JOURNAL OF APPROXIMATION THEORY 22, 360–367 (1978)

A Further Optimal Property of Natural Polynomial Splines

F. M. LARKIN

Department of Computing and Information Science, Queen's University, Kingston, Ontario, Canada

Communicated by Oved Shisha

Received October 1, 1976

A minimum mean-square-error principle is used to define optimality of rules appropriate for the estimation of linear functionals of certain nonnegative functions. It is shown that, just as in the usual linear estimation problem where nonnegativity is not a constraint, limiting forms of these optimal rules are "best" in the sense of Sard.

1. INTRODUCTION

Let $\{x_k\,; k = 1, 2,..., m\}$ be an ordered set of distinct, real abscissas, $f(\cdot)$, a member of a function space F, to be defined more precisely later, and L, a linear functional. We shall be concerned with estimation rules of the form

$$\sum_{k=1}^{m} w_k f(x_k) \simeq Lf \tag{1}$$

and

$$\sum_{k=1}^{m} v_k f^2(x_k) \simeq Lf^2, \tag{2}$$

particularly with a view to selecting sets of weights $\{w_k\,; k = 1, 2,..., m\}$ and $\{v_k\,; k = 1, 2,..., m\}$ so as to optimize their accuracy (in a sense to be defined) over F.

Rules of type (1) are familiar in the context of numerical interpolation, quadrature, etc. Consideration of rules of type (2), which have been briefly discussed in earlier papers (Larkin [4–6]), is prompted by the fact that nonnegativity is an inherent property of many functions encountered in the experimental sciences (e.g., mass, heat, or probability densities). Clearly, the construction of rules of type (2) is one way of making use of the extra, global information of nonnegativity of the subject function in a linear estimation problem.

0021-9045/78/0224–0360$02.00/0

The error functional E in a type (1) rule is defined by

$$Ef = Lf - \sum_{k=1}^{m} w_k f(x_k); \qquad \forall f \in F.$$

Suppose $s(\cdot)$ is a seminorm or norm on F, with respect to which E is a bounded, linear functional, i.e.,

$$\sup_{0 \neq f \in F} \frac{|Ef|}{s(f)} \stackrel{\text{def}}{=} \| E \| < \infty,$$

then $\| E \|$ is a function of the m weights $\{w_k\}$. If these weights are chosen so that

(i) $s(f) = 0 \Rightarrow Ef = 0$,

(ii) $\| E \|$ is minimized with respect to the remaining degrees of freedom in the $\{w_k\}$ then they are said to be "optimal" with respect to $s(\cdot)$, and rule (1) is said to be an "optimal, linear estimation rule." A rule which is optimal with respect to the norm in a space will be called "norm-optimal," to distinguish this from optimality with respect to a seminorm. In particular, let F be H_n, the Hilbert space of real functions on the interval [0, 1] having absolutely continuous $(n - 1)$th order derivative, with inner product defined by

$$(f, g) = \sum_{j=0}^{n-1} \alpha_j f^{(j)}(0)\, g^{(j)}(0) + \int_0^1 f^{(n)}(x)\, g^{(n)}(x)\, dx; \qquad \forall f, g \in H_n,$$

where the $\{\alpha_j\,; j = 0, 1, 2,..., n - 1\}$ are positive, real parameters. We presume that $1 \leqslant n < m$. In this space E has a Riesz representer $e(\cdot)$, say. A rule of type (1) which is exact for all polynomials of degree less than n, and minimizes the seminorm

$$s(e) = \left\{ \int_0^1 | e^{(n)}(x)|^2\, dx \right\}^{1/2}$$

with respect to the remaining degrees of freedom in the $\{w_k\}$, is said to be "best in the sense of Sard" (Sard [8]). Clearly such a rule is optimal.

It is well known (e.g., Handscomb [3]) that if (1) is "best" in the sense of Sard, then

$$\sum_{k=1}^{m} w_k f(x_k) = L\hat{f}; \qquad \forall f \in H_n,$$

where $\hat{f}(\cdot)$ is the natural, polynominal spline of order $2n - 1$ based on the knots $\{(x_k, f(x_k)); k = 1, 2,..., m\}$ (Ahlberg *et al.* [1]). Thus, optimal estimation of the value of a bounded, linear functional, from a set of given ordinate values, may be achieved by constructing an appropriate, natural, polynominal spline and applying the required functional to it. Furthermore, (1) is "best" in the sense of Sard if and only if it is exact for all natural, polynomial splines of degree $2n - 1$ with $\{x_k ; k = 1, 2,..., m\}$ as their knot abscissas.

The main result of this paper is that, on extending the concept of optimality in an intuitively reasonable way, so as to include estimation rules of type (2), it turns out that for the spaces $\{H_n, n = 1, 2, 3,...\}$ limiting forms of the optimal type (2) rules are identical with those of type (1), i.e., are "best" in the sense of Sard. To arrive at this result we

(a) Show how "best" linear estimation rules arise as limiting forms of rules which are optimal with respect to the norm in H_n, by exhibiting a characterizing basis of functions for which such rules are exact.

(b) Note that norm-optimal type (1) rules in a Hilbert space can be derived by minimizing the mean-square error over the space (integration being performed relative to a weak Gaussian distribution) with respect to the weights $\{w_k\}$. We then define optimality of a type (2) rule in terms of minimizing its mean-square error over the space.

(c) Exhibit a characterizing basis of functions for which a norm-optimal type (2) rule is exact and show that its limiting form is identical with the corresponding basis for a type (1) rule.

2. Characterization of "Best" Linear Rules

In an earlier paper (Larkin [5]) a more general form of the following result was proved:

Theorem 1. *If $K_n(\cdot, \cdot)$ is the reproducing kernel function for H_n and the distinct abscissa $\{x_k ; k = 1, 2,..., m\}$ all lie in* [0, 1], *then the norm-optimal estimation rule of type* (1) *is characterized by the fact that it treats the functions $\{K_n(x_k, \cdot); k = 1, 2,..., m\}$ exactly, for any bounded linear functional L on H_n.*

We now construct $K_n(x, y)$.
Let

$$\begin{aligned} J_n(x, y) &= (y - x)^{n-1}/(n - 1)!; \qquad 0 \leqslant x \leqslant y, \\ &= 0; \qquad\qquad\qquad\qquad\quad 0 \leqslant y \leqslant x, \end{aligned}$$

and define $\tilde{K}_n(x, y)$ to be the nth iterated integral with respect to x, from 0 as the lower limit, of $J_n(x, y)$. It may then be verified that

$$K_n(x, y) \overset{\text{def}}{=} \sum_{j=0}^{n-1} \frac{(xy)^j}{\alpha_j (j!)^2} + \tilde{K}_n(x, y); \qquad \forall x, y \in [0, 1],$$

regarded as a function of either variable with the other fixed, is the reproducing kernel function for H_n (Aronszajn [2]). Thus

$$K_n(\cdot, y) \in H_n ; \qquad \forall y \in [0, 1],$$
$$K_n(x, y) = K_n(y, x); \qquad \forall x, y \in [0, 1],$$

and

$$(h(\cdot), K_n(\cdot, y)) = h(y); \qquad \forall h \in H_n ; \qquad \forall y \in [0, 1].$$

For example

$$K_1(x, y) = 1/\alpha_0 + \begin{cases} x; & 0 \leqslant x \leqslant y, \\ y; & 0 \leqslant y \leqslant x, \end{cases}$$

$$K_2(x, y) = 1/\alpha_0 + xy/\alpha_1 + \begin{cases} -x^3/3! + x^2y/2!; & 0 \leqslant x \leqslant y, \\ -y^3/3! + y^2x/2!; & 0 \leqslant y \leqslant x, \end{cases}$$

$$K_3(x, y) = 1/\alpha_0 + xy/\alpha_1 + (xy)^2/4\alpha_2$$
$$+ \begin{cases} x^5/5! - x^4y/4! + x^3y^2/3!\,2!; & 0 \leqslant x \leqslant y, \\ y^5/5! - y^4x/4! + y^3x/3!\,2!; & 0 \leqslant y \leqslant x, \end{cases}$$

etc.

Now let $K_n[x_s x_{s+1} \cdots x_{s+n}, y]$ denote the nth divided difference formed from $\{(x_k, K_n(x_k, y)); k = s, s+1,..., s+n\}$, and similarly for $\tilde{K}_n[x_s x_{s+1} \cdots x_{s+n}, y]$.

THEOREM 2. *In the limit as the parameters* $\{\alpha_j ; j = 0, 1, 2,..., n-1\}$ *sequentially approach* 0 *from above, the norm-optimal type* (1) *estimation rule becomes exact for the m functions of y* $\{y^r; r = 0, 1, 2,..., n-1\}$ *and* $\{\tilde{K}[x_s x_{s+1} \cdots x_{s+n}, y]; s = 1, 2,..., m-n\}$

Proof. Form a divided difference table (whose entries will be functions of y) from the table $\{(x_k, K_n(x_k, y)); k = 1, 2,..., m\}$. From *Theorem* 1 it is clear that if (1) is norm-optimal it is exact for every function in the divided difference table. Notice that, because of the polynomial nature of the function $K_n(x, y) - \tilde{K}_n(x, y)$ the column of first differences contains no term in $1/\alpha_0$, the column of second differences contains no term in $1/\alpha_0$ or y/α_1, and so on until the column of nth differences is quite independent of the $\{\alpha_j\}$.

As $\alpha_0 \searrow 0$ the initial column of functions becomes dominated by the term

$1/\alpha_0$, so the limiting form of the norm-optimal rule will be exact for constants. As $\alpha_1 \searrow 0$ the column of first differences becomes dominated by the term y/α_1 so the limiting rule will be exact for linear functions. Similarly, as $\alpha_2, \alpha_3, ..., \alpha_{n-1}$ sequentially approach 0 from above the limiting form of the norm-optimal rule will be exact for all polynomials of degree $<n$, as well as for the functions of y

$$K_n[x_s x_{s+1} \dots x_{s+n}, y] = \tilde{K}_n[x_s x_{s+1} \dots x_{s+n}, y]; \qquad s = 1, 2, ..., m - n,$$

as required.

COROLLARY. *In the limit as the parameters* $\{\alpha_j ; j = 0, 1, 2, ..., n - 1\}$ *approach* 0 *from above, a norm-optimal type* (1) *rule becomes "best" in the sense of Sard.*

Proof. Notice that for $0 \leqslant y \leqslant x_1$ and $x_m \leqslant y$ the functions $\{\tilde{K}_n[x_s x_{s+1} \dots x_{s+n}, y]; s = 1, 2, ..., m - n\}$ are polynomials in y of degree $<n$. Hence these functions, together with $\{x^r; r = 0, 1, 2, ..., n - 1\}$ comprise a Chebyshev set of natural, polynomial splines of order $2n - 1$ whose knot abscissas are located at $\{x_k ; k = 1, 2, ..., m\}$. Thus, since the unique, natural, polynomial spline of order $2n - 1$ with knots at $\{(x_k, f(x_k)); k = 1, 2, ..., m\}$ can be expanded as a linear combination of these basis functions, for any set of finite ordinate values $\{f(x_k); k = 1, 2, ..., m\}$, the limiting form of the norm-optimal type (1) rule must be exact for this class of function, so the required result is proved.

3. Estimation Rules With Minimum Mean-Square Error

In another paper (Larkin [6]) the idea was introduced of constructing an estimation rule by minimizing the value of its squared error, averaged with respect to a weak Gaussian distribution over a Hilbert space H. In particular, it was shown that choosing the weights $\{w_k ; k = 1, 2, ..., m\}$ to minimize the mean-square error

$$\int_H \left| Lh - \sum_{k=1}^{m} w_k h(x_k) \right|^2 \mu(dh),$$

$\mu(\cdot)$ denoting the weak Guassian distribution on H, leads to a norm-optimal, linear estimation rule of type (1). Furthermore, the weights $\{v_k ; k = 1, 2, ..., m\}$ which minimize the mean-square error

$$\int_H \left| Lh^2 - \sum_{k=1}^{m} v_k h^2(x_k) \right|^2 \mu(dh)$$

satisfy the simultaneous, linear equations $C\mathbf{v} = \mathbf{d}$, where

$$C_{jk} = \int_H h^2(x_j)\, h^2(x_k)\, \mu(dh); \qquad j, k = 1, 2,..., m,$$

and

$$d_j = \int_H h^2(x_j)\, Lh^2 \mu(dh); \qquad j = 1, 2,..., m.$$

These functional integrals can be evaluated explicitly in terms of the reproducing kernel function $K(x, y)$ for H, and are proportional to $K(x_j, x_j)\, K(x_k, x_k) + 2K^2(x_j, x_k)$ and $L_x\{K(x_j, x_j)\, K(x, x) + 2K^2(x_j, x)\}$, respectively, the suffix on L indicating the independent variable through which it operates.

Thus, the weights $\{v_k ; k = 1, 2,..., m\}$ arising from this extended optimality principle satisfy the equations

$$\sum_{k=1}^{m} v_k\{K(x_j, x_j)\, K(x_k, x_k) + 2K^2(x_j, x_k)\}$$
$$= L_x\{K(x_j, x_j)\, K(x, x) + 2K^2(x_j, x)\}; \qquad j = 1, 2,..., m,$$

which is to say that the optimal type (2) rule is exact for the functions of x $\{K(x_j, x_j)\, K(x, x) + 2K^2(x_j, x); j = 1, 2,..., m\}$.

Using Δ to denote the familiar finite difference operator, we are now in a position to prove the following:

THEOREM 3. *If* $K(x, y) = 1/\alpha + \tilde{K}(x, y)$, *where* $\tilde{K}(x, y)$ *is independent of* α, *in the limit as* $\alpha \searrow 0$ *the optimal type* (2) *rule will be exact for constants, and the functions* $\{\Delta\tilde{K}(x_j, \cdot); j = 1, 2,..., m - 1\}$.

Proof. The optimal type (2) rule is exact for the functions of x

$$S_j(x) = 3/\alpha^2 + 1/\alpha\{\tilde{K}(x_j, x_j) + \tilde{K}(x, x) + 4\tilde{K}(x_j, x)\}$$
$$+ \tilde{K}(x_j, x_j)\, \tilde{K}(x, x) + 2\tilde{K}^2(x_j, x); \qquad j = 1, 2,..., m,$$

and any linear combination of them. As $\alpha \searrow 0$ the $\{S_j(\cdot)\}$ are dominated by the term in $3/\alpha^2$, so the limiting form of the rule is exact for constants.

However, the rule is also exact for the functions of x

$$(1/\alpha)\, \Delta\{\tilde{K}(x_j, x_j) + 4\tilde{K}(x_j, x)\} + \Delta\{\tilde{K}(x_j, x_j)\, \tilde{K}(x, x) + 2\tilde{K}^2(x_j, x)\};$$
$$j = 1, 2,..., m - 1,$$

and the $\{\Delta\tilde{K}(x_j, x_j); j = 1, 2,..., m - 1\}$ are constants, so its limiting form must also be exact for the functions $\{\Delta\tilde{K}(x_j, \cdot); j = 1, 2,..., m - 1\}$, as required.

COROLLARY. *When H is H_n, in the limit as the parameters $\{\alpha_j\,; j = 0, 1, 2,..., n-1\}$ sequentially approach 0 from above the optimal type (2) rule becomes identical with the type (1) rule which is "best" in the sense of Sard. That is*

$$v_k = w_k\,; \qquad k = 1, 2,..., m.$$

Proof To verify exactness for constants we merely identify α_0 with the α in *Theorem* 3. The rest of the proof follows the technique of *Theorem* 2—forming higher-order divided differences of the quantities $\{\Delta K_n(x_j\,, \cdot)/(x_{j+1} - x_j);\ j = 1, 2,..., m-1\}$ and sequentially permitting the parameters $\{\alpha_j\,; j = 1, 2,..., n-1\}$ to approach 0 from above.

4. Conclusions

We have shown how type (1) linear estimation rules which are "best" in the sense of Sard may be regarded as limiting forms of norm-optimal rules in a certain Hilbert space, and have noted that norm-optimality is obtained by minimizing the squared error of the rule, averaged with respect to a weak Gaussian distribution on the space. The analog of a norm-optimal rule was constructed for the type (2) problem, and it was shown that the limiting form of this rule is also "best" in the sense of Sard for the type (1) problem.

Thus, it appears that if operating on the natural, polynomial spline is appropriate for estimation of the value of a bounded, linear functional in a type (1) problem, it is also appropriate in the related type (2) problem. In other words, it is reasonable simply to ignore the positivity constraint. Unfortunately, this does not dispose of the objection that a natural, polynomial spline interpolated through the knots $\{(x_k\,, h^2(x_k)); k = 1, 2,..., m\}$ may cross the x-axis. One can only conclude that the requirement of positivity of an interpolating function $\hat{f}$, from which an estimate of Lf is to be computed, may be incompatible with the requirement for minimum mean-square error over H_n.

Acknowledgment

Thanks are due to John Driscoll, a graduate student in the Department of Computing and Information Science at Queen's University, for verifying the above results in two special cases.

References

1. J. H. Ahlberg, E. N. Nilson, and J. L. Walsh, "The Theory of Splines and Their Applications," Academic Press, New York, 1967.
2. N. Aronszajn, Theory of reproducing kernels, *Trans. Amer. Math. Soc.* **68** (1950), 337–404.

3. D. C. HANDSCOMB (Ed.), "Methods of numerical approximation," Pergamon, New York, 1966.
4. F. M. LARKIN, Estimation of a non-negative function, *Nordisk Tidskr. Informations-Behandling* (*BIT*) **9** (1969), 30–52.
5. F. M. LARKIN, Optimal approximation in Hilbert spaces with reproducing kernel functions, *Math. Comp.* **24** (1970), 911–921.
6. F. M. LARKIN, Gaussian measure in Hilbert space, and applications in numerical analysis, *Rocky Mt. J. Math.* **2** (1972), 379–421.
7. A. G. LAW AND B. N. SAHNEY (Eds.), Some remarks on the estimation of quadratic functionals, *in* "Theory of Approximation with Applications, pp. 43–63, Academic Press, New York, 1976.
8. A. SARD, "Linear Approximation," Math. Surveys No. 9, *Amer. Math. Soc.*, Providence, R.I., 1963.