Iterative Methods for Eigenvalues of Symmetric Matrices as Fixed Point Theorems

Student: Amanda Schaeffer
Sponsor: Wilfred M. Greenlee

December 6, 2007

1. The Power Method and the Contraction Mapping Theorem

Understanding the following derivations, definitions, and theorems may be helpful to the reader.

The Power Method

Let A be a symmetric n x n matrix with eigenvalues $|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq |\lambda_4| \geq ... \geq |\lambda_n| \geq 0$ and corresponding orthonormal eigenvectors $\vec{u}_1, \vec{u}_2, \vec{u}_3, ..., \vec{u}_n$. We call λ_1 the dominant eigenvalue, and $\vec{u}_1$ the dominant eigenvector. The Power Method is a basic method of iteration for computing this dominant eigenvector. The algorithm is as follows:

1. Choose a unit vector, $\vec{x}_0$.
2. Let $\vec{y}_1 = A\vec{x}_0$.
3. Normalize, letting $\vec{x}_1 = \frac{\vec{y}_1}{\|\vec{y}_1\|}$.
4. Repeat steps 2 and 3, letting $\vec{y}_k = A\vec{x}_{k-1}$, until $\vec{x}_k$ converges.

[4]

The following shows that the limit of this sequence is the dominant eigenvector, provided that $(\vec{x} \cdot \vec{u_1})$ is nonzero.

Let $\vec{x} \in \mathbf{R}^n$. Because the eigenvectors of A are orthonormal and therefore span $\mathbf{R}^n$, we can write: $\vec{x} = \sum_{i=1}^{n}(\alpha_i \vec{u}_i)$, where $\alpha_i \in R$. Then multiplying by the matrix, A, we find:

$$A\vec{x} = \sum_{i=1}^{n}(\alpha_i A\vec{u}_i) = \sum_{i=1}^{n}(\alpha_i \lambda_i \vec{u}_i)$$
$$A^2\vec{x} = \sum_{i=1}^{n}(\alpha_i A\lambda_i \vec{u}_i) = \sum_{i=1}^{n}(\alpha_i \lambda_i^2 \vec{u}_i)$$
$$...A^k\vec{x} = \sum_{i=1}^{n}(\alpha_i \lambda_i^k \vec{u}_i).$$

By factoring λ_1^k, we obtain:

$$A^k\vec{x} = \lambda_1^k \sum_{i=1}^{n} \alpha_i (\tfrac{\lambda_i}{\lambda_1})^k \vec{u}_i = \lambda_1^k \alpha_1 \vec{u}_1 + \lambda_1^k \sum_{i=2}^{n} \alpha_i (\tfrac{\lambda_i}{\lambda_1})^k \vec{u}_i.$$

So since $|\lambda_2| \geq |\lambda_i|$ for all $i = 2, 3, ..., n$, we have that

$$\frac{A^k\vec{x}}{\lambda_1^k} = \alpha_1 \vec{u}_1 + O(\tfrac{\lambda_2}{\lambda_1})^k$$

as $k \to \infty$.

Definition: Metric Space

A set of elements, X, is said to be a *metric space* if to each pair of elements $u, v \in X$, there is associated a real number $d(u, v)$, the distance between u and v, such that:

1. $d(u, v) > 0$ for u, v distinct,
2. $d(u, u) = 0$,
3. $d(u, v) = d(v, u)$,
4. $d(u, w) \leq d(u, v) + d(v, w)$ (The Triangle Inequality holds). [5]

Definition: Complete Metric Space

Let X be a metric space. X is said to be *complete* if every Cauchy sequence in X converges to a point of X.

Definition: Contraction

Let X be a metric space. A transformation, $T : X \to X$, is called a *contraction* if for some fixed $\rho < 1$,

$$d(Tu, Tv) \leq \rho d(u, v) \text{ for all } u, v \in X. \text{ [5]}$$

The Contraction Mapping Theorem

Let T be a contraction on a complete metric space X. Then there exists exactly one solution, $u \in X$, to $u = Tu$. [5]

Note: We leave the proof of this theorem to the discussion of our specific example below.

2. A contraction for finding the dominant eigenvector

Let A be a symmetric n x n matrix with eigenvalues $|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq |\lambda_4| \geq ... \geq |\lambda_n| \geq 0$ and corresponding orthonormal eigenvectors $\vec{u}_1, \vec{u}_2, \vec{u}_3, ..., \vec{u}_n$. Let $\vec{x}, \vec{y} \in R^n$. Because the eigenvectors of A are orthonormal and therefore span $\mathbf{R}^n$, we can write: $\vec{x} = \sum_{i=1}^{n}(\alpha_i \vec{u}_i)$ and $\vec{y} = \sum_{i=1}^{n}(\beta_i \vec{u}_i)$ with $\alpha_i, \beta_i \in \mathbf{R}$. These can also be defined as: $\alpha_i = (\vec{x} \cdot \vec{u}_i)$, $\beta_i = (\vec{y} \cdot \vec{u}_i)$.

The Contraction

Define the transformation $T = \frac{A}{\lambda_1}$. Then we have the following:

$$T\vec{x} = \alpha_1 \vec{u}_1 + \sum_{i>1} \alpha_i (\tfrac{\lambda_i}{\lambda_1}) \vec{u}_i$$
$$T\vec{y} = \beta_1 \vec{u}_1 + \sum_{i>1} \beta_i (\tfrac{\lambda_i}{\lambda_1}) \vec{u}_i$$
$$T\vec{x} - T\vec{y} = (\alpha_1 - \beta_1)\vec{u}_1 + \sum_{i>1} (\alpha_i - \beta_i)(\tfrac{\lambda_i}{\lambda_1}) \vec{u}_i.$$

Now, consider $\|T\vec{x} - T\vec{y}\|$. In order to have a contraction, we want some $0 < k < 1$ such that $\|T\vec{x} - T\vec{y}\| < k\|\vec{x} - \vec{y}\|$. To find this, we will use $\|T\vec{x} - T\vec{y}\|^2$. Then we want our k such that $\|T\vec{x} - T\vec{y}\|^2 < k^2\|\vec{x} - \vec{y}\|^2$.

Since the $\vec{u}_i$'s are orthonormal, we find that

$$\|T\vec{x} - T\vec{y}\|^2 = (\alpha_1 - \beta_1)^2 + \textstyle\sum_{i>1}((\alpha_i - \beta_i)^2(\frac{\lambda_i}{\lambda_1})^2).$$

In the case that $\beta_1 = \alpha_1$,

$$\begin{aligned}\|T\vec{x} - T\vec{y}\|^2 &= \textstyle\sum_{i>1}((\alpha_i - \beta_i)^2(\frac{\lambda_i}{\lambda_1})^2)\\ &\leq (\tfrac{\lambda_2}{\lambda_1})^2 \textstyle\sum_{i>1}((\alpha_i - \beta_i)^2)\\ &= (\tfrac{\lambda_2}{\lambda_1})^2\|\vec{x} - \vec{y}\|^2\end{aligned}$$

so, $\|T\vec{x} - T\vec{y}\|^2 \leq (\frac{\lambda_2}{\lambda_1})^2\|\vec{x} - \vec{y}\|^2$.
Then $\|T\vec{x} - T\vec{y}\| \leq |\frac{\lambda_2}{\lambda_1}|\|\vec{x} - \vec{y}\|$.

Therefore, when $\beta_1 = \alpha_1$, we have T as a contraction with $k = |\frac{\lambda_2}{\lambda_1}|$.

The Fixed Point

Let $\vec{x_0} \in R^n$ be our initial guess for finding an eigenvector. Define:

$$V = \{\vec{y} \in R^n | (\vec{y} \cdot \vec{u}_1) = (\vec{x_0} \cdot \vec{u}_1)\}.$$

(ie, V is the set of vectors such that $\beta_1 = \alpha_1$). Then T maps V back into V. Because this set is a closed subset of $\mathbf{R}^n$, and $\mathbf{R}^n$ is a complete metric space, V is also complete. Then by the Contraction Mapping Theorem, there is exactly one $\vec{y_f} \in V$ such that $\vec{y_f} = T\vec{y_f}$. It follows (from the below proof) that $\vec{y_f} = (\vec{x_0} \cdot \vec{u}_1)\vec{u}_1$ is the unique fixed point.

We now prove the Contraction Mapping Theorem in our context, as mentioned in Part 1.

Proof:

(a) "Uniqueness": The first thing we need to show is that the fixed point of our contraction is unique. Suppose $\vec{u} = T\vec{u}$ and $\vec{v} = T\vec{v}$. Then $\|\vec{v} - \vec{u}\| =$

$\|T\vec{v} - T\vec{u}\|$. But then, $\|T\vec{v} - T\vec{u}\| \leq |\frac{\lambda_2}{\lambda_1}|\|\vec{v} - \vec{u}\|$. So by substitution, $\|\vec{v} - \vec{u}\| \leq |\frac{\lambda_2}{\lambda_1}|\|\vec{v} - \vec{u}\|$. But then $\|\vec{v} - \vec{u}\| = 0$, so $\vec{v} = \vec{u}$. Thus, there is only one fixed point, $\vec{y_f}$.

(b) "Existence": Next, we want to show that there exists a fixed point at all. Again, let $\vec{x_0}$ be our initial guess. Define $\vec{x_m} = T\vec{x_{m-1}}$. Then we have:

$$\|\vec{x_m} - \vec{x_{m+1}}\| = \|T\vec{x_{m-1}} - T\vec{x_m}\| \leq |\frac{\lambda_2}{\lambda_1}|\|\vec{x_{m-1}} - \vec{x_m}\| \leq ... \leq |\frac{\lambda_2}{\lambda_1}|^m\|\vec{x_0} - \vec{x_1}\|.$$

Therefore, for $n > m$, we see that

$$\begin{aligned}
\|\vec{x_m} - \vec{x_n}\| &\leq \|\vec{x_m} - \vec{x_{m+1}}\| + ... + \|\vec{x_{n-1}} - \vec{x_n}\| \\
&\leq |\tfrac{\lambda_2}{\lambda_1}|^m\|\vec{x_0} - \vec{x_1}\| + ... + |\tfrac{\lambda_2}{\lambda_1}|^{n-1}\|\vec{x_0} - \vec{x_1}\| \\
&= \|\vec{x_0} - \vec{x_1}\||\tfrac{\lambda_2}{\lambda_1}|^m(1 + |\tfrac{\lambda_2}{\lambda_1}| + |\tfrac{\lambda_2}{\lambda_1}|^2 + ... + |\tfrac{\lambda_2}{\lambda_1}|^{n-m-1}) \\
&\leq \|\vec{x_0} - \vec{x_1}\|\frac{|\frac{\lambda_2}{\lambda_1}|^m}{1-|\frac{\lambda_2}{\lambda_1}|}.
\end{aligned}$$

Note that since $0 < |\frac{\lambda_2}{\lambda_1}| < 1$, we replaced the geometric sum with the sum of the infinite series. Now we see that $\{x_k\}$ forms a Cauchy sequence, and thus converges to an element of V since V is complete. Call the limit $\vec{x}$. Then T is a contraction, and hence continuous, so $\vec{x}$ is the fixed point, since:

$$\vec{x} = lim_{n\to\infty}\vec{x}_n = lim_{n\to\infty}T\vec{x}_{n-1} = T(lim_{n\to\infty}\vec{x}_{n-1}) = T\vec{x}.$$

(c) "The fixed point x is $\vec{y_f} = (\vec{x_0} \cdot \vec{u}_1)\vec{u}_1$": Finally, we want to find the fixed point for our particular contraction, $T = \frac{A}{\lambda_1}$. Consider $T(\vec{x_0} \cdot \vec{u}_1)\vec{u}_1$. This is $\frac{A}{\lambda_1}(\vec{x_0} \cdot \vec{u}_1)\vec{u}_1 = (\vec{x_0} \cdot \vec{u}_1)\vec{u}_1$. Then this is the fixed point.

The Schwartz Quotient

Because we do not know λ_1, we must find an adequate approximation. For this, we use the Schwartz quotient. Consider a positive symmetric matrix A. We define the Schwartz quotient, $\Lambda^{(m)}$, as follows:

$$\Lambda^{(m)} = (\frac{A^{m+1}\vec{x_0} \cdot A^m\vec{x_0}}{A^m\vec{x_0} \cdot A^m\vec{x_0}})$$

We show that this converges to λ_1, provided $\alpha_1 = (\vec{x_0} \cdot \vec{u}_1) \neq 0$.

$$\Lambda^{(m)} = (\frac{A^{m+1}\vec{x_0} \cdot A^m \vec{x_0}}{A^m \vec{x_0} \cdot A^m \vec{x_0}})$$
$$= \lambda_1 \frac{\alpha_1^2 + \sum_{i=2}^n (\frac{\lambda_i}{\lambda_1})^{2m+1} \alpha_i^2}{\alpha_1^2 + \sum_{i=2}^n (\frac{\lambda_i}{\lambda_1})^{2m} \alpha_i^2}$$
$$\leq \lambda_1,$$

since $\frac{\lambda_i}{\lambda_1} < 1$ for each $i = 2, ..., n$

But also,

$$\Lambda^{(m)} = \lambda_1 \frac{\alpha_1^2 + \sum_{i=2}^n (\frac{\lambda_i}{\lambda_1})^{2m+1} \alpha_i^2}{\alpha_1^2 + \sum_{i=2}^n (\frac{\lambda_i}{\lambda_1})^{2m} \alpha_i^2}$$
$$\geq \lambda_1 \frac{\alpha_1^2}{\alpha_1^2 + \sum_{i=2}^n (\frac{\lambda_i}{\lambda_1})^{2m} \alpha_i^2}$$
$$\geq \lambda_1 \frac{\alpha_1^2}{\alpha_1^2 + (\frac{\lambda_2}{\lambda_1})^{2m} \sum_{i=2}^n \alpha_i^2}$$
$$= \lambda_1 \frac{\alpha_1^2}{\alpha_1^2 + (\frac{\lambda_2}{\lambda_1})^{2m} (1 - \alpha_1^2)}$$

(Notice: $\alpha_1^2 + \sum_{i=2}^n \alpha_i^2 = \sum_{i=1}^n \alpha_i^2 = \|\vec{x_0}\|^2 = 1$, since $\vec{x_0}$ is normalized.)

So we find that

$$\lambda_1 \frac{\alpha_1^2}{\alpha_1^2 + (\frac{\lambda_2}{\lambda_1})^{2m} (1 - \alpha_1^2)} \leq \Lambda^{(m)} \leq \lambda_1,$$

and thus,

$$\Lambda^{(m)} = \lambda_1 + O(\frac{\lambda_2}{\lambda_1})^{2m}.$$

Notice that for a negative symmetric matrix, the inequalities are reversed, though the end result remains unchanged. Also note that the asymptotic conclusion

$$\Lambda^{(m)} = \lambda_1 + O(\frac{\lambda_2}{\lambda_1})^{2m}$$

holds without the assumption that A is definite. This follows by use of a finite geometric series argument based on the explicit formula

$$\Lambda^{(m)} = \lambda_1 \frac{\alpha_1^2 + \sum_{i=2}^n (\frac{\lambda_i}{\lambda_1})^{2m+1} \alpha_i^2}{\alpha_1^2 + \sum_{i=2}^n (\frac{\lambda_i}{\lambda_1})^{2m} \alpha_i^2}.$$

Then we define a new set of transformations $S_m : V \to V$ such that $S_m = \frac{A}{\Lambda^{(m)}}$. By applying S_m to A and normalizing, we iterate to find $\vec{y_f}$, a procedure which converges to the dominant eigenvector, $\vec{u_1}$, as in the case with the original mapping.

Note: There are two manners of doing this iteration. In the way we analyze below, call this Version I, we iterate the Schwartz quotient until it is satisfactorily close to the dominant eigenvalue, and then we iterate using the corresponding S_m to find the dominant eigenvector. However, in the example below, we use what we will call Version II of the iteration. We found the corresponding S_m for each iteration of the Schwartz quotient, and thus with each iteration, the transformation was different. While this version of the iteration still converges to the dominant eigenvector, finding the rate of convergence is more difficult than that of the first version due to the nature of this type of iteration. In the following sections, we will find the rate of convergence of Version II.

$$\text{Recall: } T\vec{x_0} = \frac{A}{\lambda_1}(\vec{x_0}) = \alpha_1\vec{u}_1 + \sum_{i>1}\alpha_i(\frac{\lambda_i}{\lambda_1})\vec{u}_i$$

Now we find that, using Version I:

$$\begin{gathered}
S_m(\vec{x_0}) = \frac{A}{\Lambda^{(m)}}(\vec{x_0}) = \alpha_1\frac{\lambda_1}{\Lambda^{(m)}}\vec{u_1} + \sum_{i=2}^{n}(\frac{\lambda_i}{\Lambda^{(m)}})\alpha_i\vec{u_i} \\
T^m(\vec{x_0}) - S_m^m(\vec{x_0}) = \alpha_1(1 - \frac{\lambda_1}{\Lambda^{(m)}})^m\vec{u}_1 + \sum_{i>1}\alpha_i[(\frac{\lambda_i}{\lambda_1})^m - (\frac{\lambda_i}{\Lambda^{(m)}})^m]\vec{u}_i \\
\|T^m(\vec{x_0}) - S_m^m(\vec{x_0})\|^2 = \alpha_1^2(1 - \frac{\lambda_1}{\Lambda^{(m)}})^{2m} + \sum_{i>1}\alpha_i^2[(\frac{\lambda_i}{\lambda_1})^m - (\frac{\lambda_i}{\Lambda^{(m)}})^m]^2 \\
= \alpha_1^2(1 - \frac{\lambda_1}{\Lambda^{(m)}})^{2m} + \sum_{i>1}\alpha_i^2(\frac{\lambda_i}{\lambda_1})^{2m}[1 - (\frac{\lambda_1}{\Lambda^{(m)}})^m]^2
\end{gathered}$$

Then

$$\|T^m(\vec{x_0}) - S_m^m(\vec{x_0})\| = O(1 - \frac{\lambda_1}{\Lambda^{(m)}})^m$$

But, recall:

$$\Lambda^{(m)} = \lambda_1 + O(\frac{\lambda_2}{\lambda_1})^{2m}$$

So

$$\|T^m(\vec{x_0}) - S_m^m(\vec{x_0})\| = O(1 - \frac{\lambda_1}{\lambda_1 + O(\frac{\lambda_2}{\lambda_1})^{2m}})^m$$
$$= O(\tfrac{\lambda_2}{\lambda_1})^{2m}$$

Now,

$$\|S_m^m(\vec{x_0}) - \alpha_1\vec{u}_1\| = \|S_m^m(\vec{x_0}) - T^m(\vec{x_0}) + T^m(\vec{x_0}) - \alpha_1\vec{u}_1\|$$
$$\leq \|S_m^m(\vec{x_0}) - T^m(\vec{x_0})\| + \|T^m(\vec{x_0}) - \alpha_1\vec{u}_1\|$$

by the triangle inequality.

$$= O(\tfrac{\lambda_2}{\lambda_1})^{2m} + O(\tfrac{\lambda_2}{\lambda_1})^m$$

So,

$$\|S_m^m(\vec{x_0}) - \alpha_1\vec{u}_1\| = O(\tfrac{\lambda_2}{\lambda_1})^m = \|T^m(\vec{x_0}) - \alpha_1\vec{u}_1\|.$$

Then these transformations have the same rate of convergence.

3. Example

1) Consider the following 3×3 matrix (found in [3], Ex. 4-2-1):

$$\begin{pmatrix} -4 & 10 & 8 \\ 10 & -7 & -2 \\ 8 & -2 & 3 \end{pmatrix}$$

This matrix has known eigenvalues

$$\lambda_1 \approx -17.895;$$
$$\lambda_2 \approx 9.470;$$
$$\lambda_3 \approx 0.425,$$

and corresponding eigenvectors

$$\vec{u_1} \approx \langle 0.6679, -0.6719, -0.32\rangle;$$
$$\vec{u_2} \approx \langle 0.6437, 0.30566, 0.7016\rangle;$$
$$\vec{u_3} \approx \langle 0.37358, 0.6746, -0.6366\rangle.$$

Using the Schwartz Quotient to find the approximate eigenvalue, $\Lambda^{(m)}$, and Version II of our fixed point mapping of the power method to find the approximate eigenvector, $\vec{v_m}$, and beginning with an initial guess $\vec{x_0} = \langle 1, 1, 1 \rangle$, we find:

m	$\Lambda^{(m)}$	$\vec{v_m}$
1	6.15827	$\langle 0.253216, 1.12, 1.133426 \rangle$
2	0.453015	$\langle 0.920897, -0.352108, 0.16724 \rangle$
3	-7.96152	$\langle 0.381492, -0.737348, -0.557 \rangle$
4	-14.1284	$\langle -0.786587, 0.594171, 0.168054 \rangle$
5	-16.7232	$\langle 0.592227, -0.701714, -0.3961 \rangle$
6	-17.5553	$\langle -0.704775, 0.652672, 0.278051 \rangle$
7	-17.7978	$\langle 0.6474, -0.6811, -0.3418 \rangle$
8	-17.8665	$\langle -0.6785, 0.6667, 0.3084 \rangle$
9	-17.8858	$\langle 0.66224, -0.674558, -0.3262 \rangle$
10	-17.8912	$\langle -0.670892, 0.6705, 0.3168 \rangle$
11	-17.8927	$\langle 0.66633, -0.6726, -0.3218 \rangle$

4. Monotonicity of Schwartz Quotients

In this section, we alter a proof found in Collatz [1] to find that in the two cases of a symmetric, positive definite matrix or negative definite matrix, the sequence of Schwartz Quotients, $\{\Lambda^{(m)}\}$, where $\Lambda^{(m)} = (\frac{A^{m+1}\vec{x_0} \cdot A^m\vec{x_0}}{A^m\vec{x_0} \cdot A^m\vec{x_0}})$, is monotone increasing or decreasing, respectively. This is also the behavior observed in the preceding example, with an indefinite matrix.

Case 1: The positive definite symmetric matrix

Consider a matrix, A, which is positive definite. The m-th term in the sequence of Schwartz Quotients is $\Lambda^{(m)} = (\frac{A^{m+1}\vec{x_0} \cdot A^m\vec{x_0}}{A^m\vec{x_0} \cdot A^m\vec{x_0}})$. Define $a_{2m+1} = A^{m+1}\vec{x_0} \cdot A^m\vec{x_0}$ and $a_{2m} = A^m\vec{x_0} \cdot A^m\vec{x_0}$. Then $\Lambda^{(m)} = \frac{a_{2m+1}}{a_{2m}}$.

Note: In a similar manner, we define a_{2m-1} as $(A^m\vec{x_0} \cdot A^{m-1}\vec{x_0})$, etc.

Then, let $Q_1 = \|a_{2m+1}A^m\vec{x_0} - a_{2m}A^{m+1}\vec{x_0}\|^2$. This value is obviously non-negative. But then

$$Q_1 = (a_{2m+1}A^m\vec{x_0} - a_{2m}A^{m+1}\vec{x_0}) \cdot (a_{2m+1}A^m\vec{x_0} - a_{2m}A^{m+1}\vec{x_0})$$

By the bilinearity of the inner product, this is

$$\begin{gathered}= (a_{2m+1}A^m\vec{x_0}) \cdot (a_{2m+1}A^m\vec{x_0}) - (a_{2m+1}A^m\vec{x_0}) \cdot (a_{2m}A^{m+1}\vec{x_0}) - \\ (a_{2m}A^{m+1}\vec{x_0}) \cdot (a_{2m+1}A^m\vec{x_0}) + (a_{2m}A^{m+1}\vec{x_0}) \cdot (a_{2m}A^{m+1}\vec{x_0}) \\ = (a_{2m+1}^2 a_{2m}) - 2(a_{2m+1}a_{2m}a_{2m+1}) + (a_{2m}^2 a_{2m+2}) \\ = (a_{2m}^2 a_{2m+2}) - (a_{2m+1}^2 a_{2m})\end{gathered}$$

Then

$$(a_{2m}^2 a_{2m+2}) - (a_{2m+1}^2 a_{2m}) \geq 0,$$

and hence

$$(a_{2m}^2 a_{2m+2}) \geq (a_{2m+1}^2 a_{2m}).$$

Using this result, we find that

$$\frac{Q_1}{(a_{2m}^2 a_{2m+1})} = (\frac{a_{2m+2}}{a_{2m+1}} - \frac{a_{2m+1}}{a_{2m}}) \geq 0 \tag{1}$$

since we assume A positive definite, and thus $a_{2m+1} > 0$.

Now define $Q_2 = (a_{2m+1}A^m\vec{x_0} - a_{2m}A^{m+1}\vec{x_0}) \cdot (a_{2m+1}A^{m-1}\vec{x_0} - a_{2m}A^m\vec{x_0})$. Since A is positive, we know Q_2 is non-negative. But then,

$$\begin{aligned}Q_2 &= a_{2m+1}^2 a_{2m-1} - 2a_{2m}^2 a_{2m+1} + a_{2m}^2 a_{2m+1} \\ &= a_{2m+1}(a_{2m+1}a_{2m-1} - a_{2m}^2)\end{aligned}$$

So, consider now $\frac{Q_2}{a_{2m-1}a_{2m}a_{2m+1}}$. This gives us:

$$\tfrac{a_{2m+1}}{a_{2m}} - \tfrac{a_{2m}}{a_{2m-1}} \geq 0,$$

and thus

$$\frac{a_{2m+1}}{a_{2m}} \geq \frac{a_{2m}}{a_{2m-1}} \tag{2}$$

Then by (1) and (2), we see that $\Lambda^{(m)} \leq \Lambda^{(m+1)}$. So we have shown that for a positive definite matrix A, the sequence of Schwartz quotients is monotone increasing.

Case 2: The negative definite symmetric matrix

An analogous result can easily be seen for a negative definite matrix, where the "$\geq$" in (1) and (2) are replaced with "$\leq$" since A is negative. Thus we find that in this case, the sequence of Schwartz quotients is monotone decreasing.

Case 3: The indefinite symmetric matrix

Unfortunately, we have yet to show a similar result for the case of the indefinite matrix. However, we know that the Schwartz quotients still converge to the dominant eigenvalue, and the mapping converges in a similar way as in the first two cases, which we now show for version II.

5. Convergence of the Fixed Point Mapping, Version II

Suppose A is a definite matrix and $\vec{x_0}$ is our initial guess. Note that, using Version II of our fixed point mapping,

$$S_1\vec{x_0} = \tfrac{A}{\Lambda^{(1)}}(\vec{x_0}) = \alpha_1\tfrac{\lambda_1}{\Lambda^{(1)}}\vec{u_1} + \textstyle\sum_{i=2}^{n}(\tfrac{\lambda_i}{\Lambda^{(1)}})\alpha_i\vec{u_i}.$$

If we let this $= \vec{x_1}$, then in the next iteration,

$$\vec{x_2} = S_2\vec{x_1} = \tfrac{A}{\Lambda^{(2)}}(\vec{x_1}) = \alpha_1\tfrac{\lambda_1^2}{\Lambda^{(1)}\Lambda^{(2)}}\vec{u_1} + \textstyle\sum_{i=2}^{n}(\tfrac{\lambda_i^2}{\Lambda^{(1)}\Lambda^{(2)}})\alpha_i\vec{u_i}.$$

Continuing in this manner, we arrive at the m-th iteration,

$$S_m\vec{x}_{m-1} = \alpha_1\tfrac{\lambda_1^m}{\Lambda^{(1)}\Lambda^{(2)}\ldots\Lambda^{(m)}}\vec{u_1} + \textstyle\sum_{i=2}^{n}(\tfrac{\lambda_i^m}{\Lambda^{(1)}\Lambda^{(2)}\ldots\Lambda^{(m)}})\alpha_i\vec{u_i}$$

Now,

$$\|\textstyle\sum_{i=2}^{n}(\tfrac{\lambda_i^m}{\Lambda^{(1)}\Lambda^{(2)}\ldots\Lambda^{(m)}})\alpha_i\vec{u_i}\|^2$$

$$= \textstyle\sum_{i=2}^{n}(\tfrac{\lambda_i^{2m}}{(\Lambda^{(1)}\Lambda^{(2)}\ldots\Lambda^{(m)})^2})\alpha_i^2$$

$$= \frac{1}{(\Lambda^{(1)}\Lambda^{(2)}...\Lambda^{(m)})^2} \sum_{i=2}^{n} \lambda_i^{2m}\alpha_i^2$$

$$\leq \frac{\lambda_2^{2m}}{(\Lambda^{(1)}\Lambda^{(2)}...\Lambda^{(m)})^2}(1-\alpha_1^2)$$

Now, let $0 < k < 1$ be such that $k|\lambda_1| > |\lambda_2|$. Because $\Lambda^{(m)}$ converges to λ_1, there exists some index, M, such that for all $m \geq M$,

$$|\Lambda^{(m)}| > k|\lambda_1| > |\lambda_2|.$$

Then for $m > M$,

$$\frac{\lambda_2^{2m}}{(\Lambda^{(1)}\Lambda^{(2)}...\Lambda^{(m)})^2} = (\frac{\lambda_2^{2M}}{(\Lambda^{(1)}\Lambda^{(2)}...\Lambda^{(M)})^2})(\frac{\lambda_2^{2m-2M}}{(\Lambda^{(M+1)}...\Lambda^{(m)})^2})$$

$$\leq (\frac{\lambda_2^{2M}}{(\Lambda^{(1)}\Lambda^{(2)}...\Lambda^{(M)})^2})(\frac{\lambda_2}{k\lambda_1})^{2m-2M} = (\frac{\lambda_2^{2M}}{(\Lambda^{(1)}\Lambda^{(2)}...\Lambda^{(M)})^2})(\frac{k\lambda_1}{\lambda_2})^{2M}(\frac{\lambda_2}{k\lambda_1})^{2m}$$

$$= O(\frac{\lambda_2}{k\lambda_1})^{2m} \text{ as } m \to \infty.$$

Since $(\frac{\lambda_2^{2M}}{(\Lambda^{(1)}\Lambda^{(2)}...\Lambda^{(M)})^2})(\frac{k\lambda_1}{\lambda_2})^{2M}$ is independent of $m > M$.

So we see that $S_m\vec{x}_{m-1}$ approaches a multiple of $\vec{u_1}$, as does the usual power method.

6. Applying Our Contraction to the Shifted Inverse Power Method

The Shifted Inverse Power Method

The Shifted Inverse Power Method is an alteration to the Power Method which finds the eigenvalue of a matrix closest to a chosen value rather than finding the dominant eigenvalue. The basic idea is to choose a value, q, and rather than using the original matrix, A, find $(A-qI)^{-1}$ and apply the Power Method to this shifted inverse matrix.

The matrix $(A-qI)^{-1}$ has eigenvalues $\frac{1}{\lambda_1-q}, \frac{1}{\lambda_2-q}, ..., \frac{1}{\lambda_n-q}$ and corresponding eigenvectors $\vec{u_1}, ..., \vec{u_n}$, where the λ_i's are again the eigenvalues of A, $|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq |\lambda_4| \geq ... \geq |\lambda_n| > 0$, and the $\vec{u_i}$'s are the corresponding orthonormal eigenvectors to the λ_i's in A.

As in the original case, the Power Method will approximate the dominant eigenvalue of $(A - qI)^{-1}$, which is $\frac{1}{\lambda_j - q}$, where λ_j is the closest eigenvalue of A to q.

Our Contraction

The mapping used in this case is completely analogous to in the case of the original Power Method.

We define T_{inv} by: $T_{inv}\vec{x} = (\lambda_j - q)(A - qI)^{-1}\vec{x}$, where again λ_j is the closest eigenvalue of A to q. For any $\vec{x}$, we again have that $\vec{x} = \sum_{i=1}^{n}(\alpha_i \vec{u}_i)$ since the eigenvectors span $\mathbf{R}^n$. Then we have that:

$$T_{inv}\vec{x} = (\lambda_j - q)\sum_{i=1}^{n}(\tfrac{1}{\lambda_i - q})(\alpha_i \vec{u}_i)$$
$$= (\tfrac{\lambda_j - q}{\lambda_j - q})\alpha_j \vec{u_j} + \sum_{i \neq j}(\tfrac{\lambda_j - q}{\lambda_i - q})(\alpha_i \vec{u}_i)$$
$$= \alpha_j \vec{u_j} + \sum_{i \neq j}(\tfrac{\lambda_j - q}{\lambda_i - q})(\alpha_i \vec{u}_i)$$

Taking the m-th iteration,

$$T_{inv}^m \vec{x} = \alpha_j \vec{u_j} + \sum_{i \neq j}(\tfrac{\lambda_j - q}{\lambda_i - q})^m(\alpha_i \vec{u}_i)$$

Since $|\lambda_j - q| < |\lambda_i - q|$ for all $i \neq j$, we have that $T_{inv}^m \to \alpha_j \vec{u_j}$ as $m \to \infty$.

Notice that if we let $(A - qI)^{-1} = B$, then $T_{inv} = \frac{B}{\gamma_1}$, where γ_1 is the dominant eigenvalue of B. Then T_{inv} is equivalent to the original mapping, T, for the matrix B.

Clearly, then, we can take the Schwartz quotients with respect to $B = (A - qI)^{-1}$, approximating the eigenvalues, $\frac{1}{\lambda_i - q}$. Call the m-th Schwartz quotient $\Lambda_B^{(m)}$. Now, with the m-th iteration, we make $T_{inv} = \frac{B}{\Lambda_B^{(m)}}$. This is completely analogous to the mapping for the original power method, and thus has convergence $O(\frac{\frac{1}{\lambda_k - q}}{\frac{1}{\lambda_j - q}})^m = O(\frac{\lambda_j - q}{\lambda_k - q})^m$ where λ_j is the closest eigenvalue of A to q and λ_k is the next closest.

References:

1. L. Collatz, Eigenwertprobleme und ihre Numerische Behandlung, Chelsea, New York, NY, 1948.

2. N. J. Higham, Review of Spectra and Pseudospectra: The Behavior of Nonnomral Matrices and Operators, by L. N. Trefethen and M. Embree, Princeton University Press, Princeton, NJ, 2005; in Bull AMS (44), 277-284, 2007.

3. A. N. Langville and C.D. Meyer, Google's Page Rank and Beyond, Princeton University Press, Princeton, NJ, 2006.

4. B.N. Parlett, The Symmetric Eigenvalue Problem, Prentice Hall, Englewood Cliffs, NJ, 1980.

5. I. Stakgold, Green's Functions and Boundary Value Problems, Wiley-Interscience, New York, NY, 1979.

6. G.W. Stewart, Introduction to Matrix Computations, Academic Press, New York, NY, 1973.