Lecture 0: Linear Algebra Background

In which we review linear algebra prerequisites.

The following background from linear algebra will be sufficient for the sake of this course: to know what is an eigenvalue and an eigenvector, to know that real symmetric matrices have real eigenvalues and their real eigenvectors are orthogonal, and to know the variational characterization of eigenvalues.

1 Basic Definitions

If $x = a + ib$ is a complex number, then we let $\bar{x} = a - ib$ denote its *conjugate*. Note that a complex number x is real if and only if $x = \bar{x}$. If $M \in \mathbb{C}^{m \times n}$ is a matrix, then M^* denotes the conjugate transpose of M, that is, $(M^*)_{i,j} = \overline{M_{j,i}}$. If the entries of M are real, then $M^* = M^T$, where M^T is the *transpose* of M, that is, the matrix such that $(M^T)_{i,j} = M_{j,i}$.

We say that a matrix M is *Hermitian* if $M = M^*$. In particular, real symmetric matrices are Hermitian.

If $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$ are two vectors, then their inner product is defined as

$$\langle \mathbf{v}, \mathbf{w} \rangle := \mathbf{v}^* \mathbf{w} = \sum_i \overline{v_i} \cdot w_i \tag{1}$$

Notice that, by definition, we have $\langle \mathbf{v}, \mathbf{w} \rangle = (\langle \mathbf{w}, \mathbf{v} \rangle)^*$ and $\langle \mathbf{v}, \mathbf{v} \rangle = ||\mathbf{v}||^2$. Note also that, for two matrices A, B, we have $(A \cdot B)^* = B^* \cdot A^*$, and that for every matrix M and every two vectors $\mathbf{x}$, $\mathbf{y}$, we have

$$\langle M\mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^* M^* \mathbf{y} = \langle \mathbf{x}, M^* \mathbf{y} \rangle$$

If $M \in \mathbb{C}^{n \times n}$ is a square matrix, $\lambda \in \mathbb{C}$ is a scalar, $\mathbf{v} \in \mathbb{C}^n - \{\mathbf{0}\}$ is a non-zero vector and we have

$$M\mathbf{v} = \lambda \mathbf{v} \tag{2}$$

then we say that λ is an *eigenvalue* of M and that $\mathbf{v}$ is *eigenvector* of M corresponding to the eigenvalue λ.

2 The Spectral Theorem

We want to prove

Theorem 1 (Spectral Theorem) *Let $M \in \mathbb{R}^{n\times n}$ be a symmetric matrix with real-valued entries, then there are n real numbers (not necessarily distinct) $\lambda_1, \ldots, \lambda_n$ and n orthonormal real vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$, $\mathbf{x}_i \in \mathbb{R}^n$ such that $\mathbf{x}_i$ is an eigenvector of λ_i.*

Assuming the fundamental theorem of algebra (that every polynomial has a complex root) and basic properties of the determinant, the cleanest proof of the spectral theorem is to proceed by induction on n, and to show that M must have a real eigenvalue λ_1 with a real eigenvector $\mathbf{v}_1$, and to show that M maps vectors orthogonal to $\mathbf{v}_1$ to vectors orthogonal to $\mathbf{v}_1$. Then one applies the inductive hypothesis to M restricted to the $(n-1)$-dimensional space of vectors orthogonal to $\mathbf{v}_1$ and one recovers the remaining $(n-1)$ eigenvalues and eigenvectors.

The cleanest way to formalize the above proof is to give all definitions and results in terms of linear operators $T : \mathcal{V} \to \mathcal{V}$ where $\mathcal{V}$ is an arbitrary vector space over the reals. This way, however, we would be giving several definitions that we would never use in the future, so, instead, the inductive proof will use a somewhat inelegant change of basis to pass from M to an $(n-1)\times(n-1)$ matrix M'.

We begin by showing that a real symmetric matrix has real eigenvalues and eigenvectors.

Theorem 2 *If $M \in \mathbb{R}^{n\times n}$ is symmetric, then there is a real eigenvalue $\lambda \in \mathbb{R}$ and a real eigenvector $\mathbf{v} \in \mathbb{R}^n$ such that $M\mathbf{v} = \lambda\mathbf{v}$.*

We begin by noting that every matrix has a complex eigenvalue.

Lemma 3 *For every matrix $M \in \mathbb{C}^{n\times n}$, there is an eigenvalue $\lambda \in \mathbb{C}$ and an eigenvector $\mathbf{v} \in \mathbb{C}^n$ such that $M\mathbf{v} = \lambda\mathbf{v}$.*

Proof: Note that λ is an eigenvalue for M if and only if

$$\exists \mathbf{x} \neq \mathbf{0}.\ (M - \lambda I)\mathbf{x} = \mathbf{0}$$

which is true if and only if the rows of $M - \lambda I$ are not linearly independent, which is true if and only if

$$\det(M - \lambda I) = 0$$

Now note that the mapping $t \to \det(M - tI)$ is a univariate polynomial of degree n in t, and so it must have a root λ by the fundamental theorem of algebra. $\square$

Next we show that if M is real and symmetric, then its eigenvalues are real.

Lemma 4 *If M is Hermitian, then, for every $\mathbf{x}$ and $\mathbf{y}$,*

$$\langle M\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, M\mathbf{y} \rangle$$

PROOF:

$$\langle M\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, M^*\mathbf{y} \rangle = \langle \mathbf{x}, M\mathbf{y} \rangle$$

$\square$

Lemma 5 *If M is Hermitian, then all the eigenvalues of M are real.*

PROOF: Let M be an Hermitian matrix and let λ be a scalar and $\mathbf{x}$ be a non-zero vector such that $M\mathbf{x} = \lambda\mathbf{x}$. We will show that $\lambda = \lambda^*$, which implies that λ is a real number.

We note that

$$\langle M\mathbf{x}, \mathbf{x} \rangle = \langle \lambda\mathbf{x}, \mathbf{x} \rangle = \lambda^* ||x||^2$$

and

$$\langle \mathbf{x}, M\mathbf{x} \rangle = \langle \mathbf{x}, \lambda\mathbf{x} \rangle = \lambda ||x||^2$$

and by the fact that $\langle M\mathbf{x}, \mathbf{x} \rangle = \langle \mathbf{x}, M\mathbf{x} \rangle$, we have $\lambda = \lambda^*$. $\square$

In order to prove Theorem 2, it remains to argue that, for a real eigenvalue of a real symmetric matrix, we can find a real eigenvector.

PROOF:[Of Theorem 2] Let $M \in \mathbb{R}^{n \times n}$ be a real symmetric matrix, then M has a real eigenvalue λ and a (possibly complex valued) eigenvector $\mathbf{z} = \mathbf{x} + i\mathbf{y}$, where $\mathbf{x}$ and $\mathbf{y}$ are real vectors. We have

$$M\mathbf{x} + iM\mathbf{y} = \lambda\mathbf{x} + i\lambda\mathbf{y}$$

from which (recalling that the entries of M and the scalar λ are real) it follows that $M\mathbf{x} = \lambda\mathbf{x}$ and that $M\mathbf{y} = \lambda\mathbf{y}$; since $\mathbf{x}$ and $\mathbf{y}$ cannot both be zero, it follows that λ has a real eigenvector. $\square$

We are now ready to prove the spectral theorem

PROOF:[Of Spectral Theorem] We proceed by induction on n. The case $n = 1$ is trivial.

Assume that the statement is true for dimension $n-1$. Let λ_1 be a real eigenvalue of M and $\mathbf{x}_1$ be a real eigenvector λ_1.

Now we claim that for every vector $\mathbf{y}$ that is orthogonal to $\mathbf{x}_1$, then $M\mathbf{y}$ is also orthogonal to $\mathbf{x}_1$. Indeed, the inner product of $M\mathbf{y}$ and $\mathbf{x}_1$ is

$$\langle \mathbf{x}_1, M\mathbf{y} \rangle = \langle M\mathbf{x}_1, \mathbf{y} \rangle = \langle \lambda \mathbf{x}_1, \mathbf{y} \rangle = 0$$

Let $\mathcal{V}$ be the $n-1$-dimensional subspace of $\mathbb{R}^n$ that contains all the vectors orthogonal to $\mathbf{x}_1$. We want to apply the inductive hypothesis to M restricted to $\mathcal{V}$; we cannot literally do that, because the theorem is not stated in terms of arbitrary linear operators over vector spaces, so we will need to do that by fixing an appropriate basis for $\mathcal{V}$.

let $B \in \mathbb{R}^{n \times (n-1)}$ be a matrix that computes a bijective map from $\mathbb{R}^{n-1}$ to $\mathcal{V}$. (If $\mathbf{b}_1, \ldots, \mathbf{b}_{n-1}$ is an orthonormal basis for $\mathcal{V}$, then B is just the matrix whose columns are the vectors $\mathbf{b}_i$.) Let also $B' \in \mathbb{R}^{(n-1) \times n}$ be the matrix such that, for every $\mathbf{y} \in \mathcal{V}$, $BB'\mathbf{y} = \mathbf{y}$. (We can set $B' = B^T$ where B is as described above.) We apply the inductive hypothesis to the matrix

$$M' := B'MB \in \mathbb{R}^{(n-1) \times (n-1)}$$

and we find eigenvalues $\lambda_2, \ldots, \lambda_n$ and orthonormal eigenvectors $\mathbf{y}_2, \ldots, \mathbf{y}_n$ for M'.

For every $i = 2, \ldots, n$, we have

$$B'MB\mathbf{y}_i = \lambda_i \mathbf{y}_i$$

and so

$$BB'MB\mathbf{y}_i = \lambda_i B\mathbf{y}_i$$

Since $B\mathbf{y}_i$ is orthogonal to $\mathbf{x}_1$, it follows that $MB\mathbf{y}_i$ is also orthogonal to $\mathbf{x}_1$, and so $BB'MB\mathbf{y}_i = MB\mathbf{y}_i$, so we have

$$MB\mathbf{y}_i = \lambda_i B\mathbf{y}_i$$

and, defining $\mathbf{x}_i := B\mathbf{y}_i$, we have

$$M\mathbf{x}_i = \lambda_i \mathbf{x}_i$$

Finally, we observe that the vectors $\mathbf{x}_i$ are orthogonal. By construction, $\mathbf{x}_1$ is orthogonal to $\mathbf{x}_2, \ldots, \mathbf{x}_n$, and, for every $2 \leq i < j \leq n$, we have that

$$\langle \mathbf{x}_i, \mathbf{x}_j \rangle = \langle B\mathbf{y}_i, B\mathbf{y}_j \rangle = \langle \mathbf{y}_i, B^T B\mathbf{y}_j \rangle = \langle \mathbf{y}_i, \mathbf{y}_j \rangle = 0$$

We have not verified that the vectors $\mathbf{x}_i$ have norm 1 (which is true), but we can scale them to have norm 1 if not. $\square$

3 Variational Characterization of Eigenvalues

We conclude these notes with the variational characterization of eigenvalues for real symmetric matrices.

Theorem 6 *Let $M \in \mathbb{R}^{n\times n}$ be a symmetric matrix, and $\lambda_1 \leq \lambda_2 \leq \cdots \lambda_n$ be the eigenvalues of M in non-increasing order. Then*

$$\lambda_k = \min_{k-\dim\ \mathcal{V}} \quad \max_{\mathbf{x}\in\mathcal{V}-\{\mathbf{0}\}} \quad \frac{\mathbf{x}^T M\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$$

The quantity $\frac{\mathbf{x}^T M\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$ is called the *Rayleigh quotient* of $\mathbf{x}$ with respect to M, and we will denote it by $R_M(\mathbf{x})$.

Proof: Let $\mathbf{v}_1,\ldots,\mathbf{v}_n$ be orthonormal eigenvectors of the eigenvalues $\lambda_1,\ldots,\lambda_n$, as promised by the spectral theorem. Consider the k-dimensional space spanned by $\mathbf{v}_1,\ldots,\mathbf{v}_k$. For every vector $\mathbf{x} = \sum_{i=1}^k a_i\mathbf{v}_i$ in such a space, the numerator of the Rayleigh quotient is

$$\sum_{i,j} a_i a_j \mathbf{v}_i^T M\mathbf{v}_j = \sum_{i,j} a_i a_j \lambda_j \mathbf{v}_i^T\mathbf{v}_j = \sum_{i=1}^k \lambda_i a_i^2 \leq \lambda_k \cdot \sum_{i=1}^k a_i^2$$

The denominator is clearly $\sum_{i=1}^k a_j^2$, and so $R_M(\mathbf{x}) \leq \lambda_k$. This shows that

$$\lambda_k \geq \min_{k-\dim\ \mathcal{V}} \quad \max_{\mathbf{x}\in\mathcal{V}-\{\mathbf{0}\}} \quad \frac{\mathbf{x}^T M\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$$

For the other direction, let $\mathcal{V}$ be any k-dimensional space: we will show that $\mathcal{V}$ must contain a vector of Rayleigh quotient $\geq \lambda_k$. Let S be the span of $\mathbf{v}_k,\ldots,\mathbf{v}_n$; since S has dimension $n-k+1$ and $\mathcal{V}$ has dimension k, they must have some non-zero vector in common. Let $\mathbf{x}$ be one such vector, and let us write $\mathbf{x} = \sum_{i=k}^n a_i\mathbf{v}_i$. The numerator of the Rayleigh quotient of $\mathbf{x}$ is

$$\sum_{i=k}^n \lambda_i a_i^2 \geq \lambda_k \sum_i a_i^2$$

and the denominator is $\sum_i a_i^2$, so $R_M(\mathbf{x}) \geq \lambda_k$. $\square$

We have the following easy consequence.

Fact 7 *If λ_1 is the smallest eigenvalue of a real symmetric matrix M, then*

$$\lambda_1 = \min_{\mathbf{x}\neq\mathbf{0}} R_M(\mathbf{x})$$

Furthermore, every minimizer is an eigenvector of λ_1.

PROOF: The identity is the $k = 1$ case of the previous theorem. For the furthermore part, let $\lambda_1 \leq \cdots \lambda_n$ be the list of eigenvalues of M in non-decreasing order, and $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be corresponding eigenvectors. If $\mathbf{x} = \sum_i a_i \mathbf{v}_i$ is any vector, then

$$R_M(\mathbf{x}) = \frac{\sum_i \lambda_i a_i^2}{\sum_i a_i^2}$$

If $R_M(\mathbf{x}) = \lambda_1$, then $a_i = 0$ for every i such that $\lambda_i > \lambda_1$, that is, $\mathbf{x}$ is a linear combination of eigenvectors of λ_1, and hence it is an eigenvector of λ_1. □

Fact 8 *If λ_n is the largest eigenvalue of a real symmetric matrix M, then*

$$\lambda_n = \max_{\mathbf{x} \neq \mathbf{0}} R_M(\mathbf{x})$$

Furthermore, every maximizer is an eigenvector of λ_n.

PROOF: Apply Fact 7 to the matrix $-M$. □

Fact 9 *If λ_1 is the smallest eigenvalue of a real symmetric matrix M, and $\mathbf{x}_1$ is an eigenvector of λ_1, then*

$$\lambda_2 = \min_{\mathbf{x} \neq \mathbf{0},\ \mathbf{x} \perp \mathbf{x}_1} R_M(\mathbf{x})$$

Furthermore, every minimizer is an eigenvector of λ_2.

PROOF: A more conceptual proof would be to consider the restriction of M to the space orthogonal to $\mathbf{x}_1$, and then apply Fact 7 to such a linear operator. But, since we have not developed the theory for general linear operators, we would need to explicitly reduce to an $(n-1)$-dimensional case via a projection operator as in the proof of the spectral theorem.

Instead, we will give a more hands-on proof. Let $\lambda_1 \leq \lambda_2 \leq \cdots \lambda_n$ be the list of eigenvalues of M, with multiplicities, and $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be orthonormal vectors as given by the spectral theorem. We may assume that $\mathbf{v}_1 = \mathbf{x}_1$, possibly by changing the orthonormal basis of the eigenspace of λ_1. For every vector $\mathbf{x} = \sum_{i=2}^k a_i \mathbf{v}_i$ orthogonal to $\mathbf{v}_1$, its Rayleigh quotient is

$$R_M(\mathbf{x}) = \frac{\sum_{i=2}^n \lambda_i a_i^2}{\sum_i a_i^2} \geq \lambda_2$$

and the minimum is achieved by vectors $\mathbf{x}$ such that $a_i = 0$ for every $\lambda_i > \lambda_2$, that is, for vectors $\mathbf{x}$ which are linear combinations of the eigenvectors of λ_2, and so every minimizer is an eigenvector of λ_2. □