Session 5: Vectors and Matrices, Matrix Algebra, Linear Combinations, Coordinate Systems, Geometric Interpretation

Lecture Notes by: J. René van Dorp

www.seas.gwu.edu/~dorpjr

1 Department of Engineering Management and Systems Engineering, School of Engineering and Applied Science, The George Washington University, 1776 G Street, N.W., Suite 110, Washington D.C. 20052. E-mail: dorpjr@gwu.edu.
Typical convention is that vectors are written as columns. The $i$-th element of a vector is indicated by $x_i$. Hence, an $n$-dimensional vector is:

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

Convention: Underline to indicate a vector or write them in a bold font.

An $m \times n$-matrix $\mathbf{A}$ may be viewed as $n$ columns each of dimension $m$. Its elements are indicated by $a_{ij}$ where the index $i$ refers to the row number and the index $j$ refers to the column number.

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & \cdots & \cdots & a_{mn} \end{pmatrix}$$
Matrix Algebra

- **Convention:** Use capital letters for matrices and often they are written in a bold font.

- If \( m = n \), then the matrix is called a **square matrix**.

- If \( a_{ij} = a_{ji} \) for all elements of a **square matrix**, then the matrix is called **symmetric**.

  \[
  A = \begin{pmatrix}
  1 & 5 & 6 & 7 \\
  5 & 2 & 8 & 9 \\
  6 & 8 & 3 & 10 \\
  7 & 9 & 10 & 4 \\
  \end{pmatrix}
  \]

- If \( a_{ij} = 0 \) for all off-diagonal elements of a **square matrix**, then the matrix is called **a diagonal matrix**.

- If \( a_{ii} = 1 \) for all on-diagonal elements of a **diagonal matrix**, then the matrix is called **the identity matrix** and is usually denoted by \( I \).

- An \( n \)-dimensional vector is an \( n \times 1 \) matrix.
• **Vector-Scalar multiplication:**

\[
\lambda \mathbf{x} = \lambda \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \vdots \\ \lambda x_n \end{pmatrix}, \text{ e.g. } 2 \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \\ 6 \\ 8 \end{pmatrix}
\]

• **Transposed column \( n \)-vector** becomes a row vector:

\[
\mathbf{x}^T = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}^T = (x_1 \ x_2 \ \cdots \ x_n)
\]

• **Conventions:** Write \( \mathbf{x}^T \), \( \mathbf{x}^t \) or \( \mathbf{x}' \) to indicate a transposed vector. A transposed column vector becomes a row vector and vice versa.

• An \( m \times n \)-matrix may also be viewed also as \( m \) row vectors each of dimension \( n \).
Matrix Algebra

- Matrix-Vector multiplication:

\[
A \mathbf{x} = \begin{pmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
a_{m1} & a_{m,n-1} & a_{m-1,n} & a_{mn}
\end{pmatrix}
\begin{pmatrix}
x_1 \\
x_2 \\
\vdots \\
x_n
\end{pmatrix}
= \begin{pmatrix}
\sum_{j=1}^{n} a_{1j}x_j \\
\sum_{j=1}^{n} a_{2j}x_j \\
\vdots \\
\sum_{j=1}^{n} a_{mj}x_j
\end{pmatrix}
\]

\((m \times n\text{-matrix}) \cdot (n\text{-vector}) = (m\text{-vector})\)

\((m \times n\text{-matrix}) \cdot (n \times 1\text{-matrix}) = (m \times 1\text{-matrix})\)

Example:

\[
\begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \end{pmatrix}
\begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 1 \cdot 2 + 2 \cdot 3 + 3 \cdot 4 \\ 2 \cdot 2 + 4 \cdot 3 + 5 \cdot 4 \end{pmatrix} = \begin{pmatrix} 20 \\ 36 \end{pmatrix}
\]
**Matrix Algebra**

- **Vector-Matrix multiplication:**

\[
\begin{align*}
\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2n} \\
\vdots & & \ddots & \vdots \\
a_{m1} & a_{m2} & \cdots & a_{mn} 
\end{pmatrix}
\begin{pmatrix}
X_1 \\
X_2 \\
\vdots \\
X_m
\end{pmatrix}
= \\
\begin{pmatrix}
\sum_{i=1}^{m} x_i a_{i1} \\
\sum_{i=1}^{m} x_i a_{i2} \\
\vdots \\
\sum_{i=1}^{m} x_i a_{in}
\end{pmatrix}
\end{align*}
\]

\[
(m\text{-vector}) \cdot (m \times n\text{-matrix}) = (n\text{-vector})
\]

Example:

\[
\begin{pmatrix} 4 & 5 \end{pmatrix}
\begin{pmatrix} 1 & 2 & 3 \\
2 & 4 & 5 \end{pmatrix}
= \\
\begin{pmatrix}
4 \cdot 1 + 5 \cdot 2 & 4 \cdot 2 + 5 \cdot 4 & 4 \cdot 3 + 5 \cdot 5
\end{pmatrix}
= \begin{pmatrix} 14 & 28 & 37 \end{pmatrix}
\]
Matrix Algebra

Matrix-Matrix multiplication:

\[ AB = \begin{pmatrix}
a_{11} & a_{12} & \ldots & a_{1n} \\
a_{21} & a_{22} & \ldots & a_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
a_{m1} & a_{m,n-1} & \ldots & a_{mn}
\end{pmatrix}
\begin{pmatrix}
b_{11} & b_{12} & \ldots & b_{1p} \\
b_{21} & b_{22} & \ldots & b_{2p} \\
\vdots & \vdots & \ddots & \vdots \\
b_{n1} & b_{n,p-1} & \ldots & b_{np}
\end{pmatrix}\]

\[ = \begin{pmatrix}
\sum_{j=1}^{n} a_{1j}b_{j1} & \sum_{j=1}^{n} a_{1j}b_{j2} & \ldots & \sum_{j=1}^{n} a_{1j}b_{jp} \\
\sum_{j=1}^{n} a_{2j}b_{j1} & \sum_{j=1}^{n} a_{2j}b_{j2} & \ldots & \sum_{j=1}^{n} a_{2j}b_{jp} \\
\vdots & \vdots & \ddots & \vdots \\
\sum_{j=1}^{n} a_{m-1,j}b_{j1} & \sum_{j=1}^{n} a_{m-1,j}b_{j2} & \ldots & \sum_{j=1}^{n} a_{m-1,j}b_{jp}
\end{pmatrix}
\]

\((m \times n\text{-matrix}) \cdot (n \times p\text{-matrix}) = (m \times p\text{-matrix})\)
Example:

\[
\begin{pmatrix}
1 & 2 & 3 \\
2 & 4 & 5 \\
\end{pmatrix}
\begin{pmatrix}
1 & 2 \\
3 & 4 \\
5 & 6 \\
\end{pmatrix}
= \\
\begin{pmatrix}
1 \cdot 1 + 2 \cdot 3 + 3 \cdot 5 & 1 \cdot 2 + 2 \cdot 4 + 3 \cdot 6 \\
2 \cdot 1 + 4 \cdot 3 + 5 \cdot 5 & 2 \cdot 2 + 4 \cdot 4 + 5 \cdot 6 \\
\end{pmatrix}
= \\
\begin{pmatrix}
22 & 28 \\
39 & 50 \\
\end{pmatrix}
\]

- **Matrix-Matrix multiplication is Non-Commutative.**
  First of all we have to consider square matrices in this case (why?). But even when we consider square matrices the following is not true in general!

\[
AB \neq BA
\]

Example:

\[
A = \begin{pmatrix}
1 & 2 \\
3 & 4 \\
\end{pmatrix},
B = \begin{pmatrix}
1 & 3 \\
2 & 4 \\
\end{pmatrix}
\]
$AB = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} 1 \cdot 1 + 2 \cdot 2 & 1 \cdot 3 + 2 \cdot 4 \\ 3 \cdot 1 + 4 \cdot 2 & 3 \cdot 3 + 4 \cdot 4 \end{pmatrix} = \begin{pmatrix} 5 & 11 \\ 11 & 25 \end{pmatrix}$

$BA = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 1 \cdot 1 + 3 \cdot 3 & 1 \cdot 2 + 3 \cdot 4 \\ 2 \cdot 1 + 4 \cdot 3 & 2 \cdot 2 + 4 \cdot 4 \end{pmatrix} = \begin{pmatrix} 10 & 14 \\ 14 & 20 \end{pmatrix}$

- Transpose of a matrix:

$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m,n-1} & a_{m-1,n} & a_{mn} \end{pmatrix}$

$A^T = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{12} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m,n-1} & a_{m-1,n} & a_{mn} \end{pmatrix}$
Example:

\[ A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}, \quad A^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix} \]

\((m \times n\text{-matrix})^T = (n \times m\text{-matrix})\)

- **Transpose of a matrix product:**

\[ (AB)^T = B^T A^T \]

\[ [(m \times n\text{-matrix}) \cdot (n \times p\text{-matrix})]^T = (p \times n\text{-matrix}) \cdot (m \times n\text{-matrix})^T = (p \times m\text{-matrix})^T = (p \times m\text{-matrix}) \]
Example:

\[
\begin{pmatrix}
  2 & 4 & 6 \\
  8 & 10 & 12 
\end{pmatrix}
\begin{pmatrix}
  1 & 2 \\
  3 & 4 \\
  5 & 6 
\end{pmatrix}
= 
\begin{pmatrix}
  44 & 56 \\
  98 & 128 
\end{pmatrix}
\]

\[
\begin{pmatrix}
  2 & 4 & 6 \\
  8 & 10 & 12 
\end{pmatrix}^T
\begin{pmatrix}
  1 & 2 \\
  3 & 4 \\
  5 & 6 
\end{pmatrix}^T
= 
\begin{pmatrix}
  1 & 2 \\
  3 & 4 \\
  5 & 6 
\end{pmatrix}^T
\begin{pmatrix}
  2 & 4 & 6 \\
  8 & 10 & 12 
\end{pmatrix}^T
= 
\begin{pmatrix}
  44 & 56 \\
  98 & 128 
\end{pmatrix}
\]

• The inverse of a square matrix is defined such that

\[
A^{-1}A = AA^{-1} = I
\]

where \(I\) is the identity matrix.
Example:

\[ A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad A^{-1} = \frac{1}{|A|} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}, \text{ where } |A| = ad - bc \]

\[ A = \begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix}, \quad A^{-1} = \frac{1}{5} \begin{pmatrix} 4 & -1 \\ -3 & 2 \end{pmatrix} \]

\[ AA^{-1} = \begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix} \frac{1}{5} \begin{pmatrix} 4 & -1 \\ -3 & 2 \end{pmatrix} = \frac{1}{5} \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \]

\[ A^{-1}A = \frac{1}{5} \begin{pmatrix} 4 & -1 \\ -3 & 2 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix} = \frac{1}{5} \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \]

• The inverse of a matrix product:

\[ (AB)^{-1} = B^{-1}A^{-1} \]
Example:

\[ A = \begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix}, \quad A^{-1} = \frac{1}{5} \begin{pmatrix} 4 & -1 \\ -3 & 2 \end{pmatrix} \]

\[ B = \begin{pmatrix} 6 & 7 \\ 3 & 5 \end{pmatrix}, \quad B^{-1} = \frac{1}{9} \begin{pmatrix} 5 & -7 \\ -3 & 6 \end{pmatrix} \]

\[ AB = \begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 6 & 7 \\ 3 & 5 \end{pmatrix} = \begin{pmatrix} 15 & 19 \\ 30 & 41 \end{pmatrix}, \]

\[ (AB)^{-1} = \frac{1}{45} \begin{pmatrix} 41 & -19 \\ -30 & 15 \end{pmatrix} \]

\[ B^{-1} A^{-1} = \frac{1}{9} \begin{pmatrix} 5 & -7 \\ -3 & 6 \end{pmatrix} \frac{1}{5} \begin{pmatrix} 4 & -1 \\ -3 & 2 \end{pmatrix} = \frac{1}{45} \begin{pmatrix} 41 & -19 \\ -30 & 15 \end{pmatrix} \]
**Definition:** Given a collection of random variables $X_1, \ldots, X_n$ and $n$ numerical constants $a_1, \ldots, a_n$, the rv

$$Y = a_1 X_1 + \ldots + a_n X_n = \sum_{i=1}^{n} a_i X_i$$

is called a **linear** combination of the $X_i$'s. Hence we can identify two vectors:

$$\mathbf{a}^T = (a_1 \ldots a_n), \quad \mathbf{X}^T = (X_1 \ldots X_n)$$

and write

$$Y = \mathbf{a}^T \mathbf{X}$$

$$(1 \times 1\text{-matrix}) = (1 \times n\text{-matrix}) \cdot (n \times 1\text{-matrix})$$

- Let $E[\mathbf{X}] = \boldsymbol{\mu}$, where $\boldsymbol{\mu}^T = (\mu_1 \ldots \mu_n)$ then

$$E[\mathbf{a}^T \mathbf{X}] = E[Y] = a_1 \mu_1 + \ldots + a_n \mu_n = \mathbf{a}^T \boldsymbol{\mu} = \mathbf{a}^T E[\mathbf{X}]$$

Recall:

$$E[aX] = a E[X]$$
STATISTICAL REVIEW

Linear Combinations

- If the $X_i$'s are mutually dependent

\[ V[Y] = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \text{COV}[X_i, X_j] \quad \text{(Note: \text{COV}[X, X] = V[X])} \]

Introducing the variance-covariance matrix of $\mathbf{X}$:

\[ \Sigma = \begin{pmatrix}
V[X_1] & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_n) \\
\text{Cov}(X_2, X_1) & V[X_2] & \cdots & \\
\vdots & \ddots & \ddots & \\
\text{Cov}(X_n, X_1) & \cdots & V[X_n] & \\
\end{pmatrix} \]

we derive:

\[ V[Y] = \mathbf{a}^T \Sigma \mathbf{a} \]

- The variance-covariance matrix is a square symmetric matrix (why?)

- The variance-covariance matrix is positive definite, i.e.

\[ \mathbf{x}^T \Sigma \mathbf{x} > 0 \quad \text{for all vectors} \quad \mathbf{x} \neq \mathbf{0} \quad \text{(Recall: V[X] > 0)} \]
For a **coordinate system** one needs three things:

1. An **origin** $(0, 0)$
2. Two lines, called **coordinate axes**, that go through the origin. In the system above each line is **perpendicular**, which makes it a **Cartesian System**.
3. One point (other than the origin) on each axis to establish scale. These points identify the standard base vectors $\mathbf{e_1}^T = (1 \ 0)$ and $\mathbf{e_2}^T = (0 \ 1)$. 
STATISTICAL REVIEW  

Coordinate Systems

• Having a coordinate system allows us to assign to each point in the system its coordinates \((a_1, a_2)\).

\[
\begin{align*}
\mathbf{a} &= \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = a_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + a_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix}
\end{align*}
\]
# Geometric Interpretation

## Heights and Weights of 20 Women

<table>
<thead>
<tr>
<th></th>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_{d1}$</th>
<th>$X_{d2}$</th>
<th>$X_{s1}$</th>
<th>$X_{s2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>57</td>
<td>93</td>
<td>-5.85</td>
<td>-30.60</td>
<td>-1.77427</td>
<td>-1.96516</td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>110</td>
<td>-4.85</td>
<td>-13.60</td>
<td>-1.47098</td>
<td>-0.87341</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>99</td>
<td>-2.85</td>
<td>-24.60</td>
<td>-0.86439</td>
<td>-1.57984</td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>111</td>
<td>-3.85</td>
<td>-12.60</td>
<td>-1.16768</td>
<td>-0.80918</td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>115</td>
<td>-1.85</td>
<td>-8.60</td>
<td>-0.56109</td>
<td>-0.55230</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>122</td>
<td>-2.85</td>
<td>-1.60</td>
<td>-0.86439</td>
<td>-0.10275</td>
<td></td>
</tr>
<tr>
<td>62</td>
<td>110</td>
<td>-0.85</td>
<td>-13.60</td>
<td>-0.25780</td>
<td>-0.87341</td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>116</td>
<td>-1.85</td>
<td>-7.60</td>
<td>-0.56109</td>
<td>-0.48808</td>
<td></td>
</tr>
<tr>
<td>62</td>
<td>122</td>
<td>-0.85</td>
<td>-1.60</td>
<td>-0.25780</td>
<td>-0.10275</td>
<td></td>
</tr>
<tr>
<td>63</td>
<td>128</td>
<td>0.15</td>
<td>4.40</td>
<td>0.04549</td>
<td>0.28257</td>
<td></td>
</tr>
<tr>
<td>62</td>
<td>134</td>
<td>-0.85</td>
<td>10.40</td>
<td>-0.25780</td>
<td>0.66790</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>117</td>
<td>1.15</td>
<td>-6.60</td>
<td>0.34879</td>
<td>-0.42386</td>
<td></td>
</tr>
<tr>
<td>63</td>
<td>123</td>
<td>0.15</td>
<td>-0.60</td>
<td>0.04549</td>
<td>-0.03853</td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>129</td>
<td>2.15</td>
<td>5.40</td>
<td>0.65208</td>
<td>0.34679</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>135</td>
<td>1.15</td>
<td>11.40</td>
<td>0.34879</td>
<td>0.73212</td>
<td></td>
</tr>
<tr>
<td>66</td>
<td>128</td>
<td>3.15</td>
<td>4.40</td>
<td>0.95538</td>
<td>0.28257</td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>135</td>
<td>4.15</td>
<td>11.40</td>
<td>1.25867</td>
<td>0.73212</td>
<td></td>
</tr>
<tr>
<td>66</td>
<td>148</td>
<td>3.15</td>
<td>24.40</td>
<td>0.95538</td>
<td>1.56699</td>
<td></td>
</tr>
<tr>
<td>68</td>
<td>142</td>
<td>5.15</td>
<td>18.40</td>
<td>1.56197</td>
<td>1.18167</td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>155</td>
<td>6.15</td>
<td>31.40</td>
<td>1.86526</td>
<td>2.01654</td>
<td></td>
</tr>
</tbody>
</table>

$X_1$ Height  
$X_2$ Weight  
$X_{d1}$ Height: mean centered  
$X_{d2}$ Weight: Mean Centered  
$X_{s1}$ Height: Standardized  
$X_{s2}$ Weight: Standardized

<table>
<thead>
<tr>
<th></th>
<th>$X_1$</th>
<th>$X_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>62.85</td>
<td>123.60</td>
</tr>
<tr>
<td>St. Dev.</td>
<td>3.30</td>
<td>15.57</td>
</tr>
</tbody>
</table>
We can now create two column vectors $\underline{x}_{s1}$ and $\underline{x}_{s2}$ that take the values of the standardized variables for height and weight.

Together, these two column vectors form a $(20 \times 2)$ matrix given by:

$$X_s = \begin{pmatrix} \underline{x}_{s1} & \underline{x}_{s2} \end{pmatrix}$$

Each row of these matrix corresponds to one object (woman) measured on each of two different characteristics (height, weight).

By displaying all points in the same coordinate system, one can clearly visualize the pattern of observations and the position of each point relative to one another. This type of representation is known as a scatter plot.
Scatter Plot of Height and Weight of 20 Women

Conclusion?
Weight and Height individually tell us less than the two combined. Sometimes, one would like to have a single **height-weight-index** describing a person's characteristics. For example,

\[ z_1 = w_1 x_{s1} + w_2 x_{s2} \]

Weights \( w_1 \) and \( w_2 \) can be represented as a vector. From the scatter plots the choice \( \mathbf{w}^T = (w_1, w_2) = (1, 1) \) seems to make sense.
STATISTICAL REVIEW

Geometric Interpretation

- Vector-Scalar multiplication

<table>
<thead>
<tr>
<th></th>
<th>y-axis</th>
<th>x-axis</th>
</tr>
</thead>
<tbody>
<tr>
<td>( w^T = (1, 1) )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length of a Vector:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( |w| = \left[ \sum_{i=1}^{n} w_i^2 \right]^{1/2} )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
\frac{w}{\|w\|} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \approx \begin{pmatrix} 0.707 \\ 0.707 \end{pmatrix}
\]

- The effect of vector-scalar multiplication is stretching or shrinking the length of a vector while maintaining its direction.
Vector-Vector multiplication

\[ a^T b = (a_1, a_2) \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \]

\[ = a_1 b_1 + a_2 b_2 \]

\[ a^T b = \| a \| \| b \| \cos(\theta) \]

The distance from the origin to the perpendicular projection of the point \( a = (a_1, a_2) \) onto the line spanned by the vector \( b \) equals \( a^T b \) provided \( \| b \| = 1 \).
Consider the following data point from our sample: a woman 5 feet tall and weighing 122 pounds.

\[
\begin{array}{|c|c|c|}
\hline
& X_1 & X_2 \\
\hline
\text{Mean} & 62.85 & 123.60 \\
\text{St. Dev.} & 3.30 & 15.57 \\
\hline
\end{array}
\]

\[
\frac{60 - 62.85}{3.30} \approx -0.86, \quad \frac{122 - 123.60}{15.57} \approx -0.10
\]

Hence, this woman is below average height \((x_{1s} = -0.86 \text{ standard deviations below the mean})\) and weight \((x_{2s} = -0.10 \text{ standard deviations below the mean})\) compared to the other women in the sample. What is the women's height-weight-index?

\[
w^T x_s = \begin{pmatrix} 0.707 \\ 0.707 \end{pmatrix} \begin{pmatrix} -0.86 \\ -0.10 \end{pmatrix} \approx -0.68
\]
• A woman of average height and average weight obtains a height-weight index of 0.
**Matrix-Vector multiplication**

We could perform this vector multiplication for every observation in the sample:

\[ z_1 = X_s w \]

\[(20 \times 1) = (20 \times 2) \cdot (2 \times 1)\]
The vector $\mathbf{z}_1$ is a linear combination of the columns of the matrix $\mathbf{X}_s$

$$\mathbf{X}_s = \begin{pmatrix} x_{s1} & x_{s2} \end{pmatrix}, \mathbf{z}_1 = \mathbf{X}_s \mathbf{w} = w_1 x_{s1} + w_2 x_{s2}$$

This is the same thing we do in multiple regression analysis when we multiply the matrix $\mathbf{X}$ (containing the values of the explanatory variables in columns) by the vector of least squares coefficients $\mathbf{\hat{b}}$ to obtain the vector of fitted values $\mathbf{\hat{y}}$.

These fitted values $\mathbf{\hat{y}}$ are next compared to observed values $\mathbf{y}$ to conclude whether an adequate model fit has been obtained.

Model fit will be described by $R^2$-values, adjusted $R^2$-values and a normal probability plot of the difference vector $(\mathbf{y} - \mathbf{\hat{y}})$.

Given an adequate model fit, hypothesis test can be formulated regarding the values of the weights $w_1$ and $w_2$. These hypothesis use the $t$-distribution and the $F$-distribution.
We could now introduce a new random variable \( Z_1 \) (describing a random height-weight index) that is a linear combination of the random variables \( X_{1s} \) (standardized height) and \( X_{2s} \) (standardized weight), where for the given data set \( Z_1 \) takes the values \( \bar{Z}_1 \).

<table>
<thead>
<tr>
<th>Xs1</th>
<th>Xs2</th>
<th>Z1</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.77427</td>
<td>-1.96516</td>
<td>-2.64418</td>
</tr>
<tr>
<td>-1.47098</td>
<td>-0.87341</td>
<td>-1.65773</td>
</tr>
<tr>
<td>-0.86439</td>
<td>-1.57984</td>
<td>-1.72833</td>
</tr>
<tr>
<td>-1.16768</td>
<td>-0.80918</td>
<td>-1.39786</td>
</tr>
<tr>
<td>-0.56109</td>
<td>-0.55230</td>
<td>-0.78729</td>
</tr>
<tr>
<td>-0.86439</td>
<td>-0.10275</td>
<td>-0.68387</td>
</tr>
<tr>
<td>-0.25780</td>
<td>-0.87341</td>
<td>-0.79988</td>
</tr>
<tr>
<td>-0.56109</td>
<td>-0.48808</td>
<td>-0.74188</td>
</tr>
<tr>
<td>0.40099</td>
<td>0.10269</td>
<td>0.29840</td>
</tr>
<tr>
<td>-0.25780</td>
<td>0.25495</td>
<td>0.07998</td>
</tr>
<tr>
<td>-0.25780</td>
<td>0.66790</td>
<td>0.28998</td>
</tr>
<tr>
<td>0.34879</td>
<td>-0.42386</td>
<td>0.05308</td>
</tr>
<tr>
<td>0.04549</td>
<td>-0.03853</td>
<td>0.00492</td>
</tr>
<tr>
<td>0.65208</td>
<td>0.34679</td>
<td>0.70631</td>
</tr>
<tr>
<td>0.34879</td>
<td>0.73212</td>
<td>0.76432</td>
</tr>
<tr>
<td>0.95538</td>
<td>-0.42386</td>
<td>0.76432</td>
</tr>
<tr>
<td>0.04549</td>
<td>-0.03853</td>
<td>0.70631</td>
</tr>
<tr>
<td>0.34879</td>
<td>0.73212</td>
<td>0.76432</td>
</tr>
<tr>
<td>1.25867</td>
<td>0.73212</td>
<td>1.40770</td>
</tr>
<tr>
<td>0.95538</td>
<td>1.56699</td>
<td>1.78358</td>
</tr>
<tr>
<td>1.56197</td>
<td>1.18167</td>
<td>1.94004</td>
</tr>
<tr>
<td>1.86526</td>
<td>2.01654</td>
<td>2.74485</td>
</tr>
</tbody>
</table>

New variable \( Z_1 \) describes a person's height-weight-index. Persons with a high value are tall and heavy, and with a low value are small and light.

To describe whether a person’s height-weight combination is above or below that of the average person with the same height-weight index, we may introduce a second variable \( Z_2 \) which we shall refer to as **body-weight-index**.
Matrix-Vector multiplication

We could perform this vector multiplication for every observation in the sample:

\[ z_1 = X_s w \quad (20 \times 1) = (20 \times 2) \cdot (2 \times 1) \]
Matrix-Matrix multiplication
A reasonable choice for the body-weight-index $Z_2$ is to search for a linear combination of $x_{s1}$ and $x_{s2}$ that is orthogonal to the height-weight-index $Z_1$. Hence, we are comparing with one another those persons which have the same height-weight index (the same value of $Z_1$). Informally, orthogonality implies that information about $Z_2$ does not provide any information about $Z_1$, which implies that $Z_1$ and $Z_2$ will be uncorrelated.

Two vectors $a$ and $b$ are orthogonal if and only if $a^T b = 0$.

Example:

$$w_1^T = (0.707 \quad 0.707)$$

To find the second vector $w_2^T = (w_{12} \quad w_{22})$ requires that we solve the equation

$$w_1^T w_2 = 0 \iff 0.707w_{12} + 0.707w_{22} = 0 \iff w_2 = \begin{pmatrix} -0.707 \\ 0.707 \end{pmatrix} \text{ or } w_2 = \begin{pmatrix} 0.707 \\ -0.707 \end{pmatrix}$$
What happens when we multiply the matrix $X_s = (x_{s1} \ x_{s2})$ with the matrix $W = (w_1 \ w_2)$?
A person with a positive body-weight-index $Z_2$ is smaller and heavier than the average person with the same height-weight-index $Z_1$.

A person with a negative body-weight-index $Z_2$ is taller and lighter than the average person with the same height-weight-index $Z_1$.

Both large positive and large negative values could be a cause for concern.
We have created two linear combinations $z_1$ and $z_2$, each of which is interpretable as the vector whose elements are projections of points onto a directed line segment.

The relative position of each data point has not changed in the new coordinate system with $z_1$ and $z_2$ as its unit base vectors.

The effect of the matrix multiplication here is a rotation while preserving size and shape.
The values $z_1$ and $z_2$ for every point form the new coordinates in the coordinate system with $\mathbf{w}_1$ and $\mathbf{w}_2$ as its base vectors.

The rotation matrix

$$W = \begin{pmatrix} 0.707 & -0.707 \\ 0.707 & 0.707 \end{pmatrix}$$

corresponds to a counterclockwise rotation of the axes by 45 degrees. In general we can accomplish a clockwise orthogonal rotation through any angle $\theta$ via the following matrix:

$$W = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}$$

An orthogonal rotation matrix has special properties:

$$WW^T = I \iff W^T = W^{-1}$$

Not all matrix operations have to involve orthogonal rotations.
• Matrix operations that are rotations are utilized in principal component analysis and factors analysis. (To be discussed in a future class).

• Recall that the original variables $X_{1s}$ and $X_{2s}$ were standardized to have variance 1. This ensures that one variable does not overshadow the other.

• The newly created variables height-weight-index $Z_1$ and body-weight-index $Z_2$ (for which we have the observations $z_1$ and $z_2$ organized in a matrix $Z = (z_1 \ z_2)$) do not have variance 1 anymore.

• Of course we can re-standardize the variables $Z_1$ and $Z_2$. This can be done by multiplying the matrix $Z = (z_1 \ z_2)$ with a diagonal matrix $D^{-1}$ with the elements $1/s_i$ on the unit diagonal, where $s_i$ is the standard deviation of the variable $Z_i$. Thus

$$D^{-1} = \begin{pmatrix} 1/s_{i1} & 0 \\ 0 & 1/s_{i2} \end{pmatrix}, \quad Z_s = Z D^{-1}$$

The standardized variables $Z_{s1}$ and $Z_{s2}$ will have standard deviation 1.
• The histogram $Z_1$ describes differences between persons in terms of their height-weight index.

• The histogram $Z_2$ describes differences between persons in terms of their body-weight index.
The next slide provides a scatter plot of the standardized data set

\[ Z_s = ZD^{-1} = \left( \frac{1}{s_1} \tilde{z}_1 \quad \frac{1}{s_2} \tilde{z}_2 \right) \]

The shape of the configuration changes from elliptical to circular.
STATISTICAL REVIEW

Singular Value Decomposition

- We started with a data matrix $X$ that contained two arbitrarily scaled (and in this case) highly correlated variables $X_{s1}$ and $X_{s2}$.

- We discovered that we could use one matrix operation to rotate the configuration and another one to change the shape by stretching and shrinking it along different axes.

- We began with an angled ellipse shaped data at a 45 degree angle and ended up with something circular in nature.

\[
Z_s = ZD^{-1} \iff Z_s = XWD^{-1} \iff Z_sD = XW \iff X = Z_sDW^T
\]

- What has been shown (although not formally), that any data matrix $X$ can be decomposed into three component parts: a matrix of uncorrelated variables ($Z_s$) that have unit variance, a stretching and shrinking transformation ($D$), and an orthogonal rotation ($W^T$).

- The process of finding these three components is called singular value decomposition (SVD) and is used in a variety of multivariate data analysis techniques.