EE 381V: Large Scale Learning **Spring 2013**

Lecture 12 — February 26 and 28

Lecturer: Caramanis & Sanghavi *Scribe: Karthikeyan Shanmugam and Natalia Arzeno*

12.1 Introduction

In this lecture, we focus on algorithms that compute the eigenvalues and eigenvectors of a real symmetric matrix. Particularly, we are interested in finding the largest and smallest eigenvalues and the corresponding eigenvectors. We study two methods: *Power method* and the *Lanczos iteration*. The first involves multiplying the symmetric matrix by a randomly chosen vector, and iteratively normalizing and multiplying the matrix by the normalized vector from the previous step. The convergence is geometric, i.e. the ℓ_1 distance between the true and the computed largest eigenvalue at the end of every step falls geometrically in the number of iterations and the rate depends on the ratio between the second largest and the largest eigenvalue. Some generalizations of the power method to compute the largest k eigenvalues and the eigenvectors will be discussed.

The second method (Lanczos iteration) terminates in n iterations where each iteration involves estimating the largest (smallest) eigenvalue by maximizing (minimizing) the Rayleigh coefficient over vectors drawn from a suitable subspace. At each iteration, the dimension of the subspace involved in the optimization increases by 1. The sequence of subspaces used are *Krylov* subspaces associated with a random initial vector. We study the relation between Krylov subspaces and tri-diagonalization of a real symmetric matrix. Using this connection, we show that an estimate of the extreme eigenvalues can be computed at each iteration which involves eigen-decomposition of a tri-diagonal matrix. At the end of the $n-$th iteration, the estimate of the eigenvalues is shown to match the true eigenvalues. Further, we study bounds on the difference between the largest (smallest) eigenvalue and the estimate computed at the end of each step and show that the estimates in each step are better compared to the power method for each step.

Notations: Let $\mathbb{S}^n$ denote the space of real symmetric matrices in $\mathbb{R}^n$. In equations and expressions involving matrices, smaller case alphabet (e.g. a) corresponds to a column vector and larger case alphabet (e.g. A) corresponds to a real matrix. $[q_1\ q_2 \dots q_n]$ denotes the matrix with columns q_i. e_i denotes the canonical unit basis vector in $\mathbb{R}^n$. Let A^T denote the transpose of the matrix A. $\|.\|_2$ denotes the ℓ_2 norm of a vector.

For the sections on the power method, we use the following specific notation. Let λ_i denote the $i-$th largest eigenvalue of a symmetric matrix A. Let, u_i denote the normalized eigenvector associated with the $i-$th eigenvalue of A. Let $\Lambda = \text{diag}(\lambda_1, \lambda_2 \dots \lambda_n)$ be the diagonal matrix with diagonal entries being the ordered eigenvalues of A. Let $U = [u_1\ u_2\ \dots u_n]$. Then, by the spectral theorem, we have $A = U\Lambda U^T$ and U is orthonormal.

12.2 Power Method

Given $A \in \mathbb{S}^n$, find the largest (and the smallest) eigenvalues and their associated eigenvectors. The power method algorithm is given in Alg. 1:

Algorithm 1 Power Method

Initialize: $q_0 =$ a random vector
for $k = 1, 2 \ldots$ **do**
  $z_k = Aq_{k-1}$
  $q_k = \frac{z_k}{\|z_k\|_2}$
  $\hat{\lambda}(k) = q_k^T A q_k$
end for

$\hat{\lambda}(k)$ denotes the estimate of the largest eigenvalue (i.e. λ_1) at the end of step k. q_k is the estimate of the eigenvector (i.e. u_1) associated with the largest eigenvalue. We will review a simple example where A is diagonal to intuitively understand why the recursive procedure should work.

12.3 Application of the power method: an example

Consider the following example.

$$A = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \; q_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \; u_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \; \lambda_1 = 3$$

Here, λ_1 and u_1 are the largest eigenvalue and corresponding eigenvector respectively.

In this example, after the $k-$th iteration, the estimate of u_1 is given by:

$$\begin{aligned} q_k &= \frac{A^k q_0}{\|A^k q_0\|_2} \\ &= \frac{\begin{bmatrix} 3^k & 0 & 0 \\ 0 & 2^k & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}}{\sqrt{3^{2k} + 2^{2k} + 1}} \\ &= \frac{\begin{bmatrix} 3^k \\ 2^k \\ 1 \end{bmatrix}}{\sqrt{3^{2k} + 2^{2k} + 1}} \\ &= \frac{\begin{bmatrix} 1 \\ (2/3)^k \\ (1/3)^k \end{bmatrix}}{\sqrt{1 + (2/3)^{2k} + (1/3)^{2k}}} \xrightarrow{k \text{ tends to } \infty} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \end{aligned}$$

The convergence to u_1 is geometric i.e. $q_k \to u_1$ at the rate $(2/3)^k$. Hence, we can expect to see a geometric convergence at the rate $\left(\frac{\lambda_2}{\lambda_1}\right)^k$. We formalize this notion in the next theorem for any general A and a suitable initial condition q_0.

Theorem 12.1. *Let $A \in \mathbb{S}^n$. Let the eigenvalues be $\lambda_1 > \lambda_2 \geq \lambda_3 \ldots \geq \lambda_n$. Let U be the unitary matrix associated with the orthonormal eigenvectors of A. Let θ_k be the angle between q_k and u_1 for all $k \in \{0, 1, 2 \ldots\}$. If $\cos(\theta_0) \neq 0$, then we have:*

1. $|\sin(\theta_k)| \leq |\tan(\theta_0)| \left(\frac{\lambda_2}{\lambda_1}\right)^k$.
2. $|\hat{\lambda}(k) - \lambda_1| \leq |\lambda_1 - \lambda_n| (\tan(\theta_0))^2 \left(\frac{\lambda_2}{\lambda_1}\right)^{2k}$.

Proof (Part 1): From previous discussions,

$$q_k = \frac{A^k q_0}{\|A^k q_0\|_2} \tag{12.1}$$

Since, columns of U form an orthonormal basis for $\mathbb{R}^n$, let $q_0 = a_1 u_1 + a_2 u_2 \ldots a_n u_n$ such that $\sum a_i^2 = 1$. Since, $\cos(\theta_0) \neq 0$, $a_1 \neq 0$. Since, u_i are orthonormal set of eigenvectors of A,

$$A^k q_0 = \sum a_i \lambda_i^k u_i \tag{12.2}$$

Now, we have:

$$\begin{aligned}
(\sin(\theta_k))^2 &= 1 - (\cos(\theta_k))^2 \\
&= 1 - \left(q_k^T u_1\right)^2 \quad \text{(by definition of } \theta_k \text{)} \\
&= 1 - \left(\frac{u_1^T A^k q_0}{\|A^k q_0\|_2}\right)^2 \quad \text{(from eqn. 12.10)} && (12.3) \\
&= 1 - \frac{a_1^2 \lambda^{2k}}{\sum_{i=1}^{n} a_i^2 \lambda^{2k}} \quad \text{(from orthonormality of } U \text{ and eqn. 12.2)} && (12.4) \\
&= \frac{\sum_{i=2}^{n} a_i^2 \lambda_i^{2k}}{\sum_{i=1}^{n} a_i^2 \lambda_i^{2k}} && (12.5) \\
&\leq \frac{\sum_{i=2}^{n} a_i^2 \lambda_i^{2k}}{a_1^2 \lambda_1^{2k}} && (12.6) \\
&\leq \frac{\sum_{i=2}^{n} a_i^2 \lambda_2^{2k}}{a_1^2 \lambda_1^{2k}} \quad (\lambda_2 \geq \lambda_i, \forall i \geq 2) && (12.7) \\
&= \frac{1 - a_1^2}{a_1^2} \left|\frac{\lambda_2}{\lambda_1}\right|^{2k} \quad (\text{since } \sum a_i^2 = 1) && (12.8) \\
&= (\tan(\theta_0))^2 \left|\frac{\lambda_2}{\lambda_1}\right|^{2k} \quad (\text{since } a_1 = \cos(\theta_0)) && (12.9)
\end{aligned}$$

Taking square root on both sides proves part 1. □

Proof (Part 2): From the results in the previous part,

$$\hat{\lambda}(k) = q_k^T A q_k = \frac{q_0^T A^{2k+1} q_0}{q_0^T A^{2k} q_0} = \frac{\sum a_i^2 \lambda_i^{2k+1}}{\sum a_i^2 \lambda_i^{2k}} \tag{12.10}$$

Therefore,

$$|\hat{\lambda}(k) - \lambda_1| = \left| \frac{\sum_{i=2}^{n} a_i^2 \lambda_i^{2k} (\lambda_i - \lambda_1)}{\sum_{i=1}^{n} a_i^2 \lambda_i^{2k}} \right| \tag{12.11}$$

$$\leq |\lambda_1 - \lambda_n| \left| \frac{\sum_{i=2}^{n} a_i^2 \lambda_i^{2k}}{\sum_{i=1}^{n} a_i^2 \lambda_i^{2k}} \right| \tag{12.12}$$

$$\leq |\lambda_1 - \lambda_n| \left| \frac{\sum_{i=2}^{n} a_i^2 \lambda_i^{2k}}{a_1^2 \lambda_1^{2k}} \right| \tag{12.13}$$

$$\leq |\lambda_1 - \lambda_n| \frac{1 - a_1^2}{a_1^2} \left(\frac{\lambda_2}{\lambda_1} \right)^{2k} \tag{12.14}$$

$$\leq |\lambda_1 - \lambda_n| \left(\tan(\theta_0) \right)^2 \left(\frac{\lambda_2}{\lambda_1} \right)^{2k} \tag{12.15}$$

This proves part 2. □

From the above proof, it is clear that the power method is ineffective if q_0 is orthogonal to u_1. But if q_0 is chosen randomly, then this happens with probability 0. Hence, choosing a random vector is a good initialization.

12.4 Variations of the Power Method

Suppose, we have identified some $\bar{\lambda}$ close to λ_i for some i, then $\left(A - \bar{\lambda} I\right)^{-1}$ has a very large eigenvalue associated with u_i since $A - \bar{\lambda} I$ is very close to being rank deficient. Hence, one can do power iteration on $\left(A - \bar{\lambda} I\right)^{-1}$.

If u is an eigenvector of the symmetric matrix A, then $Au = \lambda u$ for some λ. In other words, $\min_{\lambda} \|(A - \lambda I)\, u\|_2 = 0$ if u is an eigenvector. When u is not an eigenvector, then analogously, we can define an 'approximate eigenvalue' by:

$$r(u) = \arg\min_{\lambda} \|(A - \lambda I) u\|_2 \tag{12.16}$$

$r(u)$ is called the *Rayleigh quotient* associated with the vector u. We have the following result regarding the Rayleigh quotient.

Theorem 12.2. $r(u) = \arg\min_{\lambda} \|(A - \lambda I) u\| = \frac{u^T A u}{u^T u}$.

Proof:

$$\|(A-\lambda I)\,u\|^2 = \|\left(A-\frac{u^TAu}{u^Tu}I+\frac{u^TAu}{u^Tu}I-\lambda I\right)u\|^2 \tag{12.17}$$

$$= \|\left(A-\frac{u^TAu}{u^Tu}I\right)u\|^2+\|\left(\frac{u^TAu}{u^Tu}I-\lambda I\right)u\|^2+$$
$$2u^T\left(A^T-\frac{u^TAu}{u^Tu}I\right)u\left(\frac{u^TAu}{u^Tu}-\lambda\right) \tag{12.18}$$

$$= \|\left(A-\frac{u^TAu}{u^Tu}I\right)u\|^2+\|\left(\frac{u^TAu}{u^Tu}I-\lambda I\right)u\|^2+$$
$$2u^T\left(A-\frac{u^TAu}{u^Tu}I\right)u\left(\frac{u^TAu}{u^Tu}-\lambda\right) \quad \text{(since } A \text{ is symmetric)} \tag{12.19}$$

$$= \|\left(A-\frac{u^TAu}{u^Tu}I\right)u\|^2+\|\left(\frac{u^TAu}{u^Tu}I-\lambda I\right)u\|^2 \quad \left(\text{since } u^T\left(A-\frac{u^TAu}{u^Tu}I\right)u=0\right) \tag{12.20}$$

Therefore, minimizing the left hand side with respect to λ is equivalent to minimizing the second term on the right with respect to λ since the first term does not depend on λ. Clearly, the second term on the right hand side is minimized when $\lambda = \frac{u^TAu}{u^Tu}$. □

Combining the idea of inverse iteration and the Rayleigh quotient, one can modify the power method iteration. One such modification of power method is given in Fig. 2.

Algorithm 2 Power Method Modified using Inverse Iteration

Initialize: q_0 = a random vector
for $k = 1, 2 \ldots$ **do**
  $\mu_{k-1} = r\,(q_{k-1})$
  Solve: $(A-\mu_{k-1}I)\,z_k = q^{k-1}$
  $q_k = \frac{z_k}{\|z_k\|_2}$
  $\hat{\lambda}(k) = q_k^TAq_k$
end for

In this modified algorithm, $r(.)$ is the Rayleigh quotient. Effectively, every iteration involves a power method like iteration with $(A-\mu_{k-1}I)^{-1}$.

12.5 Generalization of the Power Method

If we are interested in the top m ($m \leq n$) eigenvalues and the corresponding eigenvectors, we can maintain m orthonormal column vectors in Q_k instead of a unit vector q_k in each iteration. Instead of normalizing z_k, we need to orthonormalize $Z_k = AQ_{k-1}$ by using Gram-Schmidt orthonormalization yielding a QR decomposition. Hence, $Z_k = Q_kR_k$ where Q_k is

an orthonormal $n \times m$ matrix and R_k is upper triangular. The next iteration continues with Q_k. It is easy to check that when $m = 1$, it reduces to the algorithm given in Fig. 1. The Generalized power method is given in Alg. 3.

Algorithm 3 Generalized Power Method

Initialize: $Q_0 =$ **a random matrix of size** $n \times m$ **with orthonormal columns**
for $k = 1, 2 \ldots$ **do**
  $Z_k = AQ_{k-1}$
  QR Decomposition: $Z_k = Q_k R_k$
end for

Q_k will be the estimate of $[u_1 \; u_2 \ldots u_m]$ after iteration m.

12.6 Introduction to Lanczos Iteration

Given $A \in \mathbb{S}^n$, we would like to find out the eigenvectors and eigenvalues. More specifically, we are interested in finding good estimates of the top and the bottom eigenvalues very quickly. Henceforth, we will study a method named Lanczos Iteration which at the end of any step gives a tri-diagonal matrix whose extreme eigenvalues approximate the extreme eigenvalues of the A. Under suitable initial conditions, the tridiagnoal matrix at the n^{th} step has the exact same eigenvalues and the eigenvectors of the original matrix. Being tri-diagonal, the eigendecomposition is easily carried out for that matrix using other methods (possibly power method).

To understand Lanczos iteration, we will review some preliminary concepts that would be required in the next section.

12.7 Preliminaries

12.7.1 Rayleigh quotient

From Sec. 12.4, we know that given any vector u, the approximate eigenvalue associated with u is given by the Rayleigh quotient $r(u) = \frac{u^T A u}{u^T u}$.

12.7.2 Krylov Subspaces

Given a vector $q \in \mathbb{R}^n$ and $A \in \mathbb{S}^n$, the Krylov subspace $\mathcal{K}(A, q, k)$ is given by:

$$\mathcal{K}(A, q, k) = \text{span}\left\{q, Aq, A^2 q, \ldots A^{k-1} q\right\} \tag{12.21}$$

and the Krylov matrix $K(A, q, k)$ is given by:

$$K(A, q, k) = \left[q \; Aq \; A^2 q \; \cdots A^{k-1} q\right] \tag{12.22}$$

12.7.3 Tri-Diagonalization

If $A \in \mathbb{S}^n$, there exists an orthogonal Q such that $Q^T AQ = T$ where T is a tri-diagonal matrix. Clearly, by the spectral theorem,there is an orthonormal U (whose columns are eigenvalues) that diagonalizes A. Since a diagonal matrix is also tri-diagonal, the statement above is true. Often, tri-diagonalization is an intermediate step to diagonalization. We will see that Lanczos iteration seeks to tri-diagonalize the matrix A at the end of n steps.

12.7.4 Connection between Tri-Diagonalization and Krylov Subspaces

Suppose the columns of $Q_k = [q_1\ q_2\ \cdots\ q_k]$ form an orthonormal basis of $\mathcal{K}(A, q, k)$, then we have the following property:

Theorem 12.3. *$Q_n^T AQ_n = T$ is tri-diagonal.*

Proof: Let us consider T_{ij} for $i > j + 1$. $T_{ij} = q_i^T Aq_j$. $Aq_j \in \text{span}\{q, Aq, A^2q \dots A^jq\} = \text{span}\{q_1, q_2 \dots q_j, q_{j+1}\}$. Since $\{q_i\}_{i=1}^n$ form an orthonormal set, q_i is orthogonal to span $\{q_1 \dots q_j, q_{j+1}\}$ for $i > j + 1$. This means $q_i^T Aq_j = 0$. Since, $Q_n^T AQ_n$ is symmetric, T is symmetric. Hence, $T_{ij} = 0$ for $i < j - 1$. Therefore, T is tri-diagonal. □

12.8 Key Ideas behind Lanczos Iteration

We focus on what the method seeks to do at every iteration. Let us consider the goal of finding/approximating the largest and the smallest eigenvalue of $A \in \mathbb{S}^n$. Let λ_i be the i^{th} largest eignevalue of A. By the variational characterization of the eigenvalues, we have the following:

Theorem 12.4. $\lambda_1 = \max_{y \neq 0} \frac{y^T Ay}{y^T y}, \quad \lambda_n = \min_{y \neq 0} \frac{y^T Ay}{y^T y}$.

Proof: By the spectral theorem, the eigenvectors $\{u_i\}_{i=1}^n$ of A form an orthonormal basis. Any $y \neq 0$ can be written as: $y = \sum_{i=1}^n a_i u_i$. Consider $r(y) = \frac{y^T Ay}{y^T y}$. We have,

$$r(y) = \frac{\left(\sum a_i u_i^T\right) A \left(\sum a_i u_i\right)}{\sum a_i^2} \quad \text{(from orthonormality of } \{u_i\}) \tag{12.23}$$

$$= \frac{\left(\sum a_i u_i^T\right) \left(\sum \lambda_i a_i u_i\right)}{\sum a_i^2} \quad (u_i \text{ is eigenvector with eigenvalue } \lambda_i) \tag{12.24}$$

$$= \frac{\sum a_i^2 \lambda_i}{\sum a_i^2} \tag{12.25}$$

Clearly, $r(y)$ is a weighted average of the eigenvalues λ_i with weights being proportional to a_i^2. Hence, $\lambda_n \leq r(y) \leq \lambda_1$. The extreme values, i.e. λ_1 and λ_n, can be attained by setting $a_1 = 1$ with all other a_i set to 0 and $a_n = 1$ with all other a_i set to 0, respectively. □

Let us define two quantities, M_k and m_k, as follows:

$$M_k = \max_{0 \neq y \in \mathcal{V}_k} \frac{y^T A y}{y^T y}, \quad m_k = \min_{0 \neq y \in \mathcal{V}_k} \frac{y^T A y}{y^T y} \tag{12.26}$$

Here, $\mathcal{V}_k$ is a k-dimensional subspace. Clearly, for any choice of increasing sequence of subspaces, i.e. $\mathcal{V}_k \subset \mathcal{V}_{k+1}$, M_k is non-decreasing and $M_n = \lambda_1$. Similarly, m_k is non-increasing and $m_n = \lambda_n$. This is because, for each k, the optimization is performed over a larger domain than the previous. This means, M_k (m_k) is an increasingly better estimate for the largest eigenvalue (smallest eigenvalue) as the step index k increases.

Lanczos iteration runs for n steps and in each step k, it augments the subspace $\mathcal{V}_k$ suitably to ensure the subsequent estimate of the largest eigenvalue is good. We will now focus on the question: What sequence of subspaces are 'good' if we want to get good estimates of the extreme eigenvalues at each step?

Let us re write M_k and m_k in terms of an orthonormal basis, given by the columns of $Q_k = [q_1\ q_2 \cdots q_k]$, of the subspace $\mathcal{V}_k$. Any vector $u \neq 0$ in $\mathcal{V}_k$ can be written as $u = Q_k y$ where $y \in \mathbb{R}^{k \times 1}$. Since $Q_k^T Q_k = I$, we have

$$M_k = \max_{0 \neq y \in \mathbb{R}^{k \times 1}} \frac{y^T Q_k^T A Q_k y}{y^T y} = \max_{0 \neq y} r\left(Q_k y\right), \quad m_k = \min_{0 \neq y \in \mathbb{R}^{k \times 1}} \frac{y^T Q_k^T A Q_k y}{y^T y} = \min_{0 \neq y} r\left(Q_k y\right) \tag{12.27}$$

Let the maximum value M_k be attained at $y = \bar{y}$ and the minimum value m_k be attained at $\underline{y}$. Let $u_k = Q_k \bar{y},\ v_k = Q_k \underline{y}$. Clearly, $u_k, v_k \in \operatorname{span}\{q_1, q_2 \ldots q_k\}$. We would like to add q_{k+1} to the set of orthonormal vectors $\{q_i\}_{i=1}^k$ such that $\{q_i\}_{i=1}^{k+1}$ is also orthonormal and the column space of Q_{k+1} will be the subspace $\mathcal{V}_{k+1}$.

Based on an optimization based idea, intuitively, $\nabla r(u_k), \nabla r(v_k) \in \operatorname{span}\{q_1, q_2 \ldots q_{k+1}\}$ where ∇ is the gradient with respect to the vector argument for $r(\cdot)$. The gradient of $r(x)$ is given by:

$$\nabla r(x) = \frac{2}{x^T x}\left(Ax - r(x)x\right) \tag{12.28}$$

This implies that, $\nabla r(x) \in \operatorname{span}\{x, Ax\}$. Since $u_k, v_k \in \operatorname{span}\{q_1, q_2 \ldots q_k\}$, we need to have $\nabla r(u_k), \nabla r(v_k) \in \operatorname{span}\{q_1, q_2 \ldots q_k, Aq_1, Aq_2 \ldots Aq_k\} = \operatorname{span}\{q_1, q_2, \ldots q_k, q_{k+1}\}$. If we choose Krylov subspaces, i.e. $\mathcal{V}_k = \mathcal{K}\left(A, q_1, k\right) = \operatorname{span}\left\{q_1, Aq_1, \ldots A^{k-1}q_1\right\}$, such that columns of Q_k is an orthonormal basis for $\mathcal{K}\left(A, q_1, k\right)$. Then clearly,

$$\operatorname{span}\{q_1, q_2 \ldots q_k, Aq_1, Aq_2 \ldots Aq_k\} = \operatorname{span}\left\{q_1, Aq_1 \ldots A^{k-1}q_1, A^{k-1}q_1\right\} \quad \text{and}$$

$$\nabla r(u_k), \nabla r(v_k) \in \operatorname{span}\left\{q_1, Aq_1 \ldots A^{k-1}q_1, A^k q_1\right\} = \operatorname{span}\{q_1, q_2 \ldots q_k, q_{k+1}\}.$$

Hence, one needs to orthonormalize the columns of Krylov matrix $K(A, q_1, k)$ to get Q_k at every step.

Also, at the end of n^{th} step, finding the eigenvalues and eigenvectors of A is equivalent to finding the eigenvalues of $Q_n^T A Q_n$. But, this is equal to a tri-diagonal matrix T by

Thm 12.3. Hence, computing eigenvalues and eigenvectors of A reduces, at the end of $n-$ steps, to finding the eigendecomposition for tri-diagonal T. Also, to find Q_k, we need to orthonormalize columns of $K(A, q_1, k)$.

In the subsequent sections, we will see that orthonormalizing $K(A, q_1, k)$ can be done iteratively through recursively trying to tri-diagonalize A using the relation $AQ_n = Q_nT$. Every step of the iteration also yields Q_k. Further, we will show that the extreme eigenvalues M_k and m_k can found by eigendecomposition of a suitable tri-diagonal matrix at every step, and that M_k and m_k form good estimates of the true extreme eigenvalues of A. The theoretical guarantees concerning these properties of Lanczos iteration is the subject of the subsequent sections.

12.9 Review of Motivation for Lanczos Iteration

We want an iterative algorithmic way to find an orthonormal basis for the Krylov subspace $\mathcal{K}(A, q_1, k)$. The key idea that we have discussed is tri-diagonalization.

We have seen that it is always possible and"easy" to tri-diagonalize a symmetric matrix A such that $Q^{\mathrm{T}}AQ = T$, where T is tri-diagonal and $q_1 = Qe_1$ is the first column of Q. Then we have $K(A, q_1, n) = Q[e_1, Te_1, \ldots, T^{n-1}e_1]$ such that T, Q give a QR-factorization of the Krylov matrix $K(A, q_1, n)$, which also gives an orthonormal basis for $\mathcal{K}(A, q_1, n)$. Thus, if there exists an orthogonal Q with $Qe_1 = q_1$ and $Q^{\mathrm{T}}AQ = T$, then we want to iteratively find this Q.

12.10 Tri-diagonalization and the Lanczos Iteration

12.10.1 Householder Reflections

One way of tri-diagonalizing a symmetric matrix using an orthogonal transformation involves using Householder reflections, which achieve tri-diagonalization by performing $n-2$ orthogonal projections. Though this method is efficient and popular, it might yield dense intermediate steps for a sparse A. Since we want to benefit from any special structure of A, including sparsity, we will not use this method.

12.10.2 Lanczos Iteration

Instead, we want a direct algorithm that can exploit the sparsity of A. Restating the problem, for $Q = [q_1 \cdots q_n]$, we want $T = Q^{\mathrm{T}}AQ$, where T is tri-diagonal and symmetric:

$$T = \begin{bmatrix} \alpha_1 & \beta_1 & & & & 0 \\ \beta_1 & \alpha_2 & \beta_2 & & & \\ & \beta_2 & \alpha_3 & \beta_3 & & \\ & & \ddots & \ddots & \ddots & \\ & & & \beta_{n-2} & \alpha_{n-1} & \beta_{n-1} \\ 0 & & & & \beta_{n-1} & \alpha_n \end{bmatrix} . \tag{12.29}$$

Since Q is orthonormal, finding $T = Q^{\mathrm{T}}AQ$ is equivalent to solving $AQ = QT$.
If we examine the k^{th} column of AQ, we have

$$\begin{aligned} Aq_k &= QT_k \\ &= \beta_{k-1}q_{k-1} + \alpha_k q_k + \beta_k q_{k+1} . \end{aligned} \tag{12.30}$$

This provides a recursion to find Q. We first set $\beta_0 q_0 = 0$. Since q_i are orthonormal, we have

$$q_k^{\mathrm{T}} A q_k = \alpha_k \|q_k\|^2 = \alpha_k , \tag{12.31}$$

which gives us the equation for α_k. Now we can solve for $\beta_k q_{k+1}$:

$$\beta_k q_{k+1} = (A - \alpha_k I)q_k - \beta_{k-1}q_{k-1} . \tag{12.32}$$

Setting r_k to be equal to the right hand side of eq. 12.32, if $r_k \neq 0$,

$$q_{k+1} = r_k/\beta_k , \tag{12.33}$$

where $\beta_k = \|r_k\|_2$.

Algorithm 4 Lanczos Iteration

Initialize: $r_0 = q_1$, $\beta_0 = 1$, $q_0 = 0$, $k = 0$
while $\beta_k \neq 0$ **do**
  $q_{k+1} = r_k/\beta_k$
  $k = k + 1$
  $\alpha_k = q_k^{\mathrm{T}} A q_k$
  $r_k = (A - \alpha_k I)q_k - \beta_{k-1}q_{k-1}$
  $\beta_k = ||r_k||_2$
end while

12.10.3 Termination of the Lanczos Iteration

If Algorithm 4 runs until $k = n$, then it will produce $Q^{\mathrm{T}}AQ = T$. However, what if the algorithm terminates at iteration $k < n$? In particular:
1. What if it terminates early because $b_k = 0$?
2. What if **we** terminate it early? Would we still get good estimates of the eigenvalues λ_1, λ_n (compared to the Power Method)?

Early termination of Lanczos iteration due to $\beta_k = 0$

Note that $\beta_k = 0$ is equivalent to $r_k = 0$. This will only happen if q_1 is constrained in a proper invariant subspace of dimension $< n$: $\{q_1, Aq_1, \dots A^k q_1\} \subseteq \{q_1, Aq_1, \dots, A^m q_1\}$ for $m < n$.

Theorem 12.5. *Let $A \in \mathbb{S}^n$, $q_1 \in \mathbb{R}^n$ and $\|q_1\| = 1$. Then, the Lanczos algorithm runs until iteration $m = \text{rank}(K(A, q_1, n))$. Moreover, for $k = 1, \dots, m$, we have $AQ_k = Q_k T_k + r_k e_k^{\mathrm{T}}$,*

where $T_k = \begin{bmatrix} \alpha_1 & \beta_1 & & & 0 \\ \beta_1 & \ddots & \beta_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \beta_{k-2} & \ddots & \beta_{k-1} \\ 0 & & & \beta_{k-1} & \alpha_k \end{bmatrix}$ *and* $Q_k = [q_1 \cdots q_k]$.

By our definitions, m is the rank of the smallest invariant subspace that contains q_1 and Q_k is the span of $\mathcal{K}(A, q_1, k)$.

Proof: We will complete the proof by induction on k. The base case, $k = 1$ is immediate from the algorithm. Now, suppose that at iteration k we have $Q_k = [q_1 \cdots q_k]$ such that $\text{range}(Q_k) = \mathcal{K}(A, q_1, k)$ and Q_k is orthogonal ($Q_k^{\mathrm{T}} Q_k = I_k$). By construction, we have $AQ_k = Q_k T_k + r_k e_k^{\mathrm{T}}$. Then,

$$Q_k^{\mathrm{T}} A Q_k = T_k + Q_k^{\mathrm{T}} r_k e_k^{\mathrm{T}} \tag{12.34}$$

Since A is symmetric, $Q_k^{\mathrm{T}} A Q_k \in \mathbb{S}^n$. Inspecting the elements on the diagonal of $(Q_k^{\mathrm{T}} A Q_k)$:

$$(Q_k^{\mathrm{T}} A Q_k)_{ii} = q_i^{\mathrm{T}} A q_i = \alpha_i \tag{12.35}$$

by the definition of α_i in the algorithm. Similarly following the algorithm updates, we have

$$q_{i+1}^{\mathrm{T}} A q_i = \beta_i \ . \tag{12.36}$$

Equations 12.35 and 12.36 imply

$$Q_k^{\mathrm{T}} A Q_k = T_k \ . \tag{12.37}$$

Equation 12.37 combined with equation 12.34 further implies

$$Q_k^{\mathrm{T}} r_k e_k^{\mathrm{T}} = 0 \ . \tag{12.38}$$

Equation 12.38 will hold for two cases:
1. If $r_k \neq 0$, we update $q_{k+1} = r_k / \|r_k\|_2$. Then $Q_k^{\mathrm{T}} r_k = 0$ requires $q_{k+1} \perp \{q_1, \dots, q_k\}$ **and** $q_{k+1} \in \text{span}\{Aq_k, q_k, q_{k-1}\}$. Therefore, $Q_{k+1}^{\mathrm{T}} Q_{k+1} = I_{k+1}$ **and** $\text{range}(Q_{k+1}) = \mathcal{K}(A, q_1, k+1)$ and our algorithm is working as claimed by the induction.
2. If $r_k = 0$, $r_k e_k^{\mathrm{T}} = 0$ and $AQ_k = Q_k T_k$. Thus, we have $k = m = \text{rank}(K(A, q_1, n))$. □

If the algorithm terminates at $m < n$, then we have found a proper invariant subspace. In this case, we can reduce the size of the problem. Note that this will never happen if we choose q_1 at random and A full-rank.

Now that we have proved that the algorithm will not terminate early, we want to examine if it is necessary to run the algorithm to completion. In particular, we are interested in looking at whether T_k will give us good estimates of the eigenvalues and eigenvalues of A, or if this is only the case for T_n.

Forced early termination of Lanczos iteration

Theorem 12.6. *Suppose we have run the Lanczos iteration for k steps without termination on its own (i.e. have produced T_k). Since $T_k \in \mathbb{S}^n$, we can find an orthonormal matrix S_k that diagonalizes T_k:*

$$S_k^{\mathrm{T}} T_k S_k = \Theta_k = \begin{bmatrix} \theta_1 & & \\ & \ddots & \\ & & \theta_k \end{bmatrix}. \tag{12.39}$$

Let $Y_k = Q_k S_k \in \mathbb{R}^{n \times k}$; $Y_k = [y_1 \cdots y_k]$, where y_i are column vectors.
Then: y_i are "close" to the eigenvectors of A and θ_i are"close" to the eigenvalues of A. Specifically: $\|Ay_i - \theta_i y_i\|_2 = |\beta_k| \cdot |S_{ki}|$, where S_{ki} is the $(k,i)^{th}$ element of S.

Proof: Recall: $AQ_k = Q_k T_k + r_k e_k^{\mathrm{T}}$. Right-multiplying by S_k, we get:

$$AQ_k S_k = Q_k T_k S_k + r_k e_k^{\mathrm{T}} S_k. \tag{12.40}$$

Using the definition $Y_k = Q_k S_k$ and the fact that S_k is orthonormal:

$$\begin{aligned} AY_k &= Q_k S_k S_k^{\mathrm{T}} T_k S_k + r_k e_k^{\mathrm{T}} S_k \\ &= Y_k \Theta_k + r_k e_k^{\mathrm{T}} S_k \end{aligned} \tag{12.41}$$

Looking at the i^{th} column:

$$Ay_i = \theta_i y_i + r_k e_k^{\mathrm{T}} S_k e_i. \tag{12.42}$$

Equation 12.42 implies $\|Ay_i - \theta_i y_i\|_2 = |\beta_k| \cdot |S_{ki}|$ since $|\beta_k| = \|r_k\|_2$.

□

12.10.4 Bounding the approximate eigenvalues after early termination of Lanczos

In the above section we showed forced early termination of Lanczos results in "close" eigenvalues and eigenvectors. However, we want to explore how "close" we can get to the true values.

Kaniel-Paige Convergence Theory

Theorem 12.7. *Let $A \in \mathbb{S}^n$. Let $\lambda_1 \geq \cdots \geq \lambda_n$ and $z_1, \ldots, z_n$ respectively be the eigenvalues and eigenvectors of A. Let T_k be the $k \times k$ tri-diagonal matrix obtained after k steps of Lanczos with eigenvalues $\theta_1, \ldots, \theta_n$.*
Then:

$$\lambda_1 \geq \theta_1 \geq \lambda_1 - \frac{(\lambda_1 - \lambda_n)\left(\tan(\phi_1)\right)^2}{\left(c_{k-1}(1+2\rho_1)\right)^2},$$

where $\cos(\phi_1) = |< q_1, z_1 >|$, $\rho_1 = (\lambda_1 - \lambda_2)/(\lambda_2 - \lambda_n)$, and $c_{k-1}(x)$ is the Chebyshev polynomial of degree $k-1$.

Proof: The upper bound is immediate from the definitions of λ_1 and θ_1. By definition of the largest eigenvalue:

$$\begin{aligned}
\lambda_1 &= \max_{||w||_2=1} w^{\mathrm{T}} A w \\
&= \max_{w \neq 0} \frac{w^{\mathrm{T}} A w}{w^{\mathrm{T}} w}, \quad \text{and} \qquad (12.43) \\
\theta_1 &= \max_{y \neq 0} \frac{y^{\mathrm{T}} T_k y}{y^{\mathrm{T}} y} \\
&= \max_{y \neq 0} \frac{y^{\mathrm{T}} Q_k^{\mathrm{T}} A Q_k y}{y^{\mathrm{T}} y} \quad (\text{since } T_k = Q_k^{\mathrm{T}} A Q_k) \\
&= \max_{0 \neq w \in \mathcal{K}(A, q_1, k)} \frac{w^{\mathrm{T}} A w}{w^{\mathrm{T}} w}. \qquad (12.44)
\end{aligned}$$

Thus, θ_1 is the maximum of a more constrained set than that in the calculation λ_1 and $\theta_1 \leq \lambda_1$.

To prove the lower bound, we will use the definition of θ_1 from eq. 12.44:

$$\theta_1 = \max_{0 \neq w \in \mathcal{K}(A, q_1, k)} \frac{w^{\mathrm{T}} A w}{w^{\mathrm{T}} w},$$

where the Krylov subspace $\mathcal{K}(A, q_1, k) = \text{span}\{q_1, Aq_1, \ldots, A^{k-1}q_1\}$.The Krylov subspace can be rewritten as

$$\mathcal{K}(A, q_1, k) = \{p(A)q_1 : p \text{ is a degree } k-1 \text{ polynomial}\}. \qquad (12.45)$$

Note that since $A \in \mathbb{S}^n$, we will also have $p(A) \in \mathbb{S}^n$. Using this new definition of $\mathcal{K}(A, q_1, k)$, we can rewrite θ_1:

$$\theta_1 = \max_{p \in \mathcal{P}_{k-1}} \frac{q_1^{\mathrm{T}} p(A) A p(A) q_1}{q_1^{\mathrm{T}} p(A)^2 q_1}, \qquad (12.46)$$

where $\mathcal{P}_{k-1}$ is the set of degree $k-1$ polynomials.

Since we are assuming A is full rank, we can rewrite q_1 as a weighted sum of the eigenvectors of A: $z_1, \ldots, z_n$ as

$$q_1 = \sum_{i=1}^{n} d_i z_i.$$

Since $\|q_1\|_2 = 1$ and the z_i are an orthonormal basis, we have $\sum_{i=1}^{n} d_i^2 = 1$. Using this decomposition of q_1, we can find a lower bound on θ_1 by:

$$\begin{aligned}
\frac{q_1^{\mathrm{T}} p(A) A p(A) q_1}{q_1^{\mathrm{T}} p(A)^2 q_1} &= \frac{\sum_{i=1}^{n} d_i^2 p(\lambda_i)^2 \lambda_i}{\sum_{i=1}^{n} d_i^2 p(\lambda_i)^2} \\
&\geq \frac{\lambda_1 d_1^2 p(\lambda_1)^2 + \lambda_n \sum_{i=2}^{n} d_i^2 p(\lambda_i)}{d_1^2 p(\lambda_1)^2 + \sum_{i=2}^{n} d_i^2 p(\lambda_i)^2} && (12.47) \\
&= \frac{\lambda_1 \left(d_1^2 p(\lambda_1)^2 + \sum_{i=2}^{n} d_i^2 p(\lambda_i)^2 \right) - \lambda_1 \sum_{i=2}^{n} d_i^2 p(\lambda_i)^2 + \lambda_n \sum_{i=2}^{n} d_i^2 p(\lambda_i)}{d_1^2 p(\lambda_1)^2 + \sum_{i=2}^{n} d_i^2 p(\lambda_i)^2} && (12.48) \\
& && (12.49) \\
&= \lambda_1 - (\lambda_1 - \lambda_n) \frac{\sum_{i=2}^{n} d_i^2 p(\lambda_i)}{d_1^2 p(\lambda_1)^2 + \sum_{i=2}^{n} d_i^2 p(\lambda_i)^2} && (12.50)
\end{aligned}$$

In order to make the bound as tight as possible, we want to design a polynomial to maximize the right-hand side of Equation 12.50. This can be achieved by using Chebyshev polynomials, which are designed to be bounded by 1 in the interval [-1,1] and grow rapidly outside of this interval. Recall that the Chebyshev polynomial is constructed as:

$$\begin{aligned}
c_0 &= 1 \\
c_1(x) &= x \\
&\vdots \\
c_k(x) &= 2x c_{k-1}(x) - c_{k-2}(x)
\end{aligned}$$

Example plots of several Chebyshev polynomials can be seen in Figure 12.1.

In order to exploit the characteristics of the Chebyshev polynomial, we need to translate and rescale so that the value is bounded by 1 not for $x \in [-1, 1]$ but for $x \in (-\lambda_n, \lambda_2)$. Since λ_1 is outside of this range, the function value will be high at λ_1. We can achieve this by defining $p(\lambda)$ as:

$$p(\lambda) = c_{k-1}(1 + 2(\lambda - \lambda_2)/(\lambda_2 - \lambda_n)) \tag{12.51}$$

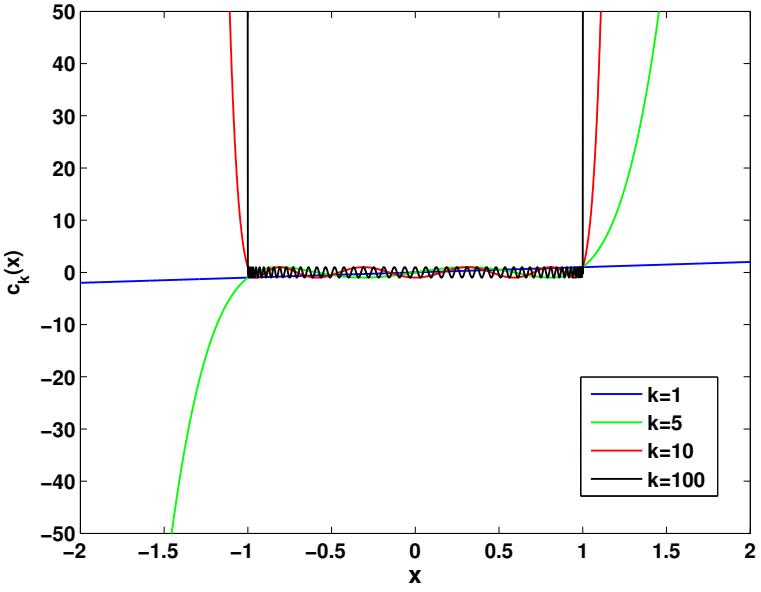

Figure 12.1. Chebyshev polynomials of varying degrees.

We can see that $|p(\lambda_i)| \leq 1$ for $i = 2, \ldots, n$ since due to the decreasing order of the eigenvalues, $(\lambda_i - \lambda_2)/(\lambda_2 - \lambda_n)$ will be a number between 0 and -1. At λ_1 we will have

$$p(\lambda_1) = c_{k-1}(1 + 2(\lambda_1 - \lambda_2)/(\lambda_2 - \lambda_n)) \; . \tag{12.52}$$

Putting together the bound from Equation 12.50 with the bounds on $c_{k-1}(\lambda_i)$, we have:

$$\begin{aligned}
\theta_1 &\geq \lambda_1 - (\lambda_1 - \lambda_n)\frac{\sum_{i=2}^{n} d_i^2 p(\lambda_i)}{d_1^2 p(\lambda_1)^2 + \sum_{i=2}^{n} d_i^2 p(\lambda_i)^2} \\
&\geq \lambda_1 - (\lambda_1 - \lambda_n)\frac{\sum_{i=2}^{n} d_i^2}{d_1^2 p(\lambda_1)^2} && (12.53) \\
&= \lambda_1 - (\lambda_1 - \lambda_n)\frac{1 - d_1^2}{d_1^2 p(\lambda_1)^2} \quad (\text{since } \sum d_i^2 = 1) && (12.54) \\
&= \lambda_1 - (\lambda_1 - \lambda_n) \cdot \frac{1 - d_1^2}{d_1^2} \cdot \frac{1}{c_{k-1}(1 + 2\rho_1)} \quad (\text{by eq. 12.52 and def. of } \rho_1) && (12.55) \\
&= \lambda_1 - (\lambda_1 - \lambda_n) \cdot \frac{(\tan(\phi_1))^2}{c_{k-1}(1 + 2\rho_1)} \quad (\text{since } (\tan(\phi_1))^2 = (1 - d_1^2)/d_1^2) && (12.56)
\end{aligned}$$

□

We can bound the lowest eigenvalue of T_k in an analogous way.

Theorem 12.8.

$$\lambda_n \leq \theta_n \leq \lambda_n + \frac{(\lambda_1 - \lambda_n)\tan((\phi_n))^2}{c_{k-1}(1+2\rho_n)^2} ,$$

where $\rho_n = \frac{\lambda_{n-1}-\lambda_n}{\lambda_1-\lambda_n}$ *and* $\cos\phi_n =< q_n, z_n >$.

Proof: The proof follows analogously to that of Theorem 12.7, but replacing A by $-A$. □

12.10.5 Lanczos vs Power Method

In the homework, we will implement Chebyshev polynomials in Matlab and compare the bounds from the Lanczos iteration to those of the power method. We will find that the bound achieved by Lanczos is much better. We can gain some intuition as to why Lanczos is better by examining the eigenvalue the approximations. Lanczos approximates λ_1 by

$$M_k = \max_{0\neq w\in(A,q_1,k)} \frac{w^{\mathrm{T}}Aw}{w^{\mathrm{T}}w} ,$$

whereas the power method approximates λ_1 by

$$\hat{M}_k = \max_{0\neq w\in \mathrm{span}\{A^{k-1}q_1\}} \frac{w^{\mathrm{T}}Aw}{w^{\mathrm{T}}w} .$$

However, it should be noted that the max in $\hat{M}_k$ is not a true max since it does not search for w in the entire Krylov space. Thus it is clear that $\hat{M}_k \leq M_k$.

Another way to compare the methods is through the proof of the bounds in the Lanczos algorithm. If we used $p(x) = x^{k-1}$ instead of cleverly designing our polynomial by using the Chebyshev polynomial, we would get the power method bound instead of the Lanczos bound.