CS-774 Spectral Graph Theory Aug 03, 2023

Lecture 4: Linear Algebra Basics

Lecturer: Akash Kumar *Scribe: Akash Kumar*

1 Linear Algebra Background

In this part of the notes, we review the basic linear algebra background you need to follow this course. These notes borrow heavily from the excellent notes by Luca Trevisan which you are invited to consult ([Tre16]). My primary objective in these notes is to present the variational characterization of eigenvalues. To describe this, I will first recap the notion eigenvalues, eigenvectors and the spectral theorem for symmetric matrices. We will also take some tiny little detours and get our feet wet with tiny little adventuorus expeditions in linear algebra. Alright, let's get started.

1.1 Eigenvalues and Eigenvectors

In this class, we will be interested in studying vector spaces over $V \subseteq \mathbb{R}^n$. Consider the following standard inner product defined on this space. It is the function

$$\bullet : V \times V \to \mathbb{R}$$

where $\bullet(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{u} \bullet \boldsymbol{v} = \sum \boldsymbol{u}_i \boldsymbol{v}_i$. For the rest of this lecture, it is convenient to think $V = \mathbb{R}^n$ and that V comes equipped with the standard inner product. Note that by definition, $\boldsymbol{u} \bullet \boldsymbol{v} = \boldsymbol{v} \bullet \boldsymbol{u}$ and that it distributes linearly "on both sides"[1]. That is,

1. $\boldsymbol{u} \bullet (\boldsymbol{v} + \boldsymbol{w}) = \boldsymbol{u} \bullet \boldsymbol{v} + \boldsymbol{u} \bullet \boldsymbol{w}$.
2. $(\boldsymbol{v} + \boldsymbol{w}) \bullet \boldsymbol{u} = \boldsymbol{v} \bullet \boldsymbol{w} + \boldsymbol{w} \bullet \boldsymbol{u}$. Note that if you define inner products to be symmetric, this second item is kind of redundant.

For a vector $\boldsymbol{u} \in \mathbb{R}^n$, we denote its length as $\|\boldsymbol{x}\|_2^2 = \boldsymbol{x} \bullet \boldsymbol{x} = \sum \boldsymbol{x}_i^2$. Now, let us take a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$. We call a matrix $\boldsymbol{M} \in \mathbb{R}^{n \times m}$ a transpose of $\boldsymbol{A}$ with respect to the standard inner product "$\bullet$" if for all pairs of vectors $\boldsymbol{u} \in \mathbb{R}^n, \boldsymbol{v} \in \mathbb{R}^m$, you have

$$\boldsymbol{A}\boldsymbol{u} \bullet \boldsymbol{v} = \boldsymbol{u} \bullet \boldsymbol{M}\boldsymbol{v}.$$

If this is the first time you are looking at this definition of a transpose, you might be wondering if this is similar to the classic definition you have seen before which requires $\boldsymbol{A}(i, j) = \boldsymbol{M}(j, i)$ to hold for all $i \in [m], j \in [n]$. Indeed, if you working with an inner product space, the two definitions

[1]This property is called *bilinearity* of the inner product.

coincide. Let us first see that our newer more abstract definition implies the more familiar definition. Note that by choosing $\boldsymbol{u} = \boldsymbol{e}_i$ and $\boldsymbol{v} = \boldsymbol{f}_j$[2], you conclude that

$$\boldsymbol{A}(i,j) = \boldsymbol{A}\boldsymbol{e}_i \bullet \boldsymbol{e}_j = \boldsymbol{e}_i \bullet \boldsymbol{M}\boldsymbol{f}_j = \boldsymbol{M}(j,i).$$

In the other direction, you know that

$$\boldsymbol{A}(i,j) = \boldsymbol{M}(j,i) \implies \boldsymbol{A}\boldsymbol{e}_i \bullet \boldsymbol{f}_j.$$

We would like to show for any fixed pair of vectors $\boldsymbol{u} \in \mathbb{R}^n, \boldsymbol{v} \in \mathbb{R}^m$ you indeed have $\boldsymbol{A}\boldsymbol{u} \cdot \boldsymbol{v}$ equals $\boldsymbol{u} \cdot \boldsymbol{M}\boldsymbol{v}$. This is done below by exploiting what we know about inner products.

$$\begin{aligned}
\boldsymbol{A}\boldsymbol{u} \bullet \boldsymbol{v} &= \left(\boldsymbol{A} \sum_{i\in[n]} u_i e_i\right) \bullet (\sum_{j\in[m]} v_j \boldsymbol{f}_j) \\
&= \sum_{\substack{i\in[n] \\ j\in[m]}} u_i v_j \boldsymbol{A}\boldsymbol{e}_i \bullet \boldsymbol{f}_j && \text{By bilinearity of inner product} \\
&= \sum_{\substack{i\in[n] \\ j\in[m]}} u_i v_j \boldsymbol{e}_i \bullet \boldsymbol{M}\boldsymbol{f}_j && \text{Because } \boldsymbol{A}\boldsymbol{e}_i \bullet \boldsymbol{f}_j = \boldsymbol{e}_i \bullet \boldsymbol{M}\boldsymbol{f}_j
\end{aligned}$$

Alright, now let us define eigenvalues and eigenvectors of a matrix $\boldsymbol{M} \in \mathbb{R}^{n\times n}$. If for some $\lambda \in \mathbb{C}$ and $\boldsymbol{x} \in \mathbb{C}^n \setminus \{0\}$, we have

$$\boldsymbol{M}\boldsymbol{x} = \lambda\boldsymbol{x}$$

we call λ the eigenvalue and $\boldsymbol{x}$ the corresponding eigenvector with eigenvalue λ. In this class, we would like to think about matrices with real valued eigenvalues and real valued eigenvectors. Fortunately, by just insisting that your matrix is symmetric, you can achieve this. This is the content of the spectral theorem which we cover next.

2 The Spectral Theorem

The main result of this section is the following theorem.

Theorem 2.1. *Let $\boldsymbol{M} \in \mathbb{R}^{n\times n}$ denote a symmetric matrix. Then there are n real numbers (not necessarily distinct) $\lambda_1, \lambda_2, \cdots \lambda_n$ and n orthonormal vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots \boldsymbol{x}_n \in \mathbb{R}^n$ such that $\boldsymbol{x}_i$ is an eigenvector with eigenvalue λ_i.*

[2] Here, $\boldsymbol{e}_i \in \mathbb{R}^n$ refers to the vector which is 1 in the ith coordinate and zero everywhere else. $\boldsymbol{f}_j \in \mathbb{R}^m$ refers to the vector which is 1 in jth coordinate and zero everywhere else.

Note that since the matrices in this section are symmetric – that is to say $\boldsymbol{M} = \boldsymbol{M}^T$ – we also know that for all pairs of vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ we have $\boldsymbol{Mu} \bullet \boldsymbol{v} = \boldsymbol{u} \bullet \boldsymbol{Mv}$. Such matrices are also called self-adjoint. The theorem above enshrines the reason why we love self-adjoint matrices. To establish this theorem, we begin with an important claim.

Claim 2.2. *Let $\boldsymbol{M} \in \mathbb{R}^{n\times n}$ denote a real symmetric matrix. Then $\boldsymbol{M}$ admits one real valued eigenvalue and one real valued eigenvector.*

Proof. Let $\boldsymbol{x} \in \mathbb{R}^n$ denote the vector which maximizes the length, $\|\boldsymbol{Mx}\|_2^2$ and let $\alpha = \|\boldsymbol{Mx}\|_2$. Note that such a maximizer exists because $\{\boldsymbol{x} \in \mathbb{R}^n \colon \|\boldsymbol{x}\|_2 = 1\}$ is a closed set. If there are many maximizers, we pick one arbitrarily. Now, note that $\boldsymbol{Mx} \bullet \boldsymbol{Mx} \geq 0$. Also,

$$\boldsymbol{Mx} \bullet \boldsymbol{Mx} = \boldsymbol{x} \bullet \boldsymbol{M}(\boldsymbol{Mx}).$$

Writing $\boldsymbol{y} = \boldsymbol{Mx}$, you get

$$\boldsymbol{Mx} \bullet \boldsymbol{Mx} = \boldsymbol{x} \bullet \boldsymbol{My} = \boldsymbol{x} \bullet \alpha^2 \boldsymbol{x}.$$

Now, two cases arise.

1. $\boldsymbol{x}$ and $\boldsymbol{y}$ are parallel. In this case, $\boldsymbol{x}$ is an eigenvector with eigenvalue $+\alpha$ or $-\alpha$. In either case, you have a real valued eigenpair – $(\alpha, \boldsymbol{x})$.
2. $\boldsymbol{x}$ and $\boldsymbol{y}$ are not parallel. In this case, $\boldsymbol{y} + \alpha\boldsymbol{x} \neq 0$. Note that

$$\boldsymbol{M}(\boldsymbol{y} + \alpha\boldsymbol{x}) = \boldsymbol{My} + \alpha\boldsymbol{Mx} = \alpha^2\boldsymbol{x} + \alpha\boldsymbol{y} = \alpha(\boldsymbol{y} + \alpha\boldsymbol{x}).$$

   Thus, in this case $\boldsymbol{y} + \alpha\boldsymbol{x}$ is a real valued eigenvector with eigenvalue α.

■

Now we are ready to prove Theorem 2.1.

Proof Sketch for Theorem 2.1. We use Claim 2.2 to get a real eigenpair $(\lambda_1, \boldsymbol{x}_1)$ of $\boldsymbol{M}_1 = \boldsymbol{M}$. Define $\boldsymbol{M}_2 = \boldsymbol{M}_1 - \lambda_1\boldsymbol{x}_1\boldsymbol{x}_1^T$ and note that being a difference of two symmetric matrices, $\boldsymbol{M}_2$ is also symmetric. Note that $(0, \boldsymbol{x}_1)$ is an eigenpair of $\boldsymbol{M}_2$ (why?). Next up, let's take any vector orthogonal $\boldsymbol{w}$ orthogonal to $\boldsymbol{x}_1$. You get

$$\boldsymbol{M}_2\boldsymbol{w} = (\boldsymbol{M}_1 - \lambda_1\boldsymbol{x}_1\boldsymbol{x}_1^T)\boldsymbol{w} = \boldsymbol{M}_1\boldsymbol{w} - \lambda_1\boldsymbol{x}_1\left(\boldsymbol{x}_1 \bullet \boldsymbol{w}\right) = \boldsymbol{M}_1\boldsymbol{w}.$$

This means $\boldsymbol{M}_2$ quashes out any component of an input vector along the direction $\boldsymbol{x}_1$ and maps any input vector to within the subspace orthogonal to $\boldsymbol{x}_1$.

Let us now take an eigenpair $(\mu, \boldsymbol{y})$ of $\boldsymbol{M}_2$. By the preceding observation, we know $\boldsymbol{y} \perp \boldsymbol{x}_1$. This gives $\boldsymbol{M}_2\boldsymbol{y} = \mu\boldsymbol{y} = \boldsymbol{M}_1\boldsymbol{y}$ which means the pair $(\mu, \boldsymbol{y})$ is an eigenpair for $\boldsymbol{M}_1$ as well. We use Claim 2.2 to obtain a real valued eigenpair $(\lambda_2, \boldsymbol{x}_2)$ of $\boldsymbol{M}_2$ (and hence $\boldsymbol{M} = \boldsymbol{M}_1$). We proceed inductively to obtain other eigenvalues and eigenvectors. ■

Next, via Lemma 2.3, we emphasize one aspect of the above theorem which asserts that eigenvectors of a symmetric matrix corresponding to distinct eigenvalues are orthogonal. This is important enough that we reiterate it once more.

Lemma 2.3. *Let $\boldsymbol{M} \in \mathbb{R}^{n \times n}$ denote a real symmetric matrix. Let $\lambda \in \mathbb{R}$ denote an eigenvalue with eigenvector $\boldsymbol{x}$ and let $\mu \neq \lambda \in \mathbb{R}$ denote a different eigenvalue of $\boldsymbol{M}$ with eigenvector $\boldsymbol{y}$. Then $\boldsymbol{x}, \boldsymbol{y}$ are orthogonal.*

Proof. Note that $\boldsymbol{Mx} \bullet \boldsymbol{y} = \lambda \cdot \boldsymbol{x} \bullet \boldsymbol{y}$. Also, $\boldsymbol{Mx} \bullet \boldsymbol{y} = \boldsymbol{x} \bullet \boldsymbol{My} = \mu \cdot \boldsymbol{x} \bullet \boldsymbol{y}$. Since $\lambda \neq \mu$, this means $\boldsymbol{x} \bullet \boldsymbol{y} = 0$. ■

3 Variational Charaterization of Eigenvalues

Following [Tre16], I conclude these notes with the variational characterization of eigenvalues for real symmetric matrices.

Theorem 3.1. *Let $\boldsymbol{M} \in Sym(\mathbb{R}^{n \times n})$ denote a real symmetric matrix and let $\lambda_1 \leq \lambda_2 \leq \dots \lambda_n$ denote the eigenvalues of $\boldsymbol{M}$ arranged in increasing order. Then, for every $k \in [n]$,*

$$\lambda_k = \min_{\substack{V \text{ subspsace of } \mathbb{R}^n \\ dim(V)=k}} \max_{\boldsymbol{x} \colon \|\boldsymbol{x}\|_2 = 1} \boldsymbol{x}^T \boldsymbol{Mx}.$$

The quantity $\boldsymbol{x}^T\boldsymbol{Mx}$ for a unit vector $\boldsymbol{x}$ is called the *Rayleigh Quotient* of $\boldsymbol{x}$ with respect to $\boldsymbol{M}$. It is denoted as $\mathcal{R}_{\boldsymbol{M}}(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{Mx}$. When the underlying matrix is clear from the context, I will write $\mathcal{R}(\boldsymbol{x})$ to refer to $\mathcal{R}_{\boldsymbol{M}}(\boldsymbol{x})$.

Proof. We will show

- $\lambda_k \geq \min_{\substack{V \text{ subspsace of } \mathbb{R}^n \\ dim(V)=k}} \max_{\boldsymbol{x} \colon \|\boldsymbol{x}\|_2 = 1} \boldsymbol{x}^T \boldsymbol{Mx}$, and
- $\lambda_k \leq \min_{\substack{V \text{ subspsace of } \mathbb{R}^n \\ dim(V)=k}} \max_{\boldsymbol{x} \colon \|\boldsymbol{x}\|_2 = 1} \boldsymbol{x}^T \boldsymbol{Mx}$.

From these two statements, the result follows.

Let us show the former result first. To this end, let $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots \boldsymbol{v}_n$ denote the orthonormal eigenvectors corresponding to the eigenvalues $\lambda_1 \leq \lambda_2 \leq \dots \lambda_n$ (which are promised to exist by the spectral theorem). Consider the k-dimensional space spanned by $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_k$. Let $\boldsymbol{x}$ denote a unit vector in this k-dimensional space. Thus, $\boldsymbol{x} = \sum_{i=1}^k a_i \boldsymbol{v}_i$ where $\sum a_i^2 = 1$. We have

$$\mathcal{R}(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{Mx} = \sum_{i,j} a_i a_j \boldsymbol{v}_i^T \boldsymbol{M} \boldsymbol{v}_j = \sum_{i,j} a_i a_j \lambda_j \boldsymbol{v}_i^T \boldsymbol{v}_j = \sum_{i=1}^k \lambda_i a_i^2 \leq \lambda_k \sum a_i^2 = \lambda_k.$$

This demonstrates that there exists a k-dimensional subspace where the Rayleigh Quotient of all unit vectors in this subspace is no larger than λ_k. Now, we would like to show the other direction. Namely, we would like to conclude that in any other k-dimensional subspace of $\mathbb{R}^n$, there exists *an offending vector* whose Rayleigh Quotient is large and thus sticks out.

To this end, let V denote any k-dimensional subspace of $\mathbb{R}^n$. Let $S = span(\boldsymbol{v}_k, \boldsymbol{v}_{k+1}, \cdots, \boldsymbol{v}_n)$. Note that $dim(V) + dim(S) = k + n - k + 1 > n$. Thus, these two spaces must have some non-zero vector in common. Take one such vector, $\boldsymbol{x}$, normalized to have length 1. Write $\boldsymbol{x} = \sum_{i=k}^n a_i \boldsymbol{v}_i$. The Rayleigh Quotient of $\boldsymbol{x}$ equals

$$\mathcal{R}(\boldsymbol{x}) = \sum_{i=k}^n \lambda_i a_i^2 \geq \lambda_k.$$

∎

Now, we finish up with a run-down of a few quickies you can show with the variational characterization in your hand.

Lemma 3.2. *Let $\boldsymbol{M} \in Sym(\mathbb{R}^{n\times n})$ denote a real symmetric matrix with eigenvalues $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$. Then*

- $\lambda_1 = \min_{\|\boldsymbol{x}\|_2=1} \mathcal{R}(\boldsymbol{x})$. *Further, every minimizer is an eigenvector of λ_1.*
- *Letting $\boldsymbol{x}_1$ denote an eigenvector of λ_1, we have $\lambda_2 = \min_{\substack{\boldsymbol{x}\perp\boldsymbol{x}_1 \\ \|\boldsymbol{x}\|_2=1}} \mathcal{R}(\boldsymbol{x})$. Further, every minimizer is an eigenvector of λ_2.*
- $\lambda_n = \max_{\|\boldsymbol{x}\|_2=1} \mathcal{R}(\boldsymbol{x})$. *Further, every maximizer is an eigenvector of λ_n.*

Proof. Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ denote an orthogonal system of eigenvectors corresponding to the given eigenvalues. Express $\boldsymbol{x}$ in the eigenbasis as $\boldsymbol{x} = \sum a_i \boldsymbol{v}_i$ and thus $\sum a_i^2 = 1$. Note that $\mathcal{R}(\boldsymbol{x}) = \sum a_i^2 \lambda_i$. This is minimized when $a_1 = 1$ and $a_2 = a_3 = \cdots = a_n = 0$. This means $\boldsymbol{x} = \boldsymbol{v}_1$ achieves the minimum in which case we have $\mathcal{R}(\boldsymbol{v}_1) = \boldsymbol{v}_1^T \boldsymbol{M} \boldsymbol{v}_1 = \lambda_1$. Also, note that if $\mathcal{R}(\boldsymbol{x}) = \lambda_1$, then $a_i = 0$ for all $i \geq 2$. This means a minimizer can only be a linear combination of eigenvectors of λ_1 and hence it is itself an eigenvector of λ_1.

For the next part, note that $\boldsymbol{x}$ has no component along $\boldsymbol{v}_1$. This means $\boldsymbol{x} = \sum_{i\geq 2} a_i \boldsymbol{v}_i$. Thus, $\mathcal{R}(\boldsymbol{x}) = \sum_{i\geq 2} a_i^2 \lambda_i$. This is minimized when $a_2 = 1$ and $a_3 = a_4 = \cdots = a_n = 0$. This means $\boldsymbol{x} = \boldsymbol{v}_2$ achieves the minimum in which case $\mathcal{R}(\boldsymbol{v}_2) = \lambda_2$. Also, note that if $\mathcal{R}(\boldsymbol{x}) = \lambda_2$ then $a_i = 0$ for all $i \geq 3$. This means minimum is achieved by those vectors $\boldsymbol{x}$ which are linear combinations of eigenvectors of λ_2 and hence every minimizer is an eigenvector of λ_2.

The last part follows by applying the result for the first part to the matrix $-\boldsymbol{M}$. ∎

References

[Tre16] Luca Trevisan Handout 00, *A course on Spectral Methods and Expanders, 2016* https://lucatrevisan.github.io/teaching/expanders2016/