On Least Squares Exponential Sum Approximation With Positive Coefficients*

By John W. Evans, William B. Gragg and Randall J. LeVeque

Abstract. An algorithm is given for finding optimal least squares exponential sum approximations to sampled data subject to the constraint that the coefficients appearing in the exponential sum are positive. The algorithm employs the divided differences of exponentials to overcome certain problems of ill-conditioning and is suitable for data sampled at noninteger times.

1. Introduction. In this paper the following least squares approximation is considered:

Let $\gamma_0, \ldots, \gamma_N$ be $N + 1$ real data points sampled at times $0 \leqslant t_0 < \cdots < t_N$, and let $w_0, \ldots, w_N > 0$ be a set of weights. For a compact subset S of the real numbers minimize

$$\sum_{n=0}^{N} w_n \left(\left(\sum_{i=1}^{r} a_i \exp(-\alpha_i t_n) \right) - \gamma_n \right)^2,$$

subject only to $a_1, \ldots, a_r > 0$ and $\alpha_1, \ldots, \alpha_r \in S$ and with no limitation on the number of terms r.

An algorithm is given which extends that of [3], [10] so that noninteger times are acceptable. The algorithm uses the exponential [7], [9] of a certain triangular matrix [8] to obtain divided differences of the exponential function. For other work on this topic the reader is referred to [1], [2], [5] and to the references in [10].

We begin by giving some definitions and results from [3] in Section 2 and then present in Section 3 a modification of the numerical procedure of that paper which applies to approximation by positive sums of exponentials. The role of divided differences in a linear least squares step of the numerical procedure is considered in Section 4. Section 5 contains the method of computing the divided differences of the exponential and other functions that is the basis of this paper. Computed examples of approximation by positive exponential sums are presented in Section 6. In Section 7 the computation of the divided differences of powers (as used in [3], [10]) is considered.

2. Preliminaries and Definitions. In this section we give some of the definitions and results of reference [3] with minor modifications to suit our present purpose.

Let E denote an expression of the form $\sum_{i=1}^{r} a_i \exp(-\alpha_i t)$, where $a_1, \ldots, a_r$, $\alpha_1, \ldots, \alpha_r$ are real. We call E an exponential polynomial. If $a_1, \ldots, a_r$, the *coefficients* of E, are nonnegative, we say that E is positive; and if $\alpha_1, \ldots, \alpha_r$, the *time constants* of E, are in S for a fixed compact subset S of the real numbers, we say that E is an *S-exponential polynomial*. For fixed times $0 \leqslant t_0 < \cdots < t_N$ let $E(n)$ denote the expression E evaluated at t_n, and let $\overline{E}$ be the column vector with entries $E(0), \ldots, E(n)$. For real α let E_α denote the expression $\exp(-\alpha t)$ so that $E_\alpha(n) = \exp(-\alpha t_n)$.

Given weights $w_0, \ldots, w_N > 0$ let $W := \operatorname{diag}(w_0, \ldots, w_N)$, where $:=$ is used as usual to indicate that the value on the right is assigned to the variable on the left. Also, let $\|\gamma\| := (\gamma^T W \gamma)^{1/2}$ define a norm $\|\cdot\|$ for a general $N + 1$ column vector γ where T denotes transpose.

Now for a fixed data vector γ with entries $\gamma_0, \ldots, \gamma_N$ let

$$(1) \qquad\qquad P_E(\alpha) := 2(\overline{E} - \gamma)^T W \overline{E}_\alpha .$$

The importance of P_E is seen in the following four propositions.

PROPOSITION 1 (OPTIMAL COEFFICIENTS). *If the coefficients $a_1, \ldots, a_r$ in $E = \sum_{i=1}^{r} a_i \exp(-\alpha_i t)$ are such that $\|\overline{E} - \gamma\|^2$ is minimized over all possible choices of coefficients, then $P_E(\alpha_i) = 0$ for $i = 1, \ldots, r$.*

Proof. The expression $\|\overline{E} + x\overline{E}_\alpha - \gamma\|^2 = \|\overline{E} - \gamma\|^2 + xP_E(\alpha) + x^2\|\overline{E}_\alpha\|^2$ must have a minimum at $x = 0$ for $\alpha = \alpha_i$ under the above assumption and thus $P_E(\alpha_i) = 0$ for $i = 1, \ldots, r$.

We say that such coefficients $a_1, \ldots, a_r$ are *optimal* for the time constants $\alpha_1, \ldots, \alpha_r$.

PROPOSITION 2 (IMPROVED APPROXIMATION). *The expression $\|\overline{E} + x\overline{E}_\alpha - \gamma\|^2$ is minimized as a function of x when $x = -P_E(\alpha)/2\|\overline{E}_\alpha\|^2$, where it takes on the value $\|\overline{E} - \gamma\|^2 - P_E(\alpha)^2/4\|\overline{E}_\alpha\|^2$.*

Proof. Using the identity in the proof of Proposition 1, set $d\|\overline{E} + x\overline{E}_\alpha - \gamma\|^2/dx$ equal to zero and solve for x. Then evaluate at this value of x.

PROPOSITION 3 (COMPARISON OF APPROXIMATIONS). *Suppose that $E = \sum_{i=1}^{r} a_i \exp(-\alpha_1 t)$ has optimal coefficients $a_1, \ldots, a_r$ as in Proposition 1 and that $F = \sum_{i=1}^{s} b_i \exp(-\beta_i t)$, then $\|\overline{E} - \gamma\|^2 - \|\overline{F} - \gamma\|^2 \leqslant -\sum_{i=1}^{s} b_i P_E(\beta_i)$.*

Proof. Observe that $2(\overline{E} - \gamma)^T W \overline{E} = \sum_{i=1}^{r} a_i P_E(\alpha_i) = 0$, while $2(\overline{E} - \gamma)^T W \overline{F} = \sum_{i=1}^{s} b_i P_E(\beta_i)$. From this it follows that

$$\|\overline{F} - \gamma\|^2 = \|(\overline{F} - \overline{E}) + (\overline{E} - \gamma)\|^2 = \|\overline{F} - \overline{E}\|^2 + \|\overline{E} - \gamma\|^2 + \sum_{i=1}^{s} b_i P_E(\beta_i).$$

PROPOSITION 4 (BEST APPROXIMATION). *The positive S-exponential polynomial $E = \sum_{i=1}^{r} a_1 \exp(-\alpha_i t)$ with $a_1, \ldots, a_r > 0$ best approximates γ in $\|\cdot\|$ over all posi-*

tive S-exponential polynomials if and only if $P_E(\alpha) \geqslant 0$ for all $\alpha \in S$ and $P_E(\alpha_i) = 0$ for $i = 1, \ldots, r$.

Proof. The necessity of $P_E(\alpha_i) = 0$ follows from Proposition 1 and that of $P_E(\alpha) \geqslant 0$ follows from Proposition 2. The sufficiency of these conditions follows from Proposition 3.

Remark. We call the exponential polynomial E of Proposition 4 a *best approximant*. That such a best approximant always exists is shown in [3] where a best approximant E is seen to be unique if $\|\bar{E} - \gamma\| > 0$.

If S is an interval and E is a best (but not perfect) approximant, then α_i is a root of even multiplicity of P_E for each time constant α_i of E which is in the interior of S. Because of this, approximants F near E (in the sense that $\|\bar{F} - \bar{E}\|$ is small) with optimal coefficients often have two time constants at simple roots of P_F near a double root in P_E (see [3], [10]).

3. The Algorithm. In this section the main features are given of our algorithm to find a best approximant among positive S-exponential polynomials to a data vector γ. Details are given in subsequent sections.

The Algorithm. 0. Given are the compact set S, $W := \text{diag}(w_0, \ldots, w_N)$ for weights $w_0, \ldots, w_N > 0$, times $0 \leqslant t_0 < \cdots < t_N$ and the data $\gamma := (\gamma_0, \ldots, \gamma_N)^T$. Initialize $r := 0$ with the result that $E := \Sigma_{i=1}^r a_i \exp(-\alpha_i t)$ is the empty sum with $E(n) = 0$ for $0 \leqslant n \leqslant N$.

1. Suppose $E = \Sigma_{i=1}^r a_i \exp(-\alpha_i t)$ is known where $r \geqslant 0$, $\alpha_1 > \cdots > \alpha_r$ and $\alpha_1, \ldots, \alpha_r \in S$ and where $a_1, \ldots, a_r$ are optimal coefficients for the $\alpha_i, \ldots, \alpha_r$ with $a_1, \ldots, a_r > 0$.

    a. $\varphi := \|\bar{E} - \gamma\|^2$.

    b. Exit if $P_E(\alpha) \geqslant 0$ for all $\alpha \in S$. In this case E is a best approximant.

    c. Find $\alpha_0 \in S$ to minimize $P_E(\alpha)$ for $\alpha \in S$.

    d. $a_0 := -P_E(\alpha_0)/2\|\bar{E}_{\alpha_0}\|^2$.

    e. $\{\bar{a}_1, \ldots, \bar{a}_{r+1}\} := \{a_0, \ldots, a_{r+1}\}$.

    f. $\{\alpha_1, \ldots, \alpha_{r+1}\} := \{\alpha_0, \ldots, \alpha_{r+1}\}$.

    g. $r := r + 1$.

    h. Renumber $\alpha_1, \ldots, \alpha_r$ and the corresponding $\bar{a}_1, \ldots, \bar{a}_r$ so that $\alpha_1 > \cdots > \alpha_r$.

2. a. Find optimal coefficients $a_1, \ldots, a_r$ for the time constants $\alpha_1, \ldots, \alpha_r$.

    b. If $a_1, \ldots, a_r > 0$, then

        (i) Exit if $r = N + 1$. In this case $E = \Sigma_{i=1}^r a_i \exp(-\alpha_i t)$ satisfies $\bar{E} = \gamma$ and E is a best approximant.

        (ii) Go to 1 if $r \leqslant N$.

3. We now have some $a_j \leqslant 0$ for some $1 \leqslant j \leqslant r$.

    a. $\tau := \min\{\bar{a}_j/(\bar{a}_j - a_j): 1 \leqslant j \leqslant r, a_j \leqslant 0\}$.

    b. $\{\bar{a}_1, \ldots, \bar{a}_r\} \leftarrow \{(1 - \tau)\bar{a}_1 + \tau a_1, \ldots, (1 - \tau)\bar{a}_r + \tau a_r\}$.

    c. Drop the zero coefficients $\bar{a}_j$ and the corresponding time constants α_j. Set r equal to the number of remaining coefficients and renumber the remaining

terms to obtain time constants $\alpha_1 > \cdots > \alpha_r$ with coefficients $\bar{a}_1, \ldots, \bar{a}_r$ > 0.

d. Go to 2.

Convergence. On each iteration we start 1 with the approximant E; and if the algroithm does not terminate, we finish with the approximant $F := \Sigma_{i=1}^{r} \bar{a}_i \exp(-\alpha_i t)$. By Proposition 2 we have $\|\bar{F} - \gamma\|^2 = \|\bar{E} - \gamma\|^2 - P_E^2(\alpha_0)/4\|\bar{E}\alpha_0\|$ so the improvement is at least as great as

$$(\min\{P_E(\alpha): \alpha \in S\})^2/4(\max\{\|\bar{E}_\alpha\|^2: \alpha \in S\}).$$

In 3 the variable τ is increased from 0 until a coefficient in $\Sigma_{j=1}^{r}((1-\tau)\bar{a}_j + \tau a_j)\bar{E}\alpha_j$ is zero. This is movement in a straight line from $(\bar{a}_1, \ldots, \bar{a}_r)$ toward the optimal $(a_1, \ldots, a_r)$ and results in no worse approximation. Along with the fact that $a_1, \ldots, a_n$ are optimal in 2 this guarantees that the improvement in 1 is not lost.

Now, if $G = \Sigma_{i=1}^{m} b_i \exp(-\beta_i t)$ is a best approximant, we have on returning to 1 from 2 that

$$\|\bar{E} - \gamma\|^2 - \|\bar{G} - \gamma\|^2 \leq \left(\sum_{i=1}^{m} b_i\right)(-\min\{P_E(\alpha): \alpha \in S\})$$

by Proposition 3 so that

$$\min\{P_E(\alpha): \alpha \in S\} \leq (\|\bar{G} - \gamma\|^2 - \|\bar{E} - \gamma\|^2)\bigg/\left(\sum_{i=1}^{m} b_i\right).$$

Thus, the assumption that $\varphi = \|\bar{E} - \gamma\|^2$ does not tend to $\|\bar{G} - \gamma\|^2$ leads to an immediate contradiction and theoretical convergence is proved. Further details are given in [3].

4. Divided Differences and the Linear Least Squares Problem. Initial implementations of the preceding algorithms for [3] revealed that the numerical aspects of the solution to the linear least squares problem in step 2 of the algroithm required close attention. For $A := (\bar{E}\alpha_1, \ldots, \bar{E}\alpha_r)$ we seek $a = (a_1, \ldots, a_r)^T$ such that $\|Aa - \gamma\|^2$ is minimized. By the remark at the end of Section 2 we often find that pairs of time constants tend to common limits as E approaches a best approximant. With standard methods of solution [6] this results in an early termination of the algorithm due to lack of improvement. The replacement of close pairs of time constants with single time constants is easily implemented, but this greatly slows the rate of convergence presumably because of the reduction of the degrees of freedom in the approximants. A solution was found in the use of divided differences to deal explicitly with the near degeneracies in A. Divided differences were used in the examples in [3]. The method is described in [10]. In these studies the calculations used are appropriate only for $t_0, \ldots, t_N$ which are integer multiples of some fixed real number and the divided differences are the divided differences of powers rather than of exponentials. (See Section 7.)

For $k \geq 1$ we define the divided differences [4] of vectors $\theta_1, \ldots, \theta_k$ with

respect to distinct real numbers $\alpha_1, \ldots, \alpha_k$ as

$$(2) \qquad [\theta_1, \ldots, \theta_k; \alpha_1, \ldots, \alpha_k] := \sum_{i=1}^{k} \left(1 \Big/ \prod_{\substack{1 \leq j \leq k \\ j \neq i}} (\alpha_i - \alpha_j) \right) \theta_i,$$

where the empty product when $k = 1$ is taken to be 1.

Note that for $k \geq 2$

$$[\theta_1, \ldots, \theta_k; \alpha_1, \ldots, \alpha_k]$$

$$(3) \qquad = \frac{[\theta_1, \ldots, \theta_{k-1}; \alpha_1, \ldots, \alpha_{k-1}] - [\theta_2, \ldots, \theta_k; \alpha_2, \ldots, \alpha_k]}{\alpha_1 - \alpha_k}.$$

We define the divided difference of a vector valued function f with respect to $\alpha_1, \ldots, \alpha_n$ by

$$(4) \qquad f[\alpha_1, \ldots, \alpha_k] := [f(\alpha_1), \ldots, f(\alpha_k); \alpha_1, \ldots, \alpha_k].$$

From (2) and (4) we obtain the matrix equation

$$(5) \qquad (f[\alpha_1] \cdots f[\alpha_1, \ldots, \alpha_r]) = (f(\alpha_1) \cdots f(\alpha_r))C,$$

where $C = (C_{ij})$ is a right triangular matrix with

$$C_{ij} = \begin{cases} 0, & j < i, \\ 1 \Big/ \displaystyle\prod_{\substack{k \neq i \\ 1 \leq k \leq j}} (\alpha_i - \alpha_k), & j \geq i, \end{cases}$$

for $1 \leq i, j \leq r$. Thus, the columns of $B := AC$ are the divided differences $(\bar{E}[\alpha_1] \cdots \bar{E}[\alpha_1, \ldots, \alpha_r])$ of $\bar{E}_\alpha$ with respect to $\alpha_1, \ldots, \alpha_r$; and to find a such that $\|Aa - \gamma\|^2$ is minimized, we may set $a := Ca'$ where a' is such that $\|Ba' - \gamma\|^2$ is minimized.

Now assuming a method for calculating the entries of B which remains accurate even when some of the $|\alpha_i - \alpha_j|$ are close to zero the problem of finding a' is computationally stable and singularities enter only in the direct calculation $a = Ca'$.

We recommend the following method for solution [10]. Use the modified Gram-Schmidt (see pp. 129–132 of [6]) method on the columns of $W^{1/2}B$ to obtain an $N + 1$ by r matrix Q with orthonormal columns and an r by r right triangular matrix R such that $W^{1/2}B = QR$. Now directly compute $Q^T W^{1/2}\gamma$, and then solve $Ra' = (Q^T W^{1/2}\gamma)$ by using back substitution on R. Finally, set $a = Ca'$.

It is easily verified that this gives the correct a' and a. In addition, for $E = \Sigma_{i=1}^{r} a_i \exp(-\alpha_i t)$ we have

$$W^{1/2}\bar{E} = W^{1/2}Aa = W^{1/2}Ba' = QRR^{-1}Q^T W^{1/2}\gamma.$$

And on setting $\rho := QQ^T W^{1/2}\gamma - W^{1/2}\gamma$, we have

$$P_E(\alpha) = 2\rho^T W^{1/2}\bar{E}_\alpha \quad \text{and} \quad \|\bar{E} - \gamma\|^2 = \rho^T \rho.$$

Thus, as a by-product of the above method we have that $P_E(\alpha)$ and $\|\bar{E} - \gamma\|^2$ are

computed without reference to $a_1, \ldots, a_r$ and so without the use of the matrix C which contains the singularities. The result is that only in step 3 is there a direct role for $a_1, \ldots, a_r$. The practical outcome is a substantial increase in the number of iterations in the program that can be carried out with continued improvement in the approximation.

Next the direct computation of the divided differences appearing in the matrix B is considered.

5. Computation of Divided Differences. The nth row of the matrix B is equal to $(g_{t_n}[\alpha_1], \ldots, g_{t_n}[\alpha_1, \ldots, \alpha_r])$ for $g_t(\alpha) := \exp(-t\alpha)$. We seek a method of calculating these terms which does not lose significance as $(\alpha_i - \alpha_j) \to 0$. We first state a proposition about divided differences in a matrix setting, and then use this proposition as the basis for our computation.

PROPOSITION 5. *Let*

$$
Z = \begin{bmatrix} \alpha_1 & \delta & & & \\ & \ddots & \ddots & & \\ & & \ddots & \delta & \\ & & & \ddots & \\ & & & & \alpha_s \end{bmatrix},
$$

and let f be any analytic function represented by a convergent power series. Then

$$
f(Z) = \begin{bmatrix} f[\alpha_1] & \delta f[\alpha_1, \alpha_2] & \cdot & \cdot & \cdot & \delta^{s-1} f[\alpha_1, \ldots, \alpha_s] \\ & f[\alpha_2] & \delta f[\alpha_2, \alpha_3] & & & \vdots \\ & & \cdot & \cdot & & \vdots \\ & & & \cdot & & \delta f[\alpha_{s-1}, \alpha_s] \\ & & & & \cdot & \\ & & & & & f[\alpha_s] \end{bmatrix}.
$$

Proof (*Opitz* [8]). Let $F = f(Z) = (F_{ij})$. Clearly, $F_{ii} = f(\alpha_i) = f[\alpha_i]$. Since F is a power series in Z, it follows that $FZ = ZF$. Equating the (i, j) elements of these products gives

$$
\delta F_{i\,j-1} + \alpha_j F_{ij} = \alpha_i F_{ij} + \delta F_{i+1\,j}
$$

or

$$
F_{ij} = \delta \frac{(F_{i+1\,j} - F_{i\,j-1})}{\alpha_j - \alpha_i},
$$

and the proposition is proved by induction on the successive superdiagonals of F using (3).

We have employed this proposition using the (p, p) Padé approximation ([7] and [9]) with $p = 12$ for the exponential function restricted to the interval $[-1, 0]$ in the following manner for a fixed $t > 0$:

For $\alpha_1 > \cdots > \alpha_r \geqslant 0$ choose an integer $k > 0$ so that $t\alpha_1/k \leqslant 1$; then set

$$\widetilde{Z} = \begin{bmatrix} -\alpha_1 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & -\alpha_r \end{bmatrix},$$

and set $\overline{Z} = t/k\widetilde{Z}$. The (p, p) Padé approximation to $\exp(\overline{Z})$ is given by $R_{pp}(\overline{Z}) = N_{pp}(\overline{Z})(D_{pp}(\overline{Z}))^{-1}$, where $N_{pp}(\overline{Z}) = \Sigma_{j=0}^{p} c_j \overline{Z}^j$ and $D_{pp}(\overline{Z}) = \Sigma_{j=0}^{p} c_j (-\overline{Z})^j$ with $c_j = (2p - j)!p!/(2p)!j!(p - j)!$.

Since the α_j are nonnegative, all terms summed in computing $D_{pp}(\overline{Z})$ have the same sign.

Once we have $R_{pp}(\overline{Z})$ approximating $\exp(\overline{Z})$, we approximate $\exp(k\overline{Z}) = \exp(t\widetilde{Z})$ by $(R_{pp}(\overline{Z}))^k$. Of course, the top row of $\exp(t\widetilde{Z})$ is $(g_t[\alpha_1] \cdots g_t[\alpha_1, \ldots, \alpha_r])$ as desired. The relation $\exp(t\widetilde{Z}) \exp(s\widetilde{Z}) = \exp((t + s)\widetilde{Z})$ can be exploited to reduce the number of times the above approximation is carried out; and if not all $\alpha_1, \ldots, \alpha_r$ are nonnegative, the relation

$$\exp(t\widetilde{Z}) = \exp(\beta t) \exp(t(\widetilde{Z} - \beta I))$$

for I the r by r identity matrix can be used.

We have found that better numerical results are obtained when $\alpha_1 > \cdots > \alpha_r$ as above.

6. Examples. The set S was chosen as $[0, 64]$ and a set of test data was generated by the model

$$T_1(t) = 0.6 \exp(-0.1t) + 0.3 \exp(-0.01t) + 0.1 \exp(-0.001t)$$

$$\equiv a_1 \exp(-\alpha_1 t) + a_2 \exp(-\alpha_2 t) + a_3 \exp(-\alpha_3 t).$$

Twenty data points were generated by setting $t_0, \ldots, t_{19}$ equal to 0, 1, 2, 3, 4, 5, 10, 30, 60, 150, 300, 400, 500, 1000, 1500, 2000, 3000, 4000, 5000, 6000 and setting $\gamma := T_1(t_n)$ rounded to four significant digits. The weight w_n was given the value $1/T_1(t_n)$ for $n = 0, \ldots, 19$. Each of these time values $t_0, \ldots, t_{19}$ was represented in seconds, minutes and hours and $\exp(\widetilde{Z})$, $\exp(60\widetilde{Z})$ and $\exp(3600\widetilde{Z})$ were estimated separately in step 2 of the procedure where

$$\widetilde{Z} = \begin{bmatrix} -\alpha_1 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & -\alpha_r \end{bmatrix},$$

as before. Suitable powers were then multiplied to give $\exp(t_n\widetilde{Z})$ for $n = 0, \ldots, 19$.

The procedure was coded in ALGOL using double precision (23 digits of significance) on the Burrough 6700 computer. The program was terminated after 80 iterations through step 1. The results are shown in Table 1, where the sum of the coefficients related to recovered time constants near each of the original time constants is given along with the sum of the coefficients of recovered time constants not near any of the original time constants.

These numerical results are comparable to those in [10] in the accuracy of parameter recovery.

TABLE 1

Original Term		Recovered Term	
coefficient	time constant	coefficient	time constant
0.6	0.1	0.47104...	0.10043...
		0.12850...	0.098566...
		Total 0.59955...	
0.3	0.01	Total 0.29980...	0.0099970...
0.1	0.001	0.071507...	0.0010002...
		0.028408...	0.00099877...
		Total 0.099916...	
		0.0000049429...	63.999...
		0.000038098...	0.43078...
		0.000081093...	0.036176...
		0.00051111...	0.035195...
		0.000089554...	0.0015239...
		Total 0.00072480...	

The residual sum of squares is $\varphi = \|\bar{E} - \gamma\|^2 = 4.3987 \times 10^{-5}$.

TABLE 2

Coefficient $(a_{20}, \ldots, a_1)^T$	Cumulative Sum $(\ldots, a_1 + \ldots + a_{20}, \ldots)^T$	Time Constant $(\alpha_{20}, \ldots, \alpha_1)^T$	Comparison term $(\ldots, \int_0^{\alpha_i} e^{-\alpha}\, d\alpha = 1 - e^{-\alpha_i}, \ldots)^T$
0.00014348...	0.00014348...	$\alpha_{20} = 0.000062625...$	0.000026231...
0.00016511...	0.00030859...	$\alpha_{19} = 0.00021750...$	0.00021747...
0.00018732...	0.00049591...	$\alpha_{18} = 0.00040736...$	0.00040727...
0.00054122...	0.0010371...	$\alpha_{17} = 0.00072805...$	0.00072778...
0.00038451...	0.0014216...	$\alpha_{16} = 0.0012662...$	0.0012653...
0.0013188...	0.0027404...	$\alpha_{15} = 0.0019696...$	0.0019676...
0.0024410...	0.0051814...	$\alpha_{14} = 0.0039520...$	0.0039442...
0.00077901...	0.0059604...	$\alpha_{13} = 0.0050879...$	0.0050749...
0.0033551...	0.0093155...	$\alpha_{12} = 0.0079146...$	0.0078833...
0.0045757...	0.013891...	$\alpha_{11} = 0.011017...$	0.010956...
0.0071382...	0.021029...	$\alpha_{10} = 0.017920...$	0.017760...
0.014811...	0.035840...	$\alpha_9 = 0.027290...$	0.026920...
0.031608...	0.067448...	$\alpha_8 = 0.051293...$	0.049999...
0.075988...	0.14343...	$\alpha_7 = 0.10536...$	0.099999...
0.11649...	0.25992...	$\alpha_6 = 0.22314...$	0.19999...
0.20203...	0.46195...	$\alpha_5 = 0.43078...$	0.34999...
0.27628...	0.73723...	$\alpha_4 = 0.91629...$	0.59999...
0.19675...	0.93398...	$\alpha_3 = 1.8971...$	0.84999...
0.036293...	0.97027...	$\alpha_2 = 2.9957...$	0.94999...
0.028701...	0.99898...	$\alpha_1 = 64.000...$	0.99999...

The residual sum of squares is $\varphi = \|\bar{E} - \gamma\|^2 = 2.3985 \times 10^{-23}$.

We note that with the computational methods given above the value of $|P_E(\alpha_i)|$ which is theoretically equal to zero was generally 10^{-22} and never larger than 10^{-19} for $i = 1, \ldots, r$ in each iteration.

As a second example with $S = [0, 64]$ and the same $t_0, \ldots, t_{19}$ data was generated by the model

$$T_2(t) = \int_0^\infty e^{-\alpha} e^{-\alpha t} d\alpha = \frac{1}{1 + t}.$$

With $\gamma_n := T_2(t_n)$ at full significance and $w_n := 1/T_2(t_n)$ for $0 \leqslant n \leqslant 19$. Here the program terminated after 38 iterations in step 2.b(i) with $r = 20$. The recovered time constants and coefficients are given in Table 2 along with the cumulative sums $a_i + \cdots + a_{20}$ and the term $\int_0^{\alpha_i} e^{-\alpha} d\alpha$ for a rough comparison for $i = 1, \ldots, 20$.

This example gives a clear demonstration of the power of the algorithm.

7. Approximation by Powers. The matrix method of divided difference calculation is also useful for approximation by sums of powers of the form $E = \sum_{z=1}^r a_i u_i^t$ for $t = $ integer $t_0, \ldots, t_N$. It follows from Proposition 5 that the top row of the t_nth power of

$$\begin{bmatrix} u_1 & 1 & & & \\ & \cdot & \cdot & & \\ & & \cdot & \cdot & \\ & & & \cdot & 1 \\ & & & & u_r \end{bmatrix}$$

gives the divided differences of the t_nth power of u with respect to $u_1, \ldots, u_r$. Considerable savings can result from the use of powers of this matrix if there are large gaps in the series $t_0, \ldots, t_N$. (If $t_n = n$, $0 \leqslant n \leqslant N$, the method in [10] is preferred.)

Department of Mathematics
University of California, San Diego
La Jolla, California 92093

1. D. BRAESS, "Approximation mit Exponentialsummen," *Computing*, v. 2, 1967, pp. 309–321.

2. D. BRAESS, "Chebyshev approximation by exponentials on finite sets," *Math. Comp.*, v. 27, 1973, pp. 327–331.

3. D. G. CANTOR & J. W. EVANS, "On approximation by positive sums of powers," *SIAM J. Appl. Math.*, v. 18, 1970, pp. 380–388.

4. S. D. CONTE & C. DE BOOR, *Elementary Numerical Analysis: An Algorithmic Approach*, McGraw-Hill, New York, 1972.

5. C. B. DUNHAM, "Nonlinear mean-square approximation on finite sets," *SIAM J. Numer. Anal.*, v. 12, 1975, pp. 105–110.

6. C. LAWSON & R. HANSON, *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliffs, N. J., 1974.

7. C. B. MOLER & C. F. VAN LOAN, "Nineteen dubious ways to compute the exponential of a matrix," *SIAM Rev.*, v. 20, 1978, pp. 801–836.

8. G. OPITZ, "Steigungsmatrizen," *Z. Angew. Math. Mech.*, v. 44, 1964, pp. T52–T54.

9. R. C. WARD, "Numerical computation of the matrix exponential with accuracy estimate," *SIAM J. Numer. Anal.*, v. 14, 1977, pp. 600–610.

10. W. J. WISCOMBE & J. W. EVANS, "Exponential sum fitting of radioactive transmission functions," *J. Computational Phys.*, v. 24, 1977, pp. 416–444.