Math Review Notes—Linear Algebra

Gregory Faletto

Contents

Last updated May 17, 2019

1 Linear Algebra

These are my notes from taking EE 588 at USC, Math 541A at USC, and various other sources which I mostly cite within the text.

1.1 Properties of Projection Matrices

i. Formula:

$$P = A(A^TA)^{-1}A^T$$

(Note that if A is an invertible (square) matrix, then $P = A(A^TA)^{-1}A^T = AA^{-1}(A^T)^{-1}A^T = I$.)

The projection matrix projects any vector b into the column space of A. In other words, $p = Pb$ is the component of b in the column space, and the error $e = b - Pb$ is the component in the orthogonal complement. ($I - P$ is also a projection matrix. It projects b onto the orthogonal complement, and the projection is $b - Pb = e$).

(Note that if A is an invertible (square) matrix, then its column space is all of $\mathbb{R}^n$, so b is already in the column space of A.)

ii. The projection matrix is **idempotent**: it equals its square–$P^2 = P$.

iii. The projection matrix is **symmetric**: it equals its transpose–$P^T = P$.

iv. Conversely, **any symmetric idempotent matrix represents a projection**. P is unique for a given subspace.

v. If A is an $m \times n$ matrix with rank n, then $\text{rank}(P) = n$. The eigenvalues of P consist of n ones and $m - n$ zeroes. P always contains n independent eigenvectors and is thus diagonalizable.

Suppose A is a square nonsingular matrix and λ is an eigenvalue of A. Then λ^{-1} is an eigenvalue of the matrix A^{-1}.

The trace of an idempotent matrix with rank r is r.

1.2 Eigenvalues, Eigenvectors, Diagonalization, Symmetric Matrices

Notes on Diagonalization

Suppose the $n \times n$ matrix A has n linearly independent eigenvectors. If these eigenvectors are the columns of a matrix S, then $S^{-1}AS$ is a diagonal matrix Λ. The eigenvalues of A are on the diagonal of Λ:

$$S^{-1}AS = \Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

We call S the **eigenvector matrix** and Λ the **eigenvalue matrix**.

1. If the matrix A has no repeated eigenvalues, then its n eigenvectors are automatically independent. Therefore **any matrix with n distinct eigenvalues can be diagonalized**.

2. **The diagonalizing matrix S is not unique**. An eigenvector x can be multiplied by a constant and remains an eigenvector. We can multiply the columns of S by any nonzero constants and produce a new diagonalizing S. Repeated eigenvalues leave even more freedom in S (columns with identical eigenvalues can be interchanged).

   (Note that for the trivial example $A = I$, any invertible S will do. $S^{-1}IS$ is always diagonal, and Λ is just I. **All vectors are eigenvectors of the identity.**)

3. **Other matrices S will not produce a diagonal** Λ. Since $\Lambda = S^{-1}AS$, S must satisfy $S\Lambda = AS$. Suppose the first column of S is y. Then the first column of $S\Lambda$ is $\lambda_1 y$. If this is to agree with the first column of AS, which by matrix multiplication is Ay, then y must be an eigenvector: $Ay = \lambda_1 y$.

   (Note that the *order* of the eigenvectors in S and the eigenvalues in Λ must match.)

4. Not all matrices posses n linearly independent eigenvectors, so **not all matrices are diagonalizable**.

   Diagonalizability of A depends on having enough (n) independent eigenvectors. Invertibility of A depends on having nonzero eigenvalues.

   There is no connection between diagonalizability (n independent eigenvectors) and invertibility (no zero eigenvalues). The only indication given by the eigenvalues is that diagonalization can fail only if there are repeated eigenvalues. (But even then, it does not always fail–e.g. I.)

   The test is to check, for an eigenvalue that is repeated p times, whether there are p independent eigenvectors–in other words, whether $A - \lambda$ has rank $n - p$.

5. **Projection matrices always contain n independent eigenvectors and thus are always diagonalizable**.

Eigenvalues of Symmetric Matrices: If A is symmetric, then it has the following properties:

1. A has exactly n (not necessarily distinct) eigenvalues

2. There exists a set of n eigenvectors, one for each eigenvalue, that are mutually orthogonal (even if the eigenvalues are not distinct).

Eigenvalues of the Inverse of a Matrix: Suppose A is a square nonsingular matrix and λ is an eigenvalue of A. Then λ^{-1} is an eigenvalue of the matrix A^{-1}. Proof: Note that since A is nonsingular, A^{-1} exists and λ is nonnegative for all eigenvalues of A. Let λ be an eigenvalue of A and let $x \neq 0$ be an eigenvector of A for λ. Suppose A is n by n. Then we have

$$A^{-1}x = A^{-1}\lambda^{-1}\lambda x = \lambda^{-1}A^{-1}\lambda x = \lambda^{-1}A^{-1}Ax = \lambda^{-1}x$$

The inverse of a symmetric matrix is symmetric. Proof: Let A be a symmetric matrix.

$$I = I'$$

$$AA^{-1} = (AA^{-1})'$$

$$A^{-1}A = (A^{-1})'A'$$

$$A^{-1}AA^{-1} = (A^{-1})'AA^{-1}$$

$$A^{-1} = (A^{-1})'$$

1.3 Positive Definite Matrices

For any real invertible matrix A, the product $A'A$ is a positive definite matrix. (Proof: Let z be a non-zero vector. We want $z'A'Az > 0 \forall z$. Note that $z'A'Az = (Az)'(Az)$. Because A is invertible and $z \neq 0$, $Az \neq 0$, so $(Az)'(Az) > 0$.)

Let $A \in \mathbb{R}^{m \times n}$ with $m \geq n$ and let $\text{rank}(A) = n$ (that is, A has full column rank). Then $A'A$ is a positive definite matrix. (Proof: Let z be a non-zero vector. We want $z'A'Az > 0 \forall z$. Note that $z'A'Az = (Az)'(Az)$. Because A has full column rank (and n linearly independent columns) and $z \neq 0$, $Az \neq 0$, so $(Az)'(Az) > 0$.)

Every positive definite matrix is invertible and its inverse is also positive definite.

1.4 Matrix Decompositions

Schur complement, Schur decomposition: For information, see Section **??**.

QR decompositon

Orthogonal Decomposition

Spectral Decomposition (eigenvalue decomposition

Generalized eigenvalue decomposition

Singular value decomposition and Pseudo-inverse

Jordan decomposition

Cholesky decomposition

1.5 Inverting Matrices

Theorem 1 (**Woodbury Matrix Identity** (or **Sherman-Morrison-Woodbury formula**))**.** For $A \in \mathbb{R}^{n \times n}$, $U \in \mathbb{R}^{n \times k}$, $C \in \mathbb{R}^{k \times k}$, and $V \in \mathbb{R}^{v \times n}$,

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}.$$

Theorem 2 (**Binomial Inverse Thereom**)**.**

1.6 Other

Frobenius norm

From appendix of Time Series:

Quadratic forms

Special matrices

Difference Equations

1.7 Practice Problems

[**The Power Method**] This exercise gives an algorithm for finding the eigenvectors and eigenvalues of a symmetric matrix. In modern statistics, this is often a useful thing to do. The Power Method described below is not the best algorithm for this task, but it is perhaps the easiest to describe and analyze.

Let A be an $n \times n$ real symmetric matrix. Let $\lambda_1 \geq \cdots \geq \lambda_n$ be the (unknown) eigenvalues of A, and let $v_1, \ldots, v_n \in \mathbb{R}^n$ be the corresponding (unknown) eigenvectors of A such that $|v_i| = 1$ and such that $Av_i = \lambda_i v_i$ for all $1 \leq i \leq n$.

Given A, our first goal is to find v_1 and λ_1. For simplicity, assume that $1/2 < \lambda_1 < 1$, and $0 \leq \lambda_n \leq \cdots \leq \lambda_2 < 1/4$. Suppose we have found a vector $v \in \mathbb{R}^n$ such that $|v| = 1$ and $|\langle v, v_1 \rangle| > 1/n$. Let k be a positive integer. Show that

$$A^k v$$

approximates v_1 well as k becomes large. More specifically, show that for all $k \geq 1$,

$$|A^k v - \langle v, v_1 \rangle \lambda_1^k v_1|^2 \leq \frac{n-1}{16^k}.$$

(Hint: use the spectral theorem for symmetric matrices.)

Solution. Since the eigenvectors for A are orthogonal, they form a basis for $\mathbb{R}^n$, so for any $v \in \mathbb{R}^n$ we have $v = \sum_{i=1}^n c_i v_i$ for some $c = (c_1, \ldots, c_n) \in \mathbb{R}^n$. It also follows then that $\langle v, v_1 \rangle = \langle \sum_{i=1}^n c_i v_i, v_1 \rangle = c_1 v_1' v_1 = c_1$. And finally, since $\|v\| = 1$ and $\|v_i\| = 1$ for all i, clearly we have $-1 \leq c_i \leq 1$. Using these facts, we have

$$\|A^k v - \langle v, v_1 \rangle \lambda_1^k v_1\|^2 = \|\sum_{i=1}^n \lambda_i^k c_i v_i - \langle v, v_1 \rangle \lambda_1^k v_1\|^2 = \|\sum_{i=1}^n \lambda_i^k c_i v_i - \lambda_1^k c_1 v_1\|^2 = \|\sum_{i=2}^n \lambda_i^k c_i v_i\|^2$$

$$= \sum_{i=2}^n \lambda_i^{2k} c_i^2 v_i' v_i = \sum_{i=2}^n \lambda_i^{2k} c_i^2$$

Since by assumption $0 \leq \lambda_n \leq \ldots \leq \lambda_2 \leq 1/4$, $\lambda_i^{2k} \leq 1/16^k$ for all i, so we have

$$\|A^k v - \langle v, v_1 \rangle \lambda_1^k v_1\|^2 \leq \frac{1}{16^k} \sum_{i=2}^n c_i^2$$

Since $-1 \leq c_i \leq 1 \implies 0 \leq c_i^2 \leq 1$, we have $\sum_{i=2}^n c_i^2 \leq n-1$, so this can be written as

$$\boxed{\|A^k v - \langle v, v_1\rangle \lambda_1^k v_1\|^2 \leq \frac{n-1}{16^k}}$$

Remark. Since $|\langle v, v_1\rangle|\lambda_1^k > 2^{-k}/n$, this inequality implies that $A^k v$ is approximately an eigenvector of A with eigenvalue λ_1. That is, by the triangle inequality,

$$|A(A^k v) - \lambda_1(A^k v)| \leq |A^{k+1}v - \langle v, v_1\rangle \lambda_1^{k+1} v_1| + \lambda_1 |\langle v, v_1\rangle \lambda_1^k v_1 - A^k v| \leq 2\frac{\sqrt{n-1}}{4^k}.$$

Moreover, by the reverse triangle inequality,

$$|A^k v| = |A^k v - \langle v, v_1\rangle \lambda_1^k v_1 + \langle v, v_1\rangle \lambda_1^k v_1| \geq \frac{1}{n}2^{-k} - \frac{\sqrt{n-1}}{4^k}.$$

If we take k to be large (say $k > 10\log n$), and if we define $z : equals A^k v$, then z is approximately an eigenvector of A, that is

$$|A\frac{A^k v||A^k v| - \lambda_1 \frac{A^k v}{|A^k v|}|}{\leq} 4n^{3/2}2^{-k} \leq 4n^{-4}.$$

And to approximately find the first eigenvalue λ_1, we simply compute

$$\frac{z^T A z}{z^T z}.$$

That is, we have approximately found the first eigenvector and eigenvalue of A.

To find the second eigenvector and eigenvalue, we can repeat the above procedure, where we start by choosing v such that $\langle v, v_1\rangle = 0$, $|v| = 1$ and $|\langle v, v_2\rangle| > 1/(10\sqrt{n})$. To find the third eigenvector and eigenvalue, we can repeat the above procedure, where we start by choosing v such that $\langle v, v_1\rangle = \langle v, v_2\rangle = 0$, $|v| = 1$ and $|\langle v, v_3\rangle| > 1/(10\sqrt{n})$. And so on.

Google's PageRank algorithm uses the power method to rank websites very rapidly. In particular, they let n be the number of websites on the internet (so that n is roughly 10^9). They then define an $n \times n$ matrix C where $C_{ij} = 1$ if there is a hyperlink between websites i and j, and $C_{ij} = 0$ otherwise. Then, they let B be an $n \times n$ matrix such that B_{ij} is 1 divided by the number of 1's in the i^{th} row of C, if $C_{ij} = 1$, and $B_{ij} = 0$ otherwise. Finally, they define

$$A = (.85)B + (.15)D/n$$

where D is an $n \times n$ matrix all of whose entries are 1.

The power method finds the eigenvector v_1 of A, and the size of the i^{th} entry of v_1 is proportional to the "rank" of website i.

12. Let A be a 2×2 matrix for which there is a constant k such that the sum of the entries in each row and each column is k. Which of the following must be an eigenvector of A ?

I. $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$

II. $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$

III. $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$

(A) I only (B) II only (C) III only (D) I and II only (E) I, II, and III

Solution 12. (C) This condition makes the matrix of the form

$$\begin{pmatrix} a & b \\ b & a \end{pmatrix}.$$

There is no reason that $a = 0$ or $b = 0$, so there is no reason $(1, 0)$ or $(0, 1)$ should be eigenvectors. But it is easy to verify that $(1, 1)$ must be.

24. Consider the system of linear equations

$$\begin{aligned} w + 3x + 2y + 2z &= 0 \\ w + 4x + y \phantom{{}+ 2z} &= 0 \\ 3w + 5x + 10y + 14z &= 0 \\ 2w + 5x + 5y + 6z &= 0 \end{aligned}$$

with solutions of the form (w, x, y, z), where w, x, y, and z are real. Which of the following statements is FALSE?

(A) The system is consistent.

(B) The system has infinitely many solutions.

(C) The sum of any two solutions is a solution.

(D) $(-5, 1, 1, 0)$ is a solution.

(E) Every solution is a scalar multiple of $(-5, 1, 1, 0)$.

Solution 24. (E) Looking at our answers, we can verify directly that $(-5, 1, 1, 0)$ is a solution. Any multiple of $(-5, 1, 1, 0)$ is also a solution, which shows that (A), (B), (C), and (D) are all true – leaving only (E). Another solution, for example, is $(0, 2, -8, 5)$

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 0 & 2 & 3 & 4 & 5 \\ 0 & 0 & 3 & 4 & 5 \\ 0 & 0 & 0 & 4 & 5 \\ 0 & 0 & 0 & 0 & 5 \end{pmatrix}$$

34. Which of the following statements about the real matrix shown above is FALSE?

(A) A is invertible.

(B) If $\mathbf{x} \in \mathbb{R}^5$ and $A\mathbf{x} = \mathbf{x}$, then $\mathbf{x} = \mathbf{0}$.

(C) The last row of A^2 is $(0 \quad 0 \quad 0 \quad 0 \quad 25)$.

(D) A can be transformed into the 5×5 identity matrix by a sequence of elementary row operations.

(E) $\det(A) = 120$

Solution 34. (B) An upper triangular matrix is easily verified to be invertible so long as its diagonal entries are all nonzero. Specifically, det A is still the product of its diagonal entries, so (E) and (D) and (A) are all true. (C) can easily be verified to be true by computing that the bottom-right corner is 25 (the product of upper triangular matrices still being upper triangular). This leaves (B). (B) can be checked directly to be false: if we let $x = (1, 0, 0, 0, 0)$, then $Ax = x$.

37. Let V be a finite-dimensional real vector space and let P be a linear transformation of V such that $P^2 = P$. Which of the following must be true?

I. P is invertible.

II. P is diagonalizable.

III. P is either the identity transformation or the zero transformation.

(A) None (B) I only (C) II only (D) III only (E) II and III

Solution 37. (C) $P^2 = P$ means that P is projection onto some subspace. There is no reason to believe that this should be invertible, but it should definitely be diagonalisable (with eigenbasis some basis of that subspace). III also need not be true if the subspace is anything proper or nontrivial.

50. Let A be a real 2×2 matrix. Which of the following statements must be true?

I. All of the entries of A^2 are nonnegative.

II. The determinant of A^2 is nonnegative.

III. If A has two distinct eigenvalues, then A^2 has two distinct eigenvalues.

(A) I only (B) II only (C) III only (D) II and III only (E) I, II, and III

Solution 50. (B) There is no reason that all the entires of A^2 need to be nonnegative. Its determinant must be nonnegative though: $\det(A^2) = (\det A)^2$. For III, suppose A is the diagonal matrix with entires $\pm\lambda$. Then those are its eigenvalues, and they are distinct so long as $\lambda \neq 0$. But A^2 has only one eigenvalue: λ^2.

51. Which of the following is an orthonormal basis for the column space of the real matrix $\begin{pmatrix} 1 & -1 & 2 & -3 \\ -1 & 1 & -3 & 2 \\ 2 & -2 & 5 & -5 \end{pmatrix}$?

(A) $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$

(B) $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$

(C) $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \\ 0 \end{pmatrix} \right\}$

(D) $\left\{ \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ -3 \\ 5 \end{pmatrix} \right\}$

(E) $\left\{ \begin{pmatrix} \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix} \right\}$

Solution 51. (E) The basis (C) is not orthogonal and (D) is not normal, so we can rule those out. We can throw out the first column, since it is the negation of the second. A little bit of math shows that the remaining 3×3 matrix has determinant 0, so the rank of our column space is 2. That leaves only (A) and (E), but (A) cannot be correct. Our column space contains vectors that have nonzero third entry, so cannot lie in the span of that basis.