Why.N.How: Linear Algebra

Adrian KC Lee ScD
Research Fellow, MGH

Martinos Center
November 6, 2008

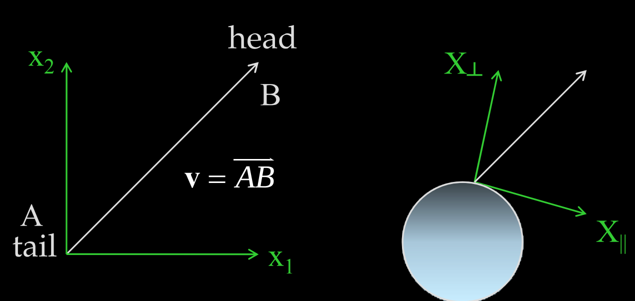
Format of tutorial

- Why do we care about Linear Algebra?
- What is a vector?
  - Geometric interpretation.
  - Vector operations, “Norms,” Projections.
  - Independence, Basis set.
- What is a matrix?
  - Determinants, Inverse.
  - Under-determined solution / null-space / MEG.
  - Over-determined solution / LMMSE / fMRI.

Why Linear Algebra?

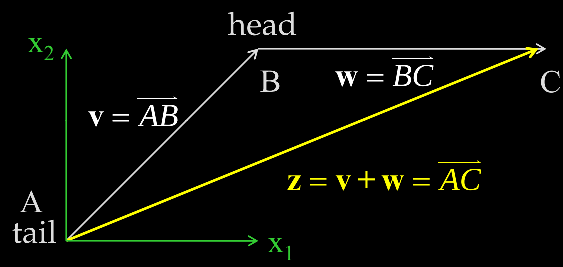
- What is linear algebra?
  - Study of vectors, systems of linear equations…
- Who cares?
  - Extensive applications in neurosciences.
  - Nice geometric interpretation of complex mathematical concepts, e.g., spectral coherence.
- How do we use MATLAB?
  - MATrix LABoratory: fundamental knowledge in linear algebra matters!!

What is a vector?



- Physical interpretation: a geometric entity characterized by a magnitude and a direction.
- Commonly encountered in Euclidean space.
  - In MEG: tangential and normal decomposition.

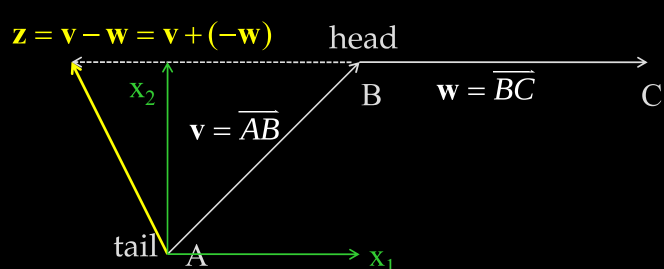
Vector Addition



$$\mathbf{v} = \begin{pmatrix} 2 \\ 2 \end{pmatrix} \qquad \mathbf{w} = \begin{pmatrix} 3 \\ 0 \end{pmatrix}$$

$$\mathbf{z} = \mathbf{v} + \mathbf{w} = \begin{pmatrix} 2 \\ 2 \end{pmatrix} + \begin{pmatrix} 3 \\ 0 \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \end{pmatrix}$$

Vector Subtraction



$$\mathbf{v} = \begin{pmatrix} 2 \\ 2 \end{pmatrix} \qquad \mathbf{w} = \begin{pmatrix} 3 \\ 0 \end{pmatrix}$$

$$\mathbf{z} = \mathbf{v} - \mathbf{w} = \mathbf{v} + (-\mathbf{w}) = \begin{pmatrix} 2 \\ 2 \end{pmatrix} + \begin{pmatrix} -3 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$$

"Norm"

$$\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + x_3^2 + \ldots + x_n^2}$$

$$= \left(|x_1|^p + |x_2|^p + |x_3|^p + \ldots + |x_n|^p\right)^{1/p}$$

$$\mathbf{v} = \begin{pmatrix} -3 \\ 4 \end{pmatrix}$$

$$\|\mathbf{v}\| = \sqrt{(-3)^2 + (4)^2} = 5$$

$$\|\mathbf{x}\|_\infty = \max\left\{|x_1|, |x_2|, \ldots, |x_p|\right\}$$

- "Norm" is a measure of distance of a vector.
  - Default norm: L2-norm
  - Other norms: L1-norm; L∞-norm.

Scalar Multiplication / Unit Vector

$$\mathbf{v} = \begin{pmatrix} -3 \\ 4 \end{pmatrix} \quad \frac{1}{2}\mathbf{v}$$

$$\frac{1}{2}\mathbf{v} = \frac{1}{2}\begin{pmatrix} -3 \\ 4 \end{pmatrix} = \begin{pmatrix} -1.5 \\ 2 \end{pmatrix}$$

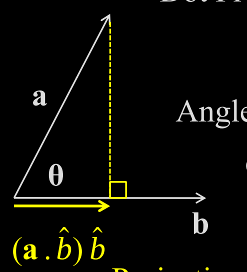
$$\hat{v} = \frac{\mathbf{v}}{\|\mathbf{v}\|}$$

$$\|\mathbf{v}\| = \sqrt{(-3)^2 + (4)^2} = 5$$

$$\hat{v} = \frac{1}{5}\begin{pmatrix} -3 \\ 4 \end{pmatrix}$$

- Unit vector factors out the magnitude,
  - Watch out for MATLAB vectors (e.g., PCA), they give you unit vectors.

Dot product / Projection

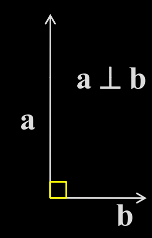


Dot Product:

$$a_1 b_1 + \ldots + a_n b_n = \|\mathbf{a}\|\|\mathbf{b}\|\cos\theta$$

Angle between 2 vectors:

$$\cos\theta = \frac{\mathbf{a} \,.\, \mathbf{b}}{\|\mathbf{a}\|\|\mathbf{b}\|}$$



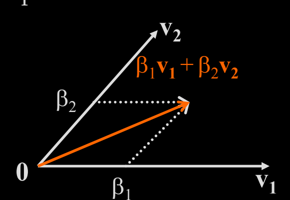
- Two vectors are orthogonal if their dot product equals to 0, i.e., If $\mathbf{a} \perp \mathbf{b}$,then $\mathbf{a} \,.\, \mathbf{b} = 0$
- Note: $\mathbf{a} \,.\, \mathbf{a} = \|\mathbf{a}\|^2$.

Independence / Basis set

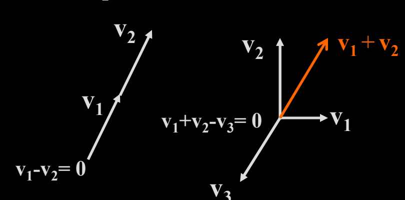
Definition: $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is independent if and only if no vector in this set can be written as a linear combination of the rest:

$${}_1\mathbf{v}_1 + \ldots + \beta_n \mathbf{v}_n = \mathbf{0} \quad \text{iff} \quad \forall \beta_i = 0$$

Independent set of vectors in $\Re^2$



Dependent set of vectors in $\Re^2$



- Unique weights for decomposition into an independent set.
- Basis set: a set of independent vectors that span V,
  - e.g., $\mathbf{v}_1$ and $\mathbf{v}_2$ form a basis set for $\Re^2$ but not $\Re^3$.

Matrix (Addition / Scalar Multiplication / Transpose)

- Matrix: rectangular table of elements
  - used to describe linear equations.

m x n Matrix

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

m rows

n columns

$$\mathbf{A} = \begin{bmatrix} 2 & 4 \\ 4 & 11 \end{bmatrix}; \mathbf{B} = \begin{bmatrix} 1 & 3 \\ 1 & 2 \end{bmatrix}$$

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} 2+1 & 4+3 \\ 4+1 & 11+2 \end{bmatrix} = \begin{bmatrix} 3 & 7 \\ 4 & 13 \end{bmatrix}$$

$$\mathbf{B}^T = \begin{bmatrix} 1 & 1 \\ 3 & 2 \end{bmatrix}$$

$$.\mathbf{A} = \begin{bmatrix} 3.2 & 3.4 \\ 3.4 & 3.11 \end{bmatrix} = \begin{bmatrix} 6 & 12 \\ 12 & 33 \end{bmatrix}$$

Note: **A** is symmetric, i.e., $\mathbf{A}^T = \mathbf{A}$

Matrix Multiplication

- Matrix multiplication is NOT commutative:
  - i.e., generally, $\mathbf{AB} \neq \mathbf{BA}$

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}; \mathbf{B} = \begin{bmatrix} 2 & 3 \\ 1 & 0 \end{bmatrix}$$

$$\mathbf{A}.\mathbf{B} = \begin{bmatrix} 1.2+2.1 & 1.3+2.0 \\ 0.2+1.1 & 0.3+1.0 \end{bmatrix} = \begin{bmatrix} 4 & 3 \\ 1 & 0 \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} 2.1+3.0 & 2.2+3.1 \\ 1.1+0.0 & 1.2+0.1 \end{bmatrix} = \begin{bmatrix} 2 & 7 \\ 1 & 2 \end{bmatrix}$$

Matrix Multiplication

- Dimension of matrices must match
  - $\mathbf{C}_{(m,p)} = \mathbf{A}_{(m,n)}\mathbf{B}_{(n,p)}$

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 2 \\ -1 & 3 & 1 \end{bmatrix}; \mathbf{B} = \begin{bmatrix} 3 & 1 \\ 2 & 1 \\ 1 & 0 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 1.3+0.2+2.1 & 1.1+0.1+2.0 \\ -1.3+3.2+1.1 & -1.1+3.1+1.0 \end{bmatrix} = \begin{bmatrix} 5 & 1 \\ 4 & 2 \end{bmatrix}$$

Solving Linear Equations

$$\begin{matrix} a_{11}x_1 + \cdots + a_{1n}x_n & = & b_1 \\ \vdots & \vdots & \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n & = & b_n \end{matrix}$$

$$\mathbf{Ax} = \mathbf{b}$$

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

m x n Matrix

m rows

n columns

Matrix (Geometric interpretation)

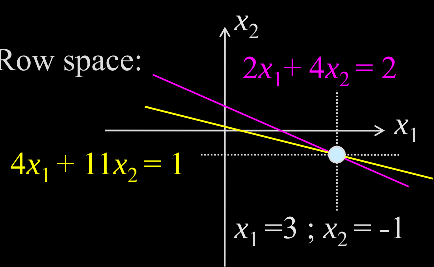
$$\mathbf{Ax} = \mathbf{b} \qquad \mathbf{A} = \begin{bmatrix} 2 & 4 \\ 4 & 11 \end{bmatrix} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

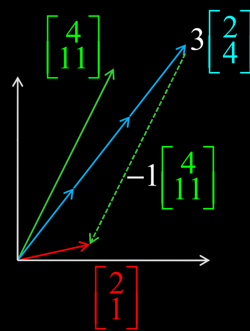
Solve for **x**:

$$\mathbf{x} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$$

Column space: $x_1 \begin{bmatrix} 2 \\ 4 \end{bmatrix} + x_2 \begin{bmatrix} 4 \\ 11 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

Row space:





How to solve for **x** in **Ax = b**?

(M equations = N unknowns)

- How many equations do I need to solve for 3 unknowns?
  - In general: M equations, N unknowns.
  - Can we guarantee a unique solution when M=N?

Demonstration:

http://demonstrations.wolfram.com/PlanesSolutionsAndGaussianEliminationOfA33LinearSystem/

Solve $\mathbf{Ax} = \mathbf{b}$ by finding $\mathbf{A}^{-1}$

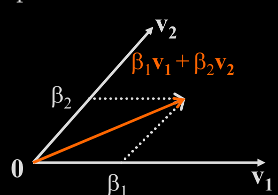
(M equations = N unknowns)

- If $\mathbf{Ax} = \mathbf{b}$ and we want to solve for $\mathbf{x}$:
  - We want $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$
- Does $\mathbf{A}^{-1}$ always exist?
  - Only if $\mathbf{A}$ is a square matrix (i.e., M=N).
  - Only if determinant of $\mathbf{A}$ is not 0.
- How do we find determinant and $\mathbf{A}^{-1}$?
  - Can use Gaussian elimination to find determinant.
  - Can use Cramer's rule to find $\mathbf{A}^{-1}$.
  - Or use MATLAB ☺!!!!

Undetermined case

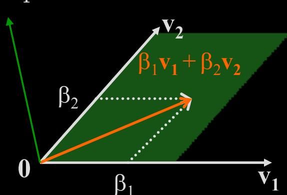
(M equations < N unknowns)

Independent set of vectors in $\Re^2$



- M = N
- Vectors span 2 dimension
  - Rank = 2.
  - Nullity = N-R = 0.

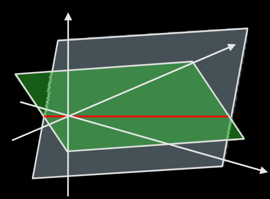
Independent set of vectors in $\Re^3$



- M < N
- Vectors span 2 dimension
  - Rank = 2.
  - Nullity = N-R = 1.

Null-space

(M equations < N unknowns)

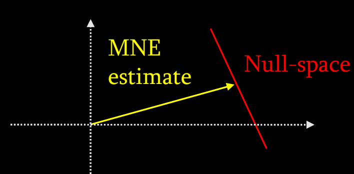


- M < N
  - Rank = 2
  - Nullity = 1

- There is no unique solution since we are in an underdetermined case.
  - Any point lying on the red line is a valid solution!
- How do we find $\mathbf{A}^{-1}$ when $\mathbf{A}$ is not square ($M \neq N$)?
  - Pseudo-inverse.
  - "Do the best we can do, given what we have."

Undetermined soln in MEG

- In MEG:
  - 306 sensors (M = 306 equations).
  - 6000 dipole estimates (N = 6000 unknowns).
  - M << N.
  - MNE: uniqueness comes from other constraints.



- Side topics:
  - How to obtain pseudo-inverse?
  - Singular Value Decomposition.
  - Eigenvalues, eigenvectors.
  - $\mathbf{Ax} = \lambda\mathbf{x}$

Overdetermined case

(M equations > N unknowns)

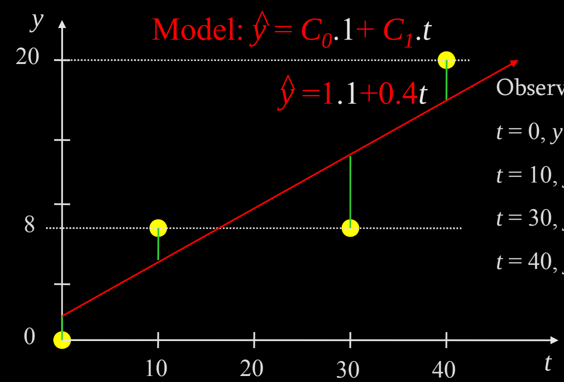
- Experiment:
  - Relationship between number of people wanting me to stop talking (y) as a function of time (t).
  - Model: $\hat{y} = C_0 + C_1 t$
    - C_0: baseline Care-Factor; C_1: drift in Attention.
    - Data collected (y: reported distractions) at time t_1=0 ($y=y_1$); t_2=10 ($y=y_2$); t_3=30 ($y=y_3$); t_4=40 ($y=y_4$).

Ax = b

$$C_0 + t_1 C_1 = y_1$$
$$C_0 + t_2 C_1 = y_2$$
$$C_0 + t_3 C_1 = y_3$$
$$C_0 + t_4 C_1 = y_4$$

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 1 & 10 \\ 1 & 30 \\ 1 & 40 \end{bmatrix} \quad \mathbf{x} = \begin{bmatrix} C_0 \\ C_1 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}$$

Results (Linear regression)



Observations:

t = 0, y = 0

t = 10, y = 8

t = 30, y = 8

t = 40, y = 20

minimize

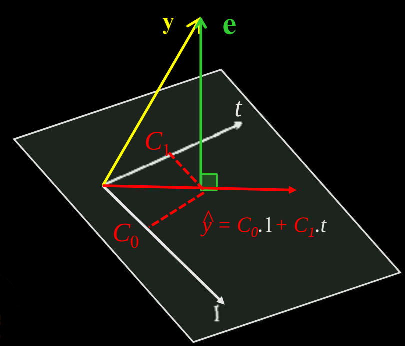
$e = y - \hat{y}$

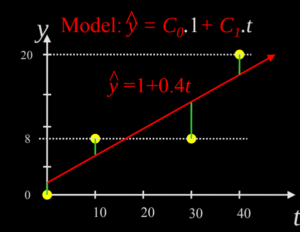
$e = \mathbf{b} - \mathbf{Ax}$

Ax = b

LMMSE

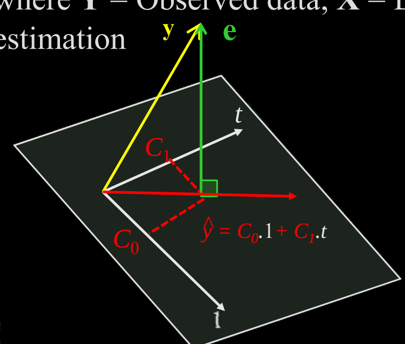
- Framework: estimate a set of weights (C_0 and C_1) such that the error (e) between our observed data (**b**) and our weighted design matrix (**A**) is minimum,
  - i.e., "minimize" $\mathbf{e} = \mathbf{b} - \mathbf{Ax} = \mathbf{y} - (C_0 \,.\, 1 + C_1 \,.\, t)$
  - <u>L</u>inearly <u>M</u>inimize by <u>M</u>ean <u>S</u>quare <u>E</u>rror: $|| \mathbf{e} ||^2$

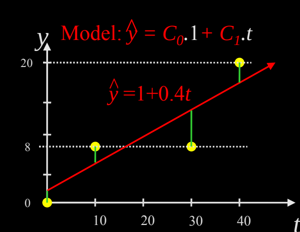




LMMSE / GLM

- Closed form solution (LMMSE)
  - Find **x** that minimizes $||\mathbf{Ax} - \mathbf{b}||^2$
  - Least square solution to $\mathbf{Ax} = \mathbf{b}$: $\mathbf{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$
- Now consider Random variables as vectors (GLM):
  - Least square solution to $\mathbf{X}\beta = \mathbf{Y}$: $\beta = (\mathbf{X}^T\mathbf{X})^{-1}\, \mathbf{X}^T\, \mathbf{Y}$
  - where **Y** = Observed data; **X** = Design matrix; β = Parameter estimation



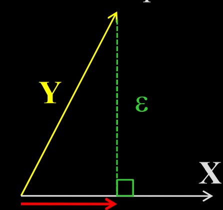


GLM in a nut-shell (fMRI context)

$$\mathrm{Y} = \mathbf{X}\beta + \varepsilon$$

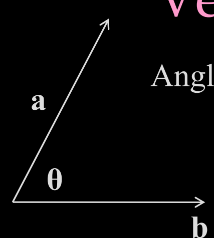
- **Y** : BOLD signal at various time at a single voxel
- **X** : Designed matrix
- β : Estimation of each component in **X** (Parameters estimation)
- ε : Noise (also known as residuals, i.e., anything that is not modeled by **X**)

Error (ε) is orthogonal to all our modeled parameters:



β= Projection onto each modeled parameter in Design matrix **X**

Vectors → Statistics

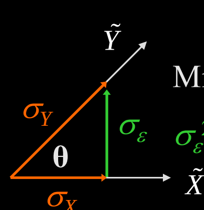


Angle between 2 vectors:

$$\cos\theta = \frac{\mathbf{a}\,.\,\mathbf{b}}{\|\mathbf{a}\|\|\mathbf{b}\|}$$

$$\sigma_{YX} = E[\tilde{Y}\tilde{X}] = \sigma_Y \sigma_X \cos\theta$$

$$\cos\theta = \frac{\sigma_{YX}}{\sigma_Y \sigma_X} = \rho$$



Min. MSE: σ_ε^2

$\sigma_\varepsilon^2 = \sigma_Y^2 - (\sigma_Y \cos\theta)^2$

$$MMSE = \sigma_Y^2(1-\rho^2)$$

ρ = correlation coefficient

Y: Data, *X*: estimate

Coherence measure (Gross *et al*, 2001):

$$\frac{|\mathbf{C}_{x,y}(f)|^2}{\mathbf{C}_{x,x}(f)\mathbf{C}_{y,y}(f)}$$

Useful Resources

- MIT OpenCourseWare:
  - 18.06 Linear Algebra (Strang's Lecture series)
  - http://ocw.mit.edu/OcwWeb/Mathematics/18-06Spring-2005/CourseHome/
- Good link between Linear Algebra and MATLAB:
  - http://www.ling.upenn.edu/courses/ling525/linear_algebra_review.html
- Why.N.How Wiki:
  - https://gate.nmr.mgh.harvard.edu/wiki/whynhow/index.php/Main_Page