Lecture Notes: Random Matrices and Vectors
Expectations and Covariance

Advanced Regression Analysis Course

Based on Lecture Transcript

Administrative Announcements

Hello everyone, I hope you all had a restful break! Please take note of the following announcements:

- **Midterm Quiz Date:** The midterm quiz has been scheduled for **Monday, May 12th**.
- **Timing:** The quiz will take place during our regular class time (2 hours). It will replace the lecture for that day.
- **Content Coverage:** The quiz will cover material presented up to approximately one week before the quiz date. We will provide a precise list of topics closer to the date.
- **Quiz Structure:** More details about the format and structure of the quiz will be announced soon.
- **Practice Materials:** Examples and past quiz questions from previous years will be uploaded to Moodle to help you prepare and practice.
- **Lecture Notes:** I aim to upload updated lecture summaries for each week at the beginning of that week. However, I often make small corrections, updates, or improvements while preparing the actual lecture.
  - **Recommendation:** Please check for the latest version of the notes shortly before the lecture.
  - **Binding Material:** As always, the definitive material for the course is what is presented on the board during the lecture.
  - **Error Reporting:** If you spot any errors or typos in the posted notes, please let me know! I appreciate your help in keeping them accurate.
  - **Location:** Updated notes for the current week will be in the weekly folder on Moodle, and the complete, cumulatively updated notes are available in the general course folder.

1 Introduction and Motivation

We are embarking on a foundational segment that builds the algebraic tools necessary for a deeper understanding of statistical models, particularly the linear model we introduced earlier. Recall that the linear model posits a relationship where the expected value of a response variable Y, given predictors $X = (X_1, \ldots, X_p)$, is a linear function of those predictors (including an intercept term):

$$\mathbb{E}[Y|X] = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

Furthermore, we often make assumptions about the errors, $\epsilon_i = Y_i - \mathbb{E}[Y_i|X_i]$. For instance, we might assume:

- $\text{Cov}(\epsilon_i, \epsilon_j) = 0$ for $i \neq j$ (uncorrelated errors).

- $\text{Var}(\epsilon_i) = \sigma^2$ for all i (constant error variance, or homoscedasticity).

Working with multiple observations $(Y_1, X_1), \ldots, (Y_n, X_n)$ naturally leads to collections of variables. To analyze such models efficiently and elegantly, especially when deriving properties of estimators or test statistics, it's incredibly useful to employ matrix notation. This chapter focuses on developing the "algebra" of expectations and covariances for random vectors and matrices, which will provide the language and machinery needed for our subsequent statistical developments.

2 Random Matrices and Expectation

Definition 2.1 (Random Matrix). A **random matrix** Z of size $n \times m$ is simply an $n \times m$ matrix whose elements, Z_{ij}, are random variables defined on some common underlying probability space.

$$Z = \begin{pmatrix} Z_{11} & Z_{12} & \ldots & Z_{1m} \\ Z_{21} & Z_{22} & \ldots & Z_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{n1} & Z_{n2} & \ldots & Z_{nm} \end{pmatrix}$$

The elements Z_{ij} (for $1 \leq i \leq n, 1 \leq j \leq m$) possess a **joint probability distribution**. This means their behavior is potentially interconnected; they are generally *not* assumed to be independent unless explicitly stated.

Remark 2.2. The only difference between a random matrix and a standard matrix of constants is that the entries here are random variables rather than fixed numbers.

Definition 2.3 (Random Vector). A particularly important special case occurs when $m = 1$. An $n \times 1$ random matrix is called a **random vector** (of dimension n). We typically denote it as:

$$Z = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{pmatrix}$$

Note that we usually omit the second index (which is always 1) for vectors. Unless otherwise specified, vectors are assumed to be **column vectors**. A random vector Z resides in $\mathbb{R}^n$.

Definition 2.4 (Expectation of a Random Matrix). The **expectation** (or expected value) of an $n \times m$ random matrix Z, denoted $\mathbb{E}[Z]$, is the $n \times m$ matrix obtained by taking the expectation of each element individually, provided each expectation exists.

$$\mathbb{E}[Z] = \begin{pmatrix} \mathbb{E}[Z_{11}] & \mathbb{E}[Z_{12}] & \ldots & \mathbb{E}[Z_{1m}] \\ \mathbb{E}[Z_{21}] & \mathbb{E}[Z_{22}] & \ldots & \mathbb{E}[Z_{2m}] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}[Z_{n1}] & \mathbb{E}[Z_{n2}] & \ldots & \mathbb{E}[Z_{nm}] \end{pmatrix}$$

In shorthand, the (i, j)-th element of $\mathbb{E}[Z]$ is $(\mathbb{E}[Z])_{ij} = \mathbb{E}[Z_{ij}]$.

Remark 2.5. Note that $\mathbb{E}[Z]$ is a **constant** matrix (a matrix of fixed numbers), not a random matrix, assuming the expectations exist.

Example 2.6 (Height and Weight Vector). Consider the case where $n = 2$ and $m = 1$. Let Z_1 be the height of a randomly selected person and Z_2 be their weight. Then $Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$ is a random vector in $\mathbb{R}^2$.

- Z_1 itself is a random variable with some distribution (e.g., perhaps approximately normal).

- Z_2 itself is a random variable with some distribution.

- Crucially, Z_1 and Z_2 have a *joint distribution*. When we sample a person, we observe *both* their height and weight simultaneously. This joint behavior is described by the distribution of the vector Z.

- Height and weight are typically *not* independent. Knowing someone's height gives us information about their likely weight. This dependence is captured in the joint distribution.

This simple example illustrates the core idea: random vectors (and matrices) group together multiple random variables whose values may be related. Our goal in regression, fundamentally, is often to understand and model such relationships.

Remark 2.7 (Joint Distributions). The concept of a joint distribution for the $n \times m$ elements of a random matrix Z is a direct generalization of the joint distribution for two random variables (X, Y), or for the n components of a random vector $(Z_1, \ldots, Z_n)$. It describes the probability of the variables simultaneously taking values in certain ranges.

3 Properties of Expectation for Random Matrices

The expectation operator for random matrices inherits the linearity property from scalar expectations. Let Z and W be random matrices of the same size ($n \times m$), and let A and B be constant matrices (not random) of appropriate sizes such that the matrix multiplications below are defined.

Theorem 3.1 (Linearity of Expectation). *1. **Additivity:** $\mathbb{E}[Z + W] = \mathbb{E}[Z] + \mathbb{E}[W]$.*

*2. **Multiplication by Constants:** $\mathbb{E}[AZB] = A(\mathbb{E}[Z])B$.*

*3. **Affine Transformation:** If C is a constant matrix of the same size as Z, then $\mathbb{E}[Z + C] = \mathbb{E}[Z] + C$. More generally, $\mathbb{E}[AZ + C] = A\mathbb{E}[Z] + C$ (assuming dimensions match for AZ).*

Proof. We prove these properties by examining the (i, j)-th element.

1. **Additivity:** Let Z and W be $n \times m$ random matrices. The (i, j)-th element of $Z + W$ is $(Z + W)_{ij} = Z_{ij} + W_{ij}$. This is a sum of two scalar random variables.

$$
\begin{aligned}
(\mathbb{E}[Z + W])_{ij} &= \mathbb{E}[(Z + W)_{ij}] &&\text{(by definition of matrix expectation)} \\
&= \mathbb{E}[Z_{ij} + W_{ij}] &&\text{(by definition of matrix addition)} \\
&= \mathbb{E}[Z_{ij}] + \mathbb{E}[W_{ij}] &&\text{(by linearity of scalar expectation)} \\
&= (\mathbb{E}[Z])_{ij} + (\mathbb{E}[W])_{ij} &&\text{(by definition of matrix expectation)} \\
&= (\mathbb{E}[Z] + \mathbb{E}[W])_{ij} &&\text{(by definition of matrix addition)}
\end{aligned}
$$

Since this holds for all (i, j), we have $\mathbb{E}[Z + W] = \mathbb{E}[Z] + \mathbb{E}[W]$.

2. **Multiplication by Constants:** It suffices to show (a) $\mathbb{E}[AZ] = A\mathbb{E}[Z]$ and (b) $\mathbb{E}[ZB] = \mathbb{E}[Z]B$. The general result $\mathbb{E}[AZB] = A\mathbb{E}[Z]B$ follows by applying these two sequentially (e.g., $\mathbb{E}[AZB] = \mathbb{E}[A(ZB)] = A\mathbb{E}[ZB] = A(\mathbb{E}[Z]B) = A\mathbb{E}[Z]B$).

Let's prove (a) $\mathbb{E}[AZ] = A\mathbb{E}[Z]$. Assume A is $p \times n$ and Z is $n \times m$. Then AZ is $p \times m$.

$$
\begin{aligned}
(\mathbb{E}[AZ])_{ij} &= \mathbb{E}[(AZ)_{ij}] && \text{(def. matrix expectation)} \\
&= \mathbb{E}\left[\sum_{k=1}^{n} A_{ik} Z_{kj}\right] && \text{(def. matrix multiplication)} \\
&= \sum_{k=1}^{n} \mathbb{E}[A_{ik} Z_{kj}] && \text{(linearity of scalar expectation)} \\
&= \sum_{k=1}^{n} A_{ik} \mathbb{E}[Z_{kj}] && \text{(since } A_{ik} \text{ are constants)} \\
&= \sum_{k=1}^{n} A_{ik} (\mathbb{E}[Z])_{kj} && \text{(def. matrix expectation)} \\
&= (A(\mathbb{E}[Z]))_{ij} && \text{(def. matrix multiplication)}
\end{aligned}
$$

Since this holds for all (i, j), we have $\mathbb{E}[AZ] = A\mathbb{E}[Z]$.

The proof for (b) $\mathbb{E}[ZB] = \mathbb{E}[Z]B$ is analogous and left as an exercise. Assume Z is $n \times m$ and B is $m \times q$. Show that $(\mathbb{E}[ZB])_{ij} = (\mathbb{E}[Z]B)_{ij}$.

3. **Affine Transformation:** Let C be a constant $n \times m$ matrix.

$$
\begin{aligned}
\mathbb{E}[Z + C] &= \mathbb{E}[Z] + \mathbb{E}[C] && \text{(by Property 1)} \\
&= \mathbb{E}[Z] + C && \text{(since } \mathbb{E}[C_{ij}] = C_{ij} \text{ for constants)}
\end{aligned}
$$

The more general form $\mathbb{E}[AZ + C] = A\mathbb{E}[Z] + C$ follows similarly using Property 1 and Property 2(a). $\qquad\square$

Remark 3.2. These properties confirm that expectation behaves linearly even when applied to matrices and vectors, which is extremely convenient for algebraic manipulations.

4 Covariance Matrices

Just as expectation generalizes to matrices, so does the concept of variance and covariance.

4.1 Review: Covariance for Scalar Random Variables

Recall for scalar random variables Z and W with means $\mu_Z = \mathbb{E}[Z]$ and $\mu_W = \mathbb{E}[W]$:

- The **covariance** is $\mathrm{Cov}(Z, W) = \mathbb{E}[(Z - \mu_Z)(W - \mu_W)]$.

- A useful computational formula is $\mathrm{Cov}(Z, W) = \mathbb{E}[ZW] - \mu_Z \mu_W = \mathbb{E}[ZW] - \mathbb{E}[Z]\mathbb{E}[W]$.

  Derivation of formula. $\mathbb{E}[(Z - \mu_Z)(W - \mu_W)] = \mathbb{E}[ZW - Z\mu_W - \mu_Z W + \mu_Z \mu_W] = \mathbb{E}[ZW] - \mathbb{E}[Z\mu_W] - \mathbb{E}[\mu_Z W] + \mathbb{E}[\mu_Z \mu_W]$ (by linearity) $= \mathbb{E}[ZW] - \mu_W \mathbb{E}[Z] - \mu_Z \mathbb{E}[W] + \mu_Z \mu_W$ (constants factor out) $= \mathbb{E}[ZW] - \mu_W \mu_Z - \mu_Z \mu_W + \mu_Z \mu_W = \mathbb{E}[ZW] - \mu_Z \mu_W$. $\qquad\square$

- The **variance** is a special case: $\mathrm{Var}(Z) = \mathrm{Cov}(Z, Z) = \mathbb{E}[(Z - \mu_Z)^2] = \mathbb{E}[Z^2] - (\mathbb{E}[Z])^2$.

- Covariance is **symmetric**: $\mathrm{Cov}(Z, W) = \mathrm{Cov}(W, Z)$.

- Covariance is **bilinear**: It is linear in each argument when the other is held fixed. For constants a, b and random variables X, Y, Z:

  - $\mathrm{Cov}(aX + bY, Z) = a\mathrm{Cov}(X, Z) + b\mathrm{Cov}(Y, Z)$

$$- \mathrm{Cov}(X, aY + bZ) = a\mathrm{Cov}(X, Y) + b\mathrm{Cov}(X, Z)$$

(The second follows from the first property and symmetry). Also note $\mathrm{Cov}(X + c, Y) = \mathrm{Cov}(X, Y)$ if c is a constant.

4.2 The Covariance Matrix for Random Vectors

Now we generalize to random vectors. Let Z be an $n \times 1$ random vector with mean vector $\mu_Z = \mathbb{E}[Z]$ (which is $n \times 1$), and let W be an $m \times 1$ random vector with mean vector $\mu_W = \mathbb{E}[W]$ (which is $m \times 1$). Both Z and W are defined on the same probability space.

Definition 4.1 (Cross-Covariance Matrix). The **covariance matrix** (or **cross-covariance matrix**) between Z and W is the $n \times m$ matrix defined as:

$$\mathrm{Cov}(Z, W) = \mathbb{E}[(Z - \mu_Z)(W - \mu_W)^T]$$

Remark 4.2 (Understanding the Definition).
- $Z - \mu_Z$ is an $n \times 1$ random vector (mean zero).

- $W - \mu_W$ is an $m \times 1$ random vector (mean zero).

- $(W - \mu_W)^T$ is its transpose, a $1 \times m$ random row vector.

- The product $(Z - \mu_Z)(W - \mu_W)^T$ is an **outer product**. It results in an $n \times m$ **random matrix**.

$$\underbrace{(Z - \mu_Z)}_{n \times 1} \underbrace{(W - \mu_W)^T}_{1 \times m} = \begin{pmatrix} Z_1 - \mu_{Z_1} \\ \vdots \\ Z_n - \mu_{Z_n} \end{pmatrix} \begin{pmatrix} W_1 - \mu_{W_1} & \cdots & W_m - \mu_{W_m} \end{pmatrix}$$

The (i, j)-th element of this random matrix is $(Z_i - \mu_{Z_i})(W_j - \mu_{W_j})$.

- Taking the expectation $\mathbb{E}[\dots]$ applies element-wise to this $n \times m$ random matrix.

- Therefore, the (i, j)-th element of the resulting (constant) matrix $\mathrm{Cov}(Z, W)$ is:

$$(\mathrm{Cov}(Z, W))_{ij} = \mathbb{E}[(Z_i - \mu_{Z_i})(W_j - \mu_{W_j})] = \mathrm{Cov}(Z_i, W_j)$$

This gives a very intuitive understanding: the covariance matrix $\mathrm{Cov}(Z, W)$ simply collects all the pairwise scalar covariances between the components of Z and the components of W.

- The resulting matrix $\mathrm{Cov}(Z, W)$ is an $n \times m$ **constant** matrix.

Definition 4.3 (Variance-Covariance Matrix). A crucial special case arises when we consider the covariance of a random vector Z with itself ($W = Z$, so $m = n$). This is called the **variance-covariance matrix** (or often just the **covariance matrix**) of Z.

$$\mathrm{Cov}(Z) := \mathrm{Cov}(Z, Z) = \mathbb{E}[(Z - \mu_Z)(Z - \mu_Z)^T]$$

This is an $n \times n$ matrix.

- The (i, j)-th element (for $i \neq j$) is $(\mathrm{Cov}(Z))_{ij} = \mathrm{Cov}(Z_i, Z_j)$.

- The (i, i)-th (diagonal) element is $(\mathrm{Cov}(Z))_{ii} = \mathrm{Cov}(Z_i, Z_i) = \mathrm{Var}(Z_i)$.

So, the covariance matrix $\text{Cov}(Z)$ has the variances of the components of Z on its diagonal and the covariances between pairs of components off the diagonal.

$$\text{Cov}(Z) = \begin{pmatrix} \text{Var}(Z_1) & \text{Cov}(Z_1, Z_2) & \ldots & \text{Cov}(Z_1, Z_n) \\ \text{Cov}(Z_2, Z_1) & \text{Var}(Z_2) & \ldots & \text{Cov}(Z_2, Z_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(Z_n, Z_1) & \text{Cov}(Z_n, Z_2) & \ldots & \text{Var}(Z_n) \end{pmatrix}$$

Since $\text{Cov}(Z_i, Z_j) = \text{Cov}(Z_j, Z_i)$, the matrix $\text{Cov}(Z)$ is always **symmetric**.

Remark 4.4 (Cov(Z, W) vs Cov(W, Z))**.** Note that $\text{Cov}(Z, W)$ is $n \times m$, while $\text{Cov}(W, Z)$ is $m \times n$. They are generally not equal, but they are transposes of each other (see Property 1 below). $\text{Cov}(Z, W)$ relates components of Z to components of W. $\text{Cov}(Z)$ relates components of Z among themselves.

4.3   Computational Formula for Covariance Matrix

Similar to the scalar case, there's a useful formula analogous to $\mathbb{E}[ZW] - \mathbb{E}[Z]\mathbb{E}[W]$.

Theorem 4.5 (Computational Formula for Covariance Matrix)**.** *For random vectors Z ($n \times 1$) and W ($m \times 1$),*

$$\text{Cov}(Z, W) = \mathbb{E}[ZW^T] - \mu_Z \mu_W^T = \mathbb{E}[ZW^T] - (\mathbb{E}[Z])(\mathbb{E}[W])^T$$

Proof. We start from the definition and expand, using linearity of expectation for matrices.

$$\begin{aligned} \text{Cov}(Z, W) &= \mathbb{E}[(Z - \mu_Z)(W - \mu_W)^T] \\ &= \mathbb{E}[(Z - \mu_Z)(W^T - \mu_W^T)] && \text{(transpose property)} \\ &= \mathbb{E}[ZW^T - Z\mu_W^T - \mu_Z W^T + \mu_Z \mu_W^T] && \text{(matrix distributivity)} \\ &= \mathbb{E}[ZW^T] - \mathbb{E}[Z\mu_W^T] - \mathbb{E}[\mu_Z W^T] + \mathbb{E}[\mu_Z \mu_W^T] && \text{(linearity of } \mathbb{E}) \end{aligned}$$

Now consider the terms involving constant vectors/matrices μ_Z, μ_W^T:

- $\mathbb{E}[Z\mu_W^T] = (\mathbb{E}[Z])\mu_W^T = \mu_Z \mu_W^T$ (since μ_W^T is constant)

- $\mathbb{E}[\mu_Z W^T] = \mu_Z (\mathbb{E}[W^T])$ (since μ_Z is constant)

- $\mathbb{E}[\mu_Z \mu_W^T] = \mu_Z \mu_W^T$ (since $\mu_Z \mu_W^T$ is constant)

We also need the property that $\mathbb{E}[W^T] = (\mathbb{E}[W])^T = \mu_W^T$.

Proof. Proof that $\mathbb{E}[W^T] = (\mathbb{E}[W])^T$: We have $W^T = \begin{pmatrix} W_1 & W_2 & \ldots & W_m \end{pmatrix}$. Then $\mathbb{E}[W^T] = \begin{pmatrix} \mathbb{E}[W_1] & \mathbb{E}[W_2] & \ldots & \mathbb{E}[W_m] \end{pmatrix}$. Also, $\mathbb{E}[W] = \begin{pmatrix} \mathbb{E}[W_1] \\ \vdots \\ \mathbb{E}[W_m] \end{pmatrix}$, so $(\mathbb{E}[W])^T = \begin{pmatrix} \mathbb{E}[W_1] & \ldots & \mathbb{E}[W_m] \end{pmatrix}$. Comparing the results, we see that $\mathbb{E}[W^T] = (\mathbb{E}[W])^T$. $\qquad \square$

Substituting back:

$$\begin{aligned} \text{Cov}(Z, W) &= \mathbb{E}[ZW^T] - \mu_Z \mu_W^T - \mu_Z (\mathbb{E}[W^T]) + \mu_Z \mu_W^T \\ &= \mathbb{E}[ZW^T] - \mu_Z \mu_W^T - \mu_Z \mu_W^T + \mu_Z \mu_W^T \\ &= \mathbb{E}[ZW^T] - \mu_Z \mu_W^T \end{aligned}$$

This formula expresses the $n \times m$ covariance matrix in terms of the expectation of the $n \times m$ outer product ZW^T and the $n \times m$ outer product of the mean vectors $\mu_Z \mu_W^T$. $\qquad \square$

5 Properties of Covariance Matrices

Covariance matrices obey several important algebraic properties, extending the bilinearity of scalar covariance. Let Z, W, R be random vectors, and A, B be constant matrices, and a be a constant vector, all with compatible dimensions. Let $\mu_Z = \mathbb{E}[Z]$, etc.

Theorem 5.1 (Properties of Covariance Matrices).   *1. **Transpose Property:** $\text{Cov}(Z, W) = (\text{Cov}(W, Z))^T$.*

2. ***Additivity (in first argument):*** $\text{Cov}(Z + R, W) = \text{Cov}(Z, W) + \text{Cov}(R, W)$. *(Assumes Z, R have same dimension).*

3. ***Linear Transformations:*** $\text{Cov}(AZ, BW) = A\text{Cov}(Z, W)B^T$.

4. ***Variance of Linear Transformation:*** *As a special case of (3) with $W = Z$ and $B = A$, $\text{Cov}(AZ) = A\text{Cov}(Z)A^T$.*

5. ***Variance of a Linear Combination:*** *For a constant vector a, $a^T Z$ is a scalar random variable, and its variance is $\text{Var}(a^T Z) = a^T \text{Cov}(Z)a$.*

6. ***Positive Semi-Definiteness:*** *The variance-covariance matrix $\text{Cov}(Z)$ is always symmetric and positive semi-definite (PSD).*

Proof. 1. **Transpose Property:**

$$
\begin{aligned}
\text{Cov}(W, Z) &= \mathbb{E}[(W - \mu_W)(Z - \mu_Z)^T] \\
(\text{Cov}(W, Z))^T &= (\mathbb{E}[(W - \mu_W)(Z - \mu_Z)^T])^T \\
&= \mathbb{E}[((W - \mu_W)(Z - \mu_Z)^T)^T] && (\text{using } (\mathbb{E}[M])^T = \mathbb{E}[M^T]) \\
&= \mathbb{E}[((Z - \mu_Z)^T)^T (W - \mu_W)^T] && (\text{using } (XY)^T = Y^T X^T) \\
&= \mathbb{E}[(Z - \mu_Z)(W - \mu_W)^T] && (\text{using } (M^T)^T = M) \\
&= \text{Cov}(Z, W)
\end{aligned}
$$

2. **Additivity:** We show this element-wise. Let Z, R be $n \times 1$ and W be $m \times 1$.

$$
\begin{aligned}
(\text{Cov}(Z + R, W))_{ij} &= \text{Cov}((Z + R)_i, W_j) && (\text{def. cov matrix element}) \\
&= \text{Cov}(Z_i + R_i, W_j) && (\text{def. vector addition}) \\
&= \text{Cov}(Z_i, W_j) + \text{Cov}(R_i, W_j) && (\text{linearity of scalar cov}) \\
&= (\text{Cov}(Z, W))_{ij} + (\text{Cov}(R, W))_{ij} && (\text{def. cov matrix element}) \\
&= (\text{Cov}(Z, W) + \text{Cov}(R, W))_{ij} && (\text{def. matrix addition})
\end{aligned}
$$

Since this holds for all (i, j), the matrix identity follows. (Linearity in the second argument follows using Property 1).

3. **Linear Transformations:** Let Z be $n \times 1$, W be $m \times 1$, A be $p \times n$, B be $q \times m$. Then AZ is $p \times 1$ and BW is $q \times 1$. $\text{Cov}(AZ, BW)$ will be $p \times q$. We use the definition and linearity of expectation. Note $\mathbb{E}[AZ] = A\mu_Z$ and $\mathbb{E}[BW] = B\mu_W$.

$$
\begin{aligned}
\text{Cov}(AZ, BW) &= \mathbb{E}[(AZ - \mathbb{E}[AZ])(BW - \mathbb{E}[BW])^T] \\
&= \mathbb{E}[(AZ - A\mu_Z)(BW - B\mu_W)^T] \\
&= \mathbb{E}[A(Z - \mu_Z)(B(W - \mu_W))^T] \\
&= \mathbb{E}[A(Z - \mu_Z)(W - \mu_W)^T B^T] && (\text{transpose property } (XY)^T = Y^T X^T) \\
&= A\left(\mathbb{E}[(Z - \mu_Z)(W - \mu_W)^T]\right) B^T && (\text{linearity } \mathbb{E}[CXD] = C\mathbb{E}[X]D) \\
&= A\text{Cov}(Z, W)B^T
\end{aligned}
$$

(The element-wise proof presented in the lecture involves manipulating sums $\sum_k A_{ik} Z_k$ and $\sum_l B_{jl} W_l$ within the scalar covariance definition $\mathrm{Cov}((\ldots)_i, (\ldots)_j)$, applying bilinearity, and then carefully recognizing the resulting double summation $\sum_k \sum_l A_{ik} \mathrm{Cov}(Z_k, W_l) B_{jl}$ as the (i,j)-th element of the matrix product $A\mathrm{Cov}(Z,W)B^T$. While more tedious, it reinforces the element-wise definitions.)

4. **Variance of Linear Transformation:** This follows directly from Property 3 by setting $W = Z$ and $B = A$.

$$\mathrm{Cov}(AZ) = \mathrm{Cov}(AZ, AZ) = A\mathrm{Cov}(Z, Z)A^T = A\mathrm{Cov}(Z)A^T$$

5. **Variance of Linear Combination:** Let a be an $n \times 1$ constant vector. Then $a^T Z$ is a 1×1 matrix (a scalar). Applying Property 4 with $A = a^T$:

$$\mathrm{Cov}(a^T Z) = a^T \mathrm{Cov}(Z)(a^T)^T = a^T \mathrm{Cov}(Z)a$$

Since variance of any (scalar) random variable $Y = a^T Z$ must be non-negative ($\mathrm{Var}(Y) \geq 0$), it follows that $x^T \mathrm{Cov}(Z)x \geq 0$ for all vectors x (of compatible dimension). This is the definition of $\mathrm{Cov}(Z)$ being positive semi-definite. We already noted $\mathrm{Cov}(Z)$ is symmetric. $\qquad\square$

Exercise 5.2. Prove Property 3, $\mathrm{Cov}(AZ, BW) = A\mathrm{Cov}(Z, W)B^T$, using the computational formula $\mathrm{Cov}(X, Y) = \mathbb{E}[XY^T] - \mathbb{E}[X](\mathbb{E}[Y])^T$. (Hint: Apply the formula with $X = AZ$ and $Y = BW$, use linearity of $\mathbb{E}$, and properties of transpose).

6 Concluding Remarks

We have established the basic definitions and properties of expectation and covariance for random vectors and matrices. These tools, particularly the linearity of expectation and the transformation properties of covariance matrices, are essential for manipulating and understanding the statistical models we will work with, especially the multivariate linear model. They provide a compact and powerful notation for expressing complex relationships and derivations. We will apply these concepts extensively in the upcoming lectures.