AIMS Mathematics, 9(10): 26499–26537.
DOI: 10.3934/math.20241290
Received: 12 April 2024
Revised: 04 September 2024
Accepted: 05 September 2024
Published: 12 September 2024

https://www.aimspress.com/journal/Math

*Research article*

# A generalized Budan-Fourier approach to generalized Gaussian and exponential mixtures

**Stefano Bonaccorsi**[1,*], **Bernard Hanzon**[2] **and Giulia Lombardi**[1]

[1] Department of Mathematics, Università degli Studi di Trento, via Sommarive 14, 38123 Trento, Italy

[2] School of Mathematical Sciences, University College Cork, Ireland

* **Correspondence:** Email: stefano.bonaccorsi@unitn.it.

**Abstract:** In the literature, finite mixture models were described as linear combinations of probability distribution functions having the form $f(x) = \Lambda \sum_{i=1}^{n} w_i f_i(x)$, $x \in \mathbb{R}$, where $w_i$ were positive weights, $\Lambda$ was a suitable normalising constant, and $f_i(x)$ were given probability density functions. The fact that $f(x)$ is a probability density function followed naturally in this setting. Our question was: *if we removed the sign condition on the coefficients $w_i$, how could we ensure that the resulting function was a probability density function?*

The solution that we proposed employed an algorithm which allowed us to determine all zero-crossings of the function $f(x)$. Consequently, we determined, for any specified set of weights, whether the resulting function possesses *no* such zero-crossings, thus confirming its status as a probability density function.

In this paper, we constructed such an algorithm which was based on the definition of a suitable sequence of functions and that we called a *generalized Budan-Fourier sequence*; furthermore, we offered theoretical insights into the functioning of the algorithm and illustrated its efficacy through various examples and applications. Special emphasis was placed on generalized Gaussian mixture densities.

**Keywords:** finite mixtures; Gaussian mixtures; Exponential-Polynomial-Trigonometric probability density functions

**Mathematics Subject Classification:** 60E05

## 1. Introduction

The first instance of a finite mixture probabilistic model goes back to the work of Pearson [27], where a two-component normal mixture was used to fit a dataset of the body lengths of crabs from the Bay of Naples. Since then, such models have been applied to various problems in statistical descriptions; for a survey of these applications, we refer, for instance, to [14, 24, 34]. The analysis of neural networks with response functions of this form-without constraints on the sign of the coefficients-goes back at least to [26] where finite Gaussian mixtures are studied. Recent applications to cluster analysis [13] and unsupervised machine learning problems [3, 4] are also worth mentioning. A recent survey of mixture-type filters and fusion approaches in the context of Bayesian filtering can be found in [23]. Although this paper focuses solely on univariate distributions, the applications of finite mixture models naturally extend to the multivariate case.

We consider a broader definition of finite mixture models beyond Gaussian mixtures, encompassing functions of the form

$$f(x) = \mathfrak{R}\left(\sum_{i=1}^{n} p_i(x)e^{q_i(x)}\right), \qquad x \in I, \tag{1}$$

where $\mathfrak{R}(z)$ denotes the real part of the complex number $z$, and $\{p_i(x),\ i = 1, \ldots, n\}$ and $\{q_i(x),\ i = 1, \ldots, n\}$ are given complex-valued polynomials on an interval $I \subset \mathbb{R}$ which may be finite, semi-infinite (such as $I = \mathbb{R}_+ := \{x \geq 0\}$), or infinite ($I = \mathbb{R}$). For example, consider a mixture of Erlang distributions $f_i(x) = x^{m_i-1}e^{-\lambda_i x}$, $x \in [0, \infty)$, which leads to the distribution

$$f(x) = \Lambda \sum_{i=1}^{n} w_i x^{m_i-1}e^{-\lambda_i x}, \qquad x \in [0, \infty),\ \lambda_i > 0, i = 1, \ldots, n \tag{2}$$

for a suitable normalizing constant $\Lambda$, where either $\lambda_{i+1} > \lambda_i$, or $\lambda_{i+1} = \lambda_i$ and $m_{i+1} > m_i$. If the coefficients $w_i$ are positive, then $f$ is a probability density function (pdf).

Suppose that the Erlang mixture in (2) has the additional property that the scale parameters are the same for all the terms: $\lambda_i = \lambda$. Then, the problem of proving that $f$ is a pdf reduces to the study of the positivity of the polynomial $p(x) = \sum_{i=1}^{n} w_i x^{m_i-1}$. This is a classical problem in the mathematical literature and can be approached in several ways.

For polynomials, there are at least *three* possible approaches. The first one is to use the fact that a nonnegative polynomial $q$ on the real line can be written as a product $q = \phi\phi^*$, where $\phi$ is a *complex-valued* polynomial. The technique used to get $\phi$ from $q$ is called spectral factorization. It could be used in the present example by taking $q(z) = p(z^2)$, accounting for the fact that $p$ is defined on the non-negative half-line only. More generally, spectral factorization and related techniques can be applied to ascertain the nonnegativity of generalized mixtures of rational densities such as Cauchy densities and Student-$t$ densities with odd degrees of freedom; see, e.g., [18]. The second method is to use the well-known Sturm chain method to determine the number of zeros of a polynomial on a given interval. This method applies the Euclidean algorithm for polynomials in various ways. The third approach is the so-called Budan-Fourier (BF) approach, which does not require the application of the Euclidean algorithm (nor spectral factorization). For the first two methods, there do *not* seem to be useful generalizations to the classes of functions that we are interested in here.

Therefore, we focus here on the BF approach for polynomials introduced in [16]; see also [8, 15]. The idea underlying this approach is that there exists a stepwise constant, and nonincreasing function $V_p(x)$, taking integer values, with $V_p(-\infty) = \kappa := \deg(p)$ and $V_p(\infty) = 0$, such that the jump points of $V_p(x)$ correspond to the *virtual roots* (which include the real roots of $p$) of $p$. The *BF theorem* states that the number of real roots of $p$ in the interval $(a, b]$ (counted with multiplicity) is at most $V_p(b) - V_p(a)$, and the defect is an even integer.

The BF approach is based on a finite sequence of functions, the BF sequence.

In the original BF approach to polynomials, this sequence is the sequence of higher-order derivatives; in [8], $\boldsymbol{f}$-derivatives are used, extending the range of application of the result. It is important to note that the BF sequence can not only be used to determine an upper bound for the number of real roots in an interval but, by repeatedly using a bisection method, can actually be used to find *all* the sign-changing zeros (i.e., zero crossings) of a polynomial on a given interval $(a, b]$. However, in general, if a function is not a polynomial, the sequence of nontrivial higher-order derivatives is not finite, and the method cannot be applied. In our approach, the sequence of higher-order derivatives is replaced by a finite sequence that we will call a *generalized sequence* (GBF sequence).

In Section 2.1 we present in detail the functioning of the GBF *algorithm* once a GBF *sequence* is available. This method was initially proposed in [17], where a GBF sequence for exponential-polynomial-trigonometric (EPT) functions was first constructed. The purpose of the GBF algorithm is to systematically compute *all* the sign-changing roots of a given function within a finite interval.

In Section 2.3 we present in detail a methodology for constructing a GBF sequence for the linear span of a finite collection of functions, and we explain the conditions under which the methodology is successful. We provide an application to EPT functions in Section 3. Next, Section 4 and Section 5 are devoted to the analysis of Gaussian mixtures. There, we construct a GBF sequence for the general framework of Gaussian mixtures with polynomial coefficients for the first time, $f(x) = \sum_{j=1}^{n} p_j(x)\, e^{q_j(x)}$, where $\{q_j(x)\}$ are second-order polynomials with real coefficients and negative leading term and $\{p_j(x)\}$ are a priori arbitrary polynomials. The GBF algorithm can check whether such a function $f$ has no zero-crossings and, hence, constitutes a pdf after scaling to ensure that the integral over the real line is one. We provide a simple, yet completely explicit, example of a polynomial Gaussian mixture depending on a parameter, and show that the behavior of the mixture may vary significantly depending on the values of this parameter.

Next, Section 5 is devoted to the special case of finite Gaussian mixtures $f(x) = \sum_{j=1}^{n} \gamma_j\, e^{q_j(x)}$. These particular densities, where the weights $\gamma_j$ are usually taken to be nonnegative, have become very popular in several fields (such as speech recognition and image analysis [31, 38]), primarily because, in many cases, data is multimodal, i.e., the underlying population is already a combination of different subpopulations. In such applications, it makes intuitive sense to model multimodal data as a mixture of unimodal Gaussian distributions to best approximate the original distribution. There is interest in generalizing the class by allowing some of the weights to be negative while keeping the overall density nonnegative (see, e.g., [19, 22, 25, 39]). Modeling with finite Gaussian mixtures retains many of the theoretical and computational benefits of Gaussian models, making them practical for efficiently modeling very large datasets. For example, the sum of a random variable with a Gaussian mixture distribution and an independent Gaussian random variable is again distributed as a Gaussian mixture, with a natural shift in the parameters.

In addition to the properties mentioned above, there is an approximation theorem [26] which proves that finite Gaussian mixtures can approximate "arbitrary" probability distribution functions well, at least for large enough $N$, where $N$ is the size of the mixture, and a good spread of means and variances, once one allows for coefficients of arbitrary sign. To summarize, although finite Gaussian mixtures are commonly represented in the literature as a linear combination of Gaussian pdfs with positive weights, where this assumption automatically ensures that the resulting Gaussian mixture does not have any sign-changing zeros along the $x$-axis, to take full advantage of this class of distributions, it seems necessary to extend the study to mixtures with coefficients of arbitrary sign. A stumbling block to their use may have been the need to certify the nonnegativity of the resulting density functions. The main aim of this paper is to provide an algorithm that can deliver that certification by determining the number of zero-crossings of the density function: if there are none and the function is normalized to integrate to one, then a probability function is obtained.

In Section 5, we also present a software tool [5], designed to work in Matlab, that allows the investigation of the number of zero-crossings in a finite Gaussian mixture. In particular, given a finite Gaussian mixture (specified by the means, variances, and mixing coefficients), the software returns a maximal interval where the sign-changing roots may appear—an interval that is always bounded. Further, if these roots exist, the software identifies them within a desired accuracy level. As a result, the algorithm will enable us to identify those cases in which the mixture is nonnegative on the real line.

One of our motivations for developing a unified procedure for finding sign-changing zeros in generalized mixtures of densities stems from the potential limitations of standard root-finder algorithms in locating the zeros of such functions. Although the numerical search for zeros of arbitrary functions is a mature field, it appears that the approximations on which many algorithms are based often fail to provide complete precision, primarily but not limited to the tails of the function. In appendix B.3, we shall compare our method with standard numerical tools available in the literature, particularly with the Chebfun software in Matlab. As we shall see in the examples in appendix B.3, there are cases where the use of Chebfun does not correctly identify mixtures as pdfs, usually due to the presence of several spurious zeros in the tails of the distribution.

Finally, Section 6 presents an application of our results to the study of the Wasserstein-1 distance between probability mixtures. In particular, we show that knowledge of the zeros of the difference between the cumulative probability distribution functions (CDFs), obtained using GBF sequences, is a necessary tool in computing the distance. To produce the relevant GBF sequence, we discuss separately the cases of EPT functions and Gaussian mixtures.

## 2. Presentation of the basic building blocks of the approach

For any sufficiently often differentiable function $f : \mathbb{R} \to \mathbb{R}$ we denote by $Df(x)$ the first order derivative in $x$, and generically $D$ the derivative operator; similarly, $D^j f(x)$ is the $j$-th order derivative, for $j \in \mathbb{N}$ and $D^j$ is the correspondent operator.
For any (real- or complex-valued) square matrix $A$, $|A|$ is the determinant of $A$.

### 2.1. Generalized BF algorithm

Let $I$ be a given interval and $f \not\equiv 0$ a function on $I$. Suppose that $f$ is real analytic and, hence, the set of zeros of $f$ is discrete and has finite cardinality on any closed and bounded subinterval $[a, b]$ of

$I$. We are interested in the *sign changing zeros* of $f$, i.e., points $x_0$ such that $f(x_0) = 0$ and for which there exists an open neighborhood $(x_0 - \epsilon, x_0 + \epsilon) \subset I, \epsilon > 0$ such that for all $y, z$ in this neighborhood with $y < x_0 < z$, it holds that $f(y) \cdot f(z) < 0$. Here, we will present a method to determine all sign-changing zeros on a given closed and bounded interval $[a, b]$. We shall denote by $R(f, [a, b])$ the set of sign-changing zeros of $f$ on the interval $[a, b]$. Note that for any given point, so also for the points $a$ and $b$, we can determine whether it is a sign-changing zero of $f$ by inspection. Therefore, we can focus on determining the sign-changing zero in the open interval $(a, b)$. An open interval is called a *simple interval* for a function $f$ if it contains at most one sign-changing root of $f$ (possibly of multiplicity greater than 1). A *simple grid* for $f$ on the interval $[a, b]$ is a finite sequence of points $G = \{x_1 < \cdots < x_n\} \subset I$ such that $(a, x_1), (x_n, b)$, and $(x_i, x_{i+1}), i = 1, 2, \ldots, n-1$ are all simple intervals for $f$.

**Remark 2.1.** *The set of sign-changing zeros for the first derivative, $R(f', [a, b])$, is a simple grid for the function $f$.*

**Definition 2.2.** *Let $f$ be a real analytic function on the interval $[a, b]$. A sequence of functions $\{\psi_i,\ i = 0, \ldots, n\}$ is called a* GBF sequence associated to *$f$ and a sequence of functions $\{\varrho_i,\ i = 1, \ldots, n\}$ is called the* associated sequence of pivots *if the following properties are satisfied:*

1. *$\psi_0 = f$;*
2. *$\psi_i$ and $\varrho_i$ are quotients of real analytic functions on $[a, b]$ for each $i = 1, \ldots, n$;*
3. *$\psi_n$ is the zero function;*
4. *For each $k = 1, 2, \ldots, n$, $D\left(\frac{\psi_{k-1}}{\varrho_k}\right) = \frac{\psi_k}{\varrho_k}$ for every value of $x$ for which both sides are defined.*

In the case $f(x)$ is a polynomial of degree $n - 1$, the BF sequence is given by setting, for any $i = 1, \ldots, n$, $\varrho_i = 1$, and $\psi_i = f^{(i)}$. Then, as noted above, a simple grid for $\psi_{i-1}$ is given by the set of sign-changing zeros of $\psi_i$. Moreover, the $n$-th derivative of $f(x)$ is the zero function. However, this same procedure does not work even for simple analytical functions, such as, for instance, $f(x) = \sin(x) + x\cos(x)$.

**Remark 2.3.** *In practical applications, the pivots are functions for which the real zeros and poles can be determined independently. This is the case, for example, for polynomials.*

**Remark 2.4.** *We shall use repeatedly in the sequel the following observation. Let $\mathcal{G} = \{x_1, \ldots, x_n\}$ be a simple grid for a function $f$ on the interval $[a, b]$, given by the sign-changing roots of a function $h$. Then, the same grid is a simple grid for the function $\lambda \cdot f$, where $\lambda$ is a continuous, nonvanishing function on $(a, b) \setminus \mathcal{G}$.*

**Remark 2.5.** *Note that if a simple interval $(c, d)$ for a function $f$ is given, then by inspecting the signs of $f$ at $c$ and $d$ (actually the signs of $f(c + \epsilon)$ and $f(d - \epsilon)$ for sufficiently small positive $\epsilon$ to be more precise), one can determine whether $f$ has a sign-changing zero in the open interval and, if so, one can determine the sign-changing zero with arbitrary precision using a bisection algorithm.*

**Remark 2.6.** *Suppose the pivot functions are such that we can independently determine their zeros and poles. Then, from the definition:*

$$D\left(\frac{\psi_{n-1}}{\varrho_n}\right) = \frac{\psi_n}{\varrho_n}$$

*and since by assumption $\psi_n \equiv 0$, it follows that $\psi_{n-1}$ is a multiple of $\varrho_n$, and, hence, $\psi_{n-1}$ inherits the poles and zeros of $\varrho_n$.*
*For any $k$, once one knows the poles as well as the sign-changing zeros of $\psi_k$, one can form a grid containing those poles and sign-changing zeros as well as all the zeros and poles of $\varrho_k$. It then follows that the grid is a simple grid for $\psi_{k-1}$. The reason is that for any interval in between two consecutive grid points, both the sign of $\varrho_k$ and that of $\psi_k$ do not change and the functions $\frac{\psi_{k-1}}{\varrho_k}$ and $\frac{\psi_k}{\varrho_k}$ are well-defined as they have no poles. Therefore, $\frac{\psi_{k-1}}{\varrho_k}$ is monotonic on such an interval and, hence, $\psi_{k-1}$ has at most one sign-changing zero (and no poles) on such an interval.*

*Therefore, using the previous remark, one can determine all the sign-changing zeros of $\psi_{n-1}$ (if they exist) using a bisection procedure, and by repeating this procedure for $k = n-2, n-3, \ldots, 1$, one can determine all the sign-changing zeros of $f$.*

### *2.2. Practical applications and theoretical limitations*

In the recent paper [6], the problem of determining zeros of EPT functions (referred to as exponential polynomials in that paper) is examined from the perspective of "computable analysis" (see, e.g., [7, 37] and the references therein). A key issue relevant to the present work is the determination of the zero points. In our algorithm, these points are obtained by bisection or similar root isolation methods, effectively as limits of some sequence, which means that, in practice, we work with approximations of these limits. Our algorithm requires determining the sign of a function at points known only through numerical approximations. If the function's sign is positive or negative at such a point, the correct sign can be recovered by using a sufficiently accurate approximation, assuming continuity. However, if the function's value is zero, it may be impossible to determine this without additional information (see, e.g., [6] for approaches to overcoming this issue under strong assumptions, including an unproven conjecture). While we do not believe this poses a significant obstacle for our algorithm, it would be interesting to investigate how the implementation could be made entirely rigorous, accounting for precision levels, round-off errors, and similar factors.

(i) In our algorithm, we work with simple intervals – intervals in which the function has at most one zero-crossing (i.e., sign-changing zero). If the function is zero at an endpoint, say $b$, a zero-crossing within the open interval cannot be ruled out, but a bisection procedure will either be able to identify it or will give a (very) small interval $[b-\epsilon, b]$, $\epsilon > 0$ small, that *may* contain a zero-crossing. The number $\epsilon$ can be made arbitrarily small by sufficiently extending the bisection process. If a zero-crossing is identified, its location is added to the grid. In cases of apparent non-sign-changing zeros, one might argue that they should be added to the grid, as they could indicate nearby, undetected zero-crossings. However, if such zero-crossings exist, they would be extremely close to each other, likely resulting in the function having the same sign at these points in the next GBF sequence (assuming nonzero values). Therefore, omitting this pair from consideration is unlikely to cause issues.

(ii) Suppose we are interested in demonstrating that a function $f_0$, with integral equal to one, is nonnegative, and hence a pdf, using the GBF algorithm. If the function is positive except at some point $x_0$ where it is zero, the GBF algorithm will identify the point $x_0$. However, as discussed above, it is not possible to guarantee, based solely on computational outcomes, that the function is nonnegative. In such a case, it may be safest to refrain from declaring the function nonnegative without further information. However, if our set contains a pdf $p$ that is positive everywhere on the domain, we can consider the mapping $\lambda \mapsto \frac{f_0+\lambda p}{1+\lambda}$, which maps $\lambda \in [-\varepsilon, \varepsilon]$, $0 < \varepsilon < 1$, to the function $f_\lambda$. By selecting

$\varepsilon$ large enough such that $f_{-\varepsilon}$ has a demonstrably negative value at $x_0$ while $f_\varepsilon$ has a demonstrably positive value at $x_0$, we can use bisection on $\lambda$ (via the GBF algorithm) to find the value $\lambda^*$ for which $f_{\lambda^*}$ is on the boundary of the set of pdfs. If $f_0$ is indeed nonnegative and satisfies $f_0(x_0) = 0$ as assumed, then $\lambda^* = 0$. In any case, $f_{\lambda^*}$ will be a nonnegative pdf on the boundary of the set of pdfs. In actual implementation, $\lambda^*$ is likely to be very small, and for theoretical purposes, one could replace $f_0$ by $f_{\lambda^*}$ to ensure a valid pdf.

In conclusion, it is important to emphasize the distinction between the *theoretical algorithm*, where we assume the ability to perfectly identify the limits of convergent sequences, and the *implementation of the algorithm* using a computer. We have included results from such implementations in later sections to demonstrate how the algorithm works, but we are aware that further improvements in implementation are possible, especially to handle exceptional cases, such as non-sign-changing zeros of high multiplicity.

### *2.3. Pivot functions and the construction of a GBF sequence*

In this subsection, we outline a strategy to generate a sequence of pivots and hence at a GBF sequence for a given function, without delving into the specifics for the time being. These details will be dealt with for several classes of functions in the following sections. Suppose that $f$ is a function in the linear span $V$ of $n$ linearly independent real analytic functions $\{h_1, \ldots, h_n\}$. Let us assume (and in the sequel we shall discuss when this assumption actually holds) that there exists an $n-$th order linear differential operator $\Phi$ such that

$$\Phi(h_j) = 0 \qquad j = 1, \ldots, n.$$

Then, by linearity, it holds that $\Phi(f) = 0$ as well.

Assume we can factorize $\Phi$ into first-order linear differential operators,

$$\Phi = \Phi_n \circ \cdots \circ \Phi_1,$$

where for any $j$ we have $\Phi_j(f) = Df + b_j f$ for some function $b_j$, for $f \in V$. Then, we can associate with this a growing sequence of subspaces of $V$, namely,

$$\begin{aligned} V_1 &= \mathrm{Ker}(\Phi_1), \\ V_2 &= \mathrm{Ker}(\Phi_2 \circ \Phi_1), \\ &\cdots \\ V_n &= \mathrm{Ker}(\Phi_n \circ \cdots \circ \Phi_1) = \mathrm{Ker}(\Phi). \end{aligned}$$

**Claim 2.7.** *Under certain regularity conditions there will exist a basis* $\{\tilde{h}_1, \ldots, \tilde{h}_n\}$ *of* $V = \mathrm{span}(h_1, \ldots, h_n)$ *such that*

$$V_1 = \mathrm{span}(\tilde{h}_1), \qquad V_2 = \mathrm{span}(\tilde{h}_1, \tilde{h}_2), \qquad \ldots \qquad V_n = V = \mathrm{span}(\tilde{h}_1, \ldots, \tilde{h}_n).$$

Notice that when $j \le k$, one has

$$\Phi_k \circ \cdots \circ \Phi_1(\tilde{h}_j) = 0.$$

Now, we set

$$\varrho_1 = \tilde{h}_1, \varrho_2 = \Phi_1(\tilde{h}_2), \quad \ldots \qquad \varrho_k = \Phi_{k-1} \circ \cdots \circ \Phi_1(\tilde{h}_k), \quad \ldots \qquad \varrho_n = \Phi_{n-1} \circ \cdots \circ \Phi_1(\tilde{h}_n).$$

We claim that under certain regularity conditions, $\{\varrho_1, \ldots, \varrho_n\}$ can play the role of the sequence of *pivot* functions in Definition 2.2.

Now, since $f \in V$, there exists a representation of $f$ in terms of the basis $\{\tilde{h}_1, \ldots, \tilde{h}_n\}$,

$$f = \sum_{j=1}^{n} f_j \tilde{h}_j, \ f_j \in \mathbb{R}, \ j = 1, 2, \ldots, n.$$

Then,

$$\begin{aligned}
\Phi_1(f) &= \sum_{j=2}^{n} f_j \Phi_1(\tilde{h}_j), \\
\Phi_2 \circ \Phi_1(f) &= \sum_{j=3}^{n} f_j \Phi_2 \circ \Phi_1(\tilde{h}_j), \\
&\vdots \\
\Phi_k \circ \ldots \circ \Phi_1(f) &= \sum_{j=k+1}^{n} f_j \Phi_k \circ \ldots \circ \Phi_1(\tilde{h}_j) \\
&\vdots \\
\Phi_{n-2} \circ \ldots \circ \Phi_1(f) &= \sum_{j=n-1}^{n} f_j \Phi_{n-2} \circ \ldots \circ \Phi_1(\tilde{h}_j) \\
\Phi_{n-1} \circ \ldots \circ \Phi_1(f) &= f_n \Phi_{n-1} \circ \ldots \circ \Phi_1(\tilde{h}_n) \\
\Phi_n \circ \ldots \circ \Phi_1(f) &= 0.
\end{aligned}$$

We notice that for each $k = 1, 2, \ldots, n$, we have $\Phi_k(\varrho_k) = 0$ by construction and, therefore, $b_k = \frac{-D\varrho_k}{\varrho_k}$. However, that means that for each function $g \in V$, one has

$$\Phi_k(g) = \varrho_k D\left(\frac{g}{\varrho_k}\right), \quad k = 1, \ldots, n \tag{3}$$

which implies, in particular,

$$\Phi_k \circ \cdots \circ \Phi_1(f) = \varrho_k D\left(\frac{\varrho_{k-1}}{\varrho_k} D\left(\frac{\varrho_{k-2}}{\varrho_{k-1}} \ldots \left(\frac{\varrho_1}{\varrho_2} D\left(\frac{f}{\varrho_1}\right)\right)\right)\right), \ k = 1, 2, \ldots, n \tag{4}$$

and that, if $\varrho_1, \varrho_2, \ldots, \varrho_n$ can indeed act as the sequence of pivots, the corresponding GBF sequence for $f$ is given by $\psi_0 = f$ and

$$\psi_k = \Phi_k \circ \Phi_{k-1} \circ \ldots \circ \Phi_1(f), \ k = 1, \ldots, n. \tag{5}$$

In the next subsection, we shall use the *Polya-Ristroph formula* to obtain such a sequence of operators $\{\Phi_j, \ j = 1, \ldots, n\}$ explicitly and show that we can actually do that in such a way that $h_j = \tilde{h}_j, \ j = 1, \ldots, n$.

*2.4. A formula by Polya and Ristroph*

We will make use of a classical formula by G. Polya [28] and results from R. Ristroph [30], which state the following.

Let $\{h_i\}$ be a finite sequence of real, linearly independent real analytic functions on a given open interval $I$. Let, for $m \in \mathbb{N}$,

$$W_m(x) = W(h_1, \dots, h_m)(x) = \begin{vmatrix} h_1 & h_2 & \dots & h_m \\ Dh_1 & Dh_2 & \dots & Dh_m \\ \dots & & & \dots \\ D^{m-1}h_1 & D^{m-1}h_2 & \dots & D^{m-1}h_m \end{vmatrix}(x)$$

be the Wronskian defined on $I$. We set $W_0(x) = 1$ and we notice that $W_1(h_1) = h_1$ and $W_2(h_1, h_2) = h_1(Dh_2) - (Dh_1)h_2$. Denote, for any real analytic function $f$,

$$K_m(f) = \frac{W(h_1, \dots, h_m, f)}{W(h_1, \dots, h_m)}. \tag{6}$$

Then, $f \mapsto K_m(f)$ is the *unique* monic[1] $m$-th order linear differential operator for which $\{h_i,\ i = 1, \dots m\}$ is a fundamental set.

**Remark 2.8.** *We introduce the following expansion for the differential operator $K_m$:*

$$K_m(f) = \sum_{j=0}^{m} a_{m,j} D^j f, \qquad a_{m,m} = 1 \tag{7}$$

*(the dependence on $x$ is suppressed for notational simplicity).*

*We remark that for $0 \le j \le m-1$, $a_{m,j}$ is the quotient of two real analytic functions and its poles are therefore the zeros of the denominator and, hence, of $W_m(x)$. Since further (cf. Polya's paper [28])*

$$DW_m(x) = -a_{m,m-1}(x) W_m(x),$$

*it follows that whenever $W_m(x)$ does not vanish inside $I$, the representation*

$$W_m(x) = \exp\left( - \int_{x_0}^{x} a_{m,m-1}(z)\, \mathrm{d}z \right) W_m(x_0) \tag{8}$$

*holds for any $x, x_0 \in I$.*

**Remark 2.9.** *The following expression of the differential operator $K_m(f)$ is proved in [28] by a direct computation involving the derivative of the Wronskian, and in [30] via a direct argument based on induction:*

$$K_m(f) = \frac{W_m}{W_{m-1}} D\left( \frac{W_{m-1}^2}{W_m W_{m-2}} D\left( \frac{W_{m-2}^2}{W_{m-1} W_{m-3}} \dots D\left( \frac{W_2^2}{W_3 W_1} D\left( \frac{W_1^2}{W_2} D\left( \frac{f}{W_1} \right) \right) \dots \right) \right) \right) \tag{9}$$

*In particular, for $m = 2$ we obtain*

$$K_2(f) = \frac{W_2}{W_1} D\left( \frac{W_1^2}{W_2} D\left( \frac{f}{W_1} \right) \right). \tag{10}$$

*Note that* (9) *and* (10) *hold for all $x$ for which none of the denominators involved are zero.*

[1] i.e., with leading coefficient equal to 1

From equation (9), it follows that in case the Wronskians $W_j,\ j = 1, \ldots, n$ are functions for which the zeros can be obtained independently (e.g., if they are the (real) zeros of polynomials), then the quotients $\varrho_k := \frac{W_k}{W_{k-1}},\ k = 1, 2, \ldots, n$, where $W_0 \equiv 1$ by convention, can play the role of the sequence of pivots and then $K_m(f),\ m = 1, 2, \ldots, n$ is the associated GBF sequence in case $f$ lies in the linear span of $\{h_1, \ldots, h_n\}$.

In the following four sections, we will use the general framework presented here to develop pivot sequences and corresponding GBF sequences for a number of classes of functions and explain how these can be used to obtain all the sign-changing zeros of a given function on a given closed interval $[a, b]$. It will also be explained how for some cases it is possible to extend the results to obtain all sign-changing zeros on the real line.

## 3. Analysis of the class of EPT functions

Consider a function $f : [0, \infty) \to \mathbb{R}$ of the form (1). We say that $f$ is a real *EPT function* if

$$f(x) = \Re\left(\sum_{k=1}^{m} p_k(x)e^{\mu_k x}\right), \qquad x \geq 0, \tag{11}$$

where $m \geq 1$, $\mu_k \in \mathbb{C}$, and $p_k(x)$ are complex-valued polynomials, for $k = 1, \ldots, m$. We shall denote by $\mathcal{B}$ the set of real EPT functions. Let us state a few alternative characterizations, that may be useful in the sequel; see, for instance, [33].

**Lemma 3.1.** *$f \in \mathcal{B}$ if, and only if, any of the following equivalent characterizations hold:*

1. *there exist a real $\tilde{n} \times \tilde{n}$ matrix $A$, a real $\tilde{n}$-dimensional column vector $b$, and a real $\tilde{n}$-dimensional row vector $c$ such that $f(x) = ce^{Ax}b$.*
   *If $n$ is the minimal possible choice for $\tilde{n}$, given $f \in \mathcal{B}$, then we say that $(A, b, c)$ is a minimal realization of $f$ and $n$ is the* order *(or* McMillan degree*) of the function $f$.*
2. *the Laplace transform of $f$ is a strictly proper rational function.*
3. *$f$ is the solution of a linear differential equation with constant coefficients (Euler-D'Alembert class).*

This last result indicates that there exists a linear differential operator $p(D)$ such that $p(D)f = 0$. The associated polynomial $p(x)$ has degree $\deg(p) \geq n$, and if we search for such an operator $p(D)$ having minimal degree we obtain $\deg(p) = n$ ($n$ is the *McMillan degree*). If we search for a minimal degree, monic polynomial, we have the uniqueness of $p(D)$ and the representation

$$p(x) = |xI - A|,$$

where $A$ is the matrix from a minimal realization $(A, b, c)$ of $f$.

Since $p(x)$ is a real polynomial of degree $n$, it can be factorized into linear and quadratic real factors

$$p(x) = p_1(x)p_2(x) \ldots p_{a+b}(x), \tag{12}$$

where $p_1, \ldots, p_a$ are linear factors $p_i = x - \lambda_i$, and $p_{a+1}, \ldots, p_{a+b}$ are irreducible quadratic factors $p_{a+i}(x) = x^2 - 2\theta_i + \rho_i^2$, with $\theta_i^2 < \rho_i^2$, [2], and $a + 2b = n$.

[2]The roots of $p_{a+i}$ are $\theta_i \pm i\sqrt{\rho_i^2 - \theta_i^2}$, which can be written as $\rho_i e^{\pm i \arccos(\theta_i/\rho_i)}$

### 3.1. GBF sequence for EPT functions

Now, we construct a GBF sequence associated with an EPT function with minimal realization $f(x) = ce^{Ax}b$. We know that $p(D)f(x) = 0$ for $p(x) = |xI - A|$. We start by setting $\psi_0 = f$.

Let $p_1(x)$ be the first factor of $p(x)$, as constructed in (12), and assume that $p_1(x) = x - \lambda_1$ is a linear factor. We set $h_1(x) = e^{\lambda_1 x}$, so that $p_1(D)h_1 = 0$. Define

$$\psi_1(x) = p_1(D)\psi_0(x) = h_1 D\left(\frac{\psi_0}{h_1}\right).$$

Then, since $h_1(x) > 0$ definitely, it follows that on any interval on which $\psi_1$ does not change sign, then $h_1^{-1}\psi_0$ is *monotonic*, hence the interval will be a *simple interval* for $\psi_0$. Therefore, the set $R(\psi_1, [a, b])$ of sign-changing zeros for $\psi_1$ defines a simple grid for $\psi_0$. We proceed recursively for any $i = 1, \ldots, a$ ($a$ being the number of linear factors in $p(x)$), thus constructing a sequence $\{\psi_0, \psi_1, \ldots, \psi_a\}$, where $\psi_i = p_i(D)\psi_{i-1} = \left(\prod_{j=1}^{i} p_j(D)\right)\psi_0$.

Now let us consider $p_{a+1}(x)$, that is, the first irreducible quadratic polynomial in the factorization of $p(x)$. It seems natural to associate to this term *two* elements of the GBF sequence. In order to explain the procedure, recall that $p_{a+1}(x) = x^2 - 2\theta_{a+1}x + \rho_{a+1}^2$ with $\theta_{a+1}^2 < \rho_{a+1}^2$. Let further $\psi_a(x) = \left(\prod_{j=1}^{a} p_j(D)\right)\psi_0$ be the last element of the GBF sequence constructed before.

Consider the fundamental set

$$h_1(x) = e^{x\theta_{a+1}} \sin\left(x\sqrt{\rho_{a+1}^2 - \theta_{a+1}^2}\right), \qquad h_2(x) = e^{x\theta_{a+1}} \cos\left(x\sqrt{\rho_{a+1}^2 - \theta_{a+1}^2}\right),$$

associated with the operator $p_{a+1}(D)$.

Recall, from formula (10) that

$$p_{a+1}(D)\psi_a(x) = D^2\psi_a(x) - 2\theta_{a+1}D\psi_a(x) + \rho_{a+1}^2\psi_a(x) = \frac{W_2}{W_1}D\left(\frac{W_1^2}{W_2}D\left(\frac{\psi_a}{W_1}\right)\right)$$

with $W_1(x) = h_1(x)$, $W_2(x) = \begin{vmatrix} h_1(x) & h_2(x) \\ h_1'(x) & h_2'(x) \end{vmatrix}$.

We can compute $W_2$ by taking into account formula (8): since $a_{2,1} = -2\theta_{a+1}$, we obtain $W_2(x) = e^{2(x-x_0)\theta_{a+1}}W_2(x_0) = ce^{2x\theta_{a+1}}$ for some nonzero constant $c$.

We define [3]

$$\psi_{a+1} = c\frac{W_1^2}{W_2}D\left(\frac{\psi_a}{W_1}\right) = \sin^2\left(x\sqrt{\rho_{a+1}^2 - \theta_{a+1}^2}\right) D\left(\frac{\psi_a}{e^{x\theta_{a+1}} \sin\left(x\sqrt{\rho_{a+1}^2 - \theta_{a+1}^2}\right)}\right),$$

$$\psi_{a+2} = \frac{W_2}{cW_1}D\psi_{a+1} = \frac{1}{e^{-x\theta_{a+1}} \sin\left(x\sqrt{\rho_{a+1}^2 - \theta_{a+1}^2}\right)}D\psi_{a+1}.$$

Note that from

$$\frac{\psi_{a+1}(x)}{\sin^2\left(x\sqrt{\rho_{a+1}^2 - \theta_{a+1}^2}\right)} = D\left(\frac{\psi_a(x)}{e^{x\theta_{a+1}} \sin\left(x\sqrt{\rho_{a+1}^2 - \theta_{a+1}^2}\right)}\right)$$

[3] One can also define $\psi_{a+1} = W_1 D\left(\frac{\psi_a}{W_1}\right)$ and $\psi_{a+2} = \frac{W_2}{W_1}D\left(\psi_{a+1}\frac{W_1}{W_2}\right)$ but the choices made in the text are a bit simpler.

it follows that if we join the set of zeros of the function $\sin\left(x\sqrt{\rho_{a+1}^2-\theta_{a+1}^2}\right)$ with the set of sign-changing zeros of $\psi_{a+1}$, we will get a simple grid for $\psi_a$, from which the sign-changing zeros of $\psi_a$ can be deduced by a bracketing procedure.

Moreover, since again

$$\psi_{a+2} = \frac{W_2}{cW_1} D\psi_{a+1},$$

it follows that a simple grid for $\psi_{a+1}$ is given by the sign-changing zeros of $\psi_{a+2}$ and the zeros of the function $\sin\left(x\sqrt{\rho_{a+1}^2-\theta_{a+1}^2}\right)$. Finally, we remark that

$$\psi_{a+2}(x) = p_{a+1}(D)\psi_a(x).$$

Note that this implies that $\psi_{a+2}$ has no poles as $\psi_a$ is real analytic.

Continuing on this way, we obtain a sequence of functions $\psi_{a+1}, \ldots, \psi_n(x)$. Notice that $n = a + 2b$ and

$$\psi_n(x) = \prod_{j=1}^{b} p_{a+j}(D) \prod_{i=1}^{a} p_i(D)f = p(D)f = 0$$

and $\{\psi_j,\ j = 0, \ldots, n\}$ forms a GBF sequence for $f$.

**Remark 3.2.** *In [17], this GBF sequence was constructed without a good understanding of* why *this works. Here, the link between the factorization of linear differential operators into first-order linear differential operators and the fundamental solutions of linear differential operators is made for the first time.*

**Remark 3.3.** *An application of this algorithm to monitor the positivity of linear combinations of exponential functions can, for instance, be found in [1].*

## 4. Analysis of Polynomial-Gaussian mixtures

In this section, we consider functions of the form

$$f(x) = \Re\left[\sum_{j=1}^{n} p_j(x)e^{q_j(x)}\right], \tag{13}$$

where we assume that

$$q_j \text{ are real second-order polynomials.} \tag{14}$$

In this case, with no loss of generality, we can also assume that the polynomials $p_j$ are real-valued as well and we denote $h_j(x) = p_j(x)e^{q_j(x)}$ for $j = 1, \ldots, n$, and also that the functions $\{h_j(x),\ j = 1, \ldots, n\}$ are linearly independent. We write

$$f(x) = \sum_{j=1}^{n} p_j(x)e^{q_j(x)} = \sum_{j=1}^{n} h_j(x). \tag{15}$$

We will call functions of this type Polynomial-Gaussian mixtures (PGMs). Note that they can also be viewed as mixtures of Hermite functions.

### 4.1. GBF sequence for PGMs

For arbitrary $m \leq n$, let us consider the Wronskian determinant $W_m$. Notice that, given a function $h_j$, its derivatives of all orders are of the same form:

$$D^k h_j(x) = p_{j,k}(x)e^{q_j(x)}$$

where $p_{j,k}(x)$ is a polynomial whose coefficients can be explicitly computed. Since

$$W(h_1, \ldots, h_m) = \begin{vmatrix} h_1 & h_2 & \ldots & h_m \\ Dh_1 & Dh_2 & \ldots & Dh_m \\ \ldots & & & \ldots \\ D^{m-1}h_1 & D^{m-1}h_2 & \ldots & D^{m-1}h_m \end{vmatrix}$$

it follows that the Wronskian $W_m$ has the form

$$W_m(x) = P_m(x)e^{\sum_{j=1}^m q_j(x)}$$

where $P_m$ is a polynomial and $P_m \not\equiv 0$. As we can determine all the zeros of a polynomial within the interval $[a, b]$ (e.g., using the BF method, the Sturm chain method, etc.), the Polya-Ristroph formula gives us a GBF sequence:

$$\begin{aligned} &\psi_0 = f = \sum_{j=1}^n p_j(x)e^{q_j(x)} && \psi_1 = W_1 D\left(\frac{\psi_0}{W_1}\right) \\ &\psi_2 = \frac{W_2}{W_1} D\left(\frac{W_1}{W_2}\psi_1\right) && \ldots \\ &\ldots && \psi_n = \frac{W_n}{W_{n-1}} D\left(\frac{W_{n-1}}{W_n}\psi_{n-1}\right) \end{aligned} \tag{16}$$

Notice that $\frac{W_j}{W_{j+1}} = \frac{P_j}{P_{j+1}}e^{q_{j+1}}$ and its inverse $\frac{W_{j+1}}{W_j}$ are rational functions times an exponential of which we can determine all the zeros and poles. Further, $\psi_n \equiv 0$ as $f$ is a solution to the linear differential equation

$$\frac{W(h_1, \ldots, h_n, f)}{W(h_1, \ldots, h_n)} = 0.$$

So, indeed $\psi_0 = f, \psi_1, \ldots, \psi_n \equiv 0$ is a GBF sequence for $f$.

### 4.2. An alternate approach

Although the above method is mathematically elegant, it requires the computation of high-order derivatives and the determinants of large matrices, which can make it seem quite elaborate and difficult to grasp at first glance. To address this, we propose an alternative, more intuitive approach for constructing the *same* GBF sequence, offering the reader a clearer structural framework for understanding the problem. As demonstrated below, this approach is based on an iterative procedure that only requires the computation of first-order derivatives at each step, rather than Wronskians of arbitrary order.

However, it is important to note that this iterative method involves multiple computations of polynomial quotients, and properly addressing the detection of common factors is crucial. This is known to be a challenging problem, especially when exact computation is not used. One potential solution is to compute symbolic expressions of the polynomials, extract common factors, and store the results in a database. These results could then be adapted to different implementations by simply substituting the symbolic values of the coefficients with numerical ones. Unfortunately, this approach was found to be computationally infeasible for larger values of $n$ due to the limitations of symbolic polynomial division.

Therefore, we opted to implement the computation in Section 5 using the GBF sequence with the Wronskians approach. Nevertheless, the alternative method is theoretically sound and provides a clear perspective on the mechanism underlying the construction of the GBF sequence, as we will illustrate in the example provided in appendix B.1. Recall the representation in (15):

$$f(x) = \sum_{j=1}^{n} p_j(x)e^{q_j(x)} = \sum_{j=1}^{n} h_j(x).$$

We write $\psi_1$ in the following form:

$$\psi_1(x) = h_1(x)D\left(\frac{f(x)}{h_1(x)}\right) = h_1(x)\sum_{j=2}^{n} D\left(\frac{h_j(x)}{h_1(x)}\right) = \sum_{j=2}^{n} \underbrace{\left(h_j'(x) - h_j(x)\frac{h_1'(x)}{h_1(x)}\right)}_{h_{1;j}}$$

A simple grid for $\psi_0 = f$ is given by the zeros and the poles of $\psi_1$ and those of $h_1$ (i.e., those of $p_1$). It shall be noticed that the functions $h_{1;j}(x)$ have a peculiar form

$$h_{1;j}(x) = r_{1;j}(x)e^{q_j(x)},$$

where $r_{1;j}(x)$ is a rational function whose poles correspond to (a subset of) the zeros of $h_1$.

Since we are interested in the construction of a GBF sequence, we write

$$\psi_2(x) = h_{1;2}(x)D\left(\frac{\psi_1(x)}{h_{1;2}(x)}\right).$$

A direct computation shows that this function coincides with the second term in the GBF sequence defined before. Therefore, just as before, a simple grid for $\psi_1$ is given by the zeros and the poles of $\psi_2$ and those of $r_{1;2}$ (the rational factor of $h_{1;2}$). However, this new formulation shows that while $\psi_1$ is defined by the sum of $n-1$ terms, now $\psi_2$ is given by the sum of $n-2$ terms, that is, with a consistent notation:

$$\psi_2(x) = \sum_{j=3}^{n}\left(h_{1;j}'(x) - h_{1;j}(x)\frac{h_{1;2}'(x)}{h_{1;2}(x)}\right) = \sum_{j=3}^{n} h_{2;j}(x).$$

therefore, the procedure finishes in $n$ steps, like before. We can identify the pivot functions as $\varrho_1 = h_1$ and $\varrho_k = h_{k-1;k}$, $k = 2, \ldots, n$. A worked-out example of this procedure is given in appendix B.1.

## 5. Analysis of the class of Gaussian mixtures with arbitrary coefficients

In this section, we use the Polya-Ristroph formula to construct a GBF sequence for finite Gaussian mixtures having coefficients of arbitrary sign

$$f(x) = \sum_{k=1}^{n} \gamma_k h_k(x) \qquad x \in \mathbb{R} \tag{17}$$

where $n$ is given, and

i) $h_k(x) = e^{q_k(x)}$ where $q_k(x)$ is a second degree, nonpositive polynomial $q_k(x) = -\frac{1}{2\sigma_k^2}(x-\mu_k)^2$ with $\mu_k \in \mathbb{R}$, $\sigma_k^2 > 0$ for all $1 \le k \le n$;

ii) for all $1 \le k \le n$, $\gamma_k \in \mathbb{R}$ satisfies $\sum_{k=1}^{n} \sigma_k \gamma_k = \frac{1}{\sqrt{2\pi}}$.

By definition, a function of the form (17) is a pdf provided that

iii) $f(x) \ge 0 \quad \forall\, x \in \mathbb{R}$.

We shall refer to functions of the form (17) as *finite Gaussian mixtures* (see, e.g., [24] and the discussion in Section 1 for an introduction to their properties and main applications in the literature).

**Remark 5.1.** *Recall that for the standard Gaussian density, we define the Hermite polynomials as*

$$H_0(x) = 1, \qquad H_n(x) = (-1)^n e^{x^2/2} D^n e^{-x^2/2}, \qquad n = 1, 2, \ldots.$$

*Then, for* $q(x) = \frac{1}{2\sigma^2}(x-\mu)^2$, *it holds*

$$D^n e^{-q(x)} = \frac{(-1)^n}{\sigma^n} e^{-q(x)} H_n\left(\frac{x-\mu}{\sigma}\right).$$

*Let us denote*

$$P_n(x) = \frac{(-1)^n}{\sigma^n} H_n\left(\frac{x-\mu}{\sigma}\right).$$

*Then, we know that* $P_n(x)$ *is a polynomial of degree n that satisfies the recurrence relation*

$$P_{n+1}(x) = -\frac{x-\mu}{\sigma^2} P_n(x) + DP_n(x).$$

With no loss of generality, we assume that $h_i \ne h_j \;\; \forall\, i \ne j$ since any mixture can be traced back to this case. Notice that for any given $h_k(x)$, its derivatives of all orders are of the form

$$D^m h_k(x) = P_{m,k}(x) h_k(x)$$

where $P_{m,k}$ is the generalized $m$-th degree Hermite polynomial with coefficients $\mu_k$ and $\sigma_k^2$:

$$P_{m,k}(x) = \frac{(-1)^m}{\sigma_k^m} H_m\left(\frac{x-\mu_k}{\sigma_k}\right), \qquad m = 1, 2, \ldots.$$

Since they play an important rôle in the sequel, we introduce the linear functions

$$\phi_k(x) = Dq_k(x) = P_{1,k}(x) = \frac{\mu_k - x}{\sigma_k^2}, \qquad k = 1, \ldots, n, \tag{18}$$

so, for instance, the recurrence relation reads

$$P_{m+1,k} = (\phi_k + D)\, P_{m,k}(x), \qquad m = 0, 1, \ldots, k = 1, \ldots, n.$$

We can further express the $m$-th order Wronskian for the sequence $\{h_j\}_{j=1}^n$ as

$$W(h_1, \ldots, h_m) = \left( \prod_{j=1}^{m} e^{-q_j} \right) \begin{vmatrix} 1 & 1 & \ldots & 1 \\ P_{1,1} & P_{1,2} & \ldots & P_{1,m} \\ \ldots & & & \ldots \\ P_{m-1,1} & P_{m-1,2} & \ldots & P_{m-1,m} \end{vmatrix}.$$

For any $j \geq 1$ and for any sequence of increasing integers $i_1, \ldots, i_j$, we introduce the polynomial

$$Q^{j-1}_{i_1,\ldots,i_j} = \begin{vmatrix} 1 & 1 & \ldots & 1 \\ P_{1,i_1} & P_{1,i_2} & \ldots & P_{1,i_j} \\ \ldots & & & \ldots \\ P_{j-1,i_1} & P_{j-1,i_2} & \ldots & P_{j-1,i_j} \end{vmatrix} \tag{19}$$

(and, in particular, $Q^0_k = 1$ for any $k \geq 1$) so that

$$W(h_1, \ldots, h_m) = \left( \prod_{j=1}^{m} e^{-q_j} \right) Q^{m-1}_{1,\ldots,m}. \tag{20}$$

**Theorem 5.2.** *The following relations hold:*

$$Q^1_{j,k} = \phi_k - \phi_j, \qquad 1 \leq j \leq k \leq n \tag{21}$$

*and*

$$Q^j_{1,\ldots,j,k} Q^{j-2}_{1,\ldots,j-1} = Q^{j-1}_{1,\ldots,j} \left[ \phi_k - \phi_j \right] Q^{j-1}_{1,\ldots,j-1,k} + Q^{j-1}_{1,\ldots,j} D_x Q^{j-1}_{1,\ldots,j-1,k} - Q^{j-1}_{1,\ldots,j-1,k} D_x Q^{j-1}_{1,\ldots,j} \tag{22}$$

*for* $2 \leq j < k \leq n$.

The proof of this result is given in appendix A.2 for the sake of clarity of exposition. Next, we can use the above relations to determine a different formulation of the sequence of Wronskians of the problem, which involves recursively defined polynomials, rather than computing derivatives of eventually high orders, which we present in the following result.

**Theorem 5.3.** *Let* $f(x) = \sum_{k=1}^{n} \gamma_k h_k(x)$ *be a finite Gaussian mixture. Then, the GBF sequence obtained using the Polya-Ristroph formula can be rewritten as*

$$\psi_j(x) = \begin{cases} f(x) & j = 0 \\ \displaystyle\sum_{k=j+1}^{n} \lambda_k \frac{Q^j_{1,\ldots,j,k}}{Q^{j-1}_{1,\ldots,j}} h_k(x) & \forall\ 1 \leq j < n. \end{cases} \tag{23}$$

The proof of this result is given in the appendix A.3.

**Remark 5.4.** *The dilemma of choosing one option over another when constructing the GBF sequence naturally arises for computational purposes. Indeed, while opting for computing the Wronskians may result in high-order derivatives and large matrices, it would also offer a cleaner and more direct definition of the objects involved in the GBF sequence construction process. On the other hand, although the polynomial recursive approach may seem faster since it involves only first-order derivatives, it should be noted that it entails hidden recursive dependencies as it requires polynomial division.*

*In the next subsection, we will outline the main steps of the implemented algorithm, which computes the GBF sequence following the Wronskians approach. This choice is made to avoid multiple executions of polynomial divisions, which lead to long execution times if performed with symbolic coefficients and incorrect detection of common factors in the case of numerical coefficients. Additionally, it is based on the availability in the literature of a fast algorithm for symbolic determinant computation using Gaussian elimination procedure [21].*

*For completeness, we also computed the Wronskians using symbolic expressions and stored them in a database (see [5]) for dimensions ranging from* 2 *to* 16*. This was done to provide the reader with ready-to-use general results, hoping to inspire future research in finding simpler recursive structures for GBF sequences.*

### *5.1. Implementation*

We have developed a software in Matlab aimed at providing readers with a hands-on tool to investigate how the GBF approach works in practice. The tool is freely available [5].

The algorithm takes the following parameters as inputs: the number of Gaussians involved in the mixture, their means, variances, the mixing coefficients, and the desired accuracy level for the searching-roots algorithm. It is important to note that the procedure does not require a bounded interval to test the mixture because such an interval is computed internally. Therefore, a searching-roots problem on the real line is reduced to a constrained searching-roots procedure on a bounded interval. As outputs, the software returns the location of any sign-changing roots and the bounded interval in which they are contained.

If the software returns no sign-changing zeros for the Gaussian mixture $f(x)$, and $f(x)$ is positive at least at one point, then it is nonnegative everywhere and defines a proper pdf.

#### *Technical aspects and execution time*

To further reduce computational time, the classical bisection method is replaced by Ridders' method which proves to be more efficient in searching-roots problems [29]. Execution times (in secs) are displayed in Table 1. They were computed by running 100 simulations for each dimension spanning from 2 to 12 where the parameters were chosen as follows:

- the means were randomly picked in [-10, 10];
- the variances were randomly picked in [0.1, 1];
- the mixing coefficients were randomly picked in [-1, 1];
- the accuracy level was set equal to $2.2204^{-16}$.

| n | nsim | min | median | average | max | % time GBF sequence |
|---|---|---|---|---|---|---|
| 2 | 100 | 0.23 | 0.60 | 0.77 | 1.55 | 28.92 |
| 3 | 100 | 0.43 | 1.16 | 1.20 | 3.98 | 29.17 |
| 4 | 100 | 0.83 | 2.79 | 2.69 | 5.74 | 28.26 |
| 5 | 100 | 1.49 | 5.83 | 5.92 | 11.14 | 25.82 |
| 6 | 100 | 5.06 | 10.85 | 11.35 | 29.91 | 24.23 |
| 7 | 100 | 7.96 | 22.73 | 22.12 | 38.28 | 23.95 |
| 8 | 100 | 21.20 | 42.15 | 42.88 | 62.92 | 37.16 |
| 9 | 100 | 50.44 | 91.45 | 91.25 | 183.57 | 36.52 |
| 10 | 100 | 111.18 | 177.90 | 177.68 | 256.83 | 49.26 |
| 11 | 100 | 338.22 | 473.11 | 486.49 | 922.00 | 68.32 |
| 12 | 100 | 1260.00 | 1466.10 | 1533.40 | 2371.40 | 79.03 |

**Table 1.** Analysis of execution times for dimensions spanning from $n = 2$ to $n = 12$.

Furthermore, average execution times for higher dimensions were estimated by fitting available data by a function of the kind $ae^{bx} + ce^{dx}$ ($a = 0.4517, b = 0.5572, c = 3.61e-05, d = 1.441$) with the Matlab function fit. The results are shown in Figure 1.

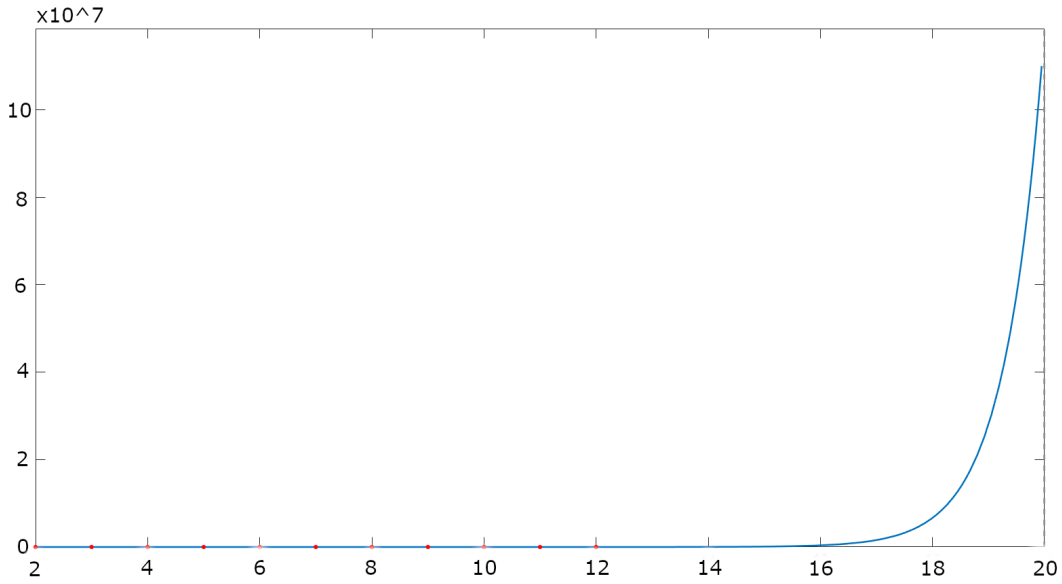


**Figure 1.** Fitted curve for average execution times up to $n = 20$.

As we can observe from the plot, average execution times tend to significantly increase from $n = 16$, reaching values in the range of $10^7$ secs for $n = 20$. The increased execution time is due to the heavier computation of the GBF sequences using the Wronskian approach - as long as to the available hardware - as we can observe from the last column of Table 1 showing the percentage of time employed in the construction of the GBF sequence. Furthermore, percentage data highlights the interplay between the

two main building blocks of the algorithm, namely, the construction of the GBF sequence and the searching-root procedure, along different dimensions. In fact, we can observe that for $2 \leq n \leq 4$, the percentage time dedicated to the construction of the GBF sequence stabilizes around $28 - 29\%$, while it decreases around $24 - 26\%$ for $5 \leq n \leq 7$ before reaching higher values spanning from 36% to 79% for $8 \leq n \leq 12$. Therefore, at least at present, we suggest that applications that require high execution times rely on computing systems with large computational power.

### *5.2. Some remarks on rotation-invariant generalized Gaussian mixtures*

Note that if we have a mixture $f(x)$ of Gaussians with mean zero, then we can (easily) find a function $g$ on the nonnegative half line such that $f(x) = g(x^2)$. Note that $g$ is a sum of exponentials, so it is an element of the family of EPT functions. Therefore, one can check nonnegativity using the GBF method for EPT functions described earlier. This actually generalizes to higher dimensional rotation-invariant Gaussian mixtures. A Gaussian density function $f$ on $\mathbf{R}^n$ is rotation invariant with respect to the origin if $f(x) = f(Qx)$ for any orthogonal matrix $Q$. This implies that one can write $f$ as $f(x) = g(\|x\|)$ for an appropriately chosen function $g$. Actually, $g$ is an exponential function. One can check the nonnegativity of any generalized mixture $\sum_{i=1}^{n} \alpha_i f_i$ of such rotation-invariant Gaussians $f_i, i = 1, 2, \ldots, n$, say, by checking whether the corresponding EPT function $\sum_{i=1}^{n} \alpha_i g_i$ is nonnegative, where $g_i$ chosen such that $f_i(x) = g_i(\|x\|)$ holds for each $i = 1, 2, \ldots, n$.

Note further that if $X \in \mathbf{R}^2$ is a bivariate, rotation-invariant Gaussian random variable, the random variable $R^2$, where $R = \|X\|$, is actually exponential; then, if we have a mixture of bivariate Gaussians that are rotation-invariant with respect to the origin, it will correspond to a mixture of exponential densities if we proceed as above. This would open up the methods for calibrating such signed mixtures of exponentials to data, as described in [32], to bivariate rotation-invariant Gaussian mixtures.

## 6. Wasserstein distance

In this section, we aim to present a different application where our algorithm can be useful since, as we shall see below, the computation of the Wasserstein distance between two Gaussian mixtures can be reduced to a simple sum, once we know the position of the zeros of the difference of the corresponding CDFs. We are not interested in discussing the topic of Wasserstein distance in general, and we refer to the existing literature for further discussion, see, e.g., [2]. In order to fix the notation, we start with a few definitions. We let $S \subset \mathbb{R}^n$ denote a general domain, endowed with the Euclidean distance. Below, we usually consider $S = \mathbb{R}$.

**Definition 6.1.** *Let $M_p(S)$ be the space of all probability measures $\mu$ on $S$ with finite $p$-th moment for some $x_0 \in S$:*

$$\int_S |x - x_0|^p \, \mu(\mathrm{d}x) < +\infty.$$

*Then, the Wasserstein-$p$ distance between two probability measures $\mu$ and $\nu$ in $M_p(S)$ is defined as*

$$W_p(\mu, \nu) = \left( \inf_{\gamma} \int_{S \times S} |x - y|^p \, \gamma(\mathrm{d}x, \mathrm{d}y) \right)^{1/p},$$

*where the infimum is taken among all two-dimensional measures $\gamma$ with marginals $\mu$ and $\nu$, respectively.*

It is proved in [11, Theorem 20.1] that in the particular case when $p = 1$, we obtain

$$W_1(\mu,\nu) = \sup\left\{\int \varphi \,\mathrm{d}(\mu-\nu) \;\mid\; \varphi : S \to \mathbb{R} \quad 1-\text{Lipschitz}\right\}.$$

An alternate form of the $W_1$ distance was given by Dall'Aglio [9]:[4]

$$W_1(\mu,\nu) = \int_0^1 \left|F^{-1}(u) - G^{-1}(u)\right| \,\mathrm{d}u, \tag{24}$$

where $F$ and $G$ are the CDFs for $\mu$ and $\nu$, respectively, and $F^{-1}$, $G^{-1}$ denote the quantile functions of $\mu$ and $\nu$, respectively. For the sake of completeness, we recall the definition below.

**Remark 6.2.** *Let us recall the definition of* generalized *inverse (compare [12]) for a distribution function $F$ (also called the* quantile function *of $F$). $F^{-1} : [0,1] \to \bar{\mathbb{R}} = [-\infty, +\infty]$ is defined by*

$$F^{-1}(u) = \inf\{x \in \mathbb{R} \,:\, F(x) \geq u\}, \qquad u \in [0,1],$$

*with the convention that* $\inf \emptyset = +\infty$. *By this definition, the 0-quantile of $F$ is always $F^{-1}(0) = -\infty$.*

For our purposes, the following result is the best way to treat the distance between one-dimensional distributions. Assume that the laws $\mu$ and $\nu$ have CDFs $F$ and $G$, respectively. Then the following result is classical [11, 36].

**Theorem 6.3.** *The Wasserstein-1 distance between $\mu$ and $\nu$ in $M_1(\mathbb{R})$ (see Definition 6.1) is equal to the $L^1$-distance between the corresponding CDFs $F$ and $G$:*

$$W_1(\mu,\nu) = \int_{\mathbb{R}} |F(x) - G(x)| \,\mathrm{d}x. \tag{25}$$

The above formula is well-defined in $M_1(\mathbb{R})$; for completeness (and since it seems a little tricky) we prove it below.

**Lemma 6.4.** *Assume that $\mu$ and $\nu$ belong to $M_1(\mathbb{R})$. Then, $w(x) = F(x) - G(x)$ is absolutely integrable.*

*Proof.* We write

$$\begin{aligned} W_1(\mu,\nu) &= \int_{-\infty}^0 |F(x) - G(x)| \,\mathrm{d}x + \int_0^{+\infty} |(1-F(x)) - (1-G(x))| \,\mathrm{d}x \\ &\leq \int_{-\infty}^0 \{|F(x)| + |G(x)|\} \,\mathrm{d}x + \int_0^{+\infty} \{|1-F(x)| + |1-G(x)|\} \,\mathrm{d}x. \end{aligned}$$

It is sufficient to prove that the following quantity is finite for arbitrary $\mu$:

$$\begin{aligned} \int_{-\infty}^0 F(x) \,\mathrm{d}x + \int_0^{+\infty} (1-F(x)) \,\mathrm{d}x &= \int_{-\infty}^0 \int_{-\infty}^x \mu(\mathrm{d}y) \,\mathrm{d}x + \int_0^{+\infty} \int_x^{\infty} \mu(\mathrm{d}y) \,\mathrm{d}x \\ &= \int_{-\infty}^0 \int_y^0 \mathrm{d}x \,\mu(\mathrm{d}y) + \int_0^{+\infty} \int_0^y \mathrm{d}x \,\mu(\mathrm{d}y) = \int_{-\infty}^0 |y| \,\mu(\mathrm{d}y) + \int_0^{+\infty} |y| \,\mu(\mathrm{d}y) = \mathbb{E}_\mu[|y|] \end{aligned}$$

and the last quantity is finite by assumption. □

[4] notice that we do not make any assumption on the form of the distribution

In the following result, we propose an analog of Markov-Chebyshev's inequality for CDFs. Here, we need to impose the existence of a finite $p$-moment, $p > 1$, for the pdf.

**Lemma 6.5.** *Let $X$ be an absolutely continuous random variable with pdf $f(x)$, such that the finite $p$-moment exists for some $p > 1$. Then, for any $L > 0$:*

$$\int_L^\infty (1 - F(x))\,\mathrm{d}x \leq \frac{1}{p-1}\,\frac{\mathbb{E}[|X|^p]}{L^{p-1}}. \tag{26}$$

*Proof.* We may compute

$$\int_L^\infty \int_x^\infty f(y)\,\mathrm{d}y\,\mathrm{d}x \leq \int_L^\infty \int_x^\infty \frac{|y|^p}{x^p} f(y)\,\mathrm{d}y\,\mathrm{d}x \leq \int_L^\infty x^{-p}\left(\int_{-\infty}^\infty |y|^p f(y)\,\mathrm{d}y\right)\mathrm{d}x = \frac{1}{p-1}\frac{1}{L^{p-1}}\mathbb{E}[|X|^p]$$

as required. □

A similar computation gives a bound for the integral on the interval $(-\infty, -L)$ for the distribution function $F(x)$. We can therefore apply this inequality to the difference $w(x)$.

**Lemma 6.6.** *Let $X$, $Y$ be random variables with pdfs $f(x)$, $g(x)$ and distribution functions $F(x)$ and $G(x)$, respectively, such that their finite $p$-moments exist for some $p > 1$. Then, there exists a constant $C$, depending on $p$ and the $p$-th moments of $X$ and $Y$, such that*

$$\int_{\mathbb{R}\setminus[-L,L]} |w(x)|\,\mathrm{d}x \leq C\frac{1}{L^{p-1}} \tag{27}$$

*where $w(x) = F(x) - G(x)$*

*Proof.* On the one hand, it is sufficient to compute

$$\int_L^\infty |w(x)|\,\mathrm{d}x = \int_L^{+\infty} |(1 - F(x)) - (1 - G(x))|\,\mathrm{d}x \leq \frac{1}{p-1}\frac{\mathbb{E}[|X|^p + |Y|^p]}{L^{p-1}}$$

Since a similar bound holds in the negative semi-axes, the thesis follows. □

### *6.1. An algorithm for the computation of Wasserstein-1 distance for EPT functions*

How to compute the quantity in (25)? Our approach will be to consider the difference $w(x) = F(x) - G(x)$ and the sign-changing zeros of $w(x)$. A nice feature of EPT functions is that they form an algebra; moreover, the cumulative density function $F(x)$ of an EPT pdf $f(x) = ce^{Ax}b$ is again an EPT function, with CDF $F(x) = 1 + cA^{-1}e^{Ax}b$. It follows that for EPT functions $F$ and $G$, the difference $w(x)$ is an EPT function as well.

**Remark 6.7.** *Notice that an EPT function, defined in* (11)*, is an entire function, hence all zeros of $w$ are isolated and $w$ is a continuous function. In particular, on any interval $[a, b]$, there will be at most finitely many sign-changing zeros.*

Let us assume that there are at most a finite number of sign-changing zeros of $w$, say $\xi_1, \ldots, \xi_n$, on the whole positive half-line, and that these have been identified using the GBF method. We use the convention $\xi_0 = 0$ and $\xi_{n+1} = +\infty$. Then, we can compute

$$\int_0^\infty |F(x) - G(x)|\,\mathrm{d}x = \sum_{j=0}^{n} \int_{\xi_j}^{\xi_{j+1}} (F(x) - G(x))\,\mathrm{sgn}\,(F(x) - G(x))\;\mathrm{d}x.$$

On each of the intervals $(\xi_j, \xi_{j+1})$, the sign is constant so it can be taken out of the integral. Further, integrals of the form

$$\int_{\xi_j}^{\xi_{j+1}} (F(x) - G(x))\, \mathrm{d}x$$

can be calculated explicitly in the class of EPT functions. In fact, we have

$$\int_{\xi_j}^{\xi_{j+1}} (F(x) - G(x))\, \mathrm{d}x = c_1 A_1^{-2} \left(e^{A_1 \xi_{j+1}} - e^{A_1 \xi_j}\right) b_1 - c_2 A_2^{-2} \left(e^{A_2 \xi_{j+1}} - e^{A_2 \xi_j}\right) b_2.$$

Since $A_1$ as well as $A_2$ are continuous-time asymptotically stable, meaning that their spectra satisfy $\sigma(A_i) \subset \{z \in \mathbb{C} : \Re(z) < 0\}$, it follows that

$$\int_{\xi_n}^{\xi_{n+1}=+\infty} (F(x) - G(x))\, \mathrm{d}x = -c_1 A_1^{-2} e^{A_1 \xi_n} b_1 + c_2 A_2^{-2} e^{A_2 \xi_n} b_2.$$

So, we obtain the following formula for the Wasserstein-1 distance between $\mu$ and $\nu$:

$$W(\mu, \nu) = \sum_{j=0}^{n-1} \left| c_1 A_1^{-2} \left(e^{A_1 \xi_{j+1}} - e^{A_1 \xi_j}\right) b_1 - c_2 A_2^{-2} \left(e^{A_2 \xi_{j+1}} - e^{A_2 \xi_j}\right) b_2 \right| + \left| c_1 A_1^{-2} e^{A_1 \xi_n} b_1 - c_2 A_2^{-2} e^{A_2 \xi_n} b_2 \right|, \quad (28)$$

where again we set $\xi_0 = 0$.

**Example** Let us consider the following pdfs (see their graphs in Figure 2):

$$f(x) = \frac{\pi}{4} \sin\left(\frac{\pi}{2} x\right), \qquad g(x) = \frac{45}{14} x - \frac{165}{28} x^2 + \frac{30}{7} x^3 - \frac{15}{14} x^4, \qquad 0 \le x \le 2.$$

It is possible to prove that both $f$ and $g$ are EPT functions, with representation, respectively,

$$f(x) \longrightarrow c_1 = \begin{pmatrix} \frac{\pi}{4} & 0 \end{pmatrix},\ A_1 = \begin{pmatrix} 0 & \frac{\pi}{2} \\ -\frac{\pi}{2} & 0 \end{pmatrix},\ b_1 = \begin{pmatrix} 1 & 0 \end{pmatrix}^\top$$

$$g(x) \longrightarrow c_2 = \begin{pmatrix} 0 & \frac{45}{14} & -\frac{165}{14} & \frac{180}{7} & -\frac{180}{7} \end{pmatrix},\ A_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix},\ b_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \end{pmatrix}^\top$$

We compute the Wasserstein distance between the CDFs $F$ and $G$ by using formula (28) (in this example, $\xi_0 = 0$, $\xi_1 = 1$, and $\xi_2 = 2$, while the last term in the formula vanishes since we are considering a finite interval):

$$W(F, G) = \left| c_1 A_1^{-2} \left(e^{A_1 \xi_1} - e^{A_1 \xi_0}\right) b_1 - c_2 A_2^{-2} \left(e^{A_2 \xi_1} - e^{A_2 \xi_0}\right) b_2 \right| + \left| c_1 A_1^{-2} \left(e^{A_1 \xi_2} - e^{A_1 \xi_1}\right) b_1 - c_2 A_2^{-2} \left(e^{A_2 \xi_2} - e^{A_2 \xi_1}\right) b_2 \right| = 2\left(\frac{1}{\pi} - \frac{31}{112}\right)$$

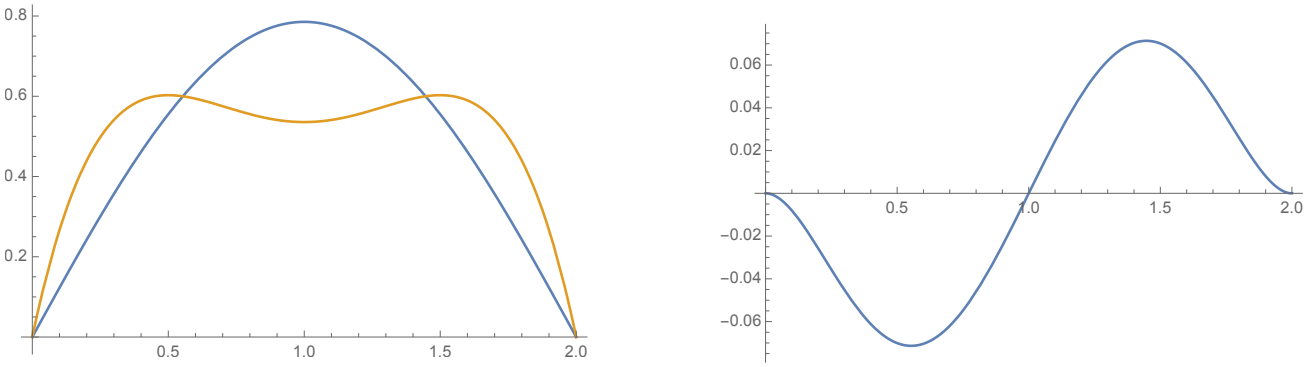


**Figure 2.** Graph of the density functions $f(x)$ and $g(x)$ (on the left) and of the function $w(x) = F(x) - G(x)$ (on the right) on the interval $[0, 2]$.

*6.2. An algorithm for the computation of Wasserstein-1 distance for Gaussian mixtures*

Here we ask the same question: *how do we compute the quantity in* (25) in the case of finite Gaussian mixtures? Again, the two crucial issues are:

- to compute the sign-changing zeros of the difference of the CDFs involved;
- to calculate the integrals of the difference of the CDFs involved over intervals on which the sign of the difference does not change.

In this case, both the distribution functions involved and their differences are of the form

$$w(x) = \sum_{i=1}^{K} \mu_i \Phi(\alpha_i x + \beta_i),$$

where $\Phi$ is the standard Gaussian distribution function.

A GBF sequence is obtained for $w$ by taking $w_0(x) = w(x)$, $w_1(x) = Dw(x)$. Then, $w_1(x)$ is a function of the form

$$w_1(x) = \sum_{i=1}^{K} \mu_i \alpha_i \phi(\alpha_i x + \beta_i),$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.
We have already developed a GBF sequence for $w_1$ in Section 5, say $\psi_0 = w_1, \psi_1, \ldots, \psi_K$. Then, a GBF sequence for $w$ is given by $w_0 = w, w_1, \psi_1, \ldots, \psi_K$.

**Lemma 6.8.** *A Gaussian mixture* $f = \sum_{i=1}^{K} \delta_i \phi(\alpha_i x + \beta_i)$*, and also* $w$*, have at most finitely many sign-changing zeros, and an interval can be identified containing all these.*

*Proof.* We emphasize that this result holds for generalized Gaussian mixtures with a similar proof. We aim to prove that there exists $x_0 > 0$ such that, for all $x > x_0$, $f(x) \neq 0$. Notice that

$$f(x) = \sum_{i=1}^{K} \delta_i \phi(\alpha_i x + \beta_i) = \frac{1}{\sqrt{2\pi}} \sum_{i=1}^{K} \delta_i e^{-\beta_i^2/2} e^{-q_i(x)}, \qquad q_i(x) = \frac{\alpha_i^2}{2} x^2 + \alpha_i \beta_i x$$

can be ordered in such a way that

- $\alpha_i < \alpha_{i+1}$, or
- $\alpha_i = \alpha_{i+1}$ and $\beta_i < \beta_{i+1}$

(a further equality cannot hold, since otherwise the two polynomials would have been equal). It follows that

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-q_1(x)} \left( \delta_1 e^{-\beta_1^2/2} + \sum_{i=2}^{K} \delta_i e^{-\beta_i^2/2} e^{q_1(x)-q_i(x)} \right)$$

has the property that the polynomial $q_1 - q_i$ has the leading (nonzero) coefficient *negative* for each $i = 2, \ldots, K$: therefore, for each $i = 2, \ldots, K$, there exists $x_i$ such that $\left|\delta_i e^{q_1(x)-q_i(x)}\right| \leq \frac{1}{2K} |\delta_1|$ for all $x > x_i$. Setting $x_0 = \max(x_2, \ldots, x_K)$, we obtain the claim.

Now we claim that it is possible to find $y_0 < 0$ such that, for all $x < y_0$, $f(x) \neq 0$. The proof is the same as before, except that we shall modify the ordering of the polynomials to take into account the correct behavior of the first-order term. Actually, we choose

- $\alpha_i < \alpha_{i+1}$, or
- $\alpha_i = \alpha_{i+1}$ and $\beta_i > \beta_{i+1}$:

then the proof follows with the same argument as above. □

To conclude this section, we recall that the integral of the Gaussian CDF can be computed explicitly in terms of the Gaussian distribution itself. The starting point is the following well-known identity:

$$\int_{-\infty}^{x} \Phi(y)\, \mathrm{d}y = x\Phi(x) + \phi(x), \tag{29}$$

where $\Phi$ is the CDF and $\phi$ the pdf for the standard Gaussian distribution. Then, from (29) it follows that

$$\begin{aligned} \int_a^b \Phi(\alpha x + \beta)\, \mathrm{d}x &= \int_{\alpha a+\beta}^{\alpha b+\beta} \frac{1}{\alpha} \Phi(y)\, \mathrm{d}y \\ &= \left(b + \frac{\beta}{\alpha}\right) \Phi(\alpha b + \beta) - \left(a + \frac{\beta}{\alpha}\right) \Phi(\alpha a + \beta) + \frac{1}{\alpha} \left[\phi(\alpha b + \beta) - \phi(\alpha a + \beta)\right]. \end{aligned}$$

## 7. Conclusions and further research

In this paper, we explored finite mixture models, focusing our efforts on two specific subclasses of the functions introduced in (1): Gaussian mixtures and EPT functions, which hold great significance in probability theory. We hope to revisit the whole class of general mixtures defined in Eq (1) in a subsequent paper.

The software [5] developed along this paper allows the investigation of the number of zero crossings of a finite Gaussian mixture. Specifically, for a given finite Gaussian mixture (defined by its means, variances, and mixing coefficients), the software provides a bounded interval where any sign-changing roots might occur. If such roots are present, the software locates them with the desired level of accuracy. In several applications there will be the necessity of suitable modifications of this software, which is constructed in such a way to be easily adapted to different settings. It is our intention to extend its current limit of elements in the mixture, which may require a different approach to the algorithm.

Furthermore, our results are explicitly formulated for the one-dimensional case. While this suffices for many applications, it poses limitations in certain contexts. Therefore, extending our results to the $d$-dimensional case is a crucial area for future investigation.

**Ackowledgements**

Stefano Bonaccorsi is a member of the group Gnampa–Indam, whose support is gratefully recognized. Bernard Hanzon acknowledges that part of this research resulted from the SFI projects 07/MI/008 and RFP2007-MATF802. Giulia Lombardi is also a member of the Gnampa–Indam group, whose support is highly recognized, and also acknowledges that part of this research resulted from the European Union - FSE REACT-EU, PON Ricerca e Innovazione 2014-2020.

**Author contributions**

Stefano Bonaccorsi and Bernard Hanzon: conceptualization, formal analysis, validation, writing; Giulia Lombardi: conceptualization, formal analysis, software, visualization, validation, writing. All authors have read and agreed to the published version of the manuscript.

**Conflict of interest**

Professor Stefano Bonaccorsi is an editorial board member for AIMS Mathematics and was not involved in the editorial review or the decision to publish this article. All authors declare that there are no competing interests.

## A. Technical results and proofs

### *A.1. Some results in bordered matrices*

Let

$$M = \begin{pmatrix} \tilde{M} & c_1 & c_2 \\ b_1^T & a_1 & a_2 \\ b_2^T & a_3 & a_4 \end{pmatrix}$$

be a square $n \times n$ matrix, partitioned in such a way that

- $\tilde{M}$ is a $(n-2) \times (n-2)$ square matrix,
- $c_1$, $c_2$, $b_1$, $b_2$ are column vectors of length $n-2$, and
- $a_1, \ldots, a_4$ are numbers.

Then, we can compute the determinant of $M$, and we have the following rule (see, e.g., [20]):

$$\det(M)\det(\tilde{M}) = \begin{vmatrix} \begin{vmatrix} \tilde{M} & c_1 \\ b_1^T & a_1 \end{vmatrix} & \begin{vmatrix} \tilde{M} & c_2 \\ b_1^T & a_2 \end{vmatrix} \\ \\ \begin{vmatrix} \tilde{M} & c_1 \\ b_2^T & a_3 \end{vmatrix} & \begin{vmatrix} \tilde{M} & c_2 \\ b_2^T & a_4 \end{vmatrix} \end{vmatrix} \tag{A.1}$$

and, if we set

$$A' = \begin{pmatrix} \tilde{M} & c_1 \\ b_1^T & a_1 \end{pmatrix}, \qquad B' = \begin{pmatrix} \tilde{M} & c_2 \\ b_1^T & a_2 \end{pmatrix}, \qquad C' = \begin{pmatrix} \tilde{M} & c_1 \\ b_2^T & a_3 \end{pmatrix}, \qquad D' = \begin{pmatrix} \tilde{M} & c_2 \\ b_2^T & a_4 \end{pmatrix}$$

we obtain

$$\det(M)\det(\tilde{M}) = \det(A')\det(D') - \det(B')\det(C'). \tag{A.2}$$

*A.2. Proof of theorem 5.2*

Relation (21) follows directly from the definition

$$Q_{i,k}^1 = \begin{vmatrix} 1 & 1 \\ P_{1,i} & P_{1,k} \end{vmatrix}$$

and the definition (18).

Next, we consider the case of $j > 2$. With a slight abuse of notation, we denote by

$$Q_{i_1,\dots,i_j}^{j-1}(y_1, \dots, y_j)$$

the determinant in (19) where in each column we consider a different variable $y_i$. Notice that we recover the polynomial in $x$ when we let $y_1 = \cdots = y_j = x$:

$$Q_{i_1,\dots,i_j}^{j-1}(x) = Q_{i_1,\dots,i_j}^{j-1}(x, \dots, x).$$

Let $D_i$ denote the differentiation with respect to the variable $y_i$. Then, the following formula holds (chain rule),

$$D_x Q_{i_1,\dots,i_j}^{j-1}(x) = [D_1 + \cdots + D_j] Q_{i_1,\dots,i_j}^{j-1}(x, \dots, x). \tag{A.3}$$

Finally, we introduce the notation

$$\Delta_j = \phi_j + D_j,$$

where $\phi_j$, defined in (18), is taken as a multiplication operator. By the results in Remark 5.1, we know that

$$\Delta_j P_{m,j} = P_{m+1,j}$$

and

$$\left[\sum_{i=1}^{j} \Delta_i\right] Q_{i_1,\dots,i_j}^{j-1} = \begin{vmatrix} 1 & 1 & \dots & 1 \\ P_{1,i_1} & P_{1,i_2} & \dots & P_{1,i_j} \\ & \dots & & \dots \\ P_{j-2,i_1} & P_{j-2,i_2} & \dots & P_{j-2,i_j} \\ P_{j,i_1} & P_{j,i_2} & \dots & P_{j,i_j}. \end{vmatrix} \tag{A.4}$$

Now, let us write $Q^j_{1,\dots,j,k}$ as the determinant of a bordered matrix in the following form

$$Q^j_{1,\dots,j,k} = \left|\begin{array}{ccc|c|c} 1 & \dots & 1 & 1 & 1 \\ P_{1,1} & \dots & P_{1,j-1} & P_{1,j} & P_{1,k} \\ & \dots & & & \dots \\ P_{j-2,1} & \dots & P_{j-2,j-1} & P_{j-2,j} & P_{j-2,k} \\ \hline P_{j-1,1} & \dots & P_{j-1,j-1} & P_{j-1,j} & P_{j-1,k} \\ \hline P_{j,1} & \dots & P_{j,j-1} & P_{j,j} & P_{j,k} \end{array}\right|.$$

Our aim is to apply formula (A.2). It is clear that

$$\det(M) = Q^j_{1,\dots,j,k}$$

as well as

$$\det(\tilde{M}) = Q^{j-2}_{1,\dots,j-1};$$

it remains to identify the other terms. Recall the definition of matrices $A', B', C'$, and $D'$ from the previous subsection. We have

$$A' = \left|\begin{array}{ccc|c} 1 & \dots & 1 & 1 \\ P_{1,1} & \dots & P_{1,j-1} & P_{1,j} \\ & \dots & & \dots \\ P_{j-2,1} & \dots & P_{j-2,j-1} & P_{j-2,j} \\ \hline P_{j-1,1} & \dots & P_{j-1,j-1} & P_{j-1,j} \end{array}\right|, \qquad B' = \left|\begin{array}{ccc|c} 1 & \dots & 1 & 1 \\ P_{1,1} & \dots & P_{1,j-1} & P_{1,k} \\ & \dots & & \dots \\ P_{j-2,1} & \dots & P_{j-2,j-1} & P_{j-2,k} \\ \hline P_{j-1,1} & \dots & P_{j-1,j-1} & P_{j-1,k} \end{array}\right|$$

$$C' = \left|\begin{array}{ccc|c} 1 & \dots & 1 & 1 \\ P_{1,1} & \dots & P_{1,j-1} & P_{1,j} \\ & \dots & & \dots \\ P_{j-2,1} & \dots & P_{j-2,j-1} & P_{j-2,j} \\ \hline P_{j,1} & \dots & P_{j,j-1} & P_{j,j} \end{array}\right|, \qquad D' = \left|\begin{array}{ccc|c} 1 & \dots & 1 & 1 \\ P_{1,1} & \dots & P_{1,j-1} & P_{1,k} \\ & \dots & & \dots \\ P_{j-2,1} & \dots & P_{j-2,j-1} & P_{j-2,k} \\ \hline P_{j,1} & \dots & P_{j,j-1} & P_{j,k} \end{array}\right|$$

hence

$$\det(A') = Q^{j-1}_{1,\dots,j}, \qquad \det(B') = Q^{j-1}_{1,\dots,j-1,k}.$$

By comparing with (A.4), we see that

$$\det(C') = \left[\sum_{i=1}^{j} \Delta_i\right] Q^{j-1}_{1,\dots,j};, \qquad \det(D') = \left[\sum_{i=1}^{j-1} \Delta_i + \Delta_k\right] Q^{j-1}_{1,\dots,j-1,k}.$$

By taking $y_1 = \dots = y_j = y_k = x$ and recalling (A.3), we finally obtain

$$Q^j_{1,\dots,j,k} = \frac{1}{Q^{j-2}_{1,\dots,j-1}} \left( Q^{j-1}_{1,\dots,j} \left[ \left( \sum_{i=1}^{j-1} \phi_i + \phi_k \right) + D_x \right] Q^{j-1}_{1,\dots,j-1,l} - Q^{j-1}_{1,\dots,j-1,k} \left[ \left( \sum_{i=1}^{j-1} \phi_i + \phi_j \right) + D_x \right] Q^{j-1}_{1,\dots,j} \right)$$

and simplifying the relevant terms we obtain the thesis. ■

*A.3. Proof of theorem 5.3*

We recall from (16) that using the Polya-Ristroph formula, we have the GBF sequence

$$\Psi_j = \frac{W_j}{W_{j-1}} D\left(\frac{W_{j-1}}{W_j}\Psi_{j-1}\right) \quad \forall\ 1 \le j \le N,$$

where for simplicity we let $W_0 = 1$; this formula can be rewritten as

$$\begin{aligned}\Psi_j =& \frac{W_j}{W_{j-1}}\left(\frac{W_j\, DW_{j-1} - W_{j-1}\, DW_j}{(W_j)^2}\Psi_{j-1} + \frac{W_{j-1}}{W_j} D\Psi_{j-1}\right)\\ =& D\Psi_{j-1} - \left(\frac{DW_j}{W_j} - \frac{DW_{j-1}}{W_{j-1}}\right)\Psi_{j-1}.\end{aligned}$$

Now, using (20) we have that

$$\frac{DW_j}{W_j} = \sum_{i=1}^{j} P_{1,i} + \frac{DQ^{j-1}_{1,\dots,j}}{Q^{j-1}_{1,\dots,j}}, \qquad 1 \le j \le N,$$

which implies that we can introduce the rational functions

$$s_1 := P_{1,1}$$

and, for $2 \le j \le N$,

$$s_j := \frac{DW_j}{W_j} - \frac{DW_{j-1}}{W_{j-1}} = \frac{DQ^{j-1}_{1,\dots,j}}{Q^{j-1}_{1,\dots,j}} - \frac{DQ^{j-2}_{1,\dots,j-1}}{Q^{j-2}_{1,\dots,j-1}} + P_{1,j} = \frac{D\left(\frac{Q^{j-1}_{1,\dots,j}}{Q^{j-2}_{1,\dots,j-1}}\right)}{\frac{Q^{j-1}_{1,\dots,j}}{Q^{j-2}_{1,\dots,j-1}}} + P_{1,j}$$

and $\Psi_j$ can be written as

$$\Psi_j(x) = (D - s_j(x))\Psi_{j-1}(x).$$

Recall that

$$\Psi_0(x) = \sum_{k=1}^{N} \lambda_k h_k(x),$$

which implies that (compare with (18))

$$\Psi_1(x) = (D - s_1(x))\Psi_0(x) = \lambda_1(D - P_{1,1})h_1 + \sum_{k=2}^{N} \lambda_k(D - P_{1,1})h_k(x) = \sum_{k=2}^{N} \lambda_k(P_{1,k} - P_{1,1})h_k(x)$$

and recalling the definition of $Q^1_{1,k}$ in (21), we obtain

$$\Psi_1(x) = \sum_{k=2}^{N} \lambda_k Q^1_{1,k} h_k(x)$$

in accordance with (23).

Now we proceed by recursion. Assume that for $2 \le j \le N$, it holds

$$\Psi_{j-1}(x) = \sum_{k=j}^{N} \lambda_k \frac{Q^{j-1}_{1,\dots,j-1,k}}{Q^{j-2}_{1,\dots,j-1}} h_k(x).$$

Then,

$$\begin{aligned}\Psi_j(x) =&(D - s_j(x))\Psi_{j-1}(x) = \sum_{k=j}^{N} \lambda_k \left( D \frac{Q^{j-1}_{1,\dots,j-1,k}}{Q^{j-2}_{1,\dots,j-1}} \right) h_k(x) \\ &+ \sum_{k=j}^{N} \lambda_k \frac{Q^{j-1}_{1,\dots,j-1,k}}{Q^{j-2}_{1,\dots,j-1}} Dh_k(x) - s_j \sum_{k=j}^{N} \lambda_k \frac{Q^{j-1}_{1,\dots,j-1,k}}{Q^{j-2}_{1,\dots,j-1}} h_k(x).\end{aligned}$$

Let us denote

$$\Lambda^{j-1}_k = \frac{Q^{j-1}_{1,\dots,j-1,k}}{Q^{j-2}_{1,\dots,j-1}}, \qquad k \ge j;$$

then we have

$$\begin{aligned}\Psi_j(x) &= \sum_{k=j}^{N} \lambda_k \left( D\Lambda^{j-1}_k \right) h_k(x) + \sum_{k=j}^{N} \lambda_k \Lambda^{j-1}_k \, P_{1,k} \, h_k(x) \\ &\quad - \frac{D\Lambda^{j-1}_j}{\Lambda^{j-1}_j} \sum_{k=j}^{N} \lambda_k \Lambda^{j-1}_k h_k(x) - P_{1,j} \sum_{k=j}^{N} \lambda_k \Lambda^{j-1}_k h_k(x) \\ &= \sum_{k=j+1}^{N} \lambda_k \left( D\Lambda^{j-1}_k - \frac{D\Lambda^{j-1}_j}{\Lambda^{j-1}_j} \Lambda^{j-1}_k \right) h_k(x) \\ &\quad + \sum_{k=j+1}^{N} \lambda_k \Lambda^{j-1}_k \left( P_{1,k} - P_{1,j} \right) h_k(x) \\ &= \sum_{k=j+1}^{N} \lambda_k \, \Lambda^{j-1}_k \left( \frac{D\Lambda^{j-1}_k}{\Lambda^{j-1}_k} - \frac{D\Lambda^{j-1}_j}{\Lambda^{j-1}_j} + \left( P_{1,k} - P_{1,j} \right) \right) h_k(x).\end{aligned}$$

In order to prove (23), it is sufficient to prove that for any $k \ge j + 1$, it holds

$$\Lambda^{j-1}_k \left( \frac{D\Lambda^{j-1}_k}{\Lambda^{j-1}_k} - \frac{D\Lambda^{j-1}_j}{\Lambda^{j-1}_j} + \left( P_{1,k} - P_{1,j} \right) \right) = \frac{Q^{j}_{1,\dots,j,k}}{Q^{j-1}_{1,\dots,j}}. \tag{A.5}$$

We will repeatedly use the identity

$$\frac{Df}{f} - \frac{Dg}{g} = \frac{D(f/g)}{f/g}$$

to compute the left-hand side of the claim (A.5): we get

$$\begin{aligned}
&\frac{Q^{j-1}_{1,\dots,j-1,k}}{Q^{j-2}_{1,\dots,j-1}}\left(\frac{D\frac{Q^{j-1}_{1,\dots,j-1,k}}{Q^{j-2}_{1,\dots,j-1}}}{\frac{Q^{j-1}_{1,\dots,j-1,k}}{Q^{j-2}_{1,\dots,j-1}}}-\frac{D\frac{Q^{j-1}_{1,\dots,j-1,j}}{Q^{j-2}_{1,\dots,j-1}}}{\frac{Q^{j-1}_{1,\dots,j-1,j}}{Q^{j-2}_{1,\dots,j-1}}}+\left(P_{1,k}-P_{1,j}\right)\right)\\
=&\frac{Q^{j-1}_{1,\dots,j-1,k}}{Q^{j-2}_{1,\dots,j-1}}\left(\frac{D\frac{Q^{j-1}_{1,\dots,j-1,k}}{Q^{j-1}_{1,\dots,j-1,j}}}{\frac{Q^{j-1}_{1,\dots,j-1,k}}{Q^{j-1}_{1,\dots,j-1,j}}}+\left(P_{1,k}-P_{1,j}\right)\right)\\
=&\frac{Q^{j-1}_{1,\dots,j-1,k}}{Q^{j-2}_{1,\dots,j-1}}\left(\frac{DQ^{j-1}_{1,\dots,j-1,k}}{Q^{j-1}_{1,\dots,j-1,k}}-\frac{DQ^{j-1}_{1,\dots,j-1,j}}{Q^{j-1}_{1,\dots,j-1,j}}+\left(P_{1,k}-P_{1,j}\right)\right)
\end{aligned}$$

Recalling (22), the quantity within bracket equals to

$$\frac{Q^{j}_{1,\dots,j,k}\,Q^{j-2}_{1,\dots,j-1}}{Q^{j-1}_{1,\dots,j-1,k}\,Q^{j-1}_{1,\dots,j-1,j}}$$

which allows us to prove (A.5). ∎

## B. Examples

### *B.1. An example for Polynomial-Gaussian mixtures*

In this example, we study the behavior of the sum of two polynomial-Gaussian functions $f(x) = h_1(x) + \alpha h_2(x)$, with $\alpha \in \mathbb{R}$. In order to provide a complete example, we fix

$$h_1(x) = (x^2+1)e^{-x^2}, \qquad h_2(x) = (x^2+4x)e^{-(x-1)^2/2}$$

so that $\mu_1 = 0$, $\sigma_1^2 = \frac{1}{2}$, $\mu_2 = 1$, $\sigma_2^2 = 1$.

The first step is the choice of the interval $[a, b]$. Since the second term has a larger variance, it is the dominant term for $|x| \to \infty$; hence, for $\alpha > 0$, we have $f(x) > 0$ for $|x| \to \infty$, and conversely for $\alpha < 0$. Therefore, we can always choose $a \ll 0$ and $b \gg 0$ such that $f(x)$ has constant sign (equal to $\mathrm{sgn}(\alpha)$) outside the interval $[a, b]$. As $\alpha$ varies in $(-10, 10)$, for instance, it is sufficient to take $a \le -5$, $b \ge 5$.

The next step is the computation of the function $\psi_1(x)$. Notice that this function (and, a fortiori, the simple grid it provides for $f(x)$) is independent of the constant $\alpha$. We obtain

$$\psi_1(x) = h_{1,2}(x) = Dh_2(x) - h_2(x)\frac{Dh_1(x)}{h_1(x)} = \frac{1}{1+x^2}e^{-(x-1)^2/2}(4+6x+x^2+5x^3+5x^4+x^5)$$

so that a simple grid for $f(x)$ is given by the real sign-changing zeros of the polynomial

$$p_{1,2}(x) = 4+6x+x^2+5x^3+5x^4+x^5,$$

and we obtain the following grid:

$$G = \{x_1 = -3.57116, x_2 = -1.72866, x_3 = -0.638509\}.$$

Finally, the existence of sign-changing zeros for $f(x)$ can be determined by computing the values of $f(x)$ in the points of $G \cup \{a, b\}$, which turns out to depend linearly on $\alpha$. Thus, we are able to construct the table in Figure A1, where the analysis of the grid (and the presence of sign-changing zeros) is shown.

According to the table, for $\alpha = 0.5$, no sign-changing zeros are present for the function $f_{0.5}(x) = h_1(x) + \frac{1}{2}h_2(x)$, while for $\alpha = 2$ there exist 4 sign-changing zeros for the function $f_2(x) = h_1(x) + 2h_2(x)$. Their plots are given in Figure A2.

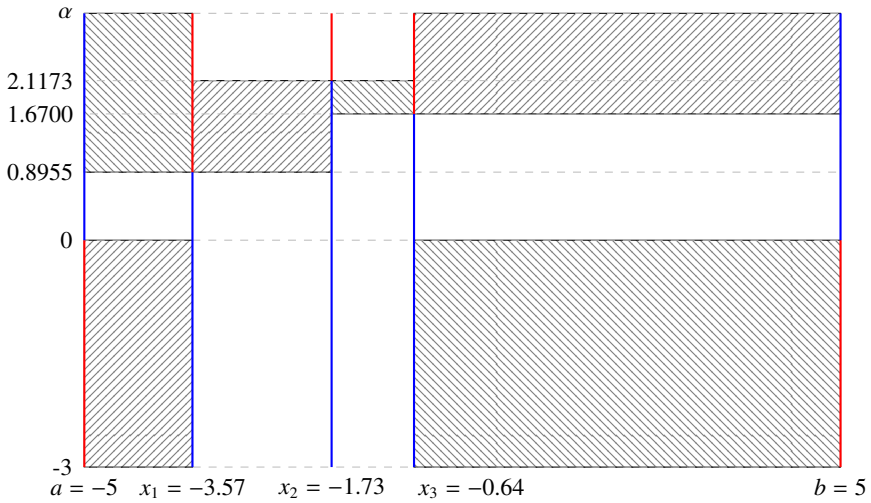


**Figure A1.** Analysis of the grid for the function $f(x)$. The parameter $\alpha$ varies in $(-3, 3)$ and $x$ in $(-5, 5)$. The orientation of the lines shows if the function is increasing or decreasing at the zero. The blue line implies that the function takes a positive value in the point of the grid, and red it is negative. No further zeros exist outside the interval $[a, b]$. Notice that for $0 < \alpha < 0.8955$, the function $f(x)$ is nonnegative on the whole real line.

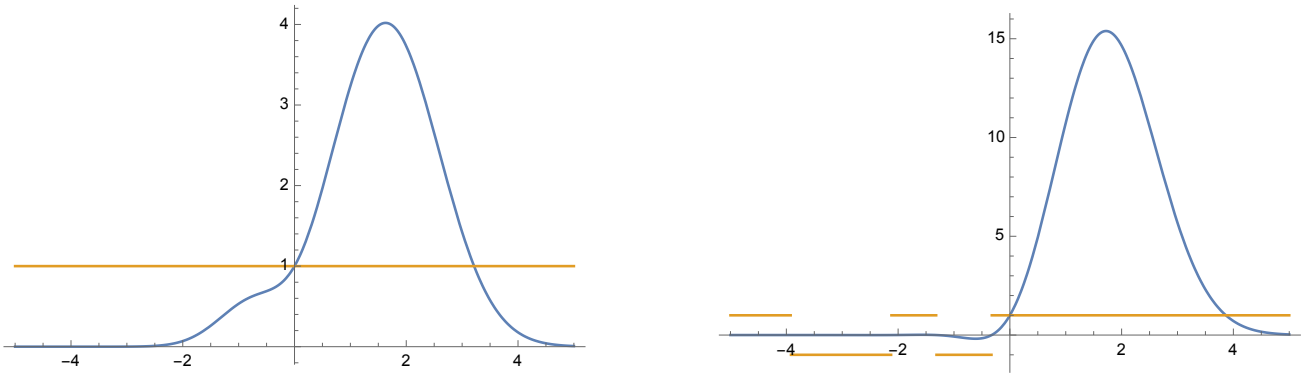


**Figure A2.** Graph of the functions $f_{0.5}(x)$ (on the left) and $f_2(x)$ (on the right). The orange line represents the sign of the function, so to help capture the sign-changes.

### *B.2. An example for finite Gaussian mixtures*

In this example, we show how the GBF algorithm computes the sign-changing roots of a Gaussian mixture with parameters

- n = 4;
- mu = [0, 1, 2, 3];
- s = [1, 0.07, 0.06, 0.8];
- lambda = [1, −0.5, −0.5, 1].

and the accuracy level for the Ridders' method set equal to eps = 2.2204e − 16. The main steps executed by the software are described as follows:

- Step 1: Computation of the GBF sequence;
- Step 2: Computation of the bounded interval $I$ which contains all the eventual sign-changing roots.
- Step 3: Iterative backward scanning of the GBF sequence embedded with the computation of grids and sign-changing roots for all levels.

We mention that the implemented version of the GBF algorithm was slightly modified from its theoretical version to reduce the execution times of the sign-changing roots algorithm. In fact, in Step 3, at each level the corresponding grid-induced partition is expanded by incorporating both previous grid points and current ones to reduce the lengths of the intervals in which the Ridders' method is applied. To capture the key idea behind this augmentation technique, we will refer to this backward recursive augmentation of grid-induced partitions as GBF cascade.

It should be noted that the version of the GBF algorithm we implemented was slightly modified from its theoretical version to reduce the execution times of the sign-changing roots algorithm. Specifically, in Step 3, at each level $j$ where $0 \leq j \leq n-1$, the partition induced by the corresponding grid is enlarged by integrating both the previous grid points and the ones from the current level to reduce the lengths of the intervals where the Ridders's method is applied. To capture the essence of this expansion strategy, the backward recursive augmentation of grid-induced partitions is referred to as the GBF cascade effect.

In the current example, the bounded interval was $I = [0.4903, 2.8369]$. The GBF procedure employed to compute the sign-changing roots of the Gaussian mixture is displayed in Figure A3. The peculiar form of this figure is meant to emphasize the GBF cascade effect and the construction of the grid, as well as the determination of the sign-changing zeros at the last level.

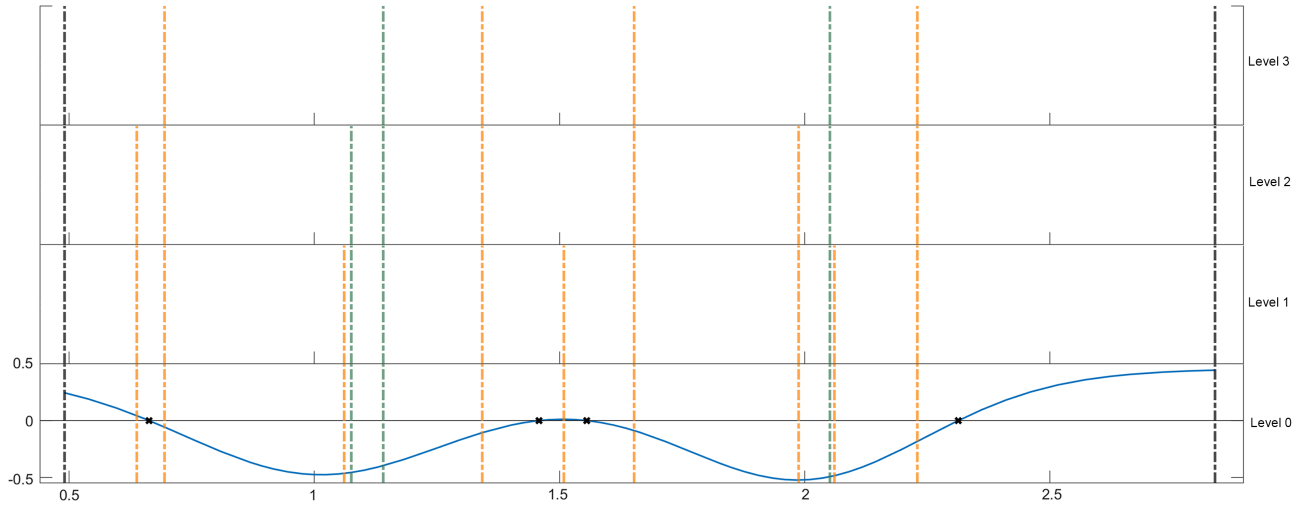


**Figure A3.** GBF procedure for the Gaussian mixture discussed in appendix B.2. Endpoints of the interval $I$ are represented as black dashed vertical lines. Grid-induced partitions are displayed as orange and green dashed vertical lines to distinguish between points stemming from zero-crossings of the GBF functions and zeros of the pivot functions, respectively. The Gaussian mixture is plotted in blue and its sign-changing roots as black x markers. The plot shows how the GBF cascade effect from level 3 down to level 0 leads to the identification of the sign-changing roots. More specifically, we observe that a grid-induced partition at a specific level $j$ is used to determine the subsequent one at level $j-1$ until the bottom level is reached and the sign-changing roots are computed.

*B.3. Numerical examples with the Chebfun package*

For a more comprehensive discussion on Chebfun, please refer to [10], [35]. In this paper, we aim to elucidate the limitations of the Chebfun method when applied to the analysis of finite Gaussian mixtures. Our primary focus is on the poor approximation of tail behavior, or, more broadly, regions we term as valleys.

A valley is defined as an interval within the function where values are minimal and exhibit negligible variation, remaining close to zero. The accurate approximation of these valleys is crucial for achieving reliable numerical approximations of finite Gaussian mixtures. In fact, sign-changing patterns often reside within valleys and manifest at a microscopic scale that is typically undetectable by standard algorithms.

In what follows we will show, through a series of examples, how Chebfun,Äôs approximation falls short when it comes to accurately representing valleys and, hence, the sign-changing roots within them.

**Example 1**

i. Consider a finite Gaussian mixture $f$ having the following parameters

$\mathtt{n} = 4$;
$\mathtt{mu} = [2, 3, 4, 5]$;
$\mathtt{s}^2 = [0.4, 0.6, 0.8, 1]$;
$\mathtt{lambda} = [0.2, 0.4, -0.2, 0.6]$.
accuracy level equal to $\mathtt{eps} = 10^{-12}$.

ii. The GBF algorithm detects no sign-changing roots on the real line;
iii. Chebfun with input interval $[-10, 20]$ identifies 69 roots: 42 negative within the interval $[-10, -3.6224]$ and 27 positive within $[13.7823, 20]$.

This discrepancy is due to the methodological choice of Chebfun, which computes a polynomial approximation of degree 154 to represent the overall Gaussian mixture. The plot of $f$ along with the roots identified by Chebfun are shown in Figure A4.

To avoid the poor approximation of the tails, one might consider reducing the width of the input interval. However, as shown in the following example, this could result in the loss of roots that lie along the tails.

**Example 2**

i. Consider a finite Gaussian mixture $f$ having the following parameters:

$\mathtt{n} = 4$;
$\mathtt{mu} = [2, 4, 6, 8]$;
$\mathtt{s}^2 = [0.2, 0.4, 0.6, 0.8]$;
$\mathtt{lambda} = [0.2, 0.4, 0.6, -0.2]$.
accuracy level equal to $\mathtt{eps} = 10^{-12}$.

ii. The GBF algorithm detects 2 sign-changing roots on the real line, specifically $x_1 = -7.3460$, $x_2 = 7.3460$ within the interval $I = [-7.7857, 7.7857]$;

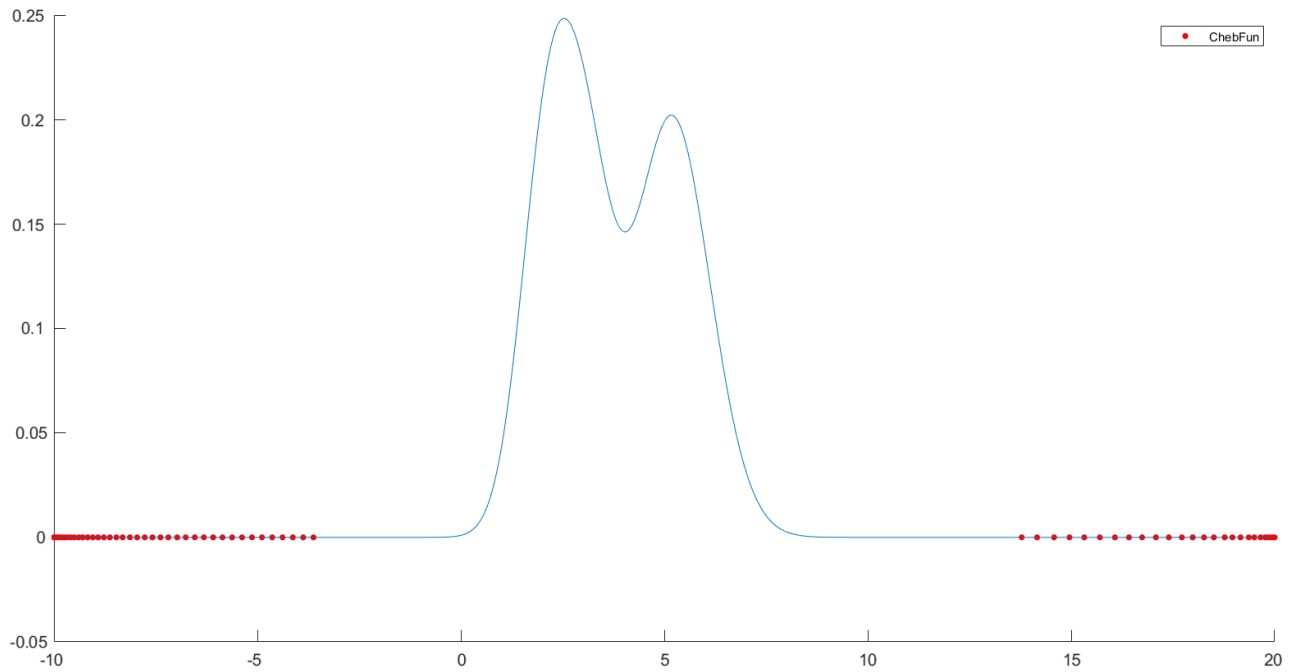


**Figure A4.** Plot of the Gaussian mixture $f$ from Example 1 in $[-10, 20]$. The roots identified by Chebfun are shown in red.

iii. Chebfun with input interval $[-10, 20]$ identifies 136 roots: 135 negative within $[-20, -1.8294]$ and one positive. Among them, one $(-7.3131)$ is very close to $x_1$ and the positive one coincides with $x_2$.

Again, the discrepancy lies in Chebfun's polynomial approximation, which is of degree 251, to represent the overall Gaussian mixture. The Gaussian mixture $f$ along with the zeros identified by Chebfun and the GBF algorithm respectively are shown in A5.

Hence, by narrowing the input interval to avoid the poor approximation of the left tail, there is a high risk of missing the correct sign-changing zero $x_1$.

However, valleys could also arise in the central body of the function, as shown in the next example.

### Example 3

i. Consider a finite Gaussian mixture $f$ having the following parameters:

$\mathtt{n} = 5;$
$\mathtt{mu} = [5, 2, 4, 6, 16];$
$\mathtt{s}^2 = [1, 1/50, 1/25, 1/25, 1/50];$
$\mathtt{lambda} = [-1.5, 2, 0.5, -0.5, 0.5].$
accuracy level equal to $\mathtt{eps} = 10^{-12}$.

ii. The GBF algorithm detects 6 sign-changing roots on the real line, specifically $x_1 = 1.4106$, $x_2 = 2.467$, $x_3 = 3.6651$, $x_4 = 4.2515, x_5 = 14.6261, x_6 = 17.8229$ within the interval $I = [1.35948, 17.8405]$;

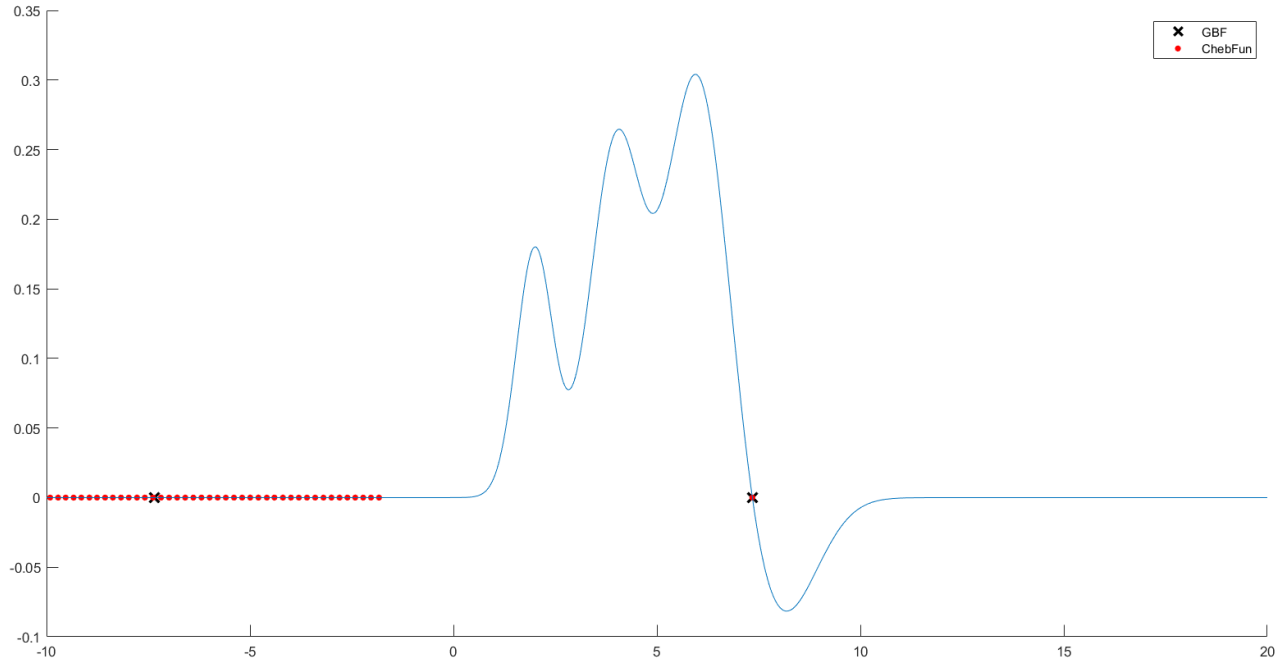


**Figure A5.** Plot of the Gaussian mixture $f$ from Example 2 in $[-10, 20]$. The roots identified by Chebfun are shown in red, while sign-changing roots identified by the GBF algorithm are indicated by black x markers.

iii. Chebfun with input interval $[-10, 20]$ identifies 247 roots: 162 negative within $[-20, -3.0649]$ and 85 positive within $[1.4106, 20]$ based on a polynomial of degree 862. Among them, only $x_1$, $x_2$, $x_3$, and $x_4$ are correctly identified. The roots $x_5$ and $x_6$ are located in a valley region, which aligns with the left tail and right tail of the last component of the mixture, respectively. Within this region, Chebfun identifies 29 roots in the proximity of $x_5$ and 52 in the proximity of $x_6$. This high density of roots makes it impractical to correctly identify the roots of the finite Gaussian mixture.

The Gaussian mixture $f$ along with the zeros identified by Chebfun and the GBF algorithm, respectively, are shown in A6.

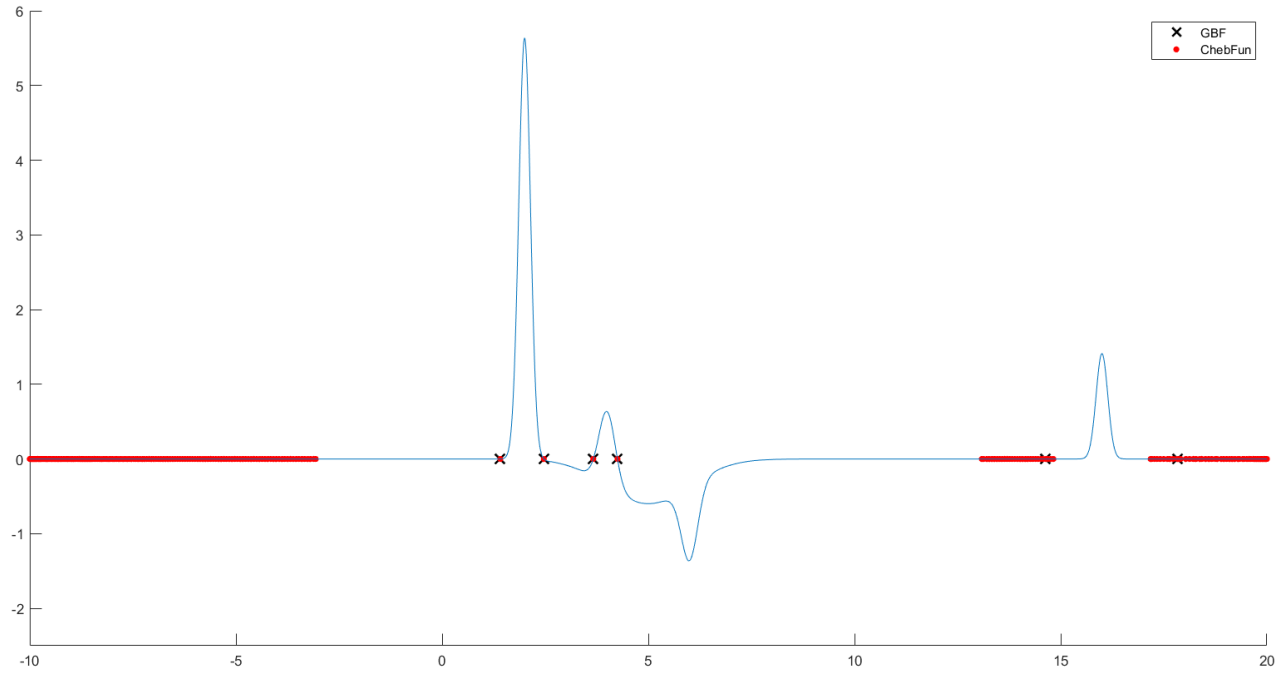


**Figure A6.** Plot of the Gaussian mixture $f$ from Example 3 in $[-10, 20]$. The roots identified by Chebfun are shown in red, while sign-changing roots identified by the GBF algorithm are indicated by black x markers.

## References


1. H. Albrecher, D. Finger, P. O. Goffard, *Empirical risk analysis of mining a proof of work blockchain*, 2023. Available from: `https://www.researchgate.net/publication/375082859`
2. L. Ambrosio, N. Gigli, G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, (2008).
3. L. Aquilanti, S. Cacace, F. Camilli, R. De Maio, A mean field games approach to cluster analysis, *Appl. Math. Optim.*, **84** (2021), 299–323. https://doi.org/10.1007/s00245-019-09646-2
4. L. Aquilanti, S. Cacace, F. Camilli, R. De Maio, A mean field games model for finite mixtures of Bernoulli and categorical distributions, *J. Dyn. Games*, **8** (2021), 35–59. https://doi.org/10.3934/jdg.2020033
5. S. Bonaccorsi, B. Hanzon, G. Lombardi, *GBF-algorithm*, 2024. Available from: `https://github.com/giuliaelle/GBF-algorithm`
6. V. Chonev, J. Ouaknine, J. Worrell, On the zeros of exponential polynomials, *J. ACM*, **70** (2023), 1–26. https://doi.org/10.1145/3603543
7. P. Collins, Computable analysis with applications to dynamical systems, *Math. Structures Comput. Sci.*, **30** (2020), 173–233. https://doi.org/10.1017/S096012952000002X

8. M. Coste, T. Lajous-Loaeza, H. Lombardi, M. F. Roy, Generalized Budan-Fourier theorem and virtual roots, *J. Complexity*, **21** (2005), 479–486. https://doi.org/https://doi.org/10.1016/j.jco.2004.11.003
9. G. Dall'Aglio, Sugli estremi dei momenti delle funzioni di ripartizione doppia, *Ann. Scuola Norm. Sup. Pisa Cl. Sci.*, **10** (1956), 35–74.
10. T. A. Driscoll, N. Hale, L. N. Trefethen, eds., *Chebfun Guide*, Pafnuty Publications, Oxford, (2014).
11. R. M. Dudley, *Probabilities and metrics*, Matematisk Institut, Aarhus Universitet, Aarhus, (1976).
12. P. Embrechts and M. Hofert, A note on generalized inverses, *Math. Methods Oper. Res.*, **77** (2013), 423–432. https://doi.org/10.1007/s00186-013-0436-7
13. B. S. Everitt, S. Landau, M. Leese, D. Stahl, *Cluster analysis*, Wiley Series in Probability and Statistics, John Wiley & Sons, Ltd., Chichester, (2011), fifth eds. https://doi.org/10.1002/9780470977811
14. S. Frühwirth-Schnatter, *Finite mixture and Markov switching models*, Springer Series in Statistics, Springer, New York, (2006).
15. A. Galligo, Budan tables of real univariate polynomials, *J. Symb. Computat.*, **53** (2013), 64–80. https://doi.org/https://doi.org/10.1016/j.jsc.2012.11.004
16. L. Gonzalez-Vega, H. Lombardi, L. Mahé, Virtual roots of real polynomials, *J. Pure Appl. Algebra*, **124** (1998), 147–166. https://doi.org/10.1016/S0022-4049(96)00102-8
17. B. Hanzon, F. Holland, Non-negativity analysis for exponential-polynomial-trigonometric functions on $[0, \infty)$, in *Spectral theory, mathematical system theory, evolution equations, differential and difference equations*, (eds. W. Arendt, J. Ball, J. Behrndt, K.H. Foerster, V. Mehrmann, C. Trunk), vol. 221 of Oper. Theory Adv. Appl., Birkhäuser/Springer Basel AG, Basel, (2012), 399–412. https://doi.org/10.1007/978-3-0348-0297-0_21
18. 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. https://doi.org/10.1137/S036301299731610X
19. R. Jiang, M. Zuo, H. X. Li, Weibull and inverse Weibull mixture models allowing negative weights, *Reliab. Eng. Syst. Safe.*, **66** (1999), 227–234. https://doi.org/https://doi.org/10.1016/S0951-8320(99)00037-X
20. A. Karapiperi, M. Redivo-Zaglia, M. R. Russo, *Generalizations of Sylvester's determinantal identity*, (2015). Available from: `https://arxiv.org/abs/1503.00519`
21. S. Kasyanov, *GaussElimination*, (2022). Available from: `https://www.mathworks.com/matlabcentral/fileexchange/90661-gausselimination`
22. W. Koch, On exploiting 'negative' sensor evidence for target tracking and sensor data fusion, *Inform. Fusion*, **8** (2007), 28–39. https://doi.org/https://doi.org/10.1016/j.inffus.2005.09.002
23. T. Li, H. Liang, B. Xiao, Q. Pan, Y. He, Finite mixture modeling in time series: A survey of Bayesian filters and fusion approaches, *Inform. Fusion*, **98** (2023), 101827. https://doi.org/10.1016/j.inffus.2023.101827
24. G. McLachlan, D. Peel, *Finite mixture models*, Wiley Series in Probability and Statistics: Applied Probability and Statistics, Wiley-Interscience, New York, (2000). https://doi.org/10.1002/0471721182

25. P. Müller, S. Ali-Löytty, M. Dashti, H. Nurminen, R. Piché, Gaussian mixture filter allowing negative weights and its application to positioning using signal strength measurements, in *9th Workshop on Positioning, Navigation and Communication*, Dresden, Germany, (2012), 71–76. https://doi.org/10.1109/WPNC.2012.6268741
26. J. Park, I. W. Sandberg, Universal approximation using radial-basis-function networks, **3** (1991), 246–257. https://doi.org/10.1162/neco.1991.3.2.246
27. K. Pearson, Contribution to the mathematical theory of evolution, *Philos. T. A*, **185** (1894), 71–110.
28. G. Pólya, On the mean-value theorem corresponding to a given linear homogeneous differential equation, *Trans. Amer. Math. Soc.*, **24** (1922), 312–324. https://doi.org/10.2307/1988819
29. C. Ridders, A new algorithm for computing a single root of a real continuous function, *IEEE T. Circuits Syst.*, **26** (1979), 979–980. https://doi.org/10.1109/TCS.1979.1084580
30. R. Ristroph, Pólya's property *W* and factorization—A short proof, *Proc. Amer. Math. Soc.*, **31** (1972), 631–632. https://doi.org/10.2307/2037591
31. D. H. H. Santosh, P. Venkatesh, P. Poornesh, L. N. Rao, N. A. Kumar, Tracking multiple moving objects using gaussian mixture model, *Int. J. Soft Comput. Eng.*, **3** (2013), 114–119.
32. C. Sexton, M. Olivi, B. Hanzon, Rational approximation of transfer functions of nonnegative ept densities, *IFAC Proceedings*, **45** (2012), 716–721. https://doi.org/10.3182/20120711-3-BE-2027.00197
33. C. Sexton, Financial modelling with 2-EPT probability density functions, *PhD Thesis*, University College Cork, (2013). `https://cora.ucc.ie/handle/10468/1430`
34. D. M. Titterington, A. F. M. Smith, U. E. Makov, *Statistical analysis of finite mixture distributions*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons, Ltd., Chichester, (1985).
35. L. N. Trefethen, *Approximation Theory and Approximation Practice*, Society for Industrial and Applied Mathematics, (2019).
36. S. S. Vallander, Calculations of the Vasserštein distance between probability distributions on the line, *Teor. Probab. Appl.*, **18** (1973), 824–827.
37. K. Weihrach, *Computable Analysis*, Texts in Theoretical Computer Science, Springer, (2000). https://doi.org/10.1007/978-3-642-56999-9
38. D. Yu, L. Deng, *Automatic speech recognition*, Signals and Communication Technology, Springer, London, (2015).
39. B. Zhang, C. Zhang, *Finite mixture models with negative components*, in *Machine Learning and Data Mining in Pattern Recognition*, (eds. P. Perner and A. Imiya), Springer, Berlin, Heidelberg, (2005), 31–41. https://doi.org/10.1007/11510888_4