Mailing list

Please register at the following mailing list if you participate in this course.

https://mailman.uni-mannheim.de/mailman/listinfo/dmm14

Low traffic, used primarily for

- Announcements
- Distributing assignments
Outline

1. Vectors

2. Matrices

3. Summary
Outline

1. Vectors
2. Matrices
3. Summary
A vector is

- A 1D array of numbers
- A geometric entity with magnitude and direction
- A matrix with exactly one row or column
  - Called row vector and column vector, resp.
  - Transpose $\mathbf{v}^T$ transposes a row vector into a column vector and vice versa
- A (latent) object or attribute

\[
\begin{pmatrix}
4 \\
-3
\end{pmatrix}
\]

\[
\begin{pmatrix}
\text{Stockholm} \\
\text{Minsk} \\
\text{London} \\
\text{Budapest} \\
\text{Paris} \\
\text{Bucharests} \\
\text{Barcelona} \\
\text{Rome} \\
\text{Lisbon} \\
\text{Athens} \\
\text{Valencia} \\
\text{Malta}
\end{pmatrix}
\begin{pmatrix}
9.95 \\
10.77 \\
14.85 \\
14.91 \\
15.46 \\
16.44 \\
19.90 \\
20.44 \\
21.36 \\
22.31 \\
22.36 \\
23.35
\end{pmatrix}
\]

\[
\begin{pmatrix}
\text{Jan} & \text{Apr} & \text{Jul} & \text{Oct} & \text{Year}
\end{pmatrix}
\begin{pmatrix}
-0.70 & 8.60 & 21.90 & 9.90 & 10.00
\end{pmatrix}
\]
Vector norm

The **norm** of vector defines its magnitude. Let \( \mathbf{v} = (v_1 \ v_2 \ \cdots \ v_n)^T \).

- **Euclidean norm**: \( \| \mathbf{v} \| = \sqrt{\sum_{i=1}^{n} v_i^2} \)
  - Corresponds to intuitive notion of length in Euclidean space

- **\( L_p \) norm** for \( 1 \leq p \leq \infty \): \( \| \mathbf{v} \|_p = \left( \sum_{i=1}^{n} |v_i|^p \right)^{1/p} \)
  - \( L_1 \) norm = sum of absolute values
    (Manhattan distance from origin)
  - \( L_2 \) norm = Euclidean norm
    (bird-fly distance from origin)
  - \( L_\infty \) norm = maximum absolute value
  - The \( L_p \) norms decrease as \( p \) increases, i.e.,
    \[ \| \mathbf{v} \|_{p+a} \leq \| \mathbf{v} \|_p \quad \text{for} \quad a \geq 0 \]

- **Properties of vector norms**
  - \( \| \mathbf{v} \| > 0 \) when \( \mathbf{v} \neq 0 \) and \( \| \mathbf{v} \| = 0 \) iff \( \mathbf{v} = \mathbf{0} \)
  - \( \| a \mathbf{v} \| = |a| \| \mathbf{v} \| \) (absolute scalability)
  - \( \| \mathbf{v}_1 + \mathbf{v}_2 \| \leq \| \mathbf{v}_1 \| + \| \mathbf{v}_2 \| \) (triangle inequality)
Norms and distances

The **distance** between two vectors \( \mathbf{u}, \mathbf{v} \in \mathbb{R}^n \) can be quantified with norm \( \| \mathbf{u} - \mathbf{v} \| \).

- Stockholm, \( \mathbf{s} = (-0.70, 8.60, 21.90, 9.90, 10.00) \)
- Minsk, \( \mathbf{m} = (-2.10, 12.20, 23.60, 10.20, 10.60) \)
- Athens, \( \mathbf{a} = (12.90, 20.30, 32.60, 23.10, 22.30) \)

<table>
<thead>
<tr>
<th></th>
<th>( L_1 )</th>
<th>( L_2 )</th>
<th>( L_\infty )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \mathbf{s} )</td>
<td>( \mathbf{s} )</td>
<td>( \mathbf{s} )</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>7.60</td>
<td>4.27</td>
<td>3.60</td>
</tr>
<tr>
<td></td>
<td>61.50</td>
<td>27.60</td>
<td>13.60</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>( \mathbf{m} )</th>
<th>( \mathbf{m} )</th>
<th>( \mathbf{m} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>7.60</td>
<td>4.27</td>
<td>3.60</td>
</tr>
<tr>
<td></td>
<td>56.70</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>25.98</td>
<td>15.00</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>27.60</td>
<td>25.98</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>25.98</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Dot product (algebraic definition)

The **dot product** of two vectors \( \mathbf{u}, \mathbf{v} \in \mathbb{R}^n \) is given by

\[
\mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^{n} u_i v_i.
\]

- Also known as **scalar product** or **inner product**
- We’ll often use matrix product notation and write \( \mathbf{u}^T \mathbf{v} \)
- Properties (with \( a, b \in \mathbb{R} \))
  - \( \mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u} \)
  - \( (a \mathbf{u}) \cdot \mathbf{v} = a(\mathbf{u} \cdot \mathbf{v}) \)
  - \( (a \mathbf{u} + b \mathbf{v}) \cdot \mathbf{w} = (a \mathbf{u})\mathbf{w} + (a \mathbf{v})\mathbf{w} \)
- Many uses, many interpretations
With dot products, we can ...

- Compute the (squared) Euclidean norm

\[ \mathbf{v} \cdot \mathbf{v} = \sum_{i=1}^{n} v_i^2 = \| \mathbf{v} \|^2 \]

- Normalize a vector to length 1 (then a unit vector)

\[ \hat{\mathbf{v}} = \mathbf{v} / \| \mathbf{v} \| \]

- Determine the value of a coordinate

\[ v_i = \mathbf{v} \cdot \mathbf{e}_i, \]

where \( \mathbf{e}_i \) denotes the \( i \)-th standard basis vector (i.e., \( [\mathbf{e}_i]_j = 1 \) if \( i = j \) else 0)

- Compute the sum of the elements of a vector

\[ \mathbf{v} \cdot \mathbf{1}_n = \sum_{i=1}^{n} v_i, \]

where \( \mathbf{1}_n \) is the all-ones vector of dimensionality \( n \)

- ...
Dot product: Weighted sum

The elements of one vector are interpreted as weights for the elements of the other vector.

Example: Anna goes shopping

<table>
<thead>
<tr>
<th>Item</th>
<th>Bread</th>
<th>Butter</th>
<th>Pizza</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price/piece</td>
<td>1 €</td>
<td>0.50 €</td>
<td>3 €</td>
</tr>
<tr>
<td>Quantity bought</td>
<td>1</td>
<td>2</td>
<td>5</td>
</tr>
</tbody>
</table>

• How much does Anna pay?
• Prices can be interpreted as “weights”: \( \mathbf{p} = (1 \ 0.5 \ 3)^T \)
• Quantities are \( \mathbf{n} = (1 \ 2 \ 5)^T \)
• Total is \( \mathbf{p} \cdot \mathbf{n} = 1 \cdot 1 + 0.5 \cdot 2 + 3 \cdot 5 = 17 \)
• Similarly: Can interpret quantities as weights for prices
Dot product: Expected value

One vector corresponds to probabilities, the other one to a random variable.

Example: Bob is gambling

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Jackpot</th>
<th>Win</th>
<th>Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.1</td>
<td>0.2</td>
<td>0.7</td>
</tr>
<tr>
<td>Amount won</td>
<td>5 €</td>
<td>1 €</td>
<td>-2 €</td>
</tr>
</tbody>
</table>

- How much does Bob win in expectation? (Should he play?)
- Probabilities $\mathbf{p} = (0.1 \ 0.2 \ 0.7)^T$
  - A non-negative vector that sums to one ($\|\mathbf{p}\|_1 = 1$) is called a **probability vector**
  - Corresponds to a probability distribution over a finite set of outcomes
- Amounts won $\mathbf{x} = (5 \ 1 \ -2)^T$
  - Corresponds to a random variable; associates a real value with each outcome
- Expected value $\mathbf{p} \cdot \mathbf{x} = 0.1 \cdot 5 + 0.2 \cdot 1 + 0.7 \cdot (-2) = -0.7$
Dot product: Sample variance

Denote by \( \bar{u} = \frac{1}{n} \sum_i u_i \) the mean of \( u \). If we treat the entries of \( u \) as samples from some distribution, then the sample variance is given by

\[
\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (u_i^2 - \bar{u})^2 = \frac{\|u - \bar{u}\|^2}{n} = \frac{(u - \bar{u}) \cdot (u - \bar{u})}{n},
\]

where \( \bar{u} \) denotes the sample mean vector, i.e., \([\bar{u}]_i = \bar{u}\) for \(1 \leq i \leq n\).

- Example
  
  - \( u = (10 \ 11 \ 12)^T \)
  
  - \( \bar{u} = 11, \ \bar{u} = (11 \ 11 \ 11)^T \)
  
  - \( u - \bar{u} = (-1 \ 0 \ 1)^T \)
  
  - \( \sigma^2 = 2/3, \ ||u||^2 = 365 \)

- Variances are thus closely related to norms; the key difference is centering and averaging

- When we center data before analyzing it, dot products are thus proportional to variances \((u \cdot u)\) or covariances \((u \cdot v)\)
Dot product: Sets and intersections

The **indicator vector** of a subset \( T \) of a set \( S = \{ s_1, \ldots, s_n \} \) is the vector \( x \) such that \( x_i = 1 \) if \( s_i \in T \) and \( x_i = 0 \) if \( s_i \notin T \). If \( u \) and \( v \) are indicator vectors for subsets \( U, V \subseteq S \), resp., then \( u \cdot v = |U \cap V| \).

- \( S = \{ \text{France, Germany, Denmark, Poland} \} \)
- Anna visited France, Germany, and Poland: \( u = (1 \ 1 \ 0 \ 1)^T \)
- Bob visited Germany, Denmark, and Poland: \( v = (0 \ 1 \ 1 \ 1)^T \)
- Number of countries visited by both:

\[
    u \cdot v = 1 \cdot 0 + 1 \cdot 1 + 0 \cdot 1 + 1 \cdot 1 = 2 = |\{ \text{Germany, Poland} \}|
\]
Dot product (geometric definition)

An alternative geometric definition of the dot product of two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ is

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos \theta,$$

where $0 \leq \theta \leq \pi$ denotes the angle between $\mathbf{u}$ and $\mathbf{v}$.
Why is this?

Let’s focus on the 2D case. Recall the law of cosines:
\[ c^2 = a^2 + b^2 - 2ab \cos \theta. \]

Now set \( u = B - C \) and \( v = A - C \) and observe that \( v - u = A - B \).

\[
\cos \theta = \frac{a^2 + b^2 - c^2}{2ab} = \frac{\|u\|^2 + \|v\|^2 - \|v - u\|^2}{2 \|u\| \|v\|}
\]

\[
= \frac{u \cdot u + v \cdot v - (v - u) \cdot (v - u)}{2 \|u\| \|v\|}
\]

\[
= \frac{u \cdot u + v \cdot v - v \cdot v + 2u \cdot v - u \cdot u}{2 \|u\| \|v\|}
\]

\[
= \frac{u \cdot v}{\|u\| \|v\|}
\]
Dot product: Test for orthogonality

Two nonzero vectors $u, v \in \mathbb{R}^n$ are orthogonal iff $u \cdot v = 0$.

- Since $0 = u \cdot v = ||u|| \cdot ||v|| \cdot \cos \theta$ and $||u||, ||v|| > 0$, we have $\cos \theta = 0$
- And this means that the angle is 90 degrees
Dot product: Cosine similarity (1)

The angle between \( u \) and \( v \) is another way to measure the similarity between two vectors. The \textbf{cosine similarity} of \( u \) and \( v \) is given by

\[
\cos(u, v) = \frac{u \cdot v}{\|u\| \|v\|}.
\]

- \(-1 \leq \cos(u, v) \leq 1\)
- Vectors that point in roughly the same direction
  \(\rightarrow\) small angle \(\rightarrow\) cosine similarity \(\approx 1\)
- Vectors that point in roughly opposite directions
  \(\rightarrow\) large angle \(\rightarrow\) cosine similarity \(\approx -1\)
- Vectors that are roughly orthogonal
  \(\rightarrow\) roughly right angle \(\rightarrow\) cosine similarity \(\approx 0\)
- Popular in IR to determine the similarity between a document and a query
Dot product: Cosine similarity (2)
The Pearson correlation coefficient is a measure of linear correlation. It is given by

\[ \rho_{x,y} = \frac{(x - \bar{x}) \cdot (y - \bar{y})}{\|x - \bar{x}\| \|y - \bar{y}\|}. \]

- Numerator proportional to the sample covariance
- Denominator proportional to sample standard deviations
- Closely related to cosine similarity but performs centering
  ▶ This is sometimes desired
  ▶ And sometimes a bad idea (e.g., example last slide)
Dot product: Similarity

The dot product itself can also be seen as a measure of similarity or compatibility. Recall

\[ u \cdot v = \|u\| \|v\| \cos \theta. \]

Example: Shopping transactions

- Like in previous example, vectors \( u \) and \( v \) correspond to persons
- Elements correspond to frequencies of buying each product
- We can think of the direction of a vector as “preference”
  - Which products are being bought?
  - \( \cos \theta \) large when \( u \) and \( u \) have similar interest
- We can think of the magnitude of a vector as “strength”
  - How much is being bought?
  - \( \|u\| \|v\| \) large when both persons buy a lot
- If \( u \cdot v \) is large, \( u \) and \( v \) have similar shopping behaviour and buy a lot
Dot product: Projection

The vector projection of $\mathbf{v}$ onto $\mathbf{u}$ is given by

$$\text{proj}_u(\mathbf{v}) = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|} \frac{\mathbf{u}}{\|\mathbf{u}\|} = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|^2} \mathbf{u}$$

- The scalar projection describes how far $\mathbf{v}$ points in the direction $\mathbf{u}$
- The vector projection is a vector pointing this far in the direction of $\mathbf{u}$
Outline

1. Vectors

2. Matrices

3. Summary
Notation

Let \( A \in \mathbb{R}^{m \times n} \) be a real \( m \times n \) matrix. We write

- \( a_{ij} \) or \( A_{ij} \) (both scalars) for the value of entry \((i, j)\)
- \( a_j \) or \( A_{\ast j} \) (both column vectors) for the \( j \)-th column of \( A \)
- \( a_i \) (column vector) of \( A_{i \ast} \) (row vector) for the \( i \)-th row of \( A \)

Thus

\[
A = \begin{pmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
a_{m1} & a_{m2} & \cdots & a_{mn}
\end{pmatrix}
\]

\[
= \left( \begin{array}{c}
A_{\ast 1} \\
A_{\ast 2} \\
\vdots \\
A_{\ast n}
\end{array} \right) = \begin{pmatrix}
A_{1 \ast} \\
A_{2 \ast} \\
\vdots \\
A_{n \ast}
\end{pmatrix}
\]
Full matrix ring (addition)

The set of all matrices in $\mathbb{R}^{n \times n}$ form a ring, the **full matrix ring**.

- Addition and substraction are element-wise

  $[A + B]_{ij} = a_{ij} + b_{ij}$
  $[A - B]_{ij} = a_{ij} - b_{ij}$

- The additive identity is the $n \times n$ zero matrix $0_{n \times n}$
- The additive inverse is $-A$ with $[-A]_{ij} = -a_{ij}$
- In general $[cA]_{ij} = ca_{ij}$ for $c \in \mathbb{R}$ (scalar multiplication)
- Addition is associative and commutative
Full matrix ring (multiplication)

- For multiplication, we take dot products

\[ [AB]_{ij} = a_i \cdot b_j = \sum_{k=1}^{n} a_{ik} b_{kj} \]

- The multiplicative identity is the \( n \times n \) identity matrix \( I_n \)

\[
I_n = \begin{pmatrix}
1 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 1
\end{pmatrix}
\]

- Multiplication is associative, but not commutative (\( AB \neq BA \) in general)

- Multiplication distributes over addition (\( A(B + C) = AB + AC \) and \( (B + C)A = BA + CA \))

- Multiplication does not always have an inverse (division)
Rectangular matrices

- We generally have rectangular matrices $A \in \mathbb{R}^{m \times n}$
- We can only add and subtract matrices of the same dimensions \((A_{m \times n} + B_{m \times n})\)
- We can only multiply matrices with a matching inner dimension
  - We can multiply $A \in \mathbb{R}^{m \times r}$ with $B \in \mathbb{R}^{r \times n}$ (inner dimension is $r$)
  - Gives an $m \times n$ matrix (outer dimensions)
  - $[AB]_{ij} = a_i \cdot b_j = \sum_{k=1}^{r} a_{ik} b_{kj}$
Interpretation for matrix multiplication (1)

When we multiply $A$ and $B$, we compute all dot products between rows of $A$ and columns of $B$.

- We can apply any of the interpretations of the dot product
- E.g., weighted sum
  - $m$ supermarkets, $r$ products, $n$ persons
  - $a_{ik} =$ price of product $k$ at supermarket $i$
  - $b_{kj} =$ quantity of product $k$ bought by person $i$
  - $[AB]_{ij} =$ how much the $j$-th person would pay when buying at the $i$-th supermarket
- E.g., covariance
  - If all rows of $A_{m\times n}$ are centered ($\sum_k a_{ik} = 0$), then $\frac{1}{m}AA^T \in \mathbb{R}^{m\times n}$ is the sample covariance matrix
  - $[\frac{1}{m}AA^T]_{ii}$ holds the sample variance of row $i$
  - $[\frac{1}{m}AA^T]_{ij}$ holds the sample covariance between rows $i$ and $j$
Interpretation for matrix multiplication (2)

We can also interpret rows $i$ of $AB$ as a linear combination of the rows of $B$ with the coefficients coming from $A$

$$[AB]_{i*} = a_{i1}B_{1*} + a_{i2}B_{2*} + \cdots + a_{ik}B_{r*},$$

and, similarly, the columns of $AB$ as linear combinations of the columns of $A$

$$[AB]_{*j} = b_{1j}A_{*1} + b_{2j}A_{*2} + \cdots + b_{kj}A_{*r}.$$
Interpretation for matrix multiplication (3)

We can view matrix $AB$ as the sum of the $r$ component matrices obtained by multiplying the $k$-th column of $A$ and the $k$-th row of $B$:

$$AB = A_{*1}B_{1*} + A_{*2}B_{2*} + \cdots + A_{*r}B_{r*}$$

- Components $A_{*k}B_{k*}$ are outer products ($m \times n$ matrices)
- Note: when $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$, the matrix product
  - $u^T v$ corresponds to a dot product (a scalar), $m = n$ required
  - $uv^T$ corresponds to an outer product (an $m \times n$ matrix)
- In our supermarket example
  - Components correspond to products
  - Entry $(i, j)$ of $k$-th component indicates how much the $j$-th person would pay for product $k$ when buying at the $i$-th supermarket

$$\begin{array}{c}
\text{\textcolor{green}{Green}} \\
\text{\textcolor{red}{Red}} \\
\text{\textcolor{blue}{Blue}} \\
\text{\textcolor{purple}{Purple}}
\end{array} = \begin{array}{c}
\text{\textcolor{green}{Green}} \\
\text{\textcolor{red}{Red}} \\
\text{\textcolor{blue}{Blue}} \\
\text{\textcolor{purple}{Purple}}
\end{array}$$
Transposes

The **matrix transpose** $A^T$ switches rows and columns, i.e.,

$$[A^T]_{ij} = a_{ji}.$$ 

The following properties hold

- $(A^T)^T = A$
- $(A + B)^T = A^T + B^T$
- $(cA)^T = cA^T$
- $(AB)^T = B^T A^T$
Summing and scaling

Let $A \in \mathbb{R}^{m \times n}$. Denote by $1_n$ the all-ones vector of dimensionality $n$. For $s \in \mathbb{R}^n$, denote by $\text{diag}(s)$ the $n \times n$ matrix with the entries of $s$ on the main diagonal:

$$\text{diag}(s) = \begin{pmatrix} s_1 & 0 & \cdots & 0 \\ 0 & s_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_n \end{pmatrix}$$

- $A1_n$ computes the row sums of $A$
- $1_m^TA$ computes the column sums of $A$
- $A\text{diag}(c)$ scales each column $j$ of $A$ by $c_j$, $c \in \mathbb{R}^n$
- $\text{diag}(r)A$ scales each row $i$ of $A$ by $r_i$, $r \in \mathbb{R}^m$
Matrices as linear maps

- A matrix $M \in \mathbb{R}^{m \times n}$ is a **linear map** from $\mathbb{R}^n$ to $\mathbb{R}^m$
  - If $x \in \mathbb{R}^n$ then $y = Mx \in \mathbb{R}^m$ is the **image** of $x$
  - $y = \sum_{j=1}^n m_j x_j$, i.e., a linear combination of the columns of $M$
- If $A \in \mathbb{R}^{m \times r}$ and $B \in \mathbb{R}^{r \times n}$, then $AB$ maps from $\mathbb{R}^n$ to $\mathbb{R}^m$
  - Product $AB$ corresponds to composition of linear maps $A$ and $B$
- Square matrix $A \in \mathbb{R}^{n \times n}$ is **invertible** (=$=$ **nonsingular**) if there is matrix $B \in \mathbb{R}^{n \times n}$ such that $AB = I$
  - Matrix $B$ is the **inverse** of $A$, denoted $A^{-1}$
  - If $A$ is invertible, then $AA^{-1} = A^{-1}A = I$
    - $AA^{-1}x = A^{-1}Ax = x$
  - Non-square matrices do not have (general) inverses but can have **left** or **right inverses**: $AR = I$ or $LA = I$
- The **transpose** of $M \in \mathbb{R}^{m \times n}$ is a linear map $M^T : \mathbb{R}^m \to \mathbb{R}^n$
  - $(M^T)_{ij} = M_{ji}$
  - Generally, transpose is **not** the inverse ($AA^T \neq I$)
Matrix norms

- Matrix norms measure the magnitude of the matrix
  - Magnitude of the values in the matrix
  - Magnitude of the image

- **Operator norms** measure how large the image of an unit vector can get
  - Induced by a vector norm
  - For $p \geq 1$, $\|M\|_p = \max\{\|Mx\|_p \mid \|x\|_p = 1\}$
  - $\|M\|_1 = \text{maximum sum of absolute values of a column}$
  - $\|M\|_\infty = \text{maximum sum of absolute values of a row}$
  - Spectral norm: $\|M\|_2 = \text{largest singular value of } A$ (more later)

- The **Frobenius norm** is the vector-$L_2$ norm applied to matrices (treating them as a vector)
  - $\|M\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} m_{ij}^2}$
  - Note: $\|M\|_F \neq \|M\|_2$ (but sometimes Frobenius norm is referred to as $L_2$ norm)
Matrix rank and linear independence

- A vector $u \in \mathbb{R}^n$ is **linearly dependent** on set of vectors $V = \{v_i\} \subset \mathbb{R}^n$ if $u$ can be expressed as a linear combination of vectors in $V$
  - $u = \sum_i a_i v_i$ for some $a_i \in \mathbb{R}$
  - Set $V$ is linearly dependent if some $v_i \in V$ is linearly dependent on $V \setminus \{v_i\}$
  - If $V$ is not linearly dependent, it is **linearly independent**

- The **column rank** of matrix $M$ is the maximum number of linearly independent columns of $M$

- The **row rank** of $M$ is the maximum number of linearly independent rows of $M$

- The **Schein rank** of $M$ is the least integer $r$ such that $M = AB$ for some $A \in \mathbb{R}^{m \times r}$ and $B \in \mathbb{R}^{r \times n}$
  - Equivalently, the least $r$ such that $M$ is a sum of $r$ vector outer products

- **All these ranks are equivalent**
  - Matrix has rank 1 iff it is an outer product of two vectors (not zero)
Matrices as systems of linear equations

- A matrix can hold the coefficients of a system of linear equations (c.f. Chinese *Nine Chapters on Arithmetic*)

\[
\begin{align*}
a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,m}x_m &= b_1 \\
a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,m}x_m &= b_2 \\
&\vdots \\
a_{n,1}x_1 + a_{n,2}x_2 + \cdots + a_{n,m}x_m &= b_n \\
\end{align*}
\]

\[
\begin{pmatrix}
a_{1,1} & a_{1,2} & \cdots & a_{1,m} \\
a_{2,1} & a_{2,2} & \cdots & a_{2,m} \\
& \vdots & \ddots & \vdots \\
a_{n,1} & a_{n,2} & \cdots & a_{n,m}
\end{pmatrix}
\begin{pmatrix}
x_1 \\
x_2 \\
\vdots \\
x_m
\end{pmatrix}
= 
\begin{pmatrix}
b_1 \\
b_2 \\
\vdots \\
b_n
\end{pmatrix}
\]

- If the coefficient matrix \(A\) is invertible, the system has exact solution \(x = A^{-1}b\)
- If \(m < n\) the system is **underdetermined** and can have infinite number of solutions
- If \(m > n\) the system is **overdetermined** and (usually) does not have an exact solution
- The **least-squares** solution is the vector \(x\) that minimizes \(\|Ax - b\|_2^2\) (cf. linear regression)
Special types of matrices

- The **diagonals** of matrix $M$ go from top-left to bottom-right
  - The main diagonal contains the elements $m_{ii}$
  - The $k$-th upper diagonal contains the elements $m_{i,(i+k)}$
  - The $k$-th lower diagonal contains the elements $m_{(i+k),i}$
  - The **anti-diagonals** go from top-right to bottom-left
- Matrix is **diagonal** if all its non-zero values are in a diagonal (typically main diagonal)
  - **Bi-diagonal** matrices have values in two diagonals, etc.
- Matrix $M$ is **upper (right) triangular** if all of its non-zeros are in or above the main diagonal
  - **Lower (left) triangular** matrices have all non-zeros in or below main diagonal
  - Upper left and lower right triangular matrices replace diagonal with anti-diagonal
- A square matrix $P$ is **permutation matrix** if each row and each column of $P$ has exactly one 1 and rest are 0s
  - If $P$ is a permutation matrix, $PM$ permutes the order of the rows and $MP$ the order of the columns
Orthogonal matrices

- A set $V = \{v_i\} \subset \mathbb{R}^n$ is **orthogonal** if all vectors in $V$ are mutually orthogonal
  - $v \cdot u = 0$ for all $v, u \in V$
  - If all vectors in $V$ also have unit norm ($\|v\|_2 = 1$), $V$ is **orthonormal**

- A square matrix $M$ is **orthogonal** if its columns are a set of orthonormal (!) vectors or equivalently
  - Its rows are orthonormal (and vice versa)
  - $M^T M = I_n$
  - $M^{-1} = M^T$

- An $m \times n$ matrix $M$ is
  - **column-orthogonal** if columns are a set of orthonormal vectors (only possible if $m \geq n$); then $M^T$ is left inverse ($M^T M = I_n$)
  - **row-orthogonal** if rows are a set of orthonormal vectors (only possible if $m \leq n$); then $M^T$ is right inverse ($MM^T = I_m$)

- If $A$ and $B$ are orthogonal, so is $AB$
  - Similarly: column-orthogonality and row-orthogonality is preserved
Lessons learned

- Many uses, many interpretations
  - Vectors
  - Matrices
  - Dot products
  - Matrix products
- Magnitudes and distances are measured by norms
- Basic concepts of linear algebra
- Special types of matrices: diagonal, triangular, orthogonal
Suggested reading

- Any (elementary) linear algebra text book
  - Carl Meyer
    *Matrix Analysis and Applied Linear Algebra*
    Society for Industrial and Applied Mathematics, 2000
    [http://www.matrixanalysis.com](http://www.matrixanalysis.com)

- [Wolfram MathWorld](http://www.mathworld.wolfram.com) articles

- [Wikipedia](http://www.wikipedia.org) articles