Lecture 2: Intro/refresher in Matrix Algebra

Bruce Walsh lecture notes
2013 Synbreed Course
version 1 July 2013

Matrix/linear algebra

- Compact way for treating the algebra of systems of linear equations
- Most common statistical methods can be written in matrix form
  - $\mathbf{y} = \mathbf{X}\beta + \mathbf{e}$ is the general linear model
    - OLS solution: $\beta = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$
  - $\mathbf{Y} = \mathbf{X}\beta + \mathbf{Za} + \mathbf{e}$ is the general mixed model

Topics

- Definitions, dimensionality, addition, subtraction
- Matrix multiplication
- Inverses, solving systems of equations
- Quadratic products and covariances
- The multivariate normal distribution
- Ordinary least squares
- Vector/matrix calculus (taking derivatives)

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**

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

Column vector (3 x 1)

Row vector (1 x 4)

Dimensionality of a matrix: r x 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)

Square matrix

(3 x 2)

Dimensionality of a matrix: r x 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 (same number of rows and columns), then matrix addition and subtraction simply follows by adding (or subtracting) on an element by element basis

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

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.

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

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

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:

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

a ·b = 1*4 + 2*5 + 3*7 + 4*9 = 60

NOT defined if a and b have different lengths

Matrices are compact ways to write systems of equations

$$5x_1 + 6x_2 + 4x_3 = 6$$
$$7x_1 - 3x_2 + 5x_3 = -9$$
$$-x_1 - x_2 + 6x_3 = 12$$

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

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

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

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

This can be more compactly written in matrix form as

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

$\mathbf{X}^T\mathbf{X}$ $\qquad$ $\boldsymbol{\beta}$ $\qquad$ $\mathbf{X}^T\mathbf{y}$

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

Matrix Multiplication:

The order in which matrices are multiplied effects the matrix product, e.g. **AB ≠ 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.

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

ij-th element of C is given by

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

Elements in the jth column of B

Elements in the ith row of matrix A

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

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

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

$$\mathbf{M} = \begin{pmatrix} \mathbf{m_1} \\ \mathbf{m_2} \\ \vdots \\ \mathbf{m_r} \end{pmatrix} \qquad \text{where} \qquad \mathbf{m_i} = (\, M_{i1} \quad M_{i2} \quad \cdots \quad M_{ic} \,)$$

Likewise express N as a row vector of column vectors

$$\mathbf{N} = (\, \mathbf{n_1} \quad \mathbf{n_2} \quad \cdots \quad \mathbf{n_b} \,) \qquad \text{where} \qquad \mathbf{n_j} = \begin{pmatrix} N_{1j} \\ N_{2j} \\ \vdots \\ N_{cj} \end{pmatrix}$$

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

$$\mathbf{L} = \begin{pmatrix} \mathbf{m_1} \cdot \mathbf{n_1} & \mathbf{m_1} \cdot \mathbf{n_2} & \cdots & \mathbf{m_1} \cdot \mathbf{n_b} \\ \mathbf{m_2} \cdot \mathbf{n_1} & \mathbf{m_2} \cdot \mathbf{n_2} & \cdots & \mathbf{m_2} \cdot \mathbf{n_b} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{m_r} \cdot \mathbf{n_1} & \mathbf{m_r} \cdot \mathbf{n_2} & \cdots & \mathbf{m_r} \cdot \mathbf{n_b} \end{pmatrix}$$

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_{3x5} D_{5x5} which gives a 3 x 5 matrix, versus D_{5x5} C_{3x5} , which is not defined

The Transpose of a Matrix

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

Useful identities

$$(AB)^T = B^T A^T$$

$$(ABC)^T = C^T B^T A^T$$

$$\mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

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

Indices match, matrices conform

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

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

Note that $b^T a = (b^T a)^T = a^T b$

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

Resulting product is an n x n matrix

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \begin{pmatrix} b_1 & b_2 & \cdots & b_n \end{pmatrix}$$

$$= \begin{pmatrix} a_1b_1 & a_1b_2 & \dots & a_1b_n \\ a_2b_1 & a_2b_2 & \dots & a_2b_n \\ \vdots & \vdots & \ddots & \vdots \\ a_nb_1 & a_nb_2 & \dots & a_nb_{bn} \end{pmatrix}$$

Solving equations

- The identity matrix **I**
  - Serves the same role as 1 in scalar algebra, e.g., a*1 = 1*a = a, with $\mathbf{AI}=\mathbf{IA}= \mathbf{A}$
- The inverse matrix $\mathbf{A}^{-1}$ (IF it exists)
  - Defined by $\mathbf{A}\ \mathbf{A}^{-1} = \mathbf{I}$, $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$
  - 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 $\mathbf{Ax} = \mathbf{c}$, $A^{-1}\mathbf{Ax} = A^{-1}\,\mathbf{c}$
    - Or $\mathbf{A}^{-1}\mathbf{Ax} = \mathbf{Ix} = \mathbf{x} = A^{-1}\,\mathbf{c}$

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.

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

$$\mathbf{I}_{3x3} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The Inverse Matrix, A^{-1}

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

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

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

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 $x_2 + x_3 = 6$)

Example: solve the OLS for β in $y = \alpha + \beta_1 z_1 + \beta_2 z_2 + e$

$\beta = V^{-1} c$

$$\mathbf{c} = \begin{pmatrix} \sigma(y, z_1) \\ \sigma(y, z_2) \end{pmatrix} \qquad \mathbf{V} = \begin{pmatrix} \sigma^2(z_1) & \sigma(z_1, z_2) \\ \sigma(z_1, z_2) & \sigma^2(z_2) \end{pmatrix}$$

It is more compact to use $\sigma(z_1, z_2) = \rho_{12}\, \sigma(z_1)\sigma(z_2)$

$$\mathbf{V}^{-1} = \frac{1}{\sigma^2(z_1)\sigma^2(z_2)\,(1-\rho_{12}^2)} \begin{pmatrix} \sigma^2(z_2) & -\sigma(z_1, z_2) \\ -\sigma(z_1, z_2) & \sigma^2(z_1) \end{pmatrix}$$

$$\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \frac{1}{\sigma^2(z_1)\sigma^2(z_2)\,(1-\rho_{12}^2)} \begin{pmatrix} \sigma^2(z_2) & -\sigma(z_1, z_2) \\ -\sigma(z_1, z_2) & \sigma^2(z_1) \end{pmatrix} \begin{pmatrix} \sigma(y, z_1) \\ \sigma(y, z_2) \end{pmatrix}$$

$$\beta_1 = \frac{1}{1-\rho_{12}^2}\left[\frac{\sigma(y,z_1)}{\sigma^2(z_1)} - \rho_{12}\,\frac{\sigma(y,z_2)}{\sigma(z_1)\sigma(z_2)}\right]$$

$$\beta_2 = \frac{1}{1-\rho_{12}^2}\left[\frac{\sigma(y,z_2)}{\sigma^2(z_2)} - \rho_{12}\,\frac{\sigma(y,z_1)}{\sigma(z_1)\sigma(z_2)}\right]$$

If $\rho_{12} = 0$, these reduce to the two univariate slopes,

$$\beta_1 = \frac{\sigma(y,z_1)}{\sigma^2(z_1)} \quad \text{and} \quad \beta_2 = \frac{\sigma(y,z_2)}{\sigma^2(z_2)}$$

Likewise, if $\rho_{12} = 1$, this reduces to a univariate regression,

Useful identities

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

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

For a diagonal matrix **D**, then Det = | **D** | = product of the diagonal elements

The determinant of any square matrix **A**, det(**A**), is the product of the **eigenvalues** λ of **A**, which statisfy

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

e is the **eigenvector** associated with λ
If *A* is n x n, solutions to λ are an n-th degree polynomial.
If any of the roots to the equation are zero, A^{-1} is not defined. Further, for some linear combination **b**, **Ab** = 0.

Variance-Covariance matrix

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

The trace

The trace, tr(**A**) or trace(**A**), of a square matrix *A* is simply the sum of its diagonal elements

The importance of the trace is that it equals the sum of the eigenvalues of **A**, tr(**A**) = $\Sigma \lambda_i$

For a covariance matrix **V**, tr(V) measures the total amount of variation in the variables

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

Quadratic and Bilinear Forms

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

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$$

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, $c_1x_1 + c_2x_2 + \ldots + c_nx_n$? (note this is a scalar)

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

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} \sigma\left(c_i\, x_i, c_j\, x_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i\, c_j\, \sigma\left(x_i, x_j\right)$$

$$= \mathbf{c}^T \mathbf{V}\, \mathbf{c}$$

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 new variables $y_1 = x_1 - 2x_2 + 5x_3$ and $y_2 = 6x_2 - 4x_3$?

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

$Var(y_1) = Var(c_1^T x) = c_1^T\, Var(x)\, c_1 = 960$

$Var(y_2) = Var(c_2^T x) = c_2^T\, Var(x)\, c_2 = 1200$

$Cov(y_1, y_2) = Cov(c_1^T x, c_2^T x) = c_1^T\, Var(x)\, c_2 = -910$

Now suppose we transform one vector of random variables into another vector of random variables

Transform $\mathbf{x}$ into

(i) $\mathbf{y}_{k \times 1} = \mathbf{A}_{k \times n} \mathbf{x}_{n \times 1}$

(ii) $\mathbf{z}_{m \times 1} = \mathbf{B}_{m \times n} \mathbf{x}_{n \times 1}$

The covariance between the elements of these two transformed vectors of the original is a $k \times m$ covariance matrix = $\mathbf{AVB}^T$

For example, the covariance between y_i and y_j is given by the ij-th element of $\mathbf{AVA}^T$

Likewise, the covariance between y_i and z_j is given by the ij-th element of $\mathbf{AVB}^T$

Positive-definite matrix

- A matrix **V** is positive-definite if for all vectors **c** contained at least one non-zero member, $\mathbf{c}^T\mathbf{V}\mathbf{c} > 0$.
- A non-negative definite matrix satisfies $\mathbf{c}^T\mathbf{V}\mathbf{c} \geq 0$.
- Any covariance-matrix is (at least) non-negative definite, as $\mathrm{Var}(\mathbf{c}^T\mathbf{x}) = \mathbf{c}^T\mathbf{V}\mathbf{c} \geq 0$.
- Any nonsingular covariance matrix is positive-definite
  - Nonsingular means $\det(V) > 0$
  - Equivalently, **all eigenvalues of V are positive**, $\lambda_i > 0$.

The Multivariate Normal Distribution (MVN)

Consider the pdf for n independent normal random variables, the ith of which has mean μ_i and variance σ^2_i

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

$$= (2\pi)^{-n/2} \left(\prod_{i=1}^{n} \sigma_i \right)^{-1} \exp\left(- \sum_{i=1}^{n} \frac{(x_i - \mu_i)^2}{2\,\sigma_i^2} \right)$$

This can be expressed more compactly in matrix form

Define the covariance matrix V for the vector x of the n normal random variable by

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

Define the mean vector μ by

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

Hence in matrix from the MVN pdf becomes

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

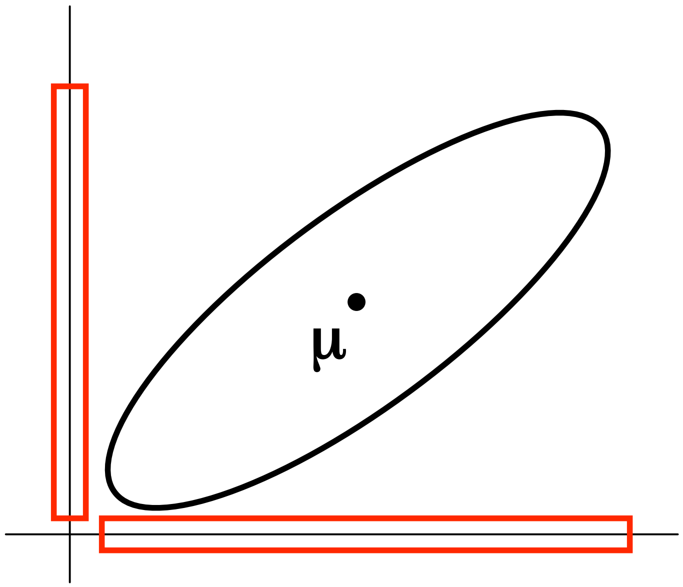
Notice this holds for any vector μ and symmetric positive-definite matrix V, as | V | > 0.

The multivariate normal

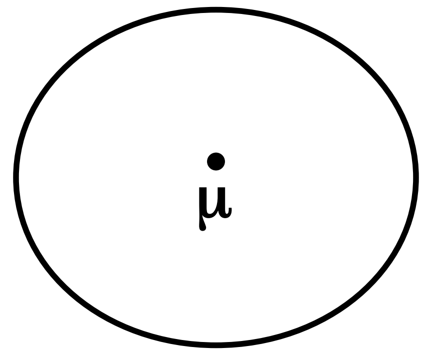
- Just as a univariate normal is defined by its mean and spread, a multivariate normal is defined by its mean vector μ (also called the centroid) and variance-covariance matrix **V**
- **V** determines the geometry (pattern of spread) of variation about the centroid. This is because $\mathbf{c}^T\mathbf{V}\mathbf{c} = a \geq 0$, the surface of all linear combinations ($\mathbf{c}^T\mathbf{x}$) with variance a defines a quadratic surface (an ellipse)

Vector of means μ determines location

Spread (geometry) about μ determined by **V**



x_1, x_2 equal variances, positively correlated

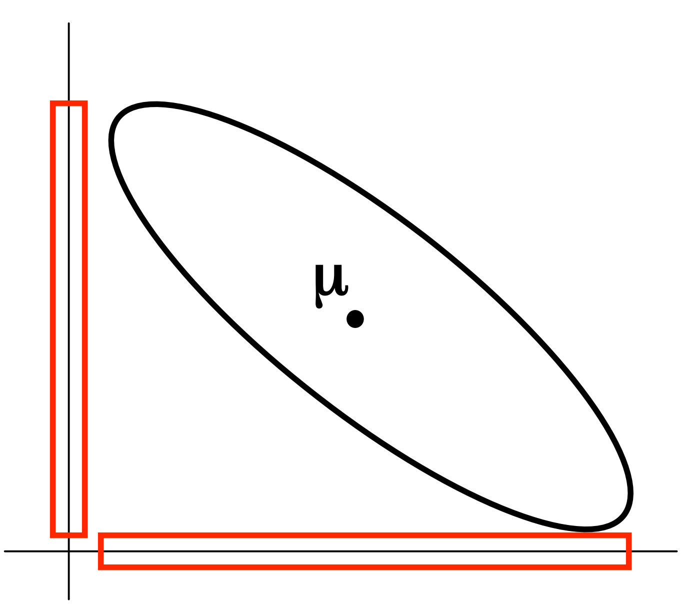


x_1, x_2 equal variances, uncorrelated

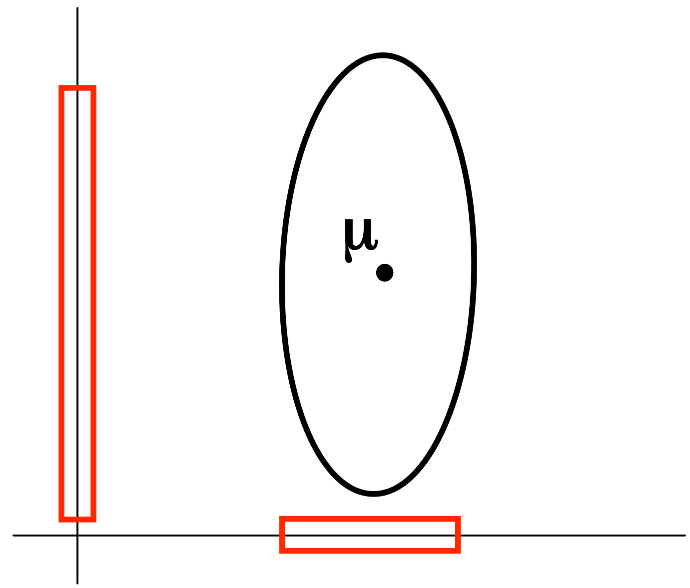
Eigenstructure (the eigenvectors and their corresponding eigenvalues) determines the *geometry* of **V**.

Vector of means μ determines location

Spread (geometry) about μ determined by **V**



x_1, x_2 equal variances, negatively correlated



$Var(x_1) < Var(x_2)$, uncorrelated

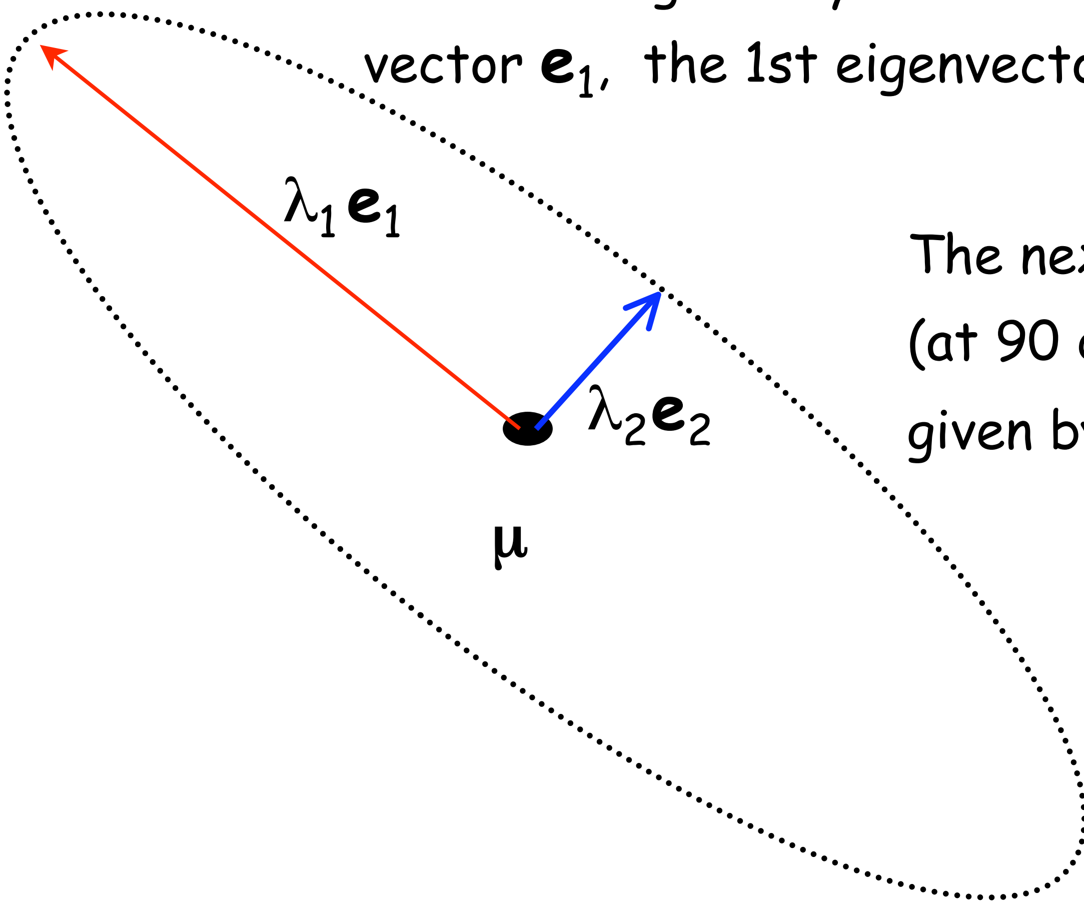
Positive tilt = positive correlations

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 $\mathbf{c}^T\mathbf{x}$) that explains the most variation.
  - PC2 is the next largest direction (at 90degree from PC1), and so on
- PCi = ith eigenvector of V
- Fraction of variation accounted for by PCi = λ_i / trace(V)
- If V has a few large eigenvalues, most of the variation is distributed along a few linear combinations (axis of variation)

Properties of the MVN - I

1) If **x** is MVN, any subset of the variables in **x** is also MVN

2) If **x** is MVN, any linear combination of the elements of **x** is also MVN. If **x** ~ MVN(μ,V)

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

$$\text{for} \quad y = \mathbf{a}^T \mathbf{x} = \sum_{k=1}^{n} a_i x_i, \qquad y \text{ is } \mathrm{N}(\mathbf{a}^T \boldsymbol{\mu}, \mathbf{a}^T \mathbf{V}\, \mathbf{a})$$

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

Properties of the MVN - II

3) Conditional distributions are also MVN. Partition x into two components, x_1 (m dimensional column vector) and x_2 (n-m dimensional column vector)

$$\mathbf{x} = \begin{pmatrix} \mathbf{x_1} \\ \mathbf{x_2} \end{pmatrix} \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu_1} \\ \boldsymbol{\mu_2} \end{pmatrix} \quad \text{and} \quad \mathbf{V} = \begin{pmatrix} \mathbf{V_{x_1 x_1}} & \mathbf{V_{x_1 x_2}} \\ \mathbf{V^T_{x_1 x_2}} & \mathbf{V_{x_2 x_2}} \end{pmatrix}$$

$x_1 \mid x_2$ is MVN with m-dimensional mean vector

$$\boldsymbol{\mu}_{\mathbf{x_1|x_2}} = \boldsymbol{\mu_1} + \mathbf{V_{x_1 x_2}} \mathbf{V^{-1}_{x_2 x_2}} (\mathbf{x_2} - \boldsymbol{\mu_2})$$

and m x m covariance matrix

$$\mathbf{V_{x_1|x_2}} = \mathbf{V_{x_1 x_1}} - \mathbf{V_{x_1 x_2}} \mathbf{V^{-1}_{x_2 x_2}} \mathbf{V^T_{x_1 x_2}}$$

Properties of the MVN - III

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

$$\mathbf{x_1} = \boldsymbol{\mu}_{\mathbf{x_1}|\mathbf{x_2}} + \mathbf{e}$$

$$= \boldsymbol{\mu_1} + \mathbf{V}_{\mathbf{x_1}\mathbf{x_2}}\mathbf{V}^{-1}_{\mathbf{x_2}\mathbf{x_2}}(\mathbf{x_2} - \boldsymbol{\mu_2}) + \mathbf{e}$$

Where e is MVN with mean vector **0** and variance-covariance matrix $\mathbf{V}_{\mathbf{x_1}|\mathbf{x_2}}$

$$\boldsymbol{\mu_1} + \mathbf{V_{x_1 x_2} V^{-1}_{x_2 x_2} (x_2} - \boldsymbol{\mu_2}) + \mathbf{e}$$

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

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

$$\mathbf{V_{x_1|x_2}} = \mathbf{V_{x_1 x_1}} - \mathbf{V_{x_1 x_2} V^{-1}_{x_2 x_2} V^{T}_{x_1 x_2}}$$

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,

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

Let $\mathbf{x_1} = (\, z_o \,), \quad \mathbf{x_2} = \begin{pmatrix} z_s \\ z_d \end{pmatrix}$

$$\mathbf{V_{x_1,x_1}} = \sigma_z^2, \quad \mathbf{V_{x_1,x_2}} = \frac{h^2\sigma_z^2}{2}(\, 1 \quad 1 \,), \quad \mathbf{V_{x_2,x_2}} = \sigma_z^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$= \boldsymbol{\mu_1} + \mathbf{V_{x_1 x_2}} \mathbf{V_{x_2 x_2}^{-1}} (\mathbf{x_2} - \boldsymbol{\mu_2}) + \mathbf{e}$$

Regression of Offspring value on Parental values (cont.)

$$= \boldsymbol{\mu_1} + \mathbf{V_{x_1x_2}V^{-1}_{x_2x_2}(x_2 - \boldsymbol{\mu_2}) + e}$$

$$\mathbf{V_{x_1,x_1}} = \sigma_z^2, \quad \mathbf{V_{x_1,x_2}} = \frac{h^2\sigma_z^2}{2}\begin{pmatrix} 1 & 1 \end{pmatrix}, \quad \mathbf{V_{x_2,x_2}} = \sigma_z^2\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Hence,

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

Where e is normal with mean zero and variance

$$\mathbf{V_{x_1|x_2}} = \mathbf{V_{x_1x_1}} - \mathbf{V_{x_1x_2}V^{-1}_{x_2x_2}V^{T}_{x_1x_2}}$$

$$\begin{aligned} \sigma_e^2 &= \sigma_z^2 - \frac{h^2\sigma_z^2}{2}\begin{pmatrix} 1 & 1 \end{pmatrix}\sigma_z^{-2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\frac{h^2\sigma_z^2}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= \sigma_z^2\left(1 - \frac{h^4}{2}\right) \end{aligned}$$

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 $Var(e) = \sigma_z^2(1-h^4/2)$

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

$$A_o = \mu_o + (A_s - \mu_s)/2 + (A_d - \mu_d)/2 + e$$
$$= A_s/2 + A_d/2 + e$$

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

Ordinary least squares

For the general linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

The predicted values given $\boldsymbol{\beta}$ and the resulting residuals are given by

$$\widehat{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta}, \quad \mathbf{e} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$$

Ordinary least squares (OLS) finds the value $\boldsymbol{\beta}$ the minimizes the sum of squared residuals

$$\sum e_i^2 = \mathbf{e}^T\mathbf{e}$$

or

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

The solution is given by setting the derivative of this function with respect to $\boldsymbol{\beta}$ equal to zero and solving.

Hence, we need to discuss vector/matrix derivatives

The gradient, the derivative of a vector-valued function

$$\nabla_{\mathbf{x}}[f] = \frac{\partial f}{\partial \mathbf{x}} = \begin{pmatrix} \dfrac{\partial f}{\partial x_1} \\ \dfrac{\partial f}{\partial x_2} \\ \vdots \\ \dfrac{\partial f}{\partial x_n} \end{pmatrix}$$

Compute the gradient for

$$f(\mathbf{x}) = \sum_{i=1}^{n} x_i^2 = \mathbf{x}^T\mathbf{x}$$

Since $\partial f / \partial x_i = 2x_i$, the gradient vector is just $\nabla_{\mathbf{x}}[f(\mathbf{x})] = 2\mathbf{x}$.

Some common derivatives

$$\nabla_{\mathbf{x}} \left[\mathbf{a}^T \mathbf{x} \right] = \nabla_{\mathbf{x}} \left[\mathbf{x}^T \mathbf{a} \right] = \mathbf{a}$$

$$\nabla_{\mathbf{x}} [\, \mathbf{A}\mathbf{x}] = \mathbf{A}^T$$

Turning to quadratic forms, if $\mathbf{A}$ is symmetric, then

$$\nabla_{\mathbf{x}} \left[\mathbf{x}^T \mathbf{A}\mathbf{x} \right] = 2 \cdot \mathbf{A}\mathbf{x}$$

$$\nabla_{\mathbf{x}} \left[(\mathbf{x} - \mathbf{a})^T \mathbf{A}(\mathbf{x} - \mathbf{a}) \right] = 2 \cdot \mathbf{A}(\mathbf{x} - \mathbf{A})$$

$$\nabla_{\mathbf{x}} \left[(\mathbf{a} - \mathbf{x})^T \mathbf{A}(\mathbf{a} - \mathbf{x}) \right] = -2 \cdot \mathbf{A}(\mathbf{a} - \mathbf{x})$$

Taking $\mathbf{A} = \mathbf{I}$,

$$\nabla_{\mathbf{x}} \left[\mathbf{x}^T \mathbf{x} \right] = \nabla_{\mathbf{x}} \left[\mathbf{x}^T \mathbf{I}\,\mathbf{x} \right] = 2 \cdot \mathbf{I}\,\mathbf{x} = 2 \cdot \mathbf{x}$$

$$\sum_{i=1}^{n} e_i^2 = \mathbf{e}^T\mathbf{e} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{x}\boldsymbol{\beta})$$
$$= \mathbf{y}^T\mathbf{y} - \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{y} - \mathbf{y}^T\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}$$
$$= \mathbf{y}^T\mathbf{y} - 2\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{y} + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}$$

where the last step follows since the matrix product $\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{y}$ yields a scaler, and hence it equals its transpose,

$$\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{y} = \left(\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{y}\right)^T = \mathbf{y}^T\mathbf{X}\boldsymbol{\beta}$$

To find the vector $\boldsymbol{\beta}$ that minimizes $\mathbf{e}^T\mathbf{e}$, taking the derivative with respect to $\boldsymbol{\beta}$ and using Equations A5.1a/c gives

$$\frac{\partial\, \mathbf{e}^T\mathbf{e}}{\partial\, \boldsymbol{\beta}} = -2\mathbf{X}^T\mathbf{y} + 2\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}$$

Setting this equal to zero gives $\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = \mathbf{X}^T\mathbf{y}$ giving

$$\boldsymbol{\beta} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$$

Simple matrix commands in R

- R is case-sensitive!
- `t(A)` = transpose of A
- `A%*%B` = matrix product AB
- `%*%` command = matrix multiplication
- `solve(A)` = compute inverse of A
- `x <-` = assigns the variable x what is to the right of the arrow
  - e.g., `x <- 3.15`

Example: Solve the equation

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

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

First, input the matrix A. In R, c denotes a list, and we built the matrix by columns,

```
> A <- matrix(c(5,7,-1,6,-3,-1,4,5,6),nrow=3)
> A
     [,1] [,2] [,3]
[1,]    5    6    4
[2,]    7   -3    5
[3,]   -1   -1    6
>
```

Next, input the vector **c**

```
> c<-matrix(c(6,-9,12),nrow=3)
> c
     [,1]
[1,]    6
[2,]   -9
[3,]   12
```

Finally, compute $\mathbf{A}^{-1}\mathbf{c}$.

```
> solve(A)%*%c
          [,1]
[1,] -2.031008
[2,]  1.426357
[3,]  1.899225
```

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

Example 2: Compute $Var(y_1)$, $Var(y_2)$, and $Cov(y_1,y_2)$, where $\mathbf{y}_i = \mathbf{c}_i^T\mathbf{x}$.

```
> 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
> c1<-matrix(c(1,-2,5),nrow=3)
> c2<-matrix(c(0,6,-4),nrow=3)
> c1
     [,1]
[1,]    1
[2,]   -2
[3,]    5
> c2
     [,1]
[1,]    0
[2,]    6
[3,]   -4
> t(c1)%*%V%*%c1
     [,1]
[1,]  960
> t(c2)%*%V%*%c2
     [,1]
[1,] 1200
> t(c1)%*%V%*%c2
     [,1]
[1,] -910
```

$\mathbf{c}_1^T\mathbf{V}\mathbf{c}_1$

$\mathbf{c}_2^T\mathbf{V}\mathbf{c}_2$

$\mathbf{c}_1^T\mathbf{V}\mathbf{c}_2$

`det(A)` returns the determinant of *A*

```
> det(V)
[1] 3250
```

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

```
> 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
```

For example, second eigenvalue is $\lambda_2 = 21.117310$, second eigenvalue

$$\mathbf{e}_2 = \begin{pmatrix} 0.2117936 \\ -0.9477830 \\ -0.2384340 \end{pmatrix}$$

Additional references

- Lynch & Walsh Chapter 8 (intro to matrices)
- Online notes:
  - Appendix 4 (Matrix geometry)
  - Appendix 5 (Matrix derivatives)

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

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