Radar Trackers

Study Guide

All chapters, problems, examples and page numbers refer to *Applied Optimal Estimation*, A. Gelb, Ed.

Chapter 1

Example 1.0-1

Problem Statement

Two sensors

- Each has a single noise measurement
- $z_i = x + v_i$
- Unknown *x* is a constant
- Measurement noises v_i are uncorrelated (generalized in Problem 1-1 pages 7 and 8)
- Estimate of x is a linear combination of the measurements that is not a function of the unknown to be estimated, *x*

We define a measurement vector $\underline{y}$,

$$\underline{y} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} x + v_1 \\ x + v_2 \end{bmatrix} = x \cdot \underline{1} + \underline{v}$$

where $\underline{1}$ is a vector with all 1's, and a linear estimator gain $\underline{k}$,

$$\underline{k} = \begin{bmatrix} k_1 \\ k_2 \end{bmatrix}$$

and an estimate is a linear combination of the measurements:

$$\hat{x} = \underline{k}^T \cdot \underline{y} .$$

We define an error (note difference in notation from the example, for consistency with later usage of similar notation)

$$x_e = \hat{x} - x .$$

We require that $\underline{k}$ be independent of *x* and that the mean of the estimate be equal to *x*:

$$E\left(\hat{x}\right) = x$$

where *E(*)* is the expectation, or ensemble mean, operator.

Solution: We substitute the equations for the estimate and measurement to find a constraint on the linear estimator gain $\underline{k}$ as follows:

$$E(\hat{x}) = E\left(\underline{k}^T \cdot \underline{y}\right) = E\left(\underline{k}^T \cdot \left(x \cdot \underline{1} + \underline{v}\right)\right) = x\,.$$

The hypothesis of zero mean measurement noise is, in equation form,

$$E(\underline{v}) = \underline{0}$$

and we have

$$\underline{k}^T \cdot \underline{1} = 1$$

as the condition on the linear estimator gain. Note that another solution is x=0, which we discard since x is an unknown and is not to be constrained, since we have required that $\underline{k}$ not depend on x.

We minimize the mean square error, and denote what we are minimizing as J, which we will call a cost function. This is a term for an optimization criterion. This is

$$J = E\left(x_e^2\right) = E\left(\left(\hat{x} - x\right)^2\right).$$

Again substituting from the above, we have

$$\begin{aligned}
J &= E\left(\left(\underline{k}^T \cdot \left(x \cdot \underline{1} + \underline{v}\right) - x\right)^2\right) \\
&= E\left(\left(x \cdot \left(\underline{k}^T \cdot 1\right) + \underline{k}^T \cdot \underline{v} - x\right)^2\right) \\
&= E\left(x^2 \cdot \underline{k}^T \cdot \left(\underline{1}^T \cdot \underline{1}\right) \cdot \underline{k}\right) + 2x \cdot E\left(\underline{k}^T \cdot \underline{1} \cdot \underline{v}^T \cdot \underline{k}\right) + \underline{k}^T \cdot E\left(\underline{v} \cdot \underline{v}^T\right) \cdot \underline{k} \\
&\quad -2x \cdot E\left(x \cdot \left(\underline{k}^T \cdot 1\right) + \underline{k}^T \cdot \underline{v}\right) + x^2 \\
&= x^2 + \underline{k}^T \cdot R_v \cdot \underline{k} - 2x^2 + x^2 \\
&= \underline{k}^T \cdot R_v \cdot \underline{k}
\end{aligned}$$

One term in this derivation contains a very special matrix, the *covariance matrix* of the measurement noise vector $\underline{v}$:

$$E\left(\underline{v} \cdot \underline{v}^T\right) = E\left(\begin{bmatrix} v_1^2 & v_1 \cdot v_2 \\ v_1 \cdot v_2 & v_2^2 \end{bmatrix}\right) = \begin{bmatrix} \sigma_1^2 & \rho \cdot \sigma_1 \cdot \sigma_2 \\ \rho \cdot \sigma_1 \cdot \sigma_2 & \sigma_2^2 \end{bmatrix} = R_v\,.$$

We will leave in the correlation coefficient ρ and not take it as zero as in the example, because, as we will quickly see, there is no immediate advantage in simplicity in dropping it out. We well make it zero when appropriate later.

We need to minimize J with respect to the linear estimator gain $\underline{k}$, subject to the condition for an unbiased estimate. We can do this directly, at the expense of the simple equation for the optimization criteria J that we have, and in the process make our solution specific to the problem statement and give up its generality for other problems. A way to keep the problem linear and in the vector domain is to add a new equation with an additional unknown and link it to the original equation. This is called the method of Lagrangian multipliers. For our problem, the optimization criteria J becomes

$$J = \underline{k}^T \cdot R_v \cdot \underline{k} - \lambda \cdot \left(\underline{k}^T \cdot \underline{1} - 1\right).$$

The new variable λ is the Lagrangian multiplier λ, and the optimization criteria is linear in λ and quadratic in $\underline{k}$. What we will do is relax the constraint on $\underline{k}$ and apply the unbiased constraint on the solution to find a value for λ to complete our solution.

We proceed by taking the gradient of the optimization criteria with respect to $\underline{k}$ and set it equal to zero and find a solution:

$$\frac{\partial J}{\partial \underline{k}} = 2 \cdot R_v \cdot \underline{k} - \lambda \cdot \underline{1} = \underline{0}$$

$$\underline{k} = \frac{\lambda}{2} \cdot R_v^{-1} \cdot \underline{1}.$$

We use the unbiased condition to find the value of the new variable λ:

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

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

Note that

$$\underline{1}^T \cdot R_v^{-1} \cdot \underline{1} = \left[\text{sum of all the terms of } R_v^{-1} \right].$$

This leaves us with an equation for the linear estimator gain k:

$$\underline{k} = \frac{R_v^{-1} \cdot \underline{1}}{\underline{1}^T \cdot R_v^{-1} \cdot \underline{1}}.$$

This gives us a minimum value of the optimization criteria of

$$J_{MIN} = \frac{1}{\underline{1}^T \cdot R_v^{-1} \cdot \underline{1}}.$$

We will close with an explicit expression for the inverse of the covariance matrix of $\underline{v}$:

$$\begin{aligned} R_v^{-1} &= \begin{bmatrix} \sigma_1^2 & \rho \cdot \sigma_1 \cdot \sigma_2 \\ \rho \cdot \sigma_1 \cdot \sigma_2 & \sigma_2^2 \end{bmatrix}^{-1} \\ &= \frac{1}{\sigma_1^2 \cdot \sigma_2^2 \cdot \left(1-\rho^2\right)} \cdot \begin{bmatrix} \sigma_2^2 & -\rho \cdot \sigma_1 \cdot \sigma_2 \\ -\rho \cdot \sigma_1 \cdot \sigma_2 & \sigma_1^2 \end{bmatrix} \\ &= \begin{bmatrix} \dfrac{1}{\sigma_1^2 \cdot \left(1-\rho^2\right)} & \dfrac{-\rho}{\sigma_1 \cdot \sigma_2 \cdot \left(1-\rho^2\right)} \\ \dfrac{-\rho}{\sigma_1 \cdot \sigma_2 \cdot \left(1-\rho^2\right)} & \dfrac{1}{\sigma_2^2 \cdot \left(1-\rho^2\right)} \end{bmatrix} \end{aligned}$$

The minimum value of the optimization criteria is

$$\begin{aligned} J_{MIN} &= \frac{\sigma_1^2 \cdot \sigma_2^2 \cdot \left(1-\rho^2\right)}{\sigma_1^2 + \sigma_2^2 - 2\rho \cdot \sigma_1 \cdot \sigma_2} \\ &= \frac{1-\rho^2}{\dfrac{1}{\sigma_1^2} + \dfrac{1}{\sigma_2^2} - \dfrac{2\rho}{\sigma_1 \cdot \sigma_2}} \end{aligned}$$

and the linear estimator gain $\underline{k}$ is

$$\underline{k} = \frac{1}{\sigma_1^2 + \sigma_2^2 - 2\rho\cdot\sigma_1\cdot\sigma_2}\cdot\begin{bmatrix}\sigma_2^2 - \rho\cdot\sigma_1\cdot\sigma_2 \\ \sigma_1^2 - \rho\cdot\sigma_1\cdot\sigma_2\end{bmatrix}$$

$$= \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} - \frac{2\rho}{\sigma_1\cdot\sigma}}\cdot\begin{bmatrix}\frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\cdot\sigma_2} \\ -\frac{\rho}{\sigma_1\cdot\sigma_2} & \frac{1}{\sigma_2^2}\end{bmatrix}.$$

This basic technique and result can be applied to higher order problems. In particular, see that problem 1-3 becomes simply a matter of writing the result, followed by a simple algebraic step. In fact, for any number of uncorrelated measurements M, the general result is obvious from

$$\underline{1}^T\cdot R_v^{-1}\cdot\underline{1} = \underline{1}^T\cdot\begin{bmatrix}\frac{1}{\sigma_1^2} & 0 & \cdots & \\ 0 & \frac{1}{\sigma_2^2} & & \\ \vdots & & \ddots & \\ & & & \frac{1}{\sigma_M^2}\end{bmatrix}\cdot\underline{1}$$

$$= \sum_{j=1}^{M}\frac{1}{\sigma_j^2}$$

and

$$R_v^{-1}\cdot\underline{1} = \begin{bmatrix}\frac{1}{\sigma_1^2} \\ \frac{1}{\sigma_2^2} \\ \vdots \\ \frac{1}{\sigma_M^2}\end{bmatrix}.$$

Gelb Problem 1-1

See the example for the estimation weights and minimum mean square error.

In this problem we see from the expression for the minimum mean square error

$$J_{MIN} = \frac{\sigma_1^2\cdot\sigma_2^2\cdot\left(1-\rho^2\right)}{\sigma_1^2 + \sigma_2^2 - 2\rho\cdot\sigma_1\cdot\sigma_2}$$

$$= \frac{1-\rho^2}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} - \frac{2\rho}{\sigma_1\cdot\sigma_2}}$$

that when ρ is large and the error variances are not the same, that the minimum mean square error is smaller. This means that we can use the information that the errors are correlated to help estimate the unknown *x*. We can investigate what happens near $\rho = 1$ by making the substitutions

$$\rho^2 = 1 - \delta$$

$$\rho = \pm\sqrt{1-\delta} \approx \pm\left(1 - \frac{\delta}{2}\right)$$

and looking at the minimum mean square error:

$$J_{MIN} = \frac{\delta}{\left(\left(\frac{1}{\sigma_1} \mp \frac{1}{\sigma_2}\right)^2 \pm \frac{\delta}{\sigma_1 \cdot \sigma_2}\right)}$$

$$\approx \frac{\delta}{\left(\frac{1}{\sigma_1} \mp \frac{1}{\sigma_2}\right)^2}$$

When $\rho = +1$ this means that the error in one of the measurements is proportional to the error in the other measurement. If the RMS measurement errors are different, we can use that knowledge to find the unknown *x* exactly. If they are equal, we have a singularity; we simply take the estimate as either of the measurements because they are equal When $\rho = -1$ we can always use the knowledge that the errors are proportional to each other to find the unknown exactly.

Gelb Problem 1-3

This problem was worked as part of Example 1.0-1 as presented avove.

Chapter 2

Gelb problem 2-1

Show that

$$\dot{P}^{-1} = -P^{-1} \cdot \dot{P} \cdot P^{-1}$$

Solution: We begin with

$$P^{-1} \cdot P = I$$

and take the derivative of both sides of this equation with respect to time. The result is, using the chain rule,

$$\dot{P}^{-1} \cdot P + P^{-1} \cdot \dot{P} = 0$$

and the result follows from right-multiplying by P^{-1} and moving the second term to the right-hand side.

Gelb problem 2-2

For the matrix

$$\begin{bmatrix} 3 & -1 & -4 \\ 2 & 2 & 2 \\ -4 & 0 & -2 \end{bmatrix}$$

show that the eigenvalues are 2, -4, and 5.

Solution: The characteristic equation is

$$\left\| \begin{bmatrix} 3-\lambda & -1 & -4 \\ 2 & 2-\lambda & 2 \\ -4 & 0 & -2-\lambda \end{bmatrix} \right\| = 0$$

Expanding the determinant by minors about the left column,

$$\begin{aligned} &(3-\lambda)\cdot(2-\lambda)\cdot(-2-\lambda) \\ &-2\cdot(-1)\cdot(-2-\lambda) \\ &+(-4)\cdot((-1)\cdot(2)-(-4)\cdot(2-\lambda)) \\ &= -\lambda^3 + 3\cdot\lambda^2 + 18\cdot\lambda - 40 \\ &= -(\lambda-2)\cdot(\lambda+4)\cdot(\lambda-5) \end{aligned}$$

Gelb Problem 2-3

Show that *A* is positive definite if, and only if, all of the eigenvalues of A are positive.

Solution: The definition of the property "positive definite" is on page 21:

$$\underline{x}^T \cdot A \cdot \underline{x} > 0$$

for all real $\underline{x}$. Note that the use of a quadratic form means that any real matrix can be replaced by the real symmetric matrix

$$A_{SYM} = \frac{1}{2}\cdot\left(A + A^T\right)$$

so that a matrix in a quadratic form is implicitly symmetric. From Gelp equation (2.1-46) page 19 and *Linear Algebra, Crash Course* page 111, an eigenvalue is a number λ such that, for some vector $\underline{x}$,

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

The vector $\underline{x}$ is the eigenvector corresponding to λ. From Theorem 7.5, *Linear Algebra, Crash Course* page 112, the eigenvectors of a real, symmetric matrix are real and orthogonal. Thus the set of eigenvectors may be considered as axes of a rotated coordinate system, and any vector is made up of a linear combination of eigenvectors. Thus, any quadratic from satisfies the inequality

$$\underline{x}^T \cdot A \cdot \underline{x} \geq \lambda_{MIN} \cdot \underline{x}^T \cdot \underline{x}$$

Thus, if $\lambda_{MIN} > 0$, the matrix is positive definite.

Gelb Problem 2-4

If *R(t)* is a time-varying orthogonal matrix, and

$$\frac{dR(t)}{dt} \cdot R^T(t) = S(t)$$

show that *S(t)* must be skew-symmetric.

Solution: Construct the product

$$R(t) \cdot R^T = I$$

where the product is the identity matrix because *R(t)* is orthogonal, or its transpose is its inverse, as defined in the paragraph following equation (2.1-35) on page 18. Taking the time derivative of this equation and using the chain rule,

$$\frac{dR(t)}{dt} \cdot R^T(t) + R(t) \cdot \frac{dR^T(t)}{dt} = 0$$

or,

$$\frac{dR(t)}{dt} \cdot R^T(t) = -R(t) \cdot \frac{dR^T(t)}{dt}$$

$$= \left(\frac{dR(t)}{dt} \cdot R^T(t) \right)^T$$

which shows that the left-hand side of the equation in the problem statement is skew-symmetric. Thus, the righ-hand side, *S(t)*, must be skew-symmetric.

Gelb Problem 2-5

Show that the matrix

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

satisfies the polynomial equation

$$A^2 - 5 \cdot A - 2 \cdot I = 0$$

and find the eigenvalues. Use the results to show the form of the solution in terms of exponential functions.

Solution: The characteristic equation is found from

$$\left| \begin{bmatrix} 1-\lambda & 2 \\ 3 & 4-\lambda \end{bmatrix} \right| = \lambda^2 - 5 \cdot \lambda - 2 = 0$$

The Cayley-Hamilton theorem (bottom of page 19) states that a matrix satisfies its own characteristic equation, so we have proven the polynomial matrix equation.

From the polynomial equation, we can use

$$A^2 = 5 \cdot A + 2 \cdot I$$

to reduce any polynomial or series in *A* to a linear combination of *A* and *I*. The eigenvalues are

$$\lambda = \frac{5 \pm \sqrt{33}}{2}$$

The remaining equation, to find the closed forms for $a_1(t)$ and $a_2(t)$, where

$$\exp(A\cdot t) = a_1(t)\cdot I + a_2(t)\cdot A$$

is done as follows. First for a two-by-two matrix, the Cayley-Hamilton theorem gives us

$$A^2 - (\lambda_1 + \lambda_2)\cdot A + \lambda_1\cdot\lambda_2\cdot I = 0$$

or,

$$A^2 = (\lambda_1 + \lambda_2)\cdot A - \lambda_1\cdot\lambda_2\cdot I$$

This equation means that the matrix exponential series as given by Equation (2.1-48) on page 20 can be collapsed to a linear combination of *I* and *A*. Since we are looking at $\exp(A\cdot t)$, we use

$$(A\cdot t)^2 = (\lambda_1 + \lambda_2)\cdot t\cdot(A\cdot t) + \lambda_1\cdot\lambda_2\cdot t^2$$

Note that the eigenvalues of $(A\cdot t)$ are the eigenvalues of *A* multiplied by *t*.

Scalar General Solution for Order Two

The Cayley-Hamilton theorem tells us that the matrix itself satisfies its characteristic equation, but we know also that both the eigenvalues also satisfy the characteristic equation. Since the characteristic equation is used to collapse the series for the exponential into a polynomial of order *N-1* (a first-order polynomial for two by two matrices), we also have

$$\exp(\lambda_i\cdot t) = a_1(t) + a_2(t)\cdot\lambda_i,\ \ i = 1,2$$

We can write this equation down for both eigenvalues and solve for $a_1(t)$ and $a_2(t)$:

$$\exp(\lambda_1\cdot t) = a_1(t) + a_2(t)\cdot\lambda_1$$

$$\exp(\lambda_2\cdot t) = a_1(t) + a_2(t)\cdot\lambda_2$$

is a set of two linear equations in $a_1(t)$ and $a_2(t)$, and can be solved for $a_1(t)$ and $a_2(t)$. The solution is

$$a_1(t) = \frac{\lambda_1\cdot\exp(\lambda_2\cdot t) - \lambda_2\cdot\exp(\lambda_1\cdot t)}{\lambda_1 - \lambda_2}$$

$$a_2(t) = \frac{\exp(\lambda_1\cdot t) - \exp(\lambda_2\cdot t)}{\lambda_1 - \lambda_2}$$

These can be substituted into the two linear equations in $a_1(t)$ and $a_2(t)$ to verify that they are correct. We can differentiate the series for $\exp(A\cdot t)$ term by term and show that

$$\frac{d}{dt}\exp(A\cdot t) = A\cdot\exp(A\cdot t)$$

and use this to verify the solution we have for $a_1(t)$ and $a_2(t)$. Note that $a_1(t)$ and $a_2(t)$ as we have found them are the same as given in the problem statement when it gives is “the form $a_1(t)$ and $a_2(t)$ (a hint that there may be a sign difference or a proportionality factor between the equations given there and the exact solution as we have found here), but the sign of the form for $a_1(t)$ given in the problem statement is incorrect.

General Matrix Solution

We can extend the result to higher order matrices by simply writing the set of two linear equations in equations in $a_1(t)$ and $a_2(t)$ as a set of N linear equations in a set of N time functions:

$$\exp\left(\lambda_i \cdot t\right) = \sum_{j=1}^{N} a_j\left(t\right) \cdot \lambda_i^{j-1},\ \ i = 1, 2 \ldots N$$

This set of equations can be written as a vector equal to a matrix times a vector:

$$\begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 & \ldots & \lambda_1^{N-1} \\ 1 & \lambda_2 & \lambda_2^2 & \ldots & \lambda_2^{N-1} \\ 1 & \lambda_3 & \lambda_3^2 & \ldots & \lambda_2^{N-1} \\ \vdots & & & \ddots & \vdots \\ 1 & \lambda_N & \lambda_N^2 & \ldots & \lambda_N^{N-1} \end{bmatrix} \cdot \begin{bmatrix} a_1\left(t\right) \\ a_2\left(t\right) \\ a_3\left(t\right) \\ \vdots \\ a_N\left(t\right) \end{bmatrix} = \begin{bmatrix} \exp\left(\lambda_1 \cdot t\right) \\ \exp\left(\lambda_2 \cdot t\right) \\ \exp\left(\lambda_3 \cdot t\right) \\ \vdots \\ \exp\left(\lambda_N \cdot t\right) \end{bmatrix}$$

The set of $a_i(t)$ are found by left-multiplying this equation by the inverse of the matrix. The solution to part (b) of the problem can easily be seen to be given by this equation for N=2.

The matrix is a special form, its rows being given by successive powers of different constants. The determinant of such a matrix is called a *Vandermonde determinant* and is very famous, and a lot of papers have been written about it. In particular, a closed form for this determinant is (see *Introduction to Matrix Analysis*, Second Edition, by Richard Bellman, page 193)

$$\left|\lambda_i^{j-1}\right| = \prod_{1 \le i < j \le N} \left(\lambda_j - \lambda_i\right)$$

Note that the matrix is singular if any two eigenvalues are equal.

Chapter 3

Example 3.1-2 Page 56

The Schuler loop is the differential equation for an inertial navigation system (INS). A simple form is to look at one axis. An INS usually needs a north-south and an east-west loop, and these loops are coupled, but we can understand some basic principles by examining this simplification.

We define the quantities of interest in terms of differential equations. The quantities are

$$\underline{x} = \begin{bmatrix} \phi \\ \delta v \\ \delta p \end{bmatrix} = \begin{bmatrix} \text{platform tilt angle in radians} \\ \text{system velocity error} \\ \text{system position error} \end{bmatrix}$$

We define the errors as

$$\underline{v} = \begin{bmatrix} \varepsilon_g \\ \varepsilon_a \end{bmatrix} = \begin{bmatrix} \text{gryo random drift rate} \\ \text{accelerometer random error} \end{bmatrix}$$

From first principles, we have the differential equations

$$\frac{d(\delta p)}{dt} = \delta v$$
$$\frac{d(\delta v)}{dt} = \varepsilon_a - g \cdot \sin(\phi) \approx \varepsilon_a - g \cdot \phi$$
$$\frac{d\phi}{dt} = \frac{\delta v}{R_e} + \varepsilon_g$$

The block diagram shown as Figure 3.1-4 on page 56 follows from these equations. Given these differential equations, we can collect them and express them in vector-matrix form as

$$\frac{d\underline{x}}{dt} = A \cdot \underline{x} + G \cdot \underline{v}$$

where

$$A = \begin{bmatrix} 0 & \frac{1}{R_e} & 0 \\ -g & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$
$$G = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

We can use Laplace transforms to analyze this equation to understand the form of the solutions. The Laplace transform of the differential equation is

$$s \cdot \underline{X}(s) - \underline{x}(0) = A \cdot \underline{X}(s) + G \cdot \underline{V}(s)$$

The solution in the *s* domain is

$$\underline{X}(s) = (s \cdot I - A)^{-1} \cdot (\underline{x}(0) + G \cdot \underline{V}(s))$$

Thus we can understand the nature of the solution by examining the eigenvalues of the matrix *A*. The characteristic equation of *A* is

$$s^2 + \frac{g}{R_e} = 0$$

This shows that the INS steady state is an undamped oscillation. You can easily show that this period is about 84 minutes, which is why an INS is sometimes called an 84-minute pendulum. Use the elementary equation for the period of a pendulum as a function of its length find the length of a “real” 84-minute pendulum.

Gelb Problem 3-1

The linear variance equation is the differential equation for propagation of errors of a noise-driven system. This is developed in Section 3.7, Propagation of Errors, pp. 75-78. The linear variance equation is given as Equation (3.7-17) at the bottom of page 77:

$$\dot{P}(t) = F(t) \cdot P(t) + P(t) \cdot F(t) + G(t) \cdot Q(t) \cdot G^T(t)$$

The problem statement is to verify a solution as given:

$$P(t) = \Phi(t,t_0)\cdot P(t_0)\cdot \Phi^T(t,t_0) + \int_{t_0}^{t} \Phi(t,\tau)\cdot G(\tau)\cdot Q(\tau)\cdot G^T(\tau)\cdot \Phi^T(t,\tau)\cdot d\tau$$

The simplest way to do this is to take the derivative of the candidate expression for *P(t)* and show that the right-hand-side of the differential equation is obtained. In doing this, we need the definition from equations (3.7-13) and (3.7-14) on page 77,

$$\dot{\Phi}(t) = F(t)\cdot \Phi(t)$$

and Leibnitz' rule for differentiation of an integral,

$$\frac{d}{dt}\left(\int_{f_1(t)}^{f_2(t)} g(t,x)\cdot dx\right) = g(t,\ f_2(t))\cdot \frac{df_2(t)}{dt} - g(t,\ f_1(t))\cdot \frac{df_1(t)}{dt} + \int_{f_1(t)}^{f_2(t)} \frac{d}{dt} g(t,x)\cdot dx$$

Because of the size of the equations, we will take one term at a time. We begin with

$$\frac{d}{dt}\left(\Phi(t,t_0)\cdot P(t_0)\cdot \Phi^T(t,t_0)\right) = F(t)\cdot \Phi(t,t_0)\cdot P(t_0)\cdot \Phi^T(t,t_0) + \Phi(t,t_0)\cdot P(t_0)\cdot \Phi^T(t,t_0)\cdot F^T(t)$$

Repeating this operation on the integrand for *P(t)* produces the remainder of the terms in $F(t)\cdot P(t) + P(t)\cdot F(t)$. Leibnitz' rule on the integral, as applied to the limits of the integral, produce the term

$$\Phi(t,t)\cdot G(t)\cdot Q(t)\cdot G^T(t)\cdot \Phi^T(t,t) = G(t)\cdot Q(t)\cdot G^T(t)$$

and we are done.

Gelb, Problem 3-2

When F and Q are constant matrices, we have

$$\Phi(t,t_0) = \exp\left(F\cdot (t-t_0)\right)$$

The given equation is equal to the linear variance equation for constant *P* and for the noise mapping matrix *G* as the identity matrix *I*, but for the time derivative of the covariance equal to zero. By looking at the case $\dot{P} = 0$, we are asking for the steady-state form for the solution, or what happens as $t \to \infty$. We use the assumption of a stable system,

$$\lim_{t\to\infty} \exp(F\cdot t) = 0$$

and the solution to the linear variance equation given in the first problem to show that the steady-state covariance for $t_0 = 0$, constant *F*, and *G* equal to the identity matrix *I* as

$$P(\infty) = 0 + \int_0^{\infty} \exp(F\cdot\tau)\cdot Q\cdot \exp\left(F^T\cdot\tau\right)\cdot d\tau$$

which is the form desired.

Chapter 4

See the separate section *Multivariate Gaussian Distribution* below.

Summary of Equations of the Kalman Filter

State Vector $\underline{x}$, measurement vector $\underline{y}$, measurement model

$$\underline{y} = \underline{h}\left(\underline{x}\right) + \underline{v}$$

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

$$R = \mathrm{Cov}\left(\underline{v}\right) = E\left(\underline{v} \cdot \underline{v}^T\right)$$

State vector extrapolation model

$$\dot{\underline{x}} = \underline{f}\left(\underline{x}\right) + \underline{w}$$

$$F = \frac{\partial \underline{f}\left(\underline{x}\right)}{\partial \underline{x}}$$

$$Q = \mathrm{Cov}\left(\underline{w}\right) = E\left(\underline{w} \cdot \underline{w}^T\right)$$

Covariance extrapolation approximation

$$\tilde{P} \approx \Phi \cdot P_{-} \cdot \Phi^T + Q \cdot T$$

$$\dot{\Phi} = F \cdot \Phi, \;\; \Phi\left(t_{-}\right) = I$$

$$\Phi \approx I + F \cdot T$$

State vector extrapolation approximation

$$\tilde{\underline{x}} = \Phi \cdot \hat{\underline{x}}_{-}$$

Kalman gain

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

or, if we have P available before K is computed,

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

State vector update

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

Covariance extrapolation short form

$$P = \left(I - K \cdot H\right) \cdot \tilde{P} \quad \text{Short form; NUMERICALLY UNSTABLE}$$

Joseph stabilized form for covariance extrapolation (good for most applications)

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

Normal form or information matrix format for covariance extrapolation – use to check for accuracy of Joseph stabilized form:

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

Other: square root filters (not treated in Gelb; overview in the next course)

Gelb Problem 4-1

Repeat example 1.0-1 with a Kalman filter, treating the measurements as sequential and simultaneous.

Sequential measurements:

We are estimating a constant so that *F=I* and *Q*=0. We have one element in the state vector so it is just a scalar. The measurement model is conventional. For the first measurement, we begin with the assumption that we have no information, so *K=1* and $P=\sigma_1^2$. Then, for the second measurement, we have

$$\tilde{x}=y_1$$
$$\tilde{P}=\sigma_1^2$$
$$R=\sigma_2^2$$
$$K=\frac{\tilde{P}}{\tilde{P}+R}=\frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}$$

The estimate is

$$\begin{aligned}\hat{x}&=\tilde{x}+K\cdot\left(y_2-\tilde{x}\right)\\&=y_1+\frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}\cdot\left(y_2-y_1\right)\\&=\frac{\sigma_2^2\cdot y_1+\sigma_1^2\cdot y_2}{\sigma_1^2+\sigma_2^2}\end{aligned}$$

and the covariance is

$$\begin{aligned}P&=\left(1-K\right)\cdot\tilde{P}\\&=\frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}\cdot\sigma_1^2\\&=\frac{\sigma_1^2\cdot\sigma_2^2}{\sigma_1^2+\sigma_2^2}\\&=\frac{1}{\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}}\end{aligned}$$

Simultaneous measurements:
Here we have a scalar state vector, two elements in the measurement vector, and we work with unknown prior data by taking $\tilde{P}$ as very large. We use the information form for the covariance matrix,

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

where we have dropped out the $\tilde{P}^{-1}$ term because we assume no prior information. We have the Kalman gain from *P* as

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

This form allows us to keep the correlations, as we did in the vector approach to Example 1.0-1, and thus to have a ready solution to problem 1-1.

The Alpha-Beta Tracker

The measurements are a sequence of data points y_i and the states are position and velocity, arranged in a state vector $\underline{x}$:

$$\underline{x} = \begin{bmatrix} \text{position} \\ \text{velocity} \end{bmatrix}$$

At the first measurement, before another is available, the measurement is used::

$$\underline{\hat{x}}_1 = \begin{bmatrix} y_1 \\ 0 \end{bmatrix}$$

The position and velocity estimates are defined at the second measurement:

$$\underline{\hat{x}}_2 = \begin{bmatrix} y_2 \\ \dfrac{y_2 - y_1}{t_2 - t_1} \end{bmatrix}$$

Extrapolation from last update (needed to prepare values for last update):

$$T_i = t_i - t_{i-1}$$

$$\Phi_i = \begin{bmatrix} 1 & T_i \\ 0 & 1 \end{bmatrix}$$

$$\underline{\tilde{x}}_i = \Phi_i \cdot \underline{\hat{x}}_{i-1}$$

Update using new data y_i:

$$K_i = \begin{bmatrix} \alpha & \dfrac{\beta}{T_i} \end{bmatrix}$$

$$h(\underline{x}) = x_1 = H \cdot \underline{x}, \;\; H = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

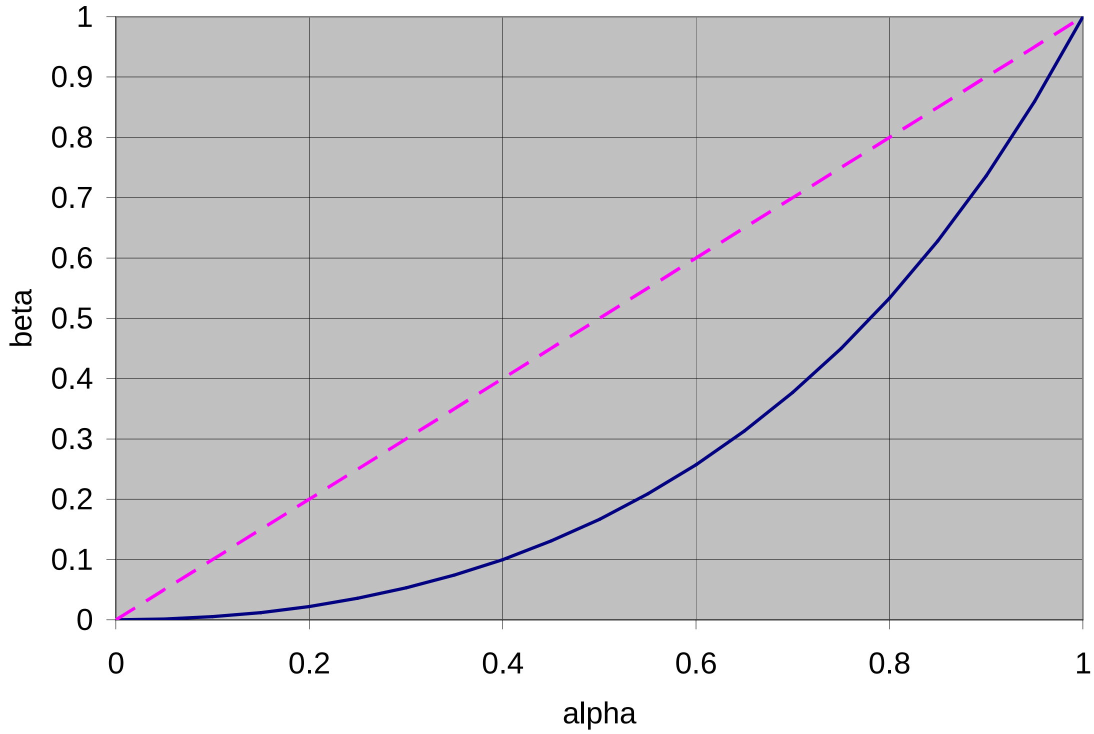
$$\underline{\hat{x}}_i = \underline{\tilde{x}}_i + K_i \cdot \left(y_i - H \cdot \underline{\tilde{x}}_i \right)$$

Relationship between α and β

The most commonly used value of β is

$$\beta = \frac{\alpha^2}{2-\alpha}$$

The choice of α is determined by signal-to-noise ratio, time between updates, target behavior, etc. A plot of β as a function of α is:



The alpha-beta tracker as a vector-matrix recursive filter:

$$\underline{x}_i = \begin{bmatrix} x_i \\ v_i \end{bmatrix}$$

$$\underline{\tilde{x}}_i = \Phi_i \cdot \underline{\hat{x}}_{i-1}, \;\; \Phi_i = \begin{bmatrix} 1 & T_i \\ 0 & 1 \end{bmatrix}$$

$$\underline{\hat{x}}_i = \underline{\tilde{x}}_i + K \cdot \left(y_i - H \cdot \underline{\tilde{x}}_i \right), \;\; K = \begin{bmatrix} \alpha \\ \dfrac{\alpha^2}{2-\alpha} \end{bmatrix}, \;\; H = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

The alpha-beta tracker as a recursive filter is

$$\underline{\hat{x}}_i = \left(I - K_i \cdot H \right) \cdot \Phi_i \cdot \underline{\hat{x}}_{i-1} + K \cdot y_i$$

The Z Transform of the Alpha-Beta Tracker

The recursive digital filter equation is of the general form

$$\underline{s}_i = A \cdot \underline{s}_{i-1} + B \cdot \underline{y}_i$$

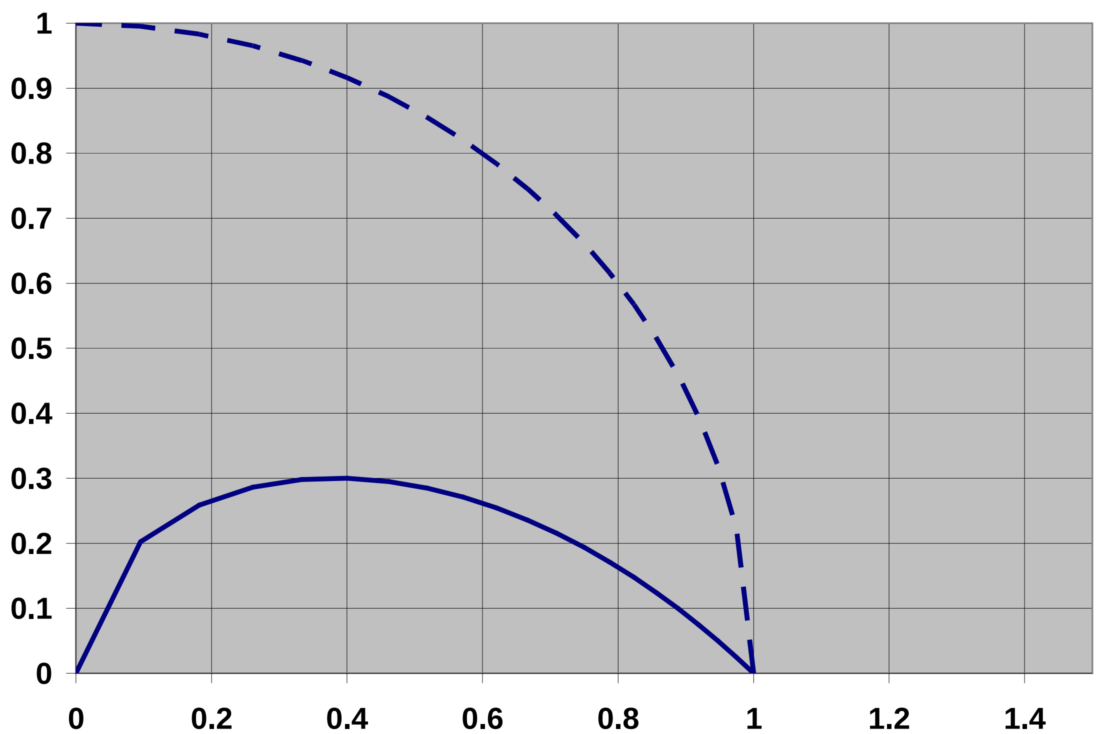
This is simple one-pole filter, posed in vector-matrix notation. For the alpha-beta tracker, when the time between updates T_i is constant, the A matrix is

$$
\begin{aligned}
A &= (I - K \cdot H) \cdot \Phi \\
&= \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} \alpha \\ \dfrac{\beta}{T} \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \end{bmatrix} \right) \cdot \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 1-\alpha & (1-\alpha)\cdot T \\ -\dfrac{\beta}{T} & 1-\beta \end{bmatrix}
\end{aligned}
$$

The characteristic values of the matrix A are

$$
\begin{aligned}
\lambda_A &= 1 - \frac{\alpha+\beta}{2} \pm j \cdot \sqrt{\beta - \left(\frac{\alpha+\beta}{2}\right)^2} \\
&= 1 - \frac{\alpha}{2-\alpha} \cdot \left(1 \pm j \cdot \sqrt{1-\alpha}\right)
\end{aligned}
$$

Note that the characteristic values are not a function of the update time T. For the value of β in terms of α as given above, the characteristic values as a function of α are as in this table: A root locus of the poles of the transfer function of the filter in the z plane is:



The Multivariate Gaussian Distribution

Gaussian Random Variables

A continuous random variable is called Gaussian with mean m and variance σ if its probability density function is

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

Multiple Correlated Gaussian Random Variables

The probability density function of multiple independent Gaussian random variables with mean zero and variance one is

$$\begin{aligned} p(\underline{y}) &= p(y_1) \cdot p(y_2) \ldots p(y_N) \\ &= \sum_{i=1}^{N} \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2} \cdot y_i^2 \right) \\ &= \frac{1}{(2\pi)^{N/2}} \cdot \exp\left(-\frac{1}{2} \cdot \sum_{i=1}^{N} y_{i2}^2 \right) \end{aligned}$$

Note that we can write the sum in the exponential as a quadratic form,

$$\sum_{i=1}^{N} y_{i2}^2 = \underline{y}^T \cdot \underline{y}$$

We define another vector of random variables $\underline{x}$ as a set of linear combinations of the random variables $\underline{y}$,

$$\underline{x} = R \cdot \underline{y}$$

where R is any real matrix. For simplicity here we will also assume that R is square and nonsingular. We find the probability density function of $\underline{x}$ by equating the differentials

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

and we use the Jacobian

$$d\underline{x} = |R| \cdot d\underline{y}$$

We also note that the covariance of $\underline{x}$, P_x, is given by

$$P_x = R \cdot R^T$$

We leave as an exercise in algebra to show that

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