ADAPTIVE SPECTRAL ESTIMATION USING THE CONJUGATE GRADIENT ALGORITHM

Pi Sheng Chang and *Alan N. Willson, Jr.*

Electrical Engineering Department
University of California, Los Angeles
Los Angeles, CA 90095-1600
e-mail: pschang@ee.ucla.edu, willson@ee.ucla.edu

ABSTRACT

A method for spectral estimation is presented, using the modified Conjugate Gradient (CG) algorithm. It implements an adaptive version of Pisarenko's harmonic retrieval method, where the estimates are updated sample-by-sample. First, a constrained unit norm CG algorithm is formulated, then it is recast into an unconstrained minimization problem. The resulting algorithms can be extended to solve the generalized eigensystem problem, when the noise covariance matrix is known a priori. It is shown that the proposed algorithms' convergence rate is comparable to that of a least-squares type algorithm, while being computationally more efficient. Performance simulations are shown, and comparisons with some existing methods are provided.

1. INTRODUCTION

Spectral estimation methods using adaptive transversal filters often give biased estimations due to the presence of noise. In [1] an adaptive scheme was proposed to estimate frequencies of sinusoids corrupted by white noise. It was shown that by using the described method the estimates were unbiased. In [2] the same problem was addressed but a least-squares type of adaptive algorithm was used. In [3] a technique using the Conjugate Gradient (CG) algorithm was presented. Although this technique was computationally more efficient than the one presented in [2], it is only suitable for block processing. The CG algorithm is also used in [4, 5] to implement an adaptive Pisarenko harmonic retrieval method. In [4] the problem of minimization of the Rayleigh quotient is addressed, while in [5] a rotational search is used. Here we use the modified CG algorithm presented in [6] to implement a frequency estimator in the presence of white noise, by using the same approach described in [1, 2], where the weight-vector of an adaptive FIR filter is constrained to have unit norm. It was shown in [1] that by using this simple criterion it is possible to obtain an adaptive version of the Pisarenko's harmonic retrieval method. The proposed method is also computationally more efficient than the least-squares type algorithms [2, 7, 8], and can update the frequency estimates at one iteration per input sample. It has the same desirable fast convergence property but without the necessity to estimate the inverse of the correlation matrix. First, a constrained CG optimization algorithm is presented. Then, by modifying the cost function, an unconstrained version is formulated. Finally, when the noise covariance matrix is known a priori, the proposed algorithms can estimate frequencies in colored noise by solving a generalized eigensystem problem.

*This work was supported by the Office of Naval Research under Grant N00014-95-1-0231.

2. THE MODIFIED CG ALGORITHM

The modified CG algorithm presented in [6] minimizes the cost function defined as $f(\mathbf{w}) = \mathbf{w}^T\mathbf{R}\mathbf{w}/2 + \mathbf{b}^T\mathbf{w}$. The algorithm is described by the following relations

Set initial conditions: $\mathbf{w}_0 = 0,\ \mathbf{g}_0 = \mathbf{b}_0,\ \mathbf{p}_1 = \mathbf{g}_0,\ k = 1.$

$$\alpha_k = \eta\frac{\mathbf{p}_k^T\mathbf{g}_{k-1}}{\mathbf{p}_k^T\mathbf{R}_k\mathbf{p}_k} \tag{1}$$

$$\mathbf{w}_k = \mathbf{w}_{k-1} + \alpha_k\mathbf{p}_k \tag{2}$$

$$\mathbf{g}_k = \lambda_f\mathbf{g}_{k-1} - \alpha_k\mathbf{R}_k\mathbf{p}_k + \mathbf{x}_k(d_k - \mathbf{x}_k^T\mathbf{w}_{k-1}) \tag{3}$$

$$\beta_k = \frac{(\mathbf{g}_k - \mathbf{g}_{k-1})^T\mathbf{g}_k}{\mathbf{g}_{k-1}^T\mathbf{g}_{k-1}} \tag{4}$$

$$\mathbf{p}_{k+1} = \mathbf{g}_k + \beta_k\mathbf{p}_k \tag{5}$$

where d_k is the desired response, α_k is the step size that minimizes the cost function $f(\mathbf{w})$, β_k provides $\mathbf{R}$-conjugacy for the direction vector $\mathbf{p}_k$, and $\mathbf{g}_k$ is the residual vector defined as $\mathbf{g}_k = -\nabla f(\mathbf{w})^T$. $\mathbf{R}_k$ is the $N \times N$ sample correlation matrix of the input data vector $\mathbf{x}_k$, and is computed as

$$\mathbf{R}_k = \lambda_f\mathbf{R}_{k-1} + \mathbf{x}_k\mathbf{x}_k^T \tag{6}$$

where λ_f is the forgetting factor of the exponentially decaying data window. The parameter η in (1) controls the convergence of the algorithm and must be set within the range $(\lambda_f - 0.5) \leq \eta \leq \lambda_f$ [6]. This method can be applied to FIR adaptive filtering problems by solving the normal equation $\mathbf{R}\mathbf{w} = \mathbf{b}$, where $\mathbf{R}$ and $\mathbf{b}$ are approximated by their time averaging versions.

3. THE PISARENKO'S HARMONIC RETRIEVAL METHOD

Pisarenko's harmonic retrieval method consists of finding the minimum eigenvalue λ_{min} and its associated eigenvector $\mathbf{q}_{min}$ of the correlation matrix $\mathbf{R}$, and then computing the roots of the polynomial whose coefficients are the components of $\mathbf{q}_{min}$. In [1] it was shown that these roots are of unit modulus and their angles are the harmonic frequencies contained in $\mathbf{R}$. Furthermore, by constraining the weight-vector of an adaptive transversal filter to unit norm, it was shown that the weight-vector converges to $\pm\mathbf{q}_{min}$.

Consider the minimization of the cost function of a deterministic least-squares type algorithm given by $1/2\sum_{i=0}^{k}\lambda_f^{k-i}e_i^2$ [9]. Minimizing this cost function is equivalent to using an exponentially decaying data window for the computation of $\mathbf{R}_k$ as shown in [6]. In this case, the function to be minimized becomes, after using the unit-norm constraint on the weight-vector,

$$\begin{aligned} f(\mathbf{w}) &= \frac{1}{2}\bar{\mathbf{w}}_k^T\mathbf{R}_k\bar{\mathbf{w}}_k && (7)\\ &= \frac{1}{2}\bar{\mathbf{w}}_k^T\sum_{i=0}^{k}\lambda_f^{k-i}\mathbf{x}_i\mathbf{x}_i^T\bar{\mathbf{w}}_k = \frac{1}{2}\sum_{i=0}^{k}\lambda_f^{k-i}e_i^2 && (8)\end{aligned}$$

where

$$e_i = \bar{\mathbf{w}}_k^T\mathbf{x}_i,\ \bar{\mathbf{w}}_k = \frac{\mathbf{w}_k}{\|\mathbf{w}_k\|}\ \ and\ \ \|\mathbf{w}_k\| = (\mathbf{w}_k^T\mathbf{w}_k)^{1/2}.$$

The gradient or residual vector is then computed as

$$\begin{aligned} \mathbf{g}_k &= -\nabla f(\mathbf{w})^T = -\mathbf{R}_k\bar{\mathbf{w}}_k \\ &= \frac{1}{\|\mathbf{w}_k\|}(\lambda_f\mathbf{g}_{k-1} - \alpha_k\mathbf{R}_k\mathbf{p}_k - \mathbf{x}_k(\mathbf{x}_k^T\bar{\mathbf{w}}_{k-1})). && (9)\end{aligned}$$

Equation (9) was obtained by using (6) and (2). The proposed algorithm becomes

Set initial cond.: $\bar{\mathbf{w}}_0 = [1,0,...,0]$, $\mathbf{g}_0 = [-1,0,...,0]$, $\mathbf{p}_1 = \mathbf{g}_0$, $k=1$.

$$\begin{aligned} \alpha_k &= \eta\frac{\mathbf{p}_k^T\mathbf{g}_{k-1}}{\mathbf{p}_k^T\mathbf{R}_k\mathbf{p}_k} && (10)\\ \mathbf{w}_k &= \bar{\mathbf{w}}_{k-1} + \alpha_k\mathbf{p}_k && (11)\\ \bar{\mathbf{w}}_k &= \mathbf{w}_k/\|\mathbf{w}_k\| && (12)\\ \mathbf{g}_k &= \frac{1}{\|\mathbf{w}_k\|}(\lambda_f\mathbf{g}_{k-1} - \alpha_k\mathbf{R}_k\mathbf{p}_k - \mathbf{x}_k(\mathbf{x}_k^T\bar{\mathbf{w}}_{k-1})) && (13)\\ \beta_k &= \frac{(\mathbf{g}_k - \mathbf{g}_{k-1})^T\mathbf{g}_k}{\mathbf{g}_{k-1}^T\mathbf{g}_{k-1}} && (14)\\ \mathbf{p}_{k+1} &= \mathbf{g}_k + \beta_k\mathbf{p}_k. && (15)\end{aligned}$$

4. PRESERVING THE R-CONJUGACY

When the unit-norm constraint and variable $\mathbf{R}$ are used, the vectors $\mathbf{p}_k$ lose the $\mathbf{R}$-conjugacy property. In order to preserve the conjugacy, more precise expressions for the parameters α_k and β_k have to be found.

Consider the Expanding Subspace Theorem [11], where $\mathbf{p}_k^T\mathbf{g}_k = 0$ should be satisfied. Premultiplying (13) by $\mathbf{p}_k^T$ we have

$$\mathbf{p}_k^T\mathbf{g}_k = \frac{\mathbf{p}_k^T}{\|\mathbf{w}_k\|}(\lambda_f\mathbf{g}_{k-1} - \alpha_k\mathbf{R}_k\mathbf{p}_k - \mathbf{x}_k(\mathbf{x}_k^T\bar{\mathbf{w}}_{k-1})) = 0. \quad (16)$$

The α_k that satisfies (16) is given by

$$\alpha_k = \frac{\lambda_f\mathbf{p}_k^T\mathbf{g}_{k-1} - \mathbf{p}_k^T\mathbf{x}_k(\mathbf{x}_k^T\bar{\mathbf{w}}_{k-1})}{\mathbf{p}_k^T\mathbf{R}_k\mathbf{p}_k}. \quad (17)$$

In order to ensure $\mathbf{R}$-conjugacy, $\mathbf{p}_k^T\mathbf{R}_k\mathbf{p}_{k+1} = 0$ should be satisfied [11]. Premultiplying (15) by $\mathbf{p}_k^T\mathbf{R}_k$, we have

$$\mathbf{p}_k^T\mathbf{R}_k\mathbf{p}_{k+1} = \mathbf{p}_k^T\mathbf{R}_k\mathbf{g}_k + \beta_k\mathbf{p}_k^T\mathbf{R}_k\mathbf{p}_k = 0 \quad (18)$$

and β_k is given by

$$\beta_k = -\frac{\mathbf{g}_k^T\mathbf{R}_k\mathbf{p}_k}{\mathbf{p}_k^T\mathbf{R}_k\mathbf{p}_k}. \quad (19)$$

We can obtain $\mathbf{R}_k\mathbf{p}_k$ from (13), resulting in

$$\beta_k = \frac{\|\mathbf{w}_k\|\mathbf{g}_k^T\mathbf{g}_k - \lambda_f\mathbf{g}_k^T\mathbf{g}_{k-1} + \mathbf{g}_k^T\mathbf{x}_k(\mathbf{x}_k^T\bar{\mathbf{w}}_{k-1})}{\lambda_f\mathbf{p}_k^T\mathbf{g}_{k-1} - \mathbf{p}_k^T\mathbf{x}_k(\mathbf{x}_k^T\bar{\mathbf{w}}_{k-1})}. \quad (20)$$

Note that the new α_k minimizes $f(\bar{\mathbf{w}}_{k-1} + \alpha_k\mathbf{p}_k)$ but, due to the normalization of $\mathbf{w}_k$, the minimum is not achieved at each iteration. The problem of inexact line search has been addressed in [6, 11, 12] and simulations have shown that it doesn't affect the convergence of the algorithm very much. Expressions of α_k, such as the one shown in (10), can be used without compromising the performance of the algorithm.

5. UNCONSTRAINED OPTIMIZATION USING THE CG ALGORITHM

It is possible to reformulate the optimization problem so that the unit-norm constraint can be incorporated in the cost function. In [7] a new cost function is formulated, taking into account the constraint. The cost function then becomes

$$f(\mathbf{w}) = \frac{\mathbf{w}_k^T\mathbf{R}_k\mathbf{w}_k}{2} + \frac{\mu(\mathbf{w}_k^T\mathbf{w}_k - 1)^2}{4}. \quad (21)$$

The gradient vector $\nabla f(\mathbf{w})^T$ associated with this cost function is given by

$$\nabla f(\mathbf{w})^T = \mathbf{R}_k\mathbf{w}_k + \mu(\mathbf{w}_k^T\mathbf{w}_k - 1)\mathbf{w}_k \quad (22)$$

and $\mathbf{g}_k = -\nabla f(\mathbf{w})^T$. A lower bound for μ can be found in [7] and is given by

$$\mu \geq tr(\mathbf{R})/2 \quad (23)$$

where $tr(\mathbf{R})$ is the trace of the matrix $\mathbf{R}$. In order to apply this cost function, the CG algorithm for nonquadratic functions presented in [10] will be used. Since the new cost function is quadractic, the global minimum will always be found. However, it does not involve the computation of the Hessian as in [7, 8]. The new algorithm is given by

Set initial cond.: $\mathbf{w}_0 = [1,0,...,0]$, $\mathbf{g}_0 = [-1,0,...,0]$, $\mathbf{p}_1 = \mathbf{g}_0$, $k=1$.

$$\begin{aligned} \sigma_k &= \frac{\sigma}{\|\mathbf{p}_k\|} && (24)\\ \mathbf{s}_k &= \frac{\nabla f(\mathbf{w}_{k-1} + \sigma_k\mathbf{p}_k)^T - \nabla f(\mathbf{w}_{k-1})^T}{\sigma_k} && (25)\end{aligned}$$

$$
\begin{aligned}
\alpha_k &= \frac{\mathbf{p}_k^T \mathbf{g}_{k-1}}{\mathbf{p}_k^T \mathbf{s}_k} && (26)\\
\mathbf{w}_k &= \mathbf{w}_{k-1} + \alpha_k \mathbf{p}_k && (27)\\
\mathbf{g}_k &= -\nabla f(\mathbf{w}_k)^T && (28)\\
\beta_k &= -\frac{\mathbf{g}_k^T \mathbf{s}_k}{\mathbf{p}_k^T \mathbf{s}_k} && (29)\\
\mathbf{p}_{k+1} &= \mathbf{g}_k + \beta_k \mathbf{p}_k && (30)
\end{aligned}
$$

where σ is a small number.

6. SOLVING GENERALIZED EIGENSYSTEM PROBLEM

If the noise covariance matrix is known a priori, we can use the CG algorithm formulated above to solve the generalized eigensystem problem, as suggested in [3, 8], when the noise is not white. Then we have

$$\mathbf{R}_k \mathbf{w}_k = \lambda \mathbf{Z} \mathbf{w}_k \qquad (31)$$

where $\mathbf{Z}$ is the covariance matrix of the colored noise. The unit-norm constraint (eq. (12)) in the algorithm described by (10)-(15) has to be modified to

$$\mathbf{w}_k^T \mathbf{Z} \mathbf{w}_k = 1. \qquad (32)$$

This can be accomplished by computing

$$\bar{\mathbf{w}}_k = \mathbf{w}_k / \mathbf{w}_k^T \mathbf{Z} \mathbf{w}_k \qquad (33)$$

instead of (12) in the constrained CG algorithm. In the unconstrained version, the cost function should be modified to [8]

$$f(\mathbf{w}) = \frac{\mathbf{w}_k^T \mathbf{R}_k \mathbf{w}_k}{2} + \frac{\mu(\mathbf{w}_k^T \mathbf{Z} \mathbf{w}_k - 1)^2}{4} \qquad (34)$$

and $\nabla f(\mathbf{w})^T$ becomes

$$\nabla f(\mathbf{w})^T = \mathbf{R}_k \mathbf{w}_k - \mu(\mathbf{w}_k^T \mathbf{Z} \mathbf{w}_k - 1) \mathbf{Z} \mathbf{w}_k. \qquad (35)$$

7. SIMULATIONS

Consider two equal-power sinusoids of normalized frequencies 0.3 and 0.4 in white noise, at 10 dB SNR, with noise variance equal to 1 and $N = 7$ (FIR filter of order 6). The mean and standard deviation shown in Table 1 were obtained by using the values of the last 100 iterations. The plots in Figs. 1 and 3 show the performance of the LMS algorithm [1], with $\mu_{LMS} = 0.001$. Figs. 2 and 4 show the improvement in performance when using the proposed constrained CG algorithm. Here $\lambda_f = \eta = 0.99$. Fig. 5 shows the performance of the least-squares algorithm described in [2], with $\lambda_f = 0.99$. Using equations (17) and (20) in the constrained CG algorithm, there was no substantial improvement in the performance, showing that the slight loss of conjugacy is well tolerated by the algorithm. Fig. 6 shows the performance of the unconstrained CG algorithm, with $\mu = 10000$ and $\sigma = 0.001$.

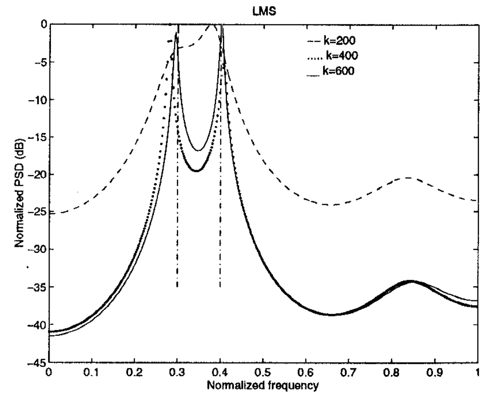

Fig. 1. Spectral estimates: LMS algorithm.

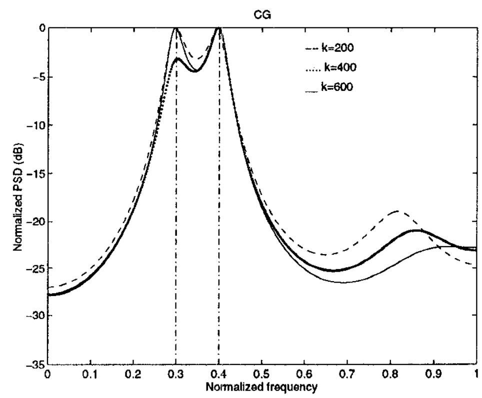

Fig. 2. Spectral estimates: constrained CG algorithm.

8. CONCLUSION

A method for spectral estimation has been presented. It is based on the adaptive version of Pisarenko's harmonic retrieval method. The modified Conjugate Gradient algorithm was used in the updating of the coefficients of the adaptive transversal filter. Two versions of the algorithm were presented and their extension for the case of a known noise-covariance matrix were shown. The proposed algorithms have a fast convergence rate, are computationally more efficient than the least-squares type algorithms [2, 7, 8], and are suitable for sample-by-sample processing. Simulations illustrating their performance have been given.

Table 1. Simulation results for the algorithms discussed in the text.

	f_1		f_2	
	mean	std	mean	std
LMS	0.2965	0.0049	0.3956	0.0055
constrained CG	0.2967	0.0024	0.3978	0.0023
Least-squares	0.3086	≈ 0	0.4238	≈ 0
unconstrained CG	0.2993	0.0015	0.4005	0.0018

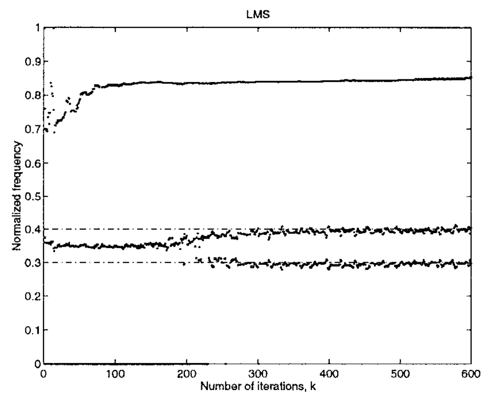

Fig. 3. Spectral estimates: LMS algorithm.

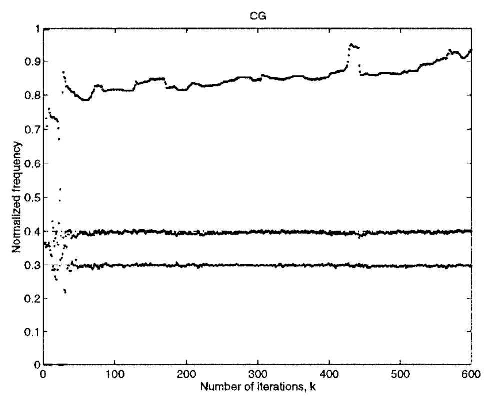

Fig. 4. Spectral estimates: constrained CG algorithm.

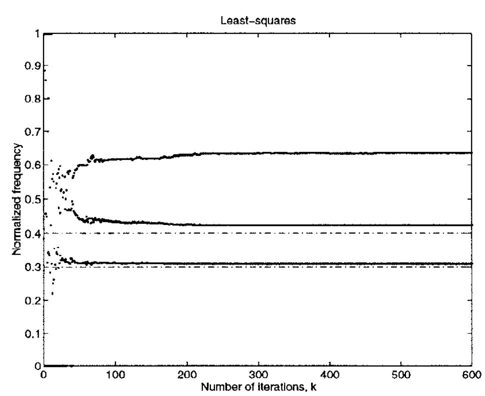

Fig. 5. Spectral estimates: least-squares type algorithm.

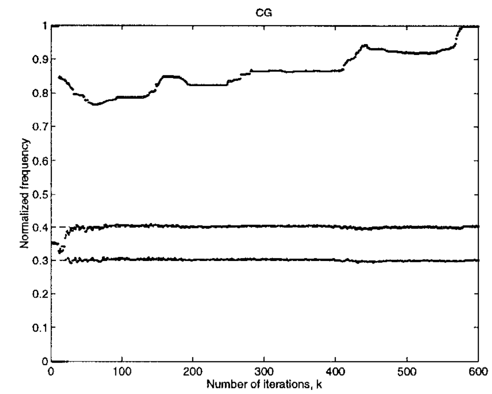

Fig. 6. Spectral estimates: unconstrained CG algorithm.

REFERENCES

[1] P. A. Thompson, "An adaptive spectral analysis technique for unbiased frequency estimation in the presence of white noise," in *Proc. 13th Asilomar Conf. on Circuits, Syst. and Comp.*, Pacific Grove, Calif., pp. 529-533, Nov. 1979.

[2] V. U. Reddy, B. Edgardt, and T. Kailath, "Least-squares type algorithm for adaptive implementation of Pisarenko's harmonic retrieval method," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, pp. 399-405, Jun. 1982.

[3] H. Chen, T. K. Sarkar, S. A. Dianat, and J. D. Brulé, "Adaptive spectral estimation by the conjugate gradient method," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 272-284, Apr. 1986.

[4] X. Yang, T. K. Sarkar, and E. Arvas, "A survey of Conjugate Gradient algorithms for solution of extreme eigenproblems of a symmetric matrix," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 1550-1556, Oct. 1989.

[5] D. R. Fuhrmann and B. Liu, "Rotational search methods for adaptive Pisarenko harmonic retrieval," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, pp. 1550-1565, Dec. 1986.

[6] P. S. Chang and A. N. Willson Jr., "Adaptive filtering using modified Conjugate Gradient," in *Proc. 38th Midwest Symp. on Circuits and Systems*, Rio de Janeiro, Brazil, in print, Aug. 1995.

[7] G. Mathew, V. U. Reddy, and S. Dasgupta, "Adaptive estimation of eigensubspace," *IEEE Trans. Signal Processing*, vol. 43, pp. 401-411, Feb. 1995.

[8] G. Mathew, V. U. Reddy, and A. Paulraj, "A Quasi-Newton adaptive algorithm for estimating generalized eigenvectors," in *Proc. 28th Asilomar Conf. on Signals, Systems and Comp.*, Pacific Grove, Calif., pp. 602-606, Oct. 1994.

[9] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1991, 2nd ed.

[10] M. Hestenes, *Conjugate Direction Methods in Optimization.* New York: Springer-Verlag, 1980.

[11] D. G. Luenberger, *Linear and Nonlinear Programming.* Reading, MA: Addison-Wesley, 1984, 2nd ed.

[12] M. Al-Baali, "Descent property and global convergence of the Fletcher-Reeves method with inexact line search," *IMA J. Num. Anal.*, vol. 5, pp. 121-124, Jan. 1985.