EE 520: Random Processes **Fall 2020**

Lecture 11
Random Vectors and Matrices

Instructor Name: John Lipor

Recommended Reading: Pishro-Nik: 6.1.1, 6.1.5; Gubner: 8.1 - 8.3

So far in this course, we have largely focused on collections of 1-3 RVs. When we have more than two but finitely many RVs, we collect them into vectors, which we call *random vectors* (RVecs). Before discussion properties of RVecs, we will first go through a brief review of linear algebra.

1 Matrix Operations

We write a matrix $A \in \mathbb{R}^{m \times n}$ as

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}.$$

Definition 1. The **transpose** of a matrix $A \in \mathbb{R}^{m \times n}$ is the $n \times m$ matrix $A^T \in \mathbb{R}^{n \times m}$ whose (i, j)th entry is the (j, i)th entry of A.

Example 1. Let

$$A = \begin{bmatrix} 1 & 0 \\ 2 & 3 \\ 1 & 2 \end{bmatrix}.$$

Then

$$A^T = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 3 & 1 \end{bmatrix}.$$

Definition 2. We say that A is **symmetric** if $A = A^T$.

Some properties of the transpose operation are as follows.

1. $\left(A^T\right)^T = A$
2. $(A + B)^T = A^T + B^T$
3. $(AB)^T = B^T A^T$ **if** $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times m}$ (otherwise $B^T A^T$ is not a valid operation)

1.1 Vector-Vector Multiplication

By convention, we assume any vector $x \in \mathbb{R}^n$ is a *column* vectors, i.e.,

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

We are interested in several multiplication operations between matrices and vectors.

Definition 3. The **dot product** or **inner product** between two vectors $x, y \in \mathbb{R}^n$ is

$$\langle x, y\rangle = x^T y = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i.$$

Definition 4. The **outer product** between two vectors $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$ is

$$xy^T = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \begin{bmatrix} y_1 & y_2 & \dots & y_n \end{bmatrix} = \begin{bmatrix} x_1 y_1 & x_1 y_2 & \dots & x_1 y_n \\ x_2 y_1 & x_2 y_2 & \dots & x_2 y_n \\ \vdots & & & \\ x_m y_1 & x_m y_2 & \dots & x_m y_n \end{bmatrix}.$$

Note that the inner product $x^T y$ is a scalar, while the outer product xy^T is a matrix of size $m \times n$.

1.2 Matrix-Vector Multiplication

We sometimes make use of Matlab notation when referring to elements of matrices. In this case, we use $A_{:,j}$ or $A_{:j}$ to denote the jth column of A and $A_{j,:}$ or $A_{j:}$ to denote the jth row of A.

Let $A \in \mathbb{R}^{m \times n}$ and $x \in \mathbb{R}^n$. We can think of matrix-vector multiplication in two ways. In the first, we view each element of the product Ax as an inner product between the corresponding row of A and the vector x

$$Ax = \begin{bmatrix} A_{1,:}x \\ A_{2,:}x \\ \vdots \\ A_{m,:}x \end{bmatrix}$$

Alternatively, we can consider the product Ax as a sum of scaled columns of A

$$Ax = A_{:,1}x_1 + A_{:,2}x_2 + \cdots + A_{:,n}x_n.$$

Deciding which view is most useful is a skill that is acquired over time, and I encourage you to begin by always writing both views when working on a problem.

1.3 Matrix-Matrix Multiplication

The standard definition of matrix-matrix multiplication defines the (i, j)th element of the product AB as

$$(AB)_{i,j} = \sum_{k=1}^{n} A_{ik} B_{kj},$$

where $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$. This is not an intuitive definition, so we instead consider two alternative views.

View 1: Inner product/Gram matrix. In this view, we think of A as a collection of row vectors and B as a collection of column vectors. Then AB is the matrix of inner products between the rows of A and the columns of B

$$AB = \begin{bmatrix} A_{1,:} \\ A_{2,:} \\ \vdots \\ A_{m,:} \end{bmatrix} \begin{bmatrix} B_{:,1} & B_{:,2} & \dots & B_{:,n} \end{bmatrix} = \begin{bmatrix} A_{1,:}B_{:,1} & \dots & A_{1,:}B_{:,n} \\ \vdots & & \\ A_{m,:}B_{:,1} & \dots & A_{m,:}B_{:,n} \end{bmatrix}.$$

When we take $B = A$, the matrix of inner products is called the *Gram matrix*.

View 2: Outer product/sample covariance matrix. In the second view, we think of A in terms of its columns and B in terms of its rows. In this case, the product AB is a sum of outer products

$$AB = \begin{bmatrix} A_{:,1} & A_{:,2} & \dots & A_{:,k} \end{bmatrix} \begin{bmatrix} B_{1,:} \\ B_{2,:} \\ \vdots \\ B_{k,:} \end{bmatrix} = \sum_{i=1}^{k} A_{:,i} B_{i,:}.$$

We will see later that this relates to the sample covariance matrix of random vectors.

1.4 More Matrix Properties

Definition 5. The **trace** of a square matrix $A \in \mathbb{R}^{n \times n}$ is the sum of its diagonal elements

$$\mathrm{tr}(A) = \sum_{i=1}^{n} a_{ii}.$$

Some facts and properties of the trace are below.

1. The trace is linear, i.e., $\mathrm{tr}\left(\alpha A + \beta B\right) = \alpha \mathrm{tr}(A) + \beta \mathrm{tr}(B)$ for scalars α, β.
2. The trace is invariant to cyclic permutations *but not all permutations*, i.e.,

$$\mathrm{tr}(ABC) = \mathrm{tr}(CAB) = \mathrm{tr}(BCA) \neq \mathrm{tr}(BAC).$$

3. $\mathrm{tr}(A) = \mathrm{tr}(A^T)$
4. The trace of a scalar is simply the scalar itself.

Definition 6. A square matrix $A \in \mathbb{R}^{n \times n}$ is **invertible** if there exists a matrix $A^{-1} \in \mathbb{R}^{n \times n}$ such that $AA^{-1} = A^{-1}A = I$, where I is the $n \times n$ identity matrix.

Definition 7. A square, symmetric matrix $A \in \mathbb{R}^{n \times n}$ is **positive semidefinite** (PSD) if

$$x^T A x \geq 0 \ \forall x \in \mathbb{R}^n$$

and is **positive definite** (PD) if

$$x^T A x > 0 \ \forall x \in \mathbb{R}^n, x \neq 0.$$

Fact 1. A matrix $A \in \mathbb{R}^{n \times n}$ is invertible if and only if A is PD.

We will sometimes be interested in the eigenvalues/eigenvectors of a matrix. These have a nice opaque definition that you probably learned poorly once. We're more interested in the eigenvalue decomposition, which exists under certain conditions.

Theorem 1 (Spectral Theorem)**.** If $A \in \mathbb{R}^{n \times n}$ is symmetric, then the following hold.

- The eigenvalues of A are all real.
- The eigenvectors of A form an orthonormal basis for $\mathbb{R}^n$, i.e., the matrix $V = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix}$ is such that $V^T V = V V^T = I$.
- A admits an eigenvalue decomposition

$$A = V \Lambda V^T,$$

where Λ is the diagonal matrix of the eigenvalues of A.

Moreover, if A is PSD, then the eigenvalues of A are also non-negative.

Definition 8. The Euclidean or **2-norm** of a vector $x \in \mathbb{R}^n$ is defined as

$$\|x\|_2 = \sqrt{x^T x} = \sqrt{\text{tr}(xx^T)}.$$

Note that the norm is a notion of length, so it is non-negative.

Definition 9. The **Cauchy-Schwarz inequality** for vectors states that

$$|\langle x, y \rangle| \leq \|x\| \, \|y\| \, ,$$

while for RVs (which form a vector space), we saw previously that

$$|\mathbb{E}\left[UV\right]| \leq \sqrt{\mathbb{E}\left[U^2\right]\mathbb{E}\left[V^2\right]}.$$

The RV version of this inequality follows from the first, since the inner product in the vector space of RVs is $\langle U, V \rangle = \mathbb{E}[UV]$.

2 Random Vectors and Matrices

Definition 10. A vector/matrix whose entries are RVs is called a **random vector/random matrix**.

Definition 11. The **expectation** of a RVec $X \in \mathbb{R}^n$, also known as the **mean vector**, is

$$\mathbb{E}[X] = \begin{bmatrix} \mathbb{E}\left[X_1\right] \\ \mathbb{E}\left[X_2\right] \\ \vdots \\ \mathbb{E}\left[X_n\right] \end{bmatrix}.$$

The expectation of a random matrix is similarly defined.

Now that we have random variables, vectors, and matrices, you need to be careful about the size of X. I will not use special notation to indicate vectors or matrices.

Fact 2. Let $X \in \mathbb{R}^{n \times m}$ be a random matrix, $A \in \mathbb{R}^{p \times n}$, $B \in \mathbb{R}^{m \times q}$, and $G \in \mathbb{R}^{p \times q}$ be fixed. Then

$$\mathbb{E}\left[AXB + G\right] = A\mathbb{E}[X]B + G.$$

Definition 12. The **correlation matrix** of a RVec $X \in \mathbb{R}^n$ is

$$R_X = \mathbb{E}\left[XX^T\right] = \mathbb{E}\begin{bmatrix} X_1^2 & \dots & X_1 X_N \\ \vdots & & \\ X_N X_1 & \dots & X_n^2 \end{bmatrix}.$$

Fact 3. Any correlation matrix R is symmetric and PSD.

Proof. Symmetry is obvious from the definition (try proving that XX^T is symmetric yourself). Let $a \in \mathbb{R}^n$ be fixed but arbitrary. Then

$$\begin{aligned} a^T R_X a &= a^T \mathbb{E}\left[XX^T\right] a \\ &= \mathbb{E}\left[a^T X X^T a\right] \\ &= \mathbb{E}\left[\left\|a^T X\right\|^2\right] \geq 0, \end{aligned}$$

where the last line follows since norms are non-negative. ☐

Definition 13. The **covariance matrix** of a RVec $X \in \mathbb{R}^n$ is

$$\begin{aligned} C_X &= \mathbb{E}\left[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T\right] \\ &= \mathbb{E}\left[XX^T\right] - (\mathbb{E}[X])(\mathbb{E}[X])^T. \end{aligned}$$

Definition 14. The **cross-correlation matrix** of a RVec $X \in \mathbb{R}^n$ is

$$R_{XY} = \mathbb{E}\left[XY^T\right]$$

and the **cross-covariance matrix** is

$$C_{XY} = \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])^T\right].$$

While R_X is always PSD, we note that R_{XY} may not be. In particular, X and Y may have different sizes, making R_{XY} not a square matrix. Looking back at the definition, we see that positive semidefiniteness is only a property of square, symmetric matrices.

2.1 Decorrelation of Random Vectors

Let $X \in \mathbb{R}^n$ have $\mathbb{E}[X] = 0$ and covariance C_X. We can "decorrelate" the elements of X by making their covariances zero, i.e., we want to find some $Y = f(X)$ such that

$$\mathbb{E}\left[Y_i^2\right] = \sigma_i^2 \text{ and } \mathbb{E}\left[Y_i Y_j\right] = 0 \;\forall i, j.$$

To find such an f, note that C_X is PSD (same proof as for R_X), so we can write

$$C_X = V\Lambda V^T \iff \Lambda = V^T C_X V.$$

Hence, taking $Y = V^T X$ gives

$$\begin{aligned} \mathbb{E}\left[YY^T\right] &= \mathbb{E}\left[\left(V^T X\right)\left(V^T X\right)^T\right] \\ &= \mathbb{E}\left[V^T X X^T V\right] \\ &= V^T \mathbb{E}\left[XX^T\right] V \\ &= V^T C_X V = \Lambda, \end{aligned}$$

which is a diagonal matrix as desired.