# Review of Matrix Algebra for Regression
Recitation, PS 7552

Drew Rosenberg

# 1 Rules for Matrices

## 1.1 Addition, Subtraction

- To add or subtract two matrices, they must have the same dimensions.
- To add (subtract) two matrices, take the sum (difference) element-wise. Where $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$, their sum is given by

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} (a_{11} + b_{11}) & (a_{12} + b_{12}) \\ (a_{21} + b_{21}) & (a_{22} + b_{22}) \end{bmatrix}.$$

  (The same goes for subtraction.)

- Matrix addition is commutative:

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$$

- Matrix addition is associative:

$$(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C}) = \mathbf{A} + \mathbf{B} + \mathbf{C}$$

## 1.2 Multiplication

- To multiply two matrices, they must be conformable, which means that the number of columns of the matrix that is premultiplied must be equal to the number of rows in the matrix that is postmultiplied. For instance, if $\dim(\mathbf{A}) = m \times n$ ($m$ rows and $n$ columns), if you are premultiplying $\mathbf{B}$ by $\mathbf{A}$, i.e., multiplying $\mathbf{AB}$, then $\dim(\mathbf{B}) = n \times p$.
- Where $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix}$, their product is given by

$$\mathbf{AB} = \begin{bmatrix} (a_{11}b_{11} + a_{12}b_{21}) & (a_{11}b_{12} + a_{12}b_{22}) & (a_{11}b_{13} + a_{12}b_{23}) \end{bmatrix}.$$

- Where $x$ is a scalar, $x\mathbf{A} = \begin{bmatrix} a_{11}x & a_{12}x \end{bmatrix}$.
- For two vectors $u$ and $v$ of equal length $n$, their dot product, or inner product returns a single number: $uv = [u_1v_1 + u_2v_2 + \ldots + u_nv_n] = \sum_{i=1}^{n} u_iv_i$.

- For two vectors $u$ and $v$ of equal length $n$, their outer product returns an $n \times n$ matrix.
- Matrix multiplication is associative:

$$(\mathbf{AB})\,\mathbf{C} = \mathbf{A}\,(\mathbf{BC}) = \mathbf{ABC}$$

- Matrix multiplication is distributive:

$$\mathbf{A}\,(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$$

  (The same goes for postmultiplication.)

- Matrix multiplication is not commutative (except for multiplication with scalars):

$$\mathbf{AB} \neq \mathbf{BA}$$

- Where $\mathbf{I}$ is the identity matrix, given by

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

  for a $3 \times 3$ matrix,

$$\mathbf{AI} = \mathbf{IA} = \mathbf{A}$$

  .

## 1.3 Transposes

- Where $\mathbf{A}$ is the following matrix

$$\mathbf{A} = \begin{bmatrix} a_{11} & \dots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nk} \end{bmatrix},$$

  its transpose $\mathbf{A}^\intercal$ is given by

$$\mathbf{A}^\intercal = \begin{bmatrix} a_{11} & \dots & a_{n1} \\ \vdots & \ddots & \vdots \\ a_{1k} & \dots & a_{nk} \end{bmatrix}.$$

- Where $\mathbf{B}$ is conformable to $\mathbf{A}$, then the following properties hold:

$$(\mathbf{A}^\intercal)^\intercal = \mathbf{A}$$
$$(\mathbf{A} + \mathbf{B})^\intercal = \mathbf{A}^\intercal + \mathbf{B}^\intercal$$
$$(\mathbf{AB})^\intercal = \mathbf{B}^\intercal\mathbf{A}^\intercal$$

- Also, $\mathbf{AA}^\intercal$ and $\mathbf{A}^\intercal\mathbf{A}$ are symmetric matrices.

## 1.4 Inverses, Determinants, and Rank

- If $\mathbf{A}$ is a square matrix ($n \times n$) and it is nonsingular, i.e. it has an inverse $\mathbf{A}^{-1}$, then $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$.
- Inverse matrices are useful for solving systems of equations. Where $\mathbf{A}$ is an $n \times k$ matrix, $\mathbf{x}$ is a $k \times 1$ column vector of unknowns, and $\mathbf{y}$ is an $n \times 1$ column vector, we can use $\mathbf{A}^{-1}$ to solve for $\mathbf{x}$:

$$\begin{aligned} \mathbf{A}\mathbf{x} &= \mathbf{y} \\ \mathbf{x} &= \mathbf{A}^{-1}\mathbf{y} \end{aligned}$$

- If the $\mathbf{A}$ has a vanishing determinant, i.e. $|\mathbf{A}| = 0$, then $\mathbf{A}$ is singular, $\mathbf{A}$ it can not be inverted, and a unique solution to $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$ does not exist.
- The rank of a matrix is the maximum number of linearly independent rows or columns (whichever is fewer) in the matrix. If your $\mathbf{X}$ matrix of observations is not of full rank, then one of the columns is a linear combination of the other columns.
- If a matrix is not of full rank, its determinant will be zero.
- For a $2 \times 2$ matrix $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, $\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$, where $|\mathbf{A}| = ad - bc$.

## 1.5 Calculus

- If $\mathbf{A}$ is an $n \times k$ matrix and $\mathbf{x}$ is a $k \times 1$ vector, then:

$$\begin{aligned} \frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} &= \mathbf{A}^\intercal \\ \frac{\partial \mathbf{x}^\intercal \mathbf{A}}{\partial \mathbf{x}} &= \mathbf{A} \\ \frac{\partial \mathbf{x}^\intercal \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} &= \left(\mathbf{A} + \mathbf{A}^\intercal\right)\mathbf{x} \end{aligned}$$

- And if $\mathbf{A}$ is symmetric, then

$$\frac{\partial \mathbf{x}^\intercal \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x}$$

## 1.6 The Linear Model

- $\mathbf{Q} = (\mathbf{X}^\intercal\mathbf{X})$ What is this? (cross-product of the X's)
- $\mathbf{A} = \mathbf{Q}^{-1}\mathbf{X}^\intercal$ is the coefficient maker ($\mathbf{A}\mathbf{y} = \hat{\boldsymbol{\beta}}$)
- $\mathbf{N} = \mathbf{X}\mathbf{A}$ is the projection matrix ($\mathbf{N}\mathbf{y} = \hat{\mathbf{y}}$)
- $\mathbf{M} = \mathbf{I}_n - \mathbf{N}$ is the residual maker ($\mathbf{M}\mathbf{y} = \hat{\boldsymbol{\varepsilon}}$)

# 2 Applying What We Know

1. Deriving parameter estimates for the OLS multivariate case.

$$
\begin{aligned}
\varepsilon_i &= y_i - \mathbf{x}_i\boldsymbol{\beta} && (1)\\
\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \varepsilon_i^2 &= \min_{\boldsymbol{\beta}} \ \boldsymbol{\varepsilon}^\intercal\boldsymbol{\varepsilon} && (2)\\
&= \min_{\boldsymbol{\beta}} \ \left[(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^\intercal (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right] && (3)\\
&= \min_{\boldsymbol{\beta}} \ \left[(\mathbf{y}^\intercal-\boldsymbol{\beta}^\intercal\mathbf{X}^\intercal) (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right] && (4)\\
&= \min_{\boldsymbol{\beta}} \ \left[\mathbf{y}^\intercal\mathbf{y}-\boldsymbol{\beta}^\intercal\mathbf{X}^\intercal\mathbf{y}-\mathbf{y}^\intercal\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\beta}^\intercal\mathbf{X}^\intercal\mathbf{X}\boldsymbol{\beta}\right] && (5)\\
\frac{\partial \text{RSS}}{\partial \boldsymbol{\beta}} &= -\mathbf{X}^\intercal\mathbf{y} - (\mathbf{y}^\intercal\mathbf{X})^\intercal + (\mathbf{X}^\intercal\mathbf{X})\,\boldsymbol{\beta} + (\mathbf{X}^\intercal\mathbf{X})^\intercal\,\boldsymbol{\beta} && (6)\\
&= -\mathbf{X}^\intercal\mathbf{y} - \mathbf{X}^\intercal\mathbf{y} + \mathbf{X}^\intercal\mathbf{X}\boldsymbol{\beta} + \mathbf{X}^\intercal\mathbf{X}\boldsymbol{\beta} && (7)\\
&= -2\mathbf{X}^\intercal\mathbf{y} + 2\mathbf{X}^\intercal\mathbf{X}\boldsymbol{\beta} && (8)\\
0 &= -2\mathbf{X}^\intercal\mathbf{y} + 2\mathbf{X}^\intercal\mathbf{X}\boldsymbol{\beta} && (9)\\
2\mathbf{X}^\intercal\mathbf{y} &= 2\mathbf{X}^\intercal\mathbf{X}\boldsymbol{\beta} && (10)\\
(\mathbf{X}^\intercal\mathbf{X})^{-1}\,\mathbf{X}^\intercal\mathbf{y} &= \boldsymbol{\beta} && (11)
\end{aligned}
$$

- In equation 1, is $\mathbf{x}_i$ a column or row vector? How can you tell?

- In equation 2, why do we write $\boldsymbol{\varepsilon}^\intercal\boldsymbol{\varepsilon}$? What is this calculation doing? What are its dimensions?

- What are the dimensions of the quantity in equation 3?

- How did we get from equation 3 to equation 4?

- What rule allowed us to get from equation 4 to equation 5?

- Explain how we got each term in equation 6. How do we know that the matrices in the terms in equation 6 are conformable?

- What rules allowed us to go from equation 6 to equation 7?
- How did we get equation 8?
- Why did we set equation 9 equal to zero?
- How did we get from equation 9 to equation 10?
- What rule did we use to get from equation 10 to equation 11?

2. Where $\mathbf{X} = \begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix}$ and $\mathbf{y}^\intercal = \begin{bmatrix} 0.3 & -0.5 \end{bmatrix}$, find $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal\mathbf{y}$.
   (a) What is $\mathbf{X}^\intercal$?
   (b) What is $\mathbf{X}^\intercal\mathbf{X}$?
   (c) What is $(\mathbf{X}^\intercal\mathbf{X})^{-1}$?
   (d) What is $(\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal$?
   (e) What is $(\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal\mathbf{y}$?