QUARTERLY OF APPLIED MATHEMATICS

Vol. XXXIII APRIL 1975 No. 1

NONLINEAR SEGMENTED FUNCTION APPROXIMATION AND ANALYSIS OF LINE PATTERNS*

By

DONALD E. McCLURE

Brown University

1. Introduction. The feature-selection problem in pattern analysis is concerned with the assignment of analytically or computationally tractable representations to patterns. Two common goals in the mathematical analysis of this problem are to make the representations precise in preserving salient characteristics of patterns, and concise in eliminating redundancy in the representations.

When the objects of the pattern analysis are curves, or line patterns, defined by functions on a subset of the real line, then the feature selection procedure is conveniently formulated in one way as an approximation problem. "Optimal" representations of patterns are derived as best approximations of the patterns by members of a suitably-chosen class of approximating functions. The motivation for the present study is the feature-selection problem for line patterns that exhibit a characteristic quality of piecewise regularity. We are led to the analysis of problems of nonlinear approximation of real-valued functions on the line. The connections between line-pattern analysis and the approximation-theoretic problems will be drawn in greater detail in the next section.

The principal mathematical results of this paper belong to the area of approximation theory. The classes of approximating functions that we consider are closely related to extended Chebyshevian splines with variable knots on a finite interval of the real line. They are defined by a linear differential operator L with constant coefficients. An approximating function ϕ will satisfy $L\phi = 0$ except at a finite number of points $t_1, \cdots, t_v$ in the domain of interest, and at the points t_v continuity constraints may be imposed on ϕ. The points t_v are referred to as knots of ϕ. The approximating classes include, in particular, finite-order algebraic and trigonometric polynomials and splines.

For a line pattern identified with a function f we consider the problem of optimal approximation of f by such a function ϕ with respect to a norm on the space of line patterns. Approximations are optimized with respect to dependence on knots and this makes the optimization problem distinctly nonlinear. We concentrate on mean-square

* Received August 1, 1973. This work was supported in part by NSF Grants GP-7347 and GJ-31007X to Brown University.

approximations, where the space of line patterns is $L_2[0, 1]$. This choice of norm is more a matter of convenience than of necessity and most of the results have immediately recognizable extensions to the other L_p spaces. These extensions are explicitly pointed out where the corresponding results for L_2 are proven, since they may be of theoretical interest. The results for least-squares and uniform approximation are the more important ones from the practical point of view.

In Sec. 3 we seek characterizations of the classes of approximating functions and consider first the question of *existence* of optimal approximations (Theorem 1). Sec. 4 illustrates the *nonuniqueness* of best approximations and directs us to questions of their characterization which may be useful for computation, especially in the troublesome area of optimizing knot locations.

In this direction Secs. 5 and 6 derive error bounds (Theorems 3 and 4) and sharp estimates on the rate of convergence (Theorem 5) as the number of knots increases for optimal approximations without continuity constraints at the knots—so-called pure-segmented approximations. These asymptotic estimates yield tight bounds on the order of convergence of optimal approximations when continuity constraints are imposed at the knots, as in the case of variable-knot spline approximations.

The convergence analysis also leads to precise asymptotic characterizations of where optimal knots are necessarily located. In Sec. 7 we use this characterization to define procedures for computing asymptotically efficient pure segmented approximations (Theorems 7 and 8).

In Sec. 8 we give the results of some computational experiments that relate the approximation-theoretic analysis back to the feature-selection problem in pattern analysis and compare nonlinear segmented approximations with certain best linear approximations.

The treatment we give to the consideration of the method of nonlinear approximation is similar in flavor and in much of its detail to previous analyses of variable-knot polynomial and Chebyshevian spline approximations. Some of the earlier work in this direction anticipates the present results. Aside from the consideration of applications to pattern analysis, the main departures in this paper from previous ones are in the classes of approximating functions considered and in the nature of the convergence analysis.

The following references are representative of related work in approximation theory. de Boor and Rice [1, 2] consider least-squares variable-knot cubic spline approximation with emphasis on algorithmic approaches to the difficult computational problems. Rice's book [21] discusses variable-knot spline approximation in both the L_2 and L_∞ norms and presents a number of interesting (and negative) results on questions of uniqueness. Powell [18] analyzes problems of L_2 spline approximation and presents characterizations of best approximations that may be useful for computation of optimal knot locations. Esch and Eastman [5] describe methods of variable-knot spline approximation in the sup norm and relate results of numerical experiments. Karlin and Ziegler [9] give an elegant and comprehensive treatment of extensions of the "classical" spline theory to classes of Chebyshevian spline functions. Schumaker [25, 26, 27] considers approximation in the uniform norm by Chebyshevian spline functions with variable knots and gives thorough treatment to problems of existence and (non)uniqueness of best approximations and their characterization by alternating properties. The existence theory of Hobby and Rice [8] and Rice [21] for γ-polynomial L_p approximation is pertinent to the similar development in Sec. 3. Meinardus [12, 13] considers segmented approximations in the

uniform norm by polynomials and obtains asymptotic results on the separation of optimal knots. Phillips [16] gives error estimates for L_p approximation by polynomials which point the direction for our discussion in Sec. 5. Burchard [3] has obtained useful error bounds for best L_p approximations by variable-knot polynomial splines; these are similar in form to but less sharp than the asymptotic estimates in Sec. 6. The early papers of Ream [19, 20] and Stone [28] consider least-squares segmented approximation by straight lines and discuss two areas of application of the technique. The sharp results of Sacks and Ylvisaker [22, 23, 24] on L_2 convergence of segmented polynomial approximations and of Wahba [30] on L_2 convergence of segmented approximations by Chebyshevian systems anticipate the present results on convergence (Theorem 5 and its Corollary in Sec. 6) and the sufficient conditions on efficient knot location (Theorem 7 in Sec. 7).

The view taken toward pattern analysis in this investigation is in the spirit of the descriptive-generative approach which underlies the formalism developed by Grenander [6, 7]. Methods of approximation similar to the one considered here have been proposed in the context of descriptive pattern analysis by Pavlidis [14]. In a recent report [15] Pavlidis has investigated piecewise-linear approximation as a technique for representing data such as the trace of an electrocardiogram and has compared algorithms for computing balanced-error segmented approximations.

2. Analysis of line patterns. Feature selection is a fundamental problem in pattern analysis. A feature-selection procedure specifies the vocabulary and grammar in terms of which individual images can be represented. The representations assigned to the patterns are the basic objects in terms of which subsequent descriptive analysis or classifications are based.

The feature-selection problem in pattern analysis is aimed at devising good feature-selection procedures and consequent pattern representations. The "goodness" of a particular procedure depends naturally on the ultimate goal of the pattern analysis, be it pure description or categorization and recognition of the patterns in the traditional sense. The mathematical analysis of the feature-selection problem is concerned with assessing the merits of alternative feature-selection procedures and with guiding the selection of procedures that are optimal within a framework that takes account of structure inherent in the patterns and of the goals of the pattern analysis. Broadly speaking, the mathematical analysis is directed to the search for pattern representations which are at the same time precise in preserving salient structural characteristics of individual patterns and concise in eliminating redundancy within the representations. The qualities of precision and conciseness in the representations are of crucial importance for efficient data compression when approximate representations are used for succinct pattern description or in the commonly large-scale computations involved in pattern classification or discrimination.

The present study is concerned with the feature-selection problem for line patterns. By "line pattern" we mean a curve in finite-dimensional Euclidean space that can be interpreted in turn through parametric representation as a vector-valued function of a real variable. On the simplest level we concentrate our attention on plane curves described a single real-valued function whose domain is a bounded interval subset of the line. The theorems we derive can be interpreted in the broader context of curves with vector-valued representations by appropriate extension of the norms to the vector-valued counterparts of the single functions we treat in detail.

For a line pattern described by a real-valued function on the line, the single real variable on which the function depends is naturally interpreted in many contexts as time. In this case, the patterns may be more appropriately termed "time patterns". Grenander [6] draws such distinctions and relates the broad background in formal pattern analyisis with numerous concrete examples.

Suppose the line patterns of interest are contained in a class $\mathfrak{F}$ of real-valued functions on the bounded interval $[a, b]$. For convenience and without restriction we fix the interval to be $[0, 1]$. On the most basic level, each pattern f is described by the values it assumes at every point in $[0, 1]$. Such a description is practically useless for analysis of the patterns and motivates consideration of alternative representations with the goals of precision and conciseness outlined above for good feature-selection procedures.

A common approach in mathematical pattern recognition identifies the class $\mathfrak{F}$ with sample functions of a random process, say with mean value function zero and a continuous covariance function R. Such assumptions on only the second-order properties of the members of $\mathfrak{F}$ already admit neat formulations of the "goodness" of *linear* feature selection procedures and lead to the derivation of *best* linear procedures.

A linear representation procedure associates with each pattern f an expansion

$$f \sim \sum_{\nu=1}^{\infty} f_\nu \psi_\nu \tag{2.1}$$

where $\Psi = \{\psi_\nu\}_{\nu=1}^{\infty}$ is a family of continuous functions on $[0, 1]$ which is complete in $\mathfrak{F}$ and orthonormal with respect to the standard L_2 inner product. The random coefficients f_ν are given by the appropriately defined stochastic integrals, the inner products (f, ψ_ν).

The crux of the linear feature-selection problem lies in choosing the family Ψ. We can measure redundancy in the representation (2.1) by the correlations of the coefficients $\{f_\nu\}_{\nu=1}^{\infty}$. The representation is not redundant (in a sense, *concise*) if the f_ν are uncorrelated. Also we can discuss how *precise* the representation (2.1) is in terms of the speed of convergence of total mean-squared-error to zero for the infinite series.

There is an optimal choice $\Phi = \{\phi_\nu\}_{\nu=1}^{\infty}$ for the class Ψ in both regards; the coefficients $\{(f, \phi_\nu)\}_{\nu=1}^{\infty}$ are uncorrelated and the total mean-squared-error of the compressed representation

$$f^*(t) = \sum_{\nu=1}^{N} (f, \psi_\nu)\psi_\nu(t) \tag{2.2}$$

is as small as possible for the choice $\psi_\nu = \phi_\nu$.

The optimal family Φ is defined through the Fredholm eigenproblem

$$\lambda\phi(t) = \int_0^1 R(t, s)\phi(s)\, ds, \tag{2.3}$$

where the kernel R is the continuous covariance function associated with the class $\mathfrak{F}$. From the symmetry, definiteness and continuity of R it follows that (2.3) admits solutions λ, ϕ with the following properties. The eigenvalues λ_ν are nonnegative, countable in number and can accumulate only at zero. The eigenfunctions ϕ_ν associated with positive eigenvalues λ_ν are continuous, only a finite number of linearly independent ones are associated with a fixed positive eigenvalue, and eigenfunctions associated with distinct eigenvalues are orthogonal with respect to the inner product (,). These properties of solutions of (2.3) allow us to index the solutions, including multiplicities, according to

diminishing magnitudes of the eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ and orthonormalize the eigenfunctions ϕ_ν.

The resulting class $\Phi = \{\phi_\nu\}_{\nu=1}^{\infty}$ obtained from (2.3) is optimal for the linear feature-selection problem. Setting $\psi_\nu = \phi_\nu$ in (2.1) yields the familiar Karhunen-Loève expansion of the random process modelling $\mathfrak{F}$. The coefficients $f_\nu = (f, \phi_\nu)$ have variance λ_ν and are mutually uncorrelated. Further, the ordering of the family Φ is the optimal one as far as the total mean squared error of compressed representations (2.2) is concerned. The class $\Phi_N = \{\phi_\nu\}_{\nu=1}^{N}$ minimizes $\int_0^1 E[f^*(t) - f(t)]^2\, dt$ and gives the minimum value

$$\min \int_0^1 E[f^*(t) - f(t)]^2\, dt = \sum_{\nu=N+1}^{\infty} \lambda_\nu . \tag{2.4}$$

This "best-linear" procedure and the expression (2.4) will be recalled in Sec. 8 as a standard for comparison of nonlinear representation methods.

By thus appealing to the rather weak assumptions on second-order structure of a class of patterns $\mathfrak{F}$ we are led to natural measures of performance of a feature-selection procedure. The precision is measured in terms of an average approximation error whose specification is reasonable in view of the assumed weak structure. Conciseness is measured in terms of correlations between representing functionals (f, ψ_ν) and, in a related manner, through the number of features necessary to attain a specified degree of precision in compressed representations. In addition, by restricting attention to feature-selection procedures of linear type, the weak structure assumed on the pattern class admits the derivation of optimal procedures of this type.

We are here concerned with feature-selection procedures for the analysis of line patterns that are based on deeper structure of a pattern class $\mathfrak{F}$ than that assumed for the familiar analysis of linear procedures sketched above. The procedures we investigate are motivated by regularity properties of individual images f from a suitable class $\mathfrak{F}$. Of particular concern are functions f on the unit interval that exhibit conspicuous localized irregularities. By this we mean isolated discontinuities in function values, isolated small intervals over which the function varies rapidly, or isolated points or small regions where low-order derivatives of the function are discontinuous or have large variations. Apart from such localized regions of irregularity the patterns globally are more regular.

A piecewise linear function on [0, 1], perhaps discontinuous, typifies the sort of behavior described above. Such line patterns would arise in parametric representations of polygons. The electrocardiogram is also a time pattern exhibiting such regularity properties and it is one particular pattern whose description and analysis is currently of concern in the area of mathematical pattern analysis [15] with the possible goal of improving automatic interpretation.

Of central concern in the present study are patterns generated by a model of the form

$$Lf(t) = \epsilon(t), \qquad 0 \leq t \leq 1, \tag{2.5}$$

where L is a linear differential operator of order n and ϵ is a (random) forcing function of impulse character. The forcing function may be a generalized function whose support is a finite discrete subset of [0, 1] and whose order is less than n, or it may be smooth function approximating such a generalized function in that its support is principally contained in a finite union of distinct small subintervals of [0, 1]. Patterns generated by the model (2.5) would describe, for example, the response of a linear system to shot noise.

A premise under which we proceed is that the irregularities of such patterns are basic

structural properties of them which should be preserved in their representation according to a prescribed feature-selection procedure. Such details of structure are important, for example, in the normal analysis and interpretation of an electrocardiogram trace. The feature-selection procedure should strive to ferret out such structural characteristics. Thus we are led to investigate pattern representations that display this same element of structure of the observable patterns and that retain as significant features the locations of the isolated regions of irregularity.

Such a feature-selection procedure might be based on the approximation of patterns by polynomial splines or extended splines with variable knots. Regions of pattern irregularity are associated with locations of knots of an approximating spline. Apart from its knots, a spline also exhibits the more globally regular behavior that is characteristic of the line patterns we consider. The paper by de Boor and Rice [2] and Rice's book [21] contain analysis and examples of variable-knot cubic spline approximations that are presented very much in this same spirit, with the choice of approximation procedure being guided by the underlying structure of the data being represented.

For the generative model (2.5) we are led to consider representation procedures which go beyond approximation by variable-knot polynomial splines. The types of pattern representations we should consider are dictated by a simple verbal description of the functions generated by a model like (2.5). If ϵ is a generalized function of order $q - 1$ whose support is a finite subset $\{t_\nu\}_{\nu=1}^{p}$ of $(0, 1)$ then a pattern f generated by Eq. (2.5) is piecewise a solution of $L\phi = 0$; at the points t_ν the pattern f has $n - q - 1$ continuous derivatives. Following this description of patterns generated by (2.5) we define the following classes $\Phi(p, q)$ of functions on $[0, 1]$ from which pattern representations will be sought.

Definition 1. Associated with a fixed constant-coefficient linear differential operator L of order n and integers p and q, where $p \geq 0$ and $1 \leq q \leq n$, $\Phi(p, q)$ is the class of functions ϕ on $[0, 1]$ such that for each ϕ there are distinct points $\{t_\nu\}_{\nu=1}^{p}$ with $0 = t_0 < t_1 < \cdots < t_p < t_{p+1} = 1$ and ϕ satisfies

i) ϕ is a C^∞ solution of $L\phi = 0$ on each open subinterval $(t_\nu, t_{\nu+1})$,

ii) ϕ is C^{n-1-q} in a neighborhood of each t_ν, $\nu = 1, \cdots, p$.[1]

The definition puts in precise terms the general description of patterns generated by the ideal model (2.5). The integer p is the number of interior knots t_ν of a function ϕ in $\Phi(p, q)$. The integer q is the number of degrees of freedom of ϕ at each of its associated knots. When $q = n$ the functions ϕ are pure-segmented; that is, the specifications of ϕ over distinct subintervals $(t_\nu, t_{\nu-1})$ are mutually independent. When $L = D^n$ and $q = 1$, the class $\Phi(p, q)$ describes the familiar polynomial splines of degree $n - 1$ with p knots in $(0, 1)$. When $q > 1$, $\Phi(p, q)$ describes the so-called extended or deficient splines.

In order to use functions in a class $\Phi(p, q)$ for pattern representation a rule is necessary which associates a member ϕ of some $\Phi(p, q)$ with each function f in the pattern class $\mathfrak{F}$. We follow the lead outlined in the description of analysis of linear feature-selection procedures above and formulate the association rule through an approximation problem. For fixed values of p and q, consider a pattern class $\mathfrak{F}$, containing $\Phi(p, q)$, which is a metric space. In particular, let the metric on $\mathfrak{F}$ be defined by a norm $||\cdot||$. We will consider the association to each pattern f in $\mathfrak{F}$ of a representation ϕ^* in a suitable extension of

[1] When $q = n$ it is understood that there are no continuity constraints on ϕ at the points t_ν.

$\Phi(p, q)$ which minimizes the norm-distance between $\Phi(p, q)$ and f:

$$||f - \phi^*|| = \inf_{\phi \in \Phi(p,q)} ||f - \phi||. \tag{2.6}$$

In the next section we will be directly concerned with characterizing the proper extensions of the classes $\Phi(p, q)$ so that the association procedure is well-posed.

Through (2.6) the representation problem, the analysis of the proposed feature-selection procedure, is translated into an approximation-theoretic problem. As phrased, the values of p and q are indeterminate and must be specified to fix the approximating set of functions. Their determination may be a nontrivial aspect of implementation. The value of q might be determined by a priori knowledge of the order of a generalized function ϵ in (2.5). Alternatively, its specification could depend on considerations such as the desired simplicity or regularity of derived representations ϕ^*.

The value of p appropriate to obtain a suitable representation for any pattern f may not, however, be so readily determined. It should depend on the number of points or localized intervals of irregularity of the pattern. Choosing a value for p is an inference problem specific to each particular pattern f; its value depends on f. We might gain insight into this inference problem, though, by considering the approximation problem (2.6) for any value of p and appealing to the description of the error $||f - \phi^*||$ in terms of p to guide the choice. The error and convergence analyses of Secs. 5 and 6 are relevant. These sections are also concerned with the problem of locating the knots of ϕ^* once their number is specified.

In the detailed analysis of the approximation problem we take $\mathfrak{F}$ to be $L_2[0, 1]$, square-integrable functions on the unit interval. This choice of pattern class and approximation norm is more for convenience than necessity. Where the proofs apply directly, the analogous results are pointed out for approximation in the other integral norms L_r, $1 \leq r < \infty$, and for the sup norm. As a practical matter the results for least-squares and uniform approximation are the more important ones.

The choice of norm will be of central importance in preserving fidelity between the observable patterns f and their approximate representations ϕ^*. The norm chosen will affect the correspondence between the knots of ϕ^* and the location of regions of irregularity of f. If the main structural characteristics of the patterns in $\mathfrak{F}$ are their very localized regions of irregularity, then the sup norm is likely to be the preferred one for producing faithful approximate representations. On the other hand, if the data describing a pattern f are inexact or noisy then a least-squares criterion will have the advantage of being less sensitive to perturbations in a single datum. Similar arguments can be put forth in favor of alternative performance criteria which would weigh, for example, errors in approximation of derivatives.

Pattern representations in a class $\Phi(p, q)$ fit rather neatly into the formalism and grammatical structures of pattern analysis developed by Grenander in [6, 7]. The formal approach describes the generation of ideal patterns in terms of combinations according to prescribed rules (syntax) of simple primitive objects (signs) and in terms of operations on the primitives and their combined forms (images and patterns). In addition to this algebraic aspect of the formalism, a probabilistic element describes the deformation of ideal images into observable ones. The description of the deformation mechanism admits precise formulation and analysis of pattern recognition and description problems as statistical inference problems.

We associate observable images with functions in the class $\mathfrak{F}$ and ideal images with members of the classes $\Phi(p, q)$, without concern at present for the deformation mechanism mapping the ideal into observable images. The ideal images are catenations on the real line of solutions on bounded intervals of the homogeneous equation $L\phi = 0$. Thus as the building blocks for the generation of members of $\Phi(p, q)$, as the *signs* s of the pattern grammar, we take any suitable representation for a solution of the differential equation on an interval. We can, for example, define a prototype sign s to be an ordered $(n + 1)$-tuple $s = (l; x_0, x_1, \cdots, x_{n-1})$ describing the unique solution ψ_0 of $L\phi = 0$ on the interval $[0, l)$ which satisfies $\psi_0^{(\nu)} = x_\nu$ for $\nu = 0, \cdots, n - 1$. Formally, $\psi_0 \sim s = (l; x_0, \cdots, x_{n-1})$.

Other signs *similar* to these prototypes are formed by translations T_τ on the real line, the background space; $\psi_\tau \sim T_\tau s$, more precisely, $\psi_\tau(t) = \psi_0(t - \tau)$. Then *configurations* c are built through catenation of signs and identified with members ϕ of some $\Phi(p, q)$, $\phi \sim c = (s_1, \cdots, s_N)$. The catenation of signs corresponds to the catenation on the real axis of the functions they represent. There are *syntactic constraints* on the catenation of signs so that the resulting configuration c is admissible, grammatical. The syntax requires that the intervals of the background space associated with the distinct signs comprising c will partition the interval $[0, 1]$ and that the continuity constraints on the members of $\Phi(p, q)$ are satisfied by the function identified with c.

Admissible configurations which represent the same function ϕ are considered equivalent. They are indistinguishable to an observer. The equivalence classes of configurations define the *images* of the pattern grammar. A function in $\Phi(p, q)$, that is with p interior knots in $[0, 1]$, admits association with a unique image and with a configuration comprising no more than $p + 1$ signs. By these associations each class $\Phi(p, q)$ is viewed as a subset of the image algebra in a formal grammar of patterns.

In passing to the analysis of the nonlinear approximation problem (2.6) we observe that it is also a standard formulation of a pattern *recognition* problem according to a *minimum-distance* criterion. The functions f in $\mathfrak{F}$ correspond to observable deformed images, and the functions ϕ in $\Phi(p, q)$ are the ideal images to be inferred. The recognition problem becomes the approximation problem where the recognition criterion is minimum distance to $\Phi(p, q)$ according to the norm on $\mathfrak{F}$. In this setting, the L_2 norm on $\mathfrak{F}$ leads to a recognition procedure corresponding to classical (but nonlinear) least-squares regression. Of course, a better norm for minimum-distance recognition may be dictated by precise specification of the deformation mechanism that transforms an ideal image ϕ into an observable image f.

3. Approximation problem: existence. Identify the space of line patterns $\mathfrak{F}$ with $L_2[0, 1]$, real-valued square-integrable functions on the unit interval with the standard L_2 norm $||\cdot||$. Let L denote a fixed nth-order linear differential operator with real constant coefficients. In factored form,

$$L = \prod_{\nu=1}^{n} (D - \alpha_\nu), \tag{3.1}$$

where D represents differentiation and the α_ν are the characteristic roots of L. $\Phi(p, q)$ denotes the class of functions generated by L, elicited in Definition 1. The following nonlinear approximation problem is posed: for f in $\mathfrak{F}$, find ϕ^* in $\Phi(p, q)$ satisfying

$$||f - \phi^*|| = \inf_{\Phi(p,q)} ||f - \phi||. \tag{3.2}$$

This problem is not well-posed. $\Phi(p, q)$ is not closed in $\mathfrak{F}$ unless $q = n$. The difficulties are essentially the same as those met in the analysis of variable-knot spline approximations. When the knots are free parameters of the approximations they must be allowed to coalesce into a multiple knot with a corresponding relaxation of continuity constraints on the spline at the multiple knot. Rice [21] describes the phenomenon for polynomial splines, as does Schumaker for Chebyshevian splines [26].

In order to assure existence of optimal approximations for arbitrary functions f in $\mathfrak{F}$ we must look at least to the closure of $\Phi(p, q)$ in $\mathfrak{F}$. In this direction we define classes of functions $\Phi^*(p, q)$ which are related to the closure of $\Phi(p, q)$ in $\mathfrak{F}$ and in which existence of best approximations is established.

Definition 2. Associated with a fixed constant-coefficient linear differential operator L of order n and integers p and q, where $p \geq 0$ and $1 \leq q \leq n$, $\Phi^*(p, q)$ is the class of functions ϕ on [0, 1] such that for each ϕ there are distinct points $\{s_\nu\}_{\nu=1}^{r}$ with $0 = s_0 < s_1 < \cdots < s_r < s_{r+1} = 1$ and associated integer multiplicities $\{m_\nu\}_{\nu=1}^{r}$ with $m_\nu \geq 1$ and $\sum_{\nu=1}^{r} m_\nu \leq p$ and ϕ satisfies

i. ϕ is a C^∞ solution of $L\phi = 0$ on each open subinterval $(s_\nu, s_{\nu+1})$,

ii. ϕ is C^{n-1-qm_ν} in a neighborhood of each s_ν, $\nu = 1, \cdots, r$.[2]

The analysis will show that $\Phi^*(p, q)$ is identified with the closure of $\Phi(p, q)$ in $\mathfrak{F}$ with respect to any L_r topology, $1 \leq r < \infty$. $\Phi^*(p, q)$ is a necessary and sufficient extension of $\Phi(p, q)$ in order to admit existence of solutions to the approximation problem (3.2).

In the existence results that follow there are strong parallels to previous results on existence of optimal variable-knot spline approximations [21, 26]. Hobby and Rice's elegent development for γ-polynomial approximations [8] is directly applicable to the present situation. The previous results do not completely cover our questions concerning the classes $\Phi(p, q)$ since (i) these classes go beyond the case of polynomial splines, (ii) they need not be Chebyshevian splines, and (iii) the arguments already presented for γ-polynomials assume the number p of interior knots is restricted by the degree of the γ-polynomial. The latter results will be adapted for our proofs.

Convenient representations for functions in the classes $\Phi(p, q)$ and $\Phi^*(p, q)$ draw on analogy with algebraic polynomials and splines. We let ϕ_ν denote the unique solution of $L\phi = 0$ which satisfies the conditions $D^{\nu-1}\phi_\nu(0) = (\nu - 1)!$ and $D^\mu\phi_\nu(0) = 0$ for $\mu = 0, 1, \cdots, n - 1$ and $\mu \neq \nu - 1$. The real-analytic function ϕ_ν has a series expansion of the form

$$\phi_\nu(t) = t^{\nu-1} + O(t^n) \tag{3.3}$$

for $\nu = 1, \cdots, n$, where $O(t^n)$ describes terms of order n and greater in the expansion. In terms of the fundamental solutions ϕ_ν, define the functions $\phi_\nu(t; s)$ which display varying degrees of regularity at a knot s by

$$\begin{aligned} \phi_\nu(t; s) &= 0, && t < s \\ &= \phi_\nu(t - s), && t \geq s \end{aligned} \quad ; \tag{3.4}$$

for $\nu = 1, \cdots, n$.

Now the members of $\Phi(p, q)$ and $\Phi^*(p, q)$ admit representations as linear combinations of the functions ϕ_ν and $\phi_\nu(\ ; s)$. If $\phi \in \Phi(p, q)$, with interior knots $t_1 < t_2 < \cdots < t_p$, then

[2] When $qm_\nu \geq n$ it is understood that there are no continuity constraints on ϕ at s_ν and, at worst, jump discontinuities are allowed.

$$\phi(t) = \sum_{\nu=1}^{n} b_\nu \phi_\nu(t) + \sum_{\mu=1}^{p} \sum_{\nu=1}^{q} a_{\mu\nu} \phi_{n+1-\nu}(t; t_\mu). \tag{3.5}$$

For any $\phi \in \Phi^*(p, q)$, with interior knots $s_1 < s_2 < \cdots < s_r$ and respective multiplicities $m_1, \cdots, m_r$,

$$\phi(t) = \sum_{\nu=1}^{n} b_\nu \phi_\nu(t) + \sum_{\mu=1}^{r} \sum_{\nu=1}^{qm_\mu} a_{\mu\nu} \phi_{n+1-\nu}(t; s_\mu), \tag{3.6}$$

where we take $a_{\mu\nu} = 0$ when $\nu > n$; the values of r and m_μ are constrained by Definition 2.

The main existence result for best approximations is the following theorem.

THEOREM 1. Let L be an nth-order constant-coefficient differential operator (3.1), let p be a nonnegative integer and q an integer with $1 \leq q \leq n$. For any function f in $L_2[0, 1]$, there exists a function ψ in $\Phi^*(p, q)$ such that $||f - \psi|| = \inf_{\phi \in \Phi(p,q)} ||f - \phi||$.

Remark. The reader will observe that the proof applies equally to establish existence of a best approximation in $\Phi^*(p, q)$ of any function f in $L_r[0, 1]$ for $1 \leq r < \infty$. The result is slightly different for approximation in the essential sup norm. For f in $L_\infty[0, 1]$, the proof establishes existence of ψ in $\Phi^*(p, q)$ for which $||f - \psi|| \leq \inf_{\phi \in \Phi(p,q)} ||f - \phi||$. For f in $C[0, 1]$ and approximation in the sup norm the equality is attainable; that is, there exists ψ in $\Phi^*(p, q)$ for which $||f - \psi|| = \inf_{\phi \in \Phi(p,q)} ||f - \phi||$; see Schumaker [26, 27] and Rice [21].

The proof of Theorem 1 is deferred so that a preliminary result relating Φ and Φ^* can be established. The first result says that we must look at least to the set $\Phi^*(p, q)$ for solutions of the optimization problem (3.2).

THEOREM 2. Let L be of the form (3.1), $p \geq 0$, and $1 \leq q \leq n$. $\Phi^*(p, q)$ is contained in the closure of $\Phi(p, q)$ with respect to $L_2[0, 1]$.

Remark. The result also holds and the proof applies for the closure of $\Phi(p, q)$ in $L_r[0, 1]$, $1 \leq r < \infty$. The result is not true for the closure of $\Phi(p, q)$ in $L_\infty[0, 1]$ unless $pq < n$.

Proof. A constructive argument shows that any function in $\Phi^*(p, q)$ can be approximated arbitrarily closely in the L_2 norm by a member of $\Phi(p, q)$. With the representation (3.6), we can reduce the proof to showing that a fun tion ϋ of the form $\sum_{\nu=1}^{mq} a_\nu \phi_{n+1-\nu}(t; s)$ is a limit of functions ϕ of the form $\sum_{\mu=1}^{m} \sum_{\nu=1}^{q} a_{\mu\nu} \phi_{n+1-\nu}(t; t_\mu)$. This result, in turn, is inferred from successive applications of the following construction which splits a single multiple knot s into two distinct ones of lesser multiplicity.

Consider a function ψ of the form

$$\psi(t) = \sum_{\nu=1}^{m} a_\nu \phi_{n+1-\nu}(t; s) \tag{3.7}$$

with a single knot s of multiplicity m, where $0 < s < 1$ and $2 \leq m \leq n$. For fixed positive integers j and k with $j + k = m$ and for a suitably small value of a positive parameter δ we construct an approximation ψ_δ of ψ which has knots of multiplicities j and k at $s - 2\delta$ and $s - \delta$, respectively. It is then argued that $||\psi - \psi_\delta|| \to 0$ as $\delta \to 0$ to complete the proof.

ψ_δ is constructed to interpolate values of right-derivatives of ψ at the knot s. Consider ψ_δ of the form

$$\psi_\delta(t) = \sum_{\nu=1}^{j} b_\nu \phi_{n+1-\nu}(t; s - 2\delta) + \sum_{\nu=1}^{k} c_\nu \phi_{n+1-\nu}(t; s - \delta). \tag{3.8}$$

Now in (3.7) the coefficients a_ν are determined by the right-derivatives of orders $n-1, \cdots, n-m$ of ψ at s; indeed, $D^{n-\mu}\psi(s+) = a_\mu(n-\mu)!$. The connection suggests choosing the coefficients b_ν and c_ν of ψ_δ so that $D^{n-\mu}\psi_\delta(s) = D^{n-\mu}\psi(s+)$, for $\mu = 1, \cdots, m$. These equations yield a linear system of equations $M_\delta y = \tilde{a}$ for the coefficients b_ν and c_ν; in this form the m-vector y has components $y_\nu = b_\nu$ for $\nu = 1$ to j and $y_{j+\nu} = c_\nu$ for $\nu = 1$ to k and the vector $\tilde{a}$ has components $\tilde{a}_\nu = a_\nu(n-\nu)!$. The entries of the $m \times m$ matrix M_δ are evaluations of $D^{n-\mu}\phi_{n+1-\nu}$ at δ and 2δ. We must establish that $M_\delta y = \tilde{a}$ admits solution for δ sufficiently small. This argument will yield estimates on the orders of the solutions b_ν and c_ν for small δ, which will be used to establish the convergence of the ψ_δ to ψ.

With the expansions (3.3) for ϕ_ν it follows that M_δ can be expressed as a product

$$M_\delta = D_1[M + O(\delta)]D_2 \tag{3.9}$$

where D_1 and D_2 are diagonal matrices depending on δ, M is a constant matrix and $O(\delta)$ denotes an $m \times m$ matrix whose entries are all of order δ. In particular, $[D_1]_{\nu\nu} = \delta^{\nu-m}$, $[D_2]_{\nu\nu} = \delta^{m-\nu}$ for $\nu = 1$ to j and $[D_2]_{(j+\nu)(j+\nu)} = \delta^{m-\nu}$ for $\nu = 1$ to k. The matrix M is most easily described in terms of the m polynomial functions $f_\nu(t) = (t+1)^{m-\nu} \cdot (n-\nu)!/(m-\nu)!$ for $\nu = 1$ to j and $f_{j+\nu}(t) = t^{m-\nu}(n-\nu)!/(m-\nu)!$ for $\nu = 1$ to k. In terms of these functions $M_{\mu\nu} = D^{m-\mu}f_\nu(1)$.

The solvability of the system $M_\delta y = \tilde{a}$ for y follows from the invertibility of the product (3.9) for sufficiently small δ. The invertibility of the product depends on the nonsingularity of M. The determinant of M is, up to a sign correction, the Wronskian of $\{f_\nu\}_{\nu=1}^{m}$ evaluated at 1. This Wronskian is nonzero since the polynomials f_ν are a fundamental set of solutions of the mth-order equation $D^m\phi = 0$. Thus M is invertible.

The solution y of $M_\delta y = \tilde{a}$ is given explicitly by

$$y = D_2^{-1}[M^{-1} + O(\delta)]D_1^{-1}\tilde{a}, \tag{3.10}$$

where as before $O(\delta)$ denotes an $m \times m$ matrix whose entries are all of infinitesimal order δ. For sufficiently small δ the desired interpolating approximations ψ_δ are well-defined.

To establish the convergence of ψ_δ to ψ as δ goes to 0, we use the explicit descriptions of D_1 and D_2 above and observe that (3.10) yields

$$\begin{aligned} y_\nu = b_\nu &= O(\delta^{\nu-m}), \qquad \nu = 1, \cdots, j, \\ y_{j+\nu} = c_\nu &= O(\delta^{\nu-m}), \qquad \nu = 1, \cdots, k. \end{aligned} \tag{3.11}$$

These order bounds on the coefficients b_ν and c_ν together with the representation (3.8) of ψ_δ imply that for sufficiently small values of δ (i) $|\psi_\delta(t)|$ is bounded by a constant B independent of δ on the interval $s - 2\delta \le t \le s$, and (ii) if $m < n$, then $D^{n-\mu}\psi_\delta(s) \to 0$ as $\delta \to 0$ for $\mu = m+1, \cdots, n$. The latter consequence, together with the continuous dependence of solutions of $L\phi = 0$ on initial conditions, imply that $\psi_\delta(t) \to \psi(t)$ uniformly on the interval $(s, 1]$ as $\delta \to 0$. Finally, recall that ψ_δ and ψ are identically zero on the interval $[0, s-2\delta)$. The separate observations on the behavior of ψ_δ on the intervals $[0, s-2\delta)$, $[s-2\delta, s]$ and $(s, 1]$ are readily combined to establish $\|\psi - \psi_\delta\| \to 0$ as $\delta \to 0$. The functions ψ_δ have the desired convergence property and the theorem is proved.

We have the necessary preliminaries to prove the main existence theorem.

Proof of Theorem 1. Let f be in $L_2[0, 1]$. We demonstrate a solution of the opti-

mization problem from the class $\Phi^*(p, q)$. Set

$$e_{p,q}(f) = \inf_{\phi \in \Phi(p,q)} ||f - \phi||. \tag{3.12}$$

The proof is given in three parts: (I) construction of a best approximation ψ, (II) proof that ψ is in $\Phi^*(p, q)$, and (III) proof that $||f - \psi|| = e_{p,q}(f)$.

(I) The construction of ψ appeals to standard compactness arguments. Let $\{\psi_i\}_{i=1}^{\infty}$ be a sequence from $\Phi(p, q)$ satisfying $\lim_{i\to\infty} ||f - \psi_i|| = e_{p,q}(f)$. For each i let $t_\nu^{(i)}$ for $\nu = 0, 1, \cdots, p + 1$ denote the distinct knots of ψ_i. Since $0 \leq t_\nu^{(i)} \leq 1$, we can extract a subsequence of $\{\psi_i\}$ for which the associated knots converge as $i \to \infty$. Assume without loss of generality that this convergence holds for the original sequence $\{\psi_i\}$; that is, for each ν, $\lim_{i\to\infty} t_\nu^{(i)} = T_\nu$ exists. Let $S_0, S_1, \cdots, S_r, S_{r+1}$ denote the distinct values in the limiting set $(T_0, T_1, \cdots, T_{p+1})$ and let m_ν, for $\nu = 1$ to r, denote the multiplicity of the value S_ν in this set. We have $0 = S_0 < S_1 < \cdots < S_r < S_{r+1} = 1$ and $\sum_{\nu=1}^{r} m_\nu \leq p$.

The function ψ is constructed to have knots at the points S_ν. Let N be an arbitrarily small closed neighborhood of the points $\{S_\nu\}_{\nu=0}^{r+1}$ and consider the sequence $\{\psi_i\}$ restricted to the complementary set $N^c = [0, 1] - N$. For i sufficiently large, ψ_i has no knots in N^c and satisfies $L\psi_i = 0$ there. The restricted ψ_i are members of a finite-dimensional vector space. Further, the sequence of norms $\{||\psi_i||\}$ is bounded since $||f - \psi_i||$ converges. Together, by the compactness argument, these observations imply that a subsequence of $\{\psi_i\}$ converges uniformly to a limit function ψ on N^c. We assume without loss of generality that the original sequence $\{\psi_i\}$ converges uniformly to the limit ψ on N^c as $i \to \infty$.

On N^c the function ψ satisfies $L\psi = 0$ and it admits a unique right-continuous extension to the interval [0, 1] that satisfies $L\psi = 0$ except at the points S_ν. Since the neighborhood N of the set $\{S_\nu\}$ is arbitrary, it follows that $\{\psi_i(t)\}$ converges to $\psi(t)$ for all t in [0, 1] except for $t = S_\nu$. This convergence is uniform on any subset of [0, 1] that deletes a neighborhood of the knots.

(II) To show that ψ is in $\Phi^*(p, q)$ we must show that $D^{n-1-qm_\nu}\psi$ is continuous at S_ν, for $\nu = 1$ to r. We first treat the case $q = 1$, analogous to spline approximations.

Let S be any interior knot of ψ and let m be its multiplicity. Assume $m \leq n - 1$; otherwise there are no continuity conditions on ψ at S. Let δ be a small positive value so that $[S - \delta, S + \delta]$ contains only one knot S of ψ.

The knot S is the limit as $i \to \infty$ of m of the interior knots of the functions ψ_i; denote these by $\tau_{1i}, \cdots, \tau_{mi}$, indexed according to their increasing magnitude. Let i be sufficiently large so that $S - \delta < \tau_{1i}$ and $\tau_{mi} < S + \delta$.

On the interval $[S - \delta, S + \delta]$ each ψ_i has a representation

$$\psi_i(t) = \phi^{(i)}(t) + \sum_{\nu=1}^{m} a_{\nu i}\phi_n(t; \tau_{\nu i}), \tag{3.13}$$

from expression (3.5), where $\phi^{(i)}$ satisfies $L\phi = 0$ on $(S - \delta, S + \delta)$. To apply the analysis of Hobby and Rice, we note that for $j = 0, 1, \cdots, n - 1$ the partials $(\partial^j/\partial s^j)\phi_n(t; s) = \phi_n^{(j)}(t; s)$ exist and are linearly independent functions of t, $\phi_n^{(n-2)}(t; s)$ is absolutely continuous, and $\phi_n^{(n-1)}(t; s)$ is bounded on $[0, 1] \times [0, 1]$. We can in fact express the partial $\phi_n^{(j)}(t; s)$ as a linear combination of $\phi_{n-j}(t; s)$, $\phi_{n-j+1}(t; s)$, $\cdots$, $\phi_n(t; s)$, from expression (3.4).

Since the ψ_i are norm-bounded on [0, 1], we know that the ψ_i restricted to $[S - \delta, S + \delta]$ are also norm-bounded. With the representation (3.13) for ψ_i, the properties of

the partials $\phi_n^{(j)}(t; s)$, and the bound on the norms of the ψ_i we can invoke the basic convergence theorem of Hobby and Rice [8] (their Theorem 2). This states that the ψ_i restricted to $[S - \delta, S + \delta]$ converge in norm to a function of the form $\phi(t) + \sum_{\nu=0}^{m-1} b_\nu \phi_n^{(\nu)}(t; S)$, where ϕ is the uniform limit of the regular components $\phi^{(i)}$ of the ψ_i. This expression must represent ψ on $[S - \delta, S + \delta]$ since ψ is the pointwise limit of $\{\psi_i\}$ there. By the observation relating the partials $\phi_n^{(j)}(t; s)$ to the functions $\phi_{n-j}(t; s), \cdots, \phi_n(t; s)$ we obtain

$$\psi(t) = \phi(t) + \sum_{\nu=1}^{m} a_\nu \phi_{n+1-\nu}(t; S), \qquad S - \delta \leq t \leq S + \delta; \tag{3.14}$$

comparable to (3.6). In particular, (3.14) implies that $D^{n-1-m}\psi$ is continuous at S. The argument applies equally at each interior knot of ψ, so ψ is in $\Phi^*(p, 1)$.

It remains to show that ψ is in $\Phi^*(p, q)$ when $q \neq 1$. Let q be any integer $1 \leq q \leq n$. We can regard the limit function ψ, constructed as a pointwise limit of the sequence $\{\psi_i\}$ from $\Phi(p, q)$, as a limit of a sequence from $\Phi(pq, 1)$. Suppose that m knots of the ψ_i converge to an interior knot S of ψ and that $mq \leq n - 1$. By the closure argument of Theorem 2, each ψ_i can be approximated arbitrarily closely by a member of $\Phi(pq, 1)$ that has q distinct knots near each knot of ψ_i. Thus, near S, ψ is a limit of a sequence from $\Phi(pq, 1)$ with mq knots converging to S. By the regularity argument above we conclude that when $mq \leq n - 1$, then $D^{n-1-mq}\psi$ is continuous at S and ψ is in $\Phi^*(p, q)$.

(III) The proof of the theorem will be completed by showing that $\|f - \psi\| = e_{p,q}(f)$.

The pointwise convergence of $\{\psi_i\}$ to ψ implies that $\|f - \psi\| \leq \lim \|f - \psi_i\| = e_{p,q}(f)$, by applying Fatou's lemma for the integral norm. By Theorem 2, ψ in $\Phi^*(p, q)$ is in the closure of $\Phi(p, q)$, so we can find a sequence $\{\chi_i\}_{i=1}^{\infty}$ from $\Phi(p, q)$ such that $\lim_{i\to\infty} \|\psi - \chi_i\| = 0$. (This is the one step in the proof that need not apply for the essential sup norm.) Necessarily, $\|f - \chi_i\| \geq e_{p,q}(f)$, so $\|f - \psi\| = \lim \|f - \chi_i\| \geq e_{p,q}(f)$. Together the inequalities yield $\|f - \psi\| = e_{p,q}(f)$. The proof is complete.

Remark. An immediate consequence of the proofs of Theorems 1 and 2 is that for $1 \leq r < \infty$, $\Phi^*(p, q)$ is the closure of $\Phi(p, q)$ in $L_r[0, 1]$. With respect to $L_\infty[0, 1]$, $\Phi^*(p, q)$ is the closure of $\Phi(p, q)$ only when $pq < n$.

For the subsequent error and convergence analyses it is convenient to observe relationships between the classes $\Phi^*(p, q)$ for different values of p and q. If m is any positive integer then $\Phi^*([p/m], \min(n, mq))$ is contained in $\Phi^*(p, q)$, where $[p/m]$ denotes the integer part of p/m. This inclusion is a direct consequence of Definition 2.

4. Nonuniqueness. A natural question to follow the demonstration of existence of best approximations in $\Phi^*(p, q)$ is that concerning their uniqueness. Uniqueness would be a desirable property for developing computational algorithms out of necessary conditions for optimality.

Unfortunately, best approximations in $\Phi^*(p, q)$ are not necessarily unique. In [2] de Boor and Rice present a neat theoretical argument based on the strict monotonicity of approximation error as a function of p, the number of interior knots, that establishes nonuniqueness for best variable-knot spline approximations. As a specific example, consider the best L_2 approximation of $f(t) = t^3$ on the interval $[-1, 1]$ by a step function with one knot. Straightforward calculations show that either $\pm 1/\sqrt{3}$ is an optimal knot location.

By thus optimizing the knot locations in approximating a function f with a member of

$\Phi^*(p, q)$ we may enounter a situation where the solution is not unique, or at least we must anticipate facing problems of local extrema in knot optimization. Powell [18] points to the negative implications of this problem of local extrema for algorithmic computational procedures and suggests using judicious starting approximations for the knot positions in order to increase the chance of obtaining good approximations through iterative knot adjustment. We obtain positive results in this direction from the convergence analysis that follows.

The iterative procedures for knot adjustment that are used or suggested in the literature [2, 5, 18] are based on necessary conditions for local minima. Their success in finding a global minimum depends on their initialization. The results of Sec. 7 on necessary and sufficient conditions on the asymptotic distribution of optimal knots will suggest a particular scheme for reasonable initialization of such iterative procedures. The theorems make precise and reinforce the successful rules of thumb for initialization described by de Boor and Rice [2].

The path to these knot-distribution results starts by examining convergence of best approximations from $\Phi(p, n)$ as p, the number of knots, increases.

5. Local error estimates. To obtain sharp estimates of approximation error and order of convergence of function approximations from the classes $\Phi(p, q)$ reflecting dependence on the number and location of interior knots we first investigate the dependence of the error on the length of the interval of approximation for function approximations by regular solutions of the equation $L\phi = 0$. Since the main results we seek are asymptotic, for p large, and since knots for optimal approximations from $\Phi(p, q)$ will be expected to cluster as p increases, we first seek estimates for approximation error on small intervals. We build on these local estimates in the next section to obtain the exact order of convergence for approximations from $\Phi(p, n)$, pure segmented approximations, and sharp upper and lower bounds on the order of convergence for approximations from the classes $\Phi(p, q)$ when $q < n$.

The following result, proven in [10], illustrates the type of local convergence results we will develop more fully in this section.

THEOREM 3. If f is in $C^{n+1}[t_0, t_0 + h]$, then the error of the best L_2 approximation of f on $[t_0, t_0 + h]$ by a solution of $L\phi = 0$ is

$$\min_{\{\phi : L\phi = 0\}} \int_{t_0}^{t_0+h} (f(t) - \phi(t))^2 \, dt = K_n (Lf(t_0))^2 h^{2n+1} + O(h^{2n+2}), \tag{5.1}$$

$$K_n = \frac{1}{(n!)^2} \left[\frac{1}{2n+1} - v^T H^{-1} v \right], \tag{5.2}$$

where H is the $n \times n$ Hilbert matrix with entries $H_{ij} = 1/(i + j - 1)$ and v is the column n-vector with components $v_i = 1/(n + i)$.

This theorem is proved by a straightforward asymptotic analysis of the problem of linear least-squares regression of the function f on the span of the functions ϕ_r (Eq. (3.3)) when the interval length h is small. The proof is omitted here as we will obtain a stronger result than (5.1) in the form of an error bound using weaker regularity assumptions on the function f. The present arguments also have the advantage of extending to yield error bounds for approximation in any L_r norm, $1 \leq r \leq \infty$. The error bounds so obtained are more useful for implementation of a variable-knot approximation procedure than is the convergence estimate (5.1).

A basic tool in deriving error bounds is Pólya's mean value theorem [17]. To invoke this result we must impose a constraint on the length h of an interval of approximation so that the operator L possesses Pólya's Property W.

Definition 3. The operator L (Eq. (3.1)) possesses *Property W* on the interval (a, b) if there are n solutions $\psi_1, \psi_2, \cdots, \psi_n$ of $L\phi = 0$ on (a, b) such that the n Wronskians $W(\psi_1)$, $W(\psi_1, \psi_2)$, $\cdots$, $W(\psi_1, \psi_2, \cdots, \psi_n)$ are all strictly positive throughout (a, b).

The interval in the definition can be closed as well as open as long as the derivatives of the function ψ_ν can be appropriately defined at or extended to include endpoints.

Any linear constant-coefficient differential operator (3.1) will possess Property W on a sufficiently small interval. To see this observe that the solutions ϕ_ν defined in Sec. 3 (Eq. (3.3)) satisfy $\phi_1(0) = 1$ and $W(\phi_1, \phi_2, \cdots, \phi_k)(0) = \prod_{j=1}^{k} (j-1)!$ for k between 2 and n. Since these Wronskians are all continuous and positive at 0, they are all strictly positive on a suitably short interval $[0, h]$. Because of the translation-invariance of solutions of the constant-coefficient homogeneous equation, whether L possesses Property W on any interval depends only on the interval length and not on its absolute location. Thus for the operator L of (3.1) there is a constant h_0, $0 < h_0 < \infty$, such that if $h < h_0$, then L possesses Property W on the interval $[t_0, t_0 + h]$.

With this notion we can establish the following result.

THEOREM 4. If the operator L possesses Property W on $[0, h]$ and if f is in $C^n[t_0, t_0 + h]$, then the error of the best L_2 approximation of f on $[t_0, t_0 + h]$ by a solution of $L\phi = 0$ is

$$\min_{\{\phi : L\phi = 0\}} \int_{t_0}^{t_0+h} (f(t) - \phi(t))^2 \, dt = K_n(h)(Lf(\xi))^2 h^{2n+1} \tag{5.3}$$

for some point ξ in $[t_0, t_0 + h]$. $K_n(h)$ is independent of f and is given by $K_n(h) = K_n + O(h)$, where K_n is given by (5.2); equivalently,

$$K_n = \left[(2n+1)\binom{2n}{n}^2 (n!)^2 \right]^{-1}. \tag{5.4}$$

Remark. Expressions analogous to (5.3) and (5.4) for L_r approximation will be observed at the end of the proof of Theorem 4.

The proof of Theorem 4 follows the development of several preliminary consequences of Property W, which are of independent interest.

LEMMA 1. If L possesses Property W on $[0, h]$ and if $\psi_1, \cdots, \psi_n$ are solutions of $L\phi = 0$ for which the Wronskians $W(\psi_1, \cdots, \psi_k)$ are strictly positive on $[0, h]$, for $1 \leq k \leq n$, then $\{\psi_\nu\}_{\nu=1}^{n}$ is a complete Chebyshev system on $[0, h]$; that is, every non-trivial linear combination $\sum_{\nu=1}^{k} c_\nu \phi_\nu$ has at most $k - 1$ zeros in $[0, h]$ for $1 \leq k \leq n$.

Proof. Meinardus [13], Theorem 70, p. 88.

From this fact, that a differential operator L with Property W on an interval has a fundamental set of solutions of the homogeneous equation $L\phi = 0$ which is a complete Chebyshev system on that interval, we can show that best approximations of a continuous function f must interpolate f at at least n distinct points. The argument from here parallels the presentation of Phillips [16] in deriving error bounds for polynomial approximations.

LEMMA 2. If L possesses Property W on $[0, h]$ and if ϕ^* is a best L_r approximation of a continuous function f on $[t_0, t_0 + h]$ by a solution of $L\phi = 0$, for $1 \leq r \leq \infty$, then there are n distinct points t_i satisfying $t_0 < t_1 < \cdots < t_n < t_0 + h$ at which $\phi^*(t_i) = f(t_i)$, for $1 \leq j \leq n$.

Proof. Without loss of generality, let $t_0 = 0$. Fix h such that L possesses Property W on $[0, h]$ and let $\psi_1, \cdots, \psi_n$ be a complete Chebyshev system of fundamental solutions of $L\phi = 0$ on $[0, h]$. Let f be continuous on $[0, h]$ and ϕ^* a best L_r approximation of f by a linear combination of the functions ψ_ν.

When $r = \infty$, the lemma is an immediate consequence of the familiar Chebyshev equioscillation theorem (see Meinardus [13], Theorem 23). We need only treat the cases $1 \leq r < \infty$.

Assume the difference $f - \phi^*$ is nonzero almost everywhere on $[0, h]$. Otherwise extract the distinct points $t_1, \cdots, t_n$ from the set of positive measure where the difference vanishes. Timan [29] shows that when $f - \phi^*$ is almost everywhere nonzero, then a necessary and sufficient condition for ϕ^* to be a best L_r approximation of f on $[0, h]$ is

$$\int_0^h \psi_\nu(t)\, |f(t) - \phi^*(t)|^{r-1} \operatorname{sign}\{f(t) - \phi^*(t)\}\, dt = 0, \tag{5.5}$$

for $\nu = 1$ to n and any r, $1 \leq r < \infty$.

Let $\nu = 1$ in (5.5). Since ψ_1 does not change sign on $[0, h]$, the vanishing integral implies that $f - \phi^*$ must change sign in $(0, h)$. Since the difference is continuous, there is at least one point t_1 in $(0, h)$ where $f(t_1) = \phi^*(t_1)$.

Now suppose that sign $\{f(t) - \phi^*(t)\}$ changes at the distinct points $t_1, \cdots, t_k$ in $(0, h)$ and nowhere else, and suppose $k < n$. We know $k \geq 1$. Since $\{\psi_1, \cdots, \psi_{k+1}\}$ is a Chebyshev system, there is a unique function of the form $\hat{\psi}(t) = \psi_{k+1}(t) - \sum_{\nu=1}^{k} b_\nu \psi_\nu(t)$ which vanishes and changes sign at the points t_j, for $j = 1$ to k. The product $\hat{\psi}$ sign $\{f - \phi^*\}$ does not change sign in $(0, h)$. Yet from (5.5),

$$\int_0^h \hat{\psi}(t) \operatorname{sign}\{f(t) - \phi^*(t)\}\, |f(t) - \phi^*(t)|^{r-1}\, dt$$

This contradiction implies $k \geq n$, and the lemma is proved.

Finally, we obtain an expression for the remainder when a smooth function f is interpolated by a solution of $L\phi = 0$.

LEMMA 3. Suppose L possesses Property W on $[0, h]$ and f is in $C^n[t_0, t_0 + h]$ and ϕ satisfies $L\phi = 0$ on $[t_0, t_0 + h]$ and $\phi(t_i) = f(t_i)$ at n distinct points $t_1, t_2, \cdots, t_n$ in $(t_0, t_0 + h)$. Then $f(t) - \phi(t) = (Lf(\xi))\hat{\phi}(t)$ for all t in $[t_0, t_0 + h]$, where $\hat{\phi}$ is the unique solution of $L\hat{\phi} = 1$ on $[t_0, t_0 + h]$ satisfying $\hat{\phi}(t_i) = 0$ for $j = 1$ to n and ξ is some intermediate point $\min(t, t_1) < \xi < \max(t, t_n)$.

Proof. This is Pólya's mean value theorem [17].

Proof of Theorem 4. The first part of this proof follows the argument of Phillips for algebraic polynomial approximations.

Fix h so that L possesses Property W on $[0, h]$, let f be in $C^n[t_0, t_0 + h]$ and let ϕ^* be a best L_2 (or L_r) approximation of f on $[t_0, t_0 + h]$ by a solution of $L\phi = 0$. By Lemma 2, there are n distinct points $t_1^*, \cdots, t_n^*$ in $(t_0, t_0 + h)$ at which $f(t_i^*) = \phi^*(t_i^*)$.

Construct the following functions on $[t_0, t_0 + h]$ related to ϕ^* and f:

(i) $\hat{\phi}$ is the unique solution of $L\hat{\phi} = 1$ which vanishes at $t_1^*, \cdots, t_n^*$.

(ii) $\bar{\phi}$ satisfies $L\bar{\phi} = 1$ and has minimum L_2 (or L_r) norm on $[t_0, t_0 + h]$ among all solutions of this equation. $\bar{\phi}$ as constructed is the remainder of a best L_r approximation of $\hat{\phi}$ by a solution of $L\phi = 0$; by Lemma 2 there are n distinct points $\bar{t}_1, \cdots, \bar{t}_n$ in $(t_0, t_0 + h)$ where $\bar{\phi}(\bar{t}_i) = 0$.

(iii) ϕ^{**} is the unique solution of $L\phi = 0$ which interpolates f at the points $\bar{t}_1, \cdots, \bar{t}_n$, $\phi^{**}(\bar{t}_i) = f(\bar{t}_i)$.

Existence of the functions $\hat{\phi}$, $\bar{\phi}$ and ϕ^{**} as described is assured by the assumption that L possesses Property W on $[0, h]$.

By Lemma 3 on the remainder and by the specifications of ϕ^* and $\hat{\phi}$, $f(t) - \phi^*(t) = (Lf(\xi))\hat{\phi}(t)$ for any t in $[t_0, t_0 + h]$ and some point ξ intermediate to $t, t_1^*, \cdots, t_n^*$. Similarly, $f(t) - \phi^{**}(t) = (Lf(\xi))\bar{\phi}(t)$. Now letting $||\cdot||$ denote the standard L_2 (or L_r) norm on $[t_0, t_0 + h]$, these remainder expressions and the properties of ϕ^*, ϕ^{**}, $\hat{\phi}$ and $\bar{\phi}$ imply

$$||f - \phi^*|| \leq ||f - \phi^{**}|| \leq (\max_{[t_0, t_0+h]} |Lf(s)|)\ ||\bar{\phi}||$$

and

$$||f - \phi^*|| \geq (\min_{[t_0, t_0+h]} |Lf(s)|)\ ||\hat{\phi}|| \geq (\min_{[t_0, t_0+h]} |Lf(s)|)\ ||\bar{\phi}||.$$

In addition, since Lf is continuous on $[t_0, t_0 + h]$ the two inequalities yield $||f - \phi^*|| = |Lf(\xi)|\ ||\bar{\phi}||$ for an intermediate point ξ in $[t_0, t_0 + h]$. For the particular case of L_2 approximation we can apply Theorem 3 directly to $\bar{\phi}$ and obtain $||\bar{\phi}||^2 = K_n h^{2n+1} + O(h^{2n+2})$. In turn this yields (5.3), $||f - \phi^*||^2 = (K_n + O(h))(Lf(\xi))^2 h^{2n+1}$.

An alternative derivation of the equivalent form (5.4) for K_n generalizes to approximations in the other L_r norms. Fix $t_0 = -h/2$ and consider the form of the minimum norm $||\bar{\phi}||$ on $[-h/2, h/2]$. Change the variable t and define $\Phi_h(\tau) = (2/h)^n \bar{\phi}(h\tau/2)$ for $-1 \leq \tau \leq 1$. Associated with the operator $L = \prod_{r=1}^{n} (D - \alpha_r)$ of (3.1) define the operator $M(h) = \prod_{r=1}^{n} (D_\tau - \alpha_r h/2)$, where D_τ denotes differentiation with respect to τ. Since $\bar{\phi}$ satisfies $L\bar{\phi} = 1$ on $[-h/2, h/2]$, through the change of variable Φ_h satisfies $M(h)\Phi_h = 1$ on $[-1, 1]$.

Also define an algebraic polynomial $\Phi_0(\tau) = \tau^n/n! - \sum_{r=0}^{n-1} c_r \tau^r$ where the coefficients c_r are chosen so that the L_r-norm of Φ_0 on $[-1, 1]$ is a minimum. Φ_0 satisfies $M(0)\Phi_0 = 1$ on $[-1, 1]$. The continuous dependence and differentiability of solutions of $M(h)\phi = 1$ with respect to the parameter h at 0 implies that $||\Phi_h - \Phi_0|| = O(h)$ for any L_r norm, $1 \leq r \leq \infty$, on $[-1, 1]$. From the relation $\Phi_h(\tau) = (2/h)^n \bar{\phi}(h\tau/2)$ we obtain $||\bar{\phi}|| = (h/2)^{n+1/r}\ ||\Phi_h||$, where the norms are over the intervals $[-h/2, h/2]$ and $[-1, 1]$ respectively. Together with $||\Phi_h|| = ||\Phi_0|| + O(h)$, this yields $||\bar{\phi}|| = (h/2)^{h+1/r}(||\Phi_0|| + O(h))$.

The norm $||\bar{\phi}||$ is thus expressed in terms of the norm of the algebraic polynomial Φ_0 satisfying $D^n\Phi_0 = 1$ and such that Φ_0 has minimum L_r deviation from zero on $[-1, 1]$. For the particular case of L_2 approximation, Φ_0 is expressed in terms of the nth Legendre polynomial P_n on $[-1, 1]$ as

$$\Phi_0 = (2^n P_n)/n! \binom{2n}{n};$$

see Davis [4]. The L_2 norm of $\bar{\phi}$ in this case is

$$\begin{aligned} ||\bar{\phi}|| &= (h/2)^{n+1/2}\left[2^{n+1/2}/n! \binom{2n}{n}(2n+1)^{1/2} + O(h)\right] \\ &= \left[n! \binom{2n}{n}(2n+1)^{1/2}\right]^{-1} h^{n+1/2} + O(h^{n+3/2}). \end{aligned}$$

Squaring this yields (5.4) and (5.3) from the equation $||f - \phi^*|| = |Lf(\xi)|\ ||\bar{\phi}||$ above. The proof is complete.

Remark. Because of the generality of the proof we can immediately state expressions for other best L_r approximations which are completely analogous to (5.3) for the L_2 case.

Under the assumptions of Theorem 4, the error $||f - \phi^*||$ of the best L_r approximation ϕ^* of f on $[t_0, t_0 + h]$ by a solution of $L\phi = 0$ is

$$||f - \phi^*|| = K_{n,r}(h)\ |Lf(\xi)|\ h^{n+1/r} \tag{5.6}$$

for some point ξ in $[t_0, t_0 + h]$. $K_{n,r}(h)$ is independent of f and, referring to the end of the proof above, is given by $K_{n,r}(h) = K_{n,r} + O(h)$ where

$$K_{n,r} = ||\Phi_0||/2^{n+1/r}; \tag{5.7}$$

here, as before, Φ_0 is the algebraic polynomial satisfying $D^n\Phi_0 = 1$ and which has minimum L_r deviation $||\Phi_0||$ from zero on $[-1, 1]$. As shown in the proof, ${K_{n,2}}^2 = K_n$ given by (5.4). In two other cases we can obtain explicit representations for $K_{n,r}$ from (5.7): (i) when $r = \infty$, $\Phi_0 = T_n/n!\,2^{n-1}$, where T_n is the nth-degree Chebyshev polynomial of the first kind, yielding $K_{n,\infty} = [n!\,2^{2n-1}]^{-1}$ (see Davis [4]); (ii) when $r = 1$, $\Phi_0 = U_n/n!\,2^n$, where U_n is the nth-degree Chebyshev polynomial of the second kind, yielding $K_{n,1} = [2^{2n}n!]^{-1}$ (see Davis [4], Timan [29] and Phillips [16]). Phillips [16] derives bounds for constants related to the $K_{n,r}$ for $1 < r < \infty$.

From Theorem 4 and the remarks above we can state approximation error bounds which are potentially useful in implementation of variable-knot approximation procedures.

COROLLARY 4.1. Under the assumptions of Theorem 4,

$$mK_{n,r}(h)h^{n+1/r} \leq \min_{\{\phi : L\phi = 0\}} ||f - \phi|| \leq MK_{n,r}(h)h^{n+1/r} \tag{5.8}$$

where the norm $||\cdot||$ is the standard L_r norm on $[t_0, t_0 + h]$ and m and M are lower and upper bounds respectively for $|Lf(t)|$ on $[t_0, t_0 + h]$.

Remark. In the proof of Theorem 4 we used the result of Lemma 2 that a best L_r approximation ϕ^* of f interpolates f at n distinct points $t_1^*, \cdots, t_n^*$ in $[t_0, t_0 + h]$. With appropriate regularity assumptions on the function f it is possible to relate these interpolating points to zeros of classical polynomials. For example, in uniform approximation they are asymptotically for small h the zeros of the nth-degree Chebyshev polynomial of the first kind. See Meinardus [13], Theorem 68 for a result of this nature. The rigorous development of this kind of result will be especially important for applications.

6. Global convergence. For sufficiently smooth functions f in a class $\mathfrak{F} = L_r[0, 1]$ and best approximations ϕ of f from the class $\Phi^*(p, q)$ we determine the asymptotic behavior of the total approximation error $e_{p,q}(f)$ (Eq. (3.12)) as the number of knots increases. The estimate take the form of limiting equations for $e_{p,n}(f)$ and limiting bounds derived from these equations for $e_{p,q}(f)$ when $q < n$.

Consider f in $L_2[0, 1]$ and let ϕ denote a best L_2 approximation of f from $\Phi(p, n)$. The L_2 deviation $e_{p,n}(f)$ is expressed as

$${e_{p,n}}^2(f) = \sum_{\nu=0}^{p} \int_{t_\nu}^{t_{\nu+1}} (f(t) - \phi(t))^2\, dt, \tag{6.1}$$

where $t_0 \leq t_1 \leq \cdots \leq t_{p+1}$ are the knots of ϕ. An expression analogous to (6.1) describes the rth power of the total error of a best L_r approximation of f on $[0, 1]$ by a member of $\Phi(p, n)$ when $1 \leq r < \infty$.

To determine the effect of knot optimization on $e_{p,n}(f)$ it is convenient to modify (6.1) slightly and work with an expression that makes the knot optimization explicit. Write

$$e_{p,n}{}^2(f) = \min_{(t_1,\cdots,t_p)} \sum_{\nu=0}^{p} \int_{t_\nu}^{t_{\nu+1}} (f(t) - \phi(t))^2 \, dt \tag{6.2}$$

where $0 = t_0 < t_1 < \cdots < t_p < t_{p+1} = 1$, and where over each separate subinterval $(t_\nu , t_{\nu+1})$ ϕ is the best L_2 approximation of f by a regular solution of $L\phi = 0$. When f is sufficiently smooth and when $t_{\nu+1} - t_\nu$ is small, the analysis of the preceding section applies to give an expression for $\int_{t_\nu}^{t_{\nu+1}} (f(t) - \phi(t))^2 \, dt$ in terms of Lf and $t_{\nu+1} - t_\nu$. We build on this analysis to determine the asymptotic behavior of (6.2) as p increases.

Problems of optimal partitioning or interval segmentation like (6.2) have previously been considered in several specific contexts. The earliest instance of such a problem of which we are aware is the problem of optimal stratification of a population for estimation by sampling of its mean, treated by Dalenius and Hodges in 1957. They obtain results on the convergence of a particular form similar to (6.2) and develop procedures for efficient stratification which parallel our results in the next section.

More recently Sacks and Ylvisaker [22, 23, 24] and Wahba [30] treated optimization problems like (6.2) in their analyses of a statistical design problem for linear regression. Their formulation in terms of reproducing-kernel Hilbert space yields results which transcend the immediate concerns with the design problem. The results anticipate the present ones on convergence of L_2 approximation error and they also apply to problems of optimal and efficient quadrature rules for numerical integration like those to which the present analysis is applied in [11].

Another area among many in which the optimal partitioning problem has occurred is that of organizing data into homogeneous groups. The analysis in this direction reflects strong similarities to the early work of Dalenius and Hodges.

The treatments of the partitioning problem in all of these particular instances have been carried out in manners specific to the contexts in which the problem arose. Indeed, in [10] we derived the asymptotic character of the approximation error $e_{p,n}(f)$ and the development was narrower and more cumbersome than necessary because of the details of the particular problem which were carried along in the analysis.

A more general result for partitioning problems like (6.2) is attainable and easier to prove when the essential features of the problem are isolated. We first prove such a result, Lemma 4 below, which applies readily to the present L_2 approximation problems, to the analysis of approximation error in the other L_r norms for $1 \leq r < \infty$, and to specific problems of optimal partitioning like the others mentioned above. The application to approximation-error convergence is elicited in Theorem 5 below and the remarks that follow.

For functions f in a set $\mathcal{G}$ of real-valued functions on [0, 1] consider the minimization with respect to p-point partitions $T = (t_0 , t_1 , \cdots , t_p , t_{p+1})$, where $0 = t_0 \leq t_1 \leq \cdots \leq t_p \leq t_{p+1} = 1$, of a functional of the form

$$E(f, T) = \sum_{\nu=0}^{p} e(f; t_\nu , t_{\nu+1}). \tag{6.3}$$

The following assumptions are imposed concerning the definition and properties of $E(f, T)$.

A1. For each f in $\mathcal{G}$ and any points $a < b$ in [0, 1] a nonnegative value $e(f; a, b)$ is defined. Assume $e(f; \ )$ is subadditive on contiguous subintervals of [0, 1]; that is, for any three points $a < b < c$ in [0, 1], $e(f; a, b) + e(f; b, c) \leq e(f; a, c)$.

Assumption A1 implies in particular that $E(f, T)$ is nonnegative. The subadditivity

of $e(f;\ )$ is equivalent to the assertion that if the partition S is contained in the partition T, that is if T is finer than S, then $E(f, T) \leq E(f, S)$.

The class $\mathcal{G}$ of functions f for which (6.3) is considered is restricted by the next assumption.

A2. There is a mapping J from $\mathcal{G}$ onto $PC^+[0, 1]$, nonnegative piecewise continuous functions on $[0, 1]$ that have at worst a finite number of jump discontinuities, and a constant exponent $m > 1$ such that $\lim_{h \downarrow 0} e(f; a, a + h)/h^m = Jf(a+)$ for all a in $[0, 1)$. Further, this limit is uniform in the sense that the difference $|Jf(a+) - e(f; a, a + h)/h^m|$ can be made uniformly small whenever $(a, a + h)$ is contained in an interval (α, β) that contains no discontinuity points of Jf.

Two additional assumptions are not necessary for the result, but they hold in the present setting and simplify the proof.

A3. If Jf vanishes identically on (a, b) then $e(f; a, b) = 0$.

A4. For any f in $\mathcal{G}$, $e(f; a, b)$ is a continuous function of the pair (a, b) on $0 \leq a \leq b \leq 1$.

With these assumptions on $e(f; a, b)$ and $E(f, T)$ we can characterize the asymptotic behavior of the functional

$$E_p(f) = \inf_T E(f, T), \tag{6.4}$$

where the infimum is taken with respect to all p-point partitions $T = (t_0, t_1, \cdots, t_{p+1})$ of $[0, 1]$.

LEMMA 4. If assumptions A1, A2, A3 and A4 hold for $e(f; a, b)$ then

$$\lim_{p \to \infty} p^{m-1} E_p(f) = \left(\int_0^1 (Jf(s))^{1/m}\, ds \right)^m. \tag{6.5}$$

Proof. Let f be in the class $\mathcal{G}$ and consider its image Jf in $PC^+[0, 1]$. With approximations of Jf from above and below by step functions, we reduce the proof of (6.5) to its proof for step functions Jg.

Let Jg be a step-function approximation of Jf from above, with $Jg > Jf$ throughout $[0, 1]$. The uniform convergence assumption A2 implies that $e(f; a, a + h) < e(g; a, a + h)$ if h is sufficiently small and $(a, a + h)$ does not contain a discontinuity point of Jf or Jg. With this we can fix a finite number of points $\tau_1 < \tau_2 < \cdots < \tau_k$ in $(0, 1)$, which include all the jumps of Jf and Jg and which are sufficiently close together, so that $E(f, T_\tau) < E(g, T_\tau)$ for any partition T_τ including the points $\tau_1, \cdots, \tau_k$. Thus for $p \geq k$, if T_τ denotes a p-partition including $\tau_1, \cdots, \tau_k$, we have

$$E_p(f) \leq \inf_{T_\tau} E(f, T_\tau) \leq \inf_{T_\tau} E(g, T_\tau). \tag{6.6}$$

To obtain a similar lower bound on $E_p(f)$, let Jg be a nonnegative step-function approximation of Jf from below so that $Jg < Jf$ on intervals where $Jg > 0$. If Jg vanishes identically on $(a, a + h)$ then $e(f; a, a + h) \geq e(g; a, a + h)$ by assumption A3. Thus, by the same argument as that leading to (6.6), we can fix a finite number of points $\tau_1 < \tau_2 < \cdots < \tau_k$ in $(0, 1)$, which include all the jumps of Jf and Jg and which are sufficiently close together, so that $E(f, T_\tau) \geq E(g, T_\tau)$ for any partition including $\tau_1, \cdots, \tau_k$. Let T be an arbitrary p-partition and let T_τ denote the finer $(p + k)$-partition that adds the points $\tau_1, \cdots \tau_k$ to T. By assumption A1, $E(f, T) \geq E(f, T_\tau)$.

Thus

$$E_p(f) \geq \inf_{T_\tau} E(f, T_\tau) \geq \inf_{T_\tau} E(g, T_\tau), \tag{6.7}$$

where the last infima are with respect to the constrained $(p + k)$-partitions T_τ.

The inequalities (6.6) and (6.7) suggest consideration of the functionals $E(g, T_\tau)$ alone, where Jg is a nonnegative step function and T_τ is a p-partition of [0, 1] that includes the discontinuity points of Jg. By assumption A4, $E(g, T_\tau)$ is a continuous function of the free partition points of T_τ. Since these free partition points range over a compact simplex, the value $\inf_{T_\tau} E(g, T_\tau)$ is attained for some p-partition T_p. For $p \geq k$ consider a sequence $\{T_p\}_{p=k}^{\infty}$ of optimal constrained p-partitions; $E(g, T_p) = \inf_{T_\tau} E(g, T_\tau)$ and each T_p includes the k fixed points $\tau_1, \cdots, \tau_k$. Let $t_0, t_1, \cdots, t_{p+1}$ denote the points of T_p. The values t_ν depend on p, but we do not make this explicit in the notation when there is no risk of ambiguity. Also denote $\tau_0 = 0$ and $\tau_{k+1} = 1$.

Now $E(g, T_p) = \sum_{\mu=0}^{k} \sum_{\tau_\mu \leq t_\nu < \tau_{\mu+1}} e(g; t_\nu, t_{\nu+1})$. First consider the contribution to $E(g, T_p)$ by the terms $e(g; t_\nu, t_{\nu+1})$ for the t_ν that lie in one of the intervals $[\tau_\mu, \tau_{\mu+1})$ and assume $p \gg k$. Let q_μ denote the number of points t_ν in this interval and let c_μ denote the constant value of Jg on $(\tau_\mu, \tau_{\mu+1})$. Let $c_\mu > 0$; otherwise $e(g; t_\nu, t_{\nu+1}) = 0$ by assumption A3. We note that q_μ must increase as p does and deduce the asymptotic behavior of

$$E_{q_\mu}(g; \tau_\mu, \tau_{\mu+1}) = \sum_{\tau_\mu \leq t_\nu < \tau_{\mu+1}} e(g; t_\nu, t_{\nu+1}). \tag{6.8}$$

By assumption A1, $e(g; a, a + h)$ is nondecreasing in h and by assumption A2 it is strictly positive when $(a, a + h) \subseteq (\tau_\mu, \tau_{\mu+1})$, since $Jg = c_\mu > 0$ on this interval. Thus if $\max_{\tau_\mu \leq t_\nu < \tau_{\mu+1}} (t_{\nu+1} - t_\nu)$ were bounded away from zero then $E_{q_\mu}(g; \tau_\mu, \tau_{\mu+1})$ would be bounded away from zero. It is not, as is easily seen from the asymptotic form of $E(g, T)$ for uniform partitions of [0, 1]. Thus $\max_{\tau_\mu \leq t_\nu < \tau_{\mu+1}} (t_{\nu+1} - t_\nu)$ must go to zero as p increases and this implies further that $q_\mu \to \infty$ as $p \to \infty$.

Since $\max_{\tau_\mu \leq t_\nu < \tau_{\mu+1}} (t_{\nu+1} - t_\nu) \to 0$ and $q_\mu \to \infty$ as $p \to \infty$, assumption A2 can be applied to obtain a lower bound on (6.8):

$$q_\mu^{m-1} E_{q_\mu}(g; \tau_\mu, \tau_{\mu+1}) = q_\mu^{m-1} \sum_{\tau_\mu \leq t_\nu < \tau_{\mu+1}} (c_\mu + o(1)) h_\nu^m \geq q_\mu^{m-1} (\min_{t_\nu} (c_\mu + o(1))) \sum_{t_\nu} h_\nu^m,$$

where $h_\nu = t_{\nu+1} - t_\nu$. Since $\sum_{t_\nu} h_\nu^m$ subject to the constraint $\sum_{t_\nu} h_\nu = \tau_{\mu+1} - \tau_\mu$ is minimized by $h_\nu = (\tau_{\mu+1} - \tau_\mu)/q_\mu$, we obtain

$$q_\mu^{m-1} E_{q_\mu}(g; \tau_\mu, \tau_{\mu+1}) \geq (\tau_{\mu+1} - \tau_\mu)^m \min_{t_\nu} (c_\mu + o(1)).$$

By assumption A2 the terms $o(1)$ are uniformly small as $\max h_\nu$ goes to zero, so the lower bound approaches a limit and yields

$$\liminf_{q_\mu \to \infty} q_\mu^{m-1} E_{q_\mu}(g; \tau_\mu, \tau_{\mu+1}) \geq c_\mu (\tau_{\mu+1} - \tau_\mu)^m. \tag{6.9}$$

The lower bound of (6.9) is actually attained as a limit when q_μ partition points are uniformly spaced in $[\tau_\mu, \tau_{\mu+1})$. Since the partitions T_p are optimal, this implies

$$\lim_{q_\mu \to \infty} q_\mu^{m-1} E_{q_\mu}(g; \tau_\mu, \tau_{\mu+1}) = c_\mu (\tau_{\mu+1} - \tau_\mu)^m. \tag{6.10}$$

The asymptotic behavior of $E(g, T_p)$ is determined from (6.10) by necessary conditions on the integers q_μ associated with the optimal partitions. In particular, $\sum_{\mu=0}^{k} q_\mu = p$ and $0 \leq q_\mu/p \leq 1$ for each μ. Suppose the ratios q_μ/p converge as $p \to \infty$ and denote

$m_\mu = \lim_{p\to\infty} q_\mu/p$. The limits must exist for a subsequence of $\{T_p\}$ and in the end we conclude that the ratios must converge for the original sequence.

Expressing $E(g, T_p)$ in terms of the $E_{q_\mu}(g; \tau_\mu, \tau_{\mu+1})$, we obtain

$$\lim_{p\to\infty} p^{m-1}E(g, T_p) = \sum_{\mu=0}^{k} \lim_{p\to\infty} (p/q_\mu)^{m-1} q_\mu^{m-1} E_{q_\mu}(g; \tau_\mu, \tau_{\mu+1})$$
$$= \sum_{\mu=0}^{k} m_\mu^{1-m} c_\mu (\tau_{\mu+1} - \tau_\mu)^m, \tag{6.11}$$

using (6.10) and the observation that $q_\mu \to \infty$ as $p \to \infty$. Since the partitions T_p are optimal, the expression on the right of (6.11) is minimal with respect to its dependence on the quantities m_μ. The m_μ satisfy $m_\mu \geq 0$ and $\sum_{\mu=0}^{k} m_\mu = 1$.

The minimum of $\sum_{\mu=0}^{k} m_\mu^{1-m} c_\mu (\tau_{\mu+1} - \tau_\mu)^m$ with respect to the m_μ is attained for $m_\mu = c_\mu^{1/m}(\tau_{\mu+1} - \tau_\mu)/[\sum_{\nu=0}^{k} c_\nu^{1/m}(\tau_{\nu+1} - \tau_\nu)]$ and the minimum value is $[\sum_{\mu=0}^{k} c_\mu^{1/m} (\tau_{\mu+1} - \tau_\mu)]^m$. Recalling that c_μ is the constant value of Jg on $\tau_{\mu+1} - \tau_\mu$, we obtain from (6.11) that

$$\lim_{p\to\infty} p^{m-1}E(g, T_p) = \left[\int_0^1 (Jg(s))^{1/m}\, ds\right]^m,$$

which proves (6.5) when Jg is a step function.

To complete the proof for arbitrary f in $\mathcal{G}$, use (6.6) to obtain

$$\limsup_{p\to\infty} p^{m-1}E_p(f) \leq \lim_{p\to\infty} p^{m-1} \inf_{T_\tau} E(g, T_\tau) = \lim_{p\to\infty} p^{m-1}E(g, T_p) = \left(\int_0^1 (Jg(s))^{1/m}\, ds\right)^m,$$

where Jg is the step function approximation of Jf from above. Thus

$$\limsup_{p\to\infty} p^{m-1}E_p(f) \leq \left(\int_0^1 (Jf(s))^{1/m}\, ds\right)^m.$$

Using (6.7) similarly, we obtain

$$\liminf_{p\to\infty} p^{m-1}E_p(f) \geq \left(\int_0^1 (Jf(s))^{1/m}\, ds\right)^m.$$

Together these inequalities imply (6.5) and the lemma is proved.

An immediate consequence of Lemma 4 with the local error estimates of Sec. 5 is the following main result on convergence of $e_{p,n}(f)$ (Eq. (6.1)).

THEOREM 5. Let f be in $C^n[0, 1]$. The error $e_{p,n}(f)$ of the best L_2 approximation of f on $[0, 1]$ in the class $\Phi(p, n)$ satisfies

$$\lim_{p\to\infty} p^n e_{p,n}(f) = K_n^{1/2}\left(\int_0^1 |Lf(s)|^{2/(2n+1)}\, ds\right)^{(2n+1)/2} \tag{6.12}$$

where K_n is given by (5.4).

Proof. The squared error $e_{p,n}^2(f)$ is expressed in (6.1) as $e_{p,n}^2(f) = \sum_{\nu=0}^{p} e(f; t_\nu, t_{\nu+1})$ where $e(f; t_\nu, t_{\nu+1}) = \min_{\{\phi: L\phi=0\}} \int_{t_\nu}^{t_{\nu+1}} (f(t) - \phi(t))^2\, dt$.

Apply Lemma 4. The local squared errors $e(f; t_\nu, t_{\nu+1})$ are nonnegative and satisfy the subadditivity and continuity assumptions A1 and A4 for any f in $L_2[0, 1]$. By Theorem 4, assumptions A2 and A3 are satisfied by $e(f; a, b)$ for any f in the class of functions $\mathcal{G}$ having $(n - 1)$ continuous derivatives, absolutely continuous $(n - 1)$st derivative and nth derivative in $PC[0, 1]$. From Theorem 4, $Jf = K_n(Lf)^2$ and the exponent is $m = 2n + 1$. The limit (6.12) follows as the square root of (6.5) and the theorem is proved.

Remarks. (I) The regularity assumption on f in Theorem 5 can clearly be relaxed to hold piecewise on [0, 1], as long as $D^n f$ admits at worst jump discontinuities.

(II) The same argument using Lemma 4 applies to give the asymptotic form of the error $e_{p,n}(f)$ for best L_r approximation when $1 \leq r < \infty$. Express $e_{p,n}{}^r(f)$ in the form (6.3) and use the estimate (5.6) for local approximation error. Immediately, if f is in $C^n[0, 1]$ then the error $e_{p,n}(f)$ of the best L_r approximation, $1 \leq r < \infty$, of f on [0, 1] in the class $\Phi(p, n)$ satisfies

$$\lim_{p\to\infty} p^n e_{p,n}(f) = K_{n,r}\left(\int_0^1 |Lf(s)|^{r/(rn+1)}\, ds\right)^{(rn+1)/r}, \tag{6.13}$$

where $K_{n,r}$ is given by (5.7).

(III) For sup norm approximation ($r = \infty$), Lemma 4 is not applied since total approximation error is not expressed as a sum of local error terms. The argument for this case is easier.

The total uniform error of a best segmented approximation is given by the minmax expression $e_{p,n}(f) = \min_{0<t_1<\cdots<t_p<1} \max_{t_\nu \leq t \leq t_{\nu+1}} |f(t) - \phi(t)|$. If f is continuous then the errors over separate subintervals are balanced by an optimal ϕ in $\Phi(p, n)$; see Meinardus [13]. That is, $e_{p,n}(f) = \max_{t_\nu \leq t \leq t_{\nu+1}} |f(t) - \phi(t)|$, independent of ν. Since the error goes to zero as $p \to \infty$, the interval lengths $t_{\nu+1} - t_\nu$ must go to zero where Lf is continuous and positive. The estimate (5.6) applies to give $e_{p,n}(f) = K_{n,\infty}(h_\nu)\, |Lf(\xi)|\, h_\nu{}^n$, where $h_\nu = t_{\nu+1} - t_\nu$ and $t_\nu \leq \xi \leq t_{\nu+1}$. Taking the nth root and summing over ν yields

$$\lim_{p\to\infty} p(e_{p,n}(f))^{1/n} = K_{n,\infty}{}^{1/n} \int_0^1 |Lf(s)|^{1/n}\, ds.$$

Thus if f is in $C^n[0, 1]$, then the error $e_{p,n}(f)$ of the best uniform approximation of f on [0, 1] in the class $\Phi(p, n)$ satisfies

$$\lim_{p\to\infty} p^n e_{p,n}(f) = K_{n,\infty}\left(\int_0^1 |Lf(s)|^{1/n}\, ds\right)^n, \tag{6.14}$$

where $K_{n,\infty} = [n!\, 2^{2n-1}]^{-1}$ from the remark following (5.7).

(IV) Sacks and Ylvisaker [24] obtain expression (6.12) for best L_2 segmented approximations by polynomials, i.e. the case $L = D^n$. They impose some unnecessary restrictions on the zeros of $D^n f$ and mention that these can be relaxed. They obtain (6.12) as the limit of the error for prescribed partitions defined by "regular sequences". In effect, the "regular sequence" construction builds on the notion of an asymptotic density for the sequences of partition points. Such constructions of asymptotically efficient partitions are further justified by the analysis of the next section.

Wahba [30], building on the results of Sacks and Ylvisaker, establishes a result which includes (6.12) when the null space of L is spanned by an extended complete Chebyshevian system and when the function Lf is strictly positive. The analysis of Sec. 5 shows that Wahba's assumption on the null space of L is satisfied in an appropriate local sense by any differential operator of the form (3.1) considered.

Though the thorough and elegant analyses of Sacks and Ylvisaker and of Wahba yield essentially the same result as Theorem 5 for L_2 approximation, their Hilbert space formulations preclude immediate extension of their arguments to obtain results like (6.13) and (6.14) for approximation in other L_r norms.

(V) Phillips [16] has obtained (6.13) as the limiting error for best balanced-error L_r approximation by segmented polynomials. A balanced-error segmented approximation is

one for which approximation errors over distinct segments of the interval are equal. The best L_r segmented approximation and the best balanced-error L_r segmented approximation need coincide only when $r = \infty$. Phillip's result together with (6.13) imply that balanced-error approximations are asymptotically efficient. This also follows from the analysis of the next section.

(VI) Lemma 4 also applies to give the precise order of convergence of the error of best weighted L_r approximations in $\Phi(p, n)$. This entails only a minor adaptation of the arguments of Sec. 5 to obtain corresponding local error estimates for weighted approximations of f.

As a corollary of Theorem 5 we obtain sharp bounds on the rate of convergence of approximations from the classes $\Phi(p, q)$ when $q < n$.

COROLLARY 5.1. Let f be in $C^n[0, 1]$, let integer q satisfy $1 \leq q \leq n$, and let m be the smallest integer satisfying $mq \geq n$. The error $e_{p,q}(f)$, defined by (3.12), of the best L_2 approximation of f on $[0, 1]$ in the class $\Phi^*(p, q)$ satisfies

$$\liminf_{p\to\infty} p^n e_{p,q}(f) \geq K_n^{1/2}\left(\int_0^1 |Lf(s)|^{2/(2n+1)}\, ds\right)^{(2n+1)/2} \tag{6.15}$$

and

$$\limsup_{p\to\infty} p^n e_{p,q}(f) \leq m^n K_n^{1/2}\left(\int_0^1 |Lf(s)|^{2/(2n+1)}\, ds\right)^{(2n+1)/2}. \tag{6.16}$$

Proof. By the remark at the end of Sec. 3, $\Phi^*(p, q) \subseteq \Phi(p, n)$ and $\Phi([p/m], n) \subseteq \Phi^*(p, q)$. Thus $e_{p,q}(f) \geq e_{p,n}(f)$ and $e_{[p/m],n}(f) \geq e_{p,q}(f)$. Applying (6.12) for the limiting forms of $e_{p,n}(f)$ and $e_{[p/m],n}(f)$ yields (6.15) and (6.16) respectively.

Remarks. (VII) The argument of Corollary 5.1 applies as well to obtain bounds like (6.15) and (6.16) on L_r convergence of approximations from $\Phi(p, q)$. For the L_∞ case we recall the observation made in Sec. 3 that when f is continuous then $e_{p,q}(f) = \inf_{\phi\in\Phi(p,q)} \|f - \phi\| = \inf_{\phi\in\Phi^*(p,q)} \|f - \phi\|$; the minimum error $e_{p,q}(f)$ is actually attained for some function in $\Phi^*(p, q)$. From (6.13) and (6.14) we infer that if f is in $C^n[0, 1]$, $1 \leq q \leq n$, and m is the smallest integer satisfying $mq \geq n$, then the error $e_{p,q}(f)$ of the best L_r approximation, $1 \leq r \leq \infty$, of f on $[0, 1]$ in the class $\Phi^*(p, q)$ satisfies

$$\liminf_{p\to\infty} p^n e_{p,q}(f) \geq K_{n,r}\left(\int_0^1 |Lf(s)|^{r/(rn+1)}\, ds\right)^{(rn+1)/r} \tag{6.17}$$

and

$$\limsup_{p\to\infty} p^n e_{p,q}(f) \leq m^n K_{n,r}\left(\int_0^1 |Lf(s)|^{r/(rn+1)}\, ds\right)^{(rn+1)/r}; \tag{6.18}$$

when $r = \infty$, the exponent $r/(rn + 1)$ becomes $1/n$.

(VIII) Sacks and Ylvisaker [24] establish a much sharper result than Corollary 5.1 for the particular case $n = 2, q = 1$ and $L = D^2$. In this case $\Phi(p, 1)$ describes *continuous* piecewise linear functions. Under appropriate regularity conditions on f, they show that the lower bound of (6.15) is actually the limit of $p^2 e_{p,1}(f)$ as $p \to \infty$; see their Theorem 4 and the related discussion in [24]. This is apparently the sharpest result of this type for convergence of spline approximations, with continuity conditions imposed on the approximating functions at their knots.

(IX) Burchard [3] has obtained an upper bound similar in form to (6.18) for L_r approximation by polynomial splines, i.e. the case $L = D^n$. This bound reflects the

dependence of the error on the integral of $(D^n f)^{r/(rn+1)}$, but it is less sharp in reflecting dependence on the order n of the operator. Burchard also refers to interesting bounds of Rice on the convergence of spline approximations to functions f with prescribed singularities.

The limiting equations (6.12), (6.13) and (6.14) have interesting consequences for the characterization of optimal partitions. We explore these in the next section.

7. Optimal and efficient partitions. From the sharp global error estimates of the preceding section we can derive precise asymptotic characterizations of optimal partitions. Such characterizations are important for the development of computational procedures for good segmented approximation since knot optimization is the most difficult computational problem. The asymptotic characterizations are useful even when the number of knots is small. They form a sound theoretical basis for algorithmic initialization of iterative knot optimization schemes.

To present an argument general enough to encompass segmented approximation in any L_r norm, we phrase the argument in terms of the functionals and assumptions underlying Lemma 4; see (6.3), (6.4) and assumptions A1 to A4 of Sec. 6.

First, from assumption A4 it follows that optimal p-partitions exist. For any f in $\mathcal{G}$, since $e(f; a, b)$ is continuous in (a, b) then also $E(f, T)$ given by (6.3) is a continuous function of the points $t_1, t_2, \cdots, t_p$ comprising a p-partition T. From the compactness of the region over which these points range, it follows that the value $E_p(f)$ given by (6.4) is attained for some p-partition $T_p(f)$; that is,

$$E(f, T_p(f)) = E_p(f). \tag{7.1}$$

When the functional $E(f, T)$ describes approximation error we can also conclude the existence of $T_p(f)$ from the existence theorem of Sec. 3.

The points in $T_p(f)$ need not be distinct, but we can assume they are, if this is convenient, from the subadditivity assumption A1 on $e(f; a, b)$. Again, when $E(f, T)$ describes approximation error then either $E_p(f) = 0$ or the points of $T_p(f)$ will be distinct. This follows from a result on strict monotonicity of error for γ-polynomial approximations; see Rice [21].

Further, an optimal partition $T_p(f)$ is not necessarily unique. The example in Sec. 4 illustrates this. But in spite of nonuniqueness we can still find necessary conditions on the distribution of knots in $T_p(f)$ as p increases. Such a characterization is suggested in the proof of Lemma 4 where we derive properties of the optimal partitions associated with step functions Jg.

Let f be a function in $\mathcal{G}$ and let $\{T_p(f)\}_{p=1}^{\infty}$ be a sequence of optimal p-partitions for minimizing the functional $E(f, T)$. Associated with each partition $T_p(f)$ define the distribution function $G_{f,p}$ of its knots so that for each t in [0, 1], $G_{f,p}(t)$ is the proportion of knots in $T_p(f)$ which are less than or equal to t. Symbolically,

$$G_{f,p}(t) = \frac{1}{p+2} \# (T_p(f) \leq t). \tag{7.2}$$

The divisor $p + 2$ normalizes $D_{f,p}$ for inclusion of the points $t_0 = 0$ and $t_{p+1} = 1$ in $T_p(f)$. We show that for each t in [0, 1] the sequence $\{G_{f,p}(t)\}_{p=1}^{\infty}$ converges to the value of a distribution function defined by the function Jf. The limiting distribution G_f reflects the form of the limit (6.5) and is given by

$$G_f(t) = \int_0^t (Jf(s))^{1/m}\, ds \Big/ \int_0^1 (Jf(s))^{1/m}\, ds, \tag{7.3}$$

when Jf is not identically zero.

A special case of the following result was proven within Lemma 4.

LEMMA 5. Let f be in $\mathcal{G}$ and suppose assumptions A1, A2, A3 and A4 of Lemma 4 hold for $e(f; a, b)$. Define $G_{f,p}$ by (7.1) and (7.2) and G_f by (7.3). If Jf is not identically zero on [0, 1], then

$$\lim_{p\to\infty} G_{f,p}(t) = G_f(t) \tag{7.4}$$

for all t in [0, 1].

Proof. By Helly's selection theorem there is a distribution function F concentrated on [0, 1] and a subsequence $\{G_{f,p_k}\}_{k=1}^{\infty}$ of $\{G_{f,p}\}$ such that $\lim_{k\to\infty} G_{f,p_k}(t) = F(t)$ at continuity points of F. Without loss of generality assume this convergence holds for the original sequence; $\lim_{p\to\infty} G_{f,p}(t) = F(t)$. (This is in fact necessary, since we show the limit F is uniquely determined.)

F is monotone nondecreasing, continuous from the right, and has at most a countable number of discontinuities in [0, 1]. Also, F is a proper distribution function with $F(0) = 0$ and $F(1) = 1$. The function G_f is also monotone nondecreasing and continuous on [0, 1], with $G_f(0) = 0$ and $G_f(1) = 1$. From these properties of F and G_f it will suffice to show that $F(t) = G_f(t)$ for all t in (0, 1) where F is continuous and $0 < G_f(t) < 1$. Fix such a point t.

Form $(p + 1)$-partitions T_p^+ by adding the point t to the p-partition $T_p(f)$. From the subadditivity assumption A1 and the optimality of the partitions $T_p(f)$ we obtain $E(f, T_{p+1}(f)) \leq E(f, T_p^+) \leq E(f, T_p)$. These bounds and Lemma 4 imply

$$\lim_{p\to\infty} p^{m-1} E(f, T_p^+) = \left(\int_0^1 (Jf(s))^{1/m}\, ds\right)^m. \tag{7.5}$$

Now decompose $E(f, T_p^+)$ into two parts,

$$E(f, T_p^+) = \sum_{t_\nu < t} e(f; t_\nu, t_{\nu+1}) + \sum_{t_\nu \geq t} e(f; t_\nu, t_{\nu+1}),$$

where the t_ν denote the points of T_p^+. From the definition of $G_{f,p}$, there are $p_1 = (p + 2) \cdot G_{f,p}(t)$ points t_ν in $(0, t)$ contributing to the first sum and $p_2 = (p + 2)(1 - G_{f,p}(t))$ points t_ν in $[t, 1)$ contributing to the second sum. The values of p_1 and p_2 could be altered insignificantly if t were already a partition point of $T_p(f)$. Since $0 < G_f(t) < 1$, both p_1 and p_2 must increase without bound as p increases; otherwise one of the two sums in the decomposition of $E(f, T_p^+)$ would be bounded away from zero, by the same argument used for this point in the proof of Lemma 4.

Apply Lemma 4 to the restrictions of f to $[0, t]$ and $[t, 1]$, respectively, to obtain lower bounds on limits of the two sums in the decomposition of $E(f, T_p^+)$. In particular,

$$\liminf_{p_1\to\infty} p_1^{m-1} \sum_{t_\nu < t} e(f; t_\nu, t_{\nu+1}) \geq \left(\int_0^t (Jf(s))^{1/m}\, ds\right)^m$$

and

$$\liminf_{p_2\to\infty} p_2^{m-1} \sum_{t_\nu \geq t} e(f; t_\nu, t_{\nu+1}) \geq \left(\int_t^1 (Jf(s))^{1/m}\, ds\right)^m.$$

From these inequalities we infer a lower bound on (7.5) involving $F(t)$:

$$\begin{aligned}\lim_{p\to\infty} p^{m-1}E(f, T_p^+) &\geq \liminf_{p\to\infty} (p/p_1)^{m-1}p_1^{m-1} \sum_{t_r<t} e(f; t_r, t_{r+1}) \\ &\quad + \liminf_{p\to\infty} (p/p_2)^{m-1}p_2^{m-1} \sum_{t_r\geq t} e(f; t_r, t_{r+1}) \\ &\geq \left(\int_0^t (Jf(s))^{1/m}\, ds\right)^m \lim_{p\to\infty} (p/p_1)^{m-1} \\ &\quad + \left(\int_t^1 (Jf(s))^{1/m}\, ds\right)^m \lim_{p\to\infty} (p/p_2)^{m-1} \\ &= \left(\int_0^t (Jf(s))^{1/m}\, ds\right)^m F(t)^{1-m} + \left(\int_t^1 (Jf(s))^{1/m}\, ds\right)^m (1 - F(t))^{1-m},\end{aligned}$$

where the last equation derives from the convergence of $G_{f,p}(t)$ to $F(t)$. Dividing both sides of this inequality by the expression on the right of (7.5) yields

$$1 \geq G_f(t)^m F(t)^{1-m} + (1 - G_f(t))^m (1 - F(t))^{1-m}. \tag{7.6}$$

Finally, as a function of the value $F(t)$ the expression on the right of (7.6) is strictly greater than one for $F(t) \neq G_f(t)$. Thus (7.6) implies $F(t) = G_f(t)$ and the lemma is proved.

As an immediate consequence of Lemma 5 we obtain a characterization of the knot distribution of optimal segmented approximations from $\Phi(p, n)$.

THEOREM 6. Let f be in $C^n[0, 1]$. For each integer $p \geq 1$ let $T_p(f)$ be a partition of $[0, 1]$ defined by the knots of a best L_2 approximation of f in $\Phi(p, n)$. The knot distributions $G_{f,p}$ defined by (7.2) converge to

$$G_f(t) = \int_0^t |Lf(s)|^{2/(2n+1)}\, ds \Big/ \int_0^1 |Lf(s)|^{2/(2n+1)}\, ds \tag{7.7}$$

for t in $[0, 1]$ as $p \to \infty$.

Proof. The proof of Theorem 5 relates the L_2 approximation error to the forms (6.3) and (6.4). The limiting distribution (7.7) follows from Lemma 5 and Theorem 5.

Remarks. (I) Another view of Theorem 6 says that optimal knots for best L_2 segmented approximations of f in $\Phi(p, n)$ must have an asymptotic density given by

$$g_f(t) = |Lf(t)|^{2/(2n+1)} \Big/ \int_0^1 |Lf(s)|^{2/(2n+1)}\, ds. \tag{7.8}$$

In loose terms, optimal knots cluster in regions where $|Lf(t)|$ is relatively large and they are sparser in regions where $|Lf(t)|$ is small. In an asymptotic sense, Theorem 6 is a verification of fidelity between piecewise regular patterns f and their representations by segmented approximations in $\Phi(p, n)$; the knots of best approximations tend to ferret out localized regions of irregularity of the functions they approximate.

(II) Results analogous to Theorem 6 follow as readily from Lemma 5 and (6.13) for knot distributions of best L_r approximations from $\Phi(p, n)$ when $1 \leq r < \infty$. If f is in $C^n[0, 1]$ and $G_{f,p}$ is the distribution function of the knots associated with a best L_r approximation of f in $\Phi(p, n)$ then $G_{f,p}(t)$ converges to

$$G_f(t) = \int_0^t |Lf(s)|^{r/(rn+1)}\, ds \Big/ \int_0^1 |Lf(s)|^{r/(rn+1)}\, ds \tag{7.9}$$

for t in [0, 1] as $p \to \infty$. From (7.9) we can describe the asymptotic density of knots of best L_r approximations from $\Phi(p, n)$ by

$$g_f(t) = |Lf(t)|^{r/(rn+1)} \Big/ \int_0^1 |Lf(s)|^{r/(rn+1)}\, ds. \tag{7.10}$$

(III) The asymptotic knot distribution for best L_∞ segmented approximations of f in $\Phi(p, n)$ is also described by (7.9), where $1/n$ replaces the exponent $r/(rn + 1)$. This does not follow from Lemma 5, but from (6.14) directly. A limiting knot distribution F exists by Helly's selection theorem. At points t where F is continuous and $0 < G_f(t) < 1$, using (6.14) and an argument similar to that in the proof of Lemma 5, we obtain $1 \geq \max\{[G_f(t)/F(t)]^n, [(1 - G_f(t))/(1 - F(t))]^n\}$. In turn, this implies that the limiting distribution F must be G_f.

(IV) The form of the asymptotic knot density (7.10) implies that optimal L_r segmented approximations are asymptotically balanced-error segmented approximations; that is, L_r deviations over distinct segments are asymptotically the same. For $1 \leq r < \infty$, the rth power of the L_r deviation over any segment of a best approximation in $\Phi(p, n)$ is $e(f: t_\nu, t_{\nu+1}) = K_{n,r}{}^r(h_\nu)\, |Lf(\xi)|^r\, h_\nu^{rn+1}$ from (5.6), where $h_\nu = t_{\nu+1} - t_\nu$ and $t_\nu \leq \xi \leq t_{\nu+1}$. The knot density near ξ is approximately $g_f(\xi)$ from (7.10). If $g_f(\xi) > 0$, then the knot spacing near ξ is approximately $(pg_f(\xi))^{-1} = h_\nu$. Substitution of this value in the local error estimate yields

$$\lim_{p\to\infty} p^{rn+1} e(f; t_\nu, t_{\nu+1}) = K_{n,r}{}^r \left(\int_0^1 |Lf(s)|^{r/(rn+1)}\, ds \right)^{(rn+1)},$$

which is independent of the interval $(t_\nu, t_{\nu+1})$.

Theorem 6 and the related remarks give necessary conditions on locations of optimal knots for best segmented approximations. The characterizations can be turned around and used to construct partitions for which associated best L_r approximations are asymptotically efficient. The characterizations are sufficient for guiding selection of "asymptotically best" knot sets. Two such constructions are presented which are based respectively on the density characterization and the balanced-error characterization of optimal partitions.

Such constructions of asymptotically efficient partitions were first presented by Dalenius in the context of the stratified sampling problem. Together with Hodges, he proved the asymptotic efficiency of stratification schemes based on analogues in the approximation problem of both the density and balanced-error constructions mentioned here. Later, in the context of the regression-design problem, Sacks and Ylvisaker demonstrated that designs constructed by the density approach are asymptotically efficient. This result was part of a constructive proof of a result like Lemma 4 in which the integral expression of (6.5) was first shown to be a lower bound and then it was shown that this lower bound is attained by a so-called "regular sequence" design.

As with the other problems already treated regarding the asymptotic behavior of approximation error, results on construction of efficient partitions follow in the more general setting described before Lemma 4. We first consider the density approach for constructing partitions within this setting.

Let g be a strictly positive, bounded, piecewise-continuous function on [0, 1] which is normalized so that $\int_0^1 g(s)\, ds = 1$. Let G denote its integral; $G(t) = \int_0^t g(s)\, ds$. For any integer $p \geq 1$, we can define a p-partition $T_{g,p}$ of [0, 1] by

$$T_{g,p} = \{t_\nu \in [0, 1]: G(t_\nu) = \nu/(p + 1), \nu = 0, 1, \cdots, p + 1\}. \tag{7.11}$$

The points t_ν of $T_{g,p}$ are uniquely defined through inversion of the distribution G.

The asymptotic efficiency of the forms $E(f, T_{g,p})$ of (6.3) relative to $E_p(f)$ of (6.4) is deduced. To avoid tedious excursions in the argument to account for possible discontinuities of the function Jf, we strengthen A2 to suppose that Jf is continuous throughout [0, 1].

LEMMA 6. Let f be in $\mathcal{G}$ and suppose assumptions A1, A2, A3 and A4 of Lemma 4 hold for $e(f; a, b)$. Further assume that Jf is continuous on [0, 1]. Let g be a positive, bounded and piecewise-continuous density on [0, 1] with associated distribution G. For the sequence of partitions $\{T_{g,p}\}_{p=1}^{\infty}$ of (7.11)

$$\lim_{p\to\infty} p^{m-1}E(f, T_{g,p}) = \int_0^1 Jf(s)[g(s)]^{1-m}\, ds. \tag{7.12}$$

If Jf is strictly positive on [0, 1] then the partitions $T_{g,p}$ associated with

$$g(t) = (Jf(t))^{1/m} \Big/ \int_0^1 (Jf(s))^{1/m}\, ds$$

are asymptotically efficient; that is,

$$\lim_{p\to\infty} p^{m-1}E(f, T_{g,p}) = \left(\int_0^1 (Jf(s))^{1/m}\, ds\right)^m. \tag{7.13}$$

Proof. The points t_ν of $T_{g,p}$ satisfy $G(t_\nu) = \nu/(p + 1)$. To obtain bounds on their spacing express $g_\nu' = \min_{t_\nu \le t \le t_{\nu+1}} g(t)$ and $g_\nu'' = \max_{t_\nu \le t \le t_{\nu+1}} g(t)$. Since $G' = g$,

$$(t_{\nu+1} - t_\nu)g_\nu' \le G(t_{\nu+1}) - G(t_\nu) \le (t_{\nu+1} - t_\nu)g_\nu''.$$

Thus, denoting $h_\nu = t_{\nu+1} - t_\nu$, $((p + 1)g_\nu'')^{-1} \le h_\nu \le ((p + 1)g_\nu')^{-1}$. The spacings h_ν are strictly positive and converge to zero as $p \to \infty$.

Apply assumption A2 to obtain

$$\begin{aligned}
\limsup_{p\to\infty} p^{m-1}E(f, T_{g,p}) &= \limsup_{p\to\infty} p^{m-1} \sum_{\nu=0}^{p} (Jf(t_\nu) + o(1))h_\nu^m \\
&\le \limsup_{p\to\infty} \sum_{\nu=0}^{p} (Jf(t_\nu) + o(1))(g_\nu')^{1-m}h_\nu \\
&= \int_0^1 Jf(s)[g(s)]^{1-m}\, ds.
\end{aligned}$$

Using the lower bound $h_\nu \ge [(p + 1)g_\nu'']^{-1}$ similarly, we obtain

$$\liminf_{p\to\infty} p^{m-1}E(f, T_{g,p}) \ge \int_0^1 Jf(s)[g(s)]^{1-m}\, ds,$$

which proves (7.12). Eq. (7.13) follows by direct substitution.

The implication for the approximation problem is

THEOREM 7. Let f be in $C^n[0, 1]$. Let g be a positive, bounded and piecewise-continuous density on [0, 1]. Let ϕ_p be the best L_2 approximation of f in $\Phi(p, n)$ with knots at the points of $T_{g,p}$ defined by (7.11). Then

$$\lim_{p\to\infty} p^n \, ||f - \phi_p|| = K_n^{1/2}\left(\int_0^1 (Lf(s))^2[g(s)]^{-2n}\, ds\right)^{1/2}. \tag{7.14}$$

If $|Lf|$ is strictly positive and $g = g_f$ of (7.8), then

$$\lim_{p\to\infty} p^n \, ||f - \phi_p|| = K_n^{1/2}\left(\int_0^1 |Lf(s)|^{2/(2n+1)}\, ds\right)^{(2n+1)/2}.$$

Proof. Lemma 6.

Remarks. (V) That partitions described by g_f are asymptotically efficient was proved in [10]. The generalization in Lemma 6 and Theorem 7 to describe the efficiency of partitions associated with an arbitrary positive density was suggested by the results of Sacks and Ylvisaker [24].

(VI) A result analogous to Theorem 7 holds for L_r segmented approximation for any r, $1 \leq r \leq \infty$. When $1 \leq r < \infty$, Lemma 6 applies. If f is in $C^n[0, 1]$ and $\{T_{g,p}\}_{p=1}^{\infty}$ is a sequence of p-partitions associated with the positive, bounded, piecewise-continuous density g and ϕ_p is the best L_r approximation of f in $\Phi(p, n)$ with knots at the points of $T_{g,p}$, then

$$\lim_{p\to\infty} p^n \, ||f - \phi_p|| = K_{n,r} \, ||g^{-n}Lf||, \tag{7.15}$$

where $||\cdot||$ denotes the L_r norm on $[0, 1]$. When $r = \infty$, the same limit (7.15) follows from a simple argument based on (5.6). When $|Lf|$ is strictly positive then it follows from (7.15) that taking $g = g_f$ of (7.10) yields an asymptotically efficient sequence of approximations ϕ_p; that is, the expression on the right of (7.15) reduces to the smallest possible limit (6.13).

(VII) Burchard [3] uses knot sets associated with densities g_f in his derivation of error bounds for polynomial spline approximations.

(VIII) The restrictions on the continuity of $D^n f$ and the positivity of g in Theorem 7 can be relaxed. This may require a slight modification in the construction of $T_{g,p}$ to include a finite number of points in the partitions that are sufficiently close to zeros of g and at the points of discontinuity of Lf.

An alternative to the density construction of partitions is suggested by the balanced-error characterization of optimal segmented approximations. We observed that in the limit the L_r deviations between a function and its best approximation over distinct segments are equal. Conversely, we will establish that balanced-error approximations are asymptotically efficient.

Again, the argument is most easily phrased in terms of the functional $E(f, T)$ of Sec. 6. Let f be in $\mathcal{G}$ and suppose assumptions A1 through A4 are satisfied by $e(f; a, b)$ that describes $E(f, T)$. Since $e(f; a, b)$ depends continuously on (a, b) and monotonically on $(b - a)$, for each p we can find a partition T_p satisfying

$$T_p = \{t_\nu \in [0, 1] : t_0 = 0, t_{p+1} = 1 \text{ and } e(f; t_{\nu-1}, t_\nu) = e(f; t_\nu, t_{\nu+1}) \text{ for } \nu = 1, \cdots, p\}. \tag{7.16}$$

T_p balances the contributions of each term $e(f; t_\nu, t_{\nu+1})$ to $E(f, T_p)$ so that $E(f, T_p) = (p + 1)e(f; t_\nu, t_{\nu+1})$. Such balanced partitions are asymptotically efficient.

LEMMA 7. Let f be in $\mathcal{G}$ and suppose assumptions A1, A2, A3, and A4 hold for $e(f; a, b)$. If partitions T_p are defined by (7.16) then

$$\lim_{p\to\infty} p_{m-1} \, E(f, T_p) = \left(\int_0^1 (Jf(s))^{1/m}\, ds\right)^m. \tag{7.17}$$

Proof. Recall $E(f, T_p) = \sum_{\nu=0}^{p} e(f; t_\nu, t_{\nu+1}) = (p + 1)e(f; t_\nu, t_{\nu+1})$. First separate the sum to isolate the terms $e(f; t_\nu, t_{\nu+1})$ for which assumption A2 applies. Fix an arbitrary

$\epsilon > 0$. For each T_p consider the intervals $(t_\nu , t_{\nu+1})$ that either include a discontinuity point of Jf or whose length $h_\nu = t_{\nu+1} - t_\nu$ exceeds ϵ; there is at most a finite number $k_{\epsilon,p}$ of such intervals. Denote their union by $N_{\epsilon,p}$. Of the remaining $(p - k_{\epsilon,p})$ intervals in T_p at least one has length $h_\nu \leq (p - k_{\epsilon,p})^{-1}$. If p is sufficiently large, assumption A2 applies on this interval to yield $e(f; t_\nu , t_{\nu+1}) \leq (Jf(t_\nu) + o(1))(p - k_{\epsilon,p})^{-m}$ and $\lim\sup_{p\to\infty} p^m e(f; t_\nu , t_{\nu+1}) \leq \max_{0\leq t\leq 1} Jf(t)$. Since the terms $e(f; t_\nu , t_{\nu+1})$ are balanced, each is $o(p^{-m})$.

Now write $E(f, T_p) = \sum_{N_{\epsilon,p}^c} e(f; t_\nu , t_{\nu+1}) + \sum N_{\epsilon,p^c} e(f; t_\nu , t_{\nu+1})$. The first sum is negligible since it is a sum of a finite number of terms of order p^{-m}. This implies in addition that the measure of $N_{\epsilon,p} \cap \{t: Jf(t) > 0\}$ goes to zero as $p \to \infty$. Otherwise each set $N_{\epsilon,p}$ properly contains an interval of positive length where Jf is continuous and positive; the subadditivity assumption A1 would imply $\sum_{N_{\epsilon,p}} e(f; t_\nu , t_{\nu+1})$ was bounded away from zero.

Assumption A2 applies to each term in the second sum when ϵ is sufficiently small. For such terms we can write $E(f, T_p) = (p + 1)(Jf(t_\nu) + o(1))h_\nu^m$, where the terms $0(1)$ are uniformly small over $N_{\epsilon,p}^c$ for ϵ small. Taking mth roots and summing over $N_{\epsilon,p}^c$ yields $(p - k_{\epsilon,p})(p + 1)^{-1}\epsilon^m E(f, T_p)^{1/m} = \sum_{N_{\epsilon,p}^c} (Jf(t_\nu) + o(1))^{1/m} h_\nu$. Since Jf is continuous on $N_{\epsilon,p}^c$ and the terms $o(1)$ are uniformly small, we obtain

$$(p - k_{\epsilon,p})(p + 1)^{-1/m} E(f, T_p)^{1/m} = \int_{N_{\epsilon,p}^c} (Jf(s))^{1/m}\, ds + o(1).$$

The integral converges as $p \to \infty$ since the measure of $N_{\epsilon,p} \cap \{t: Jf(t) > 0\}$ goes to zero. This implies $\lim\sup_{p\to\infty} p^{1-1/m} E(f, T_p)^{1/m} \leq \int_0^1 (Jf(s))^{1/m}\, ds + o(1)$. Since ϵ is arbitrary,

$$\limsup_{p\to\infty} p^{m-1} E(f, T_p) \leq \left(\int_0^1 (Jf(s))^{1/m}\, ds\right)^m.$$

The same lower bound on $\lim\inf_{p\to\infty} p^{m-1} E(f, T_p)$, implied by Lemma 4, establishes (7.17) and the lemma is proved.

THEOREM 8. Let f be in $C^n[0, 1]$. Let ϕ_p be a best L_2 approximation of f in $\Phi(p, n)$ that satisfies

$$\int_{t_{\nu-1}}^{t_\nu} (f(s) - \phi(s))^2\, ds = \int_{t_\nu}^{t_{\nu+1}} (f(s) - \phi(s))^2\, ds$$

for $\nu = 1, \cdots, p$. Then

$$\lim_{p\to\infty} p^n \, ||f - \phi_p|| = K_n^{1/2}\left(\int_0^1 |Lf(s)|^{2/(2n+1)}\, ds\right)^{(2n+1)/2}.$$

Proof. Lemma 7.

Remark. The analogous result on efficiency of balanced-error L_r segmented approximations also follows from Lemma 7. As noted previously, when $r = \infty$ best segmented approximations of continuous functions f are balanced-error approximations.

Both Theorem 7 and Theorem 8 give methods of constructing *asymptotically-best* segmented approximations. If the number of knots p of an approximation is small, there is no assurance that an approximation constructed by either of the asymptotically efficient methods will compare favorably to a best segmented approximation with the same number of knots. Knot adjustment by iterative methods, such as described by de Boor and Rice [2] or Esch and Eastman [5], is still necessary to obtain locally optimal knot positions. However, both results on asymptotic efficiency suggest methods for

initializing such iterative procedures to help assure that the locally optimal solutions obtained through iteration do compare favorably with global optima. If knot sets are initialized by the density method or the balanced-error method and then adjusted by iteration, we can guarantee at least that the approximations attained are asymptotically efficient.

The density approach to selecting knots has the advantage that it is explicit and relatively easy to implement if the function Lf is known or can be approximated. It has the disadvantage, observed through experiment, that for very small values of p it may not yield really good initial approximations. These factors are reversed for the balanced-error approach. The balanced-error approach is more difficult to implement, requiring itself an iterative step for knot adjustment; but, as the experience of Pavlidis indicates, the approximations it yields are generally quite good ones.

Precise results of the type developed in this section are not known for approximation by the classes $\Phi(p, q)$ when $q < n$. There derivation will rely on obtaining more precise convergence results than Corollary 5.1. This problem is important in contexts other than the present one. In particular, it bears on the regression design problem of Sacks and Ylvisaker. However, the results for $\Phi(p, n)$, pure segmented approximations, can guide computational procedures when continuity constraints are imposed in the classes $\Phi(p, q)$.

8. Simulation experiments. The method of L_2 segmented approximation for line pattern representation has been tested in an environment where ideal line patterns exhibit the distinct characteristic of piecewise regularity which motivated consideration of segmented approximations and where we can compare the nonlinear method with an optimal method of linear representation. Experiments in two such environments are described in [10]; one of these experiments is related here.

Ideal line patterns comprising the class $\mathfrak{F}$ are generated by a stochastic equation of the form (2.5). We have experimented with a second-order model

$$f''(t) + a_1 f'(t) + a_2 f(t) = \epsilon(t), \tag{8.1}$$

where the stochastic process $\epsilon(t)$ represents a shot-noise input to the linear system described by the operator $L = D^2 + a_1 D + a_2$.

Modelling purely random impulses imparted to the system, we define

$$\epsilon(t) = \sum_{\nu=-\infty}^{\infty} A_\nu g(t - t_\nu). \tag{8.2}$$

The points t_ν, which describe times when an impulse is imparted to the system, are generated by a Poisson renewal process. The interarrival times $t_{\nu+1} - t_\nu$ are mutually independent and have a common exponential distribution with density

$$f_{t_{\nu+1}-t_\nu}(s) = \mu \exp(-\mu s), \qquad s \geq 0. \tag{8.3}$$

The function g, which describes the form of the impulse imparted at each point t_ν, is chosen to approximate a delta function; we set

$$\begin{aligned} g(t) &= 1/\delta, \qquad |t| \leq \delta/2 \\ &= 0, \qquad |t| > \delta/2. \end{aligned} \tag{8.4}$$

The random variable A_ν describes the magnitude of the impulse imparted at time t_ν.

We assume the A_ν's are mutually independent and independent of the Poisson point process and that each A_ν is normally distributed with mean value zero and variance σ^2.

The differential operator L of (8.1) is chosen to model a harmonic oscillator. Thus we prescribe complex conjugate characteristic roots α, $\bar{\alpha}$ and fix

$$L = (D - \alpha)(D - \bar{\alpha}). \tag{8.5}$$

The real part of α is negative so that the system is stable and (8.1) admits a strictly stationary solution f.

With such a specification of the model generating ideal patterns, we can appeal to the theory underlying orthogonal expansions of weakly stationary processes to derive and implement best linear representations of the patterns, according to the procedure outlined in Sec. 2. This linear procedure depends only on the second-order properties of the f-process. Alternatively, we can exploit the deeper structure of individual sample functions, which derives from (8.1) and the special form of the ϵ-process, to seek approximate pattern representations by segmented solutions of $L\phi = 0$. These options are compared.

The f-process depends on the specification of four parameters of the ϵ-process and the operator L. We fix $\mu = 5$ in the density (8.3); this means the mean interarrival time between impulses is $1/5$ and the expected number of impulses imparted in a unit interval is 5. The duration of each impulse is fixed at $\delta = 2^{-7}$, small enough so that g is a reasonable approximation of the delta function but not so small that sample functions of the f-process are described exactly by segmented solutions of $L\phi = 0$. The standard deviation of the magnitudes A_ν is taken to be $\sigma = 2^7$. The relative results in the end do not depend on this value; this choice is a convenient one for computational considerations. Finally, the root α of L was fixed arbitrarily at $\alpha = -.5 + 2\pi i$.

To evaluate the performance of the best linear representation method, the Karhunen-Loève expansion of the f-process, the eigenvalue sums (2.4) are computed. The details of these computations are described in [10] and are omitted here because they are lengthy and reasonably straightforward. To describe the computations briefiy, we start by deriving the covariance function and spectral density of the ϵ-process. Use of the spectral representations of the f- and ϵ-processes and the known spectral density of the ϵ-process yields the spectral density of the f-process. Then Fourier inversion through contour integration and summing residues gives a closed-form expression for the covariance function $R(s, t)$ of the f-process. The eiganvalues of this kernel in (2.3) are the values of interest.

Two paths were followed for solution of the eigenproblem (2.3). First we discretize the problem by Simpson's rule and solve the symmetric matrix eigenproblem with a Jacobi algorithm. Values computed by this method for a 101-point discretization are reported in [10]. Unfortunately, the discretization error bound is intolerably high because derivatives of the kernel R are large. To overcome this difficulty, the kernel R is approximated by one of a simpler form, $R^*(s, t) = \exp(-.5\,|t - s|)(c_1 \cos(2\pi\,|t - s|) + c_2 \sin(2\pi\,|t - s|))$, where c_1 and c_2 are computed from the known expression for R. The L_2 norm of $R - R^*$ is 1.2×10^{-5} and the eigenproblem for R^* lends itself to analytical treatment. In particular, eigenfunctions of R^* satisfy a simpler fourth-order linear differential equation that depends on the parameter λ. The condition that R^*, through (2.3), will map a nontrivial solution of this equation into a constant multiple of itself gives a tractable determinant equation for the eigenvalues of R^*. The roots of this

equation are found by interval bisection. Error analysis problems like those encountered with discretization methods are avoided. Comparison of the eigenvalues found by the alternative methods reaffirms one's faith in Simpson's rule. The eigenvalues computed by the discrete method are orders of magnitude better than the large error bound might suggest.

Results of the eigenvalue calculations are given in Table 1. The first column gives the index N of the respective eigenvalues of the kernel R, starting with the largest. The second column contains the corresponding eigenvalues λ_N, in order of decreasing magnitude. The third column gives the expected squared error $\sum_{\nu=N+1}^{\infty} \lambda_\nu$ of the compressed representation (2.2) by N eigenfunctions. The maximum possible numerical error in column 2 is 1×10^{-5} and the maximum possible error in column 3 is 2.5×10^{-5}. For reference with Table 1, the sum of all the eigenvalues of R is $\sum_{\nu=1}^{\infty} \lambda_\nu = 1030.79205$.

Monte-Carlo simulations of the model (8.1) over the interval [0, 1] provide data for assessing performance of segmented pattern representations. The operator L of (8.5) generates the approximating classes of functions $\Phi(p, 2)$. Each simulated sample function is approximated by a member of the class $\Phi(p, 2)$ that assigns one knot for the approximation at the midpoint of each impulse imparted to the system. We make no subsequent

TABLE 1

Eigenvalues of R and expected squared-errors of compressed linear representations.

N	λ_N	$\sum_{\nu=N+1}^{\infty}\lambda_\nu$
1	468.07585	562.71620
2	437.60507	125.11112
3	72.50417	52.60696
4	38.56538	14.04158
5	8.73215	5.30943
6	2.76057	2.54886
7	1.13015	1.41871
8	0.54769	0.87103
9	0.29779	0.57323
10	0.17595	0.39728
11	0.11069	0.28659
12	0.07315	0.21344
13	0.05028	0.16316
14	0.03570	0.12746
15	0.02605	0.10140
16	0.01945	0.08195
17	0.01482	0.06713
18	0.01148	0.05565
19	0.00903	0.04662
20	0.00720	0.03941
21	0.00581	0.03360
22	0.00474	0.02886
23	0.00391	0.02495
24	0.00324	0.02171
25	0.00272	0.01899
26	0.00229	0.01670
27	0.00195	0.01475
28	0.00166	0.01309
29	0.00143	0.01166
30	0.00123	0.01043

attempt to optimize knot locations; the approximations computed are simply best L_2 approximations in $\Phi(p, 2)$ with the prescribed knots. Since for each simulation the equation (8.1) is easy to solve analytically, as is the equation $L\phi = 0$, we compute approximation errors from simple analytical expressions for the error. There is no discretization involved in the computational procedures.

Table 2 summarizes the results of seventeen independent simulations of this representation procedure. Each row expresses results for a single experiment. The simulations have been reordered according to the associated values of the parameter p, the common value of the number of impulses imparted to the system in the unit interval and of the number of knots of its computed representation. The first column of Table 2 gives the value of this parameter p for each simulation. The second column gives the value of the squared error $||f - \phi_p||^2$ of the approximation of f by the function ϕ_p in $\Phi(p, 2)$ described above. The third column contains the number N of parameters that completely describe ϕ_p; $N = 3p + 2$ since L is second-order. For comparison with the values in column 2, the last column gives the expected squared error $\sum_{\nu=N+1}^{\infty} \lambda_\nu$ of the best linear representation of f by the N principal eigen-functions of (2.3). These last values come from Table 1.

The relative performance of the segmented approximations compared to the best-linear approximations is outstanding. Table 1 shows that linear representations are good, as measured by expected squared error. With as few as eight approximating eigenfunctions the expected error is already less than one while the variance of the f-process exceeds 1000. But Table 2 indicates that segmented representations are significantly better, where the criterion for comparison is the approximation norm. In addition, since the segmented representations are based on the regularity properties of individual sample functions, they directly reflect such local structure of the patterns as locations of derivative discontinuities. The linear representations cannot reflect such structure directly since the eigenfunctions of (2.3) are smooth.

TABLE 2

Simulation experiments.

p	$\|\|f - \phi_p\|\|^2$	N	$\sum_{\nu=N+1}^{\infty} \lambda_\nu$
2	1.0690E-04	8	0.87103
2	1.1608E-05	8	0.87103
3	7.2649E-05	11	0.28659
4	1.3411E-04	14	0.12746
4	1.1552E-04	14	0.12746
4	1.9827E-05	14	0.12746
5	7.6063E-05	17	0.06713
5	6.0369E-05	17	0.06713
5	4.4758E-05	17	0.06713
5	7.8686E-05	17	0.06713
5	1.6974E-05	17	0.06713
6	3.2263E-04	20	0.03941
7	1.1365E-04	23	0.02495
7	1.3593E-04	23	0.02495
9	1.0167E-04	29	0.01166
9	2.8234E-04	29	0.01166
9	1.1616E-04	29	0.01166

In addition to the experiments related in Table 2, tests of the density approach of Theorem 7 for constructing approximations were carried out on the simulations of (8.1). The performance of this spproach for the model (8.1) is predictable. The knot-density function g_f concentrates its mass around the impulses imparted to the system. So eventually the partitions constructed from this density have to detect all of the impulses. But because the absolute mass of g_f near a particular impulse depends on the magnitude of that impulse, it is likely that the density construction will miss one impulse and place multiple knots near another when the total number of impulses is large. Such behavior is observed in the experiments. We would also note that the number of impulses is easily inferred in these experiments through varying the number of knots of successive approximations and observing the behavior of the approximation error for the different numbers of knots. When all impulses are detected the approximation error decreases drastically to the level indicated in Table 2.

Acknowledgment. The author thanks Professor Ulf Grenander for suggesting feature selection problems in pattern analysis, paving the way to early results and providing helpful advice in the course of this research.

Grateful acknowledgment is made to the National Science Foundation for support through the Graduate Fellowship Program and through Grants GP-7347 and GJ-31007X to Brown University.

References

[1] C. de Boor and J. R. Rice, *Least squares cubic spline approximation. I. Fixed knots*, Computer Science Dept. Tech. Rept. 20, Purdue University, 1968

[2] C. de Boor and J. R. Rice, *Least squares cubic spline approximation. II. Variable knots*, Computer Science Dept. Tech. Rept. 21, Purdue University, 1968

[3] H. G. Burchard, *Splines with optimal knots are better*, Applicable Anal. **3**, 309–319(1974).

[4] P. J. Davis, *Interpolation and approximation*, Blaisdell, Waltham, Mass., 1963

[5] R. E. Esch and W. L. Eastman, *Computational methods for best spline function approximation*, J. Approximation Theory 2, 85–96 (1969)

[6] U. Grenander, *Foundations of pattern analysis*, Quart. Appl. Math. 27, 1–55 (1969)

[7] U. Grenander, *A unified approach to pattern analysis*, in *Advances in computers* 10, Academic Press, New York, 1970

[8] C. R. Hobby and J. R. Rice, *Approximation from a curve of functions*, Arch. Rat. Mech. Anal. 24, 91–106 (1967)

[9] S. Karlin and Z. Ziegler, *Chebyshevian spline functions*, SIAM J. Numer. Anal. 3, 514–543 (1966)

[10] D. E. McClure, *Feature selection for the analysis of line patterns*, Center for Computer and Information Sciences and Division of Applied Mathematics Report, Brown University, 1970

[11] D. E. McClure, *Optimal and asymptotically efficient quadrature of stochastic integrals*, to appear

[12] G. Meinardus, *Zur Segmentapproximation mit Polynomen*, Z. Angew. Math. Mech. 46, 239–246 (1966)

[13] G. Meinardus, *Approximation of functions: theory and numerical methods*, Springer-Verlag, New York, 1967

[14] T. Pavlidis, *Linguistic analysis of waveforms*, in *Software engineering* 2 (ed. J. Tou), Academic Press, New York, 1971

[15] T. Pavlidis, *Waveform segmentation through functional approximation*, Department of Electrical Engineering Tech. Rept. 114, Princeton University, 1972

[16] G. M. Phillips, *Error estimates for best polynomial approximations*, in *Approximation theory* (ed. A. Talbot), Academic Press, New York, 1970

[17] G. Polya, *On the mean value theorem corresponding to a given linear homogeneous differential equation*, Trans. Amer. Math. Soc. 24, 312–324 (1922)

[18] M. J. D. Powell, *On best L_2 spline approximations*, Numerische Mathematik, Differentialgleichungen, Approximationstheorie (Tagungen, Oberwolfach, 1966), Birkhäuser, Basel, 1968, 317–338
[19] N. Ream, *Approximation errors in diode-function generators*, J. Electron. Control 7, 83–96 (1959)
[20] N. Ream, *Note on "Approximation of curves by line segments"*, Math. Comp. 15, 418–419 (1961)
[21] J. R. Rice, *The approximation of functions*, vol. II, Addison-Wesley, Reading, Mass., 1969
[22] J. Sacks and D. Ylvisaker, *Designs for regression problems with correlated errors*, Ann. Math. Statist. 37, 66–89 (1966)
[23] J. Sacks and D. Ylvisaker, *Designs for regression problems with correlated errors; many parameters*, Ann. Math. Statist. 39, 49–69 (1968)
[24] J. Sacks and D. Ylvisaker, *Designs for regression problems with correlated errors III*, Ann. Math. Statist. 41, 2057–2074 (1970)
[25] L. L. Schumaker, *Uniform approximation by Tchebycheffian spline functions I: Fixed knots*, MRC Tech. Rept. 768, Madison, Wis., 1967
[26] L. L. Schumaker, *Uniform approximation by Tchebycheffian spline functions II: Free knots*, MRC Tech. Rept. 810, Madison, Wis., 1967
[27] L. L. Schumaker, *On the smoothness of best spline approximations*, J. Approximation Theory 2, 410–418 (1969)
[28] H. Stone, *Approximation of curves by line segments*, Math. Comp. 15, 40–47 (1961)
[29] A. F. Timan, *Theory of approximation of functions of a real variable*, Pergamon Press, New York, 1963
[30] G. Wahba, *On the regression design problem of Sacks and Ylvisaker*, Ann. Math. Statist. 42, 1035–1053 (1971)