

# NAG Library Routine Document

## F11BDF

**Note:** before using this routine, please read the Users' Note for your implementation to check the interpretation of *bold italicised* terms and other implementation-dependent details.

### 1 Purpose

F11BDF is a setup routine, the first in a suite of three routines for the iterative solution of a real general (nonsymmetric) system of simultaneous linear equations. F11BDF must be called before F11BEF, the iterative solver. The third routine in the suite, F11BFF, can be used to return additional information about the computation.

These three routines are suitable for the solution of large sparse general (nonsymmetric) systems of equations.

### 2 Specification

```

SUBROUTINE F11BDF (METHOD, PRECON, NORM, WEIGHT, ITERM, N, M, TOL,           &
                  MAXITN, ANORM, SIGMAX, MONIT, LWREQ, WORK, LWORK,       &
                  IFAIL)
INTEGER          ITERM, N, M, MAXITN, MONIT, LWREQ, LWORK, IFAIL
REAL (KIND=nag_wp) TOL, ANORM, SIGMAX, WORK(LWORK)
CHARACTER(*)     METHOD
CHARACTER(1)     PRECON, NORM, WEIGHT

```

### 3 Description

The suite consisting of the routines F11BDF, F11BEF and F11BFF is designed to solve the general (nonsymmetric) system of simultaneous linear equations  $Ax = b$  of order  $n$ , where  $n$  is large and the coefficient matrix  $A$  is sparse.

F11BDF is a setup routine which must be called before F11BEF, the iterative solver. The third routine in the suite, F11BFF, can be used to return additional information about the computation. A choice of methods is available:

- restarted generalized minimum residual method (RGMRES);
- conjugate gradient squared method (CGS);
- bi-conjugate gradient stabilized ( $\ell$ ) method (Bi-CGSTAB( $\ell$ ));
- transpose-free quasi-minimal residual method (TFQMR).

#### 3.1 Restarted Generalized Minimum Residual Method (RGMRES)

The restarted generalized minimum residual method (RGMRES) (see Saad and Schultz (1986), Barrett *et al.* (1994) and Dias da Cunha and Hopkins (1994)) starts from the residual  $r_0 = b - Ax_0$ , where  $x_0$  is an initial estimate for the solution (often  $x_0 = 0$ ). An orthogonal basis for the Krylov subspace  $\text{span}\{A^k r_0\}$ , for  $k = 0, 1, \dots$ , is generated explicitly: this is referred to as Arnoldi's method (see Arnoldi (1951)). The solution is then expanded onto the orthogonal basis so as to minimize the residual norm  $\|b - Ax\|_2$ . The lack of symmetry of  $A$  implies that the orthogonal basis is generated by applying a 'long' recurrence relation, whose length increases linearly with the iteration count. For all but the most trivial problems, computational and storage costs can quickly become prohibitive as the iteration count increases. RGMRES limits these costs by employing a restart strategy: every  $m$  iterations at most, the Arnoldi process is restarted from  $r_l = b - Ax_l$ , where the subscript  $l$  denotes the last available iterate. Each group of  $m$  iterations is referred to as a 'super-iteration'. The value of  $m$  is chosen in advance and is fixed throughout the computation. Unfortunately, an optimum value of  $m$  cannot easily be predicted.

### 3.2 Conjugate Gradient Squared Method (CGS)

The conjugate gradient squared method (CGS) (see Sonneveld (1989), Barrett *et al.* (1994) and Dias da Cunha and Hopkins (1994)) is a development of the bi-conjugate gradient method where the nonsymmetric Lanczos method is applied to reduce the coefficients matrix to real tridiagonal form: two bi-orthogonal sequences of vectors are generated starting from the residual  $r_0 = b - Ax_0$ , where  $x_0$  is an initial estimate for the solution (often  $x_0 = 0$ ) and from the *shadow residual*  $\hat{r}_0$  corresponding to the arbitrary problem  $A^T \hat{x} = \hat{b}$ , where  $\hat{b}$  can be any vector, but in practice is chosen so that  $r_0 = \hat{r}_0$ . In the course of the iteration, the residual and shadow residual  $r_i = P_i(A)r_0$  and  $\hat{r}_i = P_i(A^T)\hat{r}_0$  are generated, where  $P_i$  is a polynomial of order  $i$ , and bi-orthogonality is exploited by computing the vector product  $\rho_i = (\hat{r}_i, r_i) = (P_i(A^T)\hat{r}_0, P_i(A)r_0)$ . Applying the ‘contraction’ operator  $P_i(A)$  twice, the iteration coefficients can still be recovered without advancing the solution of the shadow problem, which is of no interest. The CGS method often provides fast convergence; however, there is no reason why the contraction operator should also reduce the once reduced vector  $P_i(A)r_0$ : this may well lead to a highly irregular convergence which may result in large cancellation errors.

### 3.3 Bi-Conjugate Gradient Stabilized ( $\ell$ ) Method (Bi-CGSTAB( $\ell$ ))

The bi-conjugate gradient stabilized ( $\ell$ ) method (Bi-CGSTAB( $\ell$ )) (see Van der Vorst (1989), Sleijpen and Fokkema (1993) and Dias da Cunha and Hopkins (1994)) is similar to the CGS method above. However, instead of generating the sequence  $\{P_i^2(A)r_0\}$ , it generates the sequence  $\{Q_i(A)P_i(A)r_0\}$ , where the  $Q_i(A)$  are polynomials chosen to minimize the residual *after* the application of the contraction operator  $P_i(A)$ . Two main steps can be identified for each iteration: an OR (Orthogonal Residuals) step where a basis of order  $\ell$  is generated by a Bi-CG iteration and an MR (Minimum Residuals) step where the residual is minimized over the basis generated, by a method akin to GMRES. For  $\ell = 1$ , the method corresponds to the Bi-CGSTAB method of Van der Vorst (1989). For  $\ell > 1$ , more information about complex eigenvalues of the iteration matrix can be taken into account, and this may lead to improved convergence and robustness. However, as  $\ell$  increases, numerical instabilities may arise. For this reason, a maximum value of  $\ell = 10$  is imposed, but probably  $\ell = 4$  is sufficient in most cases.

### 3.4 Transpose-free Quasi-minimal Residual Method (TFQMR)

The transpose-free quasi-minimal residual method (TFQMR) (see Freund and Nachtigal (1991) and Freund (1993)) is conceptually derived from the CGS method. The residual is minimized over the space of the residual vectors generated by the CGS iterations under the simplifying assumption that residuals are almost orthogonal. In practice, this is not the case but theoretical analysis has proved the validity of the method. This has the effect of remedying the rather irregular convergence behaviour with wild oscillations in the residual norm that can degrade the numerical performance and robustness of the CGS method. In general, the TFQMR method can be expected to converge at least as fast as the CGS method, in terms of number of iterations, although each iteration involves a higher operation count. When the CGS method exhibits irregular convergence, the TFQMR method can produce much smoother, almost monotonic convergence curves. However, the close relationship between the CGS and TFQMR method implies that the *overall* speed of convergence is similar for both methods. In some cases, the TFQMR method may converge faster than the CGS method.

### 3.5 General Considerations

For each method, a sequence of solution iterates  $\{x_i\}$  is generated such that, hopefully, the sequence of the residual norms  $\{\|r_i\|\}$  converges to a required tolerance. Note that, in general, convergence, when it occurs, is not monotonic.

In the RGMRES and Bi-CGSTAB( $\ell$ ) methods above, your program must provide the **maximum** number of basis vectors used,  $m$  or  $\ell$ , respectively; however, a **smaller** number of basis vectors may be generated and used when the stability of the solution process requires this (see Section 9).

Faster convergence can be achieved using a **preconditioner** (see Golub and Van Loan (1996) and Barrett *et al.* (1994)). A preconditioner maps the original system of equations onto a different system, say

$$\bar{A}\bar{x} = \bar{b}, \quad (1)$$

with, hopefully, better characteristics with respect to its speed of convergence: for example, the condition number of the coefficients matrix can be improved or eigenvalues in its spectrum can be made to coalesce. An orthogonal basis for the Krylov subspace  $\text{span}\{\bar{A}^k\bar{r}_0\}$ , for  $k = 0, 1, \dots$ , is generated and the solution proceeds as outlined above. The algorithms used are such that the solution and residual iterates of the original system are produced, not their preconditioned counterparts. Note that an unsuitable preconditioner or no preconditioning at all may result in a very slow rate, or lack, of convergence. However, preconditioning involves a trade-off between the reduction in the number of iterations required for convergence and the additional computational costs per iteration. Also, setting up a preconditioner may involve non-negligible overheads.

A *left* preconditioner  $M^{-1}$  can be used by the RGMRES, CGS and TFQMR methods, such that  $\bar{A} = M^{-1}A \sim I_n$  in (1), where  $I_n$  is the identity matrix of order  $n$ ; a *right* preconditioner  $M^{-1}$  can be used by the Bi-CGSTAB( $\ell$ ) method, such that  $\bar{A} = AM^{-1} \sim I_n$ . These are formal definitions, used only in the design of the algorithms; in practice, only the means to compute the matrix–vector products  $v = Au$  and  $v = A^T u$  (the latter only being required when an estimate of  $\|A\|_1$  or  $\|A\|_\infty$  is computed internally), and to solve the preconditioning equations  $Mv = u$  are required, i.e., explicit information about  $M$ , or its inverse is not required at any stage.

The first termination criterion

$$\|r_k\|_p \leq \tau \left( \|b\|_p + \|A\|_p \times \|x_k\|_p \right) \quad (2)$$

is available for all four methods. In (2),  $p = 1, \infty$  or  $2$  and  $\tau$  denotes a user-specified tolerance subject to  $\max(10, \sqrt{n})$ ,  $\epsilon \leq \tau < 1$ , where  $\epsilon$  is the *machine precision*. Facilities are provided for the estimation of the norm of the coefficients matrix  $\|A\|_1$  or  $\|A\|_\infty$ , when this is not known in advance, by applying Higham's method (see Higham (1988)). Note that  $\|A\|_2$  cannot be estimated internally. This criterion uses an error bound derived from **backward** error analysis to ensure that the computed solution is the exact solution of a problem as close to the original as the termination tolerance requires. Termination criteria employing bounds derived from **forward** error analysis are not used because any such criteria would require information about the condition number  $\kappa(A)$  which is not easily obtainable.

The second termination criterion

$$\|\bar{r}_k\|_2 \leq \tau (\|\bar{r}_0\|_2 + \sigma_1(\bar{A}) \times \|\Delta\bar{x}_k\|_2) \quad (3)$$

is available for all methods except TFQMR. In (3),  $\sigma_1(\bar{A}) = \|\bar{A}\|_2$  is the largest singular value of the (preconditioned) iteration matrix  $\bar{A}$ . This termination criterion monitors the progress of the solution of the preconditioned system of equations and is less expensive to apply than criterion (2) for the Bi-CGSTAB( $\ell$ ) method with  $\ell > 1$ . Only the RGMRES method provides facilities to estimate  $\sigma_1(\bar{A})$  internally, when this is not supplied (see Section 9).

Termination criterion (2) is the recommended choice, despite its additional costs per iteration when using the Bi-CGSTAB( $\ell$ ) method with  $\ell > 1$ . Also, if the norm of the initial estimate is much larger than the norm of the solution, that is, if  $\|x_0\| \gg \|x\|$ , a dramatic loss of significant digits could result in complete lack of convergence. The use of criterion (2) will enable the detection of such a situation, and the iteration will be restarted at a suitable point. No such restart facilities are provided for criterion (3).

Optionally, a vector  $w$  of user-specified weights can be used in the computation of the vector norms in termination criterion (2), i.e.,  $\|v\|_p^{(w)} = \|v^{(w)}\|_p$ , where  $(v^{(w)})_i = w_i v_i$ , for  $i = 1, 2, \dots, n$ . Note that the use of weights increases the computational costs.

The sequence of calls to the routines comprising the suite is enforced: first, the setup routine F11BDF must be called, followed by the solver F11BEF. F11BFF can be called either when F11BEF is carrying out a monitoring step or after F11BEF has completed its tasks. Incorrect sequencing will raise an error condition.

In general, it is not possible to recommend one method in preference to another. RGMRES is often used in the solution of systems arising from PDEs. On the other hand, it can easily stagnate when the size  $m$

of the orthogonal basis is too small, or the preconditioner is not good enough. CGS can be the fastest method, but the computed residuals can exhibit instability which may greatly affect the convergence and quality of the solution. Bi-CGSTAB( $\ell$ ) seems robust and reliable, but it can be slower than the other methods: if a preconditioner is used and  $\ell > 1$ , Bi-CGSTAB( $\ell$ ) computes the solution of the preconditioned system  $\bar{x}_k = Mx_k$ : the preconditioning equations must be solved to obtain the required solution. The algorithm employed limits to 10% or less, when no intermediate monitoring is requested, the number of times the preconditioner has to be thus applied compared with the total number of applications of the preconditioner. TFQMR can be viewed as a more robust variant of the CGS method: it shares the CGS method speed but avoids the CGS fluctuations in the residual, which may give rise to instability. Also, when the termination criterion (2) is used, the CGS, Bi-CGSTAB( $\ell$ ) and TFQMR methods will restart the iteration automatically when necessary in order to solve the given problem.

## 4 References

- Arnoldi W (1951) The principle of minimized iterations in the solution of the matrix eigenvalue problem *Quart. Appl. Math.* **9** 17–29
- Barrett R, Berry M, Chan T F, Demmel J, Donato J, Dongarra J, Eijkhout V, Pozo R, Romine C and Van der Vorst H (1994) *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods* SIAM, Philadelphia
- Dias da Cunha R and Hopkins T (1994) PIM 1.1 — the parallel iterative method package for systems of linear equations user's guide — Fortran 77 version *Technical Report* Computing Laboratory, University of Kent at Canterbury, Kent, UK
- Freund R W (1993) A transpose-free quasi-minimal residual algorithm for non-Hermitian linear systems *SIAM J. Sci. Comput.* **14** 470–482
- Freund R W and Nachtigal N (1991) QMR: a Quasi-Minimal Residual Method for Non-Hermitian Linear Systems *Numer. Math.* **60** 315–339
- Golub G H and Van Loan C F (1996) *Matrix Computations* (3rd Edition) Johns Hopkins University Press, Baltimore
- Higham N J (1988) FORTRAN codes for estimating the one-norm of a real or complex matrix, with applications to condition estimation *ACM Trans. Math. Software* **14** 381–396
- Saad Y and Schultz M (1986) GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems *SIAM J. Sci. Statist. Comput.* **7** 856–869
- Sleijpen G L G and Fokkema D R (1993) BiCGSTAB( $\ell$ ) for linear equations involving matrices with complex spectrum *ETNA* **1** 11–32
- Sonneveld P (1989) CGS, a fast Lanczos-type solver for nonsymmetric linear systems *SIAM J. Sci. Statist. Comput.* **10** 36–52
- Van der Vorst H (1989) Bi-CGSTAB, a fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems *SIAM J. Sci. Statist. Comput.* **13** 631–644

## 5 Arguments

- 1: METHOD – CHARACTER(\*) *Input*  
*On entry:* the iterative method to be used.
- METHOD = 'RGMRES'  
 Restarted generalized minimum residual method.
- METHOD = 'CGS'  
 Conjugate gradient squared method.
- METHOD = 'BICGSTAB'  
 Bi-conjugate gradient stabilized ( $\ell$ ) method.

METHOD = 'TFQMR'

Transpose-free quasi-minimal residual method.

*Constraint:* METHOD = 'RGMRES', 'CGS', 'BICGSTAB' or 'TFQMR'.

2: PRECON – CHARACTER(1)

*Input*

*On entry:* determines whether preconditioning is used.

PRECON = 'N'

No preconditioning.

PRECON = 'P'

Preconditioning.

*Constraint:* PRECON = 'N' or 'P'.

3: NORM – CHARACTER(1)

*Input*

*On entry:* defines the matrix and vector norm to be used in the termination criteria.

NORM = '1'

$l_1$  norm.

NORM = 'I'

$l_\infty$  norm.

NORM = '2'

$l_2$  norm.

*Suggested value:*

if ITERM = 1, NORM = 'I';

if ITERM = 2, NORM = '2'.

*Constraints:*

if ITERM = 1, NORM = '1', 'I' or '2';

if ITERM = 2, NORM = '2'.

4: WEIGHT – CHARACTER(1)

*Input*

*On entry:* specifies whether a vector  $w$  of user-supplied weights is to be used in the computation of the vector norms required in termination criterion (2) (ITERM = 1):  $\|v\|_p^{(w)} = \|v^{(w)}\|_p$ , where  $v_i^{(w)} = w_i v_i$ , for  $i = 1, 2, \dots, n$ . The suffix  $p = 1, 2, \infty$  denotes the vector norm used, as specified by the argument NORM. Note that weights cannot be used when ITERM = 2, i.e., when criterion (3) is used.

WEIGHT = 'W'

User-supplied weights are to be used and must be supplied on initial entry to F11BEF.

WEIGHT = 'N'

All weights are implicitly set equal to one. Weights do not need to be supplied on initial entry to F11BEF.

*Suggested value:* WEIGHT = 'N'.

*Constraints:*

if ITERM = 1, WEIGHT = 'W' or 'N';

if ITERM = 2, WEIGHT = 'N'.

5: ITERM – INTEGER

*Input*

*On entry:* defines the termination criterion to be used.

ITERM = 1

Use the termination criterion defined in (2).

ITERM = 2

Use the termination criterion defined in (3).

*Suggested value:* ITERM = 1.

*Constraints:*

if METHOD = 'TFQMR' or WEIGHT = 'W' or NORM  $\neq$  '2', ITERM = 1;  
otherwise ITERM = 1 or 2.

**Note:** ITERM = 2 is only appropriate for a restricted set of choices for METHOD, NORM and WEIGHT; that is NORM = '2', WEIGHT = 'N' and METHOD  $\neq$  'TFQMR'.

6: N – INTEGER *Input*

*On entry:*  $n$ , the order of the matrix  $A$ .

*Constraint:*  $N > 0$ .

7: M – INTEGER *Input*

*On entry:* if METHOD = 'RGMRES',  $M$  is the dimension  $m$  of the restart subspace.

If METHOD = 'BICGSTAB',  $M$  is the order  $\ell$  of the polynomial Bi-CGSTAB method.

Otherwise,  $M$  is not referenced.

*Constraints:*

if METHOD = 'RGMRES',  $0 < M \leq \min(N, 50)$ ;  
if METHOD = 'BICGSTAB',  $0 < M \leq \min(N, 10)$ .

8: TOL – REAL (KIND=nag\_wp) *Input*

*On entry:* the tolerance  $\tau$  for the termination criterion. If  $TOL \leq 0.0$ ,  $\tau = \max(\sqrt{\epsilon}, \sqrt{n}\epsilon)$  is used, where  $\epsilon$  is the *machine precision*. Otherwise  $\tau = \max(TOL, 10\epsilon, \sqrt{n}\epsilon)$  is used.

*Constraint:*  $TOL < 1.0$ .

9: MAXITN – INTEGER *Input*

*On entry:* the maximum number of iterations.

*Constraint:*  $MAXITN > 0$ .

10: ANORM – REAL (KIND=nag\_wp) *Input*

*On entry:* if  $ANORM > 0.0$ , the value of  $\|A\|_p$  to be used in the termination criterion (2) (ITERM = 1).

If  $ANORM \leq 0.0$ , ITERM = 1 and NORM = '1' or 'I', then  $\|A\|_1 = \|A\|_\infty$  is estimated internally by F11BEF.

If ITERM = 2, ANORM is not referenced.

*Constraint:* if ITERM = 1 and NORM = '2',  $ANORM > 0.0$ .

11: SIGMAX – REAL (KIND=nag\_wp) *Input*

*On entry:* if ITERM = 2, the largest singular value  $\sigma_1$  of the preconditioned iteration matrix; otherwise, SIGMAX is not referenced.

If  $SIGMAX \leq 0.0$ , ITERM = 2 and METHOD = 'RGMRES', then the value of  $\sigma_1$  will be estimated internally.

*Constraint:* if METHOD = 'CGS' or 'BICGSTAB' and ITERM = 2,  $SIGMAX > 0.0$ .

- 12: MONIT – INTEGER *Input*
- On entry:* if MONIT > 0, the frequency at which a monitoring step is executed by F11BEF: if METHOD = 'CGS' or 'TFQMR', a monitoring step is executed every MONIT iterations; otherwise, a monitoring step is executed every MONIT super-iterations (groups of up to  $m$  or  $\ell$  iterations for RGMRES or Bi-CGSTAB( $\ell$ ), respectively).
- There are some additional computational costs involved in monitoring the solution and residual vectors when the Bi-CGSTAB( $\ell$ ) method is used with  $\ell > 1$ .
- Constraint:* MONIT  $\leq$  MAXITN.
- 13: LWREQ – INTEGER *Output*
- On exit:* the minimum amount of workspace required by F11BEF. (See also Section 5 in F11BEF.)
- 14: WORK(LWORK) – REAL (KIND=nag\_wp) array *Communication Array*
- On exit:* the array WORK is initialized by F11BDF. It must **not** be modified before calling the next routine in the suite, namely F11BEF.
- 15: LWORK – INTEGER *Input*
- On entry:* the dimension of the array WORK as declared in the (sub)program from which F11BDF is called.
- Constraint:* LWORK  $\geq$  100.
- Note:** although the minimum value of LWORK ensures the correct functioning of F11BDF, a larger value is required by the other routines in the suite, namely F11BEF and F11BFF. The required value is as follows:
- | <b>Method</b>       | <b>Requirements</b>  |
|---------------------|--|
| RGMRES              | LWORK = $100 + n(m + 3) + m(m + 5) + 1$ , where $m$ is the dimension of the basis.   |
| CGS                 | LWORK = $100 + 7n$ .   |
| Bi-CGSTAB( $\ell$ ) | LWORK = $100 + (2n + \ell)(\ell + 2) + p$ , where $\ell$ is the order of the method. |
| TFQMR               | LWORK = $100 + 10n$ ,  |
- where
- $p = 2n$  if  $\ell > 1$  and ITERM = 2 was supplied.
- $p = n$  if  $\ell > 1$  and a preconditioner is used or ITERM = 2 was supplied.
- $p = 0$  otherwise.
- 16: IFAIL – INTEGER *Input/Output*
- On entry:* IFAIL must be set to 0, -1 or 1. If you are unfamiliar with this argument you should refer to Section 3.4 in How to Use the NAG Library and its Documentation for details.
- For environments where it might be inappropriate to halt program execution when an error is detected, the value -1 or 1 is recommended. If the output of error messages is undesirable, then the value 1 is recommended. Otherwise, if you are not familiar with this argument, the recommended value is 0. **When the value -1 or 1 is used it is essential to test the value of IFAIL on exit.**
- On exit:* IFAIL = 0 unless the routine detects an error or a warning has been flagged (see Section 6).

## 6 Error Indicators and Warnings

If on entry  $IFAIL = 0$  or  $-1$ , explanatory error messages are output on the current error message unit (as defined by X04AAF).

Errors or warnings detected by the routine:

$IFAIL = 1$

F11BDF has been called out of sequence: either F11BDF has been called twice or F11BEF has not terminated its current task.

$IFAIL = -1$

On entry,  $METHOD = \langle value \rangle$ .  
Constraint:  $METHOD = 'RGMRES', 'CGS', 'BICGSTAB'$  or  $'TFQMR'$ .

$IFAIL = -2$

On entry,  $PRECON = \langle value \rangle$ .  
Constraint:  $PRECON = 'N'$  or  $'P'$ .

$IFAIL = -3$

On entry,  $NORM = \langle value \rangle$ .  
Constraint:  $NORM = '1', 'I'$  or  $'2'$ .

$IFAIL = -4$

On entry,  $WEIGHT = \langle value \rangle$ .  
Constraint:  $WEIGHT = 'N'$  or  $'W'$ .

$IFAIL = -5$

On entry,  $ITERM = \langle value \rangle$ .  
Constraint:  $ITERM = 1$  or  $2$ .  
  
On entry,  $ITERM = 2$  and  $METHOD = 'TFQMR'$ .  
Constraint: if  $ITERM = 2$ ,  $METHOD \neq 'TFQMR'$ .  
  
On entry,  $ITERM = 2$  and  $NORM = \langle value \rangle$ .  
Constraint: if  $ITERM = 2$ ,  $NORM = '2'$ .  
  
On entry,  $ITERM = 2$  and  $WEIGHT = \langle value \rangle$ .  
Constraint: if  $ITERM = 2$ ,  $WEIGHT = 'N'$ .

$IFAIL = -6$

On entry,  $N = \langle value \rangle$ .  
Constraint:  $N > 0$ .

$IFAIL = -7$

On entry,  $M = \langle value \rangle$ ,  $N = \langle value \rangle$  and  $METHOD = \langle value \rangle$ .  
Constraint: if  $METHOD = 'RGMRES'$ ,  $M \leq \min(N, 50)$ . If  $METHOD = 'BICGSTAB'$ ,  
 $M \leq \min(N, 10)$ .  
  
On entry,  $M = \langle value \rangle$  and  $METHOD = \langle value \rangle$ .  
Constraint: if  $METHOD = 'RGMRES'$  or  $'BICGSTAB'$ ,  $M > 0$ .

$IFAIL = -8$

On entry,  $TOL = \langle value \rangle$ .  
Constraint:  $TOL < 1.0$ .



IFAIL = -9

On entry, MAXITN =  $\langle value \rangle$ .  
Constraint: MAXITN > 0.

IFAIL = -10

On entry, ITERM = 1, NORM = '2' and ANORM =  $\langle value \rangle$ .  
Constraint: if ITERM = 1 and NORM = '2', ANORM > 0.0.

IFAIL = -11

On entry, ITERM = 2, METHOD =  $\langle value \rangle$  and SIGMAX =  $\langle value \rangle$ .  
Constraint: if ITERM = 2 and METHOD = 'CGS' or 'BICGSTAB', SIGMAX > 0.0.

IFAIL = -12

On entry, MONIT =  $\langle value \rangle$  and MAXITN =  $\langle value \rangle$ .  
Constraint: MONIT  $\leq$  MAXITN.

IFAIL = -15

On entry, LWORK =  $\langle value \rangle$ .  
Constraint: LWORK  $\geq$  100.

IFAIL = -99

An unexpected error has been triggered by this routine. Please contact NAG.

See Section 3.9 in How to Use the NAG Library and its Documentation for further information.

IFAIL = -399

Your licence key may have expired or may not have been installed correctly.

See Section 3.8 in How to Use the NAG Library and its Documentation for further information.

IFAIL = -999

Dynamic memory allocation failed.

See Section 3.7 in How to Use the NAG Library and its Documentation for further information.

## 7 Accuracy

Not applicable.

## 8 Parallelism and Performance

F11BDF is not threaded in any implementation.

## 9 Further Comments

RGMRES can estimate internally the maximum singular value  $\sigma_1$  of the iteration matrix, using  $\sigma_1 \sim \|T\|_1$ , where  $T$  is the upper triangular matrix obtained by  $QR$  factorization of the upper Hessenberg matrix generated by the Arnoldi process. The computational costs of this computation are negligible when compared to the overall costs.

Loss of orthogonality in the RGMRES method, or of bi-orthogonality in the Bi-CGSTAB( $\ell$ ) method may degrade the solution and speed of convergence. For both methods, the algorithms employed include checks on the basis vectors so that the number of basis vectors used for a given super-iteration may be less than the value specified in the input argument M. Also, if termination criterion (2) is used the CGS, Bi-CGSTAB( $\ell$ ) and TFQMR methods will restart automatically the computation from the last available iterates, when the stability of the solution process requires it.

Termination criterion (3), when available, involves only the residual (or norm of the residual) produced directly by the iteration process: this may differ from the norm of the true residual  $\tilde{r}_k = b - Ax_k$ , particularly when the norm of the residual is very small. Also, if the norm of the initial estimate of the solution is much larger than the norm of the exact solution, convergence can be achieved despite very large errors in the solution. On the other hand, termination criterion (3) is cheaper to use and inspects the progress of the actual iteration. Termination criterion (2) should be preferred in most cases, despite its slightly larger costs.

## 10 Example

This example solves an  $8 \times 8$  nonsymmetric system of simultaneous linear equations using the bi-conjugate gradient stabilized method of order  $\ell = 1$ , where the coefficients matrix  $A$  has a random sparsity pattern. An incomplete  $LU$  preconditioner is used (routines F11DAF and F11DBF).

### 10.1 Program Text

```

Program f11bdfe

!      F11BDF Example Program Text

!      Mark 26 Release. NAG Copyright 2016.

!      .. Use Statements ..
Use nag_library, Only: f11bdf, f11bef, f11bff, f11daf, f11dbf, f11xaf,    &
                        nag_wp
!      .. Implicit None Statement ..
Implicit None
!      .. Parameters ..
Integer, Parameter      :: nin = 5, nout = 6
!      .. Local Scalars ..
Real (Kind=nag_wp)      :: anorm, dtol, sigmax, stplhs, stprhs, &
                        tol
Integer                  :: i, ifail, ifail1, irevcn, iterm,      &
                        itn, la, lfill, liwork, lwork,          &
                        lwreq, m, maxitn, monit, n, nnz,        &
                        nnzc, npivm
Character (8)            :: method
Character (1)            :: milu, norm, precon, pstrat, weight
!      .. Local Arrays ..
Real (Kind=nag_wp), Allocatable :: a(:), b(:), wgt(:), work(:), x(:)
Integer, Allocatable      :: icol(:), idiag(:), ipivp(:),      &
                        ipivq(:), irow(:), istr(:), iwork(:)
!      .. Executable Statements ..
Write (nout,*) 'F11BDF Example Program Results'

!      Skip heading in data file

Read (nin,*)
Read (nin,*) n
Read (nin,*) nnz
la = 3*nnz
liwork = 7*n + 2
lwork = 200
Allocate (a(la),b(n),wgt(n),work(lwork),x(n),icol(la),idiag(n),ipivp(n), &
         ipivq(n),irow(la),istr(n+1),iwork(liwork))

!      Read or initialize the parameters for the iterative solver

Read (nin,*) method
Read (nin,*) precon, norm, weight, iterm
Read (nin,*) m, tol, maxitn
Read (nin,*) monit
anorm = 0.0E0_nag_wp
sigmax = 0.0E0_nag_wp

!      Read the parameters for the preconditioner

```

```

      Read (nin,*) lfill, dtol
      Read (nin,*) milu, pstrat

!      Read the nonzero elements of the matrix A

      Do i = 1, nnz
        Read (nin,*) a(i), irow(i), icol(i)
      End Do

!      Read right-hand side vector b and initial approximate solution

      Read (nin,*) b(1:n)
      Read (nin,*) x(1:n)

!      Calculate incomplete LU factorization

      milu = 'N'

!      ifail: behaviour on error exit
!            =0 for hard exit, =1 for quiet-soft, =-1 for noisy-soft
      ifail = 0
      Call f11daf(n,nnz,a,la,irow,icol,lfill,dtol,pstrat,milu,ipivp,ipivq,      &
        istr,idiag,nnzc,npivm,iwork,liwork,ifail)

!      Call F11BDF to initialize the solver

      ifail = 0
      Call f11bdf(method,precon,norm,weight,iterm,n,m,tol,maxitn,anorm,sigmax, &
        monit,lwreq,work,lwork,ifail)

!      Call repeatedly F11BEF to solve the equations
!      Note that the arrays B and X are overwritten

!      On final exit, X will contain the solution and B the
!      residual vector

      irevcm = 0
      lwreq = lwork

      ifail = 1
loop: Do
      Call f11bef(irevcm,x,b,wgt,work,lwreq,ifail)

      If (irevcm/=4) Then
        ifail1 = -1
        Select Case (irevcm)
          Case (-1)

            Call f11xaf('Transpose',n,nnz,a,irow,icol,'No checking',x,b,      &
              ifail1)

          Case (1)

            Call f11xaf('No transpose',n,nnz,a,irow,icol,'No checking',x,b,      &
              ifail1)

          Case (2)

            Call f11dbf('No transpose',n,a,la,irow,icol,ipivp,ipivq,istr,      &
              idiag,'No checking',x,b,ifail1)

          Case (3)

            ifail1 = 0
            Call f11bff(itn,stplhs,stprhs,anorm,sigmax,work,lwreq,ifail1)

            Write (nout,99999) itn, stplhs
            Write (nout,99998)
            Write (nout,99997)(x(i),b(i),i=1,n)
          End Select
        If (ifail1/=0) Then

```

```

        irevcm = 6
      End If
    Else If (ifail/=0) Then
      Write (nout,99993) ifail
      Go To 100
    Else
      Exit loop
    End If
  End Do loop

! Obtain information about the computation

  ifaill = 0
  Call f11bff(itn,stplhs,stprhs,anorm,sigmax,work,lwreq,ifaill)

! Print the output data

  Write (nout,99996)
  Write (nout,99995) 'Number of iterations for convergence: ', itn
  Write (nout,99994) 'Residual norm: ', stplhs
  Write (nout,99994) 'Right-hand side of termination criterion:', stprhs
  Write (nout,99994) '1-norm of matrix A: ', anorm

! Output x

  Write (nout,99998)
  Write (nout,99997)(x(i),b(i),i=1,n)
100 Continue

99999 Format (/,1X,'Monitoring at iteration no.',I4,/,1X,1P,'residual no',      &
  'rm: ',E14.4)
99998 Format (/,2X,' Solution vector',2X,' Residual vector')
99997 Format (1X,1P,E16.4,1X,E16.4)
99996 Format (/,1X,'Final Results')
99995 Format (1X,A,I4)
99994 Format (1X,A,1P,E14.4)
99993 Format (1X,/,1X,' ** F11BEF returned with IFAIL = ',I5)
      End Program f11bdfe

```

## 10.2 Program Data

### F11BDF Example Program Data

8			N
24			NNZ
	'BICGSTAB'		METHOD
	'P' '1' 'N' 1		PRECON, NORM, WEIGHT, ITERM
1	1.0D-8	20	M, TOL, MAXITN
1			MONIT
0	0.0		LFILL, DTOL
	'N' 'C'		MILU, PSTRAT
2.	1	1	
-1.	1	4	
1.	1	8	
4.	2	1	
-3.	2	2	
2.	2	5	
-7.	3	3	
2.	3	6	
3.	4	1	
-4.	4	3	
5.	4	4	
5.	4	7	
-1.	5	2	
8.	5	5	
-3.	5	7	
-6.	6	1	
5.	6	3	
2.	6	6	
-5.	7	3	
-1.	7	5	

```

6.   7   7
-1.  8   2
2.   8   6
3.   8   8      A(I), IROW(I), ICOL(I), I=1,...,NNZ
6.   8. -9. 46.
17. 21. 22. 34.  B(I), I=1,...,N
0.   0.  0.  0.
0.   0.  0.  0.  X(I), I=1,...,N

```

### 10.3 Program Results

F11BDF Example Program Results

```

Monitoring at iteration no.  1
residual norm:      1.4059E+02

```

Solution vector	Residual vector
-4.5858E+00	1.5256E+01
1.0154E+00	2.6624E+01
-2.2234E+00	-8.7498E+00
6.0097E+00	1.8602E+01
1.3827E+00	8.2821E+00
-7.9070E+00	2.0416E+01
4.4270E-01	9.6094E+00
5.9248E+00	3.3055E+01

```

Monitoring at iteration no.  2
residual norm:      3.2742E+01

```

Solution vector	Residual vector
4.1642E+00	-2.9585E+00
4.9370E+00	-5.5523E+00
4.8101E+00	8.2070E-01
5.4324E+00	-1.6828E+01
5.8531E+00	5.5975E-01
1.1925E+01	-1.9150E+00
8.4826E+00	1.0081E+00
6.0625E+00	-3.1004E+00

Final Results

```

Number of iterations for convergence:      3
Residual norm:                             1.0373E-08
Right-hand side of termination criterion:  5.5900E-06
1-norm of matrix A:                         1.1000E+01

```

Solution vector	Residual vector
1.0000E+00	-1.3554E-09
2.0000E+00	-2.6109E-09
3.0000E+00	2.2471E-10
4.0000E+00	-3.2203E-09
5.0000E+00	6.3045E-10
6.0000E+00	-5.2431E-10
7.0000E+00	9.5771E-10
8.0000E+00	-8.4890E-10

---