MA 580; Numerical Analysis I

C. T. Kelley
NC State University
tim_kelley@ncsu.edu

Version of November 14, 2016

NCSU, Fall 2016
Part VIIIb: Eigenvalue Conditioning

References

This part of the notes comes from

- Applied Numerical Linear Algebra, Demmel, SIAM 1997
- Matrix Computations, Golub and Van Loan, Johns Hopkins, 2013
- Numerical Linear Algebra, Trefethen and Bau, SIAM 1997
- Introduction to Matrix Computations, Stewart, Academic Press, 1973

Eigenvalue Conditioning

Here's some bad news.

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & & & \\ 0 & 0 & 1 & 0 & & \\ & \ddots & \ddots & \ddots & \ddots & \\ & & 0 & 0 & 1 & 0 \\ & & & 0 & 0 & 1 \\ & & & & 0 & 0 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} 0 & 1 & 0 & & & \\ 0 & 0 & 1 & 0 & & \\ & \ddots & \ddots & \ddots & \ddots & \\ & & 0 & 0 & 1 & 0 \\ & & & 0 & 0 & 1 \\ \epsilon & & & & 0 & 0 \end{pmatrix}.$$

Eigenvalues

- $\mathbf{A}$ is a Jordan block. So
  - $\sigma(\mathbf{A}) = \{0\}$
  - Algebraic multiplicity N
  - Geometric multiplicity 1

As for $\mathbf{B}$ . . .

Spectrum of B

Suppose $\mathbf{B}\mathbf{x} = \lambda \mathbf{x}$ and $x_1 = 1$ then

$$\begin{pmatrix} 0 & 1 & 0 & & & & \\ 0 & 0 & 1 & 0 & & & \\ & \ddots & \ddots & \ddots & \ddots & & \\ & & 0 & 0 & 1 & 0 & \\ & & & 0 & 0 & 1 & \\ \epsilon & & & & 0 & 0 & \end{pmatrix} \begin{pmatrix} 1 \\ x_2 \\ \vdots \\ x_{N-2} \\ x_{N-1} \\ x_N \end{pmatrix} = \begin{pmatrix} x_2 \\ \vdots \\ x_{N-2} \\ x_{N-1} \\ x_N \\ \epsilon \end{pmatrix} = \lambda \begin{pmatrix} 1 \\ x_2 \\ \vdots \\ x_{N-2} \\ x_{N-1} \\ x_N \end{pmatrix}$$

and so . . .

$$\lambda = x_2, \quad x_3 = \lambda x_2 = \lambda^2, \ldots,$$

$$x_N = \lambda x_{N-1} = \lambda^{N-1}, \epsilon = \lambda x_{N-1} = \lambda^N.$$

So $\epsilon = \lambda^N$.
We have N solutions

$$\lambda_k = \epsilon^{1/N} e^{2\pi ik/N} 1 \leq k \leq N.$$

which are evenly spaced on the circle of radius $\epsilon^{1/N}$ on the complex plane.

Condition number

Any sensible definition of condition number is the ratio of

- the (relative) size of the change in the output, which is $\epsilon^{1/N}$
- to the size of the change in the input, which is $O(\epsilon)$

So

$$\kappa = \frac{\epsilon^{1/N}}{\epsilon} = \epsilon^{\frac{N-1}{N}} \to \infty$$

as $\epsilon \to 0$.

So, Jordan blocks are bad things. Any hope for nicer problems?

The characteristic polynomial

Recall that the characteristic polynomial of **A** is

$$p(z, \mathbf{A}) = \det(z\mathbf{I} - \mathbf{A})$$

and its roots are the eigenvalues of **A**.
The roots of a polynomial are continiuous functions of the coefficients, so then

$$\sigma(\mathbf{A} + \delta\mathbf{A}) \to \sigma(\mathbf{A}) \text{ as } \|\delta\mathbf{A}\| \to 0.$$

As the Jordan block example shows, eigenvalues need not be differentiable functions of the coefficients.

Simple eigenvlaues because I'm tired of Jordan blocks

- Suppose λ is a simple eigenvalue of $\mathbf{A}$.
- Is a nearby simple eigenvalue $\lambda + \delta\lambda$ of $\mathbf{A} + \delta\mathbf{A}$ out there?
- Is there a useful definition of condition number?

Left and Right eigenvectors

- $\mathbf{Ax} = \lambda \mathbf{x}$ says "$\mathbf{x}$ is a right eigenvector.
- $\sigma(\mathbf{A}) = \sigma(\mathbf{A}^T)$, so there's also a left eigenvector

$$\mathbf{A}^T \mathbf{y} = \lambda \mathbf{y} \text{ or } \mathbf{y}^T \mathbf{A} = \lambda \mathbf{y}^T$$

- From now on, $\mathbf{x}$ will be a right eigenvector, $\mathbf{y}$ a left eigenvector, and
- $\|\mathbf{x}\| = \|\mathbf{y}\| = 1$

Perturbation theory for simple eigenvalues

Theorem: Assume

- λ is a simple eigenvalue of $\mathbf{A}$,
- $\mathbf{x}$ ($\mathbf{y}$) are normalized right (left) eigenvectors,
- $\lambda + \delta\lambda$ is the eigenvalue of $\mathbf{A} + \delta\mathbf{A}$ nearest to λ.

Let $\theta(\mathbf{y}, \mathbf{x})$ be the acute angle between $\mathbf{y}$ and $\mathbf{x}$.
Note that $\sec(\theta(\mathbf{y}, \mathbf{x})) = 1/|\mathbf{y}^T\mathbf{x}|$.
Then ...

Perturbation estimates

$$\delta\lambda = \frac{\mathbf{y}^T \delta\mathbf{A}\mathbf{x}}{\mathbf{y}^T\mathbf{x}} + O\left(\|\delta\mathbf{A}\|^2\right) \text{ and}$$

$$|\delta\lambda| \le \frac{\|\delta\mathbf{A}\|\mathbf{x}}{\mathbf{y}^T\mathbf{x}} + \sec(\theta(\mathbf{y},\mathbf{x}))\|\delta\mathbf{A}\| + O\left(\|\delta\mathbf{A}\|^2\right)$$

So $\sec(\theta(\mathbf{y},\mathbf{x})) = 1/|\mathbf{y}^T\mathbf{x}|$ is the condition number of the simple eigenvalue λ.

Proof: I

We've done things like this for equations.

$$(\mathbf{A} + \delta\mathbf{A})(\mathbf{x} + \delta\mathbf{x}) - \mathbf{A}\mathbf{x} = (\lambda + \delta\lambda)(\mathbf{x} + \delta\mathbf{x}) - \lambda\mathbf{x}$$

so,

$$\mathbf{A}\delta\mathbf{x} + \delta\mathbf{A}\mathbf{x} + \delta\mathbf{A}\delta\mathbf{x} = \lambda\delta\mathbf{x} + \delta\lambda\mathbf{x} + \delta\lambda\delta\mathbf{x}$$

Ignore for now any term with two δs in it and multiply by $\mathbf{y}^T$

$$\mathbf{y}^T\mathbf{A}\delta\mathbf{x} + \mathbf{y}^T\delta\mathbf{A}\mathbf{x} \approx \lambda\mathbf{y}^T\delta\mathbf{x} + \delta\lambda\mathbf{y}^T\mathbf{x}$$

Note that $\mathbf{y}^T\mathbf{A}\delta\mathbf{x} = \lambda\mathbf{y}^T\delta\mathbf{x}$ because $\mathbf{y}^T\mathbf{A} = \lambda\mathbf{y}^T$. So ...

Proof: II

$$\mathbf{y}^T \delta \mathbf{A} \mathbf{x} \approx \delta\lambda \mathbf{y}^T \mathbf{x}$$

so

$$\delta\lambda \approx \frac{\mathbf{y}^T \delta \mathbf{A} \mathbf{x}}{\mathbf{y}^T \mathbf{x}}$$

The terms we ingored, after multiplying by $\mathbf{y}^T$ are

$$\mathbf{y}^T \delta \mathbf{A} \delta \mathbf{x} \text{ and } \delta\lambda \mathbf{y}^T \delta \mathbf{x}$$

If we now put them back we get ...

Proof: III

$$\delta\lambda = \frac{\mathbf{y}^T \delta\mathbf{A}\mathbf{x}}{\mathbf{y}^T\mathbf{x}} + \frac{\mathbf{y}^T(\delta\mathbf{A}\delta\mathbf{x} - \delta\lambda\delta\mathbf{x})}{\mathbf{y}^T\mathbf{x}}$$

The terms we neglected are smaller than the main term, if $\delta\mathbf{A}$ is sufficiently small, so

$$\|\delta\lambda\| = O\left(\frac{\|\delta\mathbf{A}\|}{\mathbf{y}^T\mathbf{x}}\right)$$

We now assume that $\delta\mathbf{A}$ is small enough that we can ignore factors of $1/\mathbf{y}^T\mathbf{x}$ in the high order terms.
This means that

$$\frac{\mathbf{y}^T(\delta\mathbf{A}\delta\mathbf{x} - \delta\lambda\delta\mathbf{x})}{\mathbf{y}^T\mathbf{x}} = O(\|\delta\mathbf{A}\|\|\delta\mathbf{x}\|).$$

Proof: IV

The power method says that

$$\|\delta \mathbf{x}\| = O(|\delta \lambda|)$$

if $\mathbf{A}$ is sufficiently small. That's it.

Observations

- If $\mathbf{A} = \mathbf{A}^T$, then $|\mathbf{y}^T \mathbf{x}| = 1$, and the conditioning is perfect.
- For the Jordan block example,

$$\mathbf{x} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} \text{ and } \mathbf{y} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

so the condition number is infinite.

Gershgorin Theorem

Let **B** be a square matrix. The eigenvalues of **B** lie in the union of the disks

$$G_i = \left\{ z \,\middle|\, |z - b_{ii}| \le \sum_{j \ne i} |b_{ij}| \right\} \text{ for } 1 \le i \le N$$

Proof of Gershgorin Theorem: I

Let $\lambda \in \sigma(\mathbf{A})$ and let $\mathbf{x}$ be a corresponding eigenvector. Let i be such that

$$|x_i| = \|\mathbf{x}\|_\infty.$$

Since $\mathbf{Ax} = \lambda \mathbf{x}$

$$(\lambda - a_{ii})x_i = \sum_{j \neq i} a_{ij} x_j.$$

Proof of Gershgorin Theorem: II

We picked i so that $|x_j|/|x_i| \le 1$ for $j \ne i$, so

$$|\lambda - a_{ii}| \le \sum_{j \ne i} |a_{ij}||x_j|/|x_i|. \le \sum_{j \ne i} |a_{ij}|,$$

as asserted.

Diagonalizable Matrices and the Bauer-Fike Theorem

Theorem: Suppose

- $\mathbf{A}$ is diagonalizable with only simple eigenvalues $\{\lambda_i\}$
- $\mathbf{x}_i$ ($\mathbf{y}_i$) are the normalized left (right) eigenvectors corresponding to λ_i.

Then the eigenvalues of $\mathbf{A} + \delta\mathbf{A}$ lie in disks B_i where

$$B_i = \left\{ z \mid |z - \lambda_i| \leq \frac{N\|\delta\mathbf{A}\|}{|\mathbf{y}_i^T \mathbf{x}_i|} \right\}$$

Proof: preliminaries

Lemma: Diagonalization Let $\mathbf{S}$ be the matrix with the right eigenvectors as columns. Then

$$\mathbf{S}^{-1} = \left(\frac{\mathbf{y}_1}{\mathbf{y}_1^T \mathbf{x}_1}, \frac{\mathbf{y}_2}{\mathbf{y}_2^T \mathbf{x}_2}, \ldots \frac{\mathbf{y}_N}{\mathbf{y}_N^T \mathbf{x}_N} \right)^T$$

Proof: plug in.

Proof: more preliminaries

Lemma: Suppose the columns of $\mathbf{S}$ are normalized ($\|\mathbf{s}_i\| = 1$). Then $\|\mathbf{S}\| \le \sqrt{N}$.

Proof: Let $\mathbf{x}$ be the unit vector so that $\|\mathbf{Sx}\| = \|\mathbf{S}\|$. Use Cauchy-Schwarz

$$\sum_{i=1}^{N} |a_i||b_i| \le \|\mathbf{a}\|_2 \|\mathbf{b}\|_2$$

$$\begin{aligned} \|\mathbf{S}\| &= \|\mathbf{Sx}\| = \| \textstyle\sum_{i=1}^{N} \mathbf{s}_i x_i \| \le \sum_{i=1}^{N} \|\mathbf{s}_i\| |x_i| \\ &\le \sqrt{\textstyle\sum_{i=1}^{N} \|\mathbf{s}_i\|^2} \sqrt{\textstyle\sum_{i=1}^{N} x_i^2} \le \sqrt{\textstyle\sum_{i=1}^{N}} = \sqrt{N}. \end{aligned}$$

Proof of Bauer-Fike: I

Note that $\mathbf{S}$ is the diagonalizing transformation for $\mathbf{A}$, so

$$\mathbf{S}^{-1}\mathbf{A}\mathbf{S} = \Lambda.$$

Apply Gersgorin to

$$\mathbf{B} = \mathbf{S}^{-1}(\mathbf{A} + \delta\mathbf{A})\mathbf{S} = \Lambda + \mathbf{F}$$

where $\mathbf{F} = \mathbf{S}^{-1}\delta\mathbf{A}\mathbf{S}$.
Gersgorin says that the eigenvalues of $\mathbf{B}$ lie in the disks,

$$G_i = \left\{ |\lambda - (\lambda_i + f_{ii})| \le \sum_{j \ne i} |f_{ij}| \right\}$$

Proof of Bauer-Fike: II

Since

$$G_i = \left\{ |\lambda - (\lambda_i + f_{ii})| \le \sum_{j \ne i} |f_{ij}| \right\}$$

Any $\lambda \in G_i$ satisfies

$$|\lambda - \lambda_i| - |f_{ii}| \le \sum_{j \ne i} |f_{ij}| \text{ which implies that}$$

$$|\lambda - \lambda_i| \le \sum_{i=1}^{N} |f_{ij}| \le \sqrt{N} \sqrt{\sum_{i=1}^{N} |f_{ij}|^2} = \sqrt{N} \|\mathbf{F}(i,:)\|$$

Proof of Bauer-Fike: III

So we need a bound on the ith row of $\mathbf{F}$.
Note that if $\mathbf{B} = \mathbf{B}_1\mathbf{B}_2$, then

$$\|\mathbf{B}(i,:)\| \le \|\mathbf{B}_1(i,:)\|\|\mathbf{B}_2\|$$

as you can see from the rules for matrix-matrix multiply.
So, since $\mathbf{F} = \mathbf{S}^{-1}\delta\mathbf{A}\mathbf{S}$,

$$\|\mathbf{F}(i,:)\| \le \|\mathbf{S}^{-1}(i,:)\|\|\delta\mathbf{A}\|\|\mathbf{S}\|$$

and we have formulae to estimate all this stuff ...

Proof of Bauer-Fike: IV

- Since the columns of $\mathbf{S}$ are the normalized eigenvectors

$$\|\mathbf{S}\| \le \sqrt{N}$$

  by one of the lemmas.

- Use the other lemma and $\|\mathbf{y}_i\| = 1$ to get

$$\|\mathbf{S}^{-1}(i,:)\| \le \frac{1}{|\mathbf{y}_i^T \mathbf{x}_i|}$$

and glue everything together to get ...

Proof of Bauer-Fike: V

$$\|\mathbf{F}(i,:)\| \le \frac{\sqrt{N}}{|\mathbf{y}_i^T \mathbf{x}|} \|\delta \mathbf{A}\|.$$

Plug into

$$|\lambda - \lambda_i| \le \sum_{i=1}^{N} |f_{ij}| \le \sqrt{N} \sqrt{\sum_{i=1}^{N} |f_{ij}|^2} = \sqrt{N} \|\mathbf{F}(i,:)\|$$

and we're done.

ℓ^p etimates: Bauer-Fike revisited

Theorem: Suppose

- $\mathbf{A}$ is diagonalizable with eigenvalues $\{\lambda_i\}$
- $\mathbf{x}_i$ are the normalized left eigenvectors corresponding to λ_i.
- $\mu \in \sigma(\mathbf{A} + \delta\mathbf{A})$

Then

$$\min_{\lambda \in \sigma(\mathbf{A})} |\mu - \lambda| \le \kappa_p(\mathbf{S}) \|\delta \mathbf{A}\|_p$$

where $\mathbf{S}$ is the matrix whose columns are the eigenvectors of $\mathbf{A}$.

Proof: I

If $\mu \in \sigma(\mathbf{A})$, then the left side of the estimate is 0.
Here we let $\|\cdot\|$ be any ℓ^p norm.
Otherwise, the matrix $\Lambda - \mu\mathbf{I}$ is not singular, but

$$\begin{aligned} \mathbf{S}^{-1}(\mathbf{A} + \delta\mathbf{A} - \mu\mathbf{I})\mathbf{S} &= \mathbf{S}^{-1}(\mathbf{A} - \mu\mathbf{I})\mathbf{S} + \mathbf{S}^{-1}\delta\mathbf{A}\mathbf{S} \\ &= (\Lambda - \mu\mathbf{I}) + \mathbf{S}^{-1}\delta\mathbf{A}\mathbf{S} \end{aligned}$$

is singular.

Proof: II

Multiply the singular matrix by $(\Lambda - \mu \mathbf{I})^{-1}$ to see that

$$\mathbf{I} + (\Lambda - \mu \mathbf{I})^{-1}(\mathbf{S}^{-1}\delta \mathbf{A}\mathbf{S})$$

is also singular. Hence

$$1 \leq \|(\Lambda - \mu \mathbf{I})^{-1}(\mathbf{S}^{-1}\delta \mathbf{A}\mathbf{S})\| \leq \|(\Lambda - \mu \mathbf{I})^{-1}\| \|\mathbf{S}^{-1}\| \|\delta \mathbf{A}\| \|\mathbf{S}\|$$

$$= \max_{\lambda \in \sigma(\mathbf{A})} \frac{1}{|\lambda - \mu|} \kappa_p(\mathbf{S}) \|\delta \mathbf{A}\|.$$

That's it since $\max(1/x) = 1/(\min x)$.

The QR algorithm

Consider this iteration:

$\mathbf{A}_0 = \mathbf{A}$
for $k = 0, \ldots$ **do**
  Factor $\mathbf{A}_k = \mathbf{QR}$
  $\mathbf{A}_{k+1} = \mathbf{RQ}$
end for

What does this have to do with eigenvalues?

Note that

$$\mathbf{A}_{k+1} = \mathbf{Q}^T\mathbf{QRQ} = \mathbf{Q}^T\mathbf{AQ}$$

is similar to $\mathbf{A}_k$, so has the same eigenvalues.
Let's give it a shot.

```
A=[1 2 3; 4 5 6; 7 8 9];
for i=1:10
    [q,r]=qr(A); A=r*q;
end
```

Results

The eigenvalues are

```
>> eig(A)
ans =
   1.6117e+01
  -1.1168e+00
  -1.3037e-15
```

and when the loop's done

```
A =
   1.6117e+01   4.8990e+00  -6.9295e-16
  -8.0448e-11  -1.1168e+00   1.6506e-15
            0            0            0
```

What?

For diagonalizable $\mathbf{A}$ with distinct real eigenvalues

- The iteration converges to an upper triangular matrix,
- which is similar to $\mathbf{A}$,
- and therefore has the same eigenvalues.

You can understand this via the power method.
This is the core of Matlab's `eig` code.

What a real code must do

- Reduce **A** to a form with a cheap QR factorization (upper Hessenberg),
- deal with multiple eigenvalues,
- deal with complex conjugate pairs of eigenvalues,
- build in shifts, ...

A feel-good theorem

Suppose:

- $\mathbf{A}$ is symmetric.
- $\mathbf{A}$ is nonsingular.
- The QR iterations $\mathbf{A}_n$, $\mathbf{R}_n$, $\mathbf{Q}_n$ converge to $\bar{\mathbf{A}}$, $\bar{\mathbf{Q}}$, $\bar{\mathbf{R}}$.

Then $\bar{\mathbf{A}}$ is diagonal with the eigenvalues of $\mathbf{A}$ along the diagonal.

Feel-good proof: I

Convergence implies that

$$\bar{\mathbf{A}} = \bar{\mathbf{Q}}\bar{\mathbf{R}} = \bar{\mathbf{R}}\bar{\mathbf{Q}}$$

Then symmetry implies that

$$\bar{\mathbf{A}}^T = \bar{\mathbf{Q}}^T\bar{\mathbf{R}}^T = \bar{\mathbf{R}}^T\bar{\mathbf{Q}}^T = \bar{\mathbf{A}} = \bar{\mathbf{Q}}\bar{\mathbf{R}} = \bar{\mathbf{R}}\bar{\mathbf{Q}}.$$

So

$$\bar{\mathbf{R}}^T\bar{\mathbf{R}} = \bar{\mathbf{R}}^T\bar{\mathbf{Q}}^T\bar{\mathbf{Q}}\bar{\mathbf{R}} = \bar{\mathbf{A}}^T\bar{\mathbf{A}} = \bar{\mathbf{A}}^2 = \bar{\mathbf{R}}\bar{\mathbf{Q}}\bar{\mathbf{Q}}^T\bar{\mathbf{R}}^T = \bar{\mathbf{R}}\bar{\mathbf{R}}^T$$

Feel-good proof: II

Since $\bar{\mathbf{R}}$ is upper triagular and

$$\bar{\mathbf{R}}^T \bar{\mathbf{R}} = \bar{\mathbf{R}} \bar{\mathbf{R}}^T$$

$\bar{\mathbf{R}}$ is diagonal. Let's prove this.

Lemma: Suppose $\mathbf{U}$ is upper triangular and $\mathbf{U}\mathbf{U}^T = \mathbf{U}^T\mathbf{U}$. Then $\mathbf{U}$ is diagonal.

Proof of Lemma: I

The proof is via induction. It's clear for $N = 1$. Assume that the theorem holds for dimensions up to $N - 1$. Let $\mathbf{U}$ be $N \times N$ upper triangular and decompose it as

$$\mathbf{U} = \begin{pmatrix} \mathbf{U}_1 & \mathbf{x} \\ 0 & \alpha \end{pmatrix}$$

where $\mathbf{U}_1$ is $(N-1) \times (N-1)$ upper triangular, $\mathbf{x} \in R^{N-1}$, and α is real.
Assume that $\mathbf{U}\mathbf{U}^T = \mathbf{U}^T\mathbf{U}$ then ...

Proof of Lemma: II

$$\mathbf{U}\mathbf{U}^T = \begin{pmatrix} \mathbf{U}_1 & \mathbf{x} \\ 0 & \alpha \end{pmatrix} \begin{pmatrix} \mathbf{U}_1^T & 0 \\ \mathbf{x}^T & \alpha \end{pmatrix} = \begin{pmatrix} \mathbf{U}_1\mathbf{U}_1^T + \mathbf{x}\mathbf{x}^T & \alpha\mathbf{x} \\ \alpha\mathbf{x}^T & \alpha^2 \end{pmatrix} =$$

$$\mathbf{U}^T\mathbf{U} = \begin{pmatrix} \mathbf{U}_1^T & 0 \\ \mathbf{x}^T & \alpha \end{pmatrix} \begin{pmatrix} \mathbf{U}_1 & \mathbf{x} \\ 0 & \alpha \end{pmatrix} = \begin{pmatrix} \mathbf{U}_1^T\mathbf{U}_1 & \mathbf{U}_1^T\mathbf{x} \\ \mathbf{x}^T\mathbf{U}_1 & \alpha^2 + \mathbf{x}^T\mathbf{x} \end{pmatrix}$$

So $\mathbf{x} = 0$ and ...

Proof of Lemma: III

- $\mathbf{U}_1^T \mathbf{U}_1 = \mathbf{U}_1 \mathbf{U}_1^T$, so
- $\mathbf{U}_1$ is diagonal by the induction hypothesis.
- $\mathbf{x} = 0$ implies that $\mathbf{U}$ is diagonal.

We are almost done.

Feel-good proof: III

Now that $\bar{\mathbf{R}}$ is diagonal, we can use

$$\bar{\mathbf{R}}^T\bar{\mathbf{R}} = \bar{\mathbf{R}}^T\bar{\mathbf{Q}}^T\bar{\mathbf{Q}}\bar{\mathbf{R}} = \bar{\mathbf{A}}^T\bar{\mathbf{A}} = \bar{\mathbf{A}}^2 = \bar{\mathbf{R}}\bar{\mathbf{Q}}\bar{\mathbf{Q}}^T\bar{\mathbf{R}}^T = \bar{\mathbf{R}}\bar{\mathbf{R}}^T$$

to conclude that

$$\bar{\mathbf{Q}}\bar{\mathbf{R}} = \bar{\mathbf{Q}}^T\bar{\mathbf{R}}^T = \bar{\mathbf{Q}}^T\bar{\mathbf{R}}$$

since $\bar{\mathbf{A}}$ is nonsingular, we must have

$$\bar{\mathbf{Q}} = \bar{\mathbf{Q}}^T = \bar{\mathbf{Q}}^{-1}.$$

Feel-good proof: IV

So $\bar{\mathbf{R}}$ is diagonal and $\bar{\mathbf{Q}}$ is symmetric. This means that

$$\bar{\mathbf{A}}^2 = \bar{\mathbf{Q}}\bar{\mathbf{R}}\bar{\mathbf{R}}\bar{\mathbf{Q}} = \bar{\mathbf{Q}}\bar{\mathbf{R}}^2\bar{\mathbf{Q}}$$

is a spectral decomposition of $\bar{\mathbf{A}}^2$ so

- The columns of $\bar{\mathbf{Q}}$ are eigenvectors of $\bar{\mathbf{A}}^2$
- and hence they are eigenvectors of $\bar{\mathbf{A}}$ (symmetry).
- So I can order the eigenvalues of $\bar{\mathbf{A}}$ so that

$$\bar{\mathbf{A}} = \bar{\mathbf{Q}}\Lambda\bar{\mathbf{Q}}$$

is a spectral decomposition of $\bar{\mathbf{A}}$

Feel-good proof: IV

We're done because

$$\bar{\mathbf{Q}}\Lambda = \bar{\mathbf{A}}\bar{\mathbf{Q}} = \bar{\mathbf{R}}\bar{\mathbf{Q}}\bar{\mathbf{Q}} = \bar{\mathbf{R}}$$

which means that

$$\Lambda = \bar{\mathbf{Q}}\bar{\mathbf{R}} = \bar{\mathbf{A}}$$

is diagonal and the eigenvectors of $\bar{\mathbf{A}}$ (which are the eigenvectors of $\mathbf{A}$ are the entries.

Convergence Theory for Happy Matrices

Assume that $\mathbf{A}$ has real distinct eigenvalues and

$$|\lambda_1| < |\lambda_2| < \ldots |\lambda_N|.$$

Then $\mathbf{A}_n \to \bar{\mathbf{R}}$ where $\bar{\mathbf{R}}$ has the eigenvalues of $\mathbf{A}$ on the diagonal. If $\mathbf{A}$ is symmetric, then $\mathbf{A}_n \to \Lambda$. Moreover

$$\|\bar{\mathbf{R}} - \mathbf{A}_n\| = O\left(\left[\max_i \frac{|\lambda_i|}{|\lambda_{i+1}|}\right]^n\right)$$

which sure does smell like the power method.