Norm, Distance, Angle

CE40282-1: Linear Algebra
Hamid R. Rabiee and Maryam Ramezani
Sharif University of Technology

Chebyshev inequality for standard deviation

x is an n-vector with mean $\mathbf{avg}(x)$, standard deviation $\mathbf{std}(x)$

rough idea: most entries of x are not too far from the mean

by Chebyshev inequality, fraction of entries of x with

$$|x_i - \mathbf{avg}(x)| \geq \alpha \ \mathbf{std}(x)$$

is no more than $1/\alpha^2$ (for $\alpha > 1$)

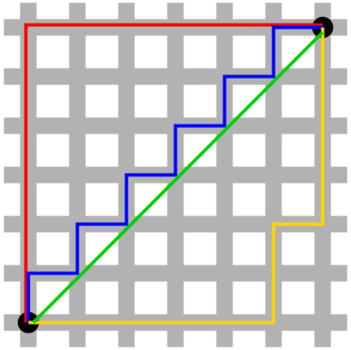
- The fraction of entries of x within θ standard deviations of avg(x) is at least $(1 - \frac{1}{\theta^2})$ for $\theta > 1$

Vector Norms

- 1-norm: $(\boldsymbol{l_1})$

$$||x||_1 = (|x_1| + |x_2| + \cdots + |x_n|)$$

- What is the shape of $||x||_1 = 1$?
- The distance between two vectors under the L1 norm is also referred to as the Manhattan distance
- Example:
  - L1 distance between (0,1) and (1,0)?



Vector Norms

- ∞-norm: ($\boldsymbol{l}_\infty$) (max norm)

$$L_\infty = max(|x_1|, |x_2|, \dots, |x_n|)$$

- What is the shape of $||x||_\infty = 1$?

Vector Norms

- $\frac{\mathbf{1}}{\mathbf{2}}$-norm: $(\boldsymbol{l}_{\frac{\mathbf{1}}{\mathbf{2}}})$

- What is the shape of $||x||_{\frac{1}{2}} = 1$?

Vector Norms

- $zero$-norm: (l_0)

$$||x||_0 = \lim_{\alpha \to 0^+} ||x||_\alpha = \left(\sum_{k=1}^{n} |x|^\alpha \right)^{1/\alpha} = \sum_{k=1}^{n} 1_{(0,\infty)}(|x|)$$

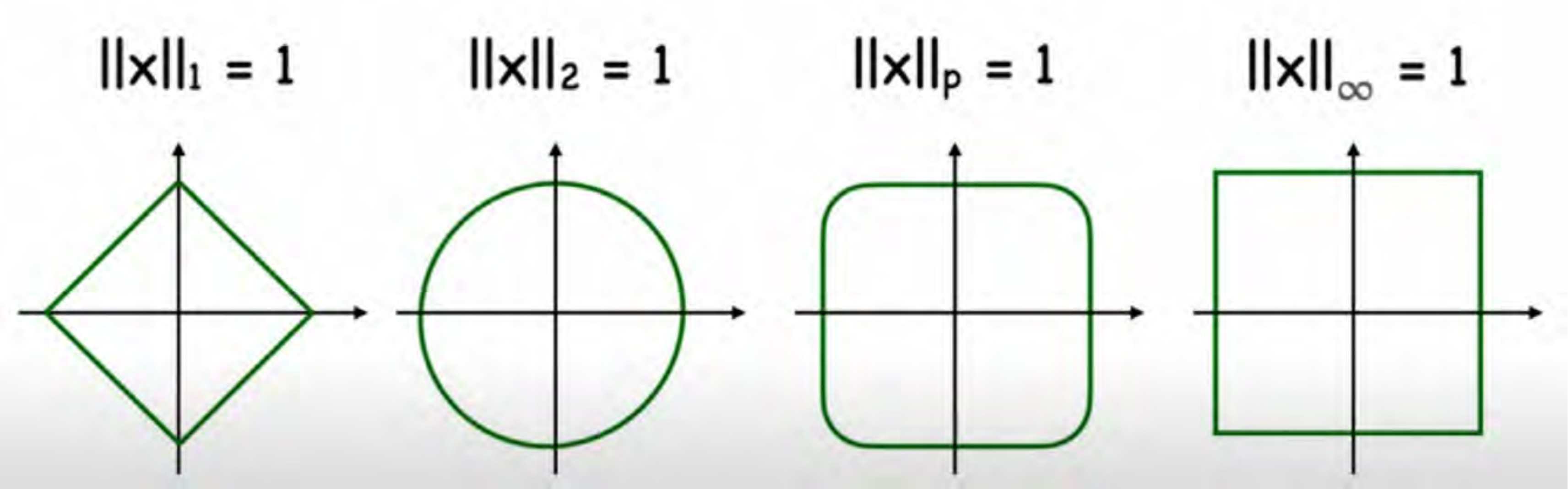
- Zero-norm, defined as **the number of non-zero elements in a vector**, is an ideal quantity for feature selection. However, minimization of zero-norm is generally regarded as a combinatorically difficult optimization
- $||x||_0 = \sum_{x_i \neq 0} 1$

Vector Norms

- Is zero-norm a norm??
- What is the shape of $||x||_0 = 1$?
- Examples:
  - L0 distance between (0,0) and (0,5)?
  - L0 distance between (1,1) and (2,2)?
  - (username,password)

Norms and Convexity

- For $p \geq 1$, l_p norm is convex



Norms and Convexity

Theorem: If A is a convex subset of a normed linear space B whose norm is strictly convex, then, for every f in B, there exists a unique best approximation a in A to f*

Norm Derivations

- Square of l_2

$$u = \begin{bmatrix} u_1 \\ u_2 \\ \cdots \\ u_n \end{bmatrix} \quad ||u||_2 = u_1^2 + u_2^2 + \cdots + u_n^2$$

$$\begin{cases} \dfrac{d||u||_2}{du_1} = 2u_1 \\ \dfrac{d||u||_2}{du_2} = 2u_2 \\ \cdots \\ \dfrac{d||u||_2}{du_n} = 2u_n \end{cases}$$

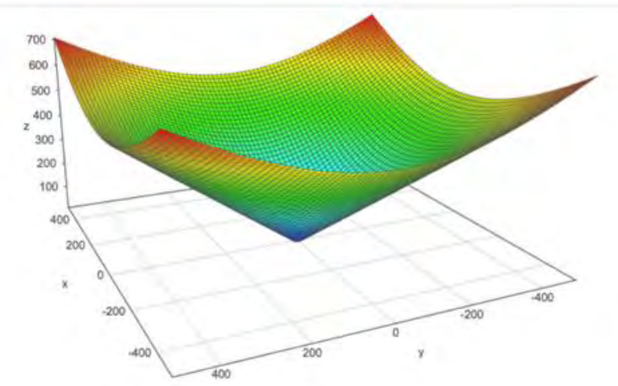
- l_2

$$||u||_2 = \sqrt{(u_1^2 + u_2^2 + \cdots + u_n^2)} = (u_1^2 + u_2^2 + \cdots + u_n^2)^{\frac{1}{2}}$$

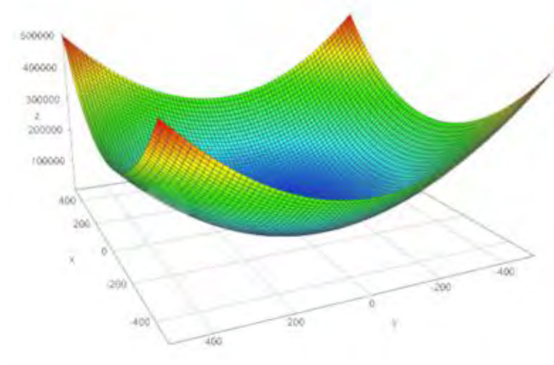
$$\begin{aligned} \frac{d||u||_2}{du_1} &= \frac{1}{2}(u_1^2 + u_2^2 + \cdots + u_n^2)^{\frac{1}{2}-1} \cdot \frac{d}{du_1}(u_1^2 + u_2^2 + \cdots + u_n^2) \\ &= \frac{1}{2}(u_1^2 + u_2^2 + \cdots + u_n^2)^{-\frac{1}{2}} \cdot \frac{d}{du_1}(u_1^2 + u_2^2 + \cdots + u_n^2) \\ &= \frac{1}{2} \cdot \frac{1}{(u_1^2 + u_2^2 + \cdots + u_n^2)^{\frac{1}{2}}} \cdot \frac{d}{du_1}(u_1^2 + u_2^2 + \cdots + u_n^2) \\ &= \frac{1}{2} \cdot \frac{1}{(u_1^2 + u_2^2 + \cdots + u_n^2)^{\frac{1}{2}}} \cdot 2 \cdot u_1 \\ &= \frac{u_1}{\sqrt{(u_1^2 + u_2^2 + \cdots + u_n^2)}} \end{aligned}$$

$$\begin{cases} \dfrac{d||u||_2}{du_1} = \dfrac{u_1}{\sqrt{(u_1^2 + u_2^2 + \cdots + u_n^2)}} \\ \dfrac{d||u||_2}{du_2} = \dfrac{u_2}{\sqrt{(u_1^2 + u_2^2 + \cdots + u_n^2)}} \\ \cdots \\ \dfrac{d||u||_2}{du_n} = \dfrac{u_n}{\sqrt{(u_1^2 + u_2^2 + \cdots + u_n^2)}} \end{cases}$$

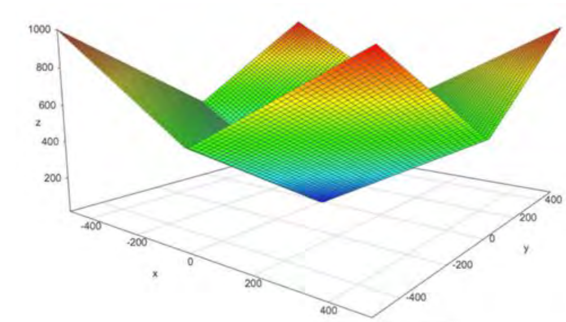
Norm Comparisons



l_2 norm

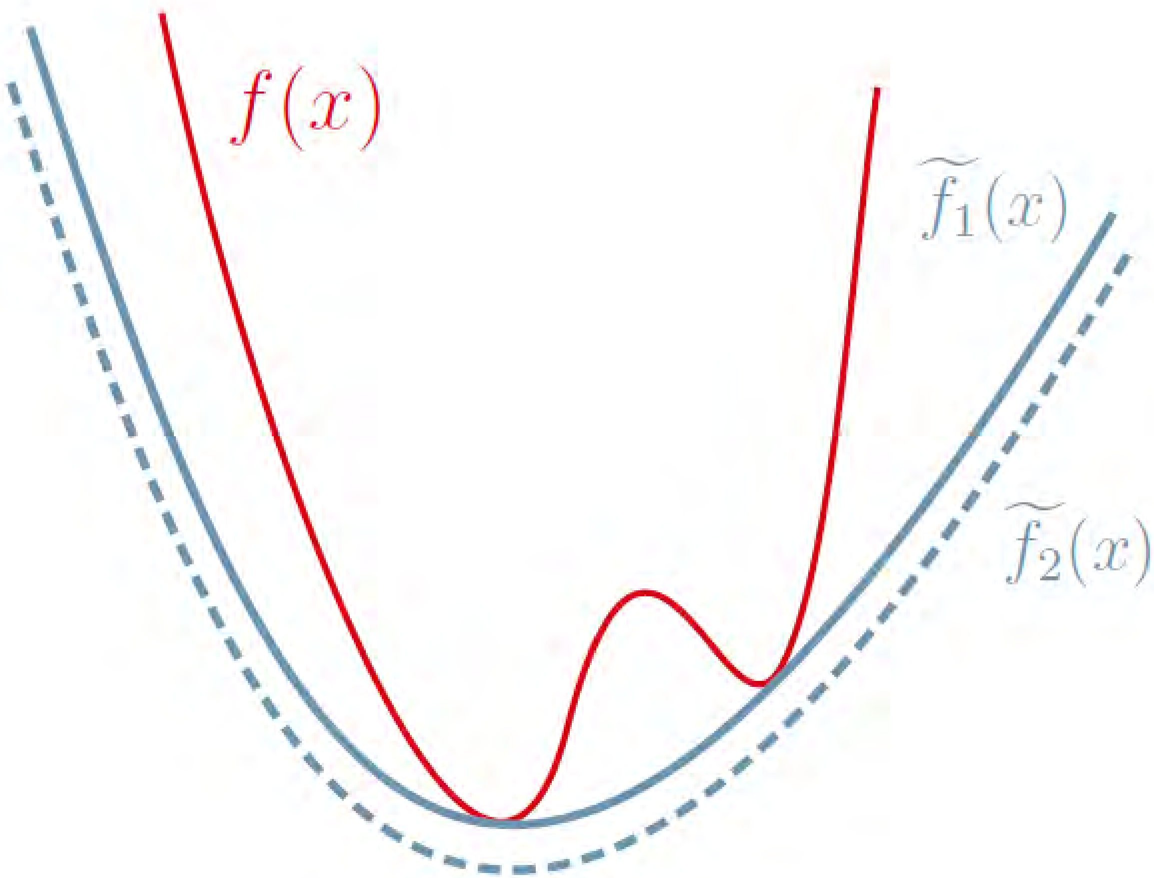


Square l_2 norm



l_1 norm

Convex Relaxation

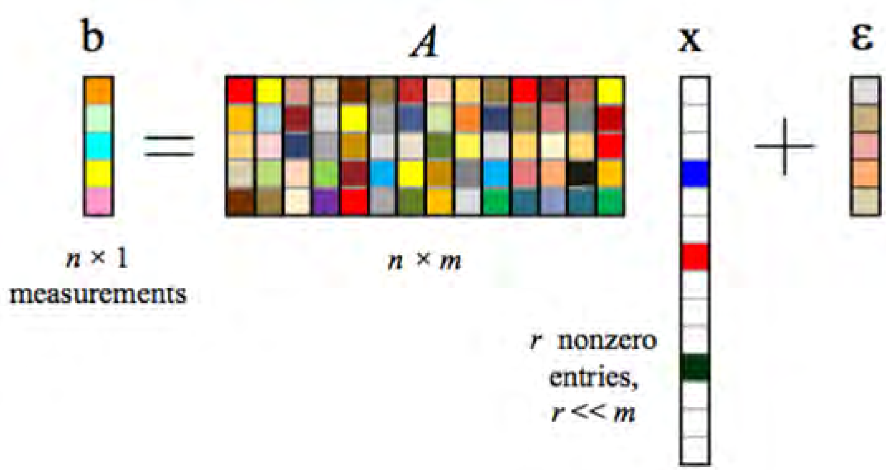


Sparse applications

- **Alternative viewpoint:** We try to find the sparsest solution which explains our noisy measurements

$$\min_{x} \|\mathbf{x}\|_0 \qquad \text{subject to } \|\mathbf{Ax}-\mathbf{b}\|_2 < \varepsilon$$

- Here, the l_0-norm is a shorthand notation for *counting the number of non-zero elements in* x.



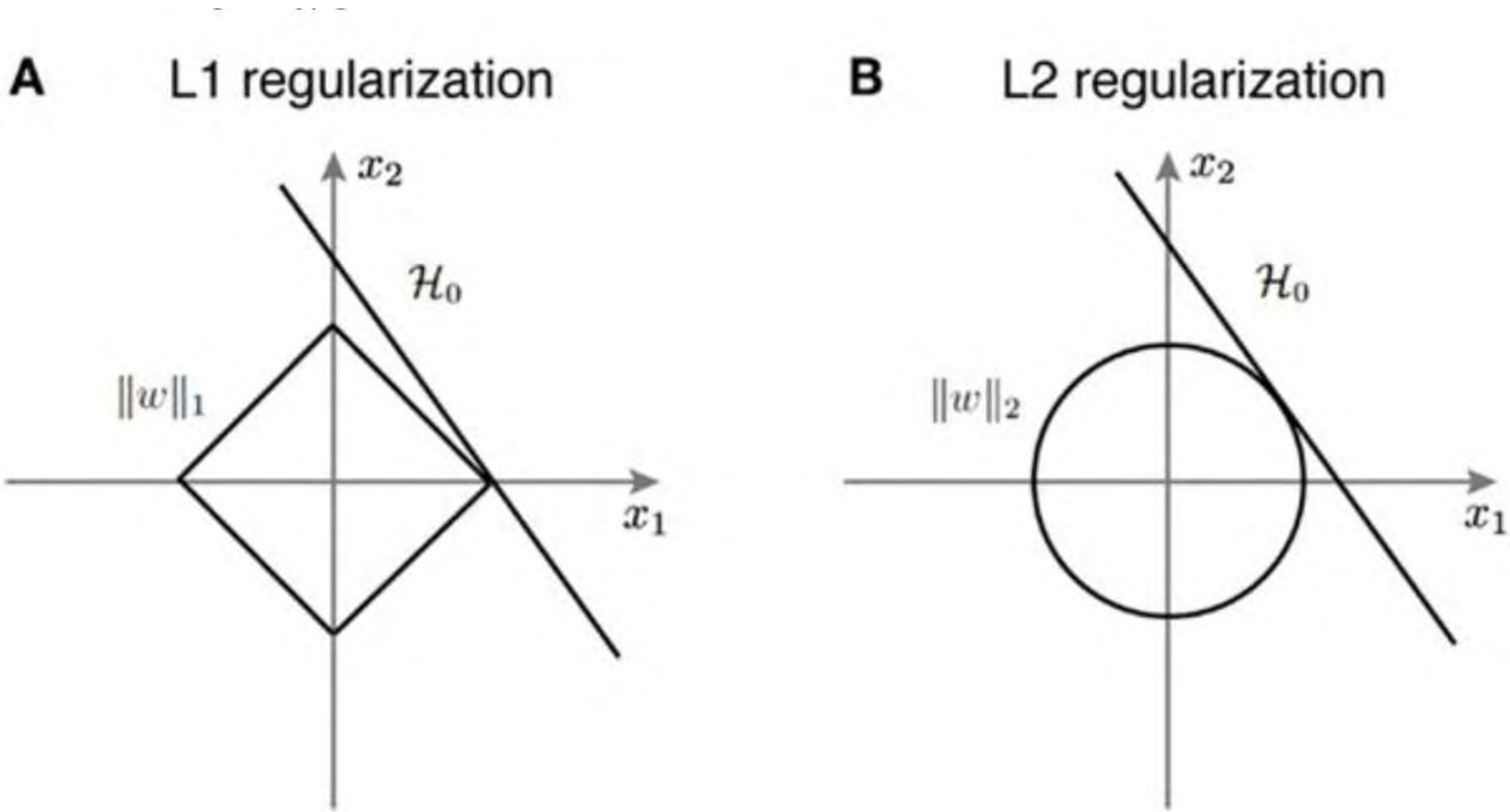
Sparse Solution

- l_0 optimization is np-hard
- Convex relaxation for solving the problem

$$\min_x \|x\|_1 \quad \text{subject to } \|Ax - b\|_2 < \varepsilon$$

$$\min_x \|x\|_0 \quad \text{subject to } \|Ax - b\|_2 < \varepsilon$$

Why is L1 supposed to lead to sparsity than L2?



Vector Norms

- Which norm is the convex hull of the intersection between the L0 norm ball and L2 norm ball?

- Any valid norm ||.|| is a convex function.
  - Proof?

- The L0 norm is not convex.
  - Proof?

Cauchy–Schwarz inequality

- for two n-vectors a and b, $|a^T b| \leq ||a|| ||b||$
- written out,

$$|a_1 b_1 + \cdots + a_n b_n| \leq \left(a_1^2 + \cdots + a_n^2\right)^{1/2} \left(b_1^2 + \cdots + b_n^2\right)^{1/2}$$

Derivation of Cauchy–Schwarz inequality

- it's clearly true if either a or b is 0

  so assume $\alpha = ||a||$ and $\beta = ||b||$ are nonzero

  we have

  $$\begin{aligned} 0 &\leq ||\beta a - \alpha b||^2 \\ &= ||\beta a||^2 - 2(\beta a)^T(\alpha b) + ||\alpha b||^2 \\ &= \beta^2 ||a||^2 - 2\beta\alpha(a^T b) + \alpha^2 ||b||^2 \\ &= 2||a||^2 ||b||^2 - 2||a||\,||b||(a^T b) \end{aligned}$$

  divide by $2||a||\,||b||$ to get $a^T b \leq ||a||\,||b||$

  apply to $-a$, b to get other half of Cauchy–Schwarz inequality

- Cauchy{Schwarz inequality holds with equality when one of the vectors is a multiple of the other

Cauchy–Schwarz inequality

- Verification of triangle inequality.

$$
\begin{aligned}
\|a+b\|^2 &= \|a\|^2 + 2a^Tb + \|b\|^2 \\
&\le \|a\|^2 + 2\|a\|\|b\| + \|b\|^2 \\
&= (\|a\| + \|b\|)^2
\end{aligned}
$$

Angle

- *angle* between two nonzero vectors a, b defined as

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

- $\angle(a,b)$ is the number in $[0,\pi]$ that satisfies

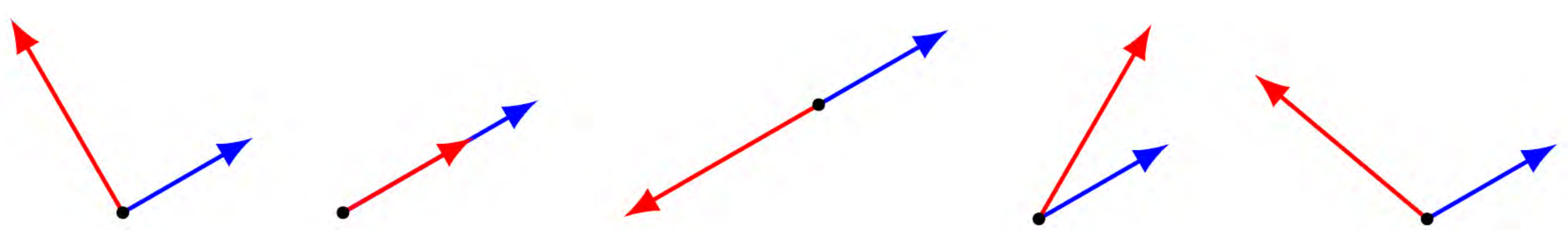
$$a^T b = \|a\|\,\|b\| \cos\left(\angle(a,b)\right)$$

- coincides with ordinary angle between vectors in 2-D and 3-D

Classification of angles

$\theta = \angle(a, b)$

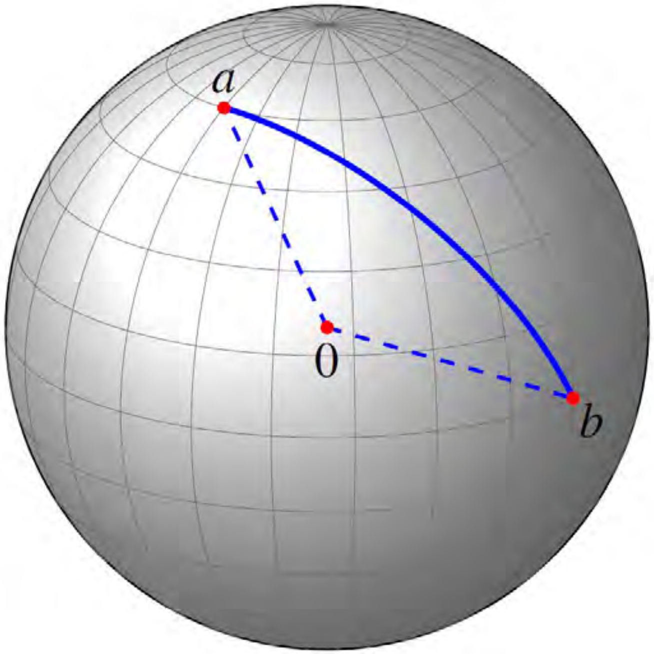
- $\theta = \pi/2 = 90^\circ$: a and b are *orthogonal*, written $a \perp b$ ($a^T b = 0$)
- $\theta = 0$: a and b are *aligned* ($a^T b = \|a\| \|b\|$)
- $\theta = \pi = 180^\circ$: a and b are *anti-aligned* ($a^T b = -\|a\|\,\|b\|$)
- $\theta \le \pi/2 = 90^\circ$: a and b make an *acute angle* ($a^T b \ge 0$)
- $\theta \ge \pi/2 = 90^\circ$: a and b make an *obtuse angle* ($a^T b \le 0$)



Applications

Spherical distance

if a, b are on sphere of radius R, distance *along the sphere* is $R\angle(a,b)$



Applications

Correlation coefficient

- vectors a and b, and de-meaned vectors

$$\tilde{a} = a - \mathbf{avg}(a)\mathbf{1}, \qquad \tilde{b} = b - \mathbf{avg}(b)\mathbf{1}$$

- *correlation coefficient* (between a and b, with $\tilde{a} \neq 0$, $\tilde{b} \neq 0$)

$$\rho = \frac{\tilde{a}^T \tilde{b}}{\|\tilde{a}\| \, \|\tilde{b}\|}$$

- $\rho = \cos \angle(\tilde{a}, \tilde{b})$
  - $\rho = 0$: a and b are *uncorrelated*
  - $\rho > 0.8$ (or so): a and b are *highly correlated*
  - $\rho < -0.8$ (or so): a and b are *highly anti-correlated*

- very roughly: highly correlated means a_i and b_i are typically both above (below) their means together

Applications

Document dissimilarity by angles

- measure dissimilarity by angle of word count histogram vectors
- pairwise angles (in degrees) for 5 Wikipedia pages shown below

	Veterans Day	Memorial Day	Academy Awards	Golden Globe Awards	Super Bowl
Veterans Day	0	60.6	85.7	87.0	87.7
Memorial Day	60.6	0	85.6	87.5	87.5
Academy A.	85.7	85.6	0	58.7	85.7
Golden Globe A.	87.0	87.5	58.7	0	86.0
Super Bowl	87.7	87.5	86.1	86.0	0

Complexity

- Norm: 2n flops. O(n)
- RMS: 2n flops. O(n)
- Distance: 3n flops. O(n)
- Angle: 6n flops. O(n)
- Standard deviation: 4n flops. O(n) can reduce to 3n flops $\mathbf{std}(x)^2 = \mathbf{rms}(x)^2 - \mathbf{avg}(x)^2,$
- Standardizing: 5n flops. O(n)
- Correlation coefficient: 10n flops. O(n)

Reference

- Linear Algebra and Its Applications David C. Lay
- Introduction to Applied Linear Algebra Vectors, Matrices, and Least Squares
- https://www.youtube.com/watch?v=76B5cMEZA4Y