The Approximation of an Eigenvector by Ritzvectors

Paul Smit

January 11, 1995

Abstract

Eigenvalue algorithms belonging to the class of the Rayleigh-Ritz methods (Krylov-space methods for example) use 'projections' on subspaces to produce approximations to eigenvalues and eigenvectors of a matrix. This paper focuses on the eigenvectors. Two angles are important when considering an eigenvector: the angle between the eigenvector and the best approximating Ritzvector and the angle between the eigenvector and the subspace involved. It is studied how an upperbound for the first angle can be expressed in terms of the second one. This results in a theoretical expression for the case of two-dimensional subspaces and a conjecture for higher dimensional subspaces supported by numerical experiments.

1 Introduction

In many problems it is important to calculate eigenvalues and/or eigenvectors of a matrix. For this purpose a variety of algorithms exists. An important class of algorithms is based on the 'projection' of the matrix on a series of subspaces. Let $\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation. $\mathcal{V} \subset \mathbb{R}^n$ is a linear subspace of dimension k, $\mathcal{A}|_\mathcal{V} : \mathcal{V} \to \mathbb{R}^n$ is the restriction of $\mathcal{A}$ to $\mathcal{V}$ and $\mathcal{P}_\mathcal{V}$ the orthogonal projection onto $\mathcal{V}$. The eigenvalues of $\mathcal{P}_\mathcal{V}\mathcal{A}|_\mathcal{V} : \mathcal{V} \to \mathcal{V}$ are called the Ritzvalues of $\mathcal{A}$ with respect to $\mathcal{V}$ and the corresponding eigenvectors are called Ritzvectors. The Ritzvalues and -vectors can be considered as approximations to the eigenvalues and -vectors of $\mathcal{A}$. If A is a matrix representing $\mathcal{A}$ and V is a matrix whose columns are an orthonormal basis of $\mathcal{V}$ then the matrix of $\mathcal{P}_\mathcal{V}\mathcal{A}|_\mathcal{V}$ with respect to the columns of V is given by V^TAV. This method is known as the Rayleigh-Ritz Procedure.
A special series of subspaces is formed by the Krylov subspaces. The k-th Krylov subspace with respect to matrix A and vector $v \in \mathbb{R}^n$ is defined by:

$$K_k(A, v) = \text{span}\{v, Av, \ldots, A^{k-1}v\}$$

Many important eigenvalue algorithms are based on series of Krylov subspaces (see [5, 1]). For an increasing row of values k, the Ritzvalues form a series of approximations to the eigenvalues with very good 'convergence' properties in many cases. The Ritzvectors also 'converge' to the eigenvectors.

In this paper we will focus on the Ritzvectors. Suppose we are interested in an eigenvector x, corresponding to an eigenvalue λ. For the series of Krylov-spaces we have that the angle $\phi^{(k)}$ between an eigenvector x and $K_k(A, v)$ decreases when k increases. Bounds for $\phi^{(k)}$ can be given, for example in the case of a symmetric matrix we have:

$$\tan \phi^{(k)} = \min_{p \in \mathcal{P}_{k-1}^{\lambda}} \frac{||p(A)(I - \pi_x)v||}{||(I - \pi_x)v||} \tan \phi^{(1)}$$

where $\mathcal{P}_{k-1}^{\lambda}$ is the set of polynomials of degree $k-1$, having the value 1 in λ and π_x is the orthogonal projection onto x (see [2]). However, in the case of approximating an eigenvector we are not interested in $\phi^{(k)}$, but in the angle $\theta^{(k)}$ between x and the best approximating Ritzvector. So we need a relation between those two angles. Such a relation is known (see 3.3), but it uses more information than ϕ_k and the eigenvalues only.
Here we will look at other bounds and do not restrict ourselves to Krylov spaces. If a result is valid for general subspaces, it is also valid for Krylov spaces, but of course it is possible that better bounds can be derived for the latter ones. This paper is concerned with the following question.

Problem 1.1 Let A be a matrix, V a subspace, x an eigenvector of A, ϕ the angle between x and V and θ the angle between x and the best approximating Ritzvector. What is the largest possible value of θ, expressed in ϕ and the eigenvalues of A?

For general matrices this is a difficult problem, therefore this paper considers symmetric matrices only. If A is symmetric, it has an orthonormal basis of eigenvectors. The corresponding matrix with respect to that basis is a diagonal matrix. Therefore it is no restriction considering diagonal matrices only.
After the notation in section 2, some upperbounds are given in section 3. In section 4 theoretical and experimental results about subspaces of dimension two are presented. Section 5 is concerned with subspaces of higher dimension, but here only experimental results are available. It contains a conjecture which remains to be proved.

2 Notation

Let $n \in \mathbb{N}$, $n \geq 3$. A denotes a $n \times n$ real diagonal matrix with n distinct eigenvalues:

$$A = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n \end{pmatrix} \tag{1}$$

We are interested in the eigenvalue λ and the corresponding eigenvector x. Without loss of generality it may be assumed that $\lambda = \lambda_1$ and $x = e_1$. The rest of the eigenvalues is ordered ascending: $\lambda_2 < \ldots < \lambda_n$. Let $k \in \mathbb{N}$, $2 \leq k \leq n$. $V \subset \mathbb{R}^n$ is a subspace of dimension k and P_V is the orthogonal projection on V. $z_1, \ldots, z_k$ are the Ritzvectors of A with respect to V. They are chosen such that their first components are nonnegative. P_i denotes the orthogonal projection on z_i. ϕ denotes the angle between x and V and θ_i the angle between x and z_i. The Ritzvectors are ordered such that $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_k$, so z_1 is the Ritzvector which approximates x best. See figure 1 for a visualization of the case $k = 2$.

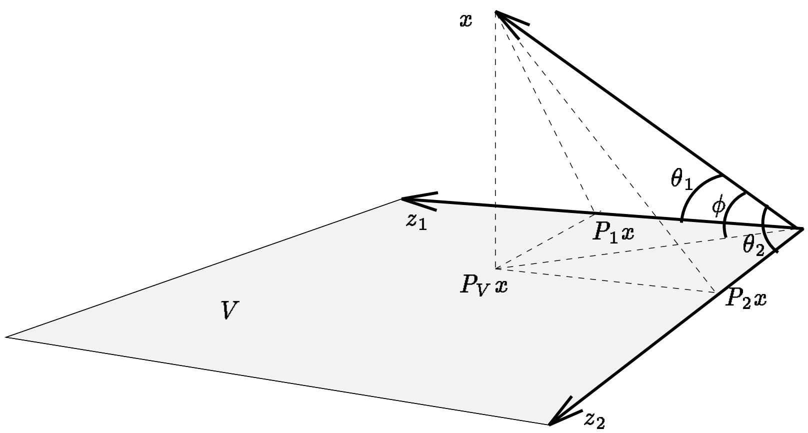

FIGURE 1: The approximation of an eigenvector by Ritzvectors.

Z is the matrix containing the Ritzvectors:

$$Z = [z_1 z_2 \cdots z_k] = \begin{pmatrix} z_{11} & \ldots & z_{1k} \\ z_{21} & \ldots & z_{2k} \\ \vdots & & \vdots \\ z_{n1} & \ldots & z_{nk} \end{pmatrix} \tag{2}$$

It should be kept in mind that ϕ and θ_i can be considered as functions of V, so they may be written as $\phi(V)$ and $\theta_i(V)$. In the next sections it occurs often that $\phi(V)$ is fixed to a certain value ϕ and V is varied over all subspaces leading to the same $\phi(V)$. So the expression $\min\limits_V E(V)$ should be interpreted as

$$\min\{E(V) | V \subset \mathbb{R}^n, \dim(V) = k, \phi(V) = \phi\}$$

The analogue holds for $\max\limits_V E(V)$. In particular expressions like $\max\limits_V \sin^2 \theta_1(V)$ will be considered. So problem 1.1 can be rewritten as:

Given $\phi \in \mathbb{R}$, $0 \leq \phi \leq \frac{\pi}{2}$, what is the value of $\max\limits_V \theta_1(V)$?

3 BOUNDS FOR θ_1

Some simple bounds can be proved without problems. The first lemma gives the relation between all the θ_i expressed in ϕ.

Lemma 3.1 $\sum_{i=1}^{k} \sin^2\theta_i = k - 1 + \sin^2\phi$

Proof:

$$\begin{aligned}
\cos^2\phi &= \|P_V x\|_2^2 = \|\sum_{i=1}^{k} P_i x\|_2^2 = \sum_{i=1}^{k} \|P_i x\|_2^2 = \sum_{i=1}^{k} \cos^2\theta_i \\
\sin^2\phi &= 1 - \cos^2\phi = 1 - \sum_{i=1}^{k}(1 - \sin^2\theta_i) = 1 - k + \sum_{i=1}^{k} \sin^2\theta_i
\end{aligned}$$

■

The next lemma gives a bound which will be sharp in a lot of cases.

Lemma 3.2 $\sin^2\phi \le \sin^2\theta_1 \le \dfrac{k-1+\sin^2\phi}{k}$

Proof: From lemma 3.1 we have:

$$\begin{aligned}
\cos^2\phi &= \sum_{i=1}^{k} \cos^2\theta_i \ge \cos^2\theta_1 \\
\sin^2\theta_1 &\ge \sin^2\phi \\
\sin^2\theta_1 &= \frac{\sum_{i=1}^{k} \sin^2\theta_1}{k} \le \frac{\sum_{i=1}^{k} \sin^2\theta_i}{k} = \frac{k-1+\sin^2\phi}{k}
\end{aligned}$$

■

Another upperbound for $\sin^2\theta_1$ is given by Parlett and Saad.

Theorem 3.3 Let $\gamma = \|P_V A(I - P_V)\|_2$ and δ the distance between λ and the set of Ritzvalues other than the one corresponding to z_1. Then:

$$\sin^2\theta_1 \le (1 + \frac{\gamma^2}{\delta^2}) \sin^2\phi$$

Proof: See [6], page 246 or [7], theorem 4.6. ■

It must be noticed that this bound depends not on ϕ only, but also on the projection P_V and the Ritzvalues. So this is not the answer to question 1.1 since it requires more information than just the angle ϕ. Later this bound is revisited and compared to another bound to be derived in the next section.
It is obvious that the sharpest bound possible for $\sin^2\theta_1$ which only depends on A and ϕ is $\max_V \sin^2\theta_1(V)$. This leads to an optimization problem which is formulated in the following theorem.

Theorem 3.4 Let $\mathcal{D}^k$ be the set of diagonal matrices of order k and Z the $n \times k$ matrix with elements (z_{ij}), then:

$$\max_V \sin^2\theta_1 = 1 - \min\{z_{11}^2 \mid \forall j : z_{11}^2 \geq z_{1j}^2, \sum_{j=1}^k z_{1j}^2 = \cos^2\phi, Z^T Z = I, Z^T A Z \in \mathcal{D}^k\}$$

Proof: Every subspace V corresponds to a matrix Z as defined in (2). Z is a matrix of Ritzvectors if and only if $Z^T Z = I$ and $Z^T A Z \in \mathcal{D}^k$.

$$\cos^2\phi = \|ZZ^T e_1\|_2^2 = \|Z^T e_1\|_2^2 = \sum_{j=1}^k z_{1j}^2$$

$$\cos^2\theta_i = \|z_i z_i^T e_1\|_2^2 = (z_i^T e_1)^2 = z_{1i}^2$$

The fact that $\forall j : \theta_1 \leq \theta_j$ corresponds to $\forall j : z_{11}^2 \geq z_{1j}^2$. This proves the theorem together with $\max_V \sin^2\theta_1 = 1 - \min_V \cos^2\theta_1$. ■

4 Subspaces of dimension two

4.1 A sharp upperbound

In theorem 3.4 a bound is given in the shape of a minimization problem which is not easy to solve. In the case of $k = 2$ however its solution can be formulated as a single expression in term of the eigenvalues of A and ϕ. To achieve this, the following lemma is needed.

Lemma 4.1 Let $n \in \mathbb{N}, n \geq 2, r, s \in \mathbb{R}, r, s \geq 0$, then:

$$\{(x_1 y_1, \ldots, x_n y_n) \in \mathbb{R}^n \mid \ \|x\|_2 = r, \|y\|_2 = s\} = \{z \in \mathbb{R}^n \mid \|z\|_1 \leq rs\}$$

Proof:

$\boxed{\subset}$ Let $x, y \in \mathbb{R}^n, \|x\|_2 = r, \|y\|_2 = s$ and $\forall i \in \{1, \ldots, n\} : z_i = x_i y_i$. Then:

$$\|z\|_1 = \sum_{i=1}^n |x_i y_i| \leq \|x\|_2 \|y\|_2 = rs$$

$\boxed{\supset}$ Let $z \in \mathbb{R}^n, \|z\|_1 \leq rs$. $z = \begin{pmatrix} \hat{z} \\ z_n \end{pmatrix}$ with $\hat{z} \in \mathbb{R}^{n-1}$ and $z_n \in \mathbb{R}$. Because:

$$|\ \|\hat{z}\|_1 \pm z_n\ | \leq \|z\|_1 \leq rs$$

$\alpha, \beta \in \mathbb{R}$ can be found such that

$$\begin{aligned} ||\hat{z}||_1 + z_n &= rs\cos\alpha \\ ||\hat{z}||_1 - z_n &= rs\cos\beta \end{aligned}$$

If $||\hat{z}||_1 = 0$, then $z_n = rs\cos\alpha$. Choose a random $\hat{x} \in \mathbb{R}^{n-1}$ satisfying $||\hat{x}||_2 = r\sin\alpha$ and let $\hat{y} \in \mathbb{R}^{n-1}$ be the zero vector. Let $x = \begin{pmatrix} \hat{x} \\ r\cos\alpha \end{pmatrix}$ and $y = \begin{pmatrix} \hat{y} \\ s \end{pmatrix}$, then $||x||_2 = r$, $||y||_2 = s$ and $\forall i \in \{1, \ldots, n\} : x_i y_i = z_i$.
If $||\hat{z}||_1 \neq 0$, let $\gamma = (\beta + \alpha)/2$ and $\delta = (\beta - \alpha)/2$.

$$\begin{aligned} ||\hat{z}||_1 &= rs(\cos\alpha + \cos\beta)/2 = rs\cos\gamma\cos\delta \\ z_n &= rs(\cos\alpha - \cos\beta)/2 = rs\sin\gamma\sin\delta \end{aligned}$$

Define for $1 \leq i \leq n-1$:

$$\begin{aligned} x_i &= \sqrt{\frac{r\cos\gamma}{s\cos\delta}}\sqrt{|z_i|} \\ y_i &= \sqrt{\frac{s\cos\delta}{r\cos\gamma}}\sqrt{|z_i|}\,\mathrm{sgn}(z_i) \end{aligned}$$

and:

$$\begin{aligned} x_n &= r\sin\gamma \\ y_n &= s\sin\delta \end{aligned}$$

then:

$$\begin{aligned} ||x||_2^2 &= \frac{r\cos\gamma}{s\cos\delta}||\hat{z}||_1 + x_n^2 = r^2\cos^2\gamma + r^2\sin^2\gamma = r^2 \\ ||y||_2^2 &= \frac{s\cos\delta}{r\cos\gamma}||\hat{z}||_1 + y_n^2 = s^2\cos^2\delta + s^2\sin^2\delta = s^2 \end{aligned}$$

and:

$$\begin{aligned} \forall i \in \{1, \ldots, n-1\} : x_i y_i &= |z_i|\,\mathrm{sgn}(z_i) = z_i \\ x_n y_n &= rs\sin\gamma\sin\delta = z_n \end{aligned}$$

■

Now the solution to the minimization problem can be given.

Theorem 4.2 Let $k = 2$ and $\kappa = \frac{(\lambda_n - \lambda_2)^2}{(\lambda_2 - \lambda_1)(\lambda_n - \lambda_1)}$. If $\lambda_1 < \lambda_2$ and $\sin^2\phi < \frac{\lambda_2 - \lambda_1}{\lambda_n - \lambda_1}$ or $\lambda_1 > \lambda_n$ and $\sin^2\phi < \frac{\lambda_1 - \lambda_n}{\lambda_1 - \lambda_2}$ then:

$$\max_V \sin^2\theta_1 = \frac{1}{2}\left(1 + \sin^2\phi - \sqrt{(1 - \sin^2\phi)^2 - \kappa\sin^2\phi}\right)$$

Otherwise:

$$\max_V \sin^2\theta_1 = \frac{1}{2}(1+\sin^2\phi)$$

Proof:
For the case of simplicity of notation we put: $c = \cos\phi$, $s = \sin\phi$, $c_i = z_{1i}$, $s_i = \sqrt{1-c_i^2}$. Let $S_1 \subset \mathbb{R}^{2n}$ be the set

$$S_1 = \{(z_{ij})_{1\le i\le n, 1\le j\le 2} | z_{11}^2 \ge z_{12}^2, z_{11}^2 + z_{12}^2 = c^2, Z^TZ = I, Z^TAZ \in \mathcal{D}^2\}$$

The we have according to theorem 3.4

$$\min_V \cos^2\theta_1 = \min\{c_1^2 | (z_{ij}) \in S_1\}$$

S_1 can be rewritten as:

$$\begin{aligned} S_1 &= \{(z_{ij}) | z_{11}^2 \ge c^2 - z_{11}^2, z_{11}^2 + z_{12}^2 = c^2, \sum_{i=1}^n z_{ij}^2 = 1, \\ &\quad \sum_{i=1}^n z_{i1}z_{i2} = 0, \sum_{i=1}^n \lambda_i z_{i1} z_{i2} = 0\} \\ &= \{(z_{ij}) | c_1^2 \ge \frac{c^2}{2}, c_1^2 + c_2^2 = c^2, \sum_{i=2}^n z_{ij}^2 = s_j^2, \\ &\quad \sum_{i=2}^n z_{i1}z_{i2} = -c_1c_2, \sum_{i=2}^n \lambda_i z_{i1}z_{i2} = -\lambda_1 c_1 c_2\} \end{aligned}$$

The constraints of this set are nonlinear. To linearize part of the problem the set $S_2 \subset \mathbb{R}^{3n-1}$ is introduced.

$$S_2 = \{((z_{ij})_{1\le i\le n, 1\le j\le 2}, (x_i)_{2\le i\le n}) | (z_{ij}) \in S_1, x_i = z_{i1}z_{i2}\}$$

The variables x_i do not occur in the objective function, so

$$\min\{c_1^2 | (z_{ij}) \in S_1\} = \min\{c_1^2 | ((z_{ij}), (x_i)) \in S_2\}$$

Then using lemma 4.1:

$$\begin{aligned} S_2 &= \{((z_{ij}), (x_i)) | c_1^2 \ge \frac{c^2}{2}, c_1^2 + c_2^2 = c^2, \sum_{i=2}^n x_i = -c_1c_2, \\ &\quad \sum_{i=2}^n \lambda_i x_i = -\lambda_1 c_1 c_2, x_i = z_{i1}z_{i2}, \sum_{i=2}^n z_{ij}^2 = s_j^2\} \\ &= \{((z_{ij}), (x_i)) | c_1^2 \ge \frac{c^2}{2}, c_1^2 + c_2^2 = c^2, \sum_{i=2}^n x_i = -c_1c_2, \\ &\quad \sum_{i=2}^n \lambda_i x_i = -\lambda_1 c_1 c_2, \sum_{i=2}^n |x_i| \le s_1 s_2\} \end{aligned}$$

The objective function and the constraints do no longer contain the z_{ij} for $i \geq 2$, so if $S_3 \subset \mathbb{R}^{n+1}$ is defined by

$$\begin{aligned} S_3 &= \{(c_1, c_2, (x_i)_{2\leq i\leq n}) | c_1^2 \geq \frac{c^2}{2}, c_1^2 + c_2^2 = c^2, \sum_{i=2}^{n} x_i = -c_1 c_2, \\ &\quad \sum_{i=2}^{n} \lambda_i x_i = -\lambda_1 c_1 c_2, \sum_{i=2}^{n} |x_i| \leq s_1 s_2\} \end{aligned}$$

Then

$$\min\{c_1^2 | ((z_{ij}), (x_i)) \in S_2\} = \min\{c_1^2 | (c_1, c_2, (x_i)) \in S_3\}$$

The constraints contain two equations which are linear in the x_i's, so two of them can be eliminated. For this purpose x_2 and x_n are selected.

$$\begin{aligned} S_3 &= \{(c_1, c_2, (x_i)) | c_1^2 \geq \frac{c^2}{2}, c_1^2 + c_2^2 = c^2, x_2 + x_n = -c_1 c_2 - \sum_{i=3}^{n-1} x_i, \\ &\quad \lambda_2 x_2 + \lambda_n x_n = -\lambda_1 c_1 c_2 - \sum_{i=3}^{n-1} \lambda_i x_i, |x_2| + |x_n| \leq s_1 s_2 - \sum_{i=3}^{n-1} |x_i|\} \end{aligned}$$

Define

$$f_i = \frac{\lambda_n + \lambda_2 - 2\lambda_i}{\lambda_n - \lambda_2}$$

then:

$$\begin{aligned} x_2 - x_n &= \frac{\lambda_n + \lambda_2}{\lambda_n - \lambda_2}(x_2 + x_n) - \frac{2}{\lambda_n - \lambda_2}(\lambda_2 x_2 + \lambda_n x_n) \\ &= \frac{\lambda_n + \lambda_2}{\lambda_n - \lambda_2}(-c_1 c_2 - \sum_{i=3}^{n-1} x_i) - \frac{2}{\lambda_n - \lambda_2}(-\lambda_1 c_1 c_2 - \sum_{i=3}^{n-1} \lambda_i x_i) \\ &= -c_1 c_2 f_1 - \sum_{i=3}^{n-1} x_i f_i \end{aligned}$$

Because

$$|a| + |b| \leq c \Leftrightarrow a + b \leq c \wedge -a - b \leq c \wedge a - b \leq c \wedge -a + b \leq c$$

we have that for the set $S_4 \subset \mathbb{R}^{n-1}$ defined by

$$\begin{aligned} S_4 &= \{(c_1, c_2, (x_i)_{3\leq i\leq n-1}) | c_1^2 \geq \frac{c^2}{2}, c_1^2 + c_2^2 = c^2, \\ &\quad -c_1 c_2 - \sum_{i=3}^{n-1} x_i \leq s_1 s_2 - \sum_{i=3}^{n-1} |x_i|, -c_1 c_2 f_1 - \sum_{i=3}^{n-1} x_i f_i \leq s_1 s_2 - \sum_{i=3}^{n-1} |x_i|, \\ &\quad c_1 c_2 + \sum_{i=3}^{n-1} x_i \leq s_1 s_2 - \sum_{i=3}^{n-1} |x_i|, c_1 c_2 f_1 + \sum_{i=3}^{n-1} x_i f_i \leq s_1 s_2 - \sum_{i=3}^{n-1} |x_i|\} \end{aligned}$$

the equality

$$\min\{c_1^2|(c_1,c_2,(x_i))\in S_3\} = \min\{c_1^2|(c_1,c_2,(x_i))\in S_4\}$$

holds.

$$\begin{aligned} S_4 &= \{(c_1,c_2,(x_i))|c_1^2 \ge \frac{c^2}{2}, c_1^2+c_2^2=c^2, \\ &\quad c_1c_2+s_1s_2 \ge \sum_{i=3}^{n-1}(|x_i|-x_i), -c_1c_2+s_1s_2 \ge \sum_{i=3}^{n-1}(|x_i|+x_i), \\ &\quad c_1c_2f_1+s_1s_2 \ge \sum_{i=3}^{n-1}(|x_i|-x_if_i), -c_1c_2f_1+s_1s_2 \ge \sum_{i=3}^{n-1}(|x_i|+x_if_i)\} \end{aligned}$$

If $S_5 \subset \mathbb{R}^2$ is the set

$$\begin{aligned} S_5 &= \{(c_1,c_2)|c_1^2 \ge \frac{c^2}{2}, c_1^2+c_2^2=c^2, c_1c_2+s_1s_2\ge 0, \\ &\quad -c_1c_2+s_1s_2\ge 0, c_1c_2f_1+s_1s_2\ge 0, -c_1c_2f_1+s_1s_2\ge 0\} \end{aligned}$$

then

$$\begin{aligned} S_4 &\supset \{(c_1,c_2,(x_i))\in S_4|\forall i\in\{3,\dots,n-1\}: x_i=0\} \\ &= \{(c_1,c_2,(x_i))|(c_1,c_2)\in S_5, \forall i\in\{3,\dots,n-1\}: x_i=0\} \end{aligned}$$

On the other hand

$$|x_i| \pm x_i \ge 0$$

and for $3 \le i \le n-1 : -1 \le f_i \le 1$, so

$$|x_i| \pm x_if_i \ge |x_i| - |x_if_i| \ge 0$$

leading to

$$S_4 \subset \{(c_1,c_2,(x_i))|(c_1,c_2)\in S_5\}$$

and

$$\min\{c_1^2|(c_1,c_2,(x_i))\in S_4\} = \min\{c_1^2|(c_1,c_2)\in S_5\}$$

eliminating the rest of the x_i. Furthermore two inequalities in the constraints of S_5 are trivial since

$$s_1s_2 = \sqrt{1-c_1^2}\sqrt{1-c_2^2} = \sqrt{1-c_1^2-c_2^2+c_1^2c_2^2} = \sqrt{s^2+c_1^2c_2^2} \ge |c_1c_2| \ge \pm c_1c_2$$

so

$$S_5 = \{(c_1,c_2)|c_1^2 \ge \frac{c^2}{2}, c_1^2+c_2^2=c^2, s_1s_2 \ge |c_1c_2f_1|\}$$

Now all variables x_i have disappeared from the constraints, so with $S_6 \subset \mathbb{R}^2$ given by

$$S_6 = \{(c_1, c_2) | c_1^2 \geq \frac{c^2}{2}, c_1^2 + c_2^2 = c^2, s_1 s_2 \geq |c_1 c_2 f_1|\}$$

we have

$$\min\{c_1^2 | (c_1, c_2, (x_i)) \in S_5\} = \min\{c_1^2 | (c_1, c_2) \in S_6\}$$

For the final reduction to a problem of one variable we use:

$$\begin{aligned}
s_1 s_2 \geq |c_1 c_2 f_1| \quad &\Leftrightarrow \quad s_1^2 s_2^2 \geq c_1^2 c_2^2 f_1^2 \Leftrightarrow s^2 + c_1^2 c_2^2 \geq c_1^2 c_2^2 (1 + 2\frac{\lambda_2 - \lambda_1}{\lambda_n - \lambda_2})^2 \\
&\Leftrightarrow \quad s^2 \geq 4 c_1^2 c_2^2 ((\frac{\lambda_2 - \lambda_1}{\lambda_n - \lambda_2})^2 + \frac{\lambda_2 - \lambda_1}{\lambda_n - \lambda_2}) \\
&\Leftrightarrow \quad s^2 \geq 4 c_1^2 c_2^2 \frac{(\lambda_2 - \lambda_1)(\lambda_n - \lambda_1)}{(\lambda_n - \lambda_2)^2} = 4\kappa^{-1} c_1^2 c_2^2 \\
&\Leftrightarrow \quad 4\kappa^{-1} c_1^2 (c_1^2 - c^2) + s^2 \geq 0 \Leftrightarrow 4\kappa^{-1} c_1^4 - 4\kappa^{-1} c^2 c_1^2 + s^2 \geq 0
\end{aligned}$$

Define $f(x) = 4\kappa^{-1} x^2 - 4\kappa^{-1} c^2 x + s^2$ and the set $S_7 \subset \mathbb{R}$ by

$$S_7 = \{x | \frac{c^2}{2} \leq x \leq c^2, f(x) \geq 0\}$$

then

$$\min\{c_1^2 | (c_1, c_2) \in S_6\} = \min S_7$$

This minimum is equal to $\frac{c^2}{2}$ if and only if $f(\frac{c^2}{2}) \geq 0$, so it is important to evaluate this value.

$$\begin{aligned}
f(\frac{c^2}{2}) \quad &= \quad \kappa^{-1} c^4 - 2\kappa^{-1} c^4 + s^2 = s^2 - \kappa^{-1} c^4 = 1 - c^2 - \kappa^{-1} c^4 \\
&= \quad -\kappa^{-1} (c^2 - \frac{\lambda_n - \lambda_2}{\lambda_1 - \lambda_2})(c^2 - \frac{\lambda_n - \lambda_2}{\lambda_n - \lambda_1})
\end{aligned}$$

We consider the following cases:

1. If $\lambda_2 < \lambda_1 < \lambda_n$ (λ_1 is not an extreme eigenvalue) then $\kappa^{-1} < 0$, so $\forall c : f(\frac{c^2}{2}) \geq 0$.

2. If $\lambda_1 < \lambda_2$ or $\lambda_1 > \lambda_n$ (λ_1 is an extreme eigenvalue) then $\kappa^{-1} > 0$.

- **2.1.** $\lambda_1 < \lambda_2$
  - **2.1.1.** If $c^2 \leq \frac{\lambda_n - \lambda_2}{\lambda_n - \lambda_1}$ then $f(\frac{c^2}{2}) \geq 0$.
  - **2.1.2.** If $c^2 > \frac{\lambda_n - \lambda_2}{\lambda_n - \lambda_1}$ then $f(\frac{c^2}{2}) < 0$.
- **2.2.** $\lambda_1 > \lambda_n$
  - **2.2.1.** If $c^2 \leq \frac{\lambda_n - \lambda_2}{\lambda_1 - \lambda_2}$ then $f(\frac{c^2}{2}) \geq 0$.
  - **2.2.2.** If $c^2 > \frac{\lambda_n - \lambda_2}{\lambda_1 - \lambda_2}$ then $f(\frac{c^2}{2}) < 0$.

The zeroes of f are given by

$$f(x)=0 \Leftrightarrow x=\frac{1}{2}(c^2 \pm \sqrt{c^4-\kappa s^2})=\frac{1}{2}(c^2 \pm \sqrt{-\kappa f(\frac{c^2}{2})})$$

If $\kappa^{-1}>0$ and $f(\frac{c^2}{2})<0$ then f has two real zeroes, situated respectively in the intervals $[0,\frac{c^2}{2}]$ and $[\frac{c^2}{2},c^2]$. In that case the minimum of S_7 is equal to the rightmost zero $\frac{1}{2}(c^2+\sqrt{c^4-\kappa s^2})$.
This leads to:

$$\min_V \cos^2\theta_1 = \begin{cases} \frac{c^2}{2} & \text{in the cases } \mathbf{1}\ ,\ \mathbf{2.1.1}\ ,\ \mathbf{2.2.1} \\ \frac{1}{2}(c^2+\sqrt{c^4-\kappa s^2}) & \text{in the cases } \mathbf{2.1.2}\ ,\ \mathbf{2.2.2} \end{cases}$$

Writing out the cosines in terms of sines proves the theorem. ■

The behaviour of the upperbound for small ϕ is described in the next corrolary.

Corrolary 4.3 If $\lambda_1<\lambda_2$ or $\lambda_1>\lambda_n$:

$$\max_V \sin^2\theta_1=(1+\frac{(\lambda_n-\lambda_2)^2}{4(\lambda_2-\lambda_1)(\lambda_n-\lambda_1)})\sin^2\phi+\mathcal{O}(\sin^4\phi)$$

Proof: Let κ, s, s_1 be as in theorem 4.2. For s small enough:

$$\begin{aligned}\max_V s_1^2 &= \frac{1}{2}(1+s^2-\sqrt{1-2s^2+\mathcal{O}(s^4)-\kappa s^2}) \\ &= \frac{1}{2}(1+s^2-(1-s^2-\frac{\kappa}{2}s^2+\mathcal{O}(s^4))) \\ &= \frac{1}{2}(2s^2+\frac{\kappa}{2}s^2+\mathcal{O}(s^4))=(1+\frac{\kappa}{4})s^2+\mathcal{O}(s^4)\end{aligned}$$

■

From theorem 4.2 a bound for $\sin^2\theta_1$ can be derived that is linear in $\sin^2\phi$.

Corrolary 4.4
If $\lambda_1<\lambda_2: \forall\phi: \sin^2\theta_1 \le (1+\frac{1}{2}\frac{\lambda_n-\lambda_2}{\lambda_2-\lambda_1})\sin^2\phi$
If $\lambda_1>\lambda_n: \forall\phi: \sin^2\theta_1 \le (1+\frac{1}{2}\frac{\lambda_n-\lambda_2}{\lambda_1-\lambda_n})\sin^2\phi$

Proof: If $\lambda_1<\lambda_2$, the line which goes through $(0,0)$ and $(\frac{\lambda_2-\lambda_1}{\lambda_n-\lambda_1},\frac{1}{2}(1+\frac{\lambda_2-\lambda_1}{\lambda_n-\lambda_1}))$ in the $(\sin^2\phi,\sin^2\theta_1)$-plane is an upperbound for $\sin^2\theta_1$:

$$\sin^2\theta_1 \le \frac{\lambda_n+\lambda_2-2\lambda_1}{2(\lambda_n-\lambda_1)}\cdot\frac{\lambda_n-\lambda_1}{\lambda_2-\lambda_1}\cdot\sin^2\phi=\frac{\lambda_n+\lambda_2-2\lambda_1}{2(\lambda_2-\lambda_1)}\cdot\sin^2\phi$$

If $\lambda_1>\lambda_n$, the line which goes through the points $(0,0)$ and $(\frac{\lambda_1-\lambda_n}{\lambda_1-\lambda_2},\frac{1}{2}(1+\frac{\lambda_1-\lambda_n}{\lambda_1-\lambda_2}))$ is an upperbound for $\sin^2\theta_1$:

$$\sin^2\theta_1 \leq \frac{2\lambda_1-\lambda_n-\lambda_2}{2(\lambda_1-\lambda_2)}\cdot\frac{\lambda_1-\lambda_2}{\lambda_1-\lambda_n}\cdot\sin^2\phi = \frac{2\lambda_1-\lambda_n-\lambda_2}{2(\lambda_1-\lambda_n)}\cdot\sin^2\phi$$

■

Example 4.5 If $n=5, k=2, \lambda_i = i, i\in\{1,\ldots,5\}$ then:

$$\sin^2\phi \leq \sin^2\theta_1 \leq \begin{cases} \frac{1}{2}(1+\sin^2\phi - \sqrt{(4-\sin^2\phi)(\frac{1}{4}-\sin^2\phi)}) & \text{if } \sin^2\phi < \frac{1}{4} \\ \frac{1}{2}(1+\sin^2\phi) & \text{if } \sin^2\phi \geq \frac{1}{4} \end{cases}$$

If $\sin^2\phi \ll 1$, then $\max\limits_V \sin^2\theta_1 \approx \dfrac{25}{16}\sin^2\phi$. $\forall\phi : \sin^2\theta_1 \leq \frac{5}{2}\sin^2\phi$. See also figure 2.

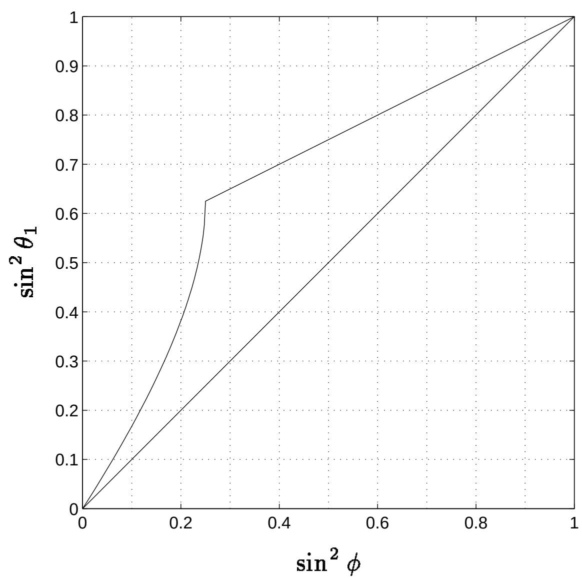

FIGURE 2: The set of points $\{(\sin^2\phi, \sin^2\theta_1) | \dim(V) = 2\}$ in the case $\lambda_i = i, i\in\{1,\ldots,5\}$.

4.2 Quality of the bound

Theorem 4.2 gives a bound for $\sin^2\theta_1$ and it is a sharp bound. This means that subspaces V exist for which $\sin^2\theta_1(V)$ is equal to the value of the bound, but the question remains how good it is on the average. It is also interesting to know how it compares to the bound in 3.3. We should keep in mind however that this latter bound uses more information than the one in 4.2, so it can be expected to be a better bound.

To check this some numerical experiments have been performed with the matrix $A = \text{diag}(1,2,3)$ ans $\lambda_1 = 1$. Random subspaces V have been generated and for each subspace the numbers $s^2 = \sin^2\phi$ and $s_1^2 = \sin^2\theta_1$ have been calculated. Now the bounds b_1 and b_2 from respectively 4.2 and 3.3 have been evaluated:

$$b_1 = \begin{cases} \frac{1}{2}(1+s^2-\sqrt{(1-s^2)^2-\kappa s^2}) & \text{if } s^2 < \frac{\lambda_2-\lambda_1}{\lambda_n-\lambda_1} \\ \frac{1}{2}(1+s^2) & \text{otherwise} \end{cases} \tag{3}$$

$$b_2 = (1+\frac{\gamma^2}{\delta^2})s^2 \tag{4}$$

The quality q_i of these bounds is defined as follows:

$$q_i = \frac{s_1^2}{b_i} \quad (i \in \{1,2\}) \tag{5}$$

All the q_i lie between 0 and 1; the better the bound b_i, the closer q_i to 1. b_1 has also been compared to b_2 by the quantity:

$$r = \frac{b_1}{b_2} \tag{6}$$

The results are shown in figure 3. In (a) the distribution of (s, s_1) was made visible by dividing the square $[0,1]\times[0,1]$ in 100×100 subsquares and storing the fraction of instances in each square in a matrix. The plot shows the contour lines of the density at the heights 0.0001, 0.0003, 0.0004, 0.0006, 0.0010, 0.0020, 0.0040. The highest density occurs near the points (0,0) and (1,1) and the lowest near the point $(\frac{1}{2}, \frac{3}{4})$. (b) and (c) show the distributions of the q_i and (d) that of r. These distributions have been approximated by plotting the fractions in certain small intervals.
As can be seen from the figures the bounds b_1 and b_2 are rather good with average qualities of respectively 0.9055 and 0.8983. The value r has an average of 0.9872, so b_1 is approximately as good as b_2, despite the fact that the latter uses more information. Of course A is a very simple matrix, so it is too early to generalize these results.

5 General subspaces

5.1 Experimental results

Now the case of subspaces of general dimension can be considered. The maximum value for $\sin^2\theta_1$ is given by the minimization problem in theorem 3.4. Unfortunately, solving this problem for $k > 2$ is much more difficult than the case $k = 2$ and a theoretical expression has not been found yet.
To compensate this lack of theoretical results a number of computer experiments have been performed. These experiments have followed the following scheme. A matrix A of size n and an extreme eigenvalue λ_1 have been chosen and a large number of subspaces V of dimension k have been generated. For each subspace the Ritzvectors and the values of $\sin^2\phi$ and $\sin^2\theta_i$, $i \in \{1,\ldots,k\}$ have been calculated. The $\sin^2\phi$ axis was divided into 100 intervals and for each interval the value of the maximal $\sin^2\theta_1$ thus far was stored, together with the corresponding $\sin^2\phi$, $\sin^2\theta_i$, $i \in \{2,\ldots,k\}$ and Z. In this way the curve connecting the maximal $\sin^2\theta_1$ can be seen as the empirical

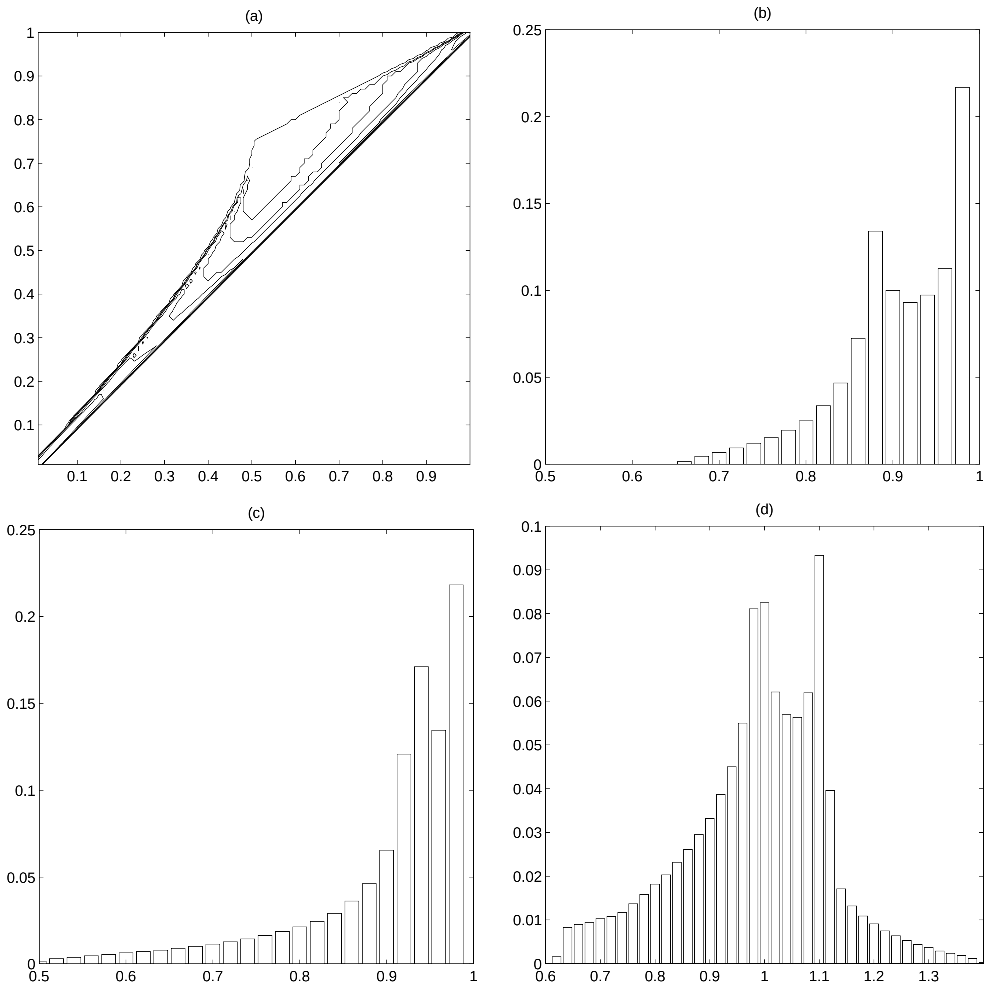

FIGURE 3: (a) Distribution of (s, s_1). (b) Distribution of q_1. (c) Distribution of q_2. (d) Distribution of r.

equivalent for the theoretical upperbound of the type shown in figure 2. At first the subspaces V were chosen completely random, but after some time there was hardly any improvement with this method. The second phase of the algorithm used small perturbations of the Z already found to improve the rate of useful subspaces.

The first experiment used $n = 5$, $A = \text{diag}(1, 2, 3, 4, 5)$, $\lambda_1 = 1$. The case $k = 2$ was considered in example 4.5 and the corresponding curve for $k = 3$ is shown in figure 4(a). The left and the right of the curve resemble the shape of the curve in figure 2, but the middle section is different.

Things get clearer if we also plot the values of $\sin^2\theta_2$ and $\sin^2\theta_3$ corresponding to the maximal $\sin^2\theta_1$. The result can be seen in figure 4(c). At each value of $\sin^2\phi$ three dots are plotted, representing the three values of the $\sin^2\theta_i$.

The following events are visible in this figure. When we look from the right to the left we find that first all $\sin^2\theta_i$ are equal to eachother and to the upperbound given in lemma 3.2. Then at a certain point there occurs a sort of bifurcation. $\sin^2\theta_1$ and $\sin^2\theta_2$ remain equal, but $\sin^2\theta_3$ takes larger values and ascents to reach the value one. At that point the $\sin^2\theta_1$ and $\sin^2\theta_2$ seperate while $\sin^2\theta_3$ remains equal to one. At $\sin^2\phi = 0$, $\sin^2\theta_2$ has also reached one and $\sin^2\theta_1$ is zero. The most remarkable thing is that the point where the second bifurcation occurs is the same point where the discontinuity occurs in the curve for the case $k = 2$. The values of $\sin^2\theta_1$ and $\sin^2\theta_2$ for $k = 2$ are plotted in figure 4(b). To the left of the bifurcation at $(\frac{1}{4}, \frac{5}{8})$ the curves of $\sin^2\theta_1$ and $\sin^2\theta_2$ are identical in (b) and (c). It is hard to believe that this is a coincidence, for $k = 3$ it indicates a reduction to the 2-dimensional case.

The experiment was repeated for $k = 4$, leading to figure 4(d). For each $\sin^2\phi$ the four values $\sin^2\theta_i$ are plotted. In the plot we see three bifurcations. To the left of the middle one is $\sin^2\theta_4 = 1$ and the graphs of $\sin^2\theta_i, i \in \{1, 2, 3\}$ are identical to those in (c). So there is a reduction to the case $k = 3$ and further to the left to $k = 2$.

The next experiment used a slightly larger matrix $n = 10$, $A = \text{diag}(1, 2, \ldots, 10)$, $\lambda_1 = 1$. The cases $k \in \{2, \ldots, 9\}$ have been considered and the figures 5 and 6 contain the results. In these experiments more or less the same behaviour as described before is visible. It should be noted however that there are more irregularities in the curves. It is true that there are always $k - 1$ bifurcations, but the reductions to lower-dimensional cases are not always obvious except for the cases $k = 2$ and $k = 3$ which are repeated in each plot. It seems that this is due to the difficulty of finding the subspace maximizing $\sin^2\theta_1$ by generating random V. During the experiments it became clear that the probability of choosing an nearly maximizing subspace is very small.

5.2 Conclusions from the experiments

In the previous section the theoretical upperbound for $\sin^2\theta_1$ was approximated from below by experimental results. Define $s = \sin^2\phi$ and for all $k \geq 2$ and $i \in \{1, \ldots, k\}$: $s_i^k(s) = \sin^2\theta_i$ where $\sin^2\theta_1$ is maximal with respect to all subspaces of dimension k. The experimental results indicate that the s_i^k are uniquely defined. They look differentiable except in $k - 1$ points $p_1^k, \ldots, p_{k-1}^k$ where bifurcations in the values of the sines occur. Figure 4 and part of the results in the figures 5 and 6 suggest that only

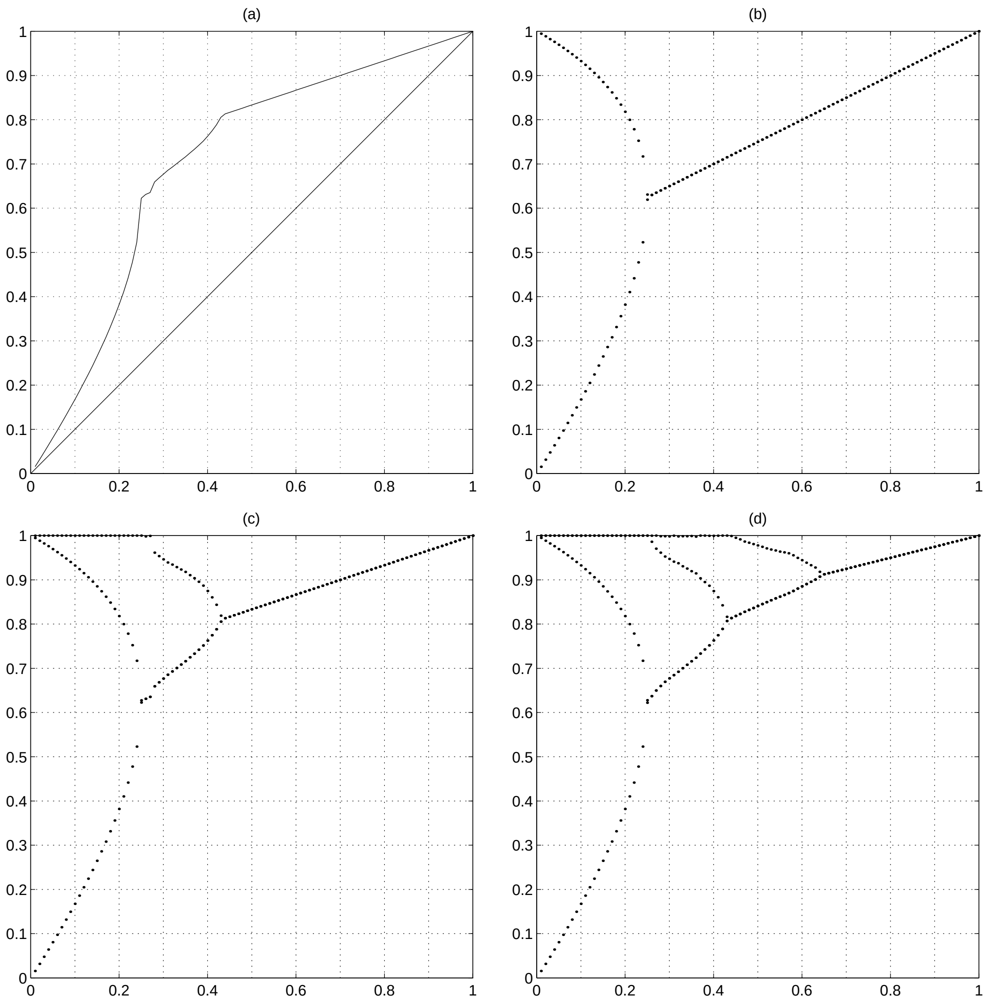

FIGURE 4: (a): Experimentally generated maximum values of $\sin^2\theta_1$ for $\dim(V) = 3$. (b), (c), (d): Experimentally generated maximum values of $\sin^2\theta_1$ together with the corresponding values of $\sin^2\theta_i (i \in \{2, \ldots, k\})$ for $\dim(V) = k$. k is respectively 2,3,4.

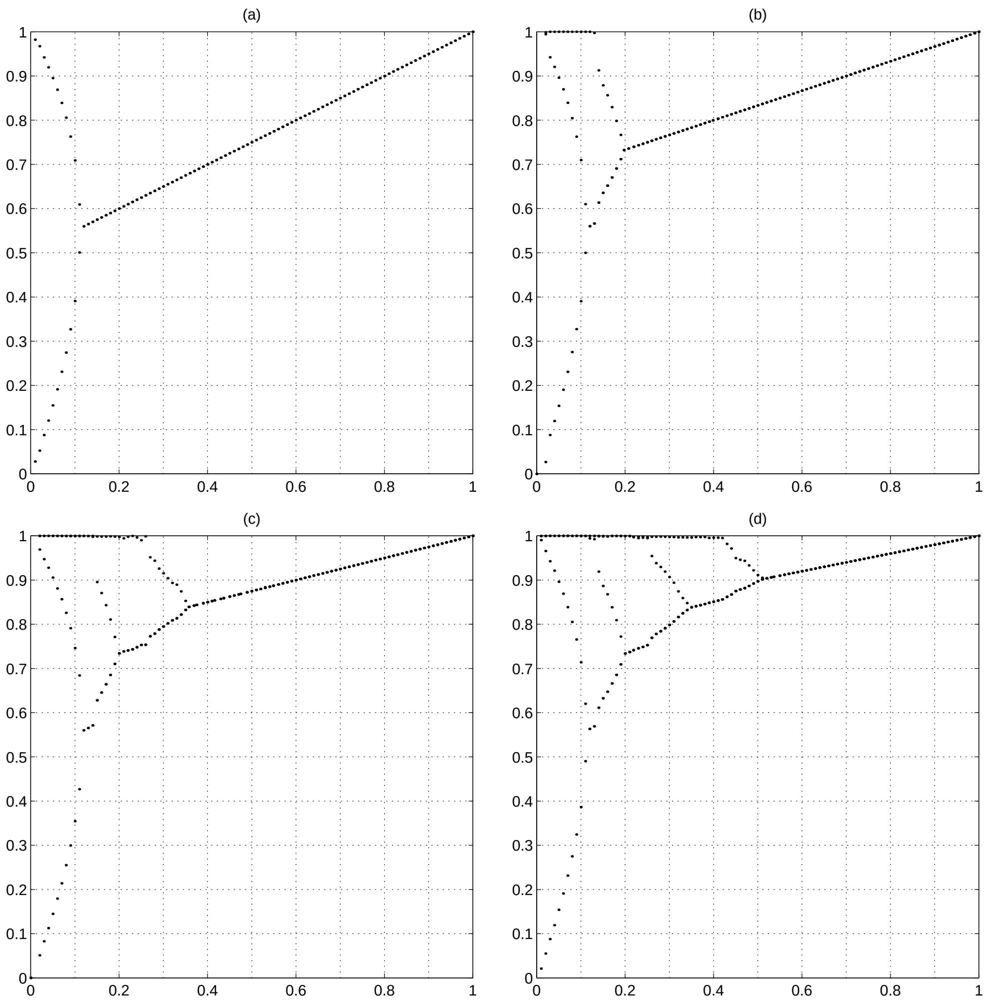

FIGURE 5: Experimentally generated maximum values of $\sin^2\theta_1$ together with the corresponding values of $\sin^2\theta_i (i \in \{2, \ldots, k\})$ for $\dim(V) = k$, $k \in \{2, 3, 4, 5\}$.

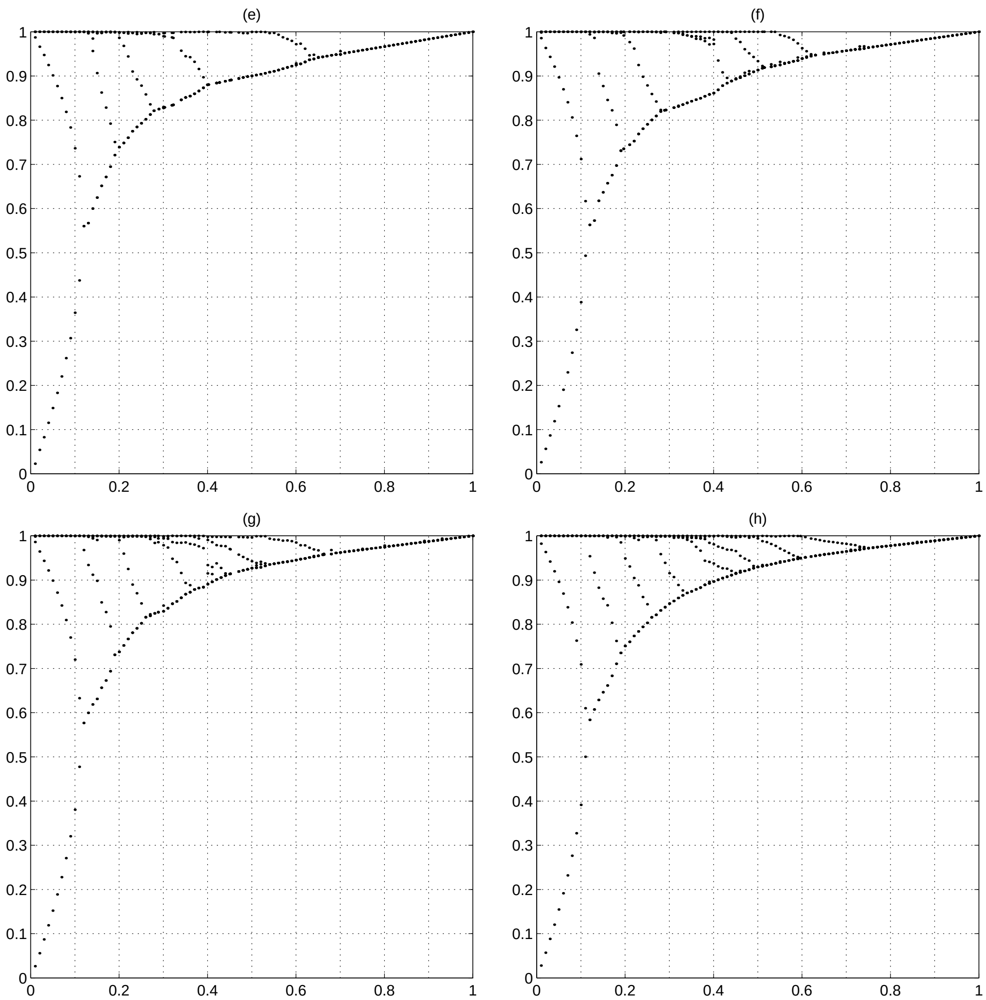

FIGURE 6: Experimentally generated maximum values of $\sin^2\theta_1$ together with the corresponding values of $\sin^2\theta_i (i \in \{2, \ldots, k\})$ for $\dim(V) = k$, $k \in \{6, 7, 8, 9\}$.

the number and not the position of these points depend on k, so their superscripts can be dropped. To the left of p_1 the curves of s_1^k and s_2^k look identical to the case $k = 2$ which is theoretically understood. In general can be said that to the left of p_j the s_i^k for $i \leq j+1$ seem to be identical to the case $k = j+1$, while the rest of the s_i^k is equal to one. To the right of p_{k-1} all s_i^k are equal to eachother, resulting in the worst possible case, described in lemma 3.2.
These observations lead to the proposal of the next conjecture which summarizes the paragraph above.

Conjecture 5.1 Let the symmetric matrix A be given and let λ_1 be an extreme eigenvalue of A. Define $s = \sin^2 \phi$ and for all $k \geq 2$ and $i \in \{1, \ldots, k\}$: $s_i^k(s) = \sin^2 \theta_i$ where $\sin^2 \theta_1$ is maximal with respect to all subspaces of dimension k.

1. The functions s_i^k are well defined and continuous.
2. There is a ascending row of points p_j $(j \geq 1)$ such that all s_i^k are differentiable on the intervals $(0, p_1)$, (p_j, p_{j+1}) $(j \in \{1, \ldots, k-2\})$ and $(p_{k-1}, 1)$.
3. For all $j \leq k-2$ on the interval $[0, p_j]$ is: $\forall i \leq j+1 : s_i^k = s_i^{j+1}$ and $\forall i > j+1 :$ $s_i^k = 1$.
4. On the interval $[0, p_1]$ is s_1^k described by theorem 4.2.
5. On the interval $[p_{k-1}, 1]$ is $\forall i : s_i^k = \frac{k-1+s}{k}$.

References

[1] W. Arnoldi, *The principle of minimized iterations in the solution of the matrix eigenvalue problem*, Quart. Appl. Math., 9 (1951), pp. 17–29.

[2] F. Chatelin, *Valeurs propres de matrices*, Masson, Paris, 1988.

[3] G. Golub and C. van Loan, *Matrix Computations*, The John Hopkins University Press, Baltimore, second ed., 1989.

[4] S. Kaniel, *Estimates for some computational techniques in linear algebra*, Math. Comp., 20 (1966), pp. 369–378.

[5] C. Lanczos, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*, J. Res. Nat. Bur. Standards, 45 (1950), pp. 255–282.

[6] B. Parlett, *The Symmetric Eigenvalue Problem*, Prentice-Hall, Englewood Cliffs, NJ, 1980.

[7] Y. Saad, *Numerical Methods for Large Eigenvalues Problems*, Manchester University Press, 1992.