Exponential Modeling with Unknown Model Order Using Structured Nonlinear Total Least Norm

Lei Zhang , * Haesun Park* † and J. Ben Rosen ‡

November, 2001

Abstract:

A new algorithm called Structured Nonlinear Total Least Norm (SNTLN) has recently been developed for obtaining an approximate solution to the structured overdetermined nonlinear system. Both theoretical justification and computational testing show that SNTLN is an efficient method for solving structured overdetermined systems.

In this paper, we present a method based on SNTLN for estimating the parameters of exponentially damped sinusoidal signals in noise when the model order is unknown. It is compared to two other existing methods to show its robustness in recovering correct values of parameters when the model order is unknown, in spite of some large errors in the measured data.

Keywords: Exponential damped signals, model order, overdetermined linear systems, Hankel matrices, singular value decomposition, structured total least norm.

*The work of Lei Zhang was supported in part by the National Science Foundation grant CCR-9509085. Verizon Labs, 40 Sylvan Rd., Waltham, MA 02451, U.S.A., lzhang@cs.umn.edu .

†The work of Haesun Park* was supported in part by the National Science Foundation grants CCR-9509085 and CCR-9901992. Department of Computer Science and Engineering, Univ. of Minnesota, Minneapolis, MN 55455, U.S.A. and Korea Institute for Advanced Study, 207-43 Cheongryangri-dong, Dongdaemun-gu, Seoul 130-012, Korea. A part of this work was carried out while H. Park was visiting the Korea Institute for Advanced Study, Seoul, Korea, for her sabbatical leave, from September 2001 to July 2002. hpark@cs.umn.edu.

‡The work of J. Ben Rosen was supported in part by NSF grants CCR-9527151, CCR-9509085 and CCR-9901992. Department of Computer Science and Engineering, Univ. of Minnesota, Minneapolis, MN 55455 and Dept. of Computer Science and Engineering, Univ. of California, San Diego, La Jolla, CA 92093, U.S.A. e-mail: jbrosen@cs.ucsd.edu.

1 Introduction

Parameter estimation of a sum of exponential dumped sinusoids embedded in noise is a kernel problem in NMR spectroscopy. The data model considered is as follows: suppose a record of a data sequence y_i, $i = 1, 2, \ldots, n$, is composed of uniformly spaced samples of a sum of exponential damped signals y_i^* and measurement noise w_i. Specifically,

$$
\begin{aligned}
y_i &= y_i^* + w_i, \\
y_i^* &= \sum_{k=1}^{p} x_k e^{s_k i \Delta t} = \sum_{k=1}^{p} (a_k e^{\sqrt{-1}\phi_k}) e^{(-d_k + \sqrt{-1} 2\pi f_k) i \Delta t}, \\
& \quad i = 0, \ldots, n-1,
\end{aligned}
\tag{1.1}
$$

where p is the model order, which is unknown, and Δt is the constant sampling interval. The values y_i^* are called true signals and y_i observed signals. The objective is to estimate the model order p and the parameters, which are frequencies f_k, damping factors d_k, amplitudes a_k, and phases ϕ_k, $k = 1, \ldots, p$.

Cramér-Rao lower bounds (CR bounds) [11, 13, 19] are often used as a standard of precision to which the result of an estimation can be compared. The existence of such bounds implies that, under the same assumptions or conditions, irrespective of the method used to estimate the values of the model parameters from a data set, there is a lower bound on the precision that can not be superseded. Thus, there are two ways to improve the precision. One way is to choose a method that yields results as close as possible to the CR bounds. Another way is to try to reduce the CR bound by redesigning the experiment, i.e. varying the noise level, number of sample times and distribution of sample times, or imposing prior knowledge on the model function. We also point out that any parameter estimation procedure eventually breaks down as the SNR is decreased. The SNR at which this takes place is called breakdown SNR.

Two of the most commonly used methods for estimating the model order and the parameters are the Kumaresan-Tufts' linear prediction (LP) method [7, 8, 21, 6] and the HSVD method [1, 9, 12, 21]. The LP method is an improvement of the Prony Method [4]. It first sets up the linear prediction equations such as backward linear prediction equations,

$$Ab = -h$$

which is

$$
\begin{pmatrix}
\overline{y_1} & \overline{y_2} & \cdots & \overline{y_m} \\
\overline{y_2} & \overline{y_3} & \cdots & \overline{y_{m+1}} \\
\vdots & \vdots & \cdots & \vdots \\
\overline{y_l} & \overline{y_{l+1}} & \cdots & \overline{y_{n-1}}
\end{pmatrix}
\begin{pmatrix}
b_1 \\ b_2 \\ \vdots \\ b_m
\end{pmatrix}
= -
\begin{pmatrix}
\overline{y_0} \\ \overline{y_1} \\ \vdots \\ \overline{y_{l-1}}
\end{pmatrix}
\tag{1.2}
$$

where $b = (\, b_1 \cdots b_m \,)^T$ is the vector of backward prediction coefficients and $\overline{y_i}$ denotes the complex conjugate of y_i. If the data is noiseless, the prediction-error filter polynomial

$$f(z) = 1 + b_1 z^{-1} + b_2 z^{-2} + \cdots + b_m z^{-m} \tag{1.3}$$

will have zeros at $e^{-\overline{s_k}\Delta t}$, $k = 1, 2, \cdots, p$, if m is chosen to satisfy the inequality $p \le m \le n-p$. In fact, from (1.1) and (1.2) we have

$$\sum_{k=1}^{p} \overline{x_k} e^{\overline{s_k} i \Delta t} + \sum_{j=1}^{m} (\sum_{k=1}^{p} \overline{x_k} e^{\overline{s_k}(i+j)\Delta t}) b_j = 0, \quad i = 0, \ldots, l-1$$

or

$$\sum_{k=1}^{p} \overline{x_k} (e^{\overline{s_k} i \Delta t} + \sum_{j=1}^{m} \overline{x_k} e^{\overline{s_k}(i+j)\Delta t} b_j) = 0, \quad i = 0, \ldots, l-1.$$

From (1.3), this is equivalent to

$$\sum_{k=1}^{p} \overline{x_k} f(e^{-\overline{s_k}\Delta t}) e^{\overline{s_k} i \Delta t} = 0, \quad i = 0, \ldots, l-1.$$

Since $e^{\overline{s_k}\Delta t}$, $k = 1, \ldots, p$, are all distinct, it follows that $f(e^{-\overline{s_k}\Delta t}) = 0$, $k = 1, \ldots, p$.

These zeros, called signal zeros, fall outside the unit circle [5, 7]. If $m > p$, $f(z)$ has $m - p$ other zeros called extraneous zeros which will always fall within the unit circle. In the presence of noise in the data, the above procedure is shown experimentally still valid to estimate s_k's. However, A will have full rank and $m - p$ columns of A tend to be closely dependent causing a close-to rank deficient least squares problem. To handle this problem and filter out some noises, the SVD of A is computed and truncated by setting the smaller singular values of A to zero [7]. The rank p' of the truncated SVD of A is said to be the estimate of the exact model order p.

The HSVD method arranges the data y_i into a Hankel matrix $K \in \mathbf{C}^{l \times m}$,

$$K = \begin{pmatrix} y_0 & y_1 & \cdots & y_{m-1} \\ y_1 & y_2 & \cdots & y_m \\ \vdots & \vdots & \cdots & \vdots \\ y_{l-1} & y_l & \cdots & y_{n-1} \end{pmatrix},$$

where $l \ge m > p$ and $n = l + m - 1$. Then the SVD of K is computed and truncated by setting the smaller singular values of K to zero, similar to the procedure of LP. Same as in the LP method, the rank p' of the truncated SVD of K is regarded as the estimate of the exact model order p. Suppose the truncated SVD of K is $U_{p'} \Sigma_{p'} V^H{}_{p'}$, where H denotes complex conjugate transpose. A matrix

$$Z = (I + uu^H/(1 - u^H u))(U_{p'})^{\downarrow H} (U_{p'})^{\uparrow} \tag{1.4}$$

is then obtained, where u is the bottom row of $U_{p'}$, I is the $p' \times p'$ identity matrix, and $(U_{p'})^{\uparrow}$, $(U_{p'})^{\downarrow}$ are matrices derived from $U_{p'}$ by deleting the first row and last row, respectively. The estimates of the frequencies and damping factors are derived by diagonalizing Z. Finally, the amplitudes and phases are estimated by linear least squares fitting of Eqn. (1.1). See [1, 9] for more details.

The LP and HSVD methods constitute the so called black-box methods: they require no prior knowledge on the parameters or model order. However, it is known that when some of the noise components w_n are large, i.e., the SNR is small, these methods may break down. The SNR at which this takes place is called breakdown SNR. In addition, the gap between small and large singular values of A becomes unclear when the SNR is small, which makes it very difficult to correctly estimate the model order. The situation can be improved by imposing prior knowledge on the model parameters, which causes the Cramer-Rao (CR) bounds to decrease and lowers the threshold SNR. For more details see [2].

Recently, a new algorithm, called Structured Nonlinear Total Least Norm (SNTLN), has been developed [15, 16, 17] for solving nonlinear parameter estimation problems $A(\alpha)x \approx b$ when the elements of A may be differentiable nonlinear function of a parameter vector α whose values need to be approximated. A theoretical justification and computational testing of the SNTLN algorithm confirm that it is an efficient method for solving structured overdetermined systems by minimizing structured perturbations in the discrete L_q norm, where $q = 1, 2,\ or\ \infty$ [16, 17, 18]. In the next section, we will discuss how SNTLN can be applied to estimation of the model order and the parameters in Eqn. (1.1). Then in Section 3, the computational results, which compare the performance of the SNTLN based methods with that of the HSVD and the LP methods, are summarized, and the advantage of the SNTLN methods discussed.

2 Formulation of the SNTLN based method

When the model order p is unknown, to apply SNTLN, we first need to choose a prediction order m, where m should not be smaller than p, i.e., $m \geq p$. We assume that the correct signal y_i^* is represented by Eqn. (1.1) with m possible terms (rather than p), but with $m-p$ of the amplitudes $a_k = 0$. The validity of this assumption is proved in Theorem 2. The model order is then determined by identifying the zero amplitudes in the computed signal. Eqn. (1.1) can be rewritten as an overdetermined system

$$A(\alpha)x \approx y, \tag{2.5}$$

where

$$\alpha = (\alpha_1, \ldots, \alpha_m)^T, \quad \alpha_k = -d_k + \sqrt{-1}2\pi f_k \quad \text{for} \quad k = 1, \ldots, m,$$

$$x = (x_1, \ldots, x_m)^T, y = (y_0, \ldots, y_{n-1})^T,$$

$$A(\alpha) = \begin{pmatrix} 1 & \cdots & 1 \\ e^{\alpha_1 \Delta t} & \cdots & e^{\alpha_m \Delta t} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ e^{\alpha_1 (n-1) \Delta t} & \cdots & e^{\alpha_m (n-1) \Delta t} \end{pmatrix}, \tag{2.6}$$

m is the prediction order, and $n > m \geq p$.

To solve this problem, we minimize the following:

$$\min_{\alpha, x} \left\| \begin{matrix} r(\alpha, x) \\ D(\alpha - \hat{\alpha}) \end{matrix} \right\|_q, \quad q = 1, 2, \infty, \tag{2.7}$$

where $r(\alpha, x) = y - A(\alpha)x$, $\hat{\alpha}$ is an initial estimate of the parameter vector α, and D is a diagonal matrix of positive weights. We compute the minimum solution to Eqn. (2.7) iteratively by linearizing $r(\alpha, x)$:

$$r(\alpha + \Delta\alpha, x + \Delta x) \approx r(\alpha, x) - A(\alpha)\Delta x - J(\alpha, x)\Delta\alpha, \tag{2.8}$$

where $\Delta\alpha$ and Δx represent small changes in the elements of α and x respectively, and $J(\alpha, x)$ is the Jacobian, with respect to α, of $A(\alpha)x$. Let $A(\alpha)_k$ represent the k^{th} column of $A(\alpha)$. Then

$$J(\alpha, x) = \nabla_\alpha (A(\alpha)x) = \sum_{k=1}^{m} x_k \nabla_\alpha (A(\alpha)_k). \tag{2.9}$$

In solving (2.5) using SNTLN, the perturbation on A will be found so that each α_k is perturbed to $\alpha_k + \Delta\alpha_k$ for some value $\Delta\alpha_k$ and the perturbed matrix $A(\alpha)$ keeps the Vandermonde structure shown in Eqn (2.6). The solution vector x and the perturbation $\Delta\alpha = (\, \Delta\alpha_1 \quad \cdots \Delta\alpha_m \,)^T$ on the parameter α which satisfy

$$A(\alpha + \Delta\alpha)x = y - r(\alpha, x)$$

will be found while minimizing Eqn. (2.6).

For the Vandermonde matrix $A(\alpha)$, the Jacobian $J(\alpha, x)$ for $A(\alpha)x$ is

$$J(\alpha, x) = \begin{pmatrix} 0 & \cdots & 0 \\ x_1 \Delta t e^{\alpha_1 \Delta t} & \cdots & x_m \Delta t e^{\alpha_m \Delta t} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ x_1 (n-1) \Delta t e^{\alpha_1 (n-1) \Delta t} & \cdots & x_m (n-1) \Delta t e^{\alpha_m (n-1) \Delta t} \end{pmatrix}. \tag{2.10}$$

Algorithm SNTLN based Parameter Estimation

Input – Initial estimate $\hat{\alpha}$ for α, matrix $A(\alpha)$, $\nabla_\alpha(A(\alpha)_k)$ for $1 \le k \le m$, diagonal weight matrix D, vector y, convergence tolerance ϵ_1, and truncation tolerance ϵ_2.
Output –parameter vectors from α and x, and the model order p

1. Set $\alpha = \hat{\alpha}$, compute x from (2.7) with $A = A(\alpha)$, $J(\alpha, x)$, and set $r = y - A(\alpha)x$.
2. **repeat**
   (a) $$\underset{\Delta x, \Delta\alpha}{\text{minimize}} \left\| G \begin{pmatrix} \Delta x \\ \Delta\alpha \end{pmatrix} + \begin{pmatrix} -r \\ D(\alpha - \hat{\alpha}) \end{pmatrix} \right\|_q,$$
   where $$G = \begin{pmatrix} A(\alpha) & J(\alpha, x) \\ 0 & D \end{pmatrix}.$$
   (b) Set $x := x + \Delta x$, $\alpha := \alpha + \Delta\alpha$.
   (c) Compute $J(\alpha, x)$, $A(\alpha)$, $r = y - A(\alpha)x$.

   until ($\|\Delta x\| \le \epsilon_1$ and $\|\Delta\alpha\| \le \epsilon_1$)
3. Set the elements of x that are smaller than ϵ_2 in magnitude to be zero.
4. Determine the value of the model order p which is the number of nonzero elements of the vector x. For each nonzero element x_j of x, compute the parameters from the relation: $x_k = a_k e^{\sqrt{-1}\phi_k}$ and $\alpha_k = -d_k + \sqrt{-1}2\pi f_k$.

The SNTLN based parameter estimation method is summarized in Algorithm SNTLN based Parameter Estimation.

In Step 2(a) of Algorithm SNTLN, we need to solve a minimization problem with the $(n+m) \times 2m$ matrix $\begin{pmatrix} A(\alpha) & J(\alpha, x) \\ 0 & D \end{pmatrix}$, where $J(\alpha, x)$ and $A(\alpha)$ are given in (2.10) and (2.6) respectively. Note that we have chosen not to perturb the first row of the matrix $A(\alpha)$.

An initial estimate of $\hat{\alpha}$ is needed to start the iterative procedure. We will show how this value can be obtained in Section 3. The values of the diagonal weight matrix D should be chosen according to the error levels in the data and how good the initial estimate of $\hat{\alpha}$ is, and they can be used to fine tune the algorithm. The computational method by which Step 2(a) is carried out depends on the value of q. For $q = 2$, the least squares problem is solved. For $q = 1$ or ∞, Step 2(a) is solved as a linear program. To further improve robustness of the

SNTLN algorithm with $q = 1$, bounds can be imposed on α, which can easily be imposed as constraints. It has been shown that this algorithm will always converge [18]. Moreover, according to the following theorem the result obtained by the algorithm is a local optimal one.

Theorem 1 *Assume that when there is no noise (or error) in $y = y_c$, there is a parameter vector α_c, and corresponding x_c, such that $A(\alpha_c)x_c = y_c$. A neighborhood Ω_c is assumed, with α_c at its center, that contains the initial estimate $\hat{\alpha}$. In addition, assume that $A(\alpha)$ has full rank, and that all the elements $a_{ij}(\alpha)$ have bounded Hessians, for all $\alpha \in \Omega_c$. Then the solution obtained by Algorithm SNTLN is a local optimum.*

Proof: The function (2.7) being minimized, when $q = 2$, is equivalent to:

$$\varphi(\alpha, x) = \frac{1}{2} r^T r + \frac{1}{2}(\alpha - \hat{\alpha})^T D^2 (\alpha - \hat{\alpha}), \tag{2.11}$$

where $r = r(\alpha, x) = y - A(\alpha)x$. The first order optimality conditions for a local optimum of $\varphi(\alpha, x)$ are the vanishing of the gradients $\nabla_\alpha \varphi$ and $\nabla_x \varphi$, where these gradients are given by:

$$\begin{array}{l} \nabla_\alpha \varphi = -J^T(\alpha, x) r + D^2(\alpha - \hat{\alpha}), \\ \nabla_x \varphi = -A(\alpha)^T r. \end{array} \tag{2.12}$$

At each iteration of the SNTLN algorithm with the L_2 norm, we compute the least squares solution to

$$\min_{\Delta x, \Delta \alpha} \| G \begin{pmatrix} \Delta \alpha \\ \Delta x \end{pmatrix} + \begin{pmatrix} -r \\ D(\alpha - \hat{\alpha}) \end{pmatrix} \|_2, \tag{2.13}$$

where

$$G = \begin{bmatrix} J(\alpha, x) & A(\alpha) \\ D & 0 \end{bmatrix}. \tag{2.14}$$

The least squares solution is given by

$$G^T G \begin{pmatrix} \Delta \alpha \\ \Delta x \end{pmatrix} = -G^T \begin{pmatrix} -r \\ D(\alpha - \hat{\alpha}) \end{pmatrix} = -\begin{pmatrix} \nabla_\alpha \varphi \\ \nabla_x \varphi \end{pmatrix}, \tag{2.15}$$

where the last equality follows from (2.12). Since G has full rank, by the assumption on $A(\alpha)$, $G^T G$ is nonsingular and $(\Delta \alpha, \Delta x)$ can only be zero when the gradient is zero. Therefore at termination, the algorithm gives an (α, x) satisfying the first-order optimality conditions.

■

The following theorem explains how the SNTLN based parameter estimation algorithm for exponential modeling with unknown model order works.

Theorem 2 *Suppose x^* is the solution to the following zero residual overdetermined system*

$$Ax = b, \tag{2.16}$$

where A is an $m \times n$, $m > n$, matrix. Now consider the following two systems

$$A_s y = b, \tag{2.17}$$

where $m \times l$ matrix A_s, $l < n$, is a submatrix of A, and

$$A_e z = b, \tag{2.18}$$

where A_e is an $m \times p$, $m > p$, matrix, and has A as its submatrix. If A and A_e have full rank then the following results hold:

(a) *Let y^* denote the vector with $n - l$ components consisting of elements of x^* that correspond to columns of A not in A_s. The Eqn. (2.17) has a solution only if $y^* = 0$;*

(b) *The elements z_j^*, $j = 1, \ldots, p$, of the solution to (2.18) are given as follows:*

$$z_j^* = \begin{cases} x_i^* & \text{if the } j^{th} \text{ column of } A_e \text{ is} \\ & \text{the } i^{th} \text{ column of } A, \\ 0 & \text{if the } j^{th} \text{ column of } A_e \text{ is not} \\ & \text{a column of } A, \end{cases}$$

where x_i^* is the i^{th} element of x^*.

Proof: Without loss of generality, we assume that A_s consists of the first l columns of A, and that the first n columns of A_e are given by A. To prove (a), note that if the last $n - l$ elements of x^* are not all zeros and (2.17) has a solution then (2.16) has two distinct solutions. This contradicts the fact that A has full rank.

Similarly, since A_e has full rank, $z^* = \begin{pmatrix} x^* \\ 0_{m\times(p-n)} \end{pmatrix}$ is the only solution to (2.18). ■

In Eqn. (2.5), the matrix $A(\alpha)$ has Vandermonde structure. It has full rank if the elements of α are distinct. Theorem 2 shows that if there is no noise in the data, after solving (2.5) the x_k corresponding to the nonexistent frequency f_k and damping factor d_k will be zero. If there is noise in the data, being applied to Eqn. (2.5), the SNTLN algorithm tries to recover the correct values of x_k. This means some of the x_k are close to zero. We can then determine the value of the model order p from the number of nonzero elements of x, and select the damping factors d_k and frequencies f_k correspondingly.

Both theoretical and computational results show that 1-norm SNTLN is able to obtain accurate approximations with data containing a relatively small number of large errors, which we call "outliers" [17]. In contrast, our results show that HSVD and LP have difficulty in estimating the model order correctly, and parameters accurately, when some of the errors are large. The SNTLN based methods can correctly estimate the model order, and the errors in parameter estimates are much smaller, as shown in the next section.

3 Computational Results

We now present the test results using 1-norm SNTLN (SNTLN1), 2-norm SNTLN (SNTLN2), the HSVD and the LP methods. To construct the test problems, we assume that the correct parameter vector α_c and amplitude vector x_c are known. This information is of course not used in the algorithms tested, but allows us to investigate the ability of the algorithms to recover the correct α and x, in spite of errors in the data. A total of eight test cases were considered.

In the first case, we used the following values: exact model order $p = 4$, the prediction order $m = 7$, $n = 128$, $\Delta t = 0.0004$. The measurement noise w_n consisted of a small random (normally distributed) error with magnitude $\tau \leq 10^{-6}$ for each n, and no large errors were added. For SNTLN1, bounds were imposed on $d = (d_1, \ldots, d_m)^T$ and $f = (f_1, \ldots, f_m)^T$:

$$\begin{array}{l} d_{lk} \leq d_k \leq d_{uk} \\ f_{lk} \leq f_k \leq f_{uk} \end{array} \qquad k = 1, \ldots, 7,$$

where

$$\begin{array}{rcl} d_l &=& (40, 40, 130, 100, 160, 190, 240), \\ d_u &=& (65, 65, 160, 130, 190, 220, 280), \\ f_l &=& (8, 18, 32, 120, 370, 530, 790), \\ f_u &=& (13, 28, 42, 140, 400, 560, 825). \end{array}$$

A total of 10 test problems were constructed. For each problem, the correct values x_c, d_c, and f_c were generated by selecting random values within the bounds. For $x_{ck} = x_{Rk} + \sqrt{-1}x_{Ik}$, $k \leq p$, each x_{Rk} and x_{Ik} were randomly selected in the range $[-10, 10]$, and x_{ck} was set to zero for $k > p$. These test problems were chosen so as to be similar to the NMR signal data problem described in [3]. For SNTLN1 and SNTLN2, we used the following initial values:

$$\begin{array}{l} d_{init} = \frac{1}{2}[d_l + d_u] \\ f_{init} = \frac{1}{2}[f_l + f_u] \end{array}$$

for each problem. For SNTLN, the tolerance value used to truncate x and estimate the model order was set to be 0.02 and for HSVD and LP, the tolerance value used to truncate the singular values of Hankel matrices formed by the signals was set to be 2.0×10^{-5}. Those tolerance values were chosen based on the knowledge of the approximate size of errors in the signals, as well as the approximate magnitude of x and y. The computed solution x for the structured overdetermined system seems to be far less sensitive to the high noise level in the data matrix, than the singular values of the data matrix. This means that the determination of the model order is also less sensitive in the SNTLN based method than others that depends on the singular values of the data matrix.

We measured the performance of each method on a test problem in terms of the relative errors:

$$\begin{aligned} RE_d &= \|d_A - d_c\|/\|d_c\|, \\ RE_f &= \|f_A - f_c\|/\|f_c\|, \\ RE_x &= \|x_A - x_c\|/\|x_c\|, \end{aligned} \tag{3.19}$$

where d_A, f_A, and x_A are parameter vectors estimated by a method. The results are summarized in Table 1. For each method, the average of relative errors over 10 different correct signals are shown. In this case, all the methods can correctly estimate the model order. However, we see that SNTLN based methods performed much better.

In the second case, we used the same data as in the first case except that in addition to the small random error, for 25 randomly selected i, an outlier $\delta = 0.01$ was added to the measurement noise w_i. Also the tolerance value used to truncate the singular values of Hankel matrices formed by the signals was set to be 0.05, while for SNTLN the tolerance value to truncate x remained the same. See Table 2 for results. Since in this case, both the HSVD and the LP methods had difficulty in estimating the correct model order, for the purpose of comparison, only those test problems for which those two methods happened to correctly estimate the model order were chosen. SNTLN based methods still estimated the correct model order easily.

In the third and fourth cases, the test problems were constructed in the same way as in the second test except that the actual model order p was set to be 5 and 6, respectively. The results are shown in Table 3 and Table 4.

Table 1: Average Relative Errors in f, d, and x for 10 runs. Magnitude of small random errors is 10^{-6}. No large errors ($\delta = 0$). Actual model order $p = 4$.

	SNTLN1	SNTLN2	HSVD	LP
RE_d	2.52e-6	1.36e-6	4.28e-4	3.13e-4
RE_f	1.38e-7	4.87e-7	1.61e-4	6.19e-5
RE_x	8.48e-6	2.63e-6	2.46e-3	5.38e-4

Table 2: Average Relative Errors in f, d, and x for 10 runs. Magnitude of small random errors is 10^{-6}. $\delta = 0.01$. Actual model order $p = 4$.

	SNTLN1	SNTLN2	HSVD	LP
RE_d	2.71e-6	6.17e-3	2.13e+0	2.69e+0
RE_f	1.54e-7	1.28e-3	9.10e+0	4.85e+0
RE_x	1.22e-5	1.53e-2	8.77e-1	1.08e+0

Table 3: Average Relative Errors in f, d, and x for 10 runs. Magnitude of small random errors is 10^{-6}. $\delta = 0.01$. Actual model order $p = 5$.

	SNTLN1	SNTLN2	HSVD	LP
RE_d	2.13e-6	5.01e-4	3.29e-1	1.70e+0
RE_f	2.33e-7	8.59e-5	3.00e+0	1.86e+0
RE_x	3.41e-5	2.71e-3	1.12e+0	1.21e+0

The fifth and sixth cases are similar to the first and second cases, respectively, except that the exact model order $p = 4$, the prediction order $m = 6$ and the bounds were changed as follows

$$\begin{aligned} d_l &= (28.5, 12, 24, 28.5, 12, 24), \\ d_u &= (33, 18, 28.5, 33, 18, 28.5), \\ f_l &= (40, 70, 100, 220, 320, 550), \\ f_u &= (60, 80, 140, 270, 380, 600). \end{aligned}$$

The results are shown in Table 5 and Table 6.

Finally, the seventh and eighth cases are similar to the first and second cases, respectively, except that the magnitude of the small errors τ ranged from 10^{-6} to 10^{-2}. The results are

Table 4: Average Relative Errors in f, d, and x for 10 runs. Magnitude of small random errors is 10^{-6}. $\delta = 0.01$. Actual model order $p = 6$.

	SNTLN1	SNTLN2	HSVD	LP
RE_d	2.38e-6	2.94e-3	2.57e-1	1.39e+0
RE_f	1.72e-7	1.45e-4	1.86e+0	1.22e+0
RE_x	2.67e-5	5.32e-3	8.43e-1	9.39e-1

Table 5: Average Relative Errors in f, d, and x for 10 runs. Magnitude of small random errors is 10^{-6}. No large errors ($\delta = 0$). Actual model order $p = 4$.

	SNTLN1	SNTLN2	HSVD	LP
RE_d	1.06e-4	1.95e-6	1.06e-3	1.00e-3
RE_f	2.29e-6	1.19e-8	3.21e-5	8.14e-6
RE_x	4.96e-5	5.04e-7	4.46e-4	1.89e-4

Table 6: Average Relative Errors in f, d, and x for 10 runs. Magnitude of small random errors is 10^{-6}. $\delta = 0.01$. Actual model order $p = 4$.

	SNTLN1	SNTLN2	HSVD	LP
RE_d	2.59e-7	2.98e-3	1.40e+1	1.05e+1
RE_f	9.72e-9	8.90e-5	4.21e+0	2.66e+0
RE_x	1.74e-7	2.45e-3	1.12e+0	6.07e-1

summarized in Table 7 and Table 8. We see that even without outliers, the HSVD and LP methods broke down when the error level reached 10^{-4}. With outlier present in the data, the HSVD and LP methods broke down from the very beginning. This is a live proof of the statement that as SNR decreases, the HSVD and LP methods will eventually break down. In contrast, the SNTLN2 method continued to do very well. The SNTLN1 method also encountered some difficulty when τ reached 10^{-3}. But when we looked carefully, out of 10 runs, some of them actually gave good results. Since we are only showing the average here, those good runs were masked.

We would like to point out that for all the cases where outliers are also present, the SNTLN1 method showed the remarkable ability to obtain the correct parameter and amplitude vectors in spite of large errors in the data and initial estimates. This is due both to the use of the L_1 norm and the imposed bounds. In fact, the errors in the parameters estimated by SNTLN1 are determined by the small ($< 10^{-6}$) data errors, and are essentially independent of the 25 outliers. More test results also show that the SNTLN1 solution changes only slightly by a change in the size of outliers, δ. In contrast, the errors in parameters estimated by other methods are determined primarily by the largest data errors, and will increase in proportion to δ, as δ is increased. While SNTLN2 is not as robust as SNTLN1, it performed significantly better than the HSVD and the LP.

To illustrate the difficulty encountered by HSVD and LP in estimating the correct model order, we show in the following the singular values of one of the Hankel matrices used in the

Table 7: Average Relative Errors in f, d, and x for 10 runs each. Magnitude of small random errors τ ranges from 10^{-6} to 10^{-2}. No large errors ($\delta = 0$). Actual model order $p = 4$.

τ		SNTLN1	SNTLN2	HSVD	LP
	RE_d	$2.23e-6$	$2.81e-6$	$7.13e-4$	$2.55e-4$
$1e-6$	RE_f	$9.26e-7$	$6.71e-7$	$1.71e-4$	$9.25e-5$
	RE_x	$1.44e-5$	$1.17e-5$	$2.44e-3$	$1.31e-3$
	RE_d	$1.81e-5$	$2.53e-5$	$1.79e-2$	$2.33e-2$
$1e-5$	RE_f	$4.69e-6$	$5.84e-6$	$4.53e-3$	$2.92e-2$
	RE_x	$6.87e-5$	$9.86e-5$	$4.67e-2$	$2.04e-2$
	RE_d	$1.88e-4$	$3.17e-4$	$1.54e-1$	$2.09e+0$
$1e-4$	RE_f	$1.15e-4$	$5.60e-5$	$3.89e-1$	$2.58e+0$
	RE_x	$8.81e-4$	$1.22e-3$	$7.03e-1$	$8.17e-1$
	RE_d	$2.19e-1$	$5.91e-3$	$4.86e-1$	$2.88e+0$
$1e-3$	RE_f	$1.35e+0$	$1.20e-3$	$3.99e+0$	$5.61e+0$
	RE_x	$4.29e-1$	$4.95e-3$	$1.21e+0$	$1.35e+0$
	RE_d	$2.46e-1$	$3.60e-2$	$4.37e+0$	$2.94e+0$
$1e-2$	RE_f	$1.88e+0$	$5.04e-3$	$7.73e+0$	$5.91e+0$
	RE_x	$8.16e-1$	$1.88e-1$	$1.22e+0$	$1.27e+0$

second test case:

$$\begin{pmatrix} 1.4825e+2 \\ 1.9843e+1 \\ 8.7398e-1 \\ 5.3410e-2 \\ 4.7958e-2 \\ 4.6210e-2 \\ 4.4174e-2 \end{pmatrix}.$$

The gap which should occur between the 4^{th} and 5^{th} singular values is not identified by this result. Now look at the x value obtained by SNTLN1 in the same test run:

$$\begin{pmatrix} 6.92e+0 & - & 3.84e+0\sqrt{-1} \\ -4.75e+0 & - & 8.07e+0\sqrt{-1} \\ -1.88e+0 & + & 4.13e+0\sqrt{-1} \\ 8.67e+0 & - & 2.18e+0\sqrt{-1} \\ 3.44e-5 & - & 7.11e-6\sqrt{-1} \\ 7.73e-6 & + & 1.05e-5\sqrt{-1} \\ -6.63e-6 & - & 2.63e-6\sqrt{-1} \end{pmatrix}.$$

Table 8: Average Relative Errors in f, d, and x for 10 runs each. Magnitude of small random errors τ increases from 10^{-6} to 10^{-2}. $\delta = 0.01$. Actual model order $p = 4$.

τ		SNTLN1	SNTLN2	HSVD	LP
	RE_d	$3.67e-6$	$7.41e-3$	$1.80e-1$	$2.67e+0$
$1e-6$	RE_f	$1.03e-6$	$2.68e-3$	$9.07e+0$	$5.49e+0$
	RE_x	$1.72e-5$	$4.42e-2$	$1.01e+0$	$1.14e+0$
	RE_d	$1.65e-5$	$6.72e-3$	$7.09e+0$	$2.75e+0$
$1e-5$	RE_f	$4.57e-6$	$9.78e-4$	$6.67e+0$	$5.41e+0$
	RE_x	$6.82e-5$	$1.48e-2$	$1.09e+0$	$1.21e+0$
	RE_d	$2.31e-4$	$2.08e-2$	$3.73e+0$	$2.86e+0$
$1e-4$	RE_f	$2.97e-5$	$2.02e-3$	$6.71e+0$	$5.36e+0$
	RE_x	$6.10e-4$	$6.48e-2$	$1.22e+0$	$1.03e+0$
	RE_d	$3.57e-1$	$6.34e-3$	$8.27e-1$	$2.63e+0$
$1e-3$	RE_f	$1.75e+0$	$2.69e-3$	$8.44e+0$	$5.34e+0$
	RE_x	$8.52e-1$	$1.43e-2$	$1.42e+0$	$1.28e+0$
	RE_d	$4.67e-1$	$1.26e-2$	$5.82e-1$	$2.47e+0$
$1e-2$	RE_f	$2.51e+0$	$3.05e-3$	$7.09e+0$	$4.63e+0$
	RE_x	$1.13e+0$	$6.08e-2$	$1.09e+0$	$1.40e+0$

We note that there is a well defined gap between the 4^{th} and 5^{th} elements. This is true for all test problems. SNTLN2 also did well in identifying the gap. Table 9 shows the minimum, average and maximum gap values in the second test case for each method over 10 runs. For the LP and HSVD methods, the gap value is the ratio of the 4^{th} to the 5^{th} singular value. With the SNTLN2 and SNTLN1 methods, the gap value is calculated as the ratio of the 4^{th} to the 5^{th} element of sorted x. Again, we can see for the LP and HSVD methods, the gap values are close to 1, making it very difficult to determine the model order correctly. Whereas for the SNTLN based methods, the gaps are significant.

Table 9: Minimum, average and maximum gap values for 10 runs in the second test case.

	SNTLN1	SNTLN2	HSVD	LP
MIN GAP	$3.82e+5$	$6.71e+2$	$1.077e+0$	$1.077e+0$
AVG GAP	$2.11e+6$	$1.81e+3$	$1.080e+0$	$1.079e+0$
MAX GAP	$4.64e+6$	$3.27e+3$	$1.082e+0$	$1.081e+0$

In Table 10 and Table 11, we show the computed parameters when the model order

Table 10: Values of d and f when the model order is *overestimated*. Comparison of SNTLN2, SNTLN1, HSVD, and LP.

d_c	151	116	61.4	57.4	
f_c	37.1	127	12.5	21.3	
d_{SNTLN2}	151	116	61.4	57.4	260
f_{SNTLN2}	37.1	127	12.5	21.3	804
d_{SNTLN1}	151	116	61.4	57.4	190
f_{SNTLN1}	37.1	127	12.5	21.3	376
d_{HSVD}	151	116	61.4	57.4	559
f_{HSVD}	37.1	127	12.5	21.3	1208
d_{LP}	151	116	61.4	57.4	-515
f_{LP}	37.1	127	12.5	21.3	-813

was estimated incorrectly in some of our tests. When the model order was overestimated, the parameters computed by each method consisted of the correct parameters as well as some nonexisting parameters. It is interesting to notice that even when the model order was underestimated, the parameters computed by the SNTLN based method made a subset of the set of correct parameters. In contrast, the HSVD and the LP methods produced completely wrong parameters when the model order was underestimated. This is because the SNTLN based method determines the model order after the algorithm converges and parameter vector α is obtained. Whereas for the HSVD and LP methods, it happens during the SVD truncation phase. If the model order is underestimated, in other words, the corresponding matrix is truncated too much, the matrix Z in (1.4) or the prediction-error filter polynomial in (1.3) will be adversely affected, which causes the HSVD or LP method to yield wrong results.

In summary, we have introduced a new, SNTLN based method for estimating the model order and parameters in exponential modeling. Then the SNTLN based method was compared to the HSVD and the LP methods, and its superior capability in estimating the correct model order, and the accurate parameters is shown. Moreover, we have shown the robustness of SNTLN with L_1 norm in recovering correct values of parameters, in spite of some large errors in the measured data. Often in real applications, estimated bounds on the parameters can be easily obtained. To fully take advantage of the prior knowledge, for the 1-norm SNTLN based method these bounds can be imposed as constraints, further improving the robustness.

Table 11: Values of d and f when the model order is *underestimated.* Comparison of SNTLN2, SNTLN1, HSVD, and LP.

d_c	155	52.1	41.1	108
f_c	37.7	26.0	9.65	134
d_{SNTLN2}	155	52.1	41.1	
f_{SNTLN2}	37.7	26.0	9.65	
d_{SNTLN1}	155	52.1	41.1	
f_{SNTLN1}	37.7	26.0	9.65	
d_{HSVD}	110	101	19.6	
f_{HSVD}	37.3	134	11.2	
d_{LP}	100	90.5	18.4	
f_{LP}	134	37.2	11.3	

References

[1] H. Barkjuijsen, R. De Beer and D. Van Ormondt. Improved algorithm for noniterative time-domain model fitting to exponentially damped magnetic resonance signals. *J. Magnetic Resonance* 73, pp. 553-557, 1987.

[2] R. De Beer and D. Van Ormondt. Analysis of NMR data using time-domain fitting procedures. *NMR Basic Principles and Progress* series 26, Springer-Verlag, Berlin Heidelberg, pp. 201-248, 1992.

[3] H. Chen, S. Van Huffel, D. Ormondt, and R. deBeer. Parameter estimation with prior knowledge of known signal poles for the quantification of NMR spectroscopy data in the time domain. *J. Magnetic Resonance*, Series A 119, 225–234, 1996.

[4] F. B. Hildebrand. *Introduction to Numerical Analysis*, New York, McGraw-Hill, 1956.

[5] R. Kumaresan. Estimating the parameters of exponentially damped / undamped sinusoidal signals in noise. Ph.D. dissertation, University of Rhode Island, Kingston, RI, Aug. 1982.

[6] R. Kumaresan and D.W. Tufts. Accurate parameter estimation of noisy speech-like signals. Presented at the *IEEE Int. Conf. on Acoust., Speech, Signal Processing*, May 3-5, 1982, Paris, France.

[7] R. Kumaresan and D.W. Tufts. Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise. *IEEE Trans. on Acoust., Speech, Signal Process.*, vol. ASSP-30, no. 6, pp. 833-840, Dec. 1982.

[8] R. Kumaresan, D.W. Tufts, and L.L. Scharf. A prony method for noisy data: choosing the signal components and selecting the order in exponential signal models. *Proc. IEEE* 72:2, pp. 230-233, February 1984.

[9] S.Y. Kung, K.S. Arun and D.V. Bhaskar Rao. State-space and singular value decomposition-based approximation methods for the harmonic retrieval problem. *J. Opt. Soc. Am.* 73, pp.1799-1811, 1983.

[10] A. Macovski and D. Spielman. *Magnetic Resonance in Medicine*, Vol 3, pp. 97, 1986.

[11] J. P. Norton. *An introduction to identification.* Academic Press, London, 1986.

[12] H. Park, S. Van Huffel and L. Eldén. Fast algorithm for exponential data modeling. *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 94)*, Adelaide, Australia, Apr. 1994, pp. 25-28.

[13] M. B. Priestley. *Spectral analysis and time series*, vols 1 and 2. Academic Press, London, 1981.

[14] J.B. Rosen, H. Park, and J. Glick. Total least norm formulation and solution for structured problems. *SIAM Journal on Matrix Anal. Appl.*, 17-1:110-128, 1996.

[15] J.B. Rosen, H. Park, and J. Glick. Total least norm for linear and nonlinear structured problems, Recent advances in total least squares techniques and errors-in-variables modeling. Ed. S. Van Huffel, SIAM, pp. 203-214, 1997.

[16] J.B. Rosen, H. Park, and J. Glick. Structured total least norm for nonlinear problems. *SIAM Journal on Matrix Anal. Appl.*, 20-1:14-30,1999.

[17] J. B. Rosen, H. Park, J. Glick, and L. Zhang. Accurate solution to overdetermined linear equations with errors, using L_1 norm minimization. *Computational Optimization and Applications*, 17, pp. 329-341, 2000.

[18] J.B. Rosen, H. Park, and J. Glick. Signal Identification Using a Least L_1 Norm Algorithm. *Optimization and Engineering* , 1, pp. 51-65, 2000.

[19] A. Tarantola. *Inverse problem theory. Methods for data fitting and model parameter estimation.* Elsevier, Amsterdam, 1987.

[20] A Van den Bos. Estimation of Fourier coefficients. *IEEE Transactions on Instrumentation and Measurement (IM)* 38, pp. 1005-1007, 1989

[21] S. Van Huffel, L. Aerts, J. Bervoets, J. Vandewalle, C. Decanniere and P. Van Hecke. Improved quantitative time-domain analysis of NMR data by total least squares. *Signal Processing VI: theories and applications.* Ed. J. Vandewalle, R. Boite, M. Moonen and A. Oosterlinck, Elsevier Science Publishers B.V. (North Holland), 1992, Vol. III, pp. 1721-1724.