

# Modeling probability densities with sums of exponentials via polynomial approximation

Bogdan Dumitrescu<sup>a,b</sup>, Bogdan C. Şicleru<sup>a</sup>, Florin Avram<sup>c</sup>

## Abstract

We propose a method for optimization with semi-infinite constraints that involve a linear combination of functions, focusing on shape-constrained optimization with exponential functions. Each function is lower and upper bounded on sub-intervals by low-degree polynomials. Thus, the constraints can be approximated with polynomial inequalities that can be implemented with linear matrix inequalities. Convexity is preserved, but the problem has now a finite number of constraints. We show how to take advantage of the properties of the exponential function in order to build quickly accurate approximations. The problem used for illustration is the least-squares fitting of a positive sum of exponentials to an empirical probability density function. When the exponents are given, the problem is convex, but we also give a procedure for optimizing the exponents. Several examples show that the method is flexible, accurate and gives better results than other methods for the investigated problems.

**Keywords:** optimization, density fitting, semi-infinite programming, sum of exponentials, polynomial approximation.

<sup>a</sup>Department of Automatic Control and Computers, University Politehnica of Bucharest, Spl. Independenței 313, Bucharest 060042, Romania. E-mail: bogdan.dumitrescu@acse.pub.ro

<sup>b</sup> Department of Signal Processing, Tampere University of Technology, Finland. E-mail: bogdan.dumitrescu@tut.fi

<sup>c</sup> Université de Pau et des Pays de l'Adour, France

# 1 Introduction

## 1.1 The problem

Our main aim is to give a detailed procedure for solving a class of semi-infinite programming (SIP) problems involving functions of the form

$$f(t) = \sum_{i=1}^n p_i(t) e^{-\lambda_i t}, \quad (1)$$

where  $0 \leq \lambda_1 < \dots < \lambda_n$  and  $p_i(t) = \sum_{j=0}^{\kappa_i} \alpha_{i,j} t^j$  are polynomials. For the sake of abbreviation, we name SOPE (sum of polynomials times exponentials) a function like (1).

At the same time, but with much less detail, we point out how the proposed approach can be extended to a much more general category of functions, with the exponentials from (1) replaced by arbitrary, typically elementary, functions.

One important source of problems involving (1) is the modeling of random variables. If  $p_i(t)$  are all nonnegative constants ( $\kappa_i = 0$ ), then  $f(t)$  is the probability density function (pdf) of a hyperexponential distribution, up to a proportionality constant. If  $n = 1$  and  $\alpha_{i,j}$  are nonnegative, then  $f(t)$  is the pdf of an Erlang mixture. If  $n > 1$  and  $\alpha_{i,j}$  are nonnegative, then  $f(t)$  is called a hyper-Erlang density. In what follows, we do not impose nonnegativity conditions on the coefficients  $\alpha_{i,j}$ . As a result, we will be able to find solutions of lower degree than competitor algorithms which make that assumption.

There are three types of optimization problems involving (1) that we tackle. They all involve SIP constraints having the form

$$f(t) \geq 0, \quad \forall t \in [t_0, t_f], \quad (2)$$

where typical choices are  $t_0 = 0$ ,  $t_f = \infty$ ; however, it is enough to take a sufficiently large  $t_f$  instead of infinity, due to the decay properties of the exponential. For simplicity of reference, we give ad hoc names to these problems and list them in increasing order of difficulty.

*Positivity check (SOPE-P).* The simplest problem is: given a function (1), decide if the inequality (2) holds. Such decision is necessary e.g. when checking if  $f(t)$  is a pdf.

*Convex constraint (SOPE-C).* More interesting is the case when the parameters of the function (1) are variables of an optimization problem. In particular, when the polynomial coefficients  $\alpha_{i,j}$  are variable, but the exponents  $\lambda_i$  are given, the constraint (2) is convex. This type of problem is our main focus. We will always assume that the number  $n$  of exponentials and the degrees  $\kappa_i$  of the polynomials are known. They could be selected using Information Theoretic Criteria, but this is beyond the scope of this paper.

The problem on which we focus is fitting a pdf to empirical data. A random variable is governed by an unknown pdf  $h(t)$ ; from empirical observations, we know the values  $h_m = h(t_m) \geq 0$ , for some times  $t_m$ ,  $m = 1 : M$ , usually equidistant. Assuming that (1) is an appropriate model, we want to fit it to the data by solving

the least squares problem

$$\begin{aligned} \min \quad & J(f) = \frac{1}{M} \sum_{m=1}^M w_m [f(t_m) - h_m]^2 \\ \text{s.t.} \quad & f(t) \geq 0, \forall t \in [t_0, t_f] \end{aligned} \quad (3)$$

where  $w_m > 0$  are weights; we implicitly take  $w_m = 1$ . If the exponents  $\lambda_i$  are given, the problem (3) is an instance of convex SIP; its difficulty comes from the infinite number of constraints hidden by the inequality (2). There is no simple way to express it equivalently using a finite number of constraints, as would be the case if  $f(t)$  was a polynomial.

With small modifications, the problem (3) can be posed not only for pdfs, but also for cumulative density functions (cdf) or complementary cdfs, see later (35).

*General constraint (SOPE-G).* The most difficult problem is when the exponents  $\lambda_i$  are also variable. In this case an optimization problem like (3) is no longer convex. Although we will present a solution to this problem also, it will have no guarantee of optimality.

An interesting particular case of the above problems is when  $\kappa_i = 0$ , hence the polynomials are reduced to constants and so

$$f(t) = \sum_{i=1}^n \alpha_i e^{-\lambda_i t}. \quad (4)$$

The name of the problems will be changed accordingly, by replacing SOPE with SOE (sum of exponentials).

## 1.2 Contribution and contents

The approach we propose is based on a polynomial approximation of the exponential function that allows the approximation of a SOPE-C problem like (3) by a polynomial optimization problem. The latter may be reformulated via linear matrix inequalities and hence solved efficiently with semidefinite programming (SDP) methods in friendly media like CVX [10] and POS3POLY [24] or GloptiPoly [12].

We split the inequality (2) in  $K$  sub-intervals

$$[t_0, t_f] = \bigcup_{k=1}^K [t_{k-1}, t_k] \quad (5)$$

and impose positivity on each sub-interval. We compute polynomials  $\hat{b}_{ik}(t)$ ,  $\check{b}_{ik}(t)$  such that

$$\hat{b}_{ik}(t) \geq e^{-\lambda_i t} \geq \check{b}_{ik}(t), \quad \forall t \in [t_{k-1}, t_k]. \quad (6)$$

Using such lower and upper approximations of the exponentials, we impose sufficient positivity conditions on (1), simultaneously on all sub-intervals. The resulting approximation of SOPE-C will be detailed in Section 2.

One-sided approximations like in (6) can be computed by solving a linear system, using results from [3] (see also [6, 16]) reviewed in Section 3 and taking advantage of the fact that the derivatives of the exponential have constant sign.

It is intuitive that by playing with the degrees of the approximation polynomials and the lengths of the sub-intervals, the approximation can be in principle as good as desired, at the price of increased complexity. This is true not only for the exponential, but for all smooth functions on a finite interval. Numerical accuracy is also an important issue, especially since the approximating polynomials from (6) are defined by their coefficients. Note that since we use the inequalities (6) to build a lower approximation of  $f(t)$  which does not violate the positivity constraint, the quality of the approximation should be comparable to the accuracy of the SDP algorithm used for solving the transformed problem.

In a practical implementation, we have reached the conclusion that it is enough to use relatively low degrees of the approximation polynomials, for example 8 or 10, in order to reach a reasonable trade-off between accuracy and complexity. Given an approximation accuracy, the computation of the sub-intervals from (5) is completely automated, using a precomputed table of approximation errors, as described in section 3.2.

Using this approach, we propose in Section 4 a complete iterative procedure for solving the SOE-G and SOPE-G problems (3). In particular, we select initial values for the exponents  $\lambda_i$  by searching sparse functions (1) whose exponents belong to an arithmetic progression. For given exponents, the coefficients  $\alpha_{i,j}$  are optimized using the polynomial approximation suggested above. Possibly better values of the exponents are sought through a descent search, the coefficients are re-optimized, etc.

Although the procedure is based on simple ideas, the results given in Section 5 show that it is competitive, giving good results on several fitting problems studied previously.

### 1.3 Previous work

The idea of using polynomial approximations in optimization is by no means new. However, as far as we know, the particular combination of ideas that lies at the foundation of our approach was not proposed. We present below a few relevant works, pointing out the distinctive features of our method.

Most of the methods use a single polynomial for approximation on the whole interval or on sub-intervals. A leading example is the library Chebfun [25], which allows numerically-performed "symbolic" computation by actually replacing *given* functions with polynomials that approximate them to machine precision on the interval of interest. However, due to numerical considerations, the polynomials are not defined by their coefficients but by their values in Chebyshev nodes and barycentric Lagrange interpolation is employed for computing the function values. Hence, Chebfun can solve a problem like SOPE-P by computing the minimum of the function, but is not able to take advantage of convexity.

In [5], functional optimization is performed by approximating the unknown function  $f(t)$  with a polynomial with unknown coefficients. It can be seen as a *non-parametric* form of our problem, where the values of  $f(t)$  are sought for all  $t$  in some interval. Here, we are interested in the parameters  $\alpha_{i,j}$  of a function with known structure, hence our method is *parametric*.

A precursor of the above approach was presented in [4], where the minimum of a

given function was obtained by computing instead the minimum of an interpolating polynomial. Polynomial positivity, used for the minimum computation, is enforced through SDP via interpolation [17].

In somewhat the same vein of non-parametric methods, but with pdf estimation as specific target, functional approximation of a pdf with polynomials, fitting given moments, is proposed in [13]. A simpler idea appeared in [23], proposing piecewise polynomial approximations for pdfs, with no interest in optimization and no positivity enforcement. More refined approaches appeal to splines [14, 18, 1] (the latter work summarizes previous work of its authors, including the use of multidimensional splines); positivity is enforced with an ad hoc method in the first work, while in the others it is imposed on each spline section via SDP.

One-sided polynomial approximations are less current; for example, in [22], they were used for relaxation in a branch and bound process for global optimization. The approximation regarded the optimization variables, not independent functions, like here. Polynomial optimization was solved via linear programming.

There is also an entire body of literature dedicated to the estimation of the parameters of a pdf, with diverse applications. Optimization is used more or less explicitly, but usually in a relatively standard way. We cite only a few methods connected to our work. SDP was used in [9] for estimating pdfs that are the product between a polynomial and a kernel with few parameters, like the Erlang mixture; polynomial positivity is imposed via SDP. In [7], the SOE-G problem is solved within the class of exponents in arithmetic progression. However, the opportunity is missed for transforming it into a polynomial positivity problem solvable by SDP; positivity is imposed instead by rather ad hoc means. In the related work [8], we searched in the more general class of exponents belonging to a sparse arithmetic progression, and also used SDP to impose the SIP constraints. However, only SOE problems were considered. In [21], matrix exponential distributions are estimated; after finding the exponents with a method for linear systems identification, the coefficients are found by optimization, positivity being ensured by a Budan-Fourier technique. Last but not least, let us mention expectation maximization [2]; this produces solutions with positive coefficients, which typically involve polynomials of higher degree.

## 2 Transformation to a polynomial problem

We discuss here the approximation of SOPE-C by an optimization problem with polynomials, thus preserving convexity, but transforming the semi-infinite constraint into a finite one.

Let us assume that the splitting (5) and the approximations (6) of the exponentials are available. We show how to impose the positivity constraint (2) on a generic sub-interval  $[t_{k-1}, t_k]$ , keeping in mind that the conditions are imposed simultaneously on all sub-intervals.

### 2.1 Known signs

Assume first that the polynomials  $p_i(t)$  have constant and known signs on  $[t_{k-1}, t_k]$ . This is usually possible only when solving SOPE-P, where  $p_i(t)$  are fixed. Let  $I_+$  be

the set of indices  $i$  for which  $p_i(t) \geq 0$  on  $[t_{k-1}, t_k]$ , and  $I_-$  be defined similarly for the negative case. In this situation, we can write

$$f(t) \geq \sum_{i \in I_+} p_i(t) \check{b}_{ik}(t) + \sum_{i \in I_-} p_i(t) \hat{b}_{ik}(t) = P_k(t). \quad (7)$$

Since  $P_k(t)$  is a polynomial whose coefficients depend linearly on those of  $p_i(t)$ , imposing  $P_k(t) \geq 0$  on  $[t_{k-1}, t_k]$  is easy and can be inserted into any convex optimization problem involving  $f(t)$ . How to transform a polynomial positivity constraint into a linear matrix inequality is discussed in [19]. However, such knowledge is not necessary when using a library facilitating the manipulation of positive polynomials, like POS3POLY.

## 2.2 Unknown signs

In general, the coefficients and the signs of  $p_i(t)$  are not known, since the coefficients of these polynomials are variables in the optimization problem SOPE-C. In this case, it is impossible to build the approximation (7). However, we can replace it with

$$P_k(t) = \sum_{i=1}^n [p_i(t) - \gamma_{ik}(t)] \hat{b}_{ik}(t) + \gamma_{ik}(t) \check{b}_{ik}(t), \quad (8)$$

with the extra conditions

$$\begin{aligned} p_i(t) - \gamma_{ik}(t) &\leq 0, \quad \forall t \in [t_{k-1}, t_k] \\ \gamma_{ik}(t) &\geq 0, \end{aligned} \quad (9)$$

where  $\gamma_{ik}(t)$  is a polynomial, typically of the same degree as  $p_i(t)$ . If  $\deg \gamma_{ik} < \deg p_i$ , then we can consider the alternative bounding polynomial

$$P_k(t) = \sum_{i=1}^n \gamma_{ik}(t) \hat{b}_{ik}(t) + [p_i(t) - \gamma_{ik}(t)] \check{b}_{ik}(t), \quad (10)$$

with the extra conditions

$$\begin{aligned} p_i(t) - \gamma_{ik}(t) &\geq 0, \quad \forall t \in [t_{k-1}, t_k] \\ \gamma_{ik}(t) &\leq 0. \end{aligned} \quad (11)$$

If  $\deg \gamma_{ik} \geq \deg p_i$ , then (8)-(9) and (10)-(11) are equivalent; otherwise, they are different, but it is hard to give general rules for choosing one over the other. In both cases, it is clear that  $f(t) \geq P_k(t)$  and so  $P_k(t) \geq 0$  is a sufficient condition for  $f(t) \geq 0$ . From now on we will work only with (8)-(9).

So, the SOPE-C problem (3) is approximated with

$$\begin{aligned} \min \quad & J(f) \\ \text{s.t.} \quad & \left. \begin{aligned} \sum_{i=1}^n [p_i(t) - \gamma_{ik}(t)] \hat{b}_{ik}(t) + \gamma_{ik}(t) \check{b}_{ik}(t) &\geq 0, \quad \forall t \in [t_{k-1}, t_k], \quad k = 1 : K \\ p_i(t) - \gamma_{ik}(t) &\leq 0 \\ \gamma_{ik}(t) &\geq 0 \end{aligned} \right\} \quad \forall t \in [t_{k-1}, t_k], \quad k = 1 : K, \quad i = 1 : n \end{aligned} \quad (12)$$

Note that the objective is unchanged. The price for generality is the apparition of the new variable polynomials  $\gamma_{ik}(t)$  and of the new constraints (9). Although there is a potentially large number of variable and constraints, namely  $nK$ , we will see later that the degrees of all polynomials involved here are generally small, and also the number  $K$  of sub-intervals is not large. So, the problem (12) does not have an excessively high complexity.

**Proposition 1** Denote  $J_o$  and  $J_\gamma$  the optimal value of the problems (3) and (12), respectively. Let  $\varepsilon_{ik}(t) \geq 0, \forall t \in [t_{k-1}, t_k]$ , be some fixed polynomials. Define

$$P_{k,\varepsilon}(t) = \sum_{i=1}^n [p_i(t) - \gamma_{ik}(t) - \varepsilon_{ik}(t)]\hat{b}_{ik}(t) + [\gamma_{ik}(t) + \varepsilon_{ik}(t)]\check{b}_{ik}(t) \quad (13)$$

and denote  $J_{\gamma,\varepsilon}$  the optimal value of (12), modified by  $P_{k,\varepsilon}(t)$  replacing  $P_k(t)$ . Then, the inequality  $J_o \leq J_\gamma \leq J_{\gamma,\varepsilon}$  holds.

*Proof.* Since  $\gamma_{ik}(t) \geq 0, \varepsilon_{ik}(t) \geq 0, \forall t \in [t_{k-1}, t_k]$ , it results that  $P_{k,\varepsilon}(t) \leq P_k(t) \leq f(t)$  on  $[t_{k-1}, t_k]$ . Hence, the feasibility domain of the modified problem (12) is included into that of (12), which in turn is included into that of (3). The objectives of the three problems being the same and a larger feasibility domain meaning the possibility of lower objective values, it follows that the inequality stated by the Proposition holds. ■

**Remark 1** The inequality  $J_o \leq J_\gamma$  is a natural consequence of the approximation (6). If there would be equality in (6), then (3) and (12) would be equivalent. By reducing the lengths of the sub-intervals  $[t_{k-1}, t_k]$  the approximation in (6) becomes better and so when  $K \rightarrow \infty$  (and the length of all sub-intervals tends to zero), the solutions of (3) and (12) are the same. ■

**Remark 2** The inequality  $J_\gamma \leq J_{\gamma,\varepsilon}$  says that, when solving (12), the polynomials  $\gamma_{ik}(t)$  will generally tend to take their smallest possible values that are allowed by the constraints (9).

So, if  $p_i(t) \leq 0$ , then it results that  $\gamma_{ik}(t) = 0$ , and (assuming  $\deg \gamma_{ik} \geq \deg p_i$ ) if  $p_i(t) \geq 0$ , then  $\gamma_{ik}(t) = p_i(t)$ ; the approximation (8) actually coincides (a posteriori, after the optimization problem is solved and  $p_i(t)$  is available) with that for known signs (7). If  $p_i(t)$  changes the sign on the current interval, then  $\gamma_{ik}(t)$  is the best upper approximation to  $\max(p_i(t), 0)$ .

The above values of  $\gamma_{ik}(t)$  are reached only if the constraint  $P_k(t) \geq 0$  is active on the interval  $[t_{k-1}, t_k]$ . However, the important conclusion is that the construction (8)-(9) naturally gives the best lower approximation of  $f(t)$  by a polynomial  $P_k(t)$ , given the bounding polynomials (6). ■

### 3 One-sided polynomial approximations of a set of exponentials

We present here the tools for finding the polynomial approximation (6) and the intervals (5) for a set of exponentials  $e^{-\lambda_i t}, i = 1 : n$ . We also suggest how this can be done for other functions.

### 3.1 Approximation of a single exponential

We start by showing how to find lower and upper polynomial approximations to a single exponential  $e^{-\lambda t}$  on an interval  $[\tau_0, \tau_1]$ . This is achieved via [3, Th.4], taking advantage of the fact that the derivatives of all orders of the exponential have constant sign.

We present in detail only one case, that of the lower approximation  $\check{b}(t)$  of odd degree  $\nu = 2\ell - 1$ , to a function  $\phi(t)$  whose derivative of order  $\nu + 1$  is nonnegative on  $[\tau_0, \tau_1]$  (which is the case of our exponential). The approximation is optimal in the sense that the norm

$$\int_{\tau_0}^{\tau_1} w(t)[\phi(t) - \check{b}(t)]dt \quad (14)$$

is minimized, where  $w(t)$  is a weight function. Denote by  $x_1, \dots, x_\ell$  the zeros of the  $\ell$ -th order polynomial from the sequence of polynomials which are orthogonal on  $[\tau_0, \tau_1]$  with respect to  $w(t)$ . Then, the coefficients of  $\check{b}(t)$  can be found by solving the linear system given by the equations

$$\check{b}(x_k) = \phi(x_k), \quad \check{b}'(x_k) = \phi'(x_k), \quad k = 1 : \ell. \quad (15)$$

So, if the zeros of  $w(t)$  are readily available, then the computation of  $\check{b}(t)$  is very simple and effective.

The other cases, corresponding to the three other combinations of the parity of  $\nu$  and the sign of the  $\nu + 1$ -th derivative of  $\phi(t)$  are similar, but involve the zeros of the polynomials that are orthogonal with respect to  $(t - \tau_0)w(t)$ ,  $(\tau_1 - t)w(t)$  and  $(t - \tau_0)(\tau_1 - t)w(t)$ , respectively, and one or both of the interval ends.

For the sake of quick computation, we choose the weight  $w(t) = \sqrt{(\tau_1 - t)(t - \tau_0)}$ , which generates Chebyshev polynomials of the first kind. Their roots are

$$x_k = \frac{\tau_1 - \tau_0}{2} \cos \frac{(2k - 1)\pi}{2\ell} + \frac{\tau_1 + \tau_0}{2}, \quad k = 1 : \ell. \quad (16)$$

(The other three weights above generate Chebyshev polynomials of the second kind and Jacobi  $(-\frac{1}{2}, \frac{1}{2})$  and  $(\frac{1}{2}, -\frac{1}{2})$  polynomials, whose roots are also available via simple formulas.)

**Example 3** To have an idea of the approximation error, we plot in Figure 1 the maximum value of  $\hat{b}(t) - e^{-t}$  and  $e^{-t} - \check{b}(t)$  for several degrees of the polynomial and intervals starting from 0 and ending in various points up to 5. One can see that, for example, a polynomial of degree 8 gives an error smaller than  $10^{-10}$  for intervals included in  $[0, 1.1]$ . The approximation becomes unreliable when the approximation error approaches  $10^{-16}$ , in the sense that it may be no longer one-sided. However, errors of order  $10^{-12}$  appear perfectly obtainable. ■

Although the infinity norm is preferable to an integral norm like (14), if the minimization of the maximum approximation error is desired, the computation of such a norm would involve a Remez-like algorithm which significantly increases the computation time with relatively small benefits (at least in the case of the exponential and our choices of intervals).



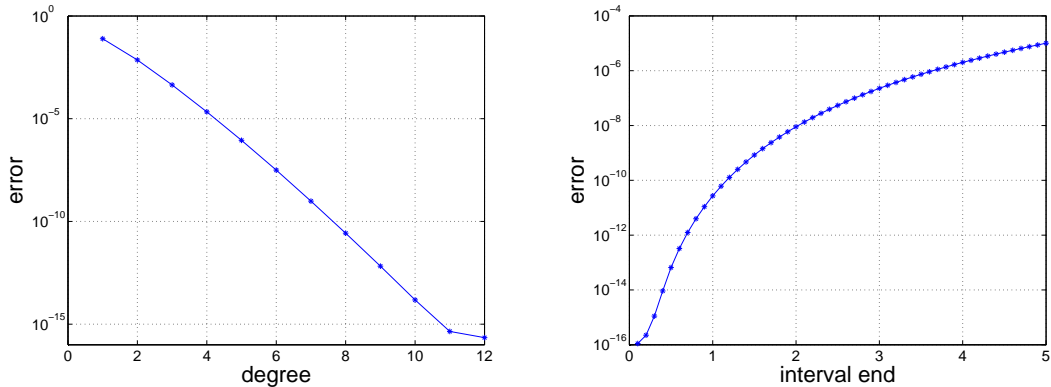


Figure 1: One-sided polynomial approximation error for  $e^{-t}$ . Left: for various degrees on the interval  $[0, 1]$ . Right: for degree 8 and intervals from 0 to the value on the horizontal axis.

An alternative to the above construction was employed in [22] and works for general functions, not only in the restrictive conditions from [3]. Using a grid like (16), one builds Chebyshev interpolation polynomials, for whom an approximation error bound is available. Subtracting or adding the bound value to the polynomial gives lower or upper one-sided approximations. A more refined approach [20], based on a sound numerical implementation of the Remez algorithm, can give nearly optimal solutions.

### 3.2 Approximations of a set of exponentials

We go now to our full problem: the approximation of a set of exponentials on a union of intervals. We assume a target approximation error  $\epsilon$ . Given the exponentials  $e^{-\lambda_i t}$ ,  $i = 1 : n$ , the problem is how to find the intervals (5) and the degrees  $\nu_{ik}$  of the polynomials (6) such that

$$\begin{aligned} \max_{t \in [t_{k-1}, t_k]} |\hat{b}_{ik}(t) - e^{-\lambda_i t}| &\leq \epsilon, & \forall i = 1 : n, \forall k = 1 : K. \\ \max_{t \in [t_{k-1}, t_k]} |e^{-\lambda_i t} - \check{b}_{ik}(t)| &\leq \epsilon, \end{aligned} \quad (17)$$

The solution is obviously not unique, but we aim to find one that is computationally advantageous. The exponential function allows a very cheap solution.

**Remark 4** Assume that the maximum approximation error of a polynomial of degree  $\nu$  approximating  $e^{-t}$  on the interval  $[0, \tau]$  is  $\epsilon$ . Then, there exist polynomials of degree  $\nu$  such that:

- $e^{-\lambda t}$  is approximated on  $[0, \tau/\lambda]$  with error at most  $\epsilon$ ;
- $e^{-t}$  is approximated on  $[x, x + \tau]$  with error at most  $\epsilon e^{-x}$ .

Combining these two results, it follows that  $e^{-\lambda t}$  is approximated on  $[x, x + \tau/\lambda]$  with error at most  $\epsilon e^{-\lambda x}$ . ■

**Example 5** Let us put the problem differently. We take  $\epsilon = 10^{-10}$  and  $\nu = 8$ . We want to approximate  $e^{-3t}$  on an interval  $[1, 1 + \theta]$  as large as possible, with approximation error less than  $\epsilon$ . From the above Remark, the error for  $e^{-t}$  on  $[0, 3\theta]$  should be less than  $e^3\epsilon \approx 2 \cdot 10^{-9}$ . Looking in Figure 1 (right), we see that this happens for  $3\theta \approx 1.6$ . Indeed, taking  $\theta = 0.55$  gives the desired accuracy. ■

The approximation procedure we propose is as follows. Besides the input data listed in the beginning of this section, we assume that a maximum degree  $\nu_{\max}$  is given.

1. We make a table  $T(\tau, \nu)$  of approximation errors given by polynomials built as described in this section for approximating  $e^{-t}$  on  $[0, \tau]$ . The errors are measured on a grid; there can be a few tens of  $\tau$  values and  $\nu$  can go from 0 to 10 or 12. (The result would be a "cartesian product" of the two graphs from Figure 1.) This table, having a few hundred entries, is built a single time. Even so, the computation time for a  $50 \times 12$  table was only 0.25 seconds on a standard desktop, which is negligible.

2. Assume that we have found  $t_{k-1}$  and we search  $t_k$ . For each  $\lambda_i$ , using the table and Remark 4, we find the interval length  $d_{ik}$  such that the approximation error for  $e^{-\lambda_i t}$  on  $[t_{k-1}, t_{k-1} + d_{ik}]$  is at most  $\epsilon$ . More precisely, we seek in the table the value  $\tau$  for which  $T(\tau, \nu_{\max})$  has the largest value smaller than  $\epsilon \cdot e^{\lambda_i t_{k-1}}$  and set  $d_{ik} = \tau/\lambda_i$ . Finally, to ensure that the approximation error is respected for all exponentials, we take the smallest interval length and put  $t_k = t_{k-1} + \min_{i=1:n} d_{ik}$ . This iterative procedure, ending when  $t_k \geq t_f$ , is extremely fast since it involves only table searches.

For the sake of numerical accuracy, we optionally can reduce the degrees of the approximating polynomials for the exponentials that are not deciding the length of an interval. (Note that the highest degree of a polynomial actually decides the complexity.) Denoting  $d_k = t_k - t_{k-1}$ , for each  $i$  for which  $d_{ik} > d_k$  we search in the table the smallest degree  $\nu_{ik}$  for which  $T(\tilde{\tau}, \nu)e^{\lambda_i \tau_{k-1}}$  is smaller than  $\epsilon$ , where  $\tilde{\tau}$  is the smallest grid value larger than  $d_k/\lambda_i$ . Remark 4 ensures that the error made by approximating  $e^{-\lambda_i t}$  on  $[t_{k-1}, t_k]$  with a polynomial of degree  $\nu_{ik}$  is at most  $\epsilon$ .

**Example 6** Let us approximate the exponentials with exponents 0.3, 1 and 3 on the interval  $[0, 10]$ , with error less than  $\epsilon = 10^{-10}$ , using polynomials of degree at most  $\nu_{\max} = 8$ . The table is built with step value 0.2 for  $\tau$ . Using the above procedure, we find that 10 sub-intervals are necessary. The first is  $[0, 0.33]$ , on which the degrees of the approximating polynomials are 5, 6 and 8; the fastest decaying exponential naturally needs the highest degree. The last sub-interval is  $[9.27, 10]$  and the degrees are 6, 5 and 1. They are smaller than  $\nu_{\max}$  because the sub-interval is cut short by  $t_f = 10$ ; for example, the previous sub-interval is  $[6.67, 9.27]$ ; since the exponentials are decaying, the sub-intervals are longer as  $t$  grows. Note that now the slowest exponential sets the degree; the fastest has almost vanished. ■

Also for improving numerical accuracy, it is useful to move the approximation on an interval centered in the origin. Similarly to a Vandermonde system, the linear system (15) tends to become ill-conditioned when the roots (16) have all the same sign and large absolute value. So, denoting  $\theta_k = t_{k-1} + d_k/2$ , instead of working with  $p_i(t)e^{-\lambda_i t}$  for  $t \in [t_{k-1}, t_k]$ , we work with  $p_i(t + \theta_k)e^{-\lambda_i \theta_k} e^{-\lambda_i t}$  for  $t \in [-d_k/2, d_k/2]$ .

Since usually the degree  $\kappa_i$  of  $p_i(t)$  is small, the bad numerical effect of computing the coefficients of  $p_i(t + \theta_k)$  (as a polynomial in  $t$ ) is much smaller than the reduction of the condition number of the system solved for approximating  $e^{-\lambda_i t}$  on  $[-d_k/2, d_k/2]$ .

It is clear that the above procedure is particularly fit for the exponential function. For other functions, the polynomial approximations must be computed explicitly for each sub-interval, in a trial-and-error procedure for finding the right polynomial degrees and the sub-intervals lengths; this is essentially what Chebfun does, but in our case more flexibility is allowed since we do not (and cannot) aim to approximation within machine precision.

It is also obvious that imposing an approximation error  $\epsilon$  for each exponential, does not make the approximation error in (2) be of the same size. However, an a posteriori analysis of the function  $f(t)$ , in particular of the values of the computed polynomials  $p_i(t)$ , can help estimate the actual error. A single new run with a smaller  $\epsilon$  would be required, since the new optimized function should be near the previous one and hence the worst-case accuracy becomes predictable.

## 4 Fitting an empirical density

Let us now come back to our prototype problem (3) of fitting an empirical density and consider first the simpler SOE case (4). We start by studying a helpful particular case.

### 4.1 Exponents in arithmetic progression

Let us assume that the exponents are known and form an arithmetic progression, meaning that

$$\lambda_i = \lambda_1 + (i - 1)q, \quad i = 1 : n, \quad (18)$$

where  $q > 0$  is the ratio. In this case, the function (4) has the form

$$f(t) = e^{-\lambda_1 t} \sum_{i=1}^n \alpha_i e^{-(i-1)qt}. \quad (19)$$

Denoting  $x = e^{-qt}$ , the condition (2) becomes

$$\sum_{i=1}^n \alpha_i x^{i-1} \geq 0, \quad \forall x \in [e^{-qt_f}, e^{-qt_0}]. \quad (20)$$

This is a polynomial positivity condition, easy to impose through LMIs [19]. Hence the SOPE-C problem (3) is equivalent to an SDP problem. No approximation is required, the problem is naturally polynomial.

The method from [7] works with models based on the progression (18) and transforms  $f(t)$  into a polynomial as above. However, positivity is imposed by quite rudimentary means (fitting a SOE to the square root of the target pdf, then squaring, which gives also a SOE, but with more terms). Moreover, the values  $\lambda_1$  and  $q$  defining the arithmetic progression are found by just trying different values until a satisfying result is obtained.

## 4.2 An algorithm for the SOE-G problem

The arithmetic progression case can be used for initialization in an iterative procedure for the general problem SOE-G. The iterative part uses the polynomial approximation described in Sections 2 and 3.

1. *Initialization using sparse arithmetic progression.* Let us assume that an estimate  $\tilde{\lambda}_1$  of  $\lambda_1$  is available. This can be obtained like in [21], using the tail of  $h(t)$ ; for large values of  $t$ , the slowest exponential dominates the others. Alternatively, we can just take  $\tilde{\lambda}_1$  sufficiently small, since it is enough to have  $\tilde{\lambda}_1 < \lambda_1$ .

We then attempt to solve SOE-G by making a more general assumption than in Section 4.1 and searching exponents estimations  $\tilde{\lambda}_i$  that *belong* to an arithmetic progression

$$\tilde{\lambda}_i = \tilde{\lambda}_1 + \mu_i q, \quad (21)$$

where  $\mu_i$  are unknown positive integers.

To this purpose, we work with the function

$$\tilde{f}(t) = \sum_{i=0}^N \tilde{\alpha}_i e^{-(\tilde{\lambda}_1 + \mu_i q)t} = e^{-\tilde{\lambda}_1 t} \sum_{i=0}^N \tilde{\alpha}_i e^{-q \mu_i t}, \quad (22)$$

with given  $q$  and  $N$ . In principle,  $q$  should be small enough to cover decently well the possible intervals where the exponents lie and  $N$  should be as large as computationally acceptable (e.g. 100 is certainly good, but one can consider going to 200 and even beyond). Since we seek a sparse solution, with only  $n$  nonzero coefficients  $\tilde{\alpha}_i$ , we modify the problem (3) by adding a sparsity-promoting term to the objective and transforming it into

$$\begin{aligned} \min \quad & J(\tilde{f}) + \beta \sum_{i=0}^N \varpi_i |\tilde{\alpha}_i| \\ \text{s.t.} \quad & \tilde{f}(t) \geq 0, \quad \forall t \in [0, \infty) \end{aligned} \quad (23)$$

where  $\beta$  and  $\varpi_i$ ,  $i = 0 : N$ , are weighting constants. The second term of the objective is a weighted 1-norm of the coefficients vector  $\tilde{\alpha}_i$ , denoted  $\|\tilde{\alpha}\|_{\varpi,1}$ ; the most meaningful choice appears to be  $\varpi_i = 1/\tilde{\lambda}_i$ , which takes into account that  $\int_0^\infty e^{-\lambda t} = 1/\lambda$  and thus implicitly normalizes the exponentials from (4).

The problem (23) is convex in the coefficients  $\tilde{\alpha}_i$ . The positivity constraint can be expressed as the positivity of a polynomial  $\tilde{f}(x)$  on  $[0, 1]$ , by substituting  $x = e^{-qt}$  as in Section 4.1. If the constant  $\beta$  is large enough, many of the coefficients  $\tilde{\alpha}_i$  are small and only few have significant values. We take the largest  $n$  of them (in weighted absolute value  $\varpi_i |\tilde{\alpha}_i|$ ) and the corresponding  $\lambda_i$  are given by their positions.

We can then re-solve (23) by imposing that only the chosen  $n$  coefficients are nonzero, finding thus the optimal coefficients  $\alpha_i$  for the selected  $\lambda_i$  (this can have the advantage of exact positivity constraint via polynomials, but can be skipped by going directly to the iterative step).

The problem (23) could be replaced with a minimization of  $\|\tilde{\alpha}\|_{\varpi,1}$  with a bounded  $J(\tilde{f})$ ; in this case, the bound should be chosen instead of  $\beta$ ; this may make more sense if some value of the objective is already available.

2. *Iterative part.* With the above initialization, we can start the iterative part of the algorithm, where the coefficients and the exponents are optimized alternatively.

For given exponents  $\lambda_i$ , we optimize the coefficients  $\alpha_i$  by solving the approximation (12) of the SOE-C problem (3), as shown in Sections 2 and 3. The auxiliary variables  $\gamma_{ik}(t)$  are just scalars.

For optimizing the exponents  $\lambda_i$  we attempt small gradient steps. The gradient is

$$\frac{\partial J(f)}{\partial \lambda_i} = -\frac{2\alpha_i}{M} \sum_{m=1}^M w_m t_m [f(t_m) - h_m] e^{-\lambda_i t_m}. \quad (24)$$

The step size is  $\varsigma$ , so the new values of the exponents are

$$\lambda_i \leftarrow \lambda_i - \varsigma \frac{\partial J(f)}{\partial \lambda_i}. \quad (25)$$

(If the resulting  $\lambda_i$  would become negative, we reduce the step size such that they stay positive.) Then, the SOE-C problem is (approximately) solved with the new exponents  $\lambda_i$ . If the value of the objective does not decrease, we restore the previous exponents, halve the step size  $\varsigma$  and recompute (25). Note that by modifying  $\lambda_i$  we usually improve the objective but may go out of the positivity domain. Solving (12) means returning back to it, but not necessarily with a better objective value.

The iterations continue as long as the improvement is significant or the step size is not very small. Of course, there is no guarantee of optimality, but our main purpose is to illustrate the kernel of our approach—the polynomial approximation idea. Even with such a simple descent procedure, the results are satisfactory in the test problems we will report in Section 5.

### 4.3 Extension to SOPE

The iterative part of the above procedure can be used for true SOPE functions (1), with some  $\kappa_i > 0$ , with the minor change of the gradient expression (24) into

$$\frac{\partial J(f)}{\partial \lambda_i} = -\frac{2}{M} \sum_{m=1}^M w_m t_m p_i(t_m) [f(t_m) - h_m] e^{-\lambda_i t_m}. \quad (26)$$

However, the arithmetic progression trick used in the SOE case is no more possible, since the transformation  $x = e^{-qt}$  no longer leads to a polynomial. Instead, we simply use the exponents resulting from a SOE solution (not necessarily fully optimized) as initialization for the SOPE problem. As confirmed by numerical evidence, this seems effective especially when the polynomials degrees  $\kappa_i$  from (1) are small.

## 5 Numerical results

The polynomial approximation method described in this paper has been implemented using POS3POLY [24] for CVX [10] and can be downloaded from <http://www.schur.pub.ro/sope>, together with the programs solving the problems presented in this section. The simplest description of the constraint (2), with a SOE function (4) characterized by the variable vector of coefficients `alpha` and the constant vector of exponents `lambda` is

```
alpha == pos_soe( lambda, t0, tf );
```

where `t0` and `tf` are the positivity interval ends. This is consistent with CVX style and the parameters of the approximation are hidden, although the user can control them if so desired.

We start illustrating the behavior of our method by solving a problem proposed in [21], that is slightly different in nature from (3), but finally has the same form.

**Example 7** Unlike an empirical pdf, the SOE

$$h(t) = 16e^{-\frac{t}{2}} - 30e^{-t} + 15e^{-2t} \quad (27)$$

has also negative values. We want to find the nearest positive SOE (4) to  $h(t)$ , by minimizing

$$J(f) = \int_0^{\infty} [f(t) - h(t)]^2 dt. \quad (28)$$

This objective is convex quadratic in  $\alpha_i$ , like (3), but is rational in  $\lambda_i$ . The expression of the gradient with respect to  $\lambda_i$  is omitted here, for brevity, but easy to obtain.

Using the tail of  $h(t)$  like in [21], the smallest exponent estimation is  $\tilde{\lambda}_1 = 0.4873$ . With  $q = 0.01$ ,  $N = 100$  and  $\beta = 0.0005$  in (22–23), the initialization step of our algorithm gives  $\lambda_1 = 0.5073$ ,  $\lambda_2 = 1.0773$ ,  $\lambda_3 = 1.4873$  and an objective value  $J = 0.0433$ . Positivity is imposed on the interval  $[0, 10]$ , which is large enough to ensure it on  $[0, \infty]$ .

The iterative step improves it to  $J = 0.0425$ , the result being

$$f(t) = 19.91e^{-0.5315t} - 48.63e^{-1.0264t} + 29.66e^{-1.5177t}. \quad (29)$$

Figure 2 presents the graph of this function and of the target  $h(t)$ . The whole design, containing 30 iterations, took less than 1 minute on a standard desktop computer. The polynomial approximations of the exponentials were made with  $\nu_{\max} = 8$  and  $\epsilon = 10^{-8}$ . The minimum value of the SOE (29), computed on a very fine grid, is  $3.5 \cdot 10^{-7}$ . Taking a smaller  $\epsilon$  reduces this value only if the accuracy of the SDP solver is also increased. For example, setting `cvx_precision best` (which means that CVX iterates as long as the objective can be decreased without numerical trouble), a value of  $\epsilon = 10^{-12}$  (which is still safe for the polynomial approximation) leads to a minimum of the SOE of  $3.6 \cdot 10^{-11}$ .

In the above case, the initialization step already gave a good result. However, the iterative step can significantly decrease the objective if the approximation is poor. For example, with  $q = 0.03$ , the initialization step gives  $J = 0.2227$  and the final result is  $J = 0.0439$ .

Solving the SOE-C problem with the exponents from (27) kept fixed, the optimal objective is 0.0712. The same value is given by a dedicated algorithm that takes into account that the exponents are actually part of an arithmetic progression and imposes positivity via polynomials; the resulting optimal SOE, whose graph is also shown in Figure 2, is

$$f(t) = 15.5243e^{-\frac{t}{2}} - 28.5073e^{-t} + 14.2410e^{-2t}. \quad (30)$$

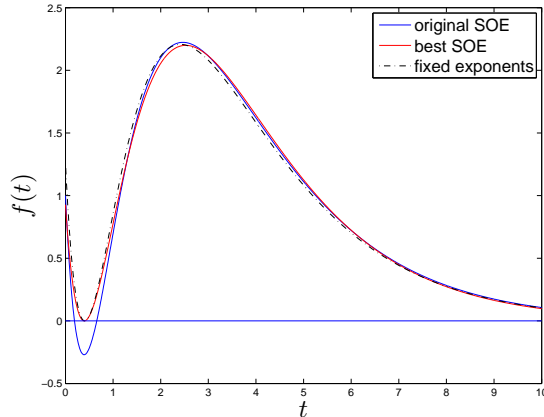


Figure 2: Graphs of the functions from Example 7. Blue: target  $h(t)$  (27). Red: our best approximation (29). Black, dashed: best approximation (30), with the same exponents as  $h(t)$ .

The result reported in [21] is

$$f(t) = 16e^{-\frac{t}{2}} - 29.946e^{-t} + 15.5385e^{-2t}. \quad (31)$$

Somewhat surprisingly, the corresponding objective is  $J = 0.0933$ . ■

Many methods in the literature are tested using standard densities, and our next two examples will be of this type. Given a continuous pdf  $h(t)$ , we simply discretize it on  $M = 100$  equidistant values in the interval  $[0, t_f]$  and optimize the objective (3) through the approximation (12). The values  $\nu_{\max} = 8$  and  $\epsilon = 10^{-8}$  are kept throughout the rest of the paper. The other parameters may change values, but we will not report them, since they can be found in the Matlab files.

**Example 8** We discuss in detail the example W1 from [11], where the target is the Weibull pdf

$$h(t) = \frac{k}{\eta} \left(\frac{t}{\eta}\right)^{k-1} e^{-(t/\eta)^k}, \quad (32)$$

with  $\eta = 1$ ,  $k = 1.5$ . We approximate it with a SOE with  $n = 4$  terms on the interval  $[0, 5]$ . The solution

$$f(t) = 130.12e^{-2.7535t} - 329.39e^{-3.1707t} + 228.96e^{-3.5850t} - 29.665e^{-4.7543t} \quad (33)$$

gives  $J = 4.82 \cdot 10^{-5}$ . The result is illustrated in Figure 3, together with the best 4-term SOE reported in [11] (where the approximation is made on the Laplace transform), for which the objective is only  $J = 5.90 \cdot 10^{-4}$ . Also, our result is visually similar to the order six model from [2, Fig. 2].

For other examples from [11] involving the Weibull and lognormal distribution, where we could compare the solutions only graphically (since the result was given only in this form), we have obtained better results with the same number of exponentials or similar results with fewer exponentials. Note also that the method

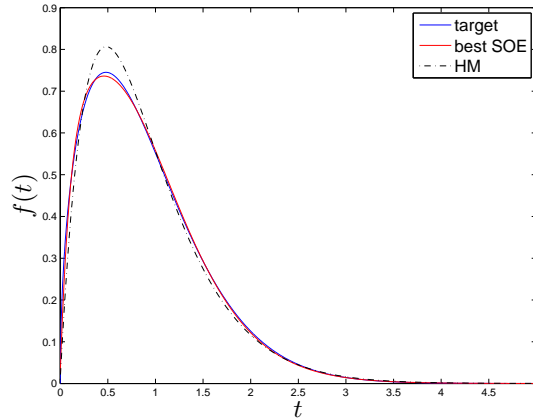


Figure 3: Graphs of the functions from Example 8. Blue: target  $h(t)$  (32). Red: our best approximation (33). Black, dashed: the approximation from [11].

from [11] allows complex values for the exponents; however, for example W1 the exponents were real. ■

**Example 9** We turn now to some examples from [7], where the complementary cumulative density function (ccdf)

$$H(t) = \int_t^\infty h(\tau) d\tau \quad (34)$$

was used instead of the pdf  $h(t)$  and modeled with a SOE (4). However, since the derivative of a SOE  $F(t)$  is also a SOE  $f(t)$ , we have solved a problem very similar to (3), namely

$$\begin{aligned} \min \quad & J(F) = \frac{1}{M} \sum_{m=1}^M w_m [F(t_m) - H_m]^2 \\ \text{s.t.} \quad & f(t) \geq 0, \quad \forall t \in [t_0, t_f] \\ & F'(t) = -f(t) \\ & F(t_0) = 1, \quad F(t_f) \geq 0 \end{aligned} \quad (35)$$

The last two constraints, that do not appear in (3), are linear in the coefficients of the SOE and hence the algorithms described in this paper can be immediately adapted.

The two ccdf discussed in this example are those of the Pareto

$$H(t) = \frac{1}{1+t} \quad (36)$$

and lognormal

$$H(t) = 1 - \frac{1}{2} \operatorname{erfc} \left( -\frac{\log(t-\mu)}{\sigma\sqrt{2}} \right) \quad (37)$$

distributions. For the lognormal, the parameters are  $\mu = 0$ ,  $\sigma = 0.5$ . The intervals on which optimization is performed are  $[0, 20]$  for Pareto and  $[0, 6]$  for the lognormal.



Target ccdf	Approximation	$n$			
		3	5	11	21
Pareto	[7], LS optimal	$1.84 \cdot 10^{-2}$	$5.02 \cdot 10^{-4}$	$1.40 \cdot 10^{-6}$	$1.52 \cdot 10^{-8}$
	Our SOE	$1.28 \cdot 10^{-4}$	$2.36 \cdot 10^{-5}$	$1.73 \cdot 10^{-6}$	$3.42 \cdot 10^{-9}$
	Our SOPE	$2.87 \cdot 10^{-6}$	$1.27 \cdot 10^{-7}$	$1.33 \cdot 10^{-6}$	$4.73 \cdot 10^{-9}$
lognormal	[7], LS optimal	$8.97 \cdot 10^{-3}$	$1.86 \cdot 10^{-3}$	$3.01 \cdot 10^{-6}$	$4.50 \cdot 10^{-7}$
	Our SOE	$1.39 \cdot 10^{-3}$	$6.68 \cdot 10^{-5}$	$3.83 \cdot 10^{-6}$	$3.75 \cdot 10^{-6}$
	Our SOPE	$1.29 \cdot 10^{-4}$	$7.14 \cdot 10^{-6}$	$4.00 \cdot 10^{-6}$	$5.26 \cdot 10^{-6}$

Table 1: Values of the objective (35) for Example 9.

We present in Table 1 the values of the objective (35) for several values of the number  $n$  of exponentials and three methods. The first is an optimized version of the method from [7], which used exponents in arithmetic progression. The first exponent and the progression ratio were found by trial and error; we use the same values as reported there. In [7], the SOE is approximated using Jacobi polynomials, but with no explicit optimization. We use instead the exact optimization method sketched in section 4.1. Hence we obtain better results than those reported in [7]. (Note that the maximum error is reported in [7], while we optimize an LS objective and still get better maximum error.)

The other two methods are ours. In addition to SOE, we report now also some results for SOPE models. The values shown in Table 1 are obtained with degrees  $\kappa_i = 1, i = 1 : n$ , in (1). For Pareto, when comparing the number  $\kappa = n + \sum_{i=1}^n \kappa_i$  of coefficients  $\alpha_{i,j}$ , the SOPE model is sometimes more efficient than SOE. For example, the objective for SOPE with  $n = 5$  and hence  $\kappa = 10$  is smaller than that for SOE with  $n = \kappa = 11$ . Also, if we take  $\kappa = 5$  via  $\kappa_1 = 0, \kappa_2 = \kappa_3 = 1$ , the objective is  $J = 5.53 \cdot 10^{-6}$ , smaller than the value for SOE and  $n = 5$ . For the lognormal (as well as for the previous examples, where SOPE was not mentioned), SOE appears more adequate than SOPE.

Comparing now our SOE method with the least-squares optimal version of the method from [7], we notice two distinct behaviors. For small  $n$ , our method is significantly better, showing that the search for good exponents is successful and that using an arithmetic progression may be quite restrictive. However, for larger  $n$ , the methods give comparable results. Now the arithmetic progression seems a satisfactory model. Numerical accuracy properties of the SDP solver may also come into play, since the fit is actually very good and the values of the objective very small. Finally, we note that our method, by its nature, targets small values of  $n$ , where the parameters of the model may have a significance. Taking exponents in a long arithmetic progression resembles more what was called non-parametric method in the introduction. ■

**Example 10** We fit now SOE and SOPE models to data belonging to a popular time series, the eruption durations of the Old Faithful geyser. The problem is notoriously difficult since the pdf appears to have two disjoint intervals as support. There are 272 values in the series, which are grouped into  $M = 40$  equally spaced

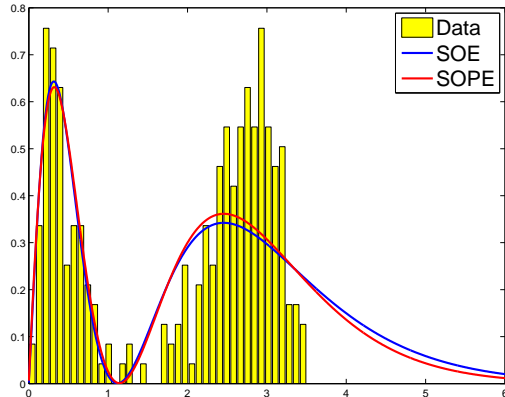


Figure 4: Graphs of the functions from Example 10.

bins and shifted towards the origin with 1.6 minutes, which is the smallest duration of an eruption. Figure 4 shows the resulting histogram and two approximations with  $\kappa = 6$  parameters, a SOE with  $n = 6$  and a SOPE with  $n = 4$  and  $\kappa_1 = 0$ ,  $\kappa_2 = \kappa_3 = 1$ ,  $\kappa_4 = 0$ . The optimization is performed on the interval  $[0, 10]$ ; zero values are appended to the histogram for times greater than the longer duration. The values of the objective are 0.0112 for SOE and 0.0104 for SOPE, so again the SOPE model is more appropriate. The first peak of the data and the valley are quite well followed, while the approximation is worse for the second peak.

A visual comparison with Fig. 4b from [15] shows that our models with only 6 parameters have a better fit than some (older) methods with order 10 models and are only slightly worse than the model proposed in [15]. All the models investigated there are more general than SOPE. Anyway, our purpose in this example is not to prove that SOPE is a good model for the Old Faithful data, but that our method is flexible enough to give reasonably good results in this case. Better approximation of the data can be obtained by shortening the optimization interval, but then, like in [15], the pdf is not strictly decreasing after the second peak, but has a third small peak. ■

## 6 Conclusion and future work

We have presented a method for solving SIP problems with inequality constraints involving functions (1) that are a sums of exponentials multiplied with polynomials with variable coefficients. The method is based on one-sided polynomial approximations of the exponentials on sub-intervals, which allow transformation of SIP into SDP and hence reliable computation of near-optimal solutions that are guaranteed to respect the constraints. We showed through several examples that our method is computationally attractive and gives good results compared to methods based on different principles.

Future work will be dedicated to the investigation of approximations for other functions. The exponential has properties that allowed us several tricks easing the computation of the polynomial approximation. Extension to general functions is not

trivial, but is certainly possible. When implemented for more general categories of functions, our method could be seen as a possible meeting point between CVX and Chebfun, by extending CVX to SIP, using polynomial approximations like Chebfun and having a simple modus operandi like both of them.

## 7 Acknowledgment

This work was supported by the Romanian National Authority for Scientific Research, CNCS - UEFISCDI, project number PN-II-ID-PCE-2011-3-0400.

## References

- [1] F. Alizadeh and D. Papp, *Estimating arrival rate of nonhomogeneous Poisson processes with semidefinite programming*, Ann. Oper. Res. (2013), in press.
- [2] S. Asmussen, O. Nerman, and M. Olsson, *Fitting Phase-Type Distributions via the EM Algorithm*, Scand. J. Stat. 23 (1996), pp. 419–441.
- [3] R. Bojanic and R. DeVore, *On Polynomials of Best One Sided Approximation*, Enseignement Math. 12 (1966), pp. 139–164.
- [4] E. de Klerk, G. Elabwabi, and D. den Hertog, *Optimization of univariate functions on bounded intervals by interpolation and semidefinite programming*, CentER discussion paper 2006-26, Tilburg Univ. (2006).
- [5] O. Devolder, F. Glineur, and Y. Nesterov, *Solving infinite-dimensional Optimization Problems by Polynomial Approximation*, in *Recent Advances in Optimization and its Applications in Engineering*, M.D. et al, ed., Springer, 2010.
- [6] R. DeVore, *One Sided Approximation of Functions*, J. Approx. Theory 1 (1968), pp. 11–25.
- [7] D. Dufresne, *Fitting Combinations of Exponentials to Probability Distributions*, Appl. Stochastic Models Bus. Ind. 23 (2007), pp. 23–48.
- [8] B. Dumitrescu and B.C. Şicleru. Optimization with sums of exponentials and applications. In *Int. Conf. Acoustics Speech Signal Proc.*, pages 370–374, Florence, Italy, 2014.
- [9] T. Fushiki, S. Horiuchi, and T. Tsuchiya, *A Maximum Likelihood Approach to Density Estimation with Semidefinite Programming*, Neural Computation 18 (2006), pp. 2777–2812.
- [10] M. Grant and S. Boyd, *CVX: Matlab software for disciplined convex programming*, <http://cvxr.com/cvx> (2010).
- [11] C. Harris and W. Marchal, *Distribution estimation using Laplace transforms*, INFORMS J. Computing 10 (1998), pp. 448–459.

- [12] D. Henrion, J. Lasserre, and J. Löfberg, *GloptiPoly 3: moments, optimization and semidefinite programming*, Optim. Meth. Software 24 (2009), pp. 761–779.
- [13] D. Henrion, J. Lasserre, and M. Mevissen, *Mean squared error minimization for inverse moment problems*, Appl. Math. Optim. (2012), submitted, <http://homepages.laas.fr/henrion/Papers/density.pdf>.
- [14] V. John, I. Angelov, A. Oncul, and D. Thevenin, *Techniques for the reconstruction of a distribution from a finite number of its moments*, Chem. Eng. Science 62 (2007), pp. 2890–2904.
- [15] K. Kim and N. Thomas, *A fitting method with generalized Erlang distributions*, Simulation Modelling Practice and Theory 19 (2011), pp. 1507–1517.
- [16] C. Lewis, *Computation of Best One-Sided  $L_1$  Approximation*, Math. Comput. 24 (1970), pp. 529–536.
- [17] J. Löfberg and P. Parrilo, *From Coefficients to Samples: a New Approach to SOS Optimization*, in *43rd IEEE Conf. Decision and Control*, Bahamas, 2004, pp. 3154–3159.
- [18] A. Monteiro, R. Tutuncu, and L. Vicente, *Recovering risk-neutral probability density functions from options prices using cubic splines and ensuring nonnegativity*, Eur. J. Oper. Res. 187 (2008), pp. 525–542.
- [19] Y. Nesterov, *Squared Functional Systems and Optimization Problems*, in *High Performance Optimization*, J. Frenk, C. Roos, T. Terlaky, and S. Zhang, eds., Kluwer Academic, 2000, pp. 405–440.
- [20] R. Pachon and L. Trefethen, *Barycentric-Remez algorithms for best polynomial approximation in the chebfun system*, BIT Numer. Math. 49 (2009), pp. 721–741.
- [21] C. Sexton, M. Olivi, and B. Hanzon, *Rational Approximation of Transfer Functions for Non-Negative EPT Densities*, in *IFAC Symposium on System Identification*, July, Brussels, Belgium, 2012, pp. 716–721.
- [22] H. Serali and H. Wang, *Global optimization of nonconvex factorable programming problems*, Math. Program., ser. A 89 (2001), pp. 459–478.
- [23] J. Shortle, *Piecewise Polynomial Approximations for Heavy-Tailed Distributions in Queueing Analysis*, Stochastic Models 21 (2005), pp. 215–234.
- [24] B. Şicleru and B. Dumitrescu, *POS3POLY – a MATLAB Preprocessor for Optimization with Positive Polynomials*, Optimization and Engineering 14 (2013), pp. 251–273.
- [25] L.N. Trefethen, *et al.*, *Chebfun Version 4.2*, The Chebfun Development Team (2011), <http://www.maths.ox.ac.uk/chebfun/>.