J Math Chem (2010) 48:3–20
DOI 10.1007/s10910-009-9648-8

ORIGINAL PAPER

Neville elimination: an efficient algorithm with application to chemistry

P. Alonso · J. Delgado · R. Gallego · J. M. Peña

Received: 2 August 2009 / Accepted: 25 November 2009 / Published online: 10 December 2009

Abstract Solving linear systems is often required in chemical problems. Besides, birth and death processes occur in many chemical phenomena and the matrices associated to these processes are totally positive, that is, all their minors are nonnegative. Neville elimination is an elimination procedure very useful when dealing with these matrices. Convergence and stability of iterative refinement using Neville elimination are analyzed, in particular when the coefficient matrix is totally positive. Other applications to chemistry are commented and numerical experiments are shown.

Keywords Iterative refinement · Neville elimination · Convergence · Stability · Total positivity

1 Introduction

Solution of the matrix equation and calculation of the matrix inverse of a square matrix are recurrent tasks handled by molecular modeling software (see [6]). The calculation of the inverse of positive matrices is a common task in Computational Chemistry. Thus, in Quantum Chemistry, inversion of the overlap matrix between basis functions is required to obtain the electronic energy and to perform charge density analyses. Similarly, inversion of the hessian matrix, which stores the second derivatives of molecular energy with respect to nuclear coordinates, is frequently demanded by optimization

P. Alonso (✉) · R. Gallego
Department of Mathematics, University of Oviedo, Campus de Viesques,
33203 Gijón, Spain
e-mail: palonso@uniovi.es

J. Delgado · J. M. Peña
Department of Applied Mathematics, University of Zaragoza,
Pedro Cerbuna 12, 50009 Zaragoza, Spain

techniques and normal mode analyses. Hence, the availability of more efficient direct methods for solving matrix equations could be of particular interest.

Birth and death processes occur in many chemical phenomena, such as in aerosol chemistry (see [20]), in applications of industrial chemistry (see [27]), or in stochastic models of chemical reactions (see [21]). A birth and death process is a stationary Markov process whose state space is the nonnegative integers and whose transition probability matrix $(P_{ij}(t))_{i,j=0,1,2,\cdots}$, $t \geq 0$, with

$$P_{ij}(t) = \Pr\{x(t) = j \mid x(0) = i\}$$

satisfies the conditions (as $t \to 0$) $P_{i,i+1}(t) = \lambda_i t + o(t)$, $P_{i,i-1}(t) = \mu_i t + o(t)$ and $P_{i,i}(t) = 1 - (\lambda_i + \mu_i)t + o(t)$, where $\lambda_i > 0$ for all $i \geq 0$, $\mu_i > 0$ for all $i \geq 1$ and $\mu_0 \geq 0$. By the results of [17] (see also [18]), the matrices associated to these processes are totally positive (TP), that is, all their minors are nonnegative. Moreover they are strictly totally positive (i.e., all their minors are positive) for all $t > 0$. In fact, in [18] it was proved that, for all $i_1 < i_2 < \cdots < i_n$ and $j_1 < j_2 < \cdots < j_n$, the determinants

$$\det \begin{pmatrix} P_{i_1,j_1}(t) & \cdots & P_{i_1,j_n}(t) \\ \vdots & & \vdots \\ P_{i_n,j_1}(t) & \cdots & P_{i_n,j_n}(t) \end{pmatrix} \tag{1}$$

have the following interpretation. Suppose that n labelled particles start out in states $i_1, \ldots, i_n$ and execute the process simultaneously and independently. Then the determinant (1) is equal to the probability that at time t the particles will be found in states $j_1, \ldots, j_n$, respectively, without any two of them ever having been coincident (in the same state) in the intervening time. See also the related papers [16,19].

On the other hand, knowledge of the behaviour exhibited by matter under extreme conditions of pressure and temperature is of capital interest in areas such as materials science, geophysics or solid state chemistry [23]. We show at the end of the paper that NE is adequate for solving linear systems arising in these fields.

Neville elimination (NE) is an alternative procedure to Gaussian elimination (GE) to transform a square matrix A into an upper triangular matrix U. Roughly speaking, NE makes zeros in a column of the matrix A by adding to each row a multiple of the previous one. It has been proved that this elimination process is very useful with totally positive matrices (see [7,10,12,14,26]).

Iterative refinement (IR) improves the accuracy of the computed solution of a linear system $Ax = b$. IR using GE has been deeply studied from several points of view: convergence (see [9,25,31]), stability (see [15,28]) and error analysis (see [22,31]). In contrast, IR using NE only has been considered very recently in [2], where only the convergence has been analyzed. In this paper, we analyze the stability of IR when using NE and we also improve the convergence conditions of [2].

In Sect. 2 we recall basic notations related to NE and IR. In Sect. 3 we apply factorization and backward error results on NE to the convergence of IR with NE, simplifying and improving the conditions obtained in [2]. Section 4 is devoted to the analysis of the stability of IR with NE. In Sect. 5 we show that the results of previous

sections can be improved when the coefficient matrix is totally positive. Finally, in Sect. 6 an application of NE to a chemical problem is illustrated.

2 Neville elimination and iterative refinement

In this section, we will describe basic notations related to NE and IR.

2.1 Neville elimination

For a nonsingular matrix A of order n the *Neville elimination* procedure consists of $n-1$ successive steps, resulting in a sequence of matrices as follows:

$$A = A^{(1)} \to \tilde{A}^{(1)} \to A^{(2)} \to \tilde{A}^{(2)} \to \cdots \to A^{(n)} = \tilde{A}^{(n)} = U,$$

where U is an upper triangular matrix. For each t, $1 \leq t \leq n$, the matrices $A^{(t)} = (a_{ij}^{(t)})_{1\leq i,j\leq n}$ and $\tilde{A}^{(t)} = (\tilde{a}_{ij}^{(t)})_{1\leq i,j\leq n}$ have zeros below their main diagonal in the first $t-1$ columns and also one has

$$\tilde{a}_{it}^{(t)} = 0, \quad i \geq t \Rightarrow \tilde{a}_{ht}^{(t)} = 0 \quad \forall h \geq i. \tag{2}$$

$\tilde{A}^{(t)}$ is obtained from the matrix $A^{(t)}$ by moving to the bottom the rows with a zero entry in the column t, if necessary, in order to get (2). The rows which are moved are placed in the same relative order they have in $A^{(t)}$. To get $A^{(t+1)}$ from $\tilde{A}^{(t)}$ we produce zeros in the column t below the main diagonal by subtracting a multiple of the ith row to the $(i+1)$th for $i = n-1, n-2, \ldots, t$, according to the following formula. For any column j

$$a_{ij}^{(t+1)} = \begin{cases} \widetilde{a}_{ij}^{(t)} & \text{if } 1 \leq i \leq t \\ \widetilde{a}_{ij}^{(t)} - \dfrac{\widetilde{a}_{it}^{(t)}}{\widetilde{a}_{i-1,t}^{(t)}} \widetilde{a}_{i-1,j}^{(t)} & \text{if } t+1 \leq i \leq n \text{ and } \widetilde{a}_{i-1,t}^{(t)} \neq 0 \\ \widetilde{a}_{ij}^{(t)} & \text{if } t+1 \leq i \leq n \text{ and } \widetilde{a}_{i-1,t}^{(t)} = 0. \end{cases}$$

Observe that in the third case $\tilde{a}_{i-1,t}^{(t)} = 0$ implies $\tilde{a}_{it}^{(t)} = 0$ by (2). In this process one has $A^{(n)} = \tilde{A}^{(n)} = U$, and when no row exchanges are needed, then $A^{(t)} = \tilde{A}^{(t)}$ for all t.

The number

$$m_{ij} = \begin{cases} \dfrac{\widetilde{a}_{ij}^{(j)}}{\widetilde{a}_{i-1,j}^{(j)}} & \text{if } \widetilde{a}_{i-1,j}^{(j)} \neq 0 \\ 0 & \text{if } \widetilde{a}_{i-1,j}^{(j)} = 0 \end{cases}$$

is called the (i, j) *multiplier*.

Matrices satisfying that NE can be performed without changes of rows will be referred to as *matrices satisfying* the WR *condition*. Alternatively, for each step t of the NE of A, we can reorder the rows $t, \ldots, n$ of the matrix $A^{(t)}$ according to a row pivoting strategy in order to get $\tilde{A}^{(t)}$ satisfying (2).

2.2 Iterative refinement

In this subsection, we recall the IR technique and apply it to NE. Let us assume that the system $Ax = b$ has been solved using NE with floating point arithmetic, obtaining an approximate solution $\hat{x} = x^{(0)}$. If the matrix A is ill conditioned, the computed solution $x^{(0)}$ might not be accurate enough. When $x^{(0)}$ has been computed, we will obtain the corresponding residual, which is defined as

$$r^{(0)} = A(x - x^{(0)}) = Ax - Ax^{(0)} = b - Ax^{(0)}.$$

It is convenient to compute this vector in extra precision in order to improve the accuracy of $x^{(0)}$. Observe that Skeel proves in his works (see [28,29]) that in the Gaussian case it is not always necessary to compute the residual in extra precision to obtain a stable procedure, because stability might be achieved in working precision.

Next we will solve the linear system $Ae = r^{(0)}$ and for this purpose we will use the matrices $\hat{L}$ and $\hat{U}$ obtained in the NE process (see (16) in Theorem 1). Then the system to be solved will be $\hat{L}\hat{U}e = r^{(0)}$ that is equivalent to $(A + E)e = r^{(0)}$. The solution obtained will not be exactly e, but an approximation, which will be denoted by $e^{(0)}$, and then $x^{(1)} = x^{(0)} + e^{(0)}$ will be a new approximation to the exact solution.

This procedure is commonly known IR and it was first presented by Wilkinson in [30,31]. In other works, for example in, [9,15,22,25,29], several features of IR have been studied.

In a general way, we will define the vectors $r^{(k)}$ and $x^{(k+1)}$ through the following equations

$$r^{(k)} = b - Ax^{(k)}, \; x^{(k+1)} = x^{(k)} + e^{(k)}, \tag{3}$$

where $e^{(k)}$ is the approximate solution obtained when solving the system $\left(\hat{L}\hat{U}\right) e = r^{(k)}$, namely $e^{(k)} = (\hat{L}\hat{U})^{-1} r^{(k)}$.

Bringing the expression of $e^{(k)}$ to (3) we have that

$$x^{(k+1)} = x^{(k)} + \left(\hat{L}\hat{U}\right)^{-1} r^{(k)}.$$

If the vector sequence $\{x^{(k)}\}$ can be formed without additional roundoff errors, we deduce that

$$x^{(k+1)} = x^{(k)} + \left(\hat{L}\hat{U}\right)^{-1} \left(b - Ax^{(k)}\right) = x^{(k)} + \left(\hat{L}\hat{U}\right)^{-1} A \left(x - x^{(k)}\right),$$

and so

$$x^{(k+1)} - x = \left(I - \left(\hat{L}\hat{U}\right)^{-1} A\right)\left(x^{(k)} - x\right),$$

and iterating the process

$$x^{(k+1)} - x = \left(I - \left(\hat{L}\hat{U}\right)^{-1} A\right)^{k}\left(x^{(1)} - x\right). \tag{4}$$

In the same way, we have that

$$r^{(k+1)} = b - Ax^{(k+1)} = Ax - Ax^{(k+1)} = A\left(x - x^{(k+1)}\right)$$

and by (4)

$$r^{(k+1)} = A\left(I - \left(\hat{L}\hat{U}\right)^{-1} A\right)^{k}\left(x - x^{(1)}\right).$$

Considering (4) we can state that

$$\|x^{(k+1)} - x\| \leq \|I - \left(\hat{L}\hat{U}\right)^{-1} A\|^{k}\|x^{(1)} - x\|,$$

so for $x^{(k+1)}$ to converge towards x, or in other words, for

$$\lim_{k\longrightarrow\infty} \|x^{(k+1)} - x\| = 0$$

it is enough that the following inequality holds:

$$\|I - \left(\hat{L}\hat{U}\right)^{-1} A\| < 1. \tag{5}$$

Let us see how we can interpret the condition (5) and with this purpose let us assume that $\hat{L}\hat{U} = A + E$. Hence

$$I - \left(\hat{L}\hat{U}\right)^{-1} A = I - (A + E)^{-1}A. \tag{6}$$

On the other hand, it is clear that

$$I - (A + E)^{-1}A = I - (I + A^{-1}E)^{-1}. \tag{7}$$

Finally, it is well known that, if $\|A^{-1}E\| < 1$, then $(I + A^{-1}E)$ is nonsingular and it holds

$$\|I - (I + A^{-1}E)^{-1}\| \leq \frac{\|A^{-1}\|\|E\|}{1 - \|A^{-1}\|\|E\|}. \tag{8}$$

Using (5) and (8) we will obtain in the following sections sufficient conditions for the convergence of IR with NE.

3 Matrix factorization and iterative refinement with Neville elimination

Let us present some results concerned with matrix factorizations associated to NE. As in [12] we denote by $E_i(\alpha)$ the bidiagonal lower triangular matrix whose (r, s) entry $(1 \leq r, s \leq n)$ is given by 1 if $r = s$, α if $(r, s) = (i, i-1)$ and 0 elsewhere, therefore

$$E_i(\alpha) = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & \alpha & 1 & & & \\ & & & & \ddots & & \\ & & & & & 1 & \end{pmatrix}.$$

In Theorem 2.5 of [12] it is proved that a nonsingular $n \times n$ matrix A satisfies the WR condition if and only if it can be factorized in the form

$$A = [E_n(m_{n1}) \dots E_2(m_{21})][E_n(m_{n2}) \dots E_3(m_{32})] \dots \dots E_n(m_{n,n-1})U \tag{9}$$

with the m_{ij}'s satisfying

$$m_{ij} = 0 \implies m_{hj} = 0, \ \forall h > i; \ 1 \leq j \leq n-1, \ i \geq j+1. \tag{10}$$

If A satisfies this condition the factorization is unique and m_{ij} is the (i, j) multiplier of the NE of A. Observe that (9) is, obviously, the same factorization LU in GE, with the matrix L decomposed in a different way, as a product of elementary bidiagonal matrices.

The factorization (9) can be reordered in several different forms. In particular, we easily get

$$A = E_n(m_{n1})[E_{n-1}(m_{n-1,1})E_n(m_{n2})] \dots [E_2(m_{21})E_3(m_{32}) \dots E_n(m_{n,n-1})]U$$

that can also be written in the form

$$[E_n(-m_{n,n-1}) \dots E_2(-m_{21})] \dots [E_n(-m_{n2})E_{n-1}(-m_{n-1,1})]E_n(-m_{n1})A = U. \tag{11}$$

This means that NE of A can be done by *subdiagonals*, that is, making zero first the $(n, 1)$ entry, then the $(n-1, 1)$ and the $(n, 2)$ ones, and so on, finishing making zeros the $(2, 1)$, $(3, 2)$, $\dots$, $(n, n-1)$ entries.

We shall denote $F_i = E_{n-(i-1)}(m_{n-(i-1),1})E_{n-(i-2)}(m_{n-(i-2),2})\cdots E_n(m_{ni})$, that is,

$$F_i = \begin{pmatrix} 1 & & & & & & & & \\ 0 & 1 & & & & & & & \\ & \ddots & \ddots & & & & & & \\ & & 0 & 1 & & & & & \\ & & & m_{n-(i-1),1} & 1 & & & & \\ & & & & m_{n-(i-2),2} & 1 & & & \\ & & & & & \ddots & \ddots & & \\ & & & & & & m_{ni} & 1 \end{pmatrix} \tag{12}$$

and so the NE process of a matrix A satisfying the WR condition, performed as in (11), can be described in following form

$$\begin{aligned} &A^{[1]} := A \\ &F_{t-1}A^{[t]} := A^{[t-1]}, \quad 2 \le t \le n-1. \end{aligned} \tag{13}$$

Here

$$A^{[t]} = \begin{pmatrix} a_{11}^{[t]} & a_{12}^{[t]} & \cdots\cdots\cdots & a_{1n}^{[t]} \\ \vdots & \vdots & & \vdots \\ a_{n-(t-1),1}^{[t]} & a_{n-(t-1),2}^{[t]} & & a_{n-(t-1),n}^{[t]} \\ 0 & a_{n-(t-2),2}^{[t]} & & a_{n-(t-2),n}^{[t]} \\ 0 & 0 & \ddots & \vdots \\ \vdots & & \ddots\ \ddots & \vdots \\ 0 & \cdots & 0\ a_{nt}^{[t]}\ \cdots & a_{nn}^{[t]} \end{pmatrix} \tag{14}$$

with

$$\begin{aligned} a_{ij}^{[t]} &:= a_{ij}^{[t-1]} && 1 \le j \le n,\ 1 \le i \le n-t-1 \\ a_{ij}^{[t]} &:= a_{ij}^{[t-1]} - m_{i,i-n+t}a_{i-1,j}^{[t-1]} && n-t+2 \le i \le n,\ i-n+t \le j \le n. \end{aligned} \tag{15}$$

Remember that this elimination process is just a reordering of the one explained in Sect. 2, with the same multipliers. More precisely, for each t the nth row of $A^{[t]}$ is the same as that of $A^{(t)}$, the $(n-1)$th row of $A^{[t]}$ is the same as the $(n-1)$th row of $A^{(t-1)}$, and so on until the $(n-t+1)$th row of $A^{[t]}$, which is the $(n-t+1)$th row of $A^{(1)}$. All the other precedent rows $1, 2, \ldots, n-t$ of $A^{[t]}$ are also the same rows as the correspondent ones of $A^{(1)}$. Obviously one has $A^{[n]} = A^{(n)} = U$.

Finite precision arithmetic produces a sequence of matrices $\hat{A}^{(t)}$ in the NE of A by columns which is different from the sequence $A^{(t)}$ obtained by exact arithmetic,

as studied in [11]. If the NE of A is performed by subdiagonals we get the sequence $\hat{A}^{[t]}$ whose relationship with the sequence $\hat{A}^{(t)}$ is the same as that of $A^{[t]}$ and $A^{(t)}$ explained before.

Remark that we are interested on matrices A satisfying the WR condition for sufficiently high precision, that is, for a unit roundoff u sufficiently small, the NE of A can be computed without row exchanges. For brevity, this condition will be referred to as WR$^+$ condition. We always assume that we are working under this condition. In Sect. 5 we show a class of matrices satisfying this condition.

In [1], Alonso et al. derived the backward error analysis of NE, according to the next result:

Theorem 1 *Let A be a nonsingular $n \times n$ matrix and assume that NE of A has been computed under the WR$^+$ condition, with unit roundoff u. Then the matrices $\hat{L} := \hat{F}_1\hat{F}_2 \ldots \hat{F}_{n-1}$, $\hat{U} := \hat{A}^{[n]}$ verify*

$$\hat{L}\hat{U} = A + E \tag{16}$$

with

$$|E| \leq u \sum_{j=1}^{n-1} |\hat{F}_1||\hat{F}_2| \ldots |\hat{F}_j||\hat{A}^{[j+1]}|. \tag{17}$$

If $\hat{L}$, $\hat{U}$ are used to compute the solution $\hat{x}$ of the system $Ax = b$ by forward and backward substitution one has

$$(A + H)\hat{x} = b \tag{18}$$

with

$$|H| \leq u \sum_{j=1}^{n-1} |\hat{F}_1||\hat{F}_2| \ldots |\hat{F}_j||\hat{A}^{[j+1]}| + \gamma_n|\hat{L}||\hat{U}| \tag{19}$$

where $\gamma_n := (nu)/(1 - nu)$.

We point out that the general condition about the convergence of IR with NE proved in [2] involves the perturbation matrix H. In this paper, we give a convergence condition involving the matrix E, instead of H.

Theorem 2 *Let A be a nonsingular $n \times n$ matrix and assume that $\widehat{L} := \widehat{F}_1\widehat{F}_2 \cdots \widehat{F}_{n-1}$ and $\widehat{U} := \widehat{A}^{[n]}$ have been computed by NE of A under the WR$^+$ condition, with unit roundoff u. If $\widehat{L}$ and $\widehat{U}$ are used to compute the solution $\widehat{x}$ of the system $Ax = b$, then one has that if*

$$\|A^{-1}\|\|E\| < \frac{1}{2} \tag{20}$$

then the IR converges to $x = A^{-1}b$.

Proof As we have pointed out, the procedure converges whenever (5) holds. By (7) and (8)

$$\|I - (A+E)^{-1}A\| \leq \frac{\|A^{-1}\|\|E\|}{1 - \|A^{-1}\|\|E\|},$$

so the condition $\|I - \left(\hat{L}\hat{U}\right)^{-1} A\| < 1$ holds, by (6), if

$$\|A^{-1}\|\|E\| < \frac{1}{2}.$$

The previous result will be used in Sect. 5 to improve and refine the convergence condition given in [2] in the case that A is a totally positive matrix.

In the general case and taking into account (17), the sufficient condition for convergence shown in the previous Theorem can be stated as follows.

Corollary 1 *Let A be a nonsingular $n \times n$ matrix and assume that $\widehat{L}$ and $\widehat{U}$ have been computed by NE of A under the WR^+ condition, with unit roundoff u. If $\widehat{L}$ and $\widehat{U}$ are used to compute the solution $\widehat{x}$ of the system $Ax = b$, then one has that if*

$$\left(u \sum_{j=1}^{n-1} \|\hat{F}_1\|\|\hat{F}_2\| \ldots \|\hat{F}_j\| \|\hat{A}^{[j+1]}\| \right) \|A^{-1}\| < \frac{1}{2} \tag{21}$$

then the IR converges to $x = A^{-1}b$.

We want to stress the fact that the condition (21) is a refinement of the result obtained in Theorem 3.4 of [2] since on the left hand side of (21) the term $\gamma_n \|\hat{L}\|\|\hat{U}\|\|A^{-1}\|$ does not appear.

4 Stability of the iterative refinement with Neville elimination

In this section, we study the stability of the IR when applied to the NE procedure. First, let us define backward error.

Definition 1 Given a linear system of equations $Ax = b$, the componentwise backward error of an approximate solution $\hat{x}$ is defined as

$$\omega_{E,f}(\hat{x}) = \min\{\epsilon : \ (A + \delta A)\hat{x} = b + \delta b, \ |\delta A| \leq \epsilon E, \ |\delta b| \leq \epsilon f\}$$

where the matrix E and the vector f have nonnegative entries.

We will adopt the usual choice for the tolerances E and f: $E = |A|$, $f = |b|$. With this choice, a small backward error $\omega_{|A|,|b|}$ means that the approximate solution $\hat{x}$ is the solution of a slightly perturbed system, in the sense of componentwise relative perturbation.

The componentwise backward error can be computed with a simple formula (see [24] and §7.2 in [15]) as given in the next result.

Theorem 3 *The componentwise backward error is given by*

$$\omega_{|A|,|b|}(\hat{x}) = \max_i \frac{|r_i|}{(|A||\hat{x}| + |b|)_i}$$

where $r = b - A\hat{x}$. *The quotient* $\alpha/0$ *is interpreted as zero if* $\alpha = 0$ *and infinity otherwise.*

An algorithm to solve a linear system of equations is said to be componentwise backward stable if the componentwise backward error is of order the unit roundoff provided that this is less than a certain threshold, that is to say

$$\omega_{|A|,|b|}(\hat{x}) = \mathcal{O}(u) \quad \text{for } u \leq \bar{u}(n).$$

Nevertheless, for practical purposes, the threshold is also allowed to depend upon the data A and b, so that $\bar{u} = \bar{u}(n, A, b)$. Then, componentwise backward stability means that the approximate solution $\hat{x}$ is the exact solution of a perturbed system where the perturbations are small componentwise.

In [29], Skeel studies the stability of single precision IR applied to GE. He proved that if $x^{(m)}$, $m = 0, 1, \ldots$, is the sequence of vectors obtained through IR, then

$$\omega_{|A|,|b|}\left(x^{(1)}\right) \leq (n+1)u + \mathcal{O}(u^2) \tag{22}$$

as long as

$$cu\,\mathrm{cond}(A^{-1})\sigma(A, x) \leq 1/2. \tag{23}$$

The quantity c is a constant bounded above by functions depending only on n, and

$$\mathrm{cond}(A^{-1}) = \||A||A^{-1}|\|, \quad \sigma(A, x) := \max \frac{e\||A||x|\|}{|A||x|}, \quad e = (1, 1, \ldots, 1)^t.$$

Therefore, one step of single precision IR makes GE stable. This result was very important because it shattered the idea that it was necessary to compute the residual in extra precision in order that the IR can be beneficial.

Remark 1 The inequality (23) determines the stability threshold, or in other words the conditions (with respect to the accuracy and the data A, b) under which the algorithm is backward stable. It holds as long as the unit roundoff is small enough or, equivalently, as long as the accuracy is high enough. It is important to notice that this inequality is very difficult to check in practice since the quantities involved are rather complex expressions. The quantity $\sigma(A, x)$ is large for ill scaled systems and so is $\mathrm{cond}(A^{-1})$ for ill conditioned systems. Hence, for ill scaled or ill conditioned systems, the inequality might not hold.

In [15] some general results about IR which do not depend on the solver used are shown. In particular we need to mention the following result:

Theorem 4 *Let A be a nonsingular $n \times n$ matrix and assume that the linear system $Ax = b$ is solved in floating point arithmetic using a solver S together with one step of fixed precision IR. Assume that the residual of the computed solution $\hat{x}$ satisfies*

$$|b - A\hat{x}| \leq u(G|A||\hat{x}| + M|b|), \tag{24}$$

where G and M are $n \times n$ matrices with nonnegative entries. We also assume that the residual is computed in the conventional way so that it satisfies:

$$|\hat{r} - r| \leq \gamma_{n+1}(|A||\hat{x}| + |b|).$$

Then, there is a function

$$f(\alpha, \beta) \approx (n+1)^{-1}\left(\frac{\beta(\alpha + n + 1)}{\mathrm{cond}(A^{-1})} + 2(\alpha + n + 2)^2(1 + u\beta)^2\right)$$

such that if

$$\mathrm{cond}(A^{-1})\sigma_R(A, x^{(1)}) \leq (f(\|G\|_\infty, \|M\|_\infty)u)^{-1} \tag{25}$$

then

$$|b - Ax^{(1)}| \leq 2\gamma_{n+1}|A||x^{(1)}|. \tag{26}$$

The quantity $\sigma_R(C, x)$ is a measure of ill scaling of the vector $|C||x|$ and it was first introduced by Skeel [28]:

$$\sigma_R(C, x) := \frac{\max_i(|C||x|)_i}{\min_i(|C||x|)_i}.$$

Formula (26) implies backward stability according to Theorem 3 as

$$\omega_{|A|,|b|}\left(x^{(1)}\right) \leq 2\gamma_{n+1} = 2(n+1)u + \mathcal{O}(u^2).$$

The previous result is equivalent to the Skeel's given in (22), but this one is more general since it does not depend on the particular solver used.

Let us now prove that one step of IR with NE is backward stable for any linear system. We can show that for NE the residual satisfies a bound of the form (24) for any matrix satisfying the WR^+ condition, so we can find a sufficient condition for the stability of IR with NE by using Theorem 4. Next we bound the residual of an approximate solution of a linear system obtained through NE.

In the Gaussian case, and with the purpose of applying Theorem 4, Higham considers in pages 238 and 239 of [15] that $A \approx \hat{L}\hat{U}$, and then he obtains that

$$|\hat{L}||\hat{U}| \approx |\hat{L}||\hat{L}^{-1}A| \leq |\hat{L}||\hat{L}^{-1}||A|. \tag{27}$$

With regard to NE, we will also take this same approach and, analogously, we will also consider that $\hat{F}_j \hat{A}^{[j+1]} \approx A^{[j]}$, so

$$\hat{A}^{[j+1]} \approx \hat{F}_j^{-1}\hat{F}_{j-1}^{-1}\ldots\hat{F}_1^{-1}A,$$

and therefore

$$|\hat{A}^{[j+1]}| \approx |\hat{F}_j^{-1}\hat{F}_{j-1}^{-1}\ldots\hat{F}_1^{-1}A| \leq |\hat{F}_j^{-1}||\hat{F}_{j-1}^{-1}|\ldots|\hat{F}_1^{-1}||A|. \tag{28}$$

Under these hypothesis, we can state that if $\hat{x}$ is the computed solution of the system $Ax = b$ with NE under the WR^+ condition with floating point arithmetic and unit roundoff u, then the following inequality holds:

$$|b - A\hat{x}| \leq uG|A||\hat{x}|, \tag{29}$$

where

$$G \approx \sum_{j=1}^{n-1} |\hat{F}_1||\hat{F}_2|\ldots|\hat{F}_j||\hat{F}_j^{-1}|\ldots|\hat{F}_2^{-1}||\hat{F}_1^{-1}| + u^{-1}\gamma_n|\hat{L}||\hat{L}^{-1}|.$$

The previous result is obtained from the fact that the computed solution is the exact solution of a perturbed system, and then from (18) we have that $|b - A\hat{x}| \leq |H||\hat{x}|$. Moreover the matrix H can be bounded according to (19).

Now, bringing (28) and (27) into (19), we obtain (29).

In the previous conditions the residual satisfies a bound of the form (24) with $M = 0$ so that we can apply Theorem 4 to conclude that one step of IR with NE is backward stable.

The stability threshold adopts a simpler expression than that of (25) when dealing with totally positive matrices as shown in the next section.

Remark 2 Although Higham does not prove explicitly the approximation considered in (27), we can check the validity of the approximation we have used, $\hat{F}_j \hat{A}^{[j+1]} \approx A^{[j]}$, by using a linear approximation and a unit roundoff small enough.

In formula (3.14) of [1] it has been proved that the intermediate matrices appearing in the NE process by subdiagonals verify

$$\begin{cases} \hat{F}_j \hat{A}^{[j+1]} = \hat{A}^{[j]} + E_j \\ \\ |E_j| \leq u|\hat{F}_j||\hat{A}^{[j+1]}|. \end{cases} \quad k = 1, \ldots, n-1 \tag{30}$$

Taking into account the previous formula, we get

$$|\hat{F}_j \hat{A}^{[j+1]} - \hat{A}^{[j]}| = |E_j| \leq u|\hat{F}_j||\hat{A}^{[j+1]}| = \mathcal{O}(u),$$

and this completes the argument. On the other hand, just taking into account that $\hat{F}_j \hat{A}^{[j+1]}$ is a linear approximation of $A^{[j]}$, we deduce that $\hat{A}^{[j+1]} \approx \hat{F}_j^{-1} A^{[j]}$. By applying the procedure recursively we get

$$\hat{A}^{[j+1]} \approx \hat{F}_j^{-1} \hat{F}_{j-1}^{-1} \dots \hat{F}_1^{-1} A,$$

and then the bound (28) is obtained.

Finally, it is convenient to observe that the approximation considered by Higham and the extension to the Neville case can be justified following a process similar to the previous one.

5 Totally positive matrices

It has been shown that NE process is very useful with TP matrices. TP matrices present many applications in Statistics, Approximation Theory, Computer Aided Geometric Design (CAGD), Economics and, as recalled in the introduction, in Chemistry.

A class of matrices which satisfies the WR condition is that of nonsingular TP matrices (see Theorem 4.1 of [12]) which include that of *strictly totally positive* (STP) matrices. The multipliers m_{ij} of TP matrices are nonnegative and satisfy (10). Those of STP matrices are all positive. Consequently, the matrices F_i of (13) are nonnegative when A is a nonsingular TP matrix:

$$F_t \geq 0, \ \forall t.$$

Observe that if A is a nonsingular TP matrix, the upper triangular matrix U obtained by NE or GE is TP (see for example Corollary 3.14 of [3]) and, therefore, nonnegative. From (14) and (15), with $m_{i,i-n+t} \geq 0$, we get

$$A = A^{[1]} \geq A^{[2]} \geq \cdots \geq A^{[n-1]} \geq A^{[n]} = U$$

and consequently all the matrices $A^{[t]}$ are nonnegative. In fact, these matrices are TP.

A nonsingular TP matrix A is said to be *almost strictly totally positive* (ASTP) if it satisfies the following condition: each minor $\det(C)$ of A es positive if and only if all the diagonal elements of C are positive. These matrices form an important class of TP matrices (see [13]) which includes that of STP matrices. For nonsingular ASTP matrices we have (see Proposition 3.5 of [11]) that for sufficiently high finite precision arithmetic the NE of A can be carried out without row exchanges (that is, ASTP matrices are WR^+ condition) and the matrices $\hat{A}^{(t)}$ $(t = 1, 2, \dots, n)$ are nonnegative.

Due to the strong relationship between $\hat{A}^{[t]}$ and $\hat{A}^{(t)}$ we have explained before, Proposition 3.5 of [11] holds with $\hat{A}^{(t)}$ replaced by $\hat{A}^{[t]}$. The multipliers $\hat{m}_{ij}$ which

appear along the process will be nonnegative and therefore the matrices $\hat{F}_t$ will be nonnegative. In summary, for a nonsingular ASTP matrix A one has

$$\hat{A}^{[t]} \geq 0, \quad \hat{F}_t \geq 0, \quad \forall t. \tag{31}$$

The next result proved in [1] provide upper bounds for the matrices E y H.

Theorem 5 *Let A be an $n \times n$ ($n \geq 2$) nonsingular TP matrix such that the computed matrices in the NE process satisfy (31). For a unit roundoff u sufficiently small the following properties hold:*

1. *The matrices $\hat{L}$ and $\hat{U}$ computed by NE satisfy $\hat{L}\hat{U} = A + E$, with*

$$|E| \leq \gamma_{n-1} A. \tag{32}$$

2. *The solution of $Ax = b$ computed by NE, $\hat{x}$, satisfies $(A + H)\hat{x} = b$ where*

$$|H| \leq [\gamma_{n-1} + \gamma_n(1 + \gamma_{n-1})]A. \tag{33}$$

Considering (32) and taking into account Theorem 1 we can obtain a condition ensuring the convergence of IR when dealing with ASTP matrices.

If A is a TP matrix satisfying (31) we have obtained in Theorem 5 a bound of the error matrix E that allows us to refine and improve the convergence condition of the general case, thus obtaining a convergence condition depending on the condition number of A.

Taking norms in (32) we have that

$$\|E\| \leq \gamma_{n-1}\|A\|,$$

and using (20) we find a sufficient condition for the convergence of IR.

Corollary 2 *Let A be an $n \times n$ ($n \geq 2$) nonsingular TP matrix and assume that $\widehat{L}$ and $\widehat{U}$ have been computed by NE of A under the WR$^+$ condition and satisfying (31). If $\widehat{L}$ and $\widehat{U}$ are used to compute the solution $\widehat{x}$ of the system $Ax = b$, then one has, for a unit roundoff u sufficiently small, that if*

$$\gamma_{n-1}\|A^{-1}\|\|A\| \leq \frac{1}{2} \tag{34}$$

then the IR converges to $x = A^{-1}b$.

We point out that the condition (34) is a refinement of the result shown in Theorem 3.7 of [2]. Furthermore, it has been obtained in a more direct way and also the constraint $\gamma_n \leq \frac{1}{4}$ has been removed.

With regard to the stability of IR for TP matrices satisfying the WR$^+$ condition, we can find a simple condition that ensures that a single step of IR with NE is backward stable.

Theorem 6 *Let A be a $n \times n$ nonsingular TP matrix and assume that NE of A has been computed under the WR^+ condition, with unit roundoff u. Then it holds that*

$$|b - Ax^{(1)}| \leq 2\gamma_{n+1}A|x^{(1)}|$$

as long as

$$cond(A^{-1})\sigma_R(A, x^{(1)}) \leq \frac{n+1}{2u\left(n+2+\frac{9\gamma_n}{4u}\right)^2}.$$

Proof The residual satisfies $|b-A\hat{x}| \leq |H||\hat{x}|$ where the matrix H is bounded according to (33). Then

$$|b - A\hat{x}| \leq [\gamma_{n-1} + \gamma_n(1+\gamma_{n-1})]A|\hat{x}|.$$

If the accuracy is large enough for $\gamma_n \leq 1/4$ (as done by de Boor and Pinkus in [5]), we have that $\gamma_{n-1} < \gamma_n$, so

$$\gamma_{n-1} + \gamma_n(1+\gamma_{n-1}) \leq \gamma_n + \gamma_n(1+\gamma_n) = \gamma_n(\gamma_n+2) \leq \frac{9}{4}\gamma_n.$$

Therefore, we have that

$$|b - A\hat{x}| \leq \frac{9}{4}\gamma_n A|\hat{x}|.$$

This bound is of the same form as that of (24) with $G = 9\gamma_n I/(4u)$ and $M = 0$, and so in this particular case the value of the function f on the right hand side of (25) can be taken as

$$f(\|G\|_\infty, \|M\|_\infty) = f\left(\frac{9\gamma_n}{4u}, 0\right) = \frac{2\left(n+2+\frac{9\gamma_n}{4u}\right)^2}{n+1}.$$

Then, by Theorem 4, we have that

$$|b - Ax^{(1)}| \leq 2\gamma_{n+1}A|x^{(1)}|$$

as long as

$$\text{cond}(A^{-1})\sigma_R(A, x^{(1)}) \leq \frac{n+1}{2u\left(n+2+\frac{9\gamma_n}{4u}\right)^2}.$$

Taking into account the results of this section, we can assert that using NE when IR is applied to TP matrices is very adequate.

Table 1 Numerical test

Degree	Method	$\|\widehat{x} - x\|_\infty / \|x\|_\infty$	$\omega_{\|A\|,\|b\|}$
6	GE	3.904661×10^{-6}	1.288651×10^{-16}
	GEPP	1.333393×10^{-6}	9.083658×10^{-17}
	NE	7.740051×10^{-7}	1.104792×10^{-16}
7	GE	4.083333×10^{-4}	0.00000
	GEPP	1.614084×10^{-4}	6.775837×10^{-17}
	NE	9.011237×10^{-5}	1.286912×10^{-16}

6 Numerical experiments

Knowledge of the behaviour exhibited by matter under extreme conditions of pressure and temperature is of capital interest in areas such as materials science, geophysics or solid state chemistry [23]. The key expression to describe this behaviour is a relationship linking pressure, temperature and volume for each given material in an analytical function called the equation of state (EOS). Many efforts to represent either experimental or simulated data in a convenient analytical EOS have traditionally been paid [8]. Fitting procedures need to resort to solving linear systems of equations in order to minimize the deviations of the proposed EOS with respect to the set of measured or calculated data. Numerical procedures for this type of fittings have to face in many instances with problems originated from the different weights that some of the points may have. Repetitions of fittings have to be undertaken before a stable final answer is obtained. In this respect, fast and effective algorithms are therefore desirable.

NE has proved to be an efficient algorithm to carry out the least-square fitting mentioned in the previous paragraph, at least for polynomials of degree not too high.

We have performed numerical experiments using data considered in Sect. 4.2 of [4], and we have obtained different approximations by least square fitting taking polynomials of degree ranging from 4 to 8. We have compared the approximate solutions obtained with NE with those obtained with GE and with *Mathematica*. In the numerical tests we perform all computations in IEEE double precision.

In Table 1, we show the results obtained for polynomials of degree 6 and 7. As it can be seen, the results for the Neville method improve those of the Gauss method with regard to the normwise relative error, with and without partial pivoting (GEPP, GE). This proves the high performance of NE for this kind of problems.

On the other hand, we stress the fact that in this case iterative refinement is not useful since the entire error arises from the ill conditioning of the problem. Moreover, the componentwise backward error (defined in Theorem 3), is of the order of the unit roundoff, which means that the method is backward stable.

Acknowledgments This work has been partially supported by the Spanish Research Grant MTM2006-03388 and under MEC and FEDER Grant TIN2007-61273.

References

1. P. Alonso, M. Gasca, J.M. Peña, Backward error analysis of Neville elimination. Appl. Numer. Math. **23**, 193 (1997)
2. P. Alonso, J. Delgado, R. Gallego, J.M. Peña, Iterative refinement for Neville elimination. Int. J. Comput. Math. **86**(2), 341 (2009)
3. T. Ando, Totally positive matrices. Linear Algebra Appl. **90**, 165 (1987)
4. M.A. Blanco, E. Francisco, V. Luaña, GIBBS: isothermal-isobaric thermodynamics of solids from energy curves using a quasi-harmonic Debey model. Comput. Phys. Commun. **158**, 57 (2004)
5. C. de Boor, A. Pinkus, Backward error analysis for totally positive linear systems. Numer. Math. **23**, 193 (1997)
6. C.J. Cramer, *Essentials of Computational Chemistry: Theories and Models* (John Wiley and Sons, England, 2004)
7. J. Demmel, P. Koev, The accurate and efficient solution of a totally positive generalized vandermonde linear system. SIAM J. Matrix Anal. Appl. **27**, 142 (2005)
8. S. Eliezer, R.A. Ricci, *High-Pressure Equations of State: Theory and Applications* (North Holland, Amsterdam, 1991)
9. G.E. Forsythe, C.B. Moler, *Computer Solutions of Linear Algebraic Systems* (Prentice-Hall, Englewood Clifs, NJ, 1967)
10. M. Gasca, J.M. Peña, Total positivity and Neville elimination. Linear Algebra Appl. **165**, 25 (1992)
11. M. Gasca, J.M. Peña, Scaled pivoting in Gauss and Neville elimination for totally positive systems. Appl. Numer. Math. **13**, 345 (1993)
12. M. Gasca, J.M. Peña, A matricial description of Neville elimination with applications to total positivity. Linear Algebra Appl. **202**, 33 (1994)
13. M. Gasca, C.A. Michelli, *Total positivity and its Applications* (Kluwer Academic Publishers, The Netherlands, 1996)
14. M. Gassó, J.R. Torregrosa, A totally positive factorization of rectangular matrices by the Neville elimination. SIAM J. Matrix Anal. Appl. **25**, 986 (2004)
15. N.J. Higham, *Accuracy and Stability of Numerical Algorithms*, 2nd edn. (SIAM, Philadelphia, 2002)
16. S. Karlin, J. McGregor, A characterization of birth and death processes. Proc. Natl. Acad. Sci. USA **45**, 375 (1959)
17. S. Karlin, J. McGregor, Coincidence properties of birth and death processes. Pac. J. Math. **9**, 1109 (1959)
18. S. Karlin, J. McGregor, Coincidence probabilities. Pac. J. Math. **9**, 1141 (1959)
19. S. Karlin, J. McGregor, Classical diffusion processes and total positivity. J. Math. Anal. Appl. **1**, 163 (1960)
20. C.M. Losert-Valiente Kroon, I.J. Ford, Stochastic birth and death equations to treat chemistry and nucleation in small systems, in *Nucleation and atmospheric aerosols*, 17th international conference (Springer, Galway, 2007), pp. 332–336
21. A.Y. Mitrophanov, Note on Zeifman's bounds on the rate of convergence for birth-death processes. J. Appl. Probab. **41**, 593 (2004)
22. C.B. Moler, Iterative refinement in floating point. J. Assoc. Comput. Mach. **14**, 316 (1967)
23. A. Mujica, A. Rubio, A. Muñoz, R.J. Needs, High-pressure phases of group-IV, III-V, and II-VI compounds. Rev. Mod. Phys. **75**, 863 (2003)
24. W. Oettli, W. Prager, Compatibility of approximate solution of linear equations with given error bounds for coefficients and right-hand sides. Numer. Math. **6**, 405 (1964)
25. J.M. Ortega, *Numerical Analysis A Second Course* (SIAM, Philadelphia, 1990)
26. J.M. Peña, *Shape preserving representations in Computer Aided-Geometric Design* (Nova Science Publishers, New York, 1999)
27. J.V. Ross, P.K. Pollett, Extinction times for a birth-death process with two phases. Math. Biosci **202**, 310 (2006)
28. R.D. Skeel, Scaling for numerical stability in Gaussian elimination. J. Assoc. Comput. Mach. **26**, 494 (1979)
29. R.D. Skeel, Iterative refinement implies numerical stability for Gaussian elimination. Math. Comput. **35**, 817 (1980)

30. J.H. Wilkinson, Progress report on the automatic computing engine. report MA/17/1024, Mathematics Division, Department of Scientific and Industrial Research, National Physical Laboratory, Teddington, UK, April 1948
31. J.H. Wilkinson, Rounding errors in algebraic processes, notes on applied science 32. Her Majesty's Stationery Office (1963)