Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 207–218

Applying the Backus-Gilbert theory to function approximation

Flavian Abramovici

Abstract. In this paper are given new results within the project I started some years ago, of using inverse problems methods for recovering the values at points $\mathbf{x}_0$ of a continuous function f with compact support $\mathbb{E} \subseteq \mathbb{R}^m$, when N of its values are given at the nodes $\mathbf{x}_i$. After showing in [1] how to obtain Shepard's formula with two different versions of the well known Backus-Gilbert process, building averaging kernels that resemble δ - "functions" centered at the nodes and consist in linear combinations of the data representers. In the present paper I am showing how to attach a spread to the Shepard formula itself, leading to a convergence theorem concerning the recovery of the considered function.

Mathematics Subject Classification (2010): 41A30.

Keywords: Backus-Gilbert theory, Shepard's formula, deltaness, inverse problems, moving least-squares.

1. Introduction

In order to use the classical Backus-Gilbert process, the data should consist in a set of bounded functionals. Not having such functionals, I tried to use the internal products

$$\int_{\mathbb{R}^m} f(\mathbf{x}) G_i^{(\lambda)}(\mathbf{x}) dV \tag{1.1}$$

between the given function f and the elements of a Dirac sequence ([9]), which for high enough λ could approximate such functionals, e.g. the elements of the Dirac sequence used in [1]

$$G^{(\lambda)}(\mathbf{x}) = \begin{cases} \lambda^m & \text{for } \mathbf{x} \in \mathbb{S} \\ \\ 0 & \text{otherwise} \end{cases}, \tag{1.2}$$

where $\mathbb{S}$ is the reqular hypercube having the center at the origin and edges of length $\frac{1}{\lambda}$, with λ a positive real parameter. The Backus-Gilbert classical theory is looking for the optimal linear combination of the form

$\sum_{i=1}^{N} a_i^{(\lambda)}(\mathbf{x}_0)G_i^{(\lambda)}(\mathbf{x})$ with $G_i^{(\lambda)}(\mathbf{x}) = G^{(\lambda)}(\mathbf{x}-\mathbf{x}_i)$, that gives the best approximation of the value of the function f at $\mathbf{x}_0$

$$\begin{aligned}\tilde{f}^{(\lambda)}(\mathbf{x}_0) &= \sum_{i=1}^{N}\int_{\mathbb{R}^m} a_i^{(\lambda)}(\mathbf{x}_0)f(\mathbf{x})G_i^{(\lambda)}(\mathbf{x})dV \\ &= \sum_{i=1}^{N}\int_{\mathbb{R}^m} a_i^{(\lambda)}(\mathbf{x}_0)f(\mathbf{x})G^{(\lambda)}(\mathbf{x}-\mathbf{x}_i)dV\ , \end{aligned} \tag{1.3}$$

with dV the volume element $dx_1,\ldots,dx_m$. Taking the limit of the result for $\lambda\to\infty$ in order to compensate for the errors involved in using only finite values of λ, I obtained the well-known Shepard's formula [7],[12],[13]:

$$\lim_{\lambda\to\infty}\tilde{f}^{(\lambda)}(\mathbf{x}_0) = \sum_{k=1}^{N}\frac{1}{||\mathbf{x}_k-\mathbf{x}_0||^2\sum\limits_{i=1}^{N}\frac{1}{||\mathbf{x}_i-\mathbf{x}_0||^2}}\ . \tag{1.4}$$

This result surprised some people working in seismology and others working in numerical analysis, as the Backus-Gilbert theory [2],[3],[4] was known mainly to geophysicists, while Shepard's formula was familar to mathematicians working in numerical analysis. In order to make sure that my results were correct, I looked for another approximation of $f(\mathbf{x}_0)$ having all the ingredients of the Backus-Gilbert theory including a spread. What I actually did was to discretize the integrals involved, obtaining the discrete version of the Backus-Gilbert theory. Following closely the way the classical theory was built, I applied the Backus-Gilbert linear representation theorem marked below as Theorem 2.1, finding that necessarily the average given by my discrete version had to be of the form obtained by discretization. After finishing the preliminary report, I passed copies to people who showed a special interest in my results and with whom I had many discussions, among them Prof. Kes Salkauskas from the University of Calgary in Canada and Prof. David Levin from the Hebrew University in Jerusalem. They extended my findings obtaining new results, e.g. making the same steps I did including taking a limit, Bos and Salkoskas obtained in [5] the moving least-squares approximation [8], a generalisation of Shepard's formula. For this purpose they defined a special type of Dirac sequences called regular, in order to handle the quadratic integrals needed for the optimal solutions. On the other hand David Levin using not only my results but also those of Bos and Salkauskas, presented an elegant way to obtain the moving least squares process using block matrices and applied with great success in [10] my discrete Backus-Gilbert process to scattered interpolation, smoothing and numerical differentition and, in [11], to numerical integration.

Once I obtained the Shepard formula using two independent Backus-Gilbert theories and being further stimulated by the results obtained by Bos, Salkauskas and Levin, I decided to try to settle an important question that was still open: is it possible to attach to the Shepard approximation also a sort of spread, which would lead to further properties? Although both Backus-Gilbert theories led to Shepard's formula, only within the discrete one I could define a spread, as for $\lambda \to \infty$ the process was divergent, the spread tending to infinity. The solution came to me while reading the paper by Bos and Salkauskas, as I realized that the simple Dirac sequence I used was regular and therefore what I had to do was "only" to modify a little the classical Backus-Gilbert spread by normalizing its integrand. As a result, the integral representing the spread for every λ became convergent, the limit having all the characteristic properties of a spread. It was therefore justified to define this limit as being the Backus-Gilbert spread of Shepard's formula. Moreover, it turned out that the limit obtained as described coincides with the spread attached to the Shepard formula within the discrete theory directly, not in combination with taking a limit and with its own justification. For the benefit of those not familiar with the classical Backus-Gilbert theory, a short description is given in Section 2. In Section 3 is shown how the Shepard formula is obtained and in the last Section is described the discrete Backus-Gilbert version.

2. The classical Backus-Gilbert theory

Clearly, if we have only a finite number of data, it is not possible to determine exactly the properties of the Earth at every location, but it may be possible to get averages of the so called "Earth models", functions f belonging to the Hilbert space $\mathcal{H} = L_2(\mathbb{E})$ with $\mathbb{E}$ the closed, connected and bounded support of f in $\mathbb{R}^m$ representing the properties of the Earth. Hence, the most we may hope to achieve using the data described in Section 1 is to find significant quantities that characterize the entire family of models, e.g. the average f_{av} at every point $\mathbf{x}_0$ of $\mathbb{E}$, corresponding to an averaging kernel $\mathcal{A}(\mathbf{x}_0, \mathbf{x})$:

$$f_{av}(\mathbf{x}_0) = \int_{\mathbb{E}} \mathcal{A}(\mathbf{x}_0, \mathbf{x}) f(\mathbf{x}) dV \simeq f(\mathbf{x}_0) \ . \tag{2.1}$$

Backus and Gilbert proved in [3] the following general result.

Theorem 2.1. *Let $f \in \mathcal{H}$ be a function for which N linearly independent bounded linear functionals γ_i on $\mathcal{H}$ are known. If it is possible to obtain a linear average $\mathcal{L}_{av}(f)$ of f at a point $\mathbf{x}_0$ using only the given N functionals, then the average $\mathcal{L}_{av}(f)$ is necessarily a linear combination of these functionals:*

$$\mathcal{L}_{av}(f) = \sum_{i=1}^{N} a_i \mathcal{L}_i(f) \tag{2.2}$$

with coefficients a_i that depend upon $\mathbf{x}_0$.

Using this result taking as $\mathcal{L}_i(f)$ the values $G_i(\mathbf{x})$, we find that

$$\mathcal{L}_{av}(f) = \sum_{i=1}^{N} a_i \int_{\mathbb{E}} G_i(\mathbf{x}) f(\mathbf{x}) dV = \int_{\mathbb{E}} \left[\sum_{i=1}^{N} a_i G_i(\mathbf{x}) \right] f(\mathbf{x}) dV \ , \tag{2.3}$$

i.e. the averaging kernels are indeed linear combination of the representers $G_i(\mathbf{x})$:

$$\mathcal{A}(\mathbf{x}_0, \mathbf{x}) = \sum_{i=1}^{N} a_i(\mathbf{x}_0), G_i(\mathbf{x}) \ . \tag{2.4}$$

Consequently, Backus and Gilbert looked for an optimal unimodular averaging kernel $\mathcal{A}(\mathbf{x}_0, \mathbf{x})$ i.e satisfying the condition

$$\int_{\mathbb{E}} \mathcal{A}(\mathbf{x}_0, \mathbf{x}) dV = 1 \tag{2.5}$$

and having the highest deltaness, that is the highest likeness to the Dirac δ-function centered at $\mathbf{x}_0$, condition checked by using the "spread" of the average kernel defined, by Backus and Gilbert as follows.

Definition 2.2. For every averaging kernel $\mathcal{A}(\mathbf{x}_0, \mathbf{x})$ on a compact set $\mathbb{E} \subseteq \mathbb{R}^m$, the function

$$s_0 = s(\mathbf{x}_0) = \frac{12}{m} \int_{\mathbb{E}} J(\mathbf{x}_0, \mathbf{x}) \mathcal{A}^2(\mathbf{x}_0, \mathbf{x}) dV \tag{2.6}$$

with $J(\mathbf{x}_0, \mathbf{x})$ a "sink" function i.e. a non-negative function that vanishes for $\mathbf{x} = \mathbf{x}_0$ and grows rapidly away from this point, is called a spread of $\mathcal{A}$ at $\mathbf{x}_0$. A typical "sink" function is $J(\mathbf{x}_0, \mathbf{x}) = ||\mathbf{x} - \mathbf{x}_0||^2$, with $||\mathbf{x} - \mathbf{x}_0||$ the Euclidean norm.

Thus, the Backus-Gilbert process solves the following variational problem: *find the coefficients $a_i(\mathbf{x}_0)$ for which the averaging kernel has the highest δ-ness, i.e. the smallest spread.*

Using a Lagrange multiplier, Backus and Gilbert solved this classical variational problem, obtaining the following relation giving the coefficients of the averaging kernel:

$$\mathbf{a}(\mathbf{x_0}) = \frac{1}{\mathbf{u}^T \left[\mathbf{Z}(\mathbf{x_0})\right]^{-1} \mathbf{u}} \left[\mathbf{Z}(\mathbf{x_0})\right]^{-1} \mathbf{u} \tag{2.7}$$

with $\mathbf{Z}(\mathbf{x}_0)$ a Gram matrix [6] of components

$$Z_{ik}(\mathbf{x}_0) = \frac{12}{m} \int_{\mathbf{E}} J(\mathbf{x}_0, \mathbf{x}) G_i(\mathbf{x}) G_k(\mathbf{x}) dV \ , \tag{2.8}$$

the corresponding spread $s(\mathbf{x}_0)$ being given by

$$s_0 = \mathbf{a}(\mathbf{x_0})^T \mathbf{Z}(\mathbf{x_0}) \mathbf{a}(\mathbf{x_0}) \ , \tag{2.9}$$

Consider now the functions $f \in \mathcal{H}$ for which the following integrals

$$\int_{\mathbf{E}} \frac{|f(\mathbf{x}) - f(\mathbf{x}_0)|}{||\mathbf{x} - \mathbf{x}_0||} dV \quad \text{and} \quad \int_{\mathbf{E}} \frac{|f(\mathbf{x}) - f(\mathbf{x}_0)|^2}{||\mathbf{x} - \mathbf{x}_0||^2} dV \tag{2.10}$$

are finite. Defining the square root of the second integral as being the "$\mathbf{x}_0$-norm" $||f(\mathbf{x})||_{\mathbf{x}_0}$ of a function f that belongs to $\mathcal{H}$ and is either identically zero or not constant, Backus and Gilbert proved the following result:

Theorem 2.3. *Under the conditions described above. if the averaging kernel $\mathcal{A}(\mathbf{x}_0, \mathbf{x})$ given by (2.4) is unimodular, the error functional*

$$\mathcal{E}(f(\mathbf{x}_0)) = \int_{\mathbf{E}} \mathcal{A}(\mathbf{x}_0, \mathbf{x}) \left[f(\mathbf{x}) - f(\mathbf{x}_0) \right] dV \tag{2.11}$$

is bounded, its x_0-norm being given by

$$||\mathcal{E}||_{\mathbf{x}_0} = \sqrt{\frac{m}{12} s(\mathbf{x}_0)} \ , \tag{2.12}$$

with $s(\mathbf{x}_0)$ the considered spread.

Corollary 2.4. *(The boundedness inequality for a given λ) (For every $\lambda > 0$ the following inequality takes place:*

$$|\mathcal{E}(f(\mathbf{x}_0))| \leq ||\mathcal{E}||_{\mathbf{x}_0} ||f||_{\mathbf{x}_0} \tag{2.13}$$

Corollary 2.5. *(A convergence theorem) Let $\mathcal{A}^{(\lambda)}(\mathbf{x}_0, \mathbf{x}) = \sum\limits_{i=1}^{N} a_i^{(\lambda)}(\mathbf{x}_0) G_i^{(\lambda)}(\mathbf{x})$ be a family of unimodular averaging kernels. The average $f^{(\lambda)}(\mathbf{x}_0)$ tends to $f(\mathbf{x}_0)$ when $\lambda \to \mu$, if and only if $\lim\limits_{\lambda \to \mu} s^{(\lambda)}(\mathbf{x}_0) = 0$, in particular $f^{(\lambda)}(\mathbf{x}_0)$ tends to $f(\mathbf{x}_0)$ when $\lambda \to \infty$ if and only if $s^{(\lambda)}(\mathbf{x}_0) \to 0$ when $\lambda \to \infty$.*

Remark 2.6. This theorem is important as a general result but in many cases the computer time needed to reach the wanted precision is very large. This is why it is important to have an efficient method, giving an effective growth of accuracy at every step, which is precisely why the Backus-Gilbert process is preferable to other methods, as one may see on the examples given by David Levin in the articles mentioned above.

3. Approximating a function with given values using the classical Backus-Gilbert theory

Using the general relations (2.7) - (2.8), we prove the following result:

Theorem 3.1. *Let f be a unimodular Earth model satisfying the conditions of Theorem 2.1. For every set of data of the form*

$$\tilde{f}^{(\lambda)}(\mathbf{x}_i) = \int_{\mathbb{E}_i} \tilde{G}_i^{(\lambda)}(\mathbf{x}) f(\mathbf{x}) dV \ , \tag{3.1}$$

the coefficients that minimize the spread of the averaging kernel are

$$\tilde{\mathbf{a}}^{(\lambda)}(\mathbf{x}_0) = \frac{1}{\left[\tilde{\mathbf{u}}^{(\lambda)}\right]^T \left[\tilde{\mathbf{Z}}^{(\lambda)}(\mathbf{x}_0)\right]^{-1} \tilde{\mathbf{u}}^{(\lambda)}} \left[\tilde{\mathbf{Z}}^{(\lambda)}(\mathbf{x}_0)\right]^{-1} \tilde{\mathbf{u}}^{(\lambda)} \ . \tag{3.2}$$

with $\tilde{u}^{(\lambda)} = 1$ for every λ and

$$\tilde{\mathbf{Z}}_{ik}^{(\lambda)}(\mathbf{x}_0) = \int_{\mathbf{R}^m} J(\mathbf{x}_0, \mathbf{x}) \tilde{G}_i^{(\lambda)}(\mathbf{x}) \tilde{G}_k^{(\lambda)}(\mathbf{x}) dV \ . \tag{3.3}$$

As for $\lambda > 0$ large enough, the matrix $\tilde{\mathbf{Z}}^{(\lambda)}(\mathbf{x}_0)$ is diagonal with diagonal elements $\lambda^2 \tilde{X}_k^{(\lambda)}(\mathbf{x}_0)$ where

$$\tilde{X}_k^{(\lambda)}(\mathbf{x}_0) = ||\mathbf{x}_k - \mathbf{x}_0||^2 + \frac{1}{\lambda^2} \ . \tag{3.4}$$

Making λ tend to infinity, we find that

$$\breve{\mathbf{a}}(\mathbf{x}_0) = \lim_{\lambda\to\infty} \tilde{a}^{(\lambda)}(\mathbf{x}_0) = \frac{1}{\sum\limits_{i=1}^{N} \frac{1}{||\mathbf{x}_i - \mathbf{x}_0||^2}} \begin{pmatrix} \frac{1}{||\mathbf{x}_1 - \mathbf{x}_0||^2} \\ \vdots \\ \frac{1}{||\mathbf{x}_N - \mathbf{x}_0||^2} \end{pmatrix} , \tag{3.5}$$

i.e. we arrive to the following result:

Corollary 3.2. *The considerate Earth model is approximated by Shepard's formula*

$$\tilde{f}(x_0) = \frac{1}{\sum\limits_{i=1}^{N} \frac{1}{||\mathbf{x}_i - \mathbf{x}_0||^2}} \sum_{k=1}^{N} \frac{f(\mathbf{x}_k)}{||\mathbf{x}_k - \mathbf{x}_0||^2} \quad . \tag{3.6}$$

Having obtained the Shepard formula using the Backus-Gilbert theory, it is only natural to try to attach to it a Backus-Gilbert spread for characterizing the way Shepard's formula approximates the function f. In order to do so, we calculate the spread of the optimal averaging kernel for a given λ:

$$\begin{aligned} \tilde{s}^{(\lambda)}(\mathbf{x}_0) = {} & \frac{12}{m} \sum_{k=1}^{N} \left[\tilde{a}_k^{(\lambda)}(\mathbf{x}_0)\right]^2 \int_{\mathbb{E}_k} J(\mathbf{x}_0, \mathbf{x}) \left[G_k^{(\lambda)}(\mathbf{x})\right]^2 dV \\ & + \frac{12}{m} \sum_{k,l=1}^{N} \tilde{a}_k^{(\lambda)}(\mathbf{x}_0)\tilde{a}_\ell^{(\lambda)}(\mathbf{x}_0) \int_{\mathbb{E}_k \bigcap \mathbb{E}_\ell} J(\mathbf{x}_0, \mathbf{x})\tilde{G}_k^{(\lambda)}(\mathbf{x}_0)\tilde{G}_\ell^{(\lambda)}(\mathbf{x}_0) dV \ . \end{aligned} \tag{3.7}$$

and see if it tends to a finite limit when λ tends to infinity. For λ large enough the intersection $\mathbb{E}_k \bigcap \mathbb{E}_\ell$ is empty, so that the double sum in the second term of the right hand side is equal to zero as it is easy to see and therefore we are left with

$$\tilde{s}^{(\lambda)}(\mathbf{x}_0) \quad = \quad \frac{12}{m} \sum_{k=1}^{N} \left[\tilde{a}_k^{(\lambda)}(\mathbf{x}_0)\right]^2 \int_{\mathbb{E}_k} J(\mathbf{x}_0, \mathbf{x}) \left[G_k^{(\lambda)}(\mathbf{x})\right]^2 dV \ , \tag{3.8}$$

a divergent integral ! However, as already explained, the simple Dirac sequence $G^{(\lambda)}(\mathbf{x})$ is regular, according to the following definition.

Definition 3.3. A Dirac sequence $G^{(\lambda)}(\mathbf{x})$ is called regular if the following conditions hold:

1. $G^{(\lambda)} \in L_2(\mathbb{R}^m)$.
2. For every bounded and continuous function $f \in L_2(\mathbb{R}^m)$

$$\lim_{\lambda\to\infty} \int_{\mathbb{R}^m} \frac{G^{(\lambda)}(\mathbf{x}-\mathbf{a})G^{(\lambda)}(\mathbf{x}-\mathbf{b})}{\kappa_m^{(\lambda)}} f(\mathbf{x})dV = \begin{cases} \mathbf{0} & \text{if} \quad \mathbf{b} \neq \mathbf{a} \\ f(\mathbf{a}) & \text{if} \quad \mathbf{b} = \mathbf{a} \end{cases} \tag{3.9}$$

with

$$\kappa_m^{(\lambda)} = \int_{\mathbb{R}^m} \left[G^{(\lambda)}(\mathbf{x})\right]^2 dV \; . \tag{3.10}$$

Dividing both sides of relation (38) by $\tilde{\kappa}_m^{(\lambda)}$ and knowing that the ratio $\dfrac{\tilde{s}^{(\lambda)}(\mathbf{x}_0)}{\tilde{\kappa}_m^{(\lambda)}}$ is convergent for $\lambda \to \infty$, we find that the right hand side is convergent and therefore the left hand side is also convergent to $\check{s}(\mathbf{x}_0)$ that may be called the normalized spread of Shepard's formula:

$$\begin{aligned} \check{s}(\mathbf{x}_0) &= \frac{12}{m}\sum_{k=1}^{N} \lim_{\lambda\to\infty}\left[\tilde{a}_k^{(\lambda)}(\mathbf{x}_0)\right]^2 \lim_{\lambda\to\infty}\int_{\mathbf{E}_k} J(\mathbf{x}_0,\mathbf{x})\frac{\left[\tilde{G}_k^{(\lambda)}(\mathbf{x})\right]^2}{\tilde{\kappa}_m^{(\lambda)}} dV \; . \\ &= \frac{12}{m}\sum_{k=1}^{N} \check{a}_k^2(\mathbf{x}_0) J(\mathbf{x}_0,\mathbf{x}_k) \; . \end{aligned} \tag{3.11}$$

Moreover, we may attach to Shepard's formula a boudedness inequality leading to a convergence property. Indeed, consider the boundedness inequality (2.13) for a the minimal solution for any λ

$$|\tilde{\mathcal{E}}^{(\lambda)}(\tilde{f}^{(\lambda)})| \le ||\tilde{\mathcal{E}}^{(\lambda)}||_{\mathbf{x}_0}\; ||\tilde{f}^{(\lambda)}||_{\mathbf{x}_0} \; , \tag{3.12}$$

the error functional $\tilde{\mathcal{E}}^{(\lambda)}(f^{(\lambda)})$ being defined by

$$\tilde{\mathcal{E}}^{(\lambda)}(f^{(\lambda)}) = \tilde{f}^{(\lambda)}(\mathbf{x}_0) - f(\mathbf{x}_0) = \sum_{i=1}^{N} \tilde{a}_i^{(\lambda)}(\mathbf{x}_0)\tilde{f}_i^{(\lambda)} - f(\mathbf{x}_0) \; , \tag{3.13}$$

with $\tilde{f}_i^{(\lambda)} = \int_{\mathbf{E}} \tilde{G}_i^{(\lambda)}(\mathbf{x}) f(\mathbf{x}) dV$. Using the expression of the averaging kernel (2.4) and its unimodularity, we find that

$$\lim_{\lambda\to\infty} \tilde{\mathcal{E}}^{(\lambda)}(f) = \lim_{\lambda\to\infty}\int_{\mathbf{E}} \tilde{\mathcal{A}}^{(\lambda)}(\mathbf{x}_0,\mathbf{x})[f(\mathbf{x}) - f(\mathbf{x}_0)]dV = \check{\mathcal{E}}(f) \; . \tag{3.14}$$

As to the right hand side, the first factor is equal to $\sqrt{\frac{m}{12}\tilde{s}^{(\lambda)}(\mathbf{x}_0)}$ and therefore tends to infinity when λ does, due to the presence of the spread according to (2.12). Dividing by κ_m the first factor of the right hand side becomes convergent but in this case we have to multiply the second factor also by κ_m, bringing the inequality to the following form:

$$|\tilde{\mathcal{E}}^{(\lambda)}(\tilde{f}^{(\lambda)})| \le \sqrt{\frac{m}{12}\frac{\tilde{s}^{(\lambda)}(\mathbf{x}_0)}{\tilde{\kappa}_m^{(\lambda)}}}\; \sqrt{\tilde{\kappa}_m^{(\lambda)}}||\tilde{f}^{(\lambda)}||_{\mathbf{x}_0} \; . \tag{3.15}$$

In order to prove that the right hand side is convergent and that its limit is also a "discrete" quantity, consider the "shrinking" function

$$F(\mathbf{x}) = \begin{cases} f(\mathbf{x}) & \text{if } \mathbf{x} = \mathbf{x}_0 \text{ or } \mathbf{x} \in \mathbb{E}_i \\ & \text{for } i = 1, \ldots, N \text{ and} \\ 0 & \text{otherwise} \end{cases} \tag{3.16}$$

and its approximation $\tilde{F}^{(\lambda)}(\mathbf{x}_0) = \sum_{i=1}^{N} \tilde{a}_i^{(\lambda)}(\mathbf{x}_0) \int_{\mathbb{E}_i} \tilde{G}_i^{(\lambda)}(\mathbf{y})F(\mathbf{y})dV$. For λ large enough this approximation also satisfies the boundedness inequality, in fact $\tilde{F}^{(\lambda)}(\mathbf{x}_0)$ is equal to $\tilde{f}^{(\lambda)}(\mathbf{x}_0)$ for every $\mathbf{x}_0$ different from any node $\mathbf{x}_i$ for $i = 1, \dots N$, so that the error functionals and the x_0-norms coincide. Taking into account that $\tilde{f}^{(\lambda)}$ is sectionally continuous we find, using the mean value theorem, that there exists at least one point $\xi_i^{(\lambda)} \in \mathbb{E}_i$ that depends upon λ, such that

$$\int_{\mathbb{E}_i} \frac{\left|\tilde{f}^{(\lambda)}(\mathbf{x}) - \tilde{f}^{(\lambda)}(\mathbf{x}_0)\right|^2}{(\mathbf{x}-\mathbf{x}_0)^2} dx = \frac{\left|\tilde{f}^{(\lambda)}(\xi_i^{(\lambda)}) - \tilde{f}^{(\lambda)}(\mathbf{x}_0)\right|^2}{(\xi_i^{(\lambda)} - \mathbf{x}_0)^2} \epsilon \tag{3.17}$$

and therefore

$$\sqrt{\tilde{\kappa}_1^{(\lambda)}} ||\tilde{F}^{(\lambda)}||_{\mathbf{x}_0} = \sqrt{\sum_{i=1}^{N} \frac{\left|\tilde{f}^{(\lambda)}(\xi|_i^{(\lambda)}) - \tilde{f}^{(\lambda)}(\mathbf{x}_0)\right|^2}{(\xi_i^{(\lambda)} - \mathbf{x}_0)^2}}\ , \tag{3.18}$$

as $\lambda\epsilon = 1$. Using this relation we get from (3.15) the inequality

$$|\tilde{\mathcal{E}}^{(\lambda)}(\tilde{f}^{(\lambda)})| \leq \sqrt{\frac{m\tilde{s}^{(\lambda)}(\mathbf{x}_0)}{12\tilde{\kappa}_1^{(\lambda)}}} \sqrt{\sum_{i=1}^{N} \frac{\left|\tilde{f}^{(\lambda)}(\xi_i^{(\lambda)}) - \tilde{f}^{(\lambda)}(\mathbf{x}_0)\right|^2}{(\xi_i^{(\lambda)} - \mathbf{x}_0)^2}}\ , \tag{3.19}$$

with the first factor of the right hand side tending to the "discrete" quantity $\sqrt{\frac{m}{12}\check{s}(\mathbf{x}_0)}$. On the other hand when λ tends to infinity, ϵ tends to zero so that $\xi_i^{(\lambda)}$ tends to the center $\mathbf{x}_i$ of $\mathbb{E}_i$ and $\tilde{f}^{(\lambda)}$ is continuous at $\mathbf{x}_i$. Therefore, $\lim_{\lambda\to\infty} \tilde{f}^{(\lambda)}(\xi) = \tilde{f}^{(\lambda)}(\mathbf{x}_i)$ implying that

$$\lim_{\lambda\to\infty} \sqrt{\sum_{i=1}^{N} \frac{\left|\tilde{f}^{(\lambda)}(\xi_i^{(\lambda)}) - \tilde{f}^{(\lambda)}(\mathbf{x}_0)\right|^2}{(\xi_i^{(\lambda)} - \mathbf{x}_0)^2}} = \sqrt{\sum_{i=1}^{N} \frac{|f(\mathbf{x}_i) - f(\mathbf{x}_0)|^2}{(\mathbf{x}_i - \mathbf{x}_0)^2}}\ . \tag{3.20}$$

As a result, we get using (3.13) the following boundedness inequality for the discrete error functional:

$$|\check{\mathcal{E}}(f)| \leq \sqrt{\frac{\check{s}(\mathbf{x}_0)}{12}} \sqrt{\sum_{i=1}^{N} \frac{|f(\mathbf{x}_i) - f(\mathbf{x}_0)|^2}{(\mathbf{x}_i - x_0)^2}}\ . \tag{3.21}$$

making possible to define the following discrete quantity.

Definition 3.4. *The discrete* $\mathbf{x}_0$*-norm of a function* $f(\mathbf{x})$ *corresponding to the points* $x_1, \dots, x_N \in \mathbb{E}$ *is*

$$\|f\|_{\mathbf{x}_0} = \sqrt{\sum_{k=1}^{N} \frac{|f(\mathbf{x}_k) - f(\mathbf{x}_0)|^2}{(\mathbf{x}_k - \mathbf{x}_0)^2}}\ . \tag{3.22}$$

As a result, the discrete boundedness inequality (3.21) becomes

$$|\check{\mathcal{E}}(f)| \leq \sqrt{\frac{\check{s}(\mathbf{x}_0)}{12}}\|f\|_{\mathbf{x}_0} \tag{3.23}$$

leading to the following conclusion:

Corollary 3.5. *The continuous boundedness inequality (*3.12*) written in the form*

$$|\tilde{\mathcal{E}}^{(\lambda)}(\tilde{f}^{(\lambda)})| \leq \sqrt{\frac{\check{s}(\mathbf{x}_0)}{12}}||\tilde{f}^{(\lambda)}||_{\mathbf{x}_0} \tag{3.24}$$

*tends to the discrete boundedness inequality (*3.23*) when $\lambda \to \infty$, i.e. the left hand side of (*3.24*) tends to the left hand side of (*3.23*) and similarly for the right hand sides.*

Corollary 3.6. *The discrete error functional tends to zero if the discrete spread tends to zero.*

4. A discrete Backus-Gilbert theory

In this Section are presented formally without going into all the proofs, the main definitions and properties of my version of the Backus-Gilbert discrete process, built by similarity with the classical process: similar discrete spread, discrete $\mathbf{x}_0$-norm and discrete boundedness inequality with similar properties. As already explained, my initial intention was just to check the results obtained applying the classical theory combined with taking the limit for λ tending to infinity, but it turned out from my own results as well as from those of other people, that the discrete theory is very effective and gives very good numerical results in many cases. Moreover, I found that the way to write the corresponding approximation is dictated by Theorem 1.2 like in the continuous case, the difference being, of course the different data set. As a result, one finds that the form one writes usually a discrete average

$$f_{av}(\mathbf{x}_0) = \sum_{i=1}^{N} a_i(\mathbf{x}_0)f(\mathbf{x}_i) \ . \tag{4.1}$$

is the only possible one under the adopted assumptions.

Definition 4.1. *Let $\mathcal{A} = \mathcal{A}(\mathbf{x}_0)$ be a set of real non-negative numbers a_i*

$$\mathcal{A} = \{a_i(\mathbf{x}_0)\}_{i=1}^{N} \ , \tag{4.2}$$

used as coefficients of the average value of a function $f \in \mathcal{H}$ at a point $\mathbf{x}_0 \in \mathbb{E} \subset \mathbb{R}^m$. The set $\mathcal{A}$ is called a discrete averaging kernel of f at $\mathbf{x}_0$ and the Euclidean norm of the vector $\mathbf{a}$ having as components the coefficients a_i is called the norm $||\mathcal{A}|| = ||\mathcal{A}(\mathbf{x}_0)||$ of the averaging kernel $\mathcal{A}$.

It is not difficult to prove the following property.

Theorem 4.2. *The $\mathbf{x}_0$-norm (*3.22*) of any sectionally continuous function f which is either zero identically or non-constant on $\mathbb{E}$, is indeed a norm.*

Definition 4.3. *For every discrete averaging kernel $\mathcal{A}$ at $\mathbf{x}_0$, one may define its discrete spread at $\mathbf{x}_0$*

$$s_0 = s(\mathbf{x}_0, \mathcal{A}) = \frac{12}{m} \sum_{i=1}^{N} J_i a_i^2(\mathbf{x}_0) \ , \tag{4.3}$$

the factors $J_i = J(\mathbf{x}_i, \mathbf{x}_0) = ||\mathbf{x}_i - \mathbf{x}_0||^2$ called the spread's coefficients, measuring the "location separation" between the nodes $\mathbf{x}_i$ and the target point $\mathbf{x}_0$.

Remark 4.4. Like with the "deltaness" property in the classical Backus-Gilbert theory, these coefficients depend upon the distances between the nodes $\mathbf{x}_i$ and the current point $\mathbf{x}_0$ through a "sink" function.

Definition 4.5. *A discrete averaging kernel $\mathcal{A} = \{a_i(\mathbf{x}_0)\}_{i=1}^{N}$ is called unimodular if $\sum\limits_{i=1}^{N} a_i(\mathbf{x}_0) = 1$.*

Theorem 4.6. *For every unimodular averaging kernel and every function f belonging to $\mathcal{C}_{\mathbb{E}}$ or to L^2, the error functional*

$$\mathcal{E}(f, \mathbf{x}_0) = f_{av}(\mathbf{x}_0) - \ f(\mathbf{x}_0) \tag{4.4}$$

with f_{av} given by (4.1) is bounded, satisfies the inequality

$$|\mathcal{E}(f, \mathbf{x}_0)|_{\mathbf{x}_0} \leq ||\mathcal{E}||_{\mathbf{x}_0} ||f||_{\mathbf{x}_0} \ , \tag{4.5}$$

and its $\mathbf{x}_0$-norm is given by

$$||\mathcal{E}||_{\mathbf{x}_0} = \sqrt{\frac{m}{12} s(\mathbf{x}_0, \mathcal{A})} \ . \tag{4.6}$$

Proof. Using (4.1) and the unimodularity of the averaging kernel, we may write

$$\mathcal{E}(f, \mathbf{x}_0) = \sum_{i=1}^{N} a_i(\mathbf{x}_0)[f(\mathbf{x}_i) - f(\mathbf{x}_0)] = \sum_{i=1}^{N} u_i v_i \tag{4.7}$$

with

$$u_i = \sqrt{J_i} a_i(\mathbf{x}_0) \quad \text{and} \quad v_i = \frac{f(\mathbf{x}_i) - f(\mathbf{x}_0)}{\sqrt{J_i}} \ , \tag{4.8}$$

so that applying the Cauchy-Schwarz inequality we get

$$|\mathcal{E}(f, \mathbf{x}_0)| \leq \sqrt{\sum_{i=1}^{N} J_i a_i^2(\mathbf{x}_0)} \sqrt{\sum_{i=1}^{N} \frac{|f(\mathbf{x}_i) - f(\mathbf{x}_0)|^2}{J_i}} \ . \tag{4.9}$$

However, the first factor on the right is the discrete spread and the second one is the $\mathbf{x}_0$-norm of f so that

$$|\mathcal{E}(f, \mathbf{x}_0)| \leq \sqrt{\frac{m}{12} s\left(\mathbf{x}_0, \mathcal{A}\right)} \ ||f||_{\mathbf{x}_0} \ , \tag{4.10}$$

Hence, the linear functional $\mathcal{E}(f, \mathbf{x}_0)$ is bounded, its norm $||\mathcal{E}||_{\mathbf{x}_0}$ being not larger than any of its upper bounds, in particular not larger than

$\sqrt{\frac{m}{12} s\left(\mathbf{x}_0, \mathcal{A}\right)}$, which proves inequality (4.5). As to (4.6), consider the function $p(\mathbf{x}) = ||\mathbf{x} - \mathbf{x}_0||^2 q(\mathbf{x})$ with $q(\mathbf{x})$ continuous in $\mathbb{E}$ and satisfying the condition $q(\mathbf{x}_i) = a_i(\mathbf{x}_0)$ for $i = 1, \ldots, N$. It turns out that the inequality (4.10) is actually an equality for $f(\mathbf{x}) = p(\mathbf{x})$:

$$|\mathcal{E}(p, \mathbf{x}_0)| = \sqrt{\frac{m}{12} s\left(\mathbf{x}_0, \mathcal{A}\right)}\ ||p||_{\mathbf{x}_0}\ . \tag{4.11}$$

Indeed, the error functional (4.4) corresponding to $f(\mathbf{x}) = p(\mathbf{x})$ is equal just to $f_{av}(\mathbf{x}_0)$ as $p(\mathbf{x}_0) = 0$, so that in this case we get using (4.1) $\mathcal{E}(p, \mathbf{x}_0) = \sum_{i=1}^{N} a_i p(\mathbf{x}_i)$. However, $p(\mathbf{x}_i) = ||\mathbf{x}_i - \mathbf{x}_0||^2 q(\mathbf{x}_i)$ and $||\mathbf{x}_i - \mathbf{x}_0||^2 = J_i$ while $q(\mathbf{x}_i) = a_i(\mathbf{x}_0)$, so that $p(\mathbf{x}_i) = J_i a_i(\mathbf{x}_0)$ and therefore

$$\mathcal{E}(p, \mathbf{x}_0) = \sum_{i=1}^{N} a_i^2(\mathbf{x}_0) J_i = \sqrt{\sum_{i=1}^{N} a_i^2(\mathbf{x}_0) J_i} \sqrt{\sum_{i=1}^{N} a_i^2(\mathbf{x}_0) J_i}\ . \tag{4.12}$$

On the other hand, using the definition of the $\mathbf{x}_0$-norm we find that

$$||p||_{\mathbf{x}_0} = \sqrt{\sum_{i=1}^{N} \frac{|p(\mathbf{x}_i) - p(\mathbf{x}_0)|^2}{J_i}} = \sqrt{\sum_{i=1}^{N} \frac{|p(\mathbf{x}_i)|^2}{J_i}} = \sqrt{\sum_{i=1}^{N} a_i^2(\mathbf{x}_0) J_1}\ , \tag{4.13}$$

enabling us to replace one of the square roots in (4.12) by the $\mathbf{x}_0$-norm $||p||_{\mathbf{x}_0}$, whereas using the definition (4.3) of the discrete spread,, we find that we may replace the second square root in the right hand side of (4.12) by $\sqrt{\frac{m}{12} s(\mathbf{x}_0, \mathcal{A})}$, obtaining precisely (4.11). □

Based on this result, we arrive to the following easy to prove pointwise convergence theorem.

Theorem 4.7. *For any sequence $\mathcal{A}^{(\nu)}$ $(\nu = 1, 2, \ldots)$ of $\mathbf{x_0}-$ restricted unimodular discrete averaging kernels, the sequence $f_{av}^{(\nu)}(\mathbf{x}_0)$ tends to the exact value $f(\mathbf{x}_0)$ for every $\mathbf{x}_0$, if and only if $\lim_{n\to\infty} s_0^{(\nu)} = \lim_{\nu\to\infty} s\left(\mathbf{x}_0, \mathcal{A}^{(\nu)}\right) = 0$.*

In order to obtain the optimal average one solve here also a variational problem using also a Lagrange multiplier, the coefficients and the multiplier satisfying the same system of equations

$$\frac{\partial \tau}{\partial a_k} = 0 \ \text{ for } \ k = 1, \ldots, N \text{ and } \frac{\partial \tau}{\partial \eta} = 0\ . \tag{4.14}$$

Hence, the solution of this equation is

$$a_k = \frac{m\eta}{24||\mathbf{x}_0 - \mathbf{x}_k||2} \quad (k = 1, \ldots, N)\ . \tag{4.15}$$

with

$$\eta = \frac{24}{m} \left(\sum_{i=1}^{N} \frac{1}{||\mathbf{x}_0 - \mathbf{x}_i||^2} \right)^{-1}\ . \tag{4.16}$$

Substituting the obtained value of η into the expression of a_k, we obtain precisely the components of $\check{\mathbf{a}}(x_0)$ given by (3.5), where the components of this vector are denoted using i as index instead of k. Hence we get again Shepard formula.

References

[1] Abramovici, F., *The Shepard interpolation as the best average of a set of data*, School of Mathematical Sciences, Tel Aviv University, Israel, 1984.

[2] Backus, G.E., Gilbert, J.F., *Numerical application of a formalism for geophysical inverse problems*, Geophys. J. R. Astr. Soc., **13**(1967), 247-276.

[3] Backus, G.E., Gilbert, J.F., *The resolving power of gross earth data*, Geophys. J. R. Astr. Soc., **16**(1968), 169-205.

[4] Backus, G.E., Gilbert, J.F., *Uniqueness in the inversion of inaccurate gross earth data*, Phil. Trans. R. Soc. London ser. A, **266**(1970), 123-192.

[5] Bos, L.P., Salkauskas, K., *Moving least-squares are Backus-Gilbert optimals*, J. Approx. Theory, **59**(1989), 267-275.

[6] Courant, R., Hilbert, D., *Methods of mathematical physics*, Vol. I, Interscience Publishers, New York, 1959.

[7] Franke, R., *Scattered data interpolation: tests of some methods*, Math. Comp., **38**(1982), 181-200.

[8] Lancaster, P., Salkaukas, *Surfaces generated by moving least squares methods*, Math. Comp., **37**(1981), 141-158.

[9] Lang, S., *Undergraduate Analysis*, 1998.

[10] Levin, D., *The approximation power of moving least - squares*, Math. Comp., **67**(1998), 1517-1531.

[11] Levin, D., *Stable integration rules with scattered integration points*, J. Comp. Appl. Math. **112**(1999), 181-187.

[12] Shepard, D., *A two-dimensional interpolation function for irregularly spaced data*, Proc. 23-rd Nat. Conf. A.C.M., 1968, 517-523.

[13] Schumaker, L.L., *Fitting surfaces to scattered data*, In: Approximation Theory ed. by G.G. Lorentz, C. K. Chui and L. L. Schumaker, **II**(1976), 203-268.

Flavian Abramovici
University of Tel-Aviv
School of Mathematical Sciences
Ramat-Aviv, Israel
e-mail: `c35@post.tau.ac.il`