# Scientific Computing for Biologists

## Introduction to matrices

Instructor: Paul M. Magwene

# Overview of Lecture

- Introduction to Matrices
  - Matrices as collections of vectors
  - Special matrices
- Matrix operations
  - Matrix addition, subtraction
  - Matrix multiplication
  - Transpose
  - More special matrices
- Matrices concepts
- Linear dependence/independence
- Matrix inverses
- Solving simultaneous linear equations
- Multiple regression

# Introduction to Matrices

- One way to think about a matrix is as a collection of vectors. This is, in essence, what a multivariate data set is.
- A matrix which has $n$ rows and $p$ columns will be referred to as a $n \times p$ matrix. $n \times p$ is the shape of the matrix.

$$\underset{(n\times p)}{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{bmatrix}$$

# Special Matrices

- Zero matrix

$$\mathbf{0} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

- Square matrix
  A matrix whose shape is is $n \times n$

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

- Ones matrix

$$\mathbf{1} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}$$

- Diagonal matrix
  A square matrix where the off-diagonal elements are zero.

$$A = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}$$

# Scalar Multiplication of a Matrix

- Let $k$ be a scalar and let $A$ be the matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{bmatrix}$$

- then

$$kA = \begin{bmatrix} ka_{11} & ka_{12} & \cdots & ka_{1p} \\ ka_{21} & ka_{22} & \cdots & ka_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ ka_{n1} & ka_{n2} & \cdots & ka_{np} \end{bmatrix}$$

# Addition and Subtraction of Matrices

- Let $A$ and $B$ be matrices that have the same shape, $n \times p$:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{bmatrix} \quad B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{np} \end{bmatrix}$$

- then

$$A + B = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1p} + b_{1p} \\ a_{21} + b_{11} & a_{22} + b_{22} & \cdots & a_{2p} + b_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} + b_{n1} & a_{n2} + b_{n2} & \cdots & a_{np} + b_{np} \end{bmatrix}$$

$$A - B = A + (-B)$$

# Multiplying a Matrix by a Vector

- Let $A$ be a $n \times p$ matrix, and let $\boldsymbol{x}$ be a $p \times 1$ column vector

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{bmatrix} \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}$$

- then

$$A\boldsymbol{x} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1p}x_p \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2p}x_p \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{np}x_p \end{bmatrix}$$

Note that $A\boldsymbol{x}$ is a vector with shape $n \times 1$. The $i$-the element of $A\boldsymbol{x}$ is equivalent to the dot product of the $i$-th row vector of $A$ with $\boldsymbol{x}$.

# General Matrix Multiplication

- Let $A$ be a $n \times p$ matrix and $B$ be a $p \times q$ matrix:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{bmatrix} \quad B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1q} \\ b_{21} & b_{22} & \cdots & b_{2q} \\ \vdots & \vdots & \vdots & \vdots \\ b_{p1} & b_{n2} & \cdots & b_{nq} \end{bmatrix}$$

- The product $AB$ is an $n \times q$ matrix whose $(i, j)$-entry is the dot product of the $i$-th row vector of $A$ and the $j$-th column vector of $B$.
- $A$ and $B$ muse be *conformable* to calculate the product $AB$, i.e. the number of columns in $A$ must be the same as the number of rows in $B$.

# Matrix Arithmetic Rules

- i $A + B = B + A$
- ii $(A + B) + C = A + (B + C)$
- iii $k(A + B) = kA + kB$
- iv $(kA)B = k(AB)$
- v $(AB)C = A(BC)$ (associative)
- vi $A(B + C) = AB + AC$ (distributive)
- vii $(A + B)C = AC + BC$ (distributive)

**Alert**

*Matrix multiplication is **not** commutative, i.e. $AB \neq BA$ in general.*

Be careful when you expand expressions like $(A + B)(A + B)$.

# Matrix Transpose

- We denote the transpose of a matrix as $A^T$
- If $A$ is an $n \times p$ matrix, then $A^T$ is a $p \times n$ matrix where $A^T_{ji} = A_{ij}$
- Transpose rules:
  - $(A^T)^T = A$
  - $(A+B)^T = A^T + B^T$
  - $(AB)^T = B^T A^T$
- Symmetric matrix- square matrix, $A$, where $A^T = A$

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{bmatrix}$$

$$A^T = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ a_{1p} & a_{12} & \cdots & a_{np} \end{bmatrix}$$

# Identity matrix

An *identity matrix* is a $p \times p$ matrix with ones on the diagonal and zeros everywhere else.

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

- $IA = AI = A$ if $I$ and $A$ are $n \times p$ matrices
- $A = I\boldsymbol{x}$ is a diagonal matrix where $a_{ii} = x_i$ if $I$ is an $n \times n$ matrix and $\boldsymbol{x}$ is a $n \times 1$ vector.

# Descriptive statistics as matrix functions

# Mean vector and mean matrix

Assume you have a data set represented as a $n \times p$ matrix, $\boldsymbol{X}$, with observations in rows and variables in columns.

## Mean vector

To calculate a row vector of means, $\boldsymbol{m} = [\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_p]$

$$\boldsymbol{m} = \frac{1}{n} \mathtt{1}^T \boldsymbol{X}$$

where $\mathtt{1}$ is a $n \times 1$ column vector of ones.

## Matrix of column means

$$\boldsymbol{M} = \mathtt{1} \boldsymbol{m}$$

# Deviation matrix

To re-express each value as the deviation from the variable means (i.e. each column is a mean centered vector) we calculate a deviation matrix:

$$\boldsymbol{D} = \boldsymbol{X} - \boldsymbol{M}$$

# Covariance and correlation matrices

## Covariance matrix

$$\boldsymbol{S} = \frac{1}{n-1}\boldsymbol{D}^T\boldsymbol{D}$$

## Correlation matrix

$$\boldsymbol{R} = \boldsymbol{V}\boldsymbol{S}\boldsymbol{V}$$

where $\boldsymbol{V}$ is a $p \times p$ diagonal matrix where $\boldsymbol{V}_{ii} = 1/\sqrt{\boldsymbol{S}_{ii}}$.

# Linear dependence and independence

# Linear dependence and independence

- You'll remember that a *linear combination* of vectors is an equation of the form $z = b_1\boldsymbol{x}_1 + b_2\boldsymbol{x}_2 + \cdots + b_p\boldsymbol{x}_p$
- A list of vectors, $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p$ , is said the be ***linearly dependent*** if there is a non-trivial combination of them which is equal to the zero vector.

$$b_1\boldsymbol{x}_1 + b_2\boldsymbol{x}_2 + \cdots + b_p\boldsymbol{x}_p = 0$$

- A list of vectors that are not linearly dependent are said to be ***linearly independent***

# Matrix Inverses

- If $A$ is a *square matrix* and $C$ is a matrix of the same size where $AC = I$ and $CA = I$ than $C$ is the inverse of $A$ and we denote is $A^{-1}$.

$$AA^{-1} = A^{-1}A = I$$

- Rules for inverses:
  - Only square matrices are invertible; but not every square matrix can be inverted
  - A matrix for which we can find an inverse is called ***invertible*** (non-singular)
  - A matrix for which no inverse exists is ***singular*** (non-invertible)
  - Any diagonal matrix, $A$, where the $a_{ii}$ are non-zero, is invertible
  - If $A$ and $B$ are both invertible $p \times p$ matrices than $(AB)^{-1} = B^{-1}A^{-1}$ (note change in order).

**Highlight**

*If a matrix is invertible than it's columns form a linearly independent list of vectors!*

# Simultaneous Linear Equations

# Simultaneous Linear Equations

- A set of simultaneous linear equations are equations like the following:

$$
\begin{aligned}
x_1 + 3x_2 + 2x_3 &= 3 \\
-x_1 + x_2 + 2x_3 &= -2 \\
2x_1 + 4x_2 - 2x_3 &= 10
\end{aligned}
$$

- Simultaneous linear equations have either:
  - No solutions
  - One solution
  - Infinitely many solutions

# Matrices and Simultaneous Linear Equations

- Matrices can be used to represent and solve simultaneous linear equations. For example,

$$
\begin{aligned}
x_1 + 3x_2 + 2x_3 &= 3 \\
-x_1 + x_2 + 2x_3 &= -2 \\
2x_1 + 4x_2 - 2x_3 &= 10
\end{aligned}
$$

Can be represented by the equation $A\boldsymbol{x} = \boldsymbol{h}$:

$$
\begin{bmatrix} 1 & 3 & 2 \\ -1 & 1 & 2 \\ 2 & 4 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3 \\ -2 \\ 10 \end{bmatrix}
$$

- Solve this equation by pre-multiplying both sides of the equation by $A^{-1}$.

$$
\begin{aligned}
A^{-1}A\boldsymbol{x} &= A^{-1}\boldsymbol{h} \\
\boldsymbol{x} &= A^{-1}\boldsymbol{h}
\end{aligned}
$$

# Simultaneous Equations and Matrix Inverses

- $A\boldsymbol{x} = \boldsymbol{h}$ has a unique solution iff $A$ is invertible.
- If $A$ is a singular matrix than $A\boldsymbol{x} = \boldsymbol{h}$ either has no solution or infinitely many solutions.

# Multiple regression

# Variable space view of multiple regression

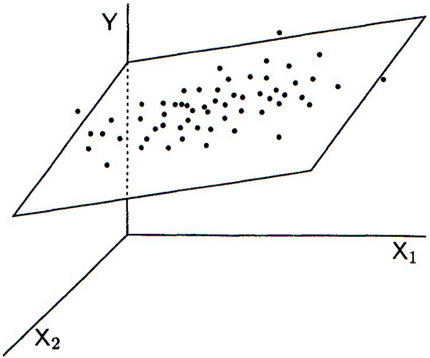
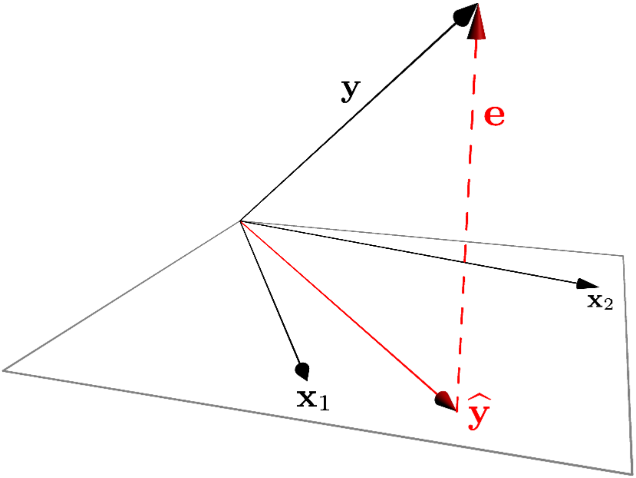



Figure 4.1: *The regression of* Y *onto* $X_1$ *and* $X_2$ *as a scatterplot in variable space.*

# Subject Space Geometry of Multiple Regression



# Multiple Regression

Let $\boldsymbol{y}$ be a vector of values for the outcome variable. Let $\boldsymbol{X}_i$ be explanatory variables.

The regression model is:

$$\boldsymbol{y} = \hat{\boldsymbol{y}} + \boldsymbol{e}$$

where

$$\hat{\boldsymbol{y}} = a\mathbf{1} + b_1\boldsymbol{X}_1 + b_2\boldsymbol{X}_2 + \cdots + b_p\boldsymbol{X}_p$$

Note that $\hat{\boldsymbol{y}}$ is a linear combination of the column vectors of $\boldsymbol{X}_i$.

# Matrix representation of multiple regression

$$\boldsymbol{y} = \hat{\boldsymbol{y}} + \boldsymbol{e}$$

looks like:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

We can solve for $\hat{\boldsymbol{y}}$ as:

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{b}$$

where

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} ; \; \boldsymbol{b} = \begin{bmatrix} a \\ b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix}$$

# Finding the Multiple Regression Coefficients

How do we solve for $\boldsymbol{b}$ in $\hat{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{b}$?

Estimate $\boldsymbol{b}$ as:

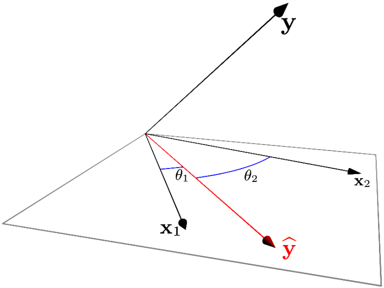
$$\boldsymbol{b} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

Compare this to our formula for estimating the coefficient $b$ in the bivariate regression of $\vec{\boldsymbol{y}}$ on a single variable $\vec{\boldsymbol{x}}$:

$$b = \frac{\vec{\boldsymbol{x}} \cdot \vec{\boldsymbol{y}}}{\vec{\boldsymbol{x}} \cdot \vec{\boldsymbol{x}}}$$

# Multiple Regression "Loadings"

In addition to the regression coefficients, the **regression "loadings"** – the cosine of the angle between vectors that represent the explanatory variables and the prediction vector – are also very useful for interpretting the regression model.



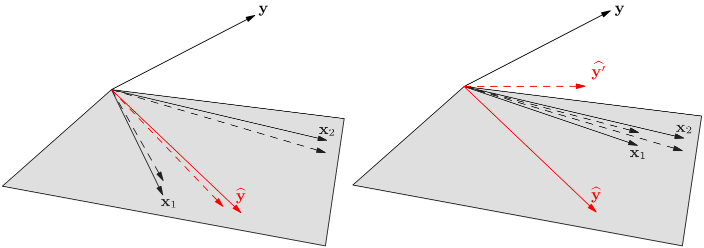

Individual loadings can be calculated as:

$$\cos\theta_{\overrightarrow{x_j},\overrightarrow{\hat{y}}} = \frac{\overrightarrow{x_j} \cdot \overrightarrow{\hat{y}}}{|\overrightarrow{x_j}||\overrightarrow{\hat{y}}|}$$

# Multiple regression: Cautions and Tips

- Comparing the size of regression coefficients only makes sense if all the predictor variables have the same scale
- The predictor variables (columns of $X$) must be linearly independent; when they're not the variables are **multicollinear**
- Predictor variables that are **nearly multicollinear** are, perhaps, even more difficult to deal with

# Why is near multicollinearity of the predictors a problem?




(a) Non-collinear predictors

(b) Non-collinear predictors

Figure: When predictors are nearly collinear, small differences in the vectors can result in large differences in the estimated regression.

# What can I do if my predictors are (nearly) collinear?

- Drop some of the linearly dependent sets of predictors.
- Replace the linearly dependent predictors with a combined variable.
- Define orthogonal predictors, via linear combinations of the original variables (PC regression approach)
- 'Tweak' the predictor variables so that they're no longer multicollinear (Ridge regression).