E04NKF – NAG Fortran Library Routine Document

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.

Note. *This routine uses* optional parameters *to define choices in the problem specification and in the details of the algorithm. If you wish to use* default *settings for all of the optional parameters, you need only read Section 1 to Section 9 of this document. Refer to the additional Section 10, Section 11 and Section 12 for a description of the algorithm, the specification of the optional parameters and a description of the monitoring information produced by the routine.*

Note. This routine was introduced into the NAG Fortran Library at Mark 19 and may therefore not be available to all users of the NAG Fortran SMP Library.

1 Purpose

E04NKF solves sparse linear programming or quadratic programming problems.

2 Specification

```
  SUBROUTINE E04NKF(N, M, NNZ, IOBJ, NCOLH, QPHX, A, HA, KA, BL, BU,
 1                   START, NAMES, NNAME, CRNAME, NS, XS, ISTATE,
 2                   MINIZ, MINZ, NINF, SINF, OBJ, CLAMDA, IZ, LENIZ,
 3                   Z, LENZ, IFAIL)
  INTEGER            N, M, NNZ, IOBJ, NCOLH, HA(NNZ), KA(N+1), NNAME,
 1                   NS, ISTATE(N+M), MINIZ, MINZ, NINF,
 2                   IZ(LENIZ), LENIZ, LENZ, IFAIL
  real               A(NNZ), BL(N+M), BU(N+M), XS(N+M), SINF, OBJ,
 1                   CLAMDA(N+M), Z(LENZ)
  CHARACTER*8        NAMES(5), CRNAME(NNAME)
  CHARACTER*1        START
  EXTERNAL           QPHX
```

3 Description

E04NKF is designed to solve a class of quadratic programming problems that are assumed to be stated in the following general form:

$$\underset{x \in R^n}{\text{minimize}} \ f(x) \ \text{subject to} \ l \le \left\{ \begin{array}{c} x \\ Ax \end{array} \right\} \le u, \tag{1}$$

where x is a set of variables, A is an m by n matrix and the objective function $f(x)$ may be specified in a variety of ways depending upon the particular problem to be solved. The optional parameter **Maximize** (see Section 11.2) may be used to specify an alternative problem in which $f(x)$ is maximized. The possible forms for $f(x)$ are listed in Table 1 below, in which the prefixes FP, LP and QP stand for 'feasible point', 'linear programming' and 'quadratic programming' respectively, c is an n element vector and H is the n by n second-derivative matrix $\nabla^2 f(x)$ (the *Hessian matrix*).

Problem type	Objective function $f(x)$	Hessian matrix H
FP	Not applicable	Not applicable
LP	$c^T x$	Not applicable
QP	$c^T x + \frac{1}{2} x^T H x$	Symmetric positive semi-definite

Table 1

For LP and QP problems, the unique global minimum value of $f(x)$ is found. For FP problems, $f(x)$ is omitted and the routine attempts to find a feasible point for the set of constraints. For QP problems, you must also provide a subroutine that computes Hx for any given vector x. (H need not be stored explicitly.)

E04NKF is intended to solve large-scale linear and quadratic programming problems in which the constraint matrix A is *sparse* (i.e., when the number of zero elements is sufficiently large that it is

worthwhile using algorithms which avoid computations and storage involving zero elements). The routine also takes advantage of sparsity in c. (Sparsity in H can be exploited in the subroutine that computes Hx.) For problems in which A can be treated as a *dense* matrix, it is usually more efficient to use E04MFF, E04NCF or E04NFF.

If H is positive-definite, then the final x will be unique. If E04NKF detects that H is indefinite, it terminates immediately with IFAIL = 5 (see Section 6). If H is the zero matrix, the routine will still solve the resulting LP problem; however, this can be accomplished more efficiently by setting NCOLH = 0 (see Section 5).

The upper and lower bounds on the m elements of Ax are said to define the *general constraints* of the problem. Internally, E04NKF converts the general constraints to equalities by introducing a set of *slack variables* s, where $s = (s_1, s_2, \ldots, s_m)^T$. For example, the linear constraint $5 \le 2x_1 + 3x_2 \le +\infty$ is replaced by $2x_1 + 3x_2 - s_1 = 0$, together with the bounded slack $5 \le s_1 \le +\infty$. The problem defined by (1) can therefore be re-written in the following equivalent form:

$$\underset{x \in R^n, s \in R^m}{\text{minimize}} \ f(x) \ \text{subject to} \ Ax - s = 0, \ l \le \begin{Bmatrix} x \\ s \end{Bmatrix} \le u.$$

Since the slack variables s are subject to the same upper and lower bounds as the elements of Ax, the bounds on Ax and x can simply be thought of as bounds on the combined vector (x, s). (In order to indicate their special role in QP problems, the original variables x are sometimes known as 'column variables', and the slack variables s are known as 'row variables'.)

Each LP or QP problem is solved using an *active-set* method. This is an iterative procedure with two phases: a *feasibility phase*, in which the sum of infeasibilities is minimized to find a feasible point; and an *optimality phase*, in which $f(x)$ is minimized by constructing a sequence of iterations that lies within the feasible region.

A constraint is said to be *active* or *binding* at x if the associated element of either x or Ax is equal to one of its upper or lower bounds. Since an active constraint in Ax has its associated slack variable at a bound, the status of both simple and general upper and lower bounds can be conveniently described in terms of the status of the variables (x, s). A variable is said to be *nonbasic* if it is temporarily fixed at its upper or lower bound. It follows that regarding a general constraint as being *active* is equivalent to thinking of its associated slack as being *nonbasic*.

At each iteration of an active-set method, the constraints $Ax - s = 0$ are (conceptually) partitioned into the form

$$Bx_B + Sx_S + Nx_N = 0,$$

where x_N consists of the nonbasic elements of (x, s) and the *basis matrix* B is square and non-singular. The elements of x_B and x_S are called the *basic* and *superbasic* variables respectively; with x_N they are a permutation of the elements of x and s. At a QP solution, the basic and superbasic variables will lie somewhere between their upper or lower bounds, while the nonbasic variables will be equal to one of their bounds. At each iteration, x_S is regarded as a set of independent variables that are free to move in any desired direction, namely one that will improve the value of the objective function (or sum of infeasibilities). The basic variables are then adjusted in order to ensure that (x, s) continues to satisfy $Ax - s = 0$. The number of superbasic variables (n_S say) therefore indicates the number of degrees of freedom remaining after the constraints have been satisfied. In broad terms, n_S is a measure of *how nonlinear* the problem is. In particular, n_S will always be zero for FP and LP problems.

If it appears that no improvement can be made with the current definition of B, S and N, a nonbasic variable is selected to be added to S, and the process is repeated with the value of n_S increased by one. At all stages, if a basic or superbasic variable encounters one of its bounds, the variable is made nonbasic and the value of n_S is decreased by one.

Associated with each of the m equality constraints $Ax - s = 0$ is a *dual variable* π_i. Similarly, each variable in (x, s) has an associated *reduced gradient* d_j (also known as a *reduced cost*). The reduced gradients for the variables x are the quantities $g - A^T \pi$, where g is the gradient of the QP objective function; and the reduced gradients for the slack variables s are the dual variables π. The QP subproblem is optimal if $d_j \ge 0$ for all nonbasic variables at their lower bounds, $d_j \le 0$ for all nonbasic variables at their upper bounds and $d_j = 0$ for all superbasic variables. In practice, an *approximate* QP solution is found

by slightly relaxing these conditions on d_j (see the description of the optional parameter **Optimality Tolerance** in Section 11.2).

The process of computing and comparing reduced gradients is known as *pricing* (a term first introduced in the context of the simplex method for linear programming). To 'price' a nonbasic variable x_j means that the reduced gradient d_j associated with the relevant active upper or lower bound on x_j is computed via the formula $d_j = g_j - a^T \pi$, where a_j is the jth column of $(A - I)$. (The variable selected by such a process and the corresponding value of d_j (i.e., its reduced gradient) are the quantities +S and dj in the monitoring file output; see Section 12.) If A has significantly more columns than rows (i.e., $n \gg m$), pricing can be computationally expensive. In this case, a strategy known as *partial pricing* can be used to compute and compare only a subset of the d_j's.

E04NKF is based on SQOPT, which is part of the SNOPT package described in Gill *et al.* [1], which in turn utilizes routines from the MINOS package (see Murtagh and Saunders [2]). It uses stable numerical methods throughout and includes a reliable basis package (for maintaining sparse LU factors of the basis matrix B), a practical anti-degeneracy procedure, efficient handling of linear constraints and bounds on the variables (by an active-set strategy), as well as automatic scaling of the constraints. Further details can be found in Section 10.

4 References

[1] Gill P E, Murray W and Saunders M A (1996) SNOPT: An SQP Algorithm for Large-scale Constrained Optimization *Numerical Analysis Report 96–2.* Department of Mathematics, University of California, San Diego

[2] Murtagh B A and Saunders M A (1995) MINOS 5.4 User's Guide *Report SOL 83-20R.* Department of Operations Research, Stanford University

[3] Gill P E and Murray W (1978) Numerically stable methods for quadratic programming *Math. Programming* **14** 349–372

[4] Gill P E, Murray W, Saunders M A and Wright M H (1989) A practical anti-cycling procedure for linearly constrained optimization *Math. Programming* **45** 437–474

[5] Gill P E, Murray W, Saunders M A and Wright M H (1991) Inertia-controlling methods for general quadratic programming *SIAM Rev.* **33** 1–36

[6] Gill P E, Murray W, Saunders M A and Wright M H (1987) Maintaining LU factors of a general sparse matrix *Linear Algebra and its Applics.* **88/89** 239–270

[7] Hall J A J and McKinnon K I M (1996) The Simplest Examples where the Simplex Method Cycles and Conditions where EXPAND Fails to Prevent Cycling *Report MS 96–010.* Department of Mathematics and Statistics, University of Edinburgh

[8] Fourer R (1982) Solving staircase linear programs by the simplex method *Math. Programming* **23** 274–313

5 Parameters

1: N — INTEGER *Input*

On entry: n, the number of variables (excluding slacks). This is the number of columns in the linear constraint matrix A.

Constraint: N ≥ 1.

2: M — INTEGER *Input*

On entry: m, the number of general linear constraints (or slacks). This is the number of rows in A, including the free row (if any; see IOBJ below).

Constraint: M ≥ 1.

3: NNZ — INTEGER *Input*

On entry: the number of non-zero elements in A.

Constraint: $1 \leq \text{NNZ} \leq \text{N} \times \text{M}$.

4: IOBJ — INTEGER *Input*

On entry: if IOBJ > 0, row IOBJ of A is a free row containing the non-zero elements of the vector c appearing in the linear objective term $c^T x$. If IOBJ $= 0$, there is no free row — i.e., the problem is either an FP problem (in which case IOBJ must be set to zero), or a QP problem with $c = 0$.

Constraint: $0 \leq \text{IOBJ} \leq \text{M}$.

5: NCOLH — INTEGER *Input*

On entry: n_H, the number of leading non-zero columns of the Hessian matrix H. For FP and LP problems, NCOLH must be set to zero.

Constraint: $0 \leq \text{NCOLH} \leq \text{N}$.

6: QPHX — SUBROUTINE, supplied by the NAG Fortran Library or the user. *External Procedure*

For QP problems, you must supply a version of QPHX to compute the matrix product Hx. If H has zero rows and columns, it is most efficient to order the variables $x = (y \ z)^T$ so that

$$Hx = \left(\begin{array}{cc} H_1 & 0 \\ 0 & 0 \end{array} \right) \left(\begin{array}{c} y \\ z \end{array} \right) = \left(\begin{array}{c} H_1 y \\ 0 \end{array} \right),$$

where the nonlinear variables y appear first as shown. For FP and LP problems, QPHX will never be called by E04NKF and hence QPHX may be the dummy routine E04NKU (NKUE04 in some implementations: see the Users' Note for your implementation for details).

Its specification is:

```
      SUBROUTINE QPHX(NSTATE, NCOLH, X, HX)
      INTEGER          NCOLH, NSTATE
      real             HX(NCOLH), X(NCOLH)
```

1: NSTATE — INTEGER *Input*

On entry: if NSTATE $= 1$, then E04NKF is calling QPHX for the first time. This parameter setting allows you to save computation time if certain data must be read or calculated only once. If NSTATE ≥ 2, then E04NKF is calling QPHX for the last time. This parameter setting allows you to perform some additional computation on the final solution. In general, the last call to QPHX is made with NSTATE $= 2 + \text{IFAIL}$ (see Section 6). Otherwise, NSTATE $= 0$.

2: NCOLH — INTEGER *Input*

On entry: this is the same parameter NCOLH as supplied to E04NKF (see above).

3: X(NCOLH) — *real* array *Input*

On entry: the first NCOLH elements of the vector x.

4: HX(NCOLH) — *real* array *Output*

On exit: the product Hx.

QPHX must be declared as EXTERNAL in the (sub)program from which E04NKF is called. Parameters denoted as *Input* must **not** be changed by this procedure.

7: A(NNZ) — *real* array *Input*

On entry: the non-zero elements of A, ordered by increasing column index. Note that elements with the same row and column indices are not allowed.

8: HA(NNZ) — INTEGER array *Input*

On entry: HA(i) must contain the row index of the non-zero element stored in A(i), for $i = 1, 2, \ldots$,NNZ. Note that the row indices for a column may be supplied in any order.

Constraint: $1 \leq$ HA(i) $\leq$ M, for $i = 1, 2, \ldots$,NNZ.

9: KA(N+1) — INTEGER array *Input*

On entry: KA(j) must contain the index in A of the start of the jth column, for $j = 1, 2, \ldots$,N. To specify the jth column as empty, set KA(j) = KA($j + 1$). Note that the first and last elements of KA must be such that KA(1) = 1 and KA(N+1) = NNZ + 1.

Constraints:

KA(1) = 1,
KA(j) ≥ 1 for $j = 2, 3, \ldots$,N,
KA(N+1) = NNZ+1,
$0 \leq$ KA($j + 1$) $-$ KA(j) $\leq$ M for $j = 1, 2, \ldots$,N.

10: BL(N+M) — ***real*** array *Input*

On entry: l, the lower bounds for all the variables and general constraints, in the following order. The first N elements of BL must contain the bounds on the variables x, and the next M elements the bounds for the general linear constraints Ax (or slacks s) and the free row (if any). To specify a non-existent lower bound (i.e., $l_j = -\infty$), set BL(j) $\leq -bigbnd$, where $bigbnd$ is the value of the optional parameter **Infinite Bound Size** (default value $= 10^{20}$; see Section 11.2). To specify the jth constraint as an *equality*, set BL(j) = BU(j) = β, say, where $|\beta| < bigbnd$. Note that the lower bound corresponding to the free row must be set to $-\infty$ and stored in BL(N+IOBJ).

Constraints:

BL(N+IOBJ) $\leq -bigbnd$ when IOBJ > 0.
(See also the description for BU below.)

11: BU(N+M) — ***real*** array *Input*

On entry: u, the upper bounds for all the variables and general constraints, in the following order. The first N elements of BL must contain the bounds on the variables x, and the next M elements the bounds for the general linear constraints Ax (or slacks s) and the free row (if any). To specify a non-existent upper bound (i.e., $u_j = +\infty$), set BU(j) $\geq bigbnd$. Note that the upper bound corresponding to the free row must be set to $+\infty$ and stored in BU(N+IOBJ).

Constraints:

BU(N+IOBJ) $\geq bigbnd$ when IOBJ > 0,
BL(j) $\leq$ BU(j), for $j = 1, 2, \ldots$,N+M,
$|\beta| < bigbnd$ when BL(j) = BU(j) = β.

12: START — CHARACTER*1 *Input*

On entry: indicates how a starting basis is to be obtained as follows.

If START = 'C', then an internal Crash procedure will be used to choose an initial basis matrix B.
If START = 'W', then a basis is already defined in ISTATE (probably from a previous call).

Constraint: START = 'C' or 'W'.

13: NAMES(5) — CHARACTER*8 *Input*

On entry: a set of names associated with the so-called MPSX form of the problem as follows:

NAMES(1) must contain the name for the problem (or be blank);
NAMES(2) must contain the name for the free row (or be blank);
NAMES(3) must contain the name for the constraint right-hand side (or be blank);
NAMES(4) must contain the name for the ranges (or be blank);
NAMES(5) must contain the name for the bounds (or be blank).

(These names are used in the monitoring file output; see Section 12.)

14: NNAME — INTEGER *Input*

On entry: the number of column (i.e., variable) and row names supplied in CRNAME as follows.

If NNAME = 1, there are no names. Default names will be used in the printed output.
If NNAME = N+M, all names must be supplied.

Constraint: NNAME = 1 or N+M.

15: CRNAME(NNAME) — CHARACTER*8 array *Input*

On entry: the optional column and row names, respectively as follows.

If NNAME = 1, CRNAME is not referenced and the printed output will use default names for the columns and rows.
If NNAME = N+M, the first N elements must contain the names for the columns and the next M elements must contain the names for the rows. Note that the name for the free row (if any) must be stored in CRNAME(N+IOBJ).

16: NS — INTEGER *Input/Output*

On entry: n_S, the number of superbasics. For QP problems, NS need not be specified if START = 'C', but must retain its value from a previous call when START = 'W'. For FP and LP problems, NS need not be initialized.

On exit: the final number of superbasics. This will be zero for FP and LP problems.

17: XS(N+M) — *real* array *Input/Output*

On entry: the initial values of the variables and slacks (x, s). (See the description for ISTATE below.)

On exit: the final values of the variables and slacks (x, s).

18: ISTATE(N+M) — INTEGER array *Input/Output*

On entry: if START = 'C', the first N elements of ISTATE and XS must specify the initial states and values, respectively, of the variables x. (The slacks s need not be initialized.) An internal Crash procedure is then used to select an initial basis matrix B. The initial basis matrix will be triangular (neglecting certain small elements in each column). It is chosen from various rows and columns of $(A - I)$. Possible values for ISTATE(j) are as follows:

ISTATE(j)	**State of XS(j) during Crash procedure**
0 or 1	Eligible for the basis
2	Ignored
3	Eligible for the basis (given preference over 0 or 1)
4 or 5	Ignored

If nothing special is known about the problem, or there is no wish to provide special information, you may set ISTATE$(j) = 0$ and XS$(j) = 0.0$ for $j = 1, 2, \ldots,$N. All variables will then be eligible for the initial basis. Less trivially, to say that the jth variable will probably be equal to one of its bounds, set ISTATE$(j) = 4$ and XS$(j) = $ BL(j) or ISTATE$(j) = 5$ and XS$(j) = $ BU(j) as appropriate.

Following the Crash procedure, variables for which ISTATE(j) = 2 are made superbasic. Other variables not selected for the basis are then made nonbasic at the value XS(j) if BL(j) $\leq$ XS(j) $\leq$ BU(j), or at the value BL(j) or BU(j) closest to XS(j).

If START = 'W', ISTATE and XS must specify the initial states and values, respectively, of the variables and slacks (x, s). If E04NKF has been called previously with the same values of N and M, ISTATE already contains satisfactory information.

Constraints:

> If START = 'C', $0 \leq$ ISTATE(j) ≤ 5 for $j = 1, 2, \ldots,$N.
> If START = 'W', $0 \leq$ ISTATE(j) ≤ 3 for $j = 1, 2, \ldots,$N+M.

On exit: the final states of the variables and slacks (x, s). The significance of each possible value of ISTATE(j) is as follows:

ISTATE(j)	**State of variable** j	**Normal value of XS**(j)
0	Nonbasic	BL(j)
1	Nonbasic	BU(j)
2	Superbasic	Between BL(j) and BU(j)
3	Basic	Between BL(j) and BU(j)

If NINF = 0, basic and superbasic variables may be outside their bounds by as much as the value of the optional parameter **Feasibility Tolerance** (default value =max($10^{-6}, \sqrt{\epsilon}$), where ϵ is the *machine precision*; see Section 11.2). Note that unless the optional parameter **Scale Option** = 0 (default value = 2; see Section 11.2) is specified, the **Feasibility Tolerance** applies to the variables of the scaled problem. In this case, the variables of the original problem may be as much as 0.1 outside their bounds, but this is unlikely unless the problem is very badly scaled.

Very occasionally some nonbasic variables may be outside their bounds by as much as the **Feasibility Tolerance**, and there may be some nonbasic variables for which XS(j) lies strictly between its bounds.

If NINF > 0, some basic and superbasic variables may be outside their bounds by an arbitrary amount (bounded by SINF if **Scale Option** = 0).

19: MINIZ — INTEGER *Output*

On exit: the minimum value of LENIZ required to start solving the problem. If IFAIL = 12, E04NKF may be called again with LENIZ suitably larger than MINIZ. (The bigger the better, since it is not certain how much workspace the basis factors need.)

20: MINZ — INTEGER *Output*

On exit: the minimum value of LENZ required to start solving the problem. If IFAIL = 13, E04NKF may be called again with LENZ suitably larger than MINZ. (The bigger the better, since it is not certain how much workspace the basis factors need.)

21: NINF — INTEGER *Output*

On exit: the number of infeasibilities. This will be zero if IFAIL = 0 or 1.

22: SINF — ***real*** *Output*

On exit: the sum of infeasibilities. This will be zero if NINF = 0. (Note that E04NKF does *not* attempt to compute the minimum value of SINF if IFAIL = 3.)

23: OBJ — ***real*** *Output*

On exit: the value of the objective function. If NINF = 0, OBJ includes the quadratic objective term $\frac{1}{2}x^T H x$ (if any). If NINF > 0, OBJ is just the linear objective term $c^T x$ (if any). For FP problems, OBJ is set to zero.

24: CLAMDA(N+M) — **real** array   *Output*

> *On exit:* a set of Lagrange multipliers for the bounds on the variables and the general constraints. More precisely, the first N elements contain the multipliers (*reduced costs*) for the bounds on the variables, and the next M elements contain the multipliers (*shadow prices*) for the general linear constraints.

25: IZ(LENIZ) — INTEGER array   *Workspace*
26: LENIZ — INTEGER   *Input*

> *On entry:* the dimension of the array IZ as declared in the (sub)program from which E04NKF is called.

> *Constraint:* LENIZ ≥ 1.

27: Z(LENZ) — **real** array   *Workspace*
28: LENZ — INTEGER   *Input*

> *On entry:* the dimension of the array Z as declared in the (sub)program from which E04NKF is called.

> *Constraint:* LENZ ≥ 1.

> The amounts of workspace provided (i.e., LENIZ and LENZ) and required (i.e., MINIZ and MINZ) are (by default) output on the current advisory message unit (as defined by X04ABF). Since the minimum values of LENIZ and LENZ required to start solving the problem are returned in MINIZ and MINZ, respectively, you may prefer to obtain appropriate values from the output of a preliminary run with LENIZ and LENZ set to 1. (E04NKF will then terminate with IFAIL = 12.)

29: IFAIL — INTEGER   *Input/Output*

> *On entry:* IFAIL must be set to 0, -1 or 1. Users who are unfamiliar with this parameter should refer to Chapter P01 for details.

> *On exit:* IFAIL = 0 unless the routine detects an error or gives a warning (see Section 6).

> **For this routine**, because the values of output parameters may be useful even if IFAIL $\neq 0$ on exit, users are recommended to set IFAIL to -1 before entry. **It is then essential to test the value of IFAIL on exit**.

> E04NKF returns with IFAIL = 0 if the reduced gradient (`Norm rg`; see Section 8.1) is negligible, the Lagrange multipliers (`Lagr Mult`; see Section 8.1) are optimal and x satisfies the constraints to the accuracy requested by the value of the optional parameter **Feasibility Tolerance** (default value $= \max(10^{-6}, \sqrt{\epsilon})$, where ϵ is the **machine precision**; see Section 11.2).

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 specified by the routine:

IFAIL = 1

> Weak solution found. The final x is not unique, although x gives the global minimum value of the objective function.

IFAIL = 2

> The problem is unbounded (or badly scaled). The objective function is not bounded below in the feasible region.

IFAIL = 3

> The problem is infeasible. The general constraints cannot all be satisfied simultaneously to within the value of the optional parameter **Feasibility Tolerance** (default value $= \max(10^{-6}, \sqrt{\epsilon})$, where ϵ is the **machine precision**; see Section 11.2).

IFAIL = 4

Too many iterations. The value of the optional parameter **Iteration Limit** (default value = $\max(50, 5(n + m))$; see Section 11.2) is too small.

IFAIL = 5

The reduced Hessian matrix $Z^T H Z$ (see Section 10.2) exceeds its assigned dimension. The value of the optional parameter **Superbasics Limit** (default value = $\min(n_H + 1, n)$; see Section 11.2) is too small.

IFAIL = 6

The Hessian matrix H appears to be indefinite. Check that subroutine QPHX has been coded correctly and that all relevant elements of Hx have been assigned their correct values.

IFAIL = 7

An input parameter is invalid.

IFAIL = 8

Numerical error in trying to satisfy the general constraints. The basis is very ill-conditioned.

IFAIL = 9

Not enough integer workspace for the basis factors. Increase LENIZ and rerun E04NKF.

IFAIL = 10

Not enough real workspace for the basis factors. Increase LENZ and rerun E04NKF.

IFAIL = 11

The basis is singular after 15 attempts to factorize it (adding slacks where necessary). Either the problem is badly scaled or the value of the optional parameter **LU Factor Tolerance** (default value = 100.0; see Section 11.2) is too large.

IFAIL = 12

Not enough integer workspace to start solving the problem. Increase LENIZ to at least MINIZ and rerun E04NKF.

IFAIL = 13

Not enough real workspace to start solving the problem. Increase LENZ to at least MINZ and rerun E04NKF.

7 Accuracy

The routine implements a numerically stable active-set strategy and returns solutions that are as accurate as the condition of the problem warrants on the machine.

8 Further Comments

This section contains a description of the printed output.

8.1 Description of the Printed Output

This section describes the (default) intermediate printout and final printout produced by E04NKF. The intermediate printout is a subset of the monitoring information produced by the routine at every iteration (see Section 12). The level of printed output can be controlled by the user (see the description of the optional parameter **Print Level** in Section 11.2). Note that the intermediate printout and final printout are produced only if **Print Level** ≥ 10 (the default).

The following line of summary output (< 80 characters) is produced at every iteration. In all cases, the values of the quantities printed are those in effect *on completion* of the given iteration.

Itn	is the iteration count.
Step	is the step taken along the computed search direction.
Ninf	is the number of violated constraints (infeasibilities). This will be zero during the optimality phase.
Sinf/Objective	is the value of the current objective function. If x is not feasible, `Sinf` gives the sum of the magnitudes of constraint violations. If x is feasible, `Objective` is the value of the objective function. The output line for the final iteration of the feasibility phase (i.e., the first iteration for which `Ninf` is zero) will give the value of the true objective at the first feasible point. During the optimality phase, the value of the objective function will be non-increasing. During the feasibility phase, the number of constraint infeasibilities will not increase until either a feasible point is found, or the optimality of the multipliers implies that no feasible point exists.
Norm rg	is $\|d_S\|$, the Euclidean norm of the reduced gradient (see Section 10.3). During the optimality phase, this norm will be approximately zero after a unit step. For FP and LP problems, `Norm rg` is not printed.

The final printout includes a listing of the status of every variable and constraint.

The following describes the printout for each variable. A full stop (.) is printed for any numerical value that is zero.

Variable	gives the name of the variable. If NNAME = 1, a default name is assigned to the jth variable for $j = 1, 2, \ldots, n$. If NNAME = N + M, the name supplied in CRNAME(j) is assigned to the jth variable.
State	gives the state of the variable (`LL` if nonbasic on its lower bound, `UL` if nonbasic on its upper bound, `EQ` if nonbasic and fixed, `FR` if nonbasic and strictly between its bounds, `BS` if basic and `SBS` if superbasic).
	A key is sometimes printed before `State` to give some additional information about the state of a variable. Note that unless the optional parameter **Scale Option** $= 0$ (default value $= 2$; see Section 11.2) is specified, the tests for assigning a key are applied to the variables of the scaled problem.

A	*Alternative optimum possible.* The variable is nonbasic, but its reduced gradient is essentially zero. This means that if the variable were allowed to start moving away from its bound, there would be no change in the value of the objective function. The values of the other free variables *might* change, giving a genuine alternative solution. However, if there are any degenerate variables (labelled D), the actual change might prove to be zero, since one of them could encounter a bound immediately. In either case, the values of the Lagrange multipliers *might* also change.
D	*Degenerate.* The variable is basic or superbasic, but it is equal to (or very close to) one of its bounds.
I	*Infeasible.* The variable is basic or superbasic and is currently violating one of its bounds by more than the value of the optional parameter **Feasibility Tolerance** (default value $= \max(10^{-6}, \sqrt{\epsilon})$, where ϵ is the ***machine precision***; see Section 11.2).
N	*Not precisely optimal.* The variable is nonbasic or superbasic. If the value of the reduced gradient for the variable exceeds the value of the optional parameter **Optimality Tolerance** (default value $= \max(10^{-6}, \sqrt{\epsilon})$; see Section 11.2), the solution would not be declared optimal because the reduced gradient for the variable would not be considered negligible.

Value	is the value of the variable at the final iterate.
Lower Bound	is the lower bound specified for the variable. `None` indicates that BL(j) $\leq -bigbnd$.
Upper Bound	is the upper bound specified for the variable. `None` indicates that BU(j) $\geq bigbnd$.
Lagr Mult	is the Lagrange multiplier for the associated bound. This will be zero if `State` is `FR`. If x is optimal, the multiplier should be non-negative if `State` is `LL`, non-positive if `State` is `UL`, and zero if `State` is `BS` or `SBS`.

Residual   is the difference between the variable `Value` and the nearer of its (finite) bounds $BL(j)$ and $BU(j)$. A blank entry indicates that the associated variable is not bounded (i.e., $BL(j) \leq -bigbnd$ and $BU(j) \geq bigbnd$).

The meaning of the printout for linear constraints is the same as that given above for variables, with 'variable' replaced by 'constraint', n replaced by m, CRNAME(j) replaced by CRNAME($n + j$), $BL(j)$ and $BU(j)$ are replaced by $BL(n + j)$ and $BU(n + j)$ respectively, and with the following change in the heading:

Constrnt   gives the name of the linear constraint.

Note that movement off a constraint (as opposed to a variable moving away from its bound) can be interpreted as allowing the entry in the `Residual` column to become positive.

Numerical values are output with a fixed number of digits; they are not guaranteed to be accurate to this precision.

9   Example

To minimize the quadratic function $f(x) = c^T x + \frac{1}{2} x^T H x$, where

$$c = (-200.0, \ -2000.0, \ -2000.0, \ -2000.0, \ -2000.0, \ 400.0, \ 400.0)^T$$

$$H = \begin{pmatrix} 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 2 & 2 \end{pmatrix}$$

subject to the bounds

$$\begin{aligned} 0 &\leq x_1 \leq 200 \\ 0 &\leq x_2 \leq 2500 \\ 400 &\leq x_3 \leq 800 \\ 100 &\leq x_4 \leq 700 \\ 0 &\leq x_5 \leq 1500 \\ 0 &\leq x_6 \\ 0 &\leq x_7 \end{aligned}$$

and to the linear constraints

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 &= 2000 \\ 0.15x_1 + 0.04x_2 + 0.02x_3 + 0.04x_4 + 0.02x_5 + 0.01x_6 + 0.03x_7 &\leq 60 \\ 0.03x_1 + 0.05x_2 + 0.08x_3 + 0.02x_4 + 0.06x_5 + 0.01x_6 &\leq 100 \\ 0.02x_1 + 0.04x_2 + 0.01x_3 + 0.02x_4 + 0.02x_5 &\leq 40 \\ 0.02x_1 + 0.03x_2 + 0.01x_5 &\leq 30 \\ 1500 \leq 0.70x_1 + 0.75x_2 + 0.80x_3 + 0.75x_4 + 0.80x_5 + 0.97x_6 & \\ 250 \leq 0.02x_1 + 0.06x_2 + 0.08x_3 + 0.12x_4 + 0.02x_5 + 0.01x_6 + 0.97x_7 &\leq 300 \end{aligned}$$

The initial point, which is infeasible, is

$$x_0 = (0.0, \ 0.0, \ 0.0, \ 0.0, \ 0.0, \ 0.0, \ 0.0)^T.$$

The optimal solution (to five figures) is

$$x^* = (0.0, \ 349.40, \ 648.85, \ 172.85, \ 407.52, \ 271.36, \ 150.02)^T.$$

One bound constraint and four linear constraints are active at the solution. Note that the Hessian matrix H is positive semi-definite.

9.1   Program Text

Note. The listing of the example program presented below uses bold italicised terms to denote precision-dependent details. Please read the Users' Note for your implementation to check the interpretation of these terms. As explained in the Essential Introduction to this manual, the results produced may not be identical for all implementations.

```
*     E04NKF Example Program Text.
*     Mark 19 Revised. NAG Copyright 1999.
*     .. Parameters ..
      INTEGER          NIN, NOUT
      PARAMETER        (NIN=5,NOUT=6)
      INTEGER          IDUMMY
      PARAMETER        (IDUMMY=-11111)
      INTEGER          NMAX, MMAX, NNZMAX, LENIZ, LENZ
      PARAMETER        (NMAX=100,MMAX=100,NNZMAX=100,LENIZ=10000,
     +                 LENZ=10000)
*     .. Local Scalars ..
      real             OBJ, SINF
      INTEGER          I, ICOL, IFAIL, IOBJ, J, JCOL, M, MINIZ,
     +                 MINZ, N, NCOLH, NINF, NNAME, NNZ, NS
      CHARACTER        START
*     .. Local Arrays ..
      real             A(NNZMAX), BL(NMAX+MMAX), BU(NMAX+MMAX),
     +                 CLAMDA(NMAX+MMAX), XS(NMAX+MMAX), Z(LENZ)
      INTEGER          HA(NNZMAX), ISTATE(NMAX+MMAX), IZ(LENIZ),
     +                 KA(NMAX+1)
      CHARACTER*8      CRNAME(NMAX+MMAX), NAMES(5)
*     .. External Subroutines ..
      EXTERNAL         E04NKF, QPHX
*     .. Executable Statements ..
      WRITE (NOUT,*) 'E04NKF Example Program Results'
*     Skip heading in data file.
      READ (NIN,*)
      READ (NIN,*) N, M
      IF (N.LE.NMAX .AND. M.LE.MMAX) THEN
*
*        Read NNZ, IOBJ, NCOLH, START and NNAME from data file.
*
         READ (NIN,*) NNZ, IOBJ, NCOLH, START, NNAME
*
*        Read NAMES and CRNAME from data file.
*
         READ (NIN,*) (NAMES(I),I=1,5)
         READ (NIN,*) (CRNAME(I),I=1,NNAME)
*
*        Read the matrix A from data file. Set up KA.
*
         JCOL = 1
         KA(JCOL) = 1
         DO 40 I = 1, NNZ
*
*           Element ( HA( I ), ICOL ) is stored in A( I ).
*
            READ (NIN,*) A(I), HA(I), ICOL
*
            IF (ICOL.LT.JCOL) THEN
*
*              Elements not ordered by increasing column index.
*
               WRITE (NOUT,99999) 'Element in column', ICOL,
```

```
     +               ' found after element in column', JCOL, '. Problem',
     +               ' abandoned.'
                 STOP
             ELSE IF (ICOL.EQ.JCOL+1) THEN
*
*               Index in A of the start of the ICOL-th column equals I.
*
                 KA(ICOL) = I
                 JCOL = ICOL
             ELSE IF (ICOL.GT.JCOL+1) THEN
*
*               Index in A of the start of the ICOL-th column equals I,
*               but columns JCOL+1,JCOL+2,...,ICOL-1 are empty. Set the
*               corresponding elements of KA to I.
*
                 DO 20 J = JCOL + 1, ICOL - 1
                    KA(J) = I
   20            CONTINUE
                 KA(ICOL) = I
                 JCOL = ICOL
             END IF
   40     CONTINUE
*
        KA(N+1) = NNZ + 1
*
        IF (N.GT.ICOL) THEN
*
*          Columns N,N-1,...,ICOL+1 are empty. Set the corresponding
*          elements of KA accordingly.
*
           DO 60 I = N, ICOL + 1, -1
              IF (KA(I).EQ.IDUMMY) KA(I) = KA(I+1)
   60      CONTINUE
        END IF
*
*       Read BL, BU, ISTATE and XS from data file.
*
        READ (NIN,*) (BL(I),I=1,N+M)
        READ (NIN,*) (BU(I),I=1,N+M)
        IF (START.EQ.'C') THEN
           READ (NIN,*) (ISTATE(I),I=1,N)
        ELSE IF (START.EQ.'W') THEN
           READ (NIN,*) (ISTATE(I),I=1,N+M)
        END IF
        READ (NIN,*) (XS(I),I=1,N)
*
*       Solve the QP problem.
*
        IFAIL = -1
*
        CALL E04NKF(N,M,NNZ,IOBJ,NCOLH,QPHX,A,HA,KA,BL,BU,START,NAMES,
     +              NNAME,CRNAME,NS,XS,ISTATE,MINIZ,MINZ,NINF,SINF,OBJ,
     +              CLAMDA,IZ,LENIZ,Z,LENZ,IFAIL)
*
      END IF
      STOP
*
99999 FORMAT (/1X,A,I5,A,I5,A,A)
```

```
      END
*
      SUBROUTINE QPHX(NSTATE,NCOLH,X,HX)
*
*     Routine to compute H*x. (In this version of QPHX, the Hessian
*     matrix H is not referenced explicitly.)
*
*     .. Parameters ..
      real          TWO
      PARAMETER     (TWO=2.0e+0)
*     .. Scalar Arguments ..
      INTEGER       NCOLH, NSTATE
*     .. Array Arguments ..
      real          HX(NCOLH), X(NCOLH)
*     .. Executable Statements ..
*
      HX(1) = TWO*X(1)
      HX(2) = TWO*X(2)
      HX(3) = TWO*(X(3)+X(4))
      HX(4) = HX(3)
      HX(5) = TWO*X(5)
      HX(6) = TWO*(X(6)+X(7))
      HX(7) = HX(6)
*
      RETURN
      END
```

9.2 Program Data

```
E04NKF Example Program Data
 7  8                 :Values of N and M
48  8  7  'C'  15 :Values of NNZ, IOBJ, NCOLH, START and NNAME
'          ' '         ' '          ' '          ' '          ' :End of NAMES
'...X1...' '...X2...' '...X3...' '...X4...' '...X5...'
'...X6...' '...X7...' '..ROW1..' '..ROW2..' '..ROW3..'
'..ROW4..' '..ROW5..' '..ROW6..' '..ROW7..' '..COST..' :End of CRNAME
      0.02   7   1
      0.02   5   1
      0.03   3   1
      1.00   1   1
      0.70   6   1
      0.02   4   1
      0.15   2   1
   -200.00   8   1
      0.06   7   2
      0.75   6   2
      0.03   5   2
      0.04   4   2
      0.05   3   2
      0.04   2   2
      1.00   1   2
  -2000.00   8   2
      0.02   2   3
      1.00   1   3
      0.01   4   3
      0.08   3   3
      0.08   7   3
      0.80   6   3
```

```
-2000.00  8   3
    1.00  1   4
    0.12  7   4
    0.02  3   4
    0.02  4   4
    0.75  6   4
    0.04  2   4
-2000.00  8   4
    0.01  5   5
    0.80  6   5
    0.02  7   5
    1.00  1   5
    0.02  2   5
    0.06  3   5
    0.02  4   5
-2000.00  8   5
    1.00  1   6
    0.01  2   6
    0.01  3   6
    0.97  6   6
    0.01  7   6
  400.00  8   6
    0.97  7   7
    0.03  2   7
    1.00  1   7
  400.00  8   7                                          :End of matrix A
  0.0      0.0      4.0E+02  1.0E+02 0.0     0.0     0.0      2.0E+03
 -1.0E+25 -1.0E+25 -1.0E+25 -1.0E+25 1.5E+03 2.5E+02 -1.0E+25 :End of BL
  2.0E+02  2.5E+03  8.0E+02  7.0E+02 1.5E+03 1.0E+25 1.0E+25  2.0E+03
  6.0E+01  1.0E+02  4.0E+01  3.0E+01 1.0E+25 3.0E+02 1.0E+25  :End of BU
  0    0    0    0    0    0    0    0   :End of ISTATE
  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0 :End of XS
```

9.3 Program Results

```
E04NKF Example Program Results

*** E04NKF
*** Start of NAG Library implementation details ***

Implementation title: Generalised Base Version
           Precision: FORTRAN double precision
        Product Code: FLBAS18D
                Mark: 18A

*** End of NAG Library implementation details ***

Parameters
----------

Frequencies.
Check frequency........       60       Expand frequency........     10000
Factorization frequency.     100

LP Parameters.
Scale tolerance........  9.00E-01      Feasibility tolerance...  1.00E-06
Iteration limit........       75       Scale option............        2
Optimality tolerance....  1.00E-06     Partial price...........       10
```

```
   Crash tolerance........ 1.00E-01        Pivot tolerance........  2.04E-11
   Crash option............       2


   QP objective.
   Objective variables.....       7        Hessian columns........        7
   Superbasics limit.......       7


   Miscellaneous.
   Variables..............        7        Linear constraints......       8
   LU factor tolerance..... 1.00E+02       LU update tolerance..... 1.00E+01
   LU singularity tolerance 2.04E-11       Monitoring file.........      -1
   EPS (machine precision). 1.11E-16       Print level.............      10
   Infinite bound size..... 1.00E+20       Infinite step size...... 1.00E+20
   COLD start.............                 MINIMIZE...............


   Workspace provided is              IZ(  10000),  Z(  10000).
   To start solving the problem we need IZ(    428),  Z(    358).



    Itn    Step  Ninf  Sinf/Objective  Norm rg
   Itn      0 -- Infeasible.
      0  0.0E+00     1    1.152891E+03   0.0E+00
      1  4.3E+02     0    0.000000E+00   0.0E+00
   Itn      1 -- Feasible point found (for     1 equality constraints).
      1  0.0E+00     0    0.000000E+00   0.0E+00
      1  0.0E+00     0    1.460000E+06   0.0E+00
   Itn      1 -- Feasible QP solution.
      2  8.7E-02     0    9.409959E+05   0.0E+00
      3  5.3E-01     0   -1.056552E+06   0.0E+00
      4  1.0E+00     0   -1.462190E+06   2.3E-12
      5  1.0E+00     0   -1.698092E+06   2.2E-12
      6  4.6E-02     0   -1.764906E+06   7.0E+02
      7  1.0E+00     0   -1.811946E+06   2.8E-12
      8  1.7E-02     0   -1.847325E+06   1.7E+02
      9  1.0E+00     0   -1.847785E+06   7.0E-12
```

Variable	State	Value	Lower Bound	Upper Bound	Lagr Mult	Residual
...X1...	LL	0.000000E+00	.	200.00	2361.	.
...X2...	BS	349.399	.	2500.0	-3.6566E-12	349.4
...X3...	SBS	648.853	400.00	800.00	-5.9240E-12	151.1
...X4...	SBS	172.847	100.00	700.00	1.9491E-12	72.85
...X5...	BS	407.521	.	1500.0	.	407.5
...X6...	BS	271.356	.	None	-3.2802E-12	271.4
...X7...	BS	150.023	.	None	-1.4130E-12	150.0

Constrnt	State	Value	Lower Bound	Upper Bound	Lagr Mult	Residual
..ROW1..	EQ	2000.00	2000.0	2000.0	-1.2901E+04	.
..ROW2..	BS	49.2316	None	60.000	.	-10.77
..ROW3..	UL	100.000	None	100.00	-2325.	.
..ROW4..	BS	32.0719	None	40.000	.	-7.928
..ROW5..	BS	14.5572	None	30.000	.	-15.44
..ROW6..	LL	1500.00	1500.0	None	1.4455E+04	.
..ROW7..	LL	250.000	250.00	300.00	1.4581E+04	.
..COST..	BS	-2.988690E+06	None	None	-1.000	-2.9887E+06

```
       Exit E04NKF - Optimal QP solution found.

       Final QP objective value =     -1847785.

       Exit from QP problem after        9 iterations.
```

The remainder of this document is intended for more advanced users. Section 10 contains a detailed algorithm description that may be needed in order to understand Section 11 and Section 12. Section 11 describes the optional parameters that may be set by calls to E04NLF and/or E04NMF. Section 12 describes the quantities that can be requested to monitor the course of the computation.

10   Algorithmic Details

This section contains a description of the method used by E04NKF.

10.1   Overview

E04NKF is based on an inertia-controlling method that maintains a Cholesky factorization of the reduced Hessian (see below). The method is similar to that of Gill and Murray [3], and is described in detail by Gill *et al.* [5]. Here we briefly summarize the main features of the method. Where possible, explicit reference is made to the names of variables that are parameters of the routine or appear in the printed output.

The method used has two distinct phases: finding an initial feasible point by minimizing the sum of infeasibilities (the *feasibility phase*), and minimizing the quadratic objective function within the feasible region (the *optimality phase*). The computations in both phases are performed by the same subroutines. The two-phase nature of the algorithm is reflected by changing the function being minimized from the sum of infeasibilities (the printed quantity `Sinf`; see Section 12) to the quadratic objective function (the printed quantity `Objective`; see Section 12).

In general, an iterative process is required to solve a quadratic program. Given an iterate (x, s) in both the original variables x and the slack variables s, a new iterate $(\bar{x}, \bar{s})$ is defined by

$$\left(\begin{array}{c} \bar{x} \\ \bar{s} \end{array} \right) = \left(\begin{array}{c} x \\ s \end{array} \right) + \alpha p, \tag{2}$$

where the *step length* α is a non-negative scalar (the printed quantity `Step`; see Section 12), and p is called the *search direction*. (For simplicity, we shall consider a typical iteration and avoid reference to the index of the iteration.) Once an iterate is feasible (i.e., satisfies the constraints), all subsequent iterates remain feasible.

10.2   Definition of the Working Set and Search Direction

At each iterate (x, s), a *working set* of constraints is defined to be a linearly independent subset of the constraints that are satisfied 'exactly' (to within the value of the optional parameter **Feasibility Tolerance**; see Section 11.2). The working set is the current prediction of the constraints that hold with equality at a solution of the LP or QP problem. Let m_W denote the number of constraints in the working set (including bounds), and let W denote the associated m_W by $(n + m)$ *working set matrix* consisting of the m_W gradients of the working set constraints.

The search direction is defined so that constraints in the working set remain *unaltered* for any value of the step length. It follows that p must satisfy the identity

$$W p = 0. \tag{3}$$

This characterization allows p to be computed using any n by n_Z full-rank matrix Z that spans the null space of W. (Thus, $n_Z = n - m_W$ and $WZ = 0$.) The null space matrix Z is defined from a sparse LU factorization of part of W (see (6) and (7) below). The direction p will satisfy (3) if

$$p = Z p_Z, \tag{4}$$

where p_Z is any n_Z-vector.

The working set contains the constraints $Ax - s = 0$ and a subset of the upper and lower bounds on the variables (x, s). Since the gradient of a bound constraint $x_j \geq l_j$ or $x_j \leq u_j$ is a vector of all zeros except for ± 1 in position j, it follows that the working set matrix contains the rows of $(A \;\; -I)$ and the unit rows associated with the upper and lower bounds in the working set.

The working set matrix W can be represented in terms of a certain column partition of the matrix $(A -I)$. As in Section 3 we partition the constraints $Ax - s = 0$ so that

$$Bx_B + Sx_S + Nx_N = 0, \tag{5}$$

where B is a square non-singular basis and x_B, x_S and x_N are the basic, superbasic and nonbasic variables respectively. The nonbasic variables are equal to their upper or lower bounds at (x, s), and the superbasic variables are independent variables that are chosen to improve the value of the current objective function. The number of superbasic variables is n_S (the printed quantity Ns; see Section 12). Given values of x_N and x_S, the basic variables x_B are adjusted so that (x, s) satisfies (5).

If P is a permutation matrix such that $(A \; -I)P = (B \; S \; N)$, then the working set matrix W satisfies

$$WP = \left(\begin{array}{ccc} B & S & N \\ 0 & 0 & I_N \end{array} \right), \tag{6}$$

where I_N is the identity matrix with the same number of columns as N.

The null space matrix Z is defined from a sparse LU factorization of part of W. In particular, Z is maintained in 'reduced gradient' form, using the LUSOL package (see Gill *et al.* [6]) to maintain sparse LU factors of the basis matrix B that alters as the working set W changes. Given the permutation P, the null space basis is given by

$$Z = P \left(\begin{array}{c} -B^{-1}S \\ I \\ 0 \end{array} \right). \tag{7}$$

This matrix is used only as an operator, i.e., it is never computed explicitly. Products of the form Zv and $Z^T g$ are obtained by solving with B or B^T. This choice of Z implies that n_Z, the number of 'degrees of freedom' at (x, s), is the same as n_S, the number of superbasic variables.

Let g_Z and H_Z denote the *reduced gradient* and *reduced Hessian* of the objective function:

$$g_Z = Z^T g \;\; \text{and} \;\; H_Z = Z^T HZ, \tag{8}$$

where g is the objective gradient at (x, s). Roughly speaking, g_Z and H_Z describe the first and second derivatives of an n_S-dimensional *unconstrained* problem for the calculation of p_Z. (The condition estimator of H_Z is the quantity Cond Hz in the monitoring file output; see Section 12.)

At each iteration, an upper triangular factor R is available such that $H_Z = R^T R$. Normally, R is computed from $R^T R = Z^T HZ$ at the start of the optimality phase and then updated as the QP working set changes. For efficiency, the dimension of R should not be excessive (say, $n_S \leq 1000$). This is guaranteed if the number of nonlinear variables is 'moderate'.

If the QP problem contains linear variables, H is positive semi-definite and R may be singular with at least one zero diagonal element. However, an inertia-controlling strategy is used to ensure that only the last diagonal element of R can be zero. (See Gill *et al.* [5] for a discussion of a similar strategy for indefinite quadratic programming.)

If the initial R is singular, enough variables are fixed at their current value to give a non-singular R. This is equivalent to including temporary bound constraints in the working set. Thereafter, R can become singular only when a constraint is deleted from the working set (in which case no further constraints are deleted until R becomes non-singular).

10.3 The Main Iteration

If the reduced gradient is zero, (x, s) is a constrained stationary point on the working set. During the feasibility phase, the reduced gradient will usually be zero only at a vertex (although it may be zero elsewhere in the presence of constraint dependencies). During the optimality phase, a zero reduced

gradient implies that x minimizes the quadratic objective function when the constraints in the working set are treated as equalities. At a constrained stationary point, Lagrange multipliers λ are defined from the equations

$$W^T \lambda = g(x). \tag{9}$$

A Lagrange multiplier λ_j corresponding to an inequality constraint in the working set is said to be *optimal* if $\lambda_j \leq \sigma$ when the associated constraint is at its *upper bound*, or if $\lambda_j \geq -\sigma$ when the associated constraint is at its *lower bound*, where σ depends on the value of the optional parameter **Optimality Tolerance** (see Section 11.2). If a multiplier is non-optimal, the objective function (either the true objective or the sum of infeasibilities) can be reduced by continuing the minimization with the corresponding constraint excluded from the working set. (This step is sometimes referred to as 'deleting' a constraint from the working set.) If optimal multipliers occur during the feasibility phase but the sum of infeasibilities is non-zero, there is no feasible point and the routine terminates immediately with IFAIL = 3 (see Section 6).

The special form (6) of the working set allows the multiplier vector λ, the solution of (9), to be written in terms of the vector

$$d = \begin{pmatrix} g \\ 0 \end{pmatrix} - (A \ -I)^T \pi = \begin{pmatrix} g - A^T \pi \\ \pi \end{pmatrix}, \tag{10}$$

where π satisfies the equations $B^T \pi = g_B$, and g_B denotes the basic elements of g. The elements of π are the Lagrange multipliers λ_j associated with the equality constraints $Ax - s = 0$. The vector d_N of nonbasic elements of d consists of the Lagrange multipliers λ_j associated with the upper and lower bound constraints in the working set. The vector d_S of superbasic elements of d is the reduced gradient g_Z in (8). The vector d_B of basic elements of d is zero, by construction. (The Euclidean norm of d_S and the final values of d_S, g and π are the quantities `Norm rg`, `Reduced Gradnt`, `Obj Gradient` and `Dual Activity` in the monitoring file output; see Section 12.)

If the reduced gradient is not zero, Lagrange multipliers need not be computed and the search direction is given by $p = Z p_Z$ (see (7) and (11)). The step length is chosen to maintain feasibility with respect to the satisfied constraints.

There are two possible choices for p_Z, depending on whether or not H_Z is singular. If H_Z is non-singular, R is non-singular and p_Z in (4) is computed from the equations

$$R^T R p_Z = -g_Z, \tag{11}$$

where g_Z is the reduced gradient at x. In this case, $(x, s) + p$ is the minimizer of the objective function subject to the working set constraints being treated as equalities. If $(x, s) + p$ is feasible, α is defined to be unity. In this case, the reduced gradient at $(\bar{x}, \bar{s})$ will be zero, and Lagrange multipliers are computed at the next iteration. Otherwise, α is set to α_M, the step to the 'nearest' constraint along p. This constraint is added to the working set at the next iteration.

If H_Z is singular, then R must also be singular, and an inertia-controlling strategy is used to ensure that only the last diagonal element of R is zero. (See Gill *et al.* [5] for a discussion of a similar strategy for indefinite quadratic programming.) In this case, p_Z satisfies

$$p_Z^T H_Z p_Z = 0 \ \text{ and } \ g_Z^T p_Z \leq 0, \tag{12}$$

which allows the objective function to be reduced by any step of the form $(x, s) + \alpha p$, where $\alpha > 0$. The vector $p = Z p_Z$ is a direction of unbounded descent for the QP problem in the sense that the QP objective is linear and decreases without bound along p. If no finite step of the form $(x, s) + \alpha p$ (where $\alpha > 0$) reaches a constraint not in the working set, the QP problem is unbounded and the routine terminates immediately with IFAIL = 2 (see Section 6). Otherwise, α is defined as the maximum feasible step along p and a constraint active at $(x, s) + \alpha p$ is added to the working set for the next iteration.

10.4 Miscellaneous

If the basis matrix is not chosen carefully, the condition of the null space matrix Z in (7) could be arbitrarily high. To guard against this, the routine implements a 'basis repair' feature in which the LUSOL package (see Gill *et al.* [6]) is used to compute the rectangular factorization

$$(B \ S)^T = LU, \tag{13}$$

returning just the permutation P that makes PLP^T unit lower triangular. The pivot tolerance is set to require $|PLP^T|_{ij} \le 2$, and the permutation is used to define P in (6). It can be shown that $\|Z\|$ is likely to be little more than unity. Hence, Z should be well-conditioned *regardless of the condition of W*. This feature is applied at the beginning of the optimality phase if a potential $B - S$ ordering is known.

The EXPAND procedure (see Gill *et al.* [4]) is used to reduce the possibility of cycling at a point where the active constraints are nearly linearly dependent. Although there is no absolute guarantee that cycling will not occur, the probability of cycling is extremely small (see Hall and McKinnon [7]). The main feature of EXPAND is that the feasibility tolerance is increased at the start of every iteration. This allows a positive step to be taken at every iteration, perhaps at the expense of violating the bounds on (x, s) by a small amount.

Suppose that the value of the optional parameter **Feasibility Tolerance** (see Section 11.2) is δ. Over a period of K iterations (where K is the value of the optional parameter **Expand Frequency**; see Section 11.2), the feasibility tolerance actually used by E04NKF (i.e., the *working* feasibility tolerance) increases from 0.5δ to δ (in steps of $0.5\delta/K$).

At certain stages the following 'resetting procedure' is used to remove small constraint infeasibilities. First, all nonbasic variables are moved exactly onto their bounds. A count is kept of the number of non-trivial adjustments made. If the count is non-zero, the basic variables are recomputed. Finally, the working feasibility tolerance is reinitialized to 0.5δ.

If a problem requires more than K iterations, the resetting procedure is invoked and a new cycle of iterations is started. (The decision to resume the feasibility phase or optimality phase is based on comparing any constraint infeasibilities with δ.)

The resetting procedure is also invoked when E04NKF reaches an apparently optimal, infeasible or unbounded solution, unless this situation has already occurred twice. If any non-trivial adjustments are made, iterations are continued.

The EXPAND procedure not only allows a positive step to be taken at every iteration, but also provides a potential *choice* of constraints to be added to the working set. All constraints at a distance α (where $\alpha \le \alpha_{\mathrm{M}}$) along p from the current point are then viewed as acceptable candidates for inclusion in the working set. The constraint whose normal makes the largest angle with the search direction is added to the working set. This strategy helps keep the basis matrix B well-conditioned.

11 Optional Parameters

Several optional parameters in E04NKF define choices in the problem specification or the algorithm logic. In order to reduce the number of formal parameters of E04NKF these optional parameters have associated *default values* that are appropriate for most problems. Therefore, the user need only specify those optional parameters whose values are to be different from their default values.

The remainder of this section can be skipped by users who wish to use the default values for *all* optional parameters. A complete list of optional parameters and their default values is given in Section 11.1.

Optional parameters may be specified by calling one, or both, of the routines E04NLF and E04NMF prior to a call to E04NKF.

E04NLF reads options from an external options file, with `Begin` and `End` as the first and last lines respectively and each intermediate line defining a single optional parameter. For example,

```
Begin
  Print Level = 5
End
```

The call

```
CALL E04NLF (IOPTNS, INFORM)
```

can then be used to read the file on unit IOPTNS. INFORM will be zero on successful exit. E04NLF should be consulted for a full description of this method of supplying optional parameters.

E04NMF can be called to supply options directly, one call being necessary for each optional parameter. For example,

```
CALL E04NMF ('Print Level = 5')
```

E04NMF should be consulted for a full description of this method of supplying optional parameters.

All optional parameters not specified by the user are set to their default values. Optional parameters specified by the user are unaltered by E04NKF (unless they define invalid values) and so remain in effect for subsequent calls unless altered by the user.

11.1 Optional Parameter Checklist and Default Values

For easy reference, the following list shows all the valid keywords and their default values. The symbol ϵ represents the **machine precision** (see X02AJF).

Optional Parameters	Default Values
Check frequency	60
Crash option	2
Crash tolerance	0.1
Defaults	
Expand frequency	10000
Factorization frequency	100
Feasibility tolerance	$\max(10^{-6}, \sqrt{\epsilon})$
Infinite bound size	10^{20}
Infinite step size	$\max(bigbnd, 10^{20})$
Iteration limit	$\max(50, 5(n+m))$
List/Nolist	**List**
LU factor tolerance	100.0
LU singularity tolerance	$\epsilon^{0.67}$
LU update tolerance	10.0
Maximize/Minimize	**Minimize**
Monitoring file	-1
Optimality tolerance	$\max(10^{-6}, \sqrt{\epsilon})$
Partial price	10
Pivot tolerance	$\epsilon^{0.67}$
Print level	10
Scale option	2
Scale tolerance	0.9
Superbasics limit	$\min(n_H + 1, n)$
Rank tolerance	100ϵ

11.2 Description of the Optional Parameters

The following list (in alphabetical order) gives the valid options. For each option, we give the keyword, any essential optional qualifiers, the default value, and the definition. The minimum abbreviation of each keyword is underlined. If no characters of an optional qualifier are underlined, the qualifier may be omitted. The letters i and r denote INTEGER and **real** values required with certain options. The default value of an option is used whenever the condition $|i| \geq 100000000$ is satisfied. The number ϵ is a generic notation for **machine precision** (see X02AJF).

<u>Check</u> Frequency i Default $= 60$

Every ith iteration after the most recent basis factorization, a numerical test is made to see if the current solution (x, s) satisfies the linear constraints $Ax - s = 0$. If the largest element of the residual vector $r = Ax - s$ is judged to be too large, the current basis is refactorized and the basic variables recomputed to satisfy the constraints more accurately. If $i < 0$, the default value is used. If $i = 0$, the value $i = 99999999$ is used and effectively no checks are made.

C̲rash O̲ption   *i*   Default = 2

Note that this option does not apply when START = 'W' (see Section 5).

If START = 'C', an internal Crash procedure is used to select an initial basis from various rows and columns of the constraint matrix $(A - I)$. The value of i determines which rows and columns are initially eligible for the basis, and how many times the Crash procedure is called. If $i = 0$, the all-slack basis $B = -I$ is chosen. If $i = 1$, the Crash procedure is called once (looking for a triangular basis in all rows and columns of the linear constraint matrix A). If $i = 2$, the Crash procedure is called twice (looking at any *equality* constraints first followed by any *inequality* constraints). If $i < 0$ or $i > 2$, the default value is used.

If $i = 1$ or 2, certain slacks on inequality rows are selected for the basis first. (If $i = 2$, numerical values are used to exclude slacks that are close to a bound.) The Crash procedure then makes several passes through the columns of A, searching for a basis matrix that is essentially triangular. A column is assigned to 'pivot' on a particular row if the column contains a suitably large element in a row that has not yet been assigned. (The pivot elements ultimately form the diagonals of the triangular basis.) For remaining unassigned rows, slack variables are inserted to complete the basis.

C̲rash T̲olerance   *r*   Default = 0.1

This value allows the Crash procedure to ignore certain 'small' non-zero elements in the constraint matrix A while searching for a triangular basis. For each column of A, if a_{max} is the largest element in the column, other non-zeros in that column are ignored if they are less than (or equal to) $a_{max} \times r$.

When $r > 0$, the basis obtained by the Crash procedure may not be strictly triangular, but it is likely to be non-singular and almost triangular. The intention is to obtain a starting basis with more column variables and fewer (arbitrary) slacks. A feasible solution may be reached earlier for some problems. If $r < 0$ or $r \geq 1$, the default value is used.

D̲efaults

This special keyword may be used to reset all optional parameters to their default values.

E̲xpand Frequency   *i*   Default = 10000

This option is part of an anti-cycling procedure (see Section 10.4) designed to allow progress even on highly degenerate problems.

For LP problems, the strategy is to force a positive step at every iteration, at the expense of violating the constraints by a small amount. Suppose that the value of the optional parameter **Feasibility Tolerance** is δ. Over a period of i iterations, the feasibility tolerance actually used by E04NKF (i.e., the *working* feasibility tolerance) increases from 0.5δ to δ (in steps of $0.5\delta/i$).

For QP problems, the same procedure is used for iterations in which there is only one superbasic variable. (Cycling can only occur when the current solution is at a vertex of the feasible region.) Thus, zero steps are allowed if there is more than one superbasic variable, but otherwise positive steps are enforced.

Increasing the value of i helps reduce the number of slightly infeasible nonbasic basic variables (most of which are eliminated during the resetting procedure). However, it also diminishes the freedom to choose a large pivot element (see **Pivot Tolerance** below).

If $i < 0$, the default value is used. If $i = 0$, the value $i = 99999999$ is used and effectively no anti-cycling procedure is invoked.

F̲actorization Frequency   *i*   Default = 100

If $i > 0$, at most i basis changes will occur between factorizations of the basis matrix. For LP problems, the basis factors are usually updated at every iteration. For QP problems, fewer basis updates will occur as the solution is approached. The number of iterations between basis factorizations will therefore increase. During these iterations a test is made regularly according to the value of **Check Frequency** (see above) to ensure that the linear constraints $Ax - s = 0$ are satisfied. If necessary, the basis will be refactorized before the limit of i updates is reached. If $i \leq 0$, the default value is used.

F̲easibility Tolerance   *r*   Default = $\max(10^{-6}, \sqrt{\epsilon})$

If $r \geq \epsilon$, r defines the maximum acceptable *absolute* violation in each constraint at a 'feasible' point (including slack variables). For example, if the variables and the coefficients in the linear constraints are

of order unity, and the latter are correct to about 5 decimal digits, it would be appropriate to specify r as 10^{-5}. If $r < \epsilon$, the default value is used.

E04NKF attempts to find a feasible solution before optimizing the objective function. If the sum of infeasibilities cannot be reduced to zero, the problem is assumed to be *infeasible*. Let `Sinf` be the corresponding sum of infeasibilities. If `Sinf` is quite small, it may be appropriate to raise r by a factor of 10 or 100. Otherwise, some error in the data should be suspected. Note that the routine does *not* attempt to find the minimum value of `Sinf`.

If the constraints and variables have been scaled (see **Scale Option** below), then feasibility is defined in terms of the scaled problem (since it is more likely to be meaningful).

<u>In</u>finite <u>B</u>ound Size r Default $= 10^{20}$

If $r > 0$, r defines the 'infinite' bound *bigbnd* in the definition of the problem constraints. Any upper bound greater than or equal to *bigbnd* will be regarded as plus infinity (and similarly any lower bound less than or equal to $-bigbnd$ will be regarded as minus infinity). If $r \leq 0$, the default value is used.

<u>In</u>finite <u>S</u>tep Size r Default $= \max(bigbnd, 10^{20})$

If $r > 0$, r specifies the magnitude of the change in variables that will be considered a step to an unbounded solution. (Note that an unbounded solution can occur only when the Hessian is not positive-definite.) If the change in x during an iteration would exceed the value of r, the objective function is considered to be unbounded below in the feasible region. If $r \leq 0$, the default value is used.

<u>It</u>eration Limit i Default $= \max(50, 5(n + m))$
<u>It</u>ers
<u>It</u>ns

The value of i specifies the maximum number of iterations allowed before termination. Setting $i = 0$ and **Print Level** > 0 means that the workspace needed to start solving the problem will be computed and printed, but no iterations will be performed. If $i < 0$, the default value is used.

<u>L</u>ist   Default $=$ **List**
<u>N</u>olist

Normally each optional parameter specification is printed as it is supplied. **Nolist** may be used to suppress the printing and **List** may be used to restore printing.

<u>LU</u> <u>F</u>actor Tolerance r_1 Default $= 100.0$
<u>LU</u> <u>U</u>pdate Tolerance r_2 Default $= 10.0$

The values of r_1 and r_2 affect the stability and sparsity of the basis factorization $B = LU$, during refactorization and updates respectively. The lower triangular matrix L is a product of matrices of the form

$$\begin{pmatrix} 1 & \\ \mu & 1 \end{pmatrix}$$

where the multipliers μ will satisfy $|\mu| \leq r_i$. The default values of r_1 and r_2 usually strike a good compromise between stability and sparsity. For large and relatively dense problems, setting r_1 and r_2 to 25 (say) may give a marked improvement in sparsity without impairing stability to a serious degree. Note that for band matrices it may be necessary to set r_1 in the range $1 \leq r_1 < 2$ in order to achieve stability. If $r_1 < 1$ or $r_2 < 1$, the default value is used.

<u>LU</u> <u>S</u>ingularity Tolerance r Default $= \epsilon^{0.67}$

If $r > 0$, r defines the singularity tolerance used to guard against ill-conditioned basis matrices. Whenever the basis is refactorized, the diagonal elements of U are tested as follows. If $|u_{jj}| \leq r$ or $|u_{jj}| < r \times \max_i |u_{ij}|$, the jth column of the basis is replaced by the corresponding slack variable. If $r \leq 0$, the default value is used.

<u>M</u>aximize   Default $=$ **Minimize**
<u>M</u>inimize

This option specifies the required direction of the optimization. It applies to both linear and nonlinear terms (if any) in the objective function. Note that if two problems are the same except that one minimizes

$f(x)$ and the other maximizes $-f(x)$, their solutions will be the same but the signs of the dual variables π_i and the reduced gradients d_j (see Section 10.3) will be reversed.

Monitoring File $\qquad\qquad\qquad\qquad\qquad\qquad i \qquad\qquad\qquad\qquad\qquad\qquad$ Default $= -1$

If $i \geq 0$ and **Print Level** > 0 (see below), monitoring information produced by E04NKF is sent to a file with logical unit number i. If $i < 0$ and/or **Print Level** $= 0$, the default value is used and hence no monitoring information is produced.

Optimality Tolerance $\qquad\qquad\qquad\qquad\quad r \qquad\qquad\qquad\qquad$ Default $= \max(10^{-6}, \sqrt{\epsilon})$

If $r \geq \epsilon$, r is used to judge the size of the reduced gradients $d_j = g_j - \pi^T a_j$. By definition, the reduced gradients for basic variables are always zero. Optimality is declared if the reduced gradients for any nonbasic variables at their lower or upper bounds satisfy $-r \times \max(1, \| \pi \|) \leq d_j \leq r \times \max(1, \| \pi \|)$, and if $|d_j| \leq r \times \max(1, \| \pi \|)$ for any superbasic variables. If $r < \epsilon$, the default value is used.

Partial Price $\qquad\qquad\qquad\qquad\qquad\qquad\qquad i \qquad\qquad\qquad\qquad\qquad\qquad\qquad$ Default $= 10$

Note that this option does not apply to QP problems.

This option is recommended for large FP or LP problems that have significantly more variables than constraints (i.e., $n \gg m$). It reduces the work required for each pricing operation (i.e., when a nonbasic variable is selected to enter the basis). If $i = 1$, all columns of the constraint matrix $(A - I)$ are searched. If $i > 1$, A and $-I$ are partitioned to give i roughly equal segments A_j, K_j, for $j = 1, 2, \ldots, p$ (modulo p). If the previous pricing search was successful on A_{j-1}, K_{j-1}, the next search begins on the segments A_j, K_j. If a reduced gradient is found that is larger than some dynamic tolerance, the variable with the largest such reduced gradient (of appropriate sign) is selected to enter the basis. If nothing is found, the search continues on the next segments A_{j+1}, K_{j+1}, and so on. If $i \leq 0$, the default value is used.

Pivot Tolerance $\qquad\qquad\qquad\qquad\qquad\qquad r \qquad\qquad\qquad\qquad\qquad\qquad$ Default $= \epsilon^{0.67}$

If $r > 0$, r is used to prevent columns entering the basis if they would cause the basis to become almost singular. If $r \leq 0$, the default value is used.

Print Level $\qquad\qquad\qquad\qquad\qquad\qquad\qquad i \qquad\qquad\qquad\qquad\qquad\qquad\qquad$ Default $= 10$

The value of i controls the amount of printout produced by E04NKF, as indicated below. A detailed description of the printed output is given in Section 8.1 (summary output at each iteration and the final solution) and Section 12 (monitoring information at each iteration). Note that the summary output will not exceed 80 characters per line and that the monitoring information will not exceed 120 characters per line. If $i < 0$, the default value is used.

The following printout is sent to the current advisory message unit (as defined by X04ABF):

i	**Output**
0	No output.
1	The final solution only.
5	One line of summary output for each iteration (no printout of the final solution).
≥ 10	The final solution and one line of summary output for each iteration.

The following printout is sent to the logical unit number defined by the optional parameter **Monitoring File** (see above):

i	**Output**
0	No output.
1	The final solution only.
5	One long line of output for each iteration (no printout of the final solution).
≥ 10	The final solution and one long line of output for each iteration.
≥ 20	The final solution, one long line of output for each iteration, matrix statistics (initial status of rows and columns, number of elements, density, biggest and smallest elements, etc.), details of the scale factors resulting from the scaling procedure (if **Scale Option** $= 1$ or 2; see below), basis factorization statistics and details of the initial basis resulting from the Crash procedure (if START $=$ 'C'; see Section 5).

If **Print Level** > 0 and the unit number defined by **Monitoring File** is the same as that defined by X04ABF, then the summary output is suppressed.

<u>Sc</u>ale <u>O</u>ption $\qquad\qquad\qquad\qquad\qquad\qquad i \qquad\qquad\qquad\qquad\qquad\qquad$ Default $= 2$

This option enables you to scale the variables and constraints using an iterative procedure due to Fourer (see [8]), which attempts to compute row scales r_i and column scales c_j such that the scaled matrix coefficients $\bar{a}_{ij} = a_{ij} \times (c_j/r_i)$ are as close as possible to unity. This may improve the overall efficiency of the routine on some problems. (The lower and upper bounds on the variables and slacks for the scaled problem are redefined as $\bar{l}_j = l_j/c_j$ and $\bar{u}_j = u_j/c_j$ respectively, where $c_j \equiv r_{j-n}$ if $j > n$.)

If $i = 0$, no scaling is performed. If $i = 1$, all rows and columns of the constraint matrix A are scaled. If $i = 2$, an additional scaling is performed that may be helpful when the solution x is large; it takes into account columns of $(A - I)$ that are fixed or have positive lower bounds or negative upper bounds. If $i < 0$ or $i > 2$, the default value is used.

<u>Sc</u>ale <u>T</u>olerance $\qquad\qquad\qquad\qquad\qquad\qquad r \qquad\qquad\qquad\qquad\qquad\qquad$ Default $= 0.9$

Note that this option does not apply when **Scale Option** $= 0$ (see above).

If $0 < r < 1$, r is used to control the number of scaling passes to be made through the constraint matrix A. At least 3 (and at most 10) passes will be made. More precisely, let a_p denote the largest column ratio (i.e., $\frac{'\text{biggest}' \text{ element}}{'\text{smallest}' \text{ element}}$ in some sense) after the pth scaling pass through A. The scaling procedure is terminated if $a_p \geq a_{p-1} \times r$ for some $p \geq 3$. Thus, increasing the value of r from 0.9 to 0.99 (say) will probably increase the number of passes through A. If $r \leq 0$ or $r \geq 1$, the default value is used.

<u>S</u>uperbasics Limit $\qquad\qquad\qquad\qquad\qquad\qquad i \qquad\qquad\qquad\qquad$ Default $= \min(n_H + 1, n)$

Note that this option does not apply to FP or LP problems.

The value of i specifies 'how nonlinear' you expect the QP problem to be. If $i \leq 0$, the default value is used.

12 Description of Monitoring Information

This section describes the intermediate printout and final printout which constitutes the monitoring information produced by E04NKF. (See also the description of the optional parameters **Monitoring File** and **Print Level** in Section 11.2). The level of printed output can be controlled by the user.

When **Print Level** $= 5$ or ≥ 10 and **Monitoring File** ≥ 0, the following line of intermediate printout (< 120 characters) is produced at every iteration on the unit number specified by **Monitoring File**. Unless stated otherwise, the values of the quantities printed are those in effect *on completion* of the given iteration.

`Itn`	is the iteration count.
`pp`	is the partial price indicator. The variable selected by the last pricing operation came from the `pp`-th partition of A and $-I$. Note that `pp` is reset to zero whenever the basis is refactorized.
`dj`	is the value of the reduced gradient (or reduced cost) for the variable selected by the pricing operation at the start of the current iteration.
`+S`	is the variable selected by the pricing operation to be added to the superbasic set.
`-S`	is the variable chosen to leave the superbasic set.
`-B`	is the variable removed from the basis (if any) to become nonbasic.
`-B`	is the variable chosen to leave the set of basics (if any) in a special basic $\leftrightarrow$ superbasic swap. The entry under `-S` has become basic if this entry is non-zero, and nonbasic otherwise. The swap is done to ensure that there are no superbasic slacks.
`Step`	is the value of the step length α taken along the computed search direction p. The variables x have been changed to $x + \alpha p$. If a variable is made superbasic during the current iteration (i.e., `+S` is positive), `Step` will be the step to the nearest bound. During the optimality phase, the step can be greater than unity only if the reduced Hessian is not positive-definite.

Pivot	is the rth element of a vector y satisfying $By = a_q$ whenever a_q (the qth column of the constraint matrix $(A - I)$) replaces the rth column of the basis matrix B. Wherever possible, Step is chosen so as to avoid extremely small values of Pivot (since they may cause the basis to be nearly singular). In extreme cases, it may be necessary to increase the value of the optional parameter **Pivot Tolerance** (default value $= \epsilon^{0.67}$, where ϵ is the ***machine precision***; see Section 11.2) to exclude very small elements of y from consideration during the computation of Step.
Ninf	is the number of violated constraints (infeasibilities). This will be zero during the optimality phase.
Sinf/Objective	is the value of the current objective function. If x is not feasible, Sinf gives the sum of the magnitudes of constraint violations. If x is feasible, Objective is the value of the objective function. The output line for the final iteration of the feasibility phase (i.e., the first iteration for which Ninf is zero) will give the value of the true objective at the first feasible point. During the optimality phase, the value of the objective function will be non-increasing. During the feasibility phase, the number of constraint infeasibilities will not increase until either a feasible point is found, or the optimality of the multipliers implies that no feasible point exists.
L	is the number of non-zeros in the basis factor L. Immediately after a basis factorization $B = LU$, this is lenL, the number of subdiagonal elements in the columns of a lower triangular matrix. Further non-zeros are added to L when various columns of B are later replaced. (Thus, L increases monotonically.)
U	is the number of non-zeros in the basis factor U. Immediately after a basis factorization, this is lenU, the number of diagonal and superdiagonal elements in the rows of an upper triangular matrix. As columns of B are replaced, the matrix U is maintained explicitly (in sparse form). The value of U may fluctuate up or down; in general, it will tend to increase.
Ncp	is the number of compressions required to recover workspace in the data structure for U. This includes the number of compressions needed during the previous basis factorization. Normally, Ncp should increase very slowly. If it does not, increase LENZ by at least L + U and rerun E04NKF (possibly using START = 'W'; see Section 5).
Norm rg	is $\|d_S\|$, the Euclidean norm of the reduced gradient (see Section 10.3) at the start of the current iteration. During the optimality phase, this norm will be approximately zero after a unit step. For FP and LP problems, Norm rg is not printed.
Ns	is the current number of superbasic variables. For FP and LP problems, Ns is not printed.
Cond Hz	is a lower bound on the condition number of the reduced Hessian (see Section 10.2). The larger this number, the more difficult the problem. For FP and LP problems, Cond Hz is not printed.

When **Print Level** ≥ 20 and **Monitoring File** ≥ 0, the following lines of intermediate printout (< 120 characters) are produced on the unit number specified by **Monitoring File** whenever the matrix B or $B_S = (B \ \ S)^T$ is factorized. Gaussian elimination is used to compute an LU factorization of B or B_S, where PLP^T is a lower triangular matrix and PUQ is an upper triangular matrix for some permutation matrices P and Q. The factorization is stabilized in the manner described under the optional parameter **LU Factor Tolerance** (default value $= 100.0$; see Section 11.2).

Factorize	is the factorization count.

Demand   is a code giving the reason for the present factorization as follows:

Code	Meaning
0	First LU factorization.
1	Number of updates reached the value of the optional parameter **Factorization Frequency** (default value $= 100$; see Section 11.2).
2	Excessive non-zeros in updated factors.
7	Not enough storage to update factors.
10	Row residuals too large (see the description for the optional parameter **Check Frequency** in Section 11.2).
11	Ill-conditioning has caused inconsistent results.

Iteration   is the iteration count.

Nonlinear   is the number of nonlinear variables in B (not printed if B_S is factorized).

Linear   is the number of linear variables in B (not printed if B_S is factorized).

Slacks   is the number of slack variables in B (not printed if B_S is factorized).

Elems   is the number of non-zeros in B (not printed if B_S is factorized).

Density   is the percentage non-zero density of B (not printed if B_S is factorized). More precisely, $\texttt{Density} = 100\times \texttt{Elems}/(\texttt{ Nonlinear } + \texttt{ Linear } + \texttt{ Slacks})^2$.

Compressns   is the number of times the data structure holding the partially factorized matrix needed to be compressed, in order to recover unused workspace. Ideally, it should be zero. If it is more than 3 or 4, increase LENIZ and LENZ and rerun E04NKF (possibly using START = 'W'; see Section 5).

Merit   is the average Markowitz merit count for the elements chosen to be the diagonals of PUQ. Each merit count is defined to be $(c-1)(r-1)$, where c and r are the number of non-zeros in the column and row containing the element at the time it is selected to be the next diagonal. $\texttt{Merit}$ is the average of $\texttt{m}$ such quantities. It gives an indication of how much work was required to preserve sparsity during the factorization.

lenL   is the number of non-zeros in L.

lenU   is the number of non-zeros in U.

Increase   is the percentage increase in the number of non-zeros in L and U relative to the number of non-zeros in B. More precisely, $\texttt{Increase} = 100 \times (\texttt{lenL} + \texttt{lenU} - \texttt{Elems})/\texttt{Elems}$.

m   is the number of rows in the problem. Note that $\texttt{m} = \texttt{Ut} + \texttt{Lt} + \texttt{bp}$.

Ut   is the number of triangular rows of B at the top of U.

d1   is the number of columns remaining when the density of the basis matrix being factorized reached 0.3.

Lmax   is the maximum subdiagonal element in the columns of L (not printed if B_S is factorized). This will not exceed the value of the **LU Factor Tolerance**.

Bmax   is the maximum non-zero element in B (not printed if B_S is factorized).

BSmax   is the maximum non-zero element in B_S (not printed if B is factorized).

Umax   is the maximum non-zero element in U, excluding elements of B that remain in U unchanged. (For example, if a slack variable is in the basis, the corresponding row of B will become a row of U without modification. Elements in such rows will not contribute to $\texttt{Umax}$. If the basis is strictly triangular, *none* of the elements of B will contribute, and $\texttt{Umax}$ will be zero.)

Ideally, $\texttt{Umax}$ should not be significantly larger than $\texttt{Bmax}$. If it is several orders of magnitude larger, it may be advisable to reset the **LU Factor Tolerance** to a value near 1.0.

$\texttt{Umax}$ is not printed if B_S is factorized.

Umin   is the magnitude of the smallest diagonal element of PUQ (not printed if B_S is factorized).

Growth   is the value of the ratio Umax/Bmax, which should not be too large.
                          Providing Lmax is not large (say < 10.0), the ratio max(Bmax, Umax)/Umin is an
                          estimate of the condition number of B. If this number is extremely large, the
                          basis is nearly singular and some numerical difficulties could occur in subsequent
                          computations. (However, an effort is made to avoid near singularity by using
                          slacks to replace columns of B that would have made Umin extremely small, and
                          the modified basis is refactorized.)
                          Growth is not printed if B_S is factorized.
Lt   is the number of triangular columns of B at the beginning of L.
bp   is the size of the 'bump' or block to be factorized nontrivially after the triangular
                          rows and columns have been removed.
d2   is the number of columns remaining when the density of the basis matrix being
                          factorized reached 0.6.

When **Print Level** ≥ 20 and **Monitoring File** ≥ 0, the following lines of intermediate printout (< 80 characters) are produced on the unit number specified by **Monitoring File** whenever START = 'C' (see Section 5). They refer to the number of columns selected by the Crash procedure during each of several passes through A, whilst searching for a triangular basis matrix.

Slacks   is the number of slacks selected initially.
Free cols   is the number of free columns in the basis.
Preferred   is the number of 'preferred' columns in the basis (i.e., ISTATE$(j) = 3$ for some
                          $j \leq n$).
Unit   is the number of unit columns in the basis.
Double   is the number of double columns in the basis.
Triangle   is the number of triangular columns in the basis.
Pad   is the number of slacks used to pad the basis.

When **Print Level** ≥ 20 and **Monitoring File** ≥ 0, the following lines of intermediate printout (< 80 characters) are produced on the unit number specified by **Monitoring File**. They refer to the elements of the NAMES array (see Section 5).

Name   gives the name for the problem (blank if none).
Status   gives the exit status for the problem (i.e., Optimal soln, Weak soln, Unbounded,
                          Infeasible, Excess itns, Error condn or Feasble soln) followed by details of
                          the direction of the optimization (i.e., (Min) or (Max)).
Objective   gives the name of the free row for the problem (blank if none).
RHS   gives the name of the constraint right-hand side for the problem (blank if none).
Ranges   gives the name of the ranges for the problem (blank if none).
Bounds   gives the name of the bounds for the problem (blank if none).

When **Print Level** $= 1$ or ≥ 10 and **Monitoring File** ≥ 0, the following lines of final printout (< 120 characters) are produced on the unit number specified by **Monitoring File**.

Let a_j denote the jth column of A, for $j = 1, 2, \ldots, n$. The following describes the printout for each column (or variable). A full stop (.) is printed for any numerical value that is zero.

Number   is the column number j. (This is used internally to refer to x_j in the intermediate
                          output.)
Column   gives the name of x_j.
State   gives the state of x_j (LL if nonbasic on its lower bound, UL if nonbasic on its upper
                          bound, EQ if nonbasic and fixed, FR if nonbasic and strictly between its bounds,
                          BS if basic and SBS if superbasic).

                          A key is sometimes printed before State to give some additional information about
                          the state of x_j. Note that unless the optional paramter **Scale Option** $= 0$ (default
                          value $= 2$; see Section 11.2) is specified, the tests for assigning a key are applied
                          to the variables of the scaled problem.

A   *Alternative optimum possible.* x_j is nonbasic, but its reduced gradient is essentially zero. This means that if x_j were allowed to start moving away from its bound, there would be no change in the value of the objective function. The values of the basic and superbasic variables *might* change, giving a genuine alternative solution. However, if there are any degenerate variables (labelled D), the actual change might prove to be zero, since one of them could encounter a bound immediately. In either case, the values of the Lagrange multipliers *might* also change.

D   *Degenerate.* x_j is basic or superbasic, but it is equal to (or very close to) one of its bounds.

I   *Infeasible.* x_j is basic or superbasic and is currently violating one of its bounds by more than the value of the optional parameter **Feasibility Tolerance** (default value $= \max(10^{-6}, \sqrt{\epsilon})$, where ϵ is the **machine precision**; see Section 11.2).

N   *Not precisely optimal.* x_j is nonbasic or superbasic. If the value of the reduced gradient for x_j exceeds the value of the optional parameter **Optimality Tolerance** (default value $= \max(10^{-6}, \sqrt{\epsilon})$), the solution would not be declared optimal because the reduced gradient for x_j would not be considered negligible.

`Activity`	is the value of x_j at the final iterate.
`Obj Gradient`	is the value of g_j at the final iterate. For FP problems, g_j is set to zero.
`Lower Bound`	is the lower bound specified for x_j. `None` indicates that $\mathrm{BL}(j) \leq -bigbnd$.
`Upper Bound`	is the upper bound specified for x_j. `None` indicates that $\mathrm{BU}(j) \geq bigbnd$.
`Reduced Gradnt`	is the value of d_j at the final iterate (see Section 10.3). For FP problems, d_j is set to zero.
`m + j`	is the value of $m + j$.

Let v_i denote the ith row of A, for $i = 1, 2, \ldots, m$. The following describes the printout for each row (or constraint). A full stop (.) is printed for any numerical value that is zero.

`Number`	is the value of $n + i$. (This is used internally to refer to s_i in the intermediate output.)
`Row`	gives the name of v_i.
`State`	gives the state of v_i (`LL` if active on its lower bound, `UL` if active on its upper bound, `EQ` if active and fixed, `BS` if inactive when s_i is basic and `SBS` if inactive when s_i is superbasic).

A key is sometimes printed before `State` to give some additional information about the state of s_i. Note that unless the optional parameter **Scale Option** $= 0$ (default value $= 2$; see Section 11.2) is specified, the tests for assigning a key are applied to the variables of the scaled problem.

A   *Alternative optimum possible.* s_i is nonbasic, but its reduced gradient is essentially zero. This means that if s_i were allowed to start moving away from its bound, there would be no change in the value of the objective function. The values of the basic and superbasic variables *might* change, giving a genuine alternative solution. However, if there are any degenerate variables (labelled D), the actual change might prove to be zero, since one of them could encounter a bound immediately. In either case, the values of the dual variables (or Lagrange multipliers) *might* also change.

D   *Degenerate.* s_i is basic or superbasic, but it is equal to (or very close to) one of its bounds.

I   *Infeasible.* s_i is basic or superbasic and is currently violating one of its bounds by more than the value of the optional parameter **Feasibility Tolerance** (default value $= 10^{-6}$; see Section 11.2).

N	*Not precisely optimal.* s_i is nonbasic or superbasic. If the value of the reduced gradient for s_i exceeds the value of the optional parameter **Optimality Tolerance** (default value $= \max(10^{-6}, \sqrt{\epsilon})$), the solution would not be declared optimal because the reduced gradient for s_i would not be considered negligible.

Activity	is the value of v_i at the final iterate.
Slack Activity	is the value by which v_i differs from its nearest bound. (For the free row (if any), it is set to `Activity`.)
Lower Bound	is the lower bound specified for v_i. `None` indicates that $\mathrm{BL}(n+j) \leq -bigbnd$.
Upper Bound	is the upper bound specified for v_i. `None` indicates that $\mathrm{BU}(n+j) \geq bigbnd$.
Dual Activity	is the value of the dual variable π_i (the Lagrange multiplier for v_i; see Section 10.3). For FP problems, π_i is set to zero.
i	gives the index i of v_i.

Numerical values are output with a fixed number of digits; they are not guaranteed to be accurate to this precision.