CENTER FOR
Brains
Minds+
Machines

Tutorial on Linear Algebra

Brains, Minds & Machines Summer School 2018

Andrzej Banburski
(based on slides of Xavier Boix, in turn based on those of Joe Olson)

Tutorial Outline

The goal is to review the parts of linear algebra necessary to understand Principal Component Analysis (PCA)

Linear Algebra

1. Matrices, vectors and products

Linear Algebra

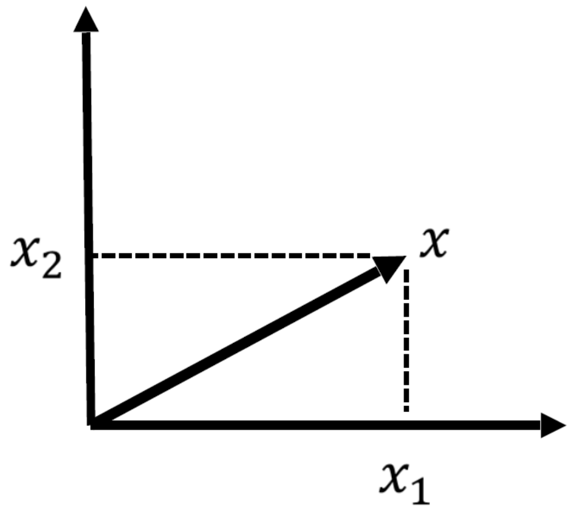
Vectors

$x = (x_1, x_2, \dots, x_n)$ and $y = (y_1, y_2, \dots, y_m)$

Example:

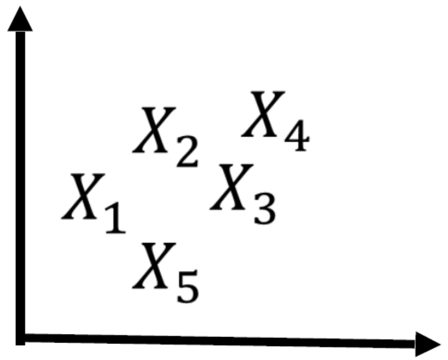
let $n = 2$ and $m = 2$

$x = (x_1, x_2)$ and $y = (y_1, y_2)$



Linear Algebra

Matrices



$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1T} \\ x_{21} & x_{22} & \cdots & x_{2T} \\ \vdots & \vdots & \cdots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NT} \end{pmatrix} = \begin{pmatrix} X_1 & X_2 & \cdots & X_T \end{pmatrix}$$

Each X_i is N-dimensional data point (vector) from trial i

Linear Algebra

Transpose

Transpose of a matrix swaps rows with columns

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \longrightarrow A^T = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$$

$$A = \begin{pmatrix} 4 & 7 \\ -2 & 1 \end{pmatrix} \longrightarrow A^T = \begin{pmatrix} 4 & -2 \\ 7 & 1 \end{pmatrix}$$

Linear Algebra

Symmetric Matrices

A matrix is symmetric if it is equal to its transpose

$A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$ is symmetric because $A^T = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$

$$\begin{pmatrix} -3 & 7 \\ 7 & 9 \end{pmatrix} \qquad \begin{pmatrix} 1 & 9 \\ 9 & 0 \end{pmatrix} \qquad \begin{pmatrix} c & m \\ m & b \end{pmatrix}$$

Linear Algebra

Matrix Multiplication

matrix*matrix $\quad W = AB$

$$\begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{pmatrix}$$

W has size (2,2)
A has size (2,2)
B has size (2,2)

In order to perform multiplication, the sizes need to match up accordingly
W has size (m, n)
A has size (m, p)
B has size (p, n)

$$(m, n) = (m, \cancel{p}) \times (\cancel{p}, n)$$

Does AB = BA ?

Linear Algebra

Matrix Multiplication

matrix*vector $y = Ax$

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \\ a_{31}x_1 + a_{32}x_2 \end{pmatrix}$$

y has size $(3{,}1)$
A has size $(3{,}2)$
x has size $(2{,}1)$

In order to perform multiplication, the sizes need to match up accordingly
y has size $(m, 1)$
A has size (m, n)
x has size $(n, 1)$

$$(m, 1) = (m, \not{n}) \times (\not{n}, 1)$$

Linear Algebra

Matrix Multiplication

vector*vector

$$x^T = (x_1 \quad x_2 \quad \cdots \quad x_n)$$

$$y^T = (y_1 \quad y_2 \quad \cdots \quad y_n)$$

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \qquad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

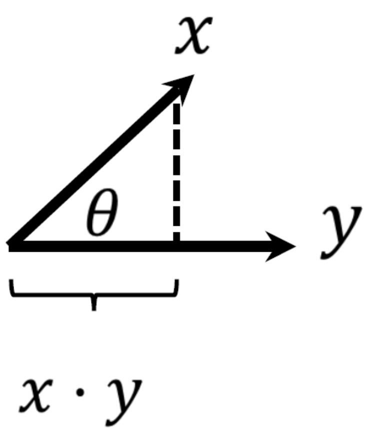
$$x \cdot y = \sum_{i=1}^{n} x_i y_i = x^T y = y^T x$$

Matrix Multiplication

vector*vector

A dot product gives the “overlap” of two vectors.

It is a *number* not a vector.

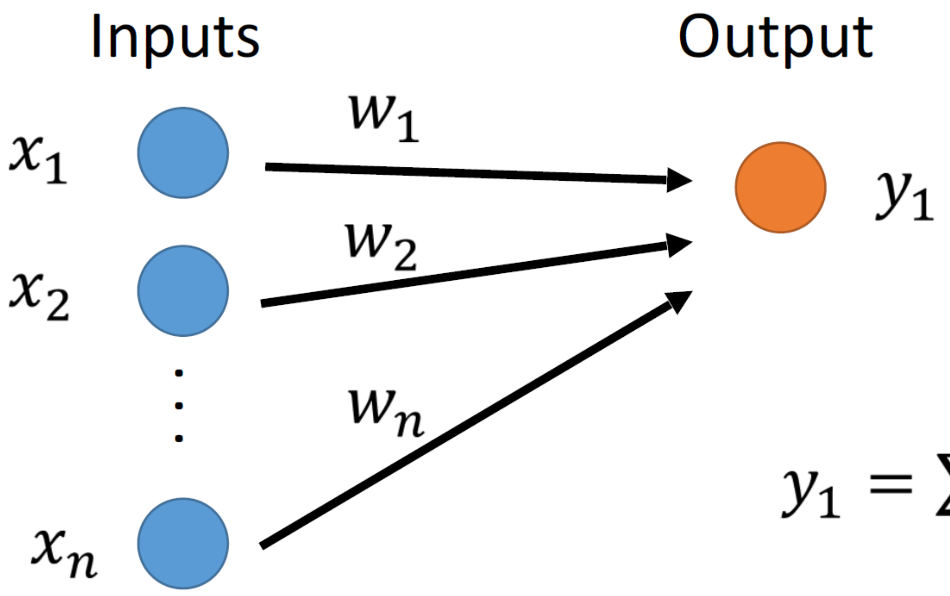


$$x \cdot y = \sum_{i=1}^{n} x_i y_i = |x||y|\cos(\theta)$$

If x and y are perpendicular (orthogonal) then $\theta = 90°$ and $\cos(\theta) = 0$.

Then $x \cdot y = 0$

Example: Linear Neuron Model

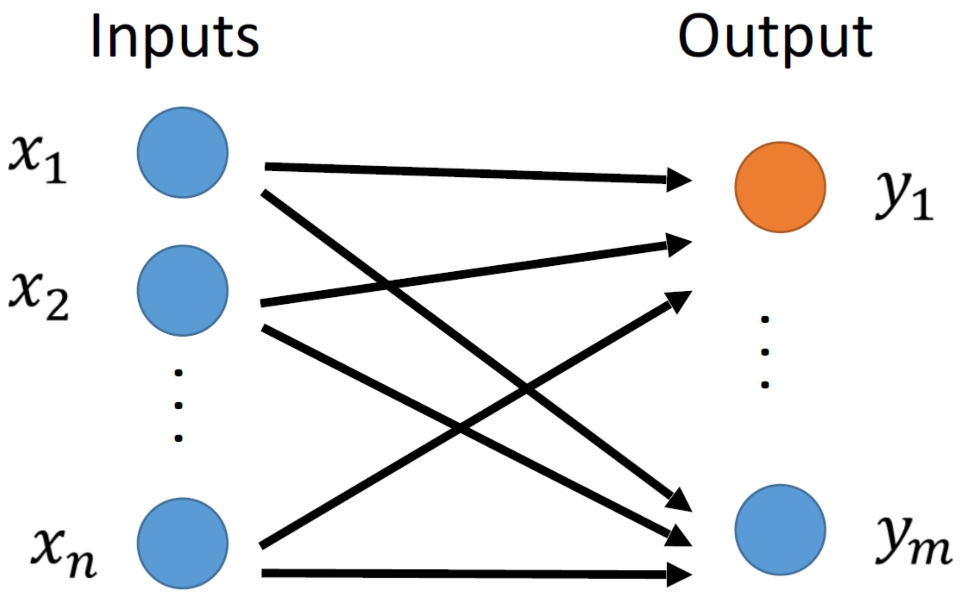


$$y_1 = \sum_{j=1}^{n} w_j x_j = \begin{pmatrix} w_1 & \cdots & w_n \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

$$= Wx$$

Linear Algebra

Example: Linear Neuron Model



$$y_i = \sum_{j=1}^{n} w_{ij} x_j$$

$$y = Wx$$

w_{ij} = weight from x_j to y_i

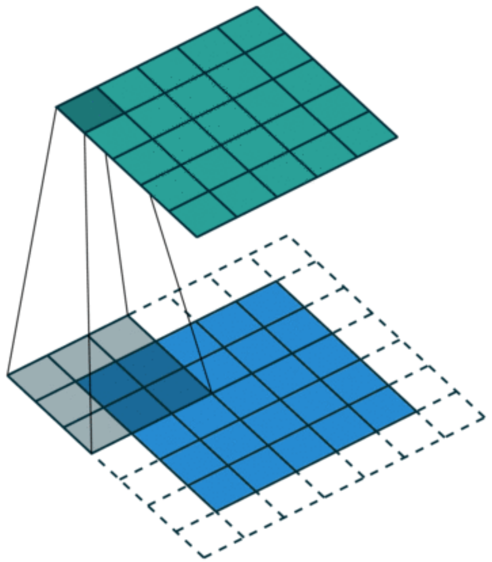
Convolution as Toeplitz matrix

For experts: even a convolution operation can be recast as matrix multiplication:

$$\begin{pmatrix} x1 & x2 & x3 \\ x4 & x5 & x6 \\ x7 & x8 & x9 \end{pmatrix} * \begin{pmatrix} k1 & k2 \\ k3 & k4 \end{pmatrix}$$

Here is a constructed matrix with a vector:

$$\begin{pmatrix} k1 & k2 & 0 & k3 & k4 & 0 & 0 & 0 & 0 \\ 0 & k1 & k2 & 0 & k3 & k4 & 0 & 0 & 0 \\ 0 & 0 & 0 & k1 & k2 & 0 & k3 & k4 & 0 \\ 0 & 0 & 0 & 0 & k1 & k2 & 0 & k3 & k4 \end{pmatrix} \cdot \begin{pmatrix} x1 \\ x2 \\ x3 \\ x4 \\ x5 \\ x6 \\ x7 \\ x8 \\ x9 \end{pmatrix}$$



which is equal to

$$\begin{pmatrix} k1\,x1 + k2\,x2 + k3\,x4 + k4\,x5 \\ k1\,x2 + k2\,x3 + k3\,x5 + k4\,x6 \\ k1\,x4 + k2\,x5 + k3\,x7 + k4\,x8 \\ k1\,x5 + k2\,x6 + k3\,x8 + k4\,x9 \end{pmatrix}.$$

https://stackoverflow.com/questions/48768555/toeplitz-matrix-with-an-image

Linear Algebra

Norms

A function that measures the “size” of a vector is called a **norm**.

The L^p norm us given by $\|x\|_p = \left(\sum_i |x_i|^p\right)^{\frac{1}{p}}$

More generally, the norm has to satisfy the following:

- $f(x) = 0 \Rightarrow x = 0$
- $f(x+y) \leq f(x) + f(y)$ (triangle inequality
- $\forall \alpha \in \mathbb{R}, f(\alpha x) = |\alpha| f(x)$

Linear Algebra

Norms

The most commonly used norm for vectors is $p = 2$, which is compatible with the inner product $\|x\|_2 = \sqrt{x \cdot x}$

Another useful norm is the L_∞ norm, also known as **max norm**:

$$\|x\|_\infty = \max_i |x_i|$$

The most natural measure of matrix “size” is the **Frobenius norm**:

$$\|A\|_F = \sqrt{\sum_{i,j} A_{ij}^2}$$

Linear Algebra

Trace

The **trace** of a matrix is the sum of its' diagonal entries

$$Tr(A) = \sum_i A_{ii}$$

Some useful properties:

- $Tr(A) = Tr(A^T)$
- $Tr(ABC) = Tr(CAB) = Tr(BCA)$ (if defined)

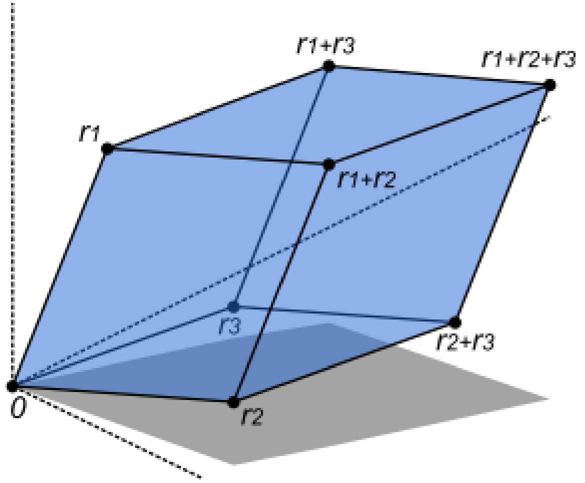
Linear Algebra

Determinant

The **determinant** is a value that can be computed for a square matrix.

For a 2x2 matrix it is given by $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$

Interpretation: volume of parallelepiped is the absolute value of the determinant of a matrix formed of row vectors r1, r2, r3.



In general for an (n,n) matrix it is given by

$$\det(A) = \sum_{\sigma \in S_n} \left(\operatorname{sgn}(\sigma) \prod_{i=1}^{n} A_{i,\sigma(i)} \right)$$

Computed over all permutations σ of the set {1,…,n}.

Linear Algebra

2. Matrices and data transformations

Linear equations

$x = (x_1, x_2, \dots, x_n)$ and $y = (y_1, y_2, \dots, y_m)$

Example:

let $n = 2$ and $m = 2$

$x = (x_1, x_2)$ and $y = (y_1, y_2)$

We have 2 linear equations:

$y_1 = a_{11} * x_1 + a_{12} * x_2$

$y_2 = a_{21} * x_1 + a_{22} * x_2$

$$y_i = \sum_{j=1}^{n} a_{ij} x_j$$

gives m linear equations

Linear Algebra

Linear equations with Matrices

We have 2 linear equations:

$$y_1 = a_{11} * x_1 + a_{12} * x_2$$
$$y_2 = a_{21} * x_1 + a_{22} * x_2$$

We write this as:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$
$$= \begin{pmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{pmatrix}$$

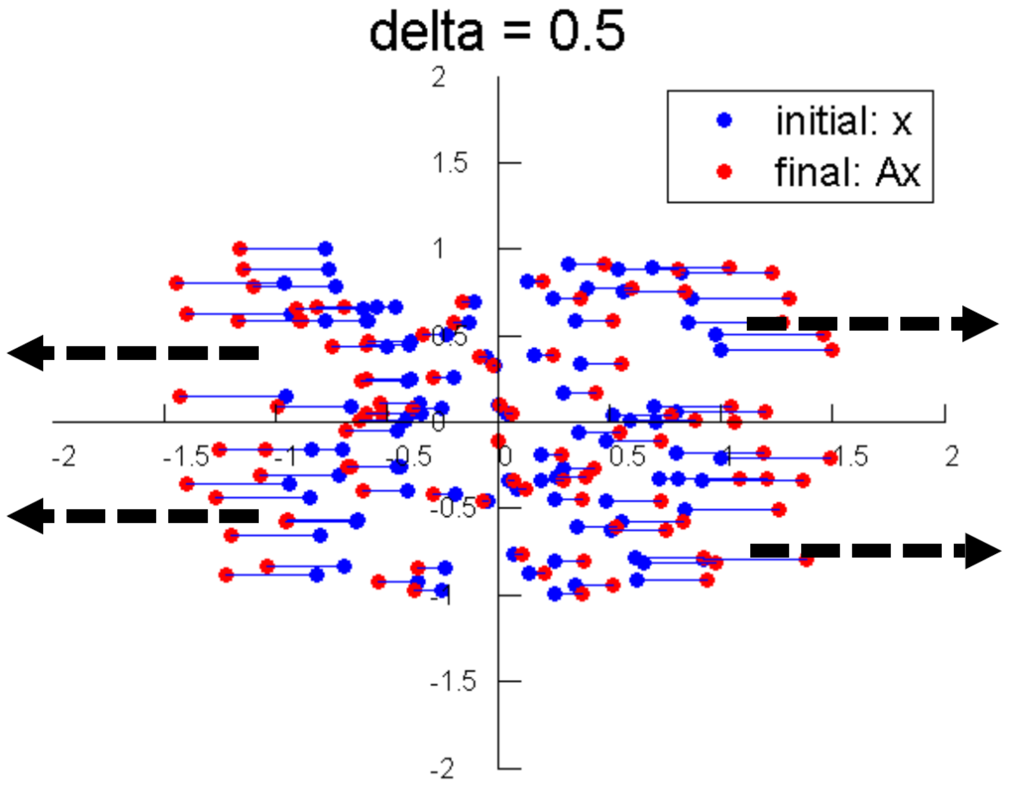
We can think of the matrix A as a function or transformation of x:

$$y = f(x) = Ax$$

Linear Algebra

Matrix Example: Stretch

$$S_1 = \begin{pmatrix} 1+\delta & 0 \\ 0 & 1 \end{pmatrix}$$

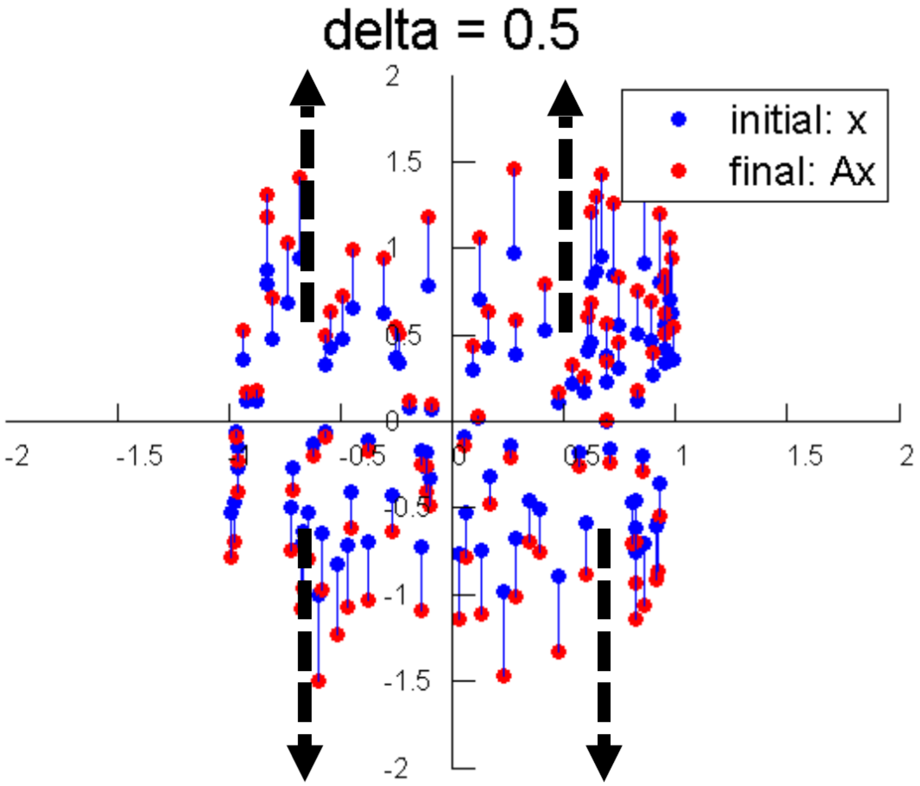


$$S_1 x = \begin{pmatrix} 1+\delta & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= \begin{pmatrix} (1+\delta)x_1 \\ x_2 \end{pmatrix}$$

Matrix Example: Stretch

$$S_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1+\delta \end{pmatrix}$$



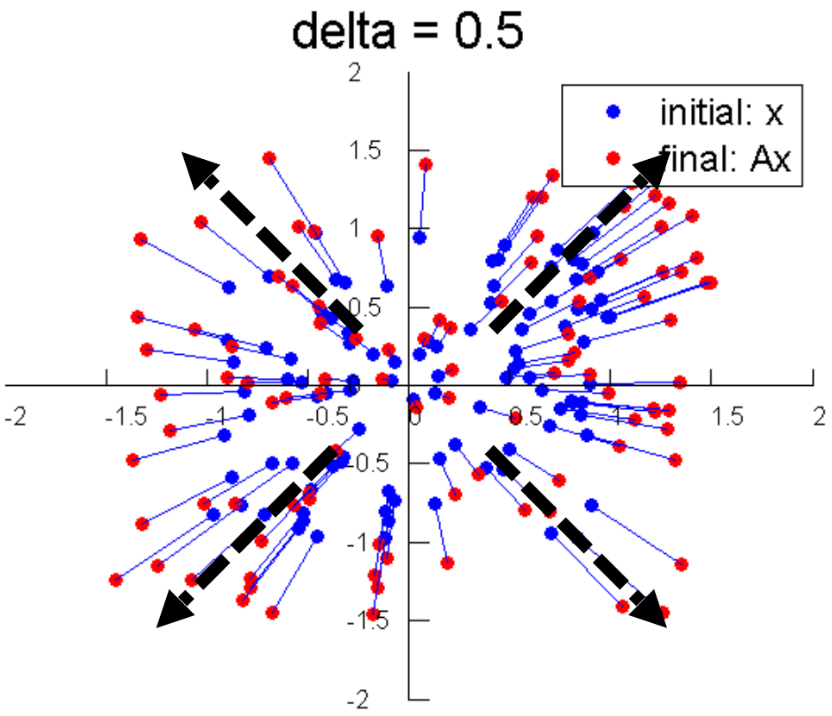
$$S_2 x = \begin{pmatrix} 1 & 0 \\ 0 & 1+\delta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= \begin{pmatrix} x_1 \\ (1+\delta)x_2 \end{pmatrix}$$

Linear Algebra

Matrix Example: Stretch

$$S_{1,2} = \begin{pmatrix} 1+\delta & 0 \\ 0 & 1+\delta \end{pmatrix}$$



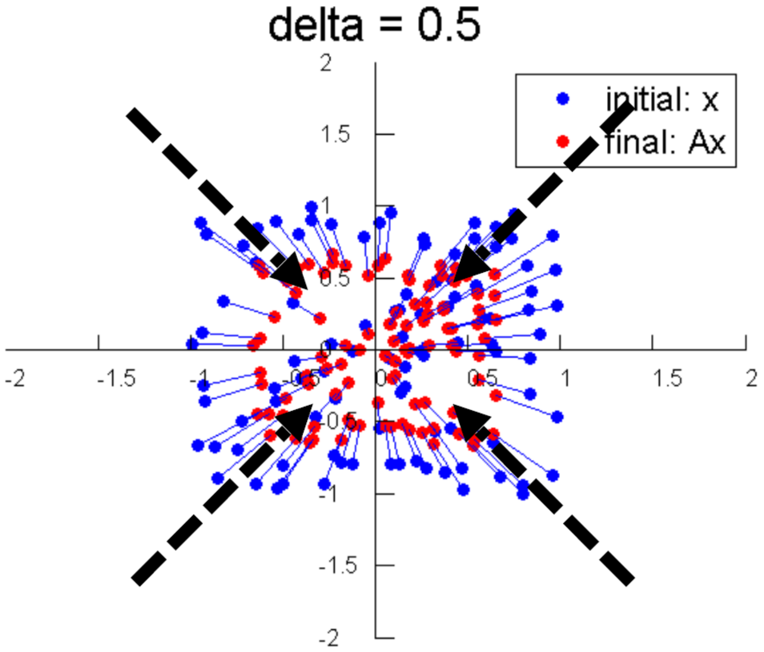
$$S_{1,2}x = \begin{pmatrix} 1+\delta & 0 \\ 0 & 1+\delta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= \begin{pmatrix} (1+\delta)x_1 \\ (1+\delta)x_2 \end{pmatrix}$$

Linear Algebra

Matrix Example: Shrink

$$A = \begin{pmatrix} 1/(1+\delta) & 0 \\ 0 & 1/(1+\delta) \end{pmatrix}$$



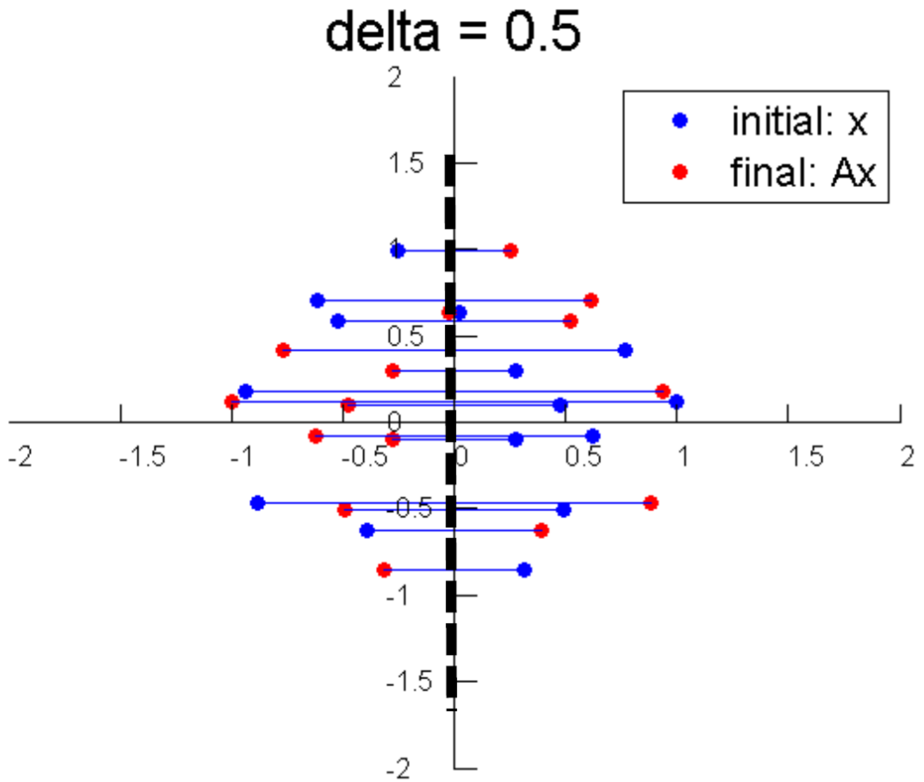
$$Ax = \begin{pmatrix} 1/(1+\delta) & 0 \\ 0 & 1/(1+\delta) \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= \begin{pmatrix} x_1/(1+\delta) \\ x_2/(1+\delta) \end{pmatrix}$$

Linear Algebra

Matrix Example: Reflection

$$R_1 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

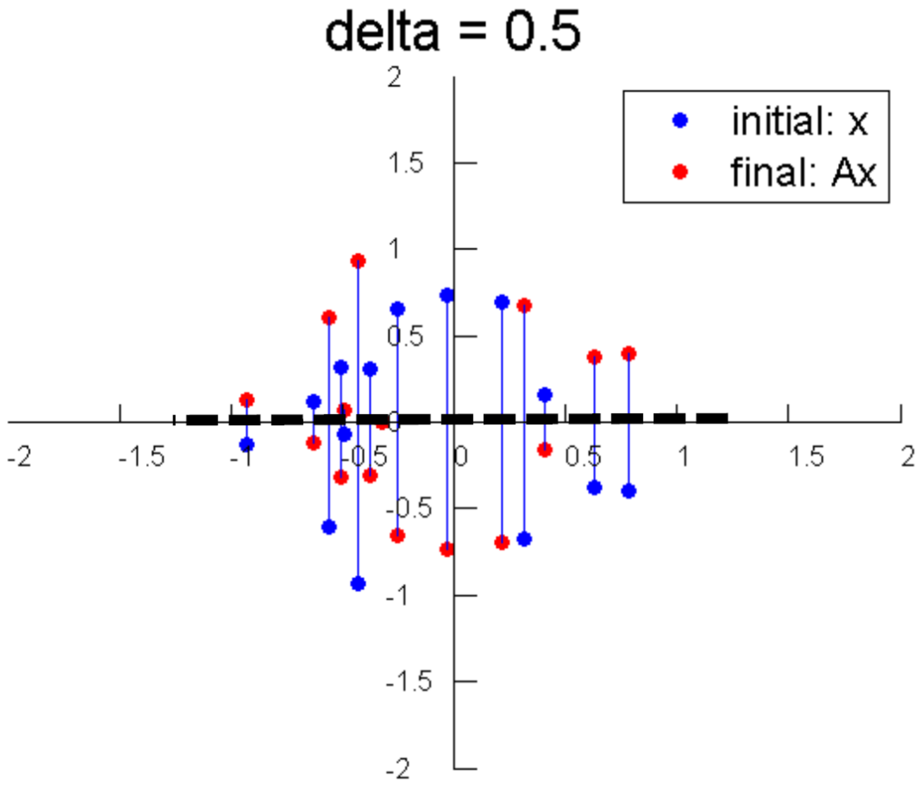


$$R_1 x = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= \begin{pmatrix} -x_1 \\ x_2 \end{pmatrix}$$

Matrix Example: Reflection

$$R_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$



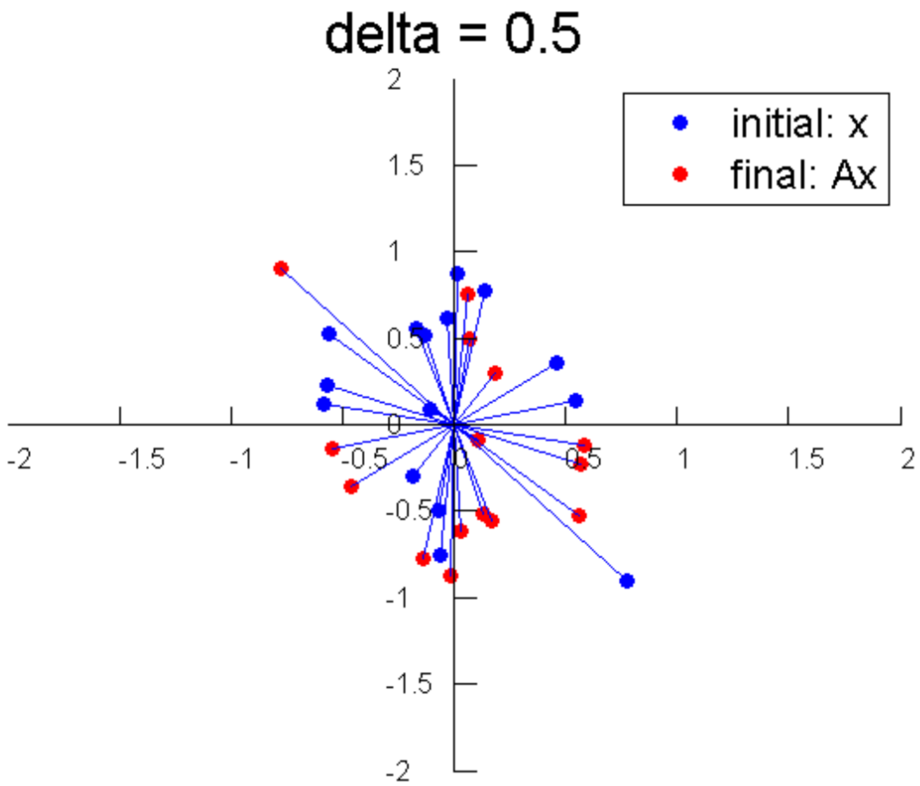
$$R_2 x = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= \begin{pmatrix} x_1 \\ -x_2 \end{pmatrix}$$

Linear Algebra

Matrix Example: Reflection

$$R_{1,2} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$$



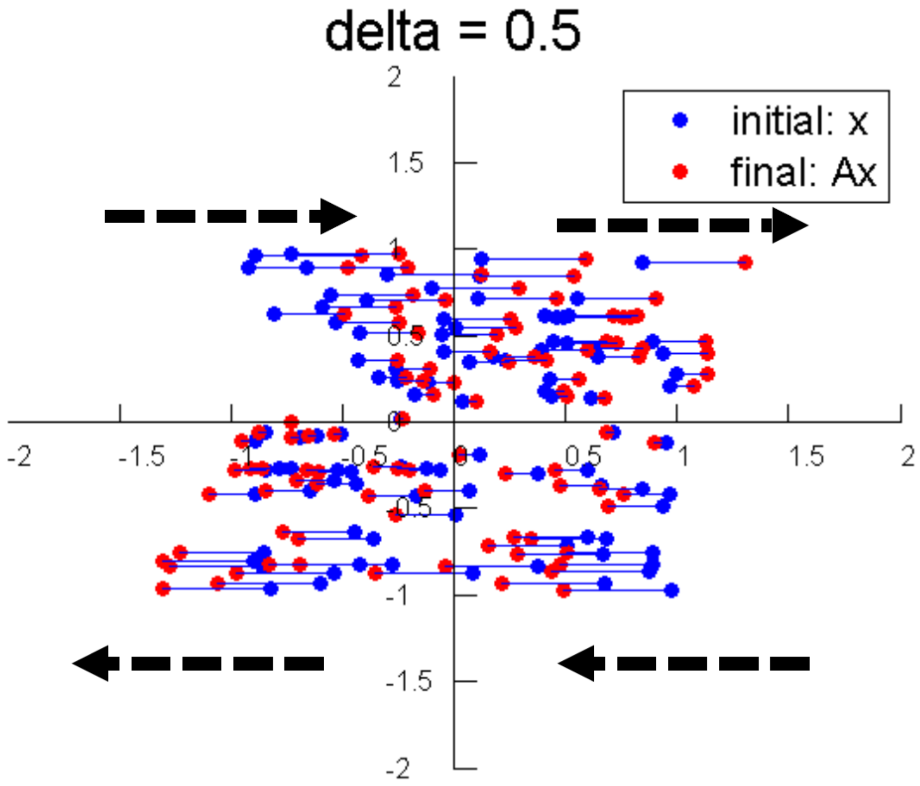
$$R_{1,2}x = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= \begin{pmatrix} -x_1 \\ -x_2 \end{pmatrix}$$

Linear Algebra

Matrix Example: Shear

$$T_1 = \begin{pmatrix} 1 & \delta \\ 0 & 1 \end{pmatrix}$$



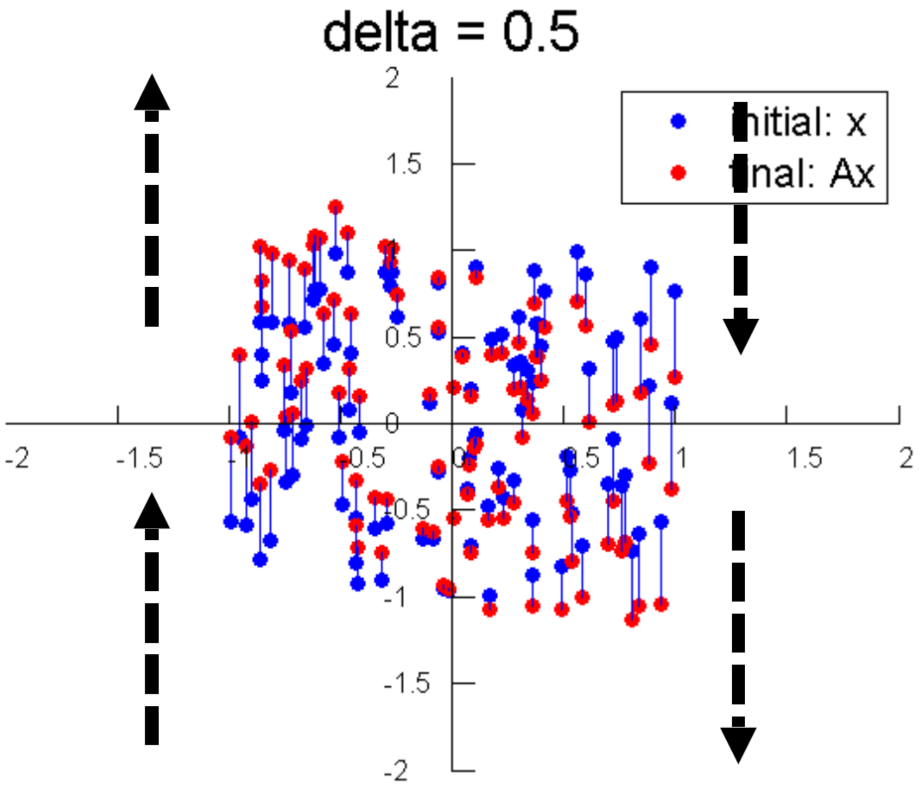
$$T_1 x = \begin{pmatrix} 1 & \delta \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= \begin{pmatrix} x_1 + \delta x_2 \\ x_2 \end{pmatrix}$$

Linear Algebra

Matrix Example: Shear

$$T_2 = \begin{pmatrix} 1 & 0 \\ -\delta & 1 \end{pmatrix}$$



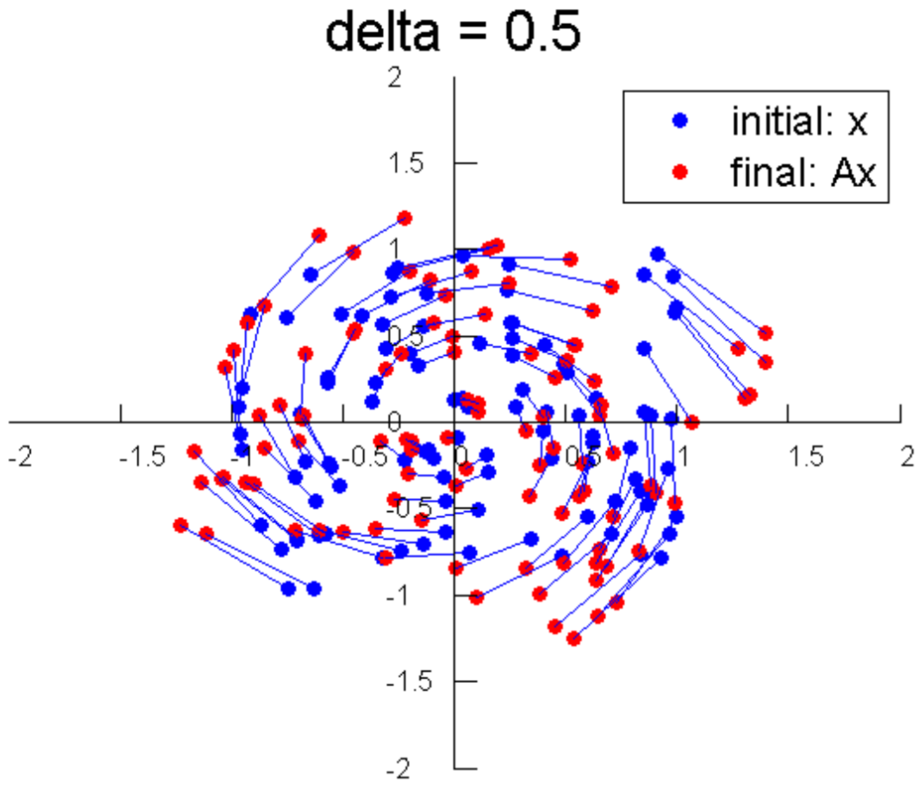
$$T_2 x = \begin{pmatrix} 1 & 0 \\ -\delta & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= \begin{pmatrix} x_1 \\ x_2 - \delta x_1 \end{pmatrix}$$

Linear Algebra

Matrix Example: Shear

$$T_{1,2} = \begin{pmatrix} 1 & \delta \\ -\delta & 1 \end{pmatrix}$$



$$T_{1,2}x = \begin{pmatrix} 1 & \delta \\ -\delta & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= \begin{pmatrix} x_1 + \delta x_2 \\ x_2 - \delta x_1 \end{pmatrix}$$

Linear Algebra

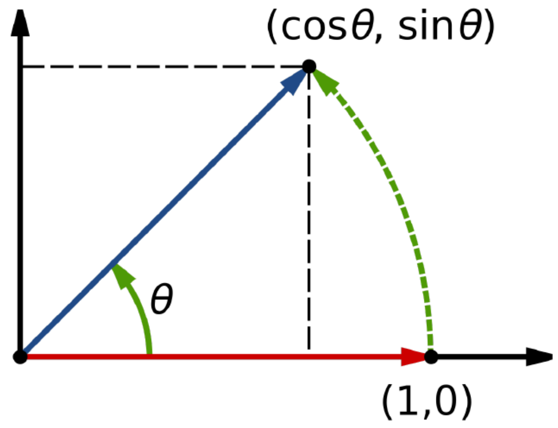
Rotation Matrices

$$R = \begin{pmatrix} 1 & -\theta \\ \theta & 1 \end{pmatrix}$$

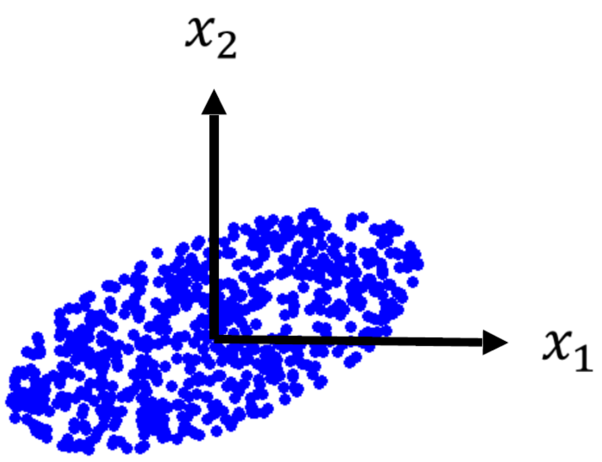
is the small-angle approximation to the true rotation matrix.

$$R = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

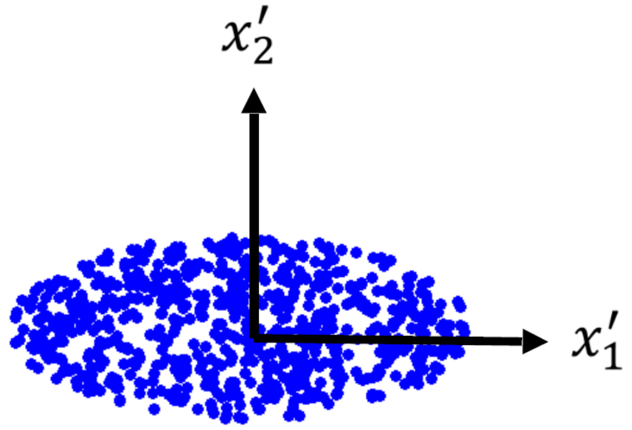
is a counter-clockwise rotation by angle θ.



Rotation Matrices



$$R = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$



Matrix Example: Inverses

The inverse of a function f, denoted by f^{-1}, satisfies:

$$f^{-1}\big(f(x)\big) = x$$

i.e. $f^{-1}(f(\cdot))$ is the identity function.

Similarly, the inverse of a matrix satisfies:

$$A^{-1}Ax = AA^{-1}x = x$$

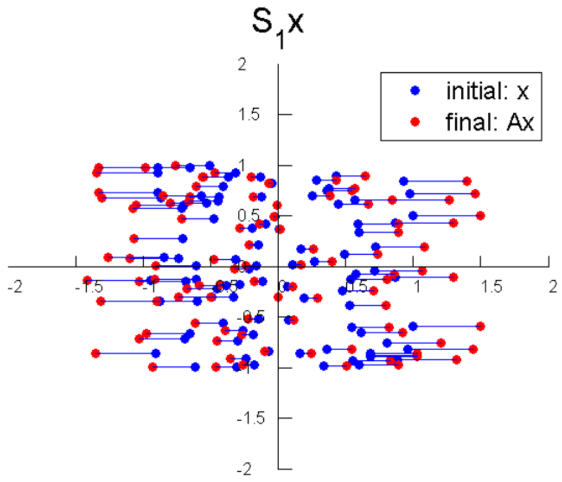
$$A^{-1}A = AA^{-1} = I$$

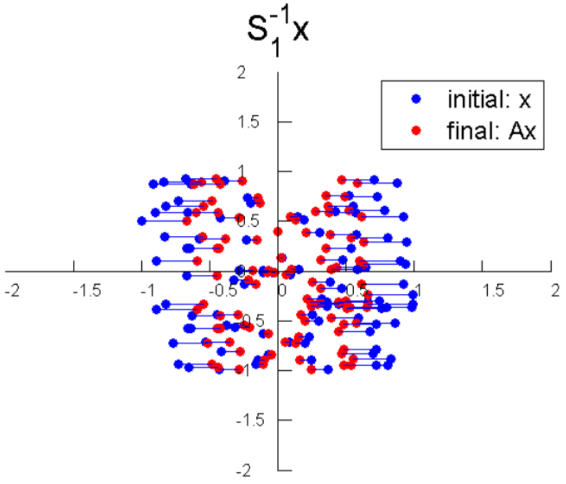
Linear Algebra

Matrix Example: Inverses $A^{-1}A = AA^{-1} = I$

$S_1 = \begin{pmatrix} 1+\delta & 0 \\ 0 & 1 \end{pmatrix}$ has inverse $S_1^{-1} = \begin{pmatrix} 1/(1+\delta) & 0 \\ 0 & 1 \end{pmatrix}$

$$S_1 S_1^{-1} = \begin{pmatrix} 1+\delta & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1/(1+\delta) & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} (1+\delta)/(1+\delta) & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$



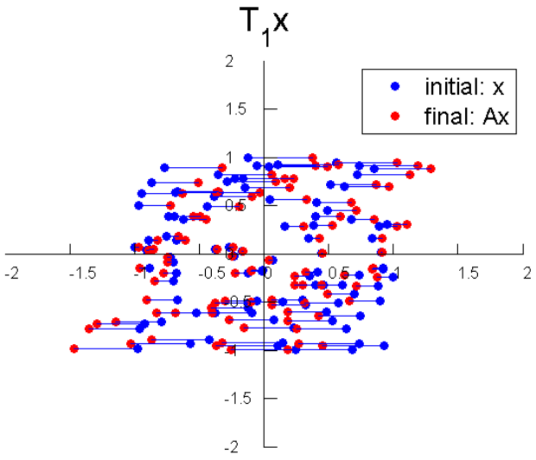


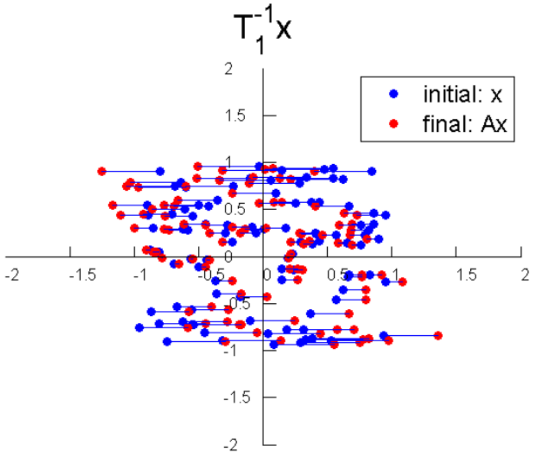
Linear Algebra

Matrix Example: Inverses $A^{-1}A = AA^{-1} = I$

$T_1 = \begin{pmatrix} 1 & \delta \\ 0 & 1 \end{pmatrix}$ has inverse $T_1^{-1} = \begin{pmatrix} 1 & -\delta \\ 0 & 1 \end{pmatrix}$

$$T_1 T_1^{-1} = \begin{pmatrix} 1 & \delta \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -\delta \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -\delta + \delta \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$



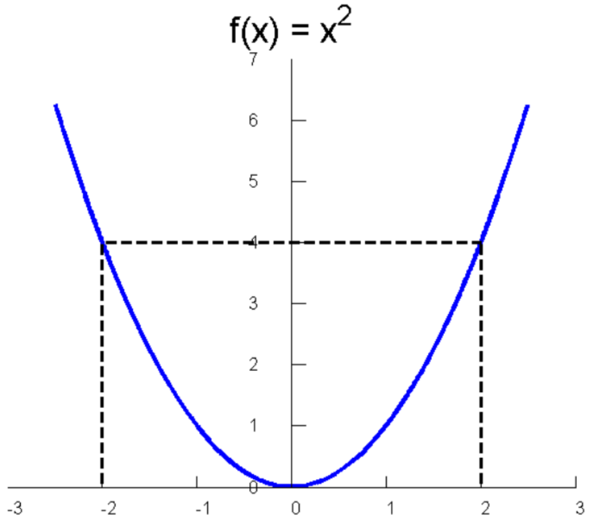


Non-Invertible

A function is non-invertible if taking the inverse would be ambiguous. Mathematically, if there are points x_1, x_2 such that $f(x_1) = f(x_2)$. Because then $f^{-1}\big(f(x_1)\big) = x_1$ or x_2.

Example: $f(x) = x^2$ has an ambiguous inverse because $f^{-1}(4) = 2$ or -2.

Thus $f(x) = x^2$ is non-invertible



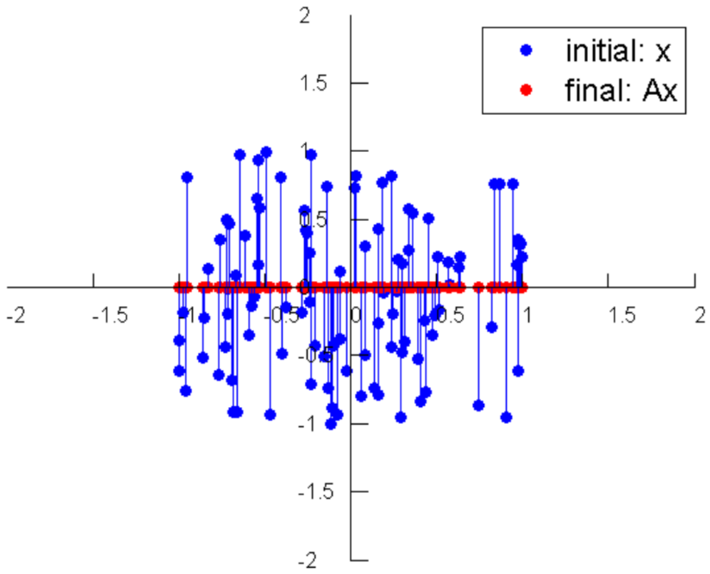
Linear Algebra

Matrix Example: Projection (no inverse)

Projection matrices project all the points to a smaller number of dimensions (dimensionality reduction).

$$P_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

$$P_1 x = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ 0 \end{pmatrix}$$



When is a matrix invertible

In general, for an inverse matrix A^{-1} to exist, A has to be square and its' columns have to form a **linearly independent** set of vectors – no column can be a linear combination of the others.

A necessary and sufficient condition is that $\det(A) \neq 0$.

Finding the inverse is usually quite arduous, even though an explicit expression exists:

$$A^{-1} = \frac{1}{\det(A)} \sum_{s=0}^{n-1} A^s \sum_{k_1, k_2, \ldots, k_n} \prod_{l=1}^{n-1} \frac{(-1)^{k_l+1}}{k_l!\, l^{k_l}} Tr\left(A^l\right)^{k_l}$$

Linear Algebra

Orthogonal Matrices

The matrix Q is an orthogonal matrix if: $Q^T Q = I$

The dot product across different column-vectors of Q are 0, ie. $q_i^T q_j = 0, i \neq j$

Or equivalently, the matrix Q is an orthogonal matrix if: $Q^T = Q^{-1}$

To show this, recall that $Q^{-1} Q = I$

Linear Algebra

Orthogonal Matrices

A square matrix U is *orthogonal* if $U^{-1} = U^T$

What matrices are orthogonal?

Rotation:

$$R = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

$$R^{-1} = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} = R^T$$

Reflection:

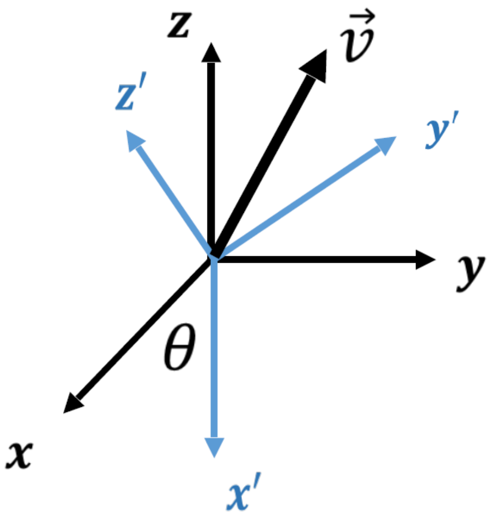
$$\mathrm{M} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$M^{-1} = M^T = M$$

Linear Algebra

Orthogonal Matrices

Orthogonal matrix changes the coordinate system



$$U' = \begin{pmatrix} \overrightarrow{x'}^T \\ \overrightarrow{y'}^T \\ \overrightarrow{z'}^T \end{pmatrix} \qquad \vec{v} = v_1\vec{x} + v_2\vec{y} + v_3\vec{z} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}_{B1}$$

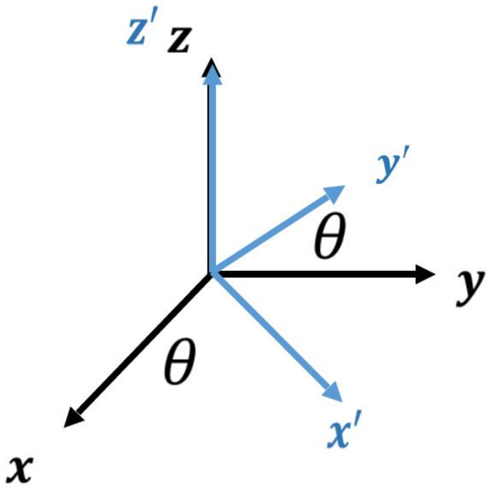
$$U'\vec{v} = \begin{pmatrix} \overrightarrow{x'} \cdot \vec{v} \\ \overrightarrow{y'} \cdot \vec{v} \\ \overrightarrow{z'} \cdot \vec{v} \end{pmatrix} = \begin{pmatrix} v_1' \\ v_2' \\ v_3' \end{pmatrix}_{B2} = v_1'\overrightarrow{x'} + v_2'\overrightarrow{y'} + v_3'\overrightarrow{z'}$$

Projection of v onto new coordinate system

Linear Algebra

Orthogonal Matrices

The rows and columns of orthogonal matrices form an **ortho-normal basis**



$$\vec{x} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \vec{y} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \vec{z} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

$$\overrightarrow{x'} = \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \\ 0 \end{pmatrix} \quad \overrightarrow{y'} = \begin{pmatrix} -\sin(\theta) \\ \cos(\theta) \\ 0 \end{pmatrix} \quad \overrightarrow{z'} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Define $U' = \begin{pmatrix} \overrightarrow{x'}^T \\ \overrightarrow{y'}^T \\ \overrightarrow{z'}^T \end{pmatrix} = \begin{pmatrix} \cos(\theta) & \sin(\theta) & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}$

3. Eigenvalues and Eigenvectors

Eigenvalues and Eigenvectors

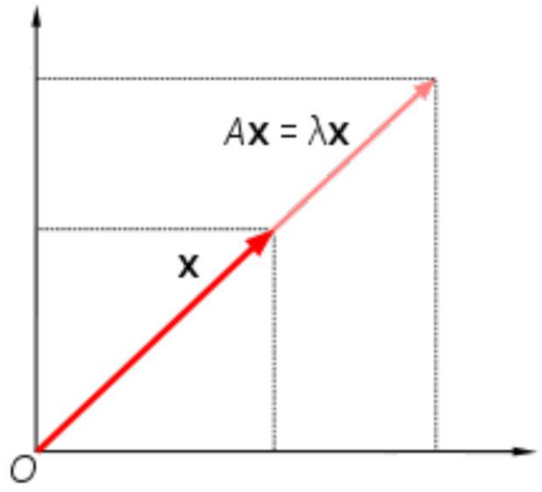
Eigen is German for "proper", "special", "characteristic"

The eigenvector x of a matrix A is a vector that satisfies the equation:

$$Ax = \lambda x$$

where λ, called the eigenvalue, is a number.

Graphically, this means that under the operation A, the vector x doesn't change direction, just magnitude.

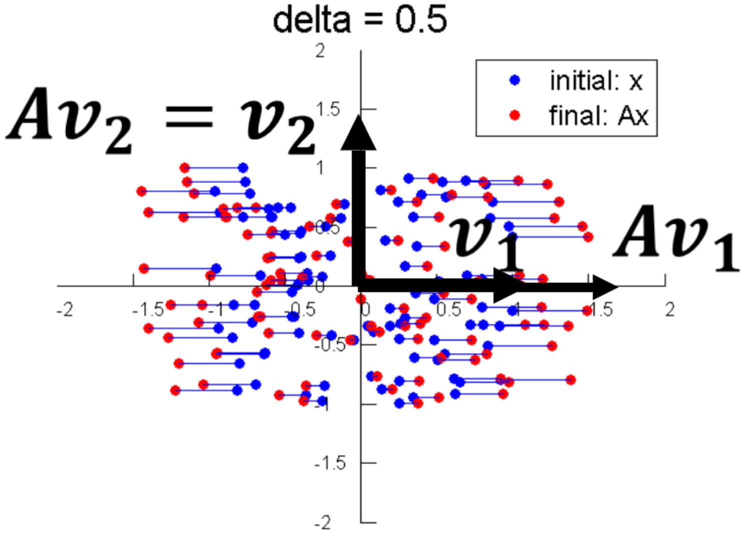


Linear Algebra

$$A = \begin{pmatrix} 1+\delta & 0 \\ 0 & 1 \end{pmatrix} \text{ has eigenvectors } v_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, v_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$Av_1 = \begin{pmatrix} 1+\delta & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1+\delta \\ 0 \end{pmatrix} = (1+\delta)v_1 \Rightarrow \lambda_1 = 1+\delta$$

$$Av_2 = \begin{pmatrix} 1+\delta & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 1 * v_1 \Rightarrow \lambda_2 = 1$$

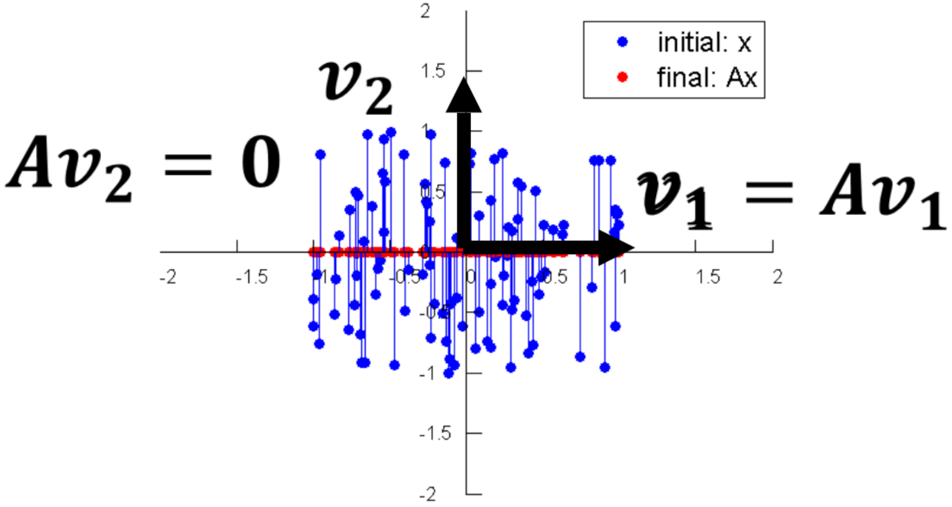


Linear Algebra

$P = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ also has eigenvectors $v_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, v_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

$$Pv_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 1 * v_1 \Rightarrow \lambda_1 = 1$$

$$Pv_2 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0 * v_1 \Rightarrow \lambda_2 = 0$$



A matrix is **non-invertible** if it has an eigenvalue $\boldsymbol{\lambda = 0}$

Property of Symmetric Matrices

The eigenvectors of any symmetric matrix A are orthogonal.

Idea of proof:

Let x and y be eigenvectors of A with eigenvalues λ, μ respectively. Assume $\lambda \neq \mu$.

$(Ax) \cdot y = (Ax)^T y = \lambda x^T y$

$(Ax) \cdot y = y^T A x = y^T A^T x = (Ay)^T x = \mu y^T x = \mu x^T y$

$\Rightarrow \lambda x^T y = \mu x^T y$

Since $\lambda \neq \mu$ then $x^T y = 0$. Thus x and y are orthogonal.

Property of Symmetric Matrices

Any **symmetric matrix** $\boldsymbol{A}$ is **diagonalizable**.

Proof:

Let A be size(n,n). Let $U_i = \begin{pmatrix} x_{1i} \\ \vdots \\ x_{ni} \end{pmatrix}$ be an eigenvector of A so that $AU_i = \lambda_i U_i$.

Assume the set of U_i form an orthonormal basis. Let $U = (U_1 \quad \ldots \quad U_n)$.

$$U^T A U = \begin{pmatrix} U_1^T \\ \vdots \\ U_n^T \end{pmatrix} A \,(U_1 \quad \ldots \quad U_n) = \begin{pmatrix} U_1^T \\ \vdots \\ U_n^T \end{pmatrix} (\lambda_1 U_1 \quad \ldots \quad \lambda_n U_n) = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix}$$

Linear Algebra

Eigendecomposition and SVD

In fact, if a square matrix has n linearly independent eigenvectors, it can always be diagonalized $A = U\ diag(\lambda)U^T$.

In this case we also immediately get the inverse matrix $A^{-1} = U\ diag\left(\frac{1}{\lambda}\right)U^T$

For a non-square $m \times n$ matrix we can at best perform the **Singular Value Decomposition**: $A = U\ D\ V^T$, where U is an $m \times m$ orthogonal matrix, D is diagonal $m \times n$ and V another n$\times$ n orthogonal matrix. Elements of D are known as **singular values**

Moore-Penrose Pseudoinverse

Matrix inversion is not defined for non-square matrices. Suppose we have a linear equation $Ax = y$ and we want to solve for x.

- If A is taller than wider, there might be no solutions
- If it is wider than taller, there might be many solutions.

The pseudoinverse is defined as $$A^+ = \lim_{\alpha \to 0^+} (A^T A + \alpha I)^{-1} A^T$$

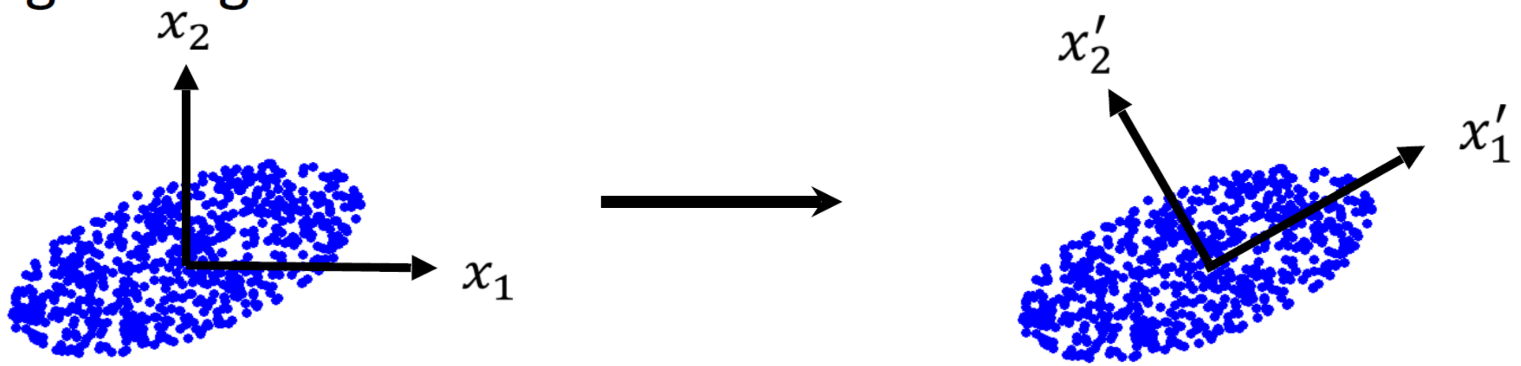
In practice we calculate it as $A^+ = VD^+U^T$, where U, D, V are the SVD of A and D^+ is calculated by taking the reciprocal of non-zero singular values and taking the transpose of the result. (Note this is clearly discontinuous)

4. Principal Component Analysis (PCA)

Linear Algebra

We are now ready for PCA

The goal of PCA is to visualize and find structure in the data. This is challenging for high dimensional data.

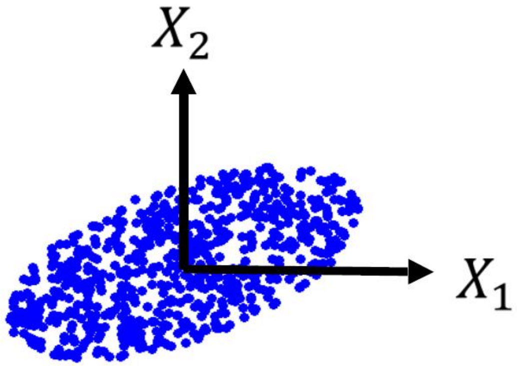


Assumption: The relevant dimensions are a linear combination of the variables we measured.

Assumption: The relevant variables are orthogonal

Linear Algebra

PCA



We measure n variables m times.

Example: $n =$ # of neurons, m= # of measurements/trials.

$X_i^T = (x_{i1} \quad \cdots \quad x_{im})$ is the measures of neuron i over all trials.

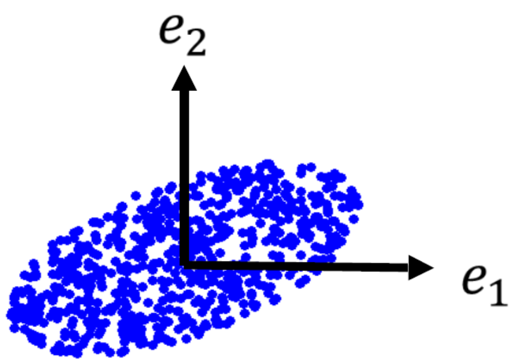
Assume X_i has zero mean.

$$var(i) = \frac{1}{m-1}\sum_{k=1}^{m} x_{ik}^2 = \frac{1}{m-1} X_i^T X_i$$

$$cov(i,j) = \frac{1}{m-1}\sum_{k=1}^{m} x_{ik}x_{jk} = \frac{1}{m-1} X_i^T X_j$$

Linear Algebra

PCA



Let $X = \begin{pmatrix} X_1^T \\ \vdots \\ X_n^T \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{pmatrix}$

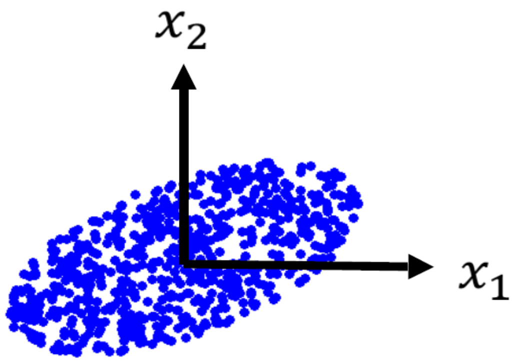
$C_X = \frac{1}{m-1} XX^T = \frac{1}{m-1} \begin{pmatrix} X_1^T \\ \vdots \\ X_n^T \end{pmatrix} \begin{pmatrix} X_1 & \dots & X_n \end{pmatrix}$ is the *Covariance Matrix*.

It is symmetric.

$$C_X^T = \frac{1}{m-1}(XX^T)^T = \frac{1}{m-1}(X^T)^T X^T = \frac{1}{m-1} XX^T = C_X$$

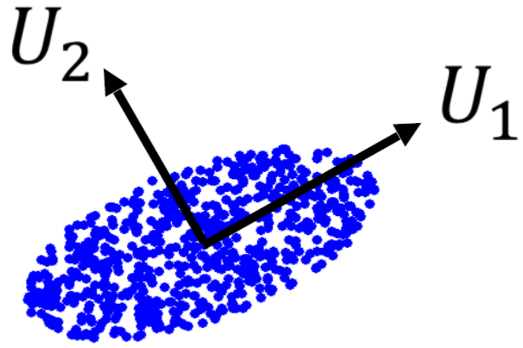
Linear Algebra

PCA



$C_X = \frac{1}{m-1} X X^T$ has orthonormal eigenvectors

Let $U = (U_1 \quad \dots \quad U_n)$ be an orthogonal matrix whose columns are eigenvectors of C_X with eigenvalues λ_i. We can chose U such that $\lambda_1 > \lambda_2 > \cdots > \lambda_n$.

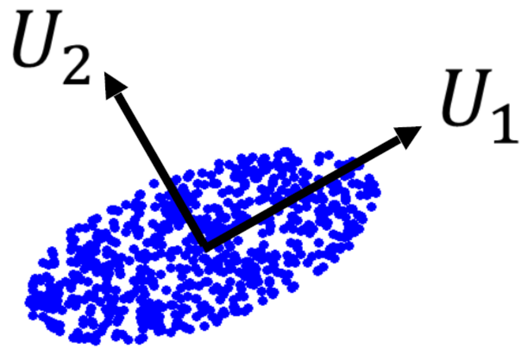


U_1 points in the directions of the greatest variance.

U_2 points in the orthogonal direction of next greatest variance. Etc.

Linear Algebra

PCA



Let $Y = U^T X$ be a transformation of the data.

The new variables Y_i are uncorrelated

$$C_Y = \frac{1}{m-1} Y Y^T = \frac{1}{m-1} U^T X X^T U = U^T C_X U = diag(\lambda_1 \quad \ldots \quad \lambda_n)$$

The eigenvectors are the principal components:

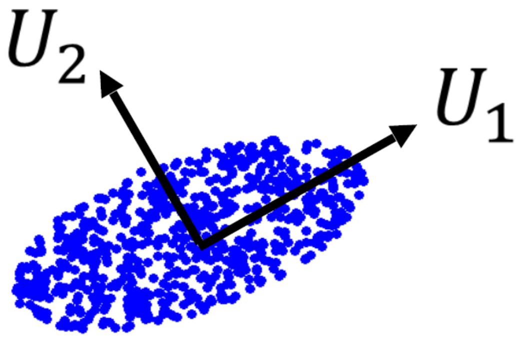
U_1 = 1st principal component

U_2 = 2nd principal component

Etc.

Linear Algebra

PCA in MATLAB




```
Cx = (1/m).*X*X';
[U,D]=eig(Cx);

% reorder things
[D,ord]=sort(diag(D));
D = flip(D); ord=flip(ord);
U=U(:,ord);    % U(:,i) is the ith principal component
```

When PCA fails

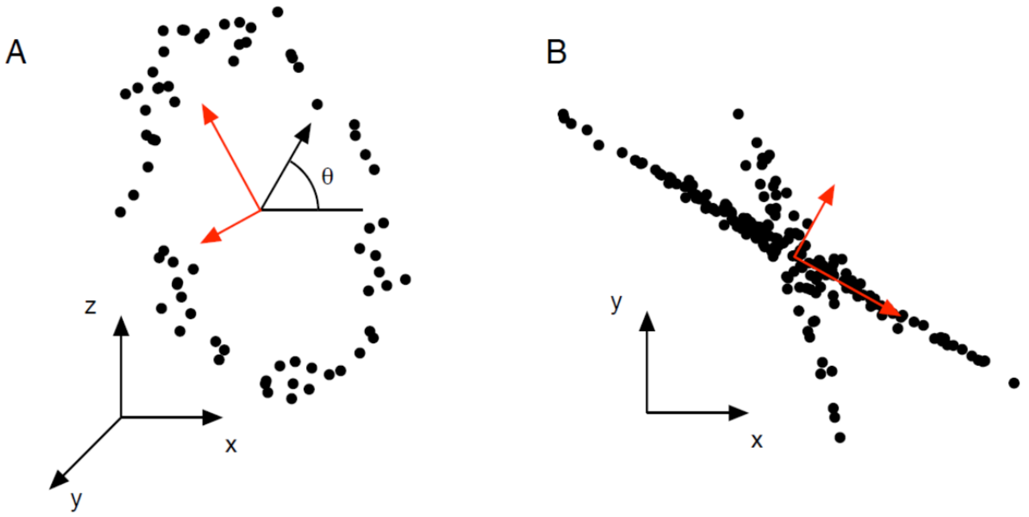
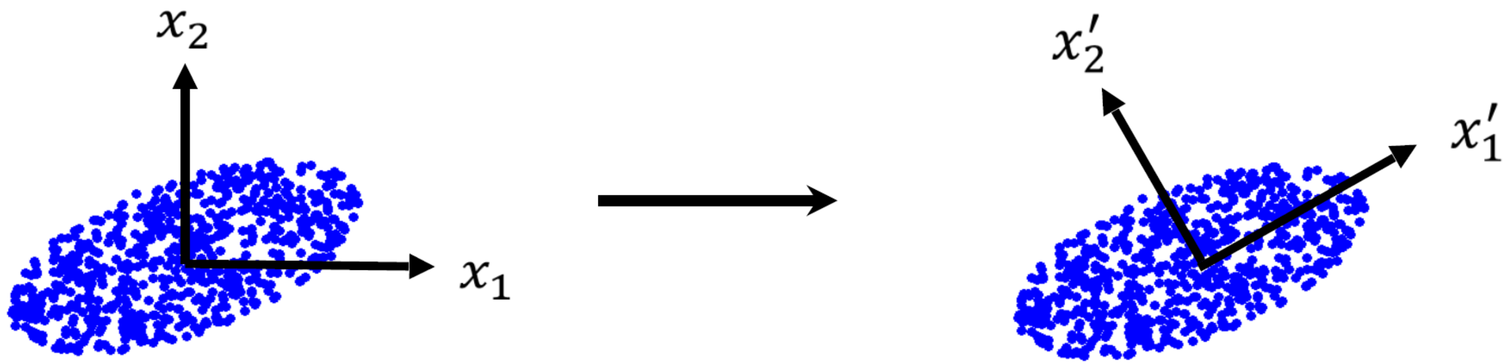


FIG. 6 Example of when PCA fails (red lines). (a) Tracking a person on a ferris wheel (black dots). All dynamics can be described by the phase of the wheel θ, a non-linear combination of the naive basis. (b) In this example data set, non-Gaussian distributed data and non-orthogonal axes causes PCA to fail. The axes with the largest variance do not correspond to the appropriate answer.

A Tutorial on Principal Component Analysis by Jonathon Shiens. Google Research 2014

Linear Algebra

PCA



Dimensionality reduction: We can use PCA to reduce the dimensions of our data to include only those dimensions which have high variance and regard the other dimensions as noise.

We assume the direction in the data which contains the most variance contains the interesting dynamics

CENTER FOR
Brains
Minds+
Machines

Thanks!