On "an apparent truth about matrices"

Bryan Gin–ge Chen
University of Pennsylvania
Department of Physics
chb@sas.upenn.edu

In the paper "Sloppy-Model Universality Class and the Vandermonde Matrix", Waterfall et al[1] conjecture the following:

"Let $S \in \mathbb{R}^{n\times n}$ be symmetric and positive definite. Let $E \in \mathbb{R}^{n\times n}$ be diagonal with $E_{ii} = \epsilon^{i-1}$ and $0 < \epsilon \ll 1$. Then the mth largest eigenvalue of ESE is $\mathcal{O}(\epsilon^{2(m-1)})$."

In this note, we characterize the asymptotics of eigenvalues (i.e. we can find formulae for the eigenvalues up to leading order in ϵ) of EME for arbitrary complex $n \times n$ matrices M and obtain the above as a corollary.

Conventions: Our indices for the rows and columns run from 0 to $n-1$, rather than from 1 to n. $[M]_{i_1,i_2,\dots,i_m}$ denotes the minor given by excluding the indices $\{i_l\}$. In other words, $[M]_{i_1,i_2,\dots,i_m}$ is the determinant of the matrix we get by excluding the rows and columns indexed by $i_1, i_2, \dots, i_m$. We have of course the convention that $[M] = \det M$, and we will also take $[M]_{0,\dots,n-1} = 1$.

The principal result of this note is:

Let $M \in \mathbb{C}^{n\times n}$, and let E be as above. The asymptotic behavior of the eigenvalues of EME for $\epsilon \ll 1$ and $\epsilon \gg 1$ may be determined by balancing terms of the characteristic polynomial against each other in the limit of large or small ϵ, respectively. The characteristic polynomial may be written as follows (this holds for any ϵ in E):

$$P_{EME}(\lambda) = \sum_{m=0}^{n} (-\lambda)^m \sum_{0\le j_1<j_2<\cdots<j_m\le n-1} [M]_{j_1,j_2,\dots,j_m} \epsilon^{2\left(\sum_{l=0}^{n-1} l - \sum_{k=1}^{m} j_k\right)}$$

So for large ϵ the asymptotic eigenvalues are given by $\frac{\det M}{[M]_0}$, $\frac{[M]_0}{[M]_{0,1}}\epsilon^2$, $\dots$, $\frac{[M]_{0,\dots,n-2}}{[M]_{0,\dots,n-1}}\epsilon^{2(n-1)}$. For small ϵ, the asymptotic eigenvalues are $\frac{\det M}{[M]_{n-1}}\epsilon^{2(n-1)}$, $\frac{[M]_{n-1}}{[M]_{n-2,n-1}}\epsilon^{2(n-2)}$, $\dots$, $\frac{[M]_{1,\dots,n-1}}{[M]_{0,\dots,n-1}} = [M]_{1,\dots,n-1} = a_{00}$, which are ratios of leading principal minors. These are valid as long as none of the determinants involved vanish.

3×3 matrix case

We work out the example of a 3×3 matrix to illustrate the ideas behind the proof. The characteristic polynomial of a matrix EME with $(M)_{ij} = a_{ij}$ is:

$$\begin{aligned} P_{EME}(\lambda) = {} & -\lambda^3 + (a_{00} + a_{11}\epsilon^2 + a_{22}\epsilon^4)\lambda^2 \\ & - \left((a_{00}a_{11} - a_{01}a_{10})\epsilon^2 + (a_{00}a_{22} - a_{02}a_{20})\epsilon^4 + (a_{11}a_{22} - a_{12}a_{21})\epsilon^6\right)\lambda \\ & + \left(a_{00}a_{11}a_{22} + a_{01}a_{12}a_{20} + a_{02}a_{10}a_{21} - a_{00}a_{12}a_{21} - a_{11}a_{02}a_{20} - a_{22}a_{01}a_{10}\right)\epsilon^6 \end{aligned}$$

The constant term is just $(\det M)\epsilon^6$, the coefficient of the λ^1 term is $\sum_{j=0}^{2}[M]_j\lambda^{2(0+1+2-j)}$ and the coefficient of the λ^2 term is $\sum_{0\le j<k\le 2}[M]_{j,k}\epsilon^{2(0+1+2-j-k)}$. Assume for the moment

that none of the principal minors vanish, which is certainly the case if M is positive definite by the Sylvester criterion[2].

How might we approximate the roots of this polynomial and thus the eigenvalues of EME? Let's divide the characteristic polynomial by ϵ^6. We need to solve:

$$0 \;=\; -\epsilon^{-6}\lambda^3 + (a_{00}\epsilon^{-6} + a_{11}\epsilon^{-4} + a_{22}\epsilon^{-2})\lambda^2 - \left([M]_2\epsilon^{-4} + [M]_1\epsilon^{-2} + [M]_0)\right)\lambda + \det M$$

For large ϵ, we can neglect all the terms except for two:

$$0 = -[M]_0\lambda + \det M$$

Now approximate λ by $\frac{\det M}{[M]_0}$. We can check that this solution will be self-consistent by plugging this into the characteristic polynomial:

$$\frac{P_{EME}}{\epsilon^6} = -\epsilon^{-6}\left(\frac{\det M}{[M]_0}\right)^3 + (a_{00}\epsilon^{-6} + a_{11}\epsilon^{-4} + a_{22}\epsilon^{-2})\left(\frac{\det M}{[M]_0}\right)^2 - \left([M]_2\epsilon^{-4} + [M]_1\epsilon^{-2} + [M]_0)\right)\frac{\det M}{[M]_0} + \det M$$

This expression is zero up to $\mathcal{O}(\epsilon^{-2})$ - which supports $\frac{\det M}{[M]_0}$ as a self-consistent solution of the characteristic polynomial.

By dividing by ϵ^4 we get the self-consistent asymptotic approximation $\lambda \sim \frac{[M]_0}{a_{22}}\epsilon^2$ and by dividing by ϵ^2 we get $\lambda \sim a_{22}\epsilon^4$ up to $\mathcal{O}(\epsilon^{-2})$ terms.

A brief mathematical aside: These approximations correspond to the finest triangulation of the configuration of 4 points 0, 1, 2, 3 on the affine line. By a certain deep connection in the theory of series solutions to polynomials, we get the above approximations from the zeroth order terms in said series[3]. The asymptotics of polynomial roots and this connection is the subject of some ongoing work by Michael Brenner and myself [4].

It's now not too hard to see that if any of the dominant principal minors did vanish, the asymptotic behavior of the eigenvalues would be modified and follow then a new leading order balance (in ϵ) in the principal minors.

Eigenvalue asymptotics

We will make a dominant balance argument on the terms of the characteristic polynomial of EME written in the following form:

$$P_{EME}(\lambda) = \sum_{m=0}^{n}(-\lambda)^m \sum_{0\le j_1<j_2<\cdots<j_m\le n-1} [M]_{j_1,j_2,\ldots,j_m}\epsilon^{2\left(\sum_{l=0}^{n-1} l - \sum_{k=1}^{m} j_k\right)}$$

We first discuss the case when M is positive definite. In this case, the Sylvester criterion ensures that all of the principal minors are positive[2].

For the large ϵ case, the dominant contribution to the coefficient of the degree m term is $[M]_{0,\ldots,m-1}\epsilon^{2\sum_{l=m}^{n-1} l}$ (Choose $j_1, j_2, \ldots, j_m = 0, 1, 2, \ldots, m-1$, to maximize the sum of ϵ). We

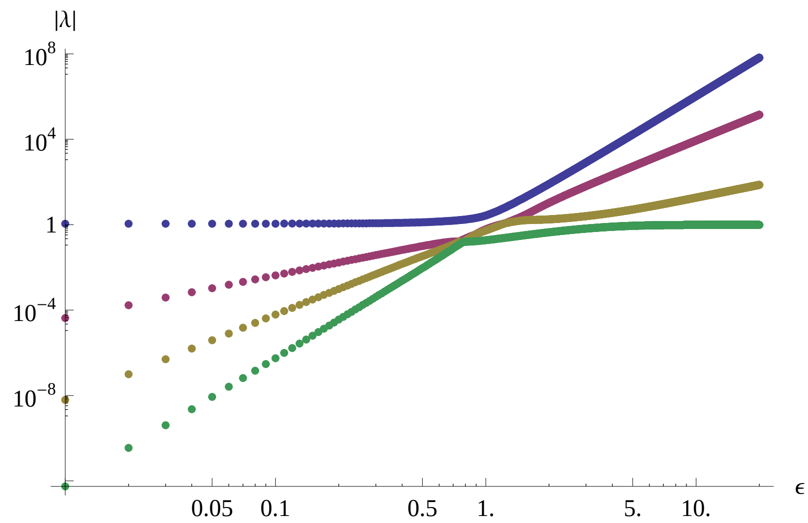

Figure 1: The absolute values of eigenvalues of some complex 4×4 matrix M sandwiched between $E = \mathrm{diag}(1, \epsilon, \epsilon^2, \epsilon^3)$ vs ϵ. Note the power law behavior in $|\lambda|$ for large and small ϵ.

now argue that the asymptotic eigenvalues are given by $\frac{\det M}{[M]_0}$, $\frac{[M]_0}{[M]_{0,1}}\epsilon^2$, $\ldots$, $\frac{[M]_{0,\ldots,n-2}}{[M]_{0,\ldots,n-1}}\epsilon^{2(n-1)}$. We obtained these by solving two term equations like:

$$[M]_{0,\ldots,m-1}\epsilon^{2\sum_{l=m}^{n-1} l}\,(-\lambda)^m + [M]_{0,\ldots,m}\epsilon^{2\sum_{l=m+1}^{n-1} l}\,(-\lambda)^{m+1} = 0$$

Consider the rth such asymptotic approximation $\frac{[M]_{0,\ldots,r-2}}{[M]_{0,\ldots,r-1}}\epsilon^{2(r-1)}$. If we put this into our expression for the characteristic polynomial, we have:

$$P_{EME}(\lambda) = \sum_{m=0}^{n}\left(-\frac{[M]_{0,\ldots,r-2}}{[M]_{0,\ldots,r-1}}\epsilon^{2(r-1)}\right)^m \sum_{0\le j_1<j_2<\cdots<j_m\le n-1}[M]_{j_1,j_2,\ldots,j_m}\epsilon^{2\left(\sum_{l=0}^{n-1} l-\sum_{k=1}^{m} j_k\right)}$$

The power of ϵ in the dominant contribution to the term of degree p (with coefficient $[M]_{0,\ldots,p-1}\epsilon^{2\sum_{l=p}^{n-1} l}$) is:

$$2\left(p(r-1) + \sum_{l=p}^{n-1} l\right) = 2p(r-1) + n(n-1) - p(p-1) = p(2r-p-1) + n(n-1)$$

This is concave down and so maximized when $\frac{d}{dp}(p(2r-p-1)+n(n-1) = 0$, or $p = r - \frac{1}{2}$. Thus the powers of ϵ in the dominant contribution to the degree r and $r-1$ are greater than the powers of ϵ in any of the other terms. If we divide the characteristic polynomial by this maximum power of ϵ, at large ϵ, all terms other than the degree r and $r-1$ will drop out. Now we recover (up to terms of $\mathcal{O}(\epsilon^{-2})$) the equation we used to derive this expression for the eigenvalue:

$$[M]_{0,\ldots,r-2}\epsilon^{2\sum_{l=r-1}^{n-1} l}(-\lambda)^{r-1} + [M]_{0,\ldots,r-1}\epsilon^{2\sum_{l=r}^{n-1} l}(-\lambda)^r = 0$$

So $\frac{[M]_{0,\ldots,r-2}}{[M]_{0,\ldots,r-1}}\epsilon^{2(r-1)}$ captures the asymptotic behavior of the rth eigenvalue at large ϵ. This argument holds for any of the expressions we wrote, so we have fully characterized the large ϵ asymptotics of the eigenvalues of EME.

We may repeat these steps for the case of small ϵ, but now the dominant contribution to the term of degree m is instead $[M]_{0,\ldots,m-1}\epsilon^{2\sum_{l=0}^{n-m-1} l}$ (choose $j_1, \ldots, j_m = n-m, \ldots, n$), and our asymptotic expressions for the eigenvalues are $\frac{\det M}{[M]_{n-1}}\epsilon^{2(n-1)}$, $\frac{[M]_{n-1}}{[M]_{n-2,n-1}}\epsilon^{2(n-2)}$, $\ldots$, $\frac{[M]_{1,\ldots,n-1}}{[M]_{0,\ldots,n-1}} = [M]_{1,\ldots,n-1} = a_{00}$.

If we sort these asymptotic expressions for the eigenvalues, we see that they agree with the conjectured bounds by Waterfall et al.

For arbitrary $n \times n$ matrices M, we may perform this procedure, but now we aren't guaranteed that any given $[M]_{j_1,\ldots,j_m}$ will be nonzero, so we must instead use whatever the dominant contribution in powers of ϵ or ϵ^{-1} happen to be.

Derivation of the formula for the characteristic polynomial of EME

The mth coefficient of the characteristic polynomial defined to be $\det(\tilde{M} - \lambda I)$ is:

$$\frac{1}{m!}\frac{d^m}{d\lambda^m}\det(\tilde{M} - \lambda I)\Big|_{\lambda=0}$$

Now we use the identity $\frac{d}{dt}\det A = \sum_j \det(A)_j$ where $(A)_j$ is the matrix formed by replacing the jth row (or column) of A with the derivative of the jth row (or column) of A with respect to t[5]. Then let $A = \tilde{M} - \lambda I$, $t = \lambda$:

$$\begin{aligned}
\frac{1}{m!}\frac{d^m}{d\lambda^m}\det(\tilde{M} - \lambda I) &= \frac{(-1)^m}{m!}\sum_{j_1}\sum_{j_2}\cdots\sum_{j_m}\det(\tilde{M} - \lambda I)_{j_1,j_2,\ldots,j_m} \\
&= \frac{(-1)^m}{m!}\sum_{j_1\neq j_2\neq\cdots\neq j_m}\det(\tilde{M} - \lambda I)_{j_1,j_2,\ldots,j_m} \\
&= (-1)^m \sum_{0\leq j_1<j_2<\cdots<j_m\leq n-1}\det(\tilde{M} - \lambda I)_{j_1,j_2,\ldots,j_m} \\
\frac{1}{m!}\frac{d^m}{d\lambda^m}\det(\tilde{M} - \lambda I)\Big|_{\lambda=0} &= (-1)^m \sum_{0\leq j_1<j_2<\cdots<j_m\leq n-1}[\tilde{M}]_{j_1,j_2,\ldots,j_m}
\end{aligned}$$

Above we've used the fact that only the entries on the diagonal contain λ, and the fact that taking the derivative on any row yields a row of just 0's except for a -1 on the diagonal element, then canceled the $\frac{1}{m!}$ factor due to the different permutations of the $\{j_l\}$, and finally used the definition of the principal minor. Our conventions for $[M] = \det M$ (when $m = 0$) and $[M]_{0,\ldots,n-1} = 1$ have been chosen so that this expression remains true at those values of m.

Let $\tilde{M} = EME$. The structure of this matrix is as follows:

$$\begin{pmatrix}
a_{00} & a_{01}\epsilon & a_{02}\epsilon^2 & \ldots & a_{0,n-1}\epsilon^{n-1} \\
a_{10}\epsilon & a_{11}\epsilon^2 & a_{12}\epsilon^3 & \ldots & a_{1,n-1}\epsilon^{n} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
a_{n-1,0}\epsilon^{n-1} & a_{n-1,1}\epsilon^{n} & a_{n-1,2}\epsilon^{n+1} & \ldots & a_{n-1,n-1}\epsilon^{2(n-1)}
\end{pmatrix}$$

EME with $j_1, j_2, \ldots, j_m$ removed looks like:

$$\begin{pmatrix}
a_{00} & \cdots & a_{0,j_1-1}\epsilon^{j_1-1} & a_{0,j_1+1}\epsilon^{j_1+1} & \cdots & a_{0,j_m-1}\epsilon^{j_m-1} & a_{0,j_m+1}\epsilon^{j_m+1} & \cdots & a_{0,n-1}\epsilon^{n-1} \\
\vdots & \ddots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots \\
a_{j_1-1,0}\epsilon^{j_1-1} & \vdots & a_{j_1-1,j_1-1}\epsilon^{2(j_1-1)} & a_{j_1-1,j_1+1}\epsilon^{2j_1} & \vdots & a_{j_1-1,j_m-1}\epsilon^{j_1+j_m-2} & a_{j_1-1,j_m+1}\epsilon^{j_1+j_m} & \vdots & a_{j_1-1,n-1}\epsilon^{j_1+n-2} \\
a_{j_1+1,0}\epsilon^{j_1+1} & \vdots & a_{j_1+1,j_1-1}\epsilon^{2j_1} & a_{j_1+1,j_1+1}\epsilon^{2(j_1+1)} & \vdots & a_{j_1+1,j_m-1}\epsilon^{j_1+j_m} & a_{j_1+1,j_m+1}\epsilon^{j_1+j_m+2} & \vdots & a_{j_1+1,n-1}\epsilon^{j_1+n} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \cdots & \cdots & \cdots & \cdots \\
a_{j_m-1,0}\epsilon^{j_m-1} & \vdots & a_{j_m-1,j_1-1}\epsilon^{j_1+j_m-2} & a_{j_m-1,j_1+1}\epsilon^{j_1+j_m} & \vdots & a_{j_m-1,j_m-1}\epsilon^{2(j_m-1)} & a_{j_m-1,j_m+1}\epsilon^{2j_m} & \vdots & a_{j_m-1,n-1}\epsilon^{j_m+n-2} \\
a_{j_m+1,0}\epsilon^{j_m+1} & \vdots & a_{j_m+1,j_1-1}\epsilon^{j_1+j_m} & a_{j_m+1,j_1+1}\epsilon^{j_1+j_m+2} & \vdots & a_{j_m+1,j_m-1}\epsilon^{2j_m} & a_{j_m+1,j_m+1}\epsilon^{2(j_m+1)} & \vdots & a_{j_m+1,n-1}\epsilon^{j_m+n} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
a_{n-1,0}\epsilon^{n-1} & \cdots & a_{n-1,j_1-1}\epsilon^{j_1-1} & a_{0,j_1+1}\epsilon^{j_1+1} & \cdots & a_{n-1,j_m-1}\epsilon^{j_m+n-2} & a_{n-1,j_m+1}\epsilon^{j_m+n} & \cdots & a_{n-1,n-1}\epsilon^{2(n-1)}
\end{pmatrix}$$

When evaluating the determinant of this (this is the principal minor excluding $j_1, \ldots, j_m$), we can factor out ϵ^l from each of the lth columns, and we get:

$$\epsilon^{\sum_{l=0}^{n-1} l - \sum_{k=1}^{m} j_k} \det \begin{pmatrix} a_{00} & a_{01} & a_{02} & \cdots & a_{0,n-1} \\ a_{10}\epsilon & a_{11}\epsilon & a_{12}\epsilon & \cdots & a_{1,n-1}\epsilon \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n-1,0}\epsilon^{n-1} & a_{n-1,1}\epsilon^{n-1} & a_{n-1,2}\epsilon^{n-1} & \cdots & a_{n-1,n-1}\epsilon^{n-1} \end{pmatrix}$$

And we may do the same with the rows:

$$\epsilon^{2\left(\sum_{l=0}^{n-1} l - \sum_{k=1}^{m} j_k\right)} \det \begin{pmatrix} a_{00} & a_{01} & a_{02} & \cdots & a_{0,n-1} \\ a_{10} & a_{11} & a_{12} & \cdots & a_{1,n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n-1,0} & a_{n-1,1} & a_{n-1,2} & \cdots & a_{n-1,n-1} \end{pmatrix}$$

We've shown that

$$[\tilde{M}]_{j_1,\ldots,j_m} = [M]_{j_1,\ldots,j_m} \epsilon^{2\left(\sum_{l=0}^{n-1} l - \sum_{k=1}^{m} j_k\right)}$$

And so the characteristic polynomial takes the form:

$$P_{EME}(\lambda) = \sum_{m=0}^{n} (-\lambda)^m \sum_{0 \leq j_1 < j_2 < \cdots < j_m \leq n-1} [M]_{j_1,j_2,\ldots,j_m} \epsilon^{2\left(\sum_{l=0}^{n-1} l - \sum_{k=1}^{m} j_k\right)}$$

Acknowledgements: Thanks to Randy Kamien for critical comments.

References

[1] J. J. Waterfall, F. P. Casey, R. N. Gutenkunst, K. S. Brown, C. R. Myers, P. W. Brouwer, V. Elser and J. P. Sethna. PRL 97, 150601 (2006).

[2] S. Perlis. Theory of Matrices. Courier Dover. 1991.

[3] B. Sturmfels. Discrete Math 210, 171-181 (2000).

[4] B. G. Chen, HCMR 1, 50 (2007) (available at http://www.hcs.harvard.edu/hcmr/issue1/bryan.pdf).
B. G. Chen, unpublished.
B. G. Chen and M. P. Brenner, forthcoming.

[5] R. A. Horn and C. R. Johnson. Topics in Matrix Analysis. Cambridge UP. 1994.