Data Fusion III – Estimation Theory
QUIZ Response

Table of Contents

Problem 1 1
- Part A (40%) 1
- Part B (40%) 2
- Part C (20%) 2

Response 2

Problem 2 4
- Part A (50%) 4
- Part B (50%) 4

Response 4

Problem 3 5
- Part A (50%) 5
- Part B (50%) 6

Response 6

Problem 4 6
- Part A (50%) 6
- Part B (50%) 7

Response 7

Problem 5 9
- Part A (50%) 9
- Part B (50%) 9

Response 10

Problem 1

Part A (40%)

Define these terms and their place in estimation theory:

- A *statistic*
- An *estimator*

- A *confidence limit*
- An *unbiased estimator*
- A *consistent estimator*
-

Part B (40%)

- Take a few minutes to explain what a *sufficient estimator* is.
- Define an *invariant* estimator.
- What is an *efficient* estimator?

Part C (20%)

The Kalman update is an estimator that uses the extrapolated state vector $\underline{\tilde{x}}$ and the measurement data vector $\underline{y}$ to compute an estimator $\underline{\hat{x}}$ for the state vector $\underline{x}$,

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

where K is the Kalman gain,

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

$\underline{y}(\underline{x})$ is an algebraic model of the measurements $\underline{y}$ in terms of the states $\underline{x}$, and H is the gradient of $\underline{y}$ with respect to $\underline{x}$

$$H = \left. \frac{\partial \underline{h}\left(\underline{x} \right)}{\partial \underline{x}} \right|_{\underline{x} = \underline{\tilde{x}}}$$

and $\tilde{P}$ is the covariance of the extrapolated state vector $\underline{\tilde{x}}$. The covariance of the updated state vector estimate $\underline{\hat{x}}$ is

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

Under what conditions is a Kalman update an efficient estimator of the state vector?

Response

From the lecture notes and handouts of February 23, 2—6:

- A *statistic* is a function of data. Examples are the sample mean, the sample variance, the sample median, and the sample quartiles.
- A function of data which approximates an unknown parameter is an *estimator*.

- A *confidence limit* is a probability that is set as part of an experiment design as an a priori threshold on the significance of results of the experiment.
- An *unbiased estimator* is one whose ensemble mean is equal to the value of the unknown parameter.
- A *consistent estimator* is one which approaches the value of the unknown parameter as the number of measurements increases without bound.
- A *sufficient estimator* is one that uses all the information in the available data that applies to estimating the value of the unknown parameters.

The *Neyman-Fisher factorization theorem* states that an estimator of a set of states $\underline{x}$ is sufficient if and only if the conditional probability density of the measurement data $\underline{y}$ given the states $\underline{x}$ can be factored as

$$p\left(\underline{y}\middle|\underline{x}\right)=g\left(\underline{x},\underline{\phi}\left(\underline{y}\right)\right)\cdot h\left(\underline{y}\right) \tag{1.1}$$

Let's use this theorem to look at our starting point for maximum likelihood estimators, $p\left(\underline{y}\middle|\underline{x}\right)$. We know from Bayes' definition of conditional probability that

$$p\left(\underline{x},\underline{y}\right)=p\left(\underline{y}\middle|\underline{x}\right)\cdot p\left(\underline{x}\right)=p\left(\underline{x}\middle|\underline{y}\right)\cdot p\left(\underline{y}\right) \tag{1.2}$$

Thus we can write

$$p\left(\underline{x}\middle|\underline{y}\right)=\frac{p\left(\underline{y}\middle|\underline{x}\right)\cdot p\left(\underline{x}\right)}{p\left(\underline{y}\right)}=\frac{g\left(\underline{x},\underline{\phi}\right)\cdot h\left(\underline{y}\right)\cdot p\left(\underline{x}\right)}{p\left(\underline{y}\right)} \tag{1.3}$$

where $p\left(\underline{y}\right)$ is defined as

$$\begin{aligned}p\left(\underline{y}\right)&=\int p\left(\underline{x},\underline{y}\right)\cdot d\underline{x}=\int p\left(\underline{y}\middle|\underline{x}\right)\cdot p\left(\underline{x}\right)\cdot d\underline{x}\\&=h\left(\underline{y}\right)\cdot\int g\left(\underline{x},\underline{\phi}\right)\cdot p\left(\underline{x}\right)\cdot d\underline{x}\end{aligned} \tag{1.4}$$

We combine (1.3) and (1.4), noting that $h\left(\underline{y}\right)$ cancels, and we have

$$p\left(\underline{x}\middle|\underline{y}\right)=\frac{g\left(\underline{x},\underline{\phi}\right)\cdot p\left(\underline{x}\right)}{\int g\left(\underline{x},\underline{\phi}\right)\cdot p\left(\underline{x}\right)\cdot d\underline{x}}\equiv p\left(\underline{x}\middle|\underline{\phi}\right) \tag{1.5}$$

That is, the *a priori* probability density function of $\underline{x}$ given the data $\underline{y}$ is the same as the *a priori* probability density given the sufficient estimator $\underline{\phi}\left(\underline{y}\right)$. Thus we have all the information about $\underline{x}$ that is contained in the data $\underline{y}$ also contained in the sufficient estimator $\underline{\phi}\left(\underline{y}\right)$.

Part C is explained in the lecture, notes, and handout of February 23. Section 6.4 showed that estimators that are derived by the method of maximum likelihood are efficient only if the states are linearly related to the measurements (but that they were asymptotically efficient in any case). Section 7 showed that the Kalman update can be derived by the method of maximum likelihood, and the covariance update is also a result of analyzing the Kalman update using the method of maximum likelihood. Thus the Kalman update is efficient only if the states are linearly related to the measurements. If the states are not linearly related to the measurements, then the Kalman update is not efficient, is not unbiased, and the covariance update is not achieved because it is the Cramer-Rao bound and achieved only by efficient estimators. Asymptotic efficiency is not a consideration for the Kalman update because only one update is done with one set of data. An exception might be taken if there is no process noise and the state extrapolation is linear, so that the Kalman updates over time are algebraically equivalent to a batch estimator.

Problem 2

Part A (50%)

Define these three estimator techniques and briefly describe the distinctions between them:

- Maximum likelihood estimator
- Maximum *a posteriori* estimator
- Bayesian mean

Part B (50%)

Given an unknown state vector $\underline{x}$ and a set of measurements $\underline{y}$, and the conditional probability density function $p\left(\underline{y}\middle|\underline{x}\right)$, write the equation for the general form for:

- The *likelihood function*
- The *log likelihood function*
- The *likelihood equation*
- The *Cramer-Rao Bound*

Response

Part A is given in the lecture, notes and handouts for February 23, 2006, section 5.

The maximum likelihood estimator poses the likelihood function of the available data $\underline{y}$ given a particular set of states $\underline{x}$ as the conditional probability density function $p\left(\underline{y}\middle|\underline{x}\right)$.

The maximum likelihood estimator is the set of states $\underline{x}$ that maximizes this probability density function for the measurements $\underline{y}$ as observed. Since the likelihood function is usually quite easily written, the method of maximum likelihood is usually very simple to apply to practical problems.

The maximum *a posteriori* (MAP) estimator finds the set of states that maximizes the *a posteriori* probability density function, which is defined as $p\left(\underline{x}\middle|\underline{y}\right)$. Bayes' theorem shows that the *a posteriori* probability density function and the likelihood function are related by

$$p\left(\underline{x}\middle|\underline{y}\right)=\frac{p\left(\underline{y}\middle|\underline{x}\right)\cdot p\left(\underline{x}\right)}{p\left(\underline{y}\right)} \tag{2.1}$$

where we define $p\left(\underline{y}\right)$ as

$$p\left(\underline{y}\right)=\int p\left(\underline{x},\underline{y}\right)\cdot d\underline{x}=\int p\left(\underline{y}\middle|\underline{x}\right)\cdot p\left(\underline{x}\right)\cdot d\underline{x} \tag{2.2}$$

This method can be applied when the probability density function of the states $p\left(\underline{x}\right)$ without knowledge of measurements can be written.

The Bayesian mean is defined as

$$\hat{\underline{x}}=\int \underline{x}\cdot p\left(\underline{x}\middle|\underline{y}\right)\cdot d\underline{x} \tag{2.3}$$

and can be used when the *a posteriori* probability density function of $\underline{x}$ can be written.

Part B is given in the lectures and notes for February 23, 2006, section 6.

- The *likelihood function* is the conditional probability density function $p\left(\underline{y}\middle|\underline{x}\right)$.
- The *log likelihood function* is the natural logarithm of the likelihood function, $\ln\left(p\left(\underline{y}\middle|\underline{x}\right)\right)$.
- The *likelihood equation* is

$$\frac{\partial \ln\left(p\left(\underline{y}\middle|\underline{x}\right)\right)}{\partial \underline{x}}=\underline{0} \tag{2.4}$$

- The *Cramer-Rao Bound* is the minimum covariance possible and is given by

$$P^{-1}=-\frac{\partial^2\left(\ln\left(p\left(\underline{y}\middle|\underline{x}\right)\right)\right)}{\partial \underline{x}^2} \tag{2.5}$$

Note the minus sign and the inverse. The inverse of the covariance matrix, when the covariance matrix is the Cramer-Rao bound, is called the Fisher information matrix.

Problem 3

Part A (50%)

Given a set of M samples of Gaussian random variables that have a mean of m and a variance of σ^2, we will write down a maximum likelihood estimator (MLE) for the mean. The measurements are of the form

$$y_i = m + v_i$$

where the measurement noises v_i are Gaussian random variables with zero mean and variance σ^2. Write down

- The likelihood function.
- The log likelihood function.
- The likelihood equation for *m*.
- The estimator for *m*.
- The variance of the estimator for *m*.

Use the Cramer-Rao bound as your estimator for the variance of *m*.

Part B (50%)

Modify the estimator for Part A to include the variance σ^2. Using the log likelihood function from Part A, write down

- The likelihood equation for σ^2.
- The estimator for σ^2.
- The variance of your estimator for σ^2.

Use the Cramer-Rao bound as your estimator for the variance of σ^2.

Response

Part A is exactly the same problem presented in detail on March 2, 2006, Section 1, with the simplification that only the sample mean is required by the problem. Please refer to those notes and handouts.

Part B asks for the estimate of the variance, and is part of the example presented on March 2, 2006, in the handouts Section 1.

Problem 4

Building a radar tracker, "Care and Feeding of Radar Trackers," is the problem we will address here.

Part A (50%)

When an estimate of SNR is available from any means, write down a working equation for the variance of a measurement. Then, give

- The lowest possible measurement variance for very high SNR
- The measurement variance for zero SNR.
- The term c_x for moderate SNR in the form

$$\sigma_y^2 \approx \frac{c_x}{SNR}$$

Part B (50%)

Use the equations for the "Snake-oil" two-state tracker presented on March 16 in this part. Using the form for σ_y^2 for moderate SNR from Part A, write

- The equation for the Kalman gain in terms of the SNR.
- The equation for the updated state covariance P in terms of the SNR.
- The ratio of the determinants of the extrapolated and updated covariance matrices in terms of the SNR.

Response

In the lecture notes and handouts for March 16, 2006, section 1.1 addresses Part A. Variances of individual measurements as a function of signal-to-noise ratio (SNR) usually follow the form

$$\sigma_y^2 = c_0 + \frac{c_1}{1 + c_2 \cdot SNR}. \tag{4.1}$$

Note that SNR is a power ratio; never use decibels in an equation. Angle measurements using monopulse need a slightly more complicated equation, but we are not concerned with them here.

- The lowest possible measurement variance for very high SNR is c_0. This term is where you put in the lower limits due to measurement quantization and other sources of noise that will always be present regardless of how high the SNR is.
- The measurement variance for zero SNR is $c_0 + c_1$. Clearly, this term will be dominated by c_1, so c_1 is where you put the number your analysis shows is the measurement variance for zero SNR. This value may not be given directly but can be part of the process of setting c_2.
- The equation given shows that the measurement variance begins to be affected by the SNR when $SNR > 1/c_2$. For $SNR \gg 1/c_2$, the measurement variance is essentially given by

  $$\sigma_y^2 \approx \frac{c_1/c_2}{SNR}, \frac{1}{c_2} \ll SNR \ll \frac{1}{c_0} \cdot \frac{c_1}{c_2} = \frac{1}{c_2} \cdot \frac{c_1}{c_0} \tag{4.2}$$

  If the SNR is always significantly greater than zero (as is usually the case) and the measurement variance is always inversely proportional to SNR, the "1" can be dropped and only the coefficient $c_x = c_1/c_2$ used.

Part B asks that we use (4.2) in the Snake Oil Tracker equations to show how SNR affects the Kalman filter parameters. Please refer to the lecture notes and handouts from March 16, 2006, section 2.3 for the Snake Oil Tracker update equations. The Kalman gain is

$$\begin{aligned} K &= \frac{1}{\tilde{p}_{11} + \sigma_y^2} \cdot \begin{bmatrix} \tilde{p}_{11} \\ \tilde{p}_{12} \end{bmatrix} = \frac{1}{\tilde{p}_{11} + c_x / SNR} \cdot \begin{bmatrix} \tilde{p}_{11} \\ \tilde{p}_{12} \end{bmatrix} \\ &\approx \begin{bmatrix} 1 - \dfrac{\sigma_y^2}{\tilde{p}_{11}} \\ \dfrac{\tilde{p}_{12}}{\tilde{p}_{11}} \end{bmatrix} = \begin{bmatrix} 1 - \dfrac{c_x}{\tilde{p}_{11} \cdot SNR} \\ \dfrac{\tilde{p}_{12}}{\tilde{p}_{11}} \end{bmatrix}, \sigma_y^2 \ll \tilde{p}_{11} \end{aligned} \tag{4.3}$$

The equation for the updated state covariance matrix P in terms of the SNR is

$$\begin{aligned} P &= \frac{1}{1 + \tilde{p}_{11} / \sigma_y^2} \cdot \begin{bmatrix} \tilde{p}_{11} & \tilde{p}_{12} \\ \tilde{p}_{12} & \tilde{p}_{22} + \left|\tilde{P}\right| / \sigma_y^2 \end{bmatrix} \\ &\approx \frac{\sigma_y^2}{\tilde{p}_{11}} \cdot \tilde{P} + \begin{bmatrix} 0 & 0 \\ 0 & \dfrac{\left|\tilde{P}\right|}{\tilde{p}_{11}} \end{bmatrix} = \frac{c_x}{\tilde{p}_{11} \cdot SNR} \cdot \tilde{P} + \begin{bmatrix} 0 & 0 \\ 0 & \dfrac{\left|\tilde{P}\right|}{\tilde{p}_{11}} \end{bmatrix} \end{aligned} \tag{4.4}$$

The ratio of the determinants is

$$\begin{aligned} \frac{\left|\tilde{P}\right|}{\left|P\right|} &= 1 + \frac{\tilde{p}_{11}}{\sigma_y^2} \\ &\approx \frac{\tilde{p}_{11} \cdot SNR}{c_x}, \sigma_y^2 \ll \tilde{p}_{11} \end{aligned} \tag{4.5}$$

We observe from (4.3) that when the SNR is high and the measurement variance is low compared to the variance of the position state, the Kalman gain essentially takes the measurement data for the position state. Further examination for high SNR for two updates with high SNR will reveal that the velocity state is essentially estimated as the change in position from the last two updates, divided by the time between updates.

We observe from (4.4) that the covariance matrix is reduced by a factor of the ratio of the measurement and position state variances, except for p_{22}, which is essentially left near $\tilde{p}_{22}$ but reduced by $\tilde{p}_{12}^2 / \tilde{p}_{11}$. For two or more successive updates, this also dramatically reduces $\tilde{p}_{22}$. This is further borne out by examining(4.5), which shows that the determinant of the covariance matrix is reduced by a factor proportional to the SNR, and $\tilde{p}_{22}$ as given by (4.4) is proportional to the determinant of the extrapolated covariance matrix.

One can conclude that nonzero plant noise must be included for the velocity state or, for high SNR, the covariance matrix will decrease geometrically and covariance collapse will inevitably result, even for the Snake-Oil tracker.

Problem 5

In this problem you will be asked for your judgement as a system engineer. Your grade will be based on your reasons as much as your conclusions.

Part A (50%)

You are asked to provide a preliminary recommendation for the configuration of a radar tracker for a new radar, not an ASR-9 or DASR but a radar that is offered for networking by another Agency. This new radar will provide azimuth tick information, plus time tag information, and horizontal monopulse azimuth data. SNR will be available but not measurement variances. The SSR data is present, and the SSR antenna is mounted on top of the PSR antenna. The azimuth bias of this radar is unknown but uniform over 360 degrees – that is, azimuth bias is not a function of aircraft azimuth. In addition to aircraft angle, aircraft range rate information is available as Doppler information. The range of this radar for transport aircraft at altitude is 120 nautical miles. It is not located near an airport but air traffic routes pass over the coverage area of the radar.

The aircraft environment at this radar location is described as sparse to moderate, with no more than 10 radar contacts. The output of this radar is used to supplement ARSR radars but is not used to merge with data of radars that track the same aircrafts simultaneously. You need to make recommendations based on the following alternatives:

a) Would you recommend a Kalman filter or an alpha-beta filter? Why?

b) Do you need to estimate an additional azimuth bias state? Why?

c) Do you recommend that IMM be incorporated into the tracker design? Why?

d) Do you recommend an MHT layer on the tracker for track-before detect? Why?

e) If you estimate azimuth bias, would you estimate is with the tracker or add a batch estimator specifically for estimation of azimuth bias? Why?

Part B (50%)

We have a radar identical to the one in Part A but at a different site. This site is near an airport and the new radar's coverage area overlaps two existing ASR radar coverage areas and some ARSR coverage areas. Data from all radars that detect a particular aircraft will be fused, and a central C2 database will support an ATC facility at the airport. The aircraft environment at this radar location is described as dense, with 100 to 300 aircraft in its coverage area. Data from existing radars is available for estimating azimuth bias. You need to make recommendations based on the following alternatives:

a) Would you recommend a Kalman filter or an alpha-beta filter? Why?

b) Do you need to estimate an additional azimuth bias state? Why?

c) Do you recommend that IMM be incorporated into the tracker design? Why?

d) Do you recommend an MHT layer on the tracker for track-before detect? Why?

e) If you estimate azimuth bias, would you estimate is with the tracker or add a batch estimator specifically for estimation of azimuth bias? Why?

Response

There is no right or wrong answer for this problem. In particular, the requirements are given in broad terms, and the details may be interpreted differently by different people. Your score is based on your interpretation of the requirements and how different solutions as spelled out in the problem will meet the requirements.

We have a typical situation where an air surveillance radar is made available from another Agency. This radar has horizontal monopulse data and Doppler information, and an SSR is incorporated into the main antenna.

For Part A, the mission of this radar is to supplement ARSR (long range route surveillance) ATC radars. Thus this new radar is basically a gap-filler and its performance requirements are similar to those of ARSR radars except possibly for detection range. Good responses to Part A are:

- An alpha-beta tracker is adequate to meet performance requirements for the ARSR mission, so this is an acceptable recommendation. However, a Kalman filter can accept Doppler measurements and better use monopulse data, and as such make it simpler to integrate SSR data, rather than pass PSR and SSR data to the C2 net separately. As we see from the Snake-Oil tracker, little additional complexity is needed for a Kalman filter over an alpha-beta tracker. In addition, estimation of biases may be possible with this radar, and as such it may provide a first step in upgrading the ARSR integration with the ATC network. Thus a Kalman filter may be more useful in upgrading ARSR support to the C2 net. An air surveillance radar with monopulse measurements, Doppler measurements, and an integrated SSR will have more than sufficient on-board data processing capacity to do data fusion between the PSR and SSR and support the C2 interface. This capability should be supported if the C2 net, or upgrades to it, require this performance.
- Estimation of azimuth bias is not required for ARSR support, so an extra bias state on a Kalman tracker is not required. In any case, bias can be estimated with an extra state on the Kalman filter but a simpler tracker architecture may result from use of a batch estimator for biases.
- En-route surveillance is not likely to need IMM, even if high-performance aircraft comprise some of the radar contacts.
- En-route surveillance of low-traffic areas would not benefit from MHT.
- As mentioned above, a batch estimator for biases would result in simpler trackers.

For part B, our new radar is to be integrated into the ASR net in a high-traffic area near an airport. Here, the azimuth monopulse, Doppler measurement, and integrated SSR capabilities are more than welcome. Acceptable responses are:

- We surely need a Kalman filter to do 2-D tracking using the azimuth monopulse and Doppler measurements. We would use polar coordinates internally to the Kalman filter to achieve efficient estimation in the Kalman updates.
- We would want to estimate biases to allow close coupling of radars in the C2 net. Our radar will have two-way C2 capability and we may want to use it to estimate biases. We may want to use an extra state or states if our own azimuth bias is a function of antenna rotation angle or weather conditions – i.e., varying with time and antenna rotation.
- Near an airport, we would want IMM to detect maneuvers on the first hit rather than wait for turns to be detected after several hits.
- In a dense environment where altitude is not measured, MHT can support continued high performance in the scenarios that inevitably result in some misassociations. This can be a real asset in the ASR missions.
- We may use a bias state or states if our own bias is a function of time (wind) and antenna angle. If we are estimating a constant angle bias – a radar site calibration setup error – then a batch estimator is probably a better answer because it provides better data and leaves the tracker without the complication of an added state.

Interpretation of the requirements and understanding of how different tracking and estimation techniques support different requirements is most important here; more than one solution is acceptable for any problem in the real world.