ATEP SYS12525

Radar Trackers and Applications for SAADS

Four-Day Short Course

Originally tailored from existing material for reciprocal technology transfer to the South Korean Air force under the SAADS program and presented October 19, October 26, November 2, and November 9 of 1999. Minor updates February 2008.

http://www.jameskbeard.com

ATEP SYS12525

Radar Trackers and Applications for SAADS

October 19, 1999

Topic 1: Course Overview

Sensor Systems Engineering for the 21st Century

ATEP SYS12525

- Radar Trackers and Applications
  - Overview of tracker technology with examples
  - Emphasis on digital trackers as a system functional block
- Text is “Applied Optimal Estimation,” by staff of TASC, edited by A. Gelb (MIT Press, 1974)
- Additional material by handouts
  - Reports, published and unpublished
  - Examples in Excel, FORTRAN

Course Overview

- Overview and Introduction
- Mission, Requirements, and Concept Definition
- Foundations
  - Probability and Statistics
  - Digital and Analog Filters
  - Linear Systems
- Simple Estimators

Course Overview (Continued 1)

- Markov Processes and Vector Estimators
- The Alpha-Beta and Kalman Filter Trackers
- Tracker Overview
  - Tracker Functions and the Tracker Manager
  - Kalman and Batch Estimators in Trackers
  - The Track File as a Sensor System Database
  - Sensor Fusion
  - Radar Resource Management

From Raytheon ATEP SYS12525 for SAADS Day 1, October 19, 1999

Course Overview (Continued 2)

- Computations Using Signal Processor Data Products
- Batch Estimators
- The Association Problem
- The Multiple Hypothesis Tracker (MHT)
  - Track Before Detect
  - Interactive Multiple Models (IMM)
  - Real World Issues

Course Overview (Concluded)

- Advanced Topics
  - Radar Resource Management
  - INS Errors
  - Range and and Doppler Ambiguity Resolution
- Monte Carlo Techniques
  - Principles and Techniques
  - Variance Reduction
- Detailed Example

Instructor Developed Materials

- Handouts
  - Material Not in Literature
  - Tutorial and Reference
- Computer Programs
  - Excel using QBASIC Macro Language
  - Tutorial Topics
  - Tracker Implementation and Tuning
  - Monte Carlo Tutorials and Examples

Requirements for Certificate

- Four Sessions Total
- Attend All Sessions
  – RLI/ATEP Criteria is 80% Attendance
  – 80% of 4 is 3.2
- Certificate
  – Raytheon Learning Institute (nee HIPD)
  – Names You, Course, Date

Your Background

- What's Your Background?
  - I'll tailor material to your background
  - Course topics can be added or deleted for you
- What's Your Interest?
  - Why are you taking this course?
  - What do you expect from this course?
- What Do You Use?
  - MATLAB
  - MathCAD
  - FORTRAN, C, C++, BASIC, etc.

Today's Topics

- System Engineering Context
  - System Mission and Requirements Context
  - The Tracker as Integral Part of Sensor System
- Building a Tracker Concept from Requirements
  - Flowdown
  - Examples
- Probability Theory
- Digital and Analog Filters

ATEP SYS12525

Radar Trackers and Applications for SAADS

October 19, 1999

Topic 2: Systems Engineering Context

Sensor Systems Engineering for the 21st Century

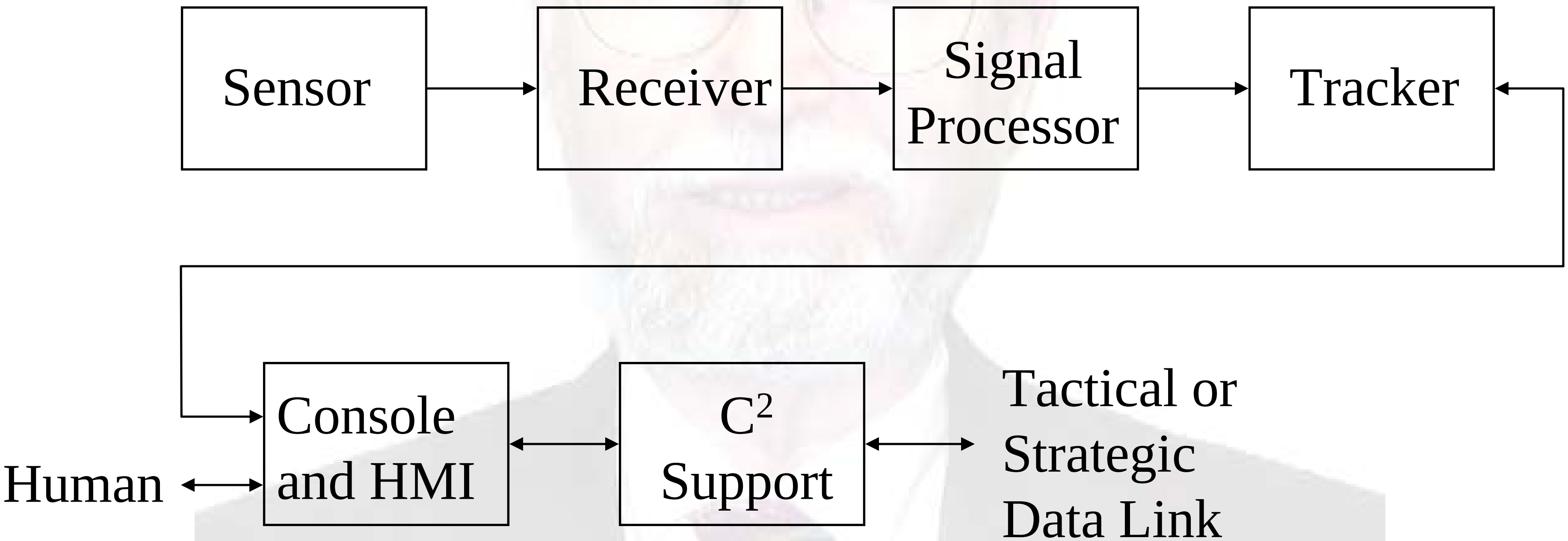
Tracker Requirements Overview

- Trackers are a System Functional Block
  – Inputs are Processed Sensor Data
  – Outputs are Descriptions of Sensed Objects
    » Support user interface
    » Provide data for command and communication links
  – Systems Engineering is the Context for Tracker Study
- The Tracker is Defined by the Requirements

Terminology

Sensor	Active or passive radar or EO/IR device
System	Sensor, processor, operator console, data links, van, etc.
Tracker	Function that supports human interface from sensor data
User Interface	System Console or other Display or HMI
C^2 Support	Capability to send or receive data over tactical links

A Tracker is a System Functional Block



Analog System Functional Blocks

- Sensor
  - Acquire data
  - Provide selectivity for desirable information
- Receiver
  - Selectively amplify signals
  - Digitize data

Digital System Functional Blocks

- Signal Processor
  - Digitize data
  - Process for detection
  - Provided detection information
- Tracker
  - Organize tracker information
  - Estimate target parameters
  - Provide data for console, C^2 support, etc.

Perspective on Sensors

- Threat, Mission and Requirements
  - Every System Addresses a Threat
  - Applies to a Mission to Address a Threat
  - Meets Requirements to Meet a Mission
- Process of DOD 5000.2
  - Define Technical Requirements
  - Define Program Schedule, Budget and Milestones

Requirements Flowdown 1

- First Derived from Mission
  - What is required from sensor?
    - » Detection and tracking range
    - » Position accuracy
    - » Environment
  - What are observables from target?
    - » Passive emissions (EO/IR, RF, other)
    - » RCS, reflectivity in laser bands
- Define Technology (Active/Passive/EO/RF)

Requirements Flowdown 2

- Target Range Drives Selection
  - – Passive or active technology
  - – EO/IR or RF Technology
- Position Accuracy Drives Design
  - – Sensor aperture
  - – Dwell time
- Stealth Drives Aperture Concept

Example: Ground Tactical

- Threat: Multiple Helicopters, Cruise Missiles
- Range: To 50 nautical miles
- Accuracy: 5% of range
- Technology: Mobile X Band ESA
- Mechanically Scanned in Azimuth
- Real-Time Link to Missile Launcher, C^2

System Design

- Define RF Tradeoff Space
  - Aperture Size
  - Two vs. Three Channel Monopulse
  - Squintback Vs. Two Axis ESA (Include Cost)
  - Cruise Missile RCS Requirements
- Define Timeline Tradeoffs
  - Scan Rate Vs. ERP, RCS, Track Accuracy
  - Revisit Time Vs. Threat Density, Track Accuracy

Threat Examples

- New Aircraft
  - New or Modified Fighter
  - New Helicopter Configuration
- New Weapon
  - Air to Air Missile
  - Ground to Air Missile
- New Environment
  - Littoral
  - Arctic

From Raytheon ATEP SYS12525 for SAADS Day 1, October 19, 1999

Requirements Examples

- Longer Maximum Range
- Greater 2D or 3D Position Accuracy
- Higher Clutter Attenuation
- Greater Velocity Accuracy
- Stealth
  - RF
  - EO/IR

Resource Examples

- Cost
  - Initial
  - Support
  - Life Cycle
- Size, Weight, Power, Cooling
- Frequency Allocation or Availability
- May Be a Hard Constraint or a Tradeoff

System Parameter Examples

- Operating Frequency
- Bandwidth
- PRF
- Waveform
- Dwell, Revisit Time
- ERP
- Aperture Size

System Configuration

- Define Cost Functions
- Define Tradeoff Functions
- Consider Arbitrary Constraints
  – Platform-Driven Constraints
  – Cost Limitations
  – Center Frequency Parameters
  – Leveraging of Existing Hardware
- Cull to Multiple Design Points

Cost Function Examples

- Tactical Level
  - Platform Survivability
  - Missile Probability of Kill
  - Search and Acquisition Volume
- System Level
  - Track Accuracy
  - Acquisition Time

BREAK

Four Functions in a Radar

- RF
  - Transmitter
  - Receiver
  - Analog Processing such as SAW line de-chirp
  - Analog De-chirp or Stretch Processing
- Digital Signal Processing
  - Range Gating
  - FFT
  - CFAR Detection

Four Functions of a Radar (Continued)

- Post-Detection Processing
  - Tracking
  - Cueing from C^2 Inputs and Data
  - Controls and Display Data Support
  - System Modes and States Control, T&C
- HMI and C^2 Functions
  - Display Tracker and C^2 Data
  - System Control from Console

Four Functions of a Tracker

- Detection Return Definition
  - Clustering Adjoining Detections
  - Defining a Position from Contiguous Detections
- Association of Detection with a Track File
  - Raw Problem is M Tracks by N Returns
  - May Be Ambiguous

Four Functions of a Tracker (Continued)

- Track File Maintenance
  - Update Track Files
  - Initiation of New Tracks
  - Drop Superfluous Tracks
  - Other
- Formatting Track File Data
  - Support of HMI Functions
  - Providing Data for C^2 Links

Two Concepts of Track Before Detect

- Averaging in Signal Processor Before Detection
  - Interpolate and sum range-Doppler map
  - SAR works this way
  - Applicable for zero process noise targets
  - Incoherent summation OK for TBD
- Track Targets, False Alarms, Missasociations
  - System detection must be done in tracker
  - MHT works this way

Track Before Detect in an MHT

- Concept in an MHT
  - Lower Signal Processor Detection Threshold
    - » Increases Sensitivity
    - » Allows Noise to be Tracked
  - Tracks Differentiated by Scoring
    - » SNR
    - » Target-Like Movements
  - System Detection Performed in the Tracker
- Trades Off Computer Capacity for Sensitivity

Radar Trackers and Applications for SAADS

October 19, 1999

Topic 3: Probability Theory

Sensor Systems Engineering for the 21st Century

Readings From Gelb

- Probability and Statistics
  - Probability Theory, pp 24-41
  - Noise and Markov Processes, pp 42-44
- Linear System Theory
  - Chapter 3, pp 51-96
  - Examples, p 61, p 87
- Matrix Theory
  - Problems 2-5 and 2-6, p 47

Probability Density Functions

- Normal, or Gaussian

$$p(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp\left(-\frac{(x-m)^2}{2 \cdot \sigma^2} \right)$$

- Raleigh

$$p(x) = \frac{1}{2 \cdot \sigma^2} \cdot \exp\left(-\frac{x}{2 \cdot \sigma^2} \right)$$

PDFs (Continued)

- Chi-Square

$$p(x) = \frac{1}{2^{\frac{\upsilon}{2}} \cdot \Gamma\left(\frac{\upsilon}{2}\right)} \cdot \exp\left(-\frac{x}{2 \cdot \sigma^2}\right)$$

- Cauchy

$$p(x) = \frac{1}{\pi \cdot \beta} \cdot \frac{1}{1 + \left(\frac{x - \alpha}{b}\right)^2}$$

PDFs (Concluded)

- Other
  - – Poisson
  - – Weibull
  - – Student T
  - – Edgeworth/Hermite/Cornish-Fisher Series
  - – Rician or Non-Central Chi-Square
  - – F or Variance Ratio

Mean and Variance

- Mean
  - Arithmetic Average Approximates Mean
  - “Center” of Distribution
  - Mode, or Point of Highest Probability Density is Same as Mean for This Distribution
- Variance
  - Mean Square Deviation from Mean
  - Square Root is Standard Deviation σ

Meaning of Variance

- Chebychev's Inequality

$$P\left(\left|x-m\right| \geq \sigma \cdot t\right) \leq \frac{1}{t^2}$$

- Bounds Probability Distribution for t>1
- Holds for Any Distribution for which the Variance Exists
- Among Distributions Given Above, Only the Cauchy Distribution Fails This Test

The Gaussian Distribution

- General Probability Density Function

$$p(x) = \frac{1}{2\pi\sigma} \cdot \exp\left(-\frac{1}{2} \cdot \frac{(x-m)^2}{\sigma^2} \right)$$

- Random variable x, mean m, variance σ^2
- Sometimes Called the Normal Distribution
  – Most Often Seen in Nature
  – Favored in Averages -- Central Limit Theorem
  – “Energy” or $(x\text{-}m)^2$ is Poisson Distributed

The Chi Square Distribution

- Poisson or Exponential Distribution is Chi Square with 1 Degree of freedom
- Sum of Squares of N Zero Mean, Unit Variance Gaussian Variables is Chi Square with N Degrees of Freedom
  – Power Spectra are Chi Square
  – Square Root of Chi Square Random Variable with 2 Degrees of Freedom is Rayleigh
  – Square Root of Chi Square Random Variable with 3 Degrees of Freedom is Maxwellian

The Multivariate Gaussian Distribution

- Given:
  - A Vector of Random Variables $\underline{x}$
  - Means of Zero
  - Covariance Matrix is $P = \mathrm{Exp}\{\underline{x} \cdot \underline{x}^T\}$
- To Find: the Joint Probability Density Function $p(\underline{x})$
- Solution: Construct a Vector of Zero Mean, Unity Variance Uncorrelated Random Vectors $\underline{y}$
- Covariance of $\underline{y}$ is $\mathrm{Exp}\{\underline{y} \cdot \underline{y}^T\} = I$

The Constructions

- Make a Variable Change $\underline{x} = C \cdot \underline{y}$
- Rule for Mapping Probability Density Functions

$$p(\underline{x}) \cdot d\,\underline{x} = p(\underline{y}) \cdot d\,\underline{y}$$

- Jacobian of the Variable Change

$$d\,\underline{x} = \left\| \left[\frac{\partial \underline{x}}{\partial \underline{y}} \right] \right\| \cdot d\,\underline{y} = |C| \cdot d\,\underline{y}$$

The Probability Density Function

- Joint Probability Density Function of y

$$p\left(\underline{y}\right)\cdot d\,\underline{y}=\frac{1}{\left(2\pi\right)^{N/2}}\cdot\exp\left(-\frac{1}{2}\cdot\underline{y}^{T}\cdot\underline{y}\right)\cdot d\,\underline{y}$$

- With Variable Change and Jacobian

$$p\left(\underline{x}\right)\cdot d\,\underline{x}=\frac{1}{\left(2\pi\right)^{N/2}\cdot\left|C\right|}\cdot\exp\left(-\frac{1}{2}\cdot\underline{x}^{T}\cdot C^{-T}\cdot C^{-1}\cdot\underline{x}\right)\cdot d\,\underline{x}$$

Covariance of x

- From Variable Change

$$Cov\left\{\underline{x}\right\} = Exp\left\{\underline{x}\cdot\underline{x}^T\right\} = Exp\left\{C\cdot\underline{y}\cdot\underline{y}^T\cdot C^T\right\} = C\cdot C^T$$

- If C is Chosen so that

$$C\cdot C^T = P$$

- Then

$$p\left(\underline{x}\right) = \frac{1}{\left(2\pi\right)^{N/2}\cdot\left|P\right|^{1/2}}\cdot\exp\left(-\frac{1}{2}\cdot\underline{x}^T\cdot P^{-1}\cdot\underline{x}\right)$$

Cholesky Factorization

- For Every Symmetric Positive Definite Matrix P there Exists an Upper Triangular Matrix C Such That $P = C \cdot C^T$
- Find C Recursively From

$$p_{ij} = \sum_{k=\max(i,j)}^{N} c_{ik} \cdot c_{jk}$$

- Begin With

$$c_{NN} = \sqrt{p_{NN}}$$

Cholesky Factorization Algorithm (Continued)

- Find Rest of Column N from $c_{i,N} = p_{i,N} / c_{N,N}$
- Work from Column N-1 Down to Column 1
- Diagonal Elements

$$c_{k,k} = \sqrt{p_{k,k} - \sum_{s=k+1}^{N} c_{k,s}^{2}}$$

- Rest of Column

$$c_{i,k} = \frac{p_{i,k} - \sum_{s=k+1}^{N} c_{i,s} \cdot c_{k,s}}{c_{k,k}}$$

Cholesky Pointers

- During Progress of Algorithm
  - Element $c_{i,j} = 0$ if $i>j$ (C is Upper Triangular)
  - Element $c_{i,j}$ Always Defined if Column j Already Computed
- Algorithmic Tricks
  - Overstore Input Matrix P
  - Accumulate Sum as Each Column is Computed
- C is Unique Within Signs of Square Roots

From Raytheon ATEP SYS12525 for SAADS Day 1, October 19, 1999

Actual Code from Bierman

```
subroutine choles(n,p) #Cholesky factorization, finds U given P, P=U*UT
# U is upper triangular
# See Bierman, p 53
# Inputs:
#   n     Number of states
#   P     Covariance matrix to be factored
# Outputs:
#   U     Upper triangular matrix factor
# Notes:  P is overstored with U
dimension p(n,n)
for(j=n;j>1;j=j-1)
  {
  p(j,j)=sqrt(p(j,j))
  alpha=1./p(j,j)
  for(k=1;k<j;k=k+1)
    {
    p(k,j)=alpha*p(k,j)
    beta=p(k,j)
    do i=1,k
      p(i,k)=p(i,k)-beta*p(i,j)
    }
  }
p(1,1)=sqrt(p(1,1))
return
end
```

Localization Ellipsoid

- Probability Density of Random Variables $\underline{x}$

$$p(\underline{x}) = \frac{1}{(2\pi)^{N/2} \cdot |P|^{1/2}} \cdot \exp\left(-\frac{1}{2} \cdot (\underline{x} - \underline{m})^T \cdot P^{-1} \cdot (\underline{x} - \underline{m}) \right)$$

- The Quadratic Form is Chi-Square Distributed

$$J = (\underline{x} - \underline{m})^T \cdot P^{-1} \cdot (\underline{x} - \underline{m})$$

Statistics of Ellipsoid

- Set J = T for Specified Confidence that the Random Variable $|\underline{x} - \underline{m}|$ is Within Ellipsoid using Chi-Square Distribution for N Degrees of Freedom
- Find Ellipsoid from Singular Value Decomposition of P (Matrix C here characteristic vector matrix, not a Cholesky factor)

$$P = C \cdot \Lambda \cdot C^T$$

Geometry of Ellipsoid

- The Chi-Square Quadratic Form on P^{-1}

$$P^{-1} = C \cdot \Lambda^{-1} \cdot C^T$$

- Make the Variable Change

$$\underline{z} = C^T \cdot \left(\underline{x} - \underline{m}\right)$$

- The Chi-Square Random Variable is Now

$$J = \sum_{i=1}^{N} \frac{z_i^2}{\lambda_i}$$

Conclusions about Ellipsoid

- The Equation is

$$\sum_{i=1}^{N} \frac{z_i^2}{\lambda_i} = T$$

- An Ellipsoid with Semi-Axes

$$a_i = \sqrt{\lambda_i \cdot T}$$

- The Variable Change $\underline{z} = C \cdot \underline{x}$
  - A Cartesian Coordinate Rotation
  - Rows of C are Directions of Semi-Axes

BREAK

Localization Ellipse

- Definition -- The Representation of a Localization Ellipsoid on a 2-Dimensional Surface
- Example -- Showing Track Accuracy on a Display
- Technique
  – Map Covariance Matrix to Display Coordinates
  – Show Resulting Ellipse

Covariance Mapping

- Mapping
  - Covariance P_x of Estimate $\underline{x}$
  - To Covariance P_y of Vector $\underline{y}$
  - Using Functional Relationship $\underline{y} = \underline{h}(\underline{x})$
- Look at Taylor Expansion

$$\underline{y}(\underline{x} + d\,\underline{x}) = \underline{h}(\underline{x}) + \frac{\partial \underline{h}(\underline{x})}{\partial \underline{x}} \cdot d\,\underline{x} + O\left(d\,\underline{x}^2\right)$$

- Define Sensitivity Matrix H

$$H = \frac{\partial \underline{h}(\underline{x})}{\partial \underline{x}}$$

Error Mapping

- Errors in $\underline{y}$

$$d\,\underline{y} = \underline{y}(\underline{x} + d\,\underline{x}) - \underline{y}(\underline{x}) \approx H \cdot d\,\underline{x}$$

- Covariance of Errors in $\underline{y}$

$$P_y = \left\langle d\,\underline{y} \cdot d\,\underline{y}^T \right\rangle \approx H \cdot \left\langle d\,\underline{x} \cdot d\,\underline{x}^T \right\rangle \cdot H^T = H \cdot P_x \cdot H^T$$

- This is the *Covariance Mapping Equation*
- On a Display
  - Vector $\underline{y}$ is a 2 X 2
  - Often, mapping is simply a submatrix of P_x

The Two Dimensional Localization Ellipse

- Equation of Ellipse is

$$\underline{y}^T \cdot P_y \cdot \underline{y} = T$$

- Problem is Finding
  - Orientation of Ellipse
  - Semiaxes of Ellipse
- Approach: Find $\underline{z}$ as Rotation of $\underline{y}$:

$$\underline{z} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \cdot \underline{y}$$

Solving for Orientation

- Principle: Covariance P_z of $\underline{z}$ is

$$P_z = \begin{bmatrix} pz_{11} & pz_{12} \\ pz_{12} & pz_{22} \end{bmatrix}$$

$$pz_{11} = p_{11} \cdot \cos(\theta)^2 - p_{12} \cdot \sin(2\theta) + p_{22} \cdot \sin(\theta)^2$$

$$pz_{22} = p_{11} \cdot \sin(\theta)^2 + p_{12} \cdot \sin(2\theta) + p_{22} \cdot \cos(\theta)^2$$

$$pz_{12} = \tfrac{1}{2}\left(p_{11} - p_{22}\right) \cdot \sin(2\theta) + p_{12} \cdot \cos(2\theta)$$

Semiaxes

- Lengths of Semiaxes are

$$\sigma_1 = T\sqrt{\lambda_1} \quad \text{(semimajor axis)}$$

$$\sigma_2 = T\sqrt{\lambda_2} \quad \text{(seminor axis)}$$

$$\lambda_1 = \frac{p_{11} + p_{22} + \sqrt{\left(p_{11} - p_{22}\right)^2 + 4p_{12}^2}}{2}$$

$$\lambda_2 = \frac{p_{11} + p_{22} - \sqrt{\left(p_{11} - p_{22}\right)^2 + 4p_{12}^2}}{2}$$

Orientation of Semimajor Axis

$$\theta\begin{cases} = \dfrac{1}{2}\tan^{-1}\left(\dfrac{2p_{12}}{p_{22}-p_{11}}\right), & p_{11} > p_{22} \\ = \dfrac{1}{2}\tan^{-1}\left(\dfrac{2p_{12}}{p_{22}-p_{11}}\right) - \dfrac{\pi}{2}, & p_{11} < p_{22} \end{cases}$$

$$\cos(2\theta) = \frac{p_{11}-p_{22}}{\sqrt{(p_{11}-p_{22})^2 + 4p_{12}^2}}$$

$$\cos(\theta)^2 = \frac{1+\cos(2\theta)}{2}, \ \sin(\theta)^2 = \frac{1-\cos(2\theta)}{2}$$

Averages of Gaussian Variables

- Sample Mean

$$\overline{m} = \frac{1}{N} \cdot \sum_{i=1}^{N} x_i$$

- Itself a Random Variable
  - Mean is m (why?)
  - Variance is σ^2/N (why?)
- An Example of a Block Average or Unweighted Average

Student's t Distribution

- Arises When Zero Mean Gaussian Variable is Divided by a Sample Standard Deviation
- Number of Degrees of Freedom M
  - Size of Sample Used in Estimating Variance
  - Reduced by 1 when Variable Itself is in Sample
- As M Increases, Approaches Gaussian
  - Mean 0
  - Variance 1

Central Limit Theorem

- Only One Case Shown Here
  - Proof is a Function of Scenario
  - Characteristic Function Used Here

$$\phi(t) = \int_{-\infty}^{\infty} \exp(+j \cdot x \cdot t) \cdot p(x) \cdot dx = \langle \exp(j \cdot x \cdot t) \rangle$$

- Characteristic Function for the Gaussian Distribution

$$\phi(t) = \exp\left(j \cdot m \cdot t - \frac{(\sigma \cdot t)^2}{2} \right)$$

Cumulant Function

- Logarithm of Characteristic Function

$$\ln\left(\phi(t)\right) = \sum_{n=1}^{\infty} \kappa_n \cdot \frac{(j \cdot t)^n}{n!}, \ \kappa_n \equiv n^{th} \text{ cumulant}$$

- First Few Cumulants

$$\kappa_1 = m, \ \kappa_2 = \sigma^2, \ \kappa_3 = \mu, \ \kappa_4 = \mu_4 - 3 \cdot \sigma^4$$

- For Gaussian Probability Density Function

$$\ln\left(\phi(t)\right) = -j \cdot m \cdot t - \frac{(\sigma \cdot t)^2}{2}$$

Mean of N Random Variables

- Characteristic Function

$$\phi(m(N),t) = \left(\phi\left(\frac{x}{N}\right),t\right)^{N} = \left(\phi\left(x,\frac{t}{N}\right)\right)^{N}$$

- Cumulant Function

$$\ln(\phi(m(N),t)) = \ln\left(\phi\left(x,\frac{t}{N}\right)^{N}\right) = N \cdot \ln\left(\phi\left(x,\frac{t}{N}\right)\right)$$

Mean of N Random Variables (Continued)

- Mean of N Random Variables

$$\ln\left(\phi\left(m(N),t\right)\right)=N\cdot\sum_{i=1}^{\infty}\frac{\kappa_i}{N^i}\cdot\frac{\left(j\cdot t\right)^i}{i!}$$

- Normalization of Probability Density Function ⇒ Define New Random Variable

$$z(N)=\left(m(N)-\kappa_1\right)\cdot\sqrt{\frac{N}{\kappa_2}}$$

The New Random Variable

- Cumulants

$$\kappa'_1 = 0,\ \kappa'_2 = 1,\ \kappa'_i = \frac{1}{N^{(i-2)/2}} \cdot \frac{\kappa_i}{\kappa_2^{i/2}},\ \ i \ge 2$$

- Conclusion
  - Random Variable z(N) Approaches Gaussian
  - Mean Zero, Unity Variance
  - N Increases Without Bound

Failure of the Central Limit Theorem

- Consider the Cauchy Probability Density Function

$$p(x) = \frac{1}{\pi \cdot D} \cdot \frac{1}{1 + \left(\frac{x}{D} \right)^2}$$

- Characteristic Function

$$\phi(x, t) = \exp(-D \cdot |t|)$$

- Characteristic Function of Mean of N Variables

$$\left(\phi\left(x, \frac{t}{N} \right) \right)^N = \exp(-D \cdot |t|) = \phi(x, t)$$

ATEP SYS12525

Radar Trackers and Applications for SAADS

October 19, 1999

Topic 4: Random Numbers

Sensor Systems Engineering for the 21st Century

Random Number Generators

- Types
  - Congruence
    - » Most Often Used
  - Fibonacci
    - » Type Used in This Course
  - Other
- Desired Properties
  - Pass Statistical Tests
  - Speed

Statistical Tests

- Test is Done on Raw Output
  - Uniform Distribution from 0 to 1
  - Other Distributions Formed by Mapping
- Mean of 0.5
- Uniformity of Distribution
- Lack of Correlations
- Length of Periodic Repeat

Statistics of Congruence Random Number Generators

- Type

$$X_{n+1} = a \cdot X_n + b \mod T$$

- Correlation

$$\rho_s = 12 \cdot \left(\left\langle X_n \cdot X_{n+s} \right\rangle - \tfrac{1}{4} \right) = \frac{1 - 6 \cdot \frac{b_s}{T} \cdot \left(1 - \frac{b_s}{T}\right)}{a^s \mod T} + e$$

$$b_s = \left(1 + a + a^2 + \ldots + a^{s-1}\right) \cdot b \mod T$$

$$|e| < \frac{a^s \mod T}{T}$$

Best Congruence Types When

- Minimum Correlation

$$a \approx T^{\frac{1}{2}}, \;\; \rho_1 \approx T^{-\frac{1}{2}}$$

- Maximum Period of T is Achieved

b relatively prime to T

$a = 1 \mod p, \;\; p$ any prime factor of T

$a = 1 \mod 4$ if 4 is a factor of T

- Example

$$T = 2^q, \;\; a = 2^s + 1 = 1 \mod 4, \;\; b \text{ odd}$$

Known Properties

- Randomness Best
  - Most significant bits
  - Significance is about the number of bits in a
- Randomness Worst
  - Least significant bits
  - Period is very short in bottom few bits
  - Randomness poor below the number of bits in a

Break

Fibonacci Type

- Basic Premise
  - Given Two Random Numbers X_1 and X_2
    - » Over Interval [0,1]
    - » Uniformly Distributed and Uncorrelated
  - New Random Number X3

$$X_3 \begin{cases} = X_2 - X_1, & X_2 - X_1 \geq 0 \\ = X_2 - X_1 + 1, & X_2 - X_1 < 0 \end{cases}$$

- X_3 Not Correlated with Either X_1 or X_2

Structure of Fibonacci Random Number Generator

- Initialization
  - 23 Random Numbers
  - Index Pointers to 1st and 12th
- Operation
  - Generate New Random Number
    - » User Pointers to Select Pair
    - » Replace One with New Random Number
  - Increment Pointers Modulo 23

Initialization and Seeding

- Algorithm
  - Use Small Integer Fibonacci Type
    - » Integer Arithmetic
    - » Arithmetic is Modulo 32768-(delta)
    - » Only Three Integers, Not 23
    - » One is the Seed
  - Run it 1,000 Times Before Using It
  - Set Each Bit of 23 Numbers Individually
- Function is Demonstrably Excellent

Extension of Fibonacci Seeding Method

- Use Long Integers
- Use More Than One Seed
  - First is Seed
  - Others are User Input (Seeds) Also
- Use More Integers
  - Existing Method Uses 3
  - Use 5, 7, More
- Allow More than 32,000 Independent Seeds

Statistics of Fibonacci Random Number Generator

- All are Excellent
  – Mean
  – Variance
  – Uniformity of Distribution
  – Correlation
  – Period
- Never Known to Fail a Randomness Test
  – All Statistics Within Expected Bounds
  – No Known Period (10^{270} Possible)

Pros and Cons

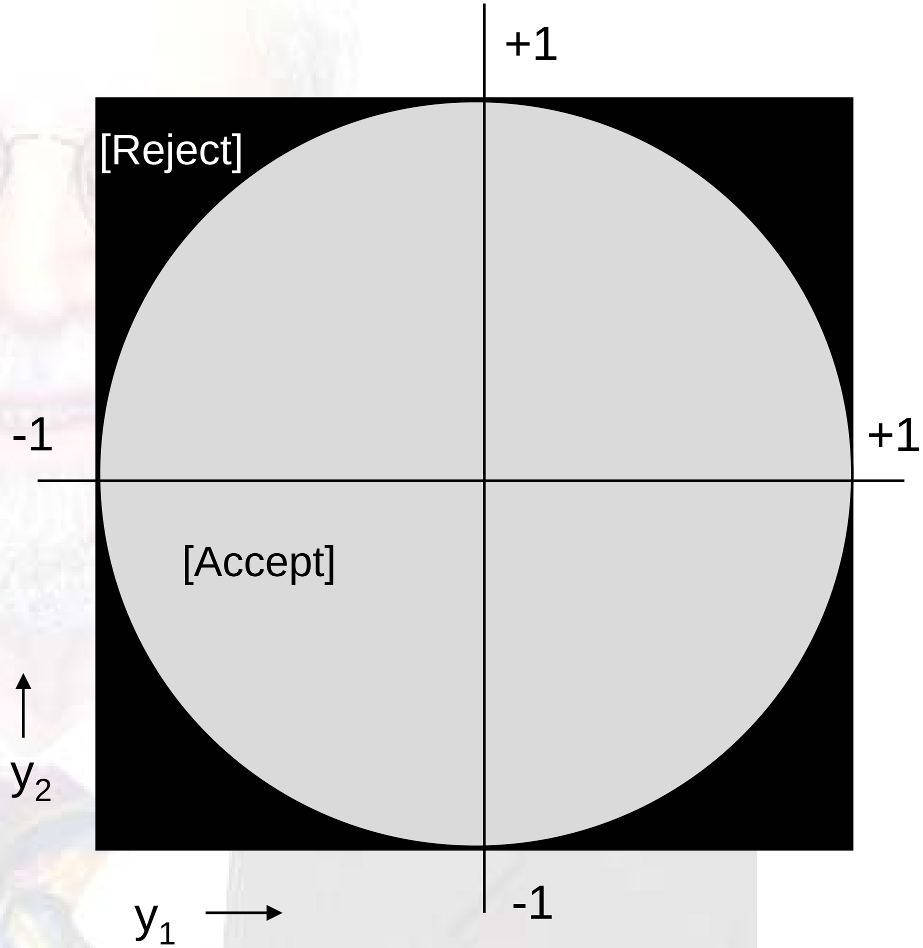
- Congruence
  - Pros
    - » Very Simple
    - » Well Understood
    - » Fast
  - Cons
    - » Correlations Can't Be Eliminated
    - » Randomness Confined to Most Significant Bits
    - » Lower Order Bits Have Short Period
- Fibonacci
  - Passes All Randomness Tests
  - Faster
  - Simpler

Generating Gaussian Random Numbers

- Use Two Uniform Random Variables
  - Transform $y=2 \cdot x-1$
  - Two Together Give Uniform Distribution over Unit Square
- Reject Pairs Outside Unit Circle
  - Start Over With New Pair of Random Numbers
  - Gives Uniform Distribution Over Unit Circle
- Transform to 2-D Gaussian

Obtaining Uniform Distribution Over the Unit Circle

- Begin with x_1, x_2
  - Uniformly distributed over [0, +1]
  - Uncorrleated
- Variables $y_i = 2 \cdot x_i - 1$
  - Uniformly distributed over [-1, +1]
  - Uncorrleated
- Result is uniform distribution over unit square
- Reject if (y1,y2) outside the unit circle



The Transformation

- Find r^2

  $$r^2 = y_1^2 + y_2^2$$

- Variable r^2 is Uniformly Distributed on [0,1]
  - Points are uniformly distributed over circle
  - Area inside r is proportional to r^2
- (z_1,z_2) is 2-D Gaussian if $r'^2=z_1^2+z_2^2$ is Poisson

  $$r' = \sqrt{-2 \cdot \ln\left(r^2\right)}$$

- Independent Gaussian Random Variables Are

  $$z_1 = r' \cdot \frac{y_1}{r}, \;\; z_2 = r' \cdot \frac{y_2}{r}$$

Radar Trackers and Applications for SAADS

October 19, 1999

Topic 5: Estimation Theory

Sensor Systems Engineering for the 21st Century

Estimation Theory Overview

- Definitions
  - Estimation: Use of data to compute estimates of parameters
  - Estimator: Procedures for Estimation
- Desirable Attributes
  - Minimum Variance
  - Unbiased
  - Consistent
  - Sufficient

Minimum Variance

- Dispersion of Estimate
  - RMS Dispersion is Standard Deviation
  - Less is Better
  - Theoretical Minimum Exists □ Cramer-Rao Bound
  - Ratio of Estimator Variance to Cramer-Rao Bound is called “Statistical Efficiency”
- Mean Square Error is More General
  - RMS Error about True Value
  - Applies with Asymptotically Unbiased Estimators

Estimator Bias

- Definitions
  - Strictly Unbiased

$$\left\langle \underline{\hat{x}} \right\rangle = \underline{x}$$

  - Asymptotically Unbiased

$$\lim_{N \to \infty} \underline{\hat{x}}\left(N\right) = \underline{x}$$

- Both Definitions are Useful
- Mean Square Error Used for Asymptotically Unbiased Estimates

Consistency

- Classical Definition

$$\lim_{N \to \infty} \underline{\hat{x}}(N) = \underline{x}$$

- Other Definitions
  - Mean Square Error is Asymptotically Zero
  - Convergence in Probability [illegible] For Any δ,

$$\lim_{N \to \infty} P\left(\left| \underline{\hat{x}} - \underline{x} \right| < \delta \right) = 1$$

Sufficiency

- Definition
  - The Estimator of $\underline{x}$ $\hat{\underline{x}}(\underline{y})$ is Sufficient if

$$\frac{\partial p\left(\underline{y} \middle| \hat{\underline{x}}\right)}{\partial \underline{x}} = \underline{0}$$

  - Equivalently, the Factorization Theorem:

$$p(\underline{y}) = g(\hat{\underline{x}}, \underline{x}) \cdot h(\underline{y})$$

- Meaning: The Estimator $\hat{\underline{x}}(\underline{y})$ Contains “all” the Information on $\underline{x}$ Available from $\underline{y}$

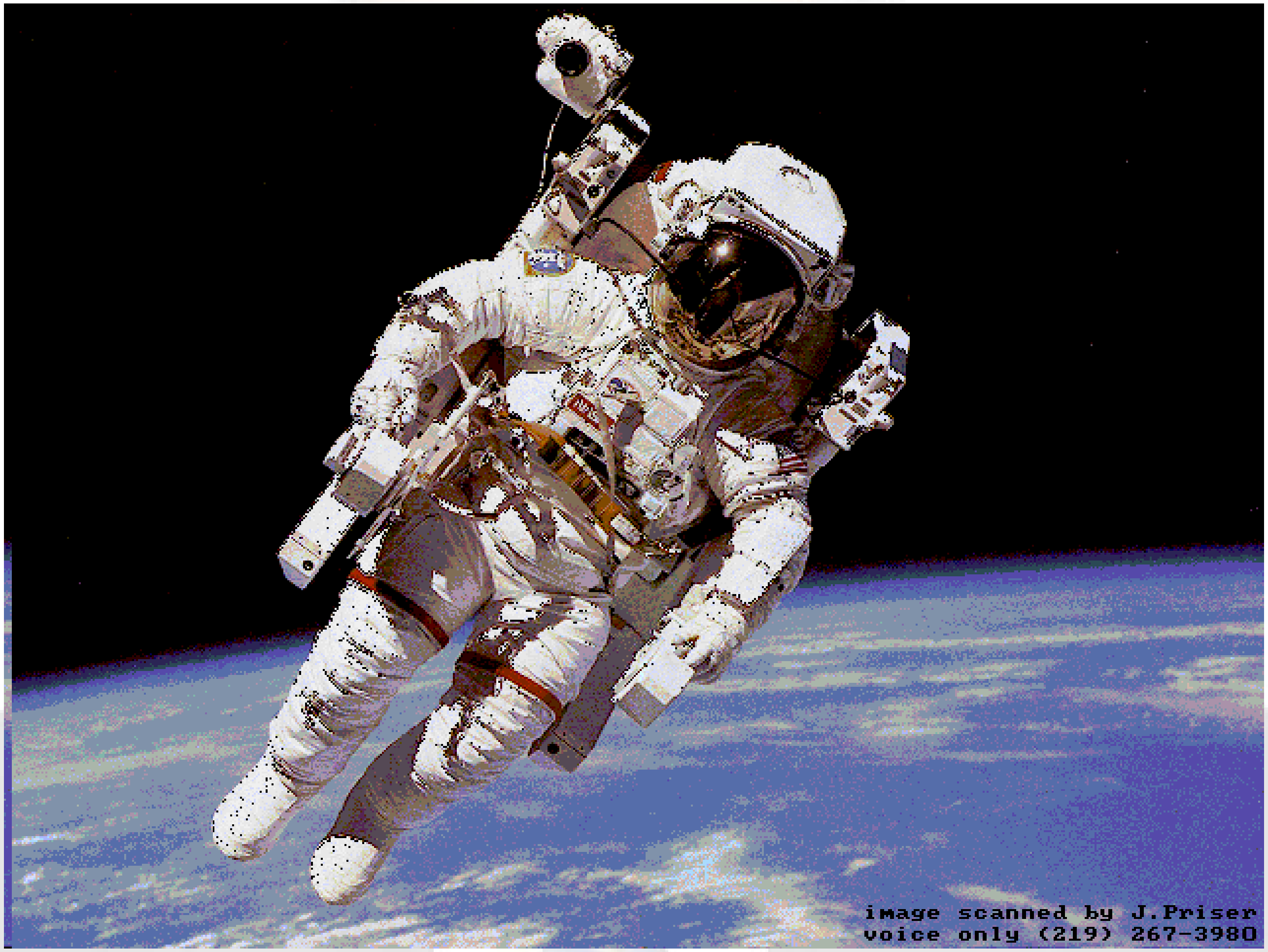
Mapping of Sufficient Estimators

- If $\underline{\hat{x}}\left(\underline{y}\right)$ is a sufficient estimator of $\underline{x}$ Then
  - The Function $\underline{f}\left(\underline{\hat{x}}\right)$ has the property that if
    - » The matrix F, defined by the gradient

$$F = \left.\frac{\partial \underline{f}\left(\underline{x}\right)}{\partial \underline{x}}\right|_{\underline{x}=\underline{\hat{x}}}$$

    - » The matrix F is square and nonsingular
  - The Function $\underline{f}\left(\underline{\hat{x}}\right)$ is Also a Sufficient Estimator
- An Estimator Must Be Consistent to be Useful

BREAK



Method of Maximum Likelihood

- Classical, Extensively Developed by R. A. Fisher in 1920 Time Frame
- General Form
  - Given States $\underline{x}$ and Measurements $\underline{y} = \underline{h}(\underline{x})+\underline{v}$
  - Log Likelihood Function Defined as

$$L\left(\underline{x}\middle|\underline{y}\right) = \ln\left(p\left(\underline{y}\middle|\underline{x}\right)\right)$$

- The Likelihood Equation is

$$L_x = l\left(\underline{x}\middle|\underline{y}\right) = \frac{\partial L\left(\underline{x}\middle|\underline{y}\right)}{\partial \underline{x}} = \underline{0}$$

- Solution is estimate $\hat{\underline{x}}$

Variance Propagation

- Gradient of the Likelihood Equation $L_x = \underline{0}$

$$L_{xx} \cdot \Delta\underline{\hat{x}} + L_{xy} \cdot \underline{v} = \frac{\partial l\left(\underline{\hat{x}} \middle| \underline{y}\right)}{\partial \underline{\hat{x}}} \cdot \Delta\underline{\hat{x}} + \frac{\partial l\left(\underline{\hat{x}} \middle| \underline{y}\right)}{\partial \underline{y}} \cdot \underline{v} = \underline{0}$$

- Covariance Mapping Equation is

$$L_{xx} \cdot P \cdot L_{xx}^T = L_{xy} \cdot R \cdot L_{xy}^T = \left\langle L_x \cdot L_x^T \right\rangle$$

- Key Lemma

$$\left\langle L_x \cdot L_x^T \right\rangle = -\left\langle L_{xx} \right\rangle$$

- Solution is

$$P = Cov\left\{\underline{\hat{x}}\right\}, \quad P^{-1} = -L_{xx} = L_{xy} \cdot R \cdot L_{xy}^T, \;\; R = Cov\left\{\underline{v}\right\}$$

If the Measurement Errors are Gaussian

- Log Likelihood Function

$$L\left(\underline{x} \middle| \underline{y}\right) = -\frac{N}{2} \cdot \ln(2\pi) - \frac{1}{2} \cdot \ln\left(|R|\right)$$

$$-\frac{1}{2} \cdot \left(\underline{y} - \underline{h}(\underline{x})\right)^T \cdot R^{-1} \cdot \left(\underline{y} - \underline{h}(\underline{x})\right)$$

- Likelihood Equation, Cramer-Rao Bound

$$H^T \cdot R^{-1} \cdot \left(\underline{y} - \underline{h}(\hat{\underline{x}})\right) = \underline{0}$$

$$P^{-1} = H^T \cdot R^{-1} \cdot H$$

Classical Example

- Measurements $\underline{y} = m \cdot \underline{1} + \underline{v},\ \text{Cov}\left\{\underline{v}\right\} = \sigma^2 \cdot I$
- Estimate I, m and σ^2, and II m and σ
- Results I:

$$L\left(m, \sigma^2\right) = -\frac{M}{2} \cdot \ln\left(2\pi\right) - \frac{M}{2} \cdot \ln\left(\sigma^2\right)$$

$$-\frac{1}{2\sigma^2} \cdot \left(\underline{y} - m \cdot \underline{1}\right)^T \cdot \left(\underline{y} - m \cdot \underline{1}\right)$$

Likelihood Equation I

- The Gradients

$$\frac{\partial L}{\partial m} = \frac{\underline{1}^T \cdot \left(\underline{y} - m \cdot \underline{1} \right)}{\sigma^2}$$

$$\frac{\partial L}{\partial \sigma^2} = -\frac{M}{2\sigma^2} + \frac{\left(\underline{y} - m \cdot \underline{1} \right)^T \cdot \left(\underline{y} - m \cdot \underline{1} \right)}{2\sigma^4}$$

- The Solutions

$$\hat{m} = \frac{\underline{1}^T \cdot \underline{y}}{M}, \quad \widehat{\sigma^2} = \frac{\left(\underline{y} - m \cdot \underline{1} \right)^T \cdot \left(\underline{y} - m \cdot \underline{1} \right)}{M}$$

Results II

- Gradient

$$\frac{\partial L}{\partial \sigma} = -\frac{M}{\sigma} + \frac{\left(\underline{y} - m \cdot \underline{1}\right)^T \cdot \left(\underline{y} - m \cdot \underline{1}\right)}{\sigma^3}$$

- Solution

$$\hat{\sigma}^2 = \frac{\left(\underline{y} - m \cdot \underline{1}\right)^T \cdot \left(\underline{y} - m \cdot \underline{1}\right)}{M}$$

- Remember: Mapping of Consistent Estimators

The Variances

- The Mean

$$-\frac{1}{\mathrm{Var}\{\hat{m}\}} = \frac{\partial^2 L}{\partial m^2} = -\frac{\underline{1}^T \cdot \underline{1}}{\sigma^2} = -\frac{M}{\sigma^2}$$

$$\mathrm{Var}\{\hat{m}\} = \frac{\sigma^2}{M}$$

- The Variance

$$-\frac{1}{\mathrm{Var}\{\widehat{\sigma^2}\}} = \frac{\partial^2 L}{\left(\partial \sigma^2\right)^2} = \frac{M}{2\sigma^4} - \frac{\left(\underline{y} - m \cdot \underline{1}\right)^T \cdot \left(\underline{y} - m \cdot \underline{1}\right)}{\sigma^6}$$

$$Var\{\widehat{\sigma^2}\} = \frac{2 \cdot \sigma^4}{M}$$

Classical Linear Regression

- States
  - Means of Two Noisy Correlated Variables $\underline{m}$
  - Variances and Correlation Coefficient R
- Measurements
  - M Pairs of the Variables with Noise

$$\underline{y}_i = \begin{bmatrix} z_{0,i} \\ z_{1,i} \end{bmatrix}, \quad i = 0, \ldots M-1$$

$$p\left(\underline{y}_i \middle| \underline{x}\right) = \frac{1}{(2\pi)\left|R\right|^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2}\left(\underline{y}_i - \underline{m}\right)^T \cdot R^{-1} \cdot \left(\underline{y}_i - \underline{m}\right)\right)$$

Maximum Likelihood

- Conditional Probability Density Function

$$p\left(\underline{y}\middle|\underline{x}\right)=\frac{1}{\left(2\pi\right)^{M}\left|R\right|^{\frac{M}{2}}}\cdot\exp\left(-\frac{1}{2}\sum_{i=0}^{M-1}\left(\underline{y}_{i}-\underline{m}\right)^{T}\cdot R^{-1}\cdot\left(\underline{y}_{i}-\underline{m}\right)\right)$$

- Log Likelihood Function

$$l\left(\underline{x}\right)=-M\cdot\ln\left(2\pi\right)+\frac{M}{2}\cdot\ln\left(\left|F\right|\right)$$

$$-\frac{1}{2}\sum_{i=0}^{M-1}\left(\underline{y}_{i}-\underline{m}\right)^{T}\cdot F\cdot\left(\underline{y}_{i}-\underline{m}\right)$$

$$F=R^{-1}$$

Matrix Gradient Lemma

- Quadratic Form J

$$J = \underline{v}^T \cdot P^{-1} \cdot \underline{v}$$

- Gradient With Respect to Matrix P
  - Not Given in Gelb
  - Not Obvious
  - Easily Found with a Trick
- Use and Extra $P \cdot P^{-1}$ in the Quadratic Form

Gradient Is

$$\frac{\partial J}{\partial P} == \frac{\partial}{\partial P}\left(\underline{x}^T \cdot P^{-1} \cdot \underline{x}\right)\frac{\partial}{\partial P}\left(\underline{x}^T \cdot P^{-1} \cdot P \cdot P^{-1} \cdot \underline{x}\right)$$

$$= \frac{\partial}{\partial P_1}\left(\underline{x}^T \cdot P_1^{-1} \cdot \left(P \cdot P^{-1}\right) \cdot \underline{x}\right)$$

$$+P^{-1} \cdot \underline{x} \cdot \underline{x}^T \cdot P^{-1}$$

$$+\frac{\partial}{\partial P_3}\left(\underline{x}^T \cdot \left(P^{-1} \cdot P\right) \cdot P_3^{-1} \cdot \underline{x}\right)$$

$$= 2 \cdot \frac{\partial J}{P} + P^{-1} \cdot \underline{x} \cdot \underline{x}^T \cdot P^{-1} = -P^{-1} \cdot \underline{x} \cdot \underline{x}^T \cdot P^{-1}$$

Likelihood Equations

- For Means

$$\frac{\partial l(\underline{x})}{\partial \underline{m}} = F \cdot \sum_{i=0}^{M-1} \left(\underline{y}_i - \hat{\underline{m}} \right) = 0$$

- For Covariances

$$\frac{\partial l(\underline{x})}{\partial F} = \frac{M}{2} \cdot F^{-T} - \frac{1}{2} \sum_{i=0}^{M-1} \left(\underline{y}_i - \underline{m} \right) \cdot \left(\underline{y}_i - \underline{m} \right)^T$$

Estimates

- For Mean

$$\underline{\hat{m}} = \frac{1}{M} \sum_{i=0}^{M-1} \underline{y}_i$$

- For Covariances

$$\hat{R} = \frac{1}{M} \sum_{i=0}^{M-1} \left(\underline{y}_i - \underline{\hat{m}} \right)^T \cdot \left(\underline{y}_i - \underline{\hat{m}} \right)$$

Variances of Estimates

- For Means

$$\frac{\partial^2 l(\underline{x})}{\partial \underline{m}^2} = -P^{-1} = -M \cdot \hat{R}^{-1}$$

$$P = \frac{1}{M} \cdot \hat{R}$$

- For Variances
  - Scalar term by term approach OK for 2 X 2
  - General Form
    - » Variances for elements of F, tensor gradients

Last Generalization

- Set of Noisy Correlated Means with Trends

$$\underline{x} = \underline{m} + \underline{r} \cdot t + \underline{v},\ \ \mathrm{Cov}\{\underline{v}\} = R$$

- Measurements for known t_i

$$\underline{y}_i = \underline{x}(t_i) = \underline{m} + \underline{r} \cdot t_i + \underline{v}$$

- States are Elements of m, r, R
- This is the Basis of the Questionnaire Analysis
- Solution is an Optional Exercise

The Bayesian Mean

- Bayes' Theorem

$$P\left(A_i \middle| B\right) = \frac{P\left(B \middle| A_i\right) \cdot P\left(A_i\right)}{\sum_k P\left(B \middle| A_k\right) \cdot P\left(A_k\right)}$$

- Bayesian Mean of State Vector (Gelb, page 104)

$$\hat{\underline{x}} = \int \underline{x} \cdot p\left(\underline{x} \middle| \underline{y}\right) \cdot d\,\underline{x}$$

$$p\left(\underline{x} \middle| \underline{y}\right) = \frac{p\left(\underline{y} \middle| \underline{x}\right) \cdot p\left(\underline{x}\right)}{p\left(\underline{y}\right)}, \quad p\left(\underline{y}\right) = \int p\left(\underline{y} \middle| \underline{x}\right) \cdot p\left(\underline{x}\right) \cdot d\,\underline{x}$$

Bayesian Kalman Update

- *A Priori* Mean and Covariance are $\tilde{\underline{x}},\ \tilde{P}$
- Known Probability Densities are

$$p\left(\underline{y}\middle|\underline{x}\right)=\frac{\exp\left(-\frac{1}{2}\left(\underline{y}-\underline{h}\left(\underline{x}\right)\right)^{T}\cdot R^{-1}\cdot\left(\underline{y}-\underline{h}\left(\underline{x}\right)\right)\right)}{\left(2\pi\right)^{\frac{M}{2}}\cdot\left|R\right|^{\frac{1}{2}}}$$

$$p\left(\underline{x}\middle|\tilde{\underline{x}}\right)=\frac{\exp\left(-\frac{1}{2}\left(\underline{x}-\tilde{\underline{x}}\right)^{T}\cdot\tilde{P}^{-1}\cdot\left(\underline{x}-\tilde{\underline{x}}\right)\right)}{\left(2\pi\right)^{\frac{N}{2}}\cdot\left|\tilde{P}\right|^{\frac{1}{2}}}$$

Bayesian (Continued)

- Distributions

$$p\left(\underline{y}\right) = \frac{\exp\left(-\tfrac{1}{2}\cdot\left(\underline{y}-\underline{h}\left(\underline{\tilde{x}}\right)\right)^T \cdot P_y^{-1}\cdot\left(\underline{y}-\underline{h}\left(\underline{\tilde{x}}\right)\right)\right)}{\left(2\pi\right)^{\frac{M}{2}}\cdot\left|P_y\right|^{\frac{1}{2}}}$$

$$P_y = H\cdot\tilde{P}\cdot H^T + R$$

$$p\left(\underline{x}\middle|\underline{y}\right) = \frac{\exp\left(-\tfrac{1}{2}\cdot Q\left(\underline{x},\underline{y}\right)\right)}{\left(2\pi\right)^{\frac{M\cdot N}{2}}\cdot\left|P_x\right|^{\frac{1}{2}}\cdot\left|P_y\right|^{\frac{1}{2}}\cdot p\left(\underline{y}\right)}$$

Kernel of p(x|y)

- The Sum of Quadratic Forms

$$Q\left(\underline{x},\underline{y}\right)=\left(\underline{y}-\underline{h}\left(\tilde{\underline{x}}\right)\right)^T\cdot P_y^{-1}\cdot\left(\underline{y}-\underline{h}\left(\tilde{\underline{x}}\right)\right)^T$$

$$+\left(\Delta\underline{x}\right)^T\cdot P_x^{-1}\cdot\left(\Delta\underline{x}\right)$$

$$\Delta\underline{x}=\underline{x}-\left(\tilde{\underline{x}}+\tilde{P}\cdot H^T\cdot R^{-1}\cdot\left(\underline{y}-\underline{h}\left(\tilde{\underline{x}}\right)\right)\right)$$

$$P_x=\left(\tilde{P}^{-1}+H^T\cdot R^{-1}\cdot H\right)^{-1}$$

Resulting Update

- Bayesian Mean

$$\underline{\hat{x}} = \underline{\tilde{x}} + K \cdot \left(\underline{y} - \underline{h}(\underline{\tilde{x}}) \right)$$

$$K = \tilde{P} \cdot H^T \cdot \left(H \cdot \tilde{P} \cdot H^T + R \right)^{-1}$$

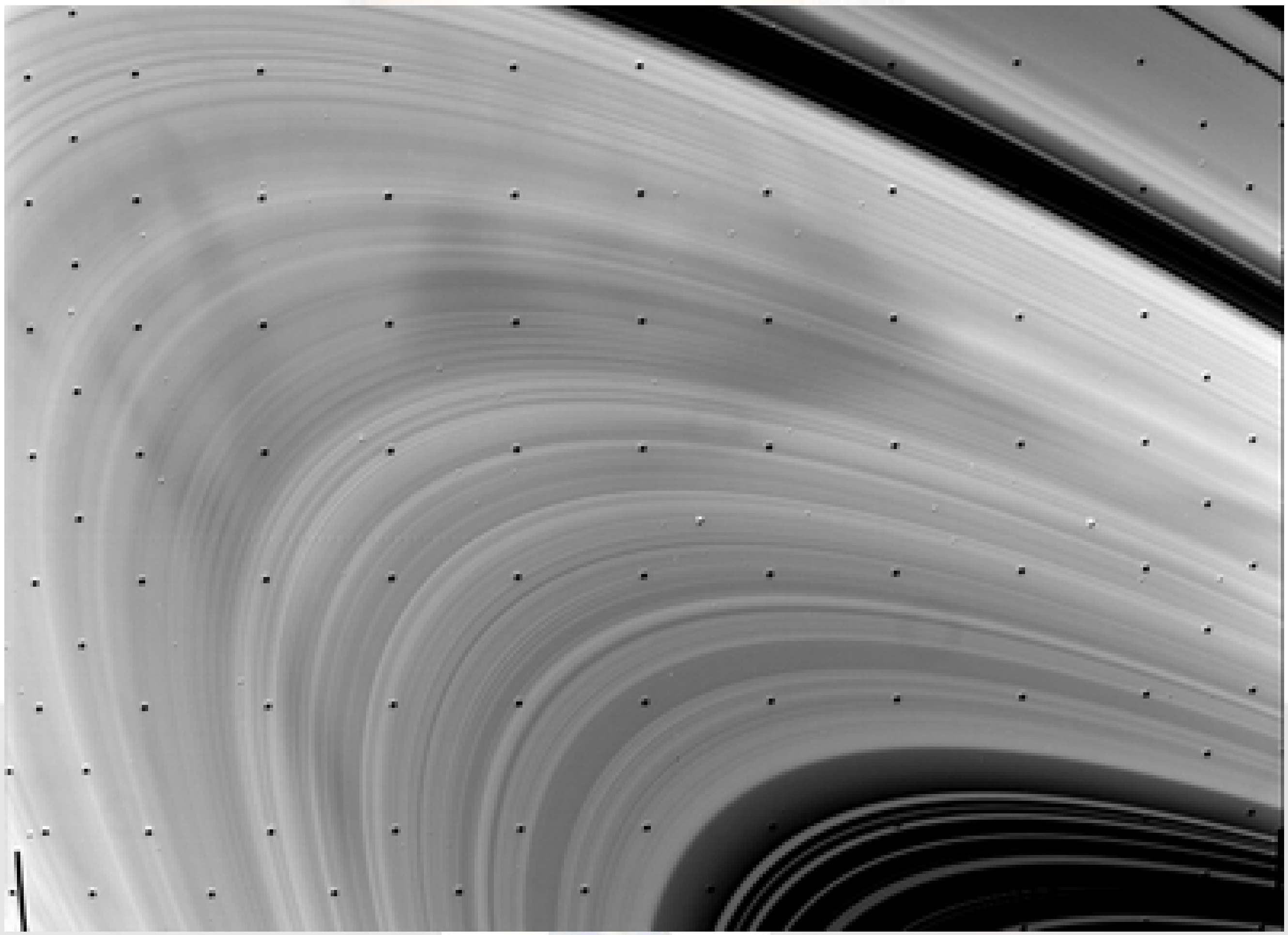
$$P = \left(\tilde{P}^{-1} + H^T \cdot R^{-1} \cdot H \right)^{-1}$$

- Maximum Likelihood gives Similar Result
  - Same Equations
  - But, Iterated A Little Batch Algorithm

Batch Estimators Compared

- Notation for Error $\underline{e} = \underline{y} - \underline{h}\left(\underline{\tilde{x}}\right)$
- Least Squares $J = \underline{e}^T \cdot W \cdot \underline{e}$
- Bayesian Weighted $J = \underline{e}^T \cdot R^{-1} \cdot \underline{e} + \left(\underline{x} - \underline{\tilde{x}}\right)^T \cdot \tilde{P}^{-1} \cdot \left(\underline{x} - \underline{\tilde{x}}\right)$
  - Maximum *A Posteriori*
  - Least Squares
- Maximum Likelihood $J = \underline{e}^T \cdot R^{-1} \cdot \underline{e} + \ln\left(\left|R\right|\right)$

Break



  From Raytheon ATEP SYS12525 for SAADS Day 1, October 19, 1999

Notation of Batch Estimators

- Method
  - Quadratic Cost Function in $\underline{y}$
  - Nonlinear Problem Statements
    - » Force Recursion □ Not One Pass
    - » Result is Asymptotically Unbiased and Efficient
    - » Multivariate Newton's Method
- Universally Found Variables

$$\underline{e} = \underline{y} - \underline{h}\left(\underline{\tilde{x}}\right)$$

The Nonlinear Recursion

- Basic Relationship

$$\underline{\hat{x}}_i = \underline{\hat{x}}_{i-1} - \left[\frac{\partial^2 J\left(\underline{\hat{x}}_{i-1}\right)}{\partial \underline{\hat{x}}_{i-1}^2}\right] \cdot \left[\frac{\partial J\left(\underline{\hat{x}}_{i-1}\right)}{\partial \underline{\hat{x}}_{i-1}}\right]$$

- Can Be Difficult
  - Increasingly Sensitive to Initialization as Amount of Data Increases
  - Selection of Coordinate System is Important

Numerical Formulation

- The Likelihood Equation

$$\underline{l}(\hat{\underline{x}}) = H^T \cdot R^{-1} \cdot \left(\underline{y} - \underline{h}(\hat{\underline{x}})\right) = \underline{0}$$

- Taylor Series

$$\underline{h}(\hat{x}) = \underline{h}(\underline{x}) + H \cdot (\hat{\underline{x}} - \underline{x}) + O\left\{\Delta \underline{x}^3\right\}$$

- Other Variable Changes

$$R^{-\frac{1}{2}} \cdot H = A, \;\; R^{-\frac{1}{2}} \cdot \left(\underline{y} - \underline{h}(\underline{x})\right) = \underline{e}$$

- Linearized Likelihood Equation

$$A^T \cdot (\hat{\underline{x}} - \underline{x}) = A^T \cdot A \cdot \underline{e}$$

Numerical Method

- Nonlinear Iteration

$$A \cdot \left(\underline{\hat{x}}_{i+1} - \underline{\hat{x}}_i \right) = \underline{e}$$

- Triangularize A with Sequence of Householder Transformations T

$$T \cdot A \cdot \left(\underline{\hat{x}}_{i+1} - \underline{x}_i \right) = \begin{bmatrix} S \\ 0 \end{bmatrix} \cdot \left(\underline{\hat{x}}_{i+1} - \underline{x}_i \right) = T \cdot \underline{e} = \begin{bmatrix} \underline{e'} \\ \underline{e''} \end{bmatrix}$$

- Solve Square Recursion

$$\underline{\hat{x}}_{i+1} = \underline{\hat{x}}_i + S^{-1} \cdot \underline{e'}$$

Use of an MLE

- Initialize Using First Few Measurements
- Update Estimate as New Data Becomes Available
- Extrapolate States Between Updates
  - If States Change with Time
  - Use Closed Form Solution to State Equation
- If Necessary to Re-Initialize Later
  - Use Minimum Sparse Data
  - Add Interspersed Data in Stages

Initialization of an MLE

- Initialization
  - Use First Few Measurements
    - » Compute State Estimates Algebraically
    - » Compute Covariance Matrix
      - State Vector Estimate Computation
      - Variance Propagation Equation
  - Initialize Unobservable States
    - » Best Guess
    - » Large Covariance
- Never Use Ad Hoc Initialization

From Raytheon ATEP SYS12525 for SAADS Day 1, October 19, 1999

Necessary *Ad* Hoc Technique: Re-Initialization

- Use Minimum Number of Measurements
  - For Observability of Most or All States
  - Use Either
    - » Algebraic Closed Form
    - » Nonlinear Recursion
- Select Data for State Observability
  - First, Quartile Points, Last Data Point
- Add Data Points in Stages
  - 1/8, 3/8, 5/8, 7/8 Stage, Etc.

Necessary *Ad Hoc* Technique: Control of First Convergence

- Put Hard Limits on State Excursion
  - Visualize Locus of Estimate in State Space During Convergence
  - Allow an Order of Magnitude Beyond Reasonable Limits for Locus
  - Preserve the Direction of $\Delta \underline{x}$
- Limit the Number of Iterations
  - Recursion May Not Converge Early in Run
  - Wait for More Data

Generalizations

- A Priori Data
  - Augment Original Data Equation
  - Treat Previous Data as Pseudo-Measurements
- Weighted Least Squares
  - Replace R^{-1} with W
  - Proceed as Above

From Raytheon ATEP SYS12525 for SAADS Day 1, October 19, 1999

Canonic Variates

- Definition
  - Linear combination of states $\underline{x}$
  - Unitless, covariance of I
- General Form

$$\underline{z} = S' \cdot (\underline{x}), \;\; S' = C \cdot S, \;\; C^T \cdot C = I$$

- Estimation in Canonic Coordinates
  - Optimal numerical conditioning
  - Natural consequence of recommended process

Gelb Example 1-1 PP. 5-6

- Two Sensors
  - Each taking a single noisy measurement
  - $z_i = x + v_i$
  - Unknown x is a constant
  - Measurement noises v_i are unbiased, uncorrleated
- Begin with Linear Estimator

$$\hat{x} = k_1 \cdot z_1 + k_2 \cdot z_2$$

Gelb Example 1-1 (Continued)

- Define error $\tilde{x} = \hat{x} - x$
- Make k_1 and k_2 Independent of x
- Make Mean of Estimate x; Result is $k_1 + k_2 = 1$
- Minimum Squared Error; Result is

$$k_1 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}, \quad k_2 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}$$

- Variance σ^2 of Optimal Estimate is

$$\frac{1}{\sigma^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}$$

The Example from Gelb

- An *ad hoc* Estimation Procedure of a Mean
- Premises are
  - Estimator is Independent of Estimate
  - Estimate is Unbiased
  - Estimate has Minimum Variance for Assumed Form
  - Variances of Measurements are Known
- Result is a Simple Weighted Average

Problem 1-1

- Same as Example Except Variables are Correlated
- Still, $k_1 + k_2 = 1$
- But,

$$E\left(\tilde{x}^2\right) = k_1^2 \cdot \sigma_1^2 + 2 \cdot k_1 \cdot k_2 \cdot \sigma_1 \cdot \sigma_2 \cdot \rho + k_2^2 \cdot \sigma_2^2$$

Solving for Gains

- The Gradient with Respect to k_1 is

$$2 \cdot k_1 \cdot \sigma_1^2 + 2 \cdot \left(k_2 - k_1\right) \cdot \sigma_1 \cdot \sigma_2 \cdot \rho - 2 \cdot k_2 \cdot \sigma_2^2 = 0$$

- The Solution for the Gain is

$$k_1 = \frac{-\sigma_1 \cdot \sigma_2 \cdot \rho + \sigma_2^2}{\sigma_1^2 - 2 \cdot \sigma_1 \cdot \sigma_2 \cdot \rho + \sigma_2^2}$$

$$k_2 = \frac{\sigma_1^2 - \sigma_1 \cdot \sigma_2 \cdot \rho}{\sigma_1^2 - 2 \cdot \sigma_1 \cdot \sigma_2 \cdot \rho + \sigma_2^2}$$

From Raytheon ATEP SYS12525 for SAADS Day 1, October 19, 1999

When ρ is 1

- What is ρ
  - Called the Correlation Coefficient
  - Is the Normalized Covariance of Two Random Variables
- Meaning of $\rho = \pm 1$

$$v_2 = \pm \frac{\sigma_2}{\sigma_1} \cdot v_1$$

Let's Use Vectors

- Measurements (unknown x is a scalar)

$$\underline{z} = x \cdot \underline{1} + \underline{v},\ \ Cov\{\underline{v}\} \equiv Exp\left\{\underline{v} \cdot \underline{v}^T\right\} = R$$

$\underline{1} \equiv$ [vector of all 1's]

- Estimate

$$\hat{x} = K \cdot \underline{z}$$

- Estimation Error

$$\tilde{x} = K \cdot \underline{z} - x = K \cdot \underline{v} + K \cdot \underline{1} \cdot x - x$$

Estimator Derivation

- Unbiased Condition

$$K \cdot \underline{1} = 1$$

- Variance of Estimate

$$Exp\left\{ \tilde{x}^{2} \right\} = K \cdot R \cdot K^{T} - \lambda \cdot \left(K \cdot \underline{1} - 1 \right)$$

- Gradient of Error Variance With Respect to Weights K

$$2 \cdot R \cdot K^{T} - \lambda \cdot \underline{1} = \underline{0}$$

The General Solution

- The Weights

$$K^T = \frac{\lambda}{2} \cdot R^{-1} \cdot \underline{1}$$

- The Unbiased Condition

$$K \cdot \underline{1} = \frac{\lambda}{2} \cdot \underline{1}^T \cdot R^{-1} \cdot \underline{1} = 1$$

- The Value of the Extra Parameter λ

$$\lambda = \frac{2}{\underline{1}^T \cdot R^{-1} \cdot \underline{1}}$$

The General Result

- The Weights

$$K^T = \frac{1}{\underline{1}^T \cdot R^{-1} \cdot \underline{1}} \cdot R^{-1} \cdot \underline{1}$$

- The Resulting Variance of the Estimate

$$Exp\left\{\tilde{x}^2\right\} = K \cdot R \cdot K^T = \frac{1}{\underline{1}^T \cdot R^{-1} \cdot \underline{1}}$$

A Well Known Result

- When All the Cross Terms of R are Zero

$$K^T = \frac{\left[\frac{1}{\sigma_i^2}\right]}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}} \cdot \underline{1}, \quad \left[\frac{1}{\sigma_i^2}\right] = Diag\left\{R^{-1}\right\}$$

$$Exp\left\{\left(\tilde{x} - x\right)^2\right\} = K \cdot R \cdot K^T = \frac{1}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}}$$

Method of Maximum Likelihood

- Correlated Noisy Measurements of Constant

$$\underline{y} = z \cdot \underline{1} + \underline{v}, \ \ \mathrm{Cov}\left\{\underline{v}\right\} = R$$

- Log Likelihood Function

$$l(x) = -\frac{N}{2} \cdot \ln(2\pi) - \frac{1}{2} \cdot \ln(|R|)$$

$$-\frac{1}{2} \cdot \left(\underline{y} - x \cdot \underline{1}\right)^T \cdot R^{-1} \cdot \left(\underline{y} - x \cdot \underline{1}\right)$$

- Likelihood Equation

$$\frac{\partial l(x)}{\partial x} = \underline{1}^T \cdot R^{-1} \cdot \left(\underline{y} - \hat{x} \cdot \underline{1}\right) = 0$$

Solution and Covariance

- Solution to Likelihood Equation

$$\hat{x} = \frac{\underline{1}^T \cdot R^{-1} \cdot \underline{y}}{\underline{1}^T \cdot R^{-1} \cdot \underline{1}}$$

- Variance of Estimate is Cramer-Rao Bound

$$\frac{\partial^2 l(x)}{\partial x^2} = -\underline{1}^T \cdot R^{-1} \cdot \underline{1} = \frac{1}{\sigma^2(\hat{x})}$$

$$\sigma^2(\hat{x}) = \frac{1}{\underline{1}^T \cdot R^{-1} \cdot \underline{1}}$$

BREAK

ATEP SYS12525

Radar Trackers and Applications for SAADS

October 19, 1999

Topic 6: Simple Filters

Sensor Systems Engineering for the 21st Century

Elementary Digital Filters

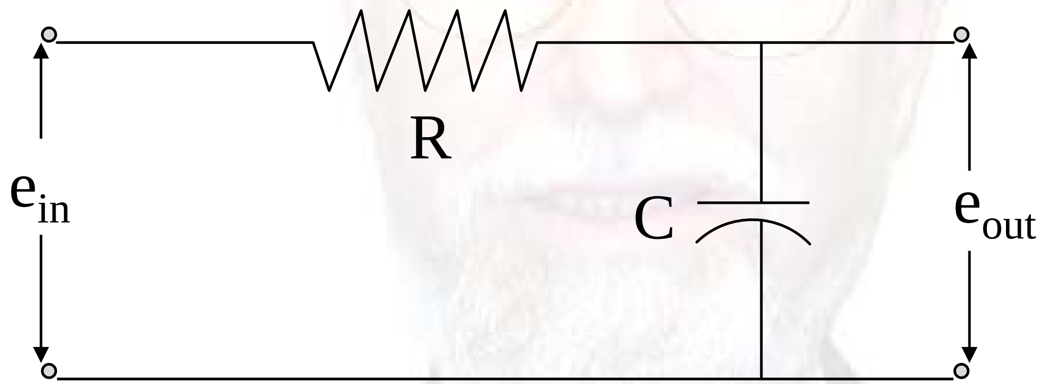
- Types
  - Simple Recursive
  - Block Averager
  - Two State
- Advantages
  - Simplicity -- Simple Recursive
  - Optimality -- Block Averager
  - Zero Latency -- Two State

Elementary Digital Filters (Continued)

- Disadvantages
  - Time Delay Vs. Bandwidth Tradeoff -- Simple Recursive
  - Complexity -- Block Averager
  - Overshoot and Ringing -- Two State
- Selection Criteria
  - Requirements Driven
  - Varies with Application

Simple Digital Filter

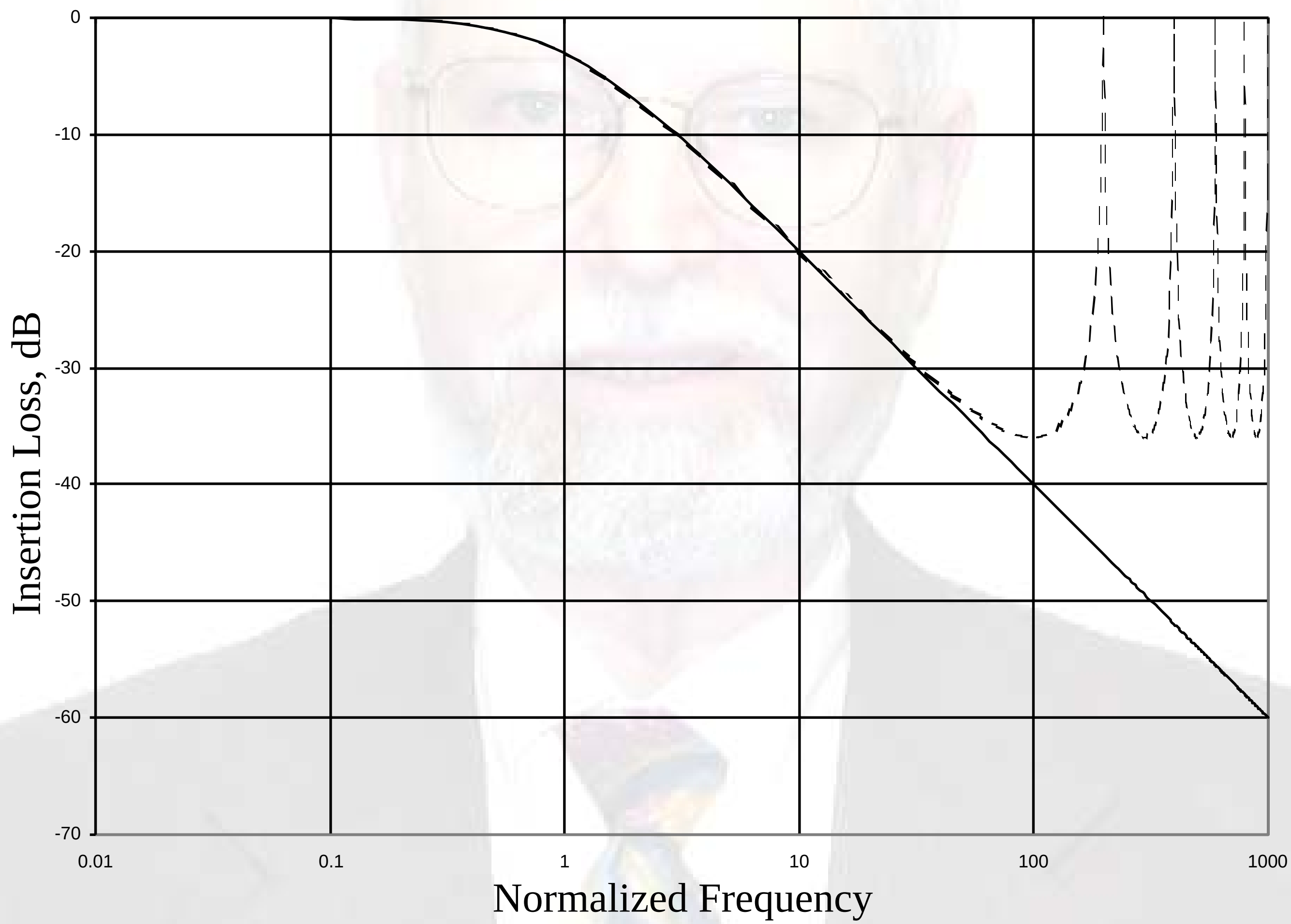
- Analog Filter Analogy



- Step Response

$$e_{out}(t) = V \cdot \left(1 - \exp\left(-\frac{t}{R \cdot C}\right)\right)$$

Frequency Response



Equivalent Digital Operation

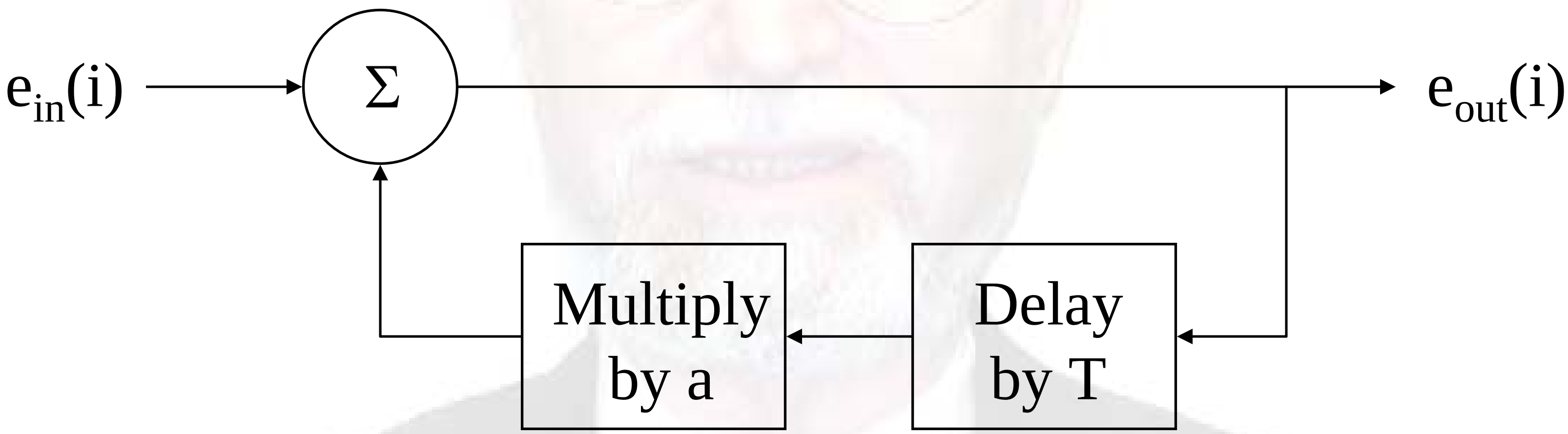
- Samples Vs Time

$$e(i) = 1 - \exp\left(-\frac{i \cdot T}{R \cdot C}\right) = 1 - a^i$$

- Analogy

$$a = \exp\left(-\frac{T}{R \cdot C}\right)$$

Digital Filter Topology



Filter Normalization

- Scale Input by Factor of (1 - a)
  - Insertion Loss at DC is Normalized
  - Scales Output Without Changing Form of Transfer Function
- Arithmetic is

$$y(i) = (1 - a) \cdot x(i) + a \cdot y(i-1)$$

- Step Response is

$$y(n) = 1 - a^{n+1}$$

Noise Variance Propagation

- Variance at Output is

$$\sigma_y^2(i) = (1-a)^2 \cdot \sigma_x^2 + a^2 \cdot \sigma_y^2(i-1)$$

- Steady State Variance is

$$\sigma_y^2 = \frac{1-a}{1+a} \cdot \sigma_x^2$$

Recursive Filter Summary

- Impulse Response Same as Analog Filter
  - Impulse Invariant Method
  - Other Methods
    - » Match Pole Positions
    - » Match Frequency Response
- Phase Delay is $R \cdot C = -\frac{T}{\ln(a)}$
- Rise Time is RC
- Bandwidth is about $\frac{1}{2\pi RC}$

Digital Filter Design Tradeoffs

- Type
  - Finite Impulse Response (FIR)
  - Infinite Impulse Response (IIR or recursive)
- Phase Delay
  - Increases with Decreasing Bandwidth
  - Linear with FIR
  - Larger near poles with IIR

Digital Filters (Continued)

- Settling Time
  - Increases as Bandwidth Decreases
  - Fill Time for FIR Filters
  - IIR Settles Exponentially
    - » Rate Determined by Highest Q Pole
    - » For simple filter, rate is $\frac{20 \cdot \log 10(a)}{T}$ dB/msec
- Noise Propagation

Summary

- Sensor Systems
  - Defined from Mission and Requirements
  - Tracker is Intrinsic in Sensor Definition
- Digital Filters
  - Many Kinds
  - Design Tradeoffs Include
    - » Bandwidth
    - » Time Delay
    - » Settling Time

ATEP SYS12525

Radar Trackers and Applications for SAADS

October 19, 1999

Topic 7: Z Transforms and Digital Filters

Sensor Systems Engineering for the 21st Century

Suggested Study from Gelb or Other References

- Readings
  - Refresher on Fundamentals of Vectors and Matrices pp 10-19
  - “Functions of Square Matrices” pp 19-20
  - “Definite Forms” p 21
  - Gradients and Least Squares, pp 21-24
- Look at the Exercises on p 46 (NOT AN ASSIGNMENT)
  - Problem 2-2 (Hint: Compute the polynomial and try the suggested values by direct substitution)
  - Problem 2-3 (Hint: Make a variable change based on a Singular Value Decomposition of the matrix, use p 21)

Z Transforms and Digital Filters

- Difference Equations
  - – Z Transforms Classically Used in Analysis
  - – Digital Filter Algebraic Definitions
- Digital Trackers
  - – Stationary Filters Easily Analyzed
  - – Adaptive Filters Characterized and Bounded
    - » Bounds of Parameters Bound Z Transform Results
    - » Z Transform Characterizations Useful in Interpretation

Z Transform Definition

- Laplace Transform Analogy for Uniformly Sampled Filters

$$F(s) = \int_0^\infty e(t) \cdot \exp(-s \cdot t) \cdot dt \qquad F(z) = \sum_{k=0}^\infty e(k) \cdot z^{-k}$$

- Function e(t) is Series of Impulses at $t_k = k \cdot T$
- Inversion Integral

$$f(n) = \frac{1}{2\pi j} \cdot \oint F(z) \cdot z^{n-1} \cdot dz$$

- Variable Change Linking the Analogy is

$$z = \exp(s \cdot T)$$

where 1/T is sample rate

Z Transform Pairs

$f(n)$	$F(z)$
a^n	$\dfrac{z}{z-a}$
$f(n+a)$	$z^a \cdot f(z) - \sum_{i=0}^{a} z^{a-i} \cdot f(i)$
$f(n-a)\cdot u(n-a)$	$z^{-a} \cdot F(z)$
$\sum_{k=0}^{\infty} f_2(k)\cdot f_1(n-k)$	$F_1(z)\cdot F_2(z)$

Laplace-Stieltjes Transforms

- Form is

$$\int_0^\infty \exp(-s \cdot t) \cdot d\Phi(t)$$

- Where $\Phi(t)$
  - Constant Except for Steps at $t_n = nT$
  - Size of Steps is f(n)
- A Z Transform for Mathematicians

Break

Digital Filters and Difference Equations

- Simple Digital Filter Revisited

$$y(n+1) = (1-a)\cdot x(n) + a\cdot y(n)$$

- Taking Z Transform of Both Sides:

$$z\cdot Y(z) - z\cdot y(0) = (1-a)\cdot X(z) + a\cdot Y(z)$$

- Solving for Y(z)

$$Y(z) = \frac{z}{z-a}\cdot y(0) + \frac{1-a}{z-a}\cdot X(z)$$

Transfer Function

- First Term
  - Ringing or Settling from Initial Conditions
  - Decays to Zero in Stable Filter (|a|<1)
  - Ignored in Transfer Function from Input
- Transfer Function

$$G(z) = \frac{Y(z)}{X(z)} = \frac{1-a}{z-a}$$

Initial and Final Values of f(n)

- From Definition

$$F(z) = \sum_{n=0}^{\infty} z^{-n} \cdot f(n)$$

- Initial Value is Easy

$$f(0) = \lim_{z \to \infty} F(z)$$

- Final Value From

$$Z\big(f(n+1) - f(n)\big) = z \cdot F(z) - z \cdot f(0) - F(z)$$

Final Value (Continued)

- The Logic is

$$(z-1)\cdot F(z)-f(0)=$$

$$\lim_{p\to\infty}\sum_{n=0}^{p}\left(f(n+1)-f(n)\right)\cdot z^{-n}$$

- Taking a Second Limit $z\to 1$

$$\lim_{n\to\infty} f(n)=\lim_{z\to 1}(z-1)\cdot F(z)$$

Noise Variance Propagation

- Variance Difference Equation

$$\sigma_y^2(i+1) = (1-a)^2 \cdot \sigma_x^2(i) + a^2 \cdot \sigma_y^2(i)$$

- Z Transform of Output Variance

$$V_y(z) = \frac{z}{z-a^2} \cdot \sigma_y^2(0) + \frac{(1-a)^2}{z-a^2} \cdot V_x(z)$$

- Steady State Output Variance (Why?)

$$\sigma_y^2 = \frac{1-a}{1+a} \cdot \sigma_x^2$$

Frequency Response

- Follows From Variable Changes

$$z = \exp(s \cdot T),\ \ s = j \cdot \omega$$

- Z Transform is Then

$$F(\omega) = \sum_{n=0}^{\infty} f(n) \cdot \exp(-j \cdot \omega \cdot T)$$

- This is a Frequency Response
  - Same Equation as Fourier Summation
  - From Z Transform for z on Unit Circle (|z|=1)

Generalization Using Vectors

- Order N Digital Filter Difference Equation

$$\underline{y}(n+1) = A \cdot \underline{y}(n) + B \cdot \underline{x}(n)$$

- And the Z Transform is

$$z \cdot \underline{Y}(z) - z \cdot \underline{y}(0) = A \cdot \underline{Y}(z) + B \cdot \underline{X}(z)$$

- Solving for Y(z)

$$\underline{Y}(z) = -\left[A - z \cdot I\right]^{-1} \cdot \left[B \cdot \underline{X}(z) + z \cdot \underline{y}(0)\right]$$

The Poles

- Denominator Polynomial
  - Determinant $|A - z \cdot I|$
  - Order N
- Poles of $\underline{Y}(z)$ are
  - Roots of Denominator Polynomial
  - The Characteristic Values of Matrix A
  - Filter Stable is All Magnitudes Less Than 1

Example

- Constant Velocity Model

$$\underline{x} = \begin{bmatrix} Position \\ Velocity \end{bmatrix}, \; A = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}$$

- Z Transform of Output is

$$\underline{Y}(z) = \begin{bmatrix} \frac{1}{z-1} & \frac{T}{(z-1)^2} \\ 0 & \frac{1}{z-1} \end{bmatrix} \cdot \left[B \cdot \underline{X}(z) + z \cdot \underline{y}(0) \right]$$

Decoupling the Matrix Equation

- Make a Variable Change

$$\underline{w}(n) = C \cdot \underline{y}(n)$$

- The New Difference Equation is

$$\underline{w}(n+1) = C \cdot A \cdot C^{-1} \cdot \underline{w}(n) + C \cdot B \cdot \underline{x}(n)$$

- Select C as the Characteristic Vector Matrix of A

$$C \cdot A \cdot C^{-1} = \Lambda$$

The Form of C and Λ

- C is Characteristic Vector Matrix
  - Columns of C^{-1} are Characteristic Vectors
  - If A is Symmetric, then $C^{-1} = C^T$
- The Matrix Λ is Diagonal if
  - No Characteristic Values are Repeated
- Repeated Characteristic Values
  - Matrix Λ is Not Diagonal
  - Submatrices of Jordan Canonical Form

Characteristic Values

- Any Characteristic Value λ and Vector c

$$A \cdot \underline{c} = \lambda \cdot \underline{c}$$

- Jordan Canonical Form Submatrices

$$\Lambda_R = \begin{bmatrix} \lambda_R & 1 & 0 \\ 0 & \lambda_R & 1 \\ 0 & 0 & \lambda_R \end{bmatrix}$$

Time Functions for Multiple Characteristic Values

- Terms Appearing in the Z Transform

$$\frac{z^{i}}{(z-a)^{i}},\ 1 \le i \le N_R$$

- Corresponding Time Series Terms

$$n^{i-1} \cdot a^{n}$$

Form of Λ Matrix

- Form of Inverse of [z·I - Λ]

$$\left[z\cdot I-\Lambda_R\right]^{-1}=\begin{bmatrix}\frac{1}{z-\lambda_R} & \frac{1}{(z-\lambda_R)^2} & \frac{1}{(z-\lambda_R)^3}\\ 0 & \frac{1}{z-\lambda_R} & \frac{1}{(z-\lambda_R)^2}\\ 0 & 0 & \frac{1}{z-\lambda_R}\end{bmatrix}$$

Useful Books

- “Introduction to Matrix Analysis” (Second Edition), Richard Bellman, McGraw-Hill, 1970.
- “Operational Mathematics,” R. V. Churchill, McGraw-Hill (1958)
- “Theory and Application of the Z Transform Method,” E. I. Jury, John Wile & Sons, Inc. (1964)

Suggested Readings

- Blackman and Popoli Introduction
- Blackman and Popoli Chapter 1
  - – Basics of Target Tracking
    - » Single Target Tracking
    - » Multiple Target Tracking
  - – Data Association: The Key Problem
  - – Sensor Issues

Discussion

- Questions About Material
  - Content
  - Organization
  - Application
  - Theory
- Questions About Course
  - This Session
  - Future Sessions

Useful Books

- “Design and Analysis of Modern Tracking Systems” by Sam Blackman and Bob Popoli, Artech (1999)
- “Linear Algebra” by Georgi E. Shilov, Dover (1977)
- “Spectral Analysis and Time Series,” M. B. Priestly, Academic Press (1989).
- “Applications of Tensor Analysis,” A. J. McConnell, Dover 0-486-60373-3

Useful Books

- "Applied Optimal Estimation," A. Gelb, Ed. MIT Press, (1974)
- "Multiple Target Tracking in a Dense Environment," S. Blackman, Artech (1988)
- "Factorization Methods in Discrete Sequential Estimation," G. Bierman, Academic Press (1977)
- "Parameter Estimation," Harold W. Sorenson, Marcel Dekker (1980)

Useful Books

- "Optimal Control and Estimation Theory," George M. Siouris, Wiley (1996).
- "Handbook of Mathematical Functions," National Bureau of Standards Applied Mathematics Series #55 (Also Available from Dover)
- "Introduction to Airborne Radar," George W. Stimson (Call Mike Tom to order)

Stimson's Book

- "Introduction to Airborne Radar"
- Second Edition (2000)
- Major Update to First Edition (1983)
- ISBN-10: 1891121154
- ISBN-13: 978-1891121159