71-23,783

KAMMLER, David William, 1940-
EXISTENCE, CHARACTERIZATION, AND COMPUTATION
OF BEST APPROXIMATIONS BY SUMS OF EXPONENTIALS.

The University of Michigan, Ph.D., 1971
Mathematics

University Microfilms, A XEROX Company, Ann Arbor, Michigan

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED

EXISTENCE, CHARACTERIZATION, AND COMPUTATION OF BEST APPROXIMATIONS BY SUMS OF EXPONENTIALS

by

David William Kammler

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Mathematics)
in The University of Michigan
1971

Doctoral Committee:

Associate Professor Cleve B. Moler, Co-Chairman
Associate Professor Jack L. Goldberg, Co-Chairman
Associate Professor Carl R. de Boor, Purdue University
Professor Donald A. Calahan
Assistant Professor James S. Milne
Professor Joseph L. Ullman

ABSTRACT

EXISTENCE, CHARACTERIZATION, AND COMPUTATION OF BEST APPROXIMATIONS BY SUMS OF EXPONENTIALS

by

David William Kammler

Chairman: Cleve B. Moler

Let $V_n(S)$ denote the set of all solutions of all n-th order linear homogeneous differential equations with constant coefficients for which the roots of the corresponding auxiliary polynomial all lie in $S \subseteq \mathbf{C}$. Given $1 \leq p \leq \infty$ and $f \in L_p[0,1]$, we show that there is a best $\| \|_p$ - approximation to f from $V_n(S)$ provided S is closed, thus extending the work of Rice, Hobby, and de Boor who establish this result when $S = \mathbf{R}$.

An element of $V_n(\mathbf{C})$ may be specified by giving the n initial conditions and the n coefficients of an n-th order differential operator which annihilates it. The tangential manifold corresponding to this parametric form does not suffer a loss of dimension when the roots of the auxiliary polynomial coalesce as is the case when one uses the roots of the auxiliary polynomial as parameters.

In general, a best $\| \|_p$ - approximation need not be unique and there may be troublesome local best $\| \|_p$ - approximations as we illustrate by means of various examples. In

particular, neither $V_n(\mathbb{R})$ nor $V_n(\mathbb{C})$ can be parametrized in such a manner as to fall within Rice's theory of varisolvent functions as was formerly believed.

Fundamental first order necessary (but not sufficient) conditions which a best $\| \|_p$ - approximation must satisfy are presented, and sufficient conditions are given. Using the basic necessary conditions, we derive generalizations of the previously known necessary conditions of Hobby and Rice, and of Braess which were formulated in $V_n(\mathbb{R})$. Conditions for a best $\| \|_\infty$ - approximation to f from $V_n(\mathbb{R})$ to be expressable as a simple sum of exponentials with positive coefficients are derived independently of the Decartes sign rule used by Braess. The use and limitations of these theorems are illustrated by means of carefully chosen examples.

When $f \in C^n[0,1]$, the distance of $f^{(n)}$ from the linear space spanned by f, f', ..., $f^{(n-1)}$ is used to bound the error $\|f-y\|_p$ in a best $\| \|_p$ - approximation to f from $V_n(\mathbb{C})$, and from this error bound we devise two simple procedures for constructing remarkably good initial estimates of f from $V_n(\mathbb{C})$. For the construction of best $\| \|_\infty$ - approximations, we also present an iterative procedure which has been found to be quite effective in the early stages of the iterative process when the parameters are not particularly close to their optimum values. Numerical results are given for several examples to illustrate these procedures.

ACKNOWLEDGEMENTS

First, I would like to express my gratitude for the many stimulating discussions, the helpful suggestions, and the patient manner in which Professors Cleve B. Moler, Carl R. de Boor, and Jack L. Goldberg have ably assisted in the development of this work. The many hours of time which they have invested in my efforts have been deeply appreciated.

Second, I would like to express my appreciation to Dr. Sidney D. Nolte for first stimulating my interest in this project (some five years ago) by challenging me to find a best approximation to $f(t)=1/t$ by a sum of two exponentials.

Third, I would like to acknowledge my debt to the University of Michigan Computing Center for providing me with the computer time used in the numerical computations.

Fourth, I would like to acknowledge my debt to the National Science Foundation for the Graduate Fellowship and to the University of Michigan for the University Fellowship by which I have been financially supported during the past three years.

Fifth and finally, I would like to express my heartfelt gratitude to my parents, Mr. and Mrs. Ruben W. J. Kammler, for instilling within me as a child a love of learning, and to my wife, Ruth, for the continuing encouragement which she has provided during these years of graduate study and for carefully typing the final version of this thesis.

TABLE OF CONTENTS

	Page
ABSTRACT	ii
ACKNOWLEDGEMENTS	iv
Chapter	
I. INTRODUCTION	1
II. EXISTENCE OF A BEST APPROXIMATION	9
III. PARAMETRIZATION OF EXPONENTIAL SUMS	37
IV. CHARACTERIZATION OF A BEST APPROXIMATION, $p < \infty$	52
V. CHARACTERIZATION OF A BEST APPROXIMATION, $p = \infty$	70
VI. CONSTRUCTION OF GOOD AND BEST APPROXIMATIONS	103
REFERENCES	141

I. INTRODUCTION

Let $S \subseteq \mathbf{C}$ where $\mathbf{C}$ is the set of complex numbers. We define $V_n(S)$, $n=1,2,\ldots$ to be the set of all complex valued functions y defined on the real line $\mathbf{R}$ which satisfy some n-th order homogeneous differential equation of the form

$$(1) \quad \{(D-\lambda_1)(D-\lambda_2)\ldots.(D-\lambda_n)\}\, y(t) \equiv 0, \quad -\infty < t < \infty$$

where $D=d/dt$ is the differential operator and where $\lambda_1, \lambda_2, \ldots \lambda_n \in S$. We define $V_o(S)$ to be the set whose only element is the zero function and we set

$$V_\infty(S) = \bigcup_{n=1}^{\infty} V_n(S)$$

We shall also define $V_n^r(S)$, $n=0,1,\ldots,\infty$, to be the set of real valued functions $y \in V_n(S)$.

The space $L_p[0,1]$ with the associated norm, $\| \|_p$, will be defined in the usual manner with the understanding that the elements of $L_p[0,1]$ may, in general, be complex valued. We may regard each y from $V_\infty(\mathbf{C})$ as being an element of each of the spaces $L_p[0,1]$, $1 \le p \le \infty$. Analogously, each y from $V_\infty^r(\mathbf{C})$ may be regarded as an element of the space $L_p^r[0,1]$ which consists of the real valued elements from $L_p[0,1]$, $1 \le p \le \infty$.

Our problem is then the following. Given $S \subseteq \mathbf{C}$, $1 \le p \le \infty$, and $f \in L_p[0,1]$ we would like to find a best

$\| \ \|_p$ - approximation to f from $V_n(S)$, i.e. we would like to find $y_o \in V_n(S)$ such that

$$(2) \quad \|f - y_o\|_p = \inf \{ \|f - y\|_p : y \in V_n(S) \}$$

We shall prove that this problem always has a solution provided that S is closed, thus extending the work of Hobby and Rice [12,p.96] where existence of a best approximation is proved for the case where S is compact, and of de Boor [2,p.171] who gives a complete existence proof for the case where $S=\mathbb{R}$.

Unfortunately, the solution to the approximation problem (2) need not be unique for any value of p, and indeed there may be numerous local best $\| \ \|_p$ - approximations which greatly complicate the problem. In particular, we shall find that $V_n^r(\mathbb{C})$ cannot be parametrized in such a fashion as to fall within Rice's theory of varisolvent approximating functions as was formerly thought to be the case, cf. Rice [21, p.42]. (It has already been pointed out by Hobby and Rice [12, p.104] and by Braess [3, p.314] that $V_n^r(\mathbb{R})$ does not give rise to a varisolvent approximating function.)

We shall present the basic first order necessary conditions which any best $\| \ \|_p$ - approximation to f from $V_n(S)$ must satisfy and we shall use these necessary conditions to derive and extend the formerly known necessary conditions of Hobby and Rice [12, p.98] and Braess [3,p.313].

Sufficient conditions will also be given, and the use and limitations of the theorems will be illustrated by means of several carefully chosen examples.

Finally, we shall suggest two relatively simple procedures for actually constructing a reasonably good $\| \|_p$ - approximation to f from the set $V_n(S)$, and we shall present an iterative algorithm which may often be used to construct best $\| \|_\infty$ - approximations as we illustrate with a few numerical results.

There are a number of basic terms and notational conventions which we shall introduce at this point and use throughout the remainder of this work. First of all, if y satisfies (1) but does not satisfy any such differential equation of lower order we shall say that y is an exponential sum with order n. The n (not necessarily distinct) complex numbers $\lambda_1, \lambda_2, \ldots, \lambda_n$ will be called the essential exponential parameters of y, and we shall refer to the set

$$(3) \quad \Lambda[y] \equiv \bigcup_{i=1}^{n} \{\lambda_i\} \qquad (\text{with } \Lambda[0] = \emptyset)$$

as the spectrum of y. For example, every polynomial of degree n is an exponential sum of order n+1 with the spectrum $\{0\}$, and thus our approximation problem includes the problem of $\| \|_p$ - approximation of f by polynomials of degree at most n in the special case where $S = \{0\}$.

When y is a solution of (1) and $\lambda_1, \lambda_2, \ldots, \lambda_n$ are all distinct, then we may find $a_1, a_2, \ldots, a_n \in \mathbf{C}$ such that

$$(4)\quad y(t) = \sum_{i=1}^{n} a_i \exp(\lambda_i t)$$

so that y actually looks like an "exponential sum". On the other hand, if λ_i occurs with multiplicity n_i, $i=1,2, \ldots, \ell$ (with $\lambda_1,\lambda_2, \ldots, \lambda_\ell$ being distinct and $n_1+n_2+\ldots+n_\ell=n$) then y may be expressed in the more general form

$$(5)\quad y(t) = \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} a_{ij}\, t^{j-1} \exp(\lambda_i t)$$

where again we have n linear parameters $a_{ij} \in \mathbf{C}$. The familiar form (5) for writing the solutions of (1) is quite convenient when $\lambda_1,\lambda_2, \ldots, \lambda_n$ are held fixed, but it is not at all convenient when these exponential parameters are allowed to vary and thus to coalesce or to separate. Consequently we shall introduce a somewhat different parametrization which will prove to be quite useful in the later development.

Given $y \in V_n(\mathbf{C})$ we may always find $\mathbf{b}=(b_1,b_2, \ldots, b_n)$ and $\mathbf{c}=(c_1,c_2, \ldots, c_n)$ from $\mathbf{C}^n$ such that

$$(6)\quad \{D^n + c_1 D^{n-1} + \ldots + c_{n-1} D + c_n\}\, y = 0$$

$$(7)\quad D^{i-1}\, y(0) = b_i, \quad i=1,2, \ldots, n$$

The initial conditions, $\mathbf{b}$, are always uniquely determined by y, i.e.

$$\mathbf{b} = (y(0), y'(0), \ldots, y^{(n-1)}(0))$$

but the differential equation (and therefore $\mathbf{c}$) is uniquely determined if and only if y has full order n. We shall let $Y_n(\mathbf{b}, \mathbf{c}, -)$ denote the solution of (6) - (7) noting that

$$V_n(\mathbf{C}) = \{Y_n(\mathbf{b},\mathbf{c},-): \mathbf{b}, \mathbf{c} \in \mathbf{C}^n\}$$

so that Y_n provides a 2n (complex) dimensional parametrization of $V_n(\mathbf{C})$. Moreover, since

$$V_n^r(\mathbf{C}) = \{Y_n(\mathbf{b},\mathbf{c},-): \mathbf{b}, \mathbf{c} \in \mathbf{R}^n\}$$

we see that Y_n also provides a 2n (real) dimensional parametrization of $V_n^r(\mathbf{C})$.

Corresponding to a given $\mathbf{c} \in \mathbf{C}^n$ is the associated spectral set $\Lambda_n(\mathbf{c})$ (to be distinguished from (3)) which is defined to be the smallest set containing all of the roots of the auxiliary polynomial

$$(8) \quad P_n(\mathbf{c},\lambda) = \lambda^n + c_1\lambda^{n-1} + \ldots + c_{n-1}\lambda + c_n$$

of the differential equation (6). It follows that

$$\Lambda[Y_n(\mathbf{b},\mathbf{c},-)] \subseteq \Lambda_n(\mathbf{c}), \quad \mathbf{b}, \mathbf{c} \in \mathbf{C}^n$$

but the reverse inclusion does not hold in general since $Y_n(\mathbf{b},\mathbf{c},-)$ may fail to have full order n. We shall also define

$$(9) \quad L_n(\mathbf{c}) = \{Y_n(\mathbf{b},\mathbf{c},-): \mathbf{b} \in \mathbf{C}^n\}$$

so that $L_n(\mathbf{c})$ is the set of solutions of the differential equation (6). $L_n^r(\mathbf{c})$ will denote real valued solutions of (6).

Finally, we shall often have occasion to parametrize $\mathbf{c}$ in terms of the roots of (8). Consequently, given $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n)$ from $\mathbf{C}^n$ we shall define $\boldsymbol{\sigma}_n(\boldsymbol{\lambda})$ such that

$$(10)\quad P_n(\boldsymbol{\sigma}_n(\boldsymbol{\lambda}), \lambda) = (\lambda-\lambda_1)(\lambda-\lambda_2)\ \ldots\ (\lambda-\lambda_n)$$

i.e. so that the components of $\boldsymbol{\sigma}_n(\boldsymbol{\lambda})$ are the elementary symmetric functions of the components of $\boldsymbol{\lambda}$. In terms of this notation, our approximation problem requires that we find $\boldsymbol{\lambda}_o \in S^n$ and $\mathbf{b}_o \in \mathbf{C}^n$ such that

$$(11)\quad \|f-Y_n(\mathbf{b}_o, \boldsymbol{\sigma}_n(\boldsymbol{\lambda}_o), -)\|_p = \inf\{\|f-Y_n(\mathbf{b}, \boldsymbol{\sigma}_n(\boldsymbol{\lambda}), -\|_p : \boldsymbol{\lambda}\in S^n, \mathbf{b}\in\mathbf{C}^n\}$$

This then completes the basic terminology which is repeatedly used henceforth.

In concluding this introduction, we might briefly mention three primary reasons why one would be interested in the theory of approximation with exponential sums. First of all, many physical and biological processes may be very well modeled by an exponential sum of the form (4) with n being relatively small, i.e. n=1,2,3, cf. [5,p.272] and [6]. For example, f(t) might represent the radiation count at time t of a sample with n types of radioactive material with a_i, λ_i being measures of the amount and half life, respectively, of the i-th species. In this case the

parameters a_i, λ_i (and sometimes even n) are to be determined from the experimentally determined data, and as is well known (cf. [14, p.272-280] and [23]) slight inaccuracies in the data may result in gross variations in the computed parameters a_i, λ_i, i.e. the problem is poorly conditioned. (It is of some historical interest to note that what appears to be the first study of approximation with exponential sums was concerned with the modeling of experimentally determined dilation versus temperature curves for common gases [17]. Today, Prony's method still survives in a few textbooks of numerical analysis, e.g. [11, p.378-382].)

A far more important reason for a study of this type lies in the potential application to the generation of a variety of special functions for use in analog and other special purpose computers, cf.[13]. In the same way that polynomials and rational functions may be naturally generated on a digital computer, exponential sums, i.e. solutions of initial value problems of the form (6) - (7), may be naturally generated on an analog computer. Thus we anticipate the eventual compilation and common use of tables of coefficients **b**, **c** of exponential sums corresponding to good and best $\| \ \|_\infty$ - approximations to commonly occurring functions in the same way that similar tables for polynomial and rational approximations, e.g.[10], are commonly found at present wherever digital computers are used.

Finally, we might observe that the problem is one of genuine mathematical interest since there are very few other nonlinear approximation problems which may be analyzed more readily. Indeed, in one sense the problem could be said to be semi-linear because $Y_n(\mathbf{b}, \mathbf{c}, -)$ depends linearly on half of the parameters and satisfies a simple differential equation which depends linearly on the remaining half of the parameters.

II. EXISTENCE OF A BEST APPROXIMATION

As a first step in the analysis of our problem we shall seek conditions which insure that a solution actually exists. It is immediately seen that a necessary condition for every $f \in L_p[0,1]$ to have a best $\| \ \|_p$ - approximation from $V_n(S)$ is that S be closed; the main objective of this section is to prove that this condition is also sufficient. In so doing it is useful to first establish existence under the much stronger hypothesis that S is compact, and then to remove the requirement that S be bounded.

When S is compact, existence can be established by means of Taylor series expansions (cf. [2,p.159-163], [12,p.92-96].) We shall find it more convenient, however, to make use of the following well known special properties of the approximating function Y_n.

Theorem 1: The approximating function $Y_n(\mathbf{b},\mathbf{c},t)$ is an entire analytic function of the 2n+1 arguments $b_1,b_2,\ldots,b_n$, $c_1,c_2,\ldots, c_n$, t which depends linearly on $\mathbf{b}$, and which for fixed $\mathbf{b},\mathbf{c}$ vanishes identically on some non-degenerate interval $t_0 \le t \le t_1$ if and only if $\mathbf{b}=\mathbf{0}$.

Proof: cf. [7,p.21,75-76]. ∎

We shall define the seminorm $\| \ \|_{p,\alpha}$, $0 \le \alpha \le 1/3$, on $L_p[0,1]$, $1 \le p \le \infty$, such that

(12) $\|f\|_{p,\alpha} = \|\chi_\alpha f\|_p$

where

$$\chi_\alpha(t) = \begin{cases} 1 & \text{for } |t-1/2| \le 1-\alpha \\ 0 & \text{otherwise} \end{cases}$$

is the characteristic function of $[\alpha,1-\alpha]$. When S is bounded, the seminorms $\| \|_{p,\alpha}$, $1 \le p \le \infty$, $0 \le \alpha \le 1/3$ are actually uniformly equivalent norms on $V_n(S)$, and the differential operator, D, is bounded on $V_n(S)$ as we see from the following lemma.

<u>Lemma 1:</u> Let $S \subset \mathbf{C}$ be bounded. Then there exists a constant M (depending only on S and n) such that

(13) $\|D^i y\|_q \le M \cdot \|y\|_{p,\alpha} \qquad i=0,1,\ldots, n$

holds true whenever $y \in V_n(S)$, $1 \le p,q \le \infty$, and $0 \le \alpha \le 1/3$.

<u>Proof:</u> In view of the monotonicity of $\| \|_{p,\alpha}$ with respect to both p and α it is sufficient to establish (13) when $q=\infty$, $p=1$, and $\alpha=1/3$. By enlarging S, if necessary, we may also assume that S is compact. We define

$$F_i(\mathbf{b},\boldsymbol{\lambda}) = \|D^i Y_n(\mathbf{b},\sigma_n(\boldsymbol{\lambda}),-)\|_\infty / \|Y_n(\mathbf{b},\sigma_n(\boldsymbol{\lambda}),-)\|_{1,1/3}$$

for $i=0,1, \ldots, n$, for $\boldsymbol{\lambda} \in S^n$, and for $\mathbf{b} \in \partial\mathbf{B}^n$ where $\partial\mathbf{B}^n$ is the surface of the unit ball in $\mathbf{C}^n$ (noting from Theorem 1 that the denominator can never vanish.) We may then choose

$$M = \max \{M_0, M_1, \ldots, M_n\}$$

where M_i is maximum value of the continuous function F_i as $\mathbf{b}$, $\boldsymbol{\lambda}$ range over the compact sets $\partial \mathbf{B}^n$ and S^n respectively, i=0,1,..., n. With this choice of M we see that (13) holds for all $y \in V_n(S)$ so that the proof is complete. □

The conclusion of the lemma is no longer valid when S is not bounded. For example, if

$$y_\nu(t) = \nu^{1/2} \exp(-\nu t), \quad \nu=1,2,\ldots$$

we find that the ratio

$$\|D^i y_\nu\|_\infty / \|y_\nu\|_1 = \nu^{i+1}/[1-\exp(-\nu)], \quad \nu=1,2,\ldots$$

is unbounded as ν becomes infinite even when i=0.

Theorem 2: Let S be a compact subset of $\mathbf{C}$, and let p, $1 \leq p \leq \infty$, be chosen. Then each closed $\| \ \|_p$ - bounded subset of $V_n(S)$ is $\| \ \|_p$ - compact, i.e. given any $\| \ \|_p$ - bounded sequence $\{y_\nu\}$ from $V_n(S)$ there exists a subsequence which $\| \ \|_p$ - converges to some $y \in V_n(S)$.

Proof: In view of Lemma 1 it is sufficient to establish the theorem for the special case $p=\infty$. Let $\{y_\nu\}$ be a $\| \ \|_\infty$ - bounded sequence from $V_n(S)$, and let $\{\mathbf{b}_\nu\}$, $\{\boldsymbol{\lambda}_\nu\}$ be chosen from $\mathbf{C}^n$, S^n, respectively, so that

$$y_\nu = Y_n(\mathbf{b}_\nu, \sigma_n(\boldsymbol{\lambda}_\nu), -), \quad \nu=1,2,\ldots$$

Since S^n is compact, we may (by taking a subsequence, if necessary) assume that $\{\boldsymbol{\lambda}_\nu\}$ converges to some $\boldsymbol{\lambda} \in S^n$. In view of Lemma 1, the $\|\ \|_\infty$ - boundedness of $\{y_\nu\}$ implies the $\|\ \|_\infty$ - boundedness of $\{D^i y_\nu\}$, i=1,2,..., n, and therefore $\{\mathbf{b}_\nu\}$ is bounded. We may therefore (by again passing to a subsequence if necessary) assume that $\{\mathbf{b}_\nu\}$ converges to some $\mathbf{b} \in \mathbf{C}^n$. Finally, from the continuous dependence of Y_n on its parameters (Theorem 1) we see that

$$y = Y_n(\mathbf{b}, \sigma_n(\boldsymbol{\lambda}), -)$$

is the $\|\ \|_\infty$ - limit of $\{y_\nu\}$. □

<u>Corollary 1:</u> Let S be a compact subset of $\mathbf{C}$, and let p, $1 \le p \le \infty$, be chosen. Then each $f \in L_p[0,1]$ has a best $\|\ \|_p$ - approximation from $V_n(S)$ and from $V_n^r(S)$.

<u>Proof:</u> Let $f \in L_p[0,1]$ be chosen, and let $\{y_\nu\}$ be a $\|\ \|_p$ - minimizing sequence for f from $V_n(S)$, i.e.

$$(14)\quad \lim \|f-y_\nu\|_p = \inf \{\|f-y\|_p : y \in V_n(S)\}$$

Then $\{y_\nu\}$ is $\|\ \|_p$ - bounded, and, in view of Theorem 2, we may (by passing to a subsequence, if necessary) assume that $\{y_\nu\}$ has a $\|\ \|_p$- limit $y \in V_n(S)$. Hence we find

$$\|f-y\|_p \le \lim [\|f-y_\nu\|_p + \|y-y_\nu\|_p] = \lim \|f-y_\nu\|$$

so that y is a best $\|\ \|_p$ - approximation to f from $V_n(S)$. Analogously, f has a best $\|\ \|_p$ - approximation from $V_n^r(S)$. □

Corollary 2: Let $\mathbf{c} \in \mathbf{C}^n$, and let p, $1 \leq p \leq \infty$, be chosen. Then each $f \in L_p[0,1]$ has a best $\| \|_p$ - approximation from the linear spaces $L_n(\mathbf{c})$, $L_n^r(\mathbf{c})$.

Proof: Given $f \in L_p[0,1]$ we may choose a $\| \|_p$ - minimizing sequence $\{y_\nu\}$ for f from $L_n(\mathbf{c})$. Then $\{y_\nu\}$ is $\| \|_p$ - bounded, and since $L_n(\mathbf{c}) \subseteq V_n(\Lambda_n(\mathbf{c}))$ we see from the theorem that we may (by passing to a subsequence, if necessary) assume that $\{y_\nu\}$ has a $\| \|_p$ - limit, y. Since $L_n(\mathbf{c})$ is $\| \|_p$ - closed, y is then a best $\| \|_p$ - approximation to f from $L_n(\mathbf{c})$. Analogously, f has a best $\| \|_p$ - approximation from $L_n^r(\mathbf{c})$. □

Corollary 3: Let $f \in L_p[0,1]$ for some p, $1 \leq p \leq \infty$. Then there exists a best $\| \|_p$ - approximation to f from the set $V_{n+1}(\{0\}) = L_{n+1}(\mathbf{0})$ of all polynomials of degree at most n and from the set $V_{n+1}^r(\{0\}) = L_{n+1}^r(\mathbf{0})$ of all real valued polynomials of degree at most n.

Proof: This follows from either of the above corollaries. □

We would now like to strenghten Corollary 1 of Theorem 2 and obtain an existence theorem when S is closed but not compact. Unfortunately, Theorem 2 cannot be extended to apply to $V_n(S)$ when S is not bounded. For example, if

$$u_\nu(t) = \exp[-\nu t] + \exp[-\nu(1-t)] \qquad 0 \leq t \leq 1,\ \nu=1,2,\ldots \tag{15}$$

$$w_\nu(t) = \sin \nu t \qquad 0 \leq t \leq 1,\ \nu=1,2,\ldots$$

then $\{u_\nu\}$, $\{w_\nu\}$ are both $\| \|_\infty$ - bounded sequences from $V_2(\mathbf{C})$, but no subsequence of either $\{u_\nu\}$ or $\{w_\nu\}$ is $\| \|_\infty$ - convergent. Indeed, these two sequences illustrate two fundamentally different ways in which noncompactness can be troublesome.

In order to make these intuitive ideas more precise, we shall introduce definitions for three distinct types of sequences which may be extracted from $V_n(\mathbf{C})$. We shall say that $\{y_\nu\}$ is a U- sequence, a V- sequence, or a W- sequence according as the corresponding sequence of spectral sets $\Lambda[y_\nu]$, $\nu=1,2,\ldots$ satisfies the respective conditions

$$\text{(16)}\quad \liminf\{|\mathrm{Re}\lambda| : \lambda \in \Lambda[y_\nu]\} = +\infty ,$$

$$\text{(17)}\quad \sup\{|\lambda| : \lambda \in \bigcup_{\nu=1}^{\infty} \Lambda[y_\nu]\} < +\infty ,$$

or both of

$$\text{(18)}\quad \begin{aligned} &\liminf\{|\mathrm{Im}\lambda| : \lambda \in \Lambda[y_\nu]\} = +\infty \\ &\sup\{|\mathrm{Re}\lambda| : \lambda \in \bigcup_{\nu=1}^{\infty} \Lambda[y_\nu]\} < +\infty \end{aligned}$$

Examples of U, W- sequences are provided by $\{u_\nu\}$, $\{w_\nu\}$, respectively, as given by (15). For large ν, the function u_ν is essentially nonzero only near the end points of $[0,1]$ (as suggested by the U - shape), and this behavior is typical of all $\| \|_p$ - bounded U- sequences which (as we shall see)

converge uniformly to zero on compact subsets of (0,1). On the other hand, for large ν the function w_ν oscillates throughout the whole interval [0,1] (as suggested by the W- shape), and this behavior is typical of all W- sequences. Indeed, (as we shall see) the seminorms $\| \|_{p,\alpha}$, $0 \leq \alpha \leq 1/3$, are uniformly equivalent on the terms of a W- sequence. Finally, we note that any sequence $\{v_\nu\}$ extracted from $V_n(S)$ is a V- sequence provided that S is bounded. From Theorem 2 we see that from any $\| \|_p$ - bounded V- sequence, we can extract a $\| \|_p$ - convergent subsequence which, in view of Lemma 1, is also $\| \|_\infty$ - convergent, $1 \leq p \leq \infty$.

Although a general sequence $\{y_\nu\}$ from $V_n(\mathbf{C})$ need not be either a U, V, or W - sequence, we may always extract from $\{y_\nu\}$ a subsequence (which we shall continue to call $\{y_\nu\}$ that may be decomposed in the form

$$(19)\quad y_\nu = u_\nu + v_\nu + w_\nu, \qquad \nu = 1,2,\ldots$$

where $\{u_\nu\}$, $\{v_\nu\}$, and $\{w_\nu\}$ are U, V, and W - sequences from $V_{n_1}(\mathbf{C})$, $V_{n_2}(\mathbf{C})$, and $V_{n_3}(\mathbf{C})$, respectively, with $n_1 + n_2 + n_3 \leq n$. Indeed, if $y_\nu = 0$ for infinitely many indices ν, than the zero sequence (which may be regarded as either a U, V, or W - sequence) may be used for $\{u_\nu\}$, $\{v_\nu\}$, and $\{w_\nu\}$ thus providing the desired decomposition of the zero subsequence of $\{y_\nu\}$. Otherwise, (by passing to a

subsequence, if necessary) we may assume that $\Lambda[y_\nu]$ has exactly m, $1 \leq m \leq n$, distinct elements for each ν. We may then write

$$\Lambda[y_\nu] = \{\lambda_{1\nu}, \lambda_{2\nu}, \ldots, \lambda_{m\nu}\}, \quad \nu=1,2,\ldots$$

and (by again passing to a subsequence, if necessary) assume that for each $\ell = 1,2,\ldots, m$ we have either

$$(20) \quad \lim |\mathrm{Re}\ \lambda_{\ell\nu}| = +\infty ,$$

$$(21) \quad \overline{\lim}\ |\lambda_{\ell\nu}| < +\infty ,$$

or both of

$$(22) \quad \begin{array}{l} \lim\ |\mathrm{Im}\ \lambda_{\ell\nu}| = +\infty \\ \overline{\lim}\ |\mathrm{Re}\ \lambda_{\ell\nu}| < +\infty \end{array}$$

Upon comparing (16), (17), and (18) with (20), (21), and (22) respectively, we see that the desired decomposition follows at once from the well known form (5) for expressing a solution of the differential equation (1).

With these concepts in mind, we may now establish the following lemma which precisely characterizes the important properties of U, V, W - sequences which are essential for the proof of the desired existence theorem.

Lemma 2: Let $\{u_\nu\}$, $\{v_\nu\}$, and $\{w_\nu\}$ denote U, V, and W - sequences from $V_n(\mathbf{C})$, n=0,1,... . Then

(i) : If $\{u_\nu + v_\nu + w_\nu\}$ is a $\| \ \|_p$ - bounded sequence from $V_n(\mathbf{C})$ for some p, $1 \leq p \leq \infty$, then the component sequences $\{u_\nu\}$, $\{v_\nu\}$, and $\{w_\nu\}$ are all $\| \ \|_p$ - bounded.

(ii): If $\{v_\nu + w_\nu\}$ is a sequence from $V_n(\mathbf{C})$, then there exists a constant $M > 0$ (depending only on the sequence) such that

(23) $\|v_\nu + w_\nu\|_p \leq M \|v_\nu + w_\nu\|_{p,\alpha}$, $\nu=1,2,...$

holds true for all p, α with $1 \leq p \leq \infty$ and $0 \leq \alpha \leq 1/3$. (Again $\| \ \|_{p,\alpha}$ is the seminorm defined by (12).)

(iii): If $\{u_\nu\}$ is a $\| \ \|_p$ - bounded sequence for some p, $1 \leq p \leq \infty$, then

(24) $\lim \|u_\nu\|_{\infty,\alpha} = 0$

for every α with $0 < \alpha \leq 1/3$.

(iv): If $\{u_\nu + w_\nu\}$ is a sequence from $V_n(\mathbf{C})$ and $f \in L_p[0,1]$ for some p, $1 \leq p \leq \infty$, then

(25) $\underline{\lim} \|f + u_\nu + w_\nu\|_p \geq \|f\|_p$

Proof: We shall give a proof by induction on n. The lemma certainly holds when n=0 since the zero sequence is the only sequence which may be extracted from $V_0(\mathbf{C})$.

We must now show that if the lemma holds in $V_{n-1}(\mathbf{C})$ then it also holds in $V_n(\mathbf{C})$, n=1,2,... . In so doing it

is important to note that in proving each of (i) - (iv) we lose no generality in passing to subsequences whenever it is convenient to do so. For example, if (i) fails for some sequence $\{u_\nu + v_\nu + w_\nu\}$ from $V_n(\mathbf{C})$, then (by passing to a subsequence, if necessary) we may assume that at least one of $\lim \|u_\nu\|_p$, $\lim \|v_\nu\|_p$, and $\lim \|w_\nu\|_p$ exists and is $+\infty$ so that (i) also fails for every subsequence. Thus it is sufficient to show that (i) holds for some subsequence of every given sequence $\{u_\nu + v_\nu + w_\nu\}$. Analogous considerations hold in each of (ii)-(iv). We shall now divide the proof of the induction step into four sections corresponding to the statements (i) - (iv).

Section (i): $\{u_\nu + v_\nu + w_\nu\}$ be a sequence from $V_n(\mathbf{C})$ with $\| \|_p$ - bound B. We must show that some subsequence satisfies (i). Now if $u_\nu + v_\nu + w_\nu = u_\nu$ for infinitely many indices ν, then we can extract from $\{u_\nu + v_\nu + w_\nu\}$ a U- subsequence which clearly satisfies (i) (the corresponding terms v_ν, w_ν being zero for all but finitely many ν.) Analogous considerations hold if $u_\nu + v_\nu + w_\nu = v_\nu$ or $u_\nu + v_\nu + w_\nu = w_\nu$ for infinitely many ν. We may therefore restrict our attention to the remaining two cases where either $u_\nu=0$, $v_\nu\neq0$, $w_\nu\neq0$ for infinitely many ν or where $u_\nu\neq0$, $v_\nu + w_\nu\neq0$ for infinitely many ν. Indeed, since neither of these situations can arise in $V_1(\mathbf{C})$ the validity of (i) in $V_1(\mathbf{C})$ is thereby established, and we shall assume in the remainder of this section that $n \geq 2$.

Suppose then that $\{v_\nu + w_\nu\}$ is a sequence from $V_n(\mathbf{C})$ with $\| \|_p$ - bound B and that $v_\nu \neq 0$, $w_\nu \neq 0$ for each ν. Since the order of $v_\nu + w_\nu$ is at most n, and since v_ν, w_ν both have order at least 1, it follows that v_ν, $w_\nu \in V_{n-1}(\mathbf{C})$ for all but a finite number of indices ν, and we may therefore assume that $\{v_\nu\}$, $\{w_\nu\}$ are V, W - sequences from $V_{n-1}(\mathbf{C})$. We shall define the auxiliary sequences

$$v_\nu^* = v_\nu / \|v_\nu\|_p , \quad \nu=1,2,\ldots$$

$$w_\nu^* = w_\nu / \|v_\nu\|_p , \quad \nu=1,2,\ldots$$

so that $\{v_\nu^*\}$ is a $\| \|_p$ - normalized V- sequence. We may therefore assume (by passing to a subsequence, if necessary) that $\{v_\nu^*\}$ $\| \|_p$ - converges to some $v^* \in V_{n-1}(\mathbf{C})$ with $\|v^*\|_p = 1$, and that the real sequence $\{\|v_\nu\|_p\}$ has some (possibly infinite) limit. Hence, by using the inductive hypothesis that (iv) holds in $V_{n-1}(\mathbf{C})$ we find

$$\begin{aligned} B &\geq \overline{\lim} \|v_\nu + w_\nu\|_p \\ &\geq \{\lim \|v_\nu\|_p\} \cdot \{\underline{\lim} \|v_\nu^* + w_\nu^*\|_p\} \\ &= \{\lim \|v_\nu\|_p\} \cdot \{\underline{\lim} \|v^* + w_\nu^*\|_p\} \\ &\geq \{\lim \|v_\nu\|_p\} \cdot \|v^*\|_p \\ &= \lim \|v_\nu\|_p \end{aligned}$$

so that $\{v_\nu\}$ and therefore $\{w_\nu\}$ must be $\| \|_p$ - bounded. Thus (i) holds in this case.

Suppose next that $\{u_\nu + v_\nu + w_\nu\}$ is a sequence from $V_n(\mathbf{C})$ with $\| \ \|_p$ - bound B, and that $u_\nu \neq 0$, $v_\nu + w_\nu \neq 0$ for each ν. Then we may assume that $\{u_\nu\}$ and $\{v_\nu + w_\nu\}$ are sequences from $V_{n-1}(\mathbf{C})$, and since (from the induction hypothesis) (ii) holds in $V_{n-1}(\mathbf{C})$ there is a constant $M > 0$ such that (23) holds with α, $0 < \alpha \leq 1/3$, being held fixed. Moreover, since (iii) holds in $V_{n-1}(\mathbf{C})$ we see from (24) that

$$\|u_\nu\|_{p,\alpha} \leq \|u_\nu\|_p \ / \ (2M)$$

for all sufficiently large ν. Using this result together with (23) and the triangle inequality we find

$$\begin{aligned} \|u_\nu\|_p &\leq \|u_\nu + v_\nu + w_\nu\|_p + \|v_\nu + w_\nu\|_p \\ &\leq B + M\|v_\nu + w_\nu\|_{p,\alpha} \\ &\leq B + M \cdot \{\|u_\nu + v_\nu + w_\nu\|_{p,\alpha} + \|u_\nu\|_{p,\alpha}\} \\ &\leq B \cdot (1+M) + \|u_\nu\|_p \ / \ 2 \end{aligned}$$

for all sufficiently large ν. Hence $\{u_\nu\}$ is $\| \ \|_p$ - bounded and therefore $\{v_\nu + w_\nu\}$ is also $\| \ \|_p$ - bounded. But since (i) holds in $V_{n-1}(\mathbf{C})$, this implies that $\{v_\nu\}$, $\{w_\nu\}$ are individually $\| \ \|_p$ - bounded and therefore (i) holds in this case also. Thus the induction step for (i) is complete. ∎

Section (ii): Let $\{v_\nu + w_\nu\}$ be a sequence from $V_n(\mathbf{C})$. In seeking to establish (ii) we may assume, with no loss of generality, that $\|v_\nu + w_\nu\|_\infty = 1$ for each ν. Next, if $\{\beta_\nu\}$ is any real number sequence and

$$\theta_\nu(t) = \exp(i\beta_\nu t), \quad 0 \le t \le 1 , \quad \nu=1,2,...$$

(where $i^2 = -1$) then $|\theta_\nu| \equiv 1$ so that the sequence $\{\theta_\nu[v_\nu+w_\nu]\}$ satisfies (ii) if and only if $\{v_\nu + w_\nu\}$ does. Hence, after properly selecting the phase parameters $\{\beta_\nu\}$ and factoring out the appropriate factors $\{\theta_\nu\}$ we may also assume that $\underline{\lim}\ \|v_\nu\|_\infty > 0$ and that $\{w_\nu\}$ is a W- sequence from $V_{n-1}(\mathbf{C})$. Moreover, since we have already shown that (i) holds in $V_n(\mathbf{C})$ the assumption that $\|v_\nu + w_\nu\|_\infty = 1$, $\nu=1,2,...$ implies that both $\{v_\nu\}$ and $\{w_\nu\}$ are $\|\ \|_\infty$ - bounded. Thus (by passing to a subsequence, if necessary) we may further assume that $\{v_\nu\}$ $\|\ \|_\infty$ - converges to some $v \in V_n(\mathbf{C})$ with $\|v\|_\infty > 0$.

Under these restrictions, we see that when n=1 we have $w_\nu=0$ for each ν so that (ii) holds by virtue of Lemma 1. On the other hand, when $n \ge 2$ we may use the induction hypothesis that (iv) holds in $V_{n-1}(\mathbf{C})$ together with the Hölder inequality to see that

$$\underline{\lim}_{\nu} \inf_{p,\alpha} \|v_\nu+w_\nu\|_{p,\alpha}/\|v_\nu+w_\nu\|_p \geq \underline{\lim}_{\nu} \inf_{\alpha} \|v_\nu+w_\nu\|_{1,\alpha}/\|v_\nu+w_\nu\|_\infty$$

$$= \underline{\lim}_{\nu} \|v_\nu+w_\nu\|_{1,1/3}$$

$$= \underline{\lim}_{\nu} \|v+w_\nu\|_{1,1/3}$$

$$\geq \|v\|_{1,1/3} > 0$$

(where the infinum is taken over the sets where $1 \leq p \leq \infty$ and $0 \leq \alpha \leq 1/3$). From this inequality we immediately infer the existence of a constant $M > 0$ such that (23) holds whenever $1 \leq p \leq \infty$ and $0 \leq \alpha \leq 1/3$. Thus the induction step is complete for (ii). ∎

Section (iii): First of all, since any $\| \|_p$ - bounded U-sequence $\{u_\nu\}$ is also $\| \|_1$ - bounded, it is sufficient to prove (iii) under the hypothesis that p=1, and this being the case we may further assume that $\|u_\nu\|_1 = 1$ for each ν. We now select the sequences $\{\gamma_\nu\}$, $\{\beta_\nu\}$ from $\mathbf{C}$ such that $\beta_\nu \in \Lambda[u_\nu]$, $\nu=1,2,\ldots$ (so that $\lim |\mathrm{Re}\ \beta_\nu| = +\infty$) and such that if

$$\theta_\nu(t) = \gamma_\nu \cdot \exp(\beta_\nu t), \quad 0 \leq t \leq 1, \quad \nu=1,2,\ldots$$

then u_ν may be decomposed in either of the forms

$$u_\nu = \theta_\nu \cdot [u_\nu^* + v_\nu^* + w_\nu^*]$$

$$= \theta_\nu \cdot v_\nu^* + u_\nu^{**}$$

with $\{v_\nu^*\}$ being a $\| \|_1$ - normalized V-sequence, with $\{u_\nu^*\}$ and $\{u_\nu^{**}\} = \{\theta_\nu \cdot [u_\nu^* + w_\nu^*]\}$ being U-sequences, with $\{w_\nu^*\}$ being a W-sequence, and with the order of u_ν being the sum of the orders of either u_ν^*, v_ν^*, w_ν^* or of v_ν^*, u_ν^{**} for each ν. Since $\{v_\nu^*\}$ is $\| \|_1$ - normalized, we may assume (by passing to a subsequence, if necessary) that $\{v_\nu^*\}$ is $\| \|_1$ - convergent to some $v^* \in V_n(\mathbf{C})$ with $\|v^*\|_1 = 1$. Finally, we may further assume that Re $\beta_\nu > 0$ for each ν (since in showing that (iii) holds we may always replace $u_\nu(t)$ by $u_\nu(1-t)$ for any ν) so that lim Re $\beta_\nu = +\infty$.

With these restrictions in mind, we now let α be chosen with $0 < \alpha \leq 1/9$. Using the inductive hypothesis that (iv) holds in $V_{n-1}(\mathbf{C})$ we find

$$
\begin{aligned}
1 &= \lim \|u_\nu\|_1 \\
&\geq \overline{\lim} \int_{1-\alpha}^{1} |u_\nu| \, dt \\
&\geq \overline{\lim} \, \{|\theta_\nu(1-\alpha)| \cdot \int_{1-\alpha}^{1} |u_\nu^* + v_\nu^* + w_\nu^*| \, dt \\
&\geq \{\overline{\lim} \, |\theta_\nu(1-\alpha)|\} \cdot \{\underline{\lim} \int_{1-\alpha}^{1} |u_\nu^* + v_\nu^* + w_\nu^*| \, dt \\
&\geq \{\overline{\lim} \, |\theta_\nu(1-\alpha)|\} \cdot \int_{1-\alpha}^{1} |v^*| \, dt
\end{aligned}
$$

and since $\|v^*\|_1 = 1$ we infer the existence of a constant $B > 0$ such that

$$|\theta_\nu(1-\alpha)| \leq B , \quad \nu=1,2,...$$

This being the case we have

$$\overline{\lim} \; \|\theta_\nu v_\nu^*\|_{\infty, 2\alpha} \leq \|v^*\|_\infty \cdot B \cdot \overline{\lim} \; |\exp(-\beta_\nu \alpha)| = 0$$

and since $\{u_\nu\}$ is $\| \; \|_1$ - bounded this implies that $\{u_\nu^{**}\}$ is $\| \; \|_{1,2\alpha}$ - bounded. But since $\{u_\nu^{**}\}$ is a U-sequence from $V_{n-1}(\mathbf{C})$ we may use the induction hypothesis that (iii) holds in $V_{n-1}(\mathbf{C})$ to see that

$$\overline{\lim} \; \|u_\nu\|_{\infty, 3\alpha} \leq \overline{\lim} \; \|\theta_\nu \cdot v_\nu^*\|_{\infty, 3\alpha} + \overline{\lim} \; \|u_\nu^{**}\|_{\infty, 3\alpha} = 0$$

Thus the induction step is complete for (iii). ∎

<u>Section (iv):</u> We shall first establish (iv) for the special case where $u_\nu=0$ for each ν. We may assume (by passing to a subsequence, if necessary) that $\{\|f+w_\nu\|_p\}$ has a finite limit so that $\{w_\nu\}$ is $\| \; \|_p$ - bounded. We now select the real number sequence $\{\rho_\nu\}$ in such a manner that if

$$\theta_{1\nu}(t) = \exp(i \; \beta_{1\nu} \; t), \quad 0 \leq t \leq 1, \quad \nu=1,2,...$$

then w_ν may be decomposed in either of the forms

$$\begin{aligned} w_\nu &= \theta_{1\nu}[v_{1\nu} + w_{1\nu}] \\ &= \theta_{1\nu} \; v_{1\nu} + w_{1\nu}^* \end{aligned}$$

where $\{v_{1\nu}\}$ is a V-sequence and $\{w_{1\nu}\}$, $\{w^*_{1\nu}\}$ are both W-sequences from $V_{n-1}(\mathbf{C})$. Since $|\theta_{1\nu}| \equiv 1$ for each ν and since we have already shown that (i) holds in $V_n(\mathbf{C})$, the sequences $\{v_{1\nu}\}$, $\{w_{1\nu}\}$, and $\{w^*_{1\nu}\}$ are all $\| \|_p$ - bounded. We may therefore assume (by passing to a subsequence, if necessary) that $\{v_{1\nu}\}$ is $\| \|_\infty$ - convergent to some v_1 from $V_n(\mathbf{C})$. Next, by using the triangle inequality (and the fact $|\theta_{1\nu}| \equiv 1$ for each ν) we have

$$|\|f+\theta_{1\nu}v_1+w^*_{1\nu}\|_p-\|f+w_\nu\|_p| \leq \|v_1-v_{1\nu}\|_p$$

so that

$$\lim \|f+\theta_{1\nu}v_1+w_{1\nu}\|_p = \lim \|f+w_\nu\|_p$$

and thus in showing that (iv) holds we lose no generality in assuming that $v_{1\nu} = v_1$ for each ν.

In a similar manner we decompose the $\| \|_p$ - bounded sequence $\{w^*_{1\nu}\}$ from $V_{n-1}(\mathbf{C})$, and by proceeding in this manner we see that it is sufficient to establish (iv) in the special case where

$$(26)\quad w_\nu = \theta_{1\nu}v_1+\theta_{2\nu}v_2+ \ldots +\theta_{k\nu}v_k\ , \qquad \nu=1,2,\ldots$$

where v_ℓ is a fixed element of $V_{n-\ell+1}(\mathbf{C})$, where

$$(27)\quad \theta_{\ell\nu}(t) = \exp(i\ \beta_{\ell\nu}t),\quad 0 \leq t \leq 1\ , \qquad \nu=1,2,\ldots$$

and where $\{\beta_{\ell\nu}\}$ is a real number sequence with

(28) $\lim |\beta_{\ell\nu}| = +\infty$

for each $\ell=1,2,\ldots,$ k with $k \leq n$.

Finally, since the class of finite linear combinations of characteristic functions of subintervals of [0,1] is dense in $L_p[0,1]$, it is sufficient to establish (iv) whenever $f, v_1, v_2, \ldots, v_k$ are all constant multiples of the characteristic function of a single subinterval of [0,1], or more simply, whenever $f, v_1, v_2, \ldots, v_k$ are all complex constants (with this last step involving a simple change of variables, if necessary.) But in this special situation we may use (26)-(28) together with Hölder's inequality to obtain

$$\lim \|f+w_\nu\|_p \geq \underline{\lim} \int_0^1 |f + \sum_{\ell=1}^{k} v_\ell \, \theta_{\ell\nu}| dt$$

$$\geq \underline{\lim} \left| \int_0^1 [f + \sum_{\ell=1}^{k} v_\ell \, \theta_{\ell\nu}] \, dt \right|$$

$$= \underline{\lim} \, |f + \sum_{\ell=1}^{k} v_\ell [\theta_{\ell\nu}(1)-1]/\beta_{\ell\nu}|$$

$$= |f| = \|f\|_p$$

Thus we have shown that (iv) holds in $V_n(\mathbf{C})$ in the special case where $u_\nu = 0$ for each ν .

We shall now remove this restriction. Indeed, in the general case we may again assume (by passing to a subsequence,

if necessary) that $\{\|f + u_\nu + w_\nu\|_p\}$ has a finite limit. Since we have shown that (i) holds in $V_n(\mathbf{C})$ this implies that $\{u_\nu\}$ and $\{w_\nu\}$ are both $\| \ \|_p$ - bounded. Finally, by using the fact that(iii) and the above special case of (iv) all hold in $V_n(\mathbf{C})$ we have

$$\lim \|f + u_\nu + w_\nu\|_p \geq \underline{\lim} \|f + u_\nu + w_\nu\|_{p,\alpha}$$

$$\geq \underline{\lim} \{\|f + w_\nu\|_{p,\alpha} - \|u_\nu\|_{p,\alpha}\}$$

$$= \underline{\lim} \|f + w_\nu\|_{p,\alpha}$$

$$\geq \|f\|_{p,\alpha}$$

for all α, $0 < \alpha \leq 1/3$, and from the arbitrariness of α we conclude that(25) must hold. This then finishes the induction for (iv) and so completes the proof of the lemma. □

It would be desirable to simplify the rather long and tedious proof of Lemma 2, and since (i), (iv) are the only sections needed for the desired existence theorem it would be nice to prove these results independently. Unfortunately, the above induction proceeds cyclically (i.e., to prove that (i) holds in $V_n(\mathbf{C})$ we make use of the fact that (i) - (iv) all hold in $V_{n-1}(\mathbf{C})$, and then use the fact that (i) holds in $V_n(\mathbf{C})$ in showing that (iii) - (iv) also hold in $V_n(\mathbf{C})$, etc.), and we have been unsuccessful in proving (i), (iv) by any other argument.

Using Lemma 2 it is now a simple matter to prove the following basic existence theorem.

Theorem 3: Let $S \subseteq \mathbf{C}$, $1 \leq p \leq \infty$, and $n=1,2,\ldots$ Then every $f \in L_p[0,1]$ has a best $\| \|_p$ - approximation from $V_n(S)$ if and only if S is closed.

Proof: Let S be closed, let $f \in L_p[0,1]$, and let $\{y_\nu\}$ be chosen from $V_n(S)$ in such a manner that

$$\lim \|f - y_\nu\|_p = \inf \{\|f-y\|_p : y \in V_n(S)\}$$

We may assume (by passing to a subsequence, if necessary) that $\{y_\nu\}$ may be decomposed in such a manner that

$$y_\nu = u_\nu + v_\nu + w_\nu , \quad \nu=1,2,\ldots$$

where $\{u_\nu\}$, $\{v_\nu\}$, $\{w_\nu\}$ are U, V, W - sequences, respectively, from $V_n(S)$. Since $\{y_\nu\}$ is $\| \|_p$ - bounded, we may infer from Lemma 2 - (i) that the component sequences $\{u_\nu\}$, $\{v_\nu\}$, $\{w_\nu\}$ are all $\| \|_p$ - bounded, and in view of Theorem 2 we may further assume (by again passing to a subsequence, if necessary) that $\{v_\nu\}$ $\| \|_p$ - converges to some $v \in V_n(S)$. This being the case we may use Lemma 2 - (iv) to obtain.

$$\lim \|f - y_\nu\|_p = \lim \|f - v - u_\nu - w_\nu\|_p \geq \lim \|f - v\|_p$$

so that v is a best $\| \|_p$ - approximation to f from $V_n(S)$. Thus the closure of S is a sufficient condition for every $f \in L_p[0,1]$ to have a best $\| \|_p$ - approximation from $V_n(S)$.

On the other hand, the necessity of this condition is readily apparent (e.g., if $\lambda \notin S$ is a limit point of S, then the function

$$f(t) = \exp(\lambda t), \quad 0 \leq t \leq 1$$

has no best $\| \|_p$ - approximation from $V_n(S)$) so that the proof is complete . □

Corollary 1: Let $S \subseteq \mathbf{C}$ be closed, let $1 \leq p \leq \infty$, and let n=1,2,.... Then every $f \in L_p[0,1]$ has a best $\| \|_p$ - approximation from $V_n^r(S)$.

Proof: Replace $V_n(S)$ by $V_n^r(S)$ in the proof of the theorem. □

For the special case where S=$\mathbb{R}$ the result of Corollary 1 was first stated by Rice [17, p.153] (for $p=\infty$) and later (and somewhat indirectly) by Hobby and Rice [12, p.97-98], with the first complete proof being given by de Boor [2, p.172]

Corollary 2: Let $S \subseteq \mathbf{C}$ be closed, let $1 \leq p \leq \infty$, and let n=1,2,.... Then $V_{n-1}(S)$, $V_{n-1}^r(S)$ are $\| \|_p$ - closed subsets of $V_n(S)$, $V_n^r(S)$, respectively.

Proof: Let $y \in V_n(S)$ have order n. In view of the theorem, we may find a best $\| \|_p$ - approximation, y_o, to y from $V_{n-1}(S)$, and since the order of y_o is at most n-1 we must have $\|y - y_o\|_p > 0$. Thus we infer the existence of an open $\| \|_p$ - ball in $V_n(S)$ which contains y but does not

meet $V_{n-1}(S)$. Hence $V_{n-1}(S)$ is a $\| \ \|_p$ - closed subset of $V_n(S)$. Analogously, $V^r_{n-1}(S)$ is a $\| \ \|_p$ - closed subset of $V^r_n(S)$. □

For practical applications one may at times wish to restrict S to be a closed disc

(29) $\mathbf{D}_\alpha = \{\lambda \in \mathbf{C} : |\lambda| \leq \alpha\}, \ \alpha > 0$

a left half plane

(30) $\mathbf{L}_\alpha = \{\lambda \in \mathbf{C} : \text{Re } \lambda \leq -\alpha\}, \quad \alpha > 0$

etc., and this being the case the best $\| \ \|_p$ - approximation to f from $V_n(S)$ may actually be a very poor representation of f when n is small. Nevertheless, we have a fortunate trade off here, since even when S is highly restricted the error, $\|f - y_n\|_p$, in a best $\| \ \|_p$ - approximation to f from $V_n(S)$ can be made arbitrarily small by choosing n to be sufficiently large, as we see from the following result.

Theorem 4: Let S be a nonvoid subset of $\mathbf{C}$ and let $1 \leq p \leq \infty$. Then $V_\infty(S)$, $V^r_\infty(S)$ are dense in $L_p[0,1]$, $L^r_p[0,1]$, respectively.

Proof: It is sufficient to prove the theorem under the assumption that $S = \{\lambda\}$, and this being the case we define

$$e_{\pm}(t) = \exp(\pm \lambda t), \quad 0 \leq t \leq 1$$

Let $f \in L_p[0,1]$ be selected and let $\varepsilon > 0$ be chosen. Since the class of polynomials is dense in $L_p[0,1]$, we may find a polynomial, q_1, with $\|q_1\|_p > 0$ such that

$$\|f - q_1\|_p < \varepsilon/2$$

Having chosen q_1 we may find a second polynomial, q_2, such that

$$\|e_- - q_2\|_p < \varepsilon/(2\|q_1 e_+\|_\infty)$$

This being the case, we have

$$\begin{aligned}\|f - q_1 q_2 e_+\|_p &\le \|f - q_1\|_p + \|q_1 - q_1 q_2 e_+\|_p \\ &\le \|f - q_1\|_p + \|q_1 e_+\|_\infty \cdot \|q_2 - e_-\|_p \\ &< \varepsilon\end{aligned}$$

and since $q_1 q_2 e_+ \in V_\infty(\{\lambda\})$ the proof is complete. □

Let $y \in V_n(\mathbf{L}_\alpha)$, $\alpha > 0$, where again $\mathbf{L}_\alpha$ is the half plane defined by (30). Then the exponential parameters of y all have negative real parts so that

$$(31)\quad |y| = \sup \{ |y(t)| : t \ge 0 \}$$

is finite. (Indeed (31) defines a norm on $V_n(\mathbf{L}_\alpha)$.) Using Lemma 2 we may now show that there is a minimum rate (which depends only on n and α) at which $y(t)$ can decay to zero as t becomes infinite.

Theorem 5: Let $\mathbf{L}_\alpha$, $\alpha > 0$, and $\| \ \|$ be defined by (30), (31), and for $t \geq 0$, $n=1,2,\ldots$ let

$$(32)\quad g_n(t) = \sup \{ |y(t)| : y \in V_n(\mathbf{L}_1),\ \|y\| = 1 \}$$

Then $g_n(t)$ is a continuous nonincreasing function of t with

$$(33)\quad 0 < g_1(t) \leq g_2(t) \leq g_3(t) \leq \ldots \leq 1\ ,\ t \geq 0$$

$$(34)\quad \lim_{t \to +\infty} g_n(t) = 0\ ,\quad n=1,2,\ldots$$

$$(35)\quad \lim_{n \to +\infty} g_n(t) = 1\ ,\quad t \geq 0$$

Moreover, for any $y \in V_n(\mathbf{L}_\alpha)$, $\alpha > 0$, we have

$$(36)\quad |y(t)| \leq \|y\| \cdot g_n(\alpha t),\quad t \geq 0.$$

Proof: Since $V_1(\mathbf{L}_1) \subset V_2(\mathbf{L}_1) \subset \ldots$ (and since the simple exponential $e(t) = \exp(-t)$ is contained in $V_1(\mathbf{L}_1)$) we see at once from the definition (32) that (33) must hold. Next, for any $\alpha > 0$ we may use (32) to obtain

$$g_n(\alpha t) = \sup \{ y(\alpha t) : y \in V_n(\mathbf{L}_1),\ \|y\| = 1 \}$$

$$= \sup \{ y(t) : y \in V_n(\mathbf{L}_\alpha),\ \|y\| = 1 \}$$

from which (36) follows at once. In addition, this identity also shows that $g_n(\alpha t) \leq g_n(t)$ when $\alpha \geq 1$ (since $V_n(\mathbf{L}_\alpha) \subset V_n(\mathbf{L}_1)$ when $\alpha \geq 1$) so that $g_n(t)$ is a nonincreasing

function of t. Next, since each $y \in V_n(\mathbf{L}_1)$ is continuous, we see from (32) that the nonincreasing function g_n has no jump discontinuities so that g_n must be continuous. Thus (34), (35) are all that remain to be shown.

We shall use Lemma 2 to establish (34). Let $\{y_\nu\}$ be chosen from $V_n(\mathbf{L}_1)$ in such a manner that $|y_\nu| = 1$ and

(37) $|y_\nu(\nu)| \geq g_n(\nu)/2, \qquad \nu=1,2,...$

Then if $\{u_\nu\}$ is defined such that

(38) $u_\nu(t) = y_\nu(2\nu t), \quad 0 \leq t \leq 1, \quad \nu=1,2,...$

we have $u_\nu \in V_n(\mathbf{L}_{2\nu})$ so that $\{u_\nu\}$ is a $\| \|_\infty$ - bounded U - sequence. This being the case we may use Lemma 2-(iii) together with (37)-(38) to conclude that

$$\lim_{\nu \to +\infty} g_n(\nu) \leq 2 \lim u_\nu(1/2) = 0$$

so that (34) holds.

Finally, we must establish (35). Let $t > 0$ be chosen, if possible, such that $g_n(t) < 1$ for each n=1,2,... For such t we may use (36) to conclude that

$$|y(s)| \leq |y| \cdot g_n(st) < |y| , \quad s \geq 1$$

and thus that

(39) $\|y\|_\infty = |y|$

for every $y \in V_n(\mathbb{L}_t)$, n=1,2,... In particular, if we let y_n be a best $\|\ \|_\infty$ - approximation to the unit function $f \equiv 1$ from the set $V_n(\mathbb{L}_t)$, n=1,2,..., then we may use (36), (39) together with Theorem 4 (and the fact that $\lim \|y_n\|_\infty = \|f\|_\infty = 1$) to obtain

$$
\begin{aligned}
0 &= \lim \|f - y_n\|_\infty \\
&\geq \lim |f(1) - y_n(1)| \\
&\geq \lim \{1 - |y_n(1)|\} \\
&\geq \lim \{1 - \|y\|_\infty \cdot g_n(t)\} \\
&= 1 - \lim g_n(t)
\end{aligned}
$$

which completes the proof of (34). □

The rate at which $g_n(t)$ decays to zero as $t \to +\infty$ is not provided by the above analysis since the critical argument which establishes (34) is based upon the nonconstructive proof of Lemma 2-(iii). A simple lower bound for g_n may be obtained in the following manner.

For each n=0,1,... let the real polynomial

$$q_n(t) = a_{n0} + a_{n1}\ t + \ldots + a_{nn}\ t^n$$

be selected in such a manner that the leading coefficient a_{nn} is maximized subject to the constraint that

$$\sup \{ |q_n(t)| \cdot \exp(-t) : t \geq 0 \} \leq 1$$

By standard arguments, we see that the polynomial q_n is uniquely defined and that for each n there exist n+1 points $0 = t_{n0} < t_{n1} < \ldots < t_{nn}$ such that

$$q_n(t_i) \exp(-t_{ni}) = (-1)^{n-i} , \quad i=0,1,\ldots, n$$

i.e., q_n is the Tchebycheff polynomial of degree n with respect to the interval $[0, +\infty)$ and the weight function $w(t) = \exp(-t)$. We then define

$$g_n(t) = \sup \{ |q_{n-1}(\alpha t)| \exp(-\alpha t) : \alpha \geq 1 \}$$

$$= \begin{cases} 1 \quad \text{for} \quad 0 \leq t \leq t_{nn} \\ \\ q_{n-1}(t) \exp(-t) \text{ otherwise} \end{cases} \qquad n=1,2,\ldots$$

noting that

$$g_n(t) \leq g_n(t), \qquad t \geq 0, \quad n=1,2,\ldots$$

(where g_n is the envelope function defined by (32).)

When n=1 we find

$$g_1(t) = g_1(t) = \exp(-t), \qquad t \geq 0$$

and we conjecture that $g_n = g_n$ for $n \geq 2$ as well. If this conjecture could be proved directly, then Lemma 2-(iii) would follow as a simple corollary and make possible an

alternative proof of the existence of a best $\| \ \|_p$ - approximation to $f \in L_p[0,1]$ from $V_n(S)$ when S is a closed subset of some strip $\{\lambda \in \mathbf{C} : |\mathrm{Im}\ \lambda| < B\}$ (in which case $V_n(S)$ contains no W-sequences.)

III. PARAMETRIZATION OF EXPONENTIAL SUMS

When one attempts to characterize a best $\| \ \|_p$ - approximation to a given f from $V_n(S)$ (i.e., to find various necessary and/or sufficient conditions which a best $\| \ \|_p$ - approximation must satisfy) or to actually compute a best $\| \ \|_p$ - approximation, it is important to have some means to parametrically represent each $y \in V_n(\mathbf{C})$. For example, y may be represented in the conventional form (5) with a_{ij}, λ_i, $j=1,2,\ldots, n_i$, $i=1,2,\ldots, \ell$ being the parameters (as is done by Rice [18,p.152]), y may be represented by (6), (7), i.e., $y = Y_n(\mathbf{b},\mathbf{c},-)$ for some choice of the parameters $b_1,b_2,\ldots, b_n$, $c_1,c_2,\ldots, c_n$, or y may be represented in the form $y = Y_n(\mathbf{b},\sigma_n(\boldsymbol{\lambda}),-)$ for some choice of the parameters $b_1, b_2,\ldots, b_n$, $\lambda_1, \lambda_2,\ldots, \lambda_n$. (Yet another parameterization is given by Braess [4,p.102] who writes y as a linear combination of the first n-1 divided differences of the exponential $\exp(\lambda t)$ with respect to the exponential parameters $\lambda_1,\lambda_2,\ldots, \lambda_n$ of y.) Indeed, we would ideally like to find some continuously differentiable homeomorphism from some familiar open subset of euclidean space onto the $\| \ \|_p$ - normed space $V_n(\mathbf{C})$. Unfortunately, no such parametric representation is known. In some respects, however, the parametric form $Y_n(\mathbf{b},\mathbf{c},-)$ comes close to being ideal as we shall show in this section.

Moreover, (as we shall later show by means of examples) a best $\| \|_p$ - approximation to f from $V_n(S)$ need not be unique, even when $S = \mathbb{R}$ or $S = \mathbb{C}$ (so that no constraints are imposed) and $p = \infty$ (when we might possibly expect to have uniqueness.) This being the case, we are forced, right from the start, to conern ourselves with "local" rather than "global" phenomena. Thus we are naturally interested in the tangential manifold of the parametric representation. In this section, we shall determine the tangential manifolds of the parametric forms $Y_n(\mathbf{b},\mathbf{c},-)$ and $Y_n(\mathbf{b},\sigma_n(\lambda),-)$, noting that the latter suffers a "loss of dimension" when the exponential parameters are allowed to coalesce. Finally we shall introduce the concept of the tangential cone which will be useful in the subsequent analysis.

We recall that for each $y \in V_n(\mathbb{C})$ we may find $\mathbf{b}, \mathbf{c} \in \mathbb{C}^n$ such that $y = Y_n(\mathbf{b},\mathbf{c},-)$ with $\mathbf{b}$ being uniquely determined. Unfortunately, $\mathbf{c}$ is uniquely determined only when y has full order n, and it is this nonuniqueness of $\mathbf{c}$ which lies behind the difficulties one encounters in attempting to describe a $\| \|_p$ - neighborhood of $y \in V_n(\mathbb{C})$ when the order of y is less than n. For example, if $\{u_\nu\}$, $\{w_\nu\}$ are given by (15), if

$$y(t) = \exp(t), \quad 0 \leq t \leq 1,$$

and if α, β are complex constants with $|\alpha|, |\beta| < \varepsilon/3$ then for every $\nu=1,2,\ldots$, $y + \alpha u_\nu + \beta w_\nu$ is contained in the open

$\| \|_p$ - ball of radius ε which is centered at y in $V_s(\mathbf{C})$. Such an unpleasant situation cannot occur when y has full order, however, as we see from the following theorem.

Theorem 6: For n=1,2,... let the set $\mathbf{F}_{2n} \subset \mathbf{C}^{2n}$ be defined by

(40) $\mathbf{F}_{2n} = \{(\mathbf{b},\mathbf{c}) : \mathbf{b},\mathbf{c} \in \mathbf{C}^n,\ Y_n(\mathbf{b},\mathbf{c},-) \in V_{n-1}(\mathbf{C})\}$

Then $\mathbf{F}_{2n}$ is a nowhere dense closed subset of $\mathbf{C}^{2n}$, and the mapping

(41) $\mathbf{C}^{2n} \setminus \mathbf{F}_{2n} \to V_n(\mathbf{C}) \setminus V_{n-1}(\mathbf{C})$ by $(\mathbf{b},\mathbf{c}) \mapsto Y_n(\mathbf{b},\mathbf{c},-)$

is a continuously differentiable homeomorphism provided that $\mathbf{C}^{2n}$ has the usual topology and $V_n(\mathbf{C})$ has the topology induced by any one of the $\| \|_p$ - norms, $1 \leq p \leq \infty$.

Proof: In view of Theorem 1 the mapping

(42) $\mathbf{C}^{2n} \to V_n(\mathbf{C})$ by $(\mathbf{b},\mathbf{c}) \mapsto Y_n(\mathbf{b},\mathbf{c},-)$

is continuously differentiable and therefore the same is true of the restriction (41). From Corollary 2 to Theorem 3 we know that $V_{n-1}(\mathbf{C})$ is a $\| \|_p$ - closed subset of $V_n(\mathbf{C})$, and therefore $\mathbf{F}_{2n}$, the inverse image of $V_{n-1}(\mathbf{C})$ under the continuous mapping (42), must be a closed subset of $\mathbf{C}^{2n}$. Next, the closed set $\mathbf{F}_{2n}$ can contain no open ball (since by slightly perturbing $\mathbf{b}$ one can insure that

$Y_n(\mathbf{b},\mathbf{c},-)$ has full order n) and therefore $\mathbf{F}_{2n}$ is nowhere dense in $\mathbf{C}^{2n}$. Moreover, from the previous discussion we see that the restriction (41) of (42) is a bijection, and thus all that remains to be shown is that the inverse mapping of (41) is continuous.

Accordingly, let the functions $Y_n(\mathbf{b},\mathbf{c},-)$ and $Y_n(\mathbf{b}_\nu,\mathbf{c}_\nu,-)$, $\nu=1,2,\ldots$ be chosen in such a manner that each has order n and that

$$(43)\quad \lim \|Y_n(\mathbf{b}_\nu,\mathbf{c}_\nu,-) - Y_n(\mathbf{b},\mathbf{c},-)\|_p = 0$$

We must show that this implies that $\{(\mathbf{b}_\nu,\mathbf{c}_\nu)\}$ converges to $(\mathbf{b},\mathbf{c})$ in $\mathbf{C}^{2n}$. In order to do this, it is sufficient to show that the sequence $\{(\mathbf{b}_\nu,\mathbf{c}_\nu)\}$ is bounded and that every convergent subsequence has the limit $(\mathbf{b},\mathbf{c})$.

As a first step, we shall show that $\{\mathbf{c}_\nu\}$ is bounded. Indeed, if this is not the case then the corresponding sequence of spectral sets, $\{\Lambda_n(\mathbf{c}_\nu)\}$, in unbounded, and in view of (43), Lemma 2-(i), and Theorem 2 we may (by passing to a subsequence, if necessary) obtain the decomposition

$$Y_n(\mathbf{b}_\nu,\mathbf{c}_\nu,-) = u_\nu + v_\nu + w_\nu\ ,\quad \nu=1,2,\ldots$$

where $\{v_\nu\}$ is a $\|\ \|_p$ - bounded sequence from $V_{n-1}(\mathbf{C})$ which $\|\ \|_p$ - converges to some $v \in V_{n-1}(\mathbf{C})$ and where $\{u_\nu\}$, $\{w_\nu\}$ are U, W - sequences, respectively, from $V_n(\mathbf{C})$. Together with (43) and Lemma 2-(iv) this gives

$$0 = \lim \|u_\nu+v_\nu+w_\nu - Y_n(\mathbf{b},\mathbf{c},-)\|_p$$

$$= \lim \|u_\nu+w_\nu+v - Y_n(\mathbf{b},\mathbf{c},-)\|_p$$

$$\geq \|v - Y_n(\mathbf{b},\mathbf{c},-)\|_p$$

which implies that $Y_n(\mathbf{b},\mathbf{c},-) \in V_{n-1}(\mathbf{C})$ contrary to hypothesis. Thus $\{\mathbf{c}_\nu\}$ is bounded.

Now since $\{\mathbf{c}_\nu\}$ is bounded, the corresponding sequence of spectral sets is bounded, and this together with (43) and Lemma 1 implies that $\{\mathbf{b}_\nu\}$ is also bounded. Hence some subsequence of $\{(\mathbf{b}_\nu,\mathbf{c}_\nu)\}$ is convergent. Let $(\mathbf{b}^*,\mathbf{c}^*)$ denote the limit of this subsequence (which we shall continue to call $\{(\mathbf{b}_\nu,\mathbf{c}_\nu)\}$.) Then by using (43) and the continuity of the mapping (42) we find

$$\|Y_n(\mathbf{b}^*,\mathbf{c}^*,-)-Y_n(\mathbf{b},\mathbf{c},-)\|_p = \lim \|Y_n(\mathbf{b}_\nu,\mathbf{c}_\nu,-)-Y_n(\mathbf{b},\mathbf{c},-)\|_p = 0$$

and since (41) is a bijection this implies that $(\mathbf{b}^*,\mathbf{c}^*) = (\mathbf{b},\mathbf{c})$. □

<u>Corollary 1</u>: For n=1,2,... let the set $\mathbf{F}^r_{2n} \subset \mathbf{R}^{2n}$ be defined by

$$(44)\quad \mathbf{F}^r_{2n} = \{(\mathbf{b},\mathbf{c}) : \mathbf{b},\mathbf{c} \in \mathbf{R}^n,\ Y_n(\mathbf{b},\mathbf{c},-) \in V^r_{n-1}(\mathbf{C})\}$$

Then $\mathbf{F}^r_{2n}$ is a nowhere dense closed subset of $\mathbf{R}^{2n}$ and the mapping

$$\mathbb{R}^{2n} \setminus \mathbb{F}_{2n}^{r} \to V_n^r(\mathbb{C}) \setminus V_{n-1}^r(\mathbb{C}) \text{ by } (\mathbf{b},\mathbf{c}) \to Y_n(\mathbf{b},\mathbf{c},-)$$

is a continuously differentiable homeomorphism provided $\mathbb{R}^{2n}$ has the usual topology and $V_n^r(\mathbb{C})$ has the topology induced by any one of the $\| \ \|_p$ - norms, $1 \leq p \leq \infty$.

Proof: After replacing $V_n(\mathbb{C})$ by $V_n^r(\mathbb{C})$ and complex numbers by real numbers we proceed exactly as in the proof of the theorem. □

In view of the above theorem, when $y = Y_n(\mathbf{b},\mathbf{c},-)$ has order n, then we can explore a neighborhood of y in $V_n(\mathbb{C})$ or $V_n^r(\mathbb{C})$ by allowing each of $\mathbf{b},\mathbf{c}$ to vary over a small open ball in $\mathbb{C}^n$ or $\mathbb{R}^n$, respectively, and we shall make use of this important fact later on in establishing sufficient conditions for y to be a local best $\| \ \|_p$ - approximation to f. When y does not have full order n, however, we cannot explore a neighborhood of y in $V_n(\mathbb{C})$ or $V_n^r(\mathbb{C})$ so simply, and indeed we do not have any means of reasonably describing such a neighborhood in terms of any presently known parametric form for y. (Some feeling for the problem can be obtained by trying to visualize the sets $\mathbb{F}_{2n}$, $\mathbb{F}_{2n}^r$, e.g., $\mathbb{F}_2^r = \{(0,c_1) : c_1 \in \mathbb{R}\}$ so that $\mathbb{F}_2^r$ consists of a single line which Y_1 maps to the zero function.)

For all $\mathbf{b}$, $\mathbf{c}$, $\mathbf{b}^*$, $\mathbf{c}^* \in \mathbb{C}^n$ we shall define

$$(45)\quad h_n(\mathbf{b},\mathbf{c},\mathbf{b}^*,\mathbf{c}^*,t) = \sum_{i=1}^{n}\left\{b_i^* \cdot \frac{\partial}{\partial b_i} + c_i^* \cdot \frac{\partial}{\partial c_i}\right\} Y_n(\mathbf{b},\mathbf{c},t)$$

and we shall refer to the set

$$(46)\quad H_n(\mathbf{b},\mathbf{c}) = \{h_n(\mathbf{b},\mathbf{c},\mathbf{b}^*,\mathbf{c}^*,-) : \mathbf{b}^*,\mathbf{c}^*, \in \mathbf{C}^n\}$$

as the tangential manifold of $y = Y_n(\mathbf{b},\mathbf{c},-)$ with respect to the parameters $\mathbf{b},\mathbf{c}$. (For $\mathbf{b},\mathbf{c} \in \mathbf{R}^n$ we shall also define $H_n^r(\mathbf{b},\mathbf{c})$ to be the set of real valued functions contained in $H_n(\mathbf{b},\mathbf{c})$.) Clearly $H_n(\mathbf{b},\mathbf{c})$ is a linear space with dimension at most 2n, and in view of the previous theorem, we expect this linear space to be of great importance in seeking to characterize and to compute best $\| \|_p$ - approximations to a given f. Consequently we have formulated the following lemma which specifies the dimension and supplies a useful alternative description of the linear space $H_n(\mathbf{b},\mathbf{c})$.

Lemma 3: Let $\mathbf{b},\mathbf{c} \in \mathbf{C}^n$ and assume that the exponential sum $Y_n(\mathbf{b},\mathbf{c},-)$ has order k. Let $\lambda_1,\lambda_2,\ldots,\lambda_k$ be the essential exponential parameters of $Y_n(\mathbf{b},\mathbf{c},-)$ so that

$$(47)\quad Q_n(\mathbf{b},\mathbf{c},D) = \begin{cases} 1 & \text{if } k = 0 \\ (D-\lambda_1)(D-\lambda_2)\ldots(D-\lambda_k), & \text{if } k \geq 1 \end{cases}$$

is the monic k-th order differential operator which annihilates $Y_n(\mathbf{b},\mathbf{c},-)$. Then the linear space $H_n(\mathbf{b},\mathbf{c})$

given by (46) has dimension n+k with

$$(48)\quad H_n(\mathbf{b},\mathbf{c}) = \{h : Q_n(\mathbf{b},\mathbf{c},D) \cdot P_n(\mathbf{c},D)h \equiv 0\}$$

(where $P_n(\mathbf{c},D)$ is again given by (8).)

Proof: From the definition of $Y_n(\mathbf{b},\mathbf{c},-)$ we know that

$$\begin{aligned} &P_n(\mathbf{c},D)\ Y_n(\mathbf{b},\mathbf{c},-) \equiv 0 \\ (49)\quad &D^{j-1}\ Y_n(\mathbf{b},\mathbf{c},0) = b_j\ , \quad j=1,2,\ldots,\ n \end{aligned}$$

Recalling from Theorem 1 that $Y_n(\mathbf{b},\mathbf{c},t)$ is analytic in all of the variables, we obtain

$$\begin{aligned} &P_n(\mathbf{c},D)\ \frac{\partial Y_n(\mathbf{b},\mathbf{c},-)}{\partial b_i} \equiv 0 \\ (50)\quad &D^{j-1}\ \frac{\partial Y_n(\mathbf{b},\mathbf{c},0)}{\partial b_i} = \delta_{ij}\ , \quad j=1,2,\ldots,\ n \end{aligned}$$

and

$$\begin{aligned} &P_n(\mathbf{c},D)\ \frac{\partial Y_n(\mathbf{b},\mathbf{c},-)}{\partial c_i} = -\ D^{n-i}\ Y_n(\mathbf{b},\mathbf{c},-) \\ (51)\quad &D^{j-1}\ \frac{\partial Y_n(\mathbf{b},\mathbf{c},0)}{\partial c_i} = 0\ , \quad j=1,2,\ldots,\ n \end{aligned}$$

by differentiating (49) with respect to b_i and c_i, respectively, i=1,2,..., n. Using (50) - (51) and the definition (45) we find that $h_n(\mathbf{b},\mathbf{c},\mathbf{b}^*,\mathbf{c}^*,-)$ satisfies the

initial value problem

$$(52)\quad P_n(\mathbf{c},D)\, h_n(\mathbf{b},\mathbf{c},\mathbf{b}^*,\mathbf{c}^*,-) = -\sum_{i=1}^{n} c_i^*\, D^{n-i}\, Y_n(\mathbf{b},\mathbf{c},-)$$

$$(53)\quad D^{j-1}\, h_n(\mathbf{b},\mathbf{c},\mathbf{b}^*,\mathbf{c}^*,0) = b_j^*\,, \quad j=1,2,\ldots,\ n$$

and hence vanishes identically if and only if the driving function on the right hand side of (52) and the initial conditions in (53) both vanish. But since $Y_n(\mathbf{b},\mathbf{c},-)$ has order k, the n functions $D^{j-1}\, Y_n(\mathbf{b},\mathbf{c},-)$, $j=1,2,\ldots,\ n$ span a linear space of dimension k, and therefore from (52)-(53) we conclude that $H_n(\mathbf{b},\mathbf{c})$ has dimension n+k. Finally, from (52) and the definition of $Q_n(\mathbf{b},\mathbf{c},D)$ we find

$$Q_n(\mathbf{b},\mathbf{c},D)\, P_n(\mathbf{c},D)\, h_n(\mathbf{b},\mathbf{c},\mathbf{b}^*,\mathbf{c}^*,-) \equiv 0$$

holds for all $\mathbf{b}^*$, $\mathbf{c}^* \in \mathbf{C}^n$ so that the n+k dimensional kernel of the differential operator $Q_n(\mathbf{b},\mathbf{c},D)\, P_n(\mathbf{c},D)$ contains the n+k dimensional linear space $H_n(\mathbf{b},\mathbf{c})$, i.e., (48) holds. □

When $y = Y_n(\mathbf{b},\mathbf{c},-)$ has full order n, then every sufficiently small $\| \ \|_p$ - neighborhood of y in $V_n(\mathbf{C})$ is homeomorphic to some neighborhood of $(\mathbf{b},\mathbf{c})$ in the 2n dimensional euclidean space $\mathbf{C}^{2n}$, and therefore we would like for the tangential manifold, $H_n(\mathbf{b},\mathbf{c})$, of y with respect to $(\mathbf{b},\mathbf{c})$ to have the full dimension 2n. Lemma 3 insures us

that this is indeed the case. If one chooses some other means to parametrize $V_n(\mathbf{C})$, however, this need no longer hold.

For example, since

$$V_n(\mathbf{C}) = \{Y_n(\mathbf{b},\sigma_n(\boldsymbol{\lambda}),-) : \mathbf{b},\boldsymbol{\lambda} \in \mathbf{C}^n\}$$

$$(54)\quad V_n^r(\mathbf{R}) = \{Y_n(\mathbf{b},\sigma_n(\boldsymbol{\lambda}),-) : \mathbf{b},\boldsymbol{\lambda} \in \mathbf{R}^n\}$$

we might choose to parametrize y in terms of $\mathbf{b}$, $\boldsymbol{\lambda}$. (The conventional representation (5) is essentially in this form.) This being the case, for all $\mathbf{b},\boldsymbol{\lambda},\mathbf{b}^*,\boldsymbol{\lambda}^* \in \mathbf{C}^n$ we define

$$(55)\quad g_n(\mathbf{b},\boldsymbol{\lambda},\mathbf{b}^*,\boldsymbol{\lambda}^*,t) = \sum_{i=1}^{n}\left\{b_i^* \cdot \frac{\partial}{\partial b_i} + \lambda_i^* \cdot \frac{\partial}{\partial \lambda_i}\right\}Y_n(\mathbf{b},\sigma_n(\boldsymbol{\lambda}),-)$$

and

$$(56)\quad G_n(\mathbf{b},\boldsymbol{\lambda}) = \{g_n(\mathbf{b},\boldsymbol{\lambda},\mathbf{b}^*,\boldsymbol{\lambda}^*,-) : \mathbf{b}^*,\boldsymbol{\lambda}^* \in \mathbf{C}^n\}$$

(with $G_n^r(\mathbf{b},\boldsymbol{\lambda})$ being defined to be the subset of real valued functions contained in $G_n(\mathbf{b},\boldsymbol{\lambda})$ when $\mathbf{b},\boldsymbol{\lambda} \in \mathbf{R}^n$.) The dimension and an alternative representation for the linear space $G_n(\mathbf{b},\boldsymbol{\lambda})$ are now given by the following lemma.

Lemma 4: Let $\mathbf{b}$, $\boldsymbol{\lambda} \in \mathbf{C}^n$ and assume that the exponential sum $y = Y_n(\mathbf{b}, \sigma_n(\boldsymbol{\lambda}),-)$ has order k, $0 \le k \le n$. Let $\lambda_1^o, \lambda_2^o, \ldots, \lambda_\ell^o$, $0 \le \ell \le k$, be the (necessarily distinct)

elements of $\Lambda[y]$ and set

$$(57)\quad Q_n^o(\mathbf{b},\boldsymbol{\lambda},D) = \begin{cases} 1 & \text{if} \quad \ell = 0 \\ (D-\lambda_1^o)(D-\lambda_2^o)\ \ldots\ (D-\lambda_\ell^o) & \text{if} \quad \ell \geq 1 \end{cases}$$

Then the linear space $G_n(\mathbf{b},\boldsymbol{\lambda})$ given by (56) has dimension $n+\ell$ with

$$(58)\quad G_n(\mathbf{b},\boldsymbol{\lambda}) = \{g : Q_n^o(\mathbf{b},\boldsymbol{\lambda},D)\ P_n(\sigma_n(\boldsymbol{\lambda}),D)\ g = 0\}$$

<u>Proof</u>: Following the same procedure used in the proof of Lemma 3 we find that $g_n(\mathbf{b},\boldsymbol{\lambda},\mathbf{b}^*,\boldsymbol{\lambda}^*,-)$ satisfies the initial value problem

$$(59)\quad P_n(\sigma_n(\boldsymbol{\lambda}),D)\,g_n(\mathbf{b},\boldsymbol{\lambda},\mathbf{b}^*,\boldsymbol{\lambda}^*,-) = \sum_{i=1}^{n}\lambda_i^* \prod_{\substack{j=1\\ j\neq i}}^{n}(D-\lambda_j)\Big\} Y_n(\mathbf{b},\sigma_n(\boldsymbol{\lambda}),-)$$

$$(60)\quad D^{j-1}g_n(\mathbf{b},\boldsymbol{\lambda},\mathbf{b}^*,\boldsymbol{\lambda}^*,0) = b_j^*\ , \quad j=1,2,\ldots,\ n$$

From the choice of $\lambda_1^o,\ \lambda_2^o,\ \ldots\ \lambda_\ell^o$ we see that (59) may also be written in the form

$$(61)\quad P_n(\sigma_n(\boldsymbol{\lambda}),D)\,g_n(\mathbf{b},\boldsymbol{\lambda},\mathbf{b}^*,\boldsymbol{\lambda}^*,t) = \begin{cases} 0 & \ell = 0 \\ \displaystyle\sum_{i=1}^{\ell}\varphi_i(\boldsymbol{\lambda}^*)\exp(\lambda_i^* t), & \ell \geq 1 \end{cases}$$

where $\varphi_1,\varphi_2,\ \ldots,\ \varphi_\ell$ are linearly independent linear functionals on $\mathbf{C}^n$. From (60), (61) we now see that the linear space $G_n(\mathbf{b},\boldsymbol{\lambda})$ has dimension $n+\ell$. Finally, from (61) it is clear that $G_n(\mathbf{b},\boldsymbol{\lambda})$ contains the kernel of the

$(n+\ell)$-th order differential operator $Q_n^0(\mathbf{b},\lambda,D)P_n(\sigma_n(\lambda),D)$ and therefore (58) holds. □

Upon comparing Lemmas 3, 4 we see that if $y = Y_n(\mathbf{b},\sigma_n(\lambda),-)$ has order $k \leq n$ with exactly $\ell \leq k$ elements in the spectrum $\Lambda[y]$, then $H_n(\mathbf{b},\sigma_n(\lambda))$ has dimension $n+k$ while $G_n(\mathbf{b},\lambda)$ only has dimension $n+\ell$, and one reason why we have chosen to work primarily with the parametrization $Y_n(\mathbf{b},\mathbf{c},-)$ is to avoid this "loss of dimension" in the tangential manifold. (This "loss of dimension" is not totally unfamiliar. For example, we might parametrize the family of straight lines on $[0,1]$ by either

$$F_1(\alpha,\beta,t) = \alpha + \beta^3 t, \quad 0 \leq t \leq 1, \quad \alpha,\beta \in \mathbb{R}$$

or

$$F_2(\alpha,\beta,t) = \alpha + \beta t, \quad 0 \leq t \leq 1, \quad \alpha,\beta \in \mathbb{R}$$

and note that the tangential manifolds of F_1, F_2 have dimensions 1, 2, respectively, when $\alpha = \beta = 0$.)

For some very important applications, however, we cannot make use of the full dimensionality of the tangential manifold, $H_n(\mathbf{b},\mathbf{c})$, due to some constraints which we might wish to impose upon the set S. For example, let

$$\begin{aligned} y(t) &= t \exp(t) \\ &= Y_2(\mathbf{b},\mathbf{c},t), \quad \mathbf{b} = (0,1), \quad \mathbf{c} = (-2,1) \end{aligned} \tag{62}$$

and suppose that we wish to choose $\mathbf{c}^* \in \mathbb{C}^2$ in such a manner that the roots

$$\lambda_{\pm} = 1 - \alpha\, c_1^*/2 \pm \{-4\alpha(c_1^*+c_2^*) + \alpha^2(c_1^*)^2/4\}^{1/2}$$

of the auxiliary polynomial

$$P_2(\mathbf{c}+\alpha\mathbf{c}^*,\lambda) = \lambda^2+(-2+\alpha c_1^*)\cdot\lambda + (1+\alpha c_2^*)$$

are real for all sufficiently small $\alpha > 0$. For this to be the case, it is both necessary and sufficient that $c_1^*, c_2^* \in \mathbb{R}$ and $c_1^*+c_2^* \leq 0$, i.e., by constraining the exponential parameters to lie in the set $S=\mathbb{R}$ we prohibit the full use of $H_n(\mathbf{b},\mathbf{c})$ through the restriction on $\mathbf{c}^*$.

In an effort to deal with situations of this type we shall find it convenient to introduce two new concepts. First of all, we shall define

(63) $$K_n(\mathbf{b},\mathbf{c},\mathbf{c}^*) = \{h_n(\mathbf{b},\mathbf{c},\mathbf{b}^*,\alpha\mathbf{c}^*) : \mathbf{b}^* \in \mathbb{C},\ \alpha \geq 0\}$$

and refer to this subset of $H_n(\mathbf{b},\mathbf{c})$ as the tangential cone of $Y_n(\mathbf{b},\mathbf{c},-)$ at $\mathbf{b},\mathbf{c}$ in the direction of $\mathbf{c}^*$. (For $\mathbf{b},\mathbf{c},\mathbf{c}^* \in \mathbb{R}$ we shall also define $K_n^r(\mathbf{b},\mathbf{c},\mathbf{c}^*)$ to be the subset of real valued functions from $K_n(\mathbf{b},\mathbf{c},\mathbf{c}^*)$.) By comparing (9), (46) and (63) we see at once that

(64) $$L_n\ (\ \mathbf{c}) = K_n(\mathbf{b},\mathbf{c},\mathbf{0})$$

$$(65)\quad H_n(\mathbf{b},\mathbf{c}) = \cup\{K_n(\mathbf{b},\mathbf{c},\mathbf{c}^*) : \mathbf{c}^* \in \mathbf{C}\}$$

Moreover, by using (45) and (55) we see that

$$g_n(\mathbf{b},\lambda,\mathbf{b}^*,\lambda^*,-) = \frac{d}{d\alpha} Y_n(\mathbf{b} + \alpha\mathbf{b}^*,\ \sigma_n(\lambda+\alpha\lambda^*),-)\Big|_{\alpha=0}$$

$$= h_n(\mathbf{b},\sigma_n(\lambda),\mathbf{b}^*,\mathbf{x}_n(\lambda,\lambda^*),-)$$

where

$$(66)\quad \mathbf{x}_n(\lambda,\lambda^*) = \frac{d}{d\alpha}\sigma_n(\lambda + \alpha\lambda^*)\Big|_{\alpha=0}$$

so that

$$(67)\quad G_n(\mathbf{b},\lambda) = \cup\{K_n(\mathbf{b},\sigma_n(\lambda),\ \mathbf{x}_n(\lambda,\lambda^*)) : \lambda^* \in \mathbf{C}^n\}$$

Results analogous to (64), (65), and (67) also hold for $L_n^r(\mathbf{c})$, $H_n^r(\mathbf{b},\mathbf{c})$, and $G_n^r(\mathbf{b},\mathbf{c})$, respectively.

The second new concept which we wish to introduce is the following. We shall say that the exponential sum y is accessible through the cone $K_n(\mathbf{b},\mathbf{c},\mathbf{c}^*)$ with respect to $V_n(S)$ provided $y = Y_n(\mathbf{b},\mathbf{c},-)$ and there exists some continuously differentiable arc $\mathbf{z} : [0,1] \to \mathbf{C}^n$ such that $\mathbf{z}(0)=\mathbf{c}$, $\mathbf{z}'(0) = \mathbf{c}^*$, and $\Lambda_n(\mathbf{z}(\alpha)) \subseteq S$, $0 \le \alpha \le 1$. For example, let y, $\mathbf{b}$, $\mathbf{c}$ be given by (62), and let $\mathbf{c}^* = (1,-1)$. Then we may set

$$\mathbf{z}(\alpha) = \mathbf{c} + \alpha\,\mathbf{c}^*, \quad 0 \le \alpha \le 1$$

and note that the roots of the polynomial $P_2(\mathbf{c} + \alpha\mathbf{c}^*, -)$ are real for $0 \leq \alpha \leq 1$. Hence y is accessible through the cone $K_2(\mathbf{b}, \mathbf{c}, \mathbf{c}^*)$ with respect to $V_2(\mathbb{R})$. (The same is true whenever $\mathbf{c}^* = (c_1^*, c_2^*) \in \mathbb{R}^2$ with $c_1^* + c_2^* \leq 0$.)

We shall find these concepts most helpful in the following sections where we formulate necessary conditions which a best $\| \ \|_p$ - approximation to $f \in L_p[0,1]$ from $V_n(S)$ must satisfy.

IV. CHARACTERIZATION OF A BEST APPROXIMATION, $p < \infty$

We are now ready to formulate theorems which may be used to characterize a local best $\| \ \|_p$ - approximation to a given $f \in L_p[0,1]$ from $V_n(S)$, and in this section we shall restrict our attention to the case $1 \leq p < \infty$. (We say that $y_o \in V_n(S)$ is a local best $\| \ \|_p$ - approximation to f provided the inequality $\|f-y\|_p \geq \|f-y_o\|_p$ holds for all y in some $\| \ \|_p$ - neighborhood of y_o in $V_n(S)$.) In so doing we shall use the tangential cone to formulate and prove what might well be called the fundamental first order necessary condition for a local best $\| \ \|_p$ - approximation and then specialize this result to the cases $S=\mathbb{R}$, $\mathbb{C}$. We shall use this necessary condition to show that (when $1 < p < \infty$ and S has an accumulation point) a local best $\| \ \|_p$ - approximation to $f \notin V_n(S)$ must always have full order n and that any sequence from $V_n(S)$ which $\| \ \|_p$ - converges to such a local best $\| \ \|_p$ - approximation must be a V-sequence. Under stronger hypotheses we formulate a sufficient condition by which one can recognize a local best $\| \ \|_p$ - approximation. Finally, by means of a simple example, we shall illustrate the use and limitations of these theorems and demonstrate the possible nonuniqueness of a best $\| \ \|_p$ - approximation.

For each $f \in L_p[0,1]$ we shall define

$$(68)\quad Z_f = \{t \in [0,1] : f(t) = 0\}$$

so that Z_f is the zero set of f. Given f, $g \in L_p[0,1]$ we also define $T[f,g;-] \in L_p[0,1]$ such that

$$(69)\quad T[f,g;t] = \begin{cases} |g(t)| & t \in Z_f \\ \mathrm{Re}\{g(t)\,\overline{f(t)}/|f(t)|\} & t \notin Z_f \end{cases}$$

(where the bar denotes a complex conjugate.) We may now state the following lemma which forms the basis for the characterization theorems of this section.

Lemma 5: Let $1 \leq p < \infty$ and let $f, g \in L_p[0,1]$ with $\|f\|_p \neq 0$. Then

$$\|f+\alpha g\|_p = \|f\|_p + \alpha \|f\|_p^{1-p} \int_0^1 |f(t)|^{p-1} T[f,g;t]\,dt + o(\alpha)$$

as α decreases to zero through positive values. (Here $o(\alpha)/\alpha \to 0$ as $\alpha \to 0$, and we set $|f(t)|^{p-1}=1$ in the integral whenever $t \in Z_f$ and $p=1$.)

Proof: Using the Lebesgue dominated convergence theorem (cf. [22,p.89]) we may show that

$$\lim_{\alpha \to 0+} \alpha^{-1}\{\|f+\alpha g\|_p^p - \|f\|_p^p\} = p \int_0^1 |f(t)|^{p-1} T[f,g;t]\,dt$$

from which the lemma follows. □

Using the lemma we may now establish the basic first order necessary condition for a local best $\|\ \|_p$ - approximation.

Theorem 7: Let $f \in L_p[0,1]$ with $1 \leq p < \infty$, let $y_o = Y_n(\mathbf{b},\mathbf{c},-)$ be a local best $\| \ \|_p$- approximation to f from $V_n(S)$, and let $\varepsilon = f - y_o$. Assume further that y_o is accessible through the tangential cone $K_n(\mathbf{b},\mathbf{c},\mathbf{c}^*)$ with respect to $V_n(S)$. Then y_o is a best $\| \ \|_p$- approximation to f from $K_n(\mathbf{b},\mathbf{c},\mathbf{c}^*)$, and, in particular,

$$(70) \quad \int_0^1 |\varepsilon(t)|^{p-1} \, T[\varepsilon,-h;t] \, dt \geq 0$$

must hold for each $h \in K_n(\mathbf{b},\mathbf{c},\mathbf{c}^*)$. Moreover, the result remains valid if $L_p[0,1]$, $V_n(S)$, and $K_n(\mathbf{b},\mathbf{c},\mathbf{c}^*)$ are replaced by $L_p^r[0,1]$, $V_n^r(S)$, and $K_n^r(\mathbf{b},\mathbf{c},\mathbf{c}^*)$, respectively, with $\mathbf{b},\mathbf{c},\mathbf{c}^* \in \mathbf{R}$ in this case.

Proof: Let $h \in K_n(\mathbf{b},\mathbf{c},\mathbf{c}^*)$ be chosen. From the definition (63) it follows that we may find some $\mathbf{b}^* \in \mathbf{C}^n$ and some $\alpha_o \geq 0$ such that

$$\begin{aligned} h &= h_n(\mathbf{b},\mathbf{c},\mathbf{b}^* + \mathbf{b}, \ \alpha_o \mathbf{c}^*) \\ &= h_n(\mathbf{b},\mathbf{c},\mathbf{b}^*, \ \alpha_o \mathbf{c}^*) + h_n(\mathbf{b},\mathbf{c},\mathbf{b},\mathbf{0}) \\ &= h_n(\mathbf{b},\mathbf{c},\mathbf{b}^*, \ \alpha_o \mathbf{c}^*) + y_o \end{aligned}$$

and we may (by rescaling $\mathbf{c}^*$, if necessary) assume $\alpha_o = 1$. Moreover, since y_o is accessible through the tangential cone $K_n(\mathbf{b},\mathbf{c},\mathbf{c}^*)$ we may find a smooth arc $\mathbf{z}(\alpha)$, $0 \leq \alpha \leq 1$, such that $\mathbf{z}(0) = \mathbf{c}$, $\mathbf{z}'(0) = \mathbf{c}^*$, and $\Lambda_n(\mathbf{z}(\alpha)) \subseteq S$ for $0 \leq \alpha \leq 1$.

This being the case we may use the fact that y_o is a local best $\|\ \|_p$ - approximation to f from $V_n(S)$ together with the mean value theorem to note that as α decreases to zero through positive values we must have

$$\begin{aligned}\|f-y_o\|_p &\le \|f-Y_n(\mathbf{b}+\alpha\mathbf{b}^*,\mathbf{z}(\alpha),-)\|_p \\ &= \|f-Y_n(\mathbf{b},\mathbf{z}(0),-)-\alpha h_n(\mathbf{b},\mathbf{z}(0),\mathbf{b}^*,\mathbf{z}'(0),-)\|_p+o(\alpha) \\ &= \|f-y_o-\alpha(h-y_o)\|_p + o(\alpha) \\ &= \|(1-\alpha)\cdot(f-y_o) + \alpha\cdot(f-h)\|_p + o(\alpha) \\ &\le \|f-y_o\|_p - \alpha\{\|f-y_o\|_p-\|f-h\|_p\} + o(\alpha)\end{aligned}$$

We conclude from this inequality that $\|f-y_o\|_p \le \|f-h\|_p$ so that y_o is a best $\|\ \|_p$- approximation to f from $K_n(\mathbf{b},\mathbf{c},\mathbf{c}^*)$.

To complete the proof we must establish (70), and since (70) holds trivially when $\|\varepsilon\|_p=0$ we shall assume $\|\varepsilon\|_p > 0$. We note that if $h \in K_n(\mathbf{b},\mathbf{c},\mathbf{c}^*)$ then $y_o+\alpha h \in K_n(\mathbf{b},\mathbf{c},\mathbf{c}^*)$ for all $\alpha \ge 0$. Using this together with Lemma 5 and the fact that y_o is a best $\|\ \|_p$-approximation to f from $K_n(\mathbf{b},\mathbf{c},\mathbf{c}^*)$, we find that

$$\begin{aligned}\|f-y_o\|_p &\le \|f-y_o-\alpha h\|_p \\ &= \|f-y_o\|_p+\alpha\|\varepsilon\|^{1-p}\int_0^1|\varepsilon(t)|^{p-1}T[\varepsilon,-h;t]dt + o(\alpha)\end{aligned}$$

must hold as α decreases to zero through positive values, and this gives (70). The real case is proved analogously. □

The following corollaries display somewhat more useful forms of this basic first order necessary condition.

Corollary 1: Let $f \in L_p[0,1]$ with $1 \leq p < \infty$, let $y_o = Y_n(\mathbf{b},\mathbf{c},-)$ be a local best $\| \ \|_p$- approximation to f from $V_n(S)$, and let $\varepsilon = f - y_o$. Then y_o is a best $\| \ \|_p$- approximation to f from the n dimensional linear space $L_n(\mathbf{c})$. In particular, for every $h \in L_n(\mathbf{c})$

$$\left| \int_{Z_\varepsilon^c} |\varepsilon(t)|^{-1} \ \overline{\varepsilon(t)} \ h(t) \ dt \right| \leq \int_{Z_\varepsilon} |h(t)| dt \quad \text{if } p=1$$

(71)

$$\int_{Z_\varepsilon^c} |\varepsilon(t)|^{p-2} \ \overline{\varepsilon(t)} \ h(t) \ dt = 0 \quad \text{if } 1 < p < \infty$$

where $Z_\varepsilon^c = [0,1] \setminus Z_\varepsilon$. Moreover, the result remains valid if $L_p[0,1]$, $V_n(S)$, and $L_n(\mathbf{c})$ are replaced by $L_p^r[0,1]$, $V_n^r(S)$, $L_n^r(\mathbf{c})$, respectively, with $\mathbf{b},\mathbf{c}, \in \mathbb{R}^n$ in this case.

Proof: Since $L_n(\mathbf{c}) \subseteq V_n(S)$ it is clear that y_o must be a best $\| \ \|_p$- approximation from $L_n(\mathbf{c})$. Moreover, since y_o is accessible through the (degenerate) cone $K_n(\mathbf{b},\mathbf{c},\mathbf{0}) = L_n(\mathbf{c})$ (e.g., take the smooth arc $\mathbf{z}(\alpha) \equiv \mathbf{c}$) we see that (70) must hold for every $h \in L_n(\mathbf{c})$. Since $L_n(\mathbf{c})$ is a linear space, (70) must also hold when h is replaced by θh where $\theta \in \mathbb{C}$ and $|\theta| = 1$. Together with (69) this gives

$$0 \leq \int_0^1 |\varepsilon(t)|^{p-1} T[\varepsilon, -\theta h; t]\,dt$$

$$= \int_{Z_\varepsilon} |\varepsilon(t)|^{p-1} |h(t)|\,dt \;-\; \mathrm{Re}\{\theta \cdot \int_{Z_f^c} |\varepsilon(t)|^{p-2}\, \overline{\varepsilon(t)} h(t)\,dt\}$$

from which we obtain (71) by appropriately choosing θ. ∎

Corollary 2: Let $f \in L_p[0,1]$ with $1 \leq p < \infty$, let $y_o = Y_n(\mathbf{b},\mathbf{c},-)$ be a local best $\| \|_p$-approximation to f from $V_n(S)$ with the order of y_o being k, and let $\varepsilon = f - y_o$. Assume further that the n roots of the auxiliary polynomial $P_n(\mathbf{c},-)$ are all interior points of S. Then y_o is a best $\| \|_p$- approximation to f from the n+k dimensional tangential manifold $H_n(\mathbf{b},\mathbf{c})$, and, in particular, (71) holds for every $h \in H_n(\mathbf{b},\mathbf{c})$. Moreover, the result remains valid if $L_p[0,1]$, $V_n(S)$, and $H_n(\mathbf{b},\mathbf{c})$ are replaced by $L_p^r[0,1]$, $V_n^r(S)$, and $H_n^r(\mathbf{b},\mathbf{c})$, respectively, with $\mathbf{b},\mathbf{c} \in \mathbb{R}^n$ in this case.

Proof: Let $\mathbf{c}^* \in \mathbb{C}^n$ be chosen and let

$$\mathbf{z}(\alpha) = \mathbf{c} + \alpha\, \mathbf{c}^*, \qquad \alpha \geq 0$$

Since the roots of $P_n(\mathbf{c},-)$ are in the interior of S we have $\Lambda_n(\mathbf{z}(\alpha)) \subseteq S$ for all sufficiently small α. Thus y_o is accessible through the tangential cone $K_n(\mathbf{b},\mathbf{c},\mathbf{c}^*)$ with respect to $V_n(S)$ so that y_o must be a best $\| \|_p$- approximation to f from $K_n(\mathbf{b},\mathbf{c},\mathbf{c}^*)$. But since $\mathbf{c}^* \in \mathbb{C}^n$ is arbitrary,

we may use (65) to conclude that y_o must also be a best $\| \|_p$-approximation to f from the whole tangential manifold $H_n(\mathbf{b},\mathbf{c})$. Finally, since $H_n(\mathbf{b},\mathbf{c})$ is a linear space, the same argument used in Corollary 1 above shows that (71) must hold for all $h \in H_n(\mathbf{b},\mathbf{c})$. The real case is handled analogously. □

Corollary 3: Let $f \in L_p[0,1]$ with $1 \leq p \leq \infty$, let $y_o = Y_n(\mathbf{b},\sigma_n(\boldsymbol{\lambda}),-)$ be a local best $\| \|_p$- approximation to f from $V_n(S)$, let $\varepsilon = f - y_o$, and assume that $\Lambda[y_o]$ contains exactly ℓ elements. Assume further that every component of $\boldsymbol{\lambda}$ is an interior point of S. Then y_o is a best $\| \|_p$-approximation to f from the $n+\ell$ dimensional linear space $G_n(\mathbf{b},\boldsymbol{\lambda})$ (defined by (56)), and, in particular, (71) must hold for all $h \in G_n(\mathbf{b},\boldsymbol{\lambda})$. Moreover, the result remains valid if $L_p[0,1]$, $V_n(S)$, and $G_n(\mathbf{b},\boldsymbol{\lambda})$ are replaced by $L_p^r[0,1]$, $V_n^r(S)$, and $G_n^r(\mathbf{b},\boldsymbol{\lambda})$, respectively, with $S \subseteq \mathbb{R}$, $\mathbf{b}, \boldsymbol{\lambda} \in \mathbb{R}^n$, and with the topology of S being taken with respect to $\mathbb{R}$ in this case.

Proof: Let $\boldsymbol{\lambda}^* \in \mathbb{C}^n$ chosen and let

$$\mathbf{z}(\alpha) = \sigma_n(\boldsymbol{\lambda} + \alpha\boldsymbol{\lambda}^*), \qquad \alpha \geq 0$$

Since the components of $\boldsymbol{\lambda}$ are in the interior of S we have $\mathbf{z}(\alpha) \in S^n$ for all sufficiently small α. Thus we may conclude that y_o is accessible through the cone $K_n(\mathbf{b},\sigma_n(\boldsymbol{\lambda}), \mathbf{z}_n(\boldsymbol{\lambda},\boldsymbol{\lambda}^*))$

with respect to $V_n(S)$ (where $\mathbf{x}_n$ is again defined by (66)) so that y_o must be a best $\| \|_p$- approximation to f from this cone. But since $\lambda^* \in \mathbb{C}^n$ is arbitrary, we may use (67) to conclude that y_o must be a best $\| \|_p$- approximation from $G_n(\mathbf{b},\lambda)$. Finally, since $G_n(\mathbf{b},\lambda)$ is a linear space, the argument of Corollary 1 shows that (71) must hold for all $h \in G_n(\mathbf{b},\lambda)$. The real case is handled analogously. □

The following example has been chosen to illustrate the use of the necessary condition of Corollary 1, to show that a best $\| \|_p$- approximation need not be unique, and to show that the first order necessary condition of Corollary 2 is not a sufficient condition for y_o to be a local best $\| \|_p$- approximation even when y_o has full order n.

<u>Example 1:</u> Let us define

$$(72)\quad f(t) = 90\ t^2 - 90\ t + 16\ , \qquad 0 \le t \le 1$$

and seek a best $\| \|_2$- approximation to f from $V_1(\mathbb{R})$. If the exponential sum

$$y(t) = a\ \exp(\lambda t), \qquad 0 \le t \le 1$$

is a best $\| \|_2$- approximation to f, then the necessary condition of Corollary 1 requires that

$$\int_0^1 [f(t) - y(t)]\ \exp(\lambda t)\ dt = 0$$

or equivalently that

$$a = \{\int_0^1 f(t)\exp(\lambda t)\,dt\} \,/\, \{\int_0^1 \exp(2\lambda t)\,dt\}$$

$$= \{\exp(\lambda/2)(\lambda/2)^{-3}B(\lambda/2)\}/\{\exp(\lambda)(\lambda/2)^{-1}\sinh(\lambda/2)\cosh(\lambda/2)\}$$

where

$$B(u) \equiv [45+16u^2]\sinh(u) - 45u\cosh(u), \quad u \in \mathbb{R}$$

Thus the necessary condition of Corollary 1 enables us to express the linear parameter, a, in terms of the nonlinear parameter, λ. Indeed, if we define

$$\psi(\lambda) = \|f-y\|_2^2$$

then we find

$$\psi(\lambda) = \int_0^1 \{f(t)^2 - y(t)^2 - 2[f(t) - y(t)]\,y(t)\}\,dt$$

$$= \int_0^1 f(t)^2\,dt - a^2\int_0^1 \exp(2\lambda t)\,dt$$

$$= 46-B(\lambda/2)^2/\{(\lambda/2)^5\sinh(\lambda/2)\cosh(\lambda/2)\}$$

and we may locate the best $\|\ \|_2$- approximations of f from $V_1(\mathbb{R})$ by finding the values of λ which minimize ψ. Moreover, since f is symmetric (i.e., $f(t) = f(1-t)$) we see that $\psi(\lambda)$ is an even function of λ, and therefore we

need only locate the minima of Ψ in $[0, +\infty)$. Hence we obtain the best $\| \ \|_2$- approximations

$$y_+ (t) = a \exp(\lambda t), \ a = .00008931..., \ \lambda = 12.3064...$$

$$y_- (t) = y_+ (1-t)$$

(and indeed these are the only best $\| \ \|_2$- approximations to f from $V_1(\mathbb{R})$.)

Suppose now that we consider the $\| \ \|_2$- approximation of f from the larger class $V_1(\mathbb{C})$. Taking

$$y_o(t) = Y_1(1,0,t) \equiv 1$$

we find that (by the construction of f)

$$\int_0^1 [f(t) - y_o(t)] \, dt = \int_0^1 [f(t) - y_o(t)] \cdot t \, dt = 0$$

so that y_o is the unique best approximation to f from the tangential manifold

$$H_1(1,0) = \{ \ h : D^2 h \equiv 0 \ \}$$

Thus y_o satisfies the necessary condition of Corollary 2, but y_o fails to be a best $\| \ \|_2$- approximation to f from $V_1(\mathbb{R})$ and hence from the larger set $V_1(\mathbb{C})$.

Finally, we note that even in the larger class $V_1(\mathbb{C})$ there is not a unique best $\| \ \|_2$- approximation to f.

Indeed, since $f(t) = \overline{f(t)} = f(1-t)$, $0 \leq t \leq 1$, (with the bar again denoting a complex conjugate) we see that the exponential sums

$$y_1(t) = a \exp[\lambda t] \qquad y_2(t) = \bar{a} \exp[\bar{\lambda} t]$$

$$y_3(t) = a \exp[\lambda(1-t)] \qquad y_4(t) = \bar{a} \exp[\bar{\lambda}(1-t)]$$

all $\| \ \|_2$- approximate f equally well, and when $\lambda \neq 0$ at least two of these four functions must be distinct. []

As we have noted earlier, when $y = Y_n(\mathbf{b},\mathbf{c},-)$ has full order n, it is a relatively simple matter to describe a small neighborhood of y in $V_n(\mathbf{C})$, and the effect of small perturbations of the parameters **b**, **c** is easily studied by means of the 2n dimensional tangential manifold $H_n(\mathbf{b},\mathbf{c})$. It is therefore a most pleasant feature of the $p < \infty$ theory that (under very mild hypotheses) every local best $\| \ \|_p$- approximation must have full order n as we see from the following theorem.

<u>Theorem 8:</u> Let $f \in L_p[0,1]$ with $1 \leq p < \infty$, let y_o be a local best $\| \ \|_p$- approximation to f, and assume that $\|f-y_o\|_p > 0$. Assume further that S possesses some finite accumulation point, i.e. that S contains some convergent sequence $\lambda_1, \lambda_2, \ldots$ with $\lambda_i \neq \lambda_j$ for $i \neq j$. Finally, in the case where p=1, assume in addition that $f(t) \neq y_o(t)$ holds for almost all $t \in [0,1]$. Then y_o has full order n.

Moreover, the result remains valid if $L_p[0,1]$ and $V_n(S)$ are replaced by $L_p^r[0,1]$ and $V_n^r(S)$, respectively.

Proof: We shall first show that the theorem holds when n=1. Indeed, if the zero function is a local best $\| \|_p$-approximation to f from $V_1(S)$ then we infer from Corollary 1 of Theorem 7 that

$$(73) \quad \int_0^1 |f(t)|^{p-2} \overline{f(t)}\, y(t)\, dt = 0$$

holds whenever $y \in V_1(S)$. This being the case, (73) also holds whenever y is a linear combination of terms from $V_1(S)$, and in particular whenever y may be expressed as the linear combination

$$y(t)=\exp(\lambda_\nu t)\int_0^t\int_0^{t_1}\cdots\int_0^{t_{n-1}} \exp\Big\{\sum_{j=1}^{n}(\lambda_{\nu+j}-\lambda_{\nu+j-1})t_j\Big\}dt_1dt_2\cdots dt_n$$

for some $\nu=1,2,\ldots$ and some $n=1,2,\ldots$. Thus if λ is the limit of the sequence $\{\lambda_\nu\}$, we see that (73) holds whenever

$$y(t) = \exp(\lambda t)\int_0^t\int_0^{t_1}\cdots\int_0^{t_{n-1}} dt_1\, dt_2 \cdots dt_n$$

$$= \exp(\lambda t)\cdot t^n/n\,!$$

for some $n=1,2,\ldots$. Finally, (73) must also hold for all linear combinations of such terms, i.e., for all $y \in V_\infty(\{\lambda\})$. Using (73) together with the Hölder inequality we find that

$$\|f\|_p^p = \int_0^1 |f|^{p-2} \cdot \bar{f} \cdot [f-y]\, dt$$

$$\leq \{ \int_0^1 |f|^p\, dt \}^{(p-1)/p} \cdot \{ \int_0^1 |f-y|^p\, dt \}^{1/p}$$

$$= \|f\|_p^{p-1} \cdot \|f-y\|_p$$

holds for every $y \in V_\infty(\{\lambda\})$, and in view of Theorem 4 this implies that $\|f\|_p = 0$. Hence the theorem holds when n=1.

The extension to n=2,3,... is now immediate. Indeed, if $y_o \in V_{n-1}(S)$ and $\|f-y_o\|_p > 0$, then the zero function cannot be a local best $\| \|_p$- approximation to $f-y_o$ from $V_1(S)$, and hence y_o cannot be a local best $\| \|_p$- approximation to f from $V_n(S)$. □

The mild requirement that S have a finite accumulation point cannot be relaxed as we see from the following example.

Example 2: Let us define

$$f(t) \equiv 1, \quad 0 \leq t \leq 1$$

and seek a best $\| \|_2$ - approximation to f from $V_1(S)$ where

$$S \subseteq \{ \nu \cdot 2\pi i : \nu = \pm 1, \pm 2, \ldots \}$$

Using Corollary 1 of Theorem 7 we see that if

$$y(t) = a \exp(\nu \cdot 2\pi i \cdot t), \qquad 0 \leq t \leq 1$$

is a best $\| \|_\infty$ - approximation to f from $V_1(S)$, then a = 0. Hence $y_o \equiv 0$ is the unique best $\| \|_2$ - approximation to f from $V_1(S)$. □

Slightly different formulations and proofs of the above theorem for the case of $V_n^r(\mathbb{R})$ are given by Hobby and Rice [12, p.99] and de Boor [2, p.179] with the case p=1 being left undecided. By examining the above proof, we see that when p=1 we can only deduce that

$$\left| \int_{Z_f^c} |f(t)|^{-1} \cdot \overline{f(t)}\, y(t)dt \right| \leq \int_{Z_f} |y(t)|\, dt$$

must hold for all $y \in V_1(S)$ whenever the zero function is a local best $\| \|_p$- approximation to f from $V_1(S)$, and this does not imply that $\|f\|_1 = 0$. Indeed, this observation has been used to construct the following example which shows that without the additional hypothesis, Theorem 8 cannot hold when p=1.

<u>Example 3</u>: Let us define

$$(74)\quad f(t) = \begin{cases} 1 & \text{for} \quad 1/3 \leq t \leq 2/3 \\ 0 & \text{otherwise} \end{cases}$$

and seek a best $\| \|_1$- approximation to f from $V_1(\mathbb{C})$. Indeed, let $y \in V_1(\mathbb{C})$ with $\|y\|_1 > 0$. Then we may write

$$|y(t)| = |y(0)| \exp(\lambda t), \quad |y(0)| > 0, \ \lambda \in \mathbb{R}$$

and for convenience in notation we shall set

$$s = \exp(\lambda/3)$$

We then find

$$\|f+y\|_1 \geq \int_0^{1/3} |y|dt + \int_{1/3}^{2/3} [f-|y|]dt + \int_{2/3}^{1} |y|dt$$

$$= \|f\|_1 + (1-s+s^2) \int_0^{1/3} |y|dt$$

$$= \|f\|_1 + [(1-s+s^2)/(1+s+s^2)] \int_0^{1} |y|dt$$

$$\geq \|f\|_1 + (1/3) \|y\|_1$$

$$> \|f\|_1$$

Hence $y_o=0$ is the unique best $\| \|_1$- approximation to f from $V_1(\mathbf{C})$. □

Two of the primary advantages of $\| \|_p$- approximation with $1 < p < \infty$ are displayed in the following corollaries of Theorem 8.

Corollary 1: Let $f \in L_p[0,1]$ with $1 < p < \infty$ and assume that S contains some finite accumulation point. Let y_o be a local best $\| \|_p$- approximation to f from $V_n(S)$ and assume that $\|f-y_o\|_p > 0$. Then any sequence $\{y_\nu\}$ from $V_n(S)$ which has the $\| \|_p$- limit y_o is a V-sequence.

Proof: The corollary follows immediately from Theorems 6 and 8. □

Corollary 2: Let $f \in L_p[0,1]$ with $1 < p < \infty$, let S be a closed subset of $\mathbf{C}$ which contains some finite accumulation point, and assume that $f \notin V_n(S)$. Then the set of all $(\mathbf{b},\mathbf{c}) \in \mathbf{C}^{2n}$ for which $Y_n(\mathbf{b},\mathbf{c},-)$ is a best $\|\ \|_p$- approximation to f from $V_n(S)$ is compact.

Proof: That the set of best parameters is bounded follows from Corollary 1 above and Lemma 1; the closure of this set follows from the continuity of Y_n and of the $\|\ \|_p$- norm. □

When $1 < p < \infty$ we see from Corollary 2 to Theorem 7 and from Theorem 8 that the exponential sum y_o can be a local best $\|\ \|_p$- approximation to f from $V_n(\mathbf{C})$ only if y_o has full order n and $\|\varepsilon + \alpha h\|_p = \|\varepsilon\|_p + o(\alpha)$ as α decreases to zero, where $\varepsilon = f - y_o$ and h is any element from the tangential manifold $H_n(\mathbf{b},\mathbf{c})$ of y_o. As we have already observed in the discussion of Example 1, these necessary conditions are not sufficient conditions by which one can identify a local best $\|\ \|_p$- approximation. One means of strengthening these conditions so as to obtain a sufficient condition is to introduce some hypothesis on the $o(\alpha)$ term, and we do this in the following theorem.

Theorem 9: Let $f \in L_p[0,1]$ with $1 \leq p < \infty$, let $y_o = Y_n(\mathbf{b},\mathbf{c},-)$ be chosen from $V_n(S)$, let y_o have full order n, and let $\varepsilon = f - y_o$. Let

$$(75)\quad \mu(\alpha) = \sup\{\|d^2 Y_n(\mathbf{b}+\alpha\mathbf{b}^*,\mathbf{c}+\alpha\mathbf{c}^*,-)/d\alpha^2\|_\infty : 0 \le |b_i^*|, |c_i^*| \le \alpha\}$$

be defined for $\alpha > 0$ and assume that

$$(76)\quad \varliminf_{\alpha\to 0+} (\|\varepsilon+\alpha h\|_p - \|\varepsilon\|_p)/\alpha^2 > (1/2) \varlimsup_{\alpha\to 0+} \mu(\alpha)$$

holds for each $h \in H_n(\mathbf{b},\mathbf{c})$ with $\|h\|_p > 0$. Then y_o is a local best $\|\ \|_p$- approximation to f from $V_n(S)$. Moreover, this result remains valid when $L_p[0,1]$, $V_n(S)$, and $H_n(\mathbf{b},\mathbf{c})$ are replaced by $L_p^r[0,1]$, $V_n^r(S)$, and $H_n^r(\mathbf{b},\mathbf{c})$, respectively, with $\mathbf{b},\mathbf{c} \in \mathbb{R}^n$ in this case.

Proof: For each $\alpha > 0$ we define

$$(77)\quad \varphi(\alpha) = \alpha^{-2} \cdot \min\{\|\varepsilon+\alpha h_n(\mathbf{b},\mathbf{c},\mathbf{b}^*,\mathbf{c}^*,-)\|_p - \|\varepsilon\|_p : (\mathbf{b}^*,\mathbf{c}^*) \in \mathbb{K}\}$$

where

$$\mathbb{K} = \{\mathbf{z} \in \mathbb{C}^{2n} : \max(|z_1|, |z_2|, \ldots, |z_{2n}|) = 1\}$$

The set $\mathbb{K}$ is compact, and this together with the continuity of h_n and the hypothesis (76) implies that

$$(78)\quad \varphi(\alpha) > (1/2)\ \mu(\alpha)$$

holds for all sufficiently small $\alpha > 0$. This being the case we may use the mean value theorem together with (75), (77), and (78) to see that for every $(\mathbf{b}^*,\mathbf{c}^*) \in \mathbb{K}$ and for some sufficiently small $\alpha > 0$ we have

$$\|f-Y_n(\mathbf{b}-\alpha\mathbf{b}^*,\mathbf{c}-\alpha\mathbf{c}^*,-)\|_p$$

$$\geq \|f-Y_n(\mathbf{b},\mathbf{c},-)+\alpha h_n(\mathbf{b},\mathbf{c},\mathbf{b}^*,\mathbf{c}^*,-)\|_p - \mu(\alpha)\ \alpha^2/2$$

$$\geq \|\varepsilon\|_p + \alpha^2[\varphi(\alpha) - (1/2)\ \mu(\alpha)]$$

$$> \|\varepsilon\|_p$$

(with the bound being independent of $\mathbf{b}^*$, $\mathbf{c}^*$.) In view of the homeomorphism established by Theorem 6, this implies that y_o is a local best $\|\ \|_p$- approximation to f from $V_n(\mathbf{C})$ and therefore from $V_n(S)$ as well. The real case is handled analogously. □

V. CHARACTERIZATION OF A BEST APPROXIMATION, $p = \infty$

In the formulation of characterization theorems which apply when $p = \infty$ we shall restrict ourselves to the case where the function to be approximated lies in $C[0,1]$. Under this restriction we shall now establish the $p = \infty$ analogues of Lemma 5, of the necessary conditions of Theorem 7 and its corollaries, and of the sufficient condition of Theorem 9. When the approximating functions are drawn from $V_n^r(\mathbb{R})$ these characterization results may also be translated into familiar statements framed in terms of the number of alternations of the error curve. We shall show by means of examples that a best $\| \ \|_\infty$ - approximation need not have full order and that even when the approximating functions are taken from $V_n^r(\mathbb{R})$ or $V_n^r(\mathbb{C})$ a best $\| \ \|_\infty$ - approximation need not be unique. Finally, by taking special advantage of some of the special properties of an exponential sum from $V_n^r(\mathbb{R})$ we shall give an alternative proof of an important theorem of Braess which provides a sufficient condition for a best $\| \ \|_\infty$ - approximation to be expressible as a simple exponential sum (4) with positive coefficients.

We shall first establish the analogue of Lemma 5. In so doing, we shall refer to the set

$$(79)\quad E_f = \{t \in [0,1] : |f(t)| = \|f\|_\infty \}$$

as the set of extreme points of the function $f \in C[0,1]$.

Lemma 6: Let $f, g \in C[0,1]$. Then

(80) $\|f+\alpha g\|_\infty = \|f\|_\infty + \alpha \max\{T[f,g;t] : t \in E_f\} + o(\alpha)$

as α decreases to zero through positive values (with T being again defined by (69).)

Proof: When $\|f\|_\infty = 0$ so that $E_f = [0,1]$ we immediately obtain (80) from the definition of T. We shall therefore assume that $\|f\|_\infty > 0$ in the remainder of the proof.

It is convenient to introduce an auxiliary function in the following manner. First of all, we define

(81) $E_f^\alpha = \{t \in [0,1] : |f(t)| \geq \|f\|_\infty - 2\alpha\|g\|_\infty\}$, $\alpha \geq 0$

and we then set

(82) $r(\alpha) = \max\{\overline{f(t)}\, g(t) + f(t)\, \overline{g(t)} : t \in E_f^\alpha\}$, $\alpha \geq 0$

Clearly r is a nondecreasing function of α. We shall show that r is continuous at $\alpha = 0$ (even though r may fail to be continuous at points $\alpha > 0$.) Indeed, if r fails to be continuous at $\alpha = 0$, then there exists some $\varepsilon > 0$ and positive numbers $\alpha_1 > \alpha_2 > \ldots$ with $\lim \alpha_\nu = 0$ such that for each ν there exists some point t_ν from $E_f^{\alpha_\nu}$ such that

$$|f(t_\nu)| \geq \|f\|_\infty - 2\,\alpha_\nu\,\|g\|_\infty$$

$$\overline{f(t_\nu)}\, g(t_\nu) + f(t_\nu)\, \overline{g(t_\nu)} \geq r(\alpha_\nu) - \varepsilon \geq r(0) + \varepsilon$$

We may assume (by passing to a subsequence, if necessary) that $\{t_\nu\}$ converges to some point $t_o \in [0,1]$. This being the case, we may pass to the limit in the above inequalities and thus obtain the contradiction

$$|f(t_o)| \geq \|f\|_\infty \qquad \text{i.e.} \quad t_o \in E_f^o$$

$$\overline{f(t_o)}\, g(t_o) + f(t_o)\, \overline{g(t_o)} \geq r(0) + \varepsilon \qquad \text{i.e.} \quad t_o \notin E_f^o$$

Hence we conclude that r is continuous at $\alpha = 0$.

We now proceed to establish (80). Using (81) we see that for $t \in [0,1] \setminus E_f^\alpha$ we have

$$\begin{aligned} |f(t) + \alpha\, g(t)| &\leq |f(t)| + \alpha|g(t)| \\ &< \|f\|_\infty - 2\alpha\|g\|_\infty + \alpha|g(t)| \\ &\leq \|f\|_\infty - \alpha\|g\|_\infty \\ &\leq \|f+\alpha g\|_\infty \end{aligned}$$

and this together with (82) gives

$$\begin{aligned} \|f+\alpha g\|_\infty^2 &= \max\{|f(t)+\alpha g(t)|^2 : t \in [0,1]\} \\ &= \max\{|f(t)+\alpha g(t)|^2 : t \in E_f^\alpha\} \\ &\leq \|f\|_\infty^2 + \alpha r(\alpha) + \alpha^2\|g\|_\infty^2 \\ &= \|f\|_\infty^2 + \alpha r(0) + o(\alpha) \end{aligned}$$

On the other hand, we also have

$$\|f+\alpha g\|_\infty^2 \geq \max\{|f(t)+\alpha g(t)|^2 : t \in E_f^o\}$$

$$\geq \|f\|_\infty^2 + \alpha \max\{\overline{f(t)}\, g(t) + f(t)\overline{g(t)} : t \in E_f^o\} - \alpha^2 \|g\|_\infty^2$$

$$= \|f\|_\infty^2 + \alpha\, r(0) - \alpha^2 \|g\|_\infty^2$$

and hence we obtain

$$\|f+\alpha g\|_\infty^2 = \|f\|_\infty^2 + \alpha\, r(0) + o(\alpha)$$

$$= \|f\|_\infty^2 + 2\alpha \|f\|_\infty \cdot \max\{T[f,g;t] : t \in E_f\} + o(\alpha)$$

from which (80) follows by the taking of square roots. ∎

There are two corollaries of Rolle's theorem which will be useful in the later development, and we shall present them at this point.

Lemma 7: Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be real numbers, let $\varphi_0 \in C^n[0,1]$ be real valued, let

$$\varphi_1 = (D-\lambda_1)\, \varphi_0$$

$$\varphi_2 = (D-\lambda_2)\, \varphi_1$$

$$\vdots$$

$$\varphi_n = (D-\lambda_n)\, \varphi_{n-1}$$

and assume that φ_0 has $m > n$ zeros in the real interval (α,β). Then φ_ν has at least $m-\nu$ real zeros in (α,β),

$\nu = 1,2,\ldots, n$. The result remains true irrespective of whether we count only distinct zeros or whether we count zeros according to their multiplicity.

Proof: If φ_0 has exactly m_r zeros counted according to their multiplicities and $m_d \leq m_r$ distinct zeros in (α,β), then the same is true for the auxiliary function

$$\psi_o(t) = \exp(-\lambda_1 t)\ \varphi_o(t), \quad \alpha < t < \beta$$

Using Rolle's theorem, we thus conclude that

$$D\ \psi_o(t) = \exp(-\lambda_1 t)\ (D-\lambda_1)\ \varphi_0(t) = \exp(-\lambda_1 t)\ \varphi_1(t)$$

has at least m_d-1 distinct zeros interlacing those of φ_0 as well as m_r-m_d zeros counted according to their multiplicity which coincide with some of the zeros of φ_0. Hence the same is true for φ_1 so that the lemma holds when n=1 and thus for n=2,3,... as well. □

Using Lemma 7, we may establish the following well known result, cf.[16, p.48].

Lemma 8: If $y \in V_n(\mathbb{R})$ has n real zeros (with zeros being counted according to their multiplicity) then $y \equiv 0$.

Proof: Assume that $y \in V_n(\mathbb{R})$ has n real zeros and let $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{R}$ be chosen in such a manner that

$$(83)\quad (D-\lambda_1)(D-\lambda_2)\ \ldots\ (D-\lambda_n)\ y = 0$$

Since (83) also holds for both the real and imaginary parts of y, we may assume with no loss of generality that y is real valued. Using Lemma 7 we conclude that for each $\nu=1,2,\ldots,$ n-1 the exponential sum $|(D-\lambda_1)(D-\lambda_2)\ldots(D-\lambda_\nu)|y$ has at least $n-\nu \geq 1$ real zeros and therefore cannot have the order 1. Together with (83) this implies that y has order 0, i.e. $y \equiv 0$. □

When both the approximating and approximated functions are real valued, the alternations of the error curve $\varepsilon = f-y$ are often used in the characterization of best $\| \|_\infty$ - approximations. We shall say that the (real valued) error curve ε alternates at least m times on [0,1] provided there exist m+1 points $0 \leq t_0 < t_1 < \ldots < t_m \leq 1$ such that

$$|\varepsilon(t_i)| = \|\varepsilon\|_\infty$$

$$\varepsilon(t_i) = -\varepsilon(t_{i-1}) , \quad i = 1,2,\ldots, m$$

Using Lemma 8 we may extend a sufficient condition of Rice [18, p.158] to apply to $V_n(\mathbb{R})$.

Theorem 10: Let $f \in C[0,1]$ be real valued, let $y_0 \in V_n^r(\mathbb{R})$ have order k, and assume that the (real valued) error curve $\varepsilon = f-y_0$ alternates at least n+k times on [0,1]. Then y_0 is the unique best $\| \|_\infty$ - approximation to f from $V_n(\mathbb{R})$.

Proof: Let y be a best $\| \|_\infty$ - approximation to f from $V_n(\mathbb{R})$ and assume that

$$y = y_r + i\, y_i$$

is the decomposition of y into real and imaginary parts. Then

$$\|f-y_r\|_\infty \leq \|f-y\|_\infty \leq \|f-y_o\|_\infty$$

so that

$$y_r(t) - y_o(t) = [f(t) - y_o(t)] - [f(t) - y_r(t)]$$

is alternatively non-positive and non-negative on the n+k+1 points of alternation of $f-y_o$. Hence $y_r - y_o$ has at least n+k real zeros (counted according to their multiplicities) in [0,1] and since $y_r - y_o \in V_{n+k}(\mathbb{R})$ we may conclude from Lemma 8 that $y_r = y_o$. Finally, since y_i must vanish at the n+k+1 points of alternation of $f-y_o$ we likewise conclude that $y_i = 0$. Hence $y = y_o$ and the proof is complete. □

The following example has been constructed to show that the sufficient condition of Theorem 10 no longer applies in the larger class $V_n^r(\mathbb{C})$. Indeed, the possession of an alternant of any fixed length (no matter how long) is not a sufficient condition to characterize a best

$\| \|_\infty$ - approximation to f from $V_n^r(\mathbf{C})$ when $n \geq 2$. (We note that this contradicts Rice [21,p.63] where it is incorrectly stated that $V_n^r(\mathbf{C})$ is varisolvent under a suitable parametrization.)

Example 4: Let us define

$$f(t) = \cos(N\pi t) \qquad 0 \leq t \leq 1$$

and seek a best $\| \|_\infty$ - approximation to f from $V_2^r(\mathbf{C})$. Since $f \in V_2^r(\mathbf{C})$, f is its own best $\| \|_\infty$ - approximation from this class. We note however, that if

$$y(t) = a \cos(N\pi t)$$

for some real $a \neq 1$, then the error curve $\varepsilon = f-y$ alternates exactly N times on [0,1]. Clearly, y fails to be even a local best $\| \|_\infty$ - approximation to f from $V_2^r(\mathbf{C})$. □

Using Lemma 6 we may now establish the basic first order necessary conditions for a local best $\| \|_\infty$ - approximation.

Theorem 11: Let $f \in C[0,1]$, let $y_o = Y_n(\mathbf{b},\mathbf{c},-)$ be a local best $\| \|_\infty$ - approximation to f from $V_n(S)$, and let $\varepsilon = f-y_o$. Assume further that y_o is accessible through the tangential cone $K_n(\mathbf{b},\mathbf{c},\mathbf{c}^*)$ with respect to $V_n(S)$. Then y_o is a best $\| \|_\infty$ - approximation to f from $K_n(\mathbf{b},\mathbf{c},\mathbf{c}^*)$, and, in particular,

(84) $\min\{\mathrm{Re}[\overline{\varepsilon(t)}\ h(t)] : t \in E_\varepsilon\} \leq 0$

must hold for each $h \in K_n(\mathbf{b},\mathbf{c},\mathbf{c}^*)$. Moreover, the result remains valid if $V_n(S)$ and $K_n(\mathbf{b},\mathbf{c},\mathbf{c}^*)$ are replaced by $V_n^r(S)$ and $K_n^r(\mathbf{b},\mathbf{c},\mathbf{c}^*)$, respectively, with $\mathbf{b},\mathbf{c},\mathbf{c}^* \in \mathbb{R}^n$ in this case.

Proof: The proof of the theorem follows from Lemma 6 in the same way that Theorem 7 follows from Lemma 5. □

The following corollaries display somewhat more useful forms of this basic first order necessary condition.

Corollary 1: Let $f \in C[0,1]$, let $y_o = Y_n(\mathbf{b},\mathbf{c},-)$ be a local best $\|\ \|_\infty$ - approximation to f from $V_n(S)$, and let $\varepsilon = f-y_o$. Then y_o is a best $\|\ \|_\infty$ - approximation to f from the n dimensional space $L_n(\mathbf{c})$, and, in particular, (84) holds for all $h \in L_n(\mathbf{c})$. Moreover, the result remains valid if $V_n(S)$ and $L_n(\mathbf{c})$ are replaced by $V_n^r(S)$ and $L_n^r(\mathbf{c})$, respectively, with $\mathbf{b},\mathbf{c} \in \mathbb{R}^n$ in this case.

Proof: The proof is analogous to that of Corollary 1 of Theorem 7. □

Corollary 2: Let $f \in C[0,1]$, let $y_o = Y_n(\mathbf{b},\mathbf{c},-)$ be a local best $\|\ \|_\infty$ - approximation to f from $V_n(S)$ with the order of y being k, and let $\varepsilon = f-y_o$. Assume further that the n roots of the auxiliary polynomial $P_n(\mathbf{c},-)$ all are

interior points of S. Then y_o is a best $\| \|_\infty$ - approximation to f from the n+k dimensional tangential manifold $H_n(\mathbf{b},\mathbf{c})$, and, in particular, (84) holds for all $h \in H_n(\mathbf{b},\mathbf{c})$. Moreover, the result remains valid if $V_n(S)$ and $H_n(\mathbf{b},\mathbf{c})$ are replaced by $V_n^r(S)$ and $H_n^r(\mathbf{b},\mathbf{c})$, respectively, with $\mathbf{b},\mathbf{c} \in \mathbb{R}^n$ in this case.

Proof: The proof is analogous to that of Corollary 2 of Theorem 7. □

Corollary 3: Let $f \in C[0,1]$, let $y_o = Y_n(\mathbf{b},\sigma_n(\boldsymbol{\lambda}),-)$ be a local best $\| \|_\infty$ - approximation to f from $V_n(S)$, let $\varepsilon = f-y_o$, and assume that $\Lambda[y_o]$ contains exactly ℓ elements. Assume further that every component of $\boldsymbol{\lambda}$ is an interior point of S. Then y_o is a best $\| \|_\infty$ - approximation to f from the $n+\ell$ dimensional linear space $G_n(\mathbf{b},\boldsymbol{\lambda})$ (defined by (56)) and, in particular, (84) must hold for all $h \in G_n(\mathbf{b},\boldsymbol{\lambda})$. Moreover, this result remains valid if $V_n(S)$ and $G_n(\mathbf{b},\boldsymbol{\lambda})$ are replaced by $V_n^r(S)$ and $G_n^r(\mathbf{b},\boldsymbol{\lambda})$, respectively, with $S \subseteq \mathbb{R}$, $\mathbf{b}$, $\boldsymbol{\lambda} \in \mathbb{R}^n$, and with the topology of S being taken with respect to $\mathbb{R}$ in this case.

Proof: The proof is analogous to that of Corollary 3 of Theorem 7. □

The result of Corollary 3 is contained in the work of de Boor [2, p.180] where it was proved within the theory

of γ - polynomials. By further specializing these corollaries to the case where $y_o \in V_n^r(\mathbb{R})$ we may make use of Lemma 8 and thereby obtain the following more familiar characterization in terms of the alternations of the error curve.

Corollary 4: Let $f \in C[0,1]$ be real valued, let $S \subseteq \mathbb{R}$, let y_o be a local best $\| \|_\infty$ - approximation to f from $V_n^r(S)$, let $\Lambda[y_o]$ have exactly ℓ elements, and assume that S contains a (real) neighborhood of each point of $\Lambda[y_o]$. Then the error curve $\varepsilon = f-y_o$ must alternate at least $n+\ell$ times on $[0,1]$. If in addition y has order ℓ, then y_o is a unique best $\| \|_\infty$ - approximation to f from $V_n(S)$.

Proof: We may choose $\mathbf{b} \in \mathbb{R}^n$ and $\mathbf{A} \in S^n$ such that $y_o = Y_n(\mathbf{b}, \sigma_n(\mathbf{A}), -)$. In view of Corollary 3, y_o must be a best $\| \|_\infty$ - approximation to f from the $n+\ell$ dimensional linear space $G_n^r(\mathbf{b}, \mathbf{A})$. But from Lemmas 4 and 8 we see that any basis of $G_n^r(\mathbf{b}, \mathbf{A})$ is a Tchebycheff set, and therefore (cf.[19,p.56]) $\varepsilon = f-y_o$ must alternate at least $n+\ell$ times on $[0,1]$. Finally, if in addition the order of y_o is also ℓ, we may use Theorem 10 to conclude that y_o is a unique best $\| \|_\infty$ - approximation to f from $V_n(S)$. []

The S=$\mathbb{R}$ version of the above corollary was first obtained by Rice [18,p.158] for the case where the essential exponential parameters of y_o are distinct and more generally by Braess [3,p.313]. By setting the result in the above

context it is easy to see why this necessary condition fails to be a sufficient condition.

We might also note that Corollary 4 (together with Corollary 1 of Theorem 3) implies the existence of a unique best $\| \ \|_\infty$ - approximation from $V_1^r(\mathbb{R})$ for every real valued $f \in C[0,1]$. In contrast, we recall that this need not be the case when $1 < p < \infty$ (cf. Example 1.)

Corollary 5: Let $f \in C[0,1]$ be real valued, let y_0 be a local best $\| \ \|_\infty$ - approximation to f from $V_n^r(S)$, let y_0 have order k, let $\Lambda[y_0] \subset \mathbb{R}$, and assume that S contains a (complex) neighborhood of each point of $\Lambda[y_0]$. Then the (real valued) error curve $\varepsilon = f - y_0$ must alternate at least n+k times on [0,1].

Proof: Let $\mathbf{b}, \mathbf{c} \in \mathbb{R}^n$ be chosen in such a manner that $y_0 = Y_n(\mathbf{b}, \mathbf{c}, -)$ and that each root of the auxiliary polynomial $P_n(\mathbf{c}, -)$ is real and lies in the interior of S. From Corollary 2 we see that y_0 must be a best $\| \ \|_\infty$ - approximation to f from the n+k dimensional linear space $H_n^r(\mathbf{b}, \mathbf{c})$, and from the choice of $\mathbf{c}$ we see that any basis of $H_n^r(\mathbf{b}, \mathbf{c})$ must be a Tchebycheff set. Hence $\varepsilon = f - y_0$ must alternate n+k times on [0,1]. ∎

We shall now present an example which has been chosen to illustrate the use and limitations of the preceding necessary conditions, to show that a best $\| \ \|_\infty$ - approximation

from $V_n^r(\mathbb{R})$ to a given real valued $f \in C[0,1]$ need not be unique when $n \geq 2$, and to demonstrate the existence of local best $\| \ \|_\infty$ - approximations to a given f which fail to be global best $\| \ \|_\infty$ - approximations.

Example 5: Let us define

$$f(t) = \sin(\pi t + \beta), \quad 0 \leq t \leq 1$$

(where β is a real parameter which we shall later specialize) and seek a best $\| \ \|_\infty$ - approximation to f from $V_n^r(\mathbb{R})$ when $n \geq 1$. Let y be a local best $\| \ \|_\infty$ - approximation to f from $V_n^r(\mathbb{R})$ (which exists by virtue of Corollary 1 to Theorem 3), let $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_k$ be the essential exponential parameters of y, and assume that exactly $\ell \leq k$ of these are distinct. Then if we define

$$\varphi = \{(D-\lambda_1)(D-\lambda_2) \ldots (D-\lambda_k)\}\ (f-y)$$

$$= \{(D-\lambda_1)(D-\lambda_2) \ldots (D-\lambda_k)\}\ f$$

we see that φ is a non-trivial linear combination of $\sin(\pi t+\beta)$ and $\cos(\pi t+\beta)$ so that φ has at most one zero in (0,1). Using Lemma 7 we conclude that the error curve $\varepsilon = f-y$ has at most k+1 zeros in (0,1) so that ε alternates at most k+1 times on [0,1]. On the other hand, from Corollary 4 of Theorem 11, we see that ε must alternate at least $n+\ell$ times. Hence from the inequalities $0 \leq \ell \leq k \leq n$

and $n+\ell \leq k+1$ we conclude that either $n=1$ and $k=\ell=0$ (i.e. $y \equiv 0$) or that $k=n$ and $\ell=1$ so that y may be parametrized in the form

$$(85)\quad y(\mathbf{a},\lambda,t) = \{a_0 + a_1 t + \ldots + a_{n-1}t^{n-1}\}\exp(\lambda t)$$

where λ, a_0, a_1, ..., a_{n-1} are real and $a_{n-1} \neq 0$. Thus for all $n \geq 2$ we have a maximum degeneracy in the sense that all of the exponential parameters coalesce. Moreover, since the error curve alternates exactly $n+1$ times with $k=n$ for $n \geq 2$, we conclude from Corollary 5 of Theorem 11 that no local best $\|\ \|_\infty$ - approximation to f from the larger class $V_n^r(\mathbb{C})$ is found in $V_n^r(\mathbb{R})$.

In order to determine $\lambda, \mathbf{a}$ we may proceed as follows. First of all, for each real λ we may uniquely determine $\mathbf{a}(\lambda)$ so that the error curve $f-y(\mathbf{a}(\lambda),\lambda,-)$ alternates at least n times, i.e., so as to satisfy the necessary condition of Corollary 1 of Theorem 11. With $\mathbf{a}(\lambda)$ thus determined, we may then compute

$$\psi(\lambda) = \|f-y(\mathbf{a}(\lambda),\lambda,-)\|_\infty$$

and, by numerically locating the minima of ψ find all of the local best $\|\ \|_\infty$ - approximations to f from $V_n^r(\mathbb{R})$.

When $\beta=0$ we find that

$$y_1(t) = .50000$$

$$y_2^+(t) = |\ .27620 - 0.28396t|\exp(+3.57163t)$$

$$y_2^-(t) = |-.27620+10.10176t|\exp(-3.57163t)$$

$$y_3^+(t) = |\ .20729 - 0.43588t + 0.22906t^2|\exp(+6.07902t)$$

$$y_3^-(t) = |\ .20729 - 9.71230t+100.00945t^2|\exp(-6.07902t)$$

$$\vdots$$

are (to five place accuracy) the local best $\|\ \|_\infty$ - approximations to f from $V_1^r(\mathbb{R})$, $V_2^r(\mathbb{R})$, $V_3^r(\mathbb{R})$,... (with $\|\varepsilon\|_\infty = |y(0)|$ in each case.) We might point out that f is symmetric, i.e. $f(t) = f(1-t)$, and therefore ψ is an even function. In particular, we note that there are two best $\|\ \|_\infty$ - approximations to f from $V_n^r(\mathbb{R})$, n=2,3,... in this case.

When $\beta = .1$, the local best $\|\ \|_\infty$ - approximations shift, and we find that

$$y_1(t) = |\ .60267|\exp(-.40242t), \quad [.50284]$$

$$y_2^+(t) = |\ .35119 - 0.36570t|\exp(+3.18685t), \quad [.25136]$$

$$y_2^-(t) = |-.20642+11.66609t|\exp(-4.01670t), \quad [.30625]$$

$$y_3^+(t) = |\ .27748 - 0.58475t + 0.30757t^2|\exp(+5.53442t), [.17765]$$

$$y_3^-(t) = |\ .34390-12.33185t+128.97369t^2|\exp(-6.69836t), [.24407]$$

$$\vdots$$

are (to five place accuracy) the local best $\|\ \|_\infty$ - approximations to f from $V_1^r(\mathbb{R})$, $V_2^r(\mathbb{R})$, $V_3^r(\mathbb{R})$,... (with $\|\varepsilon\|_\infty$ being given in the brackets in each case.) Since the symmetry of f has been destroyed, we are not surprised to find that some of the best $\|\ \|_\infty$ - approximations corresponding to the $\beta=0$ case have been perturbed into local best $\|\ \|_\infty$ - approximationswhich are not best $\|\ \|_\infty$ - approximations in the present case where $\beta=.1$. □

As we have previously noted, a best $\|\ \|_\infty$ - approximation to a given real valued $f \in C[0,1]$ from $V_1^r(\mathbb{R})$ is always unique. To illustrate just how nonunique a best $\|\ \|_\infty$ - approximation to f from $V_1(\mathbb{C})$ may be we consider the following example (corresponding to the $\beta=0$ case of Example 5.)

<u>Example 6:</u> Let us define

$$f(t) = \sin(\pi t), \quad 0 \leq t \leq 1$$

and seek a best $\|\ \|_\infty$ - approximation to f from $V_1(\mathbb{C})$. (Obviously f is its own best $\|\ \|_\infty$ - approximation from $V_n(\mathbb{C})$ when $n \geq 2$.) Let y be such a best $\|\ \|_\infty$ - approximation and let y be parametrized such that

$$y(t) = a \exp(\lambda t) \quad 0 \leq t \leq 1$$

where $a, \lambda \in \mathbb{C}$ and for convenience in notation we let

$$\gamma = a \exp(\lambda/2)$$

We then find

$$\|f-y\|_\infty \geq \max\{|f(t)-y(t)| : t = 0, 1/2, 1\}$$

$$= \max\{|\gamma|\cdot|\exp(-\lambda/2)|, |1-\gamma|, |\gamma|\cdot|\exp(\lambda/2)|\}$$

$$\geq \max\{|\gamma|, 1-|\gamma|\}$$

$$\geq 1/2$$

with equality holding only if $|\exp(\lambda(2)|=1$ and $\gamma=1/2$, i.e., only if we may reparametrize y in the form

$$y(\beta,t) \equiv (1/2)\exp[i\cdot(2t-1)\beta], \quad 0 \leq t \leq 1$$

for some real parameter β.

Now in the special case where $\rho=0$, we find

$$\|f-y(0,-)\|_\infty = \max\{|\sin(\pi t) - 1/2| : 0 \leq t \leq 1\} = 1/2$$

and hence the error in a best approximation to f has $\|\ \|_\infty$ - norm 1/2. This being the case $y(\beta,-)$ is a best $\|\ \|_\infty$ - approximation to f only if

$$1/4 \geq |f(t)-y(\beta,t)|^2 = \sin^2(\pi t)-\sin(\pi t)\cos[(2t-1)\beta]+1/4$$

i.e., only if

$$\cos[(2t-1)\beta] \geq \sin(\pi t) > 0 , \quad 0 < t < 1$$

i.e., only if $-\pi/2 \le \beta \le \pi/2$. On the other hand, for such β we find

$$|f(t)-y(\beta,t)|^2 = \sin^2(\pi t) - \sin(\pi t)\cos[(2t-1)\beta] + 1/4$$

$$\le \sin^2(\pi t) - \sin(\pi t)\cos[(2t-1)\pi/2] + 1/4$$

$$= 1/4 , \quad 0 \le t \le 1$$

and thus we conclude that y is a best $\| \|_\infty$ - approximation to f from $V_1(\mathbb{C})$ if and only if $y = y(\beta,\cdot)$ for some $\beta \in [-\pi/2, \pi/2]$. ∎

It would be nice if we could prove the $p = \infty$ analogue of Theorem 8 and its corollaries, i.e., if we could show that (under mild hypotheses on S) a local best $\| \|_\infty$ - approximation to f from $W_n(S)$ has full order n. Unfortunately, this is not at all the case as we see from the following simple example.

Example 7: Let us define (the "radical function")

$$f(t) = \begin{cases} |12t - 2| & \text{for} \quad 0 \le t \le 1/3 \\ 1 & \text{for} \quad 1/3 \le t \le 1 \end{cases}$$

and seek a best $\| \|_\infty$ - approximation to f from $N_1(\mathbb{C})$. Indeed, suppose that

$$y(t) = a \exp(\lambda t), \quad 0 \le t \le 1$$

is chosen from $V_1(\mathbb{C})$ and that

$$\|f+y\|_\infty \leq \|f\|_\infty = 1$$

Then since

$$f(0) = -f(1/6) = f(1/3) = 1$$

we must have

$$\text{Re}\{a\} \leq 0, \ \text{Re}\{a \exp(\lambda(6)\} \geq 0, \ \text{Re}\{a \exp(\lambda/3) \leq 0$$

which implies that $a=0$ or $|\text{Im}\{\lambda\}| \geq 3\pi$. On the other hand, since

$$f(t) = 1, \quad 1/3 \leq t \leq 1$$

we must also have

$$\text{Re}\{a \exp(\lambda t)\} \leq 0, \quad 1/3 \leq t \leq 1$$

which implies that either $a=0$ or else $|\text{Im}\{\lambda\}| \leq 3\pi/2$. Hence we conclude that $a=0$ so that $y_o \equiv 0$ is the unique best $\|\ \|_\infty$ - approximation to f from $V_1(\mathbb{C})$. ∎

When $p = \infty$ the basic first order necessary condition is stated in terms of the inequality (84) while when $1 < p < \infty$ this condition is stated in terms of the equality (71). We therefore expect that the $p = \infty$ analogue of the sufficient condition of Theorem 9 will be relatively easy to apply, and this is indeed the case as we see from the following theorem and the accompanying corollaries.

Theorem 12: Let $f \in C[0,1]$, let $y_0 = Y_n(\mathbf{b},\mathbf{c},-)$ be chosen from $V_n(S)$, let y_0 have full order n, and let $\varepsilon = f-y_0$. Let $\mu(\alpha)$ be given by (75) and assume that

$$(86)\quad \underline{\lim}_{\alpha\to 0+} (\|\varepsilon+\alpha h\|_\infty - \|\varepsilon\|_\infty)/\alpha^2 > (1/2)\ \overline{\lim}_{\alpha\to 0+} \mu(\alpha)$$

holds for each $h \in H_n(\mathbf{b},\mathbf{c})$ with $\|h\|_\infty > 0$. Then y_0 is a local best $\|\ \|_\infty$ - approximation to f from $V_n(S)$. Moreover, the result remains valid when $V_n(S)$ and $H_n(\mathbf{b},\mathbf{c})$ are replaced by $V_n^r(S)$ and $H_n^r(\mathbf{b},\mathbf{c})$, respectively, with $\mathbf{b},\mathbf{c}, \in \mathbb{R}^n$ in this case.

Proof: The proof is analogous to that of Theorem 9. □

In some cases the crutial hypothesis (86) of the theorem may be easily satisfied as we see from the following corollaries.

Corollary 1: Let $f \in C[0,1]$, let $y_0 = Y_n(\mathbf{b},\mathbf{c},-)$ be chosen from $V_n(S)$, let y_0 have full order n, and let $\varepsilon = f-y_0$. Assume further that

$$(87)\quad \min\{\mathrm{Re}[\overline{\varepsilon(t)}\ h(t)] : t \in E_\varepsilon\} < 0$$

holds for all $h \in H_n(\mathbf{b},\mathbf{c})$ with $\|h\|_\infty > 0$ (i.e., there is strict inequality in the necessary condition (84).) Then y_0 is a local best $\|\ \|_\infty$ - approximation to f from $V_n(S)$. Moreover, the result remains valid when $V_n(S)$ and $H_n(\mathbf{b},\mathbf{c})$

are replaced by $V_n^r(S)$ and $H_n^r(\mathbf{b},\mathbf{c})$, repsectively, with $\mathbf{b},\mathbf{c} \in \mathbb{R}^n$ in this case.

Proof: Let $h \in H_n(\mathbf{b},\mathbf{c})$ be chosen with $\|h\|_\infty > 0$. Using Lemma 6 together with (87) we find

$$\begin{aligned} &\underline{\lim}_{\alpha \to 0+} \ \alpha^{-2}[\|\varepsilon - \alpha h\|_\infty - \|\varepsilon\|_\infty] \\ &\qquad = \underline{\lim}_{\alpha \to 0+} \ \alpha^{-2}[-\alpha \min\{\mathrm{Re}\ \overline{\varepsilon(t)}\ h(t) : t \in E_\varepsilon\} + o(\alpha)] \\ &\qquad = +\infty \end{aligned}$$

In view of Theorem 1 and (75) we see that $\overline{\lim}\ \mu(\alpha)$ is finite and thus (86) holds for all $h \in H_n(\mathbf{b},\mathbf{c})$ with $\|h\|_\infty > 0$. Hence from the theorem we conclude that y_o is a local best $\|\ \|_\infty$ - approximation to f from $V_n(\mathbf{C})$ and thus from $V_n(S)$ as well. The real case is handled analogously. ∎

Corollary 2: Let $f \in C[0,1]$ be real valued, let $y_o \in V_n^r(S \cap \mathbb{R})$ have order n, and assume that S contains a (complex) neighborhood of each point of $\Lambda[y_o]$. Then a necessary and sufficient condition for y_o to be a local best $\|\ \|_\infty$ - approximation to f from $V_n^r(S)$ is that the error curve $\varepsilon = f - y_o$ alternate at least 2n times on [0,1].

Proof: Let $\mathbf{b},\mathbf{c} \in \mathbb{R}^n$ be chosen so that $y_o = Y_n(\mathbf{b},\mathbf{c},-)$. If ε alternates at least 2n times on [0,1] and if $h \in H_n^r(\mathbf{b},\mathbf{c})$ is chosen such that (87) fails to hold, then h has at least

2n zeros in (0,1) and therefore (since $h \in V_{2n}(\mathbb{R})$) we conclude from Lemma 7 that $\|h\|_\infty = 0$. Hence the sufficiency follows from Corollary 1 above. The necessity follows from Corollary 5 of Theorem 11. []

In the remainder of this section we shall be concerned only with the characterization of a best $\| \|_\infty$ - approximation to f from $V_n^r(\mathbb{R})$. In addition to the zero property of Lemma 8, the exponential sums from $V_n^r(\mathbb{R})$ have an important sign regularity property which is basic to the subsequent analysis.

Lemma 9: Let $\varphi_o \in V_n^r(\mathbb{R})$ have essential exponential parameters $\lambda_1, \lambda_2, \ldots, \lambda_n$, let

$$\varphi_i = (D-\lambda_i)\, \varphi_{i-1}, \quad i=1,2,\ldots, n-1$$

and assume that φ_o has exactly n-1 real zeros (counted according to their multiplicities) which lie in the interval $(-\infty, d]$. Then

(87) $\varphi_i(t)\, \varphi_o(t) > 0, \quad t > d$

holds true for each i=1,2,..., n-1.

Proof: Using Lemmas 7 and 8 we conclude that φ_i has exactly n-i-1 real zeros (counted according to their multiplicites) which lie in $(-\infty, d]$, i=1,2,..., n-1, and thus it is sufficient to establish (87) when i=1.

Now if we define

$$\psi_1(t) = \exp(-\lambda_1 t)\ \varphi_o(t)$$

then we find

$$D\psi_1(t) = \exp(-\lambda_1 t)\ \varphi_1(t)$$

so that ψ_1 and $D\psi_1$ have exactly n-1 and n-2 real zeros, respectively, all of which lie in $(-\infty, d]$. It follows that ψ_1 and $D\psi_1$ have the same algebraic sign for all $t > d$ and hence (87) holds when i=1. □

If the exponential sum $y \in V_n^r(\mathbb{R})$ has order n, we may uniquely define the n component sign vector $\mathbf{s} = (s_1, s_2, \ldots, s_n)$ of y such that

$$s_1 = \lim_{t \to -\infty} \operatorname{sgn}[y(t)]$$

$$s_2 = \lim_{t \to -\infty} \operatorname{sgn}[(D-\lambda_1)y(t)]$$

$$\vdots$$

$$s_n = \lim_{t \to -\infty} \operatorname{sgn}[\{(D-\lambda_1) \ldots (D-\lambda_{n-1})\}\ y(t)]$$

where $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_n$ are the essential exponential parameters of y and

$$\operatorname{sgn}[t] = \begin{cases} -1 & \text{if} \quad t < 0 \\ 0 & \text{if} \quad t = 0 \\ 1 & \text{if} \quad t > 0 \end{cases}$$

is the signum function. Now if y has n-1 real zeros we see by using Lemma 9 that

$$s_i = (-1)^{i-1} s_1, \quad i = 1,2,\ldots, n$$

More generally, it can be shown [3,p.316] that the number of real zeros of the exponential sum y can never exceed the number of sign changes in the sequence $s_1, s_2, \ldots, s_n$. Thus in this sense, Lemma 9 may be regarded as a special case of the "Decartes sign rule" for exponential sums.

Considerable practical experience with the general problem of obtaining a best $\| \|_\infty$ - approximation to a given real valued $f \in C[0,1]$ from a given class of real valued functions leads one to expect that if one has m parameters to adjust then a best error curve will normally alternate m times. For example, if we $\| \|_\infty$ - approximate $f(t) = \exp(t)$, $0 \leq t \leq 1$, by a polynomial of degree at most n-1 (so that we have n linear parameters to adjust) we find that the best error curve alternates exactly n times. (By applying Corollary 1 of Theorem 11 in the case $S = \{0\}$ we see that at least n alternations are required, and since the error function belongs to $V_{n+1}(\mathbb{R})$ we see that at most n alternations can occur.) We do not expect any restrictive conditions to be imposed on the parameters by this normal appearance of the error curve. It is therefore somewhat surprising to see the restriction which

is imposed when we $\| \|_\infty$ - approximate with the 2n parameter family $V_n^r(\mathbb{R})$.

Theorem 13: Let $f \in C[0,1]$ be real valued and assume that for some $n \geq 2$ there is a best $\| \|_\infty$ - approximation to f from $V_{n-1}^r(\mathbb{R})$ which may be written in the form

$$(88)\quad y_{n-1}(t) = a_1\exp(\lambda_1 t)+a_2\exp(\lambda_2 t) + \ldots + a_{n-1}\exp(\lambda_{n-1} t)$$

where each a_i is positive (negative) and where

$$(89)\quad \lambda_1 < \lambda_2 < \ldots < \lambda_{n-1}$$

Assume further that the error curve $\varepsilon = f-y_{n-1}$ alternates exactly 2n-2 times on $[0,1]$ and that the error is positive (negative) at the two points

$$t_- = \min E_\varepsilon$$

$$t_+ = \max E_\varepsilon$$

which may be chosen as the outermost points of the alternant. Then there is a unique best $\| \|_\infty$ - approximation, y_n, to f from $V_n^r(\mathbb{R})$ which may be written in the form

$$(90)\quad y_n(t) = a_1'\exp(\lambda_1' t)+a_2'\exp(\lambda_2' t) + \ldots + a_n'\exp(\lambda_n' t)$$

where each a_i' is positive (negative) and where

$$(91)\quad \lambda_1' < \lambda_1 < \lambda_2' < \lambda_2 < \ldots < \lambda_{n-1} < \lambda_n'$$

(i.e., the exponential parameters of y_{n-1} interlace those of y_n) with the error curve $f-y_n$ alternating at least 2n times on [0,1].

Proof: Since $f-y_{n-1}$ alternates exactly 2n-2 times we may conclude from Corollary 4 of Theorem 11 that y_{n-1} fails to be a best $\|\ \|_\infty$ - approximation to f from $V_n^r(\mathbb{R})$, and thus if y_n is a best $\|\ \|_\infty$ - approximation to f from $V_n^r(\mathbb{R})$ we must have

$$(92)\quad \|f-y_n\|_\infty < \|f-y_{n-1}\|_\infty$$

Hence the exponential sum y_n-y_{n-1} has at least 2n-2 distinct zeros in (t_-,t_+), and this (together with Lemma 8) implies that y_n must have exactly n essential exponential parameters $\lambda_1' \le \lambda_2' \le \ldots \le \lambda_n'$ with

$$(93)\quad \lambda_i \neq \lambda_j' ,\quad i=1,2,..,\ n-1 \quad \text{and } j=1,2,\ldots,\ n$$

Moreover, since we have assumed that $f-y_{n-1}$ is positive (negative) at $t=t_-,\ t_+$ and since the zeros of y_n-y_{n-1} all lie within the interval (t_-,t_+) we see from (92) that y_n-y_{n-1} must also be positive (negative) on $(t_+, +\infty)$. Hence y_n-y_{n-1} satisfies the hypotheses of Lemma 9, and we shall use Lemma 9 to show that y_n may be written in the form required by the theorem.

Indeed, let i, $1 \le i \le n-1$, be chosen and let $\lambda_1^*,\lambda_2^*,\ldots,\ \lambda_m^*$ be the collection of those exponential

parameters from among $\lambda_1,\lambda_2,\ldots,\lambda_{n-1},\lambda_1',\lambda_2',\ldots,\lambda_n'$ which exceed λ_i. Then by using Lemma 9 (and recalling (88),(89),(93)) we find

$$\begin{aligned}\operatorname{sgn}[f(t_+)-y_{n-1}(t_+)] &= \lim_{t\to+\infty}\operatorname{sgn}[y_n(t)-y_{n-1}(t)]\\ &= \lim_{t\to+\infty}\operatorname{sgn}[\prod_{j=1}^{m}(D-\lambda_j^*)\{y_n(t)-y_{n-1}(t)\}]\\ &= \lim_{t\to+\infty}\operatorname{sgn}[-\prod_{j=1}^{m}(\lambda_i-\lambda_j^*)\,a_i]\\ &= (-1)^{m+1}\operatorname{sgn}[a_i]\end{aligned}$$

and since both $f(t_+) - y_{n-1}(t_+)$ and a_i have the same algebraic sign we conclude that m must be odd. Since this is true for each i, we immediately obtain (91) and thus conclude that y_n may be written in the form (90) for some choice of the parameters a_i'. This being the case, we see from Corollary 4 of Theorem 11 that $\varepsilon-y_n$ must alternate at least 2n times on [0,1]. Finally, we may again use Lemma 9 to obtain

$$\begin{aligned}\operatorname{sgn}[f(t_+)-y_{n-1}(t_+)] &= \lim_{t\to+\infty}\operatorname{sgn}[\prod_{j=i}^{n}(D-\lambda_j)(D-\lambda_{j+1}')\{y_n(t)-y_{n-1}(t)\}]\\ &= \operatorname{sgn}[a_i'],\quad i=1,2,\ldots,n\end{aligned}$$

so that each a_i' must be positive (negative). ∎

The result of Theorem 13 is contained implicitly in a theorem of Braess [3,p.318] which provides a complete characterization of best $\| \ \|_\infty$ - approximations to f from the class $V_n^+(\mathbb{R})$ of all exponential sums expressible in the form

$$y(t) = a_1\exp(\lambda_1 t) + a_1\exp(\lambda_2 t) + \ldots + a_n\exp(\lambda_n t)$$

with $\lambda_1 < \lambda_2 < \ldots < \lambda_n$ and $a_i \geq 0$ for each i. The above proof of Theorem 13 is based upon Lemma 9 which is somewhat simpler to state and prove than the corresponding "Decartes sign rule" for exponential sums which is used by Braess.

In order to apply Theorem 13 one must have some means to verify that the error curve $\varepsilon=f-y_{n-1}$ cannot alternate more than 2n-2 times on [0,1]. Perhaps the simplest way to do this is to show that f, f', ..., $f^{(n-1)}$ form a Tchebycheff system on [0,1] (if this is possible) and then to use Lemma 7 in the natural manner. An alternative procedure is suggested by the following lemma.

Lemma 10: Let $y \in V_n^r(\mathbb{R})$, let $f \in C^{2n}[0,1]$ be real valued, let

$$(94)\quad \Delta_\nu[f,t] = \det\left|D^{i+j}f(t)\right|_{0 \leq i,j \leq \nu}$$

be the $(\nu+1)$-st order Hankel determinant associated with f, and assume that

$$(95)\quad \Delta_\nu[f,t]\cdot\Delta_0[f,t] > 0,\ \nu=1,2,\ldots,n,\quad 0 \leq t \leq 1$$

Then f-y has at most 2n zeros in [0,1] and, in particular, f-y alternates at most 2n times on [0,1].

Proof: Let $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbb{R}^n$ be chosen so that

$$\{(D-\lambda_1)(D-\lambda_2) \ldots (D-\lambda_n)\}y = 0$$

let $\mathbf{c} = (c_1, c_2, \ldots, c_n) = \sigma_n(\boldsymbol{\lambda})$, and let $c_0 = 1$. If f-y has $m \geq 2n$ zeros in [0,1] then from Lemma 7 (and the choice of $\mathbf{c}$) we see that

$$(96) \quad \sum_{i=0}^{n} \sum_{j=0}^{n} c_i \, c_j \, D^{i+j} \, f(t) = \{P_n(\mathbf{c},D)\}^2 \, f(t)$$

$$= \{P_n(\mathbf{c},D)\}^2 \, [f(t) - y(t)]$$

has at least $m-2n \geq 0$ zeros in [0,1]. But from the hypothesis (95) we see that the quadratic form associated with the determinant $\Delta_n[f,t]$ must be definite, and since $c_0 = 1$ this implies that neither member of (96) can vanish for any $t \in [0,1]$. Hence $m = 2n$ and the proof is complete. □

A somewhat stronger form of Lemma 10 (in which the relation > in the hypothesis (95) is relaxed to $\neq$) is given by Meinardus [15, p.179], with the accompanying proof being a bit more complicated than that required here.

In concluding this section we shall present the following example which illustrates the use of Theorem 13 as well as a number of the previous results.

Example 8: Let us define

$$(97)\quad f(t) = (t+a)^{-\alpha} \quad , \quad 0 \le t \le 1$$

where $a > 0$, $\alpha > 0$ and seek to characterize a best $\| \|_\infty$ - approximation to f from $V_n^r(\mathbb{R})$. (The special case when $\alpha=1$ has been considered by Braess [5,p.270], but the methods used do not apply when $\alpha \neq 1$.) We shall use an inductive argument to show that for each n=1,2,... there is a unique best $\| \|_\infty$ - approximation to f from $V_n^r(\mathbb{R})$ which may be parametrized in the form

$$y_n(t) = a_{n1}\exp(\lambda_{n1}t)+a_{n2}\exp(\lambda_{n2}t)+\ldots+a_{nn}\exp(\lambda_{nn}t)$$

$$(98)\quad a_{ni} > 0 \ , \qquad i=1,2,\ldots, n$$

$$\lambda_{n1} < \lambda_{n2} < \ldots < \lambda_{nn}$$

This assertion evidently holds for n=1. Assume then that $y_1, y_2, \ldots, y_{n-1}$ all have the desired form for some $n \ge 2$. From Corollary 4 to Theorem 11 we see that $f-y_{n-1}$ must alternate at least 2n-2 times on [0,1] so that $f-y_{n-1}$ has at least 2n-2 distinct zeros in (0,1). On the other hand, if we extend the definition (97) of f to the larger interval $(-a, +\infty)$ and note that

$$\{(D-\lambda_{n-1,1})(D-\lambda_{n-1,2}) \ldots (D-\lambda_{n-1,n-1})\}[f(t)-y(t)]$$

$$= (a+t)^{-\alpha-n+1}p_{n-1}(t)$$

for some polynomial p_{n-1} of degree n-1, we may then use Lemma 7 to conclude that $f-y_{n-1}$ has at most 2n-2 zeros on $(-a, +\infty)$. (Alternatively, in this case we may explicitly evaluate (94) with

$$\Delta_n[f,t] = (t+a)^{-(n+1)(\alpha+n)} \cdot |(-\alpha)_0(-\alpha)_1 \ldots (-\alpha)_n| \cdot 0!\ 1!\ \ldots n!$$

where

$$(s)_k = \begin{cases} 1 & \text{if} \quad k = 0 \\ s(s-1) \ \ldots \ (s-k+1) & \text{if } k = 1,2,\ldots \end{cases}$$

so that this result can also be obtained from Lemma 10). Hence we find that $f-y_{n-1}$ alternates exactly 2n-2 times on $[0,1]$ with

$$\begin{aligned} \operatorname{sgn}[f(1)-y_{n-1}(1)] &= \operatorname{sgn}[f(0)-y_{n-1}(0)] \\ &= \lim_{t \to -a} \operatorname{sgn}[f(t)-y_{n-1}(t)] \\ &= +1 \end{aligned}$$

The hypotheses of Theorem 13 being satisfied, we may now conclude that there is a unique best $\| \ \|_\infty$ - approximation to f from $V_n^r(\mathbb{R})$ which may be written in the form (98), and thus the induction is complete.

We see in this case that y_n is completely nondegenerate in the sense that the n essential exponential parameters are all distinct. (This is to be contrasted to the completely degenerate situation found in Example 5 where all

of the exponential parameters were found to coalesce.) This lack of degeneracy allows us to apply Corollary 2 of Theorem 12 and thereby conclude that y_n is actually a local best $\| \ \|_\infty$ - approximation to f from $V_n^r(\mathbf{C})$. We conjecture that y_n is actually a global best $\| \ \|_\infty$ - approximation to f from $V_n^r(\mathbf{C})$ but this has not been proved.

From the previous discussion we see that $f(t) > y_n(t)$ must hold for all $t \geq 1$, and in view of (97)-(98) we easily see that the exponential parameters of y_n must all be negative. By making use of a modification of a result of Pólya and Szegő [16,p.47 #65] it can be shown that when the n roots of the polynomial

$$p_n(t) = (t+a)^{\alpha+n} \{(D-\lambda_1) \ \ldots \ (D-\lambda_n)\} f(t)$$

all lie in the interval [0,1] then $\lambda_1, \lambda_2, \ldots, \lambda_n$ are distinct negative real numbers which may be ordered such that

$$\lambda^*_{ni}/a \leq \lambda_i \leq \lambda^*_{ni}/(1+a) \ , \quad i=1,2,\ldots, n$$

where $\lambda^*_{n1} < \lambda^*_{n2} < \ldots < \lambda^*_{nn}$ are the roots of the polynomial

$$(99) \quad p^*_n(\lambda) = \lambda^n + (n)_1(n+\alpha-1)_1\lambda^{n-1}/1! + (n)_2(n+\alpha-1)_2\lambda^{n-2}/2!$$

$$+ \ \ldots \ + (n)_n(n+\alpha-1)_n/n!$$

In particular, we may relax these bounds somewhat to show that

$$-n(n+\alpha-1)/a < \lambda_{n1} < \lambda_{n2} < \ldots < \lambda_{nn} < -\alpha/[n(a+1)]$$

and that

$$\lambda_{n1} < -(n+\alpha-1)/(a+1)$$

when λ_{ni} are the exponential parameters of y_n. In particular, this last inequality shows that $\lambda_{n1} \rightarrow -\infty$ as $n \rightarrow +\infty$. Numerical values for the parameters a_{ni}, λ_{ni} will be presented (for selected values of n, a, and α) in Table 2 at the end of the next section after we discuss procedures which may be used to compute them.

VI. CONSTRUCTION OF GOOD AND BEST APPROXIMATIONS

We turn now to the difficult problem of constructing a best $\| \ \|_p$ - approximation to a given f from a given set $V_n(S)$ or $V_n^r(S)$, (a problem which Rice [21,p.v] refers to as "one of the more challenging problems in computational methods of approximation!") For practical purposes, we are actually interested only in the special cases where f is real valued and relatively smooth, where the approximating functions are real valued, where S is either $\mathbb{R}$, $\mathbb{C}$, a half plane $\mathbb{L}_\alpha$, or a disc $\mathbb{D}_\alpha$, where p=2 or ∞, and where n is relatively small, i.e. $n \leq 20$. Unfortunately, even under these restrictions there is no presently known method which will with certainty produce a best $\| \ \|_p$ - approximation to a given f from $V_n(S)$.

There are essentially two major reasons why this computational problem is so intractable. The first, of course, arises from the fact that, in general, a given f may have a multiplicity of best or local best $\| \ \|_p$ - approximations from $V_n(S)$. This being the case, we can at best hope to find some procedure which will enable us to select some $y_1 \in V_n(S)$ which is a reasonably good $\| \ \|_p$ - approximation to f and then successively descend to y_2, y_3, ... in such a manner that the sequence $\{y_\nu\}$ has a limit which is a local best $\| \ \|_p$ - approximation to f from $V_n(S)$.

The second troublesome feature occurs when p=1 or ∞ since when this is the case a local best $\|\ \|_p$ - approximation to f may very well lie in $V_{n-1}(S)$. This being the case, the sequence $\{y_\nu\}$ which we construct need not be a V-sequence, and consequently we are forced to compute $Y_n(\mathbf{b},\mathbf{c},-)$ when some of its components are highly oscillatory or decay rapidly away from the endpointsof the interval [0,1]. At first thought one might assume that such terms could be quickly dropped from the computation, but the actual decision as to when this can reasonably be done is not an easy one to make since even in the most ideal situations (cf., Example 8) extremely large exponential parameters can and do occur. Thus one is forced to design procedures for the accurate computation of $Y_n(\mathbf{b},\mathbf{c},-)$ when large exponential parameters may be present, and this becomes increasingly unreasonable for n=4,5,... . (Even though there are practical means for accurately computing the solutions to such "stiff" ordinary differential equations, cf.,[8], the computational cost of doing so rapidly becomes prohibitive aseither n or the size of the underlying exponential parameters becomes large.)

In view of these difficulties, it is not at all surprising that very little is actually known about the computational problem. Rice [20] and Thorpe and Lewine [24] give a few suggestions for the case p=1. For the

special parametric form $a \cdot \exp(\lambda t) + b$ Rice [21,p.69-73] gives a complete discussion of a procedure to compute best $\| \ \|_\infty$ - approximations. Meinardus [15,p.181] suggests the use of the Newton iteration procedure for computing best $\| \ \|_\infty$ - approximations to f from $V_n^r(\mathbf{R})$, and for the same problem Braess [5] discusses an alternative procedure which may at times (e.g., n=1,2,3,4) be used successfully.

In this section we shall show how some of the special features of this particular nonlinear approximating problem lead to three procedures which may be used to generate an initial estimate of a best $\| \ \|_p$ - approximation to a given real valued f from $V_n^r(S)$ and which are relatively simple to use when p=2 or ∞ and when the exponential parameters are unconstrained, i.e., $S = \mathbf{C}$. We shall then describe a rather general iterative process which may often be used to generate a $\| \ \|_\infty$ - minimizing sequence in the case where the error curve is real valued and does not oscillate wildly. Finally, we shall describe a highly specialized iterative process which only applies to $V_n^r(\mathbf{R})$. Numerical results will be given to illustrate these processes (which we believe to be of genuine practical importance) but we shall not present an analysis of conditions under which convergence is guaranteed.

Although we know from Theorem 4 that $V_\infty(S)$ is dense in $L_p[0,1]$ (provided $S \neq \emptyset$) this tells us nothing about

how well we can $\| \|_p$ - approximate a given $f \in L_p[0,1]$ by elements of a given $V_n(S)$. We would naturally expect a best $\| \|_p$ - approximation to do a fairly good job of representing f only if f is "almost" in the class $V_n(S)$, i.e., only if f may be regarded as the solution of some n-th order linear inhomogeneous differential equation with constant coefficients where the auxiliary polynomial has roots which lie in S and where the inhomogenity (i.e., driving function) is small relative to the normal growth of solutions of the corresponding homogeneous differential equation. These ideas may be made precise in the sense of the following theorem.

<u>Theorem 14</u>: Let $f \in C^n[0,1]$, let $\mathbf{c} \in \mathbf{C}^n$, and let

$$(100)\quad \eta_n[f,\mathbf{c},t] = |D^n+c_1D^{n-1}+\ldots+c_{n-1}D+c_n|f(t),\ 0 \le t \le 1$$

Let $\lambda_1,\lambda_2,\ldots,\lambda_n$ be the roots of the auxiliary polynomial $P_n(\mathbf{c},-)$ and let

$$(101)\quad \rho_n(\mathbf{c}) = \max\{\mathrm{Re}\ [\lambda_i] : 1 \le i \le n\}$$

Finally, let y_o be a best $\| \|_p$ - approximation to f from the linear space $L_n(\mathbf{c})$ (given by (9)). Then

$$(102)\quad \|f-y_o\|_p \le \inf_q\ n^{1/q}\ \|\eta_n[f,\mathbf{c},-]\|_q \cdot \exp[\rho_n(\mathbf{c})]/n!$$

where the infinum is taken for $1 \le q < \infty$.

Proof: If we define y such that

$$(103)\quad f(t_0)-y(t_0) = \int_0^{t_0}\int_0^{t_1}\cdots\int_0^{t_{n-1}} \exp\Big\{\sum_{j=1}^{n}(t_{j-1}-t_j)\lambda_j\Big\}\cdot \eta_n[f,\mathbf{c},t_n]\,dt_1dt_2\ldots dt_n$$

then we may use (100) to verify that

$$P_n(\mathbf{c},D)\,y = 0$$

i.e., $y \in L_n(\mathbf{c})$. Using this together with (101) and (103) we now find

$$\|f-y_o\|_p \le \|f-y\|_\infty$$

$$\le \int_0^{1}\int_0^{t_1}\cdots\int_0^{t_{n-1}} \exp\{(1-t_n)\rho_n(\mathbf{c})\}\,|\eta_n[f,\mathbf{c},t_n]|\,dt_1dt_2\ldots dt_n$$

$$\le \Big\{\int_0^1 (1-t)^{n-1}\,|\eta_n[f,\mathbf{c},t]|\,dt\Big\}\exp[\rho_n(\mathbf{c})]/(n-1)!$$

from which we obtain (102) by using the Hölder inequality. □

Corollary 1: Let $f \in C^n[0,1]$, let $1 \le p \le \infty$, and let y_o be a best $\|\ \|_p$ - approximation to f from $V_n(S)$. Then

$$(104)\quad \|f-y_o\|_p \le \inf_{\mathbf{c}}\ \inf_{q}\ n^{1/q}\,\|\eta_n[f,\mathbf{c},-]\|_q\ \exp[\rho_n(\mathbf{c})]/n!$$

where the first infinum is taken over the set of $\mathbf{c} \in \mathbb{C}^n$ for which $\Lambda_n(\mathbf{c}) \subseteq S$ and the second is taken for $1 \leq q < \infty$. □

Although the bound (104) is extremely crude, in some ideal situations it may be used to obtain a simple a priori bound on the error. For example, if f is given by (97) we may choose $\mathbf{c} \in \mathbb{R}^n$ such that

$$\eta_n[f,\mathbf{c},t]$$

$$= (-\alpha)_n(t+\alpha)^{-\alpha-n} + c_1(-\alpha)_{n-1}(t+a)^{-\alpha-n+1} +\ldots+ c_n(t+a)^{-\alpha}$$

$$= (-\alpha)_n(t+a)^{-\alpha} \cdot T_n[-(1+2a)+2a(a+1)/(t+a)]2^{1-n}[2a(a+1)]^{-n}$$

where

$$T_n(x) = \cos\{n \cos^{-1} x\}, \quad -1 \leq x \leq 1$$

is the n-th order Tchebycheff polynomial. This being the case, $\eta_n[f,\mathbf{c},-]$ has n zeros in (0,1) and from the discussion at the end of Example 8 (p.101) it follows that $\Lambda_n(\mathbf{c}) \subset (-\infty, 0)$ so that $\rho_n(\mathbf{c}) \leq 0$. Thus if y_n is the best $\|\ \|_\infty$ - approximation to f from $V_n^r(\mathbb{R})$ we see from (104) that

$$\|f-y_n\|_\infty \leq 2a^{-\alpha}[4a(a+1)]^{-n}\ |(-\alpha)_n/n!|$$

$$\leq 2a^{-\alpha}\ [4a(a+1)]^{-n}\ \exp[(\alpha-1)\sum_{k=1}^{n} 1/k]$$

$$= 0\ (\ n^{\alpha-1}\ [4a(a+1)]^{-n})$$

as $n \to \infty$. It should be pointed out that this bound is never sharp and indeed is completely useless unless $a > (\sqrt{2}-1)/2$ in which case $4a(a+1) > 1$.

Corollary 2: Let $f \in C^n[0,1]$ be real valued and let y be the polynomial which best $\| \ \|_p$ - approximates f from the class $V_n^r(\{0\})$ of all polynomials of degree at most $n-1$. Then

(105) $\|f-y\|_p \leq \|f^{(n)}\|_\infty / n!$ □

The result of Corollary 2 is quite familiar. Indeed if y is a best $\| \ \|_p$ - approximation to f from $V_n^r(\{0\})$ then y must interpolate f in at least $n-1$ distinct points. Thus (105) follows from the well known formula for the error in polynomial interpolation (with the usual polynomial factor missing since we are working on the interval $[0,1]$.) We may thus view (102) as an appropriate generalization of (105).

In spite of the crude bounds used in the derivation of Theorem 14, the result is of genuine interest since it suggests a relatively simple means for actually constructing a good $\| \ \|_p$ - approximation to a given $f \in C^n[0,1]$ from a given $V_n(S)$. Indeed, the main idea of the theorem is that if $\mathbf{c}$ is chosen so that $\eta_n[f,\mathbf{c},-]$ is small in some sense, then the corresponding best $\| \ \|_p$ - approximation to f from the n dimensional space $L_n(\mathbf{c})$ will also be small. By disregarding the troublesome nonlinear factor $\exp[\rho_n(\mathbf{c})]$

appearing in (102) and by choosing to measure the smallness of η_n in the $\|\ \|_2$ - norm we are led to the following process for constructing an initial approximation y.

Procedure 1: Let $f \in C^n[0,1]$ be real valued, let $1 \leq p \leq \infty$, and let S be a closed nonempty subset of $\mathbf{C}$.

(i): Determine $\mathbf{c} \in \{\mathbf{c}^* \in \mathbf{R}^n : \Lambda_n(\mathbf{c}^*) \subseteq S\}$ in such a manner that $\|P_n(\mathbf{c},D)f\|_2$ is minimized.

(ii): With $\mathbf{c}$ thus chosen, determine $\mathbf{b} \in \mathbf{R}^n$ in such a manner that $\|f-Y_n(\mathbf{b},\mathbf{c},-)\|_p$ is minimized.

(iii): Take $y=Y_n(\mathbf{b},\mathbf{c},-)$ as an initial $\|\ \|_p$- approximation to f from $V_n^r(S)$. []

In order to carry out step (i) we must find the best $\|\ \|_2$ - approximation to $D^n f$ from some closed subset of the linear space spanned by f, Df, ..., $D^{n-1}f$. In general, $\mathbf{c}$ is not uniquely determined, but when $S = \mathbf{C}$ we may always find $\mathbf{c}$ by solving the system of linear equations

$$(106)\quad \sum_{i=1}^{n} c_i \int_0^1 [D^{n-i}f][D^{j-1}f]dt = -\int_0^1 [D^n f][D^{j-1}f]dt, j=1,2,\ldots,n$$

Now if $f \notin V_{n-1}^r(\mathbf{C})$, then $f,Df,\ldots,D^{n-1}f$ are linearly independent and the coefficient matrix of (106) is non-singular so that $c_1,c_2,\ldots,c_n$ are uniquely determined. On the other hand, if $f \in V_{n-1}^r(\mathbf{C})$ then Procedure 1 must always

produce y=f (since the minima discussed in (i) and (ii) both vanish in this case) and by appropriately reducing the order of (106) we may uniquely determine the coefficients of the minimal polynomial associated with f.

The system (106) may be very poorly conditioned and for computational purposes we may choose to work with a somewhat different basis, especially when the integrals appearing in (106) can be expressed in closed form. For example, we might choose to determine $\mathbf{c}^*$ by solving the linear system

$$(107)\quad \sum_{i=1}^{n} c_i^* \int_0^1 [T_{n-i}(D)f][T_{j-1}(D)f]dt = -\int_0^1 [D^n f][T_{j-1}(D)f]dt, \quad j=1,2,\ldots,n$$

(where T_j is the Tchebycheff polynomial of degree j) and then use $\mathbf{c}^*$ to find $\mathbf{c}$ such that

$$(108)\quad P_n(\mathbf{c},t) = t^n + c_1^* T_{n-1}(t) + c_2^* T_{n-2}(t) + \ldots + c_n^* T_0(t)$$

We are not particularly interested in the exact computation of $\mathbf{c}$ at this step, however, since the exponential sum y produced by the procedure will usually be regarded only as a reasonable "first guess" which may then be subjected to iterative improvement.

Having found $\mathbf{c}$, we must now solve the linear approximation problem of (ii). In so doing we first determine a suitable basis for $L_n(\mathbf{c})$ (which in practice requires either the numerical determination of the roots of the auxiliary polynomial associated with $\mathbf{c}$ or else the numerical determination of n linearly independent solutions of the associated n-th order homogeneous differential equation, perhaps only on a finite point set.) Using this basis, we then solve the linear $\| \ \|_p$- approximation problem of (ii). (When p=2 this again involves the solution of a system of linear equations and when $p = \infty$ this may be done by using the iterative process to be described in Procedure 4.)

When $1 < p < \infty$, the $\| \ \|_p$ - norm is strictly convex so that a unique y is determined. However, when p=1, ∞ the $\| \ \|_p$- norm is no longer strictly convex and Procedure 1 does not, in general, give rise to a unique $y \in V_n(S)$. Nevertheless, if it so happens in this case that $\Lambda_n(\mathbf{c}) \subset \mathbf{R}$, then it follows from Lemma 8 that any basis of $L_n(\mathbf{c})$ is a Tchebycheff set and therefore (cf., [19,p.61, 109]) there is a unique best $\| \ \|_p$- approximation to f from $L_n(\mathbf{c})$ even when p=1, ∞. In summary then, Procedure 1 gives rise to a unique y provided $f \in V_n(S)$ (in which case y=f), or else $\Lambda_n(\mathbf{c}) \subset S$ for the $\mathbf{c}$ satisfying (106) and either $1 < p < \infty$ or else $\Lambda_n(\mathbf{c}) \subset \mathbf{R}$.

In situations where f fails to have n continuous derivatives, where the derivatives of f are inconvenient to compute, or where the integrals appearing in (106) are difficult to evaluate Procedure 1 does not provide a satisfactory method for producing the initial approximation y. Consequently we shall replace the differential operator D appearing in Procedure 1-(i) with the discrete analogue D_h which (for $h < 1$) is defined in such a manner that

$$D_h f(t) = [f(t+h) - f(t)]/h \ , \quad 0 \leq t \leq 1-h$$

and thereby devise the following alternative process.

Procedure 2: Let the real valued function f be defined on $[0,1]$, let $1 \leq p \leq \infty$, and let S be a closed nonempty subset of $\mathbf{C}$.

(i): Choose $N \geq 2n-1$ and set $h = 1/N$.

(ii): Determine $\mathbf{c} \in \{\mathbf{c}^* \in \mathbf{R}^n : \Lambda_n(\mathbf{c}^*) \subseteq S\}$ in such a manner that

$$(109) \quad \sum_{j=0}^{N-n} |P_n(\mathbf{c}, D_h)\ f(jh)|^2 = \text{Minimum}$$

(iii): With $\mathbf{c}$ thus chosen, determine $\mathbf{b} \in \mathbf{R}^n$ in such a manner that $\|f - Y_n(\mathbf{b}, \mathbf{c}, -)\|_p$ is minimized.

(iv): Take $y = Y_n(\mathbf{b}, \mathbf{c}, -)$ as an initial $\| \ \|_p$- approximation to f from $V_n^r(S)$. □

When $S = \mathbf{C}$ the components of the solution $\mathbf{c}$ of the minimization problem (109) must satisfy the normal equations

$$(110)\quad \sum_{i=1}^{n} c_i \{h \sum_{m=0}^{N-n} [D_h^{n-i} f(mh)][D_h^{j-1} f(mh)]\}$$

$$= - h \sum_{m=0}^{N-n} [D_h^{n} f(mh)][D_h^{j-1} f(mh)], \quad j=1,2,\ldots, n$$

which are to be compared with (106). (This system of equations may also be poorly conditioned and although the discrete analogue of (107) might be helpful, in actual practice it is usually better to avoid the use of the normal equations altogether, cf. [9].) Now if we may further assume that $f \in C^n[0,1]$, then we have

$$\lim_{h \to 0+} \|D^k f - D_h^k f\|_\infty = 0, \quad k=0,1,\ldots, n$$

so that the equations (106) may be regarded as the limiting form of the equations (110) as $N \to \infty$ and $h=1/N \to 0$. Indeed, if the system (106) is nonsingular then for all sufficiently small h the same is true of the system (110) so that if $\mathbf{c}(h)$ denotes the solution of (110) then $\lim \mathbf{c}(h)$ is the corresponding solution of (106).

Needless to say, there are numerous variations of Procedure 2 which one might consider. For example, if $S = \mathbf{C}$

one might determine $\mathbf{c}$ using (109) and then choose y such that y satisfies the n-th order difference equation

$$(111)\quad P_n(\mathbf{c}, D_h)\, y(t) = 0\,, \qquad -\infty < t < \infty$$

subject to the constraint that

$$\sum_{m=0}^{N} |f(mh) - y(mh)|^2 = \text{Minimum}$$

(noting that every solution of (111) must belong to $V_n(\mathbf{C})$.) For large N this is essentially the same as Procedure 2 (with p=2). On the other hand, when N=2n-1 (the smallest possible value) then this modified scheme reduces to the familiar method of Prony (cf. [11,p.378-382]) with y interpolating f at 2n equally spaced points.

There is one variation of Prony's method which can sometimes be quite useful for providing an initial $\|\ \|_\infty$-approximation to y which may then be improved by some iterative process. We shall continue to assume that f is real valued, and we shall seek some $y \in V_n^r(\mathbf{C})$ and some $d \in \mathbf{R}$ such that

$$(112)\quad y(t_i) = f(t_i) + (-1)^i\, d, \qquad i=0,1,2,\ldots,\ 2n$$

where $t_0, t_1, \ldots, t_{2n}$ are equally spaced points from [0,1], e.g., $t_i = i/2n$. Since $V_n^r(\mathbf{C})$ is a 2n parameter family, it is quite natural to attempt to solve the system of equations

(112) since systems of this form (but with nonuniformly spaced t_i) are quite commonly encountered in the use of the Remez algorithm for solving $\| \ \|_\infty$ - approximation problems.

If such a pair y, d exists, then y must satisfy some difference equation of the form (111) (since every member of $V_n^r(\mathbf{C})$ satisfies such a difference equation) so that there exist real numbers $\gamma_1, \gamma_2, \ldots, \gamma_n$ such that

$$(113)\quad y(t_{i+n}) + \gamma_1 y(t_{i+n-1}) + \ldots + \gamma_n y(t_i) = 0, \quad i=0,1,\ldots,n$$

Upon comparing (112) - (113) we find that the coefficients $\gamma_1, \gamma_2, \ldots, \gamma_n$ must satisfy the system of equations

$$(114)\quad [f(t_{i+n})+f(t_{i+n-1})]+\gamma_1[f(t_{i+n-1})+f(t_{i+n-2})] + \ldots + \gamma_n[f(t_i)+f(t_{i-1})] = 0, \quad i=1,2,\ldots, n$$

which serves to determine them. In the event that the system (114) is singular, i.e. if $\{f(t_i)+f(t_{i-1})\}$ satisfies some linear difference equation of order less than n, then we may determine $\gamma_1, \gamma_2, \ldots, \gamma_n$ so that (114) holds subject to the additional constraint that the auxiliary polynomial

$$(115)\quad r(\mu) = \mu^n + \gamma_1 \mu^{n-1} + \ldots + \gamma_{n-1}\, \mu + \gamma_n$$

have as many zero roots as possible.

Now provided $\mu = -1$ is not a root of the auxiliary polynomial (115), we may find a unique solution y (defined only at $t_0, t_1, \ldots, t_{2n}$) of the n-th order difference equation (113) and a unique $d \in \mathbf{R}$ such that (112) holds at the n+1 points $t_0, t_1, \ldots, t_n$. This being the case, we may use (113) - (114) to see that (112) also holds at $t_{n+1}, t_{n+2}, \ldots, t_{2n}$. Finally, we extend y to all of [0,1] in such a manner that $y \in V_n^r(\mathbf{C})$ and thus obtain the desired pair y,d.

When $\mu = -1$ is a root of (115), we cannot in general find a solution to the system (112). For example, when n=1, when $t_i = i/2$, i=0,1,2, and when

$$f(t) = t \cdot \cos(2\pi t) , \qquad 0 \leq t \leq 1$$

we find that (114) becomes $(1/2) - (1/2)\,\gamma_1 = 0$ so that $r(\mu) = \mu+1$. Hence $y(t_i) = a\cdot(-1)^i$, i=0,1,2, is the only possible solution of (113), and this being the case we cannot find constants a,d so that (112) holds for i=0,1.2. An alternative derivation of the equations (114) is given by Meinardus [15,p.181-183] (although no mention is made there of the unpleasant fact that the system (114) may have no solution.) We may summarize the above discussion as follows.

Procedure 3: Let the real valued function f be defined on [0,1] and let S = **C**.

(i): Choose 2n+1 equally spaced points $t_0 < t_1 < \ldots t_{2n}$ from [0,1].

(ii): Determine $\gamma_1, \gamma_2, \ldots, \gamma_n$ so that the system (114) is satisfied and so that the auxiliary polynomial (115) has as many zero roots as possible subject to this constraint.

(iii): Find the solution y of the difference equation (113) and a real number d such that (112) holds for i=0,1,..., n (and thus for i=n+1, n+2,...,2n as well.) If $\mu = -1$ is a root of the auxiliary polynomial (115) this will not, in general, be possible and the procedure fails.

(iv): Extend the definition of y to all of [0,1] in such a manner that $y \in V_n^r(\mathbb{C})$, and take y as an initial approximation to f. □

We might point out that, in general, Procedure 3-(iv) cannot be carried out in a unique manner. For example, each of the functions

$$y_n(t) = \sin[(1+6n\pi)t], \quad -\infty < t < \infty, \quad n=0,1,\ldots$$

satisfies the initial value problem

$$y(t+2/3) - y(t+1/3) + y(t) = 0, \quad -\infty < t < \infty$$

$$y(0) = 0, \quad y(1/3) = \sin(\pi/3)$$

This being the case, we cannot expect Procedure 3 (or any other method based upon the solution of a difference equation initial value problem, e.g., Prony's method) to duplicate an arbitrary $f \in V_n^r(\mathbf{C})$ in the way that Procedure 1 does.

By using one of the Procedures 1-3, it is possible to construct a quite reasonable initial approximation to a given f. We shall now describe one means for simultaneously perturbing the parameters of a good approximation in such a manner as to produce a slightly better one, thereby making possible the generation of a sequence $\{y_\nu\}$ of good approximations which (hopefully) converge to a best (or at least a local best) $\| \ \|_\infty$ - approximation to the given f.

In describing the iterative process we shall assume that the error curve $\varepsilon=f-y$ has been parametrized by the real vector $\mathbf{q} = (q_1, q_2, \ldots, q_m)$. For example, if we seek a best approximation to f from $V_n^r(\mathbf{C})$ we will, in general, choose $\mathbf{q} = (\mathbf{b}, \mathbf{c})$ (so that $\varepsilon=f - Y_n(\mathbf{b}, \mathbf{c}, -)$) while if we seek a best approximation from $V_n^r(\mathbf{R})$ it is usually more convenient to use $\mathbf{q} = (b, \boldsymbol{\lambda})$ (so that $\varepsilon=f - Y_n(\mathbf{b}, \sigma_n(\boldsymbol{\lambda}), -)$) with m=2n in either case. On the other hand, if we have some a priori knowledge about the form of the best approximation which we seek, we will usually not wish to use these more general parametric forms. Thus in Example 5 where

the exponential parameters are known to coalesce we would use the n+1 dimensional vector $\mathbf{q} = (\mathbf{a}, \lambda)$ corresponding to the parametric form (85) while in Example 8 we would choose to use the 2n dimensional vector $\mathbf{q} = (\mathbf{a}, \mathbb{A})$ corresponding to the very convenient parametric form (98). In an ideal situation the error curve corresponding to a best $\| \|_\infty$ - approximation to f will alternate at least m times, and sometimes by carefully choosing the parametrization (as in the above two special examples) one can insure that this ideal situation will prevail.

As a first step in carrying out the procedure we locate the abscissa of each local extremum of the error curve $\varepsilon(\mathbf{q}, t)$, $0 \leq t \leq 1$, i.e., we find each point $\tau \in [0,1]$ with the property that $\varepsilon(\mathbf{q}, t) - \varepsilon(\mathbf{q}, \tau)$ does not change algebraic sign as t ranges over some suitably restricted neighborhood of τ in $[0,1]$. In most problems of practical importance there will only be a finite number of these local extrema, and we shall not consider the situation where this is not the case. We shall let M denote the number of these points which we index and order such that $0 = \tau_1 < \tau_2 < \ldots < \tau_M = 1$. (The interval endpoints 0,1 always correspond to local extrema so that $M \geq 2$.)

Using $\tau_1, \tau_2, \ldots, \tau_M$ we now define the average extremum

$$(116)\quad \bar{\varepsilon} = (1/M) \sum_{i=1}^{M} |\varepsilon(\mathbf{q},\tau_i)|$$

and we define the two M dimensional vectors $\boldsymbol{\varepsilon}$, $\mathbf{s}$ with components

$$(117)\quad \varepsilon_i = \varepsilon(\mathbf{q},\tau_i), \quad i=1,2,\ldots, M$$

$$(118)\quad s_i = \begin{cases} +1 & \text{if } \tau_i \text{ corresponds to a local maximum} \\ -1 & \text{if } \tau_i \text{ corresponds to a local minimum} \end{cases}$$

noting in the process that $s_1 = -s_2 = s_3 = -s_4 = \ldots$ and that $\|\boldsymbol{\varepsilon}\|_\infty = \|\varepsilon(\mathbf{q},-)\|_\infty$.

We shall also define

$$(119)\quad I_\varepsilon = \{ i : 1 \le i \le M \text{ and } |\varepsilon(\mathbf{q},\tau_i)| = \|\varepsilon(\mathbf{q},-)\|_\infty \}$$

Assuming that $\varepsilon(\mathbf{q},-)$ is continuously differentiable with respect to the parameters $\mathbf{q}$ we define the M by m matrix $\mathbf{A}$ with components

$$(120)\quad A_{ij} = \frac{\partial \varepsilon}{\partial q_j}(\mathbf{q},\tau_i), \quad i=1,2,\ldots,M, \quad j=1,2,\ldots,m$$

By using Lemma 6, we find that for any $\mathbf{q}^* \in \mathbb{R}^m$

$$(121)\quad \|\varepsilon(\mathbf{q}+\alpha\mathbf{q}^*)\|_\infty = \|\boldsymbol{\varepsilon}\|_\infty + \alpha \max_{i\in I_\varepsilon} \sum_{j=1}^{m} s_i A_{ij} q_i^* + o(\alpha)$$

as α decreases to zero through positive values, and we shall use this relation in seeking to choose $\mathbf{q}^*$ in such a manner that

$$(122)\quad \|\varepsilon(\mathbf{q}+\alpha\mathbf{q}^*,-)\|_\infty < \|\varepsilon(\mathbf{q},-)\|_\infty$$

holds for all sufficiently small $\alpha > 0$.

As we have already noted, in an ideal situation the error curve will have at least one more extremum than there are parameters, i.e., $M \geq m+1$, but unfortunately in this nonlinear approximating problem it is usually not possible to find an easily computable initial estimate of the parameters for which this is the case (although when applicable Procedure 3 will produce such an estimate.) Consequently we expect that in the early stages of the iterative process we will have $M \leq m$. When this is the case (and provided $\mathbf{A}$ has full rank m) we can always find $\mathbf{q}^* \in \mathbb{R}^m$ such that (122) holds for all sufficiently small $\alpha > 0$, and in general this may be done in a variety of ways. We shall choose to determine $\mathbf{q}^*$ in such a manner that

$$(123)\quad \|\mathbf{A}\,\mathbf{q}^* + \boldsymbol{\varepsilon} - \varkappa\,\bar{\varepsilon}\,\mathbf{s}\|_\infty = 0$$

and that subject to this constraint

$$(124)\quad \|\,\mathbf{q}^*\,\|_2 = \text{Minimum}$$

where $\varkappa \in [0,1)$ is a parameter which might well be called the "compression factor" of the process.

There are three main reasons why we use this choice of $\mathbf{q}^*$. First of all, in the special case where $\varepsilon(\mathbf{q} + \mathbf{q}^*,-)$ depends linearly on $\mathbf{q}^*$ we may use (117), (120), and (123) to see that

$$(125)\quad \varepsilon(\mathbf{q} + \mathbf{q}^*, \tau_i) = \varepsilon(\mathbf{q}, \tau_i) + \sum_{j=1}^{m} A_{ij}\, q_j^*$$

$$= [\boldsymbol{\varepsilon} + \mathbf{A}\,\mathbf{q}^*]_i$$

$$= \varkappa\, \bar{\varepsilon}\, s_i\ , \quad i=1,2,\ldots,\ M$$

so that the new error curve $\varepsilon(\mathbf{q} + \mathbf{q}^*,-)$ alternates M-1 times on the set $\{\tau_1, \tau_2, \ldots, \tau_M\}$ with the new equal ripple deviation being smaller than the average extremum, $\bar{\varepsilon}$, of the old error curve $\varepsilon(\mathbf{q},-)$ by the factor $\varkappa$. Thus when we choose $\mathbf{q}^*$ by using (123) and (124) we tend to both "level" and to uniformly "compress" the error curve. Of course $\varepsilon(\mathbf{q} + \mathbf{q}^*,-)$ does not, in general, depend linearly on $\mathbf{q}^*$ so that we cannot expect (125) to hold exactly. Nevertheless, we might expect (125) to hold approximately, especially since the departure from linearity is essentially measured by the quantity $\| \mathbf{q}^* \|_2^2$ which we have attempted to keep small by means of the auxiliary condition (124). Another important reason for using this choice of $\mathbf{q}^*$ is that the solution of (123) - (124) is readily expressed in terms of the pseudoinverse $\mathbf{A}^+$ of $\mathbf{A}$, i.e. we have

$$\mathbf{q}^* = \mathbf{A}^+ \cdot [\varkappa\, \bar{\varepsilon}\, \mathbf{s} - \boldsymbol{\varepsilon}]$$

and therefore $\mathbf{q}^*$ is relatively easy to compute, cf. [9]. Finally we also note that when $\mathbf{q}^*$ is given by (123) - (124) we have

$$(126)\quad \max_{i\in I_\varepsilon} \sum_{j=1}^{m} s_i\, A_{ij}\, q_j^* = \max_{i\in I_\varepsilon} s_i[\varkappa\, \bar{\varepsilon}\, s_i - \varepsilon_i]$$

$$= \varkappa\, \bar{\varepsilon} - \|\varepsilon\|_\infty \quad < 0$$

and this together with (121) shows that (122) holds for all sufficiently small $\alpha > 0$. This then takes care of the case where $M \leq m$.

Proceeding in this manner, we seek to continually "compress" and "level" the error curve and thereby naturally introduce additional extrema until eventually we arrive at a stage where $M > m$, and this being the case we can no longer expect to be able to continue to compress the error curve. Consequently when $M > m$ we determine $\mathbf{q}^*$ so that

$$(127)\quad \|\mathbf{A}\, \mathbf{q}^* + \boldsymbol{\varepsilon}\|_\infty = \text{Minimum}$$

(with the auxiliary condition (124) again being used in the event that (127) does not uniquely determine $\mathbf{q}^*$) noting that the solution of (127) may be computed using known procedures, cf. [1]. Moreover, since $\mathbf{0}$ is a

candidate for the solution of (127) we find

$$\max_{i\in I_\varepsilon} \sum_{j=1}^{m} s_i A_{ij} q_j^* = \max_{i\in I_\varepsilon} s_i \{[\mathbf{A}\,\mathbf{q}^*]_i + \varepsilon_i - \varepsilon_i\}$$

$$= \max_{i\in I_\varepsilon} s_i [\mathbf{A}\,\mathbf{q}^* + \boldsymbol{\varepsilon}]_i - \|\boldsymbol{\varepsilon}\|_\infty$$

$$\leq 0$$

with equality holding if and only if $\mathbf{q}^* = \mathbf{0}$ is a solution of (127). Hence if the minimum of (127) is strictly less than $\|\boldsymbol{\varepsilon}\|_\infty$, then (122) holds for all sufficiently small $\alpha > 0$.

We may summarize this discussion as follows.

Procedure 4: Let the real valued error curve $\varepsilon(\mathbf{q},-)$ be continuously differentiable with respect to $\mathbf{q}=(q_1,q_2,\ldots,q_m)$ and assume that $\varepsilon(\mathbf{q},-)$ has only a finite number of local extrema in $[0,1]$.

(i): Locate the abscissas $\tau_1 < \tau_2 < \ldots < \tau_M$ corresponding to the local extrema of $\varepsilon(\mathbf{q},-)$ and compute $\bar{\varepsilon}$, $\boldsymbol{\varepsilon}$, $\mathbf{s}$, $\mathbf{A}$ using (116), (117), (118), (120), respectively.

(ii): If $M \leq m$ determine $\mathbf{q}^*$ using (123) and (124). In the event that $\mathbf{A}$ does not have full rank this may not be possible and the procedure fails.

If $M > m$ determine $\mathbf{q}^*$ using (127) and (128). In the event that $\mathbf{q}^* = \mathbf{0}$ is a solution the procedure fails.

(iii): Beginning with $\alpha=1$, use bisection to find $\alpha \in (0,1]$ so that (122) holds.

(iv): Accept $\varepsilon(\mathbf{q} + \alpha\, \mathbf{q}^*,-)$ as the improved error curve. □

Through numerical experimentation, Procedure 4 (or more precisely the descritized counterpart which results when $[0,1]$ is replaced by a finite point set and the matrix $\mathbf{A}$ of (120) is computed by differencing) has been found to be quite useful for computing best $\| \|_\infty$ - approximations to a real valued f from $V_n^r(S)$. In the difficult initial stages of the iterative process (where $M \leq m$ and the parameters $\mathbf{q}$ are not at all close to optimum values) the "compress" and "level" approach is quite effective, especially when the approximating functions are taken from $V_n(\mathbb{R})$.

Reasonable hypotheses under which Procedure 4 will with certainty produce a convergent minimizing sequence $\{y_\nu\}$ are not presently known. Likewise the "best" value (if such exists) of the compression factor $\varkappa$ is also unknown. We might point out, however, that in the case where the partial derivatives of $\varepsilon(\mathbf{q},-)$ with respect to $q_1, q_2, \ldots, q_m$ form a Tchebycheff set (as will be the case

under the usual parametrizations when $S \subseteq \mathbb{R}$ and the corresponding exponential sum has full order) then the matrix $\mathbf{A}$ of (120) will always have full rank, and therefore the $M \leq m$ stage of Procedure 4 will never fail to improve the error curve (as we noted using (126).) Moreover, if we assume in addition that the error curve has a normal appearance with $M=m+1$ and $\varepsilon_i \cdot s_i > 0$, $i=1,2,\ldots,$ M then it can be shown that Procedure 4 actually reduces to the well known Newton iteration procedure described by Meinardus [15,p.149]. (In actual practice, however, we are primarily concerned with the early stages of the iteration where the bulk of the computational effort is expended and not in the last iteration or two to which the analysis of Newton's method might apply.)

We shall turn now to one final procedure which may at times be helpful in the construction of best $\| \ \|_\infty$-approximations to certain $f \in C^n[0,1]$ from $V_n^r(\mathbb{R})$. It will be noted that in each of the Procedures 1-3 one simultaneously determines the parameters $\mathbf{c}$ (which characterize the nonlinearities of the approximating function) and then simultaneously determines the parameters $\mathbf{b}$(which appear linearly in the approximating function.) Then by using Procedure 4 (or some other relatively general iterative scheme) one attempts to simultaneously perturb all of the 2n parameters $\mathbf{b}$, $\mathbf{c}$ in some systematic manner and thereby

generate a sequence $\{y_\nu\}$ which converges to a best $\|\ \|_\infty$-approximation. Unfortunately, when the initial estimates of the crutial parameters $\mathbf{c}$ are relatively poor or when n is relatively large, this process may converge slowly (or not at all.) Consequently we shall now describe an alternative scheme which is designed to take advantage of the special features of the approximating functions in $V_n^r(\mathbb{R})$ in a way that the more widely applicable Procedure 4 does not.

Assuming that $f \in C^n\ [0,1]$ is real valued and that $y \in V_n^r(\mathbb{R})$ has exponential parameters $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_n$ we shall define

$$
\begin{aligned}
&\eta_0(t) = f(t)-y(t)\\
&\eta_1(t) = (D-\lambda_1)\eta_0(t)\\
&\eta_2(t) = (D-\lambda_2)\eta_1(t)\\
&\ \ \vdots\\
&\eta_n(t) = (D-\lambda_n)\eta_{n-1}(t) = \{(D-\lambda_1)(D-\lambda_2)\ldots(D-\lambda_n)\}f(t)
\end{aligned}
\tag{128}
$$

It will be recalled that in Procedure 1 $\mathbf{c} = \sigma_n(\mathbb{A})$ is chosen in such a manner as to make $\|\eta_n\|_2$ small and then $\mathbf{b}$ is determined in such a manner as to make $\|\eta_0\|_\infty$ small with the intermediate members $\eta_1, \eta_2, \ldots, \eta_{n-1}$ being ignored in the process. By rewriting (128) in the alternative form

$$\eta_n(t) = \{(D-\lambda_1)(D-\lambda_2) \ldots (D-\lambda_n)\}\, f(t)$$

$$\eta_{n-1}(t) = \exp(\lambda_n t)\cdot\{\eta_{n-1}(0) + \int_0^t \exp(-\lambda_n \tau)\cdot \eta_n(\tau)d\tau\}$$

(129)

$$\eta_{n-2}(t) = \exp(\lambda_{n-1} t)\cdot\{\eta_{n-2}(0) + \int_0^t \exp(-\lambda_{n-1}\tau)\eta_{n-1}(\tau)d\tau\}$$

$$\vdots$$

$$\eta_0(t) = \exp(\lambda_1 t)\cdot\{\eta_0(0) + \int_0^t \exp(-\lambda_1 \tau)\cdot \eta_1(\tau)d\tau\}$$

an alternative scheme is suggested. We first choose $\lambda_1,\lambda_2,\ldots,\lambda_n$ so that η_n is small. We estimate the linear parameter $\eta_{n-1}(0)$ and adjust the n+1 parameters $\lambda_1,\lambda_2,\ldots,\lambda_n$, $\eta_{n-1}(0)$ so that η_{n-1} is small. Proceeding in this manner, at the i-th stage of the process we estimate $\eta_{n-i}(0)$ and then adjust each of $\lambda_1,\lambda_2,\ldots,\lambda_n$, $\eta_{n-1}(0)$, $\eta_{n-2}(0),\ldots,\eta_{n-i}(0)$ in such a manner as to make η_{n-i} small. Thus we sequentially determine the linear parameters with the critical nonlinear parameters being corrected at each stage of this procedure.

For purposes of computation the parameters $\eta_0(0)$, $\eta_1(0),\ldots,\eta_{n-1}(0)$ are somewhat inconvenient to use even when the integrals appearing in (129) are all simplified so that the explicit dependence of $\eta_i(t)$ on these parameters is displayed for each i=1,2,...,n.

On the other hand, when $\lambda_1 < \lambda_2 < \ldots < \lambda_n$ it is clear from (128) that there exist constants a_{ij}, $j=i+1,i+2,\ldots,n$, $i=0,1,\ldots,n-1$ such that

$$
\begin{aligned}
\eta_0(t) &= f(t) - \sum_{j=1}^{n} a_{0j} \exp(\lambda_j t) \\
\eta_1(t) &= (D-\lambda_1)f(t) - \sum_{j=2}^{n} a_{1j} \exp(\lambda_j t) \\
&\vdots \\
\eta_{n-1}(t) &= \{(D-\lambda_1)(D-\lambda_2)\ldots(D-\lambda_{n-1})\}f(t) - a_{n-1,n}\exp(\lambda_n t) \\
\eta_n(t) &= \{(D-\lambda_1)(D-\lambda_2)\ldots(D-\lambda_n)\}f(t)
\end{aligned}
\tag{130}
$$

where $a_{01}, a_{02}, \ldots, a_{0n}$ may be regarded as independent and where

$$a_{ij} = (\lambda_j - \lambda_i)\, a_{i-1,j}, \quad 1 \le i < j \le n \tag{131}$$

With respect to this parametrization we shall now describe the following process.

Procedure 5: Let $f \in C^n[0,1]$ be real valued and let $\eta_0, \eta_1, \ldots, \eta_n$ be parametrized using (130).

(i): Choose $\lambda_{n1} < \lambda_{n2} < \ldots < \lambda_{nn}$ so that η_n is small.

(ii): For each $i = n-1, n-2, \ldots, 1, 0$ in turn and beginning with the initial estimates

$$\lambda_{ij} = \lambda_{i+1,j} \qquad j=1,2,\ldots, n$$

$$a_{ij} = a_{i+1,j}/(\lambda_{i+1,j}-\lambda_{i+1,i+1}) \qquad j=i+2,i+3,\ldots, n$$

$$a_{i,i+1} = \{(D-\lambda_1)\ldots(D-\lambda_{i-1})\}f(0) - \{a_{i,i+2}+\ldots+a_{i,n}\}$$

use Procedure 4 to find the parameters $\lambda_{i1} < \lambda_{i2} < \ldots < \lambda_{in}$ and $a_{i,i+1}$, $a_{i,i+2}$, $\ldots a_{in}$ which locate a minimum of $\|\eta_i\|_\infty$. If this cannot be done, the procedure fails.

(iii): Take the exponential sum

$$y(t) = \sum_{j=1}^{n} a_{0j} \exp(\lambda_{0j} t)$$

as an estimate of the best $\| \|_\infty$ - approximation to f from $V_n^r(\mathbb{R})$. □

In concluding this section we shall illustrate the use of Procedures 1,2,3 and then present a few numerical results which have been obtained by using Procedure 2 and the descretized version of Procedure 4 (in which the interval [0,1] is replaced by a finite point set and the partial derivatives appearing in (120) are computed by differencing.)

First of all we shall use Procedure 1 to obtain an initial approximation, y*, to the function

$$f(t) = (1 + t)^{-1}, \quad 0 \le t \le 1$$

from $V_2^r(\mathbb{R})$. With this choice of f the integrals appearing in (106) may be explicitely evaluated and thus we find

$$-.37500\ldots\ c_1 + .50000\ldots\ c_2 = -.58333\ldots$$
$$.29166\ldots\ c_1 - .37500\ldots\ c_2 = .46875\ldots$$

so that

$$c_1 = 3.00000\ldots$$
$$c_2 = 1.08333\ldots$$

with the corresponding auxiliary polynomial being

$$\lambda^2 + c_1\lambda + c_2 = (\lambda + 2.58012\ldots)(\lambda + .41987\ldots)$$

By using the iterative method of Procedure 4 we now find the coefficients a_1, a_2 which make

$$y^*(t) = a_1 \exp(-2.58012\ldots\ t) + a_2 \exp(-.41987\ldots\ t)$$

a best $\|\ \|_\infty$ - approximation to f (on the discrete point set $t = i/34$, $i=0,1,2,\ldots,34$.) Thus we obtain

$$y^*(t) = .27044\ldots\exp(-2.58012\ldots t) + .73024\ldots\exp(-.41987\ldots t)$$

with $\|f-y^*\|_\infty = .00070$.

For purposes of comparison, we shall also make use of Procedure 2. Taking N=10, h=.1, and using (110) we obtain

$$-.37570\ldots\ c_1 + .51125\ldots\ c_2 = -.52734\ldots$$
$$.28410\ldots\ c_1 - .37570\ldots\ c_2 = .40857\ldots$$

so that

$$c_1 = 2.62672...$$
$$c_2 = .89881...$$

with the auxiliary polynomial being

$$\lambda^2 + c_1\lambda + c_2 = (\lambda + 2.22227...)(\lambda + .40445...)$$

in this case. We now determine a_1, a_2 so that

$$y^{**}(t) = a_1 \exp(-2.22227...\ t) + a_2 \exp(-.40445...\ t)$$

is a best $\| \ \|_\infty$ - approximation to f (on the same point set $t = i/34$, $i=0,1,\ldots, 34$) and thus obtain

$$y^{**}(t) = .30079...\exp(-2.22227...\ t) + .69680...\exp(-.40445...\ t)$$

as the initial approximation produced by Procedure 2 with $\|f-y^{**}\|_\infty = .0024$ for this choice of f,N,h.

Next, we shall also use Procedure 3 to obtain an initial approximation to f. Using n=2 and our particular choice of f we see from (114) that

$$[4/4 + 4/5]\gamma_2 + [4/5 + 4/6]\gamma_1 + [4/6 + 4/7] = 0$$
$$[4/5 + 4/6]\gamma_2 + [4/6 + 4/7]\gamma_1 + [4/7 + 4/8] = 0$$

so that

$$\gamma_1 = -1.45491...$$
$$\gamma_2 = .49765...$$

with the auxiliary polynomial being

$$\mu^2+\gamma_1\mu+\gamma_2 = (\mu-.54986...)(\mu_2-.90504...)$$

in this case. We now determine a_1, a_2, d so that if

$$y^{***}(i/4) = a_1(.54986...)^i + a_2(.90504...)^i \ , \ i=0,1,2,3,4$$

then

$$(.54986...)^0a_1 + (.90501...)^0a_2 - d = f(0) \quad =1.0000...$$
$$(.54986...)^1a_1 + (.90501...)^1a_2 + d = f(1/4) = .80000...$$
$$(.54986...)^2a_1 + (.90501...)^2a_2 - d = f(1/2) = .66666...$$

i.e., so that (112) holds. Hence we find

$$a_1 = .29505...$$

$$a_2 = .70481...$$

so that

$$y^{***}(t) = a_1\exp[\ell n(.54986...)t/.25]+a_2\exp[\ell n(.90504...)t/.25]$$
$$= .29505...\exp(-2.39230...t)+.70481...\exp(-.39906...t)$$

is the initial approximation to f obtained by Procedure 3. In this case we find d = .00013 but $\|f-y^{***}\|_\infty = .00043$.

Finally, by beginning with either y*, y**, or y*** we may use Procedure 4 to obtain the unique best $\|\ \|_\infty$ - approximation to f from $V_2^r(\mathbb{R})$. In actual practice we first choose

3N+2 equally spaced points $0 = t_0^0 < t_1^0 < \ldots < t_{3N+1}^0 = 1$ and use Procedure 4 to level the error curve on this finite point set. We then choose intervals $[\alpha_1^1, \beta_1^1]$, $[\alpha_2^1, \beta_2^1]$, $[\alpha_3^1, \beta_3^1]$ containing the three interior extrema of the error curve (which must alternate exactly 4 times on [0,1]), we take $t_0^1 = 0$, $\alpha_1^1 = t_1^1 < t_2^1 < \ldots < t_N^1 = \beta_1^1$, $\alpha_2^1 = t_{N+1}^1 < t_{N+2}^1 < \ldots < t_{2N}^1 = \beta_2^1$, $\alpha_3^1 = t_{2N+1}^1 < t_{2N+2}^1 < \ldots < t_{3N}^1 = \beta_3^1$, $t_{3N+1}^1 = 1$, and we use Procedure 4 to level the error curve on the new descrete point set $\{t_i^1\}$. Proceeding in this manner we successively choose the intervals $[\alpha_1^k, \beta_1^k]$, $[\alpha_2^k, \beta_2^k]$, $[\alpha_3^k, \beta_3^k]$, k=1,2,... in such a manner that $\{\alpha_i^k\}$, $\{\beta_i^k\}$ both converge to the abscissa of the i-th internal extremum of the error curve corresponding to the best $\| \ \|_\infty$ - approximation to f from $V_2^r(\mathbb{R})$. At the k-th step we use Procedure 4 to level the error curve on the set of 3N+2 points which includes 0,1 and N equally spaced points from each of the intervals $[\alpha_i^k, \beta_i^k]$, i=1,2,3. In this manner we find that

$$y(t) = .28612...\exp(-2.44280...t)+.71367...\exp(-.40726...t)$$

is the best $\| \ \|_\infty$ - approximation to f from $V_2^r(\mathbb{R})$ with the minimax error being $\|f-y\|_\infty = .000208$. We note that in this case f may be very well approximated by $V_2^r(\mathbb{R})$ and that each of Procedures 1,2,3 gives a quite reasonable initial estimate of the best approximation.

As we have noted earlier, the exponential parameters corresponding to a best $\| \ \|_\infty$ - approximation to a given f from $V_n(\mathbb{R})$ may have widely differing magnitudes. This is nicely illustrated in Table 1 where we present the parameters of the unique best $\| \ \|_\infty$ - approximation to $f(t) = \sqrt{t}$ from $V_n(\mathbb{R})$, $n=1,2,3,4$. (When rounded to three significant figures these results agree with those given by Braess [5,p.272].) Finally, as indicated at the end of the discussion of Example 8 we have computed the best $\| \ \|_\infty$ - approximations to $f(t) = (t + a)^{-\alpha}$ from $V_n(\mathbb{R})$ for selected values of n,α,a and these are presented in Table 2.

TABLE 1: Best $\| \ \|_\infty$ - Approximations to $f(t) = \sqrt{t}$ from $V_n(\mathbb{R})$

$$y_1(t) = .209954... \exp(+1.751445... t), \ \|f-y_1\|_\infty = .210$$

$$y_2(t) = -.430769... \exp(-9.471870... t)$$
$$+.460756... \exp(+0.804464... t), \ \|f-y_2\|_\infty = .0300$$

$$y_3(t) = -.098232... \exp(-142.8042... t)$$
$$-.448682... \exp(-4.722464... t)$$
$$+.555034... \exp(+0.600761... t). \ \|f-y_3\|_\infty = .00812$$

$$y_4(t) = -.034841... \exp(-1127.835... t)$$
$$-.110556... \exp(-45.70754... t)$$
$$-.473920... \exp(-3.380016... t)$$
$$+.622219... \exp(+0.493312... t). \ \|f-y_4\|_\infty = .00289$$

TABLE 2: Best $\| \ \|_\infty$ - Approximations to $f(t) = (a+t)^{-\alpha}$ from $V_n(\mathbb{R})$ for selected n, a, α

Function	Approximation	Error
a=.25 α=1	$y_1(t)$= 3.7169862...exp(- 1.9726522...t),	$2.83 \cdot 10^{-1}$
	$y_2(t)$= 1.9296407...exp(- 6.6015729...t)	
	+ 2.0566120...exp(- 0.9648758...t),	$1.38 \cdot 10^{-2}$
	$y_3(t)$= 0.6072307...exp(-12.2120639...t)	
	+ 1.9130664...exp(- 3.9137488...t)	
	+ 1.4791029...exp(- 0.6642925...t),	$6.00 \cdot 10^{-4}$
a=.25 α=2	$y_1(t)$= 15.4117524...exp(- 5.6964148...t).	$5.88 \cdot 10^{-1}$
	$y_2(t)$= 10.3249563...exp(-10.2749010...t)	
	+ 5.6050865...exp(- 2.2863430...t).	$7.00 \cdot 10^{-2}$
	$y_3(t)$= 4.2519392...exp(-16.2738136...t)	
	+ 8.8034311...exp(- 6.0454177...t)	
	+ 2.9405501...exp(- 1.5646197...t),	$4.08 \cdot 10^{-3}$

a=1. α=1 $y_1(t)= 0.9787205...\exp(-0.7151291...t)$, $2.13\cdot10^{-2}$

$y_2(t)= 0.2861206...\exp(-2.4428035...t)$
$+0.7136714...\exp(-0.4072692...t)$, $2.08\cdot10^{-4}$

$y_3(t)= 0.0459343...\exp(-4.5064864...t)$
$+0.3936634...\exp(-1.6046370...t)$
$+0.5604003...\exp(-0.2874933...t)$, $1.83\cdot10^{-6}$

a=1. α=2 $y_1(t)= 0.9700377...\exp(-1.4835359...t)$, $3.00\cdot10^{-2}$

$y_2(t)= 0.4231562...\exp(-3.4502945...t)$
$+0.5764041...\exp(-0.8924222...t)$, $4.40\cdot10^{-4}$

$y_3(t)= 0.0927886...\exp(-5.6327150...t)$
$+0.5256152...\exp(-2.3363023...t)$
$+0.3815909...\exp(-0.6518290...t)$, $5.15\cdot10^{-6}$

a=4. α=1 $y_1(t)= 0.2493038...\exp(-0.2238418...t)$, $6.96\cdot10^{-4}$

$y_2(t)= 0.0470111...\exp(-0.7643293...t)$
$+0.2029881...\exp(-0.1307534...t)$, $7.20\cdot10^{-7}$

$a=4.\ \alpha=2\quad y_1(t)= 0.0621884...\exp(-0.4491094...t),\ 3.12\cdot10^{-4}$

$$y_2(t)= 0.0173792...\exp(-1.0613572...t)$$
$$+0.0451203...\exp(-0.2833674...t),\ 4.83\cdot10^{-7}$$

REFERENCES

1. R. H. BARTELS and G. H. GOLUB, "Stable numerical methods for obtaining the Chebyshev solution to an overdetermined system of equations," Comm. ACM 11 (1968), 401-406.

2. C. de BOOR, "On the approximation by γ-polynomials," from Approximations with Special Emphasis on Spline Functions, ed. I. J. Schoenberg, Academic Press, New York, 1969, 157-183.

3. D. BRAESS, "Approximation mit Exponentialsummen," Computing 2 (1967), 309-321.

4. _____, "Uber die Vorzeichenstruktur der Exponentialsummen," J. Approximation Theory 3 (1970), 101-113.

5. _____, "Die Konstruktion der Tschebyshcheff-Approximierenden bei der Anpassung mit Exponentialsummen," J. Approximation Theory 3 (1970), 261-273.

6. J. CORNFIELD, J. STEINFIELD, and S. W. GREENHOUSE, "Models for the interpretation of experiments using tracer compounds," Biometrics 16 (1960), 212-234.

7. E. A. CODDINGTON and N. LEVINSON, Theory of Ordinary Differential Equations, McGraw-Hill, New York, 1955.

8. C. W. GEAR, The Numerical Integration of Ordinary Differential Equations, Prentice-Hall, Englewood Cliffs, New Jersey, 1971.

9. G. H. GOLUB and C. REINSCH, "Singular value decomposition and least squares solutions," Numer. Math 14 (1970), 403-420.

10. J. F. HART, et al, Computer Approximations, John Wiley & Sons, New York, 1968.

11. F. B. HILDEBRAND, Introduction to Numerical Analysis, McGraw-Hill, New York, 1956.

12. C. R. HOBBY and J. R. RICE, "Approximation from a curve of functions," Arch. Rat. Mech. Anal. 27 (1967), 91-106.

13. D. W. KAMMLER and P. H. LORRAIN, "Analog generation of $1/t$ and $1/t^2$," Simulation 12 (1969), 95.

14. C. LANCZOS, Applied Analysis, Prentice-Hall, Englewood Cliffs, New Jersey, 1956.

15. G. MEINARDUS, Approximation of Functions: Theory and Numerical Methods, Springer-Verlag, New York, 1967.

16. G. PÓLYA and G. SZEGŐ, Aufgaben und Lehrsatz aus der Analysis, Band II, Springer-Verlag, Berlin, 1954.

17. R. PRONY, "Essai experimentale et analytique," J. ecole polytech (Paris) 1 (1795), 24-76.

18. J. R. RICE, "Chebyshev approximation by exponentials," J. Soc. Indust. Appl. Math 10 (1962), 149-161.

19. _____, The Approximation of Functions: Vol I, Addison-Wesley, Reading, Mass., 1964.

20. _____, "On the computation of L_1 approximations by exponentials, rationals, and other functions," Math Comp 18 (1964), 390-396.

21. _____, The approximation of Functions: Vol II, Addison-Wesley, Reading, Mass., 1969.

22. H. L. ROYDEN, Real Analysis, Macmillan, New York, 1968.

23. F. J. SERVAIS, "On the expansion of exponential functions in series of descending exponentials of real argument," Nuc. Sci, and Eng. 34 (1968), 196-197.

24. J. S. THORPE and R. N. LEWINE, "Exponential approximation with piecewise linear error criteria," J. Franklin Inst. 286 (1968), 308-320.