On Interpolation and Integration in Finite-Dimensional Spaces of Bounded Functions

Per-Gunnar Martinsson*, Vladimir Rokhlin†, and Mark Tygert‡

Research Report YALEU/DCS/RR-1317
March 9, 2005

Abstract

We observe that, under very mild conditions, an n-dimensional space of functions (with a finite n) admits numerically stable n-point interpolation and integration formulae. The proof relies entirely on linear algebra, and is virtually independent of the domain and of the functions to be interpolated.

Approximation of functions and construction of quadrature formulae constitute an extremely well-developed area of numerical analysis; in most situations one is likely to encounter in practice, existing tools are satisfactory. Much of the research concentrates on obtaining powerful results under strong assumptions — designing interpolation and quadrature formulae for smooth functions on subspaces of $\mathbb{R}^n$, manifolds, etc. In this note, we make a very general observation that, given a finite set of bounded functions $f_1, f_2, \ldots, f_{n-1}, f_n$ (either real- or complex-valued) defined on a set S, there exists an interpolation formula that is exact on all linear combinations of $f_1, f_2, \ldots, f_{n-1}, f_n$, is numerically stable, and is based on n nodes in S (to be denoted $x_1, x_2, \ldots, x_{n-1}, x_n$). If, in addition, S is a measure space, and the functions $f_1, f_2, \ldots, f_{n-1}, f_n$ are integrable, then there exists a quadrature formula based on the nodes $x_1, x_2, \ldots, x_{n-1}, x_n$ that is exact on all functions $f_1, f_2, \ldots, f_{n-1}, f_n$, and is also numerically stable. Both of these statements are purely linear-algebraic in nature, and do not depend on the detailed properties of S, or of the functions $f_1, f_2, \ldots, f_{n-1}, f_n$.

It should be pointed out that all of the statements in this note follow easily from the analysis found in [3]; moreover, Theorem 2 can be found (in a slightly different form) in [4]. While we cannot cite any earlier works where these observations are published, it seems unlikely that they had not been made a long time ago (perhaps in contexts other than numerical analysis).

Throughout this note, S denotes an arbitrary set, n denotes a positive integer, and $f_1, f_2, \ldots, f_{n-1}, f_n$ denote bounded complex-valued functions on S (all results of this note also

*Partially supported by the U.S. DoD under AFOSR Award #5710001605.
†Partially supported by the U.S. DoD under AFOSR Grant #F49620–03–C–0041.
‡Partially supported by the U.S. NSF under Award #DMS–0139914.

apply in the real-valued case, provided that the word "complex" is replaced with "real" everywhere). For any n points $x_1,\, x_2,\, \dots,\, x_{n-1},\, x_n$ in S, we define $A = A(x_1,\, x_2,\, \dots,\, x_{n-1},\, x_n)$ to be the $n \times n$ matrix defined via the formula

$$A_{j,k} = f_j(x_k) \tag{1}$$

with $j, k = 1,\, 2,\, \dots,\, n-1,\, n$; we define the function g_k on S to be the ratio of the determinant of $A(x_1,\, x_2,\, \dots,\, x_{k-2},\, x_{k-1},\;\; x,\;\; x_{k+1},\, x_{k+2},\, \dots,\, x_{n-1},\, x_n)$ to the determinant of $A(x_1,\, x_2,\, \dots,\, x_{n-1},\, x_n)$, via the formula

$$g_k(x) = \frac{\det A(x_1,\, x_2,\, \dots,\, x_{k-2},\, x_{k-1},\;\; x,\;\; x_{k+1},\, x_{k+2},\, \dots,\, x_{n-1},\, x_n)}{\det A(x_1,\, x_2,\, \dots,\, x_{n-1},\, x_n)} \tag{2}$$

(here, the numerator is the same as the denominator, but with x in place of x_k). We define $D(x_1,\, x_2,\, \dots,\, x_{n-1},\, x_n)$ to be the modulus of the determinant of $A(x_1,\, x_2,\, \dots,\, x_{n-1},\, x_n)$, via the formula

$$D(x_1,\, x_2,\, \dots,\, x_{n-1},\, x_n) = |\det A(x_1,\, x_2,\, \dots,\, x_{n-1},\, x_n)|. \tag{3}$$

We define B to be the supremum of $D(x_1,\, x_2,\, \dots,\, x_{n-1},\, x_n)$ taken over all sets of n points $x_1,\, x_2,\, \dots,\, x_{n-1},\, x_n$ in S, via the formula

$$B = \sup_{x_1,\, x_2,\, \dots,\, x_{n-1},\, x_n \text{ in } S} D(x_1,\, x_2,\, \dots,\, x_{n-1},\, x_n). \tag{4}$$

For any x in S, we define $u = u(x)$ to be the $n \times 1$ column vector defined via the formula

$$u_k = f_k(x) \tag{5}$$

with $k = 1,\, 2,\, \dots,\, n-1,\, n$, and we define $v = v(x)$ to be the $n \times 1$ column vector defined via the formula

$$v_k = g_k(x) \tag{6}$$

with $k = 1,\, 2,\, \dots,\, n-1,\, n$.

Theorem 2 below asserts the existence of numerically stable n-point interpolation formulae for any set of n bounded functions; first we will need the following lemma.

Lemma 1 *Suppose that n is a positive integer, S is an arbitrary set containing at least n points, and $f_1,\, f_2,\, \dots,\, f_{n-1},\, f_n$ are complex-valued functions on S that are linearly independent.*

Then, there exist n points $x_1,\, x_2,\, \dots,\, x_{n-1},\, x_n$ in S such that the vectors $u(x_1)$, $u(x_2)$, $\dots$, $u(x_{n-1})$, $u(x_n)$ defined in (5) are linearly independent.

Proof. We apply the modified Gram-Schmidt process (see, for example, [2]) to the set of all $n \times 1$ column vectors $u(x)$ defined in (5) for all x in S, while ensuring that all pivot vectors are non-zero via appropriate column-pivoting. □

Theorem 2 *Suppose that S is an arbitrary set, n is a positive integer, f_1, f_2, $\ldots$, f_{n-1}, f_n are bounded complex-valued functions on S, and ε is a positive real number such that*

$$\varepsilon \le 1. \tag{7}$$

Then, there exist n points x_1, x_2, $\ldots$, x_{n-1}, x_n in S and n functions g_1, g_2, $\ldots$, g_{n-1}, g_n on S such that

$$|g_k(x)| \le 1 + \varepsilon \tag{8}$$

for all x in S and $k = 1, 2, \ldots, n-1, n$, and

$$f(x) = \sum_{k=1}^{n} f(x_k)\, g_k(x) \tag{9}$$

for all x in S and any function f defined on S via the formula

$$f(x) = \sum_{k=1}^{n} c_k\, f_k(x), \tag{10}$$

for some complex numbers c_1, c_2, $\ldots$, c_{n-1}, c_n.

Proof. We assume without loss of generality that the functions f_1, f_2, $\ldots$, f_{n-1}, f_n are linearly independent.

Then, due to Lemma 1, B defined in (4) is strictly positive. Since the functions f_1, f_2, $\ldots$, f_{n-1}, f_n are bounded, $D(x_1, x_2, \ldots, x_{n-1}, x_n)$ (defined in (3)) is also bounded, and hence B defined in (4) is not only strictly positive, but also finite. Therefore, there exist n points x_1, x_2, $\ldots$, x_{n-1}, x_n in S such that

$$B - D(x_1, x_2, \ldots, x_{n-1}, x_n) \le \frac{B}{2}\varepsilon \tag{11}$$

and $D(x_1, x_2, \ldots, x_{n-1}, x_n)$ is strictly positive.

Defining g_1, g_2, $\ldots$, g_{n-1}, g_n via (2), we obtain (9) from the Cramer rule applied to the linear system

$$Av = u, \tag{12}$$

where $A = A(x_1, x_2, \ldots, x_{n-1}, x_n)$ is defined in (1), $v = v(x)$ is defined in (6), and $u = u(x)$ is defined in (5). Due to the combination of (11) and (7),

$$\frac{B}{2} \le D(x_1, x_2, \ldots, x_{n-1}, x_n), \tag{13}$$

and due to the combination of (11) and (13),

$$\frac{B}{D(x_1, x_2, \ldots, x_{n-1}, x_n)} - 1 \le \varepsilon; \tag{14}$$

we also observe that, due to (4),

$$D(x_1, x_2, \ldots, x_{k-2}, x_{k-1},\ \ x,\ \ x_{k+1}, x_{k+2}, \ldots, x_{n-1}, x_n) \le B \tag{15}$$

for all x in S. Now, (8) is an immediate consequence of (2), (3), (15), (14). □

Remark 3 Due to (8), the interpolation formula (9) is numerically stable.

Remark 4 When calculations are performed using floating point arithmetic, it is often desirable to "normalize" the set of functions $f_1, f_2, \dots, f_{n-1}, f_n$ before applying Theorem 2, by replacing this set with the set of functions $\tilde{f}_1, \tilde{f}_2, \dots, \tilde{f}_{n-1}, \tilde{f}_n$, where $\tilde{f}_k$ is the function defined on S via the formula

$$\tilde{f}_k(x) = \frac{f_k(x)}{\sum_{j=1}^{n} |f_j(x)|}, \tag{16}$$

for example.

Remark 5 When S is compact and the functions $f_1, f_2, \dots, f_{n-1}, f_n$ are continuous, Theorem 2 holds with $\varepsilon = 0$ rather than $\varepsilon > 0$, since a continuous function D on a compact space S attains its maximal value. For the same reason, Theorem 2 holds with $\varepsilon = 0$ when $S = \mathbb{R}^d$ for some positive integer d, the functions $f_1, f_2, \dots, f_{n-1}, f_n$ are continuous, and $f_k(x) \to 0$ as $|x| \to \infty$ for all $k = 1, 2, \dots, n-1, n$.

The following theorem formalizes the obvious observation that integrating both sides of (9) yields numerically stable quadrature formulae.

Theorem 6 *Suppose that S is a measure space, w is a nonnegative real-valued integrable function on S (that serves as the weight for integration), n is a positive integer, $f_1, f_2, \dots, f_{n-1}, f_n$ are bounded complex-valued integrable functions on S, and $\varepsilon \leq 1$ is a positive real number.*

Then, there exist n complex numbers $w_1, w_2, \dots, w_{n-1}, w_n$ such that

$$|w_k| \leq (1+\varepsilon) \int w(x)\, dx \tag{17}$$

for all $k = 1, 2, \dots, n-1, n$, and

$$\int f(x)\, w(x)\, dx = \sum_{k=1}^{n} w_k\, f(x_k) \tag{18}$$

for any function f defined on S via (10), where $x_1, x_2, \dots, x_{n-1}, x_n$ are the n points in S chosen in Theorem 2.

Proof. For each $k = 1, 2, \dots, n-1, n$, we define w_k via the formula

$$w_k = \int g_k(x)\, w(x)\, dx, \tag{19}$$

where $g_1, g_2, \dots, g_{n-1}, g_n$ are defined in (2). Then, (17) is an immediate consequence of (19) and (8). Moreover, (18) is an immediate consequence of (9) and (19). $\square$

Remark 7 Needless to say, the weight function w in the above theorem is superfluous; it could be absorbed into the measure on S. However, we found the formulations of Theorems 6 and 10 involving w to be convenient in applications. While Theorems 6 and 10 require the functions $f_1, f_2, \dots, f_{n-1}, f_n$ to be bounded (perhaps after "normalizing" them as in Remark 4 or otherwise rescaling them), these theorems do not require the weight function w to be bounded.

Remark 8 Theorem 6 asserts the existence under very mild conditions of numerically stable quadratures that integrate linear combinations of n functions using the values of these linear combinations tabulated at n appropriately chosen points. In contrast, construction of optimal "generalized Gaussian" quadratures, which tabulate the linear combinations at fewer nodes than the number of functions, requires more subtle techniques (see, for example, the references cited in [5]).

Remark 9 The proof of Theorem 2 does not specify a computational means for choosing the points $x_1, x_2, \dots, x_{n-1}, x_n$ so that (11) is satisfied (that is, so that the resulting interpolation and quadrature schemes are guaranteed to be numerically stable). However, combining the algorithms described in [1], [3] with appropriate discretizations of S yields methods for choosing the points that are proven to work, both in theory and in practice.

Theorem 2 provides the bound (8) under the rather weak assumption that the functions $f_1, f_2, \dots, f_{n-1}, f_n$ are bounded (in fact, this assumption is necessary for (8)). The following theorem provides a stronger bound under the additional assumption that there exists a measure with respect to which the functions $f_1, f_2, \dots, f_{n-1}, f_n$ are orthonormal.

Theorem 10 *Suppose that n is a positive integer, S is a measure space containing at least n points, w is a nonnegative real-valued integrable function on S (that serves as the weight for integration), $f_1, f_2, \dots, f_{n-1}, f_n$ are bounded complex-valued square-integrable functions on S, and $\varepsilon \le 1$ is a positive real number. Suppose further that $f_1, f_2, \dots, f_{n-1}, f_n$ are orthonormal, that is,*

$$\int |f_k(x)|^2\, w(x)\, dx = 1 \tag{20}$$

for all $k = 1, 2, \dots, n-1, n$, and

$$\int \overline{f_j(x)}\, f_k(x)\, w(x)\, dx = 0 \tag{21}$$

whenever $j \neq k$.

Then,

$$|g_k(x)| \le (1+\varepsilon)\sqrt{\int w(y)\, dy}\, \sum_{j=1}^{n} |f_j(x)| \tag{22}$$

for all x in S and $k = 1, 2, \dots, n-1, n$, where $g_1, g_2, \dots, g_{n-1}, g_n$ are defined in (2), with the n points $x_1, x_2, \dots, x_{n-1}, x_n$ in S chosen in Theorem 2.

Proof. In order to prove (22), for each $k = 1, 2, \dots, n-1, n$, we define the function h_k on S via the formula

$$h_k(x) = \sum_{j=1}^{n} f_j(x) \int \overline{f_j(y)}\, g_k(y)\, w(y)\, dy \tag{23}$$

and demonstrate that

$$|h_k(x)| \le (1+\varepsilon)\sqrt{\int w(y)\, dy}\, \sum_{j=1}^{n} |f_j(x)| \tag{24}$$

and

$$h_k(x) = g_k(x) \tag{25}$$

for all x in S and $k = 1, 2, \ldots, n-1, n$.

Suppose that f is defined via (10). We first show that

$$f(x) = \sum_{k=1}^{n} f(x_k)\, h_k(x) \tag{26}$$

for all x in S. Substituting (23) into the right hand side of (26) and exchanging the orders of summation and integration,

$$\sum_{k=1}^{n} f(x_k)\, h_k(x) = \sum_{j=1}^{n} f_j(x) \int \overline{f_j(y)} \sum_{k=1}^{n} f(x_k)\, g_k(y)\, w(y)\, dy \tag{27}$$

for all x in S. Due to the combination of (27) and (9),

$$\sum_{k=1}^{n} f(x_k)\, h_k(x) = \sum_{j=1}^{n} f_j(x) \int \overline{f_j(y)}\, f(y)\, w(y)\, dy \tag{28}$$

for all x in S. Then, (26) is an immediate consequence of applying (10), (20), and (21) to the right hand side of (28).

We now complete the proof of (25). For any x in S, due to (26),

$$At = u, \tag{29}$$

where $A = A(x_1, x_2, \ldots, x_{n-1}, x_n)$ is defined in (1), $u = u(x)$ is defined in (5), and $t = t(x)$ is defined to be an $n \times 1$ column vector via the formula

$$t_k = h_k(x) \tag{30}$$

with $k = 1, 2, \ldots, n-1, n$; subtracting (12) from (29),

$$A(t - v) = 0, \tag{31}$$

where $v = v(x)$ is defined in (6). Since $f_1, f_2, \ldots, f_{n-1}, f_n$ are orthonormal, they are linearly independent. Thus, due to Lemma 1, B defined in (4) is strictly positive, so that A defined in (1) is invertible, and therefore, due to (31),

$$t(x) = v(x) \tag{32}$$

for all x in S. Then, (25) is an immediate consequence of (32), (30), (6).

Finally, due to the Cauchy-Schwarz inequality,

$$\left| \int \overline{f_k(y)}\, g_k(y)\, w(y)\, dy \right| \le \sqrt{\int |f_k(y)|^2\, w(y)\, dy} \sqrt{\int |g_k(y)|^2\, w(y)\, dy}, \tag{33}$$

and due to (8),

$$\sqrt{\int |g_k(y)|^2\, w(y)\, dy} \le (1 + \varepsilon) \sqrt{\int w(y)\, dy} \tag{34}$$

for all $k = 1, 2, \ldots, n-1, n$. Then, (24) is an immediate consequence of (23), (33), (20), (34), and then (22) is an immediate consequence of (24) and (25). □

Remark 11 Due to (22), the interpolation formula (9) is numerically stable. While the numerical stability guaranteed by (8) is sufficient under most conditions, sometimes the bound (22) is more useful.

Remark 12 Theorem 10 generalizes easily to the case when the functions $f_1, f_2, \ldots, f_{n-1}, f_n$ are not precisely orthonormal, but only "close" to orthonormal, in the sense that the condition number of their Gram matrix is reasonably small.

Remark 13 One often encounters infinite-dimensional spaces of functions that are finite-dimensional to any specified precision. A typical situation of this kind involves the range of a compact operator, and the usual way to construct the finite-dimensional approximation is via the Singular Value Decomposition (see, for example, [5]). When combined with this observation, the apparatus of the present note becomes applicable to many infinite-dimensional spaces of functions.

References

[1] H. Cheng, Z. Gimbutas, P.-G. Martinsson, and V. Rokhlin, *On the compression of low rank matrices*, to appear in SIAM J. Sci. Comput., (2005).

[2] G. Dahlquist and A. Bjorck, *Numerical Methods*, Dover Publications, Mineola, New York, 1974.

[3] M. Gu and S. C. Eisenstat, *Efficient algorithms for computing a strong rank-revealing QR factorization*, SIAM J. Sci. Comput., 17 (1996), pp. 848–869.

[4] R. Suda, *Stability analysis of the fast Legendre transform algorithm based on the fast multipole method*, Proc. Estonian Acad. Sci. Phys. Math., 53 (2004), pp. 107–115.

[5] N. Yarvin and V. Rokhlin, *Generalized Gaussian quadratures and singular value decompositions of integral operators*, SIAM J. Sci. Comput., 20 (1998), pp. 699–718.