L3: Review of linear algebra and MATLAB

- Vector and matrix notation
- Vectors
- Matrices
- Vector spaces
- Linear transformations
- Eigenvalues and eigenvectors
- MATLAB® primer

Vector and matrix notation

– A d-dimensional (column) vector x and its transpose are written as:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} \text{ and } x^T = \begin{bmatrix} x_1 & x_2 & \dots & x_d \end{bmatrix}$$

– An $n \times d$ (rectangular) matrix and its transpose are written as

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1d} \\ a_{21} & a_{22} & a_{23} & & a_{2d} \\ \vdots & & & \ddots & \\ a_{n1} & a_{n2} & a_{n3} & & a_{nd} \end{bmatrix} \text{ and } a^T = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & & a_{n2} \\ a_{13} & a_{23} & & a_{n3} \\ \vdots & & \ddots & \\ a_{1d} & a_{2d} & & a_{nd} \end{bmatrix}$$

– The product of two matrices is

$$AB = \begin{bmatrix} a_{11} & a_{12} & a_{13} & & a_{1d} \\ a_{21} & a_{22} & a_{23} & & a_{2d} \\ & & & & \\ a_{m1} & a_{m2} & a_{m3} & & a_{md} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & & b_{1n} \\ b_{21} & b_{22} & & b_{2n} \\ b_{31} & b_{32} & & b_{3n} \\ & & & \\ b_{d1} & b_{d2} & & b_{dn} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & & c_{1n} \\ c_{21} & c_{22} & c_{23} & & c_{2n} \\ c_{31} & c_{32} & c_{33} & & c_{3n} \\ & & & & \\ c_{m1} & c_{m2} & c_{m3} & & c_{mn} \end{bmatrix}$$

where $c_{ij} = \sum_{k=1}^{d} a_{ik} b_{kj}$

Vectors

– The <u>inner product</u> (a.k.a. dot product or scalar product) of two vectors is defined by

$$\langle x, y \rangle = x^T y = y^T x = \sum_{k=1}^{d} x_k y_k$$

– The <u>magnitude</u> of a vector is

$$|x| = \sqrt{x^T x} = \left[\sum_{k=1}^{d} x_k x_k \right]^{\frac{1}{2}}$$

– The <u>orthogonal projection</u> of vector y onto vector x is $\langle y^T u_x \rangle u_x$

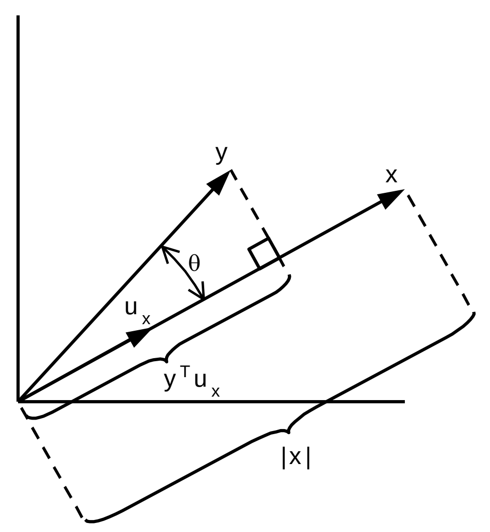
- where vector u_x has unit magnitude and the same direction as x

– The <u>angle</u> between vectors x and y is

$$cos\theta = \frac{\langle x, y \rangle}{|x||y|}$$

– Two vectors x and y are said to be

- <u>orthogonal</u> if $x^T y = 0$
- <u>orthonormal</u> if $x^T y = 0$ and $|x| = |y| = 1$



- A set of vectors $x_1, x_2, \ldots, x_n$ are said to be <u>linearly dependent</u> if there exists a set of coefficients $a_1, a_2, \ldots, a_n$ (at least one different than zero) such that

$$a_1 x_1 + a_2 x_2 \ldots a_n x_n = 0$$

- Alternatively, a set of vectors $x_1, x_2, \ldots, x_n$ are said to be <u>linearly independent</u> if

$$a_1 x_1 + a_2 x_2 \ldots a_n x_n = 0 \Rightarrow a_k = 0 \ \forall k$$

Matrices

- The determinant of a square matrix $A_{d\times d}$ is

$$|A| = \sum_{k=1}^{d} a_{ik}|A_{ik}|(-1)^{k+i}$$

  - where A_{ik} is the minor formed by removing the i^{th} row and the k^{th} column of A
  - NOTE: the determinant of a square matrix and its transpose is the same: $|A| = |A^T|$
- The trace of a square matrix $A_{d\times d}$ is the sum of its diagonal elements

$$tr(A) = \sum_{k=1}^{d} a_{kk}$$

- The rank of a matrix is the number of linearly independent rows (or columns)
- A square matrix is said to be non-singular if and only if its rank equals the number of rows (or columns)
  - A non-singular matrix has a non-zero determinant

- A square matrix is said to be orthonormal if $AA^T = A^T A = I$
- For a square matrix A
  - if $x^T Ax > 0 \quad \forall x \neq 0$, then A is said to be positive-definite (i.e., the covariance matrix)
  - $x^T Ax \geq 0 \quad \forall x \neq 0$, then A is said to be positive-semi-definite
- The inverse of a square matrix A is denoted by A^{-1} and is such that

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

  - The inverse A^{-1} of a matrix A exists if and only if A is non-singular
- The pseudo-inverse matrix $A^{\dagger}$ is typically used whenever A^{-1} does not exist (because A is not square or A is singular)

$$A^{\dagger} = [A^T A]^{-1} A^T \text{ with } A^{\dagger} A = I \quad \text{(assuming } A^T A \text{ is non-singular)}$$

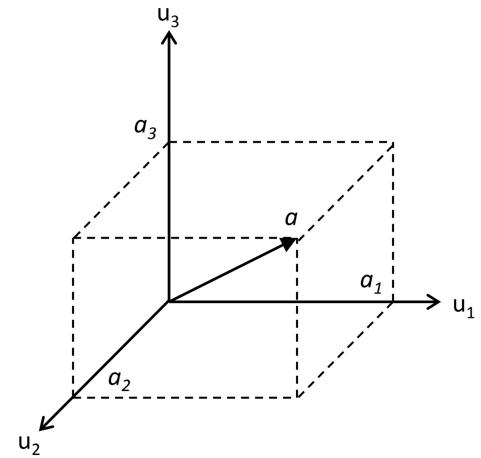
  - Note that $AA^{\dagger} \neq I$ in general

Vector spaces

- The n-dimensional space in which all the n-dimensional vectors reside is called a vector space
- A set of vectors $\{u_1, u_2, \dots\ u_n\}$ is said to form a basis for a vector space if any arbitrary vector x can be represented by a linear combination of the $\{u_i\}$

$$x = a_1 u_1 + a_2 u_2 + \cdots a_n u_n$$

  - The coefficients $\{a_1, a_2, \dots\ a_n\}$ are called the components of vector x with respect to the basis $\{u_i\}$
  - In order to form a basis, it is necessary and sufficient that the $\{u_i\}$ vectors be linearly independent



- A basis $\{u_i\}$ is said to be orthogonal if $u_i^T u_j \begin{cases} \neq 0 & i = j \\ = 0 & i \neq j \end{cases}$
- A basis $\{u_i\}$ is said to be orthonormal if $u_i^T u_j \begin{cases} = 1 & i = j \\ = 0 & i \neq j \end{cases}$
  - As an example, the Cartesian coordinate base is an orthonormal base

- Given n linearly independent vectors $\{x_1, x_2, \dots\ x_n\}$, we can construct an orthonormal base $\{\phi_1, \phi_2, \dots \phi_n\}$ for the vector space spanned by $\{x_i\}$ with the Gram-Schmidt orthonormalization procedure (to be discussed in the RBF lecture)
- The distance between two points in a vector space is defined as the magnitude of the vector difference between the points

$$d_E(x, y) = |x - y| = \left[\sum_{k=1}^{d} (x_k - y_k)^2\right]^{\frac{1}{2}}$$

  - This is also called the Euclidean distance

Linear transformations

- A linear transformation is a mapping from a vector space X^N onto a vector space Y^M, and is represented by a matrix
  - Given vector $x \epsilon X^N$, the corresponding vector y on Y^M is computed as

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & & a_{1N} \\ a_{21} & a_{22} & & a_{2N} \\ & & \ddots & \\ a_{M1} & a_{M2} & & a_{MN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$$

  - Notice that the dimensionality of the two spaces does not need to be the same
  - For pattern recognition we typically have $M < N$ (project onto a lower-dim space)
- A linear transformation represented by a square matrix A is said to be orthonormal when $AA^T = A^T A = I$
  - This implies that $A^T = A^{-1}$
  - An orthonormal xform has the property of preserving the magnitude of the vectors

$$|y| = \sqrt{y^T y} = \sqrt{(Ax)^T Ax} = \sqrt{x^T A^T Ax} = \sqrt{x^T x} = |x|$$

  - An orthonormal matrix can be thought of as a rotation of the reference frame
  - The **row vectors** of an orthonormal xform are a set of orthonormal basis vectors

$$Y_{M\times 1} = \begin{bmatrix} \leftarrow a_1 \rightarrow \\ \leftarrow a_2 \rightarrow \\ \\ \leftarrow a_N \rightarrow \end{bmatrix} X_{N\times 1} \text{ with } a_i^T a_j = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

Eigenvectors and eigenvalues

– Given a matrix $A_{N\times N}$, we say that v is an eigenvector* if there exists a scalar λ (the eigenvalue) such that

$$Av = \lambda v$$

– Computing the eigenvalues

$$Av = \lambda v \;\Rightarrow\; (A - \lambda I)v = 0 \;\Rightarrow\; \begin{cases} v = 0 & \text{Trivial solution} \\ (A - \lambda I) = 0 & \text{Non-trivial solution} \end{cases}$$

$$(A - \lambda I) = 0 \;\Rightarrow\; |A - \lambda I| = 0 \;\Rightarrow\; \underbrace{\lambda^N + a_1\lambda^{N-1} + a_2\lambda^{N-2} + +a_{N-1}\lambda + a_0 = 0}_{\textit{Characteristic equation}}$$

*The "eigen-" in "eigenvector" translates as "characteristic"

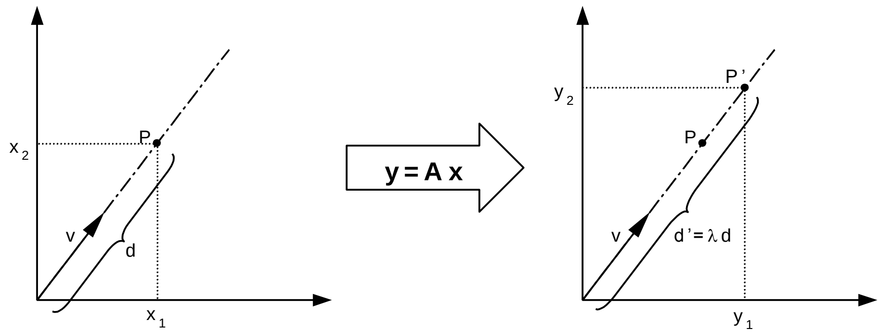
- The matrix formed by the column eigenvectors is called the modal matrix M
  - Matrix Λ is the canonical form of A: a diagonal matrix with eigenvalues on the main diagonal

$$M = \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ v_1 & v_2 & & v_N \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} \quad \Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & & \\ & & & \lambda_N \end{bmatrix}$$

- Properties
  - If A is non-singular, all eigenvalues are non-zero
  - If A is real and symmetric, all eigenvalues are real
    - The eigenvectors associated with distinct eigenvalues are orthogonal
  - If A is positive definite, all eigenvalues are positive

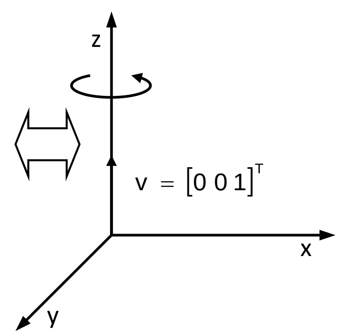
Interpretation of eigenvectors and eigenvalues

– If we view matrix A as a linear transformation, an eigenvector represents an invariant direction in vector space
  - When transformed by A, any point lying on the direction defined by v will remain on that direction, and its magnitude will be multiplied by λ



  - For example, the transform that rotates 3-d vectors about the Z axis has vector $[0\ 0\ 1]$ as its only eigenvector and $\lambda = 1$ as its eigenvalue

$$A = \begin{bmatrix} \cos\beta & -\sin\beta & 0 \\ \sin\beta & \cos\beta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

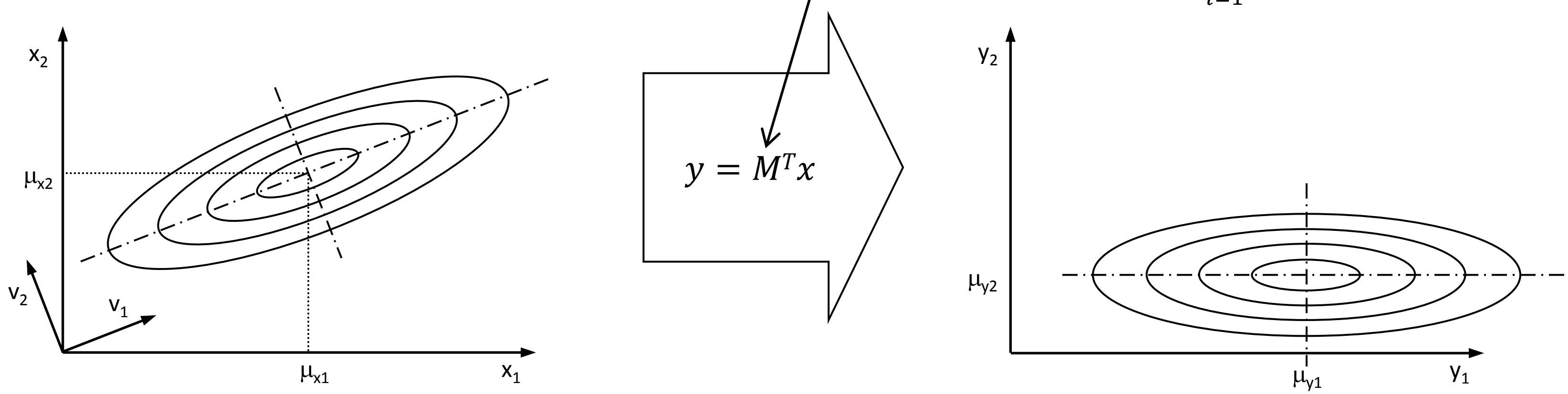


- Given the covariance matrix Σ of a Gaussian distribution
  - The eigenvectors of Σ are the principal directions of the distribution
  - The eigenvalues are the variances of the corresponding principal directions
- The linear transformation defined by the eigenvectors of Σ leads to vectors that are uncorrelated <u>regardless</u> of the form of the distribution
  - If the distribution happens to be Gaussian, then the transformed vectors will be statistically independent

$$\Sigma M = M\Lambda \text{ with } M = \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ v_1 & v_2 & & v_N \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} \quad \Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & & \\ & & & \lambda_N \end{bmatrix}$$

$$f_x(x) = \frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right]$$

$$f_y(y) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\lambda_i}} \exp\left[-\frac{\left(y_i - \mu_{y_i}\right)^2}{2\lambda_i}\right]$$



MATLAB primer

The MATLAB environment
- Starting and exiting MATLAB
- Directory path
- The startup.m file
- The help command
- The toolboxes

Basic features (help general)
- Variables
- Special variables (i, NaN, eps, realmax, realmin, pi, …)
- Arithmetic, relational and logic operations
- Comments and punctuation (the semicolon shorthand)
- Math functions (help elfun)

Arrays and matrices
- Array construction
  - Manual construction
  - The 1:n shorthand
  - The linspace command
- Matrix construction
  - Manual construction
  - Concatenating arrays and matrices
- Array and Matrix indexing (the colon shorthand)
- Array and matrix operations
  - Matrix and element-by-element operations
- Standard arrays and matrices (eye, ones and zeros)
- Array and matrix size (size and length)
- Character strings (help strfun)
  - String generation
  - The str2mat function

M-files
- Script files
- Function files

Flow control
- if..else..end construct
- for construct
- while construct
- switch..case construct

I/O (help iofun)
- Console I/O
  - The fprintf and sprintf commands
  - the input command
- File I/O
  - load and save commands
  - The fopen, fclose, fprintf and fscanf commands

2D Graphics (help graph2d)
- The plot command
- Customizing plots
  - Line styles, markers and colors
  - Grids, axes and labels
- Multiple plots and subplots
- Scatter-plots
- The legend and zoom commands

3D Graphics (help graph3d)
- Line plots
- Mesh plots
- image and imagesc commands
- 3D scatter plots
- the rotate3d command

Linear Algebra (help matfun)
- Sets of linear equations
- The least-squares solution (x = A\b)
- Eigenvalue problems

Statistics and Probability
- Generation
  - Random variables
    - Gaussian distribution: N(0,1) and N(⍰,⍰)
    - Uniform distribution
  - Random vectors
    - correlated and uncorrelated variables
- Analysis
  - Max, min and mean
  - Variance and Covariance
  - Histograms