Review of Linear Algebra
Definitions, Change of Basis, Trace, Spectral Theorem

Steven J. Miller[*]

June 19, 2004

Abstract

Matrices can be thought of as rectangular (often square) arrays of numbers, or as linear transformations from one space to another (or possibly to the same space). The former picture is the simplest starting point, but it is the latter, geometric view that gives a deeper understanding. We content ourselves in these notes with giving a brief review of some of the definitions and results of Linear Algebra, leaving many of the proofs to the reader; for more detail, the reader should consult a textbook in Linear Algebra, for example, [St].

Contents

1 Definitions

Definition 1.1 (Transpose, Complex Conjugate Transpose). *Given an $n \times m$ matrix A (where n is the number of rows and m is the number of columns), the transpose of A, denoted A^T, is the $m \times n$ matrix where the rows of A^T are the columns of A. The complex conjugate transpose, A^*, is the complex conjugate of the transpose of A.*

Exercise 1.2. *Prove $(AB)^T = B^T A^T$ and $(A^T)^T = A$.*

[*]sjmiller@math.ohio-state.edu

Definition 1.3 (Real Symmetric, Complex Hermitian). *If an $n \times n$ real matrix A satisfies $A^T = A$, then we say A is real symmetric; if an $n \times n$ complex matrix A satisfies $A = (A^*)^T$, then we say A is complex hermitian.*

Definition 1.4 (Dot or Inner Product). *The dot (or inner) product of two real vectors v and w is defined as $v^T w$; if the vectors are complex, we instead use $(v^*)^T w$. If v and w have n components, $v^T w = v_1 w_1 + \cdots v_n w_n$.*

Definition 1.5 (Orthogonality). *Two real vectors are orthogonal (or perpendicular) if $v^T w = 0$; for complex vectors, the equivalent condition is $(v^*)^T w = 0$.*

Definition 1.6 (Length of a vector). *The length of a real vector v is $|v| = \sqrt{v^T v}$; for a complex vector, we have $|v| = \sqrt{(v^*)^T v}$.*

Definition 1.7 (Eigenvalue, Eigenvector). *λ is an eigenvalue and v is an eigenvector if $Av = \lambda v$ and v is not the zero vector.*

Exercise 1.8. *If v is an eigenvector of A with eigenvalue λ, show $w = av$, $a \in \mathbb{C}$, is also an eigenvector of A with eigenvalue λ. Therefore, given an eigenvalue λ and an eigenvector v, one can always find an eigenvector w with the same eigenvalue, but $|w| = 1$.*

To find the eigenvalues, we solve the equation $\det(\lambda I - A) = 0$. This gives a polynomial $p(\lambda) = 0$. We call $p(\lambda)$ the **characteristic polynomial**.

Definition 1.9 (Degrees of freedom). *The number of degrees of freedom in a matrix is the number of elements needed to completely specify it; a general $n \times m$ real matrix has nm degrees of freedom.*

Exercise 1.10. *Show an $n \times n$ real symmetric matrix has $\frac{n(n+1)}{2}$ degrees of freedom, and determine the number of degrees of freedom of an $n \times n$ complex hermitian matrix.*

Exercise 1.11. *If A and B are symmetric, show AB is symmetric.*

2 Change of Basis

Given a matrix A, we call the element in the i^{th} row and j^{th} column a_{ij}. We represent a vector v as a column of elements with the i^{th} being v_i. A nice way to see matrix-vector multiplication is that the v_i give the *coefficients* by which the columns of A are linearly mixed together. For the product $w = Av$ to make sense, the length (dimension) of v must equal m, and the dimension of w will be n. A is therefore a linear transform from m-dimensional to n-dimensional space.

Multiple transformations appear written backwards: if we apply A then B then C to a vector, we write $w = CBAv$. Note that taking the product of two $n \times n$ matrices requires $O(n^3)$ effort.

Exercise 2.1. *Show that there are two ways to evaluate triple matrix products of the type CBA. The slow way involves $O(n^4)$ effort. How about the fast way? How do these results scale for the case of a product of k matrices?*

Definition 2.2 (Invertible Martices). *A is invertible if a matrix B can be found such that $BA = AB = I$. The inverse is then written $B = A^{-1}$.*

Exercise 2.3. *Prove if A is invertible, than A must be a square matrix.*

A matrix A is a linear transformation; to write it in matrix form requires us to choose a coordinate system (basis), and the transformation will look different in different bases. Consider the scalar quantity $x = w^T Av$, where A, v and w are written relative to a given basis, say $u_1, \ldots, u_n$. If M is an invertible matrix, we can write these quantities in a new basis, $Mu_1, \ldots, Mu_n$. We find $v' = Mv$ and $w' = Mw$. How does the matrix A look in the new basis?

For x to remain unchanged by the transformation (as any scalar must) for all choices of v and w requires that A become $A' = (M^T)^{-1}AM^{-1}$:

$$x' = w'^T A'v' = (Mw)^T((M^T)^{-1}AM^{-1})(Mv) = w^T IAIv = w^T Av = x. \tag{1}$$

This is a **similarity transformation**, and represents A in the new basis.

3 Orthogonal Matrices

Definition 3.1 (Orthogonal Matrices). *Q is an orthogonal $n \times n$ matrix if it has real entries and $Q^T Q = QQ^T = I$.*

Note Q is invertible, with inverse Q^T.

Exercise 3.2. *Show that the dot product is invariant under orthogonal transformation. That is, show that given two vectors, transforming them using the same orthogonal matrix leaves their dot product unchanged.*

Exercise 3.3. *Show the number of degrees of freedom in an orthogonal matrix is $\frac{n(n-1)}{2}$.*

The set of orthogonal matrices of order n forms a **continuous** or **topological group**, which we call $O(n)$. For the group properties:

- Associativity follows from that of matrix multiplication.
- The identity matrix acts as an identity element, since it is in the group.
- Inverse is the transpose (see above): $Q^{-1} = Q^T$.
- Closure is satisfied because any product of orthogonal matrices is itself orthogonal.

Exercise 3.4. *Prove the last assertion.*

For $n = 2$, a general orthogonal matrix can be written

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} \cos\theta & -\sin\theta \\ -\sin\theta & -\cos\theta \end{pmatrix}, \tag{2}$$

where $0 \leq \theta < 2\pi$ is a real angle. The determinant of the first is $+1$ and defines the **special** (*i.e.*, unit determinant) **orthogonal group** $SO(2)$ which is a subgroup of $O(2)$ with identity I. The second has determinant -1 and corresponds to rotations with a reflection; this subgroup is disjoint from $SO(2)$.

Note that $SO(2)$, alternatively written as the family of planar rotations $R(\theta)$, is **isomorphic** to the unit length complex numbers under the multiplication operation:

$$R(\theta) \longleftrightarrow e^{i\theta}; \tag{3}$$

(there is a bijection between the two sets). Therefore we have $R(\theta_1)R(\theta_2) = R(\theta_1 + \theta_2)$. This commutativity relation does *not* hold in higher $n > 2$.

Exercise 3.5. *In 3 dimensions a general rotation involves 3 angles (for example, azimuth, elevation, and roll). How many angles are needed in 4 dimensions? In 3 dimensions one rotates about a line (the set of points which do not move under rotation); what object do you rotate about in 4 dimensions?*

If an orthogonal matrix Q is used for conjugation of a general square matrix A, then the rule for transformation ((1)) becomes $A' = QAQ^T$.

Exercise 3.6. *Recall the Taylor expansions of* $\sin$ *and* $\cos$*:*

$$\cos(x) \;=\; \sum_{n=0}^{\infty} \frac{(-x)^{2n}}{(2n)!}, \quad \sin(x) \;=\; \sum_{n=0}^{\infty} \frac{-(-x)^{2n+1}}{(2n+1)!}. \tag{4}$$

For small x, we have

$$\cos(x) \;=\; 1 + O(x^2), \quad \sin(x) \;=\; x + O(x^3). \tag{5}$$

For small ϵ, show a rotation by ϵ is

$$\begin{pmatrix} \cos(\epsilon) & -\sin(\epsilon) \\ \sin(\epsilon) & \cos(\epsilon) \end{pmatrix} \;=\; \begin{pmatrix} 1 + O(\epsilon^2) & \epsilon + O(\epsilon^3) \\ -\epsilon + O(\epsilon^3) & 1 + O(\epsilon^2) \end{pmatrix}. \tag{6}$$

4 Trace

The **trace** of a matrix A, denote Trace(A) is the sum of the diagonal entries of A:

$$\text{Trace}(A) \;=\; \sum_{i=1}^{n} a_{ii}. \tag{7}$$

Lemma 4.1 (Eigenvalue Trace Formula). *Trace$(A) = \sum_{i=1}^{n} \lambda_i$, where the λ_is are the eigenvalues of A.*

Proof. The proof relies on writing out the characteristic equation and comparing powers of λ with the factorized version. As the polynomial has roots λ_i, we can write

$$\det(\lambda I - A) \;=\; p(\lambda) \;=\; \prod_{i=1}^{n} (\lambda - \lambda_i). \tag{8}$$

Note the coefficient of λ^n is 1, thus we have $\prod_i (\lambda - \lambda_i)$ and not $c \prod_i (\lambda - \lambda_i)$ for some constant c. By expanding out the RHS, the coefficient of λ^{n-1} is $-\sum_{i=1}^{n} \lambda_i$. Expanding the LHS, the lemma follows by showing the coefficient of λ^{n-1} in $\det(\lambda I - A)$ is $-$Trace(A). □

Lemma 4.2. $\det(A_1 \cdots A_m) = \det(A_1) \cdots \det(A_m)$. *Note* $\det(I) = 1$.

Corollary 4.3. *Trace(A) is invariant under rotation of basis: Trace$(Q^T AQ) =$ Trace(A).*

Exercise 4.4. *Prove the above corollary.*

5 Spectral Theorem for Real Symmetric Matrices

The main theorem we prove is

Theorem 5.1 (Spectral Theorem). *Let A be a real symmetric $n \times n$ matrix. Then there exists an orthogonal $n \times n$ matrix Q and a real diagonal matrix Λ such that $Q^T AQ = \Lambda$, and the n eigenvalues of A are the diagonal entries of Λ.*

This result is remarkable: any real symmetric matrix is diagonal when rotated into an appropriate basis. In other words, the operation of a matrix A on a vector v can be broken down into three steps:

$$Av = Q\Lambda Q^T v = (\text{undo the rotation})(\text{stretch along the coordinate axes})(\text{rotation})v. \tag{9}$$

We shall only prove the theorem in the case when the eigenvalues are distinct (note a generic matrix has n distinct eigenvalues, so this is not a particularly restrictive assumption). A similar theorem holds for complex hermitian matrices; the eigenvalues are again real, except instead of conjugating by an orthogonal matrix Q we must now conjugate by a unitary matrix U.

Remark 5.2. *The spectral theorem allows us to calculate the effect of high powers of a matrix quickly. Given a vector v, write v as a linear combination of the eigenvectors v_i: $v = \sum_i c_i v_i$. Then $A^m v = \sum c_i \lambda_i^m v_i$; this is significantly faster than calculating the entries of A^m.*

5.1 Preliminary Lemmas

For the Spectral Theorem we prove a sequence of needed lemmas:

Lemma 5.3. *The eigenvalues of a real symmetric matrix are real.*

Proof. Let A be a real symmetric matrix with eigenvalue λ and eigenvector v. Note that we do not yet know that v has only real coordinates. Therefore, $Av = \lambda v$. Take the dot product of both sides with the vector $(v^*)^T$, the complex conjugate transpose of v:

$$(v^*)^T Av = \lambda (v^*)^T v. \tag{10}$$

The two sides are clearly complex numbers. As A is real symmetric, taking the complex conjugate transpose of the left hand side of (10) gives $(v^*)^T Av^*$. Therefore, both sides of (10) are real. As $(v^*)^T v$ is real, we obtain λ is real. □

Lemma 5.4. *The eigenvectors of a real symmetric matrix are real.*

Proof. The eigenvectors solve the equation $(\lambda I - A)v = 0$. As $\lambda I - A$ is real, Gaussian elimination shows v is real. □

Lemma 5.5. *If λ_1 and λ_2 are two distinct eigenvalues of a real symmetric matrix A, then their corresponding eigenvectors are perpendicular.*

Proof. We study $v_1^T Av_2$. Now

$$v_1^T Av_2 \ = \ v_1^T (Av_2) \ = \ v_1^T (\lambda_2 v_2) \ = \ \lambda_2 v_1^T v_2. \tag{11}$$

Also,

$$v_1^T Av_2 \ = \ v_1^T A^T v_2 \ = \ (v_1^T A^T) v_2 \ = \ (Av_1)^T v_2 \ = \ (\lambda_1 v_1)^T v_2 \ = \ \lambda_1 v_1^T v_2. \tag{12}$$

Therefore

$$\lambda_2 v_1^T v_2 \ = \ \lambda_1 v_1^T v_2 \ \text{ or } \ (\lambda_1 - \lambda_2) v_1^T v_2 \ = \ 0. \tag{13}$$

As $\lambda_1 \neq \lambda_2$, $v_1^T v_2 = 0$. Thus, the eigenvectors v_1 and v_2 are perpendicular. □

5.2 Proof of the Spectral Theorem (Easy Case)

We prove the Spectral Theorem for real symmetric matrices **if there are n distinct eigenvectors**. Let λ_1 to λ_n be the n distinct eigenvectors, and let v_1 to v_n be the corresponding eigenvectors chosen so that each v_i has length 1. Consider the matrix Q, where the first column of Q is v_1, the second column of Q is v_2, all the way to the last column of Q which is v_n:

$$Q \;=\; \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ v_1 & v_2 & \cdots & v_n \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \tag{14}$$

The transpose of Q is

$$Q^T \;=\; \begin{pmatrix} \leftarrow & v_1 & \rightarrow \\ & \vdots & \\ \leftarrow & v_n & \rightarrow \end{pmatrix} \tag{15}$$

Exercise 5.6. *Show that Q is an orthogonal matrix. Use the fact that the v_i all have length one, and are perpendicular to each other.*

Consider $B = Q^TAQ$. To find its entry in the i^{th} row and j^{th} column, we look at

$$e_i^T B e_j \tag{16}$$

where the e_k are column vectors which are 1 in the k^{th} position and 0 elsewhere. Thus, we need only show that $e_i^T B e_j = 0$ if $i \neq j$ and equals λ_j if $i = j$.

Exercise 5.7. *Show $Qe_i = v_i$ and $Q^T v_i = e_i$.*

We calculate

$$\begin{aligned} e_i^T B e_j &= e_i^T Q^T A Q e_j \\ &= (e_i^T Q^T) A (Q e_j) \\ &= (Q e_i)^T A (Q e_j) \\ &= v_i^T A v_j \\ &= v_i^T (A v_j) \\ &= v_i^T \lambda_j v_j = \lambda_j v_i^T v_j. \end{aligned} \tag{17}$$

As $v_i^T v_j$ equals 0 if $i \neq j$ and 1 if $i = j$, this proves the claim.

Thus, the off-diagonal entries of Q^TAQ are zero, and the diagonal entries are the eigenvalues λ_j. This shows that Q^TAQ is a diagonal matrix whose entries are the n eigenvalues of A. □

Note that, in the case of n distinct eigenvalues, not only can we write down the diagonal matrix, we can easily write down what Q should be. Further, by reordering the columns of Q, we see we reorder the positioning of the eigenvalues on the diagonal.

Exercise 5.8. *Prove similar results for a complex hermitian matrix A. In particular, show the eigenvalues are real, and if the eigenvalues of A are distinct, then $A = U^* \Lambda U$ for a unitary U.*

6 Spectral Theorem for Real Symmetric Matrices (Harder Case)

Let A be a real symmetric matrix acting on $\mathbb{R}^n$. Then A has an orthonormal basis $v_1, \dots, v_n$ such that $Av_j = \lambda_j v_j$. We sketch the proof.

Write the inner or dot product $\langle v, w \rangle = v^t w$. As A is symmetric, $\langle Av, w \rangle = \langle v, Aw \rangle$.

Definition 6.1. $V^\perp = \{w : \forall v \in V, \langle w, v \rangle = 0\}$.

Lemma 6.2. *Suppose $V \subset \mathbb{R}^n$ is an* ***invariant vector subspace*** *under A ($if\, v \in V$, then $Av \in V$). Then $V^\perp$ is also A-invariant: $A(V^\perp) \subset V^\perp$.*

This proves the spectral theorem. Suppose we find a $v_0 \neq 0$ such that $Av_0 = \lambda_0 v_0$. Take $V = \{\mu v_0 : \mu \in \mathbb{R}\}$ for the invariant subspace.

By Lemma 6.2, $V^\perp$ is left invariant under A, and is one dimension less. Thus, by whatever method we used to find an eigenvector, we apply the same method on $V^\perp$.

Thus, all we must show is given an A-invariant subspace, there is an eigenvector. Consider

$$\max_{v \text{ with } \langle v,v \rangle = 1} \Big\{ \langle Av, v \rangle \Big\}. \tag{18}$$

Standard fact: every continuous function on a compact set attains its maximum (not necessarily uniquely). Note that the set of v such that $\langle v, v \rangle = 1$ is a compact set. See, for example, [**?**].

Let v_0 be a vector giving the maximum value, and denote this maximum value by λ_0. As $\langle v_0, v_0 \rangle = 1$, v_0 is not the zero vector.

Lemma 6.3. $Av_0 = \lambda_0 v_0$.

Clearly, if Av_0 is a multiple of v_0 it has to be λ_0 (from the definition of v_0 and λ_0). Thus, it is sufficient to show

Lemma 6.4. $\{\mu v_0 : \mu \in \mathbb{R}\}$ *is an A-invariant subspace.*

Proof. let w be an arbitrary vector perpendicular to v_0, and ϵ be an arbitrary small real number. Consider

$$\langle A(v_0 + \epsilon w), v_0 + \epsilon w \rangle \tag{19}$$

We need to renormalize, as $v_0 + \epsilon w$ is not unit length; it has length $1 + \epsilon^2 \langle w, w \rangle$. As v_0 was chosen to maximize $\langle Av, v \rangle$ subject to $\langle v, v \rangle = 1$, after normalizing the above cannot be larger. Thus,

$$\langle A(v_0 + \epsilon w), v_0 + \epsilon w \rangle \;=\; \langle Av_0, v_0 \rangle + 2\epsilon \langle Av_0, w \rangle + \epsilon^2 \langle w, w \rangle. \tag{20}$$

Normalizing the vector $v_0 + \epsilon w$ by its length, we see that in Equation 20, the order ϵ terms must be zero. Thus,

$$\langle Av_0, w \rangle \;=\; 0; \tag{21}$$

however, this implies Av_0 is in the space spanned by v_0 (as w was an arbitrary vector perpendicular to v_0), completing our proof. □

Corollary 6.5. *Any local maximum will lead to an eigenvalue-eigenvector pair.*

The second largest eigenvector (denoted λ_1) is

$$\lambda_1 \;=\; \max_{\langle v, v_0 \rangle = 0} \frac{\langle Av, v \rangle}{\langle v, v \rangle}. \tag{22}$$

We can either divide by $\langle v, v \rangle$, or restrict to unit length vectors.

References

[St] Strang, *Linear Algebra and Its Applications*, International Thomson Publishing, 3rd edition.

Index