t = 0$ and $t = \infty$ have opposite signs (because NS_2 is constant). Now $NS_1(0) = 1 - a$; $NS_1(\infty) = 1$, so we can distinguish:
Case A: $1 - a \geq 0$:"same signs", NS_1 has no zeros; the non-negativity requirement on NS is now that $0 \leq NS(0) = 1 - b$.
Case B: $1 - a < 0$:"opposite signs": NS_1 has a sign-changing zero; it is actually attained at $t = \log(a) > 0$. Non-negativity of NS now requires that $0 \leq NS(\log(a))$ and $0 \leq NS(0) = 1 - b$. These two inequalities can be combined into the inequality $b \leq a(1 - \log(a))$ while $a > 1$. The general case can easily be deduced from this. □
This result could be used for instance when forward rate curves are approximated by Nelson-Siegel curves, such as in Devin c.s.(2010)

References

S. Devin, B. Hanzon, Th. Ribarits, *A Finite Dimensional HJM Model: How Important is Arbitrage-Free Evolution?*, presentation 33 in this conference, and submitted to the International Journal of Theoretical and Applied Finance.

[1] B. Hanzon gratefully acknowledges support by the Science Foundation Ireland Grant 07MI008 associated with the Edgeworth Centre for Financial Mathematics