AN OPTIMIZATION BASED METHOD FOR DETERMINING EIGENPAIRS OF LARGE REAL MATRICES

ANDRÁS KOCSOR, JÓZSEF DOMBI, AND IMRE BÁLINT

Abstract. The determination of the eigenpairs of real matrices can be treated as a local optimization problem. Suitable non-negative functions are constructed with coinciding local and global minima, which are located at the points defined by the eigenvectors of the underlying matrix. Some properties of these eigenvector functions are investigated and proved.

1. Introduction

The determination of the eigenvectors and eigenvalues of large real matrices is of considerable importance in various fields of science and technology. The machinery for the solution of the problem has been worked out quite well, and we do not attempt to give an overview of the relevant literature. The methods have been devised either for determining all eigenvectors of the matrix simultaneously or one after the other. Some problems furnish large matrices whose sizes are beyond the scope of simultaneous determination of all eigenvectors and eigenvalues, but the problem may be such as to only require a few of the eigenpairs.

The proposed novel algorithm is of the iterative class type. Non-negative, homogeneous functions have been defined with coinciding local and global optima, which are located exactly at the points defined by the eigenvectors of the underlying matrix. Hence the eigenvectors of the matrix can be found using well-behaved optimization algorithms as the minima of these eigenvector function. A unique property of the algorithm is that the convergence for a selected eigenvector does not depend on the magnitude of the associated eigenvalue as is generally the case with the iterative methods [2,4,9].

Throughout this report discussion is restricted to real eigenvectors of real matrices, but it should be emphasized that non-symmetrical matrices with complex eigenpairs and complex matrices are just as applicable using our algorithm. Throughout the whole paper the $\|\ \|$ notation will be used for the Euclidean norm, while ∇f denotes the gradient of a function $f : \mathbb{R}^n \to \mathbb{R}$ and $H(f)$ denotes its Hessian. The structure of the paper is as follows. Section 2 deals with the definition and some properties of the eigenvector-functions, Section 3 is devoted to optimization properties and error bounds, while Section 4 contains numerical results and an example for optimization with closely packed eigenvalues. Section 5 rounds off with a few concluding remarks.

2. The eigenvector functions

In this section non-negative, homogeneous function are defined with the property that their local and global minima coincide with eigenvectors of a real matrix. So using a suitable optimization method one can obtain the eigenvectors one after the other (employing a deflation technique).

Definition 2.1. Given an $n \times n$ matrix $A = [a_{ij}]$ and a real normalization parameter $0 \le \omega \le 1$, the functions $f_A^{(\omega)}, g_A^{(\omega)} : \mathbb{R}^n \setminus \{\mathbf{0}\} \to \mathbb{R}$ are defined by

$$f_A^{(\omega)}(\mathbf{x}) := \frac{\mathbf{x}^\top \mathbf{x}(A\mathbf{x})^\top (A\mathbf{x}) - (\mathbf{x}^\top A\mathbf{x})^2}{||\mathbf{x}||^{4\omega}}, \quad g_A^{(\omega)}(\mathbf{x}) := \frac{\sqrt{\mathbf{x}^\top \mathbf{x}}\sqrt{(A\mathbf{x})^\top (A\mathbf{x})} \pm \mathbf{x}^\top A\mathbf{x}}{||\mathbf{x}||^{2\omega}}. \tag{2.1}$$

In $g_A^{(\omega)}(\mathbf{x})$, $\pm$ accounts for the sign of the eigenvalues, choosing the negative for positive eigenvalues and the positive for the negative ones. The following lemma refers to the basic properties of the eigenvector functions: *homogeneity*, *non-negativity* and *proportionality*.

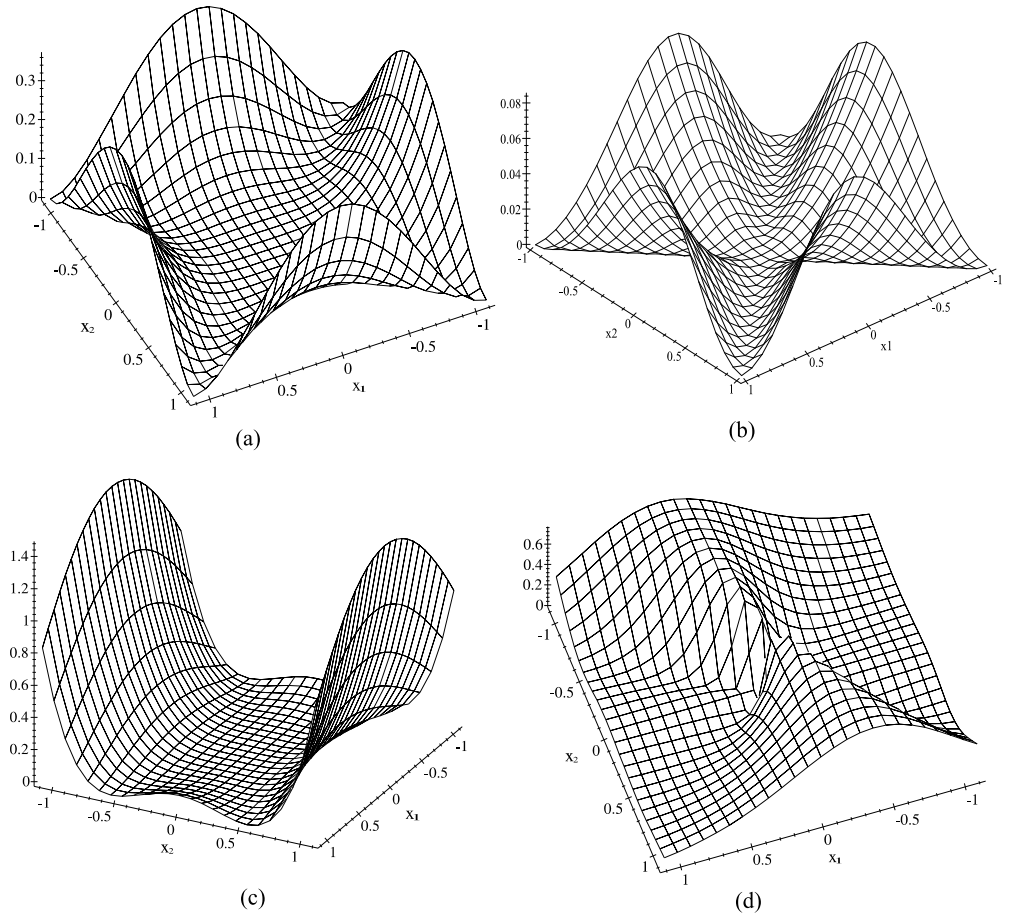

FIGURE 2.1. The graphs of $f^{(0)}_{A_0}(\mathbf{x})$, $g^{(0)}_{A_0}(\mathbf{x})$, $f^{(0)}_{A_1}(\mathbf{x})$ and $f^{(1)}_{A_1}(\mathbf{x})$.

Lemma 2.2. *Function $f^{(\omega)}_A(\mathbf{x})$ and $g^{(\omega)}_A(\mathbf{x})$, $0 \le \omega \le 1$ exhibit the following properties:*

Homogeneity. *Both function are homogeneous of degree $(4-4\omega)$ for f and $(2-2\omega)$ for g:*
$f^{(\omega)}_A(k\mathbf{x}) = k^{4-4\omega} f^{(\omega)}_A(\mathbf{x}), \qquad g^{(\omega)}_A(k\mathbf{x}) = k^{2-2\omega} g^{(\omega)}_A(\mathbf{x}), \qquad k \in \mathbb{R}.$

Non-negativity. $f^{(\omega)}_A(\mathbf{x}) \ge 0$ $(g^{(\omega)}_A(\mathbf{x}) \ge 0)$.

Proportionality. $f^{(\omega)}_A(\mathbf{x}) = 0$ $(g^{(\omega)}_A(\mathbf{x}) = 0)$ *if and only if* $\mathbf{x}$ *and* $A\mathbf{x}$ *are proportional.*

Proof. The proof of homogeneity is straightforward, while non-negativity and proportionality follow from applying the Cauchy-Schwartz inequality to the vectors $\mathbf{x}$ and $A\mathbf{x}$. Equality holds if and only if the vectors are proportional. □

Figure 2.1 shows the graphs of $f^{(0)}_{A_0}(\mathbf{x})$, $g^{(0)}_{A_0}(\mathbf{x})$, $f^{(0)}_{A_1}(\mathbf{x})$ and $f^{(1)}_{A_1}(\mathbf{x})$ associated with the following 2×2 symmetric A_0 and non-symmetric A_1 matrices

$$A_0 = \begin{pmatrix} \frac{3}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{3}{2} \end{pmatrix}, \quad A_1 = \begin{pmatrix} \frac{3}{2} & 1 \\ \frac{1}{4} & \frac{3}{2} \end{pmatrix}.$$

3. Optimization Properties of Eigenvector Functions

The following lemma describes the relationship between the gradient of eigenvector functions and their zero values.

Lemma 3.1. *Function $f^{(\omega)}_A(\mathbf{x})$ and $g^{(\omega)}_A(\mathbf{x})$, $0 \le \omega \le 1$ have the following properties:*

i. $\nabla f^{(\omega)}_A(\mathbf{x}) = \mathbf{0} \Rightarrow f^{(\omega)}_A(\mathbf{x}) = 0$ $(\nabla g^{(\omega)}_A(\mathbf{x}) = \mathbf{0} \Rightarrow g^{(\omega)}_A(\mathbf{x}) = 0)$.

ii. $\nabla f^{(\omega)}_A(\mathbf{x}) = \mathbf{0} \Leftarrow f^{(\omega)}_A(\mathbf{x}) = 0$ $(\nabla g^{(\omega)}_A(\mathbf{x}) = \mathbf{0} \Leftarrow g^{(\omega)}_A(\mathbf{x}) = 0)$.

Proof.

i. Since $f_A^{(\omega)}(\mathbf{x})$ is a homogeneous function of degree $4-4\omega$, Euler's theorem ensures that

$$\mathbf{x}^\top \nabla f_A^{(\omega)}(\mathbf{x}) = (4-4\omega) f_A^{(\omega)}(\mathbf{x}) \Longrightarrow \frac{\mathbf{x}^\top \nabla f_A^{(\omega)}(\mathbf{x})}{(4-4\omega)} = f_A^{(\omega)}(\mathbf{x}).$$

If $4-4\omega \neq 0$ the statement follows immediately, while for $\omega = 1$ L'Hospital's rule can be applied. The proof is analogous to $g_A^{(\omega)}(\mathbf{x})$.

ii. This inference follows from the nonnegative and continuously differentiable character of $f_A^{(\omega)}(\mathbf{x})$ and $g_A^{(\omega)}(\mathbf{x})$.

□

The characterization of the extremal points of eigenvector functions on the unit sphere is important. In the remaining part of this section the mathematical statements will refer to $f_A^{(0)}(\mathbf{x})$ assigned to a real symmetric matrix. However, much the same applies for general ω's ($\omega \neq 0$) as well as for $g_A^{(\omega)}$ and a similar strategy can be used for non-symmetric matrices too.

Noting Lemma 3.2, the following two theorems (3.3 and 3.4) ensure that, on the unit sphere, no local minima occur other than those associated with the eigenvectors. Moreover the peaks which separate the troughs of eigenvectors are at the bisectrices of any two eigenvectors. Now let the eigenvalues of the symmetric matrix A be denoted by $\lambda_1, \lambda_2, \ldots, \lambda_n$, and the associated orthonormal eigenvectors be denoted by $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$.

Lemma 3.2. *For the symmetric matrix A and real vector $\mathbf{x} = c_1\mathbf{u_1} + \cdots + c_n\mathbf{u_n}$,*

$$\begin{array}{lll}
i. & f_A^{(0)}(\mathbf{x}) & = \sum_{i=1}^n \sum_{j=1}^n (\lambda_i - \lambda_j)^2 c_i^2 c_j^2 \\
ii. & \nabla(f_A^{(0)}(\mathbf{x})) & = 2(\mathbf{x}^\top \mathbf{x} A^2\mathbf{x} + \mathbf{x}(A\mathbf{x})^\top A\mathbf{x} - 2\mathbf{x}^\top A\mathbf{x}(A\mathbf{x})) \\
iii. & \nabla(f_A^{(0)}(\mathbf{x})) & = 2\sum_{i=1}^n \sum_{j=1}^n (\lambda_j - \lambda_i)^2 \mathbf{u_i} c_i c_j^2 \\
iv. & H(f_A^{(0)}(\mathbf{x})) & = 2\mathbf{x}^\top\mathbf{x}A^2 + 4\mathbf{x}\mathbf{x}^\top A^2 + 4A^2\mathbf{x}\mathbf{x}^\top + 2I(A\mathbf{x})^\top A\mathbf{x} - 4\mathbf{x}^\top A\mathbf{x}A - 8A\mathbf{x}(A\mathbf{x})^\top \\
v. & H(f_A^{(0)}(\mathbf{x})) & = \sum_{i=1}^n \sum_{j=1}^n (\lambda_i - \lambda_j)^2 \left(2c_j^2 \mathbf{u_i}\mathbf{u_i}^\top + 4c_i c_j \mathbf{u_i}\mathbf{u_j}^\top\right)
\end{array}$$

Proof.

i. The first statement follows from a direct evaluation and subsequent rearrangement of the terms,

$$f_A^{(0)}(\mathbf{x}) = \mathbf{x}^\top\mathbf{x}(A\mathbf{x})^\top(A\mathbf{x}) - (\mathbf{x}^\top A\mathbf{x})^2 = \left(\sum_{i=1}^n c_i^2\right)\left(\sum_{j=1}^n \lambda_j^2 c_j^2\right) - \left(\sum_{i=1}^n \lambda_i c_i^2\right)^2 =$$

$$= \sum_{i=1}^n \sum_{j=1}^n (\lambda_i - \lambda_j)^2 c_i^2 c_j^2.$$

ii. From direct evaluation:

$$\begin{array}{rcl}
\nabla f_A^{(0)}(\mathbf{x}) & = & \nabla[\mathbf{x}^\top\mathbf{x}(A\mathbf{x})^\top(A\mathbf{x})] - \nabla(\mathbf{x}^\top A\mathbf{x})^2 \\
\nabla[\mathbf{x}^\top\mathbf{x}(A\mathbf{x})^\top(A\mathbf{x})] & = & 2(\mathbf{x}^\top\mathbf{x}A^\top(A\mathbf{x}) + \mathbf{x}(A\mathbf{x})^\top(A\mathbf{x})) \\
\nabla(\mathbf{x}^\top A\mathbf{x})^2 & = & 2(\mathbf{x}^\top A\mathbf{x})(A^\top\mathbf{x} + A\mathbf{x}).
\end{array}$$

iii. Direct evaluation and subsequent rearrangement yields,

$$1/2\nabla(f_A^{(0)}(\mathbf{x})) = \left(\mathbf{x}^\top\mathbf{x}\right)\left(A^2\mathbf{x}\right) + (\mathbf{x})\left((A\mathbf{x})^\top A\mathbf{x}\right) - 2\left(\mathbf{x}^\top A\mathbf{x}\right)(A\mathbf{x}) =$$

$$= \left(\sum_{j=1}^n c_j^2\right)\left(\sum_{i=1}^n \lambda_i^2 c_i \mathbf{u_i}\right) + \left(\sum_{i=1}^n c_i \mathbf{u_i}\right)\left(\sum_{j=1}^n \lambda_j^2 c_j^2\right) - 2\left(\sum_{j=1}^n \lambda_j c_j^2\right)\left(\sum_{i=1}^n \lambda_i c_i \mathbf{u_i}\right) =$$

$$= \sum_{i=1}^n \sum_{j=1}^n (\lambda_j - \lambda_i)^2 \mathbf{u_i} c_i c_j^2.$$

iv. To avoid unnecessary technicalities, only two steps of the evaluation are presented:

$$\begin{aligned} H(\mathbf{x}^\top\mathbf{x}(A\mathbf{x})^\top A\mathbf{x}) &= 2\mathbf{x}^\top\mathbf{x}A^2 + 4\mathbf{x}\mathbf{x}^\top A^2 + 4A^2\mathbf{x}\mathbf{x}^\top + 2I(A\mathbf{x})^\top A\mathbf{x} \\ H((\mathbf{x}^\top A\mathbf{x})^2) &= 4\mathbf{x}^\top A\mathbf{x}A + 8A\mathbf{x}(A\mathbf{x})^\top. \end{aligned}$$

v. The proof is obtained by simple algebraic rearrangement.

□

Theorem 3.3. *For a symmetric, non-singular matrix A with pairwise different eigenvalues, $\nabla f_A^{(0)}(\mathbf{x})$ and $\mathbf{x}$ are linearly dependent, if and only if, i or ii is valid.*

i. $\mathbf{x}$ *is in the linear space of a single eigenvector:* $\mathbf{x} \in \{\, c\,\mathbf{u}_i : c \in \mathbb{R},\ 1 \le i \le n\}$.

ii. $\mathbf{x}$ *is the bisectrix of two eigenvectors:* $\mathbf{x} \in \{\, c\,(\mathbf{u}_i \pm \mathbf{u}_j) : c \in \mathbb{R}, 1 \le i, j \le n\}$.

Proof. First we prove that for a non-singular, symmetric matrix A with pairwise different eigenvalues, $\nabla f_A^{(0)}(\mathbf{x})$ and $\mathbf{x}$ are linearly dependent if $\mathbf{x}$ is in the linear space of two eigenvectors:

$$\mathbf{x} \in \{c_1\,\mathbf{u}_i \pm c_2\,\mathbf{u}_j : c_1, c_2 \in \mathbb{R},\ 1 \le i, j \le n\}.$$

From Lemma 3.2 *ii* the gradient is in the subspace of three vectors $\mathbf{x}$, $A\mathbf{x}$, $A^2\mathbf{x}$, so $Rank(\mathbf{x},\ A\mathbf{x},\ A^2\mathbf{x}) < 3$. In the space of eigenvectors this is of the form

$$Rank \begin{pmatrix} c_1 & \lambda_1 c_1 & \lambda_1^2 c_1 \\ \vdots & \vdots & \vdots \\ c_n & \lambda_n c_n & \lambda_n^2 c_n \end{pmatrix} < 3. \tag{3.1}$$

Proceeding indirectly, let us assume that $c_i c_j c_k \neq 0$ for a triplet of indices providing that $\mathbf{x} = c_1\mathbf{u_1} + \cdots + c_n\mathbf{u_n}$ with at least three non-zero components. This however, leads to a contradiction, since

$$\det \begin{pmatrix} c_i & \lambda_i c_i & \lambda_i^2 c_i \\ c_j & \lambda_j c_j & \lambda_j^2 c_j \\ c_k & \lambda_k c_k & \lambda_k^2 c_k \end{pmatrix} = c_i c_j c_k (\lambda_k - \lambda_i)(\lambda_k - \lambda_j)(\lambda_j - \lambda_i) \neq 0$$

and is of rank 3.

Now without restricting the generality it can be assumed that $\mathbf{x} = c_1\mathbf{u_1} + c_2\mathbf{u_2}$. An orthogonal vector $\mathbf{d}$ is of the form $\pm(-c_2\mathbf{u_1} + c_1\mathbf{u_2} + d_3\mathbf{u_3} + \cdots + d_n\mathbf{u_n})$, where $d_3, \cdots, d_n$ are free parameters. It is enough to check the fulfilment of equality $\mathbf{d}^\top\nabla f_A^{(0)}(\mathbf{x}) = 0$. From Lemma 3.2 *iii*

$$\mathbf{d}^\top\nabla f_A^{(0)}(\mathbf{x}) = 2\mathbf{d}^\top \sum_{i=1}^{2}\sum_{j=1}^{2} (\lambda_j - \lambda_i)^2 \mathbf{u_i} c_i c_j^2 = 2c_1c_2(c_1^2 - c_2^2)(\lambda_1 - \lambda_2)^2.$$

Since $\lambda_1 \neq \lambda_2$, if $c_1 = 0$ or $c_2 = 0$, the point is at an eigenvector (i); if $c_1^2 = c_2^2$, the bisectrix of two eigenvectors is obtained $(ii.)$. □

Theorem 3.4. *If matrix A is non-singular and symmetric with pairwise different eigenvalues then:*

i. $\mathbf{d}^\top H(f_A^{(0)}(\mathbf{x}))\mathbf{d} \ge 0$ *if* $\mathbf{d}\perp\mathbf{x}$ *and* $\mathbf{x}$ *is in the linear space of an eigenvector.*

ii. $\mathbf{d}^\top H(f_A^{(0)}(\mathbf{x}))\mathbf{d} \le 0$ *if* $\mathbf{x}$ *is the bisectrix of two eigenvectors,* $\mathbf{d}\perp\mathbf{x}$*, and* $\mathbf{d}$ *is in the linear space of the same two eigenvectors.*

iii. $\mathbf{d}^\top H(f_A^{(0)}(\mathbf{x}))\mathbf{d} \le 0$ *if* $\mathbf{x}$ *is the bisectrix of two eigenvectors,* $\mathbf{d}\perp\mathbf{x}$*, and* $\mathbf{d}$ *does not contain components in the linear space of the same two eigenvectors.*

iv. $\mathbf{x}^\top H(f_A^{(0)}(\mathbf{x}))\mathbf{x} = 0$ *if* $\mathbf{x}$ *is in the linear space of an eigenvector.*

v. $\mathbf{x}^\top H(f_A^{(0)}(\mathbf{x}))\mathbf{x} > 0$ *if* $\mathbf{x}$ *is not in the linear space of an eigenvector.*

Proof. $i - iii$ It is once again presumed that $\mathbf{x} = c_1\mathbf{u_1} + c_2\mathbf{u_2}$ and $\mathbf{d} = d_1\mathbf{u_1} + d_2\mathbf{u_2} + d_3\mathbf{u_3} + \cdots + d_n\mathbf{u_n}$. Taking into account Lemma 3.2 v,

$$\mathbf{d}^\top H(f_A^{(0)}(\mathbf{x}))\mathbf{d} =$$

$$= \mathbf{d}^\top \left[4(\lambda_1 - \lambda_2)^2 c_1 c_2 (\mathbf{u_1}\mathbf{u_2}^\top + \mathbf{u_2}\mathbf{u_1}^\top) + \sum_{i=1}^{n} 2\left((\lambda_i - \lambda_1)^2 c_1^2 + (\lambda_i - \lambda_2)^2 c_2^2\right) \mathbf{u_i}\mathbf{u_i}^\top \right] \mathbf{d} =$$

$$= 4(\lambda_1-\lambda_2)^2c_1c_2d_1d_2 + \sum_{i=1}^{n} 2d_i^2\left((\lambda_i-\lambda_1)^2c_1^2 + (\lambda_i-\lambda_2)^2c_2^2\right). \tag{3.2}$$

When the statements in points *i-iii* are considered, the special cases of (3.2) imply that

$$\begin{array}{lll} i. & \mathbf{d}^\top H(f_A^{(0)}(\mathbf{x}))\mathbf{d} \geq 0 & if \quad (c_1d_2=0)\, or\, (c_2d_1=0)\,. \\ ii. & \mathbf{d}^\top H(f_A^{(0)}(\mathbf{x}))\mathbf{d} \leq 0 & if \quad \left\{\begin{array}{l} c_1=-c_2=-d_1=d_2, \quad d_3\cdots d_n=0 \\ c_1=c_2=-d_1=d_2, \quad d_3\cdots d_n=0 \\ c_1=c_2=d_1=-d_2, \quad d_3\cdots d_n=0 \end{array}\right. \\ iii. & \mathbf{d}^\top H(f_A^{(0)}(\mathbf{x}))\mathbf{d} \geq 0 & if \quad \left(c_1^2=c_2^2,\ d_1d_2=0\right). \end{array}$$

iv − v From points *i* and *v* of Lemma 3.2

$$\mathbf{x}^\top H(f_A^{(0)}(\mathbf{x}))\mathbf{x} = 6\sum_{i=1}^{n}\sum_{j=1}^{n}(\lambda_i-\lambda_j)^2c_i^2c_j^2 = 6f_A^{(0)}(\mathbf{x}),$$

which together with Lemma 2.2 (non-negativity and proportionality) yield the required proof.

□

By approximating the true eigenpairs $(\mathbf{u},\lambda)$ of a given matrix with $(\mathbf{x},\nu)$, various measures can be used for the pairwise distances of the true and approximate eigenvalues as well as those of true and approximate eigenvectors. The following lemma gives an error bound for the approximation of eigenvalues. It will be shown in the proof that the widely-used norm $\|A\mathbf{x}-\mathbf{x}\nu\|/\|\mathbf{x}\|$ [7,8] is closely related to $f_A^{(0)}(\mathbf{x})$ and the statement follows immediately from Wilkinson's result.

Lemma 3.5. *With any non-zero* $\mathbf{x}\in\mathbb{R}^{\mathbf{n}}$ *and symmetric matrix A, if*

$$f_A^{(0)}(\mathbf{x})^{1/2}/||\mathbf{x}||^2 \ \leq\ \epsilon, \tag{3.3}$$

an eigenvalue λ *necessarily exists which satisfies* $|\lambda-\sigma| \leq \epsilon$, *where* $\sigma = \mathbf{x}^\top A\mathbf{x}/\mathbf{x}^\top\mathbf{x}$ *is the Rayleigh-quotient and* ϵ *is a suitable bound.*

Proof. Firstly, for any non-zero $\mathbf{x}\in\mathbb{R}^{\mathbf{n}}$ and for any matrix A,

$$f_A^{(0)}(\mathbf{x})^{1/2}/||\mathbf{x}||^2 \ = \|A\mathbf{x}-\mathbf{x}\sigma\|/\|\mathbf{x}\|. \tag{3.4}$$

Direct computation yields the following sequence of equalities that proves this statement:

$$\|A\mathbf{x}-\mathbf{x}\sigma\|^2(\mathbf{x}^\top\mathbf{x})^2 = \left\|A\mathbf{x}-\mathbf{x}(\mathbf{x}^\top A\mathbf{x}/\mathbf{x}^\top\mathbf{x})\right\|^2(\mathbf{x}^\top\mathbf{x})^2 =$$

$$= \mathbf{x}^\top\mathbf{x}((\mathbf{x}^\top\mathbf{x})(A\mathbf{x})^\top(A\mathbf{x}) - (\mathbf{x}^\top A^\top\mathbf{x})(\mathbf{x}^\top A\mathbf{x})) = \mathbf{x}^\top\mathbf{x}f_A^{(0)}(\mathbf{x}).$$

Secondly, from Wilkinson's result, for any non-zero $\mathbf{x}\in\mathbb{R}^{\mathbf{n}}$ and any scalar ν, there is an eigenvalue λ of a symmetric matrix A which satisfies the following inequality

$$|\lambda-\nu| \leq \|A\mathbf{x}-\mathbf{x}\nu\|/\|\mathbf{x}\|. \tag{3.5}$$

Taking into account the first statement and substituting ν with the Rayleigh-quotient into the second, we obtain the above proposition. □

The following lemma furnishes a bound for the approximation of the eigenvectors. Let the eigenvalues of the symmetric matrix A be denoted by $\lambda_1,\lambda_2,\ldots,\lambda_n$, the associated normalized eigenvectors be $\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_n$, and the angles between the eigenvectors and vector $\mathbf{x}$ be $\alpha_1,\alpha_2,\ldots,\alpha_n$. The least upper bound of the angle between the eigenvector $\mathbf{u}_m$ (which is associated with the best approximated eigenvalue λ_m by Lemma 3.5) and $\mathbf{x}$ provides a measure for the accuracy of approximation.

Lemma 3.6. *For an arbitrary non-zero* $\mathbf{x}\in\mathbb{R}^n$ *and real, symmetric matrix A,*

$$f_A^{(0)}(\mathbf{x})^{1/2}/||\mathbf{x}||^2 \ \leq\ \epsilon \ \implies\ \sin^2\alpha_m \leq\ \epsilon/|\lambda_m-\mu_m|, \tag{3.6}$$

where $\lambda_m \neq 0$ is the eigenvalue nearest to σ (Rayleigh quotient) and μ_m is the convex linear combination of the complementary part of the spectrum,

$$\mu_m = \frac{\sum_{j,\ j\neq m} \lambda_j \cos^2 \alpha_j}{\sum_{j,\ j\neq m} \cos^2 \alpha_j}. \tag{3.7}$$

Proof. Taking into account $|\lambda_m - \sigma| = \min_i |\lambda_i - \sigma|$ and lemma 3.5, the bound $|\lambda_m - \sigma| \leq \epsilon$ and $\cos \alpha_j = \mathbf{u}_j^\top \frac{\mathbf{x}}{\|\mathbf{x}\|}$,

$$|\lambda_m - \sigma| = \left| \lambda_m - \left(\left(\frac{\mathbf{x}}{\|\mathbf{x}\|} \right)^\top \sum_j \lambda_j \mathbf{u}_j \mathbf{u}_j^\top \frac{\mathbf{x}}{\|\mathbf{x}\|} \right) \right| = \tag{3.8}$$

$$= \left| \lambda_m - \lambda_m \cos^2 \alpha_m - \mu_m \sum_{j,\ j\neq m} \cos^2 \alpha_j \right| = \left| (\lambda_m - \mu_m) \sin^2 \alpha_m \right| \leq \epsilon.$$

□

The essence of the lemma is that the accuracy of the approximation of an eigenvector depends on the bound ϵ and on the magnitude of the associated eigenvalue λ_m. However, it depends on the 'effective degeneracy' of the complementary part of the spectrum as well. This effective degeneracy is encoded in μ_m, which depends on the eigenvalues as well as on the relative position of $\mathbf{x}$ with respect to the eigenvectors. This relationship makes the error estimation for the eigenvectors fundamentally different from the error estimation of the approximation of eigenvalues. If the spectrum is near-degenerate, or the position vector $\mathbf{x}$ (and consequently $\cos^2(\alpha_j)$) makes it effectively near-degenerate, the bound loses its predictive power. If an estimation for $|\lambda_m - \mu_m|$ is available using other information about the spectrum, the relation offers a powerful tool for estimating a bound for the accuracy of actual calculations.

4. Numerical Results

The following results refer to numerical performance tests computed on symmetric and non-symmetric matrices of various sizes with uniformly distributed random elements in the interval $[-1, 1]$. As mentioned at the beginning, non-symmetrical and complex matrices have not been dealt with in this report, but an application for the non-symmetric case is given here for illustrative purposes. For optimization of the eigenvector functions the BFGS [1,3,6] (for huge matrices the 'limited memory' L-BFGS version is recommended) algorithm was used, which was started with a random trial vector and it was terminated by the condition $\|A\mathbf{x} - \mathbf{x}\sigma\|/\|\mathbf{x}\| < \epsilon$, where σ is the Rayleigh-quotient and ϵ is a suitable bound. Taking into account that the zero-vector is also an optimum point of $f_A^{(0)}$ (and $g_A^{(0)}$) an appropriate normalization technique was used with various normalization parameters ω.

4.1. Numerical result for the symmetric case.

The rows in the tables are arranged according to matrix size and the columns to the accuracies referred. Table elements are the average number of iteration steps necessary to reach the accuracies obtained from averaging over 100 random matrices. Tables 1 ($\omega = 0$), 2 ($\omega = 0.25$), 3 ($\omega = 0.5$) and 4 ($\omega = 0.75$) refer to $f_A^{(\omega)}(\mathbf{x})$, while table 5 refers to $g_A^{(0)}(\mathbf{x})$.

SYMMETRIC	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
n=25	5	27	37	39	40	41	42
n=50	7	53	77	83	85	85	86
n=100	18	102	141	162	163	163	164
n=200	33	140	209	222	227	230	232

Table 1. Convergence data for $f_A^{(\omega)}(\mathbf{x}),\ \omega = 0$

SYMMETRIC	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
n=25	5	21	34	39	41	43	43
n=50	13	45	58	62	65	67	68
n=100	24	80	119	132	138	141	143
n=200	41	131	220	249	261	268	269

TABLE 2. Convergence data for $f_A^{(\omega)}(\mathbf{x}),\ \omega = 0.25$

SYMMETRIC	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
n=25	4	17	25	28	29	30	31
n=50	7	30	50	52	54	55	56
n=100	10	46	92	100	103	104	105
n=200	13	69	171	196	200	203	204

TABLE 3. Convergence data for $f_A^{(\omega)}(\mathbf{x}),\ \omega = 0.5$

SYMMETRIC	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
n=25	4	16	25	28	30	31	32
n=50	5	27	47	52	55	57	59
n=100	7	42	94	105	110	112	114
n=200	10	66	171	197	203	207	210

TABLE 4. Convergence data for $f_A^{(\omega)}(\mathbf{x}),\ \omega = 0.75$

SYMMETRIC	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
n=25	3	17	25	27	29	30	30
n=50	5	31	47	52	54	55	56
n=100	8	55	93	102	105	107	107
n=200	12	92	181	200	206	208	209

TABLE 5. Convergence data for $g_A^{(0)}(\mathbf{x})$

4.2. **Numerical results for the non-symmetric case.** Although with symmetric matrices the eigenvector function $f_A^{(0)}$ rarely converges to the zero-vector, this occurs more frequently with non-symmetric matrices. To avoid this pitfall, when searching for minima with non-symmetric matrices the normalized eigenvector function $f_A^{(\omega)}$, $\omega > 0$ should be used. Tables 6 ($\omega = 0.25$), 7 ($\omega = 0.5$), 8 ($\omega = 0.75$) show the same specified data obtained with non-symmetric matrices. However, the non-normalized $g_A^{(0)}$ eigenvector function proved to be generally 'zero-vector safe' and highly effective in non-symmetric cases as well. Table 9 lists the data obtained using $g_A^{(0)}(\mathbf{x})$.

4.3. **Discussion of the results.** As the tables show, when increasing the size of the matrix the convergence speeds up in both the symmetric and non-symmetric cases. Also the degree of homogeneity affects the convergence speed, and $\omega = 0.5$ proved to be the best parameter value here. In both cases the average numbers of BFGS iteration steps were quite low. It was close to n for symmetric matrices, but about $1.5n$ for non-symmetric matrices (n is the dimension). When using the eigenvector function $g_A^{(0)}$ for symmetric matrices the results were found to be similar, but for non-symmetric matrices $g_A^{(0)}$ proved to be significantly better, with no clear indication of

NON-SYMMETRIC	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
n=25	3	31	64	71	72	74	74
n=50	12	57	108	113	116	118	119
n=100	15	71	168	186	192	196	198
n=200	25	110	274	318	322	324	325

TABLE 6. Convergence data for $f_A^{(\omega)}(\mathbf{x}),\ \omega = 0.25$

NON-SYMMETRIC	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
n=25	3	20	48	57	59	60	61
n=50	4	34	82	112	115	117	119
n=100	6	52	136	172	172	173	173
n=200	9	83	224	289	293	294	295

TABLE 7. Convergence data for $f_A^{(\omega)}(\mathbf{x}),\ \omega = 0.5$

NON-SYMMETRIC	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
n=25	3	20	42	48	51	52	53
n=50	4	33	69	80	81	84	85
n=100	6	52	128	147	150	154	154
n=200	9	83	238	288	300	305	307

TABLE 8. Convergence data for $f_A^{(\omega)}(\mathbf{x}),\ \omega = 0.75$

NON-SYMMETRIC	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
n=25	3	18	39	40	42	43	44
n=50	4	32	61	67	68	68	69
n=100	7	52	114	128	130	131	132
n=200	10	80	207	228	235	238	239

TABLE 9. Convergence data for $g_A^{(0)}(\mathbf{x})$

convergence to the zero-vector. Furthermore, the structure of $g_A^{(0)}$ is also simpler than that for $f_A^{(\omega)}$ with $\omega > 0$ so evaluation of the function value and the gradient required fewer operations. Overall, $g_A^{(0)}$ seemed to be the most efficient for the selective determination of eigenvectors.

4.4. **An example for closely packed eigenvalues.** Closely packed eigenvalues are not easily handled with most iterative methods. But the example of a 3×3 symmetric matrix illustrates the power of our algorithm in determining eigenvectors belonging to closely packed eigenvalues. The target matrix is constructed in factorized form with eigenvalues $1.002, 1.001, 1$,

$$A_3 = 1.002\mathbf{u_1}\mathbf{u_1}^\top + 1.001\mathbf{u_2}\mathbf{u_2}^\top + 1\mathbf{u_3}\mathbf{u_3}^\top, \tag{4.1}$$

and the convergence data values for the optimization of $g_A^{(0)}$ are listed in Table 10. The first column depicts the various starting vectors, the second the eigenvalues, the third the necessary steps to reach the required accuracy in the approximation of the eigenvalue, and the fourth shows the accuracy of approximation of the eigenvalue. As can be seen, the closely packed eigenvalues did not cause any special problem in the searching procedure, which is a good feature of the method

$\mathbf{x_0}$	λ	*steps*	$accuracy(\lambda)$
$[\frac{2}{\sqrt{29}}, -\frac{3}{\sqrt{29}}, \frac{4}{\sqrt{29}}]$	1	7	10^{-11}
$[\frac{2}{\sqrt{29}}, \frac{3}{\sqrt{29}}, \frac{4}{\sqrt{29}}]$	1.001	7	10^{-11}
$[\frac{2}{\sqrt{29}}, \frac{3}{\sqrt{29}}, -\frac{4}{\sqrt{29}}]$	1.002	8	10^{-9}

TABLE 10. Convergence data for the approximation of eigenvectors belonging to closely packed eigenvalues

especially since it is known that the convergence speed of methods based on the QR transformation breaks down if the ratio of eigenvalues $|\lambda_i|/|\lambda_{i-1}|$ is close to 1.

5. Conclusions

This short paper presents a novel algorithm for determining eigenvectors and eigenvalues of large real matrices. Although complex matrices and vectors are not discussed here, the method can be applied to complex matrices or complex eigenpairs of non-symmetric matrices as well. Eigenvector functions have been defined with various degree of homogeneity. It was shown that the local optima of these functions are also global optima, which coincide with the eigenvectors of the underlying matrix. Since the eigenvector functions are well-behaved, the known optimization procedures [1,3,5,6] can efficiently determine their minima, as the numerical investigation using the BFGS algorithm has clearly shown. The selection of the approximated eigenvector is independent of the distribution of the eigenvalues. The algorithm also behaves well with closely packed eigenvalues, which is generally a headache for most methods. The procedure does not assume the storage of the whole matrix in the core and requires only one matrix-vector, vector-vector and some scalar multiplications per step. The mathematical statements presented provide bounds for the accuracies of the obtained approximate eigenvalues and eigenvectors.

References

[1] Bazaraa, M. S., Sherali, H. D., Shetty, C. M. Nonlinear Programming Theory and Algorithms, John Wiley & Sons, 1993.
[2] Francis, J.F.G., The QR Transformation-A Unitary Analogue to LR Transformation, I-II. *Computer Journal*, Vol. 4 (1961/62), 265-271, 332-345.
[3] Gill, P.E., Murray, W. and Pitfield, P. A., The implementation of Two Revised Quasi-Newton Algorithms for Unconstrained Optimization, *Report NAC-11*, National Physical Lab. (1972).
[4] Lánczos, C., An Iteration Method for the Solution of the Eigenvalue Problem of Linear Differential and Integral Operators, *J. Res. Nat. Bur. Standards*, Vol. 45 (1950), 255-282.
[5] Nazareth, J. L., Conjugate Gradient Methods Less Dependency on Conjugacy, *SIAM Review*, 28(4) (1986), 501-511.
[6] Nocedal, J., The Performance of Several Algorithms for Large Scale Unconstrained Optimization, in Large-Scale Numerical Optimization, Coleman, T. F. and Li, Y. (Eds.), SIAM, Philadelphia (1990), 138-151.
[7] Parlett, B., The Symmetric Eigenvalue Problem, Prentice-Hall, Englewodd Cliffs, NJ, 1980.
[8] Wilkinson, J. H., The Algebraic Eigenvalue Problem, Oxford University Press, Oxford, 1965.
[9] Young, D. M., Gregory, R. T., A Survey of Numerical Mathematics, I-II, Addison-Wesley, Reading, 1972.

Research Group on Artificial Intelligence University of Szeged,
Szeged, Hungary, H-6720, Aradi vrt. 1.
E-mail address: kocsor@inf.u-szeged.hu

Department of Applied Informatics, University of Szeged,
Szeged, Hungary, H-6720, Árpád tér 2.
E-mail address: dombi@inf.u-szeged.hu

Department of Theoretical Physics, University of Szeged,
Szeged, Hungary, H-6720, Tisza L. krt. 84-86.
Department of Pharmaceutical Analysis, University of Szeged,
Szeged, Hungary, H-6720, Somogyi u. 4.
Department of Natural Sciences, Dunaújváros Polytechnic,
Dunaújváros, H-2400, Táncsics M. u. 1.
E-mail address: ibalint@physx.u-szeged.hu