01 Jan 1982

An Approximation Theory for Conjugate Surfaces and Solutions of Elliptic Multiple Integral Problems: Application to Numerical Solutions of Generalized Laplace's Equation

John Gregory

Ralph W. Wilkerson
Missouri University of Science and Technology, ralphw@mst.edu

Recommended Citation

J. Gregory and R. W. Wilkerson, "An Approximation Theory for Conjugate Surfaces and Solutions of Elliptic Multiple Integral Problems: Application to Numerical Solutions of Generalized Laplace's Equation," *Journal of Mathematical Analysis and Applications*, vol. 88, no. 1, pp. 231 - 244, Elsevier, Jan 1982.
The definitive version is available at https://doi.org/10.1016/0022-247X(82)90189-5

JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 88, 231–244 (1982)

An Approximation Theory for Conjugate Surfaces and Solutions of Elliptic Multiple Integral Problems: Application to Numerical Solutions of Generalized Laplace's Equation

JOHN GREGORY

Department of Mathematics, Southern Illinois University, Carbondale, Illinois 62901

AND

RALPH WILKERSON

Department of Mathematics, Winthrop College, Rock Hill, South Carolina 29733

Submitted by George Leitmann

An approximation theory is given for a class of elliptic quadratic forms which include the study of conjugate surfaces for elliptic multiple integral problems. These ideas follow from the quadratic form theory of Hestenes, applied to multiple integral problems by Dennemeyer, and extended with applications for approximation problems by Gregory.

The application of this theory to a variety of approximation problem areas in this setting is given. These include conjugate surfaces and conjugate solutions in the calculus of variations, oscillation problems for elliptic partial differential equations, eigenvalue problems for compact operators, numerical approximation problems, and, finally, the intersection of these problem areas.

In the final part of this paper the ideas are specifically applied to the construction and counting of negative vectors in order to obtain new numerical methods for solving Laplace-type equations and to obtain the "Euler-Lagrange equations" for symmetric-banded tridiagonal matrices. In this new result (which will allow the reexamination of both the theory and applications of symmetric-banded matrices) one can construct, in a meaningful way, negative vectors, oscillation vectors, eigenvectors, and extremal solutions of classical problems as well as efficient algorithms for the numerical solution of partial differential equations. Numerical examples (test runs) are given.

1. INTRODUCTION

The main purpose of this paper is to present an approximation theory of quadratic forms which is applicable to linear-elliptic multiple integral problems; that is, to quadratic forms whose Euler–Lagrange equation is

0022-247X/82/070231-14$02.00/0

$$\frac{\partial}{\partial t_i}(R_{ij}(t)\,\dot{x}_j(t)) - x(t)\left(P(t) - \sum_{i=1}^{m} \frac{\partial Q_i}{\partial t_i}\right) = 0 \quad (i,j = 1,\ldots, m).$$

In the above, $t = (t_1, t_2,\ldots, t_m)$ is in $\mathbb{R}^m$, $x(t)$ is a real-valued function, $\partial x/\partial t_j$ is written as $\dot{x}_j$, and $P(t)$, $Q_i(t)$, and $R_{ij}(t)$ satisfy smoothness and symmetric properties described below. In addition, repeated indices are summed.

Applications of our theory to approximating problems dealing with eigenvalue problems, oscillation problems or focal point problems, and numerical problems will be considered.

The fundamental quadratic form theory was given by Hestenes in 1951 [5] to handle recurring "second variation" problems in the calculus of variations. This theory was generalized by Gregory to an approximation theory of quadratic forms. In one sense this paper is an application of these ideas to a problem in partial differential equations defined by Dennemeyer [1] and Hestenes [6].

To save journal pages and expenses, many technical details will not be given here, but, instead, are carefully referenced for the reader. This is possible because our results are new, but they follow in a similar manner to the problem of ordinary differential equations given by the first author.

The outline of this paper is as follows. In Section 2 we shall present the theory of quadratic forms by Dennemeyer. The connections between conjugate surfaces, the quadratic form theory, and the Euler–Lagrange equations are the main results. In Section 3 we shall present the approximation theory of quadratic forms by Gregory which is sufficiently general to handle the quadratic forms in Section 2. The main results are given in [3] in terms of inequalities involving nonnegative indices. In particular, we shall show that the hypothesis for these inequalities are sufficiently general to include the "resolution spaces" of Hestenes [5] for focal point theories and "continuous" perturbations of coefficients of quadratic forms and partial differential equations.

In Section 4 we shall extend the approximation setting of Section 3 to obtain an approximate theory of conjugate surfaces. These results are then interpreted to obtain existence theorems and other properties for the multiple integral problem. In Section 5 we shall discuss how this theory may be applied to numerical focal point problems. In Section 6 some "test runs" are given.

The inequalities such as (4) are used on three levels in this paper. The first level leads to a theory of quadratic forms with applications given by Hestenes [5] and Dennemeyer [1]. The second level leads to an approximation theory for "level one" problems exemplified by Theorems 3–7. A third level is a numerical approximation theory for the "level two" problems (see Section 5).

2. Multiple Integral Quadratic Forms

In this section we give the quadratic form theory leading to the partial differential equation described in Section 1. We will define our fundamental Hilbert space (or Sobolev space) $\mathfrak{A}$, the quadratic form $J(x)$ to be considered, and then state a main theorem relating quadratic forms to partial differential equation. The notation and ideas are found in Dennemeyer [1]. For ease of presentation we refer the reader to this reference for technical details such as "smoothness conditions" on the coefficient functions R_{ij}, Q_i, and P, vectors $x(t)$, and on B^1 types of regions as found in the works of Calkins and Morrey.

Following Dennemeyer, we let m be a fixed positive integer, $T \subset \mathbb{R}^m$ be a fixed region of class B^1, $t = (t_1, t_2, ..., t_m)$ be a point in T, and $x(t)$ be a real-valued function defined on T. If $T_1 \subset T$, let $\bar{T}_1$ denote the closure of T and T_1^* denote the boundary of T_1. Let $\mathscr{H}$ be the Hilbert space of vectors $x(t)$ with inner product

$$(x, y) = \int_T \dot{x}_j(t)\, \dot{y}_j(t)\, dt + \int_T x(t)\, y(t)\, dt \tag{1}$$

with norm $\|x\| = (x, x)^{1/2}$, where $\dot{x}_j(t) = \partial x(t)/\partial t_j$, repeated indices are summed, and $i, j = 1, 2, ..., m$.

Our fundamental quadratic form on $\mathscr{H}$ is

$$J(x) = \int_T \{P(t)\, x^2(t) + [2Q_i(t)\, \dot{x}_i(t)]\, x(t) + R_{ij}(t)\, \dot{x}_i(t)\, \dot{x}_j(t)\}\, dt \tag{2}$$

with associated bilinear form

$$J(x, y) = \int_T \{Pxy + Q_i(x\dot{y}_i + \dot{x}_i\, y) + R_{ij}\dot{x}_i\, \dot{y}_j\}\, dt,$$

where $R_{ij}(t) = R_{ji}(t)$ and the ellipticity condition $R_{ij}(t)\, \xi_i \xi_j > 0$ holds for all t in the closure of T and $\xi = (\xi_1, \xi_2, ..., \xi_m)$ in $\mathbb{R}^n$ with $\xi \neq 0$.

The associated Euler–Lagrange equation or *extremal equation* for $J(x)$ is

$$E(x) = \frac{\partial}{\partial t_i}\left(R_{ij}\frac{\partial x}{\partial t_j}\right) - x\left(P - \sum_{i=1}^{m}\frac{\partial Q_i}{\partial t_i}\right) = 0. \tag{3}$$

For convenience, we assume additional conditions upon R_{ij}, P, and Q_i so that solutions of (2) are in $\mathscr{H} \cap C^2(T)$. In the remainder of this paper we shall assume that all function spaces are subspaces of the Hilbert space $\mathfrak{A} = \overline{C_0^\infty(T)}$ described in [1, p. 623]. That is, the vectors $x(t)$ are functions which "vanish" on the boundary $\partial T = T^*$ of T and are "smooth" on T. A

conjugate surface T_1^* of (3) is the boundary of a region $T_1 \subset T$ of class B^1 on which a nontrivial solution of (3) vanishes.

THEOREM 1. *Let $J(x)$ be the quadratic form given by* (2). *There exists a conjugate surface T_1^* with corresponding extremal solution $x(t)$ if and only if $J(x, y) = 0$ for all y in $\mathscr{H}$ which vanish in $\overline{T - T_1}$.*

This result follows by "integration by parts" and Dennemeyer's discussion [1, p. 631].

3. APPROXIMATION THEORY

In this section we give the major approximation theory and results. Much of this material is contained in Sections III and IV of [3], but with applications to integraldifferential equations. We invite those who are interested to read the omitted material for details; however, our exposition should be clear without such a reading. Thus we will omit technical details such as hypotheses (11) and (12) and Theorems 2–10 and summarize these results.

In [3] we began with a Hilbert space $\mathfrak{A}$, a metric space (Σ, ρ), a collection of quadratic forms $\{J(x; \sigma) | \sigma \in \Sigma\}$, and Hilbert subspaces $\{\mathfrak{A}(\sigma) | \sigma \in \Sigma\}$. For each $\sigma \in \Sigma$ we have a quadratic form $J(x; \sigma)$ defined on $\mathfrak{A}(\sigma)$, $J(x, y; \sigma)$ the associated (real) bilinear form, $s(\sigma)$ and $n(\sigma)$, respectively, the signature (index), and *nullity* of $J(x; \sigma)$ on $\mathfrak{A}(\sigma)$. In general the *signature* of a quadratic form $Q(x)$ on a subspace $\mathscr{L}$ of $\mathfrak{A}$ is the dimension of a maximal, linear subclass $\mathscr{E}$ of $\mathscr{L}$ such that $x \neq 0$ in $\mathscr{E}$ implies $Q(x) < 0$. The *nullity* of $Q(x)$ on $\mathscr{L}$ is the dimension of the space $\mathscr{L}_0 = \{x \in \mathscr{L} \mid Q(x, y) = 0 \text{ for all } y \in \mathscr{L}\}$.

We then defined a *resolution* $\{\mathscr{R}(\lambda) | a \leqslant \lambda \leqslant b\}$ of a space $\mathscr{R}$. That is, for each $\lambda \in [a, b]$, $\mathscr{R}(\lambda)$ is a closed subspace of $\mathscr{R}$. $\mathscr{R}(a) = 0$. $\mathscr{R}(b) = \mathscr{R}$, $a \leqslant \lambda_1 < \lambda_2 \leqslant b$ implies $\mathscr{R}(\lambda_1) \subset \mathscr{R}(\lambda_2)$.

$$\mathscr{R}(\lambda_0) = \bigcap_{\lambda_0 < \lambda \leqslant b} \mathscr{R}(\lambda) \qquad \text{whenever} \quad a \leqslant \lambda_0 < b$$

and

$$\mathscr{R}(\lambda_0) = \operatorname{cl} \bigcup_{a \leqslant \lambda < \lambda_0} \mathscr{R}(\lambda) \qquad \text{whenever} \quad a < \lambda_0 \leqslant b,$$

where $cl(S)$ denotes the closure of the set S.

This resolution concept is used to generate focal point, conjugate point or oscillation point phenomena. On one hand, we show that this concept is

contained in our approximation hypothesis of spaces $\{\mathfrak{A}(\sigma)|\sigma \in \Sigma\}$. More importantly, we show that the "σ-setting" can be generalized to the "μ-setting" of the next section and the results of the next theorem.

4. The Approximation Results

In this section we shall show that inequality involving $s(\sigma)$ and $n(\sigma)$ can be applied in a general way to obtain an approximation focal point theory. This theory can then be applied to a multitude of approximation problems in our setting.

Let $M = \Lambda \times \Sigma$ be the metric space with metric d defined by $d(\mu_1, \mu_2) = |\lambda_2 - \lambda_1| + \rho(\sigma_2, \sigma_1)$, where $\mu_1 = (\lambda_1, \sigma_1), \mu_2 = (\lambda_2, \sigma_2)$, (Σ, ρ) is a metric space, and $\Lambda = [a, b]$ with the usual absolute-valued metric. For each $\mu = (\lambda, \sigma)$ in M and $J(x;\ \sigma)$ define $J(x;\mu) = J(x;\sigma)$ on the space $\mathscr{E}(\mu) = \mathscr{A}(\sigma) \cap \mathscr{B}(\lambda)$. Let $s(\mu) = s(\lambda, \sigma)$ and $n(\mu) = n(\lambda, \sigma)$ denote the index (signature) and nullity of $J(x;\mu)$ on $\mathscr{E}(\mu)$.

In many senses Theorem 2 is the main result for applications to approximation problems. It allows us to obtain conditions (4) and (5) in very general problems.

Theorem 2. *Assume that the quadratic forms $J(x;\sigma)$ and the spaces $\mathscr{A}(\sigma)$ satisfy* (11) *and* (12) *of* [3]. *For any $\mu_0 = (\lambda_0, \sigma_0)$ in M there exists $\delta > 0$ such that if $\mu = (\lambda, \sigma)$, and $d(\mu_0, \mu) < \delta$, then*

$$s(\lambda_0, \sigma_0) \leqslant s(\lambda, \sigma) \leqslant s(\lambda, \sigma) + n(\lambda, \sigma) \leqslant s(\lambda_0, \sigma_0) + n(\lambda_0, \sigma_0). \tag{4}$$

Furthermore

$$n(\lambda_0, \sigma_0) = 0 \quad \text{implies } s(\lambda, \sigma) = s(\lambda_0, \sigma_0) \quad \text{and } n(\lambda, \sigma) = 0. \tag{5}$$

We now interpret Theorem 2 for the setting of this paper. As examples, the reader may regard $J(x;\sigma)$ as perturbations of $J(x)$ in (2) which may include an eigenvalue parameter ξ. For our numerical work $\mathfrak{A}(\sigma)$ will include doubly linear first-order spline functions described below. Resolution space examples are given in [1, pp. 629–630].

For each σ in Σ let

$$J(x;\sigma) = \int_T \{P_\sigma(t)\, x^2(t) + 2[Q_{\sigma i}(t)\, \dot{x}_i(t)]\, x(t) + R_{\sigma ij}(t)\, \dot{x}_i(t)\, \dot{x}_j(t)\}\, dt \tag{6}$$

be defined on a subspace $\mathfrak{A}(\sigma)$ of $\mathfrak{A}$ and let

$$E(x;\sigma)=\frac{\partial}{\partial t_i}\left(R_{\sigma ij}(t)\frac{\partial x}{\partial t_j}\right)-x\left(P_\sigma(t)-\sum_{i=1}^{m}\frac{\partial Q_{\sigma i}}{\partial t_i}\right)=0 \tag{7}$$

be the associated Euler–Lagrange equation. For each λ in $\Lambda=[a,b]$ let $\{\mathscr{B}(\lambda)|\lambda\in\Lambda\}$ be a resolution of $\mathfrak{A}$. For exposition purposes we assume Example 2 of [1]; namely, that T is the two-dimensional interval $[\bar{a},\bar{b}]$, where $\bar{a}=(a_1,a_2)$, $\bar{b}=(b_1,b_2)$ and $T(\lambda)=[\bar{a},\bar{b}(\lambda)]$, where $\bar{b}(\lambda)$ is linear such that $\bar{b}(a)=\bar{a}$, $\bar{b}(b)=\bar{b}$. The "bar" notation will be used in the remainder of this paper. Thus, in particular $\mathscr{B}(\lambda)$ is the set of functions $x(t)$ in $\mathfrak{A}$ with support in $T(\lambda)$. By Theorem 1 we have

THEOREM 3. *The nullity $n(\mu)=n(\lambda,\sigma)$ is the number of distinct nonzero solutions to* (7) *vanishing on $\mathscr{C}(\mu)$.*

We note that for σ_0 fixed $s(\lambda,\sigma_0)$ and $m(\lambda,\sigma_0)=s(\lambda,\sigma_0)+n(\lambda,\sigma_0)$ are nondecreasing nonnegative integer-valued functions of λ. It has been shown in [3] that $s(\lambda-0,\sigma)=s(\lambda,\sigma)$ and that the disjoint hypothesis of Theorem 4 implies $s(\lambda+0,\sigma)=s(\lambda,\sigma)+n(\lambda,\sigma)$. This hypothesis has been shown to hold in [1] in this setting. Thus $s(\lambda+0,\sigma_0)-s(\lambda-0,\sigma_0)=n(\lambda,\sigma_0)$. These results follow from (4). This disjoint hypothesis is usually called "normality" in problems of differential equations, calculus of variations, and control theory.

A point λ at which $s(\lambda,\sigma_0)$ is discontinuous will be called a *focal point* of $J(x;\sigma_0)$ relative $\mathscr{B}(\lambda)$ (λ in Λ). The difference $f(\lambda,\sigma_0)=s(\lambda+0,\sigma_0)-s(\lambda-0,\sigma_0)$ will be called the *order* of the focal point. A focal point will be counted the number of times equal to its order.

THEOREM 4. *Assume for σ_0 in Σ that $\mathscr{C}_0(\lambda_1,\sigma_0)\cap\mathscr{C}_0(\lambda_2,\sigma_0)=0$ when $\lambda_1\neq\lambda_2$. Then $f(a,\sigma_0)=0, f(\lambda,\sigma_0)=n(\lambda,\sigma_0)$ on $a\leqslant\lambda\leqslant b$. Then, if λ_0 in Λ, the following quantities are equal:*

(i) *the sum $\sum_{a\leqslant\lambda<\lambda_0}n(\lambda,\sigma_0)$,*

(ii) *the signature $s(\lambda_0,\sigma_0)$ of $J(x;\sigma)$ on $\mathscr{B}(\lambda_0)$,*

(iii) *the sum $\Sigma\, s(\lambda_i+0,\sigma_0)-s(\lambda_i,\sigma_0)$ taken over all λ_i such that $a\leqslant\lambda_i<\lambda_0$ and $s(\lambda,\sigma_0)$ discontinuous at λ_i,*

(iv) *the number of conjugate surfaces on $a\leqslant\lambda<\lambda_0$,*

(v) *the number of focal points on $a\leqslant\lambda<\lambda_0$,*

(vi) *the number of λ_i and corresponding $x\neq 0$ as described in Theorem* 3 *with $a<\lambda_i<\lambda_0$.*

For the approximation setting we can say much more. In the next two

results we assume that σ_0 in Σ satisfies $\mathscr{C}_0(\lambda_1, \sigma_0) \cap \mathscr{C}_0(\lambda_2, \sigma_0) = 0$ when $\lambda_1 \neq \lambda_2$. Since this implies that $n(\lambda, \sigma_0) = 0$ except for a finite number of points λ in Λ we have

THEOREM 5. *Assume λ' and λ'' are not focal points of σ_0 $(a \leqslant \lambda' < \lambda'' < b)$ and $\lambda_q(\sigma_0) \leqslant \lambda_{q+1}(\sigma_0) \leqslant \cdots \leqslant \lambda_{q+k-1}(\sigma_0)$ are the k focal points of σ_0 on (λ', λ''). Then there exists an $\varepsilon > 0$ such that $\rho(\sigma, \sigma_0) < \varepsilon$ implies $\lambda_q(\sigma) \leqslant \lambda_{q+1}(\sigma) \leqslant \cdots \leqslant \lambda_{q+k-1}(\sigma)$ are the k focal points of σ on (λ', λ'').*

COROLLARY 6. *The kth focal point $\lambda_k(\sigma)$ is a continuous function of $k = 1, 2, \ldots$, as is the kth conjugate surface.*

We note that in [3, Section V] we indicated that a wide variety of eigenvalue results, comparison theorems, and Sturm-type separation theorems follow from Theorem 2. Once again we refer the reader to these results. Reference [1] gives many nice comparison theorems of oscillation and conjugate surfaces also.

As an example of our methods we use Theorem 4 to generalize Corollary 8.3 of [1]. We assume that $R_{ij}(t) = R_{\sigma_0 ij}(t)$ and $P(t) = P_{\sigma_0}(t)$ are defined on T and $P(t) > 0$ on a fixed subspace $T(\lambda_0) \subset T$, where $a < \lambda_0 < b$. Then

THEOREM 7. *There exists a $\delta > 0$ such that if $\mu_0 = (\lambda_0, \sigma_0)$, $\mu = (\lambda, \sigma)$, and $|\lambda_0 - \lambda| + \rho(\sigma_0, \sigma) < \delta$, then no solution on $T(\lambda)$ of the differential equation*

$$\frac{\partial}{\partial t_i}\left(R_{\sigma ij}(t)\frac{\partial x}{\partial t_j}\right) - P_\sigma(t)\,x = 0$$

oscillates in $T(\lambda)$ in the sense that no conjugate surface is properly contained in $T(\lambda)$.

The hypothesis implies that

$$\int_{T(\lambda)} |R_{ij}(t)\,\dot{x}_j(t)\,\dot{x}_j(t) + P(t)\,x^2(t)|\,dt > 0$$

for $x(t)$ in $\mathscr{B}(\lambda_0)$ and hence that $s(\lambda_0, \sigma_0) = 0$ and $n(\lambda_0, \sigma_0) = 0$. Thus by the above there exists $\delta > 0$, such that $s(\lambda, \sigma) = 0$ and $n(\lambda, \sigma) = 0$ whenever $|\lambda_0 - \lambda| + s(\sigma_0, \sigma) < \delta$. This completes the proof.

We remark that the parameter σ above can "include" the eigenvalue parameter λ. For example, let $K(x, \sigma) = \int_T Q_\sigma(t)\,x^2(t)\,dt$ for $\sigma \in \Sigma$. Then by defining $H(x; \sigma, \xi, \lambda) = J(x, \sigma) - \xi K(x; \sigma)$, where ξ is a real parameter,

Theorems 4–6 generalize to the corresponding eigenvalue results for elliptic-type partial differential equations.

5. The Numerical Problem

In this section we give new theory, procedures, and results for the numerical computation of conjugate surfaces of the quadratic form (2) or Eq. (3). The technical theoretical results are similar to those given in [4] for the second order differential equation $(r(t)\,x'(t))' + p(t)\,x(t) = 0$ and are left, in this case, as an exercise for the reader.

To fix ideas and to make the calculations easier we considered in this exposition an elementary example; namely, $m = 2$, $R_{11}(t) = R_{22}(t) = 1$, $R_{12}(t) = R_{21}(t) = 0$, and $P(t) = 2$. It is immediate that any multiple of $x(t_1, t_2) = \sin t_1 \sin t_2$ satisfies differential equation (8) and is an extremal solution of (9) on an interval $T = [\bar{0}, \bar{b}] \subset \mathbb{R}^2$, where $\bar{b} = (b, b)$ and b is a large fixed positive number. We have considered another case with equivalent results which will not be reported here.

The major idea is as follows:

(A) the partial differential equation and initial conditions

$$\frac{\partial^2 x}{\partial t_1^2} + \frac{\partial^2 x}{\partial t_2^2} + 2x(t_1, t_2) = 0, \tag{8a}$$

$$x(t_1, 0) = 0, \qquad x(0, t_2) = 0 \qquad (0 \leqslant t_1 \leqslant b, \quad 0 \leqslant t_2 \leqslant b) \tag{8b}$$

are replaced by

(B) the quadratic form

$$J(x) = \int_T [\dot{x}_1^2(t) + \dot{x}_2^2(t) - 2x^2(t)]\, dt_1\, dt_2. \tag{9}$$

(C) A finite-dimensional quadratic form with matrix $D(\sigma)$, which is real, symmetric, and block tridiagonal, is shown to be a numerical approximation of (4) or (9).

(D) We then compute $x_\sigma(t)$ the Euler–Lagrange equation of $D(\sigma)$ and show that, if properly normalized, $x_\sigma(t)$ converges to the solution $x_0(t)$ of (3) or (9) as $\sigma \to 0$. In our case, $x_\sigma(t)$ is the discrete bilinear approximation of $x_0(t) = \sin t_1 \sin t_2$ corresponding to a "mesh size" of σ.

Unfortunately, we cannot directly compute a solution using $D(\sigma)$ as we can in the second oder ordinary case. In the situation where $D(\sigma)$ is a tridiagonal matrix, we can directly compute the numerical approximation $x_0(t)$ (see [4]). This is to be expected from the theory of elliptic partial

differential equations, the numerical results of [2], or the heuristic feeling of roundoff error, regardless of the accuracy of the computer. We will verify that $D(\sigma)$ is correct by checking the known discrete solution and by relaxation methods which are discussed below.

We begin our numerical procedure by choosing Σ to denote the set of real numbers of the form $\sigma = n^{-1}$ $(n = 1, 2, 3,...,)$ and 0. For $\sigma = n^{-1}$, define the two-dimensional partition $\pi_2(\sigma) = \pi(\sigma) \times \pi(\sigma)$ of $[\bar{0}, \bar{b}]$, where $a_k = kb/n$ $(k = 0, 1, 2,..., N_\sigma)$ and

$$\pi(\sigma) = (a_0 = 0 < a_1 < a_2 < \cdots < a_{N_\sigma} = b).$$

We assume for convenience and without loss of generality that $a_{N_\sigma} = b$. The space $\mathfrak{A}(\sigma)$ is the set of continuous bilinear functions with vertices at $\pi_2(\sigma)$. Thus $\mathfrak{A}(\sigma)$ is the vector space of bivariate splines with basis $z_{ij}(t_1, t_2) = y_i(t_1) y_j(t_2)$, where $y_k(s)$ $(k = 1,..., N_\sigma - 1)$ is the one-dimensional spline hat function

$$\begin{aligned} y_k(s) &= 1 - |s - a_k|/\sigma \qquad \text{if} \quad a_{k-1} \leqslant s \leqslant a_{k+1}, \\ &= 0, \qquad\qquad\qquad\quad \text{otherwise.} \end{aligned}$$

The basis elements $z_{ij}(t_1, t_2)$ are pyramids with apex or vertex at the point $(a_i, a_j, 1)$ in $\mathbb{R}^3$ and support in the square with corner points $P_1(a_{i-1}, a_{j-1})$, $P_2 = (a_{i-1}, a_{j+1})$, $P_3 = (a_{i+1}, a_{j-1})$, and $P_4 = (a_{i+1}, a_{j+1})$. Finally, let $\mathfrak{A}(0)$ denote the space of smooth functions described in Section 2, defined on the rectangle $T = [\bar{0}, \bar{b}] \subset \mathbb{R}^2$, and vanishing on ∂T, the boundary of T.

For each λ in $[0, b]$, let $\mathscr{H}(\lambda)$ denote the arcs $x(t)$ in $\mathfrak{A}(0)$ with support in the interval $[\bar{0}, \bar{\lambda}]$ of $\mathbb{R}^2$. If $\mu = (\lambda, \sigma)$ is in the metric space $M = [0, b] \times \Sigma$ with metric $d(\mu_1, \mu_2) = |\lambda_2 - \lambda_1| + |\sigma_2 - \sigma_1|$, let $\mathscr{B}(\mu) = \mathfrak{A}(\sigma) \times \mathscr{H}(\lambda)$. Thus an arc $x(t)$ in $\mathscr{B}(\lambda, \sigma)$ is a bivariate spline with support in $[\bar{0}, \bar{a}_k] \subset \mathbb{R}^2$, where $a_k \leqslant \lambda < a_{k+1}$.

Because of our sample problem, we define $J(x; \mu) = J(x; \sigma)$ as in (9), restricted to the class of functions $\mathfrak{A}(\sigma)$. In the more general case, we would define $J(x; \sigma)$ similar to (5) in [4], where, for example, $P_\sigma(t) = P(a_i, a_j)$ if t is in the square given by P_1, P_2, P_3, and P_4 above. A straightforward calculation in the next paragraph (for $\sigma \neq 0$) shows that $J(x; \mu) = b_\alpha b_\alpha e_{\alpha\beta}(\mu) = x^T D(\mu) x$, where $x = (b_1, b_2,...,)^T = \Sigma b_\alpha w_\alpha(t)$, $e_{\alpha\beta}(\mu) = J(w_\alpha, w_\beta; \mu)$, and $D(\mu)$ is a symmetric tridiagonal block of tridiagonal matrices "increasing" in λ so that the "upper" submatrix of $D(a_{k+1}, \sigma)$ is $D(a_k, \sigma)$. In the above, $w_\alpha(t) = z_{ij}(t)$, where the correspondence $\alpha \leftrightarrow (i, j)$ is one-to-one and given below.

To construct $D(\sigma)$, we assume the double-subscripted notation above; then

$$J(z_{ij}, z_{kl}) = \int_0^b \int_0^b \left[\left(\frac{\partial z_{ij}}{\partial t_1}\right)\left(\frac{\partial z_{kl}}{\partial t_1}\right) + \left(\frac{\partial z_{ij}}{\partial t_2}\right)\left(\frac{\partial z_{kl}}{\partial t_2}\right) - 2z_{ij}^2 z_{kl}^2 \right] dt_1\, dt_2$$

$$= \int_0^b \int_0^b [(y_i'(t_1)\, y_j(t))(y_k'(t_1)\, y_l(t_2))$$

$$+ (y_i(t_1)\, y_j'(t_2))(y_k(t_1)\, y_l'(t_2))$$

$$- 2(y_i(t_1)\, y_j(t_2))(y_k(t_1)\, y_l(t_2))]\, dt_1\, dt_2.$$

If $|i-k| > 1$ or if $|j-l| > 1$ the above is 0. Otherwise, we have

$$J(z_{ij}, z_{ij})$$

$$= \int_{a_{j-1}}^{a_{j+1}} \int_{a_{j-1}}^{a_{j+1}} [y_i'^2(t_1)\, y_j^2(t_2) + y_i^2(t_1)\, y_j'^2(t_2)$$

$$- 2y_i^2(t_1)\, y_j^2(t_2)]\, dt_1\, dt_2$$

$$= \left(\frac{2}{\sigma}\right)\left(\frac{2}{3}\sigma\right) + \left(\frac{2}{3}\sigma\right)\left(\frac{2}{\sigma}\right) - 2\left(\frac{2}{3}\sigma\right)\left(\frac{2}{3}\sigma\right) = \frac{8}{3} - \frac{8}{9}\sigma^2, \qquad (10a)$$

$$J(z_{i+1,j}, z_{ij})$$

$$= \int_{a_{j-1}}^{a_{j+1}} \int_{a_{i-1}}^{a_{i+2}} [y_{i+1}'(t_1)\, y_i'(t_1)\, y_j^2(t_2) + y_{i+1}(t_1)\, y_i(t_1)\, y_j'^2(t_2)$$

$$- 2y_{i+1}(t_1)\, y_i(t_1)\, y_j^2(t)]\, dt_1\, dt_2$$

$$= \left(-\frac{1}{\sigma}\right)\left(\frac{2}{3}\sigma\right) + \left(\frac{1}{6}\sigma\right)\left(\frac{2}{\sigma}\right) - 2\left(\frac{1}{6}\sigma\right)\left(\frac{2}{3}\sigma\right) = -\frac{1}{3} - \frac{2}{9}\sigma^2, \qquad (10b)$$

and

$$J(z_{i+1,j+1}, z_{i,j})$$

$$= \int_{a_{j-1}}^{a_{j+1}} \int_{a_{i-1}}^{a_{i+2}} [y_{i+1}'(t_1)\, y_i'(t_1)\, y_{j+1}(t_2)\, y_j(t_2)$$

$$+ y_{i+1}(t_1)\, y_i(t_1)\, y_{j+1}'(t_2)\, y_j'(t_2)$$

$$- 2y_{i+1}(t_1)\, y_i(t_1)\, y_{j+1}(t_2)\, y_j(t_2)]\, dt_1\, dt_2$$

$$= \left(-\frac{1}{\sigma}\right)\left(\frac{1}{6}\sigma\right) + \left(\frac{1}{6}\sigma\right)\left(-\frac{1}{\sigma}\right) - 2\left(\frac{1}{6}\sigma\right)\left(\frac{1}{6}\sigma\right) = -\frac{1}{3} - \frac{1}{18}\sigma^2. \qquad (10c)$$

We have carried out our calculations so that more complicated cases may be considered. Thus, for example, (10a) would become

$$R_{\sigma ij}(a_i, a_j)\left(\frac{2}{\sigma}\right)\left(\frac{2}{3}\sigma\right) + R_{\sigma ij}(a_i, a_j)\left(\frac{2}{3}\sigma\right)\left(\frac{2}{\sigma}\right) - P_{\sigma ij}(a_i, a_j)\left(\frac{2}{3}\sigma\right)\left(\frac{2}{3}\sigma\right)$$

if $J(x)$ is given in (2) with $m = 2$ and $Q_1(t) = Q_2(t) = 0$.

We now show that $D(\sigma)$ is the approximating finite-dimensional matrix to $J(x)$ on T and hence $D(\lambda, \sigma)$ is the approximating finite-dimensional matrix to $J(x)$ on $\mathscr{F}(\lambda)$. Let $\alpha = \alpha_\sigma(i,j) = N_\sigma i + j$ $(i,j = 1,..., N_\sigma)$ and $\beta = \beta_\sigma(k, l) = N_\sigma k + l$ $(k, l = 1,..., N_\sigma)$. Let $w_\alpha(t) = z_{ij}(t)$ and $x_0(t)$ be an extremal solution of (9) where we assume $x_0(t) = \sin t_1 \sin t_2$ for our example. Let $C = \{c_1, c_2, c_3,...,\}$ be a sequence of real numbers given by an algorithm similar to (8) of [4] (see Eq. (11)) and $x_\sigma(t) = \sum c_\alpha w_\alpha(t)$. This is done so that $x_0(a_i, a_j) = x_\sigma(a_i, a_j)$ if either $i = 1$ or $j = 1$. The C vector is the Euler–Lagrange solution of $D(\sigma)$, that is, $D(\sigma)\, C^T \cong 0$, where "$\cong$" is described in [4] if $D(\sigma)$ is tridiagonal, and in Section 5 of this paper if $D(\sigma)$ is block tridiagonal. Finally we have

THEOREM 8. *The vector $x_\sigma(t)$ described above converges strongly to $x_0(t)$* (as $\sigma \to 0$) *in the derivative norm sense of* (1); *that is, if*

$$g(\sigma) = \int_T \left\{ \left[\frac{\partial}{\partial t_1}(x_0(t) - x_\sigma(t))\right]^2 + \left[\frac{\partial}{\partial t_2}(x_0(t) - x_\sigma(t))\right]^2 + [x_0(t) - x_\sigma(t)]^2 \right\} dt,$$

then $g(\sigma) \to 0$ as $\sigma \to 0$.

We remark that we can derive results as in Theorems 2–7 of [4] or Theorems 4–6 of this paper width σ as the numerical parameter. This will be left to the reader.

6. Block Tridiagonal Matrices and Test Results

In this section, we shall describe "in pictures" the matrix $D(\sigma)$ and give numerical test results. This matrix is found in more classical settings of numerical solutions of partial differential equations by finite-difference approximation of the derivatives [2]. Our methods are different in that we

approximate the integration problem which will be smoother. Note that Theorem 8 gives very strong convergence results even when the coefficient functions are not very smooth.

We hope (expect) our ideas to shed more light on block tridiagonal matrices of this type. Thus we hope to show in a later paper (by "separation of variables") that $D(\sigma)$ is a linear combination of tridiagonal matrices analogous to the continuous case.

The "picture" is as follows. The Euler–Lagrange equation is

$$\begin{pmatrix} E_1 & F_1 & 0 & 0 & \\ G_2 & E_2 & F_2 & 0 & \\ 0 & G_3 & E_3 & F_3 & \cdots \\ 0 & 0 & G_4 & E_4 & \\ & & \vdots & & \ddots \end{pmatrix} \begin{pmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \\ \vdots \end{pmatrix} \cong 0. \tag{11a}$$

In the above, E_n, F_n, and G_n are $N \times N$ tridiagonal matrices, E_n is symmetric, $G_n^T = F_{n-1}$, and C_n is an $N \times 1$ column matrix corresponding to the points $\{(a_n, t_2 | t_2 \in \pi(\sigma)\}$. If $\lambda \neq b$, then the "latter elements" of E_n, F_n, G_n, and C_n contain the appropriate zeros. The "matrix equation" becomes (for $m = 1,\ldots, N$)

$$G_m C_{m-1} + E_m C_m + F_m C_{m+1} = 0 \tag{11b}$$

with associated "computer equation" (for $k = 1,\ldots, N$)

$$\begin{aligned} &g^m_{k,k-1} c_{m-1,k-1} + g^m_{k,k} c_{m-1,k} + g^m_{k,k+1} c_{m-1,k+1} + e^m_{k,k-1} c_{m,k-1} \\ &\quad + e^m_{k,k} c_{m,k} + e^m_{k,k+1} c_{m,k+1} + f^m_{k,k-1} c_{m+1,k-1} + f^m_{k,k} c_{m+1,k} \\ &\quad + f^m_{k,k+1} c_{m+1,k+1} = 0. \end{aligned} \tag{11c}$$

In all cases, a subscript of zero indicates a zero block matrix as does an index which "takes us past" the value of λ.

As we indicated above, (11) does not yield a direct numerical solution as does (8) of [4] for the second-order case. Our test results involve two different ideas which we label (A) and (B). In case A we check $D(\sigma)$ for our sample problem. In case B we use the method of overrelaxation to compute a solution.

Case A. Direct Verification. In this case we take $T = [\bar{0}, \bar{\pi}] \subset \mathbb{R}^2$ and choose a step size of $\sigma = \pi/70$. The known solution is $x_0(t) = \sin t_1 \sin t_2$. We build a numerical solution with elements $c_{ij} = \text{sn}(a_i) \sin(a_j)$, and letting C_n and $D(\sigma)$ as described above, we obtain the sum

$\sum c_\alpha c_\beta e_{\alpha\beta} = 0.952 \times 10^{-7}$. The computation was performed in double precision and is the approximation of

$$\int_0^{\pi}\int_0^{\pi} (\cos^2 t_1 \sin^2 t_2 + \cos^2 t_2 \sin^2 t_1 - 2 \sin^2 t_1 \sin^2 t_2)\, dt_1\, dt_2$$

$$= \int_0^{\pi} \cos 2t_1\, dt_1 \int_0^{\pi} \sin^2 t_2\, dt + \int_0^{\pi} \cos 2t_2\, dt_2 \int_0^{\pi} \sin^2 t_1\, dt_1$$

$$= 0. \tag{12}$$

To show that all answers are not zero when $\lambda = \pi/2 + \pi/90$, we obtain numerically $\sum c_\alpha c_\beta e_{\alpha\beta} = 0.402 \times 10^{-2}$. This is the numerical approximation of the function equal to $\sin s \sin t$ on $[\bar{0}, \bar{\pi}/2]$, bilinear in s and t on $[\bar{0}, (\pi/2) + (\pi/90)]$, and vanishing on the boundary of $[\bar{0}, \overline{(\pi/2) + (\pi/90)}]$. This number is not meaninglful except to note that it must be large and positive. Otherwise there would be a vector on $[\bar{0}, \overline{(\pi/2) + (\pi/90)}]$ vanishing on this boundary such that $J(x; \sigma) \leqslant 0$ which would imply that with $\lambda = (\pi/2) + (\pi/90)$ we have $s(\lambda, \sigma) + n(\lambda, \sigma) \geqslant 1$ which is not possible until $\lambda \geqslant \pi$. Note that in fact $\sin s \sin t$ integrated over $[\bar{0}, \bar{\pi}/2]$ in (12) would also be zero but we must "wait" for a conjugate surface, that is, until this function vanishes on the boundary of $[\bar{0}, \bar{\pi}]$.

Case B. Relaxation. By relaxation we mean a procedure where we assign initial values to the vector C of (11a) and then use (11c) to calculate the current value of $c_{m,k}$ using the "eight" neighboring points. This topic is discussed in detail in [2, Chapters 21 and 22]. One such pass with m, $k = 1,\ldots, N$ is called an *iteration.*

For the problem described above with solution $x_0(t_1, t_2) = \sin t_1 \sin t_2$ in $[\bar{0}, \bar{\pi}]$ with step size $\sigma = \pi/50$, we obtain a maximum error less than 0.2×10^{-3} after 500 iterations and less than 0.25×10^{-4} after 1000 iterations. The calculations were performed in single precision and took approximately 2 min. of computer terminal time. (We have no method of obtaining accurate timing.)

Our results were not as good when we changed the coefficient functions to nonconstant values. Thus for the equation

$$\frac{\partial x^2}{\partial t_1^2} + \frac{\partial}{\partial t_2}\left[(2 + \cos t_2)\frac{\partial x}{\partial t_2}\right] + (3 + 2\cos t_2)\, x = 0 \tag{13a}$$

or associated quadratic form

$$J(x) = \int_0^{\pi}\int_0^{\pi} [\dot{x}_1^2 + (2 + \cos t_2)\, \dot{x}_2^2 - (3 + 2\cos t_2)\, x(t)]\, dt_1\, dt_2, \tag{13b}$$

we note that as before $x_0(t) = \sin t_1 \sin t_2$ is a solution to (13a) vanishing on te boundary of $[\bar{0}, \bar{\pi}]$. In this case, with $\sigma = \pi/50$ we obtain a maximum error of 0.65×10^{-2} after 500 iterations with little improvement after 1000 iterations. With $\sigma = \pi/100$ and 2000 iterations we obtained a maximum error less than 0.35×10^{-2}.

Finally, for (13) we observed phenomena in our relaxation methods consistent with the theory [2]. If our interval is $[\bar{0}, \overline{2.5}]$, then $D(\sigma)$ is positive definite and our relaxation method drove the solution toward the zero slution (very slowly). If our interval is $[\bar{0}, \overline{3.5}]$, our solution rapidly diverges.

REFERENCES

1. R. DENNEMEYER, Conjugate surfaces for multiple integral problems in the calculus of variations, *Pacific J. Math.* **30** (3) (1969), 621–638.
2. G. E. FORSYTHE AND W. R. WASOW, "Finite difference Methods for Partial Differential Equations," Wiley, New York, 1967.
3. J. GREGORY AND G. C. LOPEZ, An approximation theory for generalized Fredholm quadratic forms and integral-differential equations, *Trans. Amer. Math. Soc.* **222** (1976), 319–335.
4. J. GREGORY, Numerical algorithms for oscillation vectors of second order differential equations including the Euler–Lagrange equation for symmetric tridiagonal matrices, *Pacific J. Math.* **76** (2) (1978), 397–406.
5. M. R. HESTENES, Application of the theory of quadratic forms in Hilbert space to the calculus of variations, *Pacific J. Math.* **1** (1951), 525–581.
6. M. R. HESTENES, Quadratic variational theory and linear elliptic partial differential equations, *Trans. Amer. Math. Soc.* **101** (1961), 306–350.