JOURNAL OF APPROXIMATION THEORY 14, 204–217 (1975)

The "Lethargy" Theorem—A Property of Approximation by γ-Polynomials

DAVID L. B. JUPP*

Macquarie University, North Ryde 2113, N.S.W. Australia

Communicated by John R. Rice

1. INTRODUCTION

Approximations to functions, and data, often have the form of the general Γ-approximation problem.

In the following, we adopt the abstract formulation due to de Boor [1], which generalized, to abstract spaces, the work of Hobby and Rice [4] and Rice [13], on γ-polynomials in L_p spaces. However, even with the more general setting, the common examples of γ-polynomials are exponential functions, spline functions, and rational functions, and the concrete examples of the smooth Banach spaces we introduce are the L_p spaces, or any Hilbert space.

Let E be a uniformly convex Banach space with twice continuously Frechet (or F-) differentiable norm, and $\gamma(x)$ be a twice (strongly) differentiable E-valued function on $[a, b] \subseteq R$.

A γ-polynomial, of order N, has the form

$$\sigma(\boldsymbol{\alpha}, \mathbf{x}) = \sum_{i=1}^{N} \alpha_i \gamma(x_i) \tag{1.1}$$

where $\boldsymbol{\alpha} \in R^N$ and $a < x_1 < x_2 \cdots < x_N < b$ is a subdivision of (a, b) by N distinct points.

The parameter space for such subdivisions of (a, b) is an N-simplex

$$s_N[a, b] = \{\mathbf{x} \in R^N;\ a < x_1 < x_2 \cdots < x_N < b\}$$

and its closure $\overline{s_N[a, b]}$ is illustrated in Fig. (1) for $N = 2$.

Let

$$\Sigma = \{\sigma(\boldsymbol{\alpha}, \mathbf{x});\quad \boldsymbol{\alpha} \in R^N;\quad \mathbf{x} \in s_N[a, b]\}.$$

* At present supported by the Australian Research Grants Committee.

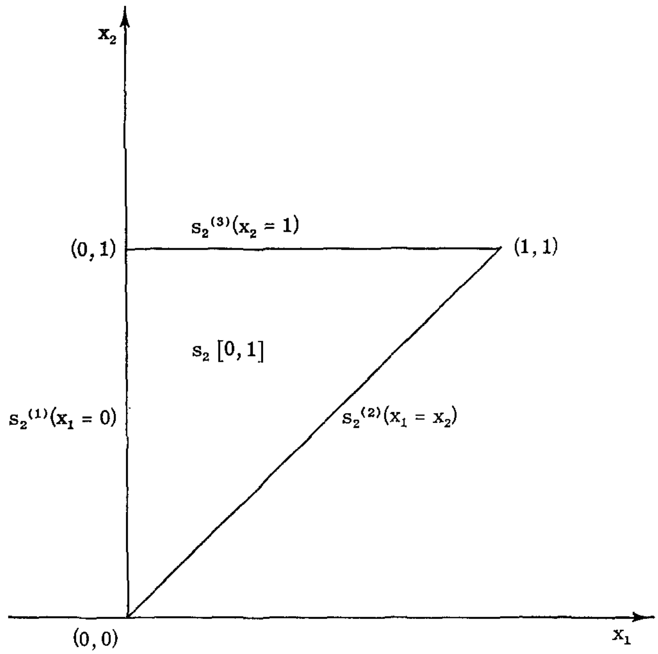

FIG. 1. The simplex $s_N[a, b]$, for $N = 2$, $a = 0$, and $b = 1$.

Then, the *Γ-approximation problem*, to $f \in E$ from Σ, is to

$$\underset{\sigma \in \Sigma}{\text{minimize}}\ \| f - \sigma \|_E \tag{1.2}$$

The problem becomes a Mathematical Programming Problem, in its parameters, when we assume that, for any choice of $\mathbf{x} \in s_N$, the N coordinate elements $(\gamma(x_1), \gamma(x_2),..., \gamma(x_N))$ are linearly independent. Then we may write (the "square" is explained in Section (3))

$$\underset{(\boldsymbol{\alpha}, \mathbf{x}) \in R^N x s_N}{\text{minimize}}\ F(\boldsymbol{\alpha}, \mathbf{x}) = \| f - \sigma \|_E^2 . \tag{1.3}$$

This problem is called the "full functional" problem as opposed to the following reduced functional problem. For each $\mathbf{x} \in s_N[a, b]$, the class

$$\Gamma(\mathbf{x}) = \left\{ \sigma(\boldsymbol{\alpha}, \mathbf{x}) = \sum_{i=1}^{N} \alpha_i \gamma(x_i);\ \boldsymbol{\alpha} \in R^N \right\}$$

is a linear subspace of E, with dimension N. The Linear approximation problem

$$\underset{\boldsymbol{\alpha} \in R^N}{\text{minimize}}\ F(\boldsymbol{\alpha}) = \left\| f - \sum_{i=1}^{N} \alpha_i \gamma(x_i) \right\|_E^2 ,$$

has a unique solution $\boldsymbol{\alpha}^* = \boldsymbol{\alpha}^*(\mathbf{x})$. The corresponding reduced functional

$$\tilde{F}(\mathbf{x}) = F(\boldsymbol{\alpha}^*, \mathbf{x}), \tag{1.4}$$

may be minimized on $s_N[a, b]$, to find a solution of the original Γ-approximation problem (1.2).

The process of reducing the problem by implicitly solving the linear problem for each set of non-linear parameters (the $\mathbf{x}$ parameters in this case) has had recent application (Osborne [10], Golub and Pereyra [3]). We show in §(5) that with our assumptions about E and $\gamma(x)$, on $s_N[a, b]$

(i) $\boldsymbol{\alpha}^*(\mathbf{x})$ is unique and differentiable, as a function of $\mathbf{x}$,
(ii) $\tilde{F}(\mathbf{x})$ is differentiable,
(iii) The $\mathbf{x}$ values of the critical points of $F(\boldsymbol{\alpha}, \mathbf{x})$ and $\tilde{F}(\mathbf{x})$ are identical.

Further, we show that by extending the γ-polynomials suitably, [1], to include multiple point "subdivisions," the statements (i), (ii) and (iii) may be continued across the bounding hyperplanes of $\overline{s_N[a, b]}$. In particular, the symmetry properties of the extended $\tilde{F}(\mathbf{x})$ imply the "natural boundary condition"

$$\mathbf{n}_p{}^T \nabla \tilde{F}(\mathbf{x}) = 0 \tag{1.5}$$

(where $\mathbf{n}_p$ is the unit normal to the respective hyperplane).

In order to solve the Γ-approximation problem numerically, it is customary to minimize $\tilde{F}(\mathbf{x})$ on $s_N[a, b]$, as a constrained Mathematical Programming Problem. However, we observed with spline functions, that numerical algorithms had very poor convergence properties near ∂s_N, and tended to "hang up" well away from meaningful solutions. This property can be directly attributed (cf. Section (6)) to the condition (1.5) which was called the "Lethargy" property [6] from this observation.

2. γ-Polynomials with Confluent Points

The boundary of $\overline{s_N[a, b]}$ consists of segments of the $(N + 1)$ hyperplanes

$$\begin{aligned} g_1(\mathbf{x}) &= x_1 - a = 0 \\ g_p(\mathbf{x}) &= x_p - x_{p-1} = 0 \qquad (p = \overline{2, N}) \\ g_{N+1}(\mathbf{x}) &= b - x_N = 0 \end{aligned} \tag{2.1}$$

That is, $\overline{s_N[a, b]}$ can be *defined* as the convex region satisfying the linear inequality constraints

$$\mathbf{g}(\mathbf{x}) \geqslant \mathbf{0}. \tag{2.2}$$

The bounding hyperplanes correspond to subdivisions of $[a, b]$ with multiple, or confluent points, and effective methods for defining the γ-polynomials corresponding to these points are given in Rice [13] or de Boor [1]. These will be called *extended* γ-polynomials on $s_N[a, b]$. A point on ∂s_N with the single active constraint $g_p(\mathbf{x}) = 0$ will be said to lie on the p'th (open) *main face* of $\overline{s_N[a, b]}$, denoted $s_N^{(p)}[a, b]$. γ-polynomials are easily extended to most of the cube $[a, b]^N$ by identifying points which are permutations. Clearly, the order of the points does not affect the γ-polynomial $\sigma(\boldsymbol{\alpha}, \mathbf{x})$ defined by (1.1).

The subset of $(a, b)^N$ generated by permutations of s_N and the main faces $s_N^{(p)}(p = \overline{2, N})$, will be denoted

$$k_N^{(1)} = \{x \in (a, b)^N\text{: for all } i \text{ there exists at most one } j \neq i \text{ such that } x_i = x_j\}$$

On $k_N^{(1)}$, a straight forward application of the results of de Boor [1] provides us with the following Lemma.

LEMMA (2.1). *If, for any two points $p \neq q$ in (a, b); $\gamma(p)$, $\gamma'(p)$, and $\gamma(q)$ are linearly independent, then the N coordinate functions*

$$(\gamma(x_1), \gamma(x_1, x_2), \ldots, \gamma(x_{N-1}, x_N))$$

form a basis for $\Gamma(\mathbf{x})$, for each $\mathbf{x}$ in $k_N^{(1)}$, where $\gamma(x_i, x_{i+1})$ is the first divided difference,

$$\gamma(x_i, x_{i+1}) = \begin{cases} \dfrac{\gamma(x_{i+1}) - \gamma(x_i)}{x_{i+1} - x_i} & x_i \neq x_{i+1} \\ \gamma'(x_i) & x_i = x_{i+1} \end{cases}$$

3. The Semi-Inner Product on E

Assuming for our purposes E to be a *real* Banach space, we can use an important tool due to Lumer [8] and Giles [2]—the *Semi-inner product.* If f is an element of E, there exists by the Hahn–Banach Theorem, at least one (in our case exactly one) linear functional μ_f in E' (the continuous dual of E), such that

(i) $\|\mu_f\|_{E'} = \|f\|_E$

and

(ii) $\langle f, \mu_f\rangle = \|f\|_E^2$.

The function on $ExE \to R$

$$[f, g] = \mu_g(f)$$

is called the (natural) Homogeneous, Semi-inner product (SIP) for E, and has the usual properties of an inner product, except for additivity in g.

Lumer and Phillips [9], and Giles [2], have used the SIP to extend Hilbert space methods to (complex) Banach spaces, and we will use it to develop expressions for the derivative of $\tilde{F}(\mathbf{x})$.

Referring to Schwartz [17] for definition of the F-derivative we conclude from Theorem (3) of Giles [2] and its proof,

LEMMA (3.1). *The F-derivative of* $\| f \|_E^2$ *in the direction g is given by*

$$d \| \cdot \|_E^2 (f, g) = 2[g, f].$$

In addition, following Schwartz [17, (1.14)] we will use the chain rule for F-derivatives.

LEMMA (3.2). *Let* $F\colon U \to V$ *be F-differentiable at* X_0, *and* $G\colon V \to W$ *be F-differentiable at* $F(X_0)$, *then* $G(F(X))$ *is F-differentiable at* X_0 *and* $d(GF)(X_0, Y) = dG(F(X_0), dF(X_0, Y))$.

From these Lemmas we get

THEOREM (3.1). *If* $\gamma(\mathbf{x})$ *is a strongly differentiable E-valued function on the open set* $\Omega \subseteq R^N$ *then*

$$\frac{\partial}{\partial x_p} \| \gamma(\mathbf{x}) \|_E^2 = 2 \left[\frac{\partial \gamma}{\partial x_p}, \gamma \right].$$

To apply Theorem (3.1), we let $\phi_p(\mathbf{x})$ denote a coordinate function for $\Gamma(\mathbf{x})$. In our case, $\phi_p(\mathbf{x})$ may be $\gamma(x_p)$ on s_N, or $\gamma(x_{p-1}, x_p)$ on $k_N^{(1)}$, and more generally, $\phi_p(\mathbf{x})$ may be a coordinate element for a general field of subspaces in the sense of Jupp [7].

Theorem (3.1) has the following corollary.

COROLLARY (3.1). *With the notation of Section* (1),

$$\left(\sigma(\boldsymbol{\alpha}, \mathbf{x}) = \sum_{i=1}^{N} \alpha_i \phi_i(\mathbf{x}) \right)$$

(i) $$\frac{\partial}{\partial \alpha_p} \| f - \sigma \|_E^2 = -2[\phi_p, f - \sigma]$$

(ii) $$\frac{\partial}{\partial x_p} \| f - \sigma \|_E^2 = -2 \sum_{i=1}^{N} \alpha_i \left[\frac{\partial \phi_i}{\partial x_p}, f - \sigma \right].$$

These results are easily verified directly when E is a Hilbert space, or an $L_p[a, b]$ space for $2 \leqslant p < \infty$. In the case of $L_p[a, b]$, (Giles [2])

$$[f, g] = \int_a^b f(t) \frac{|g|^{p-1} \operatorname{sgn}(g)\, dt}{\| g \|_p^{p-2}}$$

and Corollary (3.1) is easily verified by direct differentiation.

4. The Implicit Function Theorem

Let $F(\boldsymbol{\alpha}, \mathbf{x})$ be a function on $R^N \times \Omega$, where $\Omega \subseteq R^p$ is an open set. We assume that $F(\boldsymbol{\alpha}, \mathbf{x})$ is strictly convex as a function of $\boldsymbol{\alpha}$, for each $\mathbf{x} \in \Omega$.

Lemma (4.1). *Let $F(\boldsymbol{\alpha}, \mathbf{x})$ be differentiable in* $\mathbf{x}$, *and twice differentiable in* $\boldsymbol{\alpha}$, *on $R^N \times \Omega$. Then, $\boldsymbol{\alpha}^*(\mathbf{x})$ defined by*

$$\{\boldsymbol{\alpha}^*(\mathbf{x}) \textit{ minimizes } F(\boldsymbol{\alpha}, \mathbf{x}), \textit{for each } \mathbf{x}, \textit{ in } \boldsymbol{\alpha}\}$$

is a singly R^N-valued, differentiable function.

Proof. $\boldsymbol{\alpha}^*(\mathbf{x})$ is defined implicitly by the system of p equations

$$\nabla_{\boldsymbol{\alpha}} F(\boldsymbol{\alpha}, \mathbf{x}) = \mathbf{0}.$$

The Hessian of F, as a function of $\boldsymbol{\alpha}$,

$$H(\boldsymbol{\alpha}) = \left[\frac{\partial^2 F}{\partial \alpha_p\, \partial \alpha_q}\right]_{\overline{p,q=1,N}}$$

exists, and is positive *definite* for each $\mathbf{x} \in \Omega$, since F is strictly convex, and $C^2(R^N; R)$ in $\boldsymbol{\alpha}$. By the Basic Implicit Function Theorem of Calculus, [18] $\boldsymbol{\alpha}^*(\mathbf{x})$ will exist, and be differentiable, as required.

Lemma (4.2). *Assuming the conditions of Lemma* (4.1), *if*

$$\tilde{F}(\mathbf{x}) = F(\boldsymbol{\alpha}^*(\mathbf{x}), \mathbf{x}),$$

then $\tilde{F}(\mathbf{x})$ is differentiable, and

$$\nabla_{\mathbf{x}} \tilde{F}(\mathbf{x}) = \nabla_{\mathbf{x}} F(\boldsymbol{\alpha}, \mathbf{x})|_{(\boldsymbol{\alpha}^*, \mathbf{x})}$$

Proof. $\tilde{F}(\mathbf{x})$ is differentiable, by composition, for, since, $F(\boldsymbol{\alpha}, \mathbf{x})$ is differentiable

$$\nabla_{\mathbf{x}} \tilde{F}(\mathbf{x}) = [D\boldsymbol{\alpha}^*]^T \nabla_{\boldsymbol{\alpha}} F(\boldsymbol{\alpha}, \mathbf{x})|_{(\boldsymbol{\alpha}^*, \mathbf{x})} + \nabla_{\mathbf{x}} F(\boldsymbol{\alpha}, \mathbf{x})|_{(\boldsymbol{\alpha}^*, \mathbf{x})}$$

which is continuous since $D\boldsymbol{\alpha}^*$ is continuous by Lemma (4.1). However, $\nabla_{\alpha} F(\boldsymbol{\alpha}, \mathbf{x})|_{(\boldsymbol{\alpha}^*, \mathbf{x})} = \mathbf{0}$ *defines* $\boldsymbol{\alpha}^*(\mathbf{x})$ so that $\nabla_{\mathbf{x}} \tilde{F}(\mathbf{x}) = \nabla_{\mathbf{x}} F(\boldsymbol{\alpha}, \mathbf{x})|_{(\boldsymbol{\alpha}^*, \mathbf{x})}$ as required.

Proofs of the following Theorems appear in Jupp [7]. In a more specific context, they occur in Golub and Pereyra [3], and are included now for completeness.

THEOREM (4.1). *$(\boldsymbol{\alpha}^*, \mathbf{x}^*)$ is a strict local minimum of $F(\boldsymbol{\alpha}, \mathbf{x})$, if, and only if, $\mathbf{x}^*$ is a strict local minimum of $\tilde{F}(\mathbf{x})$, on Ω, and $\boldsymbol{\alpha}^*$ minimizes $F(\boldsymbol{\alpha}, \mathbf{x}^*)$ globally on R^N.*

Theorem (4.1) will hold when $F(\boldsymbol{\alpha}, \mathbf{x})$ is only continuous. When the conditions of Lemma (4.1) hold we have

THEOREM (4.2). *$(\boldsymbol{\alpha}^*, \mathbf{x}^*)$ is a critical point of $F(\boldsymbol{\alpha}, \mathbf{x})$ on $R^N \times \Omega$ if, and only if, $\mathbf{x}^*$ is a critical point of $\tilde{F}(\mathbf{x})$ on Ω.*

5. THE LETHARGY THEOREM

LEMMA (5.1). *$F(\boldsymbol{\alpha}, \mathbf{x}) = \|f - \sigma\|_E^2$ is strictly convex and $C^2(R^N; R)$ in $\boldsymbol{\alpha}$; and $C^1(k_N^{(1)}; R)$ in $\mathbf{x}$.*

Proof. For each $\mathbf{x}$, $\Gamma(\mathbf{x})$ is a linear subspace of E, and in Section 2, we assumed that the coordinate functions

$$(\gamma(x_1), \gamma(x_1, x_2), \ldots, \gamma(x_{N-1}, x_N))$$

where a *basis* for $\Gamma(\mathbf{x})$, for each $\mathbf{x}$ in $k_N^{(1)}$. For each $\mathbf{x}$ therefore,

(i) $\boldsymbol{\alpha}^*(\mathbf{x})$ will be defined by the nondegenerate system of N nonlinear equations

$$\frac{\partial F(\boldsymbol{\alpha}, \mathbf{x})}{\partial \alpha_p} = -2[\phi_p(\mathbf{x}), f - \sigma] = 0 \qquad \text{for} \quad p = \overline{1, N}$$

(ii) The Jacobian of this system, (the symmetric Hessian of F)

$$H(\boldsymbol{\alpha}) = \left[\frac{\partial^2 F}{\partial \alpha_p \, \partial \alpha_q}\right]_{pq=\overline{1,N}} = -2 \left(\frac{\partial}{\partial \alpha_q} [\phi_p, f - \sigma]\right)$$

will exist, and be positive definite on $k_N^{(1)}$. By Corollary (3.1),

$$\frac{\partial F(\boldsymbol{\alpha}, \mathbf{x})}{\partial x_p} = -2 \left[\frac{\partial \sigma}{\partial x_p}, f - \sigma\right] = -2 \sum_{i=1}^{N} \alpha_i \left[\frac{\partial \phi_i}{\partial x_p}, f - \sigma\right]$$

now $\phi_i(\mathbf{x}) = \gamma(x_{i-1}, x_i)$ and

$$\frac{\partial \phi_i(\mathbf{x})}{\partial x_p} = \begin{cases} 0 & \text{if} \quad i \neq p, p+1, \\ \gamma(x_p, x_p, x_{p+1}) & \text{if} \quad i = p+1, \\ \gamma(x_{p-1}, x_p, x_p) & \text{if} \quad i = p. \end{cases}$$

If $\mathbf{x} \in s_N[a, b]$ (or is some permutation of a distinct set of points) then at most $\gamma'(x_p)$ occurs in the derivative. However, if $x_p = x_{p+1}$ then

$$\frac{\partial \phi_p}{\partial x_p} = \frac{\gamma''(x_p)}{2}$$

which is still in order, since we assumed $\gamma(x)$ to be twice strongly differentiable in E.

An immediate application of §(4) to this Lemma yields

THEOREM (5.1). *On* $k_N^{(1)}$

(i) $\boldsymbol{\alpha}^*(\mathbf{x})$ *is a unique differentiable function of* $\mathbf{x}$,

(ii) $\tilde{F}(\mathbf{x})$ *is differentiable,*

(iii) *The* $\mathbf{x}$ *values of the critical points, and local minima, of* $F(\boldsymbol{\alpha}, \mathbf{x})$ *and* $\tilde{F}(\mathbf{x})$ *are the same.*

THEOREM (5.2). (*The Lethargy Theorem*). *Across the main-faces* (*cf* §(2)) $s_N^{(p)}[a, b]$, $(p = \overline{2, N})$ *of* $s_N[a, b]$

$$\mathbf{n}_p{}^T \nabla \tilde{F}(\mathbf{x}) = 0$$

where $\mathbf{n}_p$ *is the outward normal to the main face* $s_N^{(p)}$.

Proof. $\tilde{F}(\mathbf{x})$ is a symmetric, differentiable function across $s_N^{(p)}[a, b]$ (with respect to interchanges of x_{p-1} and x_p), so that its derivative across $s_N^{(p)}[a, b]$ is zero.

6. CRITICAL POINTS OF $\tilde{F}(\mathbf{x})$

The Lethargy Theorem states that the normal component of the gradient field across $s_N^{(p)}[a, b]$ is zero, that is:

COROLLARY (6.1). *The flow of the gradient field,* $\mathbf{x} \to \nabla \tilde{F}(\mathbf{x})$ *on* $k_N^{(1)}$ *is confined to the main face* $s_N^{(p)}[a, b]$, *if its initial point is on* $s_N^{(p)}[a, b]$.

It follows that the Γ-approximation problem will have *at least* one solution on each of the *closed* main faces $\overline{s_N^{(p)}[a, b]}$ for $p = \overline{2, N}$. If this solution is in $k_N^{(1)}$, by Theorem (5.1) it will be a *critical point* of $\tilde{F}(\mathbf{x})$ and $F(\boldsymbol{\alpha}, \mathbf{x})$. However, although the point is a minimum of $\tilde{F}(\mathbf{x})$ when restricted to $s_N^{(p)}$ it may be a *saddle point* of $\tilde{F}(\mathbf{x})$, (or $F(\boldsymbol{\alpha}, \mathbf{x})$) on $k_N^{(1)}$ (or $R^N x k_N^{(1)}$). The occurrence of saddle points on ∂s_N seems to be quite common, and the results on non-convexity of $\tilde{F}(\mathbf{x})$ in Jupp [7, Part V] seem to imply that saddle points of almost any complexity will occur as more points are allowed to coalesce in the search for solutions of the Γ-approximation problem.

7. Effect on Numerical Methods

Following §(1), numerical methods for the problem (1.3) treat the reduced problem (1.4) as a constrained Mathematical Programming Problem. Theorem (5.1) justifies the approach to some extent, and to avoid problems on ∂s_N we can use a Barrier Transformation Function such as one of those developed in Ryan and Osborne [14]. However, some problems occur with this approach due to the Lethargy property.

From Theorem (5.1) it follows that constrained solutions on $\overline{s_N[a, b]}$ (that is, on $s_N^{(p)}$, for some $p = \overline{2, N}$) are critical points of $\tilde{F}(\mathbf{x})$ on $k_N^{(1)}$. The consequence is that the important strict complementarity condition fails to hold for constrained solutions of the Γ-approximation problem as posed in (1.5).

Precisely, [11], strict complementarity holds for a function $G(\mathbf{x})$, if, whenever $\mathbf{x}^*$ is a local minimum of $G(\mathbf{x})$ on $s_N^{(p)}$, then

$$\mathbf{n}_p{}^T \nabla G(\mathbf{x}) < 0 \text{ (i.e., strictly).}$$

Corollary (7.1). *There is no strict complementarity on any of the main faces $s_N^{(p)}[a, b]$, $p = \overline{2, N}$ with any functional defined by the Γ-approximation problem.*

Osborne [11], and Ryan [15], show how lack of this condition will imply poor convergence (or "Lethargy") properties for any algorithm using a Barrier Transformation Function, and finding constrained solutions of (1.5).

Numerical case studies of this property, and the effects of the saddle points, (cf. §(6)) on methods of the descent type, and Gauss–Newton and Marquardt type (as developed in Golub and Pereyra [3] for the least-squares problem), are investigated in Jupp [7].

8. Differential Formulae

Without presenting the full proofs, which depend on the properties of the S.I.P. of Section (3), Corollary (3.1) and Lemma (4.2), we have

Lemma (8.1). *On $s_N[a, b]$, if*

$$\sigma^*(\boldsymbol{\alpha}^*, \mathbf{x}) = \sum_{i=1}^{N} \alpha_i^* \gamma(x_i)$$

Then

$$\frac{\partial \tilde{F}}{\partial x_p} = \frac{\partial}{\partial x_p} \| f - \sigma^* \|_E^2 = -2\alpha_p^*[\gamma'(x_p), f - \sigma^*].$$

Lemma (8.2). *On $k_N^{(1)}$, if*

$$\sigma^*(\boldsymbol{\beta}^*, \mathbf{x}) = \beta_1^* \gamma(x_1) + \sum_{i=2}^{N} \beta_i^* \gamma(x_{i-1}, x_i)$$

then

$$\frac{\partial \tilde{F}}{\partial x_p} = \begin{cases} -2 \left\{ \dfrac{\beta_p^*}{h_p} - \dfrac{\beta_{p+1}^*}{h_{p+1}} \right\} [\gamma'(x_p), f - \sigma^*] & \text{if} \quad x_{p-1} < x_p < x_{p+1} \\ -\beta_p^*[\gamma''(x_p), f - \sigma^*] & \text{if} \quad x_{p-1} = x_p < x_{p+1} \end{cases}$$

(*where $h_p = x_p - x_{p-1}$, and * denotes the optimal Linear Solution as usual*).

Let $\mathbf{x}^*$ in $k_N^{(1)}$ be a local minimum of $\tilde{F}(\mathbf{x})$. Then either $\mathbf{x}^*$ is in s_N, or else, $\mathbf{x}^*$ is in $s_N^{(p)}$ for some $2 \leqslant p \leqslant N$.

Theorem (8.1)

(i) *If $\mathbf{x}^*$ is in $s_N[a, b]$, then not only does $[\gamma(x_j), f - \sigma^*] = 0, j = \overline{1, N}$, but also either $\boldsymbol{\alpha}_j^* = 0$, or $[\gamma'(x_j), f - \sigma^*] = 0$, for each $j = \overline{1, N}$.*

(ii) *If $\mathbf{x}^* \in s_N^{(p)}$, the above holds for $j \neq p$, and at x_p, $[\gamma'(x_p), f - \sigma^*] = 0$, and either $\beta_p^* = 0$, or else $[\gamma''(x_p), f - \sigma^*] = 0$.*

Theorem (8.1) generalizes the formulae of Powell [12], and contains the important multiple interpolation, and extra precision conditions familiar in the Theory of Optimal Quadrature [16].

9. An Example

Let $\gamma(x, t) = e^{-xt}$, and consider the discrete least-squares approximation to data by exponentials. That is, let $\mathbf{f} = (f_1, f_2, ..., f_M)^T$ be M data values sampled at times $t_1 < t_2 \cdots < t_M$. The model for the data is taken to be

$$\sigma(\boldsymbol{\alpha}, \mathbf{x}, t) = \alpha_1 e^{-x_1 t} + \alpha_2 e^{-x_2 t} + \alpha_3 e^{-x_3 t}$$

which is fitted to the data in the sense of least-squares, i.e.,

$$\underset{(\alpha, \mathbf{x})}{\textit{minimize}} \|f - \sigma\|^2 = \sum_{i=1}^{M} |f_i - \sigma(\boldsymbol{\alpha}, \mathbf{x}, t_i)|^2.$$

For physical reasons, we choose to constrain x_3 to be *zero* so that there are two free non-linear parameters x_1, x_2 satisfying

$$0 < x_1 < x_2 < b \ (\mathbf{x} \in s_2[0, b])$$

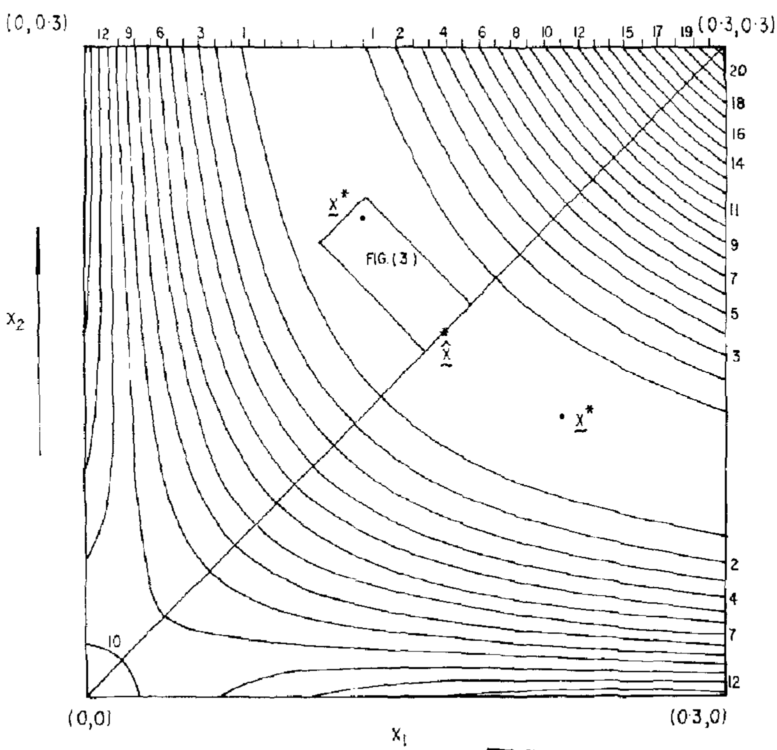

Figure 2

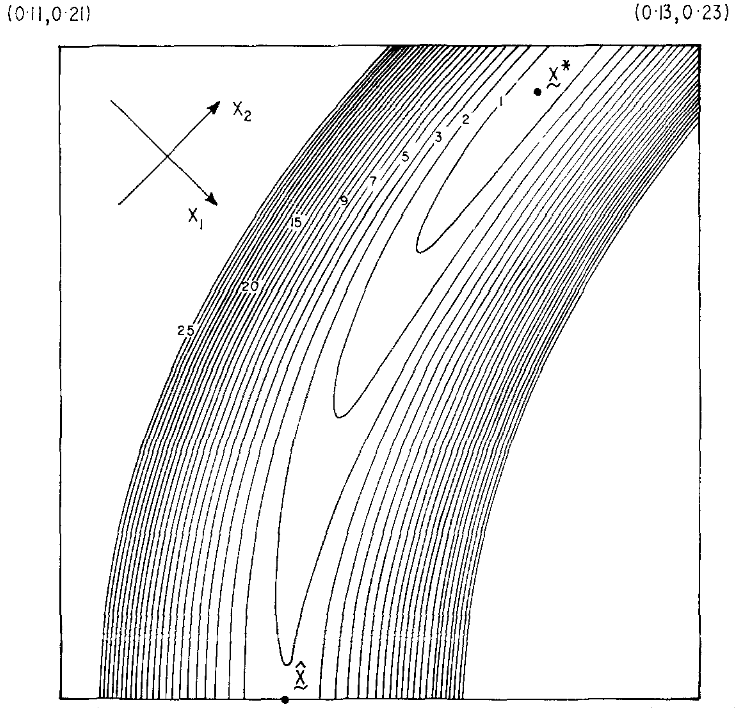

FIGURE 3

(The upper limit b can usually be placed, but is only a bound.) Taking $b = .3$, we have plotted $\tilde{F}(\mathbf{x})$, resulting from some data, in Figs. (2) and (3). The data were originally used in Jennings and Osborne [5], as an example, against which to try the Gauss-Newton-Marquardt method minimizing the full functional $F(\boldsymbol{\alpha}, \mathbf{x})$. They also appear in Golub and Pereyra [3], as a comparative example, where the Marquardt algorithm is applied to $\tilde{F}(\mathbf{x})$. Figures (2) and (3) show the strict local minimum $\mathbf{x}^*$, which is a physically meaningful solution to the problem, and a saddle point on $s_2^{(2)}$, labeled $\hat{\mathbf{x}}$. Table I summarizes the values, and it can be seen that the fit to the data, by the

TABLE 1

Point	Nonlinear parameters	Residual value
$\mathbf{x}^*$	(0.01287, 0.02212)	5.4649×10^{-5}
$\hat{\mathbf{x}}$	(0.01670, 0.01670)	7.9803×10^{-5}

TABLE II

Contours (of the Residual as a Function of $(x_1 x_2)^T$).
FIG. 2. Contour $j = j \times 5 \times 10^{-8}$ for $j = \overline{1, 21}$
FIG. 3. Contour $j = 6 \times 10^{-5} + (j-1) \times 10^{-5}$ for $j = \overline{1, 25}$

constrained solution, is quite good, in the sense of least-squares. Since the contour levels are equally spaced, (see Table II) and the large area contained by level (1) is very flat, its structure has been detailed in Fig. (3). The Lethargy property is well illustrated across the diagonal ($s_2^{(2)}$), where the contours cut at right angles and $\hat{\mathbf{x}}$ is a local minimum of $\tilde{F}(\mathbf{x})$ on $s_2^{(2)}[0, b]$, but a saddle point on $k_2^{(1)}$, which, in this case, is the open interior of the square $[0, b]^2$.

10. Acknowledgments

I wish to acknowledge, and sincerely thank, the referee for his comments which helped considerably to simplify, and generalize, the original form of the paper. I also wish to thank my supervisor Dr. J. Kautsky of Flinders University, South Australia, where most of the work in the paper was done, supported by an Australian Commonwealth Postgraduate Award.

References

1. C. de Boor, On the approximation by Γ-polynomials, *in* "Approximation with Special Emphasis on Spline Functions," (Schoenberg, Ed.), Academic Press, New York, London, 1969.
2. J. R. Giles, Classes of semi-inner product spaces, *Trans. Amer. Math. Soc.* **129** (1967), 436–446.
3. G. H. Golub and V. Pereyra, The differentiation of pseudoinverses and nonlinear least squares problems whose variables separate. *J. Siam. Num. Anal.* **10** (1973), 413–432.
4. C. R. Hobby and J. R. Rice, Approximation from a curve of functions, *Arch. Rat. Mech. Anal.* **27** (1967), 91–106.
5. L. S. Jennings and M. R. Osborne, Applications of orthogonal matrix transformations to the solution of systems of linear and nonlinear equations, Technical Report, No. 37, Computer Centre, Australian National University, 1970.
6. D. L. B. Jupp, Curve fitting by splines as an example of unconstrained optimization, *in* "Optimization" (R. S. Anderssen, L. S. Jennings, D. M. Ryan, Eds.), University of Queensland Press, 1972.
7. D. L. B. Jupp, Non-linear approximation and free knot splines, Ph.D. Thesis, The Flinders University of South Australia, August 1973.
8. G. Lumer, Semi-inner-product Spaces, *Trans. Amer. Math. Soc.* **100** (1961), 29–43.

9. G. LUMER AND R. S. PHILLIPS, Dissipative Operators in a Banach Space, *Pacific J. Math.* **11** (1961), 679–698.
10. M. R. OSBORNE, Some special nonlinear least squares problems, *J. SIAM. Num. Anal.*, (1975) (To appear).
11. M. R. OSBORNE, Topics in optimization, Tech. Rept. STAN-CS-72-279, Stanford University, 1972.
12. M. J. D. POWELL, On best L_2 spline approximations, *in* "Numerische Mathematik, Differentialgleichungen, Approximations-theorie," ISNM Vol. 9, Birkhäuser Verlag, Basel, 1968, 317–339.
13. J. R. Rice, "The Approximation of Functions," Vol. 2, Addison Wesley, Reading, MA, 1969.
14. D. M. RYAN AND M. R. OSBORNE, On Penalty Function methods for Non-Linear Programming Problems, *J. Math. Anal. Appl.* **31** (1970), 559–578.
15. D. M. RYAN, Transformation methods in non-linear programming, Ph.D. Thesis, Australia National Univ., Canberra, 1971.
16. I. J. SCHOENBERG, Monosplines and Quadrature formulae, *in* "Theory and Applications of Spline Functions," (T. N. E. Greville, ed.), Academic Press, New York, London, 1969.
17. J. T. SCHWARTZ, "Non-Linear Functional Analysis," Gordon and Breach, New York, 1969.
18. M. SPIVAK, "Calculus on Manifolds," Benjamin, New York, 1965.