OPTIMIZATION WITH SUMS OF EXPONENTIALS AND APPLICATIONS

Bogdan Dumitrescu[1,2], *Bogdan C. Şicleru*[1]

[1]Dept. of Automatic Control and Computers
Politehnica University of Bucharest
313 Spl. Independenţei, 060042 Bucharest, Romania

[2]Dept. of Signal Processing
Tampere University of Technology
P.O.Box 553, SF-33101 Tampere, Finland

e-mail: bogdan.dumitrescu@tut.fi, bogdan.sicleru@acse.pub.ro

ABSTRACT

We present a method for optimization with sums of exponentials subject to positivity constraints and apply it to the modeling of empirical probability distribution functions and to the design of IIR filters with non-negative impulse response. Our approach uses exponents in a sparse arithmetic progression and hence is able to transform the positivity condition to a polynomial form that is computationally tractable. We show how to obtain initial values for the exponents by sparsifying a full progression and then present an iterative optimization procedure using gradient steps. The modeling and design examples indicate a good behavior of our method.

Index Terms— convex optimization, semi-infinite programming, positive polynomials, density modeling, non-negative impulse response, sparse arithmetic progression

1. INTRODUCTION

The broad topic of this paper is the optimization with sum of exponentials (SOE) functions

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

with $\lambda_i \geq 0$ and (without loss of generality) $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$. The SOE is subject to the positivity constraint

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

and possibly other convex constraints. The prototype problem is the least squares optimization

$$\begin{array}{ll} \min & J(f) = \dfrac{1}{M} \displaystyle\sum_{m=1}^{M} w_m [f(t_m) - d_m]^2 \\ \text{s.t.} & f(t) \geq 0, \ \forall t \in [t_0, t_f] \end{array} \tag{3}$$

where we try to fit the SOE to a function $d(t)$, given by samples d_m at times t_m, $m = 1 : M$. The positive numbers w_m represent weights. When the points t_m are equidistant, we replace (3) with $(t_f - t_0)J(f)$, which approximates the value of an integral criterion, discretized with the rectangle method. We note that if the exponents λ_i are given, then the problem (3) is convex. It is still difficult due to its semi-infinite character, caused by the positivity constraint.

The main source of such optimization problems is the modeling of probability density functions (pdf), for which positivity is essential. In this context, a SOE pdf corresponds to a hyperexponential distribution. The values d_m from (3) can come from an empirical pdf. There are applications in various fields, including communications [1], but especially in the financial and insurance [2] domains. For continuous and discrete systems, exponentials are the basic components of non-oscillatory responses and constraints like (2) can model time domain constraints.

Our contribution here is based on a simple remark. If the exponents λ_i form a sparse arithmetic progression (SAP), case in which we will use the abbreviation SOEAP for (1), then the positivity condition (2) can be expressed as the positivity of a polynomial (with sparse coefficients). Hence, the constraint becomes finite and the problem (3) manageable. Although we work with a particular type of SOE, the class of SOEAP models is sufficiently rich for practice. (We note that SAPs are dense in the set of increasing sequences in $\mathbb{R}^n$. However, for practical reasons, the ratio of the progression cannot be arbitrarily small.)

We will show in Section 2 how to obtain a SAP from a full progression, in order to obtain a SOEAP model. Also, we will show how to find a SAP close to a given sequence, with a lower bound on the SAP ratio. These tools are employed in Section 3 for solving (3) and applied to the modeling of empirical pdfs. In section 4 we use a similar procedure for designing IIR filters with non-negative impulse response.

Relation with prior work. We are not aware of any method solving SOE problems with SAPs. A full progression was used in [3], however without taking advantage of polynomial positivity tools. A design of FIR filters with non-negative impulse response was proposed in [4]; a convex optimization approach was given in [5]. We have not found any systematic

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

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

method for the IIR case.

2. SPARSE ARITHMETIC PROGRESSIONS

If the exponents λ_i belong to an arithmetic progression, i.e. there exist integers $\kappa_i \geq 0$ and real numbers $q > 0$ and r such that

$$\lambda_i = \kappa_i q + r, \tag{4}$$

then checking the positivity condition (2) is computationally more tractable, since by denoting $x = e^{-qt}$, the condition becomes

$$\sum_{i=1}^{n} \alpha_i x^{\kappa_i} \geq 0, \ \ \forall x \in [e^{-qt_f}, e^{-qt_0}]. \tag{5}$$

This is the positivity of a polynomial on an interval and can be expressed via an LMI [6, 7]. Note that the case $t_f = \infty$ poses no difficulty and the effective degree of the polynomial, which dictates the size of the LMI and hence the complexity, is $\kappa_n - \kappa_1$, since one can factor out x^{κ_1}.

So, if (4) holds for known κ_i, q and r, the optimization problem (3) is equivalent to an SDP one and can be solved using CVX [8]. The explicit transformation from (5) to an LMI can be avoided by using the library POS3POLY [9], dedicated to optimization with positive polynomials. If the effective degree of the polynomial from (5) is reasonably small, say 100, the problem can be solved quickly and accurately.

We put the above fact to work for solving the problem (3) in two different contexts, in order to obtain good values of the exponents λ_i.

2.1. From full to sparse progression

A model (1) can be general enough only if the exponents belong to a *sparse* arithmetic progression (SAP), i.e. there is no decomposition (4) in which the integers κ_i are consecutive. Finding an optimal SAP for (3) is obviously hard: even if q and r are given, finding the best κ_i is essentially a combinatorial problem. We propose to start with a full progression and use standard tools for enforcing sparsity.

We estimate the smallest exponent with a value $\tilde{\lambda}_1$, either by choosing a sufficiently small value or using the tail of the empirical data d_m, like in [10], since the slowest exponential dominates the others for sufficiently large time values.

Instead of (1), we work with a SOE with $N \gg n$ terms and exponents in arithmetic progression with given ratio q

$$\tilde{f}(t) = \sum_{i=0}^{N} \tilde{\alpha}_i e^{-(\tilde{\lambda}_1 + qi)t} = e^{-\tilde{\lambda}_1 t} \sum_{i=0}^{N} \tilde{\alpha}_i e^{-qit}. \tag{6}$$

We choose a ratio q that is small enough to cover the possible intervals where the exponents lie and a number of terms N that gives acceptable computation times.

We add to the criterion of the optimization problem (3) a term promoting sparsity, namely the 1-norm of the vector of coefficients, transforming it into

$$\begin{array}{ll} \min & J(\tilde{f}) + \beta \sum_{i=0}^{N} \varpi_i |\tilde{\alpha}_i| \\ \text{s.t.} & \tilde{f}(t) \geq 0, \ \forall t \in [0, \infty) \end{array} \tag{7}$$

where β and ϖ_i, $i = 0 : N$, are weighting constants. We choose the value $\varpi_i = 1/\tilde{\lambda}_i$ for the coefficients weights, taking into account that $\int_0^\infty e^{-\lambda t} = 1/\lambda$ and hence normalizing the exponentials from (1). The weight β is chosen by a trial-and-error procedure.

After solving the SDP problem equivalent to (7), we choose the exponents corresponding to the largest value $|\tilde{\alpha}_i|/\tilde{\lambda}_i$ and proceed by solving (3) with these exponents that form a SAP.

2.2. Approximation with a sparse progression

There may be situations (illustrated later) when we are given a set of exponents μ_i, $i = 1 : n$, that do not form a SAP; note that this is actually the generic case. In order to obtain a directly solvable problem (3), we want to find λ_i belonging to a SAP, namely the values κ_i, q and r from (4), such that the distance

$$\delta = \sum_{i=1}^{n} (\lambda_i - \mu_i)^2 \tag{8}$$

is minimized. This distance can be made arbitrarily small (but not necessarily zero); however, to get practically useful solutions we must bound $\kappa_n - \kappa_1$ or, equivalently, to impose $q \geq q_{\min}$. If we want the maximum degree of the polynomial (5) to be at most N, then $q_{\min} = (\mu_n - \mu_1)/N$.

If q is given, then the minimization can be solved exactly. We only sketch here the solution, since the details are straightforward. It is enough to search for $r \in [-q/2, q/2)$. For a given r, the values κ_i are immediately available as $[(\mu_i - r)/q]$, where the brackets denote rounding to the nearest integer. So, the problem reduces to an unidimensional search over r.

The search can be split on $n + 1$ intervals on which the integer values κ_i cannot change their optimal value and hence the criterion (8) is a simple quadratic in r whose minimum is readily available. Let $\mu_i = k_i q + \rho_i$, with $\rho_i \in [-q/2, q/2)$; denote $\varrho_i = \rho_i - q/2$ if $\rho_i \geq 0$ and $\varrho_i = \rho_i + q/2$ otherwise. Then, the nearest SAP term to μ_i is $\kappa_i q + r$ if $r \in [-q/2, \varrho_i]$ and $(\kappa_i - 1)q + r$ if $r \in [\varrho_i, q/2)$. The (sorted) values ϱ_i, $i = 1 : n$, split $[-q/2, q/2)$ in $n + 1$ intervals on which (8) is a quadratic depending only on r.

To find a good value for q, which is a hard problem, we use a random search over the interval $[q_{\min}, 2q_{\min})$, since larger values cannot be better, as they have divisors in this interval. Since we can find quickly the best SAP for given q, we can afford running the random search for tens of q values. From our experience, this seems enough to get a near-optimal SAP approximation to a given sequence with $n \leq 10$.

3. FITTING A SUM OF EXPONENTIALS

Armed with the tools described in the previous section, we can give now the procedure for solving the problem (3).

We first find a SAP set of exponents by solving the problem (7), which allows selecting the most important terms from a full progression. With the selected λ_i, we can solve exactly (3) to get an initial SOEAP solution.

Then, we start an iterative process searching for better values of the exponents. To this purpose, we attempt gradient steps, which produce the exponents

$$\mu_i = \lambda_i - \varsigma \frac{\partial J(f)}{\partial \lambda_i}, \tag{9}$$

where ς is the step length. Since these exponents do not form a SAP, we approximate them with a set of (new) λ_i, as described in section 2.2. The problem (3) can then be solved exactly to provide a new set of coefficients α_i.

The new criterion value is not necessarily smaller than the previous; if this happens, we halve the step value and restore the previous λ_i. The iterative processes is stopped when the step size becomes very small or a maximum number of iterations is reached.

We illustrate the results given by this method with two examples. The weights in (3) are taken all equal, $w_m = 1$.

Example 1 In [10], the target function is

$$d(t) = 16e^{-\frac{t}{2}} - 30e^{-t} + 15e^{-2t} \tag{10}$$

and takes also negative values, hence positivity becomes critical when solving (3). Positivity is imposed on the interval $[0, 10]$, which is large enough to ensure it on $[0, \infty]$. We use a grid with $M = 100$ equidistant points t_m on this interval, on which we compute the values $d_m = d(t_m)$ with (10).

Using the tail of $d(t)$ like in [10], the smallest exponent estimation is $\tilde{\lambda}_1 = 0.4852$. For finding the initial SAP for the exponents, we take $q = 0.01$, $N = 100$ and $\beta = 0.00005$ in (6–7). The obtained exponents values are 0.5052, 1.0852 and 1.4852. In the iterative part of the algorithm, we take $\varsigma = 0.1$ in (9) and perform 30 gradient steps. The whole process needs about 35 seconds on a Dell M4300 laptop. The final value of the criterion is $J = 0.0419$, the SOE being

$$f(t) = 18.19e^{-0.5210t} - 56.36e^{-1.0854t} + 39.12e^{-1.4692t}. \tag{11}$$

Figure 1 presents the graph of this function and of the target (10). Also represented there is the optimal SOE function

$$f(t) = 15.5243e^{-\frac{t}{2}} - 28.5073e^{-t} + 14.2410e^{-2t} \tag{12}$$

with the same exponents as in (10) (forming a SAP), for which the criterion is 0.0736, much higher than for (11). An even higher value results for the best SOE reported in [10],

$$f(t) = 16e^{-\frac{t}{2}} - 29.946e^{-t} + 15.5385e^{-2t}, \tag{13}$$

for which the criterion is $J = 0.1111$. ■

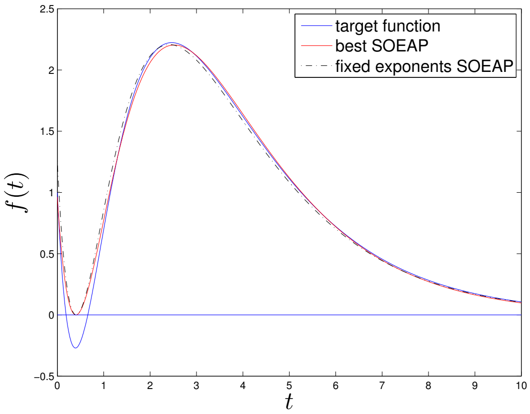

Fig. 1. The SOE functions from Example 1. Blue: target (10). Red: best SOEAP (11). Black, dashed: best approximation (12), with the same exponents as the target.

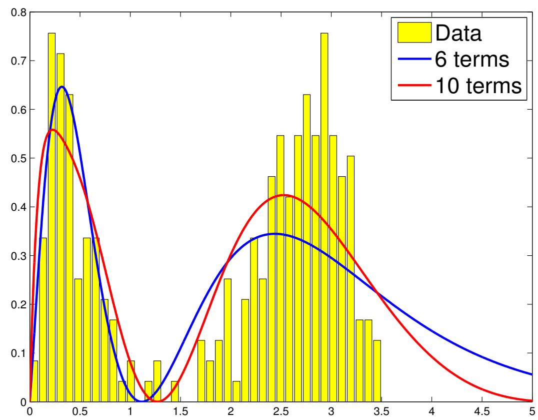

Fig. 2. Results for Example 2.

Example 2 A standard problem in modeling empirical densities is that of eruption durations of the Old Faithful geyser [11]. The data series contains 272 values, grouped into $M = 40$ equally spaced bins and shifted towards the origin with 1.6 minutes, which is the smallest duration of an eruption. Figure 2 shows the graphs of the SOEAP models obtained with our method, with $n = 6$ (blue) and $n = 10$ (red). The criterion values are 0.112 and 0.082, respectively. The optimization was performed with $q = 0.05$, $N = 100$, $\beta = 0.0001$, $\varsigma = 1$. A visual comparison with Fig. 4b from [12] (exact data are not available) shows a quite similar result with the method there (based on a more general model) and much better than in previous works. ■

In other examples (not detailed here), using standard density models (Weibull, Pareto, lognormal), our method showed better behavior, for models with the same order, than the method using exponents in arithmetic progression [3] or the classic Laplace transform method [13].

4. DESIGN OF IIR FILTERS WITH NON-NEGATIVE IMPULSE RESPONSE

A possible use of the SOE model is to impose time-domain constraints on non-oscillatory responses of linear systems. We explore here a single such application, the design of filters with non-negative impulse response [4, 14]. For FIR filters, the constraint can be put directly on the coefficients of the filter and hence is trivial. For IIR filters, the problem is much more difficult, but we can tackle the case of filters with real poles. We write the transfer function of the filter in the simple fractions decomposition

$$H(z) = \sum_{i=1}^{p} \frac{b_i}{1 - a_i z^{-1}} + \sum_{k=0}^{K} c_k z^{-k}. \tag{14}$$

Hence, the impulse response of the filter is

$$h_k = \begin{cases} \sum_{i=1}^{p} b_i a_i^k + c_k, & \text{if } k \le K \\ \sum_{i=1}^{p} b_i a_i^k, & \text{if } k > K \end{cases} \tag{15}$$

If the poles a_i are real and given, we can impose the condition $h_k \ge 0, \forall k \in \mathbb{N}$, as follows. For $0 \le k \le K$ we use directly the first expression from (15), getting $K+1$ linear inequalities in the coefficients b_i, c_k. For $k > K$ we employ the SOEAP tools developed in this paper.

Note that since the filter is stable, we have $|a_i| < 1$ and hence can write $|a_i| = e^{-\lambda_i}$, with $\lambda_i > 0$. If all $a_i > 0$, then we replace the positivity condition for $k > K$ with the continuous version

$$\sum_{i=1}^{p} b_i e^{-\lambda_i t} \ge 0, \; t \in [K+1, \infty). \tag{16}$$

This condition can be implemented as discussed in the previous sections. Note that λ_i belonging to a SAP is equivalent with the poles a_i belonging to a sparse geometric progression.

If the real poles have different signs, then we put separate conditions on even $k = 2\tau$, where $a_i^k = e^{-2\lambda_i \tau}$, and odd $k = 2\tau + 1$, where $a_i^k = a_i e^{-2\lambda_i \tau}$. Then we replace the discrete τ with a continuous t.

The optimization criterion is the distance to the frequency response of an ideal linear phase filter. Assuming that we design a lowpass filter, this response is

$$D(\omega) = \begin{cases} e^{-j\gamma\omega}, & \text{if } 0 \le \omega \le \omega_p \\ 0, & \text{if } \omega_s \le \omega \le \pi \end{cases} \tag{17}$$

The group delay γ, the passband edge ω_p and the stopband edge ω_s are given. Using a grid of discrete frequencies ω_m, $m = 1 : M$, the criterion

$$J = \frac{1}{M} \sum_{m=1}^{M} w_m [H(\omega_m) - D(\omega_m)]^2 \tag{18}$$

is quadratic in the coefficients b_i, c_k. Its optimization with positivity constraints is hence similar to the optimization of the least-squares criterion (3). Since the model (14) may be ill-conditioned, we put a cap on the maximum absolute values of the coefficients (a value of 50 seemed perfectly satisfactory in our experiments).

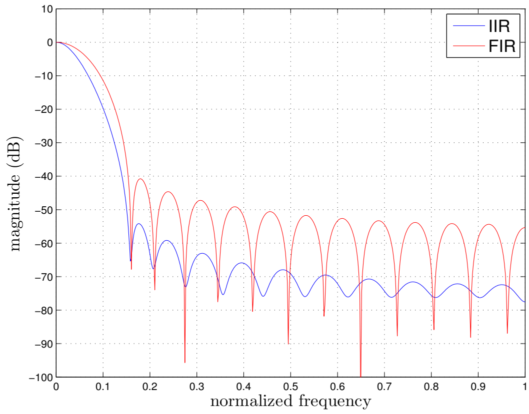

Fig. 3. Filters designed in Example 3.

Example 3 We give a single example of design, with $\omega_p = 0.1\pi$ and $\omega_s = 0.15\pi$. Since the lowpass non-negative filters have a decaying response in the passband, see [14, 15], we take weights $w_m = 1$ in the passband and $w_m = 100000$ in the stopband. The best group delay appear to be around $\gamma = 16$. Using poles in geometric progression starting at 0.9 with ratio 0.95, we note that a sparsifying criterion similar to that from (7) favors the poles with highest magnitude. Keeping $p = 4$ poles and setting $K = 20$, we obtain the filter whose response is shown in Figure 3 (in blue). For comparison, we also design an FIR filter with $p + K$ coefficients and linear phase, whose response is drawn in red (the criterion for the FIR filter takes into account its own group delay, not γ). We note that the IIR filter is sharper and has much better attenuation in the stopband, although the passband is somewhat worse. The value of the criterion is about twice smaller for the IIR filter. This behavior is typical for filters with relatively small number of coefficients. If the number of coefficients is high, then the benefits of the poles are no longer so significant. ■

5. CONCLUSIONS AND FUTURE WORK

Sparse arithmetic progressions have been proved useful in solving optimization problems involving sums of exponentials constrained to positivity. We have presented procedures for modeling an empirical probability density function and for designing IIR filters with non-negative impulse response. Further work will be directed towards applying the new tools to kernel estimation, using symmetrized exponentials centered in an arbitrary point, and also to attempts to generalize the present ideas to the case of complex exponentials.

6. REFERENCES

[1] Y. Liu and A. Tewfik, "Hyperexponential Approximation of Channel Idle Time Distribution with Implication to Secondary Transmission Strategy," in *IEEE Int. Conf. Comm.*, Ottawa, Canada, 2012, pp. 1800–1804.

[2] Z. Shang, M. Goovaerts, and J. Dhaene, "A recursive approach to mortality-linked derivative pricing," *Insurance: Mathematics and Economics*, vol. 40, no. 2, pp. 240–248, Sep. 2011.

[3] D. Dufresne, "Fitting Combinations of Exponentials to Probability Distributions," *Appl. Stochastic Models Bus. Ind.*, vol. 23, pp. 23–48, 2007.

[4] D.A. Dewasurendra, P.H. Bauer, and K. Premaratne, "Evidence Filtering," *IEEE Trans. Signal Proc.*, vol. 55, no. 12, pp. 5796–5805, Dec. 2007.

[5] X. Lai, A. Xue, Z. Lin, and C. Lai, "Minimax design of nonnegative finite impulse response filters," in *15th Int. Conf. Information Fusion*, Singapore, 2012, pp. 1441–1446.

[6] Yu. Nesterov, "Squared Functional Systems and Optimization Problems," in *High Performance Optimiation*, J.G.B. Frenk, C. Roos, T. Terlaky, and S. Zhang, Eds., pp. 405–440. Kluwer Academic, 2000.

[7] B. Dumitrescu, *Positive trigonometric polynomials and signal processing applications*, Springer, 2007.

[8] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming," http://cvxr.com/cvx, Sept. 2010.

[9] B.C. Şicleru and B. Dumitrescu, "POS3POLY – a MATLAB Preprocessor for Optimization with Positive Polynomials," *Optimization and Engineering*, vol. 14, no. 2, pp. 251–273, 2013.

[10] C. Sexton, M. Olivi, and B. Hanzon, "Rational Approximation of Transfer Functions for Non-Negative EPT Densities," in *IFAC Symposium on System Identification*, Brussels, Belgium, July 2012, pp. 716–721.

[11] "Old Faithful geyser data," `http://www.stat.cmu.edu/~larry/all-of-statistics/=data/faithful.dat`.

[12] K. Kim and N. Thomas, "A fitting method with generalized Erlang distributions," *Simulation Modelling Practice and Theory*, vol. 19, pp. 1507–1517, 2011.

[13] C.M. Harris and W.G. Marchal, "Distribution estimation using Laplace transforms," *INFORMS J. Computing*, vol. 10, no. 4, pp. 448–459, 1998.

[14] Y. Liu and P.H. Bauer, "Fundamental Properties of Non-Negative Impulse Response Filters," *IEEE Trans. Circ. Syst. I*, vol. 57, no. 6, pp. 1338–1347, June 2010.

[15] Y. Liu and P.H. Bauer, "Frequency Domain Limitations in the Design of Nonnegative Impulse Response Filters," *IEEE Trans. Signal Proc.*, vol. 58, no. 9, pp. 4535–4546, Sep. 2010.