JOURNAL OF APPROXIMATION THEORY 59, 267–275 (1989)

# Moving Least-Squares Are Backus–Gilbert Optimal

L. P. Bos and K. Salkauskas

*Department of Mathematics and Statistics, University of Calgary, Calgary, Alberta, Canada T2N 1N4*

*Communicated by E. W. Cheney*

Received March 2, 1987; revised February 2, 1988

## 1. Introduction

Moving least-squares methods for the interpolation of scattered data in the plane are well known [6]. The simplest of them is known more commonly as Shepard's method [5, 7]. The value of the interpolant at any point is obtained from a weighted least-squares polynomial approximation to the data, the weighting of a data point being inversely related to its distance from the point at which the interpolant is being evalutated. In Shepard's method, the polynomial is a constant.

The Backus–Gilbert theory [2] has been developed in a geophysical context, but it is a theory of interpolation in that the values of a number of functionals on an unknown function $f$ are given, and an approximation to the value of $f$ at some point is required. The basic principle is to optimize the approximation of the Dirac delta by a linear combination of the representers of the given functionals.

Recently, Abramovici [1] showed that Shepard's method could be obtained from the Backus–Gilbert theory. Here we will demonstrate that all moving least-squares approximants can be generated from a slightly modified Backus–Gilbert theory.

## 2. Moving Least-Squares

Let $M := \binom{n+m}{n}$, i.e., the dimension of the space of polynomials of degree at most $m$ in $n$ variables, and $\{\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x}), ..., \varphi_M(\mathbf{x})\}$ be the monomials of degree at most $m$ in $n$ real variables. Suppose that we are given $N \geqslant M$ distinct points $\mathbf{x}_i \in \mathbb{R}^n$, not all of which lie on the zero set of a polynomial of degree at most $m$. We approximate $f: \mathbb{R}^n \to \mathbb{R}$, at $\mathbf{x}_0 \in \mathbb{R}^n$, by the weighted least-squares polynomial of degree $m$ at the points $\mathbf{x}_i$, $1 \leqslant i \leqslant N$, with weights $w_i := w(\mathbf{x}_i - \mathbf{x}_0)$. Here $w: \mathbb{R}^n \to \mathbb{R}_+$. Typically, $w(\mathbf{x}) = |\mathbf{x}|^{-k}$



0021-9045/89 $3.00

for some even $k$ although many other choices are possible. This particular choice ensures that the approximation is actually an interpolant and also has a certain smoothness (see Lancaster and Salkauskas [6]).

In order to obtain an explicit expression for this approximation, let $B^T \in \mathbb{R}^{N \times M}$ be the Vandermonde matrix of the monomials of degree at most $m$ evaluated at the points $\mathbf{x}_i$. Specifically,

$$B_{ij}^T = \varphi_j(\mathbf{x}_i). \tag{2.1}$$

Further, let $W = \text{diag}(w_1, w_2, ..., w_N) \in \mathbb{R}^{N \times N}$, $\mathbf{f} = (f(\mathbf{x}_1), ..., f(\mathbf{x}_N))^T \in \mathbb{R}^N$, and $\boldsymbol{\varphi} = (\varphi_1(\mathbf{x}_0), \varphi_2(\mathbf{x}_0), ..., \varphi_M(\mathbf{x}_0))^T \in \mathbb{R}^M$. Thus the moving least-squares approximation to $f(\mathbf{x}_0)$ is $\sum_{i=1}^M c_i \varphi_i(\mathbf{x}_0)$, where $\mathbf{c} \in \mathbb{R}^M$ minimizes

$$(B^T\mathbf{c} - \mathbf{f})^T W (B^T\mathbf{c} - \mathbf{f}).$$

Under our assumptions, $W$ is positive definite and $B^T$ is of full rank. Thus $\mathbf{c}$ is the solution of the normal equations

$$(BWB^T)\,\mathbf{c} = BW\mathbf{f}, \tag{2.2}$$

and the approximation is therefore given by

$$\sum_{i=1}^{M} c_i \varphi_i(\mathbf{x}_0) = \mathbf{c}^T \boldsymbol{\varphi} = \mathbf{f}^T W B^T (BWB^T)^{-1} \boldsymbol{\varphi}. \tag{2.3}$$

For further properties of moving least-squares approximations see, for instance, [4, 6].

## 3. Backus–Gilbert Optimality

We give a brief, univariate account of those features of the Backus–Gilbert theory that apply to the problem at hand. It will be clear how this can be extended to $\mathbb{R}^n$.

Let $\{\lambda_i\}_{i=1}^N$ be a linearly independent set of bounded linear functionals on $L_2[a, b]$, and for some (unknown) $f \in L_2[a, b]$ let the values of $\lambda_i(f)$, $1 \leqslant i \leqslant N$, be given. It is required to obtain an approximation $\tilde{f}(x_0)$, $x_0 \in [a, b]$, in terms of this information. To this end we seek coefficients $a_i(x_0)$ such that $\sum_{i=1}^N a_i(x_0)\, \lambda_i(f)$ will in some sense be a good approximation to $f(x_0)$. Let $\{L_i\}_{i=1}^N$ be the representers of the functionals $\lambda_i$. Then $W_0(x) := \sum_{i=1}^N a_i(x_0)\, L_i(x)$ is to be a good approximation to the Dirac distribution, $\delta(x - x_0)$. In order to measure the "deltaness" of $W_0$, Backus and Gilbert propose that a symmetric, non-negative "sink function,"

$J(x, x_0)$, vanishing only at $x = x_0$, be selected, and that the $a_i$'s be chosen to minimize the "spread" of $W_0$:

$$S(W_0; x_0) := \int_a^b J(x, x_0)\, W_0^2(x)\, dx. \tag{3.1}$$

If $J(x, x_0)$ is chosen so as to increase as $|x - x_0|$ increases, then a $W_0$ minimizing the spread will tend to have its relatively large values concentrated near $x_0$. The choice $J(x, x_0) = (x - x_0)^2$ is typical. Clearly, the spread is just the square of a weighted $L_2$ norm of $W_0$.

As well, since the approximation

$$\tilde{f}(x_0) = \int_a^b f(x)\, W_0(x)\, dx \tag{3.2}$$

can be seen as an average of $f$ over $[a, b]$, it is natural to impose the condition

$$\int_a^b W_0(x)\, dx = 1, \tag{3.3}$$

which certainly is also satisfied by $\delta(x - x_0)$.

Abramovici [1] has applied this to the case where the functionals $\lambda_i$ are $\delta(x - x_i)$, and hence do not have representers. However, by working with a certain $\delta$-sequence of functions, he has shown that the approximation $\tilde{f}(x_0)$ is identical to the Shepard interpolant. In the sequel, we apply this technique with a broad class of $\delta$-sequences. We do not require $J(x, x_0)$ to vanish when $x = x_0$. In addition, since $W_0$ is to approximate $\delta(x - x_0)$, we impose the conditions

$$\int_a^b x^i W_0(x)\, dx = x_0^i, \qquad 0 \leqslant i \leqslant m,\ m \leqslant N - 1,$$

which forces the approximation to be exact for polynomials of degree up to $m$. The resulting schemes are shown to be moving least-squares methods of approximation, and are interpolants if $J(x, x_0)$ is chosen appropriately.

## 4. Moving Least-Squares Are Backus–Gilbert Optimal

Suppose that $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N \in \mathbb{R}^n$ form a given set of points not all of which lie on the zero set of any polynomial of degree $m$. Let $D \subset \mathbb{R}^n$ be a compact, connected set which contains all of the $\mathbf{x}_i$ in its interior. Select a continuous $J$: $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$ such that $J(\mathbf{x}, \mathbf{y}) = 0 \Rightarrow \mathbf{x} = \mathbf{y}$. Let $L_2(D)$ be, as usual $\{f: D \to \mathbb{R} \mid \int_D f^2(\mathbf{x})\, d\mathbf{x} < \infty\}$, with the usual inner product $\langle f, g\rangle := \int_D f(\mathbf{x})\, g(\mathbf{x})\, dx$. Further, for fixed $\mathbf{x}_0 \in D$, let $L_2(D;\ J; x_0) = \{f: D \to \mathbb{R} \mid \int_D J(x, x_0) f^2(x)\, d\mathbf{x} < \infty\}$, with inner product $\langle f, g\rangle_J := \int_D J(\mathbf{x}, \mathbf{x}_0) f(\mathbf{x})\, g(\mathbf{x})\, d\mathbf{x}$, and associated norm $\|\cdot\|_J$. Now, as $J$ is continuous, there is an $M > 0$ such that $\|f\|_J \leqslant M\|f\|_2$, and hence $L_2(D) \subset L_2(D; J; \mathbf{x}_0)$. Keeping in mind the form of the approximation $\tilde{f}(\mathbf{x}_0)$ of Eq. (3.2), we now make

DEFINITION 4.1. For $\mathbf{x}_0 \in D$, $W_0 \in L_2(D; J; \mathbf{x}_0)$, and $f \in C(D)$, the $W_0$-approximation to $f(\mathbf{x}_0)$ is given by

$$\tilde{f}(\mathbf{x}_0) := \int_D f(\mathbf{x})\, W_0(\mathbf{x})\, d\mathbf{x} = \langle f, W_0\rangle.$$

Now we wish our approximaton to be exact for polynomials of a given degree.

DEFINITION 4.2. $W_0 \in L_2(D; J; \mathbf{x}_0)$ has degree of precision $m$ if, for all polynomials $p$, of degree at most $m$,

$$\langle p, W_0\rangle = p(\mathbf{x}_0).$$

We set $\mathscr{A}_m := \{W_0 \in L_2(D; J; \mathbf{x}_0) \mid W_0 \text{ has degree of precision } m\}$.

As $\mathscr{A}_m$ is the intersection of a finite number of hyperplanes, it is closed and convex in $L_2(D; J; \mathbf{x}_0)$.

The fragment of the Backus–Gilbert (abbreviated B-G in the sequel) theory described in Section 3, shows that $W_0$ is drawn from $\operatorname{span}\{L_i\}_{i=1}^N$, a finite-dimensional subspace of $L_2[a, b]$. We insist on the same property of $W_0$ in our next definition.

DEFINITION 4.3. Let $V \subset L_2(D)$ be a finite-dimensional subspace, and fix $m \in \mathbb{Z}_+$. The approximation $\tilde{f}(\mathbf{x}_0) = \langle f, W_0\rangle$ is B-G optimal of degree $m$ with respect to $V$ if $W_0 \in V \cap \mathscr{A}_m$ and $\|W_0\|_J$ is a minimum. (In the terminology of B-G, $W_0$ has minimal spread.)

Note that the existence and uniqueness of such a $W_0$ is guaranteed by the fact that $V \cap \mathscr{A}_m$ is closed, convex, and finite-dimensional.

If this approximation to $f(\mathbf{x}_0)$ is to be constructed in terms of information about $f$ consisting of the values of a finite (say $N$) number of bounded, linear functionals on $L_2(D)$, with Riesz representers $L_1, ..., L_N \in L_2(D)$, then $W_0 = \sum_{i=1}^N a_i L_i$, where the $a_i$'s satisfy linear constraints inherent in Definition 4.2, and the minimization of $\|W_0\|_J$ involves nothing more than the minimization of a quadratic form with linear constraints—a standard problem in linear algebra. Our aim is to examine the existence and nature of the approximation when the given information about $f$ consists of function values at distinct points of $D$. As Dirac delta functionals are unbounded, this does not fit directly into our earlier B-G formulation. However, as delta distributions are limits of ordinary functions we are able to extend Definition 4.3 to include the notion of generalized B-G optimality of degree $m$. For this we make use of certain delta sequences defined below.

DEFINITION 4.4. $\{\Delta_\lambda(\mathbf{x}) \mid \lambda > 0\}$ is said to be a delta sequence if $\lim_{\lambda \to \infty} \int_{\mathbb{R}^n} f(\mathbf{x})\, \Delta_\lambda(\mathbf{x})\, d\mathbf{x} = f(\mathbf{0})$ for any bounded $f$ continuous at $\mathbf{0}$. We will say that the delta sequence is *regular* if, in addition, $\Delta_\lambda \in L_2(\mathbb{R}^n)$ and

$$\lim_{\lambda \to \infty} \int_{\mathbb{R}^n} f(\mathbf{x}) \frac{\Delta_\lambda(\mathbf{x}-\mathbf{a})\, \Delta_\lambda(\mathbf{x}-\mathbf{b})}{\int_{\mathbb{R}^n} \Delta_\lambda^2(\mathbf{x})\, d\mathbf{x}}\, d\mathbf{x} = \begin{cases} 0 & \text{if } \mathbf{a} \neq \mathbf{b}, \\ f(\mathbf{a}) & \text{if } \mathbf{a} = \mathbf{b}, \end{cases}$$

for any bounded $f$ continuous at $\mathbf{a}$.

DEFINITION 4.5. Suppose that $\{\mathbf{x}_1, ..., \mathbf{x}_N\} \subset \operatorname{Int}(D)$ are distinct points. Let $\{\Delta_\lambda(\mathbf{x}) \mid \lambda > 0\}$ be a regular delta sequence, and set $V^{(\lambda)} = \operatorname{span}\{\Delta_\lambda(\mathbf{x}-\mathbf{x}_1), \Delta_\lambda(\mathbf{x}-\mathbf{x}_2), ..., \Delta_\lambda(\mathbf{x}-\mathbf{x}_N)\}$. Further, let $W_0^{(\lambda)} \in V^{(\lambda)} \cap \mathscr{A}_m$ be B-G optimal of degree $m$. If $\tilde{f}(\mathbf{x}_0) := \lim_{\lambda \to \infty} \int_D f(\mathbf{x})\, W_0^{(\lambda)}(\mathbf{x})\, d\mathbf{x}$ exists for all $f \in C(D)$ we will say that $\tilde{f}(\mathbf{x}_0)$ is the generalized B-G approximation of degree $m$ to $f(\mathbf{x}_0)$.

The use of regular delta sequences is not restrictive. In fact, the common constructions of delta sequences are regular.

PROPOSITION 4.6. *Suppose that $K \in L_1(\mathbb{R}^n)$ is bounded and is such that $\int_{\mathbb{R}^n} K(\mathbf{x})\, d\mathbf{x} = 1$. Then $\{\Delta_\lambda(\mathbf{x}) := \lambda^n K(\lambda \mathbf{x}) \mid \lambda > 0\}$ is a regular delta sequence.*

*Proof.* The fact that $\{\Delta_\lambda\}$ is a delta sequence is standard (see for instance [3, Sect. 3.2]). We must show that it is regular. Now as $K \in L_1(\mathbb{R}^n)$ is bounded, $K \in L_2(\mathbb{R}^n)$ and hence, an easy calculation shows that $\Delta_\lambda \in L_2(\mathbb{R}^n)$. Further, $\Delta_\lambda^2(\mathbf{x}) / \int_{\mathbb{R}^n} \Delta_\lambda^2(\mathbf{x})\, d\mathbf{x} = \lambda^n K^2(\lambda \mathbf{x}) / \int_{\mathbb{R}^n} K^2(\mathbf{x})\, d\mathbf{x}$ and so the same calculations that show that $\{\Delta_\lambda\}$ is a delta sequence show that $\{\Delta_\lambda^2 / \int_{\mathbb{R}^n} \Delta_\lambda^2\, dx\}$ is a delta sequence. Thus

$$\lim_{\lambda \to \infty} \int_{\mathbb{R}^n} f(\mathbf{x}) \frac{\Delta_\lambda(\mathbf{x}-\mathbf{a})\, \Delta_\lambda(\mathbf{x}-\mathbf{b})}{\int_{\mathbb{R}^n} \Delta_\lambda^2(\mathbf{x})\, d\mathbf{x}}\, d\mathbf{x} = f(\mathbf{a}) \qquad \text{if } \mathbf{b} = \mathbf{a}.$$

Now if $\mathbf{b} \neq \mathbf{a}$, as we assume $f$ to be bounded, it suffices to show that

$$\lim_{\lambda \to \infty} \int_{\mathbb{R}^n} \frac{|\Delta_\lambda(\mathbf{x}-\mathbf{a})\, \Delta_\lambda(\mathbf{x}-\mathbf{b})|}{\int_{\mathbb{R}^n} \Delta_\lambda^2(\mathbf{x})\, d\mathbf{x}}\, d\mathbf{x} = 0$$

which in turn is implied by

$$\lim_{\lambda \to \infty} \int_{\mathbb{R}^n} \lambda^n |K(\lambda(\mathbf{x}-\mathbf{a}))\, K(\lambda(\mathbf{x}-\mathbf{b}))|\, d\mathbf{x} = 0.$$

To see this let $\varepsilon > 0$ be given. As $K$ is bounded, there is a constant $M > 0$ such that $|K(\mathbf{x})| \leqslant M$. Choose $R$ sufficiently large so that $\int_{|\mathbf{u}| \geqslant R} |K(\mathbf{u})|\, d\mathbf{u} \leqslant \varepsilon/(2M)$. Then, upon letting $\mathbf{u} = \lambda(\mathbf{x}-\mathbf{a})$,

$$\begin{aligned}
&\int_{\mathbb{R}^n} \lambda^n |K(\lambda(\mathbf{x}-\mathbf{a}))\, K(\lambda(\mathbf{x}-\mathbf{b}))|\, d\mathbf{x} \\
&\quad = \int_{\mathbb{R}^n} |K(\mathbf{u})| \cdot |K(\mathbf{u}+\lambda(\mathbf{a}-\mathbf{b}))|\, d\mathbf{u} \\
&\quad = \left( \int_{|\mathbf{u}| \leqslant R} + \int_{|\mathbf{u}| \geqslant R} \right) |K(\mathbf{u})| \cdot |K(\mathbf{u}+\lambda(\mathbf{a}-\mathbf{b}))|\, d\mathbf{u} \\
&\quad \leqslant \{\varepsilon/(2M)\}\, M + \int_{|\mathbf{u}| \leqslant R} |K(\mathbf{u})| \cdot |K(\mathbf{u}+\lambda(\mathbf{a}-\mathbf{b}))|\, d\mathbf{u} \\
&\quad \leqslant \varepsilon/2 + \left\{ \int_{|\mathbf{u}| \leqslant R} K^2(\mathbf{u})\, d\mathbf{u} \right\}^{1/2} \left\{ \int_{|\mathbf{u}| \leqslant R} K^2(\mathbf{u}+\lambda(\mathbf{a}-\mathbf{b}))\, d\mathbf{u} \right\}^{1/2}.
\end{aligned}$$

Now select $L$ so large that for $|\mathbf{u}| \geqslant L$, $\int_{|\mathbf{u}| \geqslant L} K^2(\mathbf{u})\, d\mathbf{u} < \varepsilon^2/4 \int_{\mathbb{R}^n} K^2(\mathbf{x})\, d\mathbf{x}$. Then for $\lambda$ sufficiently large, $|\mathbf{u}+\lambda(\mathbf{a}-\mathbf{b})| \geqslant L$ for all $|\mathbf{u}| \leqslant R$ and so $\{\int_{|\mathbf{u}| \leqslant R} K^2(\mathbf{u}+\lambda(\mathbf{a}-\mathbf{b}))\, d\mathbf{u}\}^{1/2} \leqslant \{\int_{|\mathbf{u}| \geqslant L} K^2(\mathbf{u})\ d\mathbf{u}\}^{1/2} < \varepsilon/(2\|K\|_2)$. The result now follows. ∎

We are now ready to state and prove our main result.

THEOREM 4.7. *Moving least-squares approximations by polynomials of degree m are generalized B-G optimal of degree m.*

*Proof.* We make use of the notation of Section 2. Recall that the B-G approach is to approximate $f(\mathbf{x}_0)$ by

$$\tilde{f}(\mathbf{x}_0) := \int_D f(\mathbf{x})\, W_0^{(\lambda)}(\mathbf{x})\, d\mathbf{x},$$

where $W_0^{(\lambda)}(\mathbf{x}) := \sum_{i=1}^N a_i^{(\lambda)}\, \Delta_\lambda(\mathbf{x}-\mathbf{x}_i)$. The coefficients, $a_i^{(\lambda)}$, are chosen so that the spread, $\| W_0^{(\lambda)} \|_J$, is a minimum. We have added the requirement that it reproduces all the polynomials of degree at most $m$, i.e.,

$$\int_D \varphi_i(\mathbf{x})\, W_0^{(\lambda)}(\mathbf{x})\, d\mathbf{x} = \varphi_i(\mathbf{x}_0), \qquad 1 \leqslant i \leqslant M. \tag{4.1}$$

Now,

$$\int_D J(\mathbf{x}, \mathbf{x}_0)[W_0^{(\lambda)}(\mathbf{x})]^2\, d\mathbf{x} = [\mathbf{a}^{(\lambda)}]^T A(\lambda)\mathbf{a}^{(\lambda)},$$

where

$$A_{ij}^{(\lambda)} = \int_D J(\mathbf{x}, \mathbf{x}_0)\, \Delta_\lambda(\mathbf{x} - \mathbf{x}_i)\, \Delta_\lambda(\mathbf{x} - \mathbf{x}_j)\, d\mathbf{x}, \qquad 1 \leqslant i, j \leqslant N,$$

and $\mathbf{a}^{(\lambda)} \in \mathbb{R}^N$ is the coefficient vector. Note that as $A^{(\lambda)}$ is a matrix of inner products it is non-negative definite for all $\lambda > 0$. Because of our assumption that $\{\Delta_\lambda\}$ is regular, we may assume without loss of generality that $A^{(\lambda)}$ is strictly positive definite for all $\lambda > 0$. Further, the constraints, (4.1), may be expressed in matrix form as

$$B^{(\lambda)}\mathbf{a}^{(\lambda)} = \boldsymbol{\varphi},$$

where $B^{(\lambda)} \in \mathbb{R}^{M \times N}$ and $B_{ij}^{(\lambda)} = \int_D \varphi_i(\mathbf{x}) \Delta_\lambda(\mathbf{x} - \mathbf{x}_j)\, d\mathbf{x}$.

The approximation is then

$$\tilde{f}_\lambda(\mathbf{x}_0) := \sum_{i=1}^{N} a_i^{(\lambda)} f(\mathbf{x}_i) = \mathbf{f}^T \mathbf{a}^{(\lambda)},$$

where $\mathbf{a}^{(\lambda)} \in \mathbb{R}^N$ minimizes $[\mathbf{a}^{(\lambda)}]^T A^{(\lambda)}\mathbf{a}^{(\lambda)}$ subject to the constraint $B^{(\lambda)}\mathbf{a}^{(\lambda)} = \boldsymbol{\varphi}$. Note that $B_{ij}^{(\lambda)} = \int_D \varphi_i(\mathbf{x})\, \Delta_\lambda(\mathbf{x} - \mathbf{x}_j)\, d\mathbf{x} = \int_{\mathbb{R}^n} (\chi_D(\mathbf{x})\, \varphi_i(\mathbf{x}))\, \Delta_\lambda(\mathbf{x} - \mathbf{x}_j)\, d\mathbf{x} \to \varphi_i(\mathbf{x}_j)$ as $\lambda \to \infty$. Hence

$$\lim_{\lambda \to \infty} B^{(\lambda)} = B \qquad (\text{of } (2.1)), \tag{4.2}$$

which by our assumptions is of full rank. Thus we may, without loss of generality, assume that $B^{(\lambda)}$ is of full rank. We now have a standard problem in linear algebra, whose solution is given by

LEMMA 4.8. *The* $\mathbf{a} \in \mathbb{R}^N$ *for which* $B^{(\lambda)}\mathbf{a} = \boldsymbol{\varphi}$ *and* $\mathbf{a}^T A^{(\lambda)}\mathbf{a}$ *is a minimum is given by*

$$\mathbf{a} = A^{-1}B^T(BA^{-1}B^T)^{-1}\boldsymbol{\varphi}. \tag{4.3}$$

For convenience we have suppressed the $\lambda$ superscript.

*Proof.* As $A^{-1}$ is also positive definite and $B$ is of full rank, an easy argument shows that $BA^{-1}B^T$ is non-singular. Now suppose that $\mathbf{x} \in \mathbb{R}^N$ is such that $B\mathbf{x} = \boldsymbol{\varphi}$. Then $\mathbf{x}^T A\mathbf{x} = (\mathbf{x} - \mathbf{a})^T A(\mathbf{x} - \mathbf{a}) + 2\mathbf{a}^T A(\mathbf{x} - \mathbf{a}) + \mathbf{a}^T A\mathbf{a}$.

But

$$\begin{aligned}\mathbf{a}^T A(\mathbf{x}-\mathbf{a}) &= \boldsymbol{\varphi}^T (BA^{-1}B^T)^{-1} BA^{-1}A(\mathbf{x} - A^{-1}B^T(BA^{-1}B^T)^{-1}\boldsymbol{\varphi}) \\ &= \boldsymbol{\varphi}^T(BA^{-1}B^T)^{-1}(B\mathbf{x}-\boldsymbol{\varphi}) \\ &= \boldsymbol{\varphi}^T\mathbf{0} = 0.\end{aligned}$$

Hence

$$\mathbf{x}^T A\mathbf{x} = (\mathbf{x}-\mathbf{a})^T A(\mathbf{x}-\mathbf{a}) + \mathbf{a}^T A\mathbf{a} \geqslant \mathbf{a}^T A\mathbf{a}. \quad \blacksquare$$

We now compute $\lim_{\lambda\to\infty} \mathbf{a}^{(\lambda)}$, where $\mathbf{a}^{(\lambda)}$ is given by (4.3). We may write

$$\mathbf{a}^{(\lambda)} = [A^{(\lambda)}/\kappa]^{-1}\,[B^{(\lambda)}]^T\,\{B^{(\lambda)}[A^{(\lambda)}/\kappa]^{-1}\,[B^{(\lambda)}]^T\}^{-1}\boldsymbol{\varphi},$$

where $\kappa := \int_{\mathbb{R}^n} \Delta_\lambda^2(\mathbf{x})\,d\mathbf{x}$. First consider $A^{(\lambda)}/\kappa$.

$$\begin{aligned}\lim_{\lambda\to\infty} (1/\kappa)\,A_{ii}^{(\lambda)} &= \lim_{\lambda\to\infty} (1/\kappa)\int_D J(\mathbf{x},\mathbf{x}_0)\,\Delta_\lambda^2(\mathbf{x}-\mathbf{x}_i)\,d\mathbf{x} \\ &= \lim_{\lambda\to\infty}\int_{\mathbb{R}^n} \chi_D(\mathbf{x})\,J(\mathbf{x},\mathbf{x}_0)(1/\kappa)\,\Delta_\lambda^2(\mathbf{x}-\mathbf{x}_i)\,d\mathbf{x} \\ &= J(\mathbf{x}_i,\mathbf{x}_0) \qquad \text{as} \quad \{\Delta_\lambda\} \text{ is regular.}\end{aligned}$$

Also, for $j \neq i$,

$$\begin{aligned}\lim_{\lambda\to\infty} (1/\kappa)\,A_{ij}^{(\lambda)} &= \lim_{\lambda\to\infty} (1/\kappa)\int_D J(\mathbf{x},\mathbf{x}_0)\,K(\lambda(\mathbf{x}-\mathbf{x}_i))\,K(\lambda(\mathbf{x}-\mathbf{x}_j))\,d\mathbf{x} \\ &= 0 \qquad \text{again by the regularity of } \{\Delta_\lambda\}.\end{aligned}$$

Hence $\lim_{\lambda\to\infty} (1/\kappa)\,A^{(\lambda)} = \operatorname{diag}(J(\mathbf{x}_1,\mathbf{x}_0), ..., J(\mathbf{x}_N,\mathbf{x}_0))$. Set $W := \operatorname{diag}(J^{-1}(\mathbf{x}_1,\mathbf{x}_0), ..., J^{-1}(\mathbf{x}_N,\mathbf{x}_0)) = \lim_{\lambda\to\infty}\{(1/\kappa)\,A^{(\lambda)}\}^{-1}$.

We have already seen that $\lim_{\lambda\to\infty} B^{(\lambda)} = B$ (of (2.1)). Hence

$$\lim_{\lambda\to\infty} \mathbf{a}^{(\lambda)} = WB^T(BWB^T)^{-1}\,\boldsymbol{\varphi}$$

and

$$\lim_{\lambda\to\infty} \tilde{f}_\lambda(\mathbf{x}_0) = \mathbf{f}^T\mathbf{a} = \mathbf{f}^T WB^T(BWB^T)^{-1}\,\boldsymbol{\varphi}. \tag{4.4}$$

Comparing (4.4) with (2.3) we see that the Backus–Gilbert approximation is exactly moving least-squares with weights $w_i = 1/J(\mathbf{x}_i,\mathbf{x}_0)$. $\blacksquare$

## REFERENCES

1. F. ABRAMOVICI, "The Shepard interpolation as the best average of a set of data," Personal communication.
2. G. BACKUS AND F. GILBERT, The resolving power of gross earth data, *Geophys. J. R. Astr. Soc.* **16** (1968), 169–205.
3. P. L. BUTZER AND R. J. NESSEL, "Fourier Analysis and Approximation," Vol. 1, Academic Press, San Diego, 1971.
4. R. FARWIG, Rate of convergence of moving least squares interpolation methods, submitted for publication.
5. W. J. GORDON AND W. J. WIXOM, Shepard's method of 'metric interpolation' to bivariate and multivariate data, *Math. Comp.* **32** (1978), 253–264.
6. P. LANCASTER AND K. SALKAUSKAS, Surfaces generated by moving least squares methods, *Math. Comp.* **37** (1981), 141–158.
7. D. SHEPARD, A two-dimensional interpolation function for irregularly spaced points, *in* "Proc. 1968 Assoc. Comput. Machinery National Conference," 517–524.