

Lecture 1: Intro/refresher in Matrix Algebra

Bruce Walsh lecture notes
Introduction to Mixed Models
SISG (Module 14), Seattle
19 – 21 July 2021

Topics

- Definitions, dimensionality, addition, subtraction
- Matrix multiplication
- Inverses, solving systems of equations
- Quadratic products and covariances
- The multivariate normal distribution
- Eigenstructure
- Basic matrix calculations in \mathbb{R}
- The Singular Value Decomposition (SVD)
 - First PAUSE slide 16

Matrices: An array of elements

Vectors: A matrix with either one row or one column.

Usually written in bold lowercase, e.g. **a**, **b**, **c**

$$\mathbf{a} = \begin{pmatrix} 12 \\ 13 \\ 47 \end{pmatrix} \quad \mathbf{b} = (2 \ 0 \ 5 \ 21)$$

Column vector

(3 x 1)

Row vector

(1 x 4)

Dimensionality of a matrix: $r \times c$ (rows x columns)
think of Railroad Car

General Matrices

Usually written in bold uppercase, e.g. **A**, **C**, **D**

$$\mathbf{C} = \begin{pmatrix} 3 & 1 & 2 \\ 2 & 5 & 4 \\ 1 & 1 & 2 \end{pmatrix} \quad \mathbf{D} = \begin{pmatrix} 0 & 1 \\ 3 & 4 \\ 2 & 9 \end{pmatrix}$$

(3 x 3) -- (3 x 2)

Square matrix

Dimensionality of a matrix: $r \times c$ (rows x columns)
think of Railroad Car

A matrix is defined by a list of its elements.

B has ij -th element B_{ij} -- the element in row i
and column j

Addition and Subtraction of Matrices

If two matrices have the same dimension (both are $r \times c$), then matrix addition and subtraction simply follows by adding (or subtracting) on an element by element basis

$$\text{Matrix addition: } (A+B)_{ij} = A_{ij} + B_{ij}$$

$$\text{Matrix subtraction: } (A-B)_{ij} = A_{ij} - B_{ij}$$

Examples:

$$\mathbf{A} = \begin{pmatrix} 3 & 0 \\ 1 & 2 \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$$

$$\mathbf{C} = \mathbf{A} + \mathbf{B} = \begin{pmatrix} 4 & 2 \\ 3 & 3 \end{pmatrix} \quad \text{and} \quad \mathbf{D} = \mathbf{A} - \mathbf{B} = \begin{pmatrix} 2 & -2 \\ -1 & 1 \end{pmatrix}$$

Partitioned Matrices

It will often prove useful to divide (or [partition](#)) the elements of a matrix into a matrix whose elements are itself matrices.

$$\mathbf{C} = \begin{pmatrix} 3 & 1 & 2 \\ 2 & 5 & 4 \\ 1 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 3 & \vdots & 1 & 2 \\ \dots & \dots & \dots & \dots \\ 2 & \vdots & 5 & 4 \\ 1 & \vdots & 1 & 2 \end{pmatrix} = \begin{pmatrix} \mathbf{a} & \mathbf{b} \\ \mathbf{d} & \mathbf{B} \end{pmatrix}$$

$$\mathbf{a} = (3), \quad \mathbf{b} = (1 \ 2), \quad \mathbf{d} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 5 & 4 \\ 1 & 2 \end{pmatrix}$$

One useful partition is to write the matrix as either a [row vector of column vectors](#) or a [column vector of row vectors](#)

$$\mathbf{C} = \begin{pmatrix} 3 & 1 & 2 \\ 2 & 5 & 4 \\ 1 & 1 & 2 \end{pmatrix} = \begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \mathbf{r}_3 \end{pmatrix}$$

A column vector whose elements are row vectors

$$\mathbf{r}_1 = (3 \quad 1 \quad 2)$$

$$\mathbf{r}_2 = (2 \quad 5 \quad 4)$$

$$\mathbf{r}_3 = (1 \quad 1 \quad 2)$$

$$\mathbf{C} = \begin{pmatrix} 3 & 1 & 2 \\ 2 & 5 & 4 \\ 1 & 1 & 2 \end{pmatrix} = (\mathbf{c}_1 \quad \mathbf{c}_2 \quad \mathbf{c}_3)$$

A row vector whose elements are column vectors

$$\mathbf{c}_1 = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}, \quad \mathbf{c}_2 = \begin{pmatrix} 1 \\ 5 \\ 1 \end{pmatrix}, \quad \mathbf{c}_3 = \begin{pmatrix} 2 \\ 4 \\ 2 \end{pmatrix}$$

Towards Matrix Multiplication: dot products

The **dot** (or **inner**) **product** of two vectors (both of length n) is defined as follows:

$$\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^n a_i b_i$$

Example:

$$\mathbf{a} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} \quad \text{and} \quad \mathbf{b} = (4 \ 5 \ 7 \ 9)$$

$$\mathbf{a} \cdot \mathbf{b} = 1 \cdot 4 + 2 \cdot 5 + 3 \cdot 7 + 4 \cdot 9 = 71$$

Matrices are compact ways to write systems of equations

$$5x_1 + 6x_2 + 4x_3 = 6$$

$$7x_1 - 3x_2 + 5x_3 = -9$$

$$-x_1 - x_2 + 6x_3 = 12$$

$$\begin{pmatrix} 5 & 6 & 4 \\ 7 & -3 & 5 \\ -1 & -1 & 6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 6 \\ -9 \\ 12 \end{pmatrix}$$

$$\mathbf{Ax} = \mathbf{c}, \quad \text{or} \quad \mathbf{x} = \mathbf{A}^{-1}\mathbf{c}$$

The least-squares solution for the linear model

$$y = \mu + \beta_1 z_1 + \cdots + \beta_n z_n$$

yields the following system of equations for the β_i

$$\sigma(y, z_1) = \beta_1 \sigma^2(z_1) + \beta_2 \sigma(z_1, z_2) + \cdots + \beta_n \sigma(z_1, z_n)$$

$$\sigma(y, z_2) = \beta_1 \sigma(z_1, z_2) + \beta_2 \sigma^2(z_2) + \cdots + \beta_n \sigma(z_2, z_n)$$

$$\vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \ddots \quad \quad \quad \vdots$$

$$\sigma(y, z_n) = \beta_1 \sigma(z_1, z_n) + \beta_2 \sigma(z_2, z_n) + \cdots + \beta_n \sigma^2(z_n)$$

This can be more compactly written in matrix form as

$$\begin{pmatrix} \sigma^2(z_1) & \sigma(z_1, z_2) & \cdots & \sigma(z_1, z_n) \\ \sigma(z_1, z_2) & \sigma^2(z_2) & \cdots & \sigma(z_2, z_n) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma(z_1, z_n) & \sigma(z_2, z_n) & \cdots & \sigma^2(z_n) \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix} = \begin{pmatrix} \sigma(y, z_1) \\ \sigma(y, z_2) \\ \vdots \\ \sigma(y, z_n) \end{pmatrix}$$

$X^T X \qquad \qquad \qquad \beta \qquad \qquad \qquad X^T y$

$$\text{or, } \beta = (X^T X)^{-1} X^T y$$

Matrix Multiplication:

The order in which matrices are multiplied affects the matrix product, e.g. $AB \neq BA$

For the product of two matrices to exist, the matrices must **conform**. For AB , the number of columns of A must equal the number of rows of B .

The matrix $C = AB$ has the same number of rows as A and the same number of columns as B .

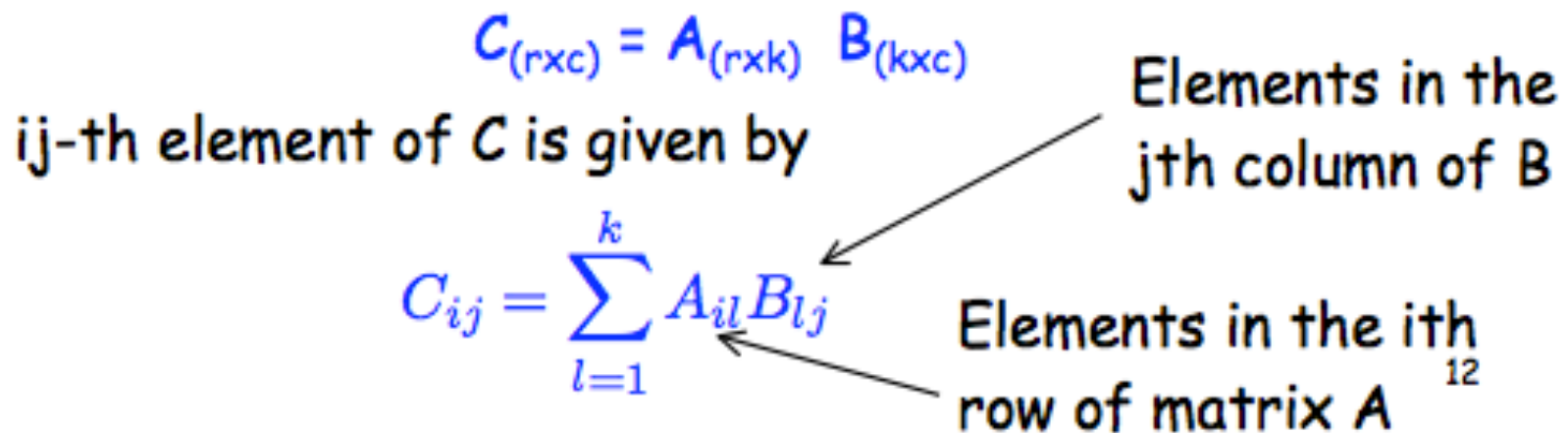
$C_{(r \times c)} = A_{(r \times k)} B_{(k \times c)}$

ij -th element of C is given by

$$C_{ij} = \sum_{l=1}^k A_{il} B_{lj}$$

Elements in the j th column of B

Elements in the i th row of matrix A ¹²



Outer indices given dimensions of resulting matrix, with r rows (A) and c columns (B)

$$C_{(rxc)} = A_{(rxk)} B_{(kxc)}$$

Inner indices must match
columns of A = rows of B

Example: Is the product ABCD defined? If so, what is its dimensionality? Suppose

$$A_{3 \times 5} B_{5 \times 9} C_{9 \times 6} D_{6 \times 23}$$

Yes, defined, as **inner indices match**. Result is a 3 x 23 matrix (3 rows, 23 columns)

More formally, consider the product $L = MN$

Express the matrix M as a column vector of row vectors

$$M = \begin{pmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \\ \vdots \\ \mathbf{m}_r \end{pmatrix} \quad \text{where} \quad \mathbf{m}_i = (M_{i1} \quad M_{i2} \quad \cdots \quad M_{ic})$$

Likewise express N as a row vector of column vectors

$$N = (\mathbf{n}_1 \quad \mathbf{n}_2 \quad \cdots \quad \mathbf{n}_b) \quad \text{where} \quad \mathbf{n}_j = \begin{pmatrix} N_{1j} \\ N_{2j} \\ \vdots \\ N_{cj} \end{pmatrix}$$

The ij -th element of L is the inner product of M 's row i with N 's column j

$$L = \begin{pmatrix} \mathbf{m}_1 \cdot \mathbf{n}_1 & \mathbf{m}_1 \cdot \mathbf{n}_2 & \cdots & \mathbf{m}_1 \cdot \mathbf{n}_b \\ \mathbf{m}_2 \cdot \mathbf{n}_1 & \mathbf{m}_2 \cdot \mathbf{n}_2 & \cdots & \mathbf{m}_2 \cdot \mathbf{n}_b \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{m}_r \cdot \mathbf{n}_1 & \mathbf{m}_r \cdot \mathbf{n}_2 & \cdots & \mathbf{m}_r \cdot \mathbf{n}_b \end{pmatrix}$$

Example

$$\mathbf{AB} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{pmatrix}$$

Likewise

$$\mathbf{BA} = \begin{pmatrix} ae + cf & eb + df \\ ga + ch & gd + dh \end{pmatrix}$$

ORDER of multiplication matters! Indeed, consider $C_{3 \times 5} D_{5 \times 5}$ which gives a 3 x 5 matrix, versus $D_{5 \times 5} C_{3 \times 5}$, which is not defined.

Matrix multiplication in R

```
> A<-matrix(c(1,2,3,4),nrow=2)
> B<-matrix(c(4,5,6,7),nrow=2)
> A
      [,1] [,2]
[1,]    1    3
[2,]    2    4
> B
      [,1] [,2]
[1,]    4    6
[2,]    5    7
> A %*% B
      [,1] [,2]
[1,]   19   27
[2,]   28   40
```

R fills in the matrix from the list `c` by filling in as columns, here with 2 rows (`nrow=2`)

Entering `A` or `B` displays what was entered (always a good thing to check)

The command `%*%` is the R code for the multiplication of two matrices

On your own: What is the matrix resulting from `BA`?
What is `A` if `nrow=1` or `nrow=4` is used?

PAUSE

- Matrix multiplication arises as a way to compactly write systems of equations
- Indeed, much of linear algebra has deep roots in systems of equations, as we will now explore.
- Next pause at slide 27


The Transpose of a Matrix

The transpose of a matrix exchanges the rows and columns, $A^T_{ij} = A_{ji}$

Useful identities

$$(AB)^T = B^T A^T$$
$$(ABC)^T = C^T B^T A^T$$
$$\mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

Inner product = $\mathbf{a}^T \mathbf{b} = \mathbf{a}^T_{(1 \times n)} \mathbf{b}_{(n \times 1)}$



Indices match, matrices conform

Dimension of resulting product is 1 X 1 (i.e. a scalar)

$$(a_1 \ \cdots \ a_n) \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \mathbf{a}^T \mathbf{b} = \sum_{i=1}^n a_i b_i$$

Note that $\mathbf{b}^T \mathbf{a} = (\mathbf{b}^T \mathbf{a})^T = \mathbf{a}^T \mathbf{b}$

$$\text{Outer product} = ab^T = a_{(n \times 1)} b^T_{(1 \times n)}$$

Resulting product is an $n \times n$ matrix

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} (b_1 \quad b_2 \quad \cdots \quad b_n)$$
$$= \begin{pmatrix} a_1 b_1 & a_1 b_2 & \cdots & a_1 b_n \\ a_2 b_1 & a_2 b_2 & \cdots & a_2 b_n \\ \vdots & \vdots & \ddots & \vdots \\ a_n b_1 & a_n b_2 & \cdots & a_n b_n \end{pmatrix}$$

R code for transposition

```
> t(A)
      [,1] [,2]
[1,]    1    2
[2,]    3    4
^
```

$t(A)$ = transpose of A

```
> a<-matrix(c(1,2,3),nrow=3) Enter the column vector a
> a
```

```
      [,1]
[1,]    1
[2,]    2
[3,]    3
```

```
> t(a) %*% a
```

Compute inner product $a^T a$

```
      [,1]
[1,]   14
```

```
> a %*% t(a)
```

Compute outer product aa^T

```
      [,1] [,2] [,3]
[1,]    1    2    3
[2,]    2    4    6
[3,]    3    6    9
```

Solving equations

- The **identity matrix** I
 - Serves the same role as 1 in scalar algebra, e.g.,
 $a*1=1*a =a$, with $AI=IA= A$
- The inverse matrix A^{-1} (IF it exists)
 - Defined by $A A^{-1} = I, A^{-1}A = I$
 - Serves the same role as scalar division
 - To solve $ax = c$, multiply both sides by $(1/a)$ to give:
 - $(1/a)*ax = (1/a)c$ or $(1/a)*a*x = 1*x = x$,
 - Hence $x = (1/a)c$
 - To solve $Ax = c$, $A^{-1}Ax = A^{-1} c$
 - Or $A^{-1}Ax = Ix = x = A^{-1} c$

The Identity Matrix, I

The identity matrix serves the role of the number 1 in matrix multiplication: $AI = A, IA = A$

I is a square diagonal matrix, with all diagonal elements being one, all off-diagonal elements zero.

$$I_{ij} = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{I}_{3 \times 3} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The Identity Matrix in R

`diag(k)`, where `k` is an integer, return the `k x k` I matrix

```
> I<-diag(4)
> I
      [,1] [,2] [,3] [,4]
[1,]    1    0    0    0
[2,]    0    1    0    0
[3,]    0    0    1    0
[4,]    0    0    0    1
> I2 <-diag(2)
> I2
      [,1] [,2]
[1,]    1    0
[2,]    0    1
```

The Inverse Matrix, A^{-1}

For a square matrix A , define its **Inverse** A^{-1} , as the matrix satisfying

$$\mathbf{A}^{-1} \mathbf{A} = \mathbf{A} \mathbf{A}^{-1} = \mathbf{I}$$

For $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ $\mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$

If this quantity (the **determinant**) is zero, the inverse does not exist.

If $\det(A)$ is not zero, A^{-1} exists and A is said to be **non-singular**. If $\det(A) = 0$, A is **singular**, and no *unique* inverse exists (**generalized inverses** do)

Generalized inverses, and their uses in solving systems of equations, are discussed in Appendix 3 of Lynch & Walsh

A^- is the typical notation to denote the G-inverse of a matrix

When a G-inverse is used, provided the system is **consistent**, then some of the variables have a family of solutions (e.g., $x_1 = 2$, but $x_2 + x_3 = 6$)

Inversion in R

`solve(A)` computes A^{-1}

`det(A)` computes determinant of A

```
> A                                     Using A entered earlier
  [,1] [,2]
[1,]  1  3
[2,]  2  4
> solve(A)                               Compute A-1
  [,1] [,2]
[1,] -2  1.5
[2,]  1 -0.5
> solve(A) %% A                           Showing that A-1 A = I
  [,1] [,2]
[1,]  1 -8.881784e-16
[2,]  0  1.000000e+00
> det(A)                                  Computing determinant of A
[1] -2
```

Homework

Put the following system of equations in matrix form, and solve using R

$$3x_1 + 4x_2 + 4x_3 + 6x_4 = -10$$

$$9x_1 + 2x_2 - x_3 - 6x_4 = 20$$

$$x_1 + x_2 + x_3 - 10x_4 = 2$$

$$2x_1 + 9x_2 + 2x_3 - x_4 = -10$$

PAUSE

- One can think of the inverse of a square matrix \mathbf{A} as the unique solution to its corresponding set of equations.
- The determinant of \mathbf{A} informs us as to whether such a unique solution exists
- As we will now see, the eigenstructure (geometry) of \mathbf{A} provides a deeper understanding of nature of solutions, especially when $\det(\mathbf{A})= 0$.
- Next pause at slide 35

Useful identities

$$(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$$

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1} \mathbf{A}^{-1}$$

For a diagonal matrix \mathbf{D} , then $\det(\mathbf{D})$, which is also denoted by $|\mathbf{D}|$, = product of the diagonal elements

Also, the determinant of any square matrix \mathbf{A} , $\det(\mathbf{A})$, is simply the product of the **eigenvalues** λ of \mathbf{A} , which satisfy

$$\mathbf{Ae} = \lambda \mathbf{e}$$

If \mathbf{A} is $n \times n$, solutions to λ are an n -degree polynomial. \mathbf{e} is the **eigenvector** associated with λ . If any of the roots to the equation are zero, \mathbf{A}^{-1} is not defined. In this case, for some linear combination \mathbf{b} , we have $\mathbf{Ab} = \mathbf{0}$.

Variance-Covariance matrix

- A very important square matrix is the **variance-covariance matrix** \mathbf{V} associated with a vector \mathbf{x} of random variables.
- $V_{ij} = \text{Cov}(x_i, x_j)$, so that the i -th diagonal element of \mathbf{V} is the variance of x_i , and off-diagonal elements are covariances
- \mathbf{V} is a symmetric, square matrix

The trace

The **trace**, $\text{tr}(\mathbf{A})$ or **trace**(\mathbf{A}), of a square matrix \mathbf{A} is simply the sum of its diagonal elements

The importance of the trace is that it equals the sum of the eigenvalues of \mathbf{A} , $\text{tr}(\mathbf{A}) = \sum \lambda_i$

For a covariance matrix \mathbf{V} , $\text{tr}(\mathbf{V})$ measures the total amount of variation in the variables

$\lambda_i / \text{tr}(\mathbf{V})$ is the fraction of the total variation in \mathbf{x} contained in the linear combination $\mathbf{e}_i^T \mathbf{x}$, where \mathbf{e}_i , the i -th **principal component** of \mathbf{V} is also the i -th eigenvector of \mathbf{V} ($\mathbf{V}\mathbf{e}_i = \lambda_i \mathbf{e}_i$)

Eigenstructure in R

`eigen(A)` returns the eigenvalues and vectors of A

```
> V<-matrix(c(10,-5,10,-5,20,0,10,0,30),nrow=3)
```

```
> V
```

```
      [,1] [,2] [,3]
[1,]  10   -5  10
[2,]  -5   20   0
[3,]  10   0   30
```

```
> eigen(V)
```

```
$values
[1] 34.410103 21.117310 4.472587
```

```
$vectors
```

```
      [,1] [,2] [,3]
[1,] 0.3996151 0.2117936 0.8918807
[2,] -0.1386580 -0.9477830 0.2871955
[3,] 0.9061356 -0.2384340 -0.3493816
```

PC 1 PC 2

Trace = 60

PC 1 accounts for $34.4/60 = 57\%$ of all the variation

$$0.400 * x_1 - 0.139 * x_2 + 0.906 * x_3$$

$$0.212 * x_1 - 0.948 * x_2 - 0.238 * x_3$$

Quadratic and Bilinear Forms

Quadratic product: for $A_{n \times n}$ and $x_{n \times 1}$

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j \quad \text{Scalar (1 x 1)}$$

Bilinear Form (generalization of quadratic product)

for $A_{m \times n}$, $a_{n \times 1}$, $b_{m \times 1}$ their bilinear form is $b^T_{1 \times m} A_{m \times n} a_{n \times 1}$

$$\mathbf{b}^T \mathbf{A} \mathbf{a} = \sum_{i=1}^m \sum_{j=1}^n A_{ij} b_i a_j$$

Note that $b^T A a = a^T A^T b$

Covariance Matrices for Transformed Variables

What is the variance of the linear combination,
 $c_1x_1 + c_2x_2 + \dots + c_nx_n$? (note this is a scalar)

$$\begin{aligned}\sigma^2(\mathbf{c}^T \mathbf{x}) &= \sigma^2\left(\sum_{i=1}^n c_i x_i\right) = \sigma\left(\sum_{i=1}^n c_i x_i, \sum_{j=1}^n c_j x_j\right) \\ &= \sum_{i=1}^n \sum_{j=1}^n \sigma(c_i x_i, c_j x_j) = \sum_{i=1}^n \sum_{j=1}^n c_i c_j \sigma(x_i, x_j) \\ &= \mathbf{c}^T \mathbf{V} \mathbf{c}\end{aligned}$$

Likewise, the covariance between two linear combinations can be expressed as a bilinear form,

$$\sigma(\mathbf{a}^T \mathbf{x}, \mathbf{b}^T \mathbf{x}) = \mathbf{a}^T \mathbf{V} \mathbf{b}$$

Example: Suppose the variances of x_1 , x_2 , and x_3 are 10, 20, and 30. x_1 and x_2 have a covariance of -5, x_1 and x_3 of 10, while x_2 and x_3 are uncorrelated.

What are the variances of the indices
 $y_1 = x_1 - 2x_2 + 5x_3$ and $y_2 = 6x_2 - 4x_3$?

$$\mathbf{V} = \begin{pmatrix} 10 & -5 & 10 \\ -5 & 20 & 0 \\ 10 & 0 & 30 \end{pmatrix}, \quad \mathbf{c}_1 = \begin{pmatrix} 1 \\ -2 \\ 5 \end{pmatrix}, \quad \mathbf{c}_2 = \begin{pmatrix} 0 \\ 6 \\ -4 \end{pmatrix}$$

$$\text{Var}(y_1) = \text{Var}(\mathbf{c}_1^T \mathbf{x}) = \mathbf{c}_1^T \text{Var}(\mathbf{x}) \mathbf{c}_1 = 960$$

$$\text{Var}(y_2) = \text{Var}(\mathbf{c}_2^T \mathbf{x}) = \mathbf{c}_2^T \text{Var}(\mathbf{x}) \mathbf{c}_2 = 1200$$

$$\text{Cov}(y_1, y_2) = \text{Cov}(\mathbf{c}_1^T \mathbf{x}, \mathbf{c}_2^T \mathbf{x}) = \mathbf{c}_1^T \text{Var}(\mathbf{x}) \mathbf{c}_2 = -910$$

Homework: use R to compute the above values

PAUSE

- The concepts of inverses, quadratic products, eigenstructure, and covariance matrices form the foundation to explore the multivariate normal (MVN) distribution
- This distribution underpins much of linear and mixed models theory and we will extensively use its properties
 - PCs
 - Regressions and conditional expectations
 - Next pause at slide 53

The Multivariate Normal Distribution (MVN)

Consider the pdf for n independent normal random variables, the i th of which has mean μ_i and variance σ_i^2

$$\begin{aligned} p(\mathbf{x}) &= \prod_{i=1}^n (2\pi)^{-1/2} \sigma_i^{-1} \exp\left(-\frac{(x_i - \mu_i)^2}{2\sigma_i^2}\right) \\ &= (2\pi)^{-n/2} \left(\prod_{i=1}^n \sigma_i\right)^{-1} \exp\left(-\sum_{i=1}^n \frac{(x_i - \mu_i)^2}{2\sigma_i^2}\right) \end{aligned}$$

This can be expressed more compactly in matrix form

Define the **covariance matrix** \mathbf{V} for the vector \mathbf{x} of the n normal random variable by

$$\mathbf{V} = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \sigma_n^2 \end{pmatrix} \quad |\mathbf{V}| = \prod_{i=1}^n \sigma_i^2$$

Define the mean vector $\boldsymbol{\mu}$ by gives

$$\sum_{i=1}^n \frac{(x_i - \mu_i)^2}{\sigma_i^2} = (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{V}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}$$

Hence in matrix form the MVN pdf becomes

$$p(\mathbf{x}) = (2\pi)^{-n/2} |\mathbf{V}|^{-1/2} \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{V}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right]$$

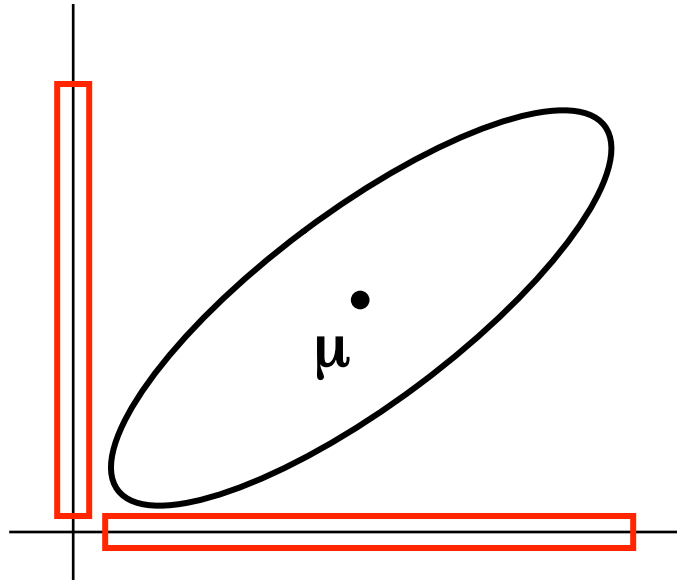
Notice this holds for any vector $\boldsymbol{\mu}$ and symmetric **positive-definite** matrix \mathbf{V} , as $|\mathbf{V}| > 0$.

The multivariate normal

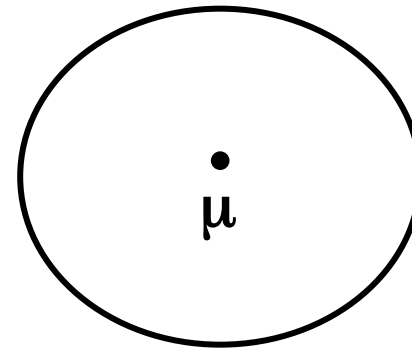
- Just as a univariate normal is defined by its mean and spread, a multivariate normal is defined by its mean vector $\boldsymbol{\mu}$ (also called the centroid) and variance-covariance matrix \mathbf{V}

Vector of means μ determines location

Spread (geometry) about μ determined by V



x_1, x_2 equal variances,
positively correlated

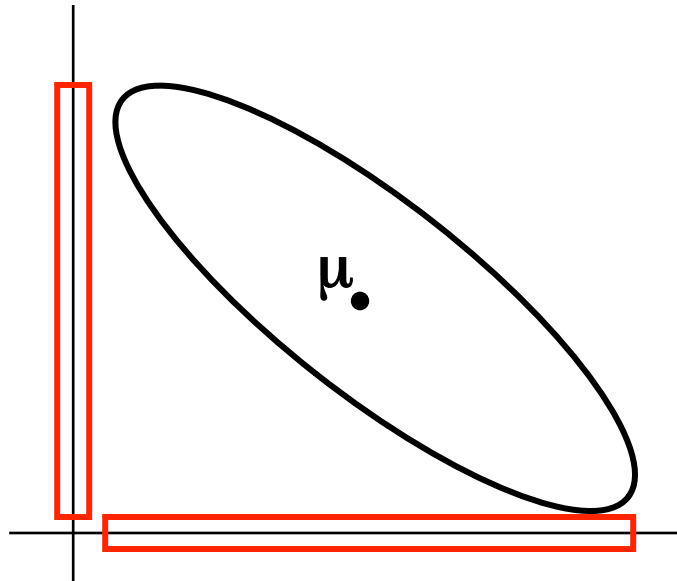


x_1, x_2 equal variances,
uncorrelated

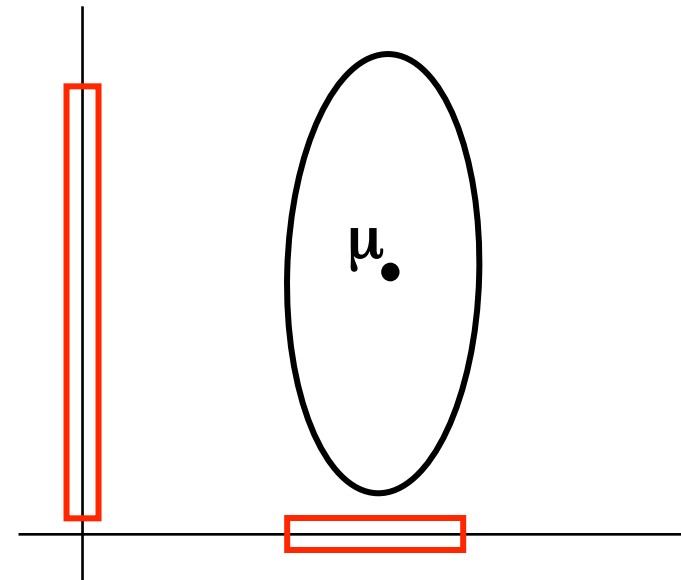
Eigenstructure (the eigenvectors and their corresponding eigenvalues) determines the geometry of V .

Vector of means μ determines location

Spread (geometry) about μ determined by V



x_1, x_2 equal variances,
negatively correlated



$\text{Var}(x_1) < \text{Var}(x_2)$,
uncorrelated

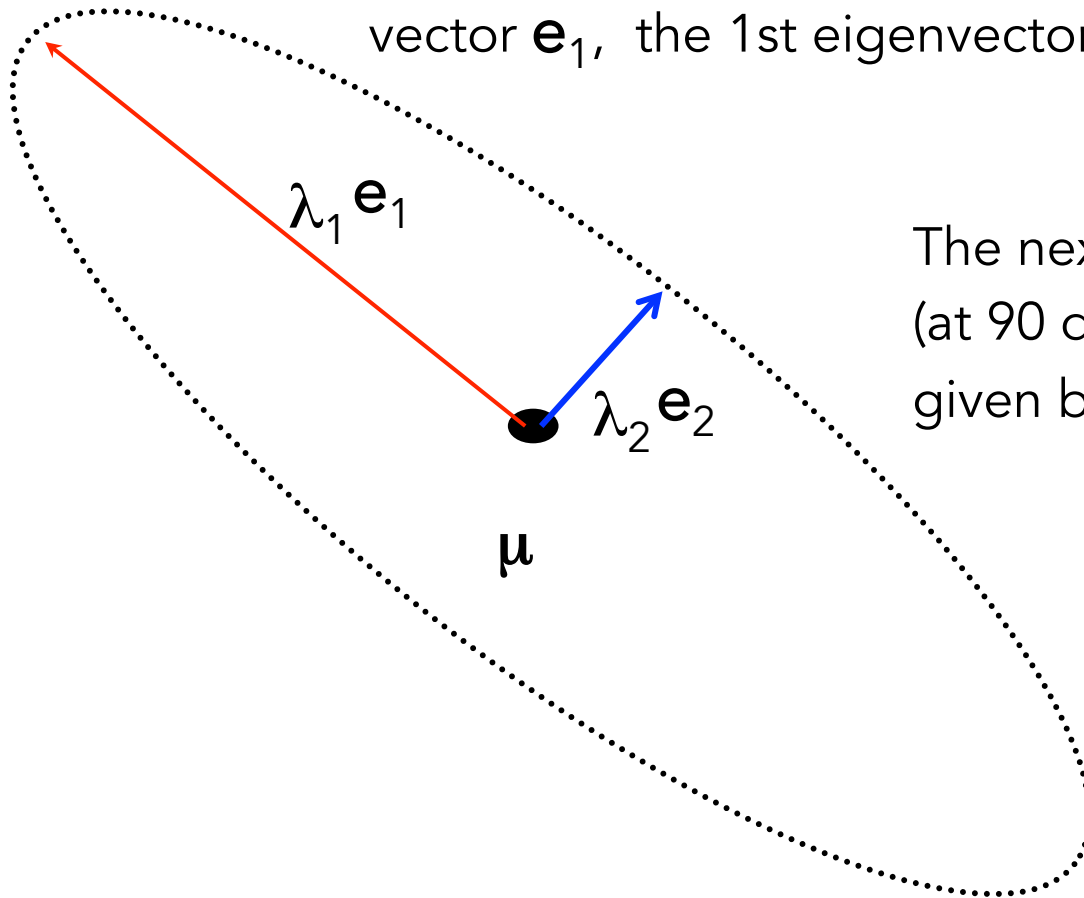
Positive tilt = positive correlations

Negative tilt = negative correlation

No tilt = uncorrelated

Eigenstructure of V

The direction of the largest axis of variation is given by the unit-length vector \mathbf{e}_1 , the 1st eigenvector of V .



The next largest axis of orthogonal (at 90 degrees from) \mathbf{e}_1 , is given by \mathbf{e}_2 , the 2nd eigenvector

Principal components

- The principal components (or PCs) of a covariance matrix define the axes of variation.
 - PC1 is the direction (linear combination $c^T x$) that explains the most variation.
 - PC2 is the next largest direction (at 90 degree from PC1), and so on
- $PC_i =$ i th eigenvector of V
- Fraction of variation accounted for by $PC_i = \lambda_i / \text{trace}(V)$
- If V has a few large eigenvalues, most of the variation is distributed along a few linear combinations (axis of variation)
- The singular value decomposition is the generalization of this idea to nonsquare matrices

Properties of the MVN - I

1) If \mathbf{x} is MVN, **any subset** of the variables in \mathbf{x} is also MVN

2) If \mathbf{x} is MVN, **any linear combination** of the elements of \mathbf{x} is also MVN. If $\mathbf{x} \sim \text{MVN}(\boldsymbol{\mu}, \mathbf{V})$

for $\mathbf{y} = \mathbf{x} + \mathbf{a}$, \mathbf{y} is $\text{MVN}_n(\boldsymbol{\mu} + \mathbf{a}, \mathbf{V})$

for $y = \mathbf{a}^T \mathbf{x} = \sum_{k=1}^n a_k x_k$, y is $\text{N}(\mathbf{a}^T \boldsymbol{\mu}, \mathbf{a}^T \mathbf{V} \mathbf{a})$

for $\mathbf{y} = \mathbf{A}\mathbf{x}$, \mathbf{y} is $\text{MVN}_m(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}^T \mathbf{V} \mathbf{A})$

Properties of the MVN - II

3) Conditional distributions are also MVN. Partition \mathbf{x} into two components, \mathbf{x}_1 (m dimensional column vector) and \mathbf{x}_2 ($n-m$ dimensional column vector)

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \quad \text{and} \quad \mathbf{V} = \begin{pmatrix} \mathbf{V}_{\mathbf{x}_1\mathbf{x}_1} & \mathbf{V}_{\mathbf{x}_1\mathbf{x}_2} \\ \mathbf{V}_{\mathbf{x}_1\mathbf{x}_2}^T & \mathbf{V}_{\mathbf{x}_2\mathbf{x}_2} \end{pmatrix}$$

$\mathbf{x}_1 | \mathbf{x}_2$ is MVN with m -dimensional mean vector

$$\boldsymbol{\mu}_{\mathbf{x}_1|\mathbf{x}_2} = \boldsymbol{\mu}_1 + \mathbf{V}_{\mathbf{x}_1\mathbf{x}_2} \mathbf{V}_{\mathbf{x}_2\mathbf{x}_2}^{-1} (\mathbf{x}_2 - \boldsymbol{\mu}_2)$$

and $m \times m$ covariance matrix


$$\mathbf{V}_{\mathbf{x}_1|\mathbf{x}_2} = \mathbf{V}_{\mathbf{x}_1\mathbf{x}_1} - \mathbf{V}_{\mathbf{x}_1\mathbf{x}_2} \mathbf{V}_{\mathbf{x}_2\mathbf{x}_2}^{-1} \mathbf{V}_{\mathbf{x}_1\mathbf{x}_2}^T$$

Properties of the MVN - III

- 4) If x is MVN, the regression of any subset of x on another subset is **linear** and **homoscedastic**

$$\begin{aligned}\mathbf{x}_1 &= \mu_{\mathbf{x}_1|\mathbf{x}_2} + \mathbf{e} \\ &= \mu_1 + \mathbf{V}_{\mathbf{x}_1\mathbf{x}_2} \mathbf{V}_{\mathbf{x}_2\mathbf{x}_2}^{-1} (\mathbf{x}_2 - \mu_2) + \mathbf{e}\end{aligned}$$

Where \mathbf{e} is MVN with mean vector $\mathbf{0}$ and variance-covariance matrix $\mathbf{V}_{\mathbf{x}_1|\mathbf{x}_2}$

$$\mu_1 + \mathbf{V}_{\mathbf{X}_1\mathbf{X}_2} \mathbf{V}_{\mathbf{X}_2\mathbf{X}_2}^{-1} (\mathbf{x}_2 - \mu_2) + \mathbf{e}$$


The regression is **linear** because it is a linear function of x_2

The regression is **homoscedastic** because the variance-covariance matrix for \mathbf{e} does not depend on the value of the x 's

$$\mathbf{V}_{\mathbf{X}_1|\mathbf{X}_2} = \mathbf{V}_{\mathbf{X}_1\mathbf{X}_1} - \mathbf{V}_{\mathbf{X}_1\mathbf{X}_2} \mathbf{V}_{\mathbf{X}_2\mathbf{X}_2}^{-1} \mathbf{V}_{\mathbf{X}_1\mathbf{X}_2}^T$$

All these matrices are constant, and hence the same for any value of x

Example: Regression of Offspring value on Parental values

Assume the vector of offspring value and the values of both its parents is MVN. Then from the correlations among (outbred) relatives,

$$\begin{pmatrix} z_o \\ z_s \\ z_d \end{pmatrix} \sim \text{MVN} \left[\begin{pmatrix} \mu_0 \\ \mu_s \\ \mu_d \end{pmatrix}, \sigma_z^2 \begin{pmatrix} 1 & h^2/2 & h^2/2 \\ h^2/2 & 1 & 0 \\ h^2/2 & 0 & 1 \end{pmatrix} \right]$$

Let $\mathbf{x}_1 = (z_o)$, $\mathbf{x}_2 = \begin{pmatrix} z_s \\ z_d \end{pmatrix}$

$$\mathbf{V}_{\mathbf{x}_1, \mathbf{x}_1} = \sigma_z^2, \quad \mathbf{V}_{\mathbf{x}_1, \mathbf{x}_2} = \frac{h^2 \sigma_z^2}{2} (1 \quad 1), \quad \mathbf{V}_{\mathbf{x}_2, \mathbf{x}_2} = \sigma_z^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$= \mu_1 + \mathbf{V}_{\mathbf{x}_1, \mathbf{x}_2} \mathbf{V}_{\mathbf{x}_2, \mathbf{x}_2}^{-1} (\mathbf{x}_2 - \mu_2) + \mathbf{e}$$

Regression of Offspring value on Parental values (cont.)

$$= \mu_1 + \mathbf{V}_{\mathbf{X}_1\mathbf{X}_2} \mathbf{V}_{\mathbf{X}_2\mathbf{X}_2}^{-1} (\mathbf{x}_2 - \mu_2) + \mathbf{e}$$

$$\mathbf{V}_{\mathbf{X}_1\mathbf{X}_1} = \sigma_z^2, \quad \mathbf{V}_{\mathbf{X}_1\mathbf{X}_2} = \frac{h^2\sigma_z^2}{2} (1 \ 1), \quad \mathbf{V}_{\mathbf{X}_2\mathbf{X}_2} = \sigma_z^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Hence,

$$z_o = \mu_o + \frac{h^2\sigma_z^2}{2} (1 \ 1) \sigma_z^{-2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} z_s - \mu_s \\ z_d - \mu_d \end{pmatrix} + e$$

$$= \mu_o + \frac{h^2}{2} (z_s - \mu_s) + \frac{h^2}{2} (z_d - \mu_d) + e$$

Where e is normal with mean zero and variance

$$\mathbf{V}_{\mathbf{X}_1|\mathbf{X}_2} = \mathbf{V}_{\mathbf{X}_1\mathbf{X}_1} - \mathbf{V}_{\mathbf{X}_1\mathbf{X}_2} \mathbf{V}_{\mathbf{X}_2\mathbf{X}_2}^{-1} \mathbf{V}_{\mathbf{X}_1\mathbf{X}_2}^T$$

$$\sigma_e^2 = \sigma_z^2 - \frac{h^2\sigma_z^2}{2} (1 \ 1) \sigma_z^{-2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \frac{h^2\sigma_z^2}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$= \sigma_z^2 \left(1 - \frac{h^4}{2} \right)$$

Hence, the regression of offspring trait value given the trait values of its parents is

$$z_o = \mu_o + h^2/2(z_s - \mu_s) + h^2/2(z_d - \mu_d) + e$$

where the residual e is normal with mean zero and $\text{Var}(e) = \sigma_z^2(1-h^4/2)$

Similar logic gives the regression of offspring breeding value on parental breeding value as

$$\begin{aligned} A_o &= \mu_o + (A_s - \mu_s)/2 + (A_d - \mu_d)/2 + e \\ &= A_s/2 + A_d/2 + e \end{aligned}$$

where the residual e is normal with mean zero and $\text{Var}(e) = \sigma_A^2/2$

Additional R matrix commands

Operator or Function	Description
<code>A * B</code>	Element-wise multiplication
<code>A ** B</code>	Matrix multiplication
<code>A %o% B</code>	Outer product. AB'
<code>crossprod(A,B)</code> <code>crossprod(A)</code>	$A'B$ and $A'A$ respectively.
<code>t(A)</code>	Transpose
<code>diag(x)</code>	Creates diagonal matrix with elements of x in the principal diagonal
<code>diag(A)</code>	Returns a vector containing the elements of the principal diagonal
<code>diag(k)</code>	If k is a scalar, this creates a $k \times k$ identity matrix. Go figure.
<code>solve(A, b)</code>	Returns vector x in the equation $b = Ax$ (i.e., $A^{-1}b$)
<code>solve(A)</code>	Inverse of A where A is a square matrix.
<code>ginv(A)</code>	Moore-Penrose Generalized Inverse of A . <code>ginv(A)</code> requires loading the <i>MASS</i> package.
<code>y<-eigen(A)</code>	$y\$val$ are the eigenvalues of A $y\$vec$ are the eigenvectors of A
<code>y<-svd(A)</code>	Single value decomposition of A . $y\$d$ = vector containing the singular values of A $y\$u$ = matrix with columns contain the left singular vectors of A $y\$v$ = matrix with columns contain the right singular vectors of A

Additional R matrix commands (cont)

<code>R <- chol(A)</code>	Choleski factorization of A . Returns the upper triangular factor, such that $R'R = A$.
<code>y <- qr(A)</code>	QR decomposition of A . <code>y\$qr</code> has an upper triangle that contains the decomposition and a lower triangle that contains information on the Q decomposition. <code>y\$rank</code> is the rank of A . <code>y\$qraux</code> a vector which contains additional information on Q. <code>y\$pivot</code> contains information on the pivoting strategy used.
<code>cbind(A,B,...)</code>	Combine matrices(vectors) horizontally. Returns a matrix.
<code>rbind(A,B,...)</code>	Combine matrices(vectors) vertically. Returns a matrix.
<code>rowMeans(A)</code>	Returns vector of row means.
<code>rowSums(A)</code>	Returns vector of row sums.
<code>colMeans(A)</code>	Returns vector of column means.
<code>colSums(A)</code>	Returns vector of column means.

Additional references

- Lynch & Walsh Chapter 8 (intro to matrices)
- Walsh and Lynch,
 - Appendix 5 (Matrix geometry)
 - Appendix 6 (Matrix derivatives)

PAUSE

- Many of the key results in linear and mixed models arise by considering the regression of one subset of a vector of random variable on another (slides 44-46)
- We conclude with a few optional slides on the singular value decomposition, the generalization of eigenstructure to any matrix (such as nonsquare matrices).
- The SVD arises when consider certain $G \times E$ problems.

The Singular-Value Decomposition (SVD)

An $n \times p$ matrix \mathbf{A} can always be decomposed as the product of three matrices: an $n \times p$ diagonal matrix $\mathbf{\Lambda}$ and two unitary matrices, \mathbf{U} which is $n \times n$ and \mathbf{V} which is $p \times p$. The resulting **singular value decomposition (SVD)** of \mathbf{A} is given by

$$\mathbf{A}_{n \times p} = \mathbf{U}_{n \times n} \mathbf{\Lambda}_{n \times p} \mathbf{V}_{p \times p}^T \quad (39.16a)$$

We have indicated the dimensionality of each matrix to allow the reader to verify that each matrix multiplication conforms. The diagonal elements $\lambda_1, \dots, \lambda_s$ of $\mathbf{\Lambda}$ correspond to the **singular values** of \mathbf{A} and are ordered by decreasing magnitude. Returning to the unitary matrices \mathbf{U} and \mathbf{V} , we can write each as a row vector of column vectors,

$$\mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_i, \dots, \mathbf{u}_n), \quad \mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_i, \dots, \mathbf{v}_p) \quad (39.16b)$$

where \mathbf{u}_i and \mathbf{v}_i are n and p -dimensional column vectors (often called the **left** and **right singular vectors**, respectively). Since both \mathbf{U} and \mathbf{V} are unitary, by definition (Appendix 4) each column vector has length one and are mutually orthogonal (i.e., if $i \neq j$, $\mathbf{u}_i \mathbf{u}_j^T = \mathbf{v}_i \mathbf{v}_j^T = 0$). Since $\mathbf{\Lambda}$ is diagonal, it immediately follows from matrix multiplication that we can write any element in \mathbf{A} as

$$A_{ij} = \sum_{k=1}^s \lambda_k u_{ik} v_{kj} \quad (39.16c)$$

where λ_k is the k th singular value and $s \leq \min(p, n)$ is the number of non-zero singular values.

The importance of the singular value decomposition in the analysis of $G \times E$ arises from the **Eckart-Young theorem** (1938), which relates the best approximation of a matrix by some lower-rank (say k) matrix with the SVD. Define as our measure of goodness of fit between a matrix \mathbf{A} and a lower rank approximation $\hat{\mathbf{A}}$ as the sum of squared differences over all elements,

$$\sum_{ij} (A_{ij} - \hat{A}_{ij})^2$$

Eckart and Young show that the best fitting approximation $\hat{\mathbf{A}}$ of rank $m < s$ is given from the first m terms of the singular value decomposition (the **rank- m SVD**),

$$\hat{A}_{ij} = \sum_{k=1}^m \lambda_k u_{ik} v_{kj} \quad (39.17a)$$

For example, the best rank-2 approximation for the $G \times E$ interaction is given by

$$GE_{ij} \simeq \lambda_1 u_{i1} v_{j1} + \lambda_2 u_{i2} v_{j2} \quad (39.17b)$$

where λ_i is the i th singular value of the \mathbf{GE} matrix, \mathbf{u} and \mathbf{v} are the associated singular vectors (see Example 39.3). The fraction of total variation of a matrix accounted for by taking the first m terms in its SVD is

$$\sum_{k=1}^m \lambda_k^2 / \sum_{ij} A_{ij}^2 = \frac{\lambda_1^2 + \dots + \lambda_m^2}{\lambda_1^2 + \dots + \lambda_s^2}$$

A data set for soybeans grown in New York (Gauch 1992) gives the GE matrix as

$$\mathbf{GE} = \begin{pmatrix} 57 & 176 & -233 \\ -36 & -196 & 233 \\ -45 & -324 & 369 \\ -66 & 178 & -112 \\ 89 & 165 & -254 \end{pmatrix}$$

Where GE_{ij} = value for Genotype i in enviro. j

In \mathbf{R} , the compact SVD (Equation 39.16d) of a matrix X is given by $\mathbf{svd}(X)$, returning the SVD of \mathbf{GE} as

$$\begin{pmatrix} 0.40 & 0.21 & 0.18 \\ -0.41 & 0.00 & 0.91 \\ -0.66 & 0.12 & -0.30 \\ 0.26 & -0.83 & 0.11 \\ 0.41 & 0.50 & 0.19 \end{pmatrix} \begin{pmatrix} 746.10 & 0 & 0 \\ 0 & 131.36 & 0 \\ 0 & 0 & 0.53 \end{pmatrix} \begin{pmatrix} 0.12 & 0.64 & -0.76 \\ 0.81 & -0.51 & -0.30 \\ 0.58 & 0.58 & 0.58 \end{pmatrix}$$

The first singular value accounts for $746.10^2 / (746.10^2 + 131.36^2 + 0.53^2) = 97.0\%$ of the total variation of \mathbf{GE} , while the second singular value accounts for 3.0%, so that together they account for essentially all of the total variation. The rank-1 SVD approximation of \mathbf{GE} is given by setting all of the diagonal elements of $\mathbf{\Lambda}$ except the first entry to zero,

$$\mathbf{GE}_1 = \begin{pmatrix} 0.40 & 0.21 & 0.18 \\ -0.41 & 0.00 & 0.91 \\ -0.66 & 0.12 & -0.30 \\ 0.26 & -0.83 & 0.11 \\ 0.41 & 0.50 & 0.19 \end{pmatrix} \begin{pmatrix} 746.10 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0.12 & 0.64 & -0.76 \\ 0.81 & -0.51 & -0.30 \\ 0.58 & 0.58 & 0.58 \end{pmatrix}$$

Similarly, the rank-2 SVD is given by setting all but the first two singular values to zero,

$$\mathbf{GE}_2 = \begin{pmatrix} 0.40 & 0.21 & 0.18 \\ -0.41 & 0.00 & 0.91 \\ -0.66 & 0.12 & -0.30 \\ 0.26 & -0.83 & 0.11 \\ 0.41 & 0.50 & 0.19 \end{pmatrix} \begin{pmatrix} 746.10 & 0 & 0 \\ 0 & 131.36 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0.12 & 0.64 & -0.76 \\ 0.81 & -0.51 & -0.30 \\ 0.58 & 0.58 & 0.58 \end{pmatrix}$$

For example, the rank-1 SVD approximation for \mathbf{GE}_{32} is

$$g_{31}\lambda_1 e_{12} = 746.10 * (-0.66) * 0.64 = -315$$

While the rank-2 SVD approximation is $g_{31}\lambda_2 e_{12} + g_{32}\lambda_2 e_{22} = 746.10 * (-0.66) * 0.64 + 131.36 * 0.12 * (-0.51) = -323$

Actual value is -324

Generally, the rank-2 SVD approximation for \mathbf{GE}_{ij} is

$$g_{i1}\lambda_1 e_{1j} + g_{i2}\lambda_2 e_{2j}$$