Chapter 5
Optimal Estimation

Part 1
5.1 Random Variables, Processes and Transformation
Outline

• 5.1 Random Variables, Processes and Transformation
  – 5.1.1 Characterizing Uncertainty
  – 5.1.2 Random Variables
  – 5.1.3 Transformation of Uncertainty
  – 5.1.4 Random Processes
  – Summary
Outline

• 5.1 Random Variables, Processes and Transformation
  – 5.1.1 Characterizing Uncertainty
  – 5.1.2 Random Variables
  – 5.1.3 Transformation of Uncertainty
  – 5.1.4 Random Processes
  – Summary
5.1.1 Characterizing Uncertainty

• Uncertainty =
  – Not Known: Bias, scale systematic error
    • E.g. Temperature sensitivity
  – Not Knowable: Noise, randomness, unpredictability
    • E.g. “drift”

• Fact of life:
  – Some randomness is fundamental
  – It can’t be measured.

• Humans do a good job coping...
Modeling Uncertainty

• An oxymoron?
• **Distributions** are models.
• Algebraic and differential equations are models.
• We can “pass distributions through” equations to get other distributions.
  – one point at a time, or...
  – as a complete distribution
5.1.1.1 Types of Uncertainty

• We usually consider it to be additive:

\[ x_{\text{meas}} = x_{\text{true}} + \varepsilon \]
\[ \hat{x} = x_{\text{true}} + \varepsilon \]

Hat means estimate

• \( \varepsilon \) may be zero, a constant, or a function of anything.

• \( \varepsilon \) may be:
  – Systematic (= “deterministic”)
  – Random (= “stochastic”)
  – a combination of both.

• Most of all \( \varepsilon \) is unknown. Otherwise we would take it out.
5.1.1.1 Real and Ideal Signals

- Below: bias, scale errors, and two “outliers”.

![Graph showing real and ideal signals with bias, scale errors, and outliers.](image-url)
5.1.1.1 Real and Ideal Signals

- Below: Saturation, nonlinearity, deadband
5.1.1.1 Real and Ideal Signals

• Might model the errors like so:

\[ \varepsilon = a + b \theta + N(\mu, \sigma) \]

• Note the appearance of model parameters of both kinds:
  – systematic \((a,b)\)
  – stochastic \((\mu, \sigma)\)
Removing errors

• Systematic $\rightarrow$ calibration:
  – Fit a line to the last graph

• Stochastic $\rightarrow$ filtering
  – Smooth out the wiggles

• Correlation $\rightarrow$ differential measurement
  – Reject the common component
Know Your Model

• You can fit a line to anything.
Know Your Model

• You can fit a line to anything.

data is a perfect parabola
Outline

• 5.1 Random Variables, Processes and Transformation
  – 5.1.1 Characterizing Uncertainty
  – 5.1.2 Random Variables
  – 5.1.3 Transformation of Uncertainty
  – 5.1.4 Random Processes
  – Summary
Probability as Frequency Distribution

- Events that occur randomly may nonetheless have a knowable probability distribution.

- It’s not unusual to know the distribution but never be able to perfectly predict an individual event.

- Knowing one distribution allows you to compute others.
  - Math on distributions is well defined.
Continuous Random Variable

- Pdf – probability density function.

\[ p(a \leq x \leq b) = \int_{a}^{b} f(u) \, du \]

A “multimodal” distribution has more than one bump in it.

- Describes probability of each possible outcome of a single experiment.
Discrete Random Variable

• Pf – probability function.

For the sum to hold, the outcomes must be mutually exclusive.

Big $p$

$$P(a \leq x \leq b) = \sum_{i} P(x_i)$$

• Describes probability of each possible outcome of a single event.
(Joint) 2D Distributions

\[ p(a \leq x \leq b \land c \leq y \leq d) = \int_{a}^{b} \int_{c}^{d} f(u, v) \, du \, dv \]
(Conditional) 2D Distributions

\[ p(x, y) \]

\[ p(x \mid y) = \frac{\int_{a}^{b} f(u, y) du}{\int_{-\infty}^{\infty} f(u, y) du} \]

- Take a slice and renormalize.
Gaussian Pdf

\[ p(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) \]
N Dimensional Gaussian

• Formula:

\[
p(x) = \frac{1}{(2\pi)^{n/2} \sqrt{\det(C)}} \exp\left(-\frac{[x - \mu]^T C^{-1} [x - \mu]}{2}\right)
\]

• Mahalanobis distance:

\[
[x - \mu]^T C^{-1} [x - \mu]
\]

• C is “covariance matrix” defined later.
5.1.2.2 Expectation

• For any function of x, this is just a weighted average where the pdf is the weight.

\[
\begin{align*}
\text{Exp}[h(x)] &= \int_{-\infty}^{\infty} h(x)p(x)dx \quad \text{scalar-scalar continuous} \\
\text{Exp}[h(x)] &= \int_{-\infty}^{\infty} h(x)p(x)dx \quad \text{vector-scalar continuous} \\
\text{Exp}[h(x)] &= \int_{-\infty}^{\infty} h(x)p(x)dx \quad \text{vector-vector continuous} \\
\text{Exp}[h(x)] &= \sum_{i=1}^{n} h(u_i)P(u_i) \quad \text{vector-vector discrete}
\end{align*}
\]

• This is a functional or moment (with infinite limits of integration) so you need the entire pdf to work it out.
5.1.2.2 Expectation

• Properties inherited from integrals.

\[
\begin{align*}
\text{Exp}[k] &= k \\
\text{Exp}[kh(x)] &= k \text{Exp}[h(x)] \\
\text{Exp}[h(x) + g(x)] &= \text{Exp}[h(x)] + \text{Exp}[g(x)]
\end{align*}
\]

Expectation is a linear operator over functions.
5.1.2.2 Mean

• Set \( h(x) \rightarrow x \) etc.

\[
\mu = Exp[x] = \int_{-\infty}^{\infty} [xp(x)] dx \quad \text{scalar continuous}
\]

\[
\mu = Exp(x) = \int_{-\infty}^{\infty} xp(x) dx \quad \text{vector continuous}
\]

\[
\mu = Exp[x] = \sum_{i=1}^{n} x_iP(x_i) \quad \text{vector discrete}
\]

• This is a property of the distribution of the population.
5.1.2.2 Mean and Most Likely Value

- Expected value is a centroid.
- It is not always the most likely value to occur.
Variance of a Random Scalar

• Set $h(x) \rightarrow [x-\mu]2$.

\[
\sigma_{xx} = \int_{-\infty}^{\infty} [(x - \mu)^2 \cdot p(x)] \, dx
\]

• Alternative notation: $\sigma^2(x)$

• Standard deviation defined as:

\[
\sigma_x = \sigma(x) = \sqrt{\sigma_{xx}}
\]
Recall: “Outer” Product

• Opposite of “inner” or dot product.

\[ \mathbf{x} = \begin{bmatrix} x & y & z \end{bmatrix}^T \]

• Generates a symmetric matrix from a vector.

\[ \mathbf{x} \mathbf{x}^T = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \begin{bmatrix} x & y & z \end{bmatrix} = \begin{bmatrix} xx & xy & xz \\ yx & yy & yz \\ zx & zy & zz \end{bmatrix} \]
Co-Variance of a Random Vector

• Continuous and discrete cases.

\[ \Sigma = E( [x - \mu] [x - \mu]^T ) = \int_{-\infty}^{\infty} [x - \mu] [x - \mu]^T f(x) \, dx \]

\[ \Sigma = E( [x - \mu] [x - \mu]^T ) = \sum_{i=1}^{n} [x_i - \mu] [x_i - \mu]^T p(x_i) \]

• Integral of a matrix is the matrix of the integrals.
Sample Statistics

• Mean:

\[ \overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \]

• Sample Covariance.

\[ S = \frac{1}{n} \sum_{i=1}^{n} [x_i - \mu] [x_i - \mu]^T \]

• Elemental variances and co variances

\[ s_{ii} = \frac{1}{n} \sum_{i=1}^{n} [x_i - \mu_i] [x_i - \mu_i] \quad s_{ij} = \frac{1}{n} \sum_{i=1}^{n} [x_i - \mu_i] [x_j - \mu_j] \]
5.1.2.3 Sampling Distributions and Statistics

• “Batch” Methods:

\[
\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i
\]

\[
S = \frac{1}{n} \sum_{i=1}^{n} [x_i - \mu][x_i - \mu]^T
\]

• Not feasible computationally for continuous update when N is large.
5.1.2.4 Computing Sample Statistics

• “Recursive” Methods:

\[
\hat{x}_{k+1} = \frac{(k\hat{x}_k + \hat{x}_{k+1})}{(k+1)} \\
S_{k+1} = \frac{kS_k + [\hat{x}_{k+1} - \mu][\hat{x}_{k+1} - \mu]^T}{(k+1)}
\]

• Related to the Kalman Filter.
Computing Sample Statistics

- “Calculator” Methods use accumulators:

\[
\begin{align*}
\text{mean} & \quad T_{k+1} &= T_k + x_{k+1} \\
& \quad \bar{x}_{k+1} &= \frac{T_{k+1}}{k+1}
\end{align*}
\]

when data arrives

when answer necessary

\[
\begin{align*}
\text{covariance} & \quad Q_{k+1} = Q_k + [x_{k+1} - \mu][x_{k+1} - \mu]^T \\
S_{k+1} &= \frac{Q_k}{k+1}
\end{align*}
\]

when data arrives

when answer necessary

- Used in ... you guessed it ... hand calculators.
Contours of Constant Probability

- Consider the probability contained within a symmetric interval on the x axis.
5.1.2.6 Contours of Constant Probability

• In 2D, consider contours of constant exponent.

• These are ellipsoids in n dimensions:

\[
(x - \mu)^T \Sigma^{-1} (x - \mu) = k^2(p)
\]
Outline

• 5.1 Random Variables, Processes and Transformation
  – 5.1.1 Characterizing Uncertainty
  – 5.1.2 Random Variables
  – 5.1.3 Transformation of Uncertainty
  – 5.1.4 Random Processes
  – Summary
Transformation

• “Pass covariance through a function”:

• **Suppose** $y = 2x$ and $x$ is random.
  - 1st point: if $x$ is random, $y$ must be random – even if “2” is not.
  - 2nd point: if we **know** $\text{cov}(x)$ we can find $\text{cov}(y)$. How? Here’s the hard way.

• This works even for nonlinear functions $y = f(x)$ but there is a simpler way.

$\text{Turns out } \sigma_y = 2 \sigma_x$
Linearization

• The Taylor series allows us to extend any function into a neighborhood around a given point if we know the derivatives at that point:

\[ f(x + \Delta x) = f(x) + f'(x) \Delta x + f''(x) \frac{\Delta x^2}{2} + \ldots \]

• Error involved in truncation is related to magnitude of first neglected term.

• We linearize like so:

\[ f(x + \Delta x) \approx f(x) + f'(x) \Delta x \]

• Errors involved are “second order”
5.1.3.1 Linear Transformation: Mean

• Suppose we know $\mu_x$ and want $\mu_y$ where:

$$y = Fx$$

• Because expectation is an integral and hence a linear operator:

$$\mu_y = \text{Exp}(Fx) = F\text{Exp}(x)$$

• In other words

$$\mu_y = F\mu_x$$
5.1.3.1 Linear Transformation: Covariance

• Suppose we know $\sigma_x$ and want $\sigma_y$ where:

$$y = Fx$$

F independent of x

• Because covariance is an integral and hence a linear operator:

$$\Sigma_{yy} = \text{Exp}(Fxx^TF^T) = F\text{Exp}(xx^T)F^T$$

• In other words

$$\Sigma_{yy} = F\Sigma_{xx}F^T$$
5.1.3.2 Variance of a Sum of RVs

• Suppose there are \( n \) random variables \( x_i \) of same distribution.

• Define a new variable \( y \) as the sum of these:

\[
y = \sum_{i=1}^{n} x_i
\]

• What is the variance of \( y \)?

\( x_i \sim N(\mu, \sigma) \), \( i = 1, n \)

Variance of x’es known and equal.
5.1.3.2 Variance of a Sum of RVs

• By our rules for uncertainty transforms:

\[ \Sigma_{yy} = F \Sigma_{xx} F^T \]

• Where, in this case:

\[ F = \begin{bmatrix} 1 & 1 & \ldots & 1 \end{bmatrix} \]

• Hence:

\[ \sigma_y^2 = F \Sigma_{xx} F^T = \sum_{i=1}^{n} \sigma_{x_i}^2 \]

• IOW:

\[ \sigma_y^2 = n \sigma_x^2 \]
5.1.3.3 Variance of an Average of RVs

• Suppose there are \( n \) random variables \( x_i \) of same distribution.

\[ x_i \sim \mathcal{N}(\mu, \sigma), \quad i = 1, n \]

Variance of x’es known and equal.

• Define a new variable \( y \) as the average of these:

\[ y = \frac{1}{n} \sum_{i=1}^{n} x_i \]

• What is the variance of \( y \)?
5.1.3.3 Variance of an Average of RVs

• By our rules for uncertainty transforms:

\[ \sum_{yy} = F \sum_{xx} F^T \]

• Where, in this case:

\[ F = \frac{1}{n} \begin{bmatrix} 1 & 1 & \ldots & 1 \end{bmatrix} \]

• Hence:

\[ \sigma_y^2 = J \sum_{xx} J^T = \frac{1}{n^2} \sum_{i=1}^{n} \sigma_{x_i}^2 \]

• IOW:

\[ \sigma_y^2 = \frac{1}{n} \sigma_x^2 \]
5.1.3.4 Coordinate Transformations

- Know covariance in one frame (because it's easy to express there).
- Want to know it in another frame.
5.1.3.4 Coordinate Transformations

• If the transform between frames is:

\[ b\bar{x} = R^a \bar{x} + \bar{t} \]

• The transformed mean and covariance are:

\[ b\bar{x} = R^a \bar{x} + \bar{t} \]
\[ b\Sigma = R^a \Sigma R^T \]

Translation part does not affect variance so it's irrelevant.
5.1.3.5 Nonlinear Transformation: Mean

• Suppose we know \( \mu_x \) and want \( \mu_y \) where:

\[
y = f(x)
\]

• Write \( x \) in terms of a deviation from a reference \( x' \):

\[
x = x' + \varepsilon
\]

• Can use Jacobian to linearize:

\[
y = f(x) = f(x' + \varepsilon) \approx f(x') + J\varepsilon
\]

• The mean of the distribution of \( y \) is...

• \( x' \) is not random, so...

• If \( \varepsilon \) is unbiased, then..........

• And if we choose ...........

\[
\mu_y = \text{Exp}(y) = f(\mu_x)
\]

Mean of the \( f() \) is the \( f() \) of the mean.

"to first order"

"for unbiased error"
5.1.3.5 NonLinear Transformation: Covariance

• Rewriting:

\[ y = f(x) = f(x' + \varepsilon) \approx f(x') + J_\varepsilon \varepsilon \]
\[ y - y' = J_\varepsilon \varepsilon \]

• By definition:

\[ \Sigma_{yy} = \text{Exp} \left([y - y'][y - y']^T\right) \]

• Which is:

\[ \Sigma_{yy} = \text{Exp}(J_\varepsilon \varepsilon^T J^T) \]
Linearization: Again

• Whenever you write:

\[ \Sigma_{yy} = J \Sigma_{xx} J^T \]

• Unless all derivatives beyond J vanish (i.e. unless the mapping from x to y really is linear)
  – You have written an approximation.
5.1.3.6 Covariance with Partitioned Inputs

• Suppose we have: \( y = f(\bar{x}) \), \( \bar{x} = [x_1 \ x_2]^T \)

• Partition the Jacobian and the Covariance:

\[
J_x = \begin{bmatrix} J_1 & J_2 \end{bmatrix} \quad \Sigma_{xx} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}
\]

• We already know that the covariance of \( y \) is:

\[
\Sigma_{yy} = J_x \Sigma_{xx} J_x^T \quad \Rightarrow \quad \Sigma_{yy} = \begin{bmatrix} J_1 & J_2 \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} J_1^T \\ J_2^T \end{bmatrix}
\]
Uncorrelated Partitioned Inputs

• Suppose we have:

\[ \Sigma_{12} = \Sigma_{21} = [0] \]

\[ \therefore \Sigma_{yy} = \begin{bmatrix} J_1 & J_2 \end{bmatrix} \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{bmatrix} \begin{bmatrix} J_1^T \\ J_2^T \end{bmatrix} \]

• Hence:

\[ \Sigma_y = J_1 \Sigma_{11} J_1^T + J_2 \Sigma_{22} J_2^T \]

Uncertainties of uncorrelated inputs add to produce the output uncertainty in \( y=f(x_1,x_2) \)
Box 5.1 Formulae for Transformation of Uncertainty

For the following nonlinear transformation relating random vector $\vec{x}$ to random vector $\vec{y}$:

$$\vec{y} = f(\vec{x})$$

The mean and covariance of $\vec{y}$ are related to those of $\vec{x}$ by:

$$\mu_y = f(\mu_x) \quad \Sigma_{yy} = J \Sigma_{xx} J^T$$

Remember that, when using this result, unless all derivatives beyond $J$ vanish (unless the original mapping really was linear), the result is a linear approximation to the true mean and covariance.

When $\vec{x}$ can be partitioned into two uncorrelated components, then:

$$\Sigma_{yy} = J_1 \Sigma_{11} J_1^T + J_2 \Sigma_{22} J_2^T$$
5.1.3.7 Example: Azimuth Scanner
Transforming Uncertainty from ‘s’ to ‘w’

Know This

Want to Know This
(world coords)
Step 1: From i to s

\[ \mathbf{v}_s = \begin{bmatrix} x_s \\ y_s \\ z_s \end{bmatrix} = \begin{bmatrix} Rc\psi c\theta \\ -Rs\psi \\ -Rc\psi s\theta \end{bmatrix} \]

- Differentiate:

\[ J_i^s = \begin{bmatrix} R & \psi & \theta \\ x & c\psi c\theta & -Rs\psi c\theta & -Rc\psi s\theta \\ y & -s\psi & -Rc\psi & 0 \\ z & -c\psi s\theta & Rs\psi s\theta & -Rc\psi c\theta \end{bmatrix} \]
Step 1: Transformation

• Assume we know:

\[ \Sigma_i = \begin{bmatrix} \sigma_{RR} & 0 & 0 \\ 0 & \sigma_{\theta\theta} & 0 \\ 0 & 0 & \sigma_{\psi\psi} \end{bmatrix} \]

• Diagonal = “uncorrelated”.

• So…….

\[ \Sigma_S = J_i^S \Sigma_i (J_i^S)^T \]
Step 2: From s to w

- $T_s^w$ matrix relates s to w.
- Translation part is additive and irrelevant, so....

\[
\Sigma_w = R_s^w \Sigma_s \left( R_s^w \right)^T
\]

\[
\Sigma_w = R_s^w J_i \Sigma_i (J_i^s)^T \left( R_s^w \right)^T
\]

\[
T_s^w = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & h \\
0 & 0 & 0 & 1 \\
\end{bmatrix}
\]

\[
R_s^w = \begin{bmatrix}
c\theta & 0 & s\theta & 0 \\
0 & 1 & 0 & 0 \\
-s\theta & 0 & c\theta & 0 \\
0 & 0 & 0 & 1 \\
\end{bmatrix}
\]
5.1.3.8 Example: Attitude from Terrain Map

\[ \theta = \frac{(z_f - z_r)}{L} \]

- Find uncertainty in computed pitch angle given uncertainty in terrain
  - Which came from uncertainty in sensor.
5.1.3.8 Example: Attitude from Terrain Map

• Suppose the uncertainty in elevation is:

• Variance of computed pitch angle is:

• Where the Jacobian in this case is a gradient:

• The result is:

\[ \Sigma_z = \begin{bmatrix} \sigma_f^2 & 0 \\ 0 & \sigma_r^2 \end{bmatrix} \]

\[ \Sigma_\theta = J \Sigma_z J^T \]

\[ J = \begin{bmatrix} \frac{\partial \theta}{\partial z_f} & \frac{\partial \theta}{\partial z_r} \end{bmatrix} = \begin{bmatrix} \frac{1}{L} & -\frac{1}{L} \end{bmatrix} \]

\[ \sigma_{\theta}^2 = \frac{1}{L^2} [\sigma_f^2 + \sigma_r^2] \]
5.1.3.9 Example: Range Error in Rangefinders

• Where do the input variances come from?
• Variance in measured range depends on Range (R) reflectance (r) and incidence angle (a).

\[ \sigma_R \propto \left[ \frac{\lambda R^2}{\rho \cos \alpha} \right] \]
5.1.3.9 Example: Range Error in Rangefinders (Real Data)

- Normal incidence, various reflectance.
- Dark surfaces are 10 X noisier.
- Fit lines of the form:

\[ \sigma_R = k_1 (\rho) R^2 \]
5.1.3.9 Example: Range Error in Rangefinders (Real Data)

- White target, various ranges.
- Fit lines of the form:

\[
\sigma_R = \frac{k_2(R)}{\cos \alpha}
\]
5.1.3.10 Example: Stereo Vision

• From similar triangles:

\[
\frac{Y_L}{X_L} = \frac{Y_L}{X} = \frac{y_l}{f} \quad \frac{Y_R}{X_R} = \frac{Y_R}{X} = \frac{y_r}{f}
\]

• Subtract:

\[
Y_L - Y_R = \frac{X[x_l - x_r]}{f}
\]

• Hence:

\[
X = \frac{bf}{d} \quad b = \frac{Xd}{f}
\]
5.1.3.11 Example: Stereo Uncertainty

- Define the normalized disparity:
  \[ \delta = \frac{d}{f} = \frac{b}{X} \]

- Now triangulation looks like:
  \[ X = \frac{b}{\delta} \]

- Uncertainty transformation:
  \[ \sigma_{xx} = J \sigma_{\delta \delta} J^T \]

- Jacobian is the scalar:
  \[ J = \frac{\partial X}{\partial \delta} = \frac{-b}{\delta^2} \]

- Variance goes with 4th power of range:
  \[ \sigma_{xx} = \begin{bmatrix} \frac{b^2}{\delta^4} \end{bmatrix} \sigma_{\delta \delta} = \begin{bmatrix} \frac{X^4}{b^2} \end{bmatrix} \sigma_{\delta \delta} \]

- Standard deviation with the square of range.
  – Famous result.
Outline

• 5.1 Random Variables, Processes and Transformation
  – 5.1.1 Characterizing Uncertainty
  – 5.1.2 Random Variables
  – 5.1.3 Transformation of Uncertainty
  – 5.1.4 Random Processes - SKIP
  – Summary
Outline

• 5.1 Random Variables, Processes and Transformation
  – 5.1.1 Characterizing Uncertainty
  – 5.1.2 Random Variables
  – 5.1.3 Transformation of Uncertainty
  – 5.1.4 Random Processes
  – Summary
Summary

• There are many kinds of error.
  – They can be removed with calibration, filtering.
• Covariance measures spread. Level curves are ellipsoids.
• Covariance is transformed with a matrix quadratic form.
• Variance of a random walk process grows linearly with time.
• Stochastic Diff Eqs are almost as easy to solve as deterministic ones.