AN OPTIMAL LINEAR ESTIMATION APPROACH TO THE PARALLEL SOLUTION OF LINEAR ALGEBRAIC SYSTEMS OF EQUATIONS

Atair Rios Neto* & Wilson Rios Neto**

*Universidade do Vale do Paraíba-UNIVAP

Instituto de Pesquisa e Desenvolvimento-IP&D

12244-000 São Jose dos Campos, SP

e-mail: atair@univap.br

**Virginia Polytechnic Institute and State University

Center for Intelligent Material Systems and Structures - CIMSS

Blacksburg, Virginia 24061-0261

e mail: wriosnet@vt.edu

Resumo Estimação linear ótima estocástica de parâmetros é utilizada para gerar um novo método iterativo, para a solução paralela de sistemas algébricos de equações. No limite, a abordagem proposta leva a algoritmos iterativos que resolvem em paralelo sistemas lineares, variável por variável. É mostrado, tanto no caso de sistemas determinados como no de sistemas indeterminados, que o método iterativo de solução paralela desenvolvido é equivalente, em cada iteração, a se aplicar a uma função objetivo quadrática, e definida positiva, um método de Newton modificado, com convergência garantida. A motivação é se ter um método que explore as possibilidades oferecidas por processamento paralelo, que pode ser útil na solução eficiente de sistemas de larga escala, especialmente aqueles envolvendo matrizes de coeficientes esparsas.É mostrado também que a abordagem mais geral por estimação estocástica leva a um método que generaliza e que, em consequência, se espera ter melhor desempenho que o método usual de Jacobi. Embora sejam apresentados exemplos numéricos, este trabalho não trata do aspecto de testes e avaliação de desempenho numérico e está focado no desenvolvimento heurístico e na verificação de convergência do método proposto.

Palavras Chaves: Sistemas Lineares, Solução Paralela de Sistemas Lineares, Método Generalizado de Jacobi.

Abstract Stochastic optimal linear estimation of parameters is used to generate a new iterative method for the parallel solution of systems of linear algebraic equations. In the limit the approach proposed leads to iterative algorithms which in parallel can solve linear systems, variable by variable. It is shown, for both determined and undetermined systems, that the parallel iterative method developed is in each iteration equivalent to apply to a quadratic positive definite functional a modified Newton method, with guaranteed convergence. The motivation is to have a method which explores the possibilities offered by parallel processing and that can be useful in the efficient solution of large scale systems, especially those with sparse coefficient matrices. It is shown also that the more general stochastic approach taken leads to a method which generalizes and which as a consequence is expected to perform better than the usual Jacobi method. Though numerical examples are presented, this paper does not yet address the aspect of testing and evaluation of numerical performance and is focused in presenting the heuristic development and verification of convergence of the proposed method.

Keywords: Algebraic linear systems; Parallel solution of linear systems; Generalized Jacobi method.

1 INTRODUCTION

The sequential solution, equation by equation, of linear algebraic equations has been treated with success using both deterministic approaches (Huang, 1975; Abbaffy et alii, 1984; Spedicato, 1995) and stochastic optimal linear estimation approaches (Rios Neto, 1981; Pinto and Rios Neto, 1990). The resulting methods are appropriate for the existing sequential processing machines and have been shown to be not only effective but also efficient in iterative schemes to deal with ill conditioned systems (Wederley, 1997).

The advent of computational parallel processing means justify the research effort to develop parallel solution type methods. The motivation is to explore the new possibilities available to gain in efficiency and to facilitate the treatment of large scale systems, searching for methods which iteratively solve a system of linear equations by processing in parallel groups of its variables or, in the limiting case, variable by variable, as the Jacobi method does.

The use of least squares (e.g., Chen and Billings, 1992) and Kalman filtering (e.g., Rios Neto, 1995; 1997) to generate parallel processing algorithms in the supervised training of artificial neural networks has called the attention upon the possibility of also using stochastic optimal linear estimation for generating methods for the parallel solution of linear algebraic equations. In what follows, this possibility is explored developing a method with these characteristics.

2 PROPOSED METHOD

2.1 Determined Systems

Consider the system of linear algebraic equations:

$$b = Ax + \varepsilon \tag{1}$$

where A is a given nxn full rank real matrix; x is the nx1 vector of unknown variables; b is a given nx1 real vector; and ε is to represent the numerical accuracy expected to be attained.

In an iterative solution scheme, the problem can be viewed , in a typical iteration, as:

$$\alpha(b - A\bar{x}) = A(x - \bar{x}) + \varepsilon \tag{2}$$

where $\bar{x}$ is the initial guess or the value from the previous iteration; and $0 < \alpha \leq 1$ is to be chosen to adjust the solution search step size.

If $x_k, k = 1,2,...,n, k \neq j$ is viewed in each iteration as given by:

$$x_k = \bar{x}_k + \bar{\eta}_k \tag{3}$$

where $\bar{\eta}_k$ is the error which represents the quality of $\bar{x}_k$, and taking this information back in Eq.(2), there results the following set of approximate and redundant equations in x_j:

$$\alpha(b_i - \sum_{l=1}^{n} a_{il} \bar{x}_l) = a_{ij}(x_j - \bar{x}_j) + \bar{v}_i^{\ j} \tag{4}$$

$$\bar{v}_i^{\ j} = \sum_{k=1}^{n} a_{ik} \bar{\eta}_k + \varepsilon_i, k \neq j \tag{5}$$

where $i=1,2,...,n$. If now the linear problem of Eqs.(4) is treated as one of unbiased parameter estimation, the errors $\bar{v}_i^{\ j}$ can be viewed as outcomes of zero mean, normally distributed random errors. In order to evaluate their dispersion, it seems reasonable to model them as zero mean, not correlated random variables with variances which are an approximation of the order of magnitude of the dispersion of their most probable realizations around zero. Since the first hand terms in Eqs.(4) are residues, with lower boundaries limited by the accuracy to be attained, it seems reasonable to consider:

$$E[(\bar{v}_i^{\ j})^2] = \bar{\beta}_i^2 \sum_{k=1}^{n} a_{ik}^2 e_k^2 = \bar{r}_i \tag{6a}$$

where $E[.]$ is the expected value operator; e_k are the standard deviations of the random errors specified to model the final accuracy sought for x_k; and $\bar{\beta}_i$ is a factor to adjust the order of magnitude of the most probable dispersion of $\bar{v}_i^{\ j}$. A natural and certainly not unique and conservative way of adjusting this dispersion is to consider a maximum likelihood statistical consistency type of criterion (see, for example, Jazwinki, 1970):

$$\beta = max\{ max\{ \beta_i : \beta_i^2 \sum_{j=1}^{n} a_{ij}^2 e_j^2 = \tag{6b}$$

$$(b_i - \sum_{k=1}^{n} a_{ik} \bar{x}_k)^2 , i = 1,2,...,n \}, 1 \}, \bar{\beta} \hat{=} \alpha\beta$$

With this modeling approach, the problem of solving for the generic component x_j in Eqs.(4) can be biewed as one of stochastic linear parameter estimation, in each iteration. For $j = 1,2,...,n$, the observation like Eqs.(4) can then in each iteration be processed in parallel, using a without a priori information Gauss-Markov estimator (see, for example, Liebelt, 1967), to get an estimate of the components x_j:

$$\hat{x}_j = \bar{x}_j + \alpha[A_j^T \bar{R}^{-1} A_j]^{-1} A_j^T \bar{R}^{-1}(b - A\bar{x}) \tag{7}$$

where A_j is the jth column of A; and

$$\bar{R} = diag.[\bar{r}_i = \alpha^2 \beta^2 \sum_{k=1}^{n} a_{ik}^2 e_k^2 = \tag{8}$$

$$\alpha^2 \beta^2 r_i : i = 1,2,...,n] = \alpha^2 \beta^2 R$$

Noticing that the without a priori information form of the estimator in Eq. (7) allows to cancel out the factor $\alpha^2 \beta^2$ in $\bar{R}$, and that R is the same for $j=1,2,...,n$, it is then possible to combine the estimates $\hat{x}_j$, of the parallel processing estimation, to get the equivalent following estimator for the whole vector x:

$$\hat{x} = \bar{x} - \alpha S \nabla f^T(\bar{x}) = \bar{x} - \alpha S g(\bar{x}) \tag{9}$$

$$S = diag.[(A_j^T R^{-1} A_j)^{-1} : \ j = 1,2,...,n] ;$$
$$f(x) = 1/2[Ax - b]^T R^{-1}[Ax - b] \tag{10}$$

$$\nabla f^T(\bar{x}) = A^T R^{-1}(A\bar{x} - b) \hat{=} g(\bar{x}) \tag{11}$$

and $0 < \alpha$ can be chosen such as to minimize $f(x)$ in a given iteration, if it is taken as (Luenberger, 1984):

$$\alpha = (\bar{g}^T S\bar{g})/(\bar{g}^T SA^T R^{-1} AS\bar{g}), \bar{g} = g(\bar{x}) \tag{12}$$

Since S and $A^T R^{-1} A$ are positive definite matrices, the estimator of Eq.(9) is equivalent to a modified Newton method applied to the functional of Eq.(10).

Convergence of the parallel processing method of Eqs.(7) can now be verified considering its Newton method equivalent form, using the Kantorovich inequality and concluding that in each iteration (see for example Luenberger, 1984, pp. 261-262):

$$Q[\hat{x}] \leq \beta^2 Q[\bar{x}] \tag{13}$$

$$Q[x] = 1/2(x - x^*)^T A^T R^{-1} A(x - x^*) \tag{14}$$

$$\beta = (\lambda_M - \lambda_m)/(\lambda_M + \lambda_m) \qquad (15)$$

$x*$ being the value of x that leads to the minimum of $f(x)$ in Eq.(10); and λ_M, λ_m the largest and smallest eigenvalues of the positive definite matrix $SA^T R^{-1}A$.

In what concerns numerical performance, the proposed method of Eqs. (9)-(11) is expected to perform better than the same purpose existing Jacobi's method, for at least the reasons that follow.

(i) Even in the most recent and elaborated form of Jacobi's method, as presented in Björck (1996), the relative influence of errors is not taken in account. The weight matrix R (in Eq.(10)) is particularized to be the indentity matrix and thus the information about accuracies sought is not considered in the relative weighting and priority to be given to the components of Eq.(1) in the search process.

(ii) In Jacobi's method, α in Eq.(10) is simplified to always take the value 1.0 and thus the full rank of the matrix of coefficients A is not sufficient to guarantee convergence.

(iii) The stochastic modelling of the accuracies e_j in x_j together with the statistical consistency criterion of Eqs.(6) lead to a way of evaluating and distiguishing the numerical zeros in the components of Eqs.(1). Thus the square roots of the diagonal terms of R (Eq.(8)) and the e_j can be used to get realistic stopping conditions, in the numerical search.

2.2 Undetermined Systems

Consider now the system of linear algebraic equations as in Eq.(1), where now A is a mxn, rank m matrix, with m<n; and b as a consequence is a mx1 real vector. In this case a unique solution does not exist, unless a criterion to be satisfied for the solution is imposed. If a norm of x closest to the origin is sought, then the problem is that of a constrained minimization, and the following solution results:

$$\hat{x} = A^T [AA^T]^{-1}(b+\varepsilon) \qquad (16)$$

where a given and arbitrary occurrence of the zero mean error ε is considered. Thus, at the cost of increasing the condition number of the matrix of coefficients (Stoer and Bulirsch, 1980), the solution is reduced to the one developed for the determined case, noticing that:

$$(AA^T)y = b+\varepsilon , \qquad \hat{x} = A^T \hat{y} \qquad (17)$$

This undetermined problem can also be viewed and treated in a suboptimal way as the following stochastic linear estimation problem:

$$0 = x+\overline{\eta} \quad , \ E[\overline{\eta}] = 0, E[\overline{\eta}\,\overline{\eta}^T] = \overline{P} = I_n \qquad (18)$$

$$b = Ax+\varepsilon \ , \ E[\varepsilon] = 0, E[\varepsilon\varepsilon^T] = R_\varepsilon \qquad (19)$$

where I_n is the identity matrix of order n; and R_ε is a diagonal matrix with variances considered to be negligible as compared to the dispersion caused by the variances of $\overline{P}$.

In a parallel iterative solution scheme, this problem can be viewed , in a typical iteration, as one of optimal linear estimation , with a priori information given by the jth component of :

$$-\alpha\overline{x} = (x - \overline{x})+\eta \qquad (20a)$$

and observations given by, for $i=1,2,...,m$:

$$\alpha(b_i - \sum_{l=1}^{n} a_{il}\overline{x}_l) = a_{ij}(x_j - \overline{x}_j)+v_i \qquad (20b)$$

where, based on the experience with the determined case, and unless of constant factors, v_i is modeled as zero mean, with uncorrelated components and covariance matrix R as defined in Eq. (8).

The solution to this problem is now given by (see, for example, Liebelt, 1967):

$$\hat{x}_j = \overline{x}_j +\alpha[I + A_j^T R^{-1} A_j]^{-1}[A_j^T R^{-1}(b-A\overline{x})-\overline{x}_j] \qquad (21)$$

where R is defined as before, in Eq. (8). Combining the estimates for $\hat{x}_j, j =1,2,...,n$, there results an estimator $\hat{x}$ for x of the form of Eq.(9), but where:

$$S = diag.[[I + A_j^T R^{-1} A_J]^{-1} : j =1,2,...,n] \qquad (22a)$$

$$f(x) = 1/2\{[Ax-b]^T R^{-1}[Ax-b]+x^T x \} \qquad (22b)$$

where α as before can be chosen in each iteration such as to minimize $f(x)$ in Eq.(22b):

$$\alpha = (\overline{g}^T S\overline{g})/(\overline{g}^T S(A^T R^{-1}A + I_n)S\overline{g}),$$
$$\overline{g} = \overline{g}(\overline{x})=\nabla f^T (\overline{x}). \qquad (23)$$

Since S and $(A^T R^{-1}A + I_n)$ are positive definite matrices, it results that the parallel iterative approximation given by Eq.(21) is again equivalent to a modified Newton method applied now to the functional of Eq.(22). Convergence to the solution of the $f(x)$ of Eq. (22b) can be demonstrated following steps analogous to those of Section 2.1.

2.3 Preliminary Numerical Testing

For the case of determined systems, the proposed method is a generalization of the unique of this type Jacobi's method, thus it is relative to this method that an evaluation has to be made.

The following results illustrate the numerical behavior of the method when applied to the solution of simple and well conditioned determined and undetermined problems. Tests conducted with an ill conditioned problem (condition number

1441) were also done. As expected, Jacobi's method diverged. Though the proposed method also converged in this case, it did with too many iterations (thousands), indicating the need of further research to explore the possibility of getting more efficient algorithms.

In all the cases tested, the numerical calculation of the optimal α using Eq.(12), when the gradient $\overline{g}$ gets near zero, and to avoid a numerical ill behavior, is done as follows:

$$\overline{g}_i = sign(\overline{g}_i).e_i, \quad if \quad /\overline{g}_i/ \le e_i.$$

(i) Determined Case: Well-Conditioned

$$A = \begin{bmatrix} 10 & 2 & 1 \\ 1 & 5 & 1 \\ 2 & 3 & 10 \end{bmatrix} \qquad b = \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}$$

$$x_0 = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}^T \qquad e = \begin{bmatrix} 1.0e\text{-}5 & 1.0e\text{-}5 & 1.0e\text{-}5 \end{bmatrix}$$

Figure 1 – Variables convergence for the proposed method

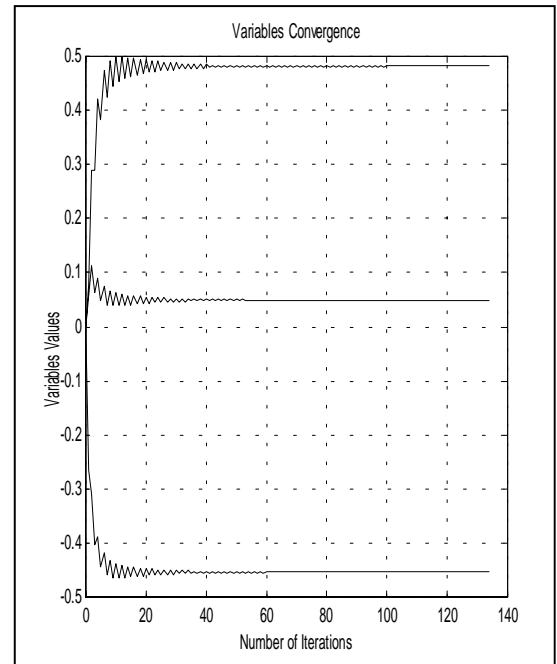**Figure 2 – Numerical residue for the proposed method**

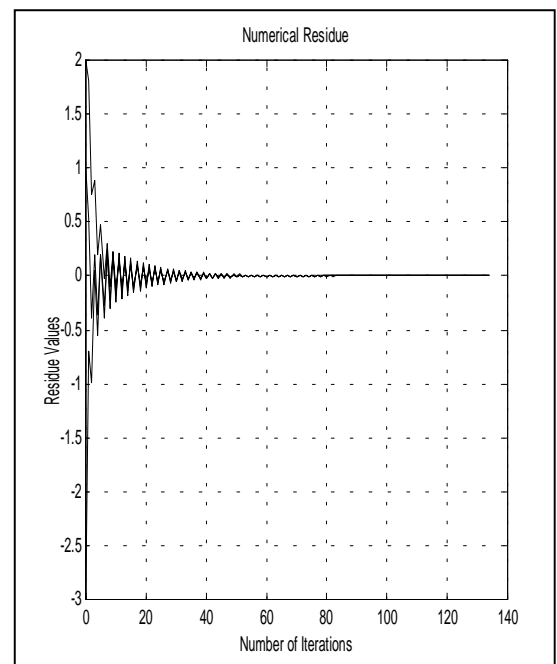**Figure 3 – Variables convergence for the Jacobi method**

Figure 4 – Numerical residue for the Jacobi method

Figures 1 and 2 illustrate a convergence behavior for the proposed method which is significantly better than that illustrated in Figures 3 and 4 for the Jacobi's method when applied to the same problem.

(ii) Determined Case: Ill-Conditioned

$$A = \begin{bmatrix} 6 & 13 & -17 \\ 13 & 29 & -38 \\ -17 & -38 & 50 \end{bmatrix} \qquad b = \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}$$
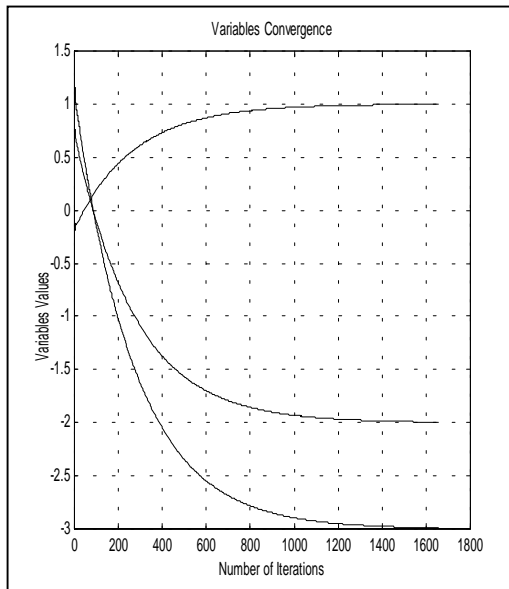
$$x_0 = [1\ 1\ 1]^T \qquad e = [1.0e\text{-}5 \quad 1.0e\text{-}5 \quad 1.0e\text{-}5]^T$$

(iii) Undetermined Case: Optimal Solution

$$A = \begin{bmatrix} 10 & 2 & 1 \\ 1 & 5 & 1 \end{bmatrix} \qquad b = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$x_0 = [0\ 0\ 0]^T \qquad e = [1.0e\text{-}5 \quad 1.0e\text{-}5 \quad 1.0e\text{-}5]^T$$

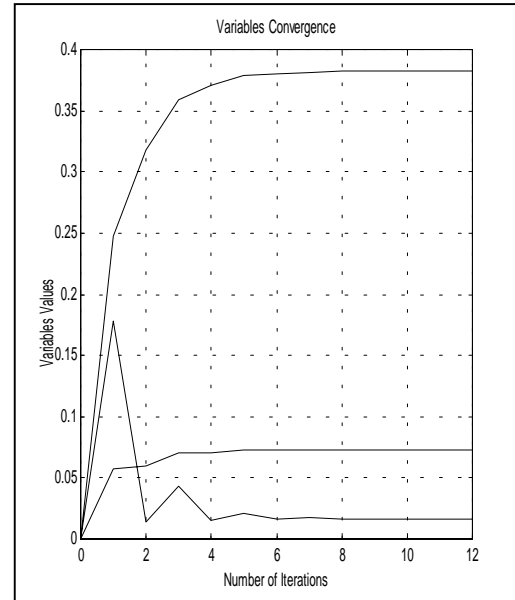**Figure 5 –Variable convergence for the proposed method: ill-conditioned case**

Figure 7 – Variables convergence for the proposed method: optimal solution

Figure 6 – Numerical residue for the proposed method: ill-conditioned
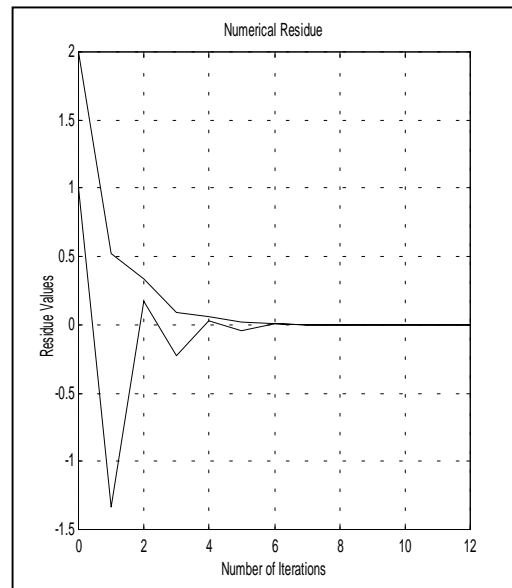

Figure 8 – Numerical residue for the proposed method: optimal solution

(iv) Undetermined Case: Suboptimal Solution

$$A = \begin{bmatrix} 10 & 2 & 1 \\ 1 & 5 & 1 \end{bmatrix} \qquad b = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$x_0 = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}^T \qquad e = \begin{bmatrix} 1.0e\text{-}5 & 1.0e\text{-}5 & 1.0e\text{-}5 \end{bmatrix}^T$$
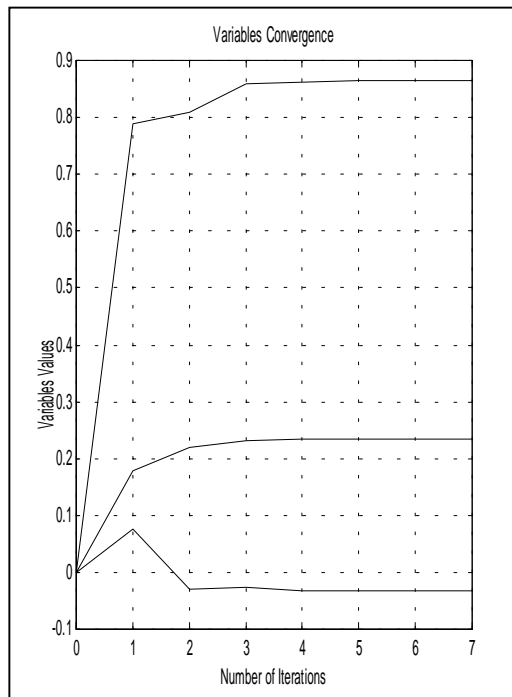

Figure 9 – Variables convergence for the proposed method: suboptimal solution
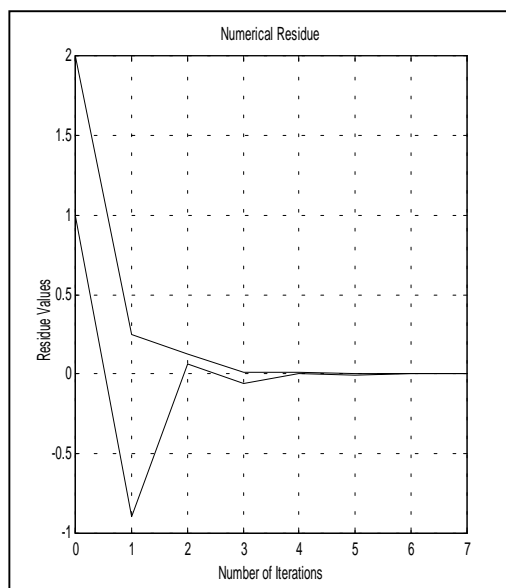

Figure 10 – Numerical residue for the proposed method: suboptimal solution

As it would be expected, the suboptimal solution does not coincide with the optimal one, and can be viewed as a the pseudo inverse generated by the criterion corresponding to the minimization of the functional of Eq.(22b).

3   CONCLUSIONS

Stochastic optimal linear estimation was used to develop a new method for the parallel solution of linear algebraic systems of equations. This was done exploring analogies with the problem of supervised training of artificial neural networks when local training Kalman algorithms are used.

For the case of determined systems, the method resulted to be a generalization of the only existing method of the type, which is the Jacobi's method.

Due to the particular diagonal structure of matrix S (Eqs. (10) and (22a)) the method was also shown to be in each iteration numerically equivalent to a modified Newton method which attains convergence.

Since the time spent to converge in each variable is of about the same order, then it is expected that, except for small differences due to the calculation of the α´s, this should also be the order of the time spent to solve the whole system, independently of the number of equations, as long as there are enough parallel processors.

An interesting feature that results from adopting the more general approach of stochastic optimal linear estimation is that the S matrix in the resulting modified Newton Method (see Eq.(9)) can be naturally and easily obtained using a criterion of statistical consistency . As a matter of fact, with the particular criterion chosen S resulted to be iteration independent.

In this phase of research, the objective was to explore the possibilities of optimal linear estimation to have a method of parallel processing solution of linear systems developed. Further studies shall address the numerical testing and evaluation of efficiency of algorithms based on this method. The higher level of sophistication of the developed method as compared to the Jacobi method and the good performance of algorithms derived from its analogous, Kalman filtering based, parallel processing, neural networks training method, rises the expectation of a satisfactory performance.

ACKNOWLEDGEMENT

REFERENCES

Abaffy, J. and Spedicato, E. (1984). A Generalization of Huang's Method for Solving Systems of Linear Algebraic Equations. *Bull. Un. Mat. Ital.,*6, 517-519.

Björck, A. (1996) *Numerical Methods for Least Squares Problems. SIAM-* Society for Industrial and Applied Mathematics, Philadelphia.

Chen, S. and Billings, S.A. (1992). Neural Networks for Nonlinear Dynamic System Modelling and Identification. *Int. J. Control,* 56(2), 319-346.

Huang, H.Y. (1975). A Direct Method for the General Solution of a System of Equations. *J. Optim. Theory Appl., 16(5/6).*

Jazwinski, A.H. (1970). *Stochastic Process and Filtering Theory.* Academic Press, New York.

Liebelt, P.B. (1967). *An Introduction to Optimal Estimation.* Addison-Wesley, Reading, MA.

Luenberger, D.G. (1984). *Linear and Nonlinear Programming.* Addison-Wesley, Reading, MA.

Miranda, W.M. (1997). *Associação de Técnicas de Tratamento de Resíduo a Algoritmos da Classe ABS para a Solução de Sistemas Lineares.* M.S. Thesis. Universidade Federal de Minas Gerais, Belo Horizonte, M.G., Brazil.

Pinto, R.L.U.F. and Rios Neto, A. (1990). An Optimal Linear Estimation Approach to Solve Systems of Linear Algebraic Equations. *Journal of Computational and Applied Mathematics,* 33, 261-268.

Spedicato, E. and Bodon, E. (1993). Solution of Linear Least Squares Via the ABS Algorithm. *Mathematical Programming,* 58, 111-136.

Spedicato,E. (1995). ABS Algorithms for Linear Systems and Linear Least Squares: Theoretical and Computational Performance. *Scientia Iranica,* 1(4), 289-303.

Stoer, J. and Bulirsch, R. (1980). *Introduction to Numerical Analysis.* Springer Verlag, N. Y.

Rios Neto, A. (1981). Filtro de Kalman Aplicado a Solução de Sistemas Lineares e Inversão de Matrizes. *33a Reunião Anual da SBPC,* Salvador, Bahia, Brazil, communication no. 19-A.8.1, p. 94. Supplement of *Ciência e Cultura,* 33(7).

Rios Neto, A. (1995). Kalman Filtering Stochastic Optimal Estimation Algorithms and Usual Backpropagation in Neural Nets Training. *Proc. Second Brazilian Congress in Neural Networks,* Curitiba, PR, Brazil, 139-144.

Rios Neto, A. (1997). Stochastic Optimal Linear Parameter Estimation and Neural Nets Training in Systems Modelling. *RBCM- J. of the Braz. Soc. Mechanical Sciences,* XIX(2),138-146.