87-091

Journal of Computational and Applied Mathematics 20 (1987) 175–187
North-Holland

The fit of a sum of exponentials to noisy data

P. DE GROEN
Department of Mathematics, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel, Belgium

B. DE MOOR *
ESAT, Department Elektrotechniek, Katholieke Universiteit Leuven, B-3030 Heverlee, Belgium

Received 24 July 1986
Revised 11 February 1987

Abstract: Given a set of noisy data $y_0, \ldots, y_m$, measured at equidistant points on the time axis, find n $(n < \frac{1}{2}m)$ exponents b_i and coefficients c_i such that $\sum_i c_i \exp(b_i t)$ yields a good approximation of the data. We study two algorithms for the reconstruction of the exponents and coefficients. We show that the error in the reconstructed exponents is of the order of the gap $\sigma_{n+1}/(\sigma_n - \sigma_{n+1})$ between the smallest accepted singular value σ_n and the largest rejected singular value σ_{n+1} of a Hankel matrix constructed from the data. Moreover, we provide a statistical analysis, which gives some insight in the optimal choice of the dimensions of the Hankel matrix to be used, given a fixed number of datapoints, and we illuminate this by several clarifying examples.

Keywords: Exponential approximation, systems realization.

1. Introduction

An important problem in experimental sciences is the fit of a weighted sum of exponentials

$$f(t) := \sum_{i=1}^{n} c_i \exp(b_i t) \tag{1.1}$$

with unknown exponents b_i and coefficients c_i to a given set $\{ y_j \mid j = 0 : 2m - 1 \}$ of $2m$ $(m > n)$ noisy data points obtained from some experiment.

The first solution to the problem of fitting experimental data by a sum of real exponentials is due to De Prony [6]. Variations on his method are described in several books on numerical analysis [10] ("problem of 'weighted moments'"). The problem occurs in fitting radioactive decay measurements (see ref. in [12]), transmission function analysis in atmospheric research [15] compartmental analysis [8], electrical network analysis (Heaviside's expansion theorem [10]), analysis of human lung exhalation and chemical rate constant estimation (see ref. in [12] and [15]). In many of these applications, the amplitudes c_i are expected to be positive. In [15] a

* Sponsored by a grant of the IWONL.

survey of various methods is given. Some fundamental theorems guaranteeing existence and uniqueness of positive sums of a 'best fit' (provided the number of exponents is not constrained a priori) are treated by Cantor and Evans [2]. Algorithms based upon these results are described in [15] and [7]. In [12] a non-linear weighted least squares approach is studied while in [12] the problem is considered as a separable non-linear least squares problem, since it is linear in the amplitudes c_i once the exponents b_i are known. A Remes type algorithm is developed in [1] while a Padé-approximant approach together with conditions for existence and uniqueness are discussed in [13].

As pointed out by several authors [10,15,12] the fit of data by a sum of exponentials and in particular by *real* exponents may be very badly conditioned; small changes in the data can cause a large change in the best-fit parameters. Varah [14] illuminates this by plots of the landscape of the Euclidean distance in several examples

$$J(\boldsymbol{b}, \boldsymbol{c}) := \sum_{j=0}^{m} \left(y_j - f(t_j)\right)^2, \tag{1.2}$$

which show very flat valleys and very steep slopes and in which the minimum seems badly determined. This suggest that minimization of the object function (1.2) is a difficult task, which is aggravated by the fact that J is invariant for interchanges of pairs of unknowns (b_i, c_i), which may cause endless loops in an iterative minimization procedure; moreover, such a method always poses the problem of a suitable starting point.

For the case, where the data are sampled at equidistant time intervals ($t_j := t_0 + jd$, $j = 0, \ldots, m$) the values $f(t_j)$ of the model admit a description as outputs of an n-dimensional linear system. Reliable reconstructions can be made by the algorithms of Zeiger and McEwen [16] and Kung [9]. In both algorithms the data are put into a Hankel matrix which is decomposed in signal and noise parts via a singular value decomposition. In the first algorithm the unknown exponents emerge as the eigenvalues of the restriction of the shifted Hankel matrix to the signal space and in the second as the eigenvalues of a matrix defined by shifts in the orthogonal projection onto the signal space. The coefficients can either be derived directly from the eigenvectors or indirectly from a linear least squares problem, the latter usually resulting in a somewhat smaller residue. The nonlinear part of the approximation problem is in this approach reduced to a singular value problem and an eigenvalue problem, which are both well-conditioned and which can be solved nowadays by reliable and fast algorithms. This method is easily implemented and works well in the presence of noise.

In our opinion the assets of this algorithm over some Newton-like method are:

(1) shorter computation time;
(2) no need to provide a suitable starting point for the iteration;
(3) less chance of breakdown if two exponents are near to each other;
(4) the algorithm provides a good answer to the question whether the quality of the data justifies a fit by a sum of—say—three exponentials.

Finally, we remark, that this algorithm works equally well in the (less sensitive) case of complex exponents, and that it can be competitive with FFT, if only the main modes are to be determined, since it requires much less data points for an accurate estimate of this mode.

The aim of this paper is to show, that the error in the computed exponents is of the order $O(\sigma_{n+1}/(\sigma_n - \sigma_{n+1}))$, the quotient of the largest noise singular value and the gap between the signal and noise singular values of the Hankel matrix. A necessary condition for such an estimate

is that the last column of the shifted Hankel matrix on $(y_1, \ldots, y_{2m-1})$ does not contain significant extra information on the signal, that is not already contained in the remaining columns or, in experimental terms, that the phenomenon to be modelled has been measured during a sufficiently long time. It does however not imply that the signal has to have died out. The key that leads to our error estimate is the observation that the singular vectors of the Hankel and shifted Hankel matrices are approximately parallel. Our error estimate is much better than Kung's [9], since the constant in the order term does not contain the (rather large) norm of the pseudo inverse of the Hankel matrix.

The error estimate is given in terms of the gap between the singular values. If a set of data can be modelled accurately by a sum of n exponentials (n being small w.r.t. the number of data) the data vectors have to span approximately an n-dimensional subspace. Hence the matrix consisting of those vectors can be approximated well by a matrix of rank n and not by one of smaller rank. This implies that there is a gap between the signal and noise singular values.

2. The algorithms of Zeiger–McEwen and Kung

Let us assume, that we are given the outputs $\{y_k\}_{k=0}^m$ of the non-degenerate n-dimensional linear system

$$\begin{aligned} &x_{k+1} := Ax_k, \qquad x_0^{\mathrm{T}} := (1, \ldots, 1), \\ &y_k = c^{\mathrm{T}} x_k, \qquad c_i \neq 0, \quad i = 1, \ldots, n \end{aligned} \tag{2.1}$$

with x_k and c in $\mathbb{R}^n$ and with A a diagonal matrix. In this section we describe the realisation algorithms of Zeiger–McEwen [16] and Kung [9]. Let H be the $p \times q$ Hankel matrix with $p > n$, $q > n$

$$H := \begin{pmatrix} y_0 & \cdots & y_{q-1} \\ \vdots & & \vdots \\ y_{p-1} & \cdots & y_{p+q-2} \end{pmatrix}, \qquad H_{i,j} = y_{i+j-1}, \tag{2.2}$$

whose rows and columns consist of consecutive outputs of the system (2.1). Since

$$y_k = c^{\mathrm{T}} x_k = c^{\mathrm{T}} A^k x_0,$$

the Hankel matrix admits the factorization into an observability matrix F and a controllability matrix G,

$$H = \begin{pmatrix} c^{\mathrm{T}} \\ c^{\mathrm{T}} A \\ \vdots \\ c^{\mathrm{T}} A^{p-1} \end{pmatrix} \left(x_0 \mid Ax_0 \mid \cdots \mid A^{q-1} x_0 \right) =: FG. \tag{2.3}$$

If the system (2.1) is non-degenerate (the geometric multiplicity of each eigenvalue of A is one and c has no zero components), both factors in (2.3) are of full rank n.

In the approach of Zeiger and McEwen [16] the system matrix A can be reconstructed, using the fact, that the shifted Hankel matrix $\overline{H}$ with entries $\overline{H}_{ij} := y_{i+j}$ satisfies the analogous factorisation

$$\overline{H} = FAG,$$

where F and G are as defined in (2.3). Since $F: \mathbb{R}^n \to \mathbb{R}^p$ and $G: \mathbb{R}^q \to \mathbb{R}^n$ have full rank, any factorisation $H = \tilde{F}\tilde{G}$ with $\tilde{F}: \mathbb{R}^n \to \mathbb{R}^p$ and $\tilde{G}: \mathbb{R}^q \to \mathbb{R}^n$ satisfies

$$I_n = \tilde{F}^+ H \tilde{G}^+ = \tilde{F}^+ F G \tilde{G}^+. \tag{2.4}$$

where F^+ denotes the pseudo-inverse of F. This implies that the matrix M,

$$M := \tilde{F}^+ H \tilde{G}^+ = \tilde{F}^+ F A G \tilde{G}^+ \tag{2.5}$$

is similar to the system-matrix A. The factorization can be obtained from the singular value decomposition $H = U\Sigma V^{\mathrm{T}}$. So we find

$$M := \Sigma^{-1/2} U^{\mathrm{T}} \overline{H} V \Sigma^{-1/2} \tag{2.6}$$

as a suitable reconstruction of the system-matrix.

In the approach of Kung [9] the system matrix is derived from the shift properties of the rows and columns in the left and right factors F and G in (2.3) respectively; e.g. the jth row of F multiplied by A yields the $(j+1)$st row. If we denote the first and last rows of F by $\boldsymbol{c}^{\mathrm{T}}$ and $\boldsymbol{d}^{\mathrm{T}}$ respectively and if we partition the matrix F in two ways,

$$F = \begin{pmatrix} \boldsymbol{c}^{\mathrm{T}} \\ F_\ell \end{pmatrix} = \begin{pmatrix} F_a \\ \boldsymbol{d}^{\mathrm{T}} \end{pmatrix}, \tag{2.7}$$

then the lower and upper parts F_ℓ and F_a are both of full column rank n and satisfy the relation

$$F_\ell = F_a A. \tag{2.8}$$

For any left factor $\tilde{F}$ in the factorisation $H = \tilde{F}\tilde{G}$ with $\tilde{F}: \mathbb{R}^n \to \mathbb{R}^p$ and $\tilde{G}: \mathbb{R}^q \to \mathbb{R}^n$ we find analogously, using (2.4),

$$\tilde{F}_\ell = F_\ell G \tilde{G}^+ = F_a A G \tilde{G}^+ = \tilde{F}_a (G\tilde{G}^+)^{-1} A G \tilde{G}^+ = \tilde{F}_a M, \tag{2.9}$$

where M is similar to A; hence M can be computed from the equation

$$M = \tilde{F}_a^+ \tilde{F}_\ell.$$

In particular, we can use the factorisation, given by the SVD $H = U\Sigma V^{\mathrm{T}}$, where the columns of U form an orthonormal basis in the range of H. We find

$$\begin{aligned} M &= U_a^+ U_\ell = \left(U_a^{\mathrm{T}} U_a\right)^{-1} U_a^{\mathrm{T}} U_\ell \\ &= \left(I_n - \boldsymbol{d}\boldsymbol{d}^{\mathrm{T}}\right)^{-1} U_a^{\mathrm{T}} U_\ell = \left(I_n + \frac{\boldsymbol{d}\boldsymbol{d}^{\mathrm{T}}}{1 - \boldsymbol{d}^{\mathrm{T}}\boldsymbol{d}}\right) U_a^{\mathrm{T}} U_\ell, \end{aligned} \tag{2.10}$$

where $\boldsymbol{d}^{\mathrm{T}}$ is the last row of U. The assumption that (2.1) is non-degenerate, implies that both factors F and G in (2.3) have full column rank n. Because $p > n$, this applies equally well to the parts F_ℓ and F_a in the partitions (2.7) and hence also to $\tilde{F}_\ell$ and $\tilde{F}_a$ in (2.9) and to U_a and U_ℓ in (2.10). Since the columns of U are orthonormal, we can extend U to an orthogonal matrix $(U \mid U_2)$. If $\boldsymbol{d}$ has length one, the last row of U_2 is zero and the orthogonality of the extension implies, that $\boldsymbol{d}$ must be orthogonal to the remaining rows of U, and hence rank $(U_a) < n$. So we find that the denominator $1 - \boldsymbol{d}^{\mathrm{T}}\boldsymbol{d}$ in (2.10) cannot vanish.

Formula (2.10) offers an easy way to reconstruct a system matrix for (2.1). Given a system matrix, the coefficient vector $\boldsymbol{c}$ can be determined from $\{y_k\}$ by a linear least squares fit.

3. Error estimates and noise

Let us now assume that $\{y_k\}_{k=0}^{m}$ are the outputs of the system (2.1), perturbed by noise. We shall consider the question how close an approximation to a system-matrix can be obtained by both algorithms. For comparisons we denote by $\{\hat{y}_k\}_{k=0}^{m}$ the unperturbed outputs, and more generally we shall denote by $\hat{X}$ the unperturbed counterpart of the object X. So we have the perturbed, unperturbed, and noise Hankel matrices H, $\hat{H}$ and W,

$$H = \hat{H} + W, \quad \|W\| < \epsilon; \qquad \overline{H} = \hat{\overline{H}} + \overline{W}, \quad \|\overline{W}\| < \epsilon. \tag{3.1}$$

In general the rank of H will be larger than n and we have to approximate it by a best rank n approximation. We can partition the singular value decomposition

$$H = U\Sigma V^{\mathrm{T}} = (U_1 \mid U_2)\begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix}\begin{pmatrix} V_1^{\mathrm{T}} \\ V_2^{\mathrm{T}} \end{pmatrix} \tag{3.2}$$

in such a way, that Σ_1 contains the n largest singular values, and we can approximate H by $U_1\Sigma_1 V_1^{\mathrm{T}}$. The reconstruction of Zeiger and McEwen is then given by

$$M = \Sigma_1^{-1/2} U_1^{\mathrm{T}} \overline{H} V_1 \Sigma_1^{-1/2} \tag{3.3}$$

and Kung's realisation by

$$M = \left(I_n + \frac{d_1 d_1^{\mathrm{T}}}{1 - d_1^{\mathrm{T}} d_1}\right) U_{1\nearrow} U_{1\swarrow}, \tag{3.4}$$

where d_1^{T} is the last row of U_1.

Both methods of approximate realisation yield an approximate system matrix M, whose error can be estimated by the norm of the noise matrix W and the gap between the singular values of Σ_1 and Σ_2:

Theorem. *For each of the approximate system matrices* M (3.2)–(3.3) *a similarity transformation* S *and a constant* C *exists, such that*

$$\|M - SAS^{-1}\| < C(\|W\| + \sigma_{n+1})/(\sigma_n - \sigma_{n+1}), \tag{3.5}$$

where $\{\sigma_j\}$ *denote the singular values of* H, *ordered in decreasing order, provided* $\|W\| + \sigma_{n+1} < 2(\sigma_n - \sigma_{n+1})$.

Proof. As is well known, we have

$$\|\hat{H} - U_1\Sigma_1 V_1^{\mathrm{T}}\| < \|W\| + \sigma_{n+1};$$

moreover the SVD of $\hat{H}$ can be chosen such that

$$\hat{H} = (\hat{U}_1 \mid \hat{U}_2)\begin{pmatrix} \hat{\Sigma}_1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} \hat{V}_1^{\mathrm{T}} \\ \hat{V}_2^{\mathrm{T}} \end{pmatrix}$$

with $\hat{U}_1$ satisfying

$$\|U_1 - \hat{U}_1\| < \frac{\|W\| + \sigma_{n+1}}{\sigma_n - \sigma_{n+1}} =: \epsilon$$

Obviously the same estimate holds for the errors in the parts d_1, U_{1u} and $U_{1\ell}$ in the partition of U_1; e.g.

$$\|d_1 - \hat{d}\| = \left\| \begin{pmatrix} 0 \\ d_1^T - \hat{d}^T \end{pmatrix} \right\| < \left\| \begin{pmatrix} U_{1u} - \hat{U}_u \\ d_1^T - \hat{d}^T \end{pmatrix} \right\| < \epsilon.$$

This immediately establishes the validity of (3.5) for Kung's realization (3.4), provided $\| d_1 \|$ is bounded away from 1. As is easily seen from (3.2), we have

$$d_1 = \Sigma_1^{-1} V_1^T h_p$$

where h_p^T is the last row of H. If all eigenvalues of the system matrix A are inside the unit circle (decaying exponents) and if the signal is measured for long enough time, $\| d_1 \|$ can easily be made smaller than (say) $\frac{1}{2}$, yielding $C \leqslant 3$ in (3.5).

The proof of the validity of (3.5) for realisation (3.3) is more involved and we refer the reader to [3] for the details. The basic idea is to insert in (3.3) the SVD of the shifted Hankel matrix $\overline{H}$, split into parts analogously to (3.2),

$$M = \Sigma_1^{-1/2} U_1^T \overline{U}_1 \overline{\Sigma}_1 \overline{V}_1^T V_1 \Sigma_1^{-1/2} + \Sigma_1^{-1/2} U_1^T \overline{U}_2 \overline{\Sigma}_2 \overline{V}_2^T V_1 \Sigma_1^{-1/2}.$$

The second term is in norm smaller than $\bar{\sigma}_{n+1}/\sigma_n$ and can be neglected from the outset. The first part can be factored as KL,

$$K := \Sigma_1^{-1/2} U_1^T \overline{U}_1 \overline{\Sigma}_1^{1/2}, \qquad L := \overline{\Sigma}_1^{1/2} \overline{V}_1^T V_1 \Sigma_1^{1/2}.$$

Under the condition that the gap between the columns of U_1 and the last column of $\overline{H}$ is small (the tail of the signal $\{y_k\}_{k=0}^{m}$ does not contain essentially new information), the (i, j)-terms with $i > j$ of $U_1^T \overline{U}_1$ are small enough to annihilate the possibly large factor $\sigma_i^{-1/2} \bar{\sigma}_j^{1/2}$ in the (i, j)-term of K, and analogously for L. Standard error estimates for spectral projections then yield the estimate (3.5) for realisation (3.3). □

Remark. In practice the value of $\| d_1 \|$ is easily monitored, such that appropriate action can be taken, in case it comes close to one.

4. Statistical analysis of Kung's algorithm and some numerical experiments

In this section we provide some more insight in the influence of additive noise for Kung's algorithm, leading to some important considerations on the choice of the optimal dimensions (p, q) of the Hankel matrices for a given number $m = p + q$ of data with respect to the accuracy of the results. Concerning the noise we shall assume

- the noise is additive ($y_k = \hat{y}_k + w_k$), has zero mean and variance σ^2;
- the noise samples are interindependent ($E(w_i w_j) = 0$) and do not depend on the data ($E(w_i \hat{y}_j) = 0$).

This implies that the noisy, exact and pure noise Hankel matrices H, $\hat{H}$ and W (cf. 3.1) with $H = \hat{H} + W$ satisfy $E(\hat{H}^T W) = 0$ and hence

$$E(H^T H) = \hat{H}^T \hat{H} + E(W^T W). \tag{4.1}$$

Because noise samples are uncorrelated, we have $E(W^{\mathrm{T}}W) = p\sigma^2 I_q$. Using the SVD of $\hat{H}$, $\hat{H} = \hat{U}_1\hat{\Sigma}_1\hat{V}_1$ with $\hat{\Sigma}_1 : \mathbb{R}^n \to \mathbb{R}^n$, we find

$$E(H^{\mathrm{T}}H) = \begin{pmatrix}\hat{U}_1 & \hat{U}_2\end{pmatrix}\begin{pmatrix}\hat{\Sigma}_1 + p\sigma^2 I_n & 0 \\ 0 & p\sigma^2 I_{q-n}\end{pmatrix}\begin{pmatrix}\hat{V}_1^{\mathrm{T}} \\ \hat{V}_2^{\mathrm{T}}\end{pmatrix}. \tag{4.2}$$

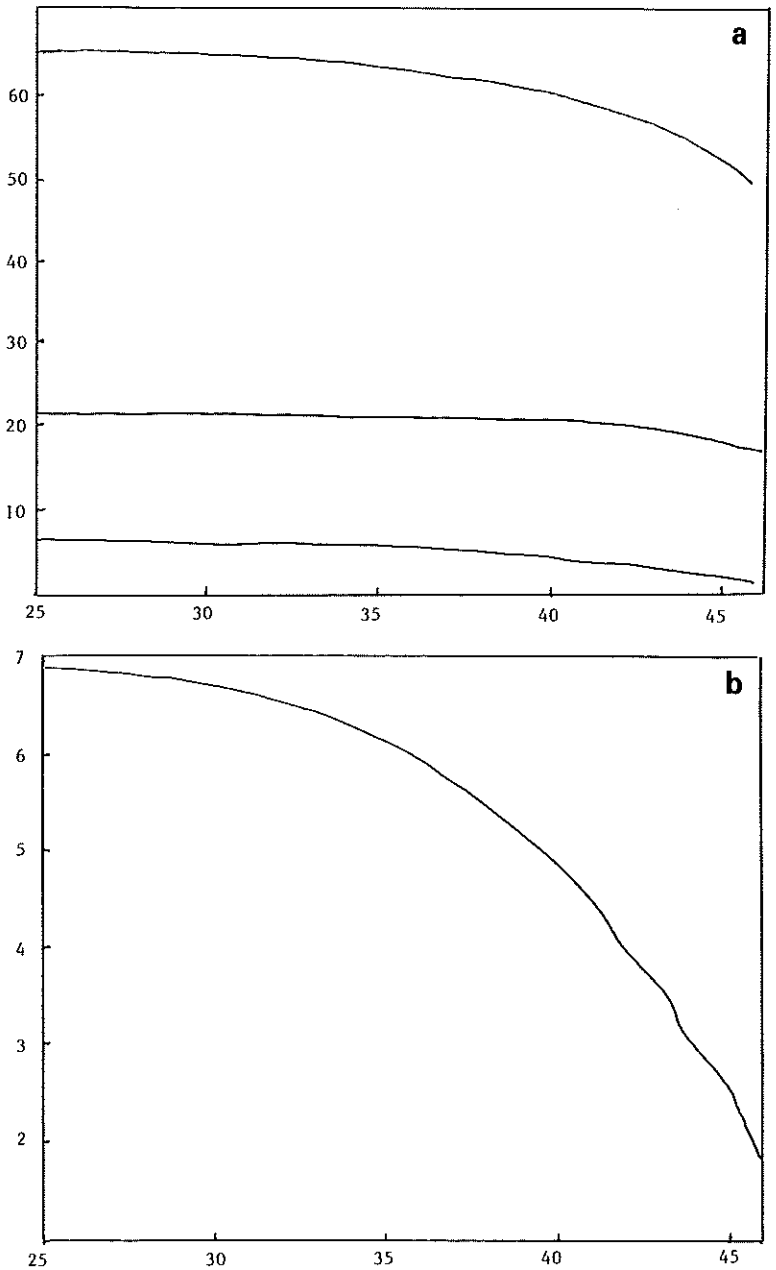

Fig. 1. Singular values of the $p \times (50 - p)$ Hankel matrix of rank 3 on the exact data $y_k = 5\ (0.95^k + 6(-0.85)^k + 10(0.77)^k,\ k = 0,\ldots,48$ as a function of the number of rows p. (a) all three singular values. (b) third singular value enlarged.

Due to the law of large numbers, the singular values of $H^{\mathrm{T}}H$ will satisfy

$$\sigma_i^2 \cong \hat{\sigma}_i^2 + p\sigma^2, \quad i = 1, \ldots, q, \tag{4.3}$$

provided p in sufficiently large in comparison to q. Moreover, if the gap between $\hat{\sigma}_n^2 + p\sigma^2$ and $p\sigma^2$ is large enough, i.e. if $\hat{\sigma}_n^2/p\sigma^2 \gg 1$, then the left singular subspace spanned by the columns of U_1 is well-determined. Hence, increasing overdetermination (i.e. an increase of p) may increase the accuracy of U_1, which is of crucial importance in Kung's algorithm, where only the left singular vectors of H are used, as was argued by De Moor and Vandewalle [5].

In most practical situations the exponents to be recovered are negative (the system is stable, $y_k \to 0$ for $k \to \infty$) and hence the amount of meaningful datapoints in the sequence $\{y_k\}$ is limited. Suppose now that we have $m+1$ datapoints $\{y_k\}_{k=0}^{m}$ available. Those can be put into a $p \times q$ Hankel matrix H_{pq} with $p + q = m$. Kung's algorithm requires only $p \geq q \geq n$ (and not all three equal) and we can ask the question which choice is best. The statistical analysis presented above suggests choosing p as large as possible. However, the relative gap

$$\left(\hat{\sigma}_n^2(p, q) + p\sigma^2\right)^{1/2} - p^{1/2}\sigma\big)/\left(p^{1/2}\sigma\right) \tag{4.4}$$

may play an antagonistic role.

We have studied both aspects in a number of simulations using Pc Matlab [11] and we have observed the following phenomena:

(1) The singular values of the Hankel matrix H_{pq} on $m+1$ 'exact' datapoints decrease if p is increased ($p > q$, $p + q = m$), the smallest one quite strongly as can be seen in Figs. 1(a) and (b). See also the remark at the end of this section.

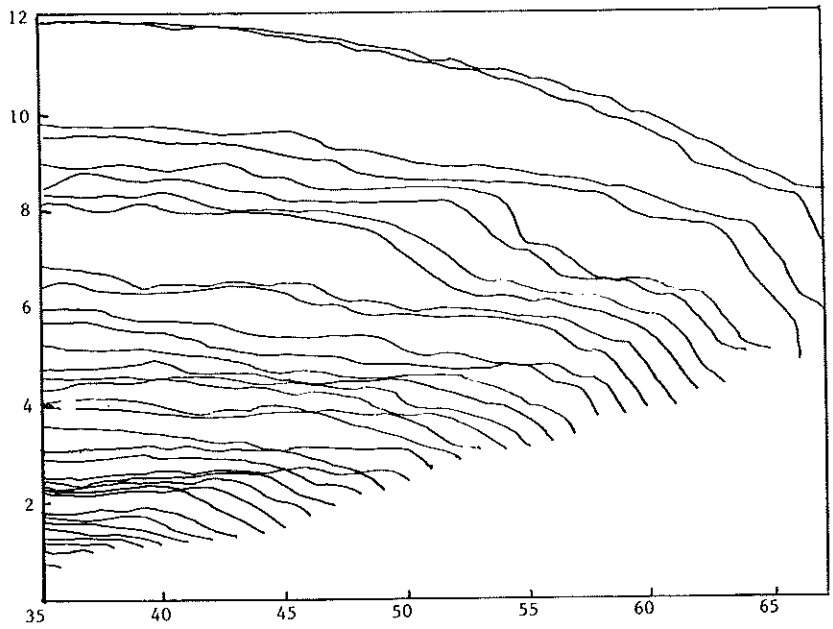

Fig. 2. Singular values of a pure noise $p \times (70 - p)$ Hankel matrix, where the noise is distributed normally with zero mean and unit variance, as a function of the number of rows $p = 35, \ldots, 67$.

(2) The singular values of the pure noise Hankel matrix W_{pq} on $m+1$ noise samples decrease weakly if p is increased ($p > q$, $p + q = m$) until their existence ends by lack of dimensions. The remaining ones come close to each other, making the noise spectrum approximately isotropic; an example is displayed in Fig. 2.

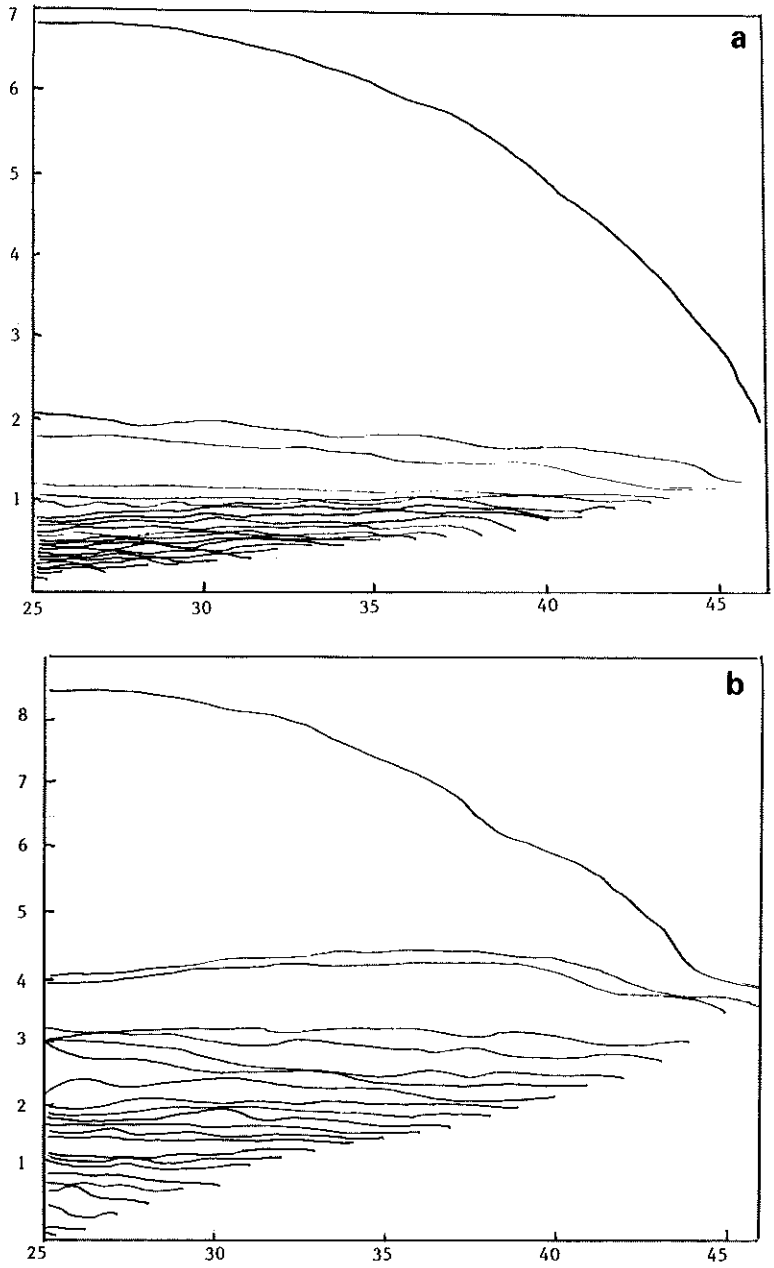

Fig. 3. Singular values except largest two of the $p \times (50-p)$ Hankel matrix on the noisy data (4.5) as a function of p for $p = 25,\ldots,46$ for three values of the noise multiplier. (a) noise multiplifier 0.01 (b) noise multiplier 0.025 (c) noise multiplier 0.05.

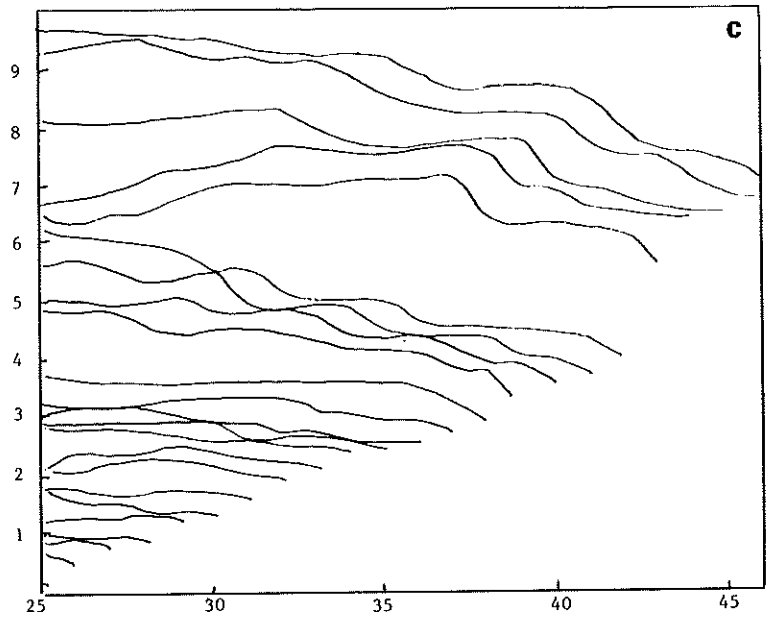

Fig. 3. continued.

Both phenomena have an opposite effect. In our simulations the deteriorating effect of the quick decrease of the smallest signal singular value $\sigma_n(p, q)$ for increasing p on the accuracy of the approximate signal space spanned by the columns of U_1 was much more important than the advantage of increasing overdetermination. For that reason, we conclude that the choice $p \simeq q$ is most available.

As an example we display the results of the following simulation. We generate 49 datapoints ($k = 0, \ldots, 48$)

$$y_k := 5 * (0.95)^k + 6 * (-0.85)^k + 10 * (0.77)^k + w_k, \tag{4.5}$$

where the noise term w_k is given by

$$w_k := \rho * y_0 * (\text{random number}),$$

and the (pseudo) random number is normally distributed with zero mean and unit variance. For the noise multiplier ρ we have chosen three values, namely 0.01, 0.025 and 0.05. In the first two cases we observe for $p = 25$ a substantial gap between the smallest signal singular value σ_3 and the largest noise singular value σ_4 plotted in Fig. 3; this gap diminishes for increasing p as does the accuracy of the corresponding reconstructed exponents in Fig. 4. In the case of noise multiplier $\rho = 0.05$ the gap between σ_3 and σ_4 in Fig. 3(c) is so small, that the third exponent is hardly recoverable; in Fig. 4(d) we see indeed that the error in the third exponent is quite large for all values of p. On the other hand, the gap between σ_2 and σ_3 remains quite large, such that the first two exponents emerge with a much smaller error. This effect was indeed predicted by the error analysis in Section 3. If p increases the errors in first and second recovered exponents remain more or less constant, until the number of columns ($= 50 - p$) becomes too small and the error increases drastically.

Remark. In simulations we have already observed that the singular values of the Hankel matrices $H_{p,m-p}$ on the same data set $\{y_0, \ldots, y_{m-2}\}$, generated by a sum of decaying exponentials, decrease if $p > m - p$ and p increases. For their sum of squares, which is equal to the Frobenius

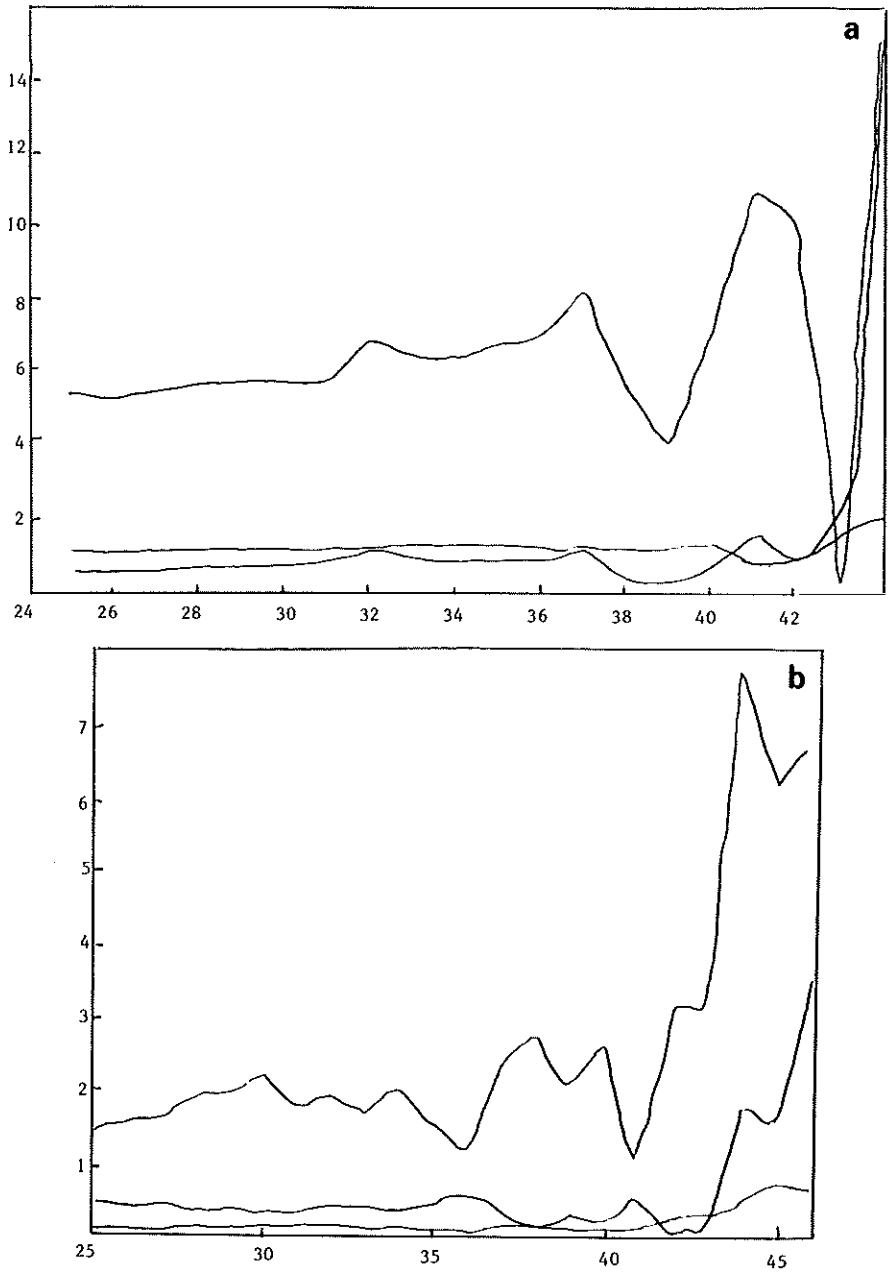

Fig. 4. Relative error in percents in the exponents reconstructed from the $p \times (50 - p)$ Hankel matrices on the noisy data (4.5) by the algorithm of Kung, plotted versus p for three values of the noise multiplier. (a) noise multiplier 0.01. (b) noise multiplier 0.025. (c^1) noise multiplier 0.05, third exponent. (c^2) noise multiplier 0.05, first and second exponent.

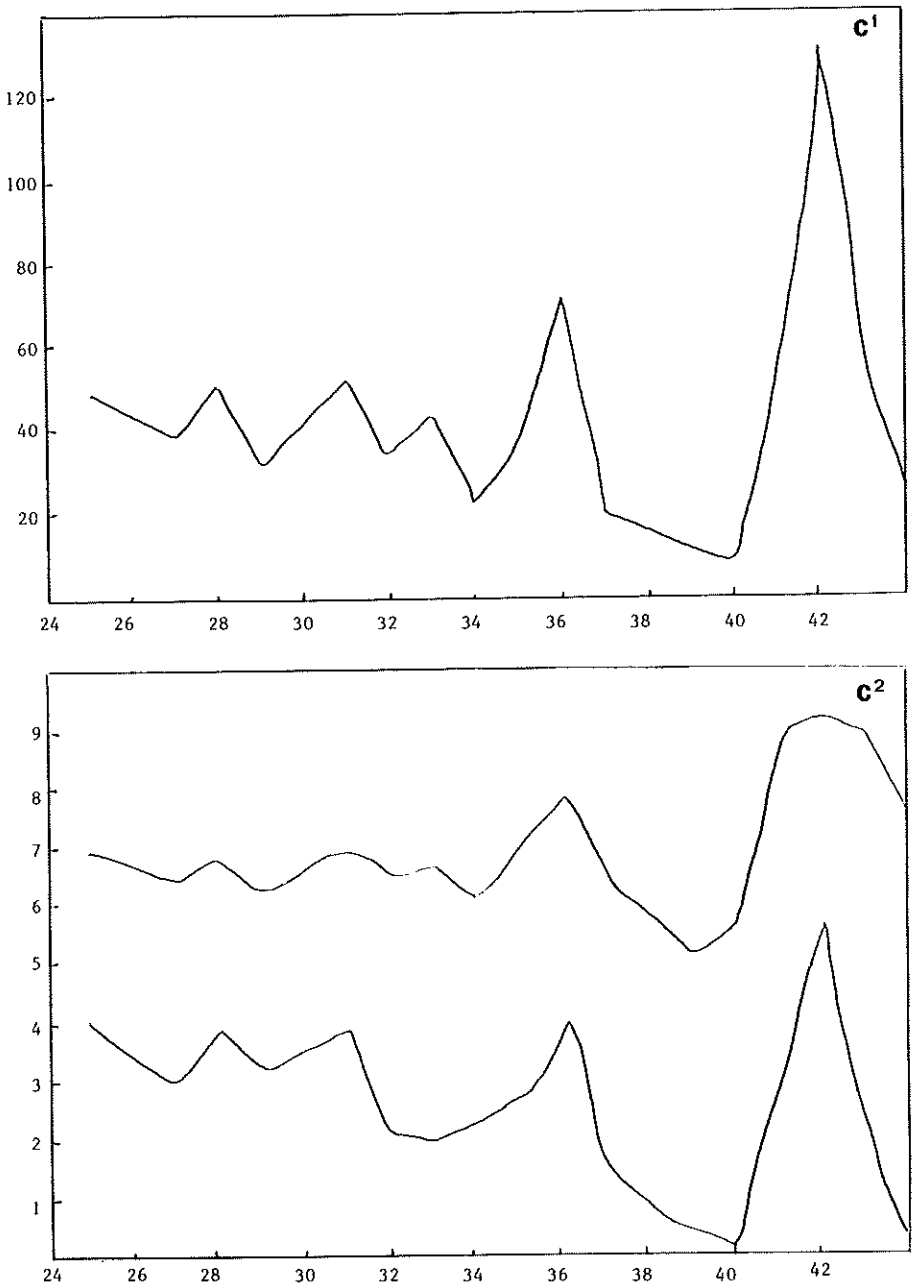

Fig. 4. continued.

norm of the matrix, this property is obvious; however, we do not know whether it is true for each singular value individually. For more general data sets, e.g. randomly generated data, this is no longer true, as can be seen in Fig. 3(c).

5. Conclusions

In this paper error estimates were derived for the fit of a sum of exponentials to a set of equidistant noisy data using the algorithms of Zeiger and McEwen [16], and Kung [9]. We have shown that, if we recover n exponents from the data, the error in those exponents is of the order $\sigma_{n+1}/(\sigma_n - \sigma_{n+1})$, where $\{\sigma_i\}$ are the singular values of the associated Hankel matrices. Although in experiments, described in De Moor and Vandewalle [5], de Groen [3] and in this paper, this bound appears somewhat pessimistic, it nevertheless yields a good criterion whether or not it makes sense to try a fit by a sum of n exponentials.

Moreover, we have made a study of the accuracy obtainable for a given set of data, If we vary the dimensions of the Hankel matrix constructed with these data. We have given an argument for strong overdetermination and another against it. From experiments we have concluded that (for the case of a sum of *real* exponents at least) a square Hankel matrix is preferable in general.

References

[1] D. Braess, Die Konstruktion der Tschebyscheff Approximierenden bei der Anpassung mit exponentialsummen, *J. Approx. Theory* **3** (1970) 261–273.

[2] D.G. Cantor and J.W. Evans, On approximations by positive sums of powers, *Siam J. Appl. Math.* **18** (1970) 380.

[3] P.P.N. de Groen, The fit of a sum of exponentials to data, Submitted for publication.

[4] B. De Moor and J. Vandewalle, A numerically reliable algorithm for fitting a sum of exponentials or sinusoids to noisy data, Proc. 3rd IFAC/IFIP international symposium on computer aided design in control and engineering system, Lyngby, Copenhagen, Denmark, 1985.

[5] B. De Moor and J. Vandewalle, A geometrical strategy for the identification of state space models of linear multivariable systems with singular value decomposition, submitted to the 10th World Congress of IFAC, July 1987, Munich.

[6] R. de Prony, Essay expérimentale et analytique, *J. Ec. Polytech. Paris* **1** (1795) 24–76.

[7] J.W. Evans, W.B. Gragg and R.J. Leveque, On least squares exponential sum approximation with positive coefficients, *Math. Comp.* **34** (1980) 203–212.

[8] J.A. Jacquez, *Compartmental Analysis in Biology and Medicine* (Elsevier, Amsterdam, 1972).

[9] S.Y. Kung, A new identification and model reduction algorithm via singular value decomposition, Proc. 12th Asilomar Conf. Circuits, Syst. Computer, Pacific Grove, CA, 1978, pp. 705–715.

[10] C. Lanczos, *Applied Analysis* (Prentice Hall, Englewood Cliffs, NJ, 1956).

[11] Pc Matlab, A software package for matrix computations, The Matworks Inc., Sherborn, MA, 1986.

[12] A. Ruhe and P.A. Wedin, Algorithms for separable nonlinear least squares problems, *Siam Rev.* **22** (3) (1980).

[13] A. Sidi, Interpolation at equidistant points by a sum of exponential functions, *J. Approx. Theory* **34** (1982) 194–210.

[14] J.M. Varah, On fitting exponentials by nonlinear least squares, *SIAM J. Sci. Statist. Comput* **6** (1985) 30–44.

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

[16] H.P. Zeiger and A.J. McEwen, Approximate linear realization of given dimension via Ho's algorithm, *IEEE Trans. Aut. Control* **AC-19** (1953) (1974).