A note on the equivalence of Broyden and Block ABS methods

Wesley V. I. Shirabayashi
Department of Mathematics, State University of Maringá,
87020-900, Maringá, PR, Brazil
wesleyvagner@gmail.com

Ronaldo Lopes
IMECC, University of Campinas, 13083-859, Campinas, SP, Brazil
ronaldolps3@gmail.com

Emerson Vitor Castelani*
Department of Mathematics, State University of Maringá
87020-900, Maringá, PR, Brazil
evcastelani@uem.br

Abstract

In this paper we are considering a scaled linear system and presenting a result of equivalence between Broyden's Method and Block ABS Method applied to this system. As consequence, we have an alternative proof of finite termination for Broyden's method [6].

keywords. ABS Block Methods, Broyden's Method, Scaled Linear System

1 Introduction

The study about methods for solving linear systems is not new, but the growing emergence of real-world problems formulated under this optics has made these studies focused on practical efficacy and theoretical development of these methods. There are several work about this subject and we recommend [7, 5, 10, 8] and the references therein.

In this paper we are approaching one generalization of ABS methods. This class was proposed to solve determined and undetermined linear systems by J. Abaffy, C. Broyden and E. Spedicato [3], in 1984.

In [2], Abaffy and Spedicato proved that ABS methods (introduced in [3]) when used for n-squared systems and by certain choices of parameters have the property that 2 steps of Broyden's method are equivalent to one step of ABS method. Since ABS methods have finite termination, they showed an alternative proof of finite termination for Broyden's methods, complementing the early results proposed by Gay [6] in 1979.

Recently, Adib [4] by considering a scaled version of ABS methods for systems of type:

$$V^T Ax = V^T b, \tag{1}$$

*Corresponding author

where $V = V_{n\times n}$ is nonsingular, proved that by certain choice of parameters this version of ABS algorithm is related to Broyden's method too. Scaled linear systems appears naturally in ill-conditioned problems so it is a relevant topic in several practical algorithms.

Another important contribution given in [2] is the introduction of Block ABS methods class for solving scaled linear systems. In i-step ABS method we find the solution for the $i-1$ first equations. The Block ABS method is more flexible than ABS method because we can choose many equations to solve in each step. This flexible feature is obtained by segmentation of certain matrix by certain submatrix.

The goal of this paper is to consider the scaled system 1 and show an equivalence between Broyden's algorithm and Block ABS methods. Consequently we are complementing the results given in [4].

2 Background Material

Consider the linear system of equation $Ax = b$, where $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$ and $x \in \mathbb{R}^n$. Let $g : \mathbb{R}^n \to \mathbb{R}$ defined by $g(x) = Ax - b$. We begin this section by describing the Broyden's method (Algorithm 1).

Algorithm 1 Broyden's algorithm (see [6])

Choose a vector $x_0 \in \mathbb{R}^n$ as an initial guess and $H_0 \in \mathbb{R}^{n\times n}$ as nonsingular matrix
for $k = 0, 1, 2, ...$ **do**
    Compute $s_k = -H_k g(x^k)$
    Update $x_{k+1} = x_k + s_k$
    Compute $c_k = g(x_{k+1}) - g(x_k)$
    if $c_k = 0$ **then**
        $H_{k+1} = H_k$
    else
        Choose $z_k \in \mathbb{R}^n$ such that ${z_k}^T c_k = 1$ and ${z_k}^T H^{-1} s_k \neq 0$
        Update $H_{k+1} = H_k + (s_k - H_k y_k){z_k}^T$
    end if
end for

Note that H_k is a nonsingular matrix for all k. Furthermore, considering the difference between $g(x_{k+1})$ and $g(x_k)$, we have:

$$c_k = g(x_{k+1}) - g(x_k) = Ax_{k+1} - b - (Ax_k - b) = As_k = -AH_k g(x_k).$$

Consequently,

$$g(x_{k+1}) = g(x_k) + c_k = g(x_k) - AH_k g(x_k) = (I - AH_k)g(x_k).$$

Let us define $F_k = I - AH_k$ and to present an important property related to this sequence.

Lemma 1. *Let us suppose $k \leq 2n-1$ odd. Let $c_k \neq 0, z_k^T c_{j-1} \neq 0$, $j = 1, ..., k$, $z_0 \in Im(F_0^T)$, $F_0 = I - AH_0 \in \mathbb{R}^{n\times n}$ is nonsingular and $l \leq 2n-1$ are odd indices. Then there exits a sequence $\{t_l\}$ of linearly independent vectors satisfying*

$$t_1^T F_j = 0, \forall j \geq 1,$$

$$t_l^T F_j = t_{l-2}^T, l = 3, 5, 7, ..., k, \forall j \geq l.$$

Moreover $t_i^T g(x_j) = 0, j = i+1, i+2, ..., k.$

Proof. We recommend [9] for a detailed proof. □

To elucidate the idea of Block scaled ABS method let us consider the Algorithm 2.

Algorithm 2 Block ABS algorithm (see [2])

Choose a vector $y_0 \in \mathbb{R}^n$ as an initial guess, $K_0 \in \mathbb{R}^{n\times n}$, $V = [V_0\ V_1 ... V_r] \in \mathbb{R}^{n\times n}$ as nonsingular matrix and $V_k \in \mathbb{R}^{n\times s_k}, k = 1, .., r$ where $s_1 + ... + s_r = n$
for $k = 0, 1, 2, ..., r$ **do**
  Choose $Z_k \in \mathbb{R}^{n\times n}$ and $q_k \in \mathbb{R}^{s_k}$ such that

$$V_k^T A P_k q_k = V_k^T \overline{g}_k \text{ where}$$

$$P_k = K_k^T Z_k \text{ and } \overline{g}_k = Ay_k - b \in \mathbb{R}^n$$

  Update $y_{k+1} = y_k - P_k q_k$
  Choose $W_k \in \mathbb{R}^{n\times s_k}$ satisfying $[W_k^T K_k A^T V_k] = I_{s_k\times s_k}$
  Update K_k by

$$K_{k+1} = K_k - K_k A^T V_k W_k^T K_k \in \mathbb{R}^{n\times n}.$$

end for

By simplicity we are considering $g_k = g(x_k)$. Related to Algorithm 2, we have important properties.

Lemma 2. *The following conditions hold:*

i. $K_j A^T V_k = 0$, $\forall k < j$;

ii. $K_i K_j = K_j K_i = K_j$ *if* $i \leq j$;

iii. If A *is full rank then*

$$rank(K_k A^T V_k | K_k A^T V_{k+1} | \dots | K_k A^T V_r) = n - (s_0 + s_1 + \dots s_{k-1});$$

iv. If $A \in R^{n\times n}$ *is nonsingular, then*
$K_k A^k V_j, j > k$ *spans* $Im(K_k)$ *and*
$A^T V_j, j > k$ *spans* $Kernel(K_k)$.

Proof. We recommend [1]. □

3 Main Result

This section is devoted to show an important relation between Broyden's Algorithm and Block ABS Algorithm. This relationship is formalized by the following theorem.

Theorem 1. *Consider* $Ax = b$ *a linear system where* $A \in \mathbb{R}^{n\times n}$ *is nonsingular. Let* z_k *and* c_k *as defined in Broyden's algorithm satisfying the Lemma 1. Let* (x_0, H_0) *and* (y_0, K_0) *initial choices for Algorithm 1 and 2 respectively with* $x_0 = y_0$. *Then, for* $2m_j$ *steps of Algorithm 1 it is possible to find* Z_j, W_j *and* V_j *such that*

$$x_{2m_j} = y_{j+1}$$

where m_j *is the number of columns of* $\overline{V}_j = [V_0\ V_1 ... V_j]$. *Consequently,*

$$x_{2m_r} = y_{r+1}, \text{ where } m_r = n.$$

Proof. Let $V = (t_1 \ t_3 ... t_{2n-1})$ with vectors $t_i' s$ linearly independent which satisfy the Lemma 1. The proof will be made using induction arguments on indices of x and y. First, we will show there exist some choices of parameters in Algorithm 2 that obtain:

$$x_{2m_0} = y_1, \tag{2}$$

where m_0 represents the number of columns of $\overline{V_0}$ defined by $V_0 \in \mathbb{R}^{n \times m_0}$. Now, using Algorithm 1 we have

$$x_{p+1} = x_p + s_p \text{ with } s_p = -H_p g_p. \tag{3}$$

Consequently

$$x_{2m_0} = x_0 - d_1, \text{ where } d_1 = \sum_{i=1}^{2m_0-1} H_i g_i. \tag{4}$$

By step 2 of Algorithm 2 we have

$$y_1 = y_0 - K_0^T Z_0 q_0. \tag{5}$$

So, we need to show that there exist choices for Z_0, q_0 such that

$$K_0^T Z_0 q_0 = d_1 \tag{6}$$

and

$$V_0^T A K_0 Z_0 q_0 = V_0^T \overline{g}_0. \tag{7}$$

The choices to satisfy (6) follow immediately from nonsingularity of K_0. Thus we need to show that these choices satisfy (7). In this case, we need to prove that

$$V_0^T A d_1 = V_0^T \overline{g}_0.$$

In fact,

$$V_0^T A d_1 = V_0^T A \left(\sum_{i=0}^{2m_0-1} H_i g_i \right) = \sum_{i=0}^{2m_0-1} V_0^T A H_i g_i \tag{8}$$

or equivalently

$$V_0^T A d_1 = \begin{bmatrix} \sum_{i=0}^{2m_0-1} t_1^T A H_i g_i \\ \sum_{i=0}^{2m_0-1} t_3^T A H_i g_i \\ \vdots \\ \sum_{i=0}^{2m_0-1} t_{2m_0-1}^T A H_i g_i \end{bmatrix}.$$

Using Lemma 1 we have

$$t_1^T = t_1^T A H_j, \forall j \geq 1, \tag{9}$$

$$\vdots$$

$$(t_k^T - t_{k-2}^T) = t_k^T A H_j, \forall j \geq k \geq 3, \tag{10}$$

$$t_k^T g_j = 0, \forall j > k. \tag{11}$$

Consequently,

$$\sum_{i=0}^{2m_0-1} t_1^T AH_i g_i = t_1^T AH_0 g_0 + t_1^T AH_1 g_1 + \sum_{i=2}^{2m_0-1} t_1^T AH_i g_i \tag{12}$$

$$= t_1^T AH_0 g_0 + t_1^T AH_1 g_1 + \sum_{i=2}^{2m_0-1} t_1^T g_i \tag{13}$$

$$= t_1^T AH_0 g_0 + t_1^T AH_1 g_1. \tag{14}$$

By Algorithm 1 we have that $g_{k+1} = (I - AH_k)g_k$, for all $k \geq 0$. Thus

$$\begin{aligned} \sum_{i=0}^{2m_0-1} t_1^T AH_i g_i &= t_1^T AH_0 g_0 + t_1^T AH_1 (I - AH_0) g_0 \\ &= t_1^T AH_0 g_0 + t_1^T (I - AH_0) g_0 \\ &= t_1^T AH_0 g_0 + t_1^T g_0 - t_1^T AH_0 g_0 \\ &= t_1^T g_0. \end{aligned} \tag{15}$$

For the general case, let's consider $j \leq 2m_0 - 1$ and odd. Therefore

$$\begin{aligned} \sum_{i=0}^{2m_0-1} t_j^T AH_i g_i &= \sum_{i=0}^{j} t_j^T AH_i g_i + \sum_{i=j+1}^{2m_0-1} t_j^T AH_i g_i \\ &= \sum_{i=0}^{j} t_j^T AH_i g_i + \sum_{i=j+1}^{2m_0-1} (t_j^T - t_{j-2}^T) g_i \\ &= \sum_{i=0}^{j} t_j^T AH_i g_i. \end{aligned} \tag{16}$$

Using (9) and (10) we have

$$\begin{aligned} \sum_{i=0}^{j} t_j^T AH_i g_i &= \sum_{i=0}^{j-2} t_j^T AH_i g_i + t_j^T AH_{j-1} g_{j-1} + t_j^T AH_j g_j \\ &= \sum_{i=0}^{j-2} t_j^T AH_i g_i + t_j^T AH_{j-1} g_{j-1} + t_j^T g_j - t_{j-2}^T g_j \\ &= \sum_{i=0}^{j-2} t_j^T AH_i g_i + t_j^T AH_{j-1} g_{j-1} + t_j^T g_j \\ &= \sum_{i=0}^{j-2} t_j^T AH_i g_i + t_j^T AH_{j-1} g_{j-1} + t_j^T g_{j-1} - t_j^T AH_{j-1} g_{j-1} \\ &= \sum_{i=0}^{j-2} t_j^T AH_i g_i + t_j^T g_{j-1}. \end{aligned} \tag{17}$$

With similar ideas of (17) we can show that:

$$\sum_{i=0}^{l} t_j^T AH_i g_i + t_j^T g_{l+1} = \sum_{i=0}^{l-1} t_j^T AH_i g_i + t_j^T g_l, \tag{18}$$

for all $l \leq j$. In fact,

$$\begin{aligned}
\sum_{i=0}^{l} t_j^T AH_i g_i + t_j^T g_{l+1} &= \sum_{i=0}^{l-1} t_j^T AH_i g_i + t_j^T AH_l g_l + t_j^T g_{l+1} \\
&= \sum_{i=0}^{l-1} t_j^T AH_i g_i + t_j^T AH_l g_l + t_j^T (I - AH_l) g_l \\
&= \sum_{i=0}^{l-1} t_j^T AH_i g_i + t_j^T AH_l g_l + t_j^T g_l - t_j^T AH_l g_l \\
&= \sum_{i=0}^{l-1} t_j^T AH_i g_i + t_j^T g_l. \qquad (19)
\end{aligned}$$

Now, using (18) recursively we obtain

$$\begin{aligned}
\sum_{i=0}^{2m_0-1} t_j^T AH_i g_i &= t_j^T AH_0 g_0 + t_j^T g_1 \\
&= t_j^T AH_0 g_0 + t_j^T (I - AH_0) g_0 \\
&= t_j^T AH_0 g_0 + t_j^T g_0 - t_j^T AH_0 g_0 \\
&= t_j^T g_0. \qquad (20)
\end{aligned}$$

Consequently,

$$V_0^T Ad_1 = \begin{bmatrix} \sum_{i=0}^{2m_0-1} t_1^T AH_i g_i \\ \sum_{i=0}^{2m_0-1} t_3^T AH_i g_i \\ \vdots \\ \sum_{i=0}^{2m_0-1} t_{2m_0-1}^T AH_i g_i \end{bmatrix} = \begin{bmatrix} t_1^T g_0 \\ t_3^T g_0 \\ \vdots \\ t_{2m_0-1}^T g_0 \end{bmatrix} = V_0^T g_0 = V_0^T \overline{g}_0,$$

terminating the first part of the induction argument.

Let's assume that

$$x_{2m_k} = y_{k+1}, \forall \quad 1 \leq k < r,$$

where m_k represents the number of columns of $\overline{V}_k = [V_0 \ V_1 ... V_k]$. We will show $x_{2m_r} = y_{r+1}$.

Considering the Algorithm 1, we have

$$x_{2m_r} = x_{2m_{(r-1)}} - d_r, \text{ where } d_r = \sum_{i=2m_{(r-1)}}^{2m_r-1} H_i g_i \qquad (21)$$

and considering the Algorithm 2, we have

$$y_{r+1} = y_r - K_r^T Z_r q_r. \qquad (22)$$

We need to show that exist Z_r and q_r such that

$$K_r^T Z_r q_r = d_r, \tag{23}$$

with $V_r^T A K_r^T Z_r q_r = V_r^T \overline{g}_r$. By Lemma 2, we have the equivalence:

$$\begin{aligned} K_r^T Z_r q_r = d_r \quad &\Leftrightarrow \quad d_r \in Im(K_r^T) \\ &\Leftrightarrow \quad d_r \perp \{A^T V_0, A^T V_1, \ldots A^T V_{r-1}\} \\ &\Leftrightarrow \quad d_r \perp A^T V_k, \text{ for all } 0 \leq k \leq r-1. \end{aligned} \tag{24}$$

Consequently it is sufficient to show

$$V_k^T A d_r = 0, \text{ for all } 0 \leq k \leq r-1. \tag{25}$$

Observe that

$$\overline{V_{r-1}}^T A d_r = \begin{bmatrix} V_0^T A d_r \\ V_1^T A d_r \\ \vdots \\ V_{r-1}^T A d_r \end{bmatrix} = \begin{bmatrix} t_1^T A d_r \\ t_3^T A d_r \\ \vdots \\ t_{2m_{(r-1)}-1}^T A d_r \end{bmatrix}. \tag{26}$$

Using, again, (9)-(11) we can obtain

$$t_1^T A d_r = t_1^T \left(\sum_{i=2m_{(r-1)}}^{2m_r-1} H_i g_i \right) = \sum_{i=2m_{(r-1)}}^{2m_r-1} t_1^T A H_i g_i = \sum_{i=2m_{(r-1)}}^{2m_r-1} t_1^T g_i = 0 \tag{27}$$

and with similar arguments, for $j \leq 2m_{(r-1)} - 1$ odd we have

$$t_j^T A d_r = t_j^T \left(\sum_{i=2m_{(r-1)}}^{2m_r-1} H_i g_i \right) = \sum_{i=2m_{(r-1)}}^{2m_r-1} t_j^T A H_i g_i = \sum_{i=2m_{(r-1)}}^{2m_r-1} t_j^T g_i = 0. \tag{28}$$

Thus, from (26), (27) and (28), we conclude (25). We need to prove that the choices of Z_r and q_r satisfy $V_r^T A K_r^T Z_r q_r = V_r^T \overline{g}_r$, or equivalently, $V_r^T A d_r = V_r^T \overline{g}_r$.

Considering $V_r = [t_{2m_{(r-1)}+1} \; t_{2m_{(r-1)}+3} \ldots t_{2m_r-1}]$ we have

$$V_r^T A d_r = \begin{bmatrix} t_{2m_{(r-1)}+1}^T A d_r \\ t_{2m_{(r-1)}+3}^T A d_r \\ \vdots \\ t_{2m_r-1}^T A d_r \end{bmatrix}. \tag{29}$$

Since $d_r = \sum_{i=2m_{(r-1)}}^{2m_r-1} H_i g_i$ we can obtain

$$\begin{aligned} t_{2m_{(r-1)}+1}^T A d_r &= t_{2m_{(r-1)}+1}^T A \left(\sum_{i=2m_{(r-1)}}^{2m_r-1} H_i g_i \right) \\ &= \sum_{i=2m_{(r-1)}}^{2m_r-1} t_{2m_{(r-1)}+1}^T A H_i g_i \\ &= \sum_{i=2m_{(r-1)}}^{2m_{(r-1)}+1} t_{2m_{(r-1)}+1}^T A H_i g_i + \sum_{i=2m_{(r-1)}+2}^{2m_r-1} t_{2m_{(r-1)}+1}^T A H_i g_i. \end{aligned} \tag{30}$$

Note that $2m_{(r-1)}+1 < 2m_{(r-1)}+2$ and using Lemma 1 we get from (30) that

$$\begin{aligned} t^T_{2m_{(r-1)}+1}Ad_r &= \sum_{i=2m_{(r-1)}}^{2m_{(r-1)}+1} t^T_{2m_{(r-1)}+1}AH_ig_i + \sum_{i=2m_{(r-1)}+2}^{2m_r-1}(t_{2m_{(r-1)}+1}-t_{2m_{(r-1)}-1})g_i \\ &= \sum_{i=2m_{(r-1)}}^{2m_{(r-1)}+1} t^T_{2m_{(r-1)}+1}AH_ig_i \\ &= t_{2m_{(r-1)}+1}AH_{2m_{(r-1)}}g_{2m_{(r-1)}} + (t_{2m_{(r-1)}+1}-t_{2m_{(r-1)}-1})g_{2m_{(r-1)}+1} \\ &= t_{2m_{(r-1)}+1}AH_{2m_{(r-1)}}g_{2m_{(r-1)}} + t_{2m_{(r-1)}+1}(I-AH_{2m_{(r-1)}})g_{2m_{(r-1)}} \\ &= t^T_{2m_{(r-1)}+1}g_{2m_{(r-1)}}. \end{aligned} \tag{31}$$

With similar argument given in (16)-(20) we can conclude

$$t^T_kAd_r = t^T_kg_{2m_{(r-1)}}, \text{ for all } 2m_{(r-1)}+1 \le k \le 2m_r-1. \tag{32}$$

Since $g_{2m_{(r-1)}} = \overline{g}_r$, we obtain

$$V_r^TAd_r = \begin{bmatrix} t^T_{2m_{(r-1)}+1}Ad_r \\ t^T_{2m_{(r-1)}+3}Ad_r \\ \vdots \\ t^T_{2m_r-1}Ad_r \end{bmatrix} = \begin{bmatrix} t^T_{2m_{(r-1)}+1}g_{2m_{(r-1)}} \\ t^T_{2m_{(r-1)}+3}g_{2m_{(r-1)}} \\ \vdots \\ t^T_{2m_r-1}g_{2m_{(r-1)}} \end{bmatrix}. \tag{33}$$

□

4 Conclusion

This work has the same direction of [4]. However, we consider the Block ABS Method instead of ABS method and we show that $2n$ steps of Broyden's Method is equivalent to one step of the ABS Block Method, when the block size is n. Since the ABS Block Method has finite termination, we provide another proof for Gay's theorem [6].

References

[1] J. Abaffy and A. Galantai. Conjugate Direction Methods for Linear and Nonlinear Systems of Algebraic Equations. In *Colloquia Mathematica Societatis Janos Bolyai*, pages 481–502, 1986.

[2] J. Abaffy and E. Spedicato. *ABS Projection Algorithms: mathematical techniques for linear and nonlinear equations*. Halsted Press, 1989.

[3] Joszef Abaffy, Charles Broyden, and Emilio Spedicato. A class of direct methods for linear systems. *Numerische Mathematik*, 45(3):361–376, 1984.

[4] Majid Adib. A handy proof of Gays theorem. *Numerical Linear Algebra with Applications*, 15(April):637–642, 2008.

[5] Howard C Elman. *Iterative methods for large, sparse, nonsymmetric systems of linear equations*. PhD thesis, Yale University New Haven, Conn, 1982.

[6] David M Gay. Some convergence properties of broyden's method. *SIAM Journal on Numerical Analysis*, 16(4):623–630, 1979.

[7] Frank L Lewis. A survey of linear singular systems. *Circuits, Systems and Signal Processing*, 5(1):3–36, 1986.

[8] J M Martinez. Practical quasi-Newton methods for solving nonlinear systems. *Journal of Computational and Applied Mathematics*, 124:97–121, 2000.

[9] Dianne P O'Leary. Why broyden's nonsymmetric method terminates on linear equations. *SIAM Journal on Optimization*, 5(2):231–235, 1995.

[10] Valeria Simoncini and Daniel B Szyld. Recent computational developments in krylov subspace methods for linear systems. *Numerical Linear Algebra with Applications*, 14(1):1–59, 2007.