Opuscula Mathematica • Vol. 31 • No. 4 • 2011

http://dx.doi.org/10.7494/OpMath.2011.31.4.481

A MIXED INTEGER NONLINEAR PROGRAMMING FORMULATION FOR THE PROBLEM OF FITTING POSITIVE EXPONENTIAL SUMS TO EMPIRICAL DATA

Adalys Alvarez, Hugo Lara

Abstract. In this work we deal with exponential sum models coming from data acquisition in the empirical sciences. We present a two step approach based on Tikhonov regularization and combinatorial optimization, to obtain stable parameter estimations, which fit the data. We develop properties of the solutions, based on their optimality conditions. Some numerical experiments are shown to illustrate our approach.

Keywords: mixed integer nonlinear programming, regularization, nonlinear least squares.

Mathematics Subject Classification: 90C11, 90C46, 65F22.

1. INTRODUCTION

Exponential sum models are frequently used, for example, in problems coming from heat diffusion, diffusion of chemical compounds, time series in medicine, economics, physics sciences and technology. For instance, in a positron annihilation lifetime experiment (see [13]) the data collected is a multi-exponential decay spectrum. From this data we wish to extract the intensity spectrum, that is the intensities of the lifetimes present in a given decay curve of the form

$$D(t) = \sum_{l=1}^{k} \rho_l \exp(-\tau_l t).$$

Various approaches have been used to analyze lifetime spectra, and other multi-exponential decay curves, including simple graphical methods [14, 15]; linear and nonlinear least squares fitting [8]; Montecarlo methods and information theory [2, 11]; Bayesian parameter estimation methods [10]; filtering methods by Fourier transform [10, 13], among other. Most of these methods involve some kind of regularization technique, used to deal with the ill-posedness of the problem. A classical regularization approach called Tikhonov regularization (see for example [6]) requires minimizing the nonlinear least squares error plus a term involving the size of the solutions. The aim is to build a better suited minimization problem with a unique solution, and still having valuable information coming from the data. Two approaches have been used to deal with Tikhonov regularization: A dimensional control approach which leads us to a nonlinear regularized least squares problem, where the parameters to be calculated are both the linear intensities ρ and the nonlinear lifetimes τ (see [8]); and a more general regularization approach where a discretization of the τ domain is taken, and a related regularized linear least squares problem is built, where only the intensities ρ should be calculated. This approach seems to be more flexible to add prior information to the problem (see [10]). In this work we develop a two step approach for the treatment of the exponential sums problem by solving above regularized linear least squares problem on the first step. When solving these regularized linear least squares problems, we obtain inaccurate intensity values, because we are ignoring the combinatorial nature of these solutions (see section 3 for details). To overcome those difficulties, we construct a related regularization problem by adding constraints imposing these combinatorial conditions on our problem. The resulting model is a mixed integer nonlinear programming problem, which is harder to solve, but more realistic given the inherent problem conditions. Solving these problems leads us to the second step of our approach. The remainder of this work is as follows: In the next section we establish the problem of fitting k exponential sums to empirical data as a parameter estimation problem to be solved by the Tikhonov regularization technique. Section 3 is devoted to setting up a combinatorial refinement of the Tikhonov problem to be used on the second step. In section 4 we establish some properties of the solutions for the intensities, by studding optimality conditions of the combinatorial problem given in section 3. A linearization of the combinatorial problem, and properties of their solutions are the topics of section 5. In section 6 we consolidate the two step approach, while numerical results ran in test problems are in section 7. In the last section we give some concluding remarks.

2. FITTING THE SUM OF K EXPONENTIALS

In some physics experiments the data collected $\{t_i, y_i\}_{i=1}^n$ consists of a decay curve which can be modeled as a multi-exponential decay function (see for instance Figure 1).

$$D(t) = \sum_{l=1}^{k} \rho_l \exp\left(-\tau_l t\right).$$

We wish to estimate $\hat{\rho}_l, \hat{\tau}_l, l = 1 \ldots, k$ and k to fit the data, that is, we want to solve the parameter estimation problem

$$y_i = D(t_i) + \epsilon_i \quad i = 1, \ldots, n. \tag{2.1}$$

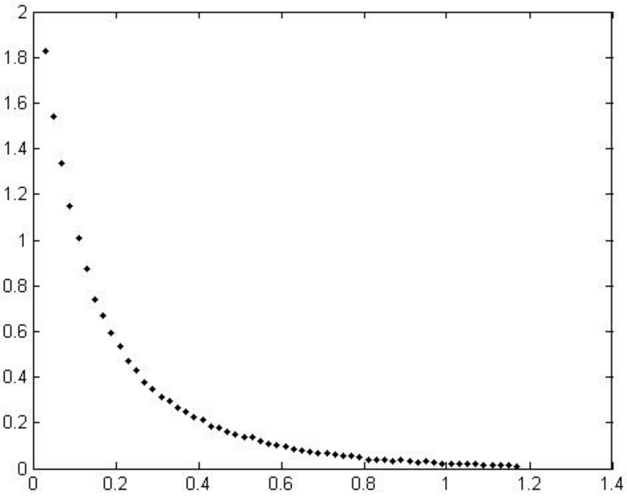

Fig. 1. Data experimental Steyn-Wyk

Two approaches have been used to deal with these kind of problems: (a) a dimensional control approach which consists of a suitable choice of k, and then the linear and nonlinear parameter ρ_l, τ_l are calculated to fit the data, by solving iteratively a nonlinear least squares problem, like in [8]; or (b) the more general regularization approach where N basis functions $\exp(-\bar{\tau}_j t)$, $j = 1 \ldots, N$ are chosen, each one defined at points $\{\bar{\tau}_1, \ldots, \bar{\tau}_N\}$, a discretization of the τ domain; and then substituting $D(t)$ by $\bar{D}(t) = \sum_{j=1}^{N} \rho_j \exp(-\bar{\tau}_j t)$ which is for fixed t a linear function on the ρ parameters. Ideally, if we choose the index set $\{j_1, \ldots, j_k\} \subseteq \{1, \ldots, N\}$ such that $\tau_{j_l} \approx \hat{\tau}_l$ for $l = 1, \ldots, k$ then the corresponding linear parameter values ρ_{j_l} should be set to $\hat{\rho}_l$, and for $j \notin \{j_1, \ldots, j_k\}$, $\rho_j \approx 0$. This observation says that for good choices of the parameters, function $\bar{D}(t)$ should be close to $D(t)$. The linearity of the resulting parameter estimation problem allows us to use robust tools to deal linear inverse problems; and according to [10], this second approach is more flexible for adding general prior information on the parameters. Our fitting problem is known to be ill posed [6, 8, 10, 13]. In this work we use the second approach.

Let us denote by M the matrix defining the linear relation in $\bar{D}(t_i)$, that is $M_{ij} = \exp(-\bar{\tau}_j t_i)$ with $j = 1, \ldots, N$. Now problem (2.1) becomes

$$y_i = \sum_{j=1}^{N} \rho_j M_{ij} + \epsilon_i \quad i = 1, \ldots, n. \tag{2.2}$$

We can write (2.2) in a compact way as

$$y = M\rho + \epsilon. \tag{2a}$$

It is known [8, 10, 13] that the sum of exponentials of the fitting problem is ill posed, and regularization techniques to obtain approximate solutions are necessary. In a regularization scheme, like Tikhonov [6, 7, 12], we choose ρ in a such way that

$$\frac{1}{2}\left\|\begin{pmatrix} M \\ \lambda L \end{pmatrix}\rho - \begin{pmatrix} y \\ \lambda L\bar{\rho} \end{pmatrix}\right\|^2, \tag{2.3}$$

is minimized, where λ is the Tikhonov regularization parameter, L is a known as the scaling matrix, possibly the identity matrix, and $\bar{\rho}$ is a default solution representing our previous knowledge about the problem. Denote by $\rho(\lambda)$ the solution of (2.3). If λ approaches zero, then $\rho(\lambda)$ tends to the least squares solution of (2a). If $\lambda \to \infty$, then $\rho(\lambda) \to \bar{\rho}$, the default solution. In classic regularization, the parameter choice plays a central role, because it controls the importance given to the regularization term in the minimization; that is the quality of the regularized solution is tuned by λ. An optimal regularization parameter value should adequately balance the importance of the perturbation error and the regularization term in the minimization problem. There are several strategies to define a selection criteria for the regularization parameter, as the discrepancy principle, the cross-validation method and the maxima curvature of the L-curve. The L-curve is a plot of the norm of the regularized solution versus the residual norm [4–7, 12]. Typically this L-shaped parametrized curve has smooth decreasing regions at the ends; an almost vertical part followed by an almost horizontal part connected by a corner, where the curvature is a maxima (see Figure 2). Such criteria looks for a value of the parameter λ next to the corner, where the importance of the residual and the regularized solution norm is balanced.

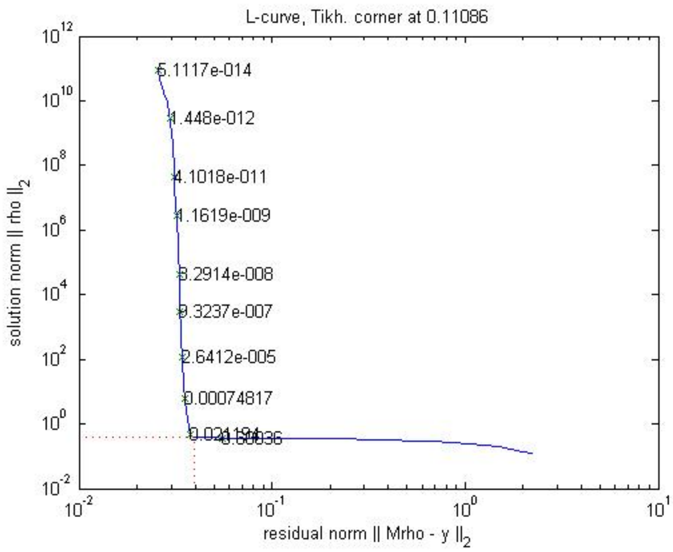

Fig. 2. L-Curve

3. REGULARIZED LEAST SQUARES SOLUTION

If we have the data $\{t_i, y_i\}_{i=1}^{n}$ and no additional information about the solutions is available, then the Tikhonov problem we should solve is

$$\text{minimize} \quad \frac{1}{2}\left\| \begin{pmatrix} M \\ \lambda I \end{pmatrix} \rho \; - \; \begin{pmatrix} y \\ 0 \end{pmatrix} \right\|^2 = \frac{1}{2}\|y - M\rho\|^2 + \frac{\lambda^2}{2}\|\rho\|^2 \qquad (3.1)$$

and by choosing the value of the regularization parameter by the L-curve criterium, we obtain a regularized least squares solution, denoted by $\bar{\rho}(\bar{\lambda}) = \bar{\rho}$. Solutions for (3.1) can be found by solving a system of linear equations, the so called normal equations. In this work we choose $\bar{\lambda}$ by using the regularization toolbox of Hansen [4, 5], and $\bar{\rho}$ by using the nonnegative least squares solver from the Matlab optimization toolbox. If we plot $\bar{\rho}$ versus the index set corresponding to the domain discretization of τ we obtain a smooth curve, with amplitude values close to zero in points far from $\hat{\tau}_j$, and peaks around points $\hat{\tau}_l$, with amplitudes $\hat{\rho}_l$ given by the height of the peaks. Even though this curve gives us some useful information, because approximated nonlinear parameter values $\hat{\tau}_l$ are identified, the amplitude values $\hat{\rho}_l$ result far from the true values, because points near $\hat{\tau}_l$ are expected to obtain zero amplitude we have positive values, which are possibly contributions for the amplitude at $\hat{\tau}_l$ disseminated at the neighbours (see Figure 3). Some other considerations should be taken to prevent from this behavior. In the next part we propose a combinatorial model to improve the accuracy when calculating these parameters.

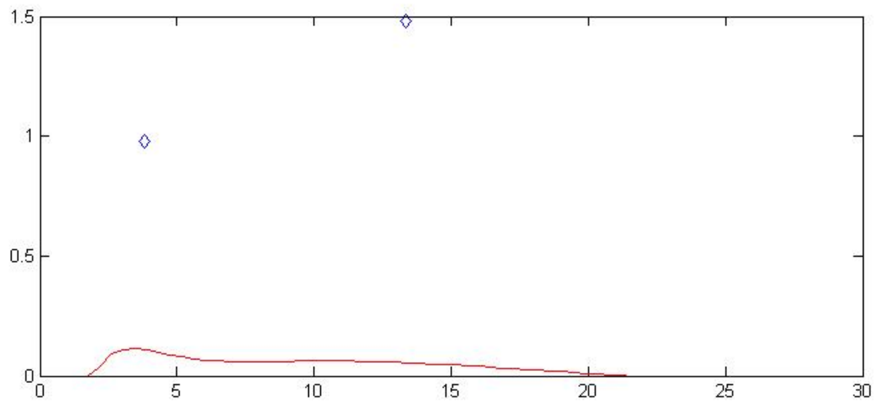

Fig. 3. Regularized least squares vs true solution

3.1. A COMBINATORIAL MODEL

In the spirit of the Tikhonov regularization scheme, with previous or historical information about the problem, we get additional elements in order to improve our model (3.1). We know that the amplitudes ρ_j should be non negatives and that only k of them are strictly positive (represented by peaks at $\bar{\rho}$). This behavior can be modeled by imposing to each point of the sequence ρ_j the condition $\rho_j = x_j q_j$ with

$$x_j \geq 0, \quad q_j \in \{0,1\} \quad \text{and} \quad e^T q \leq k,$$

like in [3]. In this way the solutions should be less smooth than $\bar{\rho}$, but closer to the characteristics of the desired solutions; because in the case of $q_j = 0$, the amplitude should be zero, but if $q_j = 1$ more accuracy of ρ_j should be achieved. We deal with the condition that only k of the variables should be positives by imposing $\sum q_i \leq k \quad (e^T q \leq k)$ as a constraint. In addition, if we calculate in a first step $\bar{\rho}(\bar{\lambda}) = \bar{\rho}$, then we can use this approximate solution as a default solution for the Tikhonov problem, resulting

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|y - M\rho\|^2 + \frac{\lambda^2}{2}\|\rho - \bar{\rho}\|^2 \\ \text{subject to} & \rho = x \bullet q \\ & e^T q \leq k \\ & x \geq 0, \rho \geq 0, q \in \{0,1\}^N, \end{array} \tag{3.2}$$

where $\bullet$ denotes the componentwise product between vectors. Problem (3.2) is a mixed integer nonlinear programming problem (MINLP).

In practice, constraints of the type $\rho_j = x_j q_j$ can produce tricky values. For instance, if ρ_j is calculated as zero, this could happen either because $q_j = 0$ or because $q_j = 1$, but x_j approaches zero. The last behavior is undesired, since we should be detecting a peak at q_j, but with the x_j value near zero. On the other hand, if $q_j = 0$, then ρ_j is also calculated as zero, but this time with a large value for the corresponding x_j, distorting the relationship among the variables. In order to prevent the problem of these undesired behaviors, we can model the regularization term by adding $\frac{\lambda^2}{2}\|x - \bar{\rho}\|^2 + \frac{\lambda^2}{2}\|q - \bar{\rho}\|^2$ to the objective function at (3.2). This allows us to keep values of x and q close to a known approximate solution, and so avoid those tricky behaviors. We establish our improved MINLP model as

$$\begin{array}{ll} \text{minimize} & f(\rho, x, q) = \frac{1}{2}\|y - M\rho\|^2 + \frac{\lambda^2}{2}\|\rho - \bar{\rho}\|^2 + \frac{\lambda^2}{2}\|x - \bar{\rho}\|^2 + \frac{\lambda^2}{2}\|q - \bar{\rho}\|^2 \\ \text{subject to} & \rho = x \bullet q \\ & e^T q \leq k \\ & x \geq 0, \rho \geq 0, q \in \{0,1\}^N. \end{array} \tag{3.3}$$

Note that this formulation leads to an NP-hard problem. Since the original problem is essentially easier, we can ask, why to construct a harder one? Two answers appear: First, we get more accuracy in calculating the amplitudes, by avoiding assigning positive values to amplitudes corresponding to the nonlinear parameter values close to the chosen ones; and secondly, we can use heuristics, or approximating solvers to deal with the combinatorial problem (see for instance [1]), which are computationally less expensive, providing efficient tools to cope with the problem.

3.2. CHARACTERIZATION OF THE SOLUTIONS

In this part we study some properties of the solutions of problem (3.3), and characterize the amplitude values in terms of only continuous variables. To do so, we consider a dualization of problem (3.3). Let us denote by $\psi = (x^T, \rho^T, s)^T$ the continuous variables, by $g(\psi, q) \in \mathbf{R}^{N+1}$, the equality constraints after inclusion of the necessary slack variable s and the objective function by $\hat{\phi}(\psi, q) = \phi(\psi) + \frac{\lambda^2}{2}\|q - \bar{\rho}\|^2$, where $\phi(\psi)$ is the portion of f independent of q. Let $\mathcal{X} = \mathbf{R}_+^{2N+1} \times \{0,1\}^N$ be the primal feasible set. Then we write (3.3) as

$$\begin{array}{ll} \text{minimize} & \phi(\psi) + \frac{\lambda^2}{2}\|q - \bar{\rho}\|^2 \\ \text{subject to} & g(\psi, q) = 0 \\ & \psi \geq 0 \\ & (\psi, q) \in \mathcal{X}. \end{array} \tag{3.4}$$

We also denote by $\mu = ({\mu^1}^T, \mu^2, {\mu^3}^T)^T \in \mathcal{Y} = \mathbf{R}^N \times \mathbf{R} \times \mathbf{R}_+^{2N}$ the dual variables and dual feasible set respectively. Then we establish the Lagrangean function as

$$\mathcal{L}((\psi, q), \mu) = \phi(\psi) + \frac{\lambda^2}{2}\|q - \bar{\rho}\|^2 - \begin{pmatrix} \mu^1 \\ \mu^2 \end{pmatrix}^T g(\psi, q) - {\mu^3}^T \psi,$$

therefore the dual Lagrangean function is defined for $\mu \in \mathcal{Y}$ as

$$\mathcal{L}_*(\mu) = \inf_{(\psi,q)\in\mathcal{X}} \mathcal{L}((\psi, q), \mu),$$

and the Lagrangean dual problem for

$$\underset{\mu\in\mathcal{Y}}{\text{maximize}}\, \mathcal{L}_*(\mu). \tag{3.5}$$

The next result characterizes the Lagrangean dual function.

Lemma 3.1. *Given $\mathcal{X}, \mathcal{Y}$ and $\mathcal{L}, \mathcal{L}_*$ defined as above. Then for $\mu \in \mathcal{Y}$*

$$\mathcal{L}_*(\mu) = \inf_{(\psi,q)\in\mathcal{X}} f_1(\psi) + \mu^T f_2(\psi) + \left(f_3(\psi)^T \mu - \lambda^2\bar{\rho} + \frac{\lambda^2}{2} q\right)^T q,$$

where

$$f_1(\psi) = \phi(\psi) + \frac{\lambda^2}{2}\bar{\rho}^T\bar{\rho} \in \mathcal{R}, \tag{3.6}$$

$$f_2(\psi) = (-\rho^T, s - k, -\psi^T)^T \in \mathcal{R}^N \times \mathcal{R} \times \mathcal{R}^{2N}, \tag{3.7}$$

and

$$f_3(\psi) = [X, e, 0]^T \in \mathcal{R}^{N\times N} \times \mathcal{R}^{N\times 1} \times \mathcal{R}^{N\times 2N}, \tag{3.8}$$

with $X := diag(x)$ stands for the $N \times N$ diagonal matrix which diagonal entries are the components of vector x.

Proof. First we write the Lagrangean function as

$$\begin{aligned}
\mathcal{L}((\psi,q),\mu) &= \phi(\psi)+\frac{\lambda^2}{2}\|q-\bar{\rho}\|^2-({\mu^1}^T\mu^2)g(\psi,q)-{\mu^3}^T\psi=\\
&=\phi(\psi)+\frac{\lambda^2}{2}(\bar{\rho}^T\bar{\rho}-2\bar{\rho}^Tq+q^Tq)-({\mu^1}^T,\mu^2)\begin{pmatrix}\rho-xq\\ k-e^Tq-s\end{pmatrix}-{\mu^3}^T\psi=\\
&=\phi(\psi)+\frac{\lambda^2}{2}\bar{\rho}^T\bar{\rho}+({\mu^1}^T,\mu^2,{\mu^3}^T)\begin{pmatrix}-\rho\\ s-k\\ -\psi\end{pmatrix}+\\
&\quad+\left[\frac{\lambda^2}{2}q^T-\lambda^2\bar{\rho}^T+\mu^T\begin{pmatrix}X\\ e^T\\ 0\end{pmatrix}\right]q=\\
&=f_1(\psi)+\mu^Tf_2(\psi)+\left[\frac{\lambda^2}{2}q^T-\lambda^2\bar{\rho}^T+\mu^Tf_3(\psi)\right]q.
\end{aligned}$$

By taking infimum on this expression for fixed μ we get the result. $\square$

We now analyze the minimization at the evaluation of $\mathcal{L}_*$ with respect to the binary variables. Suppose that $(\bar{\psi},\bar{q})$ is an optimal solution of $\inf\mathcal{L}\left((\psi,q),\mu\right)$ for fixed μ, and look at the minimization with respect to q. It is clear that

$$\min_q \mathcal{L}\left(\left(\bar{\psi},q\right),\mu\right)=\mathcal{L}_*(\mu). \tag{3.9}$$

Let us study this minimization: The quadratic term

$$\left(f_3(\bar{\psi})^T\mu-\lambda^2\bar{\rho}+\frac{\lambda^2}{2}q\right)^Tq$$

is minimized componentwise for $q_i=0$ if $(\mu^Tf_3(\bar{\psi})_i-\lambda^2\bar{\rho}_i+\frac{\lambda^2}{2})>0$ and $q_i=1$ in the case $(\mu^Tf_3(\bar{\psi})_i-\lambda^2\bar{\rho}_i+\frac{\lambda^2}{2})\le 0$. Therefore, the dual function can be evaluated by

$$\mathcal{L}_*(\mu)=f_1(\bar{\psi})+\mu^Tf_2(\bar{\psi})+\sum_{i\in I(\bar{\psi},\mu)}\left(\mu^Tf_3(\bar{\psi})_i-\lambda^2\bar{\rho}_i+\frac{\lambda^2}{2}\right)$$

for $I(\psi,\mu)=\left\{i\in\{1,\dots,N\}:\mu^Tf_3(\psi)_i-\lambda^2\bar{\rho}_i+\frac{\lambda^2}{2}\le 0\right\}$, where $f_3(\psi)_i$ stands for the i-th column of $f_3(\psi)$. This proves the following lemma:

Lemma 3.2. *Given* $(\bar{\psi},\bar{q})$ *the minimizer of* (3.9)*;*

$$I(\psi,\mu)=\left\{i\in\{1,\dots,N\}:\mu^Tf_3(\psi)_i-\lambda^2\bar{\rho}_i+\frac{\lambda^2}{2}\le 0\right\},$$

for f_3 *defined as above, and fixed* $\mu>0$*, then*

$$\mathcal{L}_*(\mu)=f_1(\bar{\psi})+\mu^Tf_2(\bar{\psi})+\sum_{i\in I(\bar{\psi},\mu)}\left(\mu^Tf_3(\bar{\psi})_i-\lambda^2\bar{\rho}_i+\frac{\lambda^2}{2}\right).$$

A simple consequence establishes that

$$\mathcal{L}_*(\mu) = \min_{\psi} f_1(\psi) + \mu^T f_2(\psi) + \sum_{i\in I(\psi,\mu)} \left(\mu^T f_3(\psi)_i - \lambda^2\bar{\rho}_i + \frac{\lambda^2}{2}\right).$$

These relation expresses the dual objective function in terms of only continuous variables ψ and μ.

We now study the optimal solutions of problem (3.3) in terms of the continuous primal and dual variables.

Theorem 3.3. *Given* $(\psi^*, q^*), \mu^*$ *primal and dual optimal solutions of* (3.4)*, and* (3.5) *respectively, then*

$$\rho_i^* = \begin{cases} x_i^* & \text{if } \frac{\lambda^2}{2} \le \lambda^2\bar{\rho}_i + x_i^*\mu_i^{1*} + \mu^{2*}, \\ 0 & \text{if } \frac{\lambda^2}{2} > \lambda^2\bar{\rho}_i + x_i^*\mu_i^{1*} + \mu^{2*}. \end{cases}$$

Proof. Observe that $q_i^* = 1$ if $\mu^T f_3(\psi)_i - \lambda^2\bar{\rho}_i + \frac{\lambda^2}{2} \le 0$ which is equivalent to $\frac{\lambda^2}{2} \le \lambda^2\bar{\rho}_i + x_i^*\mu_i^{1*} + \mu^{2*}$; and $q_i^* = 0$ otherwise. Since $\rho_i^* = x_i^* q_i^*$ then we have the result. □

4. LINEARIZATION

In this section we transform problem (3.3) to a quadratic problem with linear constraints, in order to get access to more efficient solvers. The constraints $\rho_j = x_j q_j$ are quadratic, but with the following substitution we can model it with linear constraints:

$$\begin{gathered} -\gamma q_j \le \rho_j \le \gamma q_j, \\ -(1-q_j)\gamma \le \rho_j - x_j \le (1-q_j)\gamma \end{gathered}$$

with γ a large penalization value. If $q_j = 0$, then the first group of constraints ensures that $\rho_j = 0$. If $q_j = 1$, then $\rho_j - x_j = 0$ is a consequence of the second group of constraints. In this way the constraints become linear, with continuous and binary variables. Since $\rho_j \ge 0$, then the first group of constraints can be written as $0 \le \rho_j \le \gamma q_j$. Finally, we can express problem (3.3) equivalently as

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|y - M\rho\|^2 + \frac{\lambda^2}{2}\|\rho - \bar{\rho}\|^2 + \frac{\lambda^2}{2}\|x - \bar{\rho}\|^2 + \frac{\lambda^2}{2}\|q - \bar{\rho}\|^2 \\ \text{subject to} & \rho - \gamma q \le 0 \\ & -\rho + x + \gamma q \le \gamma e \\ & \rho - x + \gamma q \le \gamma e \\ & e^T q \le k \\ & x, \rho \ge 0, q \in \{0,1\}^N, \end{array} \tag{4.1}$$

which is a mixed integer quadratic program with linear constraints.

For problem (4.1) we denote $\bar{\phi}(\psi) = \frac{1}{2}\|y - M\rho\|^2 + \frac{\lambda^2}{2}\|\rho - \bar{\rho}\|^2 + \frac{\lambda^2}{2}\|x - \bar{\rho}\|^2$, the portion of the objective function depending only on continuous variables ψ;

$$A = \begin{bmatrix} I & 0 \\ -I & I \\ I & -I \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} -\gamma I \\ \gamma I \\ \gamma I \\ e^T \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ \gamma e \\ \gamma e \\ k \end{bmatrix} \quad \text{and} \quad \psi = \begin{bmatrix} \rho \\ x \end{bmatrix},$$

denotes the matrices defining the linear constraints in the above problem; obtaining the following quadratic program with linear constraints:

$$\begin{aligned} &\text{minimize} \quad \bar{\phi}(\psi) + \frac{\lambda^2}{2}\|q - \bar{\rho}\|^2 \\ &\text{subject to} \quad A\psi + Bq \leq b, \\ &\qquad\qquad\quad \psi \geq 0, q \in \{0,1\}^N. \end{aligned} \tag{4.2}$$

In order to dualize this program, we consider $\mathcal{X} = \mathbf{R}_+^{2N} \times \{0,1\}^N$ as the primal feasible set; we define $\mu \in \mathcal{Y} = \mathbf{R}_+^{3N+1}$ the Lagrange multipliers associated with linear constraints $A\psi + Bq \leq b$, and obtain the associated Lagrangean function:

$$\mathcal{L}((\psi, q), \mu) = \phi(\psi) + \frac{\lambda^2}{2}\|q - \bar{\rho}\|^2 - \mu^T(b - A\psi - Bq) \quad \text{for} \quad (\psi, q) \in \mathcal{X} \quad \text{and} \quad \mu \in \mathcal{Y}.$$

We define the Lagrangean dual function as

$$\mathcal{L}_*(\mu) = \inf_{(\psi,q)\in\mathcal{X}} \mathcal{L}((\psi, q), \mu),$$

for each $\mu \in \mathcal{Y}$. The associated dual problem is then:

$$\sup_{\mu\in\mathcal{Y}} \mathcal{L}_*(\mu). \tag{4.3}$$

Let us observe in detail the optimization problem in the evaluation of $\mathcal{L}_*(\mu)$:

$$\begin{aligned} \mathcal{L}_*(\mu) &= \inf_{(\psi,q)\in\mathcal{X}} \phi(\psi) + \frac{\lambda^2}{2}\|q - \bar{\rho}\|^2 - \mu^T(b - A\psi - Bq) = \\ &= \inf_{(\psi,q)\in\mathcal{X}} \phi(\psi) + \frac{\lambda^2}{2}[\bar{\rho}^T\bar{\rho} - 2\bar{\rho}^T q + q^T q] - \mu^T b + \mu^T A\psi + \mu^T Bq = \\ &= \inf_{(\psi,q)\in\mathcal{X}} \phi(\psi) + \frac{\lambda^2}{2}\bar{\rho}^T\bar{\rho} + \mu^T(-b + A\psi) + \left(-\lambda^2\bar{\rho} + (\mu^T B)^T + \frac{\lambda^2}{2}q\right)^T q, \end{aligned}$$

which proves the following result:

Lemma 4.1. *Given problem* (4.2) *and its dual* (4.3), *we have*

$$\mathcal{L}_*(\mu) = \inf_{(\psi,q)\in\mathcal{X}} g_1(\psi, \mu) + \left(g_2(\mu) + \frac{\lambda^2}{2}q\right)^T q, \tag{4.4}$$

where $g_1(\psi, \mu) = \phi(\psi) + \frac{\lambda^2}{2}\bar{\rho}^T\bar{\rho} + \mu^T(-b + A\psi)$ *and* $g_2(\mu) = B^T\mu - \lambda^2\bar{\rho}$.

In the expression (4.4) we split the continuous and binary parts, to expose better the structure of the problem. We use the lemma above to study the inherent minimization. It is easy to verify that if the infimum is achieved at a point, say (ψ_μ, q_μ) then

$$\mathcal{L}_*(\mu) = \min_{\psi \in \mathbf{R}_+^{2N}} g_1(\psi, \mu) + \left(g_2(\mu) + \frac{\lambda^2}{2} q_\mu\right)^T q_\mu \tag{4.5}$$

and also

$$\mathcal{L}_*(\mu) = \min_{q \in \{0,1\}^N} g_1(\psi_\mu, \mu) + \left(g_2(\mu) + \frac{\lambda^2}{2} q\right)^T q. \tag{4.6}$$

The minimization in (4.6) is achieved componentwise with $q_i = 0$ if $g_2(\mu) + \frac{\lambda^2}{2} > 0$, and $q_i = 1$ in the case that $g_2(\mu) + \frac{\lambda^2}{2} \leq 0$. Therefore, the dual function can be evaluated as:

$$\begin{aligned}\mathcal{L}_*(\mu) &= g_1(\psi_\mu, \mu) + \sum_{i \in I(\mu)} \left(g_2(\mu)_i + \frac{\lambda^2}{2}\right) = \\ &= \min_{\psi \in \mathbf{R}_+^{2N}} g_1(\psi, \mu) + \sum_{i \in I(\mu)} \left(g_2(\mu)_i + \frac{\lambda^2}{2}\right)\end{aligned}$$

with $I(\mu) = \left\{ i \in \{1, \ldots, N\} : g_2(\mu) + \frac{\lambda^2}{2} \leq 0 \right\}$. In this way, the dual function is evaluated by choosing the values of q_i, according to ψ and μ. We shall study now a characterization of the intensity values ρ_i^* in terms of the primal-dual optimal values for (4.2) and (4.3).

Theorem 4.2. *Let us consider* (ψ^*, q^*) *an optimal solution for* (4.2)*, and* μ^* *an optimal solution for* (4.3)*. Then:*

(i)

$$\begin{gathered}\rho_i^* = x_i^* > 0 \textit{ if and only if } \mu_i^{1^*} = 0,\ \mu_i^{2^*}, \mu_i^{3^*} \geq 0 \\ \textit{and } \lambda^2 \left(\tfrac{1}{2} - \bar{\rho}_i\right) < \gamma(\mu_i^{2^*} + \mu_i^{3^*}) + \mu^{4^*}.\end{gathered} \tag{4.7}$$

(ii)

$$\begin{gathered}\rho_i^* = 0 \textit{ if and only if } \mu_i^{1^*} > 0,\ \mu_i^{2^*}, \mu_i^{3^*} = 0 \\ \textit{and } \lambda^2 \left(\tfrac{1}{2} - \bar{\rho}_i\right) \geq -\gamma \mu_i^{1^*} + \mu^{4^*}.\end{gathered} \tag{4.8}$$

Proof. Let us take fixed q^* and μ^*; and define the minimization problem in ψ.

$$\underset{\psi \in \mathbf{R}_+^{2N}}{\text{minimize}}\, g_1(\psi, \mu^*) + \left(g_2(\psi, \mu^*) + \frac{\lambda^2}{2} q^*\right)^T q^*. \tag{4.9}$$

Complementarity at the optimality conditions for problem (4.1) establish that

$$\begin{aligned}(\rho_i - \gamma q_i^*)\mu_i^1 &= 0, \quad i = 1, \ldots, N. \\ (x_i - \rho_i + \gamma(q_i^* - 1))\mu_i^2 &= 0, \quad i = 1, \ldots, N. \\ (-x_i + \rho_i + \gamma(q_i^* - 1))\mu_i^3 &= 0, \quad i = 1, \ldots, N,\end{aligned}$$

but if $\mu = \mu^*$ is an optimal solution for (4.3) then it should also verify these complementarity conditions. Let us prove (i). If $\rho_i^* = x_i^* > 0$ then for the second and third inequality in (4.1), we conclude that $q_i^* = 1$. Therefore, by the first complementarity equation, $\mu_i^{1^*} = 0$, and for dual feasibility $\mu_i^{2^*}, \mu_i^{3^*} \geq 0$. The relation $g_2(\mu)_i + \frac{\lambda^2}{2} q_i$ is equivalent to

$$\gamma(-\mu_i^1 + \mu_i^2 + \mu_i^3) + \mu^4 - \lambda^2 \bar{\rho}_i + \frac{\lambda^2}{2}. \tag{4.10}$$

Since $\mu_i^{1^*} = 0$, then the expression (4.10) evaluated at μ^* becomes

$$\gamma(\mu_i^2 + \mu_i^3) + \mu^4 - \lambda^2 \bar{\rho}_i + \frac{\lambda^2}{2}$$

if this expression was strictly positive, then we could find a better solution with $q_i^* = 0$, therefore, it should be less than 0, which gives us

$$\gamma(\mu_i^2 + \mu_i^3) + \mu^4 < \lambda^2 \left(\bar{\rho}_i - \frac{1}{2} \right). \tag{4.11}$$

Reciprocally, suppose that (4.11) holds and that $\mu_i^{1^*} = 0$, $\mu_i^{2^*}, \mu_i^{3^*} \geq 0$, then

$$g_2(\mu^*)_i + \frac{\lambda^2}{2} q_i^* < 0;$$

therefore $q_i^* = 1$ and for feasibility conditions $\rho_i^* = x_i^* > 0$. A proof for (ii) is similar, taking into account that in this case

$$g_2(\mu^*)_i + \lambda^2 q_i^* = \gamma(-\mu_i^1) + \mu^4 - \lambda \bar{\rho} + \frac{\lambda^2}{2}.$$

□

5. TWO STEPS APPROACH

In this section we consolidate the two steps approach to deal with the exponential sum estimation problem. For the first step we solve the linear least squares problem with Tikhonov regularization (3.1), with the parameter value given by the L-curve criterion [4–7], fixing an approximated stable solution $\bar{\rho}$ and also the regularization parameter value $\hat{\lambda}$. At the second step we get the information obtained at the first step, that is the approximated solution $\bar{\rho}$ and parameter value λ to then solve problem (4.1) in order to refine values of ρ, which gives us a non smooth solution, but more realistic than the one given by regularized least squares only. We present an algorithmic two step scheme:

ALGORITHM

Step I

Input: $\{t_i, y_i\}_{i=1}^n$; data of model (3.1), $\quad M \in \mathbf{R}^{n \times N}$.

– For $\bar{\rho} = 0, L = I$ calculate $\hat{\lambda}$ by the L-curve criterion, and $\rho(\hat{\lambda})$ as the solution of (3.1).

Output: $\hat{\lambda}, \bar{\rho} = \rho(\hat{\lambda})$.

Step II

Input: $M, y, k, \hat{\lambda}, \bar{\rho}, \gamma$.

– Solve problem (4.1).

Output: $\rho^*, \tau^*, x^*, q^*, \mu^*$.

6. NUMERICAL EXPERIMENTS

We have perform several numerical experiments on problems involving parameter estimation of exponential functions models, by using a Matlab code involving the two step approach mentioned at the previous section. At the first for a given data and upper bound on the nonlinear parameter, once the discretization of the nonlinear parameter is done, the regularization parameter $\hat{\lambda}$ is chosen by the corner of the L-curve criterion. Without prior information about the intensities, we solve the Tikhonov problem (3.1), obtaining intensities $\bar{\rho}$ which constitute the entries for the second step. In this second phase we use a quadratic mixed integer optimization routine given by Bemporad and Mignone [1], which allows us to refine the values of ρ and τ coming from the first step.

For this presentation we use four data sets. The first two are obtained by simulating, for $t \in [0, 5]$ and $\Delta t = 0.0505$, from

$$f(t) = \exp(-1.5t) + 2\exp(-3t) + 4\exp(-2t),$$

and

$$f(t) = \exp(-1.5t) + 4\exp(-3t) + 2\exp(-2t) + 3\exp(-6t).$$

The third one is the classical Lanczos serie [9], generated in time steps $\Delta t = 0.05$ at the interval $[0, 1.15]$ from

$$f(t) = 0.0951\exp(-t) + 0.8607\exp(-3t) + 1.5576\exp(-5t).$$

The fourth problem corresponds with the empirical data given by Steyn and Wyk [14], this data is shown in Figure 1, where 58 observations of the intensities of secondary particles of neutrons coming from cosmic rays. Here the variation of the time is 20 ms in a range of 30 to 1170 ms.

The last two problems have been studied by Petersson and Holmström by using the TOMLAB optimization environment [8].

7. RESULTS

In this section we present results for our four numerical experiments in two tables. At the first column we identify the problem, give the size of the discretization and the used regularization parameter. The normalized true linear values are at the second column and at the third one the nonlinear true values. Results of step one are given at columns 4, 5 and 6; showing the indices of the discretization that results in positive values for the linear parameters and values of both of the parameters at these indices. Columns 7, 8 and 9 show second step results, and finally the two last columns contain relative errors with respect to the true values. We also provide figures showing comparative results. The diamond shaped mark indicates the true values of τ and ρ. The pointed line showing the first step solution (RLSS) and the continuous line the refinement of the second step (MINLS). At Table 1 we show results for problems 1 and 2 corresponding to models with three and four exponential functions respectively.

Table 1. Problems 1 and 2

Problem	ρ	τ	j	τ_j	$\bar{\rho}_j$	l	ρ_l^*	τ_l^*	elin	enl
Problem 1	0.1429	1.5000	18	1.4167	0.0676				1.0e-008 *	
Data 3 exp -N 55	0.5714	2.0000	21	1.6667	0.0625	19	0.1429	1.5000	0.2042	0
	0.2857	3.0000	24	1.9167	0.4782					
$\lambda = 1.5984e-11$			30	2.4167	0.1397	25	0.5714	2.0000	0.3898	0
			36	2.9167	0.1510					
			39	3.1667	0.1010	37	0.2857	3.0000	0.1937	0
Problem 2	0.1000	1.5000	4	1.0800	0.0003					
Data 4 exp -N 26	0.2000	2.0000	5	1.4400	0.0535	5	0.0908	1.4400	0.0169	0.0084
	0.4000	3.0000	6	1.8000	0.1849					
$\lambda = 5.1560e-014$	0.3000	6.0000	8	2.5200	0.0772	7	0.3053	2.1600	0.1922	0.0223
			9	2.8800	0.2802					
			10	3.2400	0.1013	10	0.3237	3.2400	0.1392	0.0335
			17	5.7600	0.1158					
			18	6.1200	0.1868	18	0.2803	6.1200	0.0359	0.0168

For the first problem we obtain the exact solution at the second step, improving substantially the regularized least squares solution (see Table 1 and Figure 4).

For the second problem (Table 1) we observe that the estimation clearly is better at the second step. We get good approximated solutions for the first and fourth components (see Figure 5). In general, models with four or more exponentials are difficult to solve. Results for the Lanczos' series (problem 3) are given in Table 2 and Figure 6.

In general we obtain small relative errors, and particularly, at the third component the exact solution was found. Our linear estimates are more precise when compared with the best solution $[0.0226; 0.2088; 0.7686]$, obtained by Holmström and Petersson [8]. It is worth mentioning that for this problem that at the second step we perform different simulations changing the values of k, the upper bound on the number of exponential functions, testing with $k = 3, 4, 5, 6$, obtaining the same solution. This means that the number of exponential functions probably does not depends on this bound.

The data for problem 4, see Figure 1, comes from the Steyn-Wyk experiment. The theoretical optimum is unknown, and we only can compare results with others

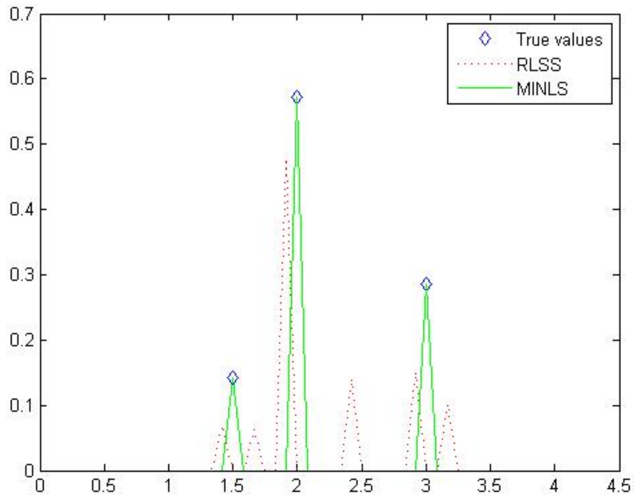

Fig. 4. Problem 1: Simulated data 3 exponentials $N = 55$

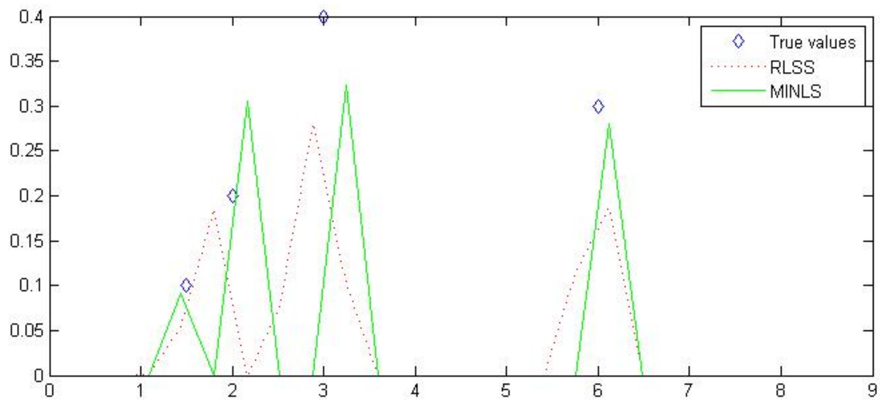

Fig. 5. Problem 2: Simulated data 4 exponentials $N = 26$

Table 2. Problems 3 and 4

Problem	ρ	τ	j	τ_j	$\bar{\rho}_j$	l	ρ_l^*	τ_l^*	elin	enl
Problem 3	0.0378	1.0000	4	0.9375	0.0276					
Data Lanczos	0.3424	3.0000	5	1.2500	0.0107	5	0.0573	1.2500	0.0275	0.0423
3 exp - N 25	0.6197	5.0000	10	2.8125	0.1190					
			11	3.1250	0.2272	11	0.3309	3.1250	0.0163	0.0211
$\lambda = 4e-7$			17	5.0000	0.6076					
			18	5.3125	0.0079	17	0.6118	5.0000	0.0111	0.0000
Problem 4			6	0.0342	2.1624					
Data Experimental	*	*	7	0.0905	2.5949	10	0.9356	3.8923	0.0288	0.0043
Steyn-Wyk N=60			8	0.1114	3.0274					
			9	0.1137	3.4599	22	1.0464	9.0821	0.3082	0.3073
$\lambda = 0.1109$			10	0.1072	3.8923					
			11	0.0974	4.3248					
			12	0.0875	4.7573					
			13	0.0787	5.1898					
			14	0.0717	5.6223					
			15	0.0666	6.0548					
			16	0.0631	6.4872					
			17	0.0610	6.9197					
			18	0.0600	7.3522					
			19	0.0597	7.7847					
			20	0.0599	8.2172					
			21	0.0604	8.6497					
			22	0.0610	9.0821					
			23	0.0615	9.5146					
			24	0.0618	9.9471					
			25	0.0620	10.3796					
			26	0.0618	10.8121					
			27	0.0614	11.2446					
			28	0.0606	11.6770					
			29	0.0596	12.1095					
			30	0.0582	12.5420					
			31	0.0566	12.9745					
			32	0.0546	13.4070					
			33	0.0525	13.8395					
			34	0.0501	14.2719					
			35	0.0476	14.7044					
			36	0.0449	15.1369					
			37	0.0420	15.5694					
			38	0.0391	16.0019					
			39	0.0360	16.4344					
			40	0.0329	16.8668					
			41	0.0298	17.2993					
			42	0.0266	17.7318					
			43	0.0234	18.1643					
			44	0.0202	18.5968					
			45	0.0170	19.0293					
			46	0.0139	19.4617					
			47	0.0108	19.8942					
			48	0.0078	20.3267					
			49	0.0048	20.7592					
			50	0.0019	21.1917					

* theoretical optimum unknown

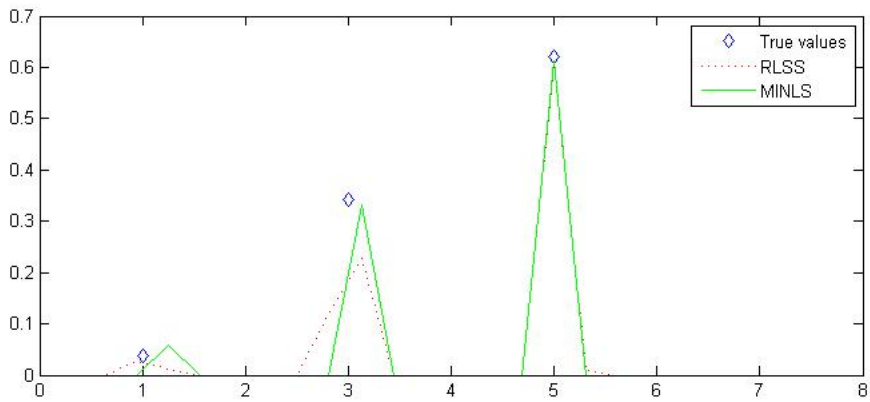

Fig. 6. Problem 3: Simulated data Lanczos $N = 25$

authors. By using graphical methods, Steyn and Wyk [14] found the two components solution for nonlinear parameter, $\tau^* = [3.858; 17.011]^t$.

Holmström and Petersson [8] reported the values $\rho^* = [0.976; 1.479]^t$ and $\tau^* = [3.832; 13.350]^t$ also detecting two exponential components. Our results for this problem are shown at Table 2. In Figure 7 we plot the first step curve (smooth curve) and our solution with two peaks. Our result is closer to the Holmström-Petersson solution when compared with that reported by Steyn-Wyk. We ran this model with $k = 3$ the bound on the number of exponential components, but for the second detected component, the linear parameter value was zero. This suggests that the correct model for this series should have two exponential terms.

It is worth mentioning that the values we obtain for the intensities in each of the studied problems satisfy the characterization established in the Theorem 4.2 in terms of the respective primal-dual optimal values.

8. CONCLUDING REMARKS

We propose a two step procedure to find estimations of linear and nonlinear parameters in exponential sum models. For the first step we solve a Tikhonov regularized problem which gives us the information to be refined in a second step. At this second stage we solve a mixed integer nonlinear programming problem obtaining good estimates for the parameters. We study properties of the solutions for this problem by exposing the combinatorial nature of the problem. Our procedure requires less initial information to successfully estimate the solution. It is enough to have the data, and an upper bound on the nonlinear parameters. We do not require the initial values to be known.

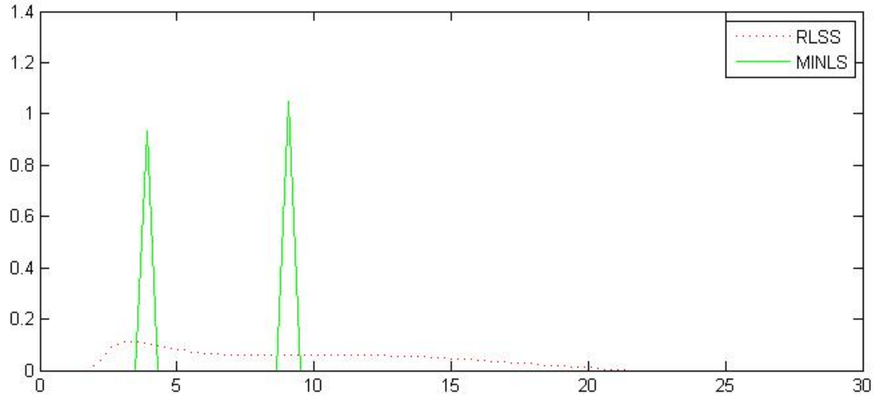

Fig. 7. Problem 4: Experimental data Steyn-Wyk $N = 58$

Sometimes all the data we have is a decay curve, and so estimating the number of exponential functions can be necessary. Our approach allows us to estimate this number by just giving a possible large upper bound, and the selecting k as the number of positive values for the intensities. As for further work we believe that our approach can be complemented with efficient algorithms to deal with the combinatorial nature of the problem, and exploiting the dual properties of the solutions to develop an algorithm which uses the special structure of the problem.

Acknowledgements
This work was partially supported by CDCHT-UCLA-Venezuela and the Alma Mater Project – OPSU – Venezuela.

REFERENCES

[1] A. Bemporad, D. Mignone, *MIQP.M: A Matlab function for solving Mixed Integer Quadratic Programs*, Technical Report, vol. Autoo-22, Switzerland, 2001.

[2] J. Burstein, *Approximations by exponentials, their extensions and differential equations*, Metric Press, Boston, 1997.

[3] C. Chong-Yung, *A Fast Maximum Likelihood Estimation and Detection Algorithm for Bernoulli-Gaussian Proccess*, IEEE Transactions on Acuoustics, Speech and Signal Processing, vol. ASSP-35, no. 11, November 1987.

[4] P.C. Hansen, *Regularization Tools: A Matlab package for analysis and solution of discrete ill-posed problems*, Numer. Algorithms **6** (1994), 1–35.

[5] P.C. Hansen, *Regularization Tools Version 4.0 for Matlab 7.3*, Numer. Algorithms **46** (2007), 189–194.

[6] P.C. Hansen, *The L-curve and its use in the numerical treatment of inverse problems*, Adv. Comput. Bioengineering, vol. 5, 2001.

[7] P.C. Hansen, D.P. O'Leary, *The use of the L-curve in the regularization of discrete ill-posed problems*, SIAM J. Sci. Comput. **14** (1993) 6, 1487–1503.

[8] K. Holmström, J. Petersson, *A review of the parameter estimation problem of fitting positive exponential sums to empirical data*, Appl. Math. Comput. **126** (2002) 1, 31–61.

[9] C. Lanczos, *Applied Analysis*, Prentice Hall, Englewood Cliffs, 272–280.

[10] A. Mohammad-Djafari, J.F. Giovannelli, G. Demoment, J. Idier, *Regularization, maximum entropy and probabilistic methods in mass spectrometry data processing problems*, Internat. J. Mass Spectrometry, **215** (2002) 1–3, 175–193.

[11] P.F. Price, *A comparison of the Least-Squares and Maximum likelihood estimators for counts of radiation quanta which follow a poisson distribution*, Acta Cryst. Sect. A, (1979), 3557–3560.

[12] T. Reginska, *A regularization parameter in discrete ill-posed problems*, SIAM J. Sci. Comput. **17** (1996) 3, 740–749.

[13] A. Shukla, M. Peter, L. Hoffmann, *Analysis of positron lifetime spectra using quantified maximum entropy and general linear filter*, Nuclear Instruments Methods Phys. Res. Sect. A **335** (1993) 1–2, 310–317.

[14] H.S. Steyn, W.J. Van Wyk, *Some methods for fitting compartment models to data*, Technical Report, Wetenskaplike Bydraes van die pu vir cho, Porcherstroomse Universiteit vir CHO, 1977.

[15] W.J. Wiscombe, J.W. Evans, *Exponential sum fitting of radioactive transmission function*, Comput. Phys. **24** (1977) 4, 416–444.

Adalys Alvarez
adalys.alvarez@gmail.com
adalvarez@una.edu.ve

Universidad Nacional Abierta
Área de Matemática
Centro Local Lara (Barquisimeto), Venezuela

Hugo Lara
hugol@ucla.edu.ve

Universidad Centroccidental Lisandro Alvarado
Decanato de Ciencia y Tecnología (Barquisimeto), Venezuela

Received: July 20, 2010.
Revised: November 12, 2010.
Accepted: November 26, 2010.