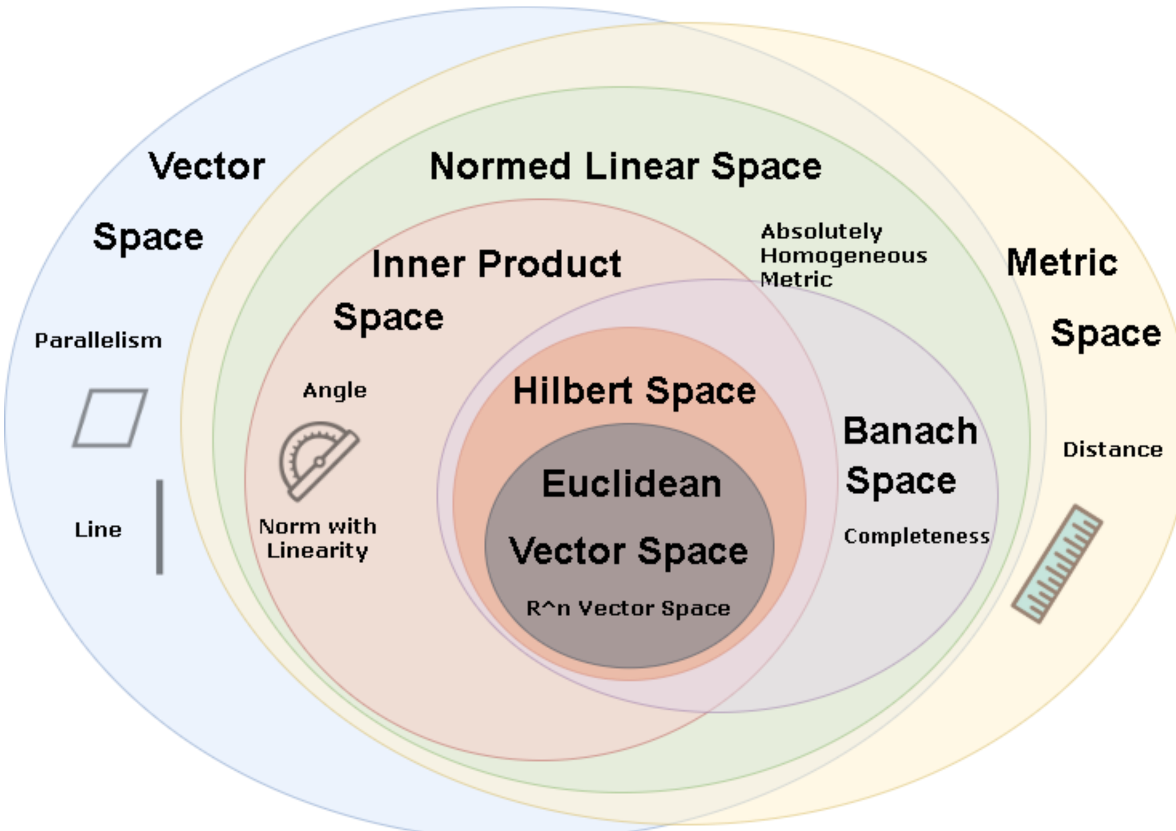
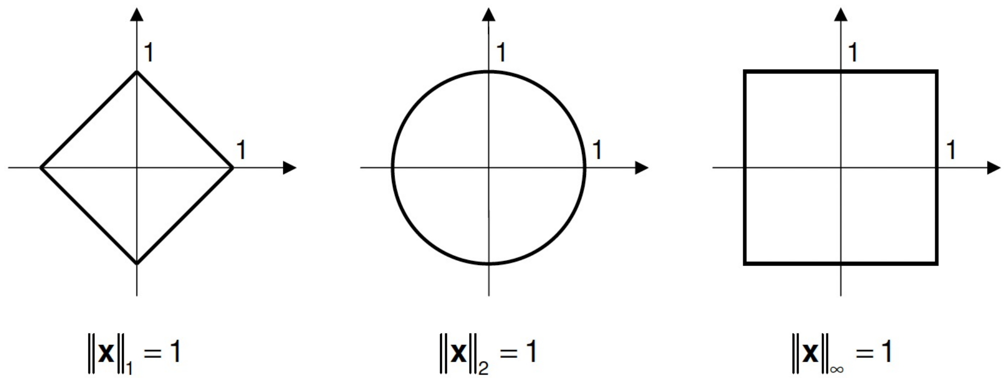
MATH 5110 – Applied Linear Algebra and Matrix Analysis

More examples on inner product, Norm and Metric



l_p − norms



Example

Find the l_1 , l_2 , l_5 , l_∞ −norms of the vector $\vec{x} = \begin{bmatrix} 3 \\ -5 \\ 4 \\ -1 \end{bmatrix}$

Solutions:

$$\|\mathbf{x}\|_1 = |3| + |-5| + |4| + |-1| = 13$$

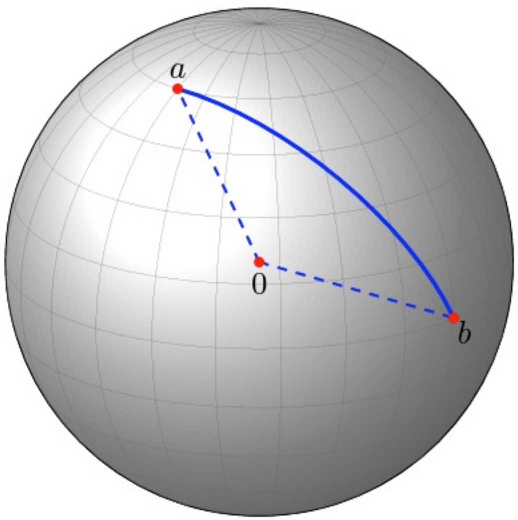
$$\|\mathbf{x}\|_2 = \left[3^2 + (-5)^2 + 4^2 + (-1)^2\right]^{1/2} = \sqrt{51} = 7.14$$

$$\|\mathbf{x}\|_5 = \left[|3|^5 + |-5|^5 + |4|^5 + |-1|^5\right]^{1/5} = \sqrt[5]{4393} = 5.35$$

$$\|\mathbf{x}\|_\infty = \max\left(|3|,\ |-5|,\ |4|,\ |-1|\right) = 5$$

Spherical distance. (Metric on a set)

Suppose $\vec{u}$ and $\vec{v}$ are 3-vectors that represent two points that lie on a sphere of radius R (for example, locations on earth). The spherical distance between them, measured along the sphere



Two points a and b on a **sphere** with radius R and center at the origin. The **spherical distance** between the points is equal to $R\angle(a, b)$

Example.

Distance from Boston to Beijing. The surface of the earth is reasonably approximated as a sphere with radius R = 6367.5km. A location on the earth's surface is traditionally given by its latitude θ and its longitude λ, which correspond to angular distance from the equator and prime meridian, respectively. The 3-D coordinates of the location are given by

$$\begin{bmatrix} R\sin\lambda\cos\theta \\ R\cos\lambda\cos\theta \\ R\sin\theta \end{bmatrix}.$$

(In this coordinate system (0, 0, 0) is the center of the earth, R(0, 0, 1) is the North pole, and R(0, 1, 0) is the point on the equator on the prime meridian, due south of the Royal Observatory outside London.)

The distance through the earth between two locations (3-vectors) a and b is $||a - b||$. The distance along the surface of the earth between points a and b is $R\angle(a, b)$. Find these two distances between **Boston** and **Beijing**, with latitudes and longitudes given below.

Beijing: Latitude $\theta = 39.914°$ and Longitude $\lambda = 116.392°$
Boston: Latitude $\theta = 42.3601°$ and Longitude $\lambda = -71.0589°$

Document similarity via angles.

If n-vectors $\vec{x}$ and $\vec{y}$ represent the **word counts** for two documents, their angle $\angle(\vec{x}, \vec{y})$ can be used as a measure of document dissimilarity.

	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

Pairwise angles (in degrees) between word histograms of five Wikipedia articles.

Some distances examples:

- **Minkowski** distance. ($l_p - norm$)

 (*Manhattan Distance, Euclidean Distance, Chebychev Distance*

- **Cosine Distance**

- ***Mahalanobis distance:*** *a measure of the distance between a point P and a distribution. The idea of measuring is, how many standard deviations away P is from the mean $\vec{\mu}$ of the distribution. (*where, S is the covariance metrics. *)*

$$D_M(\vec{x}) = \sqrt{(\vec{x} - \vec{\mu})^T S^{-1} (\vec{x} - \vec{\mu})}.$$

Application of **distance/metric** of two vectors:

- Feature distance
- Prediction error
- Nearest neighbor
- Document dissimilarity
- Clustering

Standard deviation of a data vector as norm

$$std(\vec{v}) := \frac{\left\| \vec{v} - \overline{\vec{v}}\,\vec{1} \right\|}{\sqrt{n}}$$

Here, $\overline{\vec{v}} = \dfrac{v_1 + \cdots + v_n}{n}$

$$\vec{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

Standardization

$$\frac{\vec{v} - \overline{\vec{v}}\,\vec{1}}{std(\vec{v})}$$

Correlation coefficient

$$\rho = coor\left(\vec{a}\,, \vec{b}\right) = \cos\angle(\vec{a}\,, \vec{b})$$

Standard deviation of Sum:

$$std\left(\vec{a} + \vec{b}\right) = \sqrt{std(\vec{a})^2 + 2\rho std(\vec{a}) std\left(\vec{b}\right) + std\left(\vec{b}\right)^2}$$

Hedging investments.

Suppose that vectors $\vec{a}$ and $\vec{b}$ are time series of returns for two assets with the same **return** (average) μ and **risk** (standard deviation) σ, and **correlation coefficient** $\rho = coor(\vec{a}, \vec{b}) = \cos\angle(\vec{a}, \vec{b})$. (These are the traditional symbols used.) The vector

$$\vec{c} = \frac{\vec{a} + \vec{b}}{2}$$

is the time series of returns for an investment with 50% in each of the assets. This blended investment has the **same return** as the original assets, since

$$avg(\vec{c}) = avg\left(\frac{\vec{a} + \vec{b}}{2}\right) = \frac{\left(avg(\vec{a}) + avg(\vec{b})\right)}{2} = \mu.$$

The **risk** (standard deviation) of this **blended investment** is

$$std(\vec{c}) = \frac{\sqrt{2\sigma^2 + 2\rho\sigma^2}}{2} = \sigma\sqrt{\frac{1+\rho}{2}}$$

From this we see that the **risk** of the blended investment is never more than the risk of the original assets, and is smaller when the correlation of the original asset returns is smaller.

When the returns are uncorrelated, the risk is a factor $\frac{1}{\sqrt{2}} = 0.707$ smaller than the risk of the original assets.

If the asset returns are strongly negatively correlated (i.e., ρ is near -1), the risk of the blended investment is much smaller than the risk of the original assets.

Investing in two assets with uncorrelated, or negatively correlated, returns is called **hedging** (which is short for ‘**hedging your bets**’).

Hedging reduces risk.

Matrix norms

Given an $m \times n$ matrix $A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$, how to define a norm of A?

For example,

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

1. Frobenius norm:

$$\|A\|_{Fro} := \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2}$$

We could place all the entries of the matrix into a single vector, then apply the vector l_p-norm $\|A\|_p$ introduced above.

2. Induced matrix l_p norm.

An $\mathrm{m}\times\mathrm{n}$ matrix A can be viewed as a linear operator that maps a vector $\vec{x}$ in $\mathbb{R}^n$ to a vector $A\vec{x}$ *in* $\mathbb{R}^m$

If we take the norm of the 'output' vector $\mathrm{A}\vec{\mathrm{x}}$ and divide it by the norm of the 'input' vector x, we should get some clue as to the 'magnifying power' of the matrix A (and thus its inherent 'size' or 'magnitude').

We are therefore motivated to calculate the ratio

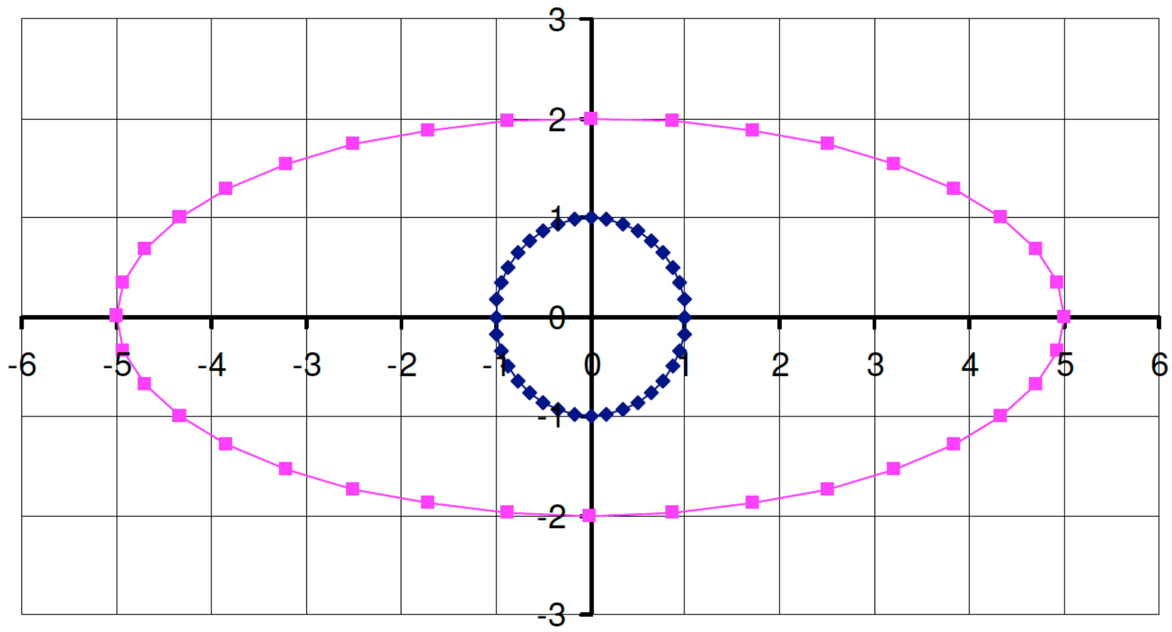
$$\frac{||A\vec{x}||_p}{||\vec{x}||_p}$$

The trouble is that, this result is not unique for different $\vec{x}$.

For example,

$$A = \begin{bmatrix} 5 & 0 \\ 0 & 2 \end{bmatrix}$$

Magnifying power



This can be clarified by looking at the image of the unit circle under the action of A. It is mapped to an ellipse with semi-axes of length 5 and 2:

Given that a general matrix A is likely to exert widely varying **'magnifying powers'** on different input vectors $\vec{x}$, we had better be conservative and define the **matrix p-norm of *A*** as

$$\|A\|_p := \max_{\vec{x}\neq\vec{0}} \frac{\|A\vec{x}\|_p}{\|\vec{x}\|_p}$$

Equivalently,

$$\|A\|_p := \max_{\|\vec{x}\|_p=1} \|A\vec{x}\|_p$$

With this definition, the matrix p-norm provides an upper bound on the 'magnifying power' of A (as measured by applying the vector p-norm to the input vector x and output vector Ax). It means that we are assured of satisfying the following important inequality for any vector x, regardless of its direction.

Theorem:

Let **A** be a matrix with *n* columns, and let **x** be a vector in $\mathbb{R}^n$. If the inequality

$$\|\mathbf{Ax}\| \leq \|\mathbf{A}\| \|\mathbf{x}\|$$

is satisfied for all vectors **x** in $\mathbb{R}^n$, then the matrix norm $\|\mathbf{A}\|$ is said to be compatible with, or induced from, the vector norm used as a metric for **x** and **Ax**.

For $p = 1$, the compatible matrix norm is

$$\|A\|_1 := \max_j \left(\sum_{i=1}^{m} |a_{ij}| \right)$$

which is the *maximum absolute* ***column*** *sum* of A.

it would be nice to verify that the inequality

$$\|A\vec{x}\|_1 \leq \|A\|_1 \|\vec{x}\|_1$$

$$\begin{aligned}
\|\mathbf{Ax}\|_1 &= \sum_i \left| \sum_j A_{ij} x_j \right| \\
&\leq \sum_i \sum_j \left| A_{ij} x_j \right| \\
&= \sum_i \sum_j \left| A_{ij} \right| \left| x_j \right| \\
&= \sum_j \sum_i \left| A_{ij} \right| \left| x_j \right|
\end{aligned}$$

$$\begin{aligned}
&= \sum_j \left| x_j \right| \sum_i \left| A_{ij} \right| \\
&\leq \sum_j \left| x_j \right| \cdot \max_j \left(\sum_i \left| A_{ij} \right| \right) \\
&= \|\mathbf{x}\|_1 \|\mathbf{A}\|_1 \\
&= \|\mathbf{A}\|_1 \|\mathbf{x}\|_1
\end{aligned}$$

For $p = \infty$, the compatible matrix norm is

$$\|A\|_\infty := \max_i \left(\sum_{j=1}^{m} |a_{ij}| \right)$$

which is the maximum absolute **row** sum of A.

Verify that the inequality $\|A\vec{x}\|_\infty \leq \|A\|_\infty \|\vec{x}\|_\infty$

$$\begin{aligned}
\|\mathbf{Ax}\|_\infty &= \max_i \left(\left| \sum_j A_{ij} x_j \right| \right) \\
&\leq \max_i \left(\sum_j \left| A_{ij} x_j \right| \right) \quad * \\
&= \max_i \left(\sum_j \left| A_{ij} \right| \left| x_j \right| \right) \\
&\leq \max_i \left(\sum_j \left| A_{ij} \right| \right) \cdot \max_j \left(\left| x_j \right| \right) \\
&= \|\mathbf{A}\|_\infty \|\mathbf{x}\|_\infty
\end{aligned}$$

For $p = 2$, the matrix 2-norm is widely used, but can be difficult to calculate.

If A happens to be square and symmetric, the 2-norm is the **supremum** of the absolute values of the **eigenvalues**:

$$\|\mathbf{A}\|_2 = \max_i \left(|\lambda_i|\right)$$

More generally, the 2-norm of A turns out to be its largest **singular value.**

Example:

Find the 1-norm, ∞-norm and Frobenius norm of the matrix

$$\mathbf{A} = \begin{bmatrix} 3 & -2 & 6 \\ 5 & -8 & 4 \\ -4 & 1 & -3 \end{bmatrix}$$

Solution

The 1-norm is the maximum absolute <u>column</u> sum: $\|\mathbf{A}\|_1 = \max(12,\ 11,\ 13) = 13$

The ∞-norm is the maximum absolute <u>row</u> sum: $\|\mathbf{A}\|_\infty = \max(11,\ 17,\ 8) = 17$

The Frobenius norm is the square root of the sum of squares:

$$\|\mathbf{A}\|_{\mathrm{Fro}} = \left[3^2 + (-2)^2 + \ldots + (-3)^2\right]^{1/2} = \sqrt{180} = 13.42$$

Example:

Find the 2-norm of the matrix

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

Solution:

The matrix is square and symmetric, so its 2-norm can be calculated from the eigenvalues.

$$\det \begin{bmatrix} 1-\lambda & 2 \\ 2 & -3-\lambda \end{bmatrix} = 0$$

$$(1-\lambda)(-3-\lambda) - 4 = 0$$

$$\lambda^2 + 2\lambda - 7 = 0$$

Solving the quadratic gives $\lambda = -1 \pm \sqrt{8} = 1.83, \ -3.83$. Hence

$$\|\mathbf{A}\|_2 = \max(|1.83|, \ |-3.83|) = 3.83$$

Example:

Given the matrix and vector

$$\mathbf{A} = \begin{bmatrix} 1.8 & 1.4 \\ -0.1 & 0.6 \end{bmatrix}, \qquad \mathbf{x} = \begin{bmatrix} 0.3 \\ 0.2 \end{bmatrix}$$

investigate the satisfaction of the ‘compatibility inequality’

$$\|\mathbf{Ax}\|_p \le \|\mathbf{A}\|_q \|\mathbf{x}\|_p$$

when p and q take values 1 and ∞ in all combinations.

Solution:

$$\mathbf{Ax} = \begin{bmatrix} 1.8 & 1.4 \\ -0.1 & 0.6 \end{bmatrix} \begin{bmatrix} 0.3 \\ 0.2 \end{bmatrix} = \begin{bmatrix} 0.82 \\ 0.09 \end{bmatrix}$$

We therefore have

$$\|\mathbf{x}\|_1 = 0.5, \ \|\mathbf{x}\|_\infty = 0.3 \qquad \text{(vector norms of } \mathbf{x}\text{)}$$

$$\|\mathbf{A}\|_1 = 2.0, \ \|\mathbf{A}\|_\infty = 3.2 \qquad \text{(matrix norms of } \mathbf{A}\text{)}$$

$$\|\mathbf{Ax}\|_1 = 0.91, \ \|\mathbf{Ax}\|_\infty = 0.82 \qquad \text{(vector norms of } \mathbf{Ax}\text{)}$$

Combination	$\lVert \mathbf{Ax} \rVert_p \leq \lVert \mathbf{A} \rVert_q \lVert \mathbf{x} \rVert_p$	Inequality satisfied?
$p = 1,\ q = 1$	$0.91 \leq 2.0 \times 0.5$	YES (expected)
$p = \infty,\ q = \infty$	$0.82 \leq 3.2 \times 0.3$	YES (expected)
$p = 1,\ q = \infty$	$0.91 \leq 3.2 \times 0.5$	YES (lucky)
$p = \infty,\ q = 1$	$0.82 \leq 2.0 \times 0.3$	NO

Don’t mix and match incompatible matrix and vector norms

Reference:

- Introduction to Applied Linear Algebra Vectors, Matrices, and Least Squares

- https://en.wikipedia.org/wiki/Great-circle_distance
- https://en.wikipedia.org/wiki/Geodesics_on_an_ellipsoid