Statistical Complexity of Dominant Eigenvector Calculation

Eric Kostlan*

Received July 12, 1990

Abstract

We show that the number of arithmetic operations required to calculate a dominant ϵ-eigenvector of a real symmetric or complex Hermitian $n \times n$ matrix, when averaged over any density invariant under linear transformations that preserve the Frobenius norm, is bounded above by a polynomial in the size of the matrix. In fact, a specific upper bound is given in terms of n and ϵ. We also describe an estimate of the distance between an arbitrary complex $n \times m$ matrix and its rank one approximation.

Introduction

We present upper bounds on the statistical complexity of dominant ϵ-eigenvector calculation. We assume exact arithmetic throughout. See Blum *et al.* (1989), Smale (1985), and Traub *et al.* (1988) for a general discussion of this methodology. We restrict our attention to real symmetric and complex Hermitian matrices until Section 4. However, Definitions 2.1-2.4 apply to any matrix. Analogous results for stochastic matrices can be found in Wright (1989). Our definition of a dominant ϵ-eigenvector is a root error definition as opposed to a residual error definition:

Definition. *A vector v is a dominant ϵ-eigenvector of a matrix $\mathbf{M}$ if there exists a dominant eigenvector w of $\mathbf{M}$ with the property that $||v - w||_2 \leq \epsilon ||v||_2$. Whenever dominant ϵ-eigenvectors are discussed, it is assumed that $0 < \epsilon \leq 1$.*

This definition is better adapted to the exact arithmetic model than to models that include error. One of the referees described the limitations of the real number model and the chosen definition of the ϵ-eigenvector as follows.

"[The author] assumes exact arithmetic, which makes his model easier to compute in than the usual floating point one. But his stopping criterion is quite strong, that the dominant eigenvector be computed with accuracy ϵ (in the norm sense) no matter how close λ_1 and λ_2 are. This is more than is asked in the usual situation, since if we compute with arithmetic with relative accuracy ϵ_a, then perturbation theory implies that we expect uncertainty in the dominant eigenvector of about $\epsilon_a/(1 - |\lambda_2/\lambda_1|)$. To meet the author's stopping criterion, we would need at least that $\epsilon_a < \epsilon_a/(1 - |\lambda_2/\lambda_1|)$. Thus the work would grow at least like $\log \epsilon_a \log \log \epsilon_a \log \log \log \epsilon_a$, (via Schönhage-Strassen multiplication) where $\log \epsilon_a$ grows like $|\log \epsilon| + \log(\lambda_2/\lambda_1 - 1)$. This is much more pessimistic than the author's bound, which grows like $\log |\log \epsilon|$, although the λ_2/λ_1 dependence is similar.

"This may be a situation where the real number model studied by Smale and others is definitely 'too strong' to accurately model actual approximate computation, and this may be the most interesting conclusion."

The author will only add that the propagation of error inherent to the algorithm can be controlled by occasionally multiplying the iterates by the original matrix and then symmetrizing. A detailed error analysis is not include in this paper.

We assume that the density chosen for the space of matrices is invariant under linear transformations that preserver the Frobenius norm. Examples of such densities are given in Section 1. However, as is implicit in Section 4 of Kostlan (1985), Wishart matrices would give similar results.

*Department of Mathematics, University of Hawaii, Honolulu, Hawaii 97822

A well-known algorithm for finding dominant eigenvectors of matrices consists of repeatedly squaring a matrix, renormalizing the matrix after each squaring. See Wilkinson (1965). In order to find an upper bound on the complexity of dominant eigenvector calculation, we need a criterion for when to terminate the algorithm. This is discussed in Section 2, and is based upon a Gerschgorin-like estimate developed in Section 4 (in particular, Corollary 4.2). The available statistical information about the eigenvalues of real symmetric and complex Hermitian $n \times n$ matrices allows us to show that the average number of iterations required by the algorithm is $O(\ln\ n\ +\ \ln|\ln\ \epsilon|)$. A specific upper bound is given in Theorem 3.1.

The average complexity of an algorithm can be misleading. For example, Theorem 9.2 of Kostlan (1985) shows that the average complexity of the power method is infinite. We thus include an estimate of the *cost distribution function* of the algorithm; this is, we give an upper bound on the probability that the algorithm will take more than any prescribed number of iterations to produce a dominant ϵ-eigenvector (Theorem 3.2).

Throughout the paper, the cumulative distribution function of a random variable X, that is, $\text{Prob}[X \leq z]$, is denoted by $F_X(z)$. The eigenvalues of a matrix are denoted by λ_i, in decreasing order of magnitude. The i-th element of the standard basis of $\mathbf{R}^n$ or $\mathbf{C}^n$ is denoted by e_i. The Frobenius norm of a matrix $\mathbf{M}$ is denoted by $||\mathbf{M}||_F$.

1 Random Matrices

We consider densities invariant under linear transformations that preserve the Frobenius norm. For simplicity we assume that $\text{Prob}[\mathbf{M} = 0] = 0$. Since we are only concerned with the ratio of the eigenvalues, all such densities give the same results. Thus, for example, our results apply to matrices uniformly distributed over the set $\{||\mathbf{M}||_F \leq c\}$, for any constant $c > 0$. Alternately, we would consider the $\mu = 0$ case of the two-parameter family of densities

$$C \exp\left[-\frac{1}{2\sigma^2}||\mathbf{M} - \mu\mathbf{I}||_F^2\right] , \tag{1.1}$$

where $C = (\sqrt{\pi}\sigma)^{-n(n+1)/2}2^{-n/2}$ [resp. $(\sqrt{\pi}\sigma)^{-n^2}2^{-n/2}$] for real symmetric [resp. complex Hermitian] $n \times n$ matrices. These densities have the property of being invariant under conjugation by orthogonal [resp. unitary] matrices. For a detailed study of the properties of these matrices and their eigenvalues, see Mehta (1967). It is easily checked that these densities are invariant under linear transformations that preserve the Frobenius norm if and only if $\mu = 0$.

Theorem 1.1 *For matrices distributed as described above the cumulative distribution function of the random variable $|\lambda_1/\lambda_2|$ satisfies*

$$F_{|\lambda_1/\lambda_2|}(z) \ = \ \text{Prob}[|\lambda_1/\lambda_2| \leq z] \ \leq \ \left\{ \begin{array}{cl} 0 & \text{if} \ \ z \leq 1 \\ \frac{z_0(z-1)}{(z_0-1)z} & \text{if} \ \ 1 < z \leq z_0 \\ 1 & \text{if} \ \ z > z_0 \end{array} \right\} ,$$

where $z_0 = 3\sqrt{2}(\beta n(n-1)/2 + n)[3\sqrt{2}(\beta n(n-1)/2 + n) - 1]$, and where $\beta = 1$ for the real symmetric case and $\beta = 2$ for the complex Hermitian case.

Note that $z_0/(z_0 - 1)$ is $O(n^2)$.

Proof. This is Theorem 4.3 of Kostlan (1985). This is also equivalent to Theorem 4.4 of Kostlan (1988), which is stated somewhat differently, and which contains a minor error – there is a division by $\sqrt{2}$ that should be a multiplication by $\sqrt{2}$. □

Note that similar bounds may be obtained by applying the theory of volumes of tubes of real algebraic varieties.

2 The Algorithm

Definition 2.1 *For any matrix $\mathbf{M}$, define I and J by $(\forall i, j)|\mathbf{M}_{IJ}| \geq |\mathbf{M}_{ij}|$. I and J are uniquely defined with probability one, and if they are not unique, break the tie by any method.*

Definition 2.2 *For any matrix* $\mathbf{M}$*, define* K *by* $(\forall i)||\mathbf{M}e_K||_2 \geq ||\mathbf{M}e_i||_2$*.* K *is uniquely defined with probability one, and if it is not unique, break the tie by any method.*

Definition 2.3 *For any matrix* $\mathbf{M}$*, define* $\mathbf{M}^{\#}$ *to be* $\mathbf{M}e_J e_I^\dagger \mathbf{M}/\mathbf{M}_{IJ}$*.*

Definition 2.4 *For any matrix* $\mathbf{M}$*, define* $\mathbf{M}^*$ *to be the rank one approximation of* $\mathbf{M}$*, that is, the closest rank one matrix to* $\mathbf{M}$ *in the Frobenius norm.* $\mathbf{M}^*$ *is uniquely defined with probability one, and if it is not unique, break the tie by any method.*

Note that if $\mathbf{M}$ is a rank one matrix $\mathbf{M} = \mathbf{M}^{\#} = \mathbf{M}^*$.

For the remainder of this section, assume $\mathbf{M}$ is a complex Hermitian $n \times n$ matrix. We consider the following algorithm:
1. $\mathbf{M} := \mathbf{M}^2$
2. $\mathbf{M} := \mathbf{M} \ / \ \mathbf{M}_{IJ}$
3. If $n||\mathbf{M} - \mathbf{M}^{\#}||_F^2 \leq \epsilon^2 ||\mathbf{M}||_F^2$, stop.
4. goto 1

The algorithm will stop if $|\lambda_1| > |\lambda_2|$. Thus the algorithm will stop with probability one. If the algorithm has stopped, $\mathbf{M}e_K$ is a dominant ϵ-eignevector – see Lemmas 2.1 and 2.2. In practice, norms other than the Frobenius norm are easier to calculate, but for the Frobenius norm the geometric picture is clearer. Step 2 could be omitted for our purposes, but in practice it must be included to prevent under/overflow. We place the stopping criterion at the end of the iteration to simplify the iteration count.

Lemma 2.1 *If* $n||\mathbf{M} - \mathbf{M}^*||_F^2 \leq \epsilon^2 ||\mathbf{M}||_F^2$*, then* $\mathbf{M}e_K$ *is a dominant* ϵ*-eigenvector.*

Proof. If $n||\mathbf{M}-\mathbf{M}^*||_F^2 \leq \epsilon^2||\mathbf{M}||_F^2$, then $n||\mathbf{M}e_K - \mathbf{M}^* e_K||_2^2 \leq \epsilon^2 ||\mathbf{M}||_F^2$. Thus $\mathbf{M}e_K$ is within $||\mathbf{M}||_F \epsilon/\sqrt{n}$ of a dominant eigenvector. But $||\mathbf{M}e_K||_2 \geq ||\mathbf{M}||_F/\sqrt{n}$, and therefore, $\mathbf{M}e_K$ is a dominant ϵ-eigenvector. □

Lemma 2.2 *If the algorithm has stopped,* $n||\mathbf{M} - \mathbf{M}^*||_F^2 \leq \epsilon^2 ||\mathbf{M}||_F^2$*, and if* $n^3||\mathbf{M} - \mathbf{M}^*||_F^2 \leq \epsilon^2||\mathbf{M}||_F^2$*, the algorithm will stop on the current iteration.*

Proof. The result follows immediately from $||\mathbf{M} - \mathbf{M}^*||_F \leq ||\mathbf{M} - \mathbf{M}^{\#}||_F \leq n||\mathbf{M} - \mathbf{M}^*||_F$, which is a special case of Corollary 4.2. □

Lemma 2.3 *For* $\delta > 0$*, if* $|\lambda_1/\lambda_2|^2 \geq (n-1)(\delta^{-2} - 1)$*, then* $||\mathbf{M} - \mathbf{M}^*||_F \leq \delta ||\mathbf{M}||_F$*.*

Proof. For $\delta \geq 1$ the result is trivial. For $0 < \delta < 1$,

$$|\lambda_1/\lambda_2|^2 \geq (n-1)(\delta^{-2} - 1) \quad \leftrightarrow \quad (n-1)|\lambda_2|^2 \leq \delta^2 [|\lambda_1|^2 + (n-1)|\lambda_2|^2] \ .$$

Since $\delta < 1$ this implies that

$$\sum_{i \neq 1} |\lambda_i|^2 \leq \delta^2 \sum_{i=1,n} |\lambda_i|^2 \quad \leftrightarrow \quad ||\mathbf{M} - \mathbf{M}^*||_F \leq \delta ||\mathbf{M}||_F \ . \ \square$$

Theorem 2.1 *For any complex Hermitian* $n \times n$ *matrix the number of iterations the algorithm requires is less than or equal to*

$$\text{Max} \left\{ \frac{1}{\ln 2} [\ln(2 \ln n + |\ln \epsilon|) - \ln \ln |\lambda_1/\lambda_2|] \ , \ 0 \right\} \ + \ 1 \ .$$

In particular, the algorithm exhibits quadratic convergence.

Proof. For $\delta = \epsilon n^{-3/2}$, Lemma 2.3 implies that if $|\lambda_1/\lambda_2|^2 > n^4/\epsilon^2 > (n-1)(n^3/\epsilon^2 - 1)$, then $n^3||\mathbf{M} - \mathbf{M}^*||_F^2 \leq \epsilon^2 ||\mathbf{M}||_F^2$, and therefore, by Lemman 2.2, the algorithm will stop on the current interation. So for any matrix $\mathbf{M}$, and for any $s \geq 0$, $|\lambda_1/\lambda_2|^{2^s} > n^2/\epsilon$ implies that the algorithm will stop in at most $s+1$ iterations – we must add one to s because s may not be an integer. Solve for s. □

3 The Statistical Complexity of the Algorithm

Theorem 3.1 *The average number of iterations required by the algorithm to find a dominant ϵ-eigenvector is less than or equal to*

$$\frac{1}{\ln 2}\{\ln[3\sqrt{2}(\beta n(n-1)/2+n)] \ + \ 1 \ + \ \ln[2\ln n + |\ln \epsilon|]\} \ + \ 1 \ ,$$

where $\beta = 1$ for the real symmetric case and $\beta = 2$ for the complex Hermitian case.

Proof. We need to integrate the expression in Theorem 2.1 against the distribution function of $|\lambda_1/\lambda_2|$. Therefore, it suffices to bound

$$\int_1^e -\ln\ln(z)dF_{|\lambda_1/\lambda_2|}(z) \tag{3.1}$$

from above. By Theorem 1.1, (3.1) is bounded above by

$$\frac{z_0}{z_0-1}\int_1^{z_0} -\ln\ln(x)\frac{dx}{x^2} \ < \ \frac{z_0}{z_0-1}\int_1^{z_0} -\ln\left(\frac{x-1}{x}\right)\frac{dx}{x^2} \ = \ \ln\left(\frac{z_0}{z_0-1}\right) \ + \ 1 \ ,$$

where $z_0/(z_0-1) \ = \ 3\sqrt{2}(\beta n(n-1)/2+n)$. □

Corollary. *For real symmetric and complex Hermitian matrices, the dominant eigenvector can be estimated to any accuracy in a number of arithmetic operations that, on the average, grows slower than some polynomial of the size of the matrix.*

Proof. Each iteration of the algorithm can be performed in a number of operations that grows as a polynomial in the size of the matrix. □
For a discussion of the complexity of matrix squaring, see Pan (1984).

Theorem 3.2 *Let N be the number of iterations required by the algorithm to find a dominant ϵ-eigenvector. Then for $z \geq 1$,*

$$1-F_N(z) \ = \ \text{Prob}[N > z] \ \leq \ 3\sqrt{2}(\beta n(n-1)/2+n)[1-(\epsilon/n^2)^{2^{1-z}}],$$

where $\beta = 1$ for the real symmetric case and $\beta = 2$ for the complex Hermitian case.

Proof. By Theorem 2.1,

$$\text{Prob}[N > z] \ \leq \ \text{Prob}\left[\text{Max}\left\{\frac{1}{\ln 2}[\ln(2\ln n + |\ln\epsilon|) - \ln\ln|\lambda_1/\lambda_2|] \ , \ 0\right\} + 1 \ > \ z\right] \ . \tag{3.2}$$

Solving for $|\lambda_1/\lambda_2|$, the right-hand side of (3.2) becomes $\text{Prob}[|\lambda_1/\lambda_2| < (n^2/\epsilon)^{2^{1-z}}]$. But by Theorem 1.1, this is less than or equal to $3\sqrt{2}(\beta n(n-1)/2+n)[1-(\epsilon/n^2)^{2^{1-z}}]$. □

Note that we could have derived Theorem 3.1 by estimating the integral $\int_{x=0}^{\infty} x dF_N(x)$, using Theorem 3.2.

4 An Inequality for Complex $n \times m$ Matrices

In this section, we prove an inequality that allows us to estimate the distance between a matrix and its rank one approximation. This result relates information about the singular values of a matrix to information about its entries, and is therefore in the spirit of the Gerschgorin Circle Theorem for eigenvalues.

Theorem 4.1 *For any $n \times m$ matrix $\mathbf{M}$, and for any $v \in \mathbf{C}^m$ and any $w \in \mathbf{C}^n$,*

$$||(w^\dagger \mathbf{M} v)\mathbf{M} - \mathbf{M} v w^\dagger \mathbf{M}||_F \ \leq \ ||\mathbf{M}||_2 ||\mathbf{M} - \mathbf{M}^*||_F ||v||_2 ||w||_2 \ .$$

Note that the left-hand side of this inequality depends only on the entries of the matrix $\mathbf{M}$, while the right-hand side depends only on the singular values of $\mathbf{M}$.

Proof. All sums appearing in this proof are *single sums*; the index of summation is always the first variable appearing under the summation sign. Both sides of the inequality are invariant under two-sided unitary coordinate transformations, so without loss of generality assume that $\mathbf{M}$ is diagonal, positive semidefinite, with singular values $\mu_1 \geq \cdots \geq \mu_k > 0$, where k is the rank of $\mathbf{M}$. The desired inequality reduces to

$$\sum_{i=1,k} \left| \mu_i \sum_{j=1,k} v_j \overline{w}_j \mu_j - v_i \overline{w}_i \mu_i^2 \right|^2 + \sum_{i=1,k} \sum_{j \neq i} |v_i \overline{w}_j \mu_i \mu_j|^2 \leq \mu_1^2 \sum_{i \neq 1} \mu_i^2 ||v||_2^2 ||w||_2^2 . \tag{4.1}$$

The left-hand side of (4.1) can be rewritten as

$$\sum_{i=1,k} \left| \mu_i \sum_{j \neq i} v_j \overline{w}_j \mu_j \right|^2 + \sum_{i=1,k} \sum_{j \neq i} |v_i \overline{w}_j \mu_i \mu_j|^2 ,$$

or, equivalently,

$$\sum_{i=1,k} \mu_i^2 \left[\left| \sum_{j \neq i} v_j \overline{w}_j \mu_j \right|^2 + |v_i|^2 \sum_{j \neq i} |\overline{w}_j \mu_j|^2 \right] . \tag{4.2}$$

But by the Cauchy-Schwartz inequality,

$$\left| \sum_{j \neq i} v_j \overline{w}_j \mu_j \right|^2 \leq \sum_{j \neq i} |v_j|^2 \sum_{j \neq i} |\overline{w}_j \mu_j|^2 ,$$

so we see that (4.2) is less than or equal to

$$||v||_2^2 \sum_{i=1,k} \mu_i^2 \left[\sum_{j \neq i} |\overline{w}_j \mu_j|^2 \right] = ||v||_2^2 \sum_{i=1,k} \sum_{j \neq i} |\overline{w}_j \mu_i \mu_j|^2 = ||v||_2^2 \sum_{j=1,k} |\overline{w}_j|^2 \left[\mu_j^2 \sum_{i \neq j} \mu_i^2 \right] . \tag{4.3}$$

But for all j,

$$\mu_j^2 \sum_{i \neq j} \mu_i^2 \leq \mu_i^2 \sum_{i \neq 1} \mu_i^2 .$$

Therefore the right-hand side of (4.3) is less than or equal to

$$||v||_2^2 ||w||_2^2 \mu_1^2 \sum_{i \neq 1} \mu_i^2 ,$$

and we have established the desired inequality. □

Corollary 4.1 *For any complex $n \times m$ matrix $\mathbf{M}$,*

$$|\mathbf{M}_{IJ}| ||\mathbf{M} - \mathbf{M}^{\#}||_F \leq ||\mathbf{M}||_2 ||\mathbf{M} - \mathbf{M}^*||_F .$$

Proof. This is just Theorem 4.1, with $v = e_J$ and $w = e_I$. □

Corollary 4.2 *For any complex $n \times m$ matrix $\mathbf{M}$ of rank at least two,*

$$1 \leq \frac{||\mathbf{M} - \mathbf{M}^{\#}||_F}{||\mathbf{M} - \mathbf{M}^*||_F} < \sqrt{nm} .$$

These inequalities are sharp, even when restricted to real (and when $n = m$, symmetric) matrices.

Proof. The left-hand inequality holds because $\mathbf{M}^{\#}$ is of rank one and $\mathbf{M}^*$ is the closest rank one matrix to $\mathbf{M}$, and is sharp because equality holds for diagonal matrices. The right-hand inequality follows immediately from Corollary 4.1, and from the following trivial inequality for matrices of rank at least two:

$$||\mathbf{M}||_2 < ||\mathbf{M}||_F \leq |\mathbf{M}_{IJ}|\sqrt{nm} \ .$$

Sharpness of the inequality can be shown as follows. For $t > 0$ define matrices $\mathbf{M}^t$ by

$$\mathbf{M}^t_{11} = 1 + t, \quad \mathbf{M}^t_{ij} = 1 \ \text{ for all other } ij.$$

Then a straightforward calculation shows that

$$\lim_{t \to 0^+} \frac{||\mathbf{M}^t - \mathbf{M}^{t\#}||_F}{||\mathbf{M}^t - \mathbf{M}^{t*}||_F} \ = \ \sqrt{nm} \ . \ \square$$

Acknowledgments

The study of the statistical complexity of dominant eigenvector calculation was suggested by Steve Smale. Mike Shub suggested studying the squaring algorithm. I appreciate the hospitality of IMPA, Rio de Janeiro, where the initial phase of this investigation was conducted.

References

[1] L. Blum, M. Shub, and S. Smale, On the theory of computation and complexity over the real numbers, NP-competeness, recursive functions and universal machines, *Bull. Amer. Math. Soc.* **21** (1989), 1-46.

[2] E. Kostlan, *Statistical Complexity of Numerical Linear Algebra*, Ph.D. thesis, Department of Mathematics, University of California, Berkeley, 1985.

[3] E. Kostlan, Complexity theory of numerical linear algebra, *J. Comput. Appl. Math.* **22** (1988), 219-230.

[4] M.L. Mehta, *Random Matrices and the Statistical Theory of Energy Levels*, Academic Press, New York, 1967.

[5] V. Pan, *How to Multiply Matrices Faster*, Lecture Notes in Computer Science, Vol. 179, Springer-Verlag, Berlin, 1984.

[6] S. Smale, On the efficiency of algorithms of analysis, *Bull. Amer. Math. Soc.* **13** (1985), 87-121.

[7] J.F. Traub, G.W. Wasilkowski, and H. Woźniakowski, *Information-Based Complexity*, Academic Press, San Diego, 1988.

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

[9] P. Wright, Statistical complexity of the power method for Markov chains, *J. Complexity* **5** (1989), 119-143.