2. Norm, distance, angle

- norm
- distance
- $k$-means algorithm
- angle
- hyperplanes
- complex vectors
Euclidean norm

(Euclidean) norm of vector \( a \in \mathbb{R}^n \):

\[ \| a \| = \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2} \]
\[ = \sqrt{a^T a} \]

- if \( n = 1 \), \( \| a \| \) reduces to absolute value \( |a| \)
- measures the magnitude of \( a \)
- sometimes written as \( \| a \|_2 \) to distinguish from other norms, e.g.,

\[ \| a \|_1 = |a_1| + |a_2| + \cdots + |a_n| \]
Properties

Positive definiteness

\[ \|a\| \geq 0 \quad \text{for all } a, \quad \|a\| = 0 \quad \text{only if } a = 0 \]

Homogeneity

\[ \|\beta a\| = |\beta|\|a\| \quad \text{for all vectors } a \text{ and scalars } \beta \]

Triangle inequality

\[ \|a + b\| \leq \|a\| + \|b\| \quad \text{for all vectors } a \text{ and } b \text{ of equal length} \]

(proof on page 2-7)
Cauchy-Schwarz inequality

\[ |a^T b| \leq \|a\| \|b\| \quad \text{for all } a, b \in \mathbb{R}^n \]

moreover, equality \( |a^T b| = \|a\| \|b\| \) holds if:

- \( a = 0 \) or \( b = 0 \); in this case \( a^T b = 0 = \|a\| \|b\| \)
- \( a \neq 0 \) and \( b \neq 0 \), and \( b = \gamma a \) for some \( \gamma > 0 \); in this case
  \[ 0 < a^T b = \gamma \|a\|^2 = \|a\| \|b\| \]
- \( a \neq 0 \) and \( b \neq 0 \), and \( b = -\gamma a \) for some \( \gamma > 0 \); in this case
  \[ 0 > a^T b = -\gamma \|a\|^2 = -\|a\| \|b\| \]
Proof of Cauchy-Schwarz inequality

1. trivial if $a = 0$ or $b = 0$

2. assume $\|a\| = \|b\| = 1$; we show that $-1 \leq a^T b \leq 1$

$$0 \leq \|a - b\|^2 = (a - b)^T (a - b) = \|a\|^2 - 2a^T b + \|b\|^2 = 2(1 - a^T b)$$

with equality only if $a = b$

$$0 \leq \|a + b\|^2 = (a + b)^T (a + b) = \|a\|^2 + 2a^T b + \|b\|^2 = 2(1 + a^T b)$$

with equality only if $a = -b$

3. for general nonzero $a$, $b$, apply case 2 to the unit-norm vectors

$$\frac{1}{\|a\|} a, \quad \frac{1}{\|b\|} b$$
Average and RMS value

let $a$ be a real $n$-vector

- the *average* of the elements of $a$ is

$$
\text{avg}(a) = \frac{a_1 + a_2 + \cdots + a_n}{n} = \frac{1^T a}{n}
$$

- the *root-mean-square* value is the root of the average squared entry

$$
\text{rms}(a) = \sqrt{\frac{a_1^2 + a_2^2 + \cdots + a_n^2}{n}} = \frac{\|a\|}{\sqrt{n}}
$$

**Exercises**

- show that $| \text{avg}(a) | \leq \text{rms}(a)$

- show that average of $b = (|a_1|, |a_2|, \ldots, |a_n|)$ satisfies $\text{avg}(b) \leq \text{rms}(a)$
Triangle inequality from Cauchy-Schwarz inequality

for vectors $a$, $b$ of equal size

$$\|a + b\|^2 = (a + b)^T(a + b)$$
$$= a^T a + b^T a + a^T b + b^T b$$
$$= \|a\|^2 + 2a^T b + \|b\|^2$$
$$\leq \|a\|^2 + 2\|a\|\|b\| + \|b\|^2 \quad \text{(by Cauchy-Schwarz)}$$
$$= (\|a\| + \|b\|)^2$$

- taking squareroots gives the triangle inequality
- triangle inequality is an equality if and only if $a^T b = \|a\|\|b\|$ (see page 2-4)
- also note from line 3 that $\|a + b\|^2 = \|a\|^2 + \|b\|^2$ if $a^T b = 0$
Outline

- norm
- **distance**
- $k$-means algorithm
- angle
- hyperplanes
- complex vectors
Distance

the (Euclidean) distance between vectors $a$ and $b$ is defined as $\|a - b\|$

- $\|a - b\| \geq 0$ for all $a$, $b$ and $\|a - b\| = 0$ only if $a = b$

- triangle inequality

$$\|a - c\| \leq \|a - b\| + \|b - c\| \quad \text{for all } a, b, c$$

- RMS deviation between $n$-vectors $a$ and $b$ is $\text{rms}(a - b) = \frac{\|a - b\|}{\sqrt{n}}$

Norm, distance, angle 2-8
Standard deviation

let $a$ be a real $n$-vector

- the *de-meaned* vector is the vector of deviations from the average

$$a - \text{avg}(a) \mathbf{1} = \begin{bmatrix} a_1 - \text{avg}(a) \\ a_2 - \text{avg}(a) \\ \vdots \\ a_n - \text{avg}(a) \end{bmatrix} = \begin{bmatrix} a_1 - (\mathbf{1}^T a)/n \\ a_2 - (\mathbf{1}^T a)/n \\ \vdots \\ a_n - (\mathbf{1}^T a)/n \end{bmatrix}$$

- the *standard deviation* is the RMS deviation from the average

$$\text{std}(a) = \text{rms}(a - \text{avg}(a) \mathbf{1}) = \frac{\|a - ((\mathbf{1}^T a)/n) \mathbf{1}\|}{\sqrt{n}}$$

- the de-meaned vector in *standard units* is

$$\frac{1}{\text{std}(a)}(a - \text{avg}(a) \mathbf{1})$$
Mean return and risk of investment

- Vectors represent time series of returns on an investment (as a percentage).
- Average value is \textit{(mean) return} of the investment.
- Standard deviation measures variation around the mean, \textit{i.e.}, \textit{risk}.

![Graphs showing mean return and risk](image)
Exercise

show that

\[ \text{avg}(a)^2 + \text{std}(a)^2 = \text{rms}(a)^2 \]

Solution

\[
\text{std}(a)^2 = \frac{\|a - \text{avg}(a) \cdot 1\|^2}{n}
\]

\[
= \frac{1}{n} \left( a - \frac{1^T a}{n} \cdot 1 \right)^T \left( a - \frac{1^T a}{n} \cdot 1 \right)
\]

\[
= \frac{1}{n} \left( a^T a - \frac{(1^T a)^2}{n} - \frac{(1^T a)^2}{n} + \left( \frac{1^T a}{n} \right)^2 n \right)
\]

\[
= \frac{1}{n} \left( a^T a - \frac{(1^T a)^2}{n} \right)
\]

\[
= \text{rms}(a)^2 - \text{avg}(a)^2
\]
Exercise: nearest scalar multiple

given two vectors $a, b \in \mathbb{R}^n$, with $a \neq 0$, find scalar multiple $ta$ closes to $b$

Solution

- squared distance between $ta$ and $b$ is

$$\|ta - b\|^2 = (ta - b)^T(ta - b) = t^2a^Ta - 2ta^ Tb + b^Tb$$

a quadratic function of $t$ with positive leading coefficient $a^Ta$

- derivative with respect to $t$ is zero for

$$\hat{t} = \frac{a^Tb}{a^Ta} = \frac{a^Tb}{\|a\|^2}$$
Exercise: average of collection of vectors

given \( N \) vectors \( x_1, \ldots, x_N \in \mathbb{R}^n \), find the \( n \)-vector \( z \) that minimizes

\[
\| z - x_1 \|^2 + \| z - x_2 \|^2 + \cdots + \| z - x_N \|^2
\]

\( z \) is also known as the centroid of the points \( x_1, \ldots, x_N \)
Solution: sum of squared distances is

\[ \| z - x_1 \|^2 + \| z - x_2 \|^2 + \cdots + \| z - x_N \|^2 \]

\[ = \sum_{i=1}^{n} \left( (z_i - (x_1)_i)^2 + (z_i - (x_2)_i)^2 + \cdots + (z_i - (x_N)_i)^2 \right) \]

\[ = \sum_{i=1}^{n} \left( N z_i^2 - 2z_i ((x_1)_i + (x_2)_i + \cdots + (x_N)_i) + (x_1)_i^2 + \cdots + (x_N)_i^2 \right) \]

here \((x_j)_i\) is \(i\)th element of the vector \(x_j\)

- term \(i\) in the sum is minimized by
  \[ z_i = \frac{1}{N} ((x_1)_i + (x_2)_i + \cdots + (x_N)_i) \]

- solution \(z\) is component-wise average of the points \(x_1, \ldots, x_N\):
  \[ z = \frac{1}{N} (x_1 + x_2 + \cdots + x_N) \]
Outline

- norm
- distance
- $k$-means algorithm
- angle
- hyperplanes
- complex vectors
$k$-means clustering

a popular iterative algorithm for partitioning $N$ vectors $x_1, \ldots, x_N$ in $k$ clusters
Algorithm

choose initial ‘representatives’ $z_1, \ldots, z_k$ for the $k$ groups and repeat:

1. assign each vector $x_i$ to the nearest group representative $z_j$
2. set the representative $z_j$ to the mean of the vectors assigned to it

• as a variation, choose a random initial partition and start with step 2
• initial representatives are often chosen randomly
• solution depends on choice of initial representatives or partition
• can be shown to converge in a finite number of iterations
• in practice, often restarted a few times, with different starting points
Example: first iteration

assignment to groups

updated representatives

Norm, distance, angle
Example: iteration 2

assignment to groups

updated representatives

Norm, distance, angle
Example: iteration 3

assignment to groups  
updated representatives

Norm, distance, angle
Example: iteration 9

assignment to groups

updated representatives

Norm, distance, angle
Example: iteration 10

assignment to groups
updated representatives

Norm, distance, angle
Example: iteration 11

assignment to groups  updated representatives

Norm, distance, angle
Example: iteration 12

assignment to groups  updated representatives
Image clustering

- MNIST dataset of handwritten digits
- $N = 60,000$ grayscale images of size $28 \times 28$ (vectors $x_i$ of size $28^2 = 784$)
- 25 examples:
Group representatives \((k = 20)\)

- \(k\)-means algorithm, with \(k = 20\) and randomly chosen initial partition
- 20 group representatives
Group representatives \((k = 20)\)

result for another initial partition
Document topic discovery

- \( N = 500 \) Wikipedia articles, from weekly most popular lists (9/2015–6/2016)
- dictionary of 4423 words
- each article represented by a word histogram vector of size 4423
- result of \( k \)-means algorithm with \( k = 9 \) and randomly chosen initial partition

Cluster 1

- largest coefficients in cluster representative \( z_1 \)

<table>
<thead>
<tr>
<th>word</th>
<th>fight</th>
<th>win</th>
<th>event</th>
<th>champion</th>
<th>fighter</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>coefficient</td>
<td>0.038</td>
<td>0.022</td>
<td>0.019</td>
<td>0.015</td>
<td>0.015</td>
<td>...</td>
</tr>
</tbody>
</table>

- documents in cluster 1 closest to representative

  “Floyd Mayweather, Jr”, “Kimbo Slice”, “Ronda Rousey”, “José Aldo”, “Joe Frazier”, ...
Cluster 2

• largest coefficients in cluster representative $z_2$

<table>
<thead>
<tr>
<th>word</th>
<th>holiday</th>
<th>celebrate</th>
<th>festival</th>
<th>celebration</th>
<th>calendar</th>
</tr>
</thead>
<tbody>
<tr>
<td>coefficient</td>
<td>0.012</td>
<td>0.009</td>
<td>0.007</td>
<td>0.006</td>
<td>0.006</td>
</tr>
</tbody>
</table>

• documents in cluster 2 closest to representative


Cluster 3

• largest coefficients in cluster representative $z_3$

<table>
<thead>
<tr>
<th>word</th>
<th>united</th>
<th>family</th>
<th>party</th>
<th>president</th>
<th>government</th>
</tr>
</thead>
<tbody>
<tr>
<td>coefficient</td>
<td>0.004</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
</tr>
</tbody>
</table>

• documents in cluster 3 closest to representative

“Mahatma Gandhi”, “Sigmund Freund”, “Carly Fiorina”, “Frederick Douglass”, “Marco Rubio”, …
Cluster 4

- largest coefficients in cluster representative $z_4$

<table>
<thead>
<tr>
<th>word</th>
<th>album</th>
<th>release</th>
<th>song</th>
<th>music</th>
<th>single</th>
</tr>
</thead>
<tbody>
<tr>
<td>coefficient</td>
<td>0.031</td>
<td>0.016</td>
<td>0.015</td>
<td>0.014</td>
<td>0.011</td>
</tr>
</tbody>
</table>

- documents in cluster 4 closest to representative


Cluster 5

- largest coefficients in cluster representative $z_5$

<table>
<thead>
<tr>
<th>word</th>
<th>game</th>
<th>season</th>
<th>team</th>
<th>win</th>
<th>player</th>
</tr>
</thead>
<tbody>
<tr>
<td>coefficient</td>
<td>0.023</td>
<td>0.020</td>
<td>0.018</td>
<td>0.017</td>
<td>0.014</td>
</tr>
</tbody>
</table>

- documents in cluster 5 closest to representative

  “Kobe Bryant”, “Lamar Odom”, “Johan Cruyff”, “Yogi Berra”, “José Mourinho”, …
Cluster 6

- largest coefficients in representative $z_6$

<table>
<thead>
<tr>
<th>word</th>
<th>series</th>
<th>season</th>
<th>episode</th>
<th>character</th>
<th>film</th>
<th>…</th>
</tr>
</thead>
<tbody>
<tr>
<td>coefficient</td>
<td>0.029</td>
<td>0.027</td>
<td>0.013</td>
<td>0.011</td>
<td>0.008</td>
<td>…</td>
</tr>
</tbody>
</table>

- documents in cluster 6 closest to cluster representative

“The X-Files”, “Game of Thrones”, “House of Cards”, “Daredevil”, “Supergirl”, …

Cluster 7

- largest coefficients in representative $z_7$

<table>
<thead>
<tr>
<th>word</th>
<th>match</th>
<th>win</th>
<th>championship</th>
<th>team</th>
<th>event</th>
<th>…</th>
</tr>
</thead>
<tbody>
<tr>
<td>coefficient</td>
<td>0.065</td>
<td>0.018</td>
<td>0.016</td>
<td>0.015</td>
<td>0.015</td>
<td>…</td>
</tr>
</tbody>
</table>

- documents in cluster 7 closest to cluster representative

Cluster 8

- largest coefficients in representative $z_8$

<table>
<thead>
<tr>
<th>word</th>
<th>film</th>
<th>star</th>
<th>role</th>
<th>play</th>
<th>series</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>coeff.</td>
<td>0.036</td>
<td>0.014</td>
<td>0.014</td>
<td>0.010</td>
<td>0.009</td>
<td>...</td>
</tr>
</tbody>
</table>

- documents in cluster 8 closest to cluster representative

  “Ben Affleck”, “Johnny Depp”, “Maureen O’Hara”, “Kate Beckinsale”, “Leonardo DiCaprio”, ...

Cluster 9

- largest coefficients in representative $z_9$

<table>
<thead>
<tr>
<th>word</th>
<th>film</th>
<th>million</th>
<th>release</th>
<th>star</th>
<th>character</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>coeff.</td>
<td>0.061</td>
<td>0.019</td>
<td>0.013</td>
<td>0.010</td>
<td>0.006</td>
<td>...</td>
</tr>
</tbody>
</table>

- documents in cluster 9 closest to cluster representative

Outline

- norm
- distance
- $k$-means algorithm
- angle
- hyperplanes
- complex vectors
Angle between vectors

the angle between nonzero real vectors $a$, $b$ is defined as

$$\arccos \left( \frac{a^T b}{\|a\| \|b\|} \right)$$

• this is the unique value of $\theta \in [0, \pi]$ that satisfies $a^T b = \|a\| \|b\| \cos \theta$

• Cauchy-Schwarz inequality guarantees that

$$-1 \leq \frac{a^T b}{\|a\| \|b\|} \leq 1$$
Terminology

\[ \theta = 0 \quad a^T b = \|a\|\|b\| \quad \text{vectors are aligned or parallel} \]

\[ 0 \leq \theta < \pi/2 \quad a^T b > 0 \quad \text{vectors make an acute angle} \]

\[ \theta = \pi/2 \quad a^T b = 0 \quad \text{vectors are orthogonal} \quad (a \perp b) \]

\[ \pi/2 < \theta \leq \pi \quad a^T b < 0 \quad \text{vectors make an obtuse angle} \]

\[ \theta = \pi \quad a^T b = -\|a\|\|b\| \quad \text{vectors are anti-aligned or opposed} \]
Orthogonal decomposition

given a nonzero $a \in \mathbb{R}^n$, every $n$-vector $x$ can be decomposed as

$$x = ta + y \quad \text{with } y \perp a$$

- proof is by inspection
- decomposition (i.e., $t$ and $y$) exists and is unique for every $x$
- $ta$ is projection of $x$ on the line through $a$ and the origin (see page 2-12)
- since $y \perp a$, we have $\|x\|^2 = \|ta\|^2 + \|y\|^2$

$$t = \frac{a^T x}{\|a\|^2}, \quad y = x - \frac{a^T x}{\|a\|^2}a$$
Correlation coefficient

the *correlation coefficient* between non-constant vectors $a, b$ is

\[
\rho_{ab} = \frac{\tilde{a}^T \tilde{b}}{\|\tilde{a}\| \|\tilde{b}\|}
\]

where $\tilde{a} = a - \text{avg}(a) \mathbf{1}$ and $\tilde{b} = b - \text{avg}(b) \mathbf{1}$ are the de-meaned vectors

- only defined when $a$ and $b$ are not constant ($\tilde{a} \neq 0$ and $\tilde{b} \neq 0$

- $\rho_{ab}$ is the cosine of the angle between the de-meaned vectors

- a number between $-1$ and $1$

- $\rho_{ab}$ is the average product of the deviations from the mean in standard units

\[
\rho_{ab} = \frac{1}{n} \sum_{i=1}^{n} \frac{(a_i - \text{avg}(a)) \cdot (b_i - \text{avg}(b))}{\text{std}(a) \cdot \text{std}(b)}
\]
Examples

\[ \rho_{ab} = 0.968 \]

\[ \rho_{ab} = -0.988 \]

\[ \rho_{ab} = 0.004 \]
Regression line

- scatter plot shows two $n$-vectors $a$, $b$ as $n$ points $(a_k, b_k)$
- straight line shows affine function $f(x) = c_1 + c_2x$ with

$$f(a_k) \approx b_k, \quad k = 1, \ldots, n$$
Least squares regression

use coefficients $c_1$, $c_2$ that minimize $J = \frac{1}{n} \sum_{k=1}^{n} (f(a_k) - b_k)^2$

• $J$ is a quadratic function of $c_1$ and $c_2$:

$$J = \frac{1}{n} \sum_{k=1}^{n} (c_1 + c_2 a_k - b_k)^2$$

$$= \left( nc_1^2 + 2n \text{avg}(a) c_1 c_2 + \|a\|^2 c_2^2 - 2n \text{avg}(b) c_1 - 2a^T b c_2 + \|b\|^2 \right) / n$$

• to minimize $J$, set derivatives with respect to $c_1$, $c_2$ to zero:

$$c_1 + \text{avg}(a) c_2 = \text{avg}(b), \quad n \text{avg}(a) c_1 + \|a\|^2 c_2 = a^T b$$

• solution is

$$c_2 = \frac{a^T b - n \text{avg}(a) \text{avg}(b)}{\|a\|^2 - n \text{avg}(a)^2}, \quad c_1 = \text{avg}(b) - \text{avg}(a) c_2$$
Interpretation

slope \( c_2 \) can be written in terms of correlation coefficient of \( a \) and \( b \):

\[
c_2 = \frac{(a - \text{avg}(a) \mathbf{1})^T (b - \text{avg}(b) \mathbf{1})}{\|a - \text{avg}(a) \mathbf{1}\|^2} = \rho_{ab} \frac{\text{std}(b)}{\text{std}(a)}
\]

- hence, expression for regression line can be written as

\[
f(x) = \text{avg}(b) + \frac{\rho_{ab} \text{std}(b)}{\text{std}(a)} (x - \text{avg}(a))
\]

- correlation coefficient \( \rho_{ab} \) is the slope after converting to standard units:

\[
\frac{f(x) - \text{avg}(b)}{\text{std}(b)} = \rho_{ab} \frac{x - \text{avg}(a)}{\text{std}(a)}
\]
Examples

\[ \rho_{ab} = 0.91 \]
\[ \rho_{ab} = -0.89 \]
\[ \rho_{ab} = 0.25 \]

- dashed lines in top row show average ± standard deviation
- bottom row shows scatter plots of top row in standard units
Outline

- norm
- distance
- $k$-means algorithm
- angle
- hyperplanes
- complex vectors
Hyperplane

one linear equation in $n$ variables $x_1, x_2, \ldots, x_n$:

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

in vector notation: $a^T x = b$

let $H$ be the set of solutions: $H = \{x \in \mathbb{R}^n \mid a^T x = b\}$

- $H$ is empty if $a_1 = a_2 = \cdots = a_n = 0$ and $b \neq 0$
- $H = \mathbb{R}^n$ if $a_1 = a_2 = \cdots = a_n = 0$ and $b = 0$
- $H$ is called a hyperplane if $a = (a_1, a_2, \ldots, a_n) \neq 0$
- for $n = 2$, a straight line in a plane; for $n = 3$, a plane in 3-D space, …
Example

\[ a = (2, 1) \]
\[ b = -15 \]
\[ b = -10 \]
\[ b = -5 \]
\[ b = 0 \]
\[ b = 5 \]
\[ b = 10 \]
\[ b = 15 \]
Geometric interpretation of hyperplane

- recall formula for orthogonal decomposition of $x$ with respect to $a$ (page 2-34):
  \[
x = \frac{a^T x}{\|a\|^2} a + y \quad \text{with } y \perp a
  \]

- $x$ satisfies $a^T x = b$ if and only if
  \[
x = \frac{b}{\|a\|^2} a + y \quad \text{with } y \perp a
  \]

- point $(b/\|a\|^2) a$ is the intersection of hyperplane with line through $a$

- add arbitrary vectors $y \perp a$ to get all other points in hyperplane
Exercise: projection on hyperplane

• show that the point in $H = \{x \mid a^T x = b\}$ closest to $c \in \mathbb{R}^n$ is

$$\hat{x} = c + \frac{b - a^T c}{\|a\|^2} a$$

• show that the distance of $c$ to the hyperplane $H = \{x \mid a^T x = b\}$ is

$$\frac{|a^T c - b|}{\|a\|}$$
Solution

\[ H = \{ x \mid a^T x = b \} \]

\[ \hat{x} = c + \frac{b - a^T c}{\|a\|^2} a \]

Norm, distance, angle
Solution

• general point $x$ in $H$ is

$$x = \frac{b}{\|a\|^2}a + y, \quad y \perp a$$

• decomposition of $c$ with respect to $a$ is

$$c = \frac{a^Tc}{\|a\|^2}a + d \quad \text{with} \quad d = c - \frac{a^Tc}{\|a\|^2}a$$

• squared distance between $x$ and $c$ is

$$\|c - x\|^2 = \left\| \frac{a^Tc - b}{a^Ta}a + d - y \right\|^2 = \frac{(a^Tc - b)^2}{\|a\|^2} + \|d - y\|^2$$

(2nd step because $d - y \perp a$); distance is minimized by choosing $y = d$
Kaczmarz algorithm

**Problem:** find (one) solution of set of linear equations

\[ a_1^T x = b_1, \quad a_2^T x = b_2, \quad \ldots, \quad a_m^T x = b_m \]

- here \( a_1, a_2, \ldots, a_m \) are nonzero \( n \)-vectors
- we assume the equations are solvable (have at least one solution)
- \( n \) is huge, so we can only use simple vector operations

**Algorithm:** start at some initial \( x \) and repeat the following steps

- pick an index \( i \in \{1, \ldots, m\} \), for example, cyclically or randomly
- replace \( x \) with projection on hyperplane \( H_i = \{ \tilde{x} \mid a_i^T \tilde{x} = b_i \} \)

\[
x := x + \frac{b_i - a_i^T x}{\|a_i\|^2} a_i
\]
Tomography

reconstruct unknown image from line integrals

- $x$ represents unknown image with $n$ pixels
- $a_{ij}$ is length of intersection of ray $i$ and pixel $j$
- $b_i$ is a measurement of the line integral $\sum_{j=1}^{n} a_{ij} x_j$ along ray $i$

Kaczmarz alg. is also known as *Algebraic Reconstruction Technique* (ART)
Outline

- norm
- distance
- $k$-means algorithm
- angle
- hyperplanes
- complex vectors
Norm

norm of vector $a \in \mathbb{C}^n$:

$$\|a\| = \sqrt{|a_1|^2 + |a_2|^2 + \cdots + |a_n|^2}$$

$$= \sqrt{a^H a}$$

- positive definite:

$$\|a\| \geq 0 \quad \text{for all } a, \quad \|a\| = 0 \quad \text{only if } a = 0$$

- homogeneous:

$$\|\beta a\| = |\beta|\|a\| \quad \text{for all vectors } a, \text{ complex scalars } \beta$$

- triangle inequality:

$$\|a + b\| \leq \|a\| + \|b\| \quad \text{for all vectors } a, b \text{ of equal size}$$
Cauchy-Schwarz inequality for complex vectors

\[ |a^H b| \leq \|a\| \|b\| \quad \text{for all } a, b \in \mathbb{C}^n \]

moreover, equality \( |a^H b| = \|a\| \|b\| \) holds if:

- \( a = 0 \) or \( b = 0 \)
- \( a \neq 0 \) and \( b \neq 0 \), and \( b = \gamma a \) for some (complex) scalar \( \gamma \)

- exercise: generalize proof for real vectors on page 2-4
- we say \( a \) and \( b \) are orthogonal if \( a^H b = 0 \)
- we will not need definition of angle, correlation coefficient, … in \( \mathbb{C}^n \)