1   Review of Linear Algebra

1. Observe that
$$(AB)^T = B^T A^T,$$

   since the i, j'th entry on the left is equal to the j, i'th entry of AB, which is the j'th row of A dot-producted with the i'th column of B. On the other hand, the (i, j)'th entry on the right is the i'th row of B^T dot producted with the j'th column of A^T. Since a row of X is the same as a column of X^T, the two are the same.

2. An implication is that zz^T is symmetric:

$$(zz^T)^T = (z^T)^T z^T = zz^T$$

3. Suppose you have matrix A with columns $a_1, \ldots, a_n$ and matrix B with rows $b_1^T, \ldots, b_n^T$. Then,

$$AB = \begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix} \begin{pmatrix} b_1^T \\ \vdots \\ b_n^T \end{pmatrix} = \sum_{k=1}^{n} a_k b_k^T$$

   Indeed, compare the (i, j)'th entry on both sides.

2   Symmetric and Definite Matrices

Definition 1. *Let A be a square $(n \times n)$* **symmetric** *matrix.*

(a) We say that A is **positive definite**, *and write $A > 0$, if $x^T A x > 0$, for every nonzero $x \in \mathbb{R}^n$.*

(b) We say that A is **nonnegative definite**, *and write $A \geq 0$, if $x^T A x \geq 0$, for every $x \in \mathbb{R}^n$.*

It is known (e.g., see any basic linear algebra text) that:

(a) The eigenvalues of a symmetric matrix are real.

Proof: Recall that given a complex number

$$z = a + bi,$$

we define its complex conjugate

$$\bar{z} = a - bi,$$

and that conjugation obeys the following laws:

$$
\begin{aligned}
\overline{z_1 + z_2} &= \bar{z}_1 + \bar{z}_2 \\
\overline{z_1 z_2} &= \bar{z}_1 \bar{z}_2
\end{aligned}
$$

The absolute value of a complex number can be written as

$$|z|^2 = z\bar{z}.$$

Now suppose

$$Ax = \lambda x.$$

Observe that this equation still holds if we replace x by multiple of x; thus, we can assume without loss of generality that $||x||_2 = 1$. Then,

$$\bar{x}^T A x = \lambda \bar{x}^T x = \lambda \sum_{i=1}^{n} |x_i|^2 = \lambda ||x||_2^2 = \lambda \tag{1}$$

Conjugating both sides of the equation,

$$\bar{\lambda} = \overline{\bar{x}^T A x} = x^T A \bar{x},$$

where we used the fact that the conjugate of a product/sum is the product/sum of the conjugates, and that A is a real matrix. But $\bar{\lambda}$ is a scalar, so we can transpose the expression for it and get the same result:

$$\bar{\lambda} = (x^T A \bar{x})^T = \bar{x}^T A^T x = \bar{x}^T A x \tag{2}$$

Comparing Eq. (1) with Eq. (2), we see that

$$\lambda = \bar{\lambda},$$

which means that λ is real.

(b) To each eigenvalue of a symmetric matrix, we can associate a real eigenvector. Eigenvectors associated with distinct eigenvalues are orthogonal; eigenvalues associated with repeated eigenvalues can always be taken to be orthogonal. Without loss of generality, all these eigenvectors can be normalized so that they have unit length, resulting in an orthonormal basis.

Proof: To see that eigenvectors can be taken to be real, take the real component of the equation

$$Ax = \lambda x$$

To show the orthogonality of eigenvectors, suppose v_1, v_2 are two eigenvectors corresponding to distinct eigenvalues of A. Then,

$$
\begin{aligned}
\lambda_1 v_1^T v_2 &= (Av_1)^T v_2 \\
&= v_1^T A^T v_2 \\
&= v_1^T A v_2 \\
&= \lambda_2 v_1^T v_2
\end{aligned}
$$

where we used the fact that $A = A^T$. Thus, if $\lambda_1 \neq \lambda_2$, it follows that $v_1 \cdot v_2 = 0$.

Thus, if A has n distinct eigenvalues, it has n real eigenvectors which are orthogonal; these eigenvectors are therefore a basis for $\mathbb{R}^n$.

We will skip the case of repeated eigenvalues.

(c) A positive definite matrix has n real and positive eigenvalues.

(d) A nonnegative definite matrix has n real and nonnegative eigenvalues.

Proof: Suppose

$$Ax = \lambda x,$$

Then, as we argued before, λ is real, and x can be taken to be real with $||x||_2 = 1$. Then,

$$x^T A x = \lambda x^T x = \lambda,$$

and the nonnegative (positive) definiteness of A implies that $\lambda \geq 0$ ($\lambda > 0$).

(e) The above essentially states that a symmetric definite matrix becomes diagonal after a suitable orthogonal change of basis.

Proof: Let $v_1, \ldots, v_n$ be the orthonormal eigenvectors of A corresponding to eigenvalues $\lambda_1, \ldots, \lambda_n$. Let U be the matrix whose i'th row is v_i. Then,

$$U^T U = I,$$

since the v_i are orthogonal. Moreover,

$$AU = U \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

Let us call the diagonal matrix on the right hand side D. We have:

$$U^T AU = D. \qquad (3)$$

Now consider the change of basis defined by $y = Ux$. Given $x = Uy$, we want to find a representation of Ax in the new coordinates, i.e. we are looking for z such that

$$AUy = Uz,$$

which means

$$U^T AUy = z,$$

or

$$Dy = z.$$

(f) The lecture notes make the following claim: as $||x||_2 \to \infty$, so does $x^T Ax$.

To prove this, decompose x as

$$x = a_1 v_1 + a_2 v_2 + \cdots a_n v_n,$$

where v_i is the eigenvector of A corresponding to λ_i. We assume that we numbered the eigenvalues so that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$.

Multiplying by A, we have

$$Ax = a_1 \lambda_1 v_1 + a_2 \lambda_2 v_2 + \cdots + \cdots \lambda_n a_n v_n,$$

which implies

$$x^T Ax = \lambda_1 a_1^2 + \lambda_2 a_2^2 + \cdots \lambda_n a_n^2$$

We lower bound this as

$$x^T Ax \geq \lambda_n(a_1^2 + a_2^2 + \cdots a_n^2) = \lambda_n ||x||_2^2,$$

so if the right-hand side approaches infinity, so does the left-hand side.

4

3 Square roots

Let us rewrite Eq. (3) as

$$A = UDU^T$$

Now observe that the columns of U are the eigenvectors v_i; and the rows of U^T are $\lambda_i v_i$. Therefore,

$$A = \sum_{i=1}^{n} \lambda_i v_i v_i^T.$$

For nonnegative definite matrices, we have $\lambda_i \geq 0$, which allows us to take square roots and define

$$B = \sum_{i=1}^{n} \sqrt{\lambda_i} v_i v_i^T.$$

We then observe that:

(a) The matrix B is symmetric.

Proof: Since $v_i v_i^T$ is symmetric, B is the sum of symmetric matrices and therefore symmetric.

(b) We have $B^2 = A$ (this is an easy calculation). Thus B is a **symmetric square root** of A.

Proof: We have that

$$B = U \hat{D} U^T,$$

where $\hat{D}$ denotes the diagonal matrix whose ii'th entry is $\sqrt{\lambda_i}$. Then,

$$B^2 = U \hat{D} U^T U \hat{D} U^T = U \hat{D}^2 U^T = U D U^T = A.$$

(c) The matrix B has eigenvalues $\sqrt{\lambda_i}$. Therefore, it is positive (respectively, nonnegative) definite if and only if A is positive (respectively, nonnegative) definite.

Proof: Since

$$B = U \hat{D} U^T,$$

it follows that

$$BU = U \hat{D},$$

which suggests that each column of U is an eigenvalue of B with eigenvector $\sqrt{\lambda_i}$.

4   Covariance Matrices

- The matrix $\text{Cov}(X, X)$ is nonnegative definite.

  Indeed,

  $$
  \begin{aligned}
  z^T \text{Cov}(X, X) z &= z^T E[(X - E[X])(X - E[X])^T] z \\
  &= E[z^T (X - E[X])(X - E[X])^T z]
  \end{aligned}
  $$

  Now define
  $$
  y = (X - E[X])^T z,
  $$
  and observe that we have shown that
  $$
  z^T \text{Cov}(X, X) z = E[y^T y] = E[\|y\|_2^2].
  $$

  Since the expectation of a nonnegative random variable is nonnegative, it follows that
  $$
  z^T \text{Cov}(X, X) z \geq 0,
  $$
  for all z.

- Suppose $Z = AY$. Then,
  $$
  \text{Cov}(Z, Z) = A\text{Cov}(Y, Y) A^T
  $$

  Indeed, since $E[Y] = AE[Z]$,
  $$
  Z - E[Z] = A(Y - E[Y]),
  $$

  and therefore

  $$
  \begin{aligned}
  \text{Cov}(Z, Z) &= E[(Z - E[Z])(Z - E[Z])^T) \\
  &= E[A(Y - E[Y])(Y - E[Y])^T A^T] \\
  &= A\text{Cov}(Y, Y) A^T
  \end{aligned}
  $$

- Suppose $Y_i, i = 1, \ldots, n$ are uncorrelated. What is the variance of $Z = a^T Y$, where a is some row vector?

  Since Z is a scalar, its variance is equal to its covariance matrix (which is of course 1×1). By the above formula,

  $$
  \sigma_Z^2 = a^T \sigma^2 I a = \sigma^2 \sum_{i=1}^{n} a_i^2.
  $$

5 Zero correlation versus independent

We have proved in lecture that if (X, Y) has a multivariate normal distribution, and if X is uncorrelated from Y, then X and Y are independent.

Consider now the following example. Let X be a standard mormal. Let U be a discrete random variable, with $\mathbb{P}(U = -1) = \mathbb{P}(U = 1) = 1/2$. Let $Y = UX$. It can be seen that the conditional distribution of Y, given either value of U, is $N(0, 1)$. Thus, the unconditional distribution of Y is also $N(0, 1)$. Furthermore, $\mathbf{E}[XY] = \mathbf{E}[UX^2] = \mathbf{E}[U]\,\mathbf{E}[X^2] = 0$, so that X and Y are uncorrelated. On the other hand, we always have $|Y| = |X|$, which shows that X and Y are not independent.

Is this example a contradiction of the earlier fact? No. The explanation is that in this example, X is normal, Y is normal, but (X, Y) is not multivariate normal. (The multivariate normal property is a property of the joint distribution; normality of the marginals is not enough.)

6 A generating function exercise

The transform associated with N, the total number of living groups contacted about the MIT blood drive, is

$$M_N(s) = \left(\frac{1}{3} + \frac{2}{3}e^s\right)^{10}.$$

(a) Determine the probability mass function (PMF) of N, i.e. $\mathbb{P}(N = n)$.

(b) Let the number K of people in any particular living group, be an independent random variable with associated transform

$$M_K(s) = \frac{\frac{1}{5}e^{4s}}{1 - \frac{4}{5}e^s}.$$

Find $p_K(k) = \mathbb{P}(K = k)$, $\mathbf{E}[K]$, and $\mathrm{var}(K)$.

(c) Let L be the total number of people whose living groups are contacted about the blood drive. Determine the transform, the mean, and the variance associated with L.

(d) Suppose that any particular person, whose living group is contacted, donates blood with probability $1/4$, and that all such individuals make their decisions independently. Let D denote the total number of blood donors from the contacted living groups. Calculate the transform and mean associated with D, and the probability that there will be no donors at all.

Solution: (a) From the transform tables, N is binomial with PMF

$$p_N(n) = \binom{10}{n} \left(\frac{2}{3}\right)^n \left(\frac{1}{3}\right)^{10-n}, \qquad n = 0, 1, \ldots, 10.$$

(b) The given transform $M_K(s)$ is e^{3s} times the transform associated with a geometric PMF with parameter $p = 1/5$. Thus $K = 3 + G$ where G is a geometric random variable with parameter $p = 1/5$. We have

$$p_K(k) = p(1-p)^{k-4} = \left(\frac{1}{5}\right) \left(\frac{4}{5}\right)^{k-4}, \qquad k = 4, 5, \ldots$$

$$\mathbf{E}[K] = 3 + \frac{1}{p} = 8, \qquad \mathrm{var}(K) = \frac{1-p}{p^2} = 20.$$

(c) L is the sum of a random number N of independent random variables each with transform $M_K(s)$. Hence $M_L(s)$ is obtained by replacing in $M_N(s)$ each occurrence of e^s by $M_K(s)$:

$$M_L(s) = \left(\frac{1}{3} + \frac{2}{3}M_K(s)\right)^{10} = \left(\frac{1}{3} + \frac{2}{3}\left(\frac{\frac{1}{5}e^{4s}}{1 - \frac{4}{5}e^s}\right)\right)^{10}.$$

We have

$$\mathbf{E}[L] = \mathbf{E}[K] \cdot \mathbf{E}[N] = 8 \cdot \left(10 \cdot \frac{2}{3}\right) = 53.33,$$

$$\mathrm{var}(L) = \mathrm{var}(K)\mathbf{E}[N] + (\mathbf{E}[K])^2 \, \mathrm{var}(N) = 20 \cdot \frac{20}{3} + 8^2 \cdot \left(10 \cdot \frac{2}{3} \cdot \frac{1}{3}\right) = \frac{2480}{9}.$$

(d) D is the sum of a random number L of random variables each of which is Bernoulli with parameter $1/4$. Hence $M_D(s)$ is obtained by replacing in $M_L(s)$ each occurrence of e^s by the Bernoulli transform $M_B(s) = 3/4 + (1/4)e^s$. So

$$M_D(s) = \left(\frac{1}{3} + \frac{2}{3}\left(\frac{\frac{1}{5}(\frac{3}{4} + \frac{1}{4}e^s)^4}{1 - \frac{4}{5}(\frac{3}{4} + \frac{1}{4}e^s)}\right)\right)^{10}.$$

We have

$$\mathbf{E}[D] = \mathbf{E}[L] \cdot \mathbf{E}[B] = 53.33 \cdot \frac{1}{4} = 13.33.$$

The probability $\mathbb{P}(D = 0)$ is obtained as

$$\lim_{s \to -\infty} M_D(s) = \left(\frac{1}{3} + \frac{1}{3}\left(\frac{3}{4}\right)^4\right)^{10}.$$

8

6.436J / 15.085J Fundamentals of Probability
Fall 2008