Numerical analysis: Homework 3

Student: Pratyush Sudhakar (ps2245)

Question 1:

Assume that we are given $A \in \mathbb{R}^{n\times n}$, $A = A^T$, and A has eigenvalue and vector pairs $\{(v_i, \lambda_i)\}_{i=1}^n$. Furthermore, assume that $|\lambda_1| = |\lambda_2| > |\lambda_3| \geq |\lambda_4| \geq \cdots$.

(a) Prove that for any initial guess $v^{(0)}$ such that $v^{(0)}$ is not simultaneously orthogonal to both v_1 and v_2 the power method yields iterates $v^{(k)}$ that converge to lie in the span of v_1 and v_2.

Solution: We know that v_0 is not orthogonal to both v_1 and v_2. Thus, we can write

$$v_0 = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n,$$

where not both c_1 and c_2 are zero.

Now, by the power method, we have

$$\begin{aligned} v^{(k)} &= A^k v^{(0)} \\ &= A^k(c_1 v_1 + c_2 v_2 + \cdots + c_n v_n) \\ &= c_1\lambda_1^k v_1 + c_2\lambda_2^k v_2 + \cdots + c_n \lambda_n^k v_n. \\ (\because A^k v_i &= \lambda_i^k v_i) \end{aligned}$$

We can write this as,

$$\begin{aligned} v^{(k)} &= c_1\lambda_1^k \left(v_1 + \frac{c_2}{c_1}\left(\frac{\lambda_2}{\lambda_1}\right)^k v_2 + \cdots + \frac{c_n}{c_1}\left(\frac{\lambda_n}{\lambda_1}\right)^k v_n \right); \\ \implies v^{(k)} &= c_1\lambda_1^k \left(v_1 + \frac{c_2}{c_1}\left(\frac{\lambda_2}{\lambda_1}\right)^k v_2 + \sum_{i=3}^n \frac{c_i}{c_1}\left(\frac{\lambda_i}{\lambda_1}\right)^k v_i \right). \end{aligned}$$

Since, $|\lambda_1| = |\lambda_2| > |\lambda_3| \geq |\lambda_4| \geq \cdots$, we have $\left(\frac{\lambda_2}{\lambda_1}\right) = \pm 1$ and $\left(\frac{\lambda_i}{\lambda_1}\right)^k \to 0$ as $k \to \infty$. Thus, we have

$$\begin{aligned} v^{(k)} &= c_1\lambda_1^k \left(v_1 + \frac{c_2}{c_1}\left(\frac{\lambda_2}{\lambda_1}\right)^k v_2 + \sum_{i=3}^n \frac{c_i}{c_1}\left(\frac{\lambda_i}{\lambda_1}\right)^k v_i \right) \\ &\approx c_1\lambda_1^k \left(v_1 \pm \frac{c_2}{c_1} v_2 \right); \;\; (\text{as } k \to \infty) \\ &= c_1\lambda_1^k \cdot v_1 \pm c_2\lambda_1^k \cdot v_2; \end{aligned}$$

where the $\pm$ sign depends on the sign of $\left(\frac{\lambda_2}{\lambda_1}\right)^k$. In both cases, we see that $v^{(k)}$ converges to the span of v_1 and v_2.

(b) What is the rate of convergence of

$$\left(1 - \left\| \left(v^{(k)}\right)^T \begin{bmatrix} v_1 & v_2 \end{bmatrix} \right\|_2^2 \right)^{1/2} ?$$

Solution: We know that $v^{(k)}$ converges to the span of v_1 and v_2. Thus, we can write $v^{(k)}$ as

$$v^{(k)} = c_1\lambda_1^k v_1 + c_2\lambda_2^k v_2 + \cdots + c_n\lambda_n^k v_n.$$

Now, we can write $v^{(k)}$ as

$$\begin{aligned}
v^{(k)} &= c_1\lambda_1^k\left(v_1 + \frac{c_2}{c_1}\left(\frac{\lambda_2}{\lambda_1}\right)^k v_2 + \cdots + \frac{c_n}{c_1}\left(\frac{\lambda_n}{\lambda_1}\right)^k v_n\right) \\
&= c_1\lambda_1^k\left(v_1 + \frac{c_2}{c_1}\left(\frac{\lambda_2}{\lambda_1}\right)^k v_2 + \sum_{i=3}^{n}\frac{c_i}{c_1}\left(\frac{\lambda_i}{\lambda_1}\right)^k v_i\right).
\end{aligned}$$

We know that the eigenvectors of a symmetric matrix are orthogonal. Thus, we can write the dot product of $v^{(k)}$ with v_1 and v_2 as

$$\begin{aligned}
\left\|\left(v^{(k)}\right)^T \begin{bmatrix} v_1 & v_2 \end{bmatrix}\right\|_2^2 &= \left\|\begin{bmatrix} c_1\lambda_1^k v_1 & c_2\lambda_2^k v_2 & \cdots & c_n\lambda_n^k v_n \end{bmatrix} \begin{bmatrix} v_1 & v_2 \end{bmatrix}\right\|_2^2 \\
&= \left\|\begin{bmatrix} c_1\lambda_1^k v_1^T v_1 & c_2\lambda_2^k v_2^T v_1 & \cdots & c_n\lambda_n^k v_n^T v_1 \end{bmatrix}\right\|_2^2 \\
&= \left\|\begin{bmatrix} c_1\lambda_1^k v_1^T v_1 & c_2\lambda_2^k v_2^T v_2 & 0 & \cdots & 0 \end{bmatrix}\right\|_2^2 && (\because v_i^T v_j = 0 \text{ for } i \neq j)^1 \\
&= \|c_1\lambda_1^k \quad c_2\lambda_2^k\|_2^2 && (\because v_i^T v_i = 1;\ \forall i)^2
\end{aligned}$$

$$= \frac{\left(c_1\lambda_1^k\right)^2 + \left(c_2\lambda_2^k\right)^2}{||v_k||_2^2}$$

We need to determine the norm of $v^{(k)}$:

$$\|v^{(k)}\|_2^2 = \left\|c_1\lambda_1^k v_1 + c_2\lambda_2^k v_2 + \sum_{i=3}^{n} c_i\lambda_i^k v_i\right\|_2^2$$

Since eigenvectors are orthogonal:

$$\|v^{(k)}\|_2^2 = \left(c_1\lambda_1^k\right)^2 + \left(c_2\lambda_2^k\right)^2 + \sum_{i=3}^{n}\left(c_i\lambda_i^k\right)^2$$

The term we are interested in is:

$$\left(1 - \left\|\left(v^{(k)}\right)^T \begin{bmatrix} v_1 & v_2 \end{bmatrix}\right\|_2^2\right)^{1/2}$$

Substituting the expressions derived above:

$$\left\|\left(v^{(k)}\right)^T \begin{bmatrix} v_1 & v_2 \end{bmatrix}\right\|_2^2 = \frac{\left(c_1\lambda_1^k\right)^2 + \left(c_2\lambda_2^k\right)^2}{\left(c_1\lambda_1^k\right)^2 + \left(c_2\lambda_2^k\right)^2 + \sum_{i=3}^{n}\left(c_i\lambda_i^k\right)^2}$$

Therefore:

$$1 - \left\|\left(v^{(k)}\right)^T \begin{bmatrix} v_1 & v_2 \end{bmatrix}\right\|_2^2 = 1 - \frac{\left(c_1\lambda_1^k\right)^2 + \left(c_2\lambda_2^k\right)^2}{\left(c_1\lambda_1^k\right)^2 + \left(c_2\lambda_2^k\right)^2 + \sum_{i=3}^{n}\left(c_i\lambda_i^k\right)^2}$$

Simplifying the fraction:

$$1 - \left\| \left(v^{(k)}\right)^T \begin{bmatrix} v_1 & v_2 \end{bmatrix} \right\|_2^2 = \frac{\sum_{i=3}^n \left(c_i \lambda_i^k\right)^2}{\left(c_1 \lambda_1^k\right)^2 + \left(c_2 \lambda_2^k\right)^2 + \sum_{i=3}^n \left(c_i \lambda_i^k\right)^2}$$

As $k \to \infty$, the terms involving λ_i (for $i \geq 3$) will dominate the denominator:

$$\frac{\sum_{i=3}^n \left(c_i \lambda_i^k\right)^2}{\left(c_1 \lambda_1^k\right)^2 + \left(c_2 \lambda_2^k\right)^2}$$

Since $|\lambda_3| < |\lambda_1|$ and $|\lambda_3| < |\lambda_2|$, this fraction tends to zero as $k \to \infty$.

The rate of convergence is dominated by the largest λ_i such that $|\lambda_i| < |\lambda_1|$. This gives the rate of convergence as:

$$\left(1 - \left\| \left(v^{(k)}\right)^T \begin{bmatrix} v_1 & v_2 \end{bmatrix} \right\|_2^2\right)^{1/2} \approx \left(\frac{|\lambda_3|}{|\lambda_1|}\right)^k$$

Thus, the solution correctly concludes that the rate of convergence is $\left(\frac{|\lambda_3|}{|\lambda_1|}\right)^k$, which is the slowest decaying term due to $|\lambda_3|$ being the largest eigenvalue less than $|\lambda_1|$.

(c) Does the associated eigenvalue estimate via the Rayleigh quotient necessarily converge in this setting? what about if $\lambda_1 = \lambda_2$?

Solution: The Rayleigh quotient is defined as

$$\rho(v) = \frac{v^T A v}{v^T v} = \frac{(\sum_{i=1}^n c_i \lambda_i^k v_i)^T \cdot A \cdot \sum_{i=1}^n c_i \lambda_i^k v_i}{\sum_{i=1}^n c_i^2 \cdot \lambda_i^{2k}}.$$

As $k \to \infty$, we have $\lambda_i^k \to 0$ for $i > 2$. Thus, the Rayleigh quotient converges to $\rho(v) = \frac{c_1^2 \lambda_1^k \lambda_1^{k+1} + c_2^2 \lambda_2^k \lambda_2^{k+1}}{c_1^2 \lambda_1^{2k} + c_2^2 \lambda_2^{2k}}$. Since, the first two eigen values will dominate the Rayleigh quotient.

If $\lambda_1 = -\lambda_2$, then the Rayleigh quotient will be,

$$\rho(v) = \frac{c_1 \lambda_1^k \lambda_1^{k+1} + c_2 \lambda_2^k \lambda_2^{k+1}}{c_1^2 \lambda_1^{2k} + c_2^2 \lambda_2^{2k}} = \frac{c_1 \lambda_1^k \lambda_1^{k+1} - c_1 \lambda_1^k \lambda_1^{k+1}}{c_1^2 \lambda_1^{2k} + c_1^2 \lambda_1^{2k}} = \frac{\lambda(c_1^2 - c_2^2)}{c_1^2 + c_2^2}$$

We know that it will converge to 0 when $|c_1| \approx |c_2|$. Thus, the Rayleigh quotient will not converge in this setting.

If $\lambda_1 = \lambda_2$, then the Rayleigh quotient will be,

$$\rho(v) = \frac{c_1 \lambda_1^k \lambda_1^{k+1} + c_2 \lambda_2^k \lambda_2^{k+1}}{c_1^2 \lambda_1^{2k} + c_2^2 \lambda_2^{2k}} = \frac{c_1 \lambda_1^k \lambda_1^{k+1} + c_2 \lambda_1^k \lambda_1^{k+1}}{c_1^2 \lambda_1^{2k} + c_1^2 \lambda_1^{2k}} = \lambda_1$$

Thus, the Rayleigh quotient will accurately converge to λ_1 when $\lambda_1 = \lambda_2$.

Question 2:

(a) If we denote $\hat{\lambda}_1^{(k)}$ as our guess for λ_1 at iteration k, show that $\lambda_1 \geq \hat{\lambda}_1^{(k)}$ for all k. I.e., our guess for λ_1 converges from below.

Solution: We will use Rayleigh quotient to show that $\lambda_1 \geq \hat{\lambda}_1^{(k)}$ for all k.

We know that the Rayleigh quotient is defined as

$$\rho(v) = \frac{v^T A v}{v^T v}$$

We know that the Rayleigh quotient converges to the largest eigenvalue, λ_1, when v is the eigenvector corresponding to λ_1. We know that λ_1 is the largest eigenvalue of A. Thus, the Rayleigh quotient's maximum value is λ_1.

$$\lambda_1 = max(\rho(v)) = max(\frac{v^T A v}{v^T v}), \ and$$
$$\hat{\lambda}_1^{(k)} = \frac{v^{(k)T} A v^{(k)}}{v^{(k)T} v^{(k)}};$$

where $v^{(k)}$ is the estimated eigenvector at kth iteration.

Since the Rayleigh quotient gives an estimate of the eigenvalue that is bounded above by the largest eigenvalue, we have:

$$\rho(v^{(k)}) \leq \lambda_1$$

Therefore:

$$\hat{\lambda}_1^{(k)} = \rho(v^{(k)}) \leq \lambda_1$$

As $v^{(k)}$ approaches the eigenvector corresponding to λ_1, it yields an increasingly accurate estimate of λ_1. The Rayleigh quotient's maximum value for any vector v is λ_1. If $\hat{\lambda}_1^{(k)}$ were to exceed λ_1, it would contradict λ_1 being the largest eigenvalue. Thus, our iterative approximation improves and converges to λ_1 without exceeding it.

Thus, we have $\lambda_1 \geq \hat{\lambda}_1^{(k)}$ for all k.

(b) If we instead assume $\lambda_1 > \lambda_2 > \cdots > \lambda_\ell > \lambda_{\ell+1} \geq \cdots \geq \lambda_n \geq 0$, would we expect the columns of $V^{(k)}$ to converge to individual eigenvectors (in an appropriate sense)? If so, what might be expect the asymptotic rates of convergence to be?

Solution:

Assuming $\lambda_1 > \lambda_2 > \cdots > \lambda_\ell > \lambda_{\ell+1} \geq \cdots \geq \lambda_n \geq 0$, we have distinct first ℓ eigenvalues. In this scenario, the columns of $V^{(k)}$ will indeed converge to individual eigenvectors. The orthogonal iteration ensures that the vectors remain orthogonal through each iteration, progressively aligning the subspace with the eigenvectors.

The convergence rate for each eigenvector depends on the gaps between the eigenvalues. Specifically, the rate of convergence for the i-th eigenvector is influenced by the ratio between the $(i+1)$-th eigenvalue and the i-th eigenvalue. Mathematically, the convergence rate for the i-th eigenvector is determined by:

$$\left| \frac{\lambda_{i+1}}{\lambda_i} \right|$$

For example, if there is a significant difference between the first two eigenvalues λ_1 and λ_2, the first column of $V^{(k)}$ will converge more rapidly to the first eigenvector. Conversely, if the eigenvalues λ_ℓ and $\lambda_{\ell+1}$ are close in value, the convergence of the ℓ-th column to the ℓ-th eigenvector will be slower.

The larger the gap between successive eigenvalues, the faster the convergence. This is because a larger gap ensures that the component of the iteration vector in the direction of the i-th eigenvector grows more rapidly relative to the components in the directions of other eigenvectors, thus accelerating convergence.

In conclusion, the columns of $V^{(k)}$ will converge to the individual eigenvectors at rates proportional to the ratios of successive eigenvalues.