Comput. Methods Appl. Mech. Engrg. 197 (2008) 4007–4015

Contents lists available at ScienceDirect

Comput. Methods Appl. Mech. Engrg.

journal homepage: www.elsevier.com/locate/cma

A robust algorithm for finding the eigenvalues and eigenvectors of 3×3 symmetric matrices

W.M. Scherzinger *, C.R. Dohrmann

Sandia National Laboratories, Albuquerque, NM 87185-0372, USA

ARTICLE INFO

Article history:
Received 23 October 2006
Received in revised form 21 March 2008
Accepted 25 March 2008
Available online 10 June 2008

Keywords:
Eigenvalues
Eigenvectors
Principal coordinates

ABSTRACT

Many concepts in continuum mechanics are most easily understood in principal coordinates; using these concepts in a numerical analysis requires a robust algorithm for finding the eigenvalues and eigenvectors of 3×3 symmetric matrices. A robust algorithm for solving this eigenvalue problem is presented along with an analysis of the algorithm. The special case of two or three nearly identical eigenvalues is examined in detail using an asymptotic analysis. Numerical results are shown that compare this algorithm with existing methods found in the literature. The behavior of this algorithm is shown to be more reliable than the other methods with a minimal computational cost.

1. Introduction

Many concepts in continuum mechanics are simplified when they are formulated and presented in principal coordinates. Hill [4] makes use of the "method of principal axes" in presenting a number of important concepts for solid mechanics, constitutive modeling and bifurcation theory. The method of principal axes is used to provide "a sure way through tensor algebra which can otherwise relapse into labyrinthine complexity". This statement is as true today as it was when it was published. Numerical methods for solid mechanics generally use Cartesian coordinate systems, and in doing so the form of the computational problem is quite simple. However, some of the mechanics that goes into developing the numerical problem, in particular kinematics and constitutive modeling, may be better suited to a formulation in principal coordinates. Some of these problems can be solved in other ways (see [1,3]), and some can be approximated (see [7]), but the best solution is to be able to solve the eigenvalue problem. To use principal coordinates in computational mechanics, a robust algorithm – one that gives accurate results for all conceivable cases – for finding the eigenvalues and eigenvectors of a symmetric, second-order tensor (or equivalently a 3×3 matrix) must be developed.

Since accurate solutions to the eigenvalue problem for 3×3 symmetric matrices are valuable for computational mechanics, a number of authors have developed algorithms for finding the eigenvalues and eigenvectors for these problems. Hartmann [2] provides a good review of what currently exists. The review includes two analytical methods presented by Simo and Hughes [8] and calculations are compared with a numerical solution from the LAPACK library. It is shown that while the analytical methods are faster than the numerical routine from LAPACK, for many problems the eigenvalues returned by the numerical algorithm in LAPACK are much more accurate than the analytical algorithms in Simo and Hughes [8]. Hartmann concludes that "the analytical solutions of the eigenvalue problem of symmetric second-order tensors should only be of interest in theoretical calculations". The method presented in this paper is based on an analytical solution and the accuracy exceeds that of the LAPACK library while at the same time being comparable in speed to other analytical methods. The method presented here is slightly more computationally expensive than other analytical methods, but not to the point of being a serious disadvantage. Another drawback to some of the work that has been presented in the literature is that it is not general. For example, finding the square root of a 3×3 symmetric positive definite matrix, as in [1] or [3], does not allow one to find the logarithm of that matrix. The algorithm presented here is extremely general, allowing one to calculate square roots or any other isotropic tensor function once the eigenvalues and eigenvectors are found.

In this paper we present a robust algorithm for finding the eigenvalues and eigenvectors of a 3×3 symmetric matrix. The algorithm is based on an analytical solution of the problem presented in Malvern [6]. Section 2 reviews the general algorithm from Malvern. The algorithm is analyzed by looking at the asymptotic behavior of the solution when either 2 or 3 eigenvalues are nearly identical. Based on this analysis an accurate algorithm is

* Corresponding author. Tel.: +1 505 284 4866; fax: +1 505 284 0135.
E-mail address: wmscher@sandia.gov (W.M. Scherzinger).

0045-7825/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.cma.2008.03.031

developed that solves the eigenvalue and eigenvector problem simultaneously. Furthermore, no tolerances are necessary to determine when two eigenvalues are "close". Section 3 provides an analysis of the algorithm, comparing it to a method based on Malvern, a LAPACK routine, and the algorithms presented in Hartmann [2] and Simo and Hughes [8]. The algorithm in this paper is the only algorithm that rivals the numerical results of the LAPACK routine. Furthermore, it is shown that this algorithm is much faster than the LAPACK routine and comparable to the speed of other analytical solutions. Conclusions and areas of application are presented in Section 4.

2. Eigenvalue problem

2.1. Solution of eigenvalue problem for a 3×3 matrix

The eigenvalue problem for a real symmetric 3×3 matrix is well known

$$\mathbf{A} \cdot \mathbf{v} = \lambda \mathbf{v}, \tag{1}$$

where $\mathbf{A} \in \mathrm{R}^{3\times 3}$, the eigenvector $\mathbf{v} \in \mathrm{R}^3$ and the eigenvalue $\lambda \in \mathrm{R}$. There are three eigenvalues, λ_i, and three associated eigenvectors, $\mathbf{v}_i$.[1] We are interested in symmetric matrices, $\mathbf{A} = \mathbf{A}^{\mathrm{T}}$, so the eigenvectors can be normalized such that $\mathbf{v}_i \cdot \mathbf{v}_j = \delta_{ij}$.[2]

The solution of the eigenvalue problem is based on the presentation by Malvern [6, pp. 91–93]. Instead of solving the original problem (1), the eigenvalue problem is solved for the deviatoric matrix. The deviatoric matrix, $\mathbf{A}'$, associated with $\mathbf{A}$ is

$$\mathbf{A}' = \mathbf{A} - \frac{1}{3}\mathbf{I}\mathrm{tr}\mathbf{A}, \tag{2}$$

where tr$\mathbf{A}$ is the trace of $\mathbf{A}$ and $\mathbf{I}$ is the identity matrix. It is easy to see that the eigenvalues of the deviatoric matrix are just shifted from those of the original matrix, $\mathbf{A}$, and that the eigenvectors for the two matrices are the same. The eigenvalue problem for the deviatoric matrix given by (2) is

$$\begin{aligned} &\mathbf{A}' \cdot \mathbf{v} = \eta \mathbf{v}, \\ &\eta_i = \lambda_i - \tfrac{1}{3}\mathrm{tr}\mathbf{A}. \end{aligned} \tag{3}$$

Since tr$\mathbf{A}' = 0$, the characteristic equation for the eigenvalues, η_i, is

$$\eta^3 - J_2\eta - J_3 = 0, \tag{4}$$

where J_2 and J_3 are the second and third invariants of the deviatoric matrix $\mathbf{A}'$[3]

$$J_2 = \frac{1}{2}\mathrm{tr}(\mathbf{A}' \cdot \mathbf{A}'); \quad J_3 = \det \mathbf{A}'. \tag{5}$$

The characteristic equation is a reduced cubic equation – it is missing its quadratic term – and a solution can be found using the substitution

$$\eta = 2\sqrt{\frac{J_2}{3}}\cos\alpha \tag{6}$$

in (4) and solving the resulting equation for the angle α. The solution for α is

$$\cos 3\alpha = \frac{J_3}{2}\left(\frac{3}{J_2}\right)^{3/2}. \tag{7}$$

There are three solutions to (7), α, $\alpha + 2\pi/3$ and $\alpha + 4\pi/3$, and they are shown in Fig. 1. Solving for the three angles and substituting them into (6) gives the three eigenvalues of $\mathbf{A}'$; the eigenvalues of the original matrix, $\mathbf{A}$, are found using (3).

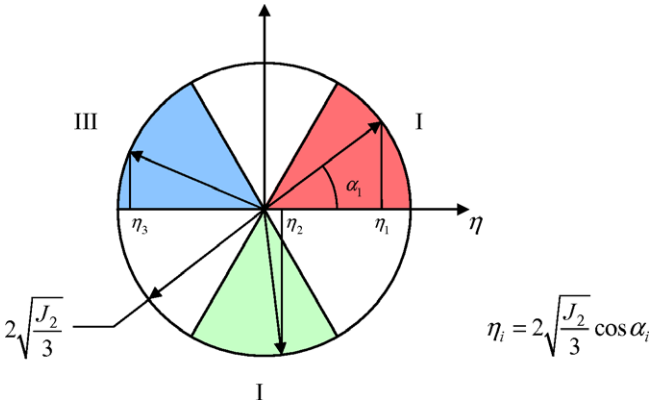

Fig. 1. The general case for finding three distinct eigenvalues. The special case of two nearly identical eigenvalues occurs as $\alpha_1 \to 0$ and $\alpha_1 \to \pi/3$. The special case of three nearly identical eigenvalues is approached as $J_2 \to 0$. As $J_2 \to 0$ the radius of the circle approaches zero.

Looking at Fig. 1 it is easy to see that the three angles associated with the roots of (5) result in eigenvalues that sum to zero, a requirement since tr$\mathbf{A}' = 0$. Furthermore, the three roots lie in distinct sectors of the circle, each angle separated by $2\pi/3$, and by finding one root we could easily find the other two. In fact, this is entirely valid theoretically. However, in a numerical algorithm when two eigenvalues are nearly identical it is difficult to retain accuracy solving the problem in this manner. The following analysis provides a method of finding an accurate numerical solution when two eigenvalues are "close".

2.2. Analysis of two nearly identical eigenvalues

Numerically, the solution presented above breaks down when we have two nearly identical eigenvalues. This is seen by making an asymptotic expansion of the eigenvalues. In addition to highlighting the problem when two eigenvalues are nearly identical, the expansion also provides a simple solution.

The analysis starts with two distinct eigenvalues, λ_1 and λ_2. Define the third eigenvalue relative to λ_2, e.g. $\lambda_3 = \lambda_2 + \Delta\lambda$. Substituting this representation into (3) the eigenvalues of the deviatoric matrix, $\mathbf{A}'$, are[4]

$$\begin{aligned} &\bar{\eta}_1 = \bar{\eta}(1-\varepsilon); \quad \bar{\eta}_2 = -\tfrac{1}{2}\bar{\eta}(1+2\varepsilon); \quad \bar{\eta}_3 = -\tfrac{1}{2}\bar{\eta}(1-4\varepsilon), \\ &\bar{\eta} = \tfrac{2}{3}(\lambda_1 - \lambda_2); \quad \varepsilon = \tfrac{\Delta\lambda}{3\bar{\eta}}. \end{aligned} \tag{8}$$

We will examine the case where $|\varepsilon| \ll 1$.

Substituting (8) into (5) we have

$$\begin{aligned} J_2 &= \tfrac{3}{4}\bar{\eta}^2(1 - 2\varepsilon + 4\varepsilon^2), \\ J_3 &= \tfrac{1}{4}\bar{\eta}^3(1 - 3\varepsilon - 6\varepsilon^2 + 8\varepsilon^3). \end{aligned} \tag{9}$$

These expressions are used in (7) to find an asymptotic expansion for $\cos 3\alpha$ in terms of ε

$$\cos 3\alpha = \mathrm{sgn}(\bar{\eta})\left(1 - \frac{27}{2}\varepsilon^2 - 27\varepsilon^3 + \mathrm{O}(\varepsilon^4)\right). \tag{10}$$

Since the leading order term in the expansion is O(1) and $\varepsilon \ll 1$, we can truncate the terms of order ε^2 and higher on the right-hand side of (10). When considering numerical applications, this is valid since adding a small number to an O(1) term will result in a loss of precision. Therefore, to first order, when $\varepsilon \ll 1$ the solutions of (10) are

[1] Latin indices, $i,j,k,\ldots$, have values from 1 to 3.

[2] δ_{ij} is the Kronecker delta. $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$.

[3] Notice that there is a sign difference in the definition of J_2 compared to the standard definition of the second invariant. This is to ensure that $J_2 > 0$ which simplifies some of the expressions that follow.

[4] The superimposed bar on the eigenvalues in (8) denote the exact solutions to the eigenvalue problem for $\mathbf{A}'$.

$$\begin{aligned} &\alpha = 0, \tfrac{2\pi}{3}, \tfrac{4\pi}{3} \quad \text{if } \bar{\eta} > 0, \\ &\alpha = \tfrac{\pi}{3}, \pi, \tfrac{5\pi}{3} \quad \text{if } \bar{\eta} < 0. \end{aligned} \tag{11}$$

This gives us the angle, α, that is used in (6). We can find the eigenvalues by substituting (9) into (6)

$$\eta = 2\sqrt{\frac{J_2}{3}}\cos\alpha = |\,\bar{\eta}\,|\,(1-\varepsilon+\mathrm{O}(\varepsilon^2))\cos\alpha. \tag{12}$$

Now two cases need to be examined, $\bar{\eta} > 0$ ($\lambda_1 > \lambda_2$) and $\bar{\eta} < 0$ ($\lambda_1 < \lambda_2$). Neglecting higher order terms, if $\bar{\eta} > 0$ the solutions to (12) are

$$\begin{aligned} &\eta_1 = \bar{\eta}(1-\varepsilon), \\ &\eta_2 = \eta_3 = -\tfrac{1}{2}\bar{\eta}(1-\varepsilon), \end{aligned} \tag{13}$$

Comparing these results to (8) we have $\eta_1 = \bar{\eta}_1$ while

$$\eta_2 = \eta_3 = \frac{\bar{\eta}_2 + \bar{\eta}_3}{2}. \tag{14}$$

If $\bar{\eta} < 0$, then the three eigenvalues are

$$\begin{aligned} &\eta_1 = \eta_2 = -\tfrac{1}{2}\bar{\eta}(1-\varepsilon) = \frac{\bar{\eta}_2 + \bar{\eta}_3}{2}, \\ &\eta_3 = \bar{\eta}(1-\varepsilon) = \bar{\eta}_1. \end{aligned} \tag{15}$$

Clearly only one accurate eigenvalue is found from this analysis: η_1 if $\bar{\eta} > 0$ and η_3 if $\bar{\eta} < 0$. In either case the other two eigenvalues are equal to each other. Furthermore they are equal to the average of the exact values, $\bar{\eta}_2$ and $\bar{\eta}_3$. In other words, looking at Fig. 1, when two eigenvalues are nearly identical, *only the most distinct eigenvalue can be found accurately by this method.*

2.3. Analysis of three nearly identical eigenvalues

The case of three nearly identical eigenvalues can be examined in a manner similar to the case of two nearly identical eigenvalues. For this case we will start with one eigenvalue, $\lambda_3 = \lambda$, and let the other two eigenvalues be "close" to λ, e.g. $\lambda_1 = \lambda + \Delta\lambda_1$ and $\lambda_2 = \lambda + \Delta\lambda_2$. Substituting into (3) we get the eigenvalues of the deviatoric matrix $\mathbf{A}'$

$$\begin{aligned} &\bar{\eta}_1 = 2\varepsilon_1 - \varepsilon_2; \quad \bar{\eta}_2 = 2\varepsilon_2 - \varepsilon_1; \quad \bar{\eta}_3 = -(\varepsilon_1 + \varepsilon_2), \\ &\varepsilon_1 = \tfrac{\Delta\lambda_1}{3\lambda}; \quad \varepsilon_2 = \tfrac{\Delta\lambda_2}{3\lambda}. \end{aligned} \tag{16}$$

We will examine the case where $|\varepsilon_1| \ll 1$ and $|\varepsilon_2| \ll 1$.

Substituting (16) into (5) we have

$$\begin{aligned} &J_2 = 3\varepsilon_1^2(1 - 2r + 4r^2), \\ &J_3 = 2\varepsilon_1^3(1 - 3r - 6r^2 + 8r^3), \quad \varepsilon_2 = 2r\varepsilon_1. \end{aligned} \tag{17}$$

Substituting these expressions into (7) we find an expression for the angle

$$\cos 3\alpha = \mathrm{sgn}(\varepsilon_1)\frac{1 - 3r - 6r^2 + 4r^3}{1 - 2r + 4r^2}. \tag{18}$$

For most values of r the equation for the angle is easy to solve and is equivalent to solving the problem for distinct eigenvalues. This brings up the idea of *relatively distinct* eigenvalues. If the eigenvalues are distinct relative to, say $\sqrt{J_2}$, then the algorithm is certainly accurate. In the case where $|r| \ll 1$, (18) reduces to the case of two nearly identical eigenvalues.

2.4. A numerical algorithm for the eigenvalue problem

Based on the analysis above, a numerical algorithm is presented that is accurate for all cases. The analysis shows that an algorithm based on the approach in Malvern [6] must take into account the special case of two nearly identical eigenvalues and this is done by *finding the most distinct eigenvalue first*. Looking at Fig. 1, this amounts to finding either η_1 or η_3 depending on the principal value of α. Eq. (7) is solved for α. If $\alpha < \pi/6$, then we find η_1 first; if $\alpha > \pi/6$, then we find η_3 first. When $\alpha = \pi/6$ the eigenvalues are equally distinct and there are no concerns with nearly identical eigenvalues in the algorithm.

Once the most distinct eigenvalue is found the next step in the algorithm is to solve for the eigenvector corresponding to this eigenvalue. Without loss of generality, assume η_1 is the most distinct eigenvalue. The eigenvector $\mathbf{v}_1$ corresponding to η_1 lies in the null space of the operator $\mathbf{A}' - \eta_1\mathbf{I}$. Using this property, we can construct the following three vectors that lie in the plane orthogonal to $\mathbf{v}_1$, the orthogonal complement to the null space of the operator $\mathbf{A}' - \eta_1\mathbf{I}$

$$\mathbf{r}_i = (\mathbf{A} - \eta_1\mathbf{I})\cdot\mathbf{e}_i, \tag{19}$$

where the vectors $\mathbf{e}_i$ are the three Cartesian basis vectors. Since the three vectors $\mathbf{r}_i$ lie in a plane, they are linearly dependent. This is shown schematically in Fig. 2. We can choose the vector $\mathbf{r}_i$ that is the largest – without loss of generality assume it is $\mathbf{r}_1$ – and normalize it, $\mathbf{s}_1 = \mathbf{r}_1/|\mathbf{r}_1|$. With the vector $\mathbf{s}_1$ we can form two vectors that are orthogonal to both $\mathbf{v}_1$ and $\mathbf{s}_1$

$$\begin{aligned} &\mathbf{t}_2 = \mathbf{r}_2 - (\mathbf{s}_1\cdot\mathbf{r}_2)\mathbf{s}_1, \\ &\mathbf{t}_3 = \mathbf{r}_3 - (\mathbf{s}_1\cdot\mathbf{r}_3)\mathbf{s}_1, \end{aligned} \tag{20}$$

as shown in Fig. 3. Normalizing the largest of these – without loss of generality assume it is $\mathbf{t}_2$ – gives an orthonormal triad of vectors: $\mathbf{v}_1$, $\mathbf{s}_1$ and $\mathbf{s}_2 = \mathbf{t}_2/|\mathbf{t}_2|$. So far we have only solved for $\mathbf{s}_1$ and $\mathbf{s}_2$, $\mathbf{v}_1$ is still unknown. But $\mathbf{v}_1$ is simply the cross product of $\mathbf{s}_1$ and $\mathbf{s}_2$

$$\mathbf{v}_1 = \mathbf{s}_1 \times \mathbf{s}_2. \tag{21}$$

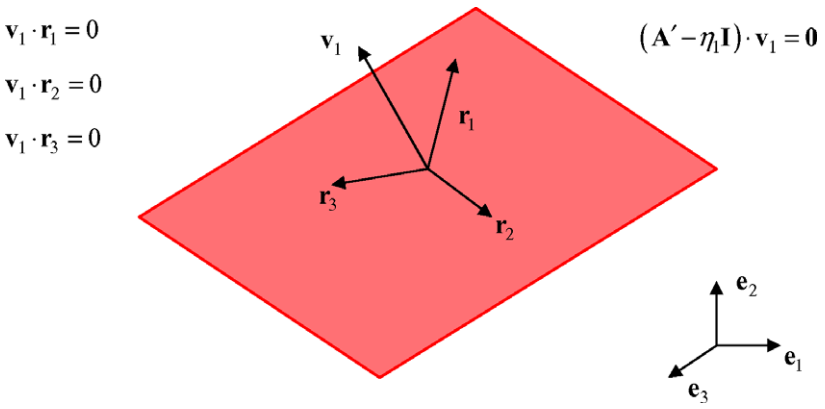

Fig. 2. The null space of the operator $\mathbf{A}' - \eta_1\mathbf{I}$ and the projections of the Cartesian basis vectors on the orthogonal complement of the null space. The three vectors $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{r}_3$ lie in the orthogonal complement of the null space and are linearly dependent.

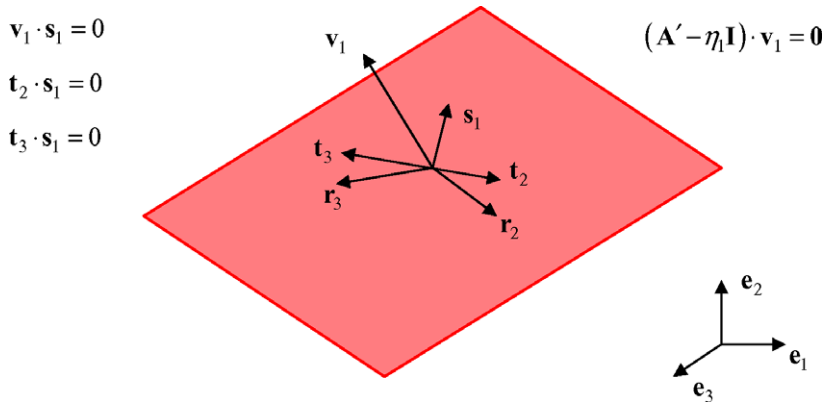

Fig. 3. The normalization of $\mathbf{r}_1$, the largest vector lying in the orthogonal complement to the null space of $\mathbf{A}' - \eta_1 \mathbf{I}$. The vectors $\mathbf{t}_2$ and $\mathbf{t}_3$ are formed so that they are orthogonal to both $\mathbf{v}_1$ and $\mathbf{s}_1$. As a result they are co-linear.

At this point in the analysis the most distinct eigenvalue and its corresponding eigenvector have been found. The eigenvector gives us a direction in which we have solved our problem. Now the problem of finding the remaining two eigenvalues that are "close", and their corresponding eigenvectors, remains. To accomplish this final task the components of $\mathbf{A}'$ are re-written in the $\mathbf{v}_1, \mathbf{s}_1, \mathbf{s}_2$ basis. The components of the deviatoric matrix in this basis are

$$[\bar{A}'_{ij}] = \begin{bmatrix} \eta_1 & 0 & 0 \\ 0 & \mathbf{s}_1 \cdot \mathbf{A}' \cdot \mathbf{s}_1 & \mathbf{s}_1 \cdot \mathbf{A}' \cdot \mathbf{s}_2 \\ 0 & \mathbf{s}_2 \cdot \mathbf{A}' \cdot \mathbf{s}_1 & \mathbf{s}_2 \cdot \mathbf{A}' \cdot \mathbf{s}_2 \end{bmatrix}. \tag{22}$$

From (22) it is easy to see that the original three-dimensional problem has been reduced to an equivalent two-dimensional problem.

The first step in solving the two-dimensional problem is to find the remaining two eigenvalues. These two eigenvalues are "close" relative to the first eigenvalue and any solution to this problem must take this into account. The remaining eigenvalues are the roots to the following quadratic equation:

$$\eta^2 - (\bar{A}'_{22} + \bar{A}'_{33})\eta + (\bar{A}'_{22}\bar{A}'_{33} - \bar{A}'_{23}\bar{A}'_{32}) = 0. \tag{23}$$

Solving this directly poses problems in the case of two nearly identical eigenvalues. This can be seen using an asymptotic analysis similar to what was done above. When the remaining two eigenvalues are nearly identical we have from (8)

$$\bar{\eta}_2 = -\tfrac{1}{2}\bar{\eta}(1 + 2\varepsilon); \quad \bar{\eta}_3 = -\tfrac{1}{2}\bar{\eta}(1 - 4\varepsilon), \tag{24}$$

which are the roots of the following quadratic equation:

$$\eta^2 + \eta\bar{\eta}(1 - \varepsilon) + \frac{1}{4}\bar{\eta}^2(1 - 2\varepsilon + 8\varepsilon^2) = 0. \tag{25}$$

Writing out the solution to (25) using the quadratic formula we have

$$\eta = -\frac{1}{2}\bar{\eta}(1 - \varepsilon) \pm \frac{1}{2}\bar{\eta}\sqrt{(1 - \varepsilon)^2 - (1 - 2\varepsilon + 8\varepsilon^2)}. \tag{26}$$

Since we are squaring terms that look like $1 + O(\varepsilon)$, the discriminant will numerically evaluate to zero. The numerical result will be two eigenvalues that are equal to each other. In fact, the solution to (26) when $\varepsilon \ll 1$ is exactly the same as (14). So if the quadratic formula is used to find the last two eigenvalues we have gained nothing by solving for the most distinct eigenvalue first.

Rather than solving (23) directly, the Wilkinson shift is used (see [5, p. 279]). This method gives the following theoretically equivalent solution for η_2 and η_3

$$\begin{aligned} \eta_2 &= \tfrac{\bar{A}'_{22} + \bar{A}'_{33}}{2} - \tfrac{1}{2}\operatorname{sgn}(\bar{A}'_{22} - \bar{A}'_{33})\sqrt{(\bar{A}'_{22} - \bar{A}'_{33})^2 + 4\bar{A}'_{23}\bar{A}'_{32}}, \\ \eta_3 &= \bar{A}'_{22} + \bar{A}'_{33} - \eta_2. \end{aligned} \tag{27}$$

Analysis of (27) shows

$$\eta_2 = -\frac{\bar{\eta}}{2}(1 - \varepsilon) + \frac{\bar{\eta}}{2}\operatorname{sgn}(\varepsilon)\sqrt{9\varepsilon^2}. \tag{28}$$

Since the value under the square root is $O(\varepsilon^2)$, and not $1 + O(\varepsilon^2)$, no precision is lost. The final two eigenvalues are

$$\begin{aligned} &\eta_2 = -\tfrac{1}{2}\bar{\eta}(1 - 4\varepsilon), \quad \eta_3 = -\tfrac{1}{2}\bar{\eta}(1 + 2\varepsilon) \ \text{ if } \varepsilon > 0, \\ &\eta_2 = -\tfrac{1}{2}\bar{\eta}(1 + 2\varepsilon), \quad \eta_3 = -\tfrac{1}{2}\bar{\eta}(1 - 4\varepsilon) \ \text{ if } \varepsilon < 0. \end{aligned} \tag{29}$$

Comparing (29) with (24) it is seen that these solutions are exact. The Wilkinson shift gives extremely accurate values for the remaining eigenvalues, even if the two eigenvalues are close relative to the first.

Finally, the eigenvectors for the last two eigenvalues are found. We form the following two vectors, orthogonal to the null space of $\mathbf{A}' - \eta_2\mathbf{I}$:

$$\begin{aligned} \mathbf{u}_1 &= (\mathbf{A}' - \eta_2\mathbf{I}) \cdot \mathbf{s}_1, \\ \mathbf{u}_2 &= (\mathbf{A}' - \eta_2\mathbf{I}) \cdot \mathbf{s}_2 \end{aligned} \tag{30}$$

determine which one is larger – without loss of generality assume it is $\mathbf{u}_1$ – and normalize it, $\mathbf{w}_1 = \mathbf{u}_1/|\mathbf{u}_1|$. Note that since we have reduced the problem to finding the null space of an operator acting on a two-dimensional space, the vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ are co-linear. Since $\mathbf{v}_1$ is already orthogonal to the null space of $\mathbf{A}' - \eta_2\mathbf{I}$, we find

$$\begin{aligned} \mathbf{v}_2 &= \mathbf{w}_1 \times \mathbf{v}_1, \\ \mathbf{v}_3 &= \mathbf{v}_1 \times \mathbf{v}_2. \end{aligned} \tag{31}$$

This completes the eigenvalue/eigenvector analysis for the deviatoric, symmetric 3×3 matrix.

The solution of the original problem is found by shifting the eigenvalues using (3). The accuracy of the computed eigenvalues is easily computed if the exact eigenvalues are known. Another way that the accuracy of the entire problem can easily be checked is to re-assemble the original matrix

$$\mathbf{A} = \lambda_1 \mathbf{v}_1\mathbf{v}_1 + \lambda_2 \mathbf{v}_2\mathbf{v}_2 + \lambda_3 \mathbf{v}_3\mathbf{v}_3 \tag{32}$$

and compare the matrix components term by term.

With the results presented here, a robust algorithm can be developed that finds the eigenvalues and eigenvectors of symmetric 3×3 matrices. The important steps in the algorithm are as follows:

1. Construct the deviatoric matrix (Eq. (2)).
2. Find the invariants, J_2 and J_3, of $\mathbf{A}'$ (Eq. (5)).
3. Find the angle α (Eq. (7)).
4. Find the most distinct eigenvalue based on α (is $\alpha < \pi/6$ or is $\alpha > \pi/6$).

5. Find the eigenvector corresponding to the most distinct eigenvalue.
6. Find the reduced form of the matrix (Eq. (22)).
7. Find the remaining two eigenvalues using the Wilkinson shift (Eq. (27)).
8. Find the eigenvectors corresponding to the last two eigenvalues.

The asymptotic analysis shows that this algorithm is robust. The robustness will be demonstrated with numerical examples in the next section.

3. Numerical examples

The previous section presented a robust algorithm based on an asymptotic analysis for finding the eigenvalues of a symmetric 3×3 matrix. In this section we present numerical results that demonstrate the accuracy and speed of this algorithm. The accuracy of the eigenvalue/eigenvector algorithm is tested in two ways. First, the original eigenvalues are compared to the computed eigenvalues. Second, the original matrix components are compared to the matrix components calculated using (32). If an algorithm provides accurate eigenvalues *and* can reconstruct the original matrix, then the algorithm is useful for numerical applications.

In this section we review a general testing methodology that will be used for the analyses that follow. The specialization of the methodology to the case of two or three nearly identical eigenvalues is also presented. Next, results showing the accuracy of the algorithms for two and three nearly identical eigenvalues are presented. In addition to the algorithm presented in this paper, results are generated for the algorithm in Malvern [6] and the three algorithms presented in Hartmann [2]: the LAPACK routine DSYEV and two analytical algorithms (an arccosine method and an arctangent method) from the work of Simo and Hughes [8]. This is followed by a brief discussion of how accuracy in the case of two nearly identical eigenvalues can be guaranteed by the algorithm proposed in this paper. These results support the analysis of Section 2. Finally, timing results are presented that show that the proposed algorithm is competitive in terms of speed.

3.1. General testing

Numerical testing of the algorithms requires randomly generated symmetric 3×3 matrices. The general routine for creating these matrices begins by finding three random eigenvalues. These eigenvalues are placed on the diagonal of a 3×3 matrix

$$[\mathbf{\Lambda}] = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}, \tag{33}$$

where λ_1, λ_2 and λ_3 are the three randomly generated eigenvalues. Next, this matrix is pre- and post-multiplied with a randomly generated rotation matrix, $\mathbf{Q}$, to generate a random, symmetric 3×3 matrix

$$\mathbf{A} = \mathbf{Q}^{\mathrm{T}} \cdot \mathbf{\Lambda} \cdot \mathbf{Q}, \tag{34}$$

that has the eigenvalues λ_1, λ_2 and λ_3. The rotation matrices, $\mathbf{Q}$, are generated by computing a randomly oriented axis of rotation and a random angle of rotation about that axis. The randomly generated 3×3 symmetric matrix, $\mathbf{A}$, is passed to the algorithms where the eigenvalues and eigenvectors are calculated.

The accuracy of a particular eigenvalue/eigenvector algorithm can be measured in two ways. First, the calculated eigenvalues can be compared to the original eigenvalues; this measures the ability of an algorithm to calculate the eigenvalues to a given degree of accuracy. Second, the calculated eigenvectors can be compared to the original eigenvectors – i.e. the columns of $\mathbf{Q}$. However, this approach has problems since nearly identical eigenvalues may have a plane, or all of R^3, where the eigenvectors may lie. Rather than compare eigenvectors term by term to the columns of $\mathbf{Q}$, we can compare the components of the re-assembled matrix (32) to the original components of (34). If the eigenvalues are accurate this is a good measure of the accuracy of their corresponding eigenvectors. Additionally this exercises the eigenvalues and eigenvectors in a way we would use them in a numerical simulation – i.e. when we use them to compute isotropic tensor functions of the original symmetric tensor.

3.2. Generation of two and three nearly identical eigenvalues

For the accuracy calculations two cases are presented. The first case is that of two nearly identical eigenvalues and the second case is that of three nearly identical eigenvalues.

For the case of two nearly identical eigenvalues, two eigenvalues are randomly generated and the second eigenvalue is perturbed to generate the third eigenvalue

$$\begin{aligned} \lambda_1 &= 5(2m_1 - 1), \\ \lambda_2 &= 5(2m_2 - 1), \\ \lambda_3 &= \lambda_2 + \varepsilon(2m_3 - 1), \end{aligned} \tag{35}$$

where $\varepsilon \ll 1$ is the perturbation amplitude and $m_i \in [0,1)$ are random numbers. For the calculations presented here $\varepsilon = 10^n$ where n is an integer such that $n \in [-15,-1]$. The "closeness" of the eigenvalues increases with decreasing n. In the case of three nearly identical eigenvalues, the eigenvalues are generated in a similar manner. One eigenvalue is randomly generated and the other two are perturbations about the first eigenvalue. The three eigenvalues are

$$\begin{aligned} \lambda_1 &= 5(2m_1 - 1), \\ \lambda_2 &= \lambda_1 + \varepsilon(2m_2 - 1), \\ \lambda_3 &= \lambda_1 + \varepsilon(2m_3 - 1). \end{aligned} \tag{36}$$

Once again ε is the perturbation amplitude, $\varepsilon = 10^n$ where n is an integer such that $n \in [-15,-1]$ and $m_i \in [0,1)$ are random numbers.

For both cases – two and three nearly identical eigenvalues – random rotation matrices are also generated and a symmetric 3×3 matrix is constructed using (34).

3.3. Accuracy for two nearly identical eigenvalues

The accuracy of the algorithms when two eigenvalues are close relative to the third is shown in Fig. 4. This shows the average absolute error (dashed lines) and the maximum absolute error (solid lines) between the original eigenvalue and the computed eigenvalue as a function of the perturbation amplitude, ε. The plot is a result of 10^9 calculations *per* perturbation amplitude.

The algorithm presented in this paper has excellent performance for all orders of perturbation amplitude. The absolute error in the eigenvalue calculations is always on the order of machine precision for this algorithm. This level of accuracy is matched only by the LAPACK algorithm. The error in the other algorithms is unacceptable, as shown by Hartmann [2].

The accuracy of the eigenvector calculations is shown in Fig. 5. Clearly we can re-assemble the original matrix (or form functions of matrices) with a great deal of accuracy using either the algorithm presented in this paper or the eigenvalue/eigenvector solver from LAPACK. The other algorithms show significant errors here as well, although the errors might be from the eigenvalue calculations. The errors make these algorithms unacceptable for use as a general purpose numerical algorithm. The only acceptable

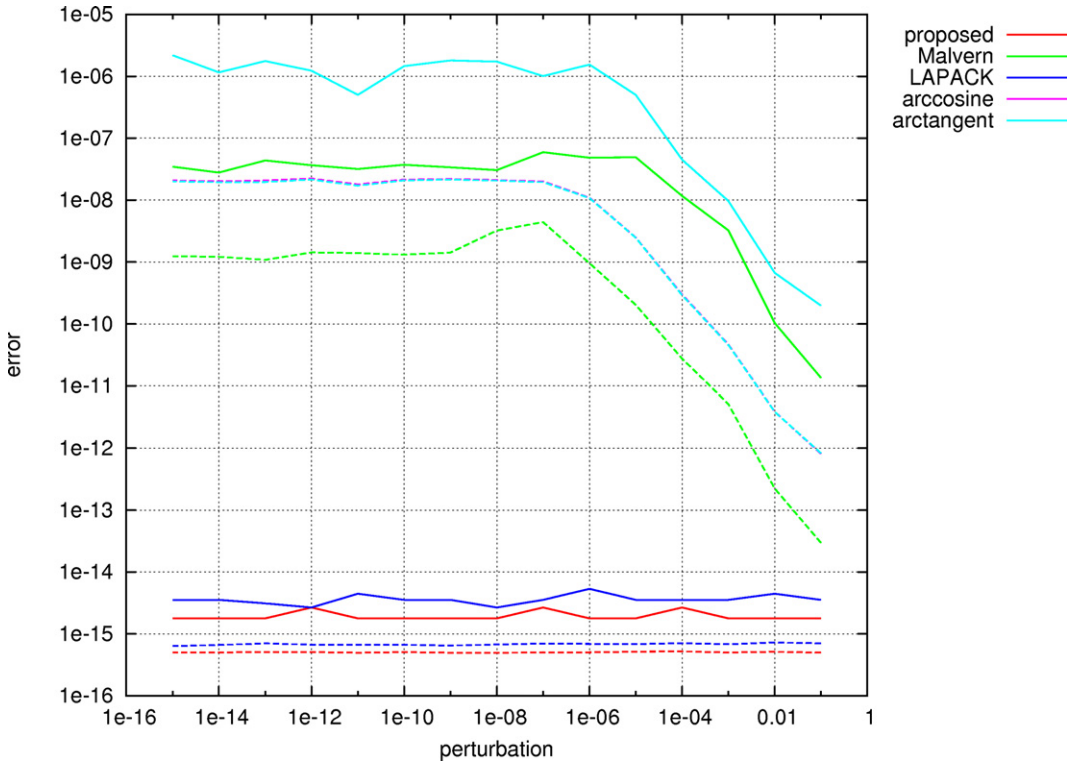

Fig. 4. The absolute error in the eigenvalue calculations with respect to perturbation amplitude for two nearly identical eigenvalues. Five algorithms are examined: the algorithm from Malvern [6], the LAPACK algorithm, the arccos and arctan algorithms and the algorithm proposed in this paper. The dashed lines show the average errors for 10^9 calculations. The solid lines indicate the maximum error over those 10^9 calculations. The errors for the arccos and arctan algorithms lie on top of each other. The errors in the proposed algorithm are on the order of machine precision.

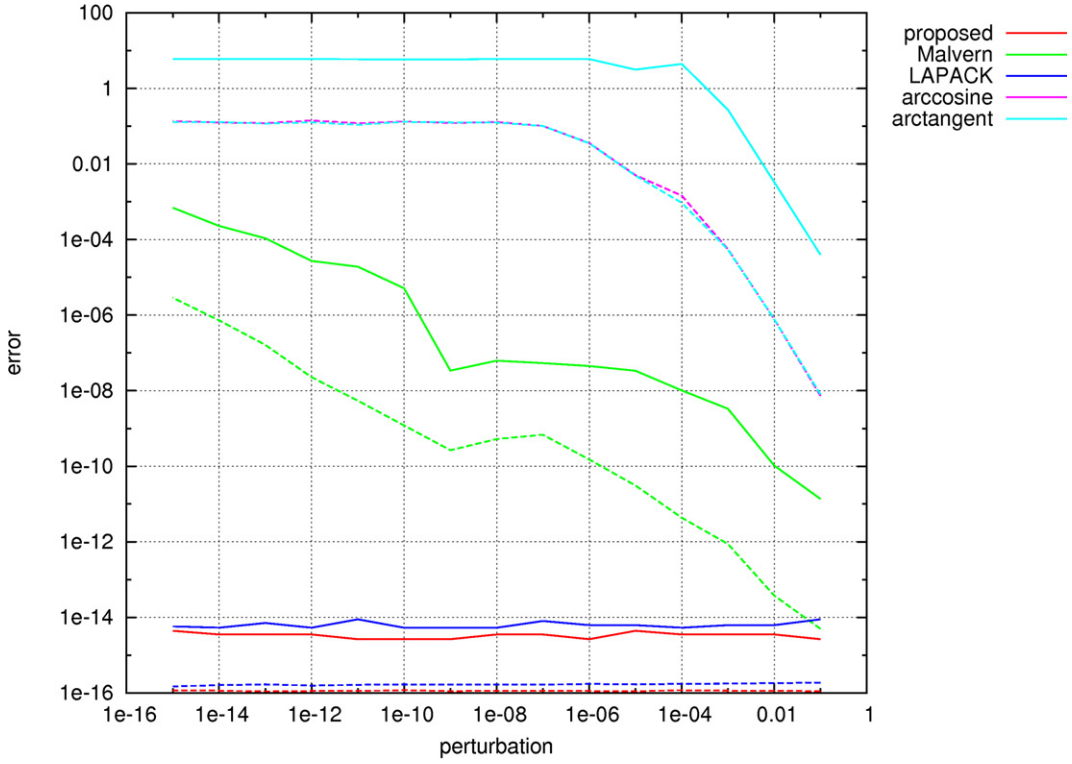

Fig. 5. The absolute error in the components of the re-assembled matrix with respect to perturbation amplitude for two nearly identical eigenvalues. Five algorithms are examined: the algorithm from Malvern [6], the LAPACK algorithm, the arccos and arctan algorithms and the algorithm proposed in this paper. The dashed lines show the average errors for 10^9 calculations. The solid lines indicate the maximum error over those 10^9 calculations. The errors in the proposed algorithm are on the order of machine precision.

algorithms are the algorithms in LAPACK and the one presented in this paper.

3.4. Accuracy for three nearly identical eigenvalues

The accuracy of the eigenvalue calculations of the three algorithms when all three eigenvalues are nearly identical is shown in Fig. 6. The results are generated for 10^9 cases per perturbation amplitude.

These results show that, on average (dashed lines), three of the algorithms – the proposed algorithm, the LAPACK algorithm and the Malvern algorithm – perform well when all three eigenvalues are nearly identical. The magnitude of the maximum error for these three algorithms is independent of perturbation amplitude. This is understood for two of the cases since the case of three nearly identical eigenvalues has been shown to be no different from three distinct eigenvalues for both the method presented in Malvern and the method presented in this paper. The accuracy of the other two algorithms is variable and depends on tolerances that can be set for each algorithm.

3.5. Ensuring accuracy of solutions

The accuracy of the eigenvalues is ensured by two things: finding the most distinct eigenvalue first and using a Wilkinson shift to find the remaining two eigenvalues. The analysis showing the accuracy of the algorithm presented in this paper was presented in Section 2. One aspect of this analysis can be verified by modifying the algorithm so that the most distinct eigenvalue is *not* found first. Instead of finding the most distinct eigenvalue first we find the largest eigenvalue first – i.e. the eigenvalue found using the principal value of α, α_1 (see Fig. 1).

Statistics are generated for the errors in the eigenvalue calculations for various perturbation amplitudes. The results are shown in Fig. 7; they are the result of 10^9 calculations for each perturbation amplitude. The results show the exact behavior predicted from the analysis in Section 2. With a large perturbation – essentially three "distinct" eigenvalues – the errors are distributed around numerical precision. As the perturbation amplitude decreases two distinct behaviors are seen: half of the calculations begin to show significant errors and the other half of the calculations continue to give accurate results. The reason for this is that half of the time, the largest eigenvalue will also be the most distinct (when $\alpha_1 < \pi/6$). When this is the case, the results are the same as using the algorithm as presented in this paper. In fact, for all of the cases shown in Fig. 7, the distributions for the accurate results look exactly the same as the distribution for the case with distinct eigenvalues, but they are half the amplitude. For the other half of the cases, where the largest eigenvalue is not the most distinct, significant errors are seen and the error grows as the perturbation amplitude decreases.

3.6. Timing studies

If an eigenvalue/eigenvector calculation will be used in the numerical analysis of a mechanics problem it must be fast. Defining speed constraints explicitly, however, is difficult. Furthermore, any speed constraints must also be weighed against accuracy constraints. Since there is no absolute speed criterion, a relative speed comparison of the various routines is presented here.

Table 1 shows the timing data for the algorithms. This data was gathered for the simple case of three distinct eigenvalues. A symmetric matrix was created, as above, and the matrix was sent to each routine 10^9 times. Two things are evident from the timing re-

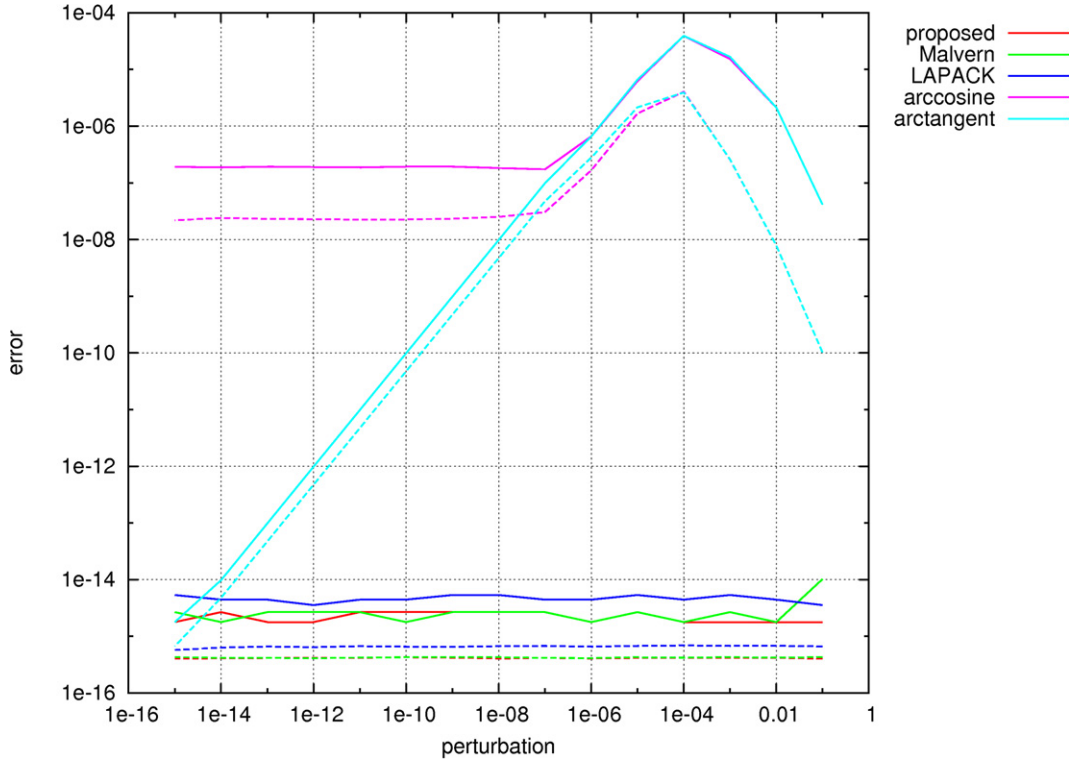

Fig. 6. The absolute error in the eigenvalue calculations with respect to perturbation amplitude for three nearly identical eigenvalues. Five algorithms are examined: the algorithm from Malvern [6], the LAPACK algorithm, the arccos and arctan algorithms and the algorithm proposed in this paper. The dashed lines show the average errors for 10^9 calculations. The solid lines indicate the maximum error over those 10^9 calculations. The errors in the proposed algorithm are on the order of machine precision.

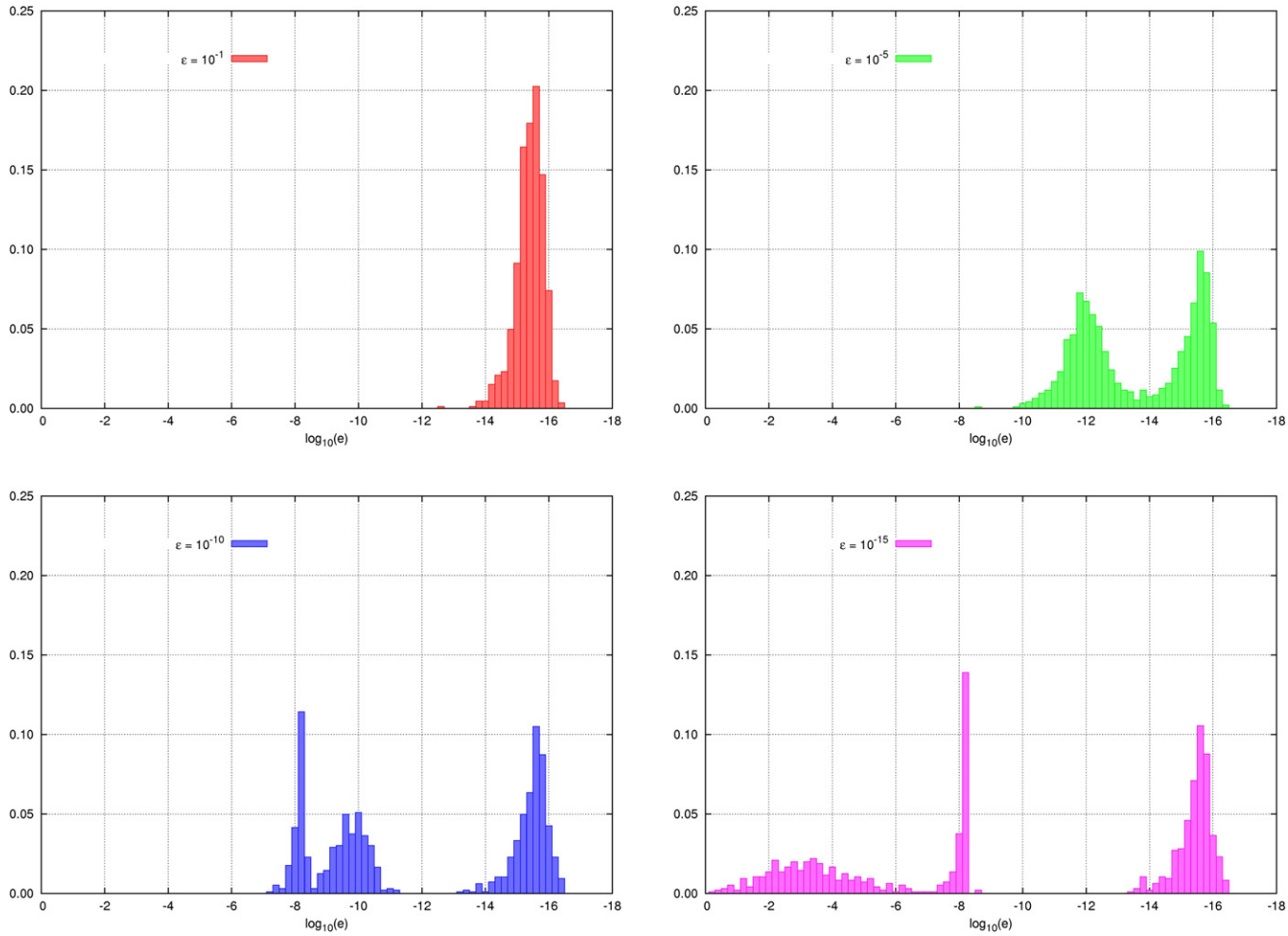

Fig. 7. Distributions of error in the eigenvalue calculations for the proposed algorithm when the largest eigenvalue, η_1, is found first rather than the most distinct eigenvalue. These results are for two nearly identical eigenvalues with 10^9 calculations run for four perturbation amplitudes: $\varepsilon = 10^{-1}$, 10^{-5}, 10^{-10} and 10^{-15}. As the perturbation amplitude decreases two distinct distributions form. Since half of the time the largest eigenvalue will also be the most distinct eigenvalue, the distribution around 10^{-15} keeps the same shape and is half the amplitude of the distribution for O(1) perturbations. The other half of the distribution corresponds to cases when the largest eigenvalue is not the most distinct eigenvalue.

Table 1
Comparison of solution times for 10^9 trials for each method

Method	Time (min)
Malvern	26.1
LAPACK	97.5
arccos	21.6
arctan	23.1
proposed	27.6

The proposed algorithm is comparable in terms of speed to the other analytical methods. The general numerical approach using the DSYEV routine in LAPACK is much slower due to the more general nature of the algorithm.

sults. First, the LAPACK algorithm is much slower than the algorithms based on analytical methods. This is due to the generality of this algorithm as mentioned in Hartmann [2]. This fact makes it inappropriate for this numerical application – it is too general. However, it is conceivable that a specialized iterative solver for 3×3 symmetric matrices could be developed that would also compete with the speed of analytically based methods. Second, the algorithm proposed in this paper is comparable in speed to other, much less accurate, analytically based methods. Since the method is both accurate and fast it is well suited to algorithms in computational mechanics.

4. Conclusions

An algorithm for computing the eigenvalues and eigenvectors of 3×3 symmetric matrices has been presented. This algorithm is extremely accurate – especially for the case of nearly identical eigenvalues. The algorithm is based on an analytical solution of the eigenvalue problem. Asymptotic analysis of the analytical solution leads to a very simple and robust numerical algorithm. The algorithm is comparable in speed to other analytical algorithms with accuracy that is only achieved with fully numerical treatments.

In computational mechanics both accuracy and speed are important. Since many rank two symmetric tensors – represented in a numerical code as a 3×3 matrix – will often have eigenvalues close to unity, finding these eigenvalues as accurately as possible is extremely important if they are to be used in an analysis. An example is the right Cauchy–Green tensor. Calculating the eigenvalues and eigenvectors of the right Cauchy–Green tensor numerically requires extreme accuracy if one is to calculate other kinematic quantities from it, e.g. the right stretch tensor or the logarithmic strain. If accuracy is not obtained then a small strain problem could become difficult to solve. The algorithm must also be fast. If it is used for constitutive modeling, then it will be called thousands of times for each material point – millions or possibly billions of times in an analysis.

The algorithm has many applications in numerical analysis for solid mechanics problems. Applications include, but are not limited to the calculation of: principal stretches and directions, principal stresses and directions, the polar decomposition, the logarithmic strain and strain measures of the Seth–Hill type. Most important is the potential change in how the fundamental mechanics are formulated and calculated computationally. Computational mechanics will be positively affected by the ability to accurately solve kinematics and constitutive modeling problems in principal coordinates.

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under contract DE-AC04-94AL85000.

References

[1] L.P. Franca, An algorithm to compute the square root of a 3×3 positive definite matrix, Comput. Math. Appl. 18 (5) (1989) 459–466.
[2] S. Hartmann, Computational aspects of the symmetric eigenvalue problem of second order tensors, Tech. Mech. 23 (2–4) (2003) 283–294.
[3] N.J. Higham, Stable iterations for the matrix square root, Numer. Algor. 15 (1997) 227–242.
[4] R. Hill, Aspects of invariance in solid mechanics, Adv. Appl. Mech. 18 (1978) 1–75.
[5] G.H. Golub, C.F. van Loan, Matrix Computations, The Johns Hopkins University Press, Baltimore, MD, 1983.
[6] L.E. Malvern, Introduction to the Mechanics of a Continuous Medium, Prentice-Hall, Englewood Cliffs, NJ, 1969.
[7] M.M. Rashid, Incremental kinematics for finite element applications, Int. J. Numer. Methods Engrg. 36 (1993) 3937–3956.
[8] J.C. Simo, T.J.R. Hughes, Computational Inelasticity, Springer-Verlag, New York, NY, 1998.