Topic Two:
Matrix Algebra and Random Vectors

Preview

- Motivation:
  - A multivariate measurement will be represented as a vector.
  - A sample of multivariate measurements will be represented as a matrix.
- Goals:
  - Review some important topics from linear algebra.
  - Introduce the random vector, define its expectation and variance.

Preview

- Motivation:
  - A multivariate measurement will be represented as a vector.
  - A sample of multivariate measurements will be represented as a matrix.
- Goals:
  - Review some important points from linear algebra.
  - Introduce the random vector, define its expectation and variance.

For a more comprehensive review of relevant linear algebra, see Supplement 2A in book.

Vectors

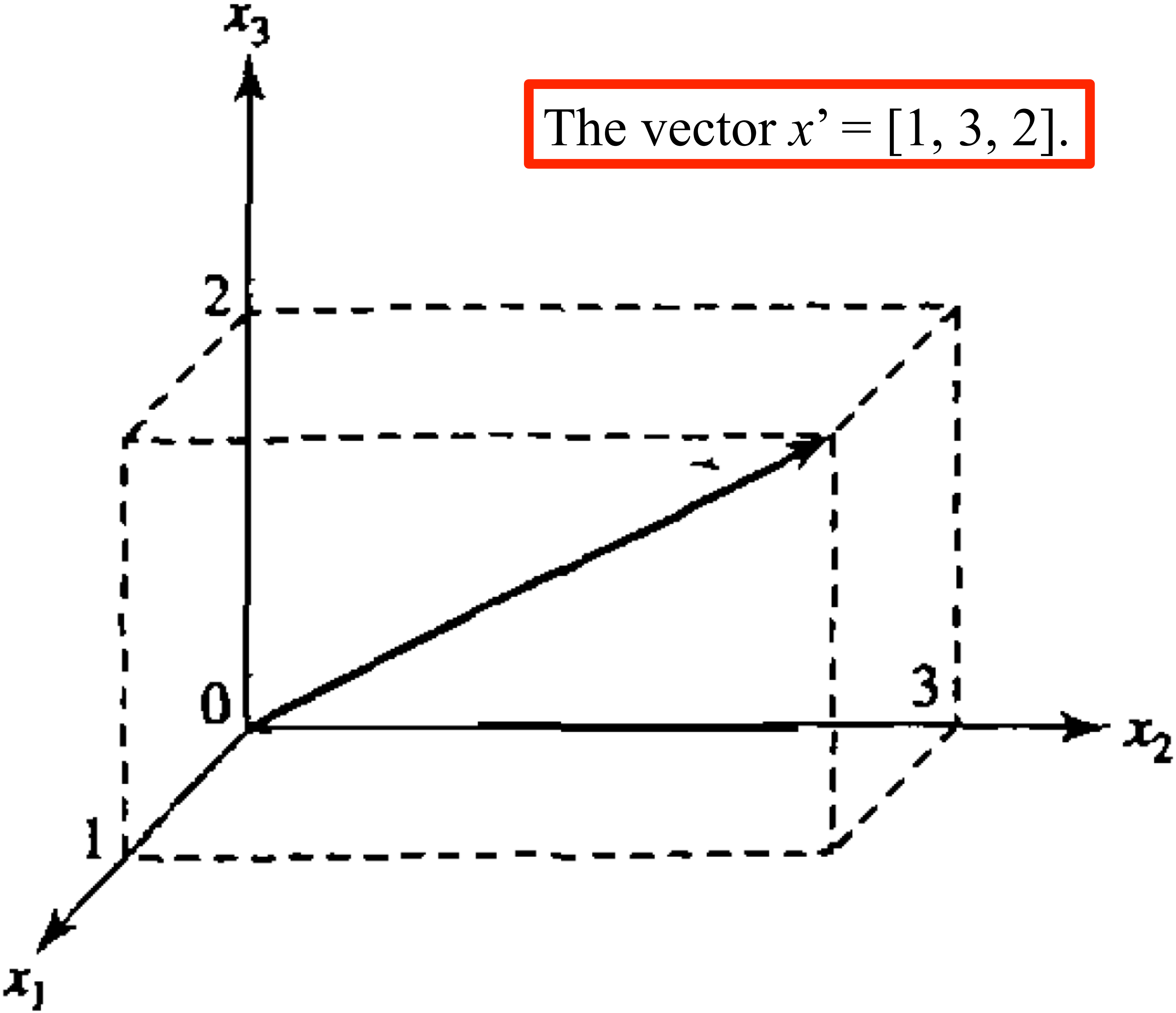
Vectors: An array $\mathbf{x}$ of n real numbers $x_1, x_2, \ldots, x_n$ is called a *vector*, and it is written as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Can also be written as $\mathbf{x}' = [x_1, x_2, \ldots, x_n]$, where the prime means transpose.

A vector $\mathbf{x}$ can be represented geometrically as a directed line in n dimensions with component x_1 along the first axis, x_2 along the second axis, ..., and x_n along the nth axis.

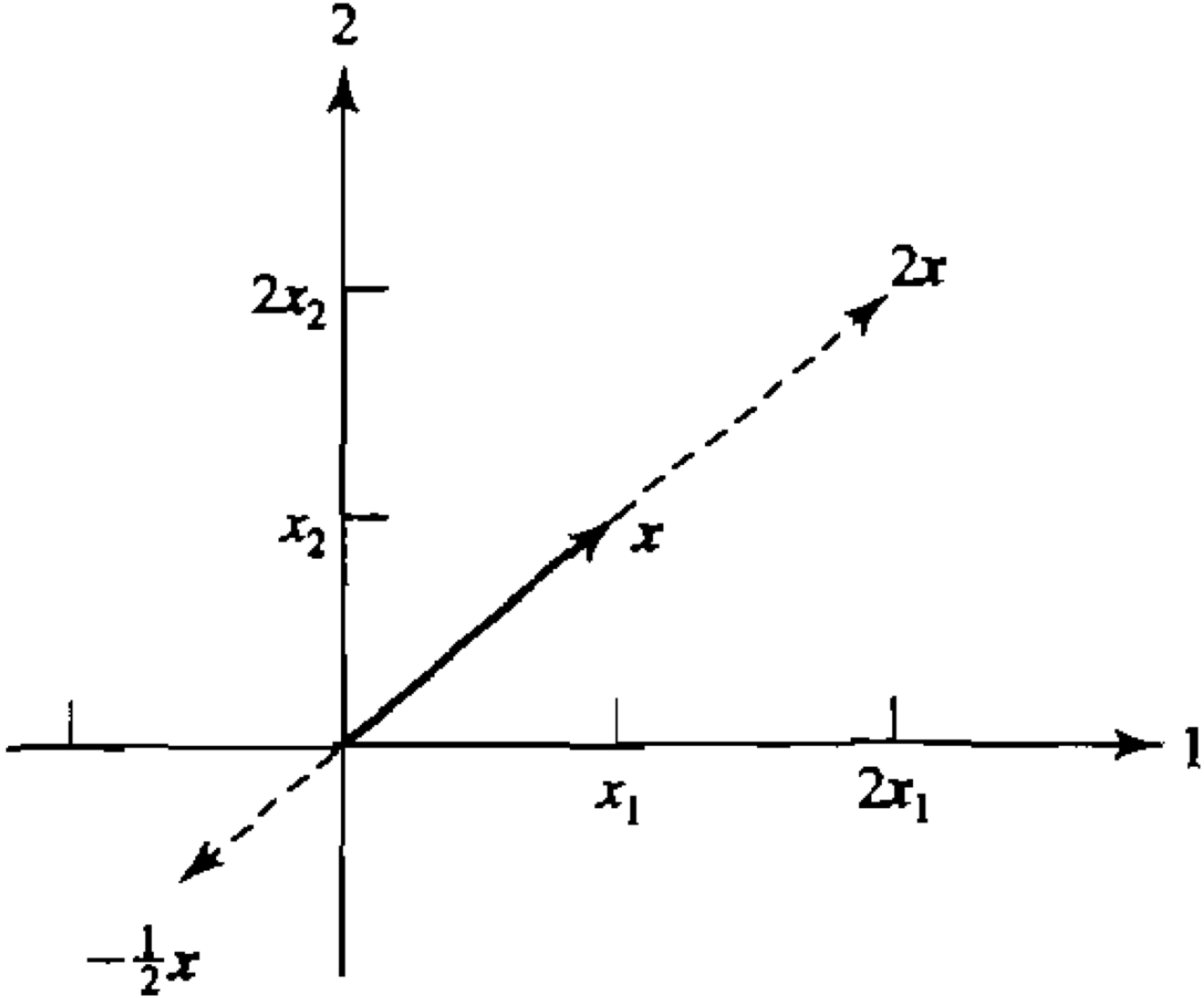
The vector $x' = [1, 3, 2]$.



Expansion / Contraction: A vector can be expanded / contracted by multiplication with a constant c:

$$c\mathbf{x} = \begin{bmatrix} cx_1 \\ cx_2 \\ \vdots \\ cx_n \end{bmatrix}$$

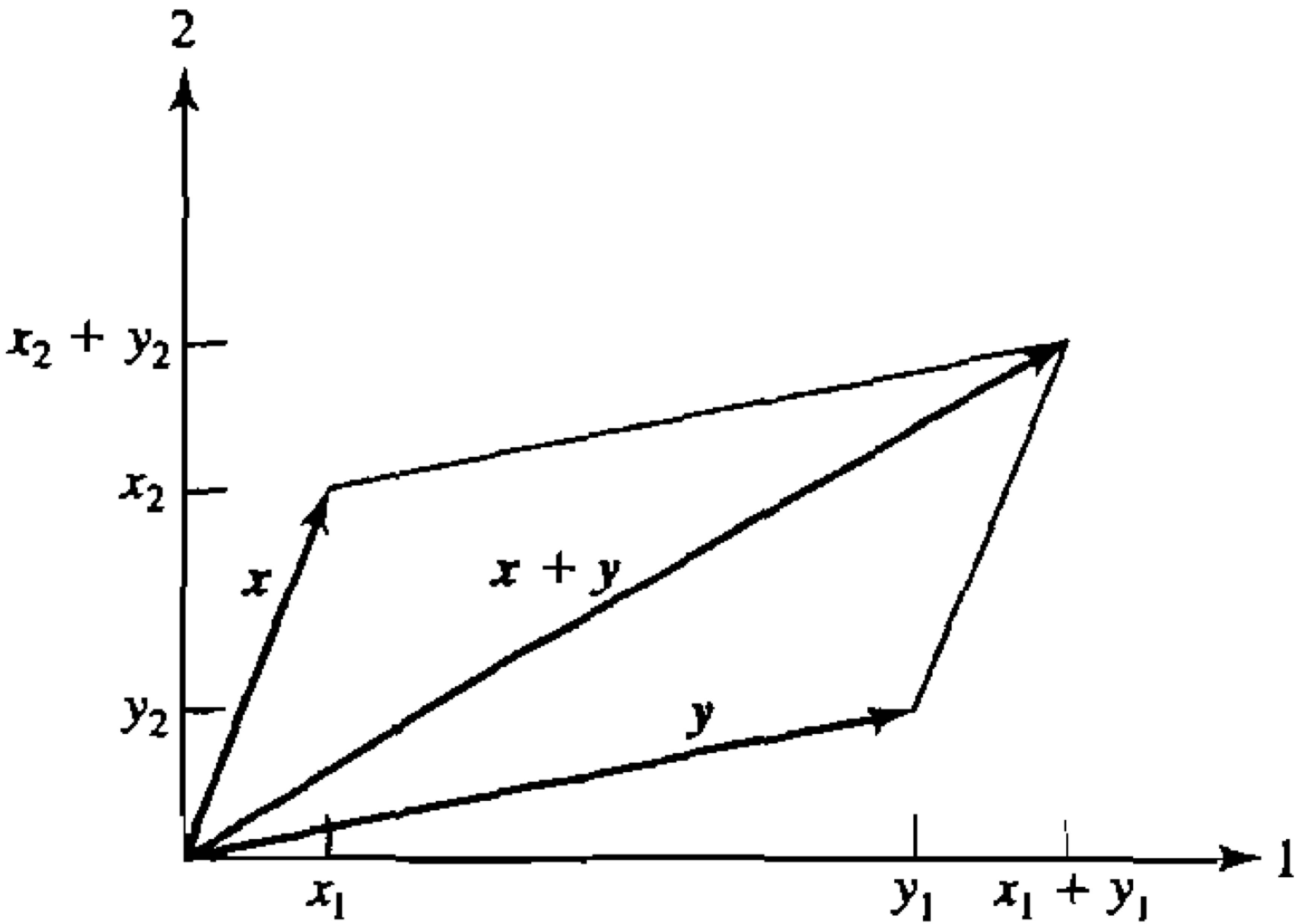
Each element is multiplied separately by c.


2
$2x$
$2x_2$
x_2
x
1
x_1
$2x_1$
$-\frac{1}{2}x$

<u>Addition</u>: The vectors $\mathbf{x}$ and $\mathbf{y}$ can be added:

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}$$

The sum of two vectors originating at the origin is the diagonal of the parallelogram formed using the two original vectors as adjacent sides.


2
$x_2 + y_2$
x_2
y_2
x
$x + y$
y
1
x_1
y_1
$x_1 + y_1$

Length and Inner Product: The *length* of $\mathbf{x}$ is

$$L_{\mathbf{x}} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

Multiplying by a scalar c changes the length:

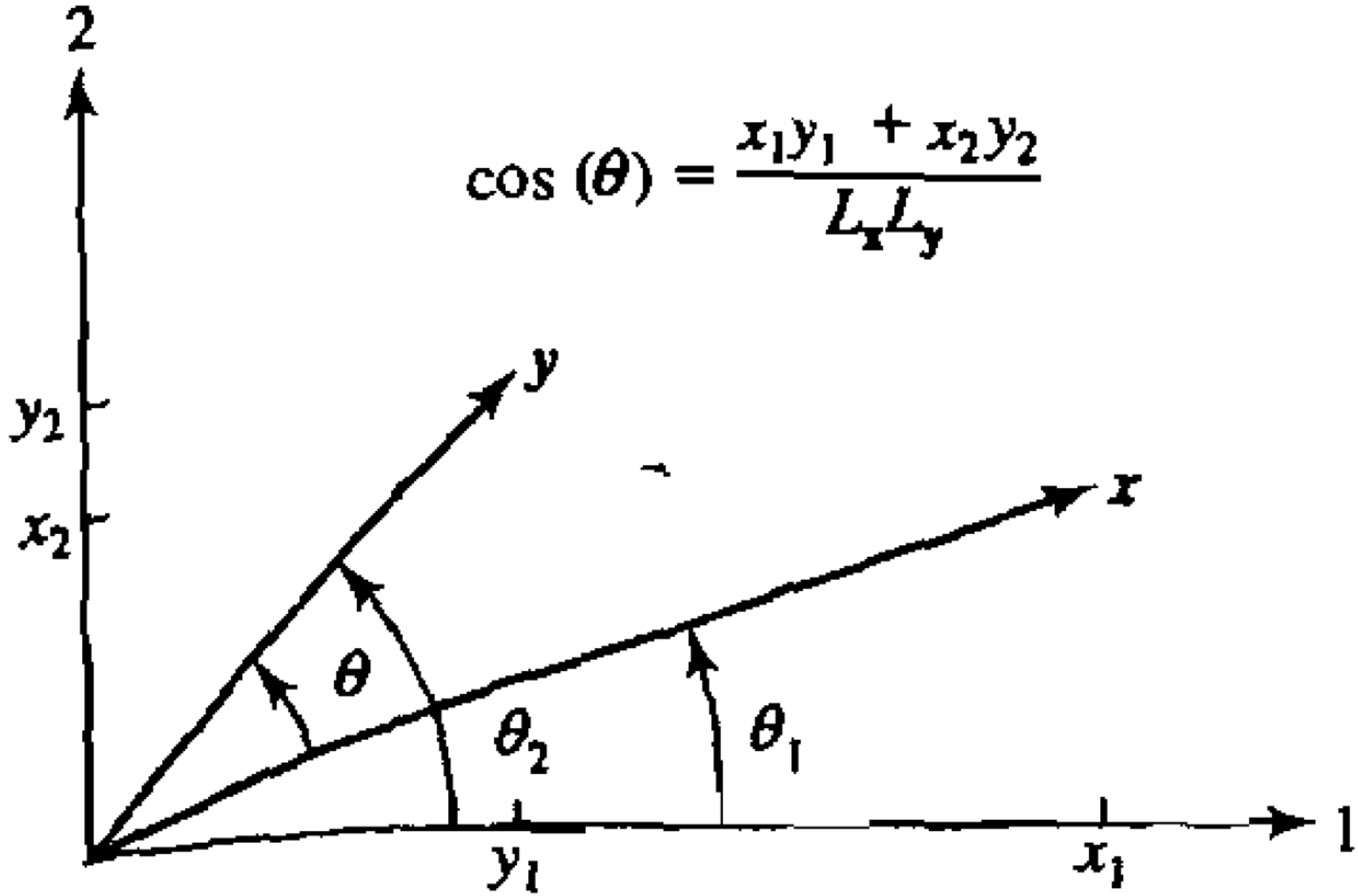
$$L_{c\mathbf{x}} = |c| L_{\mathbf{x}}$$

Note that if $c = L_{\mathbf{x}}^{-1}$, we obtain the *unit vector* $L_{\mathbf{x}}^{-1}\mathbf{x}$. The unit vector has length one and lies in the direction of $\mathbf{x}$.

The *inner product* of $\mathbf{x}$ with $\mathbf{y}$ is

$$\mathbf{x}'\mathbf{y} = x_1 y_1 + x_2 y_2 + \cdots x_n y_n$$

With this definition, $L_{\mathbf{x}} = \sqrt{\mathbf{x}'\mathbf{x}}$.



$\cos(\theta) = \frac{x_1 y_1 + x_2 y_2}{L_x L_y}$
2
1
y
x
y_2
x_2
y_1
x_1
θ
θ_2
θ_1

Angles: Recall that $\cos(\theta_1) = x_1/L_{\mathbf{x}}$, $\cos(\theta_2) = y_1/L_{\mathbf{y}}$, $\sin(\theta_1) = x_2/L_{\mathbf{x}}$, and $\sin(\theta_2) = y_2/L_{\mathbf{y}}$. Also

$$\cos(\theta) = \cos(\theta_2 - \theta_1) = \cos(\theta_2)\cos(\theta_1) + \sin(\theta_2)\sin(\theta_1)$$

So

$$\cos(\theta) = \left(\frac{y_1}{L_{\mathbf{y}}}\right)\left(\frac{x_1}{L_{\mathbf{x}}}\right) + \left(\frac{y_2}{L_{\mathbf{y}}}\right)\left(\frac{x_2}{L_{\mathbf{x}}}\right) = \frac{x_1y_1 + x_2y_2}{L_{\mathbf{x}}L_{\mathbf{y}}}$$

In general, with two n-dimensional vectors,

$$\cos(\theta) = \frac{\mathbf{x}'\mathbf{y}}{L_{\mathbf{x}}L_{\mathbf{y}}} = \frac{\mathbf{x}'\mathbf{y}}{\sqrt{\mathbf{x}'\mathbf{x}}\sqrt{\mathbf{y}'\mathbf{y}}}$$

One implication is that if $\mathbf{x}'\mathbf{y} = 0$, then $\mathbf{x}$ and $\mathbf{y}$ are perpendicular.

Example: Let $\mathbf{x}' = [2, -1, 1]$ and $\mathbf{y}' = [3, 0, 1]$.

Addition:

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} + \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ -1 \\ 2 \end{bmatrix}$$

Scalar multiplication:

$$2\mathbf{x} = 2\begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ -2 \\ 2 \end{bmatrix}$$

Dot products: $\mathbf{x}'\mathbf{x} = 2^2 + (-1)^2 + 1^2 = 6$, $\mathbf{y}'\mathbf{y} = 3^2 + 0^2 + 1^2 = 10$, and $\mathbf{x}'\mathbf{y} = (2)(3) + (-1)(0) + (1)(1) = 7$. So, $L_{\mathbf{x}} = \sqrt{6} = 2.45$, $L_{\mathbf{y}} = 3.16$. Also, $L_{2\mathbf{x}} = 4.90 = 2L_{\mathbf{x}}$.

Angle:

$$\cos(\theta) = \frac{\mathbf{x}'\mathbf{y}}{L_x L_y} = \frac{7}{2.45 \times 3.16} \approx 0.90$$

and $\theta = 25.3^o$.

Example: Let $\mathbf{x}' = [2, -1, 1]$ and $\mathbf{y}' = [3, 0, 1]$.

Addition:

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} + \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ -1 \\ 2 \end{bmatrix}$$

Scalar multiplication:

$$2\mathbf{x} = 2 \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ -2 \\ 2 \end{bmatrix}$$

Dot products: $\mathbf{x}'\mathbf{x} = 2^2 + (-1)^2 + 1^2 = 6$, $\mathbf{y}'\mathbf{y} = 3^2 + 0^2 + 1^2 = 10$, and $\mathbf{x}'\mathbf{y} = (2)(3) + (-1)(0) + (1)(1) = 7$. So, $L_{\mathbf{x}} = \sqrt{6} = 2.45$, $L_{\mathbf{y}} = 3.16$. Also, $L_{2\mathbf{x}} = 4.90 = 2L_{\mathbf{x}}$.

Angle:

$$\cos(\theta) = \frac{\mathbf{x}'\mathbf{y}}{L_x L_y} = \frac{7}{2.45 \times 3.16} \approx 0.90$$

and $\theta = 25.3^o$.

Linear Independence: A set of vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ (all of the same dimension) are *linearly dependent* if there exist constants $c_1, c_2, \ldots, c_n$, not all zero, such that

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots c_n\mathbf{x}_n = \mathbf{0}$$

If the vectors are linearly dependent, then at least one of them can be written as a linear combination of the others. If the vectors are not linearly dependent, they are said to be *linearly independent*.

Example: Let

$$\mathbf{x}_1 = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} \qquad \mathbf{x}_2 = \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} \qquad \mathbf{x}_3 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$$

Setting

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + c_3\mathbf{x}_3 = \mathbf{0}$$

implies that

$$\begin{array}{rcrcrcl} 2c_1 & - & c_2 & + & c_3 & = & 0 \\ 3c_1 & & & + & c_3 & = & 0 \\ c_1 & & & + & 2c_3 & = & 0 \end{array}$$

The only solution to this system is $c_1 = c_2 = c_3 = 0$, so $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$ are *linearly independent*. If, e.g., we replaced $\mathbf{x}_3$ with $\mathbf{x}_1 - 0.5\mathbf{x}_2$, the system would have *no* solution, making the vectors *linearly dependent*.

Example: Let

$$\mathbf{x}_1 = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} \qquad \mathbf{x}_2 = \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} \qquad \mathbf{x}_3 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$$

Setting

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + c_3\mathbf{x}_3 = \mathbf{0}$$

implies that

$$\begin{array}{rcrcrcl} 2c_1 & - & c_2 & + & c_3 & = & 0 \\ 3c_1 & & & + & c_3 & = & 0 \\ c_1 & & & + & 2c_3 & = & 0 \end{array}$$

The only solution to this system is $c_1 = c_2 = c_3 = 0$, so $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ are *linearly independent*. If, e.g., we replaced $\mathbf{x}_3$ w[illegible]2, the system would have *no* solution, making the vect[illegible]e-*pendent*.

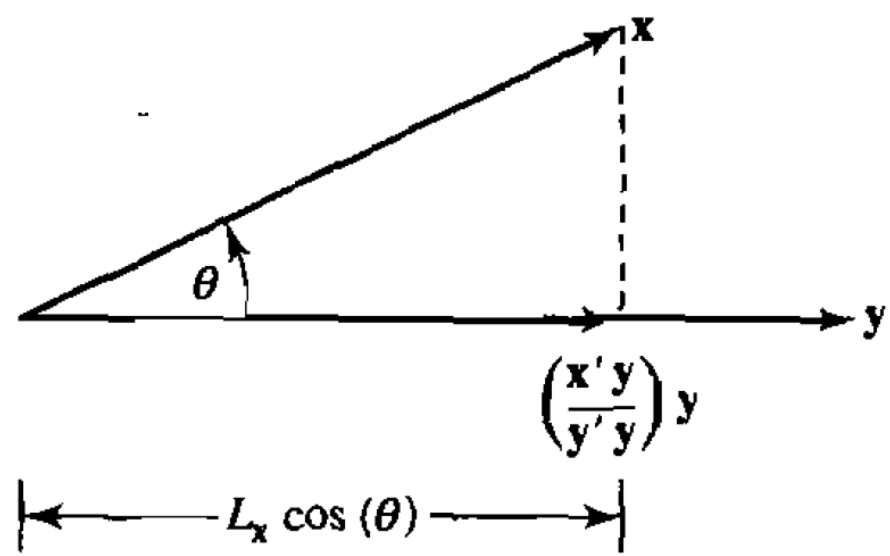
Projection: The *projection* of a vector $\mathbf{x}$ on a vector $\mathbf{y}$ is

$$\frac{\mathbf{x}'\mathbf{y}}{\mathbf{y}'\mathbf{y}}\mathbf{y} = \frac{\mathbf{x}'\mathbf{y}}{L_{\mathbf{y}}}\frac{1}{L_{\mathbf{y}}}\mathbf{y}$$

The length of the projection is

$$\frac{|\mathbf{x}'\mathbf{y}|}{L_{\mathbf{y}}} = L_{\mathbf{x}}\left|\frac{\mathbf{x}'\mathbf{y}}{L_{\mathbf{x}}L_{\mathbf{y}}}\right| = L_{\mathbf{x}}|\cos(\theta)|$$

where θ is the angle between $\mathbf{x}$ and $\mathbf{y}$.



Matrices

Matrices: A *matrix* is an array of real numbers with n rows and p columns

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

The most basic operations:

- The *transpose* $\mathbf{A}'$ transforms the matrix so that its columns become its rows.
- If $\mathbf{A}$ and $\mathbf{B}$ are matrices of the same dimension, the sum $\mathbf{A}+\mathbf{B}$ has (i,j)th entry $a_{ij}+b_{ij}$.
- Multiplying $\mathbf{A}$ by a constant c results in (i,j)th entry equal to ca_{ij}.
- If $\mathbf{A}$ is $(n \times k)$ and $\mathbf{B}$ is $(k \times p)$, then $\mathbf{AB}$ is the $(n \times p)$ product with (i,j)th entry equal to the inner product of the ith row of $\mathbf{A}$ with the jth column of $\mathbf{B}$.

<u>Example (Part 1)</u>: Let

$$\mathbf{A} = \begin{bmatrix} 2 & 3 \\ 0 & -2 \\ 4 & 1 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} -1 & 2 \\ 3 & 5 \end{bmatrix} \qquad \mathbf{c} = \begin{bmatrix} -3 \\ 5 \end{bmatrix} \qquad \mathbf{d} = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}$$

Then

$$\mathbf{A}' = \begin{bmatrix} 2 & 0 & 4 \\ 3 & -2 & 1 \end{bmatrix}$$

and

$$2\mathbf{A} = \mathbf{A} + \mathbf{A} = \begin{bmatrix} (2+2) & (3+3) \\ (0+0) & (-2-2) \\ (4+4) & (1+1) \end{bmatrix}$$

$$= \begin{bmatrix} 4 & 6 \\ 0 & -4 \\ 8 & 2 \end{bmatrix}$$

Example (Part 2): Matrix products:

$$\underset{(3\times 2)}{\mathbf{A}}\underset{(2\times 2)}{\mathbf{B}} = \begin{bmatrix} 2 & 3 \\ 0 & -2 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} -1 & 2 \\ 3 & 5 \end{bmatrix} = \begin{bmatrix} 2(-1)+3(3) & 2(2)+3(5) \\ 0(-1)-2(3) & 0(2)-2(5) \\ 4(-1)+1(3) & 4(2)+1(5) \end{bmatrix}$$

$$= \begin{bmatrix} 7 & 19 \\ -6 & -10 \\ -1 & 13 \end{bmatrix}$$

and

$$\mathbf{Ac} = \begin{bmatrix} 2 & 3 \\ 0 & -2 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} -3 \\ 5 \end{bmatrix} = \begin{bmatrix} 9 \\ -10 \\ -7 \end{bmatrix}$$

and

$$\mathbf{dc}' = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} \begin{bmatrix} -3 & 5 \end{bmatrix} = \begin{bmatrix} -3 & 5 \\ 6 & -10 \\ -9 & 15 \end{bmatrix}$$

and

$$\mathbf{d}'\mathbf{Ac} = \begin{bmatrix} 1 & -2 & 3 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 0 & -2 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} -3 \\ 5 \end{bmatrix} = [8]$$

Example (Part 2): Matrix products:

$$\underset{(3\times 2)}{\mathbf{A}}\underset{(2\times 2)}{\mathbf{B}} = \begin{bmatrix} 2 & 3 \\ 0 & -2 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} -1 & 2 \\ 3 & 5 \end{bmatrix} = \begin{bmatrix} 2(-1)+3(3) & 2(2)+3(5) \\ 0(-1)-2(3) & 0(2)-2(5) \\ 4(-1)+1(3) & 4(2)+1(5) \end{bmatrix}$$

$$= \begin{bmatrix} 7 & 19 \\ -6 & -10 \\ -1 & 13 \end{bmatrix}$$

and

$$\mathbf{Ac} = \begin{bmatrix} 2 & 3 \\ 0 & -2 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} -3 \\ 5 \end{bmatrix} = \begin{bmatrix} 9 \\ -10 \\ -7 \end{bmatrix}$$

and

$$\mathbf{dc}' = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} \begin{bmatrix} -3 & 5 \end{bmatrix} = \begin{bmatrix} -3 & 5 \\ 6 & -10 \\ -9 & 15 \end{bmatrix}$$

and

$$\mathbf{d}'\mathbf{Ac} = \begin{bmatrix} 1 & -2 & 3 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 0 & -2 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} -3 \\ 5 \end{bmatrix} = [8$$

Square Matrices: A *square* matrix $\mathbf{A}$ has the same number of columns as rows:

- A *symmetric* square matrix has $a_{ij} = a_{ji}$ for all i and j, so that $\mathbf{A} = \mathbf{A}'$.
- When two square matrices $\mathbf{A}$ and $\mathbf{B}$ are of the same dimension, both $\mathbf{A}B$ and $\mathbf{B}A$ are defined, though the two products need not be equal.
- The *identity matrix* $\mathbf{I}$ has ones on the diagonal and zeros everywhere else. We have

$$\underset{(k\times k)}{\mathbf{I}}\underset{(k\times k)}{\mathbf{A}} = \underset{(k\times k)}{\mathbf{A}}\underset{(k\times k)}{\mathbf{I}}$$

- The *inverse* of $\mathbf{A}$, if it exists, satisfies

$$\underset{(k\times k)}{\mathbf{A}^{-1}}\underset{(k\times k)}{\mathbf{A}} = \underset{(k\times k)}{\mathbf{A}}\underset{(k\times k)}{\mathbf{A}^{-1}} = \underset{(k\times k)}{\mathbf{I}}$$

  The inverse exists if the columns of $\mathbf{A}$ are linearly independent. Note that the inverse of a diagonal matrix has diagonal entries equal to $1/a_{11}$, $1/a_{22}$, ..., $1/a_{kk}$.

Example: Let

$$\mathbf{A} = \begin{bmatrix} 3 & 2 \\ 4 & 1 \end{bmatrix}$$

Ignoring for now the method of computing the inverse, we can verify that

$$\mathbf{A}^{-1} = \begin{bmatrix} -0.2 & 0.4 \\ 0.8 & -0.6 \end{bmatrix}$$

since

$$\begin{bmatrix} -0.2 & 0.4 \\ 0.8 & -0.6 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 4 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Orthogonality: A square matrix $\mathbf{A}$ is *orthogonal* if

$$\mathbf{A}\mathbf{A}' = \mathbf{A}'\mathbf{A} = \mathbf{I}$$

so that $\mathbf{A}' = \mathbf{A}^{-1}$. With $\mathbf{a}_i'$ the ith row of $\mathbf{A}$, the definition implies that $\mathbf{a}_i'\mathbf{a}_i = 1$ and $\mathbf{a}_i'\mathbf{a}_j = 0$ for $i \neq j$, so the rows are unit length and are mutually orthogonal; the same property holds for the columns.

Example: The matrix

$$\mathbf{A} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

is orthogonal since

$$\mathbf{A}\mathbf{A}' = \mathbf{A}'\mathbf{A} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I}$$

Eigenvectors and Eigenvalues: A square matrix $\mathbf{A}$ has *eigenvalue* λ, with corresponding *eigenvector* $\mathbf{x} \neq \mathbf{0}$, if

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

Typically, we use the normalized version of $\mathbf{x}$, so that $\mathbf{x}'\mathbf{x} = 1$, in which case we refer to the eigenvector as $\mathbf{e}$.

If $\mathbf{A}$ is $k \times k$ and symmetric, it has k pairs of eigenvalues and eigenvectors:

$$\lambda_1, \mathbf{e}_1 \qquad \lambda_2, \mathbf{e}_2 \quad \ldots \quad \lambda_k, \mathbf{e}_k$$

The eigenvectors can be chosen to satisfy $1 = \mathbf{e}_1'\mathbf{e}_1 = \cdots = \mathbf{e}_k'\mathbf{e}_k$ and be mutually perpendicular. The eigenvectors are unique unless two or more eigenvalues are equal.

<u>Example (Short Version)</u>: Let

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

Then, since

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = 3 \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

$\lambda_1 = 3$ is an eigenvalue, and

$$\mathbf{e}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

is its corresponding eigenvector. The other eigenvalue, eigenvector pair is $\lambda_2 = 1$ and $\mathbf{e}_2' = [-1/\sqrt{2}, 1/\sqrt{2}]$.

Example (Long Version): According to Supplement 2A in the textbook, we can derive the eigenvalues as the solutions to the *characteristic equation*

$$|\mathbf{A} - \lambda\mathbf{I}| = 0$$

where $|\cdot|$ is the *determinant*. Using the simple formula for the determinant of a 2×2 matrix, we have

$$\lambda^2 - 4\lambda + 3 = 0$$

with solutions $\lambda_1 = 3$ and $\lambda_2 = 1$. Solving

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

for eigenvectors, we have two sets of linear equations

$$\begin{aligned} -x_{11} + x_{21} &= 0 \\ x_{11} - x_{21} &= 0 \end{aligned}$$

and

$$\begin{aligned} x_{12} + x_{22} &= 0 \\ x_{12} + x_{22} &= 0 \end{aligned}$$

Both have many solutions. Arbitrarily setting $x_{11} = x_{12} = 1$, then normalizing the resulting vectors to have unit length, we have $\mathbf{e}_1' = [1/\sqrt{2}, 1/\sqrt{2}]$ and $\mathbf{e}_2' = [1/\sqrt{2}, -1/\sqrt{2}]$.

Example (Long Version): According to Supplement 2A in the textbook, we can derive the eigenvalues as the solutions to the *characteristic equation*

$$|\boldsymbol{A} - \lambda \boldsymbol{I}| = 0$$

where $|\cdot|$ is the *determinant*. Using the simple formula for the determinant of a 2×2 matrix, we have

$$\lambda^2 - 4\lambda + 3 = 0$$

with solutions $\lambda_1 = 3$ and $\lambda_2 = 1$. Solving

$$\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}$$

for eige

Note that the signs of $\boldsymbol{e}_2$ have been swapped on this slide relative to the previous one (whose results came from `R`'s **`eigen`** function). This is fine, since both parameterizations satisfy the relationship $\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}$.

and

$$x_{12} + x_{22} = 0$$

Both have many solutions. Arbitrarily setting $x_{11} = x_{12} = 1$, then normalizing the resulting vectors to have unit length, we have $\boldsymbol{e}_1' = [1/\sqrt{2}, 1/\sqrt{2}]$ and $\boldsymbol{e}_2' = [1/\sqrt{2}, -1/\sqrt{2}]$.

Example (Long Version): According to Supplement 2A in the textbook, we can derive the eigenvalues as the solutions to the *characteristic equation*

$$|\boldsymbol{A} - \lambda \boldsymbol{I}| = 0$$

where $|\cdot|$ is the *determinant*. Using the simple formula for the determinant of a 2×2 matrix, we have

$$\lambda^2 - 4\lambda + 3 = 0$$

with solutions $\lambda_1 = 3$ and $\lambda_2 = 1$. Solving

$$\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}$$

for eigenvectors, we have two sets of linear equations

$$\begin{aligned} -x_{11} + x_{21} &= 0 \\ x_{11} - x_{21} &= 0 \end{aligned}$$

and

$$\begin{aligned} x_{12} + x_{22} &= 0 \\ x_{12} + x_{22} &= 0 \end{aligned}$$

Both have many solutions. Arbitrarily setting $x_{11} = x_{12} =$ normalizing the resulting vectors to have unit length, we h $[1/\sqrt{2}, 1/\sqrt{2}]$ and $\boldsymbol{e}_2' = [1/\sqrt{2}, -1/\sqrt{2}]$.

Spectral Decomposition: The *spectral decomposition* of a $k \times k$ symmetric matrix $\mathbf{A}$ is

$$\underset{(k\times k)}{\mathbf{A}} = \lambda_1 \underset{(k\times 1)}{\mathbf{e}_1}\underset{(1\times k)}{\mathbf{e}_1'} + \cdots + \lambda_k \underset{(k\times 1)}{\mathbf{e}_k}\underset{(1\times k)}{\mathbf{e}_k'}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are the (normalized) eigenvalues of $\mathbf{A}$ and $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_k$ are the eigenvectors.

Example (Revisited): Again let

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

The eigenvalues and eigenvectors were $\lambda_1 = 3$, $\lambda_2 = 1$, $\mathbf{e}_1' = [1/\sqrt{2}, 1/\sqrt{2}]$, and $\mathbf{e}_2' = [-1/\sqrt{2}, 1/\sqrt{2}]$. The spectral decomposition of $\mathbf{A}$ is

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = 3 \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} + \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$= 3 \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} + \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

Quadratic Forms and Positive Definiteness: The product $\mathbf{x}'\mathbf{A}\mathbf{x}$, for $\mathbf{A}$ $k \times k$ symmetric, is called a *quadratic form* because it has only squared terms and product terms. For example, with $k = 2$,

$$\begin{aligned} \mathbf{x}'\mathbf{A}\mathbf{x} &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 \end{aligned}$$

If $\mathbf{A}$ satisfies

$$0 \le \mathbf{x}'\mathbf{A}\mathbf{x}$$

for all $\mathbf{x}$, then $\mathbf{A}$ is said to be *nonnegative definite*. If equality holds only for $\mathbf{x} = \mathbf{0}$, then $\mathbf{A}$ is said to be *positive definite*.

Example: Show that the following matrix is positive definite:

$$\mathbf{A} = \begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix}$$

By the spectral decomposition, we can write $\mathbf{A} = \lambda_1 \mathbf{e}_1 \mathbf{e}_1' + \lambda_2 \mathbf{e}_2 \mathbf{e}_2'$ so the quadratic form $\mathbf{x}'\mathbf{A}\mathbf{x}$ can be written as

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \lambda_1 \mathbf{x}'\mathbf{e}_1\mathbf{e}_1'\mathbf{x} + \lambda_2 \mathbf{x}'\mathbf{e}_2\mathbf{e}_2'\mathbf{x} = \lambda_1 y_1^2 + \lambda_2 y_2^2$$

where $y_1 = \mathbf{x}'\mathbf{e}_1 = \mathbf{e}_1'\mathbf{x}$ and $y_2 = \mathbf{x}'\mathbf{e}_2 = \mathbf{e}_2'\mathbf{x}$. Now we can write

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1' \\ \mathbf{e}_2' \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{P}'\mathbf{x}$$

where $\mathbf{P}$ is the 2×2 matrix with jth column equal to $\mathbf{e}_j$. Since the eigenvectors have been normalized to be unit length and mutually perpendicular, $\mathbf{P}$ is orthogonal, which means its inverse is $\mathbf{P}'$. So $\mathbf{x} = \mathbf{P}\mathbf{y}$. Since $\mathbf{x} \neq \mathbf{0}$, we know $\mathbf{y} \neq \mathbf{0}$. Furthermore, the eigenvalues of $\mathbf{A}$ are positive, so we must have that $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$.

It turns out this is a general condition: If every eigenvalue is positive, **A** is positive definite.

the eigenvalues of $\mathbf{A}$ are positive

Notes on "Distance":

- Based on what we saw in Topic One, distance can be expressed as a quadratic form.
- If $\mathbf{A}$ is a positive definite matrix such that the distance from $\mathbf{x}$ to the origin is $\mathbf{x}'\mathbf{A}\mathbf{x}$, then the distance from $\mathbf{x}$ to an arbitrary fixed point $\boldsymbol{\mu}'$ is $(\mathbf{x}-\boldsymbol{\mu})'\,\mathbf{A}\,(\mathbf{x}-\boldsymbol{\mu})$.
- We will repeatedly use quadratic forms to represent statistical distance throughout the semester.
  - One very important case will be in defining the *multivariate normal* distribution which underlies much of multivariate statistics.

Ellipsoids of Constant Distance: Let $p = 2$ and define the distance from the point $\mathbf{x}' = [x_1, x_2]$ to the origin as the quadratic form $\mathbf{x}'\mathbf{A}\mathbf{x}$ for some positive definite (2×2) matrix $\mathbf{A}$. Then the points that are a constant distance c from the origin satisfy

$$\mathbf{x}'\mathbf{A}\mathbf{x} = c^2$$

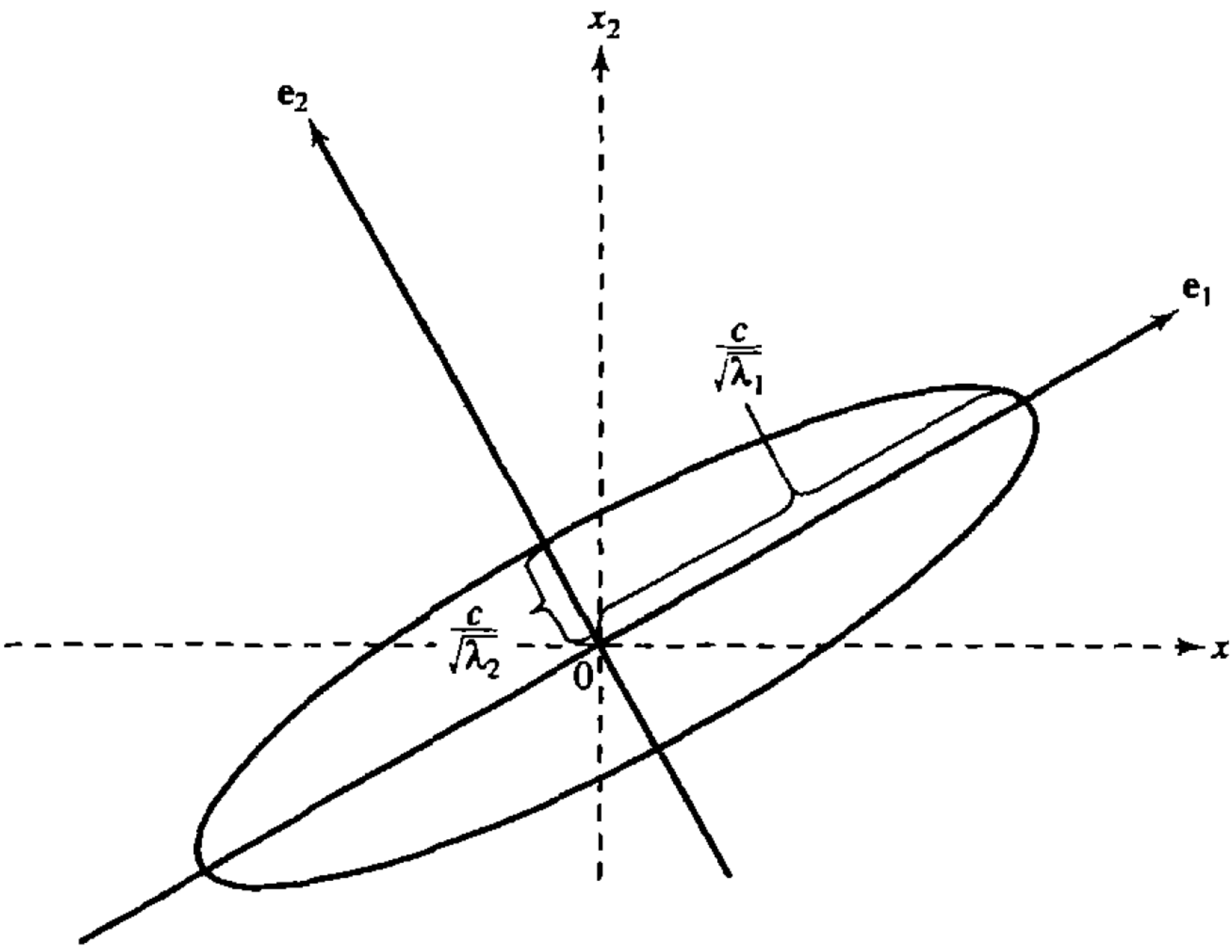
Applying the spectral decomposition, we have

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \lambda_1(\mathbf{x}'\mathbf{e}_1)^2 + \lambda_2(\mathbf{x}'\mathbf{e}_2) = \lambda_1 y_1^2 + \lambda_2 y_2^2$$

where $y_1 = \mathbf{x}'\mathbf{e}_1 = \mathbf{e}_1'\mathbf{x}$ and $y_2 = \mathbf{x}'\mathbf{e}_2 = \mathbf{e}_2'\mathbf{x}$. Since $\mathbf{A}$ is positive definite, $\lambda_1 > 0$ and $\lambda_2 > 0$, so $c^2 = \lambda_1 y_1^2 + \lambda_2 y_2^2$ is an ellipse for which the axes are $\mathbf{e}_1$ and $\mathbf{e}_2$. Choosing $\mathbf{x} = c\lambda_1^{-1/2}\mathbf{e}_1$ satisfies $\mathbf{x}'\mathbf{A}\mathbf{x} = \lambda_1\left(c\lambda_1^{-1/2}\mathbf{e}_1'\mathbf{e}_1\right)^2 = c^2$, as does $\mathbf{x} = c\lambda_2^{-1/2}\mathbf{e}_2$. The situation is illustrated by the following figure for $p = 2$. For arbitrary p, we have the analogous result: Points $\mathbf{x}' = [x_1, x_2, \ldots, x_p]$ a constant distance $c = \sqrt{\mathbf{x}'\mathbf{A}\mathbf{x}}$ from the origin lie on the hyperellipsoid $c^2 = \lambda_1(\mathbf{x}'\mathbf{e}_1)^2 + \cdots + \lambda_p(\mathbf{x}'\mathbf{e}_p)^2$. The axes of the hyperellipsoid are the $\mathbf{e}_i$, and the half-length in the direction of $\mathbf{e}_i$ is $c/\sqrt{\lambda_i}$, $i = 1, 2, \ldots, p$.

$c^2 = \lambda_1 y_1^2 + \lambda_2 y_2^2$ is an ellipse for which the axes are $\mathbf{e}_1$ and $\mathbf{e}_2$.

$\boxed{\text{You can see this by inspection of the following figure.}}$


x_2
$\mathbf{e}_2$
$\mathbf{e}_1$
$\frac{c}{\sqrt{\lambda_1}}$
$\frac{c}{\sqrt{\lambda_2}}$
0
x_1

<u>The Inverse and Square Root Matrices</u>: Let $\mathbf{A}$ be $k \times k$ positive definite. Then

$$\underset{(k\times k)}{\mathbf{A}} = \sum_{i=1}^{k} \lambda_i \underset{(k\times 1)}{\mathbf{e}_i} \underset{(1\times k)}{\mathbf{e}_i'} = \underset{(k\times k)}{\mathbf{P}} \underset{(k\times k)}{\mathbf{\Lambda}} \underset{(k\times k)}{\mathbf{P}'}$$

where $\mathbf{\Lambda} = \text{Diag}(\boldsymbol{\lambda})$ is the diagonal matrix with (i,i)th element λ_i, $i = 1, 2, \ldots, k$. Note that we can write the inverse of $\mathbf{A}$ as

$$\mathbf{A}^{-1} = \mathbf{P}\mathbf{\Lambda}^{-1}\mathbf{P}' = \sum_{i=1}^{k} \frac{1}{\lambda_i}\mathbf{e}_i\mathbf{e}_i'$$

since $(\mathbf{P}\mathbf{\Lambda}\mathbf{P}')^{-1} = (\mathbf{P}')^{-1}\mathbf{\Lambda}^{-1}(\mathbf{P})^{-1} = \mathbf{P}\mathbf{\Lambda}^{-1}\mathbf{P}'$. We can also define the *square root* of $\mathbf{A}$ as

$$\mathbf{A}^{1/2} = \sum_{i=1}^{k} \sqrt{\lambda_i}\mathbf{e}_i\mathbf{e}_i' = \mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{P}'$$

for which we have the following properties:

- $\mathbf{A}^{1/2}$ is symmetric.
- $\mathbf{A}^{1/2}\mathbf{A}^{1/2} = \mathbf{A}$.
- $\mathbf{A}^{-1/2} = \left(\mathbf{A}^{1/2}\right)^{-1} = \sum_{i=1}^{k} \frac{1}{\sqrt{\lambda_i}}\mathbf{e}_i\mathbf{e}_i' = \mathbf{P}\mathbf{\Lambda}^{-1/2}\mathbf{P}'$ where $\mathbf{\Lambda}^{-1/2} = \text{Diag}(\boldsymbol{\lambda}^{-1/2})$.
- $\mathbf{A}^{1/2}\mathbf{A}^{-1/2} = \mathbf{A}^{-1/2}\mathbf{A}^{1/2} = \mathbf{I}$, and $\mathbf{A}^{-1/2}\mathbf{A}^{-1/2} = \mathbf{A}^{-1}$.

Random Vectors

Random Variables: Consider a discrete random variable X with probability mass function $p(x)$. The mean or expected value of X is

$$\mu = E(X) = \sum_{\text{all } x} xp(x)$$

The variance of X is

$$\sigma^2 = E(X-\mu)^2 = \sum_{\text{all } x} (x-\mu)^2 p(x)$$

If instead X is a continuous random variable with probability density function $f(x)$, the mean and variance are

$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

$$\sigma^2 = E(X-\mu)^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

Covariance and Correlation: Consider two random variables X_1 and X_2 with *joint* probability mass function $p_{12}(x_1, x_2)$ (if discrete) or *joint* probability density function $f_{12}(x_1, x_2)$ (if continuous). The *covariance* between X_1 and X_2 is

$$\mathrm{Cov}(X_1, X_2) = \sigma_{12} = E(X_1 - \mu_1)(X_2 - \mu_2)$$

where $\mu_i = E(X_i)$, $i = 1, 2$. If the variables are discrete, we have

$$\sigma_{12} = \sum_{\text{all } x_1} \sum_{\text{all } x_2} (x_1 - \mu_1)(x_2 - \mu_2) p_{12}(x_1, x_2)$$

In the continuous case, we have

$$\sigma_{12} = \int_{-\infty}^{\infty} (x_1 - \mu_1)(x_2 - \mu_2) f_{12}(x_1, x_2) dx_1 dx_2$$

The *correlation* between X_1 and X_2 can then be defined in terms of the covariance as

$$\rho_{12} = \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}}$$

where $\sigma_{ii} = \mathrm{Var}(X_i) = E(X_i - \mu_i)^2$, $i = 1, 2$.

Independence: If the joint probability mass / density function can be factored into the product of the *marginal* probability mass / density functions, we say the random variables are *statistically independent*. That is, independence means

$$p_{12}(x_1, x_2) = p_1(x_1)p_2(x_2)$$

in the discrete case, or

$$f_{12}(x_1, x_2) = f_1(x_1)f_2(x_2)$$

in the continuous case. An important consequence of independence is that X_1 and X_2 have zero covariance (and are hence uncorrelated):

$$\mathrm{Cov}(X_1, X_2) = 0 \qquad\qquad \text{if } X_1 \text{ and } X_2 \text{ are independent}$$

To illustrate, suppose X_1 and X_2 are discrete and independent. Then

$$\begin{aligned}
\mathrm{Cov}(X_1, X_2) = E(X_1 - \mu_1)(X_2 - \mu_2) &= \sum_{\text{all } x_1}\sum_{\text{all } x_2}(x_1 - \mu_1)(x_2 - \mu_2)p_{12}(x_1, x_2) \\
&= \sum_{\text{all } x_1}\sum_{\text{all } x_2}(x_1 - \mu_1)(x_2 - \mu_2)p_1(x_1)p_2(x_2) \\
&= \left[\sum_{\text{all } x_1}(x_1 - \mu_1)p_1(x_1)\right]\left[\sum_{\text{all } x_2}(x_2 - \mu_2)p_2(x_2)\right] \\
&= [E(X_1 - \mu_1)]\,[E(X_2 - \mu_2)] = 0
\end{aligned}$$

An important consequence of independence is that X_1 and X_2 have zero covariance (and are hence uncorrelated):

$$\mathrm{Cov}(X_1, X_2) = 0 \qquad \text{if } X_1 \text{ and } X_2 \text{ are independent}$$

$\boxed{\text{The reverse is not necessarily true.}}$

<u>Random Vectors</u>: A *random vector* $\mathbf{X}' = [X_1, X_2, \ldots, X_p]$ is a vector whose elements are random variables. It has expected value

$$E(\mathbf{X}) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{bmatrix}$$

Also, it has (symmetrix) variance-covariance matrix

$$\boldsymbol{\Sigma} = \mathrm{Cov}(\mathbf{X}) = E(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})' = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp} \end{bmatrix}$$

And correlation matrix

$$\boldsymbol{\rho} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{12} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1p} & \rho_{2p} & \cdots & 1 \end{bmatrix}$$

where

$$\rho_{ik} = \frac{\sigma_{ik}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{kk}}}$$

__**Random Vectors**__: A *random vector* $\mathbf{X}' = [X_1, X_2, \ldots, X_p]$ is a vector whose elements are random variables. It has expected value

$$E(\mathbf{X}) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{bmatrix}$$

Also, it has (symmetrix) variance-covariance matrix

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \end{bmatrix}$$

Note the relationship

$$\boldsymbol{\Sigma} = \mathbf{V}^{1/2}\boldsymbol{\rho}\mathbf{V}^{1/2}$$

where the *standard deviation matrix* $\mathbf{V}^{1/2} = \text{Diag}(\sqrt{\sigma_{11}}, \sqrt{\sigma_{22}}, \ldots, \sqrt{\sigma_{pp}})$, which gives

$$\boldsymbol{\rho} = \left(\mathbf{V}^{1/2}\right)^{-1} \boldsymbol{\Sigma} \left(\mathbf{V}^{1/2}\right)^{-1}$$

And cor

$$\begin{bmatrix} \rho_{1p} & \rho_{2p} & \cdots & 1 \end{bmatrix}$$

where

$$\rho_{ik} = \frac{\sigma_{ik}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{kk}}}$$

Example (Part 1): Consider the discrete random vector $\mathbf{X}' = [X_1, X_2]$ with joint probability mass function $p_{12}(x_1, x_2)$, given by

x_1 \ x_2	0	1
-1	0.24	0.06
0	0.16	0.14
1	0.40	0.00

By summing over the second variable, we can obtain the marginal probability mass functions as

x_1	-1	0	1
$p_1(x_1)$	0.30	0.30	0.40

and

x_2	0	1
$p_2(x_2)$	0.8	0.2

Example (Part 2): From the marginal *pmf*s, we have

$$\mu_1 = E(X_1) = \sum_{\text{all } x_1} x_1 p_1(x_1)$$

$$= (-1)(0.30) + (0)(0.30) + (1)(0.40) = 0.10$$

and

$$\sigma_{11} = E(X_1 - \mu_1)^2 = \sum_{\text{all } x_1} (x_1 - 0.10)^2 p_1(x_1)$$

$$= (-1 - 0.10)^2(0.30) + (0 - 0.10)^2(0.30) + (1 - 0.10)^2(0.40) = 0.69$$

Also, $\mu_2 = 0.20$ and $\sigma_{22} = 0.16$. Similar calculations show that

$$\sigma_{12} = \sigma_{21} = E(X_1 - \mu_1)(X_2 - \mu_2) = -0.08$$

Example (Part 3): The mean and variance / covariance parameters from the previous slide can be represented in vector / matrix form as $E(\mathbf{X}') = [0.10, 0.20]$ and

$$\mathbf{\Sigma} = \begin{bmatrix} 0.69 & -0.08 \\ -0.08 & 0.16 \end{bmatrix}$$

The correlation matrix $\boldsymbol{\rho}$ can be obtained from $\mathbf{\Sigma}$ and the standard deviation matrix $\mathbf{V}$ as

$$\begin{aligned} \boldsymbol{\rho} &= \left(\mathbf{V}^{1/2}\right)^{-1} \mathbf{\Sigma} \left(\mathbf{V}^{1/2}\right)^{-1} \\ &= \begin{bmatrix} \frac{1}{\sqrt{0.69}} & 0 \\ 0 & \frac{1}{\sqrt{0.16}} \end{bmatrix} \begin{bmatrix} 0.69 & -0.08 \\ -0.08 & 0.16 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{0.69}} & 0 \\ 0 & \frac{1}{\sqrt{0.16}} \end{bmatrix} \\ &= \begin{bmatrix} 1.00 & -0.24 \\ -0.24 & 1.00 \end{bmatrix} \end{aligned}$$

Partitioning a Random Vector: Suppose we partition $\mathbf{X}$ into its first q variables and last $p-q$ variables:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{bmatrix} = \left[\begin{array}{c} X_1 \\ \vdots \\ X_q \\ \hdashline X_{q+1} \\ \vdots \\ X_p \end{array}\right]$$

Then $E(\mathbf{X}') = [(\boldsymbol{\mu}^{(1)})', (\boldsymbol{\mu}^{(2)})']$, where $(\boldsymbol{\mu}^{(1)})' = [\mu_1, \ldots, \mu_q]$ and $(\boldsymbol{\mu}^{(2)})' = [\mu_{q+1}, \ldots, \mu_p]$. Also,

$$\underset{(p\times p)}{\boldsymbol{\Sigma}} = E(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})' = \begin{bmatrix} \underset{(q\times q)}{\boldsymbol{\Sigma}_{11}} & \underset{(q\times(p-q))}{\boldsymbol{\Sigma}_{12}} \\ \underset{((p-q)\times q)}{\boldsymbol{\Sigma}_{21}} & \underset{((p-q)\times(p-q))}{\boldsymbol{\Sigma}_{22}} \end{bmatrix}$$

$$= \left[\begin{array}{ccc:ccc} \sigma_{11} & \cdots & \sigma_{1q} & \sigma_{1,q+1} & \cdots & \sigma_{1p} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{q1} & \cdots & \sigma_{qq} & \sigma_{q,q+1} & \cdots & \sigma_{qp} \\ \hdashline \sigma_{q+1,1} & \cdots & \sigma_{q+1,q} & \sigma_{q+1,q+1} & \cdots & \sigma_{q+1,p} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pq} & \sigma_{p,q+1} & \cdots & \sigma_{pp} \end{array}\right]$$

Partitioning a Random Vector: Suppose we partition $\mathbf{X}$ into its first q variables and last $p-q$ variables:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{bmatrix} = \begin{bmatrix} X_1 \\ \vdots \\ X_q \\ \hline X_{q+1} \\ \vdots \\ X_p \end{bmatrix}$$

Then $E(\mathbf{X}') = [(\boldsymbol{\mu}^{(1)})', (\boldsymbol{\mu}^{(2)})']$, where $(\boldsymbol{\mu}^{(1)})' = [\mu_1, \ldots, \mu_q]$ and $(\boldsymbol{\mu}^{(2)})' = [\mu_{q+1}, \ldots, \mu_p]$. Also,

$$\underset{(p\times p)}{\Sigma} = E(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})' = \begin{bmatrix} \underset{(q\times q)}{\Sigma_{11}} & \underset{(q\times(p-q))}{\Sigma_{12}} \\ \underset{((p-q)\times q)}{\Sigma_{21}} & \underset{((p-q)\times(p-q))}{\Sigma_{22}} \end{bmatrix}$$

Analogous results hold for partitioning the sample mean vector $\bar{\mathbf{x}}$ and sample covariance matrix $\mathbf{S}_n$.

$$\sigma_{p1} \quad \cdots \quad \sigma_{pq} \;\vdots\; \sigma_{p,q+1} \quad \cdots \quad \sigma_{pp}$$

Linear Combinations: Consider a linear combination of two random variables $\mathbf{X}' = [X_1, X_2]$, $\mathbf{c}'\mathbf{X} = aX_1 + bX_2$, where $\mathbf{c}' = [a, b]$. We have

$$E(\mathbf{c}'\mathbf{X}) = E(aX_1 + bX_2) = aE(X_1) + bE(X_2) = a\mu_1 + b\mu_2 = \mathbf{c}'\boldsymbol{\mu}$$

And

$$\begin{aligned}
\text{Var}(\mathbf{c}'\mathbf{X}) &= E\left[(aX_1 + bX_2) - (a\mu_1 + b\mu_2)\right]^2 \\
&= E\left[a(X_1 - \mu_1) + b(X_2 - \mu_2)\right]^2 \\
&= E\left[a^2(X_1 - \mu_1)^2 + b^2(X_2 - \mu_2)^2 + 2ab(X_1 - \mu_1)(X_2 - \mu_2)\right] \\
&= a^2\text{Var}(X_1) + b^2\text{Var}(X_2) + 2ab\text{Cov}(X_1, X_2) \\
&= a^2\sigma_{11} + b^2\sigma_{22} + 2ab\sigma_{12} \\
&= \mathbf{c}'\boldsymbol{\Sigma}\mathbf{c}
\end{aligned}$$

since

$$\mathbf{c}'\boldsymbol{\Sigma}\mathbf{c} = \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = a^2\sigma_{11} + 2ab\sigma_{12} + b^2\sigma_{22}$$

In general, for p-dimensional $\mathbf{X}$, $E(\mathbf{c}'\mathbf{X}) = \mathbf{c}'\boldsymbol{\mu}$ and $\text{Var}(\mathbf{c}'\mathbf{X}) = \mathbf{c}'\boldsymbol{\Sigma}\mathbf{c}$.

Linear Combinations: Consider a linear combination of two random variables $\mathbf{X}' = [X_1 \ \ X_2]$, ...

$E(\mathbf{c}'\mathbf{X}$ ...

And

$\mathrm{Var}(\mathbf{c}'\mathbf{X}$ ...

In the case of q linear combinations of $\mathbf{X}$, define $\mathbf{Z} = \mathbf{CX}$, where $\mathbf{C}$ is a $q \times p$ matrix. We have

$$E(\mathbf{Z}) = E(\mathbf{CX}) = \mathbf{C}\boldsymbol{\mu}$$

and

$$\mathrm{Cov}(\mathbf{Z}) = \mathrm{Cov}(\mathbf{CX}) = \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}'$$

$$= a^2\mathrm{Var}(X_1) + b^2\mathrm{Var}(X_2) + 2ab\mathrm{Cov}(X_1, X_2)$$
$$= a^2\sigma_{11} + b^2\sigma_{22} + 2ab\sigma_{12}$$
$$= \mathbf{c}'\boldsymbol{\Sigma}\mathbf{c}$$

since

$$\mathbf{c}'\boldsymbol{\Sigma}\mathbf{c} = \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = a^2\sigma_{11} + 2ab\sigma_{12} + b^2\sigma_{22}$$

In general, for p-dimensional $\mathbf{X}$, $E(\mathbf{c}'\mathbf{X}) = \mathbf{c}'\boldsymbol{\mu}$ and $\mathrm{Var}(\mathbf{c}'\mathbf{X}) = \mathbf{c}'\boldsymbol{\Sigma}\mathbf{c}$.

<u>Example</u>: Consider the sum and difference of two random variables. Let $\mathbf{X}' = [X_1, X_2]$ and

$$\mathbf{Z} = \begin{bmatrix} X_1 - X_2 \\ X_1 + X_2 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \mathbf{C}\mathbf{X}$$

Then

$$E(\mathbf{Z}) = \mathbf{C}\boldsymbol{\mu} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} \mu_1 - \mu_2 \\ \mu_1 + \mu_2 \end{bmatrix}$$

and

$$\begin{aligned} \mathrm{Cov}(\mathbf{Z}) = \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}' &= \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \sigma_{11} - 2\sigma_{12} + \sigma_{22} & \sigma_{11} - \sigma_{22} \\ \sigma_{11} - \sigma_{22} & \sigma_{11} + 2\sigma_{12} + \sigma_{22} \end{bmatrix} \end{aligned}$$

Note that the sum and difference are uncorrelated if $\sigma_{11} = \sigma_{22}$.

Important Inequality and Maximization Results

Cauchy-Schwarz Inequality: Let $\mathbf{b}$ and $\mathbf{d}$ be any two $p \times 1$ vectors. Then

$$(\mathbf{b}'\mathbf{d})^2 \leq (\mathbf{b}'\mathbf{b})(\mathbf{d}'\mathbf{d})$$

with equality if and only if $\mathbf{b} = c\mathbf{d}$ or $\mathbf{d} = c\mathbf{b}$ for some constant c.

Extended Cauchy-Schwarz Inequality: Let $\mathbf{b}$ and $\mathbf{d}$ be any two $p \times 1$ vectors, and let $\mathbf{B}$ be a $p \times p$ positive definite matrix. Then

$$(\mathbf{b}'\mathbf{d})^2 \leq (\mathbf{b}'\mathbf{B}\mathbf{b})\left(\mathbf{d}'\mathbf{B}^{-1}\mathbf{d}\right)$$

with equality if and only if $\mathbf{b} = c\mathbf{B}^{-1}\mathbf{d}$ or $\mathbf{d} = c\mathbf{B}\mathbf{b}$ for some constant c.

See textbook for proofs.

Maximization Lemma: Let $\underset{(p\times p)}{\mathbf{B}}$ be positive definite and $\underset{(p\times 1)}{\mathbf{d}}$ be a given vector. Then, for an arbitrary nonzero vector $\underset{(p\times 1)}{\mathbf{x}}$,

$$\max_{\mathbf{x}\neq\mathbf{0}} \frac{(\mathbf{x}'\mathbf{d})^2}{\mathbf{x}'\mathbf{B}\mathbf{x}} = \mathbf{d}'\mathbf{B}^{-1}\mathbf{d}$$

with the maximum attained when $\mathbf{x} = c\mathbf{B}^{-1}\mathbf{d}$ for any constant $c \neq 0$.

Maximization of Quadratic Forms for Points on the Unit Sphere: Let $\underset{(p\times p)}{\mathbf{B}}$ be a positive definite matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ and associated normalized eigenvectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_p$. Then

$$\max_{\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}'\mathbf{B}\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \lambda_1 \qquad \text{and} \qquad \min_{\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}'\mathbf{B}\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \lambda_p$$

The maximum and minimum are attained when $\mathbf{x} = \mathbf{e}_1$ and $\mathbf{x} = \mathbf{e}_p$, respectively. Also,

$$\max_{\mathbf{x}\perp\mathbf{e}_1,\cdots,\mathbf{e}_k} \frac{\mathbf{x}'\mathbf{B}\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \lambda_{k+1}$$

See textbook for proofs.

and this is attained when $\mathbf{x} = \mathbf{e}_{k+1}$, $k = 1, 2, \ldots, p-1$.