Estimating the Largest Eigenvalue by the Power and Lanczos Algorithms with a Random Start

J. Kuczyński
Institute of Computer Science, Polish Academy of Sciences

H. Woźniakowski*
Department of Computer Science, Columbia University
and
Institute of Informatics, University of Warsaw

July, 1989

CUCS-026-92

Abstract

Our problem is to compute an approximation to the largest eigenvalue of an $n \times n$ large symmetric positive definite matrix with relative error at most ε. We consider only algorithms that use Krylov information $[b, Ab, \ldots, A^k b]$ consisting of k matrix-vector multiplications for some unit vector b. If the vector b is chosen deterministically then the problem cannot be solved no matter how many matrix-vector multiplications are performed and what algorithm is used. If, however, the vector b is chosen randomly with respect to the uniform distribution over the unit sphere, then the problem can be solved on the average and probabilistically. More precisely, for a randomly chosen vector b we study the power and Lanczos algorithms. For the power algorithm (method) we prove sharp bounds on the average relative error and on the probabilistic relative failure. For the Lanczos algorithm we present only upper bounds. In particular, $\ln(n)/k$ characterizes the average relative error of the power algorithm,

*Supported in part by the National Science Foundation under Grant DCR-86-03674.

whereas $O((\ln(n)/k)^2)$ is an upper bound on the average relative error of the Lanczos algorithm. In the probabilistic case, the algorithm is characterized by its probabilistic relative failure which is defined as the measure of the set of vectors b for which the algorithm fails. We show that the probabilistic relative failure goes to zero roughly as $\sqrt{n}(1-\varepsilon)^k$ for the power algorithm and at most as $\sqrt{n}\, e^{-(2k-1)\sqrt{\varepsilon}}$ for the Lanczos algorithm. These bounds are for a worst case distribution of eigenvalues which may depend on k. We also study the behavior in the average and probabilistic cases of the two algorithms for a fixed matrix A as the number of matrix-vector multiplications k increases. The bounds for the power algorithm depend then on the ratio of the two largest eigenvalues and their multiplicities. The bounds for the Lanczos algorithm depend on the ratio between the difference of the two largest eigenvalues and the difference of the largest and the smallest eigenvalues.

1 Introduction

In this paper we address the problem of approximating the largest eigenvalue λ_1 of an $n \times n$ large symmetric positive definite matrix A. We wish to compute an approximation ξ with relative error at most ε, i.e.,$|\lambda_1 - \xi| \leq \varepsilon\, \lambda_1$. Typically the matrix A is sparse and it is reasonable to use Krylov information consisting of k matrix-vector multiplications, $[b, Ab, \dots, A^k b]$, for some unit vector b. Examples of algorithms for this problem are the power algorithm which has rather limited practical value and the far superior Lanczos algorithm. It is well known that convergence of both algorithms depends on the distribution of eigenvalues and on the angle between the vector b and the eigenvector η_1 corresponding to the largest eigenvalue, see Section 2 for references. In particular, if the vector b is chosen deterministically and independently on the matrix A then it may happen that b is orthogonal to η_1. In such a case the two algorithms fail to approximate the largest eigenvalue. It is easy to extend this negative result by showing that as long as Krylov information is used with a deterministic unit vector b, then there exists no algorithm which can approximate the largest eigenvalue for all symmetric positive matrices, see Section 2 for details. Also if Krylov information is replaced by any k matrix-vector multiplications then the problem cannot be solved for all symmetric positive matrices as long as $k \leq n-1$ since all the vectors might be orthogonal to η_1, see Remark 7.1 of Section 7.

On the other hand, a closer look at the analysis of convergence of the power or Lanczos algorithm yields the impression that it is very unlikely that the position of the vector b will be so unfortunate and that it should not really happen with a randomly chosen vector b. This is exactly the point of departure of our paper. We assume that the vector b is chosen randomly with uniform distribution over the unit sphere of n dimensional space. Then we define the average relative error of an algorithm as the expected relative error while integrating over the vectors b of the unit sphere. We also analyze the probabilistic relative failure which is defined as the measure of the set of vectors b for which the algorithm fails to approximate the largest eigenvalue with relative error at most ε.

For the average case we find sharp bounds on the relative error of the power algorithm, see Theorem 3.1. Namely, no matter what the distribution of eigenvalues of the matrix A, the relative error is bounded from above, for large n, by roughly $0.564\ \ln(n)/(k-1)$. This bound is sharp in the sense that for each k there exists a symmetric positive definite matrix A for which the relative error is at least roughly $0.5\ \ln(n)/(k-1)$. Hence, the relative error of the power algorithm tends to zero as k goes to $+\infty$, although the speed of convergence is quite slow. Observe that the dimension n of the problem affects the speed of convergence only logarithmically.

For the Lanczos algorithm we are only able to present upper bounds on its average relative error, see Theorem 3.2. We show that independently of the distribution of eigenvalues of the matrix A, the relative error is bounded by $2.575\,(\ln(n)/(k-1))^2$ for $k \in [4, n-1]$, and that the relative error is zero if k is no less than the total number of distinct eigenvalues. To check the quality of this upper bound we performed many numerical tests. They are reported in Section 6. Numerical tests for the matrix whose eigenvalues are shifted zeros of the Chebyshev polynomial of the first kind of degree n seem to indicate that the relative error of the Lanczos algorithm behaves like k^{-2}. If so then the factor $\ln^2(n)$ in our upper bound is an overestimate.

Comparing the two algorithms we see, not surprisingly, the superiority of the Lanczos algorithm. The ratio of steps of the power and Lanczos algorithms needed to achieve error at most ε is roughly at least equal to $0.35\,\varepsilon^{-1/2}$. Thus, the smaller ε the more superior the Lanczos algorithm.

So far we have discussed the bounds for a worst case distribution of eigenvalues. We also study the behavior of the average relative errors for a fixed matrix A and increasing k. For the power algorithm, we obtain formulas for the rate of convergence which depends on the ratio ρ of the two largest eigenvalues and on their multiplicities, see part (c) of Theorem 3.1. In particular, the best rate is obtained if the multiplicity p of the largest eigenvalue is at least 3 and then it is equal to $\rho^{2(k-1)}$. For $p = 1$, the rate is ρ^{k-1}. Observe that for a deterministic vector b which is not orthogonal to the eigenvector η_1, the rate is $\rho^{2(k-1)}$. In Section 3 we explain why for $p \leq 2$ the rate decreases in the average case. For the Lanczos algorithm we obtain only an upper bound on the ratio which depends on the difference of the two largest eigenvalues over the difference of the largest and the smallest eigenvalues, see part (b) of Theorem 3.2.

We now turn to the probabilistic case. As before, we find sharp bounds for the probabilistic relative failure of the power algorithm which are independent of the distribution of eigenvalues, see Theorem 4.1. The failure goes to zero roughly as $\sqrt{n}(1-\varepsilon)^k$. Note that now the dimension n affects the failure much more substantially than in the average case. Although the failure goes to zero exponentially, for small ε the speed of convergence is quite slow.

The failure of the Lanczos algorithm is zero if k is no less than the total number of distinct eigenvalues, and is bounded by roughly $1.648\,\sqrt{n}e^{-\sqrt{\varepsilon}(2k-1)}$ for any k, see Theorem 4.2. Hence, we have the same dependence on the dimension n, but the dependence on ε is much improved.

If we compare the number of steps needed to obtain a failure of at most δ, then the ratio between the steps of the power and Lanczos algorithms is independent of δ and is roughly at least $2\,\varepsilon^{-1/2}$. Thus, in both the average and probabilistic cases the ratio is proportional to $\varepsilon^{-1/2}$.

We also study the probabilistic relative failure for a fixed matrix A and increasing k. The rate of convergence of the power algorithm depends on multiplicity p and is given by $\rho^{p(k-1)}$. Hence, the rate increases with multiplicity. On the other hand, the asymptotic constant for large p and small ε is huge, see part(c) of Theorem 4.1. As before, for the Lanczos algorithm we only obtain an upper bound on the ratio which depends on the two largest and the smallest eigenvalues.

The proofs of theorems from Sections 3 and 4 are presented in Section 5. It turns out that the proof technique for the power algorithm can be applied for the Lanczos algorithm with the use of Chebyshev polynomials of the first kind for the average case and of the second kind for the probabilistic case. We think that getting a sharp lower bound on the error or failure of the Lanczos algorithm will require a more sophisticated analysis.

In Remark 7.3 of Section 7 we briefly mention a modified power algorithm which was analyzed in the probabilistic case by Dixon [83]. We extend his analysis to the average case and conclude that the power algorithm is better.

In this paper we do not address the termination criterion. Termination is inherently hard due to the negative result for deterministic vectors b. Furthermore, for the Lanczos algorithm a "misconvergence phenomenon" takes place as indicated in Parlett, Simon and Stringer [82]. We also experienced this in our tests as reported in Section 6. Nevertheless we hope that average and probabilistic bounds can be useful in deriving a reliable termination criterion for which one can prove how the algorithm works on the average or probabilistically. It should be added that it is often the case in engineering that the quality of the computed approximation ξ can be verified for moderate n by performing triangular factorization of $\xi_1 I - A$ and checking that no negative pivot occurs. Here, ξ_1 is a computed upper bound on the largest eigenvalue λ_1. For example, if one believes that ξ is an approximation to λ_1 with relative error at most ε then $\lambda_1 \leq \xi/(1-\varepsilon)$, and one can set $\xi_1 = \xi/(1-\varepsilon)$.

Of course, approximating the largest eigenvalue is only one of many interesting eigenvalue problems. To list a few, we mention approximating the mth largest eigenvalue, the smallest eigenvalue, or corresponding eigenvectors. Since the negative result for deterministic vectors b extends also for these new problems, it is quite natural to use random vectors and, hopefully, to get positive results on the average or probabilistically. In particular, it seems to us that a similar proof technique can work for approximating the smallest eigenvalue and the condition number of a symmetric positive definite matrix. We hope to report this in the near future.

Finally we add a remark on replacing the standard relative error criterion by one depending on the eigenvalue spread $\lambda_1 - \lambda_n$, where λ_1 is the rightmost and λ_n is the leftmost eigenvalue, see Parlett [89].

The gap ratio is defined, see Parlett [80], as $(\lambda_1 - \lambda_2)/(\lambda_2 - \lambda_n)$, where λ_2 is the second largest eigenvalue, and it occurs in the analysis of the rate of convergence of Krylov subspace methods to λ_1. The natural extension for our problem is to seek approximation ξ to λ_1 that satisfies

$$|\lambda_1 - \xi| \leq \varepsilon(\lambda_1 - \lambda_n).$$

Since this error criterion for the Lanczos algorithm is shift invariant, the bounds presented in this paper for standard relative error also hold for the error relative to the eigenvalue spread $\lambda_1 - \lambda_n$ as given above. The significant advantage of using the spread is that our results apply to all symmetric matrices, positive definite or not. In contrast to the Lanczos algorithm, the error of the power algorithm relative to the eigenvalue spread is not shift invariant and there is no sense in modifying the standard relative error criterion. Details are given in Remark 7.5 of Section 7.

2 Definition of the Problem

Let A be an $n \times n$ large symmetric positive definite matrix. Let $\lambda_i = \lambda_i(A)$ denote the eigenvalues of the matrix A, $\lambda_1(A) \geq \lambda_2(A) \geq \ldots \geq \lambda_n(A) > 0$. We want to compute an approximation to the largest eigenvalue $\lambda_1(A)$. More precisely, for a given (presumably small) positive number ε we want to compute a number $\xi = \xi(A)$ such that the relative error between $\lambda_1(A)$ and $\xi(A)$ does not exceed ε,

$$\left|\frac{\xi(A) - \lambda_1(A)}{\lambda_1(A)}\right| \leq \varepsilon. \tag{1}$$

Obviously, if $\varepsilon \geq 1$, $\xi(A) = 0$ satisfies (1). To avoid this trivial case, we assume that $\varepsilon \in [0, 1)$.

If n is large, say, of order 10^{+3} or 10^{+4} then it is prohibitively expensive to use well known algorithms such as QR or QL. Instead, it is reasonable to assume that the information about the matrix A is supplied by a subroutine that computes Az for any vector z. If A is sparse, which often is the case, the time and storage needed to perform the matrix-vector multiplication Az is proportional to n.

We therefore assume that *Krylov* information consisting of k matrix-vector multiplications, $k \geq 1$,

$$N_k(A, b) = [b, Ab, \ldots, A^k b], \tag{2}$$

is used to compute the approximation $\xi(A)$. That is, $\xi(A) = \phi_k(N_k(A, b))$ for some mapping $\phi_k : \mathbb{R}^{n(1+k)} \to \mathbb{R}$. Here, b is a nonzero vector which, without loss of generality, may be normalized such that $||b|| = 1$, where $||\cdot||$ stands for the Euclidean norm of vectors.

Krylov information can be written as $[z_1, z_2, \ldots, z_{k+1}]$ with $z_1 = b$ and $z_i = Az_{i-1}$. This shows that it can be computed in time of k matrix-vector multiplications.

Examples of algorithms that use Krylov information include the power and (simple) Lanczos algorithms. For the power algorithm ξ^{pow} we have

$$\xi(A) = \xi^{pow}(A, b, k) = \frac{(Ax, x)}{(x, x)} \quad \text{with} \quad x = A^{k-1}b = z_k, \tag{3}$$

whereas for the Lanczos algorithm ξ^{Lan} we have

$$\xi(A) = \xi^{Lan}(A, b, k) = \max\left\{\frac{(Ax, x)}{(x, x)} : 0 \neq x \in \mathrm{span}(b, \ldots, A^{k-1}b)\right\}. \tag{4}$$

The analysis of convergence of the power algorithm is straightforward and may be found in most books on numerical analysis. The analysis of convergence of the Lanczos algorithm is more complex and some of it may be found in e.g., Wilkinson [65], Kaniel [66], Paige [71,72], Kahan and Parlett [76], Scott [78], Parlett [80] and Saad [80].

In both cases, convergence depends on distributions of eigenvalues of the matrix A and on the vector b. In particular, if b is orthogonal to the eigenvector η_1, $A\eta_1 = \lambda_1\eta_1$, then both algorithms fail to converge to λ_1. This means that (1) cannot always be satisfied.

It is then natural to ask if there exists an algorithm using Krylov information (with sufficiently large k) for which (1) is satisfied for some ε and for *all* symmetric and positive definite matrices. It is easy to verify that, unfortunately, this is *not* the case.

We now present a simple argument why this is so, see also Remark 7.1 in Section 7, where further discussion may be found. For arbitrary A, b and k, let

$$d = d(A, b, k) = \dim \operatorname{span}(b, Ab, \dots, A^{k-1}b).$$

Clearly, $1 \le d \le \min(k, n)$ and both bounds can be achieved. Let $\xi(A) = \phi_k(N_k(A, b))$, where ϕ_k is an arbitrary mapping.

Assume that $d \le n-1$. Then there exists a matrix $\bar{A}$, $\bar{A} = \bar{A}^T > 0$, such that $\xi(\bar{A}) = \xi(A)$ and

$$\left| \frac{\xi(\bar{A}) - \lambda_1(\bar{A})}{\lambda_1(\bar{A})} \right| > \varepsilon. \tag{5}$$

That is, $\xi(\bar{A})$ does *not* satisfy (1) for the matrix $\bar{A}$. The matrix $\bar{A}$ is of the form $\bar{A} = A + \alpha\, u\, u^T$, where α is a positive constant and u is a nonzero vector orthogonal to $b, Ab, \dots, A^{k-1}b$. Such a vector exists since $d \le n-1$. By induction we get

$$\bar{A}^j b = A^j b \quad \text{for} \quad j = 0, 1, \dots k.$$

Thus, $N_k(\bar{A}, b) = N_k(A, b)$ and therefore $\xi(\bar{A}) = \xi(A)$. Observe that the trace of $\bar{A}$ is given by

$$\operatorname{trace}(\bar{A}) = \operatorname{trace}(A) + \alpha\, ||u||^2$$

and it goes to infinity as $\alpha \to +\infty$. Therefore the largest eigenvalue $\lambda_1(\bar{A})$ goes to infinity as well. We thus have

$$\left| \frac{\xi(\bar{A}) - \lambda_1(\bar{A})}{\lambda_1(\bar{A})} \right| = \left| \frac{\xi(A) - \lambda_1(\bar{A})}{\lambda_1(\bar{A})} \right| \to 1, \quad \text{as} \quad \alpha \to +\infty.$$

Hence, there exists a positive α for which (5) holds, as claimed.

Observe that for large α, the largest eigenvalue $\lambda_1(\bar{A})$ of $\bar{A}$ is close to α and the eigenvector corresponding to $\lambda_1(\bar{A})$ is close to u. The vector u is orthogonal to all but last vectors of Krylov information. Thus, $N_k(\bar{A}, b)$ contains almost no information on the vector u and therefore no matter how ϕ_k is chosen, $\xi(\bar{A}) = \phi_k(N_k(\bar{A}, b))$ cannot approximate $\lambda_1(\bar{A})$ with relative error at most ε.

To prove (5) we needed to assume that $d(A, b, k) \le n-1$. Observe that this inequality holds for all A and b as long as $k \le n-1$. Thus, if one performs fewer than n matrix-vector multiplications, there always exists a symmetric and positive definite matrix $\bar{A}$ which shares the same information as A and for which it is impossible to approximate its largest eigenvalue with relative error at most ε. We stress that ε needs not be small. The only assumption is $\varepsilon < 1$.

Clearly, if for any A we have $d(A, b, k) = n$ then it is possible to satisfy (1). Indeed, the vectors $b, Ab, \dots, A^{k-1}b$ span the whole space and the matrix A can be uniquely recovered from the computed Krylov information $N_k(A, b)$. Knowing A, we have, at least conceptually, enough information to recover the largest eigenvalue $\lambda_1(A)$ even exactly.

Can we thus guarantee that $d(A,b,k) = n$ for some $k \geq n$? Clearly, not always. For any vector b, there exists a matrix $A = A^T > 0$ such that b is its eigenvector, say, $Ab = \alpha\, b$. Then $d(A,b,k) \equiv 1$ for all k, and no matter how many matrix-vector multiplications are performed, (1) cannot be satisfied for some symmetric and positive definite matrices. It can also happen that $d(A,b,p) = d(A,b,p+1)$ for some p, where $1 \leq p \leq n-1$. Then $d(A,b,k) = d(A,b,p)$ for all $k \geq p$, and still the problem (1) cannot always be solved. We have

$d(A,b,k) = n$ iff $k \geq n$ and vectors $b, Ab, \ldots, A^{n-1}b$ are linearly independent.

Observe that $b, Ab, \ldots, A^{n-1}b$ are linearly independent iff all the eigenvalues of A are distinct and the projections of the vector b onto the eigenvectors η_i of the matrix A are nonzero. That is, $(b, \eta_i) \neq 0$ for $i = 1, 2, \ldots, n$. This property is guaranteed if, for example, A is unreduced tridiagonal and $b = [1, 0, \ldots, 0]$.

Although it is impossible to guarantee that $(b, \eta_i) \neq 0$ for all $i \in [1, n]$, it is intuitively clear that $(b, \eta_i) \neq 0$ should hold for "almost all" vectors b. This is definitely the case if the vector b is chosen randomly, say, with uniform distribution μ on the n-dimensional sphere of radius one. The reader may consult Knuth [81,p.130], where it is explained how such a vector can be generated computationally. Then $(b, \eta_i) = 0$ holds with probability zero and $d(A,b,k) = k$ with probability 1 iff A has at least k distinct eigenvalues.

The last fact follows by noting that $[b, Ab, \ldots, A^{k-1}b]$ in the basis of eigenvectors of A is equal to the product of the diagonal matrix D whose entries are components of b, and the Vandermonde matrix V whose entries are powers of eigenvalues of A. The matrix D is nonsingular with probability one whereas the matrix V has rank k iff A has at least k distinct eigenvalues.

This discussion suggests that although (1) cannot be satisfied for *all* symmetric and positive definite matrices with a deterministically chosen vector b, there is hope this problem can be solved by introducing a random initial vector b of Krylov information. That is, for *all* symmetric and positive definite matrices we wish to have the average relative error with respect to vectors b to be at most ε. Or we may wish to solve the problem with high probability, i.e., for vectors b which form a set of measure close to one.

We now formalize this idea. Let μ be a distribution of the initial vectors b. How should we select μ for the power and Lanczos algorithms? Observe that for both of them we have

$$\begin{aligned} \xi(A,b,k) &= \xi(A, \alpha\, b, k), \quad \forall \alpha \neq 0, \\ \xi(A,b,k) &= \xi(Q^T A Q, Q^T b, k), \quad \forall Q \text{ -orthogonal.} \end{aligned}$$

It is then natural to choose μ which shares these two properties. That is, μ should be concetrated on the unit sphere of $\mathbb{R}^\kappa$ and should be orthogonally invariant. Therefore, we assume in this paper that μ is a *uniform* distribution over the unit sphere of $\mathbb{R}^n$, $\mu(\{b \in \mathbb{R}^n : \|b\| = 1\}) = 1$.

For any symmetric and positive definite matrix A, we select a *random* vector b according to the distribution μ. Then we compute Krylov information $N_k(A,b)$ and the approximation $\xi(A,b,k)$ of the largest eigenvalue $\lambda_1(A)$ by the power or Lanczos algorithm (3) or (4). Then

$$e^{avg}(\xi, A, k) = \int_{\|b\|=1} \left| \frac{\xi(A,b,k) - \lambda_1(A)}{\lambda_1(A)} \right| \mu(db) \tag{6}$$

denotes the average relative error. Let

$$f^{prob}(\xi, A, k, \varepsilon) = \mu \left\{ b \in \mathbb{R}^n : \|b\| = 1, \ \left| \frac{\xi(A,b,k) - \lambda_1(A)}{\lambda_1(A)} \right| > \varepsilon \right\} \tag{7}$$

denote the probability that the algorithm fails to approximate the largest eigenvalue with relative error at most ε. We call (7) the probabilistic relative failure of ξ.

In Remark 7.2 of Section 7 we show that (6) and (7) remain the same if we integrate over the unit ball B_n with respect to normalized Lebesgue measure.

3 Average Case

In this section we present bounds on the average relative error (6) for both the power and Lanczos algorithms. Proofs are given in Section 5. To simplify some estimates we assume that $n \geq 8$. We begin with the power algorithm.

Theorem 3.1 *Let ξ^{pow} be the power algorithm defined by (3).*

(a) For any symmetric positive definite matrix A and for any $k \geq 2$ we have

$$e^{avg}(\xi^{pow}, A, k) \leq \alpha(n) \, \frac{\ln n}{k-1},$$

where $\pi^{-1/2} \leq \alpha(n) \leq 0.871$ and for large n, $\alpha(n) \simeq \pi^{-1/2} = 0.564\ldots$.

(b) For any $k > 1 + \frac{1}{2}\ln(n/\ln n)$, let A be any symmetric matrix with exactly two distinct eigenvalues $\lambda_1 > 0$ and $\lambda_i = \lambda_1(1 - \ln(n/\ln\, n)/(2(k-1))$, for $i = 2, 3, \ldots, n$. Then for large n and k,

$$e^{avg}(\xi^{pow}, A, k) \geq 0.5 \, \frac{\ln n}{k-1} \, (1 + o(1)) \, .$$

(c) For any symmetric positive definite matrix A, let p, $p < n$, and q denote

the multiplicities of the two largest eigenvalues λ_1 *and* λ_{p+1}. *Then*

$$\begin{aligned}
\lim_{k\to+\infty} \frac{e^{avg}(\xi^{pow},A,k)}{(\lambda_{p+1}/\lambda_1)^{2(k-1)}} &= \frac{q}{p-2}\left(1-\frac{\lambda_{p+1}}{\lambda_1}\right) && \text{for } p\ge 3,\\
\lim_{k\to+\infty} \frac{e^{avg}(\xi^{pow},A,k)}{(k-1)\,(\lambda_3/\lambda_1)^{2(k-1)}} &= q\left(1-\frac{\lambda_3}{\lambda_1}\right)\ln\frac{\lambda_1}{\lambda_3} && \text{for } p=2,\\
\lim_{k\to+\infty} \frac{e^{avg}(\xi^{pow},A,k)}{(\lambda_2/\lambda_1)^{k-1}} &= \sqrt{\pi}\,\frac{\Gamma((q+1)/2)}{\Gamma(q/2)}\left(1-\frac{\lambda_2}{\lambda_1}\right) && \text{for } p=1.
\end{aligned}$$

Part (a) of Theorem 3.1 states that no matter what the distribution of eigenvalues of A nor how poorly the dominant eigenvalue is separated from the next largest eigenvalue, the average relative error of the power algorithm is bounded by $0.871\ \ln(n)/(k-1)$. For large n, the constant 0.871 can be replaced by roughly 0.564.

Part (b) of Theorem 3.1 states that this upper bound is essentially sharp since for each k there exists a matrix $A=A^T>0$ with only two distinct eigenvalues for which the average relative error of the power algorithm is at least roughly $0.5\ \ln(n)/(k-1)$.

The average relative error of the power algorithm depends only logarithmically on the dimension n. Thus, even for large n, the constant $0.564\ \ln(n)$ is quite moderate and the error is a modest multiple of $(k-1)^{-1}$. Of course, $(k-1)^{-1}$ tends to zero slowly and to guarantee

$$e^{avg}(\xi^{pow},A,k)\le\varepsilon \quad \forall\ A=A^T>0$$

we have to perform roughly $k=\lceil 1+0.564\ \ln(n)/\varepsilon\rceil$ steps. For small ε, such a number of steps cannot be realistically done. As we shall see in Theorem 3.2, the Lanczos algorithm is, not surprisingly, much better and therefore the power algorithm is of limited value in numerical practice.

We now comment on the paper of O'Leary, Stewart and Vandergraft [79]. They analyzed the power algorithm for fixed eigenvectors $\eta_1,\eta_2,\dots,\eta_n$ and for a fixed vector b, $||b||=1$. They showed that for a worst case distribution of eigenvalues, the power algorithm takes roughly $k=\ln(\tau)/\varepsilon$ steps to compute an ε-approximation to the largest eigenvalue. Here $\tau=\tan|\theta|$, where θ is the angle between b and η_1. If all $b_i=(b,\eta_i)$ are more or less equal then $\tau\simeq\sqrt{n}$ and $k\simeq\frac{1}{2}\ln(n)/\varepsilon$. Hence, also in this case $\ln(n)/\varepsilon$ exhibits the behavior of the power algorithm.

We turn to part (c) of Theorem 3.1 which explains the asymptotic behavior of the average relative errors of the power algorithm. The rate of convergence depends on the multiplicity p of the largest eigenvalue. We assumed that $p<n$. Note that the case $p=n$ is not interesting since then A is proportional

to the identity matrix and one step of the power algorithm recovers exactly the largest eigenvalue.

The worst rate is for $p = 1$ and in this case is proportional to $(\lambda_2/\lambda_1)^{k-1}$. This should be compared with the deterministic case for which the rate is proportional to $(\lambda_2/\lambda_1)^{2(k-1)}$ whenever $b_1 = (b, \eta_1) \neq 0$. More precisely, for any vector b, let $\rho_k(b) = (\lambda_1 - \xi^{pow}(A, b, k))/\lambda_1$. As before, let $b_i = (b, \eta_i)$. Assuming that $b_1 \neq 0$ we have

$$\rho_k(b) = \left(\frac{\lambda_2}{\lambda_1}\right)^{2(k-1)} \left(\frac{b_2^2 + \ldots + b_{q+1}^2}{b_1^2}\right) \left(1 - \frac{\lambda_2}{\lambda_1}\right) \; + \; o\left(\left(\frac{\lambda_2}{\lambda_1}\right)^{2(k-1)}\right),$$

where q is the multiplicity of the second largest eigenvalue.

To explain the difference in the rate of convergence, note that the average value of $\rho_k(b)$ with respect to b cannot be proportional to $(\lambda_2/\lambda_1)^{2(k-1)}$ since

$$\int_{||b||=1} \frac{b_2^2 + \ldots + b_{q+1}^2}{b_1^2} \, \mu(db) \;=\; +\infty.$$

The complete analysis shows that we lose a factor $(\lambda_2/\lambda_1)^{k-1}$ when integrating $\rho_k(b)$, and therefore the average value of $\rho_k(b)$ is proportional to $(\lambda_2/\lambda_1)^{k-1}$, as claimed in part (c) for $p = 1$.

For $p \geq 2$, the situation is different since

$$\rho_k(b) = \left(\frac{\lambda_{p+1}}{\lambda_1}\right)^{2(k-1)} \left(\frac{b_{p+1}^2 + \ldots + b_{p+q}^2}{b_1^2 + \ldots + b_p^2}\right) \left(1 - \frac{\lambda_{p+1}}{\lambda_1}\right) \; + \; o\left(\left(\frac{\lambda_{p+1}}{\lambda_1}\right)^{2(k-1)}\right)$$

whenever $b_1^2 + \ldots + b_p^2 \neq 0$. For $p \geq 3$, the integral

$$\int_{||b||=1} \frac{b_{p+1}^2 + \ldots + b_{p+q}^2}{b_1^2 + \ldots + b_p^2} \, \mu(db) \;<\; +\infty$$

which explains why the rate of convergence is proportional to $(\lambda_{p+1}/\lambda_1)^{2(k-1)}$.

For $p = 2$, the integral above is "barely" infinite and the complete analysis shows that we lose the factor $\ln(\lambda_3/\lambda_1)^{2(k-1)} \;=\; 2(k-1)\ln(\lambda_3/\lambda_1)$ when integrating $\rho_k(b)$. As claimed in part (c) for $p = 2$, the rate of convergence is therefore proportional to $(k-1)\,(\lambda_3/\lambda_2)^{2(k-1)}$.

Part (c) of Theorem 3.1 shows that the asymptotic constant depends also on the multiplicity q of the second largest eigenvalue and on the ratio λ_{p+1}/λ_1. The multiplicity q may depend on the dimension n, and it can happen that $q = n - p$. In this case and for λ_{p+1}/λ_1 not too close to one, the asymptotic constant is huge.

We wish to add that a similar analysis may be performed for a modified power algorithm ξ^{mpow}, where

$$\xi^{mpow}(A, b, k) \;=\; (A^k b, b)^{1/k}, \quad ||b|| = 1.$$

For the modified power algorithm, $\ln(n)/(k-1)$ is a sharp upper bound on the average relative error which is roughly 1.8 times worse than the corresponding error bound of the power algorithm. Unlike the power algorithm, $\ln(n)/(k-1)$ is also a sharp upper bound on the asymptotic behavior of the average relative error of the modified power algorithm. This shows that the power algorithm is superior to the modified power algorithm. Details are presented in Remark 7.3 of Section 7.

We now proceed to the Lanczos algorithm. The analysis of this algorithm is much more complex and we are able to present only upper bounds. We verify some of our estimates by numerical tests which will be reported here and in more detail in Section 6. Obviously

$$e^{avg}(\xi^{Lan}, A, k) \leq e^{avg}(\xi^{pow}, A, k) \quad \forall A \text{ and } k. \tag{8}$$

Therefore one can apply estimates of the power algorithm also to the Lanczos algorithm. Of course, since the Lanczos algorithm is much more powerful than the power algorithm we hope to get much better estimates of convergence. This will be confirmed by the following theorem. To simplify some formulas we assume that $k \geq 4$, and (as before) that $n \geq 8$.

Theorem 3.2 *Let ξ^{Lan} be the Lanczos algorithm defined by (4).*

(a) For any symmetric positive definite matrix A, let m denote the number of distinct eigenvalues of A. Then
for $k \geq m$,

$$e^{avg}(\xi^{Lan}, A, k) = 0,$$

for $k \in [4, m-1]$,

$$e^{avg}(\xi^{Lan}, A, k) \leq 0.103 \left(\frac{\ln\left(n(k-1)^4\right)}{k-1}\right)^2 \leq 2.575 \left(\frac{\ln n}{k-1}\right)^2.$$

(b) For any symmetric positive definite matrix A, let p, $p < n$, denote the multiplicity of the largest eigenvalue λ_1, and let λ_{p+1} and λ_n be the second largest and the smallest eigenvalue of A. Then

$$e^{avg}(\xi^{Lan}, A, k) \leq 2.589\sqrt{n} \left(\frac{1-\sqrt{(\lambda_1-\lambda_{p+1})/(\lambda_1-\lambda_n)}}{1+\sqrt{(\lambda_1-\lambda_{p+1})/(\lambda_1-\lambda_n)}}\right)^{k-1}.$$

Theorem 3.2 states that the Lanczos algorithm converges in m steps, $m \leq n$, which confirms our intuition that it can fail only on a set of vectors b of measure zero. For k essentially less than n, the average relative error of the Lanczos algorithm is roughly bounded by $0.1\,(\ln(n)/k)^2$. Since $\ln(n)/k$ is a sharp estimate of the average relative error of the power algorithm, we see that the Lanczos algorithm is far superior. If we want to guarantee that $e^{avg}(\xi, A, k) \le \varepsilon$, then the power algorithm needs to perform roughly $k^{pow} = 0.564\, \ln(n)/\varepsilon$ steps, whereas the Lanczos algorithm will take roughly $k^{Lan} \le 1.605\, \ln(n)/\sqrt{\varepsilon}$ steps. Thus

$$\frac{k^{pow}}{k^{Lan}} \ge \frac{0.35}{\sqrt{\varepsilon}}.$$

As already indicated we do not know if the upper bound for the Lanczos algorithm presented in part (a) is sharp. We verify the sharpness of this bound by many numerical tests. These tests seem to indicate that

$$e^{avg}(\xi^{Lan}, A, k) = \Theta(k^{-2})$$

with the constant in the Θ notation independent of n. If this is the case then the bound in part (a) is an overestimate by the factor $\ln^2 n$. Details of numerical tests are reported in Section 6.

Part (b) of Theorem 3.2 yields a non-asymptotic estimate in terms of the two largest eigenvalues and the smallest eigenvalue of A. Observe that the bound in part (b) is better than the bound in part (a) if $(\lambda_1 - \lambda_{p+1})/(\lambda_1 - \lambda_n)$ is not too close to zero.

4 Probabilistic Case

In this section we present bounds for the probabilistic relative failure (7) for the power and Lanczos algorithms. Proofs are given in Section 5. As in Section 3 we begin with the power algorithm.

It is easy to check that for $\varepsilon = 0$, the probabilistic relative failure of the power algorithm $f^{prob}(\xi^{pow}, A, k, 0) = 1$ for all matrices A with at least two distinct eigenvalues, and $f^{prob}(\xi^{pow}, A, k, 0) = 0$ for all matrices A having only one distinct eigenvalue. That's why we assume in Theorem 4.1 that $\varepsilon > 0$. The probabilistic relative failure of the power algorithm depends on the function g defined by

$$g(\varepsilon, k) = \frac{(1-\varepsilon)^{k-1/2}\,(1 - 1/(2k-1))^{k-1}}{\sqrt{(1-\varepsilon)^{2k-1}\,(1-1/(2k-1))^{2(k-1)} + 2(k-1)\varepsilon}}\,, \qquad k \ge 2. \quad (9)$$

Note that

$$\begin{aligned} g(\varepsilon, k) &\le (1-\varepsilon)^{k-1/2}, \\ g(\varepsilon, k) &= \frac{1}{\sqrt{2e}}\,\frac{(1-\varepsilon)^{k-1/2}}{\sqrt{\varepsilon(k-1)}}\,(1+o(1)) \quad \text{as } k \to +\infty \text{ for } \varepsilon > 0, \end{aligned}$$

with a negative $o(1)$ term.

Theorem 4.1 *Let ξ^{pow} be the power algorithm defined by (3) and let $\varepsilon > 0$.*

(a) For any symmetric positive definite matrix A and for any $k \geq 2$ we have

$$\begin{aligned} f^{prob}(\xi^{pow}, A, k, \varepsilon) &\leq 0.824\,\sqrt{n} \int_0^{g(\varepsilon,k)} (1-t^2)^{(n-1)/2}\, dt \\ &\leq \min\left\{0.824, \frac{0.354}{\sqrt{\varepsilon(k-1)}}\right\}\ \sqrt{n}\,(1-\varepsilon)^{k-1/2}. \end{aligned}$$

(b) For any integer $k \geq 2$, let $\bar{A}$ be any symmetric matrix with two distinct eigenvalues $\lambda_1 > 0$ and $\lambda_i = \lambda_1(1-\varepsilon)(1-1/(2k-1))$ for $i = 2, 3, \ldots, n$. Then

$$\begin{aligned} \max_{A=A^T>0} f^{prob}(\xi^{pow}, A, k, \varepsilon) &= f^{prob}(\xi^{pow}, \bar{A}, k, \varepsilon) \\ &\geq 0.797\,\sqrt{n}\,(1-1/n) \int_0^{g(\varepsilon,k)} (1-t^2)^{(n-1)/2}\, dt, \end{aligned}$$

and for large n and k,

$$f^{prob}(\xi^{pow}, \bar{A}, k, \varepsilon) = a\,\sqrt{\frac{n}{\varepsilon}}\ \frac{(1-\varepsilon)^{k-1/2}}{\sqrt{k-1}}\ (1+o(1)),$$

where $a = 1/\sqrt{\pi e} = 0.342\ldots$.

(c) For any symmetric and positive definite matrix A, let p, $p < n$, and q denote the multiplicities of the two largest eigenvalues λ_1 and λ_{p+1}. If $\lambda_{p+1}/\lambda_1 < 1-\varepsilon$ then

$$\lim_{k\to+\infty} \frac{f^{prob}(\xi^{pow}, A, k, \varepsilon)}{(\lambda_{p+1}/\lambda_1)^{p(k-1)}} = \frac{2\,(1-\varepsilon-\lambda_{p+1}/\lambda_1)^{p/2}}{p\,\varepsilon^{p/2}}\ \frac{\Gamma\left((p+q)/2\right)}{\Gamma\left(p/2\right)\,\Gamma\left(q/2\right)}.$$

Parts (a) and (b) of Theorem 4.1 present sharp bounds on the probabilistic relative failure of the power algorithm. The failure tends to zero with the rate of convergence roughly $(1-\varepsilon)^{k-1/2}$. For small ε, this is quite unsatisfactory. On the other hand, if one is interested in a rough estimate of the largest eigenvalue, say $\varepsilon = 0.5$, then the rate is quite good.

The dependence of the probabilistic relative failure on the dimension n is through $\sqrt{n}$. This shows that the dimension n affects the probabilistic case for the power algorithm in a much more substantial way than the average case which depends only through $\ln n$.

Consider now the minimal number of steps needed to get

$$f^{prob}(\xi^{pow}, A, k, \varepsilon) \le \delta, \quad \forall A = A^T > 0,$$

where δ denotes the measure of a set for which the power algorithm may fail.

Then $k \simeq \ln(n/\delta^2)/(2\varepsilon)$. Hence, the dimension n and the parameter δ affect the number of steps only logarithmically. Even for huge n and very small δ, the factor $\ln(n/\delta^2)/2$ is quite moderate. The dependence on ε is much more crucial since k goes linearly to infinity with ε^{-1}. Observe that the dimension n and the parameter ε affect the number of steps in the same way in the average and probabilistic cases.

Part (c) of Theorem 4.1 presents the asymptotic behavior of the probabilistic relative failure of the power algorithm. The rate of convergence depends on the multiplicity p of the largest eigenvalue, and the rate improves as p increases. On the other hand, the asymptotic constant gets huge for large p and small ε.

Part (c) holds under the assumption that the ratio of two largest eigenvalues is not too close to one, $\lambda_{p+1}/\lambda_1 < 1 - \varepsilon$. Of course, this holds for sufficiently small ε. If, however, $\lambda_{p+1}/\lambda_1 \ge 1 - \varepsilon$, then we do not know the asymptotic behavior of the probabilistic relative failure of the power algorithm and we suspect that its behavior may be quite different from that presented in part (c).

We wish to add that the modified power algorithm in the probabilistic case was analyzed by Dixon [83]. In Remark 7.3 of Section 7 we present his result.

We now turn to the Lanczos algorithm. As was the case for the average case we are able to present only upper bounds. Also in the probabilistic case we have

$$f^{prob}(\xi^{Lan}, A, k, \varepsilon) \le f^{prob}(\xi^{pow}, A, k, \varepsilon) \quad \forall A \text{ and } k \tag{10}$$

and upper bounds of Theorem 4.1 can be used for the Lanczos algorithm. The following theorem presents some better bounds.

Theorem 4.2 *Let ξ^{Lan} be the Lanczos algorithm defined by (4) and let $\varepsilon \in [0, 1)$.*

(a) For any symmetric positive definite matrix A, let m denote the number of distinct eigenvalues of A. Then

for $k \ge m$,

$$f^{prob}(\xi^{Lan}, A, k, \varepsilon) = 0,$$

for any k,

$$f^{prob}(\xi^{Lan}, A, k, \varepsilon) \le 1.648\,\sqrt{n}\,e^{-\sqrt{\varepsilon}(2k-1)}.$$

(b) For any symmetric positive definite matrix A, let p, $p < n$, denote the multiplicity of the largest eigenvalue λ_1, and let λ_{p+1} and λ_n be the second largest and the smallest eigenvalues of A. Then for $\varepsilon > 0$,

$$f^{prob}(\xi^{Lan}, A, k, \varepsilon) \le 1.648\sqrt{\frac{n}{\varepsilon}}\left(\frac{1-\sqrt{(\lambda_1-\lambda_{p+1})/(\lambda_1-\lambda_n)}}{1+\sqrt{(\lambda_1-\lambda_{p+1})/(\lambda_1-\lambda_n)}}\right)^{k-1}.$$

Theorem 4.2 states that also in the probabilistic case the Lanczos algorithm converges in m steps. For any k and for small ε the probabilistic relative failure of the Lanczos algorithm is roughly bounded by $\sqrt{n}\exp(-\sqrt{\varepsilon}(2k-1))$. This should be compared with a sharp bound for the power algorithm given by $\sqrt{n/\varepsilon}(1-\varepsilon)^k/\sqrt{k}$. Once more we see the superiority of the Lanczos algorithm. If we want to guarantee a δ-failure, $f^{prob}(\xi, A, k, \varepsilon) \le \delta$, then we have to perform roughly $k^{pow} = \ln(n/(\delta^2))/(2\varepsilon)$ steps by the power algorithm and roughly $k^{Lan} \le \ln(n/\delta^2)/(4\sqrt{\varepsilon})$ by the Lanczos algorithm. Thus

$$\frac{k^{pow}}{k^{Lan}} \ge \frac{2}{\sqrt{\varepsilon}}.$$

Observe a weak dependence on δ which only logarithmically affects the number of steps. The dependence on ε is more crucial.

As in the average case, part (b) of Theorem 4.2 presents a non-asymptotic bound on the probabilistic relative failure of the Lanczos algorithm. Observe that the bound in part (b) is better then the bound in part (a) if $(\lambda_1 - \lambda_{p+1})/(\lambda_1 - \lambda_n) > \varepsilon$.

5 Proofs of Theorems

In this section we present proofs of theorems from Sections 3 and 4. We begin with the first theorem which deals with convergence of the power algorithm in the average case.

Proof of Theorem 3.1

Let A be any symmetric positive matrix with eigenpairs (λ_i, η_i), where the eigenvectors η_i form an orthonormal basis of $\mathbb{R}^\kappa$ and $\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_n > 0$. That is,

$$A\eta_i = \lambda_i\,\eta_i, \quad (\eta_i, \eta_j) = \delta_{i,j}, \quad i, j = 1, 2, \ldots, n.$$

Let $b = \sum_{i=1}^n b_i\,\eta_i$. From (3) we get

$$\xi^{pow} = \xi^{pow}(A, b, k) = \sum_{i=1}^n b_i^2\,\lambda_i^{2k-1} \Big/ \sum_{i=1}^n b_i^2\,\lambda_i^{2(k-1)}.$$

Let $x_i = \lambda_i/\lambda_1 \in (0,1]$. Then

$$\frac{\lambda_1 - \xi^{pow}}{\lambda_1} = \frac{\sum_{i=2}^n b_i^2\, x_i^{2(k-1)}(1-x_i)}{b_1^2 + \sum_{i=2}^n b_i^2\, x_i^{2(k-1)}}.$$

From Remark 7.2 of Section 7 we know that the average relative error can be defined through the integration over the unit ball B_n,

$$e_k^{pow} = e^{avg}(\xi^{pow}, A, k) = \frac{1}{c_n} \int_{B_n} \frac{\lambda_1 - \xi^{pow}(A,b,k)}{\lambda_1}\, db, \tag{11}$$

where $c_n = \pi^{n/2}/\Gamma(1+n/2)$ is the Lebesgue measure of the unit ball B_n. Since Lebesgue measure is orthogonally invariant we can integrate in (11) with respect to b_i,

$$\begin{aligned} e_k^{pow} &= \frac{1}{c_n} \int_{B'} \sum_{i=2}^n b_i^2\, x_i^{2(k-1)}(1-x_i) \left(\int_{b_1^2 \le 1-||b||_{n-1}^2} \frac{db_1}{b_1^2 + \sum_{i=2}^n b_i^2\, x_i^{2(k-1)}} \right) d\vec{b} \\ &= \frac{2}{c_n} \int_{B'} \sum_{i=2}^n b_i^2\, x_i^{2(k-1)}(1-x_i) \left(\sum_{i=2}^n b_i^2\, x_i^{2(k-1)} \right)^{-1/2} \arctan\left(h(b)\right) d\vec{b}, \end{aligned}$$

where B' is the $(n-1)$-dimensional unit ball, $||b||_{n-1}^2 = \sum_{i=2}^n b_i^2$, $d\vec{b}$ stands for $db_2 \cdots db_n$ and $h(b) = \sqrt{(1-||b||_{n-1}^2)/\sum_{i=2}^n b_i^2\, x_i^{2(k-1)}}$.

Schwartz's inequality for sums, $\sum_{i=2}^n y_i\, z_i \le \left(\sum_{i=2}^n y_i^2\right)^{1/2} \left(\sum_{i=2}^n z_i^2\right)^{1/2}$, with $y_i = b_i\, x_i^{k-1}$ and $z_i = b_i\, x_i^{k-1}(1-x_i)$, yields

$$e_k^{pow} \le \frac{2}{c_n} \int_{B'} \left(\sum_{i=2}^n b_i^2\, x_i^{2(k-1)}\,(1-x_i)^2 \right)^{1/2} \arctan\left(h(b)\right) d\vec{b}.$$

Using now Schwartz's inequality for integrals we get

$$\begin{aligned} e_k^{pow} &= \frac{2}{c_n} \left(\int_{B'} d\vec{b} \right)^{1/2} \left(\int_{B'} \sum_{i=2}^n b_i^2\, x_i^{2(k-1)}(1-x_i)^2 \arctan^2\left(h(b)\right) d\vec{b} \right)^{1/2} \\ &= \frac{2\, c_{n-1}}{c_n} \left(\frac{1}{c_{n-1}} \int_{B'} \arctan^2\left(h(b)\right) \left(\sum_{i:\, x_i \le \beta} b_i^2\, x_i^{2(k-1)}(1-x_i)^2 \right.\right. \\ &\qquad\qquad \left.\left. + \sum_{i:\, x_i > \beta} b_i^2\, x_i^{2(k-1)}(1-x_i)^2 \right) d\vec{b} \right)^{1/2}, \end{aligned}$$

for any number $\beta \in [0,1]$. Here, $c_{n-1} = \pi^{(n-1)/2}/\Gamma(1+(n-1)/2)$ is the Lebesgue measure of the $(n-1)$-dimensional unit ball.

Consider the function $H(t) = (1-t)^2\, t^{2(k-1)}$. The maximum of H is attained at $t_0 = 1 - 1/k$ and H is increasing in $[0, t_0]$. Let $\beta \le t_0$. Since $\arctan z \le \pi/2$ and $\arctan z \le z$, then

$$
\begin{aligned}
\arctan^2(h(b)) \sum_{i: x_i \le \beta}^{n} b_i^2\, x_i^{2(k-1)} (1-x_i)^2 &\le \left(\frac{\pi}{2}\right)^2 \sum_{i=2}^{n} b_i^2\, \beta^{2(k-1)} (1-\beta)^2, \\
\arctan^2(h(b)) \sum_{i: x_i > \beta}^{n} b_i^2\, x_i^{2(k-1)} (1-x_i)^2 &\le h^2(b) \sum_{i=2}^{n} b_i^2 x_i^{2(k-1)} (1-\beta)^2 \\
&= (1 - ||b||_{n-1}^2)\,(1-\beta)^2.
\end{aligned}
$$

Combining these bounds we obtain

$$
\begin{aligned}
e_k^{pow} &\le \frac{2c_{n-1}}{c_n} \left(\frac{1}{c_{n-1}} \int_{B'} (1-\beta)^2 \left(\frac{\pi^2}{4} ||b||_{n-1}^2\, \beta^{2(k-1)} + (1 - ||b||_{n-1}^2) \right) d\vec{b} \right)^{1/2} \\
&= \frac{2c_{n-1}}{c_n} (1-\beta) \left(\frac{1}{c_{n-1}} \left(\frac{\pi^2}{4} \beta^{2(k-1)} - 1 \right) \int_{B'} ||b||_{n-1}^2\, d\vec{b} + 1 \right)^{1/2}.
\end{aligned}
$$

Recall that for any measurable function $f : [0, r] \to \mathbb{R}$, we have

$$
\int_{b \in \mathbb{R}^i, ||b|| \le r} f(||b||)\, db = i\, c_i \int_0^r t^{i-1} f(t)\, dt, \tag{12}
$$

where $c_i = \pi^{i/2} / \Gamma(1 + i/2)$, see e.g., Gradshteyn and Ryzhik [80, 4.642]. For $f(t) = t^2$ we get

$$
\int_{B'} ||b||_{n-1}^2\, d\vec{b} = \frac{n-1}{n+1}\, c_{n-1}.
$$

From this we have

$$
e_k^{pow} \le \frac{2c_{n-1}}{c_n} (1-\beta) \sqrt{\frac{\pi^2}{4} \beta^{2(k-1)} \frac{n-1}{n+1} + \frac{2}{n+1}} \le \frac{2c_{n-1}}{c_n} (1-\beta) \sqrt{\frac{\pi^2}{4} \beta^{2(k-1)} + \frac{2}{n}}.
$$

Observe that

$$
\sqrt{\frac{n}{2\pi}} \le \frac{c_{n-1}}{c_n} = \sqrt{\frac{n}{2\pi}}\, \frac{\sqrt{n/2}\, \Gamma(n/2)}{\Gamma(n/2 + 1/2)} \le \sigma \sqrt{\frac{n}{2\pi}}, \quad \sigma = \frac{192}{105\sqrt{\pi}} \le 1.032. \tag{13}
$$

Indeed, it is enough to show that $\sqrt{x}\, \Gamma(x) / \Gamma(x + 1/2) \in [1, \sigma]$ for $x = n/2 \ge 4$. To show this consider

$$
H(x) = \ln\, \Gamma(x) \; + \; \frac{1}{2} \ln\, x \; - \; \ln\, \Gamma(x + 1/2).
$$

From Gradshteyn and Ryzhik [80, 8.360, 8.365, 8.372] we have $H'(x) = \psi(x) + 1/(2x) - \psi(x + 1/2) \le 0$ for the psi function ψ. Thus,

$$
0 = H(+\infty) \le H(x) \le H(4) = \ln\, \sigma,
$$

as claimed.

Note that for $k-1 \le \pi^{-1/2} \ln n$, part (a) of Theorem 3.1 trivially holds since

$$e_k^{pow} \le 1 \le \pi^{-1/2} \ln n/(k-1).$$

Assume thus that $k-1 > \pi^{-1/2} \ln n$. Take now $\beta = 1 - \ln n/(2(k-1))$. Then $\beta \in (0, t_0]$ for $n \ge 8$ and $\beta^{2(k-1)} \le 1/n$. Thus, we have

$$e_k^{pow} \le 2\sigma \sqrt{\frac{n}{2\pi}}\, \frac{\ln n}{2(k-1)} \sqrt{\frac{\pi^2}{4n} + \frac{2}{n}} = \sigma \sqrt{\frac{\pi^2+8}{8\pi}}\, \frac{\ln n}{k-1} \le 0.871\, \frac{\ln n}{k-1}.$$

This proves that $\alpha(n) \le 0.871$ for all $n \ge 8$.

For large n, take $\beta = 1 - a \ln n/(2(k-1))$ with $a = 1 + 1/\ln \ln n$. Then $\beta^{2(k-1)} \le 1/n^a$ and

$$e_k^{pow} \le \frac{2c_{n-1}}{c_n}\, a\, \frac{\ln n}{2(k-1)} \sqrt{\frac{\pi^2}{4n^a} + \frac{2}{n}} = \alpha(n)\, \frac{\ln n}{k-1}.$$

Since $c_{n-1}/c_n \simeq \sqrt{n/2\pi}$ and $1/n^{a-1}$ goes to zero, we have

$$\alpha(n) \simeq \frac{1}{\sqrt{\pi}} = 0.564\ldots,$$

which completes the proof of part (a).

We now prove part (b) of Theorem 3.1. Consider the matrix A from part (b). Then

$$x_i = \frac{\lambda_i}{\lambda_1} = 1 - \frac{\alpha}{2(k-1)}, \quad i = 2, 3, \ldots, n, \quad \alpha = \ln\left(\frac{n}{\ln n}\right).$$

We have $1 - x_i = \alpha/(2(k-1))$ and $\beta = x_i^{2(k-1)} = \ln(n)/n(1+o(1))$. For the matrix A, (11) takes the form

$$\begin{aligned}
e^{avg}(\xi^{pow}, A, k) &= \frac{1}{c_n}\, \frac{\alpha}{2(k-1)} \int_{B_n} \frac{\beta \sum_{i=2}^n b_i^2 + b_1^2 - b_1^2}{b_1^2 + \beta \sum_{i=2}^n b_i^2}\, db \\
&= \frac{1}{c_n}\, \frac{\alpha}{2(k-1)} \left(c_n - \int_{B_n} \frac{b_1^2}{b_1^2 + \beta \sum_{i=2}^n b_i^2}\, db \right) \\
&= \frac{1}{c_n}\, \frac{\alpha}{2(k-1)} \left(c_n - \int_{B_n} \frac{\beta^{-1} b_1^2}{\beta^{-1} b_1^2 + \sum_{i=2}^n b_i^2}\, db \right).
\end{aligned}$$

Since $\beta^{-1} \ge 1$ then

$$\begin{aligned}
e^{avg}(\xi^{pow}, A, k) &\ge \frac{1}{c_n}\, \frac{\alpha}{2(k-1)} \left(c_n - \beta^{-1} \int_{B_n} \frac{b_1^2}{\sum_{i=1}^n b_i^2}\, db \right) \\
&= \frac{1}{c_n}\, \frac{\alpha}{2(k-1)} \left(c_n - \beta^{-1} \frac{c_n}{n} \right) \\
&= \frac{\ln(n/\ln n)}{2(k-1)} \left(1 - \frac{1+o(1)}{\ln n} \right) = 0.5\, \frac{\ln n}{k-1}\, (1+o(1)),
\end{aligned}$$

as claimed.

We proceed to prove part (c) of Theorem 3.1. Recall that p and q are multiplicities of the two largest distinct eigenvalues of A. From (11) we can write

$$e_k^{pow} = \frac{1}{c_n} \int_{B_n} \frac{x_{p+1}^{2(k-1)}(1-x_{p+1}) \sum_{i=p+1}^{p+q} b_i^2}{\sum_{i=1}^{p} b_i^2 + x_{p+1}^{2(k-1)} \sum_{i=p+1}^{p+q} b_i^2}\, db \;\; (1+o(1)) \quad \text{as } k \to +\infty.$$

Let $a = x_{p+1}^{2(k-1)} = (\lambda_{p+1}/\lambda_1)^{2(k-1)}$ and $\alpha = (1-x_{p+1})\, a$. Integrating with respect to $b_{p+q+1}, \dots, b_n$ we get

$$\frac{e_k^{pow}}{1+o(1)} = \frac{\alpha\, c_{n-p-q}}{c_n} \int_{B_{p+q}} \frac{\sum_{i=p+1}^{p+q} b_i^2 \left(1-\sum_{j=1}^{p+q} b_i^2\right)^{(n-p-q)/2}}{\sum_{i=1}^{p} b_i^2 + a \sum_{i=p+1}^{p+q} b_i^2}\, db,$$

where B_i is the i-dimensional unit ball and c_i is its measure. We rewrite the last integral as an integral over the unit ball B_p and the ball $\sum_{i=p+1}^{p+q} b_i^2 \le 1-\sum_{i=1}^{p} b_i^2$. Let $t_i = b_i/(1-\sum_{j=1}^{p} b_j^2)^{1/2}$ for $i = p+1, \dots, p+q$ and let $||b||^2 = \sum_{i=1}^{p} b_i^2$, $\;\; ||t||^2 = \sum_{i=p+1}^{p+q} t_i^2$. Then we have

$$\begin{aligned} & \frac{e_k^{pow}}{1+o(1)}\, \frac{c_n}{\alpha\, c_{n-p-q}} = \\ = \;\; & \int_{B_p} \int_{B_q} \frac{(1-||b||^2)^{1+q/2}\, ||t||^2\, (1-||b||^2-(1-||b||^2)\, ||t||^2)^{(n-p-q)/2}}{||b||^2 + a\,(1-||b||^2)\, ||t||^2}\, dt\, db. \end{aligned}$$

Using (12) first for the second integral and then for the first integral we get

$$\begin{aligned} \frac{e_k^{pow}}{1+o(1)} &= \gamma_1 \int_{B_p} \int_0^1 \frac{(1-||b||^2)^{1+(n-p)/2}\, t^{q+1}\, (1-t^2)^{(n-p-q)/2}}{||b||^2 + a\,(1-||b||^2)\, t^2}\, dt\, db \\ &= \gamma_2 \int_0^1 \int_0^1 \frac{(1-x^2)^{1+(n-p)/2}\, x^{p-1}\, t^{q+1}\, (1-t^2)^{(n-p-q)/2}}{x^2 + a\,(1-x^2)\, t^2}\, dt\, dx, \end{aligned}$$

where $\gamma_1 = (\alpha\, q\, c_{n-p-q}\, c_q)/c_n$ and $\gamma_2 = (\alpha\, p\, q\, c_{n-p-q}\, c_p\, c_q)/c_n$.

Consider now the case $p \ge 3$. Then the last double integral is finite even for $a = 0$. Recalling the definition of the beta function,

$$B(i,j) = 2 \int_0^1 t^{2i-1}(1-t^2)^{j-1}\, dt = \frac{\Gamma(i)\, \Gamma(j)}{\Gamma(i+j)}, \tag{14}$$

see e.g., Gradshteyn and Ryzhik [80,8.380 and 8.384], we have

$$\begin{aligned} \frac{e_k^{pow}}{1+o(1)} &= \gamma_2 \left(\int_0^1 (1-x^2)^{1+(n-p)/2}\, x^{p-3}\, dx\right) \left(\int_0^1 (1-t^2)^{(n-p-q)/2}\, t^{q+1}\, dt\right) \\ &= \frac{\gamma_2}{4}\, B\left(\frac{p-2}{2}, \frac{n-p}{2}+2\right)\, B\left(\frac{q}{2}+1, \frac{n-p-q}{2}+1\right). \end{aligned}$$

Expressing c_i's and B's in terms of the gamma function we finally get

$$\frac{e_k^{pow}}{1+o(1)} = \frac{\alpha\, p\, q}{4}\,\frac{\Gamma(-1+p/2)}{\Gamma(1+p/2)} = \left(1-\frac{\lambda_{p+1}}{\lambda_1}\right)\,\left(\frac{\lambda_{p+1}}{\lambda_1}\right)^{2(k-1)}\,\frac{q}{p-2},$$

which proves part (c) for $p \geq 3$.

Assume now that $p = 2$. Observe that for $a \to 0$ we have

$$\begin{aligned}
&\int_0^1 \frac{x(1-x^2)^{n/2}}{x^2 + a\,(1-x^2)\,t^2}\,dx \\
= &\int_0^1 \frac{x}{(1-a\,t^2)x^2 + at^2}\,dx + O\left(\int_0^1 \frac{x^3}{(1-\,at^2)x^2+at^2}\,dx\right) \\
= &\int_0^1 \frac{x}{x^2+a\,t^2}\,dx + O\left(\int_0^1 x\,dx\right) = \frac{1}{2}\,\ln(x^2+a\,t^2)\,\Big|_0^1 + O(1) \\
= &\ln\left(\frac{1}{t\sqrt{a}}\right) + O(1).
\end{aligned}$$

Therefore we have

$$\begin{aligned}
\frac{e_k^{pow}}{1+o(1)} &= \gamma_2 \int_0^1 t^{q+1}\,(1-t^2)^{\frac{n-q-2}{2}}\,\ln\left(\frac{1}{t\sqrt{a}}\right)\,dt \\
&= \frac{\gamma_2}{2}\,\ln\left(\frac{1}{\sqrt{a}}\right)\,B\left(\frac{q}{2}+1,\frac{n-q}{2}\right)\,(1+o(1))\,,
\end{aligned}$$

where now $\gamma_2 = (2\,q\,\alpha\,c_{n-q-2}\,c_q\,c_2)/c_n$. This yields

$$\frac{e_k^{pow}}{1+o(1)} = q\,\alpha\,\ln\,\frac{1}{\sqrt{a}} = q\,\left(1-\frac{\lambda_3}{\lambda_1}\right)\,\left(\frac{\lambda_3}{\lambda_1}\right)^{2(k-1)}\,(k-1)\,\ln\,\frac{\lambda_1}{\lambda_3},$$

which proves part (c) for $p = 2$.

Finally assume that $p = 1$. Then for $a \to 0$ we have

$$\begin{aligned}
\int_0^1 \frac{(1-x^2)^{(n+1)/2}}{x^2+a\,(1-x^2)\,t^2}\,dx &= \int_0^1 \frac{1}{x^2+at^2}\,dx + O(1) \\
= \frac{1}{t\sqrt{a}}\,\arctan\,\frac{x}{t\sqrt{a}}\,\Big|_0^1 + O(1) &= \frac{1}{t\sqrt{a}}\,\frac{\pi}{2} + O(1).
\end{aligned}$$

Thus, we have

$$\begin{aligned}
&\frac{e_k^{pow}}{1+o(1)} = \gamma_2\,\frac{\pi}{2\sqrt{a}}\,\int_0^1 t^q\,(1-t^2)^{(n-q-1)/2}\,dt \\
= &\gamma_2\,\frac{\pi}{4\sqrt{a}}\,B\left(\frac{q+1}{2},\frac{n-q+1}{2}\right) = \sqrt{\pi}\left(1-\frac{\lambda_2}{\lambda_1}\right)\,\left(\frac{\lambda_2}{\lambda_1}\right)^{k-1}\,\frac{\Gamma((q+1)/2)}{\Gamma(q/2)},
\end{aligned}$$

with $\gamma_2 = (\alpha\, q\, c_{n-q-1}\, c_q\, c_1)/c_n$. This completes the proof of Theorem 3.1.

Proof of Theorem 3.2

The Lanczos algorithm takes the maximum of $(Ax, x)/(x, x)$ for $0 \neq x \in \text{span}(b, Ab, \ldots, A^{k-1}b)$, see (4). This means that $x = P(A)b$ for a nonzero polynomial from the class $\mathcal{P}_k$ of polynomials of degree $\leq k-1$. We have

$$\xi^{Lan} = \xi^{Lan}(A, b, k) = \max_{P \in \mathcal{P}_k} \frac{\sum_{i=1}^n b_i^2\, \lambda_i\, P^2(\lambda_i)}{\sum_{i=1}^n b_i^2\, P^2(\lambda_i)}.$$

The relative error of the Lanczos algorithm is given by

$$\frac{\lambda_1 - \xi^{Lan}}{\lambda_1} = \min_{P \in \mathcal{P}_k} \frac{\sum_{i=2}^n b_i^2\, P^2(\lambda_i)\,(1 - \lambda_i/\lambda_1)}{\sum_{i=1}^n b_i^2\, P^2(\lambda_i)}.$$

Using a continuity argument we may restrict ourselves to polynomials P such that $P(\lambda_1) \neq 0$. Let $Q(t) = P(\lambda_1\, t)/P(\lambda_1)$. Then $Q \in \mathcal{P}_k$ and $Q(1) = 1$. Let $\mathcal{P}_k(1)$ denote such polynomials. Thus, for $x_i = \lambda_i/\lambda_1 \in (0, 1]$ we have

$$\frac{\lambda_1 - \xi^{Lan}}{\lambda_1} = \inf_{Q \in \mathcal{P}_k(1)} \frac{\sum_{i=2}^n b_i^2\, Q^2(x_i)\,(1 - x_i)}{b_1^2 + \sum_{i=2}^n b_i^2\, Q^2(x_i)}. \tag{15}$$

As for the power algorithm, we conclude that the average relative error of the Lanczos algorithm is given by

$$e_k^{Lan} = e_k^{avg}(\xi^{Lan}, A, k) = \frac{1}{c_n} \int_{B_n} \inf_{Q \in \mathcal{P}_k(1)} \frac{\sum_{i=2}^n b_i^2\, Q^2(x_i)\,(1 - x_i)}{b_1^2 + \sum_{i=2}^n b_i^2\, Q^2(x_i)}\, db. \tag{16}$$

Assume first that $k \geq m$. This means that the set $\{x_1, x_2, \ldots, x_n\}$ contains m distinct elements $\{t_1, t_2, \ldots, t_m\}$ with $t_1 = 1$. Take

$$Q(x) = \prod_{i=2}^m (x - t_i)/(1 - t_i).$$

Then $Q \in \mathcal{P}_k(1)$ and the integrand in (16) vanishes for $b_1 \neq 0$. Since $b_1 = 0$ for a set of measure zero, we have $e_k^{Lan} = 0$, as claimed.

Assume now that $k \in [4, m-1]$. We find an upper bound on e_k^{Lan} by changing the order of integration and taking the infimum,

$$e_k^{Lan} \leq \frac{1}{c_n} \inf_{Q \in \mathcal{P}_k(1)} \int_{B_n} \frac{\sum_{i=2}^n b_i^2\, Q^2(x_i)\,(1 - x_i)}{b_1^2 + \sum_{i=2}^n b_i^2\, Q^2(x_i)}\, db.$$

Observe that to estimate the integral we can repeat the same reasoning as for the power algorithm with the polynomial Q instead of x^{k-1}. Therefore, for any $\beta \in [0,1]$ we have

$$e_k^{Lan} \le \frac{2\,c_{n-1}}{c_n} \inf_{Q\in\mathcal{P}_k(1)} \left(\frac{1}{c_{n-1}} \frac{\pi^2}{4} \int_{B'} \sum_{i:\,x_i\le\beta} b_i^2\, Q^2(x_i)\,(1-x_i)^2\, d\vec{b} \right.$$
$$\left. + \frac{1}{c_{n-1}}(1-\beta)^2 \int_{B'} (1-||b||_{n-1}^2)\, d\vec{b} \right)^{1/2}.$$

Let

$$w(\beta) = \inf_{Q\in\mathcal{P}_k(1)} \max_{0\le x\le\beta} Q^2(x)\,(1-x)^2. \tag{17}$$

Then

$$e_k^{Lan} \le \frac{2c_{n-1}}{c_n} \left(\frac{\pi^2}{4}\, w(\beta) \;+\; \frac{2(1-\beta)^2}{n} \right)^{1/2}$$

and using (13) we have

$$e_k^{Lan} \le 0.412\sqrt{\pi^2\, n\, w(\beta) \;+\; 8\,(1-\beta)^2}. \tag{18}$$

To get an upper bound on e_k^{Lan} we thus need to find an upper bound on $w(\beta)$ and select a proper β, see also Remark 7.4 in Section 7. Take

$$Q(x) = T_{k-1}\left((2/\beta)\,x-1\right) \,/\, T_{k-1}\left((2/\beta)-1\right),$$

where T_{k-1} is the Chebyshev polynomial of the first kind of degree $k-1$. Then

$$w(\beta) \le T_{k-1}^{-2}\left((2/\beta)\,-1\right) \le 4\left(\frac{1-\sqrt{1-\beta}}{1+\sqrt{1-\beta}}\right)^{2(k-1)} \le 4\,e^{-4(k-1)\sqrt{1-\beta}}. \tag{19}$$

Let $\gamma = \sqrt{1-\beta} \in [0,1]$. Then

$$e_k^{Lan} \le 0.824\sqrt{\pi^2\, n\, e^{-4(k-1)\gamma} \;+\; 2\,\gamma^4}.$$

Note that for $k-1 \le \sqrt{0.103}\, \ln(n(k-1)^4)$, part (a) of Theorem 3.2 trivially holds since

$$e_k^{Lan} \le 1 \le 0.103\, \left(\frac{\ln\, n(k-1)^4}{k-1}\right)^2.$$

Assume thus that $k-1 > \sqrt{0.103}\, \ln\, n(k-1)^4$. Take now

$$\gamma = \frac{1}{4(k-1)}\, \ln \frac{128\,\pi^2\, n\,(k-1)^4}{(\,\ln\, n(k-1)^4\,)^4}.$$

Since $128\pi^2 \leq (\ln n(k-1)^4)^4$ for $n \geq 8$ and $k \geq 4$, we have $\gamma \leq 1$. Clearly, $\gamma \geq 0$. A simple calculation yields

$$e_k^{Lan} \leq 0.103 \left(\frac{\ln n(k-1)^4}{k-1} \right)^2,$$

as claimed in part (a).

To prove part (b), define $\beta_1 = \lambda_n/\lambda_1$ and $\beta_2 = \lambda_{p+1}/\lambda_1$. Repeating the same reasoning that led to (18) we conclude that the sum for $x_i > \beta_2$ of the upper bound on e_k^{Lan} disappears and

$$e_k^{Lan} \leq 0.412\sqrt{\pi^2\, n\, w(\beta_1, \beta_2)},$$

where

$$w(\beta_1, \beta_2) = \inf_{Q \in \mathcal{P}_k(1)} \max_{\beta_1 \leq x \leq \beta_2} Q^2(x)\,(1-x)^2.$$

For $\beta = (\lambda_{p+1} - \lambda_n)/(\lambda_1 - \lambda_n)$ take

$$Q(x) = T_{k-1}\left(\frac{2(x-\beta_1)}{\beta_2 - \beta_1} - 1 \right) \,/\, T_{k-1}\left((2/\beta) - 1\right).$$

Then $w(\beta_1, \beta_2) \leq T_{k-1}^{-2}\left((2/\beta) - 1\right)$ and using the second inequality of (19), we get part (b). This completes the proof of Theorem 3.2

Proof of Theorem 4.1

We need to find the measure of the set

$$
\begin{aligned}
Z &= \left\{ b \in \mathbb{R}^n : \ \|b\| = 1, \ \frac{\sum_{i=2}^n b_i^2 \, x_i^{2(k-1)} (1 - x_i)}{b_1^2 + \sum_{i=2}^n b_i^2 \, x_i^{2(k-1)}} > \varepsilon \right\} \\
&= \left\{ b \in \mathbb{R}^n : \ \|b\| = 1, \ \sum_{i=2}^n b_i^2 \, (1 - \varepsilon - x_i) \, x_i^{2(k-1)} > \varepsilon \, b_1^2 \right\},
\end{aligned}
$$

where, as before, $x_i = \lambda_i / \lambda_1$.

Note that $H(x) = (1 - \varepsilon - x)\, x^{2(k-1)}$ for $x \in [0, 1]$ attains its maximum value at $x^* = (1 - \varepsilon)\,(1 - 1/(2k-1))$ and $H(x^*) = (1-\varepsilon)^{2k-1}\,(1 - 1/(2k-1))^{2(k-1)}/(2k-1)$. Then

$$
\sum_{i=2}^n b_i^2 \, (1 - \varepsilon - x_i) \, x_i^{2(k-1)} \le H(x^*) \sum_{i=2}^n b_i^2,
$$

and $Z \subset Z^*$, where

$$
Z^* = \left\{ b \in \mathbb{R}^n : \ \|b\| = 1, \ \sum_{i=2}^n b_i^2 > \alpha \, b_1^2 \right\}
$$

with

$$
\alpha = \frac{\varepsilon}{H(x^*)} = \frac{(2k-1)\,\varepsilon}{(1-\varepsilon)^{2k-1}\,(1 - 1/(2k-1))^{2(k-1)}}.
$$

Obviously,

$$
f^{prob}(\xi^{pow}, A, k, \varepsilon) = \mu(Z) \le \mu(Z^*).
$$

We have

$$
\begin{aligned}
1 - \mu(Z^*) &= \frac{2}{c_n} \int_0^1 \int_{\sum_{i=2}^n b_i^2 \le \min\{1 - b_1^2, \alpha \, b_1^2\}} db \\
&= \frac{2 \, c_{n-1}}{c_n} \int_0^1 \min\{1 - t^2, \alpha \, t^2\}^{(n-1)/2} \, dt.
\end{aligned}
$$

Observe that $\min\{1 - t^2, \alpha\, t^2\} = \alpha\, t^2$ for $t \le 1/\sqrt{1+\alpha} = g(k, \varepsilon)$, see (9) for the definition of g, and $\min\{1 - t^2, \alpha\, t^2\} = 1 - t^2$ for $t \ge g(k, \varepsilon)$. Therefore

$$
\begin{aligned}
1 \quad & - \mu(Z^*) = \gamma \left(\int_0^{g(k,\varepsilon)} (\alpha\, t^2)^j \, dt + \int_{g(k,\varepsilon)}^1 (1 - t^2)^j \, dt \right) \\
= \quad & \gamma \left(\frac{1}{n\sqrt{1+\alpha}} \left(\frac{\alpha}{1+\alpha} \right)^j + \int_0^1 (1 - t^2)^j \, dt - \int_0^{g(k,\varepsilon)} (1 - t^2)^j \, dt \right),
\end{aligned}
$$

where $j = (n-1)/2$ and $\gamma = 2\,c_{n-1}/c_n$. Since $c_n = 2\,c_{n-1}\int_0^1 (1-t^2)^{(n-1)/2}\,dt$, we get

$$\mu(Z^*) = \frac{2\,c_{n-1}}{c_n}\left(\int_0^{g(k,\varepsilon)} (1-t^2)^{(n-1)/2}\,dt - \frac{g(k,\varepsilon)}{n}\left(\frac{\alpha}{1+\alpha}\right)^{(n-1)/2}\right). \quad (20)$$

From (13) we have $\gamma \le 2.064\sqrt{n/(2\pi)}$ and

$$\mu(Z^*) \le 0.824\sqrt{n}\int_0^{g(k,\varepsilon)} (1-t^2)^{(n-1)/2}\,dt \le 0.824\sqrt{n}\,g(k,\varepsilon). \quad (21)$$

This and (9) complete the proof of part (a).

We proceed to part (b). It is clear that

$$f^{prob}(\xi^{pow}, \bar{A}, k, \varepsilon) = \mu(Z^*) = \max_{A=A^T>0} f^{prob}(\xi^{pow}, A, k, \varepsilon).$$

To estimate $\mu(Z^*)$ from below, note that $\gamma \ge \sqrt{2\,n/\pi}$ due to (13), and

$$\int_0^{g(k,\varepsilon)} (1-t^2)^{(n-1)/2}\,dt \ge g(k,\varepsilon)\left(\frac{\alpha}{1+\alpha}\right)^{(n-1)/2}.$$

Therefore

$$\mu(Z^*) \ge 0.797\sqrt{n}\left(1-\frac{1}{n}\right)\int_0^{g(k,\varepsilon)} (1-t^2)^{(n-1)/2}\,dt,$$

as claimed. The asymptotic formula follows from the estimates of (9).

To prove part (c), note that we need to find the measure of the set

$$W = \left\{ b \in \mathbb{R}^n : \|b\| = 1,\ \sum_{i=p+1}^{p+q} b_i^2\,(1-\varepsilon-x_{p+1})\,x_{p+1}^{2(k-1)} > \varepsilon \sum_{i=1}^{p} b_i^2 \right\}$$

since $f^{prob}(\xi^{pow}, A, k, \varepsilon) = \mu(W)(1+o(1))$ as $k \to \infty$.

Denote by $\beta = \varepsilon/((1-\varepsilon-x_{p+1})\,x_{p+1}^{2(k-1)})$, $a_p = \sum_{i=1}^p b_i^2$, $a_{p+q} = \sum_{i=1}^{p+q} b_i^2$, $a_p' = \sum_{i=p+1}^{p+q} b_i^2$. We have

$$\begin{aligned} 1-\mu(W) &= \frac{1}{c_n}\int_{a_p\le 1}\int_{a_p'\le \min\{1-a_p,\beta\,a_p\}}\int_{\sum_{i=p+q+1}^n b_i^2 \le 1-a_{p+q}} db \\ &= \frac{c_{n-p-q}}{c_n}\int_{a_p\le 1}\int_{a_p'\le \min\{1-a_p,\,\beta\,a_p\}} (1-a_p-a_p')^{(n-p-q)/2}\,db', \end{aligned}$$

where $db' = db_1\cdots db_{p+q}$. Using (12) twice we get

$$\begin{aligned} 1-\mu(W) &= \frac{q\,c_q\,c_{n-p-q}}{c_n}\int_{a_p\le 1}\int_0^{\min\{1-a_p,\beta\,a_p\}^{1/2}} t^{q-1}\,(1-a_p-t^2)^{(n-p-q)/2}\,dt \\ &= \omega\int_0^1 x^{p-1}\int_0^{\min\{1-x^2,\beta\,x^2\}^{1/2}} t^{q-1}\,(1-x^2-t^2)^{(n-p-q)/2}\,dt\,dx, \end{aligned}$$

with $\omega = p\,q\,c_p\,c_q\,c_{n-p-q}/c_n$.

Observe that by formally setting $\beta = +\infty$ we get $\mu(W) = 0$ and

$$\omega \int_0^1 x^{p-1} \left(\int_0^{\sqrt{1-x^2}} t^{q-1}\,(1-x^2-t^2)^{(n-p-q)/2}\,dt \right)\,dx = 1. \tag{22}$$

We thus have for $h(t,x) = t^{q-1}(1-x^2-t^2)^{(n-p-q)/2}$ and $a = (1-\mu(W))/\omega$,

$$\begin{aligned} a &= \int_0^{1/\sqrt{1+\beta}} x^{p-1} \int_0^{x\sqrt{\beta}} h(t,x)\,dt\,dx + \int_{1/\sqrt{1+\beta}}^1 x^{p-1} \int_0^{\sqrt{1-x^2}} h(t,x)\,dt\,dx \\ &= \int_0^1 x^{p-1} \int_0^{\sqrt{1-x^2}} h(t,x)\,dt\,dx - \int_0^{1/\sqrt{1+\beta}} x^{p-1} \int_{x\sqrt{\beta}}^{\sqrt{1-x^2}} h(t,x)\,dt\,dx. \end{aligned}$$

Due to (22) we get

$$\mu(W) = \omega \int_0^{1/\sqrt{1+\beta}} x^{p-1} \int_{x\sqrt{\beta}}^{\sqrt{1-x^2}} t^{q-1}(1-x^2-t^2)^{(n-p-q)/2}\,dt\,dx.$$

Changing variables by $\nu = x\,\sqrt{1+\beta}$, we obtain

$$\mu(W) = \frac{\omega}{(1+\beta)^{p/2}} \int_0^1 \nu^{\,p-1} \int_{\nu\sqrt{\beta/(1+\beta)}}^{\sqrt{1-\nu^2/(1+\beta)}} t^{q-1} \left(1 - \frac{\nu^2}{1+\beta} - t^2\right)^{(n-p-q)/2} dt\,d\nu.$$

Note that $\beta \to +\infty$ as $k \to +\infty$. Therefore we have

$$\begin{aligned} \frac{\mu(W)}{1+o(1)} &= \omega\,\beta^{-p/2} \int_0^1 \nu^{p-1} \int_\nu^1 t^{q-1}(1-t^2)^{(n-p-q)/2}\,dt\,d\nu \\ &= \omega\,\beta^{-p/2} \int_0^1 \left(\int_0^t \nu^{\,p-1}\,d\nu \right) t^{q-1}(1-t^2)^{(n-p-q)/2}\,dt \\ &= \frac{\omega}{p\,\beta^{p/2}} \int_0^1 t^{p+q-1}(1-t^2)^{(n-p-q)/2}\,dt \\ &= \frac{\omega}{2\,p\,\beta^{p/2}}\, B\left(\frac{p+q}{2}, \frac{n-p-q}{2}+1\right), \end{aligned}$$

the last equality due to (14). To complete the proof it is enough to observe that

$$\begin{aligned} \frac{\omega}{2\,p} B\left(\frac{p+q}{2}, \frac{n-p-q}{2}+1\right) &= \frac{p\,q\,\Gamma(1+n/2)\,\Gamma((p+q)/2)\,\Gamma(j)}{2p\Gamma(1+p/2)\Gamma(1+q/2)\Gamma(j)\Gamma(1+n/2)} \\ &= \frac{q\,\Gamma((p+q)/2)}{2\,\frac{p}{2}\,\Gamma(p/2)\,\frac{q}{2}\,\Gamma(q/2)} = \frac{2}{p}\,\frac{\Gamma((p+q)/2)}{\Gamma(p/2)\,\Gamma(q/2)}, \end{aligned}$$

where $j = 1 + (n-p-q)/2$.

Proof of Theorem 4.2

We need to find an upper bound on the measure of the set

$$Z = \{\, b: \ ||b|| = 1, \ \lambda_1 - \xi^{Lan}(A,b,k) \ > \ \varepsilon\,\lambda_1 \,\}.$$

Due to (15) we have

$$Z = \{\, b: \ ||b|| = 1, \ \inf_{Q \in \mathcal{P}_k(1)} \sum_{i=2}^{n} b_i^2\, Q^2(x_i)\,(1-\varepsilon-x_i) \ > \ b_1^2\,\varepsilon \,\}.$$

Obviously

$$f_k^{Lan} = f^{prob}(\xi^{Lan}, A, k, \varepsilon) = \mu(Z).$$

Assume first that $k \geq m$. As in the proof of Theorem 3.2, $\{x_1, x_2, \ldots, x_n\} = \{t_1, t_2, \ldots, t_m\}$ with distinct t_i and $t_1 = 1$. Setting $Q(x) = \prod_{i=2}^{m}(x-t_i)/(1-t_i)$ we get $Z = \emptyset$. Thus $f_k^{Lan} = 0$, as claimed.

Take now an arbitrary k. For $\varepsilon = 0$, the remaining bound of part (a) of Theorem 4.2 trivially holds. (In fact, it is easy to see that for $k < m$, we have $f^{prob}(\xi^{Lan}, A, k, 0) = 1$.) Assume thus that $\varepsilon > 0$ and let

$$w_k \ = \ \inf_{Q \in \mathcal{P}_k(1)} \ \max_{0 \leq x \leq 1} \ Q^2(x)\,(1-\varepsilon-x). \tag{23}$$

Then

$$Z \subset Z^* \ = \ \{\, b: \ ||b|| = 1, \ \sum_{i=2}^{n} b_i^2 \ > \ b_i^2\,\varepsilon / w_k \,\}$$

and $f_k^{Lan} \leq \mu(Z^*)$.

Observe that an upper bound on the measure of the set Z^* was found in (21),

$$f_k^{Lan} \leq 0.824\,\sqrt{n}\,g(k,\varepsilon) \ = \ 0.824\,\sqrt{\frac{n}{1+\varepsilon/w_k}}, \tag{24}$$

where now $g(k,\varepsilon) = 1/\sqrt{1+\alpha}$ with $\alpha = \varepsilon/w_k$. We prove that

$$w_k \ = \ 4\,\varepsilon\,\left(\frac{1-\sqrt{\varepsilon}}{1+\sqrt{\varepsilon}}\right)^{2k-1} \left(1 - \left(\frac{1-\sqrt{\varepsilon}}{1+\sqrt{\varepsilon}}\right)^{2k-1}\right)^{-2}.$$

Let $U_{2(k-1)}$ be the Chebyshev polynomial of the second kind of degree $2(k-1)$. Consider

$$Q(x) \ = \ U_{2(k-1)}\big(\sqrt{x/(1-\varepsilon)}\big) \,/\, U_{2(k-1)}(1/\sqrt{1-\varepsilon}), \quad x \in [0,1].$$

Since $U_{2(k-1)}$ is even, Q is a polynomial of degree $k-1$. Clearly, $Q(1) = 1$, so $Q \in \mathcal{P}_k(1)$. Let

$$H(x) = \sqrt{1-\varepsilon-x}\,Q(x), \quad x \in [0, 1-\varepsilon].$$

For $t_i = (1-\varepsilon)\cos^2\frac{(2i-1)\pi}{2(2k-1)}$, $i = 1, 2, \ldots, k$, the extremal points of $U_{2(k-1)}$ yield

$$H(t_i) = \frac{\sqrt{1-\varepsilon}}{U_{2(k-1)}(1/\sqrt{1-\varepsilon})}\,(-1)^{i-1}.$$

Note that

$$w_k \le a := \max_{0\le x\le 1-\varepsilon} H^2(x) = \frac{1-\varepsilon}{U^2_{2(k-1)}(1/\sqrt{1-\varepsilon})} = 4\varepsilon\, c\,(1-c)^{-2},$$

where $c = (\,(1-\sqrt{\varepsilon})/(1+\sqrt{\varepsilon})\,)^{2k-1}$.

Assume that $w_k < a$. Then there exists a polynomial $P \in \mathcal{P}_k(1)$ such that $\max_{x\in[0,1]} P^2(x)(1-\varepsilon-x) < w_k$. The sign of the function

$$h(x) = \sqrt{1-\varepsilon-x}\,(Q(x)-P(x))\,, \quad x \in [0, 1-\varepsilon],$$

alternates at t_i for $i = 1, 2, \ldots, k$. Thus, $Q - P$ has at least $k-1$ zeros in $[0, 1-\varepsilon)$. Since $x = 1$ is also a zero of $Q - P$ we conclude that $Q = P$, which is a contradiction. Hence $w_k = a$, as claimed.

From this and (24) we finally get

$$\begin{aligned} f_k^{Lan} &\le 0.824\sqrt{4n/(4+(1-c)^2/c)} \\ &\le 0.824\sqrt{4n/(2+1/c)} \le 1.648\sqrt{c\,n}. \end{aligned}$$

Part (a) follows by noting that $\sqrt{c} \le \exp(-\sqrt{\varepsilon})$.

To prove part (b), let $\beta_1 = \lambda_n/\lambda_1$, $\beta_2 = \lambda_{p+1}/\lambda_1$ and

$$u(\beta_1, \beta_2) = \inf_{Q\in\mathcal{P}_k(1)} \max_{\beta_1\le x\le\beta_2} Q^2(x)\,(1-\varepsilon-x).$$

(Observe that $u(0, \beta_2) = w_k$ for $\beta_2 \ge 1-\varepsilon$.) Then

$$Z \subset \{\, b:\ ||b|| = 1,\ \sum_{i=2}^{n} b_i^2 \,>\, b_1^2\,\varepsilon/u(\beta_1,\beta_2)\,\}$$

and $f_k^{Lan} \le 0.824\sqrt{n/(1+\varepsilon/u(\beta_1,\beta_2))} \le 0.824\sqrt{n\,u(\beta_1,\beta_2)/\varepsilon}$. We need to estimate $u(\beta_1, \beta_2)$. Changing the variables $x = (1-\beta_1)t + \beta_1$ we get

$$u(\beta_1, \beta_2) \le \max_{0\le t\le 1-\lambda^*} Q^2(t) \quad \forall Q \in \mathcal{P}_k(1),$$

where $\lambda^* = (\lambda_1 - \lambda_{p+1})/(\lambda_1 - \lambda_n)$. We can use now the estimate (19) with $\beta = 1 - \lambda^*$ to get

$$u(\beta_1, \beta_2) \le 4\,\left(\frac{1-\sqrt{\lambda^*}}{1+\sqrt{\lambda^*}}\right)^{2(k-1)}$$

which yields part (b) and completes the proof.

6 Numerical Tests

We have tested the Lanczos algorithm with reorthogonalization in single precision for several matrices and many (pseudo) random vectors b. We report numerical results for one matrix A for which the relative errors of the Lanczos algorithm were the largest. The matrix A was chosen as follows. Observe that for any orthogonal matrix Q we have $\xi^{Lan}(Q^T A\,Q, Q^T b, k) = \xi^{Lan}(A, b, k)$. Note also that the uniform distribution on the unit sphere of vectors b implies the same distribution of vectors $Q^T b$. This shows that without loss of generality we can restrict ourselves to diagonal matrices while testing the Lanczos algorithm. Therefore the matrix A was taken as diagonal. We chose the dimension $n = 250$ and the eigenvalues of A as

$$\lambda_i = 1 + \cos\frac{(2i-1)\pi}{2n}, \quad i = 1, 2, \ldots, n.$$

That is, the eigenvalues of A are shifted zeros of the Chebyshev polynomial T_n and $\lambda_1 = 1 + \cos(\,\pi/(2n)\,) \simeq 2$. (The shift by 1 is needed to guarantee that A is positive definite.)

We have performed numerical tests for this matrix with 30 pseudo-random vectors b uniformly distributed over the unit sphere of $\mathbb{R}^{\kappa}$. To get such a distribution we used the fact that if $X = (X_1, X_2, \ldots, X_n)$ is a random variable whose components are independent random variables with a normal distribution $N(0,1)$ then $X/||X||$ is uniformly distributed over the unit sphere, see Knuth [81, p.116]. The normal distribution was in turn generated from the uniform distribution over $(0,1)$ using the formula $Z = (-2\,\ln\,R_1)^{1/2}\cos 2\pi R_2$, where R_1 and R_2 are independent random variables uniformly distributed over $(0,1)$, see Box and Muller [58]. The variables R_i were produced using a number generator similar to that one used for testing EISPACK procedures, see Smith et al [74].

For each pseudo-random vector b we performed the Lanczos algorithm for $k = 1, 2, \ldots, k^*$, where k^* was chosen as the minimal k for which the relative error $(\,\lambda_1 - \xi^{Lan}(A, b, k)\,)/\lambda_1$ was no greater than ε. For some tests k^* was around 150. We compared the relative error with k^{-2}. For all tested b and k we obtained

$$0.1241 \;\le\; \frac{\lambda_1 - \xi^{Lan}(A, b, k)}{\lambda_1}\,k^2 \;\le 1.62.$$

In fact, in most cases $(\,\lambda_1 - \xi^{Lan}(A, b, k)\,)/\lambda_1\,k^2$ was between 0.286 and 1.25.

In the table below we report the average errors achieved after $k-1$ steps of the Lanczos algorithm for ten different values of k which are listed in the first column. The second column contains the average errors defined as

$$\varepsilon^{ave} = \frac{1}{30}\sum_{i=1}^{30}\frac{\lambda_1 - \xi^{Lan}(A, b_i, k)}{\lambda_1},$$

where b_i is the ith pseudo-random vector. The third column presents upper bounds on the Lanczos errors from Theorem 3.2, i.e.,

$$\varepsilon^{up} = 0.103 \left(\frac{\ln \left(n(k-1)^4 \right)}{k-1} \right)^2 .$$

We compute the ratios between the observed errors and their upper bounds in the fourth column, $r_1 = \varepsilon^{up}/\varepsilon^{ave}$. The fifth column displays how r_1 is related to the possibly unnecessary factor in the theoretical bound,

$$r_2 = 0.103\, r_1 / \ln^2(n(k-1)^4).$$

The last two columns show the minimal ε^{min} and maximal ε^{max} relative errors over the set of initial vectors at each step $k-1$ respectively.

$k-1$	ε^{ave}	ε^{up}	r_1	r_2	ε^{min}	ε^{max}
10	0.004830	0.2235	46.273	2.070	0.003782	0.006868
20	0.001333	0.0789	59.185	1.875	0.000853	0.002109
30	0.000623	0.0419	67.288	1.786	0.000329	0.001454
40	0.000360	0.0265	73.632	1.739	0.000145	0.000943
50	0.000226	0.0185	81.714	1.770	0.000080	0.000463
60	0.000156	0.0137	87.596	1.773	0.000051	0.000295
70	0.000115	0.0107	92.897	1.779	0.000040	0.000230
80	0.000090	0.0086	95.875	1.752	0.000034	0.000182
90	0.000072	0.0070	97.765	1.716	0.000028	0.000159
100	0.000061	0.0059	97.360	1.649	0.000028	0.000132

The fifth column of the table seems to suggest that the error of the Lanczos algorithm for the matrix with Chebyshevian distribution of eigenvalues behaves like k^{-2} and the factor $0.103 \ln^2(n(k-1)^4)$ is probably an overestimate of the upper bound.

In the next table we indicate how many steps were needed to achieve relative error no greater than ε for six different values of ε. The values of ε are displayed in the first row of the table. The second row of the table shows the average number k^{ave} of performed steps with $k^{ave} = \sum_{i=1}^{30} k(A, b_i)/30$, where $k(A, b_i)$ was the number of steps needed for the pseudo-random vector b_i. The third row gives the minimal $k = k^{up}$ such that

$$0.103 \left(\frac{\ln(n(k-1)^4)}{k-1} \right)^2 \le \varepsilon,$$

which is one of the two theoretical bounds for the Lanczos algorithm, see part (a) of Theorem 3.2. The fourth row presents the ratios between these two numbers, $r = k^{up}/k^{ave}$.

ε	$5.0_{10}-4$	$2.5_{10}-4$	$2.0_{10}-4$	$1.5_{10}-4$	$1.0_{10}-4$	$5.0_{10}-5$
k^{ave}	35.27	48.03	53.13	62.27	76.33	110.67
k^{up}	428	638	724	853	1075	1590
r	12.13	13.28	13.63	13.70	14.08	14.38

As we see the theoretical bound exceeds the actual value by a factor of at most 15. This indicates once more that the factor $\ln^2(n(k-1)^4)$ may be an overestimate in the theoretical bound. Observe also that all k^{up}s are greater than the dimension $n = 250$ and the second bound of part (a) of Theorem 3.2 gives a better estimate.

We complete this section by reporting an interesting property of the computed sequences $\xi_k = \xi^{Lan}(A, b, k)$ of the Lanczos algorithm. In some cases they have a "misconvergence" phenomenon, see Parlett, Simon and Stringer [82]. That is, before reaching the largest eigenvalue λ_1, the sequence ξ_k remained constant (within to a machine accuracy) for some consecutive steps, $\xi_k = \xi_{k+1} = \ldots = \xi_{k+t}$ and the value of t was sometimes quite large. The misconvergence phenomenon occurred when the sequence ξ_k approached the second largest eigenvalue λ_2 and sometimes even when ξ_k passed the third largest eigenvalue λ_3. For instance, for some vectors b the sequence ξ_k stabilized close to λ_2 for 28 consecutive steps. The table below shows the percentage (p) of the vectors b for which the misconvergence phenomenon occurred before the relative error reached ε.

ε	$2.5_{10}-4$	$2_{10}-4$	$1.5_{10}-4$	$1.0_{10}-4$	$5.0_{10}-5$
p	0	6.67	46.67	80	100

7 Remarks

Remark 7.1

As we know from Section 2 it is impossible to compute an ε-approximation to the largest eigenvalue by algorithms using Krylov information with a deterministically chosen vector b. One may interpret this by saying that Krylov information is poor and hope that more general information may lead to a positive result. Indeed, using matrix-vector multiplications we may compute $[Az_1, Az_2, \ldots, Az_k]$, where $z_1 = b$ and z_i can be an arbitrary function of the already computed $Az_1, Az_2, \ldots, Az_{i-1}$. Is it possible to define vectors z_i such that $\phi(Az_1, Az_2, \ldots, Az_k)$, for some ϕ, yields an ε-approximation to the largest eigenvalue of any symmetric positive definite matrix A? The answer is still *no* as long as $k \leq n-1$, see Traub, Wasilkowski and Woźniakowski [88, p.183-186] for this and related results. Thus, Krylov information as well as any other deterministic information with $k \le n-1$ does not supply enough knowledge of A to compute an ε-approximation to the largest eigenvalue.

On the other hand, if we are willing to settle for an ε-approximation to any eigenvalue, which is not necessarily the largest, then it can be done by using $\min\{n, \lceil \varepsilon^{-1} \rceil\}$ matrix-vector multiplications. This can be achieved by using deterministic Krylov information and the generalized minimal residual algorithm, see Kuczyński [86]. The number $\min\{n, \lceil \varepsilon^{-1} \rceil\}$ is within a factor of at most 2 of being minimal as shown by Chou [87] whose analysis is based on Nemirovsky and Yudin [83].

Remark 7.2

As before B_n is the unit ball of $\mathbb{R}^n$. Let $f : B_n \to \mathbb{R}$ be a measurable function such that f does not depend on the norm of b, $f(b) = f(\alpha\, b)$, $\forall \alpha > 0$, and f does not depend on signs of b_i, $f(s_1 b_1, s_2 b_2, \dots, s_n b_n) = f(b_1, b_2, \dots, b_n)$ for all $s_i \in \{-1, 1\}$. The error of the power or Lanczos algorithm as a function of b satisfies these properties.

For such functions f, the average value of f over the unit sphere is the same as the average value over the unit ball, i.e.,

$$\int_{||b||=1} f(b)\, \mu(db) = \frac{1}{c_n} \int_{B_n} f(b)\, db,$$

where c_n is the measure of the unit ball in $\mathbb{R}^n$.

Indeed, using the polar coordinates $b = \phi(t) = [\phi_1(t), \dots, \phi_n(t)]$ with $t = [r, t_1, \dots, t_{n-1}] \in [0,1] \times [0, \pi]^{n-1}$ and

$$\begin{aligned} \phi_1(t) &= r \cos t_1 \cos t_2 \cdots \cos t_{n-1}, \\ \phi_{i+1} &= r \sin t_i \cos t_{i+1} \cdots \cos t_{n-1}, \quad i = 1, 2, \dots, n-1, \end{aligned}$$

we have $|\det(\phi')| = r^{n-1} |\cos t_2 \cos^2 t_3 \cdots \cos^{n-2} t_{n-1}| = r^{n-1}\, g(t)$ and

$$a\, c_n := \int_{B_n} f(b)\, db = \int_{[0,1]\times[0,\pi]^{n-1}} f(\phi(t))\, r^{n-1}\, g(t)\, dr\, dt_{(n)},$$

where $dt_{(n)}$ stands for $dt_1 \cdots dt_{n-1}$. Since $f(\phi(t))$ does not depend on r, we can integrate over r to get

$$a = \frac{1}{n\, c_n} \int_{[0,\pi]^{n-1}} f(\phi(t))\, g(t)\, dt_{(n)}.$$

Change the variables once more by setting $b_i = \phi_i(t)/r$ for $i = 2, 3, \dots, n$. Then for $b_1 = \sqrt{1 - \sum_{i=2}^n b_i^2}$ we have

$$f(\phi(t)) = f(\phi(t)/r) = f(\pm b_1, b_2, \dots, b_n) = f(b_1, b_2, \dots, b_n)$$

and

$$db_{(n)} = db_2 \cdots db_n = |\cos t_1 \; \cos^2 t_2 \cdots \cos^{n-1} t_{n-1}| \, dt_{(n)}.$$

Therefore $g(t)\, dt_{(n)} = |\cos t_1 \cdots \cos t_{n-1}|^{-1} db_{(n)} = (1 - \sum_{i=2}^n b_i^2)^{-1/2} db_{(n)}$ and

$$a = \frac{2}{n\, c_n} \int_{b_2^2+\ldots+b_n^2 \le 1} \frac{f(\sqrt{1-\sum_{i=2}^n b_i^2}, b_2, \ldots, b_n)}{\sqrt{1-\sum_{i=2}^n b_i^2}} \, db_{(n)} = \int_{||b||=1} f(b)\, \mu(db),$$

as claimed.

Remark 7.3

The modified power algorithm is defined by

$$\xi^{mpow}(A, b, k) = (A^k b, b)^{1/k}, \quad ||b|| = 1.$$

We show that

$$\sup_{A=A^T>0} e^{avg}(\xi^{mpow}, A, b) = e^{avg}(\xi^{mpow}, A^*, k) = 1 - \frac{\Gamma(n/2)\Gamma(0.5+1/k)}{\Gamma(n/2+1/k)\Gamma(0.5)} \simeq \frac{\ln\, n}{k},$$

where A^* is a symmetric matrix with eigenvalues $\lambda_1 > 0$, and $\lambda_i = 0$ for $i \ge 2$.

Indeed, for any $A = A^T > 0$ with $x_i = \lambda_i/\lambda_1$ we have

$$\begin{aligned}
& e^{avg}(\xi^{mpow}, A, k) = 1 - \int_{||b||=1} (\sum_{i=1}^n b_i^2 x_i^k)^{1/k} \mu(db) \le e^{avg}(\xi^{mpow}, A^*, k) = \\
1 & - \int_{||b||=1} b_1^{2/k} \, \mu(db) = 1 - \frac{2}{n\, c_n} \int_{B_{n-1}} (1 - b_2^2 - \ldots - b_n^2)^{\frac{1}{k}-\frac{1}{2}} \, db_2 \cdots db_n = \\
1 & -2\alpha \int_0^1 t^{n-2} (1-t^2)^{\frac{1}{k}-\frac{1}{2}} \, dt = 1 - \alpha\, B\left(\frac{n-1}{2}, \frac{1}{k} + \frac{1}{2}\right) = \\
1 & - \frac{\Gamma(n/2)\Gamma(0.5+1/k)}{\Gamma(n/2+1/k)\Gamma(0.5)},
\end{aligned}$$

where $\alpha = (n-1)c_{n-1}/(nc_n)$. For large k and any a we have

$$\frac{\Gamma(a+1/k)}{\Gamma(a)} = 1 + \frac{\Gamma'(a)}{\Gamma(a)} \frac{1}{k} (1 + o(1)).$$

For $a = n/2$ and $a = 1/2$ we have from Gradshteyn and Ryzhik [80, 8.360, 8.362 and 8.366]

$$\frac{\Gamma'(n/2)}{\Gamma(n/2)} \simeq \ln\, n/2, \qquad \frac{\Gamma'(1/2)}{\Gamma(1/2)} = -C - 2\,\ln\, 2 = -1.9635\ldots,$$

where C is the Euler constant. This implies the error behavior $\ln(n)/k$, as claimed.

Comparing this bound with parts (a) and (b) of Theorem 3.1, we see that the power algorithm has an error bound roughly 1.8 times smaller.

One can also compare the algorithms ξ^{pow} and ξ^{mpow} asymptotically. Assume for simplicity that the largest eigenvalue is of multiplicity $p = 1$. Then part (c) of Theorem 3.1 yields that the rate of convergence of the power algorithm is exponential and proportional to $(\lambda_2/\lambda_1)^{k-1}$. For the modified power algorithm it is easy to show that the rate is only linear and roughly equal to $\ln(n)/k$. Thus, the power algorithm is far superior asymptotically in the average case to the modified power algorithm.

We now turn to the probabilistic case. The modified power algorithm was analyzed in this case by Dixon [83] who proved that

$$\sup_{A=A^T>0} f^{prob}(\xi^{mpow}, A, k, \varepsilon) = f^{prob}(\xi^{mpow}, A^*, k, \varepsilon) \le 0.8\,\sqrt{n}\,(1-\varepsilon)^{k/2}.$$

For large n and k we have

$$f^{prob}(\xi^{mpow}, A^*, k, \varepsilon) = \sqrt{2/\pi}\;\sqrt{n}\;(1-\varepsilon)^{k/2}\,(1+o(1)),$$

and $\sqrt{2/\pi} = 0.7978\ldots$.

This should be compared with the power algorithm whose rate of convergence is roughly the square of the rate of the modified power algorithm.

It is easy to check that the asymptotic behavior of ξ^{mpow} does not depend on the distribution of eigenvalues but depends on the multiplicity p of the largest eigenvalue,

$$f^{prob}(\xi^{mpow}, A, k, \varepsilon) = \frac{n-p}{n}\,\frac{\Gamma(1+n/2)}{\Gamma(1+p/2)\Gamma(1+(n-p)/2)}\,(1-\varepsilon)^{pk/2}(1+o(1)).$$

For the power algorithm with $\lambda_{p+1}/\lambda_1 < 1-\varepsilon$, the asymptotic rate of convergence is proportional to $(\lambda_{p+1}/\lambda_1)^{p(k-1)}$ which obviously tends to zero faster.

Remark 7.4

For β close to one it is easy to find the exact value $w(\beta)$, see (17). Namely, we have

$$w(\beta) = \frac{1}{k^2}\,\frac{\sin^2(\pi/(2k))}{(1+\cos(\pi/(2k))^2}\;\simeq\;\frac{\pi^2}{16\,k^4}$$

for $\beta \in [\cos^2(\pi/(2k))/\cos^2(\pi/(4k)), 1]$.

Indeed, let $\zeta_k = \cos(\pi/(2k))$ denote the largest zero of T_k. Take

$$Q(x) = \frac{1}{x-1}\,\frac{T_k((\zeta_k+1)x-1)}{(\zeta_k+1)\,T_k'(\zeta_k)}.$$

Note that Q is a polynomial of degree $\leq k-1$ and $Q(1)=1$. For $i=1,2,\ldots,k$, let $x_i = (1+\cos(i\pi/k))/(1+\zeta_k) = \cos^2(i\pi/(2k))/\cos^2(\pi/(4k))$. Then $x_i \in [0,\beta]$ and

$$(x_i-1)\,Q(x_i) = \frac{T_k(\cos(i\pi/k))}{(\zeta_k+1)T_k'(\zeta_k)} = \frac{(-1)^i}{1+\cos(\pi/(2k))}\,\frac{\sin(\pi/(2k))}{k}.$$

Suppose there exists $P \in \mathcal{P}_k(1)$ such that

$$\max_{x\in[0,\beta]} P^2(x)(1-x)^2 \;<\; \max_{x\in[0,\beta]} Q^2(x)(1-x)^2.$$

Then $h(x) = (1-x)(Q(x)-P(x))$ has a double zero at one. Since sign $h(x_i) = (-1)^i$, h has at least $k-1$ zeros in $[0,\beta)$. Thus $h\equiv 0$, which is a contradiction. Hence,

$$w(\beta) = \max_{x\in[0,\beta]} Q^2(x)(1-x)^2 = \frac{\sin^2(\pi/(2k))}{(1+\cos(\pi/(2k)))^2}\,\frac{1}{k^2},$$

as claimed.

Remark 7.5

We now consider an error relative to spread, see Parlett [89], instead of the relative error as the error criterion. That is, we wish to compute ξ such that

$$|\lambda_1(A)-\xi| \;\leq\; \varepsilon\,(\lambda_1(A)-\lambda_n(A)),$$

where, as before, $\lambda_1(A)$ and $\lambda_n(A)$ denote the largest and the smallest eigenvalues of A.

This error is a natural error criterion for the Lanczos algorithm since $\xi^{Lan}(A+\alpha\,I,b,k) = \xi^{Lan}(A,b,k)+\alpha$ and

$$\frac{\lambda_1(A+\alpha\,I)-\xi^{Lan}(A+\alpha\,I,b,k)}{\lambda_1(A+\alpha\,I)-\lambda_n(A+\alpha\,I)} = \frac{\lambda_1(A)-\xi^{Lan}(A,b,k)}{\lambda_1(A)-\lambda_n(A)}.$$

Thus, for the Lanczos algorithm the error relative to spread is shift invariant.

It is easy to see that the bounds for the Lanczos algorithm presented in Theorems 3.2 and 4.2 also hold for the error relative to spread. This follows by noting that

$$\frac{\lambda_1(A)-\xi^{Lan}(A,b,k)}{\lambda_1(A)-\lambda_n(A)} = \frac{\lambda_1(B)-\xi^{Lan}(B,b,k)}{\lambda_1(B)},$$

where $B = A \;-\; \lambda_n\,I$ and $B=B^*\geq 0$.

Although B is not positive definite, a continuity argument yields that we can use estimates of Theorems 3.2 and 4.2 for the matrix B. Parts (a) of these theorems will give estimates independent of eigenvalue distributions of B (or A). Parts (b) present estimates which are shift invariant and therefore are the same for the matrix B as well as for the matrix A. Observe also that for the error relative to spread we need only to assume that A is symmetric but not necessarily positive definite.

For the power algorithm we have different results since, in general, $\xi^{pow}(A+\alpha I, b, k) \neq \xi^{pow}(A, b, k) + \alpha$. To derive bounds for the power algorithm under the error relative to spread, consider the average case and the matrix A from part (b) of Theorem 3.1. That is, A has exactly two distinct eigenvalues λ_1 and $\lambda_n = \lambda_1(1 - \ln(n/\ln n)/(2(k-1)))$. Then the estimate of part (b) of Theorem 3.2 yields for large k and n,

$$\int_{||b||=1} \frac{\lambda_1 - \xi^{pow}(A, b, k)}{\lambda_1 - \lambda_n} \, \mu(db) = \frac{e^{avg}(\xi^{pow}, A, k)}{1 - \lambda_n/\lambda_1} = 1 + o(1).$$

Thus, no matter how many matrix-vector multiplications are performed, there exists a matrix A for which the average error of the power algorithm under the error relative to spread is about 1.

Similarly one can check that in the probabilistic case, the failure of the power algorithm under the error relative to spread for the matrix A with the two distinct eigenvalues λ_1 and $\lambda_1(1 - 1/(2k-1))$ is equal to $1 + o(1)$.

Obviously, the asymptotic bounds for the power algorithm under the error relative to spread can be easily obtained from parts (c) of Theorems 3.1 and 4.1. For the average case, the only difference is to multiply the asymptotic constants by $1 - \lambda_n/\lambda_1$, whereas for the probabilistic case, ε should be replaced by $\varepsilon\,(1 - \lambda_n/\lambda_1)$.

Acknowledgment

We appreciate comments from J. F. Traub, G. W. Wasilkowski and A. G. Werschulz.

Our special thanks are to B. N. Parlett for raising the issue of using error relative to spread and many valuable suggestions.

We also thank the anynomous referee for the carefull checking of the paper and numerical tests.

References

[1] G. E. P. Box and M. E. Muller. A Note on the Generation of Random Normal Deviates. *Ann. Math. Statist.* **29** (1958), 610–611.

[2] A. W. Chou. On the Optimality of Krylov Information. *J. Complexity* **3** (1987), 26–40.

[3] J. D. Dixon. Estimating Extremal Eigenvalues and Condition Numbers of Matrices. *SIAM J. Numer. Anal.* **20** (1983), 812–814.

[4] I. S. Gradshteyn and I. W. Ryzhik. *Table of Integrals, Series, and Products.* Academic Press, New York, 1980.

[5] W. Kahan and B. N. Parlett. How Far Should We Go with the Lanczos Process? in *Sparse Matrix Computations* eds. J. Bunch and D. Rose, Academic Press, New York, (1976), 131–144.

[6] S. Kaniel. Estimates for Some Computational Techniques in Linear Algebra. *Math. Comp.* **20** (1966), 369–378.

[7] D. E. Knuth. *The Art of Computer Science, Vol. 2: Seminumerical Algorithms.* 2nd ed. Addison-Wesley, Reading, MA., 1981.

[8] J. Kuczyński. On the Optimal Solution of Large Eigenpair Problems. *J. Complexity* **2** (1986), 131–162.

[9] A. S. Nemirovsky and D. B. Yudin. *Problem Complexity and Method Efficiency in Optimization.* Wiley-Interscience, New York, 1983.

[10] D. P. O'Leary, G. W. Stewart and J. S. Vandergraft. Estimating the Largest Eigenvalue of a Positive Definite Matrix. *Math. Comput.* **33** (1979), 1289–1292.

[11] C. C. Paige. *The Computation of Eigenvalues and Eigenvectors of Very Large Sparse Matrices.* Ph. D. Thesis, University of London, 1971.

[12] C. C. Paige. Computational Variants of the Lanczos Method for the Eigenproblem. *J. Inst. Math. Appl.* **10** (1972), 373–381.

[13] B. N. Parlett. *The Symmetric Eigenvalue Problem.* Prentice-Hall, Inc., Englewood Cliffs, N. J., 1980.

[14] B. N. Parlett. Private communication.

[15] B. N. Parlett, H. Simon and L. M. Stringer. On Estimating the Largest Eigenvalue with the Lanczos Algorithm. *Math. Comput.* **38** (1982), 153–165.

[16] Y. Saad. On the Rates of Convergence of the Lanczos and the Block Lanczos Methods. *SIAM J. Numer. Anal.* **17** (1980), 687–706.

[17] D. S. Scott. *Analysis of the Symmetric Lanczos Process.* Ph. D. Thesis, University of California at Berkeley, Berkeley, CA., Memorandum NCB/ERLM 78/40, 1978.

[18] B. T. Smith, J. M. Boyle, B. S. Garbov, Y. Ikebe, V. C. Klema, C. B. Moler. *Matrix Eigensystem Routines-EISPACK Guide.* Springer Verlag, Berlin, 1974.

[19] J. F. Traub, G. W. Wasilkowski and H. Woźniakowski. *Information-Based Complexity.* Academic Press, New York, 1988.

[20] J. H. Wilkinson. *The Algebraic Eigenvalue Problem.* Oxford Univ. Press, London and New York, 1965.