James Hardy Wilkinson
Note on inverse iteration and III-conditioned eigensystems

Acta Universitatis Carolinae. Mathematica et Physica, Vol. 15 (1974), No. 1-2, 173--177

Persistent URL: http://dml.cz/dmlcz/142352

1974 ACTA UNIVERSITATIS CAROLINAE - MATHEMATICA ET PHYSICA NO. 1-2 PAG. 173-177

Note on Inverse Iteration and Ill-Conditioned Eigensystems

J. H. WILKINSON
National Physical Laboratory, Teddington

Inverse iteration is one of the most powerful tools in numerical analysis. Used directly or in a concealed form as in the QR or LR algorithms it is at the heart of many of the most successful algorithms for computing eigensystems.

The motivation for it is the following trivial observation. Suppose A has a complete system of normalized eigenvectors u_i corresponding to eigenvalues λ_i. If x_0 is an arbitrary vector then we may write

$$x_0 = \sum \alpha_i u_i$$

using the u_i as a basis. We then have

$$x_r \equiv (A - pI)^{-r} x_0 = \sum \alpha_i u_i / (\lambda_i - p)^r \,. \tag{1}$$

If $|\lambda_s - p| = \min_i |\lambda_i - p|$ and the minimum is achieved for a single s then as r is increased, x_r is increasingly dominated by its component in the direction of u_s. In fact

$$x_r / \|x_r\|_2 \to k u_s \tag{2}$$

where k is a scalar of modulus unity. If $|\lambda_s - p| << |\lambda_i - p|$ $(i \neq s)$ then convergence is very fast. Conditions are very favourable if p has been determined as a computed eigenvalue using a stable algorithm.

From (1) it is clear that x_r steadily improves as an approximation to u_s, though if p is exceptionally close to λ_s and λ_s is well separated from the other eigenvalues one might expect one iteration to produce a fully acceptable approximation to u_s unless α_s happened to be pathologically small as a result of an unfortunate choice of x_0. In fact if $\lambda_s - p = \varepsilon$ we have

$$x_r = \frac{1}{\varepsilon^r} \left[\alpha_s u_s + \sum_{i \neq s} \frac{\alpha_i \varepsilon^r u_i}{(\lambda_i - p)^r} \right] \tag{3}$$

which presents the situation in stark relief. Assuming exact computation there appears to be everything to be gained by continuing iteration as far as accuracy is concerned since from (3) the unwanted components suffer a diminution relative to that of u_s at each iteration.

In practice inverse iteration is usually carried out in the steps

$$(A - pI)\, y_{r+1} = x_r\,, \quad x_{r+1} = y_{r+1}/\|y_{r+1}\|_2\,. \tag{4}$$

The unnormalized vector y_{r+1} is determined first at each stage and then the normalized version x_{r+1} is derived from it. $(A - pI)^{-1}$ is not computed directly but rather y_{r+1} is derived from x_r by solving the linear system in (4), usually by decomposing $(A - pI)$ into its LU factors (with pivoting) or its QR factors. Although LR factorization is usually used in practice we shall assume a QR factorization since then we can make a categorical statement about stability when rounding errors are involved.

It is interesting at each stage to form the residual vector r_r defined by $r_r = (A - pI)\, x_r$. (Here we are assuming that p does come from a stable eigenvalue routine so that it is natural to use it as the approximate eigenvalue when forming the residual). If this is done in practice one finds that when p is an approximation to an ill-conditioned eigenvalue, then after one iteration the residual r_1 is usually negligible but all subsequent r_s are far larger! This contradicts our natural expectations since the x_r should be "improving". Of course, when rounding errors are made x_r cannot continue to improve indefinitely since we are allowing only a fixed number of digits for its representation. Nevertheless the large *increase* in the size of the residual calls for an explanation.

Although the phenomenon is now fully understood and the analysis shows that a large increase in the residual would be common even if inverse iteration were performed exactly, it does not seem to be fully appreciated that the process of inverse iteration itself does not play a very important role. The main contributary factor is the nature of the basis $u_1, u_2, \ldots, u_n$ in the case when A has an ill-conditioned eigensystem.

An eigenvalue λ_1 is ill-conditioned if $s_1 = y_1^H x_1$ is small, where y_1 and x_1 are normalized left-hand and right-hand eigenvectors corresponding to λ_1. It is well known that if s_1 is small then at least one other s_i is small. We cannot have just one ill-conditioned eigenvalue! A group of k associated eigenvalues, say $\lambda_1, \ldots, \lambda_k$ will have a set of normalized eigenvectors $u_1, \ldots, u_k$ which, although independent, will be *almost* linearly dependent in the sense that there will be a unit vector β of order k such that $\|\beta_1 u_1 + \ldots + \beta_k u_k\|$ is small. Although an arbitrary unit vector x will always be expressible in the form $\sum \alpha_i u_i$ the components $\alpha_1, \ldots, \alpha_k$ will be rather special for almost all x.

We may illustrate this by an example in which $n = 3$, $k = 3$ and the u_i are given by

u_1	u_2	u_3
.701 645 809	.701 091 297	.702 148 790
.589 614 626	.589 943 936	.589 315 138
.400 059 689	.400 546 058	.399 618 250

The three vectors are parallel to three decimal places since $|u_i - u_j|$ is of the order of 10^{-3}. The trio of vectors is much more nearly linearly dependent since $\|.816\ 181\ 505u_1 - .388\ 450\ 021u_2 - .427\ 731\ 607u_3\|$ is of the order of 10^{-6}. If we choose $\alpha_1, \alpha_2, \alpha_3$ at random with $\|\alpha\|_2 = 1$ then for almost all choices the vector $y = \alpha_1 y_1 + \alpha_2 u_2 + \alpha_3 u_3$ will be in the same general direction as the u_i. It is only if we choose the α_i rather specially that we shall obtain a vector in the rest of the 3-space; when such special α_i are chosen there will be a great deal of cancellation when y is computed. On the other hand if we take a random unit vector and express it in terms of u_1, u_2, u_3 then unless the vector happens to be roughly in the direction of the u_i it must be associated with a set of α_i which are special in some sense.

If we derive a set of orthonormal vectors v_i from the u_i we find that

$$\begin{aligned} v_1 &= u_1 \\ v_2 &= (1.238)10^3 u_2 - (1.238)10^3 u_1 \\ v_3 &= (2.786)10^6 u_3 - (2.529)10^6 u_2 - (2.57)10^5 u_1 . \end{aligned}$$

A random unit x will be of the form $\beta_1 v_1 + \beta_2 v_2 + \beta_3 v_3$ where $\|\beta\|_2 = 1$ and since the v_i are orthogonal the β_i will be random; the probability that $|\beta_3| < 10^{-k}$ is of the order of 10^{-k}. Now we have

$$\begin{aligned} x = \sum \beta_i v_i &= \beta_1 u_1 + 10^3 \beta_2 [1.238u_2 - 1.238u_1] + 10^6 \beta_3 [2.786u_3 - 2.529u_2 - 0.257u_1] \\ &= u_1[\beta_1 - (1.238)10^3 \beta_2 - (0.257)10^6 \beta_3] \\ &\quad + u_2[(1.238)10^3 \beta_2 - (2.529)10^6 \beta_3] \\ &\quad + u_3[(2.786)10^6 \beta_3] \\ &= \alpha_1 u_1 + \alpha_2 u_2 + \alpha_3 u_3 . \end{aligned} \tag{5}$$

The components of u_1, u_2, u_3 are dominated by the terms in β_3 unless β_3 happened to be small. Loosely speaking almost all unit vectors when expressed in the form $\sum \alpha_i u_i$ have large components $\alpha_1, \alpha_2, \alpha_3$ which are roughly in the ratio -0.257, -2.529, 2.786.

Now let k be a random unit vector and consider the vector y defined by

$$y = k_1 \alpha_1 u_1 + k_2 \alpha_2 u_2 + k_3 \alpha_3 u_3 . \tag{6}$$

The coefficient $k_1\alpha_1$, $k_2\alpha_2$, $k_3\alpha_3$ will in general be of the order of magnitude 10^6 but they will not be in the special ratio of the α_i themselves. No cancellation will take place and $\|y\|$ will be of order 10^6 for almost all choices of k. This simple transformation will have the effect of changing the norm from 1 to 10^6.

However if we make a second transformation of the same kind to give z we have

$$z = k_1^2 \alpha_1 u_1 + k_2^2 \alpha_2 u_2 + k_3^2 \alpha_3 u_3 . \tag{7}$$

In general $\|z\|$ will be again of order 10^6. The first transformation will, for almost all k, give an increase in norm of 10^6; the second transformation will not increase the norm at all. Further transformations of the same type will not as a rule have any startling effect on the norm though if $k_1^r = k_2^r = k_3^r = k$ for some r then

at this stage the coefficients will again have the special ratio and the norm of the vector will again drop to unity. In the next step it will again display the enormous increase in norm! It is this simple mechanism which is at work in inverse iteration. *The phenomenon has nothing to do with rounding errors.*

The first step of inverse iteration is defined by

$$(A - \lambda I)\, y_1 = x_0 \quad \text{where} \quad \|x_0\|_2 = 1\,, \tag{8}$$

$$x_1 = y_1/\|y_1\|_2 \quad \text{where} \quad \|x_1\|_2 = 1\,. \tag{9}$$

The first residual r_1 corresponding to the normalized x_1 is given by

$$r_1 = (A - \lambda I)\, x_1 = x_0/\|y_1\|_2\,; \quad \|r_1\|_2 = 1/\|y_1\|_2\,. \tag{10}$$

Hence r_1 is small if $\|y_1\|_2$ is large. Suppose the value of λ we are using is associated with a group of ill-conditioned eigenvalues so that the vectors $u_1, u_2, \ldots, u_k$ are nearly linearly dependent. A random unit vector x_0 will be expressible in the form

$$x_0 = \sum_{i=1}^{n} \alpha_i u_i = \sum_{i=1}^{k} \alpha_i u_i + \sum_{i=k+1}^{n} \alpha_i u_i \tag{11}$$

and for almost all x_0 the $\alpha_1, \ldots, \alpha_k$ will be very large and specially related. y_1 is given by

$$y_1 = \sum_{i=1}^{k} \frac{\alpha_i}{(\lambda_i - \lambda)}\, u_i + \sum_{i=k+1}^{n} \frac{\alpha_i}{(\lambda_i - \lambda)}\, u_i\,. \tag{12}$$

Suppose λ is an approximation to λ_1; we may write $\lambda_i - \lambda = \varepsilon_i\ (i = 1, \ldots, k)$. where ε_1 will certainly be fairly small; it will not usually be pathologically small since, although we are assuming that it is an exact eigenvalue of $A + E$ where $\|E\|/\|A\|_2$ is of the order of the machine precision, we are interested in the case when λ_1 is ill-conditioned. Usually the other $\varepsilon_i\ (i = 2, \ldots, k)$ will also be moderately small since this is an ill-conditioned *group*. The remaining $\lambda_i - \lambda$ will not be at all small. Equation (12) may therefore be expressed in the form

$$y_1 = \frac{1}{\varepsilon_1}\left[\sum_{i=1}^{k} k_i \alpha_i u_i + \sum_{i=k+1}^{n} \frac{\alpha_i\, \varepsilon_i}{(\lambda_i - \lambda)}\, u_i\right] \tag{13}$$

where $k_i = \varepsilon_i/\varepsilon_1$. The expression $\sum_{i=1}^{k} \alpha_i u_i$ has therefore undergone just the type of transformation we discussed above. Although $\sum_{i=1}^{k} \alpha_i u_i$ is of order unity, $\sum_{i=1}^{k} k_i \alpha_i u_i$ will almost certainly be large; its size will be related to the nearness of the set of vectors $u_1, \ldots, u_k$ to linear dependence. $\|y_1\|$ will therefore be large for two reasons. First because $1/\varepsilon_1$ is fairly large and secondly because $\sum_{1}^{k} k_i \alpha_i u_i$ is large. The first residual will therefore be very small.

y_1 is now normalized to give x_1. The coefficients of u_i $(i = 1, \ldots, k)$ will not now be large and when we perform the second iteration the only factor contributing to $\|y_2\|_2$ being large is $1/\varepsilon_1$. Except in the rare cases when $[(\lambda_i - \lambda)/(\lambda_1 - \lambda)]^r = 1$ $(i = 1, \ldots, k)$ for some r we shall never again get a very large $\|y_r\|$. The subsequent residuals corresponding to λ and the successive x_i will be much larger than for λ and x_1. Of course with exact computation x_r would converge slowly to u_1 and hence ultimately x_r would give a negligible residual corresponding to λ_1 but not to λ! In fact

$$Au_1 - \lambda u_1 = Au_1 - \lambda_1 u_1 + (\lambda_1 - \lambda)\, u_1 = (\lambda_1 - \lambda)\, u_1$$

giving

$$\|Ax - \lambda u_1\| = \varepsilon_1 \tag{14}$$

confirming that the smallness of the later residual is related only to ε_1 and is not reinforced by the near linear dependence of the $u_1, \ldots, u_k$.

The condition $[(\lambda_i - \lambda)/(\lambda_1 - \lambda)]^r = 1$ discussed above might seem to be so improbable as to be not worth discussing. This is not quite true. When A has a non-linear divisor such a situation usually arises when the eigenvector is found via a transformation of A which has involved rounding errors. If the original A has a cubic divisor for example one usually finds that the residuals are pathologically small every third iteration.

Rounding errors in general affect the above arguments only very slightly in spite of the fact that $(A - \lambda I)$ is almost singular. This is because we obtain exact solutions of $(A + F - \lambda I)\, y = x$ when we solve $(A - \lambda I)\, y = x$, where $\|F\|/\|A\|$ is of machine procision. However when we include rounding errors there is no possibility of the x_r tending steadily to u_1 since at each stage we effectively iterate with some $(A + F - \lambda I)^{-1}$ and the F is different each time. In the ill-conditioned case the first iteration is the only one giving a related x and λ.

References

[1] FADDEEV, D. K., FADDEEVA, V. N.: Vychislitel'nye Metody Lineinoy Algebry. Gos. Izdat. Fiz.-Mat. Lit. Moscow (1965).

[2] WILKINSON, J. H.: The Algebraic Eigenvalue Problem. Oxford University Press, London (1965).

[3] WILKINSON, J. H.: Inverse Iteration in Theory and in Practice. Symposia Mathematica. Monograf. pp. 361—379, Bologna (1972).

[4] WILKINSON, J. H., REINSCH, C. H.: Handbook for Automatic Computation. Vol. II, Linear Algebra. Springer-Verlag, Berlin (1971).