Review of Linear Algebra

Max Turgeon

STAT 7200–Multivariate Statistics

Eigenvalues

- Let $\mathbf{A}$ be a square $n \times n$ matrix.
- The equation
$$\det(\mathbf{A} - \lambda I_n) = 0$$
is called the *characteristic equation* of $\mathbf{A}$.
- This is a polynomial equation of degree n, and its roots are called the *eigenvalues* of $\mathbf{A}$.

Example i

```
(A <- matrix(c(1, 2, 3, 2), ncol = 2))
```

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

```
eigen(A)$values
```

```
## [1]  4 -1
```

A few properties

Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $\mathbf{A}$ (with multiplicities).

1. $\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^n \lambda_i$;
2. $\det(\mathbf{A}) = \prod_{i=1}^n \lambda_i$;
3. The eigenvalues of $\mathbf{A}^k$ are $\lambda_1^k, \ldots, \lambda_n^k$, for k a nonnegative integer;
4. If $\mathbf{A}$ is invertible, then the eigenvalues of $\mathbf{A}^{-1}$ are $\lambda_1^{-1}, \ldots, \lambda_n^{-1}$.
5. If $\mathbf{A}$ is symmetric, all eigenvalues are *real*. (**Exercise**: Prove this.)

Eigenvectors

- If λ is an eigenvalue of $\mathbf{A}$, then (by definition) we have $\det(\mathbf{A} - \lambda I_n) = 0$.
- In other words, the following equivalent statements hold:
  - The matrix $\mathbf{A} - \lambda I_n$ is singular;
  - The kernel space of $\mathbf{A} - \lambda I_n$ is nontrivial (i.e. not equal to the zero vector);
  - The system of equations $(\mathbf{A} - \lambda I_n)v = 0$ has a nontrivial solution;
  - There exists a nonzero vector v such that

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

- Such a vector is called an *eigenvector* of $\mathbf{A}$.

Example (cont'd)

```
eigen(A)$vectors
```

```
##            [,1]       [,2]
## [1,] -0.7071068 -0.8320503
## [2,] -0.7071068  0.5547002
```

Spectral Decomposition

Theorem
Let $\mathbf{A}$ be an $n \times n$ symmetric matrix, and let $\lambda_1 \geq \cdots \geq \lambda_n$ be its eigenvalues (with multiplicity). Then there exist vectors $v_1, \ldots, v_n$ such that

1. $\mathbf{A}v_i = \lambda_i v_i$, i.e. v_i is an eigenvector, for all i;
2. If $i \neq j$, then $v_i^T v_j = 0$, i.e. they are orthogonal;
3. For all i, we have $v_i^T v_i = 1$, i.e. they have unit norm;
4. We can write $\mathbf{A} = \sum_{i=1}^n \lambda_i v_i v_i^T$.

In matrix form: $\mathbf{A} = \mathbf{V}\Lambda\mathbf{V}^T$, where the columns of $\mathbf{V}$ are the vectors v_i, and Λ is a diagonal matrix with the eigenvalues λ_i on its diagonal.

Positive-definite matrices

Let $\mathbf{A}$ be a real symmetric matrix, and let $\lambda_1 \geq \cdots \geq \lambda_n$ be its (real) eigenvalues.

1. If $\lambda_i > 0$ for all i, we say $\mathbf{A}$ is *positive definite*.
2. If the inequality is not strict, if $\lambda_i \geq 0$, we say $\mathbf{A}$ is *positive semidefinite*.
3. Similary, if $\lambda_i < 0$ for all i, we say $\mathbf{A}$ is *negative definite*.
4. If the inequality is not strict, if $\lambda_i \leq 0$, we say $\mathbf{A}$ is *negative semidefinite*.

Note: If $\mathbf{A}$ is *positive-definite*, then it is invertible!

Matrix Square Root i

- Let $\mathbf{A}$ be a positive semidefinite symmetric matrix.
- By the Spectral Decomposition, we can write

$$\mathbf{A} = P\Lambda P^T.$$

- Since $\mathbf{A}$ is positive-definite, we know that the elements on the diagonal of Λ are positive.
- Let $\Lambda^{1/2}$ be the diagonal matrix whose entries are the square root of the entries on the diagonal of Λ.
- For example:

$$\Lambda = \begin{pmatrix} 1.5 & 0 \\ 0 & 0.5 \end{pmatrix} \Rightarrow \Lambda^{1/2} = \begin{pmatrix} 1.2247 & 0 \\ 0 & 0.7071 \end{pmatrix}.$$

Matrix Square Root ii

- We define the square root $\mathbf{A}^{1/2}$ of $\mathbf{A}$ as follows:

$$\mathbf{A}^{1/2} := P\Lambda^{1/2}P^T.$$

- *Check*:

$$\begin{aligned}
\mathbf{A}^{1/2}\mathbf{A}^{1/2} &= (P\Lambda^{1/2}P^T)(P\Lambda^{1/2}P^T) \\
&= P\Lambda^{1/2}(P^TP)\Lambda^{1/2}P^T \\
&= P\Lambda^{1/2}\Lambda^{1/2}P^T \qquad (P \text{ is orthogonal}) \\
&= P\Lambda P^T \\
&= \mathbf{A}.
\end{aligned}$$

Matrix Square Root iii

- *Be careful*: your intuition about square roots of positive real numbers doesn't translate to matrices.
    - In particular, matrix square roots are **not** unique (unless you impose further restrictions).

Cholesky Decomposition

- Another common way to obtain a square root matrix for a positive definite matrix $\mathbf{A}$ is via the *Cholesky decomposition*.
- There exists a unique matrix L such that:
  - L is lower triangular (i.e. all entries above the diagonal are zero);
  - The entries on the diagonal are positive;
  - $\mathbf{A} = LL^T$.
- For matrix square roots, the Cholesky decomposition should be prefered to the eigenvalue decomposition because:
  - It is computationally more efficient;
  - It is numerically more stable.

Example i

```
A <- matrix(c(1, 0.5, 0.5, 1), nrow = 2)

# Eigenvalue method
result <- eigen(A)
Lambda <- diag(result$values)
P <- result$vectors
A_sqrt <- P %*% Lambda^0.5 %*% t(P)

all.equal(A, A_sqrt %*% A_sqrt) # CHECK
```

```
## [1] TRUE
```

Example ii

```r
# Cholesky method
# It's upper triangular!
(L <- chol(A))
```

```
##      [,1]      [,2]
## [1,]    1 0.5000000
## [2,]    0 0.8660254
```

```r
all.equal(A, t(L) %*% L) # CHECK
```

```
## [1] TRUE
```

Singular Value Decomposition i

- We saw earlier that real symmetric matrices are *diagonalizable*, i.e. they admit a decomposition of the form $P\Lambda P^T$ where
    - Λ is diagonal;
    - P is orthogonal, i.e. $PP^T = P^TP = I$.
- For a general $n \times p$ matrix $\mathbf{A}$, we have the *Singular Value Decomposition* (SVD).
- We can write $\mathbf{A} = UDV^T$, where
    - U is an $n \times n$ orthogonal matrix;
    - V is a $p \times p$ orthogonal matrix;
    - D is an $n \times p$ diagonal matrix.

Singular Value Decomposition ii

- We say that:
    - the columns of U are the *left-singular vectors* of $\mathbf{A}$;
    - the columns of V are the *right-singular vectors* of $\mathbf{A}$;
    - the nonzero entries of D are the *singular values* of $\mathbf{A}$.

Example i

```r
set.seed(1234)
A <- matrix(rnorm(3 * 2), ncol = 2, nrow = 3)
result <- svd(A)
names(result)
```

```
## [1] "d" "u" "v"
```

```r
result$d
```

```
## [1] 2.8602018 0.6868562
```

Example ii

```r
result$u
```

```
##              [,1]         [,2]
## [1,] -0.9182754 -0.359733536
## [2,]  0.1786546 -0.003617426
## [3,]  0.3533453 -0.933048068
```

```r
result$v
```

```
##             [,1]       [,2]
## [1,] 0.5388308 -0.8424140
## [2,] 0.8424140  0.5388308
```

Example iii

```r
D <- diag(result$d)
all.equal(A, result$u %*% D %*% t(result$v)) #CHECK
```

```
## [1] TRUE
```

Example iv

```r
# Note: crossprod(A) == t(A) %*% A
# tcrossprod(A) == A %*% t(A)
U <- eigen(tcrossprod(A))$vectors
V <- eigen(crossprod(A))$vectors

D <- matrix(0, nrow = 3, ncol = 2)
diag(D) <- result$d

all.equal(A, U %*% D %*% t(V)) # CHECK
```

```
## [1] "Mean relative difference: 1.95887"
```

Example v

```r
# What went wrong?
# Recall that eigenvectors are unique
# only up to a sign!

# These elements should all be positive
diag(t(U) %*% A %*% V)
```

```
## [1] -2.8602018  0.6868562
```

Example vi

```
# Therefore we need to multiply the
# corresponding columns of U or V
# (but not both!) by -1
cols_flip <- which(diag(t(U) %*% A %*% V) < 0)
V[,cols_flip] <- -V[,cols_flip]

all.equal(A, U %*% D %*% t(V)) # CHECK
```

```
## [1] TRUE
```