ISSN No. 0737-5352-55

Derivation and Analysis of a Computationally Efficient Discrete Backus-Gilbert Footprint-Matching Algorithm

Philip J. Stephens and Andrew S. Jones

Research Supported by the DoD Center for Geosciences/Atmospheric Research at
CSU under Copperative Agreement # DAAL01-98-0078
with the Army Research Laboratory

CIRA Cooperative Institute for Research in the Atmosphere

Colorado State University

ISSN No. 0737-5352-55

Derivation and Analysis of a Computationally Efficient Discrete Backus-Gilbert Footprint-Matching Algorithm

Philip J. Stephens and Andrew S. Jones

Research Supported by the DoD Center for Geosciences/Atmospheric Research
at CSU under Cooperative Agreement # DAAL01-98-0078
with the Army Research Laboratory

Table of Contents

ABSTRACT

A computationally efficient discrete Backus-Gilbert (BG) method is derived that is subject to minimization constraints appropriate for footprint-matching applications. The method is flexible, since computational cost can be traded for accuracy. A comparison of the discrete BG method with a non-discrete BG method shows that the new method can be 250% more efficient while maintaining the same accuracy as traditional approaches.

In addition, optimization approaches are used to further enhance the computational performance of the discretized BG method. A singular value decomposition approximation is applied that increases the computational efficiencies 43% to 106% while maintaining similar accuracies to the original discretized algorithm. Accuracies of the optimization were found to be scene dependent. In addition, alternative quadrature methods were also tested for several idealized simulated scenes. The results suggest that accuracy improvements could be made using customized quadrature methods that would be employed along known physical data discontinuities (such as along coastlines in microwave imagery data). In addition, regularization behaviors are also discussed; with a particular emphasis on the extension of the method for use with unnormalized gain functions. This work demonstrates that for some gain function configurations local biases can be intrinsic to the system.

The flexibility of the discrete BG method allowed for several of the optimizations to be performed in a straightforward manner. Many additional optimizations are likely possible. Due to the lower computational cost of the method, this work is applicable toward applications in which noise may vary dynamically (such as in RFI-contaminated environments). The computational flexibility of the method also makes it well suited to computationally constrained problems such as 4D data assimilation of remote sensing observations.

1. Introduction

The Backus-Gilbert (BG) method (Backus and Gilbert, 1967; 1968; 1970) has been employed by various authors to spatially co-register and invert various data sets while accounting for different spatial and error propagation behaviors (Twomey 1996). In particular, the earth science remote sensing community has employed BG for footprint matching between various satellite data channels (usually in the microwave spectrum) (Stogryn 1976; Poe 1990; Farrar and Smith 1992; Robinson et al. 1992; Bennartz 2000). Fundamentally, this is because microwave sensors typically use a single large antenna across a relatively wide spectrum. Depending on the particular sensor configuration (e.g., sampling rates and illumination characteristics), this results in footprints that may overlap considerably and have spatial half-power beam-widths that are substantially different from one another. The reader is referred to Poe (1990) for a more detailed technical explanation of the instrument design constraints that lead to this common microwave sensor condition.

The Defense Meteorological Satellite Program (DMSP) Special Sensor Microwave/Imager (SSM/I) instrument is an example of a sensor that exhibits this type of behavior (see Table 1 (Hollinger et al., 1990)). For this sensor, the ratio of the effective-field-of-views (EFOVs) at 85H GHz and 19H GHz is 4.6 × 3.3 (along-track and cross-track, respectively) between co-located sensor footprints. This can create a substantial problem with regards to the correct interpretation of the brightness temperatures since nonuniform fields can exist within a single footprint. The challenge of matching various data footprints will tend to increase as the number of coincident channels is increased. Stogryn (1976) and Poe (1990) provide additional discussion and graphical depictions

that might be helpful to readers unfamiliar with the application of BG theory to the footprint-matching problem. In the interest of brevity, this paper does not attempt to replicate that discussion.

The original work of BG (1970) provides a rigorous mathematical basis for the inversion of inaccurate data. Later Stogryn (1976) applied it to the specific problem of microwave footprint matching, and further developed concepts from BG that are the basis of most BG footprint-matching applications today. A key feature of the BG method is that it can be used effectively to trade instrument noise for spatial resolution and vice versa. This flexibility is a fundamental strength of the BG approach. The earth science remote sensing community has expanded upon the original BG theory through various specific applications. However, other research communities have implemented various versions and modifications of the original BG method for spatial data analysis work. In particular, while initially not very well known among mathematicians (Kirsch et al. 1988), the mathematics community has subsequently expanded the body of knowledge in this area through various techniques (e.g., Snieder 1991; Hansen 1994) some of which will be discussed and applied later in this paper. Specifically, this paper combines the work of Stogryn (1976) and Poe (1990) with that of Hansen (1994).

This work has implications for the practical utilization of BG methods within the earth sciences community. For example, a long-standing problem with the application of BG to earth science remote sensing has been the computational expense of calculating the coefficients necessary for the method (Galantowicz and England 1991). Current applications typically assume that the sensor and noise contributions are stable and that the coefficients can therefore be assumed static. However, in an era of increasing radio frequency interference (RFI), the relatively benign radiometric operating conditions that

the remote sensing community has enjoyed may be part of a passing era. Thus, methods that are more dynamic are needed to cope with such possible changes that threaten the performance of more traditional BG implementations. Additionally, certain computationally intensive applications involving remote sensing (e.g., 4D data assimilation) may impose demanding computational restrictions that traditional BG methods are not able to accommodate. In this vein, a new discrete BG (DBG) method is created which is computationally more efficient, and operationally flexible in its configuration. In the new method, it will be shown that computational performance can be dynamically traded for method accuracy. This allows the method to expend CPU cycles where the spatial data analysis is most critical and vice versa.

Our purpose in this paper is to introduce the mathematical theory and optimization techniques that allow for this type of flexibility and efficiency, and to apply it to several simulated data set series. The impact of several integration quadrature rules is also explored. In addition, aspects of the regularization behaviors are discussed.

2. Theoretical Background and Derivation

As mentioned previously, the new BG method is based on Stogryn (1976) and Poe (1990) using the approach of Hansen (1994) (hereafter referred to as S76, P90, and H94). Before deriving the new method, a common mathematical framework is introduced below to reconcile the notational differences employed by the former works.

All of the methods find a set of coefficients, **a**, that provide a linear combination of the measurements to estimate the value at a point, x_0. These coefficients are used to define an averaging kernel or interpolated gain function,

$$A(x, x_0) = \sum_{i=1}^{M} a_i(x_0) G_i(x), \tag{1}$$

where $G_i(x)$ is the gain function centered at x_i, evaluated at position, x, and i is the i^{th} measurement out of M total measurements. If it were possible, the ideal interpolated gain function would be a delta function. The interpolated gain function is used to estimate a measurement at a point x_0, $T_{est}(x_0)$, by convolving it with the spatially-integrated measurement, $T_{meas}(x)$, so that,

$$T_{est}(x_0) = \int A(x, x_0) T_{meas}(x) dx . \tag{2}$$

As an approximation to a typical microwave antenna beam pattern, this paper uses a truncated cosine function as the gain function. A 3D plot of a truncated cosine function, $G_{ij}(\mathbf{x}) = \cos(\|\mathbf{x}\|)$ for $0 \le \|\mathbf{x}\| \le \pi/2$, is shown in Figure 1. In practice, the pre-launch antenna gain patterns, or empirically estimated antenna gain patterns would be employed. However, the idealized gain function suffices for the method intercomparison analysis contained in this work.

a. Backus Gilbert Method

It is shown in Backus and Gilbert (1970) (equation 4.20), that the coefficients we are seeking, a_i, are given by,

$$\mathbf{a} = \frac{\mathbf{S}^{-1}\mathbf{u}}{\mathbf{u}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{u}}, \tag{3}$$

where **S** is a $M \times M$ matrix, and **u** is a vector of length M, with the condition

$$\mathbf{a} \cdot \mathbf{u} = 1 \tag{4}$$

imposed to keep the averaging kernel, A, unimodular (integration area equal to 1), i.e.,

$$\int A\,dx = 1. \tag{5}$$

For (1) to be true, (i.e., its linear independence maintained with regards to the gain function, $G_i(x)$), **S** must be symmetric and positive definite. It should be pointed out that if G_i were a delta function, then at point x_i, $a_j = 0$ for all $j \neq i$. In Backus and Gilbert (1970), **S** and **u** are defined as

$$S_{ij} = \int G_i(x) G_j(x) J(x, x_0)\,dx, \tag{6}$$

and

$$u_i = \int G_i(x)\,dx, \tag{7}$$

where S_{ij} is the i^{th} and j^{th} element of a $M \times M$ matrix, M is the total number of measurements in both spatial directions, u_i is the i^{th} element of a vector of length M, and J is the criteria function, or penalty function. This function adjusts how much influence or weight nearby points have on the solution. In the original BG work, J was set to be $(x_0 - x)^2$. It was later changed to an arbitrary function in both Hansen's and Stogryn's work. The larger J is at a point, the less that value is weighted in the solution (due to the inverse

of **S**). P90 set J to unity in order to put all the points on an equal footing (i.e., no sidelobe distortions or corrections are used).

b. Hansen Method

H94 discretizes the BG method using the simple rule:

$$\int G_i(x)\,dx \approx \sum_{k=1}^{N} w_k G_i(x_k)\,, \tag{8}$$

where N is the number of discrete integration intervals, and w_k are the integration weights. Hansen does not necessarily define x_k to be the same set of points as x_i used in the kernel or gain function, G_i. In our analysis, it was found that a simple trapezoidal rule for the weights worked well for a grid of points. The set of points, x_k, can be adjusted to increase the accuracy or speed. For real data observations that contain repetitive or known spatial data distributions, it is possible that a more sophisticated technique for determining the integration weights would be beneficial. Discretization of the BG integration space is accomplished by substituting equation (8) into the **S** matrix (6) and **u** vector (7) BG definitions. The discrete BG definitions of **S** and **u** then take the following form,

$$S_{ij} = \int G_i(x) G_j(x) J(x,x_0)\,dx \approx \sum_{k=1}^{N} G_i(x_k) G_j(x_k) J(x_k,x_0)\, w_k\,, \tag{9}$$

and

$$u_i = \int G_i(x)\,dx \approx \sum_{k=1}^{N} G_i(x_k) w_k\,. \tag{10}$$

If an identity vector of length N is defined, **e**, then the summations over k can be forced so that a matrix element form is obtained, so that

$$S_{ij} = G_{ik} G_{jk} (J(x_0))_{kk} W_{kk} e_k , \tag{11}$$

and

$$u_i = G_{ik} W_{kk} e_k . \tag{12}$$

where the following H94 matrix element definitions are used,

$$(G)_{ij} = G_i(x_j) , \tag{13}$$

$$(J(x_0))_{ij} = J(x_j, x_0)\delta_{ij} , \tag{14}$$

and

$$(W)_{ij} = w_j \delta_{ij} , \tag{15}$$

where δ_{ij} is the Kronecker delta, defined as:

$$\delta_{ij} = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases} . \tag{16}$$

Since **J** and **W** are diagonal matrices by definition ((14) and (15)), (11) and (12) can be expressed in matrix notation as

$$\mathbf{S} = \mathbf{G}\,\mathbf{J}(x_0)\,\mathbf{W}\,\mathbf{G}^{\mathrm{T}} , \tag{17}$$

and

$$\mathbf{u} = \mathbf{G}\,\mathbf{W}\,\mathbf{e} , \tag{18}$$

where **S** is a $M \times M$ matrix, **G** is a $M \times N$ matrix, **J** and **W** are $N \times N$ diagonal matrices, and **e** and **u** are vectors of length N and M respectively.

The last change that is made by H94 is to apply the extremum condition described in Appendix A. This leads to the following solution for **a** at position x_0 (combining (17) and (18) with (3)),

$$\mathbf{a}(x_0) = \frac{(\mathbf{S} + \lambda^2 \mathbf{I})^{-1} \mathbf{G}\mathbf{W}\mathbf{e}}{(\mathbf{G}\mathbf{W}\mathbf{e})^{\mathrm{T}} (\mathbf{S} + \lambda^2 \mathbf{I})^{-1} \mathbf{G}\mathbf{W}\mathbf{e}}, \tag{19}$$

where λ is a chosen regularization parameter. Appendix B contains a discussion from Lagendijk and Biemond (1991) on the constraints on λ when error minimization is considered. However, in practice, numerical techniques are employed to find an optimal λ. Since **S** may be singular, depending on the choice of $\mathbf{J}(x_0)$, selection of an optimal λ is a matter of iteratively converging to a value which provides error minimization.

In the results of this paper, the optimal λ was selected numerically. An arbitrary point was selected, x_λ, and an iterative minimization technique was used to minimize the square of the difference between the estimate and the truth at x_λ. The optimal λ is then applied globally. The estimated brightness temperature is not very sensitive to the λ parameter, so determining an optimal λ to within a few hundredths was sufficient. For each simulation, λ was recalculated, since it is dependent on the input data, and the number of integration and measurement points, N and M.

c. Stogryn Method

S76 adds another criterion to the BG method. Stogryn presents the idea of minimizing, with respect to **a**, another function of the form,

$$Q_0 = \int \left[\sum_{i=1}^{M} a_i G_i(x) - F(x, x_0) \right]^2 J(x, x_0) dx, \tag{20}$$

with a normalization contraint,

$$\int \sum_{i=1}^{M} a_i G_i(x) dx = 1, \tag{21}$$

where F is a feature extraction function chosen to produce the desired analysis (the reader is referred to S76 for discussion related to the feature extraction function's use and design). S76 also introduces an optional additional minimization constraint on the noise amplification so that resolution and noise are both minimized simultaneously to various degrees by varying a parameter, γ. For this work, the case where $\gamma = 1$ is used to simplify the intercomparison analysis. Physically this corresponds to a pure resolution minimization constraint. The inclusion of $\gamma \neq 1$ would require the use of a generalized SVD optimization approach (H94), and is discussed in section 0 as possible future work that could be performed. Thus, the functional form represented by (20) suffices for this work. Unless otherwise specified, F was set equal to a normalized uniform function within the interval of interest. This choice of F creates a uniformly spaced regular grid of square spatial averages that are similar in concept to a numerical weather prediction grid-volume average (e.g., see Pielke 1994).

This leads to the condition, $\int \sum_{i=1}^{M} a_i G_i(x) G_j(x) J(x, x_0) dx = \int G_j(x) F(x, x_0) J(x, x_0) dx$.(22)

Equation (22) can be expressed more compactly by defining a vector, **v**, with the following elements, $v_j = \int G_j(x) F(x, x_0) J(x, x_0) dx$, (23)

substituting (6) into (22), and rearranging the summation and integration order, so that the following form is achieved,

$$\sum_{i=1}^{M} a_i \int G_i(x) G_j(x) J(x, x_0) dx = \sum a_i S_{ij} = v_j \,. \tag{24}$$

Solving (24) for **a** in matrix form produces

$$\mathbf{a}_{\mathrm{St}} = \mathbf{S}^{-1}\mathbf{v}\,, \tag{25}$$

which is the new condition of the Stogryn minimization. The final solution is a combination of BG coefficients (3), $\mathbf{a}_{\mathbf{BG}}$, and Stogryn coefficients (25), $\mathbf{a}_{\mathbf{St}}$, so that $\mathbf{a}\cdot\mathbf{u}=1$ (4) is valid. The result is of the form $\mathbf{a}=\mathbf{a}_{\mathrm{BG}}+\mathbf{a}_{\mathrm{St}}+f(\mathbf{u},\mathbf{v},\mathbf{a}_{\mathrm{st}},\mathbf{a}_{\mathrm{BG}})$. Taking $\mathbf{a}\cdot\mathbf{u}$ and realizing that the first term is already equal to 1, then $\mathbf{a}_{\mathrm{St}}\cdot\mathbf{u}=-f(\mathbf{u},\mathbf{v},\mathbf{a}_{\mathrm{st}},\mathbf{a}_{\mathrm{BG}})\cdot\mathbf{u}$ must also be true, which results in the following expression for **a**,

$$\mathbf{a}=\mathbf{a}_{\mathrm{BG}}+\mathbf{a}_{\mathrm{St}}-(\mathbf{u}\cdot\mathbf{a}_{\mathrm{St}})\,\mathbf{a}_{\mathrm{BG}}\,. \tag{26}$$

Substituting (3) and (25) into (26) produces the BG coefficients with the Stogryn minimization constraints imposed (i.e., equations (20) and (21)),

$$\mathbf{a}(x_0)=\mathbf{S}^{-1}\left[\mathbf{v}+\frac{1-\mathbf{u}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{v}}{\mathbf{u}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{u}}\mathbf{u}\right]. \tag{27}$$

d. Modified Hansen Method with the Stogryn Minimization Constraints

The Hansen method uses the approximation given in (8) to reduce the BG method into a product of vectors and matrices. Using the same approximation as H94, and the same conditions that S76 and P90 impose (i.e., (20) and (21)), a modified set of coefficients, a_i, can be defined.

Applying the Hansen discretization of the integration space (8) to the Stogryn minimization constraint (20) yields,

$$Q_0 = \sum_{k=1}^{N} \left[\sum_{i=1}^{M} a_i G_i(x_k) - F(x_k, x_0) \right]^2 J(x_k, x_0) w_k \,. \tag{28}$$

This is minimized with respect to **a**, which results in the following relationship,

$$\sum_{i=1}^{M} a_i \sum_{k=1}^{N} G_{kk} w_k (J(x_0))_{kk} G_{ik} = \sum_{k=1}^{N} G_{kk} (J(x_0))_{kk} (F(x_0))_{kk} w_k \,, \tag{29}$$

where the previous Hansen definitions are used (13–15) with the addition of the following feature extraction term definition,

$$(F(x_0))_{ij} = F(x_j, x_0) \delta_{ij} \,. \tag{30}$$

We now define matrices for **F**, **G**, **J**, and **W**, as before, where the new matrix, **F**, is defined as a $N \times N$ diagonal matrix (30). Thus, (29) becomes

$$\mathbf{S}\,\mathbf{a} = \mathbf{v} \,, \tag{31}$$

where **S** and **v** are defined as,

$$\mathbf{S} = \mathbf{G}(\mathbf{J}(x_0))\,\mathbf{W}\,\mathbf{G}^{\mathrm{T}} \,, \tag{32}$$

and

$$\mathbf{v} = \mathbf{G}\,\mathbf{J}\,\mathbf{F}\,\mathbf{W}\mathbf{e} \,. \tag{33}$$

The same extremum condition for **a** is applied (see also Appendix A). However, **u** is replaced by **v** and β is set to 1 to match the S76 form, given by (25). This yields the following solution for **a** with the extremum condition,

$$\mathbf{a}_{\mathrm{St}} = (\mathbf{S} + \lambda^2 \mathbf{I})^{-1} \mathbf{v} \,. \tag{34}$$

Again, we must ensure a proper combination of the BG coefficients and the Stogryn coefficients are used so that $\mathbf{a} \cdot \mathbf{u} = 1$ (4) is valid. The combination that satisfies these conditions is given by (26). Therefore, the modified Hansen BG coefficients with the Stogryn minimization constraints can be written in matrix form as

$$\mathbf{a}(x_0) = \mathbf{M}\left[\mathbf{v} + \frac{1 - \mathbf{u}^{\mathrm{T}}\mathbf{M}\mathbf{v}}{\mathbf{u}^{\mathrm{T}}\mathbf{M}\mathbf{u}}\mathbf{u}\right], \tag{35}$$

where

$$\mathbf{M} = (\mathbf{S} + \lambda^2 \mathbf{I})^{-1}. \tag{36}$$

As will be seen in the next section, this form ((35) and (36)) has several desirable features; in particular, the diagonal matrices allow for some rather straightforward gains in computational efficiency. As discussed by H94, this form can also lead to an alternative fast approximate implementation under certain conditions. However, for the remainder of this paper we will focus on the more general capabilities of the new method for footprint-matching applications. The new method is referred to as the DBG method throughout the remainder of this paper.

3. Method Intercomparison Results

The S76 BG method and the DBG method are applied to three simulated truth scenes to explore the various behaviors of the methods. The truth scenes are a) a constant temperature, $T = 200$ K, b) the same as (a) but with $T = 300$ K, for all $x > 0$, and c) the same as (a) but with $T = (200 + 50 \sin x)$ K, for all $x > 0$. The truth scenes were chosen to represent various extremes. Scene A represents the simplest case, and tests the methods for correct implementation details that manifest themselves as biases and random noise behaviors. Scene B is the most harsh example, in that a large discontinuity is present in the scene. The range of 100 K is representative of the magnitude of the brightness temperature discontinuity observed when transitioning from ocean to land surfaces in the littoral regions. As will be discussed later, this case exhibits some of the more significant differences in behavior. The last idealized scene, Scene C, represents a contrast between a homogeneous region and a region with significant spatial variations. This scene, while still significantly idealized, presents a stronger challenge to the BG methods than Scene A, and yields insights into the abilities of the methods to perform in regions where periodicity of the scene may be important.

In the following simulations, a hypothetical microwave sensor is assumed to have a known antenna gain function that is represented using a truncated cosine function (Figure 1). The overlap between EFOVs is determined by the maximum extent of the truncated cosine function which is defined to be $\pi/2$. This results in a small overlap between adjacent FOVs, since each measurement is "observed" at each whole interval position (e.g., $k = 1, 2, \ldots, M$). It should be noted that the design and configuration of the hypothetical sensor used in the following simulations is not the focus of this work. The hypothetical sensor is merely meant to exercise the various BG methods in a manner in

which the methods themselves can be objectively evaluated. Particular sensor configurations may perform better or worse than the results shown here. However, the general conclusions regarding the BG method behaviors should be transferable to other sensor configurations.

e. Stogryn Method Results

The Stogryn method (as discussed in Section c) is applied to the three simulated truth scenes (A – C) where 5 K of random noise has been added to the truth scene, thus simulating a relatively noisy hypothetical microwave sensor. The estimated temperatures from the Stogryn method are shown in Figure 2a-c, using a penalty function of $J(x,x_0) = 1$, which is a reasonable choice for satellite footprint-matching applications (S76; P90). In all of the following simulations, the feature extraction function, F, is set to be the gain function centered at x_0. In Figure 2d-f, the results are shown for the case when the penalty function is $J(x,x_0) = (x - x_0)^2$. Simulation results are presented in Table 2 for Scenes A – C as root-mean-squared (RMS) differences from the "truth" scene (i.e., the scene values before the addition of 5 K of random noise). Standard deviations of the RMS values are calculated by aggregating results from several simulation instances. Aside from the conclusion that the method is performing correctly for all simulated scenes, there is an obvious scene distinction in regards to the RMS performances (Table 2). For example, the Scene A Stogryn method estimate (with $J(x,x_0) = 1$) exhibits an RMS of 0.71 K, while the RMS results from Scenes B and C are 12.2 K, and 4.4 K, respectively. Thus, results from Scene A are considerably better in terms of RMS performance. The reason for this behavior is that all points in Scene A share a common mean. The theoretical justification for this statement will be discussed in more detail in

the lateral boundary condition discussions that follow. The RMS performance for Scene A is substantially lower than the initial 5 K random noise component that was added; hence the spatial overlap of the gain function is working to the method's advantage. In comparison to the other scenes (see Table 2), Scene A is the easiest since it has no spatial structures, Scene B is the most difficult to estimate correctly since it has a single large discontinuity at $x = 0$, and Scene C is the next most difficult case of the three. Scene C is not as difficult as Scene B since the discontinuity at $x = 0$ within Scene C has a finite slope. The periodicity of Scene C for $x > 0$ is well represented by the Stogryn method.

The Stogryn method results for $J(x,x_0) = (x - x_0)^2$ (Figure 2d-f) are similar to the $J(x,x_0) = 1$ simulations (see also Table 2). A major distinction between the results is the lateral boundary condition (LBC) behaviors. In particular, Scene B shows much improved LBC behaviors for the $J(x,x_0) = (x - x_0)^2$ case, and Scene C exhibits a slight improvement for only the right-most LBC estimate. This is exhibited in the RMS values as well (Table 2), with a 4 K RMS improvement for Scene B, which is the extreme example. However, the RMS results for Scene C were slightly worse. This behavior suggests that a modified J functional form could be used as the data approach the edges of a scan. For example, the center of the scan could implement the $J(x,x_0) = 1$ form and transition to the $J(x,x_0) = (x - x_0)^2$ form near the edges. An alternate solution would be to discard the data near the LBC if $J(x,x_0) = 1$.

f. Discrete Backus-Gilbert Method Results

Similarly, the DBG method results using penalty functions of $J(x,x_0) = 1$ and $J(x,x_0) = (x - x_0)^2$ are presented in Figure 3 (corresponding to Figure 2). Again, a simple comparison demonstrates that the DBG method is also able to estimate the truth scenes

with some fidelity. The LBC behaviors in Figure 3 are a significant issue for the $J(x,x_0) = 1$ case, as was shown for the Stogryn results in Figure 2. Like the Stogryn LBC behaviors, the DBG LBC effect is minimized for Scenes A and B with the choice of $J(x,x_0) = (x - x_0)^2$. This suggests that this is a fundamental BG behavior with the Stogryn minimization constraints. An RMS comparison between the DBG method and the Stogryn method (Table 2) shows that the DBG method is outperforming the Stogryn method for the simplest scenes (A and C), and is underperforming it for the most challenging scene (B). However, it should be noted that the DBG method simulations are for an integration configuration where $N = M = 50$, thus approximating the discontinuity at $x = 0$ with a single integration interval.

1) The Effect Of Lateral Boundary Conditions

As seen in Figures 2 and 3, the LBC results for the step function scene (Scene B) are not well behaved when $J(x,x_0) = 1$. The LBC artifacts can be removed by increasing the input range, but not the output range. This has the same effect as discarding the estimates near the edges of the scan. For example, if the boundaries are moved farther out (e.g., to $x = \pm 15$), the estimates are computed for the range, $x = \pm 12.5$. This procedure effectively removes the LBC issues and, for the case when $J(x,x_0) = 1$, reduces the RMS values when applied to the previous simulations (Table 3). The results for Scene B are presented in Figure 4. Clearly, the LBC effects can be a significant issue if not handled correctly. As mentioned previously, an ideal solution would be to transition the functional form of J to a non-unity-based form near the edges of the scan, thus retaining the desirable $J(x,x_0) = 1$ analysis properties for footprint-matching applications.

2) Special Lateral Boundary Condition Considerations For The $J(X,X_0) = 1$ Case

It is important to see that setting $J(x,x_0) = 1$ in the original BG derivation does not give a spatially dependent estimate of the brightness temperatures. It is seen from (6) that the **S** matrix is independent of x_0 when J is constant. This makes the coefficients, **a** (3), also independent of x_0. If the coefficients are independent of x_0, then the estimated temperature (2) will also be independent of x_0, and the BG derivation becomes nothing more then an averaging technique. The LBC artifacts arise because the $\mathbf{a_{BG}}$ term (see equation 26) gives the mean of the input while the $\mathbf{a_{St}}$ term produces the deviations from the mean, providing the correct estimations for the Stogryn minimization constraints. Again, if **S** is independent of x_0 then the only dependence for $\mathbf{a_{St}}$ comes from **v** (23). When F is defined to be the gain function, this term decreases rapidly near the boundaries. This rapid decrease results in a decrease in the $\mathbf{a_{St}}$ term. That is why the LBC estimates approach the mean for the $J(x,x_0) = 1$ cases shown in Figures 2 and 3.

3) Computational Performance vs. Accuracy Considerations

Having shown that the LBC effects are a product of the Stogryn minimization constraints applied with the condition that $J(x,x_0) = 1$, the biggest practical benefit of setting J to unity, other than its physical intuitiveness for sensor applications, is that **S** only needs to be calculated once, and can be reused for each estimate, which decreases the computational time. This optimization was not implemented for the results presented in this paper. Thus, all time intercomparisons within this paper are also generally applicable for other J definitions.

The Hansen and DBG methods are quite flexible. Since the number of integration points is adjustable, this allows for faster, but less accurate, implementations of the technique. This would be particularly useful in the presence of large signal to noise ratios, or when computational requirements are for a particularly fast implementation.

The Stogryn implementation is optimized using library routines for numerical integration and matrix inversion. The DBG method also uses the same matrix inversion routine. However, due to its diagonal matrix form, the DBG method can be further optimized to speed the matrix multiplications. The ability to exploit the diagonal matrices is one of the primary computational advantages of the DBG method.

A series of simulations were performed in which the number of integration intervals, N, was modified to determine the computational cost of the method versus the method's RMS accuracy. All results shown in Figures 5-7 have the LBC windowing filter applied to remove the LBC artifacts for the $J(x,x_0) = 1$ cases. In each of the figures, the number of measures, M, is set to 100. The number of integration points, N, is adjustable.

Figure 5 presents results for the uniform scene (Scene A). The DBG method RMS values are generally lower than the Stogryn RMS values. This is likely due to a more optimal integration pattern for the random distribution, since the DBG integration is evenly spaced. The numerically optimized integration weights used in the Stogryn method may also be amplifying the effect of the various random patterns within the data set series. The effect of the uniform placement of integration intervals is even more pronounced for the case when $J(x,x_0) = (x - x_0)^2$, where the DBG method is approximately twice as accurate as the Stogryn method (Table 2). This suggests that the DBG method's integration pattern flexibility could be used to exploit known structures or features within data sets.

The trend in the DBG results (Figure 5) is toward lower RMS values with larger N. This corresponds to improved integration accuracy with larger N. The CPU consumption for the Stogryn method is constant since its integration is not explicitly adjustable. The CPU consumption for the DBG method increases with N as $\sim N^2$. The computational costs between the methods are approximately equivalent at $N \cong 220$.

The step function of Scene B is more of a challenge for both the DBG method and the Stogryn method (Figure 6). Both the DBG RMS values and the Stogryn RMS results are greater than the original noise. The DBG RMS results are not significantly improved by going to larger values of N. As can be seen in Figure 3, most errors are concentrated at the discontinuity. It is rather obvious that the step function is the harshest test of the DBG method. The computational cost for the DBG method is approximately the same as for the simple scene (Scene A) results presented in Figure 5.

The results for Scene C (Figure 7) are similar to that of Scene B (Figure 6), but with reduced severity. The RMS ranges are below the 5 K random noise level, but Scene C RMS values are still larger than the RMS values for Scene A. Most error is again concentrated at the discontinuity at $x = 0$. The DBG method RMS tendency is to decrease with greater values of N, hence additional integration intervals are proving to be somewhat helpful with this particular scene. Again, the computational cost behaviors to process Scene C are similar to that of Scenes A and B.

4. SVD Analysis

The SVD of a general real M-by-N matrix allows the target matrix to be separated into left and right singular vectors with the definition of appropriate singular values (Anderson et al. 1999). The practical consequence of this is that it allows the matrix to be handled via summations, and additional approximations become available to increase the overall performance of the method. The following SVD derivation is a modification of the approach used by H94 to use the DBG method with the Stogryn minimization constraints.

A new matrix and vector, $\widetilde{\mathbf{G}}$ and $\widetilde{\mathbf{e}}$ are defined as,

$$\widetilde{\mathbf{G}} = \mathbf{G}\,\mathbf{J}^{1/2}\mathbf{W}^{1/2}, \tag{37}$$

$$\widetilde{\mathbf{e}} = \mathbf{J}^{-1/2}\mathbf{W}^{1/2}\mathbf{e}, \tag{38}$$

where $\widetilde{\mathbf{G}}$ is a $M \times N$ matrix, and $\widetilde{\mathbf{e}}$ is a vector of length N. This allows the following form for $\mathbf{S}$, which subsequently allows $(\mathbf{S} + \lambda^2\mathbf{I})^{-1}$ to be written in a notation more suitable for the SVD technique. Thus, (32) - (33) are now,

$$\mathbf{S} = \widetilde{\mathbf{G}}\widetilde{\mathbf{G}}^{\mathrm{T}}, \tag{39}$$

$$\mathbf{u} = \widetilde{\mathbf{G}}\widetilde{\mathbf{e}}, \tag{40}$$

and

$$\mathbf{v} = \widetilde{\mathbf{G}}\mathbf{J}\mathbf{F}\widetilde{\mathbf{e}}. \tag{41}$$

SVD is now used to decompose $\widetilde{\mathbf{G}}$ into,

$$\widetilde{\mathbf{G}} = \mathbf{B}\boldsymbol{\Sigma}\mathbf{C}^{\mathrm{T}}, \tag{42}$$

with $\mathbf{B}^T\mathbf{B} = \mathbf{I}$, $\mathbf{C}^T\mathbf{C} = \mathbf{I}$ and $\Sigma = \text{diag}(\sigma_i)$, where $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{\min(m,n)} \geq 0$ are the singular values of $\boldsymbol{\Sigma}$ (Demmel and Kahan 1990). If the noise covariance matrix differs from $\mathbf{I}$, then a generalized SVD analysis could be performed (Paige 1986; Bai and Demmel 1993; Bai and Zha 1993;Christensen-Dalsgaard et al. 1993). The SVD decomposition (42) allows the object of the matrix inversion in (36) to be written as,

$$\mathbf{S} + \lambda^2\mathbf{I} = \mathbf{B}\left(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T + \lambda^2\mathbf{I}\right)\mathbf{B}^T. \tag{43}$$

The inverse is then,

$$\left(\mathbf{S} + \lambda^2\mathbf{I}\right)^{-1} = \mathbf{B}\left(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T + \lambda^2\mathbf{I}\right)^{-1}\mathbf{B}^T. \tag{44}$$

Following H94, a portion of the numerator term of (35) is defined such that,

$$\mathbf{p}(x_0) = \left(\mathbf{S} + \lambda^2\mathbf{I}\right)^{-1}\mathbf{u} = \mathbf{B}\left(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T + \lambda^2\mathbf{I}\right)^{-1}\mathbf{B}^T\mathbf{B}\boldsymbol{\Sigma}\mathbf{C}^T\tilde{\mathbf{e}}, \tag{45}$$

where (40) and (42) have been substituted into the expression. Through SVD, (45) can be expressed as,

$$\mathbf{p}(x_0) = \sum_i \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2}\frac{\mathbf{c}_i^T\tilde{\mathbf{e}}}{\sigma_i}\mathbf{b}_i, \tag{46}$$

where $\mathbf{p}$, is a vector of length M, and $\mathbf{c}_i$ and $\mathbf{b}_i$ are the right and left singular vectors of $\tilde{\mathbf{G}}$ with length N and M, respectively, where $\mathbf{b}_i$ are the eigenvectors of $(\tilde{\mathbf{G}}\tilde{\mathbf{G}}^T + \lambda^2\mathbf{I})$. This leads to the following solution for $\mathbf{a}$,

$$\mathbf{a} = \mathbf{r}(x_0) + \frac{1 - \mathbf{p}(x_0)^T\mathbf{r}(x_0)}{\mathbf{p}(x_0)^T\tilde{\mathbf{G}}\tilde{\mathbf{e}}}\mathbf{p}(x_0), \tag{47}$$

where,

$$\mathbf{p}(x_0)^{\mathrm{T}}\tilde{\mathbf{G}}\tilde{\mathbf{e}} = \sum_i \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2}\left(\mathbf{c}_i^{\mathrm{T}}\tilde{\mathbf{e}}\right)^2, \tag{48}$$

$$\mathbf{r}(x_0) = \sum_i \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2}\frac{\mathbf{c}_i^{\mathrm{T}}\dot{\mathbf{e}}}{\sigma_i}\mathbf{b}_i, \tag{49}$$

and

$$\dot{\mathbf{e}} = \mathbf{F}\mathbf{J}\tilde{\mathbf{e}}. \tag{50}$$

Equations (46) – (50) represent the SVD form of the DBG method.

In summary, the SVD form of the DBG method converts the matrix form of the DBG method into a specialized group of summations. As will be shown later, this results in significant computational savings. It should be noted that the SVD computational savings are in addition to the computational savings that result from the inherent form of the DBG method over traditional methods. The SVD form of the DBG is an explicit optimization technique that specifically exploits the DBG diagonal matrix form.

5. SVD Results

The summations in the SVD of the DBG method can be truncated at any desired point. Figure 8 shows a sequence of results in which additional terms are progressively added to the SVD summation, for the case where M = 100. It can be seen that the effect is similar to that of a Fourier series. The first few terms give the general average and scaling of the spatial features, and additional terms refine the spatial structure of the results.

The RMS performances of the SVD results are calculated for three simulated truth scenes as before. The results (Table 4) are defined as ratios in which the SVD DBG RMS results are normalized relative to the non-SVD DBG RMS results. Corresponding standard deviations are also calculated. RMS ratio results less than one indicate an improvement using the SVD approach. The computational cost ratio is defined similarly, with the cost ratio being defined as the SVD cost relative to the non-SVD cost. The computational costs are independent of the specific scene. These results show that the RMS errors decrease as the number of summation terms is increased. Table 4 also shows a linear relationship between computational costs and the number of summation terms, which is expected due to the summation nature of each term. In practice, 20% of the terms generally gave sufficient structure, low RMS values, as well as minimizing computational costs for the particular simulated scenes explored here. However, it should be pointed out that performance was dependent on the particular scene. For the uniform scene, all SVD RMS results were significantly improved, with RMS values less than 50% of the non-SVD RMS results, while the step function scene results were only marginally improved. The sine wave scene was the most difficult scene for the SVD approach and resulted in errors that were 30% greater than the non-SVD results. The choice of where to

truncate the SVD series will in general be dependent on the specific application and its associated error tolerance requirements. It is also worth noting that if $J(x,x_0)$ is constant, then the SVD of $\widetilde{\mathbf{G}}$ only needs to be performed once. This improves the run times by a factor of 3 or more, and does not change the output.

6. Quadrature Sensitivities

In this work, the integrations in the BG method were discretized using equation (8). Since (8) uses the same set of points, x_k, for the integral approximation, and is independent of the choice of the discrete gain positions, x_i, the total summation for each gain function may actually differ slightly, though the non-discretized integral would not. This can lead to certain areas being emphasized more or less than other areas. This emphasis can be especially useful to improve the estimates at certain locations.

The quadrature points, x_k, and the quadrature weights, w_k, can be selected using simple integration estimates, such as the trapezoidal rule, or by more advanced techniques, like Gaussian quadrature, or Fejér quadrature (Davis and Rabinowitz 1975). Here we compare, both graphically and statistically, the results of using three different quadrature rules on a step function (200 K $x < 0$, 300 K $x > 0$). The gain function used is a truncated cosine function with width of 2π. The width is wider than the prior analysis configuration to enhance the region of transition from 200 K to 300 K. This allows the differences between the quadrature methods to become more pronounced.

The three different quadratures are:

1. *Trapezoidal,* d^{-1} for all points except the end points, which are $(2d)^{-1}$, where d is the separation between neighboring points. All points are equally separated.
2. *Fejér-1,* with a weight of 1: $(1-(x/b)^2)^{1/2}$, where b is the maximum value of x.
3. *Fejér-2,* with a weight of (x-a)(b-x): $(1-(x/b)^2)^{1/2} \cdot weight$, where a is the minimum value of x and b is the maximum value of x.

Graphical depictions of the three weighting functions are shown in Figure 9. It can be seen from these weighting functions that the *Fejér-2* quadrature has a much sharper peak at x = 0 than either the of the other quadrature methods.

DBG estimated temperature results using each quadrature rule are presented in Figure 10. The Fejér-2 quadrature results have a much sharper slope at the discontinuity than the other quadrature methods. This increase in slope definition is exchanged for a decrease in the accuracy of the uniform parts of the scene. The RMS results for each quadrature method are presented in Table 5 for the three idealized scenes. It is important to note that the choice of the quadrature technique can improve the RMS performance. This is significant and suggests that customized quadrature methods could be employed along known physical data discontinuities (such as along coastlines in microwave imagery data).

7. Regularization Behaviors Due to Unnormalized Gain Functions

If the gain function is unnormalized (not the averaging gain function, which must be normalized) biases are introduced into the estimated temperatures. An example of this type of bias is shown in Figure 11, where the step function truth is $T = 200$ for $x \leq 0$, and $T = 300$ for $x > 0$, and where a gain function area of 2 is used. When λ is set to 0 (Figure 11a), neither the left side nor the right side of the step reaches the value it should (200 K and 300 K respectively), instead both sides overshoot the truth. If λ is allowed to vary, the bias can be removed. For example, given the case in Figure 11, where the gain function has an area of 2, if λ is set to 4, then the results no longer contain a bias (see Figure 11b). However, if the gain function is normalized, $\lambda = 0$ removes any bias. The λ regularization parameter, therefore, acts to unbias the estimates.

The effect that different λ values have on the temperature estimates when different width gain functions are used is shown in Figure 12. It can be seen in Figure 12 that if the gain function has an area less than 1, then no matter what value λ has, there is a bias. It is also important to note that as λ increases, the estimated temperature approaches the average temperature of all the measurements. For the step case, the biases are independent of the gain function width; however, in general, this is not the case. If a sine function is used instead of a step function, then the results are dependent on the gain function width. Figure 13 contains the temperature estimate results at the local minimum of the sine function as a function of λ and of the gain function width (for reference the local minimum truth value is 150K). Every curve corresponds to a gain function area of 1, but with varied gain function widths. It can be seen that as the gain function width increases, the minimum gets pushed closer to the average (200K). This makes intuitive

sense, since the local minimum is estimated from surrounding points, and the addition of more points to the average (i.e., a wider gain function) results in the estimate being moved away from the true minimum and toward the global average. While this is an extreme example, this demonstrates that for some gain function configurations, local biases can be intrinsic to the system. This behavior will be most apparent for certain spatial structures or patterns such as the sine wave scene.

8. Conclusions

A DBG method has been created that is a modification of the H94 BG method with the Stogryn minimization constraints. Several behaviors were observed. In particular, the LBC RMS behaviors, for the $J(x,x_0) = 1$ case for both the Stogryn and the DBG methods, were explained as a consequence of the Stogryn minimization. It is suggested that alternative forms for J should be explored to minimize this effect, so that data near the scan edge is not simply discarded. A composite functional form for J in which the center of the scan implements a $J(x,x_0) = 1$ functional, and toward the edges of the domain the $J(x,x_0) = (x - x_0)^2$ form becomes dominant, was suggested as a possible candidate. Additionally, the RMS performance of each method was shown to be dependent on the individual scene structure. The DBG method with regularly spaced integration weights was able to outperform the traditional Stogryn method for a uniform scene with random noise. For the more challenging scenes, the Stogryn method performed slightly better due to the application of a more sophisticated spatial integration method.

As a practical consequence, the DBG method can be employed as a less costly footprint-matching algorithm. If necessary to increase the overall method's performance, modifications to the distribution and number of integration weights could be made. In a similar vein, if conditions warrant, the Stogryn method could be applied for structures that are particularly difficult to spatially resolve. Additional work is needed to examine the criteria and conditions for when such a switch between methods should be made.

The DBG method was also optimized using a singular value decomposition method. The result is increased computational efficiencies between 43% and 106% while maintaining similar accuracies. By incorporating the Stogryn γ parameter (Stogryn 1976),

the optimization approach is also be able incorporate the simultaneous minimization of resolution and noise; however, doing so would require using a generalized SVD approach, since several key matrices are no longer purely diagonal, and become interdependent on the gain function. In this paper, the additional complexity of a GSVD method was avoided, but it deserves future exploration.

The examination of several alternative quadrature methods demonstrates the sensitivity of the DBG method to the choice of the quadrature rules employed, and suggests that further optimization methods may yield additional increased performance, if some a priori information exists regarding the overall scene structure. An example of where these a priori conditions may exist include microwave imagery data sets in littoral zones where a high signal contrast exists due to the radiometric behaviors of the geographic features.

The regularization sensitivity studies indicate the potential for improper regularization when used with unnormalized gain functions. This is a caution worth noting, in that data analysis operators used for data assimilation purposes may diverge from traditional remote sensing footprint-matching gain function behaviors, which are commonly normalized functions (Stogryn, 1976; Poe, 1990). This work also demonstrates a formalism to identify and constrain such possible biases. In practice, when using unnormalized gain functions, it is suggested that simulations be used to verify the accuracy of the method with the particular gain function. Developing a formalism that rigorously considers all possible functional forms for the gain function was beyond the scope of this work.

The flexibility of the DBG method allows it to trade computational cost for accuracy in most conditions, thus lending it to several challenging research application areas. In

particular, the use of a more flexible method would serve well in applications where the BG coefficients need to be routinely recalculated, depending on conditions, such as in a RFI contaminated environment. The flexibility of the DBG method also allowed for several of the optimizations to be performed in a rather straightforward manner. Many additional optimizations are likely possible.

Future work will include investigating additional computational enhancements and testing the scope and validity of those assumptions. Additional research will also be conducted to exploit the flexibility of the DBG Method for dynamic resolution and noise adjustments based on the signal to noise content. The new work will be applicable toward dynamic RFI minimization techniques, and dynamic enhancement of littoral features (based on content). The computational flexibility of the DBG method is well suited for computationally constrained problems such as 4D data assimilation. Our immediate plans include the incorporation of a temporal analysis component into the DBG spatial filter for use within a functional data assimilation system to investigate the remaining practical implementation challenges specific to data assimilation use. It is our hope that the flexible framework of the DBG method will foster additional BG implementations that resolve some of these remaining challenges.

References

Anderson, E., Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, D. Sorensen, 1999: *LAPACK User's Guide,* SIAM, Philidelphia, pp. 407.

Backus, G., and F. Gilbert, 1967: Numerical applications of a formalism for geophysical inverse problems, *Geophys. J. R. Astron. Soc., 13,* 247-276.

Backus, G., and F. Gilbert, 1968: The resolving power of gross Earth data, *Geophys. J. R. Astron. Soc., 16,* 169-205.

Backus, G., and F. Gilbert, 1970: Uniqueness in the inversion of inaccurate gross Earth data, *Philos. Trans. Roy. Soc. London Ser., 266, No. 1173,* 123-192.

Bai, Z., J. W. Demmel, 1993: Computing the generalized singular value decomposition, *SIAM, J. Sci. Comp.,* **14,** 1464-1486.

Bai, Z., H. Zha, 1993: A new preprocessing algorithm for the computation of the generalized singular value decomposition, *SIAM J. Sci. Comp.,* **14,** 1007-1012.

Bennartz, R., 2000: Optimal Convolution of AMSU-B to AMSU-A, J. *Atmos. Oceanic Technol., 17,* 1215-1225.

Christensen-Dalsgaard, J., P. C. Hansen, and M. Thompson, 1993: GSVD analysis of heiloseismic inversions, *Mon. Not. R. Astron. Soc.,* **264,** 541-564.

Davis, P. J., P. Rabinowitz, 1975: *Methods of Numerical Integration,* pp. 459, Orlando, FL, Academic Press.

Demmel, J. W., and W. Kahan, 1990: Accurate singular values of bidiagonal matricies, *SIAM J. Sci. Stat. Comput.,* **11,** 873-912.

Farrar, M. R., and E. A. Smith, 1992: Spatial resolution enhancement of terrestrial features using deconvolved SSM/I microwave brightness temperatures, *IEEE Trans. Geosci. Remote Sens., 30,* 349-355.

Galantowicz, J. F, and A. W. England, 1991: The Michigan Earth Grid: description, registration method for SSM/I data and derivative map projections, *Rep. 027396-2-T,* 19 pp., Radiation Laboratory, Univ. of Michigan, Ann Arbor, Michigan.

Hansen, P., 1994: The Backus-Gilbert method: SVD analysis and fast implementation, *Inverse Problems, 10,* 895-904.

Hollinger, J. P., J. L. Peirce, and G. A. Poe, 1990: SSM/I instrument evaluation. *IEEE Trans. Geosci. Remote Sensing, 28,* 781–790.

Kirsch, A., B. Schomburg, and G. Berendt, 1988: The Backus-Gilbert method, *Inverse Problems, 4,* 771-783.

Lagendijk, R., and Biemond J, 1991: *Iterative Identification and Restoration of Images*, pp. 208, Boston, MA, Kluwer Academic Publishers.

Paige, C., 1986: Computing the generalized singular value decomposition, *SIAM J. Sci. Stat.,* **7,** 1126-1146.

Pielke, R. A., 1994: *Mesoscale Meteorological Modeling,* pp. 612, Orlando, FL, Academic Press.

Poe, G., 1990: Optimum interpolation of imaging microwave radiometer data, *IEEE Trans. Geosci. Remote Sens., 28,* 800-810.

Robinson, W. D., C. Kummerow, and W. S. Olson, 1992: A technique for enhancing and matching the resolution of microwave measurements from the SSM/I, *IEEE Trans. Geosci. Remote Sens., 30,* 419-428.

Snieder, R., 1991: An extension of Backus-Gilbert theory to nonlinear inverse problems, Inverse Problems, 7, 409-433.

Stogryn, A., 1976: Estimates of brightness temperatures from scanning radiometer data, *IEEE Trans. Antennas Propagat., AP-26, No. 5,* 720-726.

Twomey, S., 1996: *Introduction to the Mathematics of Inverse Remote Sensing and Indirect Measurements,* pp. 160-169, Mineola, NY: Dover Publications, Inc.

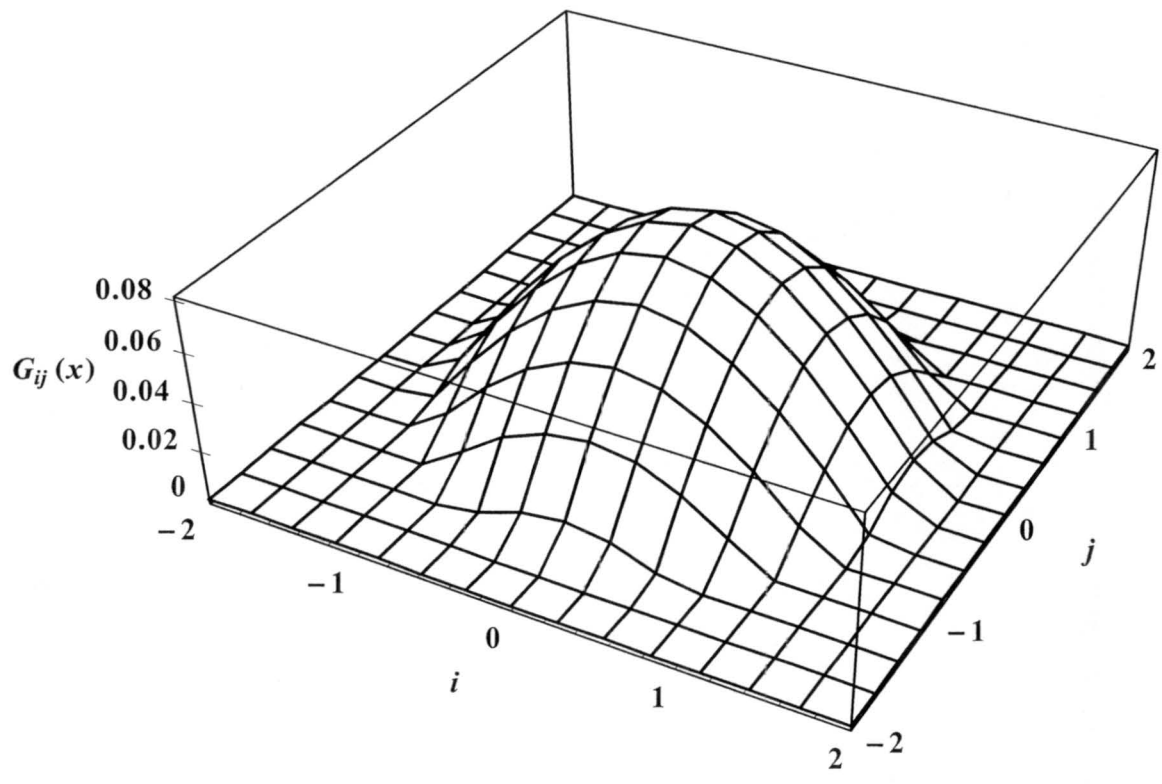

FIG. 1. 3-D plot of a truncated cosine function, $G_{ij}(\mathbf{x})$, within a 2D measurement position space, **x**.

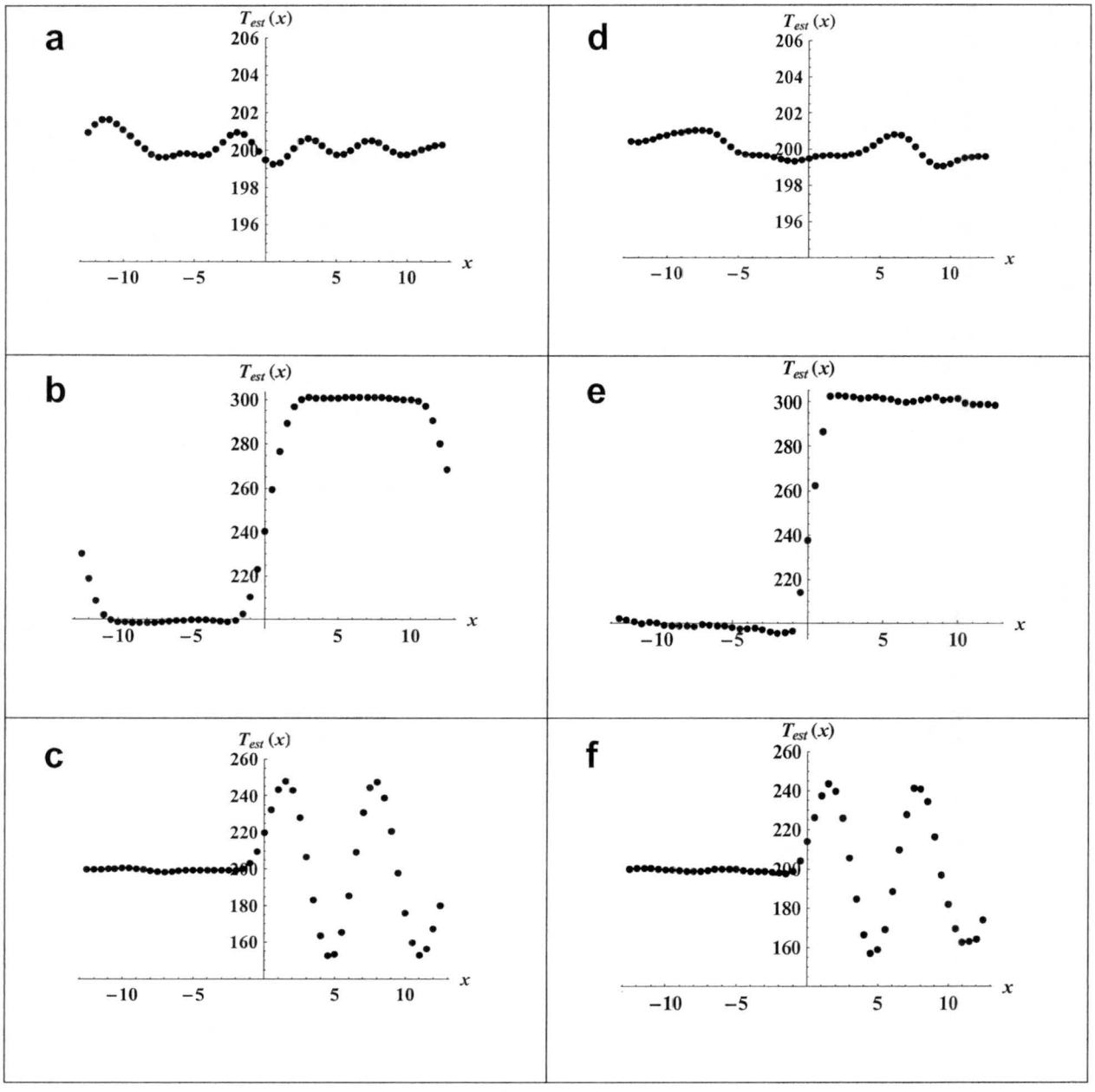

FIG. 2. Three scenes as estimated by the Stogryn method with $J(x,x_0) = 1$, with a random measurement error of 5 K imposed on the "true" scene functions which are: a) a constant temperature, $T = 200$ K, b) the same as (a) but with $T = 300$ K, for all $x > 0$, and c) the same as (a) but with $T = (200 + 50 \sin x)$ K, for all $x > 0$, d), e), and f) are the results that correspond to a), b) and c), respectively, except that $J = (x - x_0)^2$.

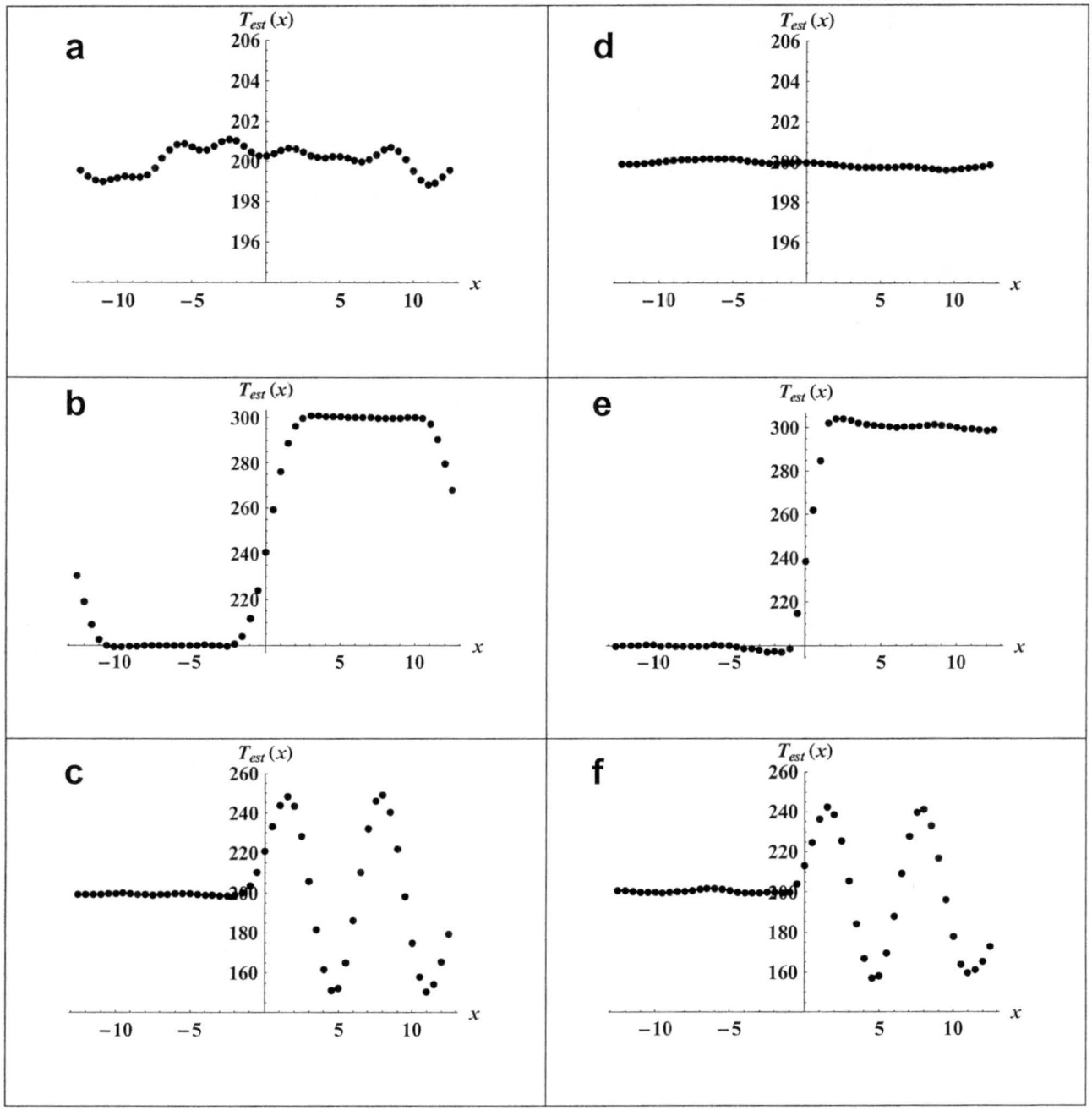

FIG. 3. The same as Figure 2, except as estimated by the DBG method, where $N = M = 50$.

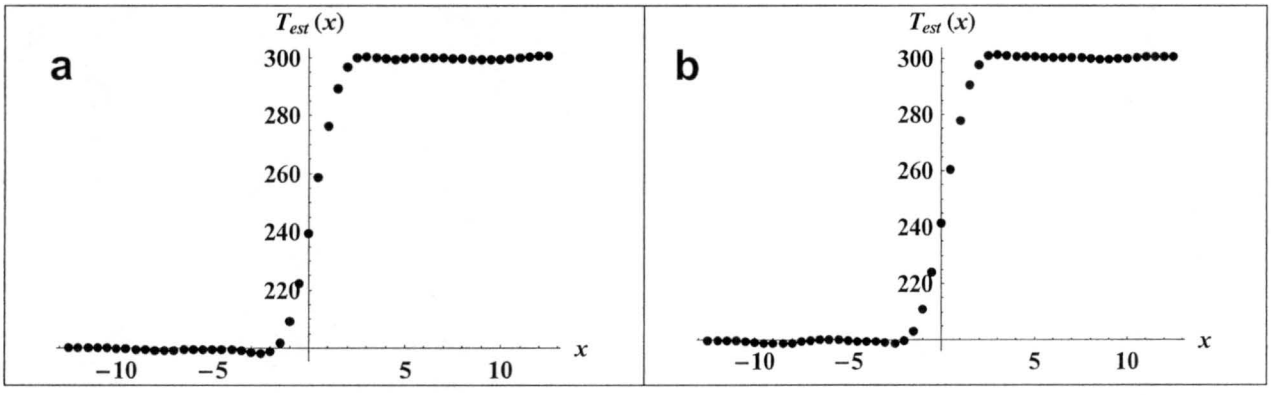

FIG. 4. The same as panel (b) of Figures 2 and 3, except a) is the DBG method, and b) is the Stogryn method. The results shown have the lateral boundary conditions moved farther away from the region of analysis by the filtering procedure discussed in the text.

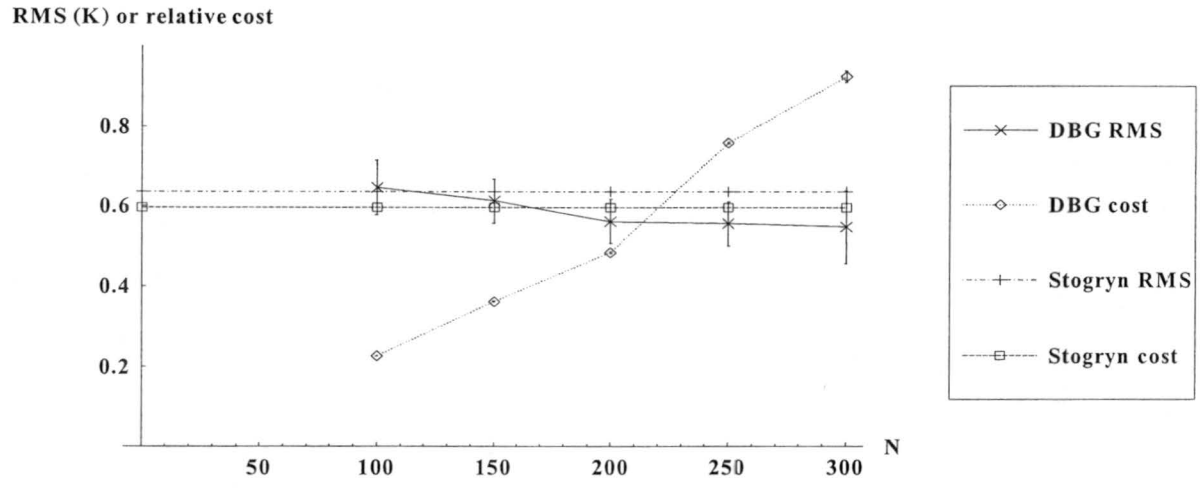

FIG. 5. This is the RMS performance and relative computational costs for the uniform scene simulations (Scene A) for the DBG method (DBG) and the Stogryn method versus the number of integration points, N, for the case where $J = 1$. The lateral boundary condition filter described in the text has been applied to the results shown.

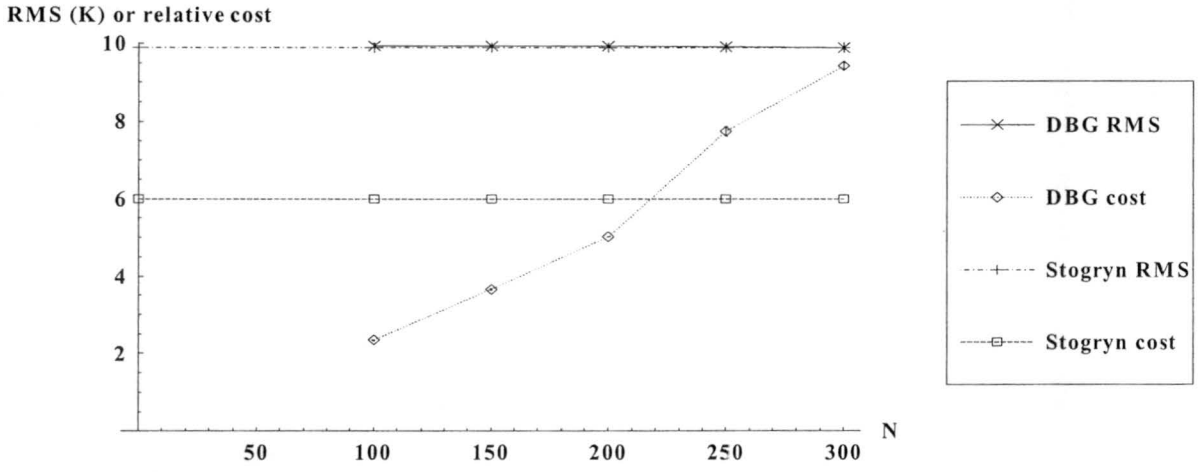

FIG. 6. Same as Figure 5, except for the step function scene (Scene B).

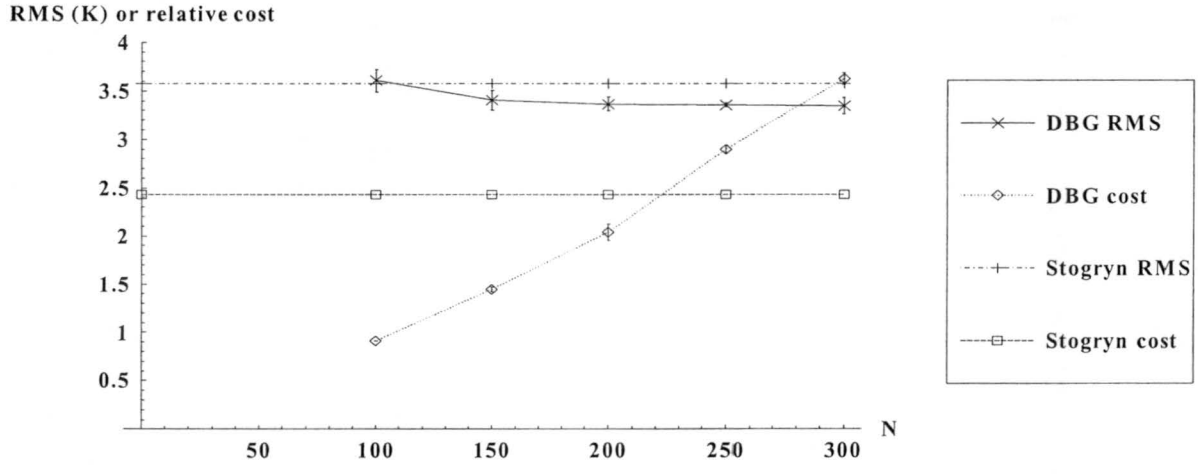

FIG. 7. Same as Figure 5, except for the sine function scene (Scene C).

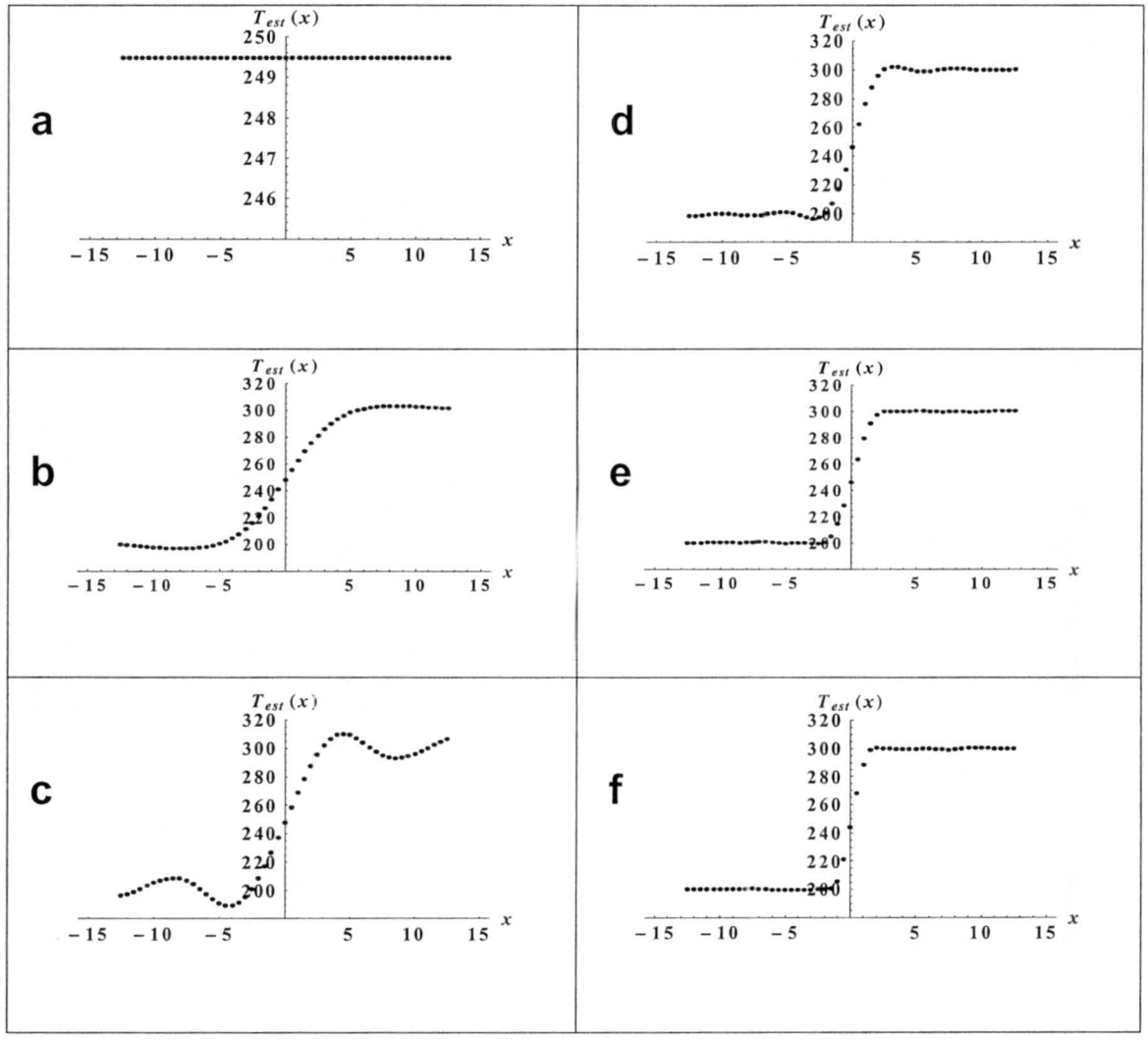

FIG. 8. The summations in the SVD of the DBG method can be truncated at any desired point with reduced fidelity. A sequence of results are shown in which additional terms are progressively added to the SVD summation, where M = 100, and a) has 1% of M terms, b) 3%, c) 5%, d) 10%, e) 15%, and f) 100%.

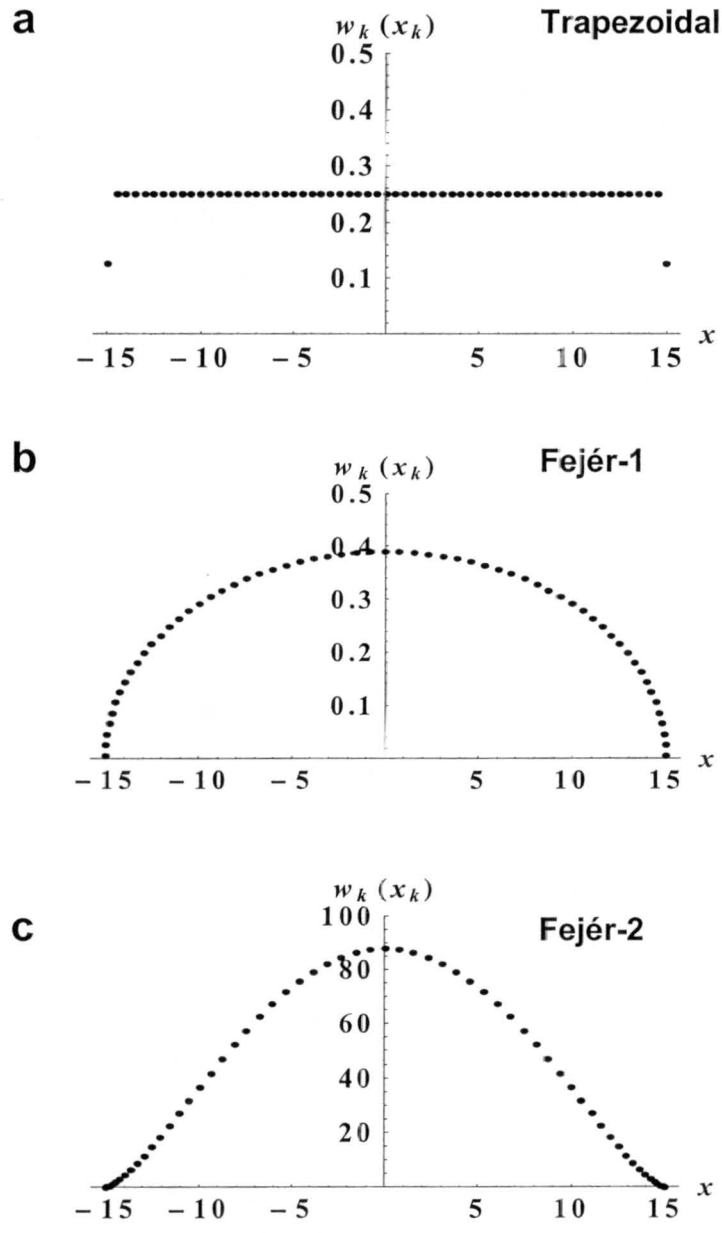

FIG. 9. Quadrature weights, w_k, and positions, x_k, for a) Trapezoidal, b) Fejér-1, and c) Fejér-2 quadrature.

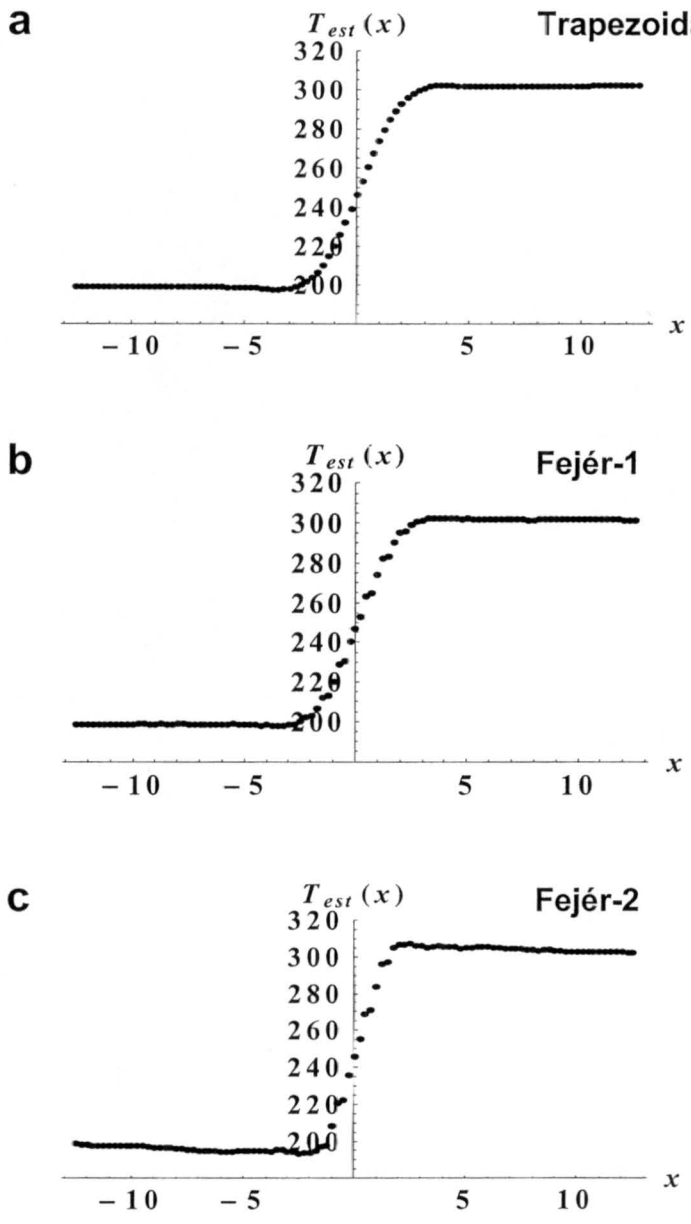

FIG. 10. DBG temperature estimates of the step function scene using a) Trapezoidal, b) Fejér-1, and c) Fejér-2 quadrature.

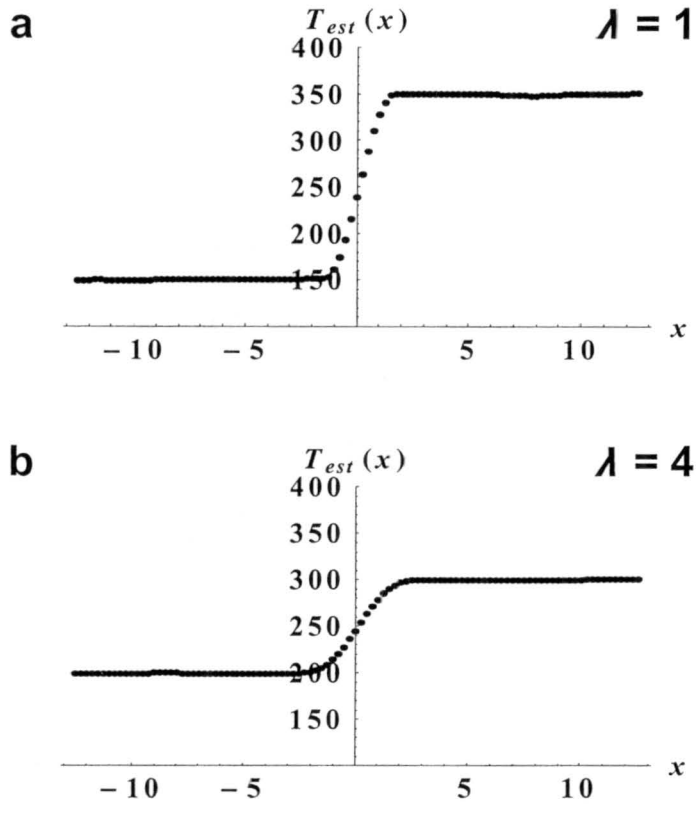

FIG. 11. DBG temperature estimates of the step function scene using an unnormalized gain function that has a gain function area of 2, with a) $\lambda = 0$, and b) $\lambda = 4$. This demonstrates how the λ regularization parameter acts to unbias the estimates for unnormalized gain functions.

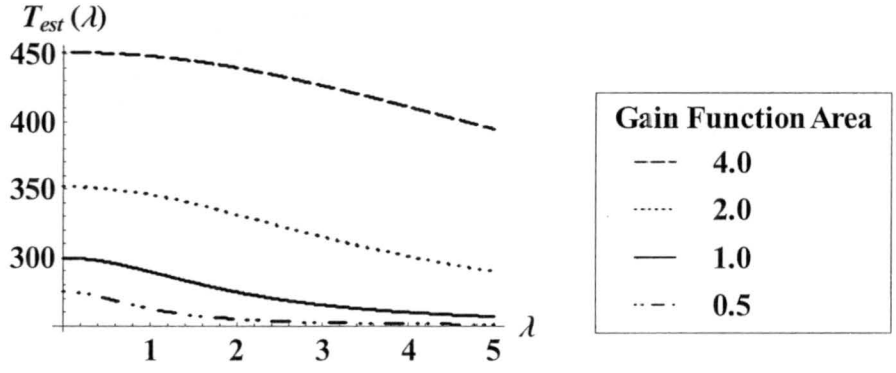

FIG. 12. DBG temperature estimate results at the maximum of the step function scene as a function of λ and of the gain function area. For reference, the maximum truth-value is 300 K. From top to bottom, the gain functions areas are 4, 2, 1, and 0.5, respectively.

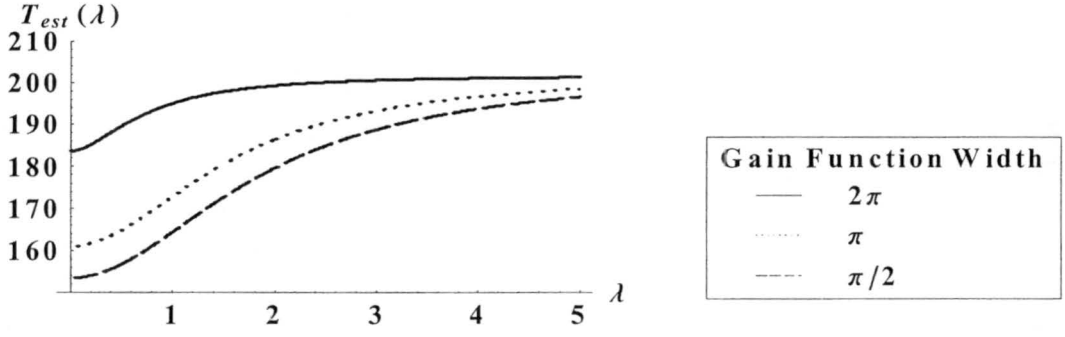

FIG. 13. DBG temperature estimate results at the local minimum of the sine function scene as a function of λ and of the gain function width. For reference, the local minimum truth-value is 150 K. From top to bottom the widths are 2π, π, and π/2 respectively. While this is an extreme example, this demonstrates that for some gain function configurations, local biases can be intrinsic to the system.

TABLE 1. SSM/I instrument characteristics (adapted from *Hollinger et al.,* 1990).

Channel number	Frequency (GHz)	Polar-ization (H or V)	Effective-field-of-view (EFOV)* (km)	EFOV Ratio with respect to Channel 7	Sensitivity (NEΔT)** (K)
1	19.35	V	69 × 43	4.6 × 3.3	0.45
2	19.35	H	69 × 43	4.6 × 3.3	0.42
3	22.235	V	60 × 40	4.0 × 3.1	0.73
4	37.0	V	37 × 28	2.5 × 2.2	0.37
5	37.0	H	37 × 29	2.5 × 2.2	0.38
6	85.5	V	15 × 13	1.0 × 1.0	0.69
7	85.5	H	15 × 13	1.0 × 1.0	0.73

* 3 dB limits.

** Average of laboratory measurements.

TABLE 2. Simulation results for Scenes A – C. Results are presented as root-mean-squared (RMS) differences from the "truth" scene before 5 K of random noise was added. Standard deviations of the RMS values are calculated by aggregating results from several simulation instances. For the scene simulations within this table, $M = 50$, and $N = 50$.

Method	Penalty Function $J(x_0)$	Scene A Homogenous Case		Scene B Step Function Case		Scene C Sine Function Case	
		RMS (K)	Std. Dev. (K)	RMS (K)	Std. Dev. (K)	RMS (K)	Std. Dev. (K)
Stogryn	1	0.71	0.14	12.2	0.08	4.4	0.08
DBG		0.61	0.14	12.3	0.07	4.5	0.15
Stogryn	$(x - x_0)^2$	0.51	0.14	8.2	0.06	5.8	0.13
DBG		0.33	0.06	10.4	0.08	5.3	0.13

TABLE 3. Same as Table 2, except that a lateral boundary condition spatial filter has been applied to the results. The spatial filter has the effect of "windowing" the results to a region such that $-12.5 < x < 12.5$. These results minimize the lateral boundary condition effects for the $J = 1$ simulations, and are therefore more indicative of the discontinuity at $x = 0$ for Scenes B and C. For the scene simulations within this table, $M = 100$, and $N = 100$. The data within this table corresponds to Figures 5-7.

Method	Penalty Function $J(x_0)$	Scene A Homogenous Case		Scene B Step Function Case		Scene C Sine Function Case	
		RMS (K)	Std. Dev. (K)	RMS (K)	Std. Dev. (K)	RMS (K)	Std. Dev. (K)
Stogryn	1	0.64	0.07	9.9	0.03	3.6	0.06
DBG		0.65	0.07	9.9	0.06	3.6	0.12

TABLE 4. RMS performance ratios are defined as the SVD DBG RMS results relative to the non-SVD DBG RMS results. Corresponding standard deviations are also calculated. RMS ratio results less than one indicate an improvement using the SVD approach. The computational cost ratio is defined similarly, with the cost ratio being defined as the SVD cost relative to the non-SVD cost. The computational costs are independent of the specific scene.

	Uniform		**Step**		**Sine**		**Cost**	
%M	RMS	σ	RMS	σ	RMS	σ	AVERAGE	σ
100	0.50	0.09	0.991	0.003	1.30	0.03	0.699	0.009
50	0.47	0.05	0.993	0.004	1.31	0.03	0.580	0.006
20	0.43	0.07	0.992	0.004	1.31	0.03	0.519	0.010
10	0.43	0.07	1.062	0.004	1.46	0.03	0.486	0.002

TABLE 5. DBG RMS performance using three different quadrature rules for three simulated scenes.

Scene	Trapezoidal		Fejér-1		Fejér–2	
	RMS	σ	RMS	σ	RMS	σ
Uniform	0.40	0.09	0.43	0.11	0.50	0.09
Step	11.24	0.04	11.37	0.04	10.26	0.09
Sine	10.00	0.08	10.00	0.07	6.11	0.10

Appendix A: Backus-Gilbert Inclusion of λ Regularization

This appendix is after *Twomey* (1996), pages 160-162. The following constraints are used in BG: 1) $\mathbf{a}\cdot\mathbf{a} \le const$, to prevent error magnification, 2) $\int G_i(x)dx = 1 = \mathbf{a}\cdot\mathbf{u} = \mathbf{u}\cdot\mathbf{a}$, to make the scanning function unitary, and 3) $G_i(x)$ is centered around x_i to approximate a delta function. With the above mathematical properties, the extremum condition is:

$$\frac{\partial}{\partial a_k}\left[\mathbf{a}^{\mathrm{T}}\mathbf{S}\mathbf{a} - \beta\,\mathbf{a}\cdot\mathbf{u} - \beta\,\mathbf{u}\cdot\mathbf{a} + \lambda^2\,\mathbf{a}\cdot\mathbf{a}\right] = 0, \tag{A1}$$

where two Lagrangian multipliers are employed, -2β, and λ^2. This condition yields,

$$\mathbf{e}^{\mathrm{T}}\mathbf{S}\mathbf{a} + \mathbf{a}^{\mathrm{T}}\mathbf{S}\mathbf{e} - \beta\,\mathbf{e}\cdot\mathbf{u} - \beta\,\mathbf{u}\cdot\mathbf{e} + \lambda^2\,\mathbf{e}\cdot\mathbf{a} + \lambda^2\,\mathbf{a}\cdot\mathbf{e}, \tag{A2}$$

where **e** is an identity vector. This simplifies to

$$\mathbf{S}\mathbf{a} + \lambda^2\,\mathbf{a} = \beta\,\mathbf{u}, \tag{A3}$$

which yields,

$$(\mathbf{S} + \lambda^2\,\mathbf{I})\,\mathbf{a} = \beta\,\mathbf{u}\ , \tag{A4}$$

or alternatively,

$$\mathbf{a} = \beta\,(\mathbf{S} + \lambda^2\,\mathbf{I})^{-1}\mathbf{u}\,. \tag{A5}$$

β is found by using the knowledge that $\mathbf{u}\cdot\mathbf{a} = 1$, so that

$$\beta = \frac{\mathbf{1}}{\mathbf{u}^{\mathrm{T}}(\mathbf{S} + \lambda^2\,\mathbf{I})^{-1}\mathbf{u}}. \tag{A6}$$

From (A5), **a** is then

$$\mathbf{a} = \frac{(\mathbf{S} + \lambda^2 \mathbf{I})^{-1}\mathbf{u}}{\mathbf{u}^T(\mathbf{S} + \lambda^2 \mathbf{I})^{-1}\mathbf{u}}. \tag{A7}$$

The λ regularization parameter dictates the error magnification and the averaging gain function width. When λ is large, the width is large and error is small. When λ is small, the width is small, but the error is large. According to the purpose of the analysis, λ should be adjusted to optimize these behaviors.

Appendix B: Regularization Analysis

This Appendix is based on the discussion in *Lagendijk and Biemond* (1991) pages 38 and 39. If there exist measurements, $T_{meas}(x)$, with noise w, and an estimate given by $T_{est}(x)$. The real function given by $T_{\text{exact}}(x)$, and a gain function, $G(x,x_i)$, maps T_{exact} into the same space as T_{meas}. This leads to

$$T_{meas}(x_i) = \int G(x,x_i) T_{exact}(x) dx + w . \tag{B1}$$

To keep the estimate within the errors gives the condition

$$\Phi(T_{est}(x_i)) = \left\| T_{meas}(x_i) - \int G(x,x_i) T_{est}(x) dx \right| \le \|w\| = \varepsilon , \tag{B2}$$

where Φ is the error measurement function which measures the error, ε.

A regularization operator, C, is defined with the function, Ω,

$$\Omega(T_{est}(x_i)) = \|C T_{est}(x_i)\| , \tag{B3}$$

so that the regularization solution now reduces to minimizing Ω with the constraint $\Phi = \|w\|$. Using the method of undetermined Lagrange multipliers, the following relationship is to be minimized

$$\Phi(T_{est}(x_i)) = \left\| T_{meas}(x_i) - \int G(x,x_i) T_{est}(x) dx \right\|^2 + \lambda^2 \|C T_{est}(x_i)\|^2 \tag{B4}$$

with respect to T_{est}. The Tikhonov-Miller Regularization is then based on replacing the minimization above with an upper bound on the error, E, so that

$$\Omega(T_{est}(x_i)) = \|C T_{est}(x_i)\| \le E . \tag{B5}$$

This leads to the following constraint in the context of the Lagrange multipliers,

$$\Phi(T_{est}(x_i)) = \left\| T_{meas}(x_i) - \int G(x, x_i) T_{est}(x) dx \right\|^2 + (\varepsilon / E)^2 \left\| C T_{est}(x_i) \right\|^2 \leq 2\varepsilon^2 \qquad \text{(B6)}$$

which yields,

$$\lambda = (\varepsilon/E). \qquad \text{(B7)}$$

When C = I and $\lambda = 0$, the filter becomes the pseudo-inverse filter, which minimizes the norm.

Acknowledgments. The authors would like to thank Dr. Tom Greenwald for a helpful review of this paper. This work was supported by the DoD Center for Geosciences/Atmospheric Research at Colorado State University under the Cooperative Agreement #DAAL01-98-2-0078 with the Army Research Laboratory.