More Linear Algebra[1]

STA 431: Fall 2023

[1] See Appendix A for more detail. This slide show is an open-source document. See last slide for copyright information.

Overview

You already know about

- Matrices $\mathbf{A} = [a_{ij}]$
- Matrix addition and subtraction $\mathbf{A} + \mathbf{B} = [a_{ij} + b_{ij}]$
- Column vectors $\mathbf{v} = [v_j]$
- Scalar multiplication $a\mathbf{B} = [a\, b_{ij}]$
- Matrix multiplication $\mathbf{AB} = \left[\sum_k a_{ik} b_{kj} \right]$

  In words: The i, j element of $\mathbf{AB}$ is the inner product of row i of $\mathbf{A}$ with column j of $\mathbf{B}$.
- Inverse $\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$
- Transpose $\mathbf{A}^\top = [a_{ji}]$
- Symmetric matrices $\mathbf{A} = \mathbf{A}^\top$
- Determinants
- Linear independence

Three mistakes that will get you a zero

Numbers are 1×1 matrices, but larger matrices are not just numbers.

You will get a zero if you

- Write $\mathbf{AB} = \mathbf{BA}$. It's not true in general.
- Write $\mathbf{A}^{-1}$ when $\mathbf{A}$ is not a square matrix. The inverse is not even defined.
- Represent the inverse of a matrix (even if it exists) by writing it in the denominator, like $\mathbf{a}^\top \mathbf{B}^{-1} \mathbf{a} = \frac{\mathbf{a}^\top \mathbf{a}}{\mathbf{B}}$. Matrices are not just numbers.

If you commit one of these crimes, the mark for the question (or part of a question, like 3c) is zero, regardless of what else you write.

Half marks off, at least

You will lose *at least* half marks for writing a product like **AB** when the number of columns in **A** does not equal the number of rows in **B**.

Trace of a square matrix: Sum of the diagonal elements

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

- Obvious: $tr(\mathbf{A}+\mathbf{B}) = tr(\mathbf{A}) + tr(\mathbf{B})$.
- Not obvious: $tr(\mathbf{AB}) = tr(\mathbf{BA})$
- Even though $\mathbf{AB} \neq \mathbf{BA}$

Example

Let $\mathbf{A} = \begin{pmatrix} 2 & 1 & 0 \\ 5 & -4 & 3 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 1 & 0 \\ 2 & 3 \\ -1 & 3 \end{pmatrix}$

$$\mathbf{AB} = \begin{pmatrix} 4 & 3 \\ -6 & -3 \end{pmatrix}$$

$$\mathbf{BA} = \begin{pmatrix} 2 & 1 & 0 \\ 19 & -10 & 9 \\ 13 & -13 & 9 \end{pmatrix}$$

And $tr(\mathbf{AB}) = tr(\mathbf{BA})$.

Eigenvalues and eigenvectors

Let $\mathbf{A} = [a_{i,j}]$ be a square matrix. $\mathbf{A}$ is said to have an *eigenvalue* λ and *eigenvector* $\mathbf{x} \neq \mathbf{0}$ corresponding to λ if

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}.$$

Recall

- Eigenvalues are the λ values that solve the determinantal equation $|\mathbf{A} - \lambda\mathbf{I}| = 0$.
- The determinant is the product of the eigenvalues: $|\mathbf{A}| = \prod_{i=1}^{n} \lambda_i$

Spectral decomposition of symmetric matrices

The *Spectral decomposition theorem* says that every square and symmetric matrix $\mathbf{A} = [a_{i,j}]$ may be written

$$\mathbf{A} = \mathbf{C}\mathbf{D}\mathbf{C}^\top,$$

where the columns of $\mathbf{C}$ (which may also be denoted $\mathbf{x}_1, \ldots, \mathbf{x}_n$) are the eigenvectors of $\mathbf{A}$, and the diagonal matrix $\mathbf{D}$ contains the corresponding eigenvalues.

$$\mathbf{D} = \left(\begin{array}{cccc} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{array} \right)$$

- If the elements of $\mathbf{A}$ are real, the eigenvalues are real.
- The eigenvectors may be chosen to be orthonormal, so that $\mathbf{C}$ is an orthogonal matrix. That is, $\mathbf{C}\mathbf{C}^\top = \mathbf{C}^\top\mathbf{C} = \mathbf{I}$.

Inverse of a diagonal matrix

Suppose the eigenvalues are all non-zero. Let

$$\mathbf{D}^{-1} = \begin{pmatrix} \frac{1}{\lambda_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\lambda_n} \end{pmatrix}$$

It works because

$$\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \begin{pmatrix} \frac{1}{\lambda_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\lambda_n} \end{pmatrix} = \mathbf{I}$$

Square root of a diagonal matrix

Suppose the eigenvalues are non-negative. Let

$$\mathbf{D}^{1/2} = \begin{pmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\lambda_n} \end{pmatrix}$$

It works because

$$\begin{aligned} \mathbf{D}^{1/2}\mathbf{D}^{1/2} &= \begin{pmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\lambda_n} \end{pmatrix} \begin{pmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\lambda_n} \end{pmatrix} \\ &= \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} = \mathbf{D} \end{aligned}$$

Using $\mathbf{A} = \mathbf{C}\mathbf{D}\mathbf{C}^\top$

Where $\mathbf{A}$ is a symmetric matrix

$$
\begin{aligned}
\mathbf{A}^{-1} &= \mathbf{C}\mathbf{D}^{-1}\mathbf{C}^\top \\
\mathbf{A}^{1/2} &= \mathbf{C}\mathbf{D}^{1/2}\mathbf{C}^\top \\
\mathbf{A}^{-1/2} &= \mathbf{C}\mathbf{D}^{-1/2}\mathbf{C}^\top
\end{aligned}
$$

Positive definite matrices

The $n \times n$ matrix $\mathbf{A}$ is said to be *positive definite* if

$$\mathbf{y}^\top \mathbf{A}\mathbf{y} > 0$$

for *all* $n \times 1$ vectors $\mathbf{y} \neq \mathbf{0}$.

It is called *non-negative definite* (or sometimes positive semi-definite) if $\mathbf{y}^\top \mathbf{A}\mathbf{y} \geq 0$.

Some properties of symmetric positive definite matrices

Variance-covariance matrices are often assumed positive definite.

For a symmetric matrix,

Positive definite

$\Downarrow$

All eigenvalues positive

$\Downarrow$

Inverse exists $\Leftrightarrow$ Columns (rows) linearly independent.

If a real symmetric matrix is also non-negative definite (as a variance-covariance matrix *must* be) Linear independence $\Rightarrow$ Positive definite.

Extras

You may not know about these, and we may use them occasionally

- Rank
- Partitioned matrices

Rank

- Row rank is the number of linearly independent rows.
- Column rank is the number of linearly independent columns.
- Rank of a matrix is the minimum of row rank and column rank.

- $rank(\mathbf{AB}) = \min\left(rank(\mathbf{A}), rank(\mathbf{B})\right)$.

Partitioned matrix

- A matrix of matrices

$$\left[\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline \mathbf{C} & \mathbf{D} \end{array}\right]$$

- Row by column (matrix) multiplication works, provided the matrices are the right sizes.

Matrix calculation with R

```
> is.matrix(3)  # Is the number 3 a 1x1 matrix?

[1] FALSE
```

```
> treecorr = cor(trees); treecorr

           Girth    Height    Volume
Girth  1.0000000 0.5192801 0.9671194
Height 0.5192801 1.0000000 0.5982497
Volume 0.9671194 0.5982497 1.0000000
```

```
> is.matrix(treecorr)

[1] TRUE
```

Creating matrices
Bind rows into a matrix

```
> # Bind rows of a matrix together
> A = rbind(  c(3, 2, 6,8),
+             c(2,10,-7,4),
+             c(6, 6, 9,1)   ); A

     [,1] [,2] [,3] [,4]
[1,]    3    2    6    8
[2,]    2   10   -7    4
[3,]    6    6    9    1


> # Transpose
> t(A)

     [,1] [,2] [,3]
[1,]    3    2    6
[2,]    2   10    6
[3,]    6   -7    9
[4,]    8    4    1
```

Matrix multiplication

Remember, **A** is 3×4

```
> # U = A A' (3x3), V = A' A (4x4)
> U = A %*% t(A)
> V = t(A) %*% A; V

     [,1] [,2] [,3] [,4]
[1,]   49   62   58   38
[2,]   62  140   -4   62
[3,]   58   -4  166   29
[4,]   38   62   29   81
```

Determinants

$\mathbf{A}$ is 3×4

```
> # U = A A' (3x3), V = A' A (4x4)
> # So rank(V) cannot exceed 3 and det(V)=0
> det(U); det(V)

[1] 1490273
[1] -3.622862e-09
```

Inverse of $\mathbf{U}$ exists, but inverse of $\mathbf{V}$ does not.

Inverses

- The `solve` function is for solving systems of linear equations like $\mathbf{Mx} = \mathbf{b}$.
- Just typing `solve(M)` gives $\mathbf{M}^{-1}$.

```
> # Recall U = A A' (3x3), V = A' A (4x4)
> solve(U)

             [,1]          [,2]          [,3]
[1,]  0.0173505123 -8.508508e-04 -1.029342e-02
[2,] -0.0008508508  5.997559e-03  2.013054e-06
[3,] -0.0102934160  2.013054e-06  1.264265e-02


> solve(V)


Error in solve.default(V) :
  system is computationally singular: reciprocal condition
  number = 6.64193e-18
```

Eigenvalues and eigenvectors

```
> # Recall  U = A A' (3x3), V = A' A (4x4)
> eigen(U)

$values
[1] 234.01162 162.89294  39.09544

$vectors
           [,1]       [,2]        [,3]
[1,] -0.6025375  0.1592598  0.78203893
[2,] -0.2964610 -0.9544379 -0.03404605
[3,] -0.7409854  0.2523581 -0.62229894
```

$\mathbf{V}$ should have at least one zero eigenvalue

Because $\mathbf{A}$ is 3×4, $\mathbf{V} = \mathbf{A}^\top \mathbf{A}$, and the rank of a product is the minimum rank of the matrices.

```
> eigen(V)

$values
[1]  2.340116e+02  1.628929e+02  3.909544e+01 -1.012719e-14

$vectors
           [,1]         [,2]       [,3]         [,4]
[1,] -0.4475551  0.006507269 -0.2328249  0.863391352
[2,] -0.5632053 -0.604226296 -0.4014589 -0.395652773
[3,] -0.5366171  0.776297432 -0.1071763 -0.312917928
[4,] -0.4410627 -0.179528649  0.8792818  0.009829883
```

Spectral decomposition $\mathbf{V} = \mathbf{C}\mathbf{D}\mathbf{C}^\top$

```
> eigenV = eigen(V)
> C = eigenV$vectors; D = diag(eigenV$values); D

         [,1]     [,2]     [,3]          [,4]
[1,] 234.0116   0.0000  0.00000  0.000000e+00
[2,]   0.0000 162.8929  0.00000  0.000000e+00
[3,]   0.0000   0.0000 39.09544  0.000000e+00
[4,]   0.0000   0.0000  0.00000 -1.012719e-14
```

```
> # C is an orthoganal matrix
> C %*% t(C)

              [,1]         [,2]         [,3]          [,4]
[1,]  1.000000e+00 5.551115e-17 0.000000e+00 -3.989864e-17
[2,]  5.551115e-17 1.000000e+00 2.636780e-16  3.556183e-17
[3,]  0.000000e+00 2.636780e-16 1.000000e+00  2.558717e-16
[4,] -3.989864e-17 3.556183e-17 2.558717e-16  1.000000e+00
```

Verify $\mathbf{V} = \mathbf{C}\mathbf{D}\mathbf{C}^\top$

```
> V;    C %*% D %*% t(C)

     [,1] [,2] [,3] [,4]
[1,]   49   62   58   38
[2,]   62  140   -4   62
[3,]   58   -4  166   29
[4,]   38   62   29   81

     [,1] [,2] [,3] [,4]
[1,]   49   62   58   38
[2,]   62  140   -4   62
[3,]   58   -4  166   29
[4,]   38   62   29   81
```

Square root matrix $\mathbf{V}^{1/2} = \mathbf{C}\mathbf{D}^{1/2}\mathbf{C}^\top$

```
> sqrtV = C %*% sqrt(D) %*% t(C)

Warning message:
In sqrt(D) : NaNs produced

> # Multiply to get V
> sqrtV %*% sqrtV;  V

     [,1] [,2] [,3] [,4]
[1,]  NaN  NaN  NaN  NaN
[2,]  NaN  NaN  NaN  NaN
[3,]  NaN  NaN  NaN  NaN
[4,]  NaN  NaN  NaN  NaN
     [,1] [,2] [,3] [,4]
[1,]   49   62   58   38
[2,]   62  140   -4   62
[3,]   58   -4  166   29
[4,]   38   62   29   81
```

What happened?

```
> D; sqrt(D)

         [,1]     [,2]     [,3]          [,4]
[1,] 234.0116   0.0000  0.00000  0.000000e+00
[2,]   0.0000 162.8929  0.00000  0.000000e+00
[3,]   0.0000   0.0000 39.09544  0.000000e+00
[4,]   0.0000   0.0000  0.00000 -1.012719e-14

         [,1]     [,2]     [,3] [,4]
[1,] 15.29744  0.00000 0.000000    0
[2,]  0.00000 12.76295 0.000000    0
[3,]  0.00000  0.00000 6.252635    0
[4,]  0.00000  0.00000 0.000000  NaN


Warning message:
In sqrt(D) : NaNs produced
```

Copyright Information

This slide show was prepared by Jerry Brunner, Department of Statistical Sciences, University of Toronto. It is licensed under a Creative Commons Attribution - ShareAlike 3.0 Unported License. Use any part of it as you like and share the result freely. The LaTeX source code is available from the course website:

http://www.utstat.toronto.edu/brunner/oldclass/431s23