Computer Sciene 2: Advanced Algorithms Fall 2026

Lecture 3 — September 3rd, 2026

Prof. Jarosław Błasiok *Scribe: David Kühnemann*

1 Overview

In the last lecture we saw the Fast Fourier Transform for integer multiplication and more generally convolution.

In this lecture we cover *power iteration* to approximate the largest eigenvector or eigenvalue of a matrix. We will see that the effectiveness of this method depends on the *spectral gap*, that is the relative difference between the largest and second-largest eigenvalues, as well as the choice of starting vector.

2 Preliminaries

Let $A \in \mathbb{R}^{n \times n}$. Recall that $u \in \mathbb{R}^n$ is an *eigenvector of* A with *eigenvalue* $\lambda \in \mathbb{R}$ if $Au = \lambda \cdot u$. We say A is symmetric if $A^T = A$. We say A is *orthogonal* if $A^T A = AA^T = I$, meaning the columns $u_1, \ldots, u_n$ (and rows) of A are orthonormal vectors (i.e. $\|u_i\| = 1$ and $\langle u_i, u_j \rangle = 0$ for $i \neq j$).

3 Power Iteration

Let $A \in \mathbb{R}^{n \times n}$ be a real symmetric matrix. The *Spectral Theorem* lets us decompose such a matrix into a diagonal and an orthogonal matrix.

Theorem 1 (Spectral Theorem)**.** *There exists an orthogonal matrix $U \in \mathbb{R}^{n \times n}$ and a diagonal matrix $\Lambda \in \mathbb{R}^{n \times n}$ such that*

$$A = U \Lambda U^T,$$

the columns of U are the eigenvectors of A and the diagonal entries of Λ are the corresponding eigenvalues.

Note that this makes no claim about *computing* U and Λ. Let $|\lambda_1| \geq \cdots \geq |\lambda_n|$ be the eigenvalues of A in decreasing magnitude and $u_1, \ldots, u_n$ the corresponding eigenvectors. This lecture is dedicated to efficiently finding the (by magnitude) largest eigenvalue λ_1 and the corresponding top eigenvector u_1 by means of *Power Iteration*.

3.1 Motivation

Determining the top eigenvector or its corresponding eigenvalue λ_1 is useful in a number of different settings. For example, the *operator norm* $\|A\|_{\text{op}}$ of a matrix A measures the maximum scaling

performed by the linear operation. Formally, it is defined as

$$\|A\|_{\text{op}} := \max_{v \neq 0} \frac{\|Av\|}{\|v\|}.$$

If A is symmetric, then $\|A\|_{\text{op}}$ is equal to $|\lambda_1|$.

3.2 Finding the largest eigenvector

For now, we assume that there is a gap between the first and second eigenvalues of A. Concretely, assume $|\lambda_1| > (1+\varepsilon)|\lambda_2|$ for some $\varepsilon > 0$. This also means that u_1 is uniquely determined (up to normalization). This is in contrast to the case of $\lambda_1 = \lambda_2$ where the corresponding eigenvectors form a subspace of at least two dimensions.

The idea of power iteration is simple: starting with a (random) vector z and repeatedly mutliplying by A will likely quickly converge to the largest eigenvector u_1. Concretely, we choose some starting vector $z \in \mathbb{R}^n$ and set $v_0 := \frac{z}{\|z\|}$. We then iteratively compute

$$\begin{aligned} y_{t+1} &= Ay_t, \\ \text{and } v_{t+1} &= \frac{y_{t+1}}{\|y_{t+1}\|}. \end{aligned}$$

We repeatedly normalize to prevent overflow or underflow in the size of the vector. Theoretically, it would suffice to iterate multiplication with A and only normalize once at the very end.

By the Spectral Theorem, there exists an orthogonal matrix $U \in \mathbb{R}^{n \times n}$, whose columns are orthonormal eigenvectors of A, such that $A = U\Lambda U^T$ where Λ a diagonal matrix containing the eigenvalues corresponding to the eigenvectors. Using the fact that $U^T U = I$, we have

$$A^m z = (U\Lambda U^T)^m z = U\Lambda^m U^T z.$$

Since $U^T = U^{-1}$, mutliplying z by U^{-1} yields the same vector but in the basis given by the orthonormal eigenvectors $u_1, \ldots, u_n$ that make up the columns of U. Analogously, multiplying by U undoes this transformation and returns the vector to the original basis. We can therefore view the algorithm as repeatedly multiplying by the diagonal matrix Λ *in the basis given by the eigenvectors* $u_1, \ldots, u_n$ (and eventually normalizing). In each iteration, the component of the vector along u_1 will be amplified the most since it has the largest eigenvalue by magnitude. Therefore, after enough iterations, we can expect all the other components to be insignificantly small in comparison and for the final (normalized) vector to be close to u_1.

We want to show that this algorithm is effective by upper bounding the distance between the final unit vector $v_m = \frac{A^m z}{\|A^m z\|}$ and the eigenvector u_1. To make the analysis easier, we will actually consider the distance between u_1 and $\tilde{v}_m = \frac{A^m z}{\lambda_1^m \alpha_1}$, i.e. a vector that points in the same direction as v_m but is in general not normalized and instead scaled by a different factor. The following lemma shows that showing $\tilde{v}_m$ to be a good approximation for u_1 is good enough.

Lemma 2. *Let $\delta > 0$. If $\|\tilde{v}_m - u_1\| \leq \delta$, then $\|v_m - u_1\| \leq 2\delta$.*

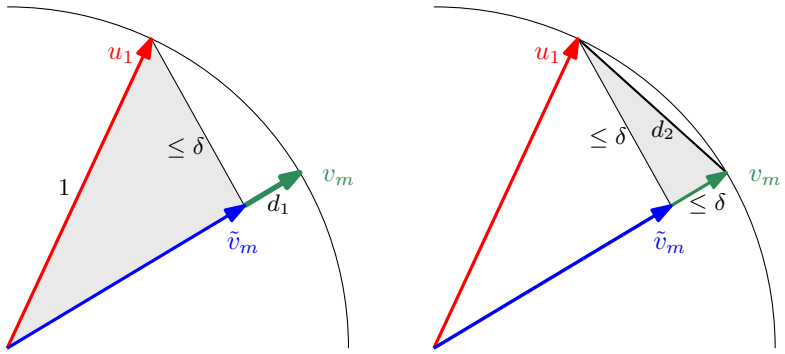


Figure 1: We apply the reverse triangle inequality twice to show that $d_1 \leq \delta$ and $d_2 \leq 2\delta$.

Proof. The reverse triangle inequality states that every side of a triangle is longer than the absolute difference (in length) between the other two sides. Figure 1 illustrates how we can apply this inequality with two different triangles to prove first that

$$\|\tilde{v}_m - u_1\| \geq \big|\|\tilde{v}_m\| - \|u_1\|\big| = \big|\|\tilde{v}_m\| - 1\big| = \|\tilde{v}_m - w_m\|,$$

and secondly that

$$\begin{aligned} &\|\tilde{v}_m - v_m\| \geq \|v_m - u_1\| - \|\tilde{v}_m - u_1\| \\ \Rightarrow\ &\|v_m - u_1\| \leq \underbrace{\|\tilde{v}_m - v_m\|}_{\leq\delta} + \underbrace{\|\tilde{v}_m - u_1\|}_{\leq\delta} \leq 2\delta. \end{aligned}$$

□

Thus, normalizing the vector $\tilde{v}_m$ to v_m will increase the distance to u_1 by a factor of at most two.

We proceed to show that $\tilde{v}_m$ (and in turn v_m) is close to the top eigenvector u_1 if the following two conditions are met.

1. There is a gap between the two largest eigenvalues by magnitude, that is to say $(1-\gamma)\cdot|\lambda_1| \geq |\lambda_2|$ for some $0 < \gamma \leq 1$.
2. The starting vector z is not orthogonal to the top eigenvector u_1, i.e. $\alpha_1 = \langle u_1, z\rangle \neq 0$.

Theorem 3. *Let $\delta > 0$ and $0 < \gamma \leq 1$. Let $z = \sum_i \alpha_i u_i \in \mathbb{R}^n$. If $(1-\gamma)\cdot|\lambda_1| \geq |\lambda_2|$ and $\alpha_1 \neq 0$, then for*

$$m \geq \frac{\ln C + \ln(1/\delta)}{\gamma}, \text{ where } C := \sqrt{\sum_{i>1}\left|\frac{\alpha_i}{\alpha_1}\right|^2},$$

we have $\|\tilde{v}_m - u_1\| \leq \varepsilon$.

Proof. By the Spectral Theorem we can decompose A into a diagonal matrix Λ of eigenvalues and

an orthogonal matrix U whose columns form an eigenbasis of A. Then,

$$\begin{aligned}\tilde{v}_m &= \frac{A^m z}{\lambda_1^m \alpha_1} = \frac{U \Lambda^m U^T z}{\lambda_1^m \alpha_1} \\ &= \frac{1}{\lambda_1^m \alpha_1} \sum_i \lambda_i^m \alpha_i u_i \\ &= \sum_i \left(\frac{\lambda_i}{\lambda_1}\right)^m \cdot \frac{\alpha_i}{\alpha_1} \cdot u_i \\ &= u_1 + \sum_{i>1} \left(\frac{\lambda_i}{\lambda_1}\right)^m \cdot \frac{\alpha_i}{\alpha_1} \cdot u_i.\end{aligned}$$

Notably, each iteration decreases the magnitude of the component vectors $u_{i>1}$ by a factor of $\left|\frac{\lambda_i}{\lambda_1}\right| \leq \left|\frac{\lambda_2}{\lambda_1}\right| \leq 1-\gamma$, that is to say they decrease exponentially in the number of iterations m. Since all u_i are orthogonal, we have

$$\|\tilde{v}_m - u_1\|^2 = \left\| \sum_{i>1} \left(\frac{\lambda_i}{\lambda_1}\right)^m \cdot \frac{\alpha_i}{\alpha_1} \cdot u_i \right\|^2 = \sum_{i>1} \left|\frac{\lambda_i}{\lambda_1}\right|^{2m} \left|\frac{\alpha_i}{\alpha_1}\right|^2 \leq (1-\gamma)^{2m} \sum_{i>1} \left|\frac{\alpha_i}{\alpha_1}\right|^2.$$

Define $C := \sqrt{\sum_{i>1} \left|\frac{\alpha_i}{\alpha_1}\right|^2}$. Using $(1-\gamma)^m \leq e^{-\gamma m}$, we get

$$\|\tilde{v}_m - u_1\| \leq (1-\gamma)^m C \leq e^{-\gamma m} C.$$

Picking m as above ensures that $\|\tilde{v}_m - u_1\| \leq \delta$. □

3.3 Choice of starting vector z

Looking at the term C from the previous section, it is easy to see that this term grows larger the smaller the component α_1 of z in the direction of the top eigenvector is. This makes sense, as we are effectively amplifying the amplitude of this component in each step of power iteration. In the extreme case of $\alpha_1 = 0$ (where z is orthogonal to u_1) no number of iterations will get us closer to u_1.

Without any prior knowledge about u_1, a good approach is to use a random starting vector z. In particular, sampling each component of z from a Gaussian distribution $\mathcal{N}(0,1)$ has the nice property that the resulting distribution is invariant under multiplication with orthogonal matrices. This ensures that while we do not know U given by the spectral decomposition of A, we know that $U^T z$ is still distributed according to a multivariate Gaussian distribution. This guarantees that for $U^T z = \sum_{i=1}^n \alpha_i u_i$ we have both

$$|\alpha_1| \geq 0.1 \text{ and } \|z\|^2 - \alpha_1^2 \leq 4n$$

with constant probability over the choice of z. Therefore, after $\Omega(\log n)$ (independent) runs of power iteration we will choose at least one good starting vector except with probability at most $1/2^{\Omega(n)}$.

3.4 Finding the largest eigenvalue

We now turn to finding a the largest Eigenvalue λ_1 of a *positive-semidefinite* matrix $A \in \mathbb{R}^{n \times n}$, which means that all its eigenvalues are non-negative. Finding only the top eigenvalue without the corresponding eigenvector is in a way easier as we do not run into problems if there is no unique top eigenvector.

We estimate λ_1 using the *Rayleigh quotient*, defined as

$$R_A(v) := \frac{v^T A v}{v^T v}.$$

If v is normalized, the Rayleigh quotient is a weighted sum of A's eigenvalues.

Lemma 4. *Let $v = \sum_i \beta_i u_i \in \mathbb{R}^n$ be a vector written in the eigenbasis. If $\|v\| = 1$, then*

$$R_A(v) = \sum_i \lambda_i \beta_i^2.$$

In particular, its value is bounded by the largest eigenvalue λ_1, i.e. $R_A(v) \leq \lambda_1$.

Proof. Since the u_is are orthonormal, we have

$$\begin{aligned} R_A(v) &= \frac{1}{v^T v} \left(\sum_i \beta_i u_i \right)^T A \sum_i \beta_i u_i \\ &= \frac{1}{\|v\|^2} \left(\sum_i \beta_i u_i \right)^T \sum_i \lambda_i \beta_i u_i \\ &= \sum_{i,j} \lambda_i \beta_i \beta_j \langle u_i, u_j \rangle \\ &= \sum_i \lambda_i \beta_i^2. \end{aligned}$$

□

The idea is that since the Rayleigh quotient is a weighted sum of the eigenvalues and after power iteration most of the mass lies with the top eigenvectors, the Rayleigh quotient will yield a good approximation for λ_1.

Theorem 5. *Let $\varepsilon \in (0, 1)$ be the desired level of accuracy. And $z = \sum_i \alpha_i u_i$ the power iteration starting vector. If $\alpha_1 \neq 0$ and*

$$m \geq \frac{\ln(\|z\|/|\alpha_1|) + \frac{1}{2}\ln(1/\varepsilon)}{\varepsilon},$$

then $(1 - 2\varepsilon) \cdot \lambda_1 \leq R_A(v_m) \leq \lambda_1$.

Proof. The upper bound $R_A(v_m) \leq \lambda_1$ follows directly from Lemma 4. To show the lower bound, $(1 - 2\varepsilon)\lambda_1$, define $S_\varepsilon := \{j \mid \lambda_j \geq (1 - \varepsilon) \cdot \lambda_1\}$ as the set of indices of eigenvalues within a factor $1 - \varepsilon$ of λ_1. Consider now the vector

$$w_m = \frac{A^m z}{\lambda_1^m} = \sum_j \left(\frac{\lambda_j}{\lambda_1} \right) \alpha_j u_j$$

which is the result from power iteration scaled by $1/\lambda_1^m$. Observe that

$$\sum_{j \notin S_\varepsilon} \langle u_j, w_m \rangle^2 = \sum_{j \notin S_\varepsilon} \left(\frac{\lambda_j}{\lambda_1} \right)^{2m} \cdot \alpha_j^2 \leq \sum_{j \notin S_\varepsilon} (1-\varepsilon)^{2m} \cdot \alpha_j^2 \leq (1-\varepsilon)^{2m} \sum_j \alpha_j^2 \leq e^{-2\varepsilon m} \|z\|^2.$$

That is to say, the sum of squares of the tail coefficients of w_m are bounded by $e^{2\varepsilon m}\|z\|^2$. Now let $v_m = w_m/\|w_m\| = \sum_j \beta_j u_j$ be the normalized vector. Then,

$$\sum_{j \notin S_\varepsilon} \beta_j^2 = \sum_{j \notin S_\varepsilon} \left(\frac{\langle u_j, w_m \rangle}{\|w_m\|} \right)^2 = \frac{1}{\|w_m\|^2} \sum_{j \notin S_\varepsilon} \langle u_j, w_m \rangle^2 \leq \frac{e^{-2\varepsilon m}\|z\|^2}{\|w_m\|^2}$$

At the same time, since the first component of z does not decay during power iteration, $\|\tilde{v}_m\|^2$ is lower bounded by $\langle u_i, z \rangle^2 = \alpha_1^2$. Hence, the above term is upper bounded by

$$\frac{e^{-2\varepsilon m}\|z\|^2}{\alpha_1^2}.$$

Picking m as above ensures that this is at most ε. Therefore

$$\sum_{j \in S_\varepsilon} \beta_j^2 = \underbrace{\left(\sum_j \beta_j^2 \right)}_{=1} - \sum_{j \notin S_\varepsilon} \beta_j^2 \geq 1 - \varepsilon$$

and

$$\begin{aligned} R_A(v_m) &= \sum_j \lambda_j \beta_j^2 \geq \sum_{j \in S_\varepsilon} \lambda_j \beta_j^2 \\ &\geq (1-\varepsilon)\lambda_1 \sum_{j \in S_\varepsilon} \beta_j^2 \\ &\geq (1-\varepsilon)^2 \lambda_1 \geq (1-2\varepsilon)\lambda_1. \end{aligned}$$

□