Optimal Estimation Methods

(Lecture 1 – Introduction and Matrix Algebra Review)

Dr. John L. Crassidis

University at Buffalo – State University of New York
Department of Mechanical & Aerospace Engineering
Amherst, NY 14260-4400
johnc@buffalo.edu
http://www.buffalo.edu/~johnc

University at Buffalo The State University of New York

Topics (i)

- Introduction to Estimation
- Parameter Optimization Review
- Least Squares Estimation
  - Linear (Batch and Sequential) and Nonlinear
  - Simple Examples
  - Basis Functions
- Basic Probability Concepts
  - Discrete and Continuous Variables
  - Gaussian Random Variables
  - Multidimensional Variables, Covariance
- Probability Concepts in Least Squares
  - Minimum Variance Estimation
  - Maximum Likelihood Estimation
  - Unbiased Estimates and Cramér-Rao Inequality
  - Maximum *A Posteriori* Estimation

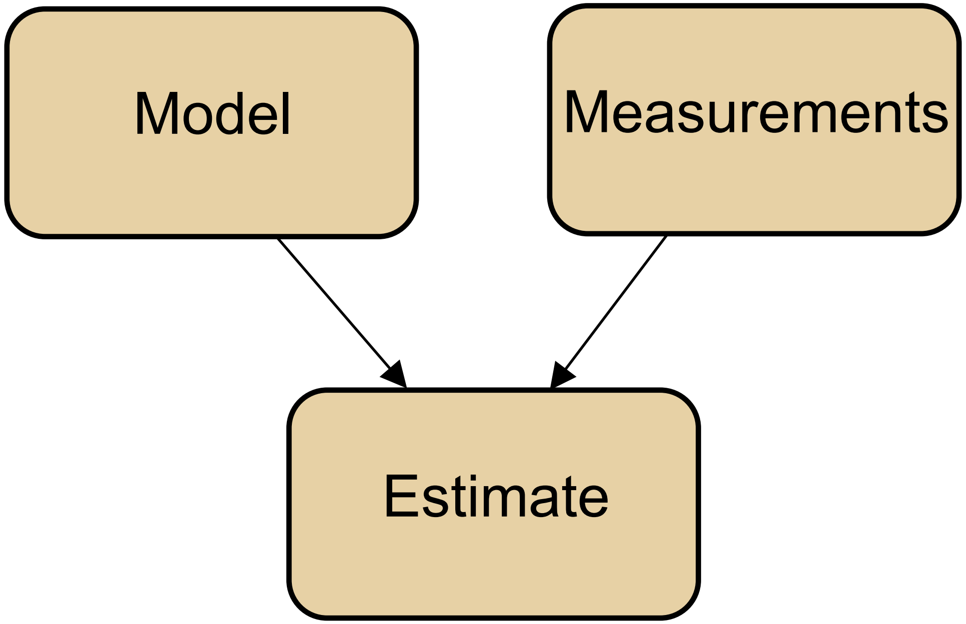
Topics (ii)

- Review of Dynamics Systems
- Kalman Filtering
  - Linear and Nonlinear (Extended) Kalman Filtering
- Sigma Point (Unscented) Filtering
  - Example and Comparison to Extended Kalman Filter
- Batch State Estimation
- Particle Filtering
  - Monte Carlo Integration
  - Importance Sampling
  - Simple Parameter Estimation Examples
  - Optimal and Approximate Filtering
  - Sequential Importance Sampling
  - Degeneracy Problem
    - Resampling and Regularization
  - Selection of the Importance Density

Topics (iii)

- Advanced Topics
  - Multiple-Model Adaptive Filtering
    - Example for Filter Tuning
  - Decentralized Filtering
    - Covariance Intersection
    - Example
  - Ensemble Kalman Filtering
    - Example
  - Introduction to Stochastic Processes
- Applications
  - Target Tracking
  - Orbit Determination
- MATLAB Code
  - Codes will be provided but use them for help only
    - Best to program up your own to learn the material

Introduction (i)



- Model usually inaccurate
- Measurements usually inaccurate and incomplete (can’t measure all states)
- Estimation combines model and measurements to obtain best of both

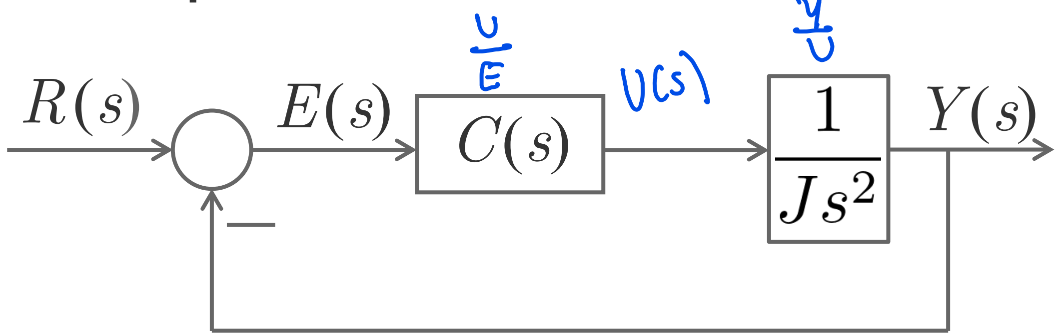
- Examples
  - Bias estimation in sensors, e.g., gyros, accelerometers, DC gains, etc.
  - Aircraft parameter estimation, e.g., ID parameters, fault detection
  - Spacecraft attitude estimation
  - Vibration modal identification
  - Others, such as forecasting stocks and health insurance rates

Introduction (ii)

$E(\,) = R(s) - Y(s)$

$Y = \left(\frac{Y}{U}\right)\left(\frac{U}{E}\right)E = \left(\frac{1}{Js^2}\right)(C)\,(R-Y)$

- Consider spacecraft attitude control system



$\frac{C}{Js^2}R - \frac{C}{Js^2}Y = Y$

$Y\left(1 + \frac{C}{Js^2}\right) = \frac{C}{Js^2}R$

$\frac{Y}{R} = \frac{\left(\frac{C}{Js^2}\right)}{1 + \frac{C}{Js^2}}$

$\frac{C}{Js^2} \cdot \frac{1}{1 + \frac{C}{Js^2}}$

$\frac{C}{Js^2 + C} = \frac{Y}{R}$

- Determine the overall transfer function (in class)

$$Y(s) = \frac{C(s)}{Js^2}E(s)$$

$$E(s) = R(s) - Y(s)$$

$$Y(s) = \frac{C(s)}{Js^2}R(s) - \frac{C(s)}{Js^2}Y(s)$$

$$[Js^2 + C(s)]Y(s) = C(s)R(s)$$

$$\frac{Y(s)}{R(s)} = \frac{C(s)}{Js^2 + C(s)}$$

- Overall closed-loop transfer function

$$\frac{Y(s)}{R(s)} = \frac{C(s)}{Js^2 + C(s)}$$

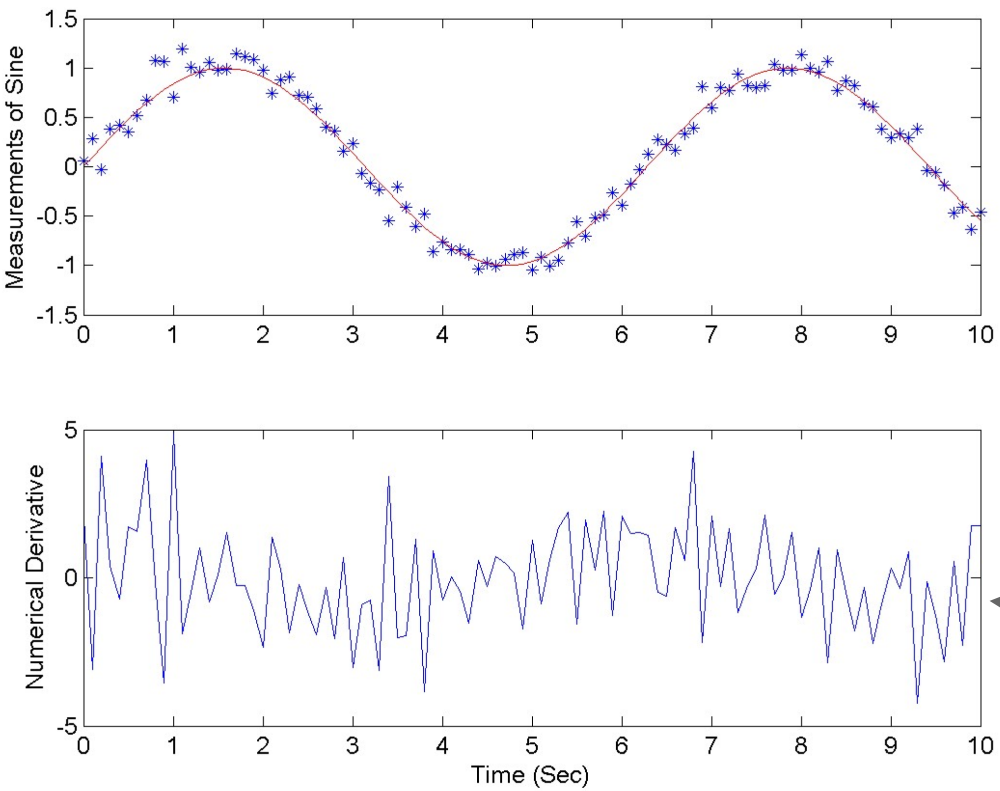
- If $C(s) = K_p$ then the closed-loop roots are given by $s_{1,2} = \pm(K_p/J)^{1/2}\ j$
  - Cannot provide an asymptotically stable response
  - Why?
- If $C(s) = K_p + K_d s$ then an asymptotically stable response can be provided
  - Why? Need angular rate information
- Bottom line is that we cannot stabilize spacecraft without some kind of rate information
  - Some researchers claim this can be done, but there is always a "derivative" hidden in their algorithm

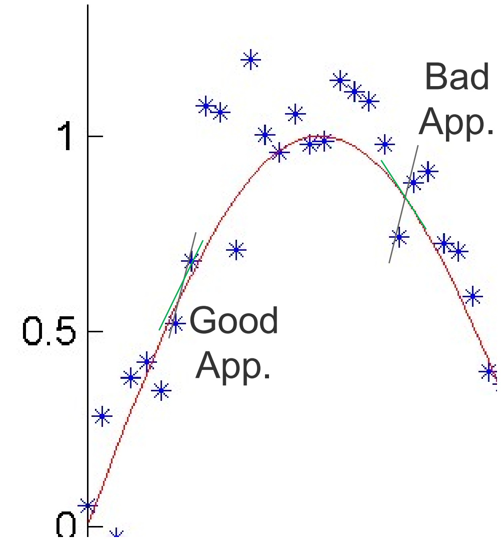
Introduction (iv)

- Spacecraft attitude control system
  - Now know can't be stabilized without rate information
- How do we obtain rate information?
  - Straight differentiation (noisy), or
  - Model for rate (inaccurate), or
  - Gyros (expensive and they drift over time)
- What is the best way?
  - Filter noisy derivatives, or
  - Use model with attitude measurements to "update model," or
  - Use gyros with a drift estimator
- All of these options require estimation principles
  - For example, a low-pass filter to filter noisy derivative "estimates" is a simple Kalman filter
    - Bandwidth set by the Kalman gain

Introduction (v)

- Numerical derivative of a noisy signal as estimate - High frequency is amplified (positive dB)
  - Enhances noise due to "slope" approximation
  - Look at a simple sine signal with noise





Can't "see" the cosine well in this signal

- Set of linear equations

$$y_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n$$

$$y_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n$$

$$\vdots$$

$$y_m = a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n$$

$$\boxed{\mathbf{y} = A\,\mathbf{x}}$$

with

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

- Matrix Addition, Subtraction and Multiplication
  - Addition/subtraction of two matrices $C = A \pm B$ where each element is given by $c_{ij} = a_{ij} \pm b_{ij}$
  - Multiplication is more complicated $C = AB$
    - Number of columns of A must be equal to the number of rows of B, so $C_{m \times p} = A_{m \times n} B_{n \times p}$
    - Elements of C are given by

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

    - Multiplication is associative $A(BC) = (AB)C$ and distributive $A(B + C) = AB + AC$
    - Multiplication is not commutative in general $AB \neq BA$
    - If $AB = BA$ is true then the matrices A and B are said to *commute*

- The transpose of a matrix, denoted by A^T, has rows that are the columns of A and the columns that are the rows of A
  - Some properties

$$(\alpha A)^T = \alpha A^T, \text{ where } \alpha \text{ is a scalar}$$
$$(A + B)^T = A^T + B^T$$
$$(A\,B)^T = B^T A^T$$

  - A symmetric matrix is define by

$$A = A^T$$

  - A skew symmetric matrix is defined by

$$A = -A^T$$

- Matrix inverse is only valid for square ($n \times n$) matrices
- The following statements are equivalent:

  null(A) - n

  linearly independent equations

  - A has linearly independent columns
  - A has linearly independent rows
  - The inverse satisfies $A^{-1}A = A\,A^{-1} = I$

  where I is the identity matrix
- A nonsingular matrix is a matrix whose inverse exists

$$(A^{-1})^{-1} = A$$

$$(A^T)^{-1} = (A^{-1})^T \equiv A^{-T}$$

- Also, if A and B are square and nonsingular then

$$(A\,B)^{-1} = B^{-1}A^{-1}$$

- The matrix inverse is given by

$$A^{-1} = \frac{\mathrm{adj}(A)}{\det(A)}$$

where $\mathrm{adj}(A)$ is the adjoint of A and $\det(A)$ is the determinant of A

- Clearly, the inverse exists if and only if the determinant of A is not zero
- The adjoint is given by the transpose of the cofactor matrix

$$\mathrm{adj}(A) = [\mathrm{cof}(A)]^T$$

- The cofactor is given by

$$C_{ij} = (-1)^{i+j} M_{ij}$$

where M_{ij} is the minor, which is the determinant of the resulting matrix given by crossing out the row and column of the matrix A element a_{ij}

- The determinant can be determined using an expansion about row i or column j

$$\det(A) = \sum_{k=1}^{n} a_{ik} C_{ik} = \sum_{k=1}^{n} a_{kj} C_{kj}$$

  - Some useful properties include

$$\det(I) = 1$$

$$\det(A\,B) = \det(A)\,\det(B)$$

$$\det(A\,B) = \det(B\,A)$$

$$\det(A) = \det(A^T)$$

$$\det(A\,B + I) = \det(B\,A + I)$$

$$\det(A)\,\det(D + C\,A^{-1}B) = \det(D)\,\det(A + B\,D^{-1}C)$$

$$\det(A^\alpha) = [\det(A)]^\alpha,\ \alpha \text{ must be positive if } \det(A) = 0$$

$$\det(\alpha\,A) = \alpha^n \det(A)$$

Block Structures

- Assume that A is an $n \times n$ matrix and that C is an $m \times m$ matrix. Then,

$$\det \begin{bmatrix} A & B \\ 0 & C \end{bmatrix} = \det \begin{bmatrix} A & 0 \\ B & C \end{bmatrix} = \det(A)\,\det(C)$$

$$\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det(A)\,\det(P) = \det(D)\,\det(Q)$$

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} Q^{-1} & -Q^{-1}B\,D^{-1} \\ -D^{-1}C\,Q^{-1} & D^{-1}(I + C\,Q^{-1}B\,D^{-1}) \end{bmatrix}$$

$$= \begin{bmatrix} A^{-1}(I + B\,P^{-1}C\,A^{-1}) & -A^{-1}B\,P^{-1} \\ -P^{-1}C\,A^{-1} & P^{-1} \end{bmatrix}$$

where P and Q are Schur complements of A and D

$$P \equiv D - C\,A^{-1}B, \quad Q \equiv A - B\,D^{-1}C$$

must exist

Useful Identities

- Assume that A is an $n \times n$ matrix and that C is an $m \times m$ matrix
  - Sherman-Morrison lemma is given by

$$(I + A\,B)^{-1} = I - A\,(I + B\,A)^{-1}B$$

  - Matrix inversion lemma is given by

$$(A + B\,C\,D)^{-1} = A^{-1} - A^{-1}B\left(D\,A^{-1}B + C^{-1}\right)^{-1}D\,A^{-1}$$

  where A is an arbitrary $n \times n$ matrix and C is an arbitrary $m \times m$ matrix

  - Matrix inversion lemma is used often in this course
    - Its proof is shown in the book

Matrix Trace

- Only valid for square matrices

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} a_{ii}$$

  - Some useful properties

$$\mathrm{Tr}(\alpha A) = \alpha \mathrm{Tr}(A)$$

$$\mathrm{Tr}(A + B) = \mathrm{Tr}(A) + \mathrm{Tr}(B)$$

$$\mathrm{Tr}(A\, B) = \mathrm{Tr}(B\, A)$$

$$\mathrm{Tr}(\mathbf{x}\, \mathbf{y}^T) = \mathbf{x}^T \mathbf{y}$$

$$\mathrm{Tr}(A\, \mathbf{y}\, \mathbf{x}^T) = \mathbf{x}^T A\, \mathbf{y}$$

$$\mathrm{Tr}(A\, B\, C\, D) = \mathrm{Tr}(B\, C\, D\, A) = \mathrm{Tr}(C\, D\, A\, B) = \mathrm{Tr}(D\, A\, B\, C)$$

  - The operation $\mathbf{y}\, \mathbf{x}^T$ is known as the *outer product* (also $\mathbf{y}\, \mathbf{x}^T \neq \mathbf{x}\, \mathbf{y}^T$ in general)

- Any scalar λ and nonzero x that satisfy the following are the eigenvalue and associated (right) eigenvector of the matrix A

$$A\,\mathbf{x} = \lambda\,\mathbf{x} \quad \text{or} \quad (\lambda\,I - A)\mathbf{x} = \mathbf{0}$$

  - Note that x is not unique (often given as a normalized vector)
  - Since x is nonzero it is in the null space of $(\lambda I - A)$, then $\lambda I - A$ must be singular, so

$$\det(\lambda\,I - A) = 0$$

  - Gives the characteristic equation (look at 2×2 case)

$$\det(\lambda\,I - A) = \det \begin{bmatrix} \lambda - a_{11} & -a_{12} \\ -a_{21} & \lambda - a_{22} \end{bmatrix} = 0$$

$$\lambda^2 - (a_{11} + a_{22})\lambda + a_{11}a_{22} - a_{12}a_{21} = 0$$

  - Special case: if the matrix A is real and symmetric then the eigenvectors are orthogonal

- A real and square matrix A is
  - *Positive definite* if $\mathbf{x}^T A\,\mathbf{x} > 0$ for all nonzero $\mathbf{x}$
  - *Positive semi-definite* if $\mathbf{x}^T A\,\mathbf{x} \geq 0$ for all nonzero $\mathbf{x}$
  - *Negative definite* if $\mathbf{x}^T A\,\mathbf{x} < 0$ for all nonzero $\mathbf{x}$
  - *Negative semi-definite* if $\mathbf{x}^T A\,\mathbf{x} \leq 0$ for all nonzero $\mathbf{x}$
  - *Indefinite* when no definiteness can be asserted
  - A real symmetric matrix is positive definite if and only if all its eigenvalues are greater than zero. For non-symmetric real matrices check the definiteness of

$$B = \frac{A + A^T}{2}$$

Check eigenvalues of B

Definiteness (ii)

- First show that real symmetric matrix has real eigenvalues
  - Standard decomposition $A\mathbf{x} = \lambda\mathbf{x}$

$(Ax)^T = x^T A'$

  - Complex conjugate for eigenvalues, denoted by λ^*, and eigenvectors follow the equation

$$\mathbf{x}^* A^T = \lambda^* \mathbf{x}^*, \text{ where } \mathbf{x}^* \equiv \bar{\mathbf{x}}^T$$

  - Left multiply first equation by $\mathbf{x}^*$ and right multiply second equation by $\mathbf{x}$ to obtain

$$\mathbf{x}^* A\mathbf{x} = \lambda\mathbf{x}^*\mathbf{x}$$

$$\mathbf{x}^* A^T\mathbf{x} = \lambda^*\mathbf{x}^*\mathbf{x}$$

  - If A is symmetric then $\lambda = \lambda^*$ must be true
    - Only way this can happen is when all eigenvalues are real

Definiteness (iii)

- A real symmetric positive definite matrix has orthogonal eigenvectors

$(Ax)^T = x^T A^T$

  - Proof is straightforward
  - Assume that eigenvalues are not equal (can be generalized if not)
  - Let $A\mathbf{x}_1 = \lambda_1 \mathbf{x}_1$ (1) and $A\mathbf{x}_2 = \lambda_2 \mathbf{x}_2$ (2) with $\lambda_1 \neq \lambda_2$
  - Take transpose of (1) and right multiply by $\mathbf{x}_2$

$$\mathbf{x}_1^T A^T \mathbf{x}_2 = \lambda_1 \mathbf{x}_1^T \mathbf{x}_2 \quad (3)$$

  - Left multiply (2) by $\mathbf{x}_1^T$

$$\mathbf{x}_1^T A \mathbf{x}_2 = \lambda_2 \mathbf{x}_1^T \mathbf{x}_2 \quad (4)$$

  - Now look at (4) – (3)

$$\mathbf{x}_1^T A \mathbf{x}_2 - \mathbf{x}_1^T A^T \mathbf{x}_2 = (\lambda_2 - \lambda_1)\mathbf{x}_1^T \mathbf{x}_2$$

Definiteness (iv)

- Assume $A = A^T$ then — symmetric

$$\mathbf{x}_1^T A \mathbf{x}_2 - \mathbf{x}_1^T A^T \mathbf{x}_2 = (\lambda_2 - \lambda_1)\mathbf{x}_1^T \mathbf{x}_2 = 0$$

$x_1 \cdot x_2 = 0$

- Since $\lambda_2 - \lambda_1 \neq 0$ then $\mathbf{x}_1$ and $\mathbf{x}_2$ must be orthogonal
- Can keep going to show that all eigenvectors are orthogonal
- Let V contain all the eigenvectors $V = [\mathbf{x}_1 \ \ldots \ \mathbf{x}_n]$ and Λ be a diagonal matrix of the eigenvalues $\lambda_1 \ \ldots \ \lambda_n$, then for a real symmetric matrix, the eigenvalue/eigenvector decomposition becomes

$$A = V \Lambda V^T$$

- For general matrices $A = V \Lambda V^{-1}$
- Note that the following notation: $B > A$ implies that $(B - A) > 0$ is a positive definite matrix — Just notation, not true element by element
  - This notation will be extensively used

Definiteness (v)

- Theorem: A real symmetric matrix is positive definite iff all of its eigenvalues are positive
- Proof begins with the eigenvalue/eigenvector decomposition

eisenvector matrix

$$A\,V = V\,\Lambda$$

- Right multiply by a vector $\mathbf{y}$: $A\,V\mathbf{y} = V\Lambda\,\mathbf{y}$
- Define $\mathbf{x} \equiv V\,\mathbf{y}$, so $\mathbf{x}^T A\,\mathbf{x}$ becomes

diag(eige values)

$$\mathbf{x}^T A\,\mathbf{x} = \mathbf{y}^T V^T A\,V\,\mathbf{y} = \mathbf{y}^T \Lambda\,\mathbf{y}$$

- Since $\mathbf{x}$ and $\mathbf{y}$ are both arbitrary then the only way that $\mathbf{y}^T \Lambda\,\mathbf{y} > 0$ can be true is if all the eigenvalues of A are positive
- Also, if all eigenvalues are negative then it is negative definite
- If some are positive and some are zero it is positive semi-definite
- If some are negative and some are positive then it is indefinite

Orthogonal Matrix

- Matrix inverse for a 2×2 matrix is simple

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad A^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

- Special case: A is orthogonal if

$$A^T A = A\,A^T = I$$

  - Let's show that $\det(A) = \pm 1$ for this case
  - Take the determinant of both sides of the equation

$$A^T A = A\,A^T = I$$

$$\det(A^T A) = \det(A\,A^T) = \det(I)$$

$$\det(A^T)\det(A) = 1$$

$$\det(A)^2 = 1$$

$$\det(A) = \pm 1$$

- Singular value decomposition of an $m \times n$ matrix A

complex equivalent or transpose

* - transpose

$$A = \mathcal{U}\,\mathcal{S}\,\mathcal{V}^*$$

where $\mathcal{U}$ is an $m \times m$ unitary matrix, $\mathcal{S}$ is an $m \times n$ diagonal matrix such that $\mathcal{S}_{ij} = 0$ for $i \neq j$, and $\mathcal{V}$ is an $n \times n$ unitary matrix

Note that the zeros below the diagonal in $\mathcal{S}$ (with $m > n$) imply that the elements of columns $(n+1), (n+2), \ldots, m$ of $\mathcal{U}$ are arbitrary, so

$$A = U\,S\,V^*$$ - Skinny Form

where U is the $m \times n$ subset matrix of $\mathcal{U}$ (with the $(n+1), (n+2), \ldots, m$ columns eliminated), S is the upper $n \times n$ matrix of $\mathcal{S}$, and $V = \mathcal{V}$. Note that $U^*U = I$, but it is no longer possible to make the same statement for $U\,U^*$.

→ must be non-zero

The elements of $S = \mathrm{diag}\begin{bmatrix} s_1 & \cdots & s_n \end{bmatrix}$ are known as the *singular values* of A, which are ordered from the smallest singular value to the largest singular value.

Condition Number

diag(S) from SVD

- Singular values are extremely important since they can give an indication of "how well" we can invert a matrix
- A common measure of the invertability of a matrix is the condition number
  - Usually defined as the ratio of its largest singular value to its smallest singular value

$$\text{Condition Number} = \frac{s_n}{s_1}$$

  - Large condition numbers may indicate a near singular matrix, and the minimum value of the condition number is unity (which occurs when the matrix is orthogonal)
  - The rank of A is given by the number of nonzero singular values , also non-zero null ectors

- The LU decomposition factors an $n \times n$ matrix A into a product of a lower triangular matrix L and an upper triangular matrix U

$$A = LU$$

  - Gaussian elimination is a foremost example of LU decompositions
- The Cholesky decomposition is possible only for symmetric positive definite matrices

non-singular

$$A = \mathcal{L}\mathcal{L}^T$$

  where $\mathcal{L}$ is known as the matrix square root
  - Careful with MATLAB, which uses the transpose of this matrix to define its Cholesky decomposition

Vector and Matrix Norms

- Many definitions, some of which are summarized by the following table

Norm	Vector	Matrix
One-norm	$\|\|\mathbf{x}\|\|_1 = \sum_{i=1}^{n} \|x_i\|$	$\|\|A\|\|_1 = \max_j \sum_{i=1}^{n} \|a_{ij}\|$
Two-norm	$\|\|\mathbf{x}\|\|_2 = \left[\sum_{i=1}^{n} x_i^2\right]^{1/2}$	$\|\|A\|\|_2 =$ max singular value of A
Frobenius norm	$\|\|\mathbf{x}\|\|_F = \|\|\mathbf{x}\|\|_2$	$\|\|A\|\|_F = \sqrt{\mathrm{Tr}(A^* A)}$
Infinity-norm	$\|\|\mathbf{x}\|\|_\infty = \max_i \|x_i\|$	$\|\|\|A\|\|_\infty = \max_i \sum_{j=1}^{n} \|a_{ij}\|$

- The cross product of two 3×1 vectors is given by

$$\mathbf{z} = \mathbf{x} \times \mathbf{y} = \begin{bmatrix} x_2 y_3 - x_3 y_2 \\ x_3 y_1 - x_1 y_3 \\ x_1 y_2 - x_2 y_1 \end{bmatrix}$$

  - Note that this follows the right-hand-rule (RHR)
  - Another way to find the cross product is by

$$\mathbf{z} = [\mathbf{x}\times]\,\mathbf{y}$$

  where the cross product matrix is defined by

$$[\mathbf{x}\times] \equiv \begin{bmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{bmatrix}$$

  - Note that $[\mathbf{x}\times]$ is a skew symmetric matrix

Cross Product Matrix (ii)

- Some properties of cross product matrices

$$[\mathbf{x}\times]^T = -[\mathbf{x}\times]$$

$$[\mathbf{x}\times]\,\mathbf{y} = -[\mathbf{y}\times]\,\mathbf{x}$$

$$[\mathbf{x}\times]\,[\mathbf{y}\times] = -\left(\mathbf{x}^T\mathbf{y}\right) I + \mathbf{y}\,\mathbf{x}^T$$

$$[\mathbf{x}\times]^3 = -\left(\mathbf{x}^T\mathbf{x}\right)[\mathbf{x}\times]$$

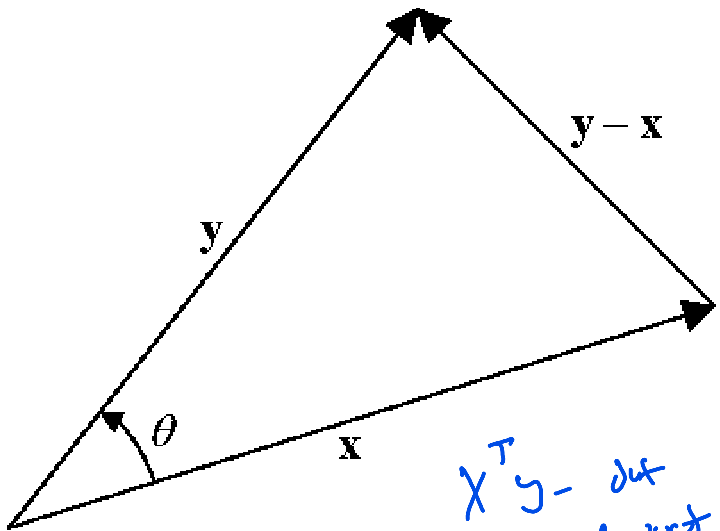
$$[\mathbf{x}\times]\,[\mathbf{y}\times] - [\mathbf{y}\times]\,[\mathbf{x}\times] = \mathbf{y}\,\mathbf{x}^T - \mathbf{x}\,\mathbf{y}^T = [(\mathbf{x}\times\mathbf{y})\times]$$

$$\mathbf{x}\,\mathbf{y}^T[\mathbf{w}\times] + [\mathbf{w}\times]\,\mathbf{y}\,\mathbf{x}^T = -[\{\mathbf{x}\times(\mathbf{y}\times\mathbf{w})\}\times]$$

$$(I - [\mathbf{x}\times])(I + [\mathbf{x}\times])^{-1} = \frac{1}{1+\mathbf{x}^T\mathbf{x}}\left\{(1-\mathbf{x}^T\mathbf{x})I + 2\mathbf{x}\,\mathbf{x}^T - 2[\mathbf{x}\times]\right\}$$

$$[\mathbf{x}\times]^2 = -\left(\mathbf{x}^T\mathbf{x}\right) I + \mathbf{x}\,\mathbf{x}^T$$

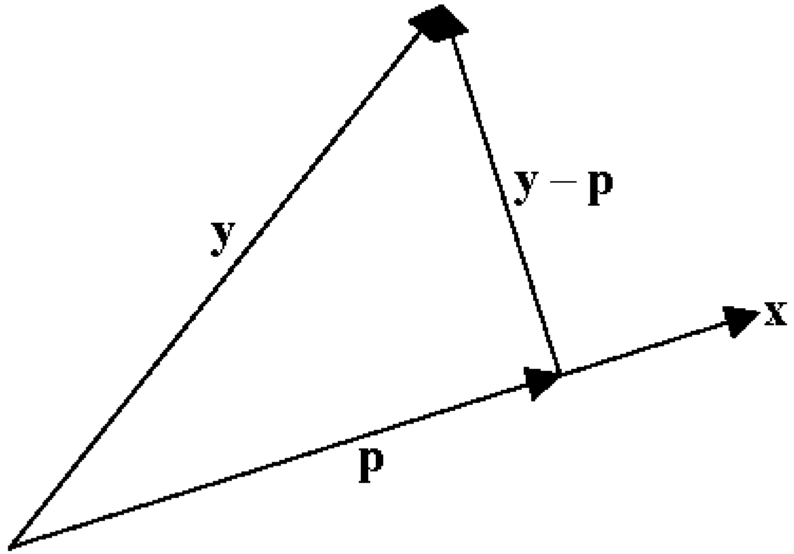
Angle and Orthogonal Projection



$x^T y$ - dot product

$$\cos(\theta) = \frac{\mathbf{x}^T\mathbf{y}}{||\mathbf{x}||\,||\mathbf{y}||}$$

$$\sin(\theta) = \frac{||\mathbf{x}\times\mathbf{y}||}{||\mathbf{x}||\,||\mathbf{y}||}$$



Orthogonal projection of $\mathbf{y}$ to $\mathbf{x}$

$$\mathbf{p} = \frac{\mathbf{x}^T\mathbf{y}}{||\mathbf{x}||^2}\mathbf{x}$$

This projection yields

$$(\mathbf{y}-\mathbf{p})^T\mathbf{x} = 0$$

- Jacobian of a scalar function $f(\mathbf{x})$ is an $n \times 1$ vector, would

be 1x

$$\nabla_{\mathbf{x}} f \equiv \frac{\partial f}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} \qquad (1)$$

- Hessian of a scalar function $f(\mathbf{x})$ is an $n \times n$ matrix (note it is a symmetric matrix)

$$\nabla^2_{\mathbf{x}} f \equiv \frac{\partial^2 f}{\partial \mathbf{x}\, \partial \mathbf{x}^T} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{bmatrix}$$

Matrix Calculus (ii)

- For a general $n \times 1$ vector $\mathbf{x}$ and $m \times 1$ vector $\mathbf{y}$ we have

$$\frac{\partial^2 f}{\partial \mathbf{x}\, \partial \mathbf{y}^T} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial y_1} & \frac{\partial^2 f}{\partial x_1 \partial y_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial y_m} \\ \frac{\partial^2 f}{\partial x_2 \partial y_1} & \frac{\partial^2 f}{\partial x_2 \partial x_y} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial y_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial y_1} & \frac{\partial^2 f}{\partial x_n \partial y_2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial y_m} \end{bmatrix}$$

- If $\mathbf{f}(\mathbf{x})$ is an $m \times 1$ vector and $\mathbf{x}$ is an $n \times 1$ vector, the Jacobian is given by

$$\nabla_{\mathbf{x}} \mathbf{f} \equiv \frac{\partial \mathbf{f}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \tag{2}$$

- Note that the Jacobian matrix is an $m \times n$ matrix

Matrix Calculus (iii)

- There is a slight inconsistency between Eq. (1) and Eq. (2) when $m = 1$, since Eq. (1) gives an $n \times 1$ vector, while Eq. (2) gives a $1 \times n$ vector
  - This should pose no problems, though, since the context of this notation is clear for the particular system shown in this course
  - We will generally be dealing with $n \times 1$ vectors only