DATA FUSION II, Radar Trackers

Date: February 2, 2006 **Time:** 5:00 – 7:30 PM
Instructor: Dr. James K Beard
Credits: 1 **Course Code:** 0901-504-04 **Registration Number:** 13459

Today's topics:

1 A Few Identities of Gradients of Determinants

These gradients are useful in understanding determinants, inverses, and in performing minimizations and optimizations. All of them can be verified by looking at the definitions of the trace and term-by-term evaluation of the terms in the gradients.

1.1.1 Definition and Notation of the Gradient

The gradient of a scalar with respect to a vector is a vector:

$$\frac{\partial f(\underline{x})}{\partial \underline{x}} = \begin{bmatrix} \frac{\partial f(\underline{x})}{\partial x_1} \\ \frac{\partial f(\underline{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\underline{x})}{\partial x_N} \end{bmatrix} \tag{1.1}$$

The gradient of a vector with respect to a vector is a matrix:

$$\frac{\partial \underline{f}(\underline{x})}{\partial \underline{x}} = \left[\frac{\partial \underline{f}(\underline{x})}{\partial x_1}, \quad \frac{\partial \underline{f}(\underline{x})}{\partial x_2} \quad \cdots \quad \frac{\partial \underline{f}(\underline{x})}{\partial x_N} \right] \tag{1.2}$$

The gradient of a scalar with respect to a matrix is a matrix:

$$\frac{\partial f(A)}{\partial A} = \left[\frac{\partial f(A)}{\partial a_{i,j}} \right] \tag{1.3}$$

Gradients of the Trace of a Matrix:

$$\frac{\partial \operatorname{tr}\{A\}}{\partial A} = I \tag{1.4}$$

A generalization that is useful when the trace is of the products of matrices is

$$\frac{\partial \operatorname{tr}\{B \cdot A \cdot C\}}{\partial A} = B^T \cdot C^T \tag{1.5}$$

Gradient of the determinant of a matrix:

$$\frac{\partial |A|}{\partial A} = |A| \cdot A^{-T} \tag{1.6}$$

A generalization that is useful when the determinant is of the products of matrices is

$$\frac{\partial |B \cdot A \cdot C|}{\partial A} = |B \cdot A \cdot C| \cdot A^{-T} \tag{1.7}$$

2 Probability and Statistics Overview

2.1 Names of Probability Distribution Functions

2.1.1 Probability Density Functions for Continuous Variables

- Gaussian, or normal
- Rayleigh, or square root of sum of two Gaussian variables
- Chi-square, or sum of squares of Poisson
- Poisson, or exponential

- Cauchy, or arc-tangent
- Weibull, or two-parameter exponential
- Student T, or normalized sample Gaussian
- Edgeworth/Hermite/Cornish-Fisher series, or expansion in Hermite polynomials
- Rician or non-central chi-square, or envelope of signal in noise
- F or variance ratio

2.1.2 Probability Density Functions for Discrete Variables

- Binomial
- Poisson, or exponential

2.2 Definition of Discrete Probability

The probability that an event number *i* of *N* will occur is $P(E_i)$

If the events are mutually exclusive, such as whether a coin toss will be heads or tails, the sum of the probabilities of all possibilities is 1. Thus each individual probability must be between zero and 1, inclusive.

The concept of independent events is fundamentally different from the concept of exclusive events. Independent events are two coin tosses made at different times or places.

2.3 Independent Probabilities

The probability of simultaneous independent events is

$$P(A \& B \& C \ldots) = P(A) \cdot P(B) \cdot P(C) \ldots \tag{2.1}$$

The probability of either of two events occurring is

$$P(A+B) = P(A) + P(B) - P(A \& B) \tag{2.2}$$

When the events are mutually exclusive, $P(A \& B) = 0$ and the probability of either of the events is simply the sum of the probabilities of the individual events. When the probabilities are not mutually exclusive, the concept of conditional probability is important. We define this probability as

$$P(A|B) = \frac{P(A \& B)}{P(B)} \tag{2.3}$$

2.4 Continuous Random Variables and the Probability Density Function

The probability that a continuous random variable such as a noisy voltage *s* is less than a given value *x* is defined as

$$P(s \le x) \tag{2.4}$$

and is between zero and one, inclusive, as is any probability. The *probability density function p(x)* is defined as

$$p(x) = \frac{dP\left(s \le x\right)}{dx} \tag{2.5}$$

The differential relationship

$$P\left(x < x \le x + dx\right) = p\left(x\right) \cdot dx \tag{2.6}$$

reveals that the probability density function *p(x)* is, in general, nonnegative but not bounded above.

2.5 Gaussian Random Variables

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

$$p\left(x\right) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp\left(-\frac{1}{2} \cdot \frac{\left(x-m\right)^2}{\sigma^2}\right) \tag{2.7}$$

2.6 Multiple Correlated Gaussian Random Variables

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

$$\begin{aligned} p\left(\underline{y}\right) &= p\left(y_1\right) \cdot p\left(y_2\right) \ldots p\left(y_N\right) \\ &= \sum_{i=1}^{N} \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2} \cdot y_i^2\right) \\ &= \frac{1}{\left(2\pi\right)^{N/2}} \cdot \exp\left(-\frac{1}{2} \cdot \sum_{i=1}^{N} y_{i2}^2\right) \end{aligned} \tag{2.8}$$

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

$$\sum_{i=1}^{N} y_{i2}^2 = \underline{y}^T \cdot \underline{y} \tag{2.9}$$

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

$$\underline{x} = R \cdot \underline{y} \tag{2.10}$$

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

$$p\left(\underline{x}\right) \cdot d\,\underline{x} = p\left(\underline{y}\right) \cdot d\,\underline{y} \tag{2.11}$$

and we use the Jacobian

$$d\,\underline{x} = \left|R\right| \cdot d\,\underline{y} \tag{2.12}$$

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

$$P_x = R \cdot R^T \tag{2.13}$$

We leave as an exercise in algebra to show that

$$p\left(\underline{x}\right)=\frac{1}{\left(2\pi\right)^{N/2}\cdot\left|P_x\right|^{1/2}}\cdot\exp\left(-\frac{1}{2}\cdot\underline{x}^T\cdot P_x^{-1}\cdot\underline{x}\right) \tag{2.14}$$

3 Gelb, Example 1-1 pages 5 and 6

3.1 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 $\underline{y}$,

$$\underline{y}=\begin{bmatrix} z_1 \\ z_2 \end{bmatrix}=\begin{bmatrix} x+v_1 \\ x+v_2 \end{bmatrix}=x\cdot\underline{1}+\underline{v} \tag{3.1}$$

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

$$\underline{k}=\begin{bmatrix} k_1 \\ k_2 \end{bmatrix} \tag{3.2}$$

and an estimate is a linear combination of the measurements:

$$\hat{x}=\underline{k}^T\cdot\underline{y}\,. \tag{3.3}$$

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

$$x_e=\hat{x}-x\,. \tag{3.4}$$

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

$$E\left(\hat{x}\right)=x \tag{3.5}$$

where *E(*)* is the expectation, or ensemble mean, operator. We substitute the equations for the estimate and measurement to find a constraint on the linear estimator gain $\underline{k}$ as follows:

$$E\left(\hat{x}\right)=E\left(\underline{k}^T\cdot\underline{y}\right)=E\left(\underline{k}^T\cdot\left(x\cdot\underline{1}+\underline{v}\right)\right)=x\,. \tag{3.6}$$

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

$$E\left(\underline{v}\right)=\underline{0} \tag{3.7}$$

and we have

$$\underline{k}^T\cdot\underline{1}=1 \tag{3.8}$$

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

$$J = E\left(x_e^2\right) = E\left(\left(\hat{x} - x\right)^2\right). \tag{3.9}$$

Again substituting from the above, we have

$$\begin{aligned} J &= E\left(\left(\underline{k}^T \cdot \left(x \cdot \underline{1} + \underline{v}\right) - x\right)^2\right) \\ &= E\left(\left(x \cdot \left(\underline{k}^T \cdot 1\right) + \underline{k}^T \cdot \underline{v} - x\right)^2\right) \\ &= E\left(x^2 \cdot \underline{k}^T \cdot \left(\underline{1}^T \cdot \underline{1}\right) \cdot \underline{k}\right) + 2x \cdot E\left(\underline{k}^T \cdot \underline{1} \cdot \underline{v}^T \cdot \underline{k}\right) + \underline{k}^T \cdot E\left(\underline{v} \cdot \underline{v}^T\right) \cdot \underline{k} \\ &\quad -2x \cdot E\left(x \cdot \left(\underline{k}^T \cdot 1\right) + \underline{k}^T \cdot \underline{v}\right) + x^2 \end{aligned} \tag{3.10}$$

This is three quadratic forms in $\underline{k}$, plus a couple of other terms linear in $\underline{k}$, and a term independent of $\underline{k}$. Let's look at the three quadratic terms separately. The first term has no random variables, so the expected value is simply the quantity in the expectation operator:

$$E\left(x^2 \cdot \underline{k}^T \cdot \left(\underline{1}^T \cdot \underline{1}\right) \cdot \underline{k}\right) = x^2 \cdot \underline{k}^T \cdot \left(\underline{1}^T \cdot \underline{1}\right) \cdot \underline{k} \tag{3.11}$$

There is a very interesting matrix in this equation:

$$\underline{1} \cdot \underline{1}^T = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \tag{3.12}$$

Expressions of the form $\underline{x}^T \cdot A \cdot \underline{x}$ are called *quadratic forms* because they are the sum of terms that are quadratic in the elements of $\underline{x}$ – they include the products of two elements of $\underline{x}$ and one element of the matrix A. Let's look at the quadratic form for this very interesting matrix and the linear estimator gain $\underline{k}$:

$$\underline{k}^T \cdot \left(\underline{1}^T \cdot \underline{1}\right) \cdot \underline{k} = \begin{bmatrix} k_1 & k_2 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \left(\underline{k}^T \cdot \underline{1}\right)^2 = \left(k_1 + k_2\right)^2 \tag{3.13}$$

which we can see is consistent with our scalar problem statement and our other equations.

The second quadratic form is linear in the elements of $\underline{v}$ and this means that the expectation or mean of each term is zero, so we can drop it out now.

The third term contains a very special matrix, the *covariance matrix* of the measurement noise vector $\underline{v}$:

$$E\left(\underline{v} \cdot \underline{v}^T\right) = E\left(\begin{bmatrix} v_1^2 & v_1 \cdot v_2 \\ v_1 \cdot v_2 & v_2^2 \end{bmatrix}\right) = \begin{bmatrix} \sigma_1^2 & \rho \cdot \sigma_1 \cdot \sigma_2 \\ \rho \cdot \sigma_1 \cdot \sigma_2 & \sigma_2^2 \end{bmatrix} = R_v. \tag{3.14}$$

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 can always make it zero in our result later.

The linear terms are

$$-2x \cdot E\left(x \cdot \left(\underline{k}^T \cdot 1\right) + \underline{k}^T \cdot \underline{v}\right) + x^2 = -2x^2 \cdot \underline{k}^T \cdot \underline{1} + x^2. \tag{3.15}$$

Since we have

$$\underline{k}^T \cdot \underline{1} = 1 \tag{3.16}$$

we note that the terms involving x^2 sum to zero, so we can drop them and have a very simple equation for the mean square error:

$$J = \underline{k}^T \cdot R_v \cdot \underline{k}\,. \tag{3.17}$$

We need to minimize this equation 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 an unknown and link it to the original equation. This is call the method of Lagrangian multipliers. For our problem, the optimization criteria J becomes

$$J = \underline{k}^T \cdot R_v \cdot \underline{k} - \lambda \cdot \left(\underline{k}^T \cdot \underline{1} - 1\right). \tag{3.18}$$

The new variable 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:

$$\begin{aligned} \frac{\partial J}{\partial \underline{k}} &= 2 \cdot R_v \cdot \underline{k} - \lambda \cdot \underline{1} = \underline{0} \\ \underline{k} &= \frac{\lambda}{2} \cdot R_v^{-1} \cdot \underline{1}\,. \end{aligned} \tag{3.19}$$

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

$$\begin{aligned} \underline{k}^T \cdot \underline{1} &= \frac{\lambda}{2} \underline{1}^T \cdot R_v^{-1} \cdot \underline{1} = 1 \\ \frac{\lambda}{2} &= \frac{1}{\underline{1}^T \cdot R_v^{-1} \cdot \underline{1}} \end{aligned} \tag{3.20}$$

Note that

$$\underline{1}^T \cdot R_v^{-1} \cdot \underline{1} = \left[\text{sum of all the terms of } R_v^{-1}\right]. \tag{3.21}$$

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

$$\underline{k} = \frac{R_v^{-1} \cdot \underline{1}}{\underline{1}^T \cdot R_v^{-1} \cdot \underline{1}}. \tag{3.22}$$

This gives us a minimum value of the optimization criteria of

$$J_{MIN} = \frac{1}{\underline{1}^T \cdot R_v^{-1} \cdot \underline{1}}\,. \tag{3.23}$$

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

$$R_v^{-1} = \begin{bmatrix} \sigma_1^2 & \rho\cdot\sigma_1\cdot\sigma_2 \\ \rho\cdot\sigma_1\cdot\sigma_2 & \sigma_2^2 \end{bmatrix}^{-1} = \frac{1}{\sigma_1^2\cdot\sigma_2^2\cdot\left(1-\rho^2\right)}\cdot\begin{bmatrix} \sigma_2^2 & -\rho\cdot\sigma_1\cdot\sigma_2 \\ -\rho\cdot\sigma_1\cdot\sigma_2 & \sigma_1^2 \end{bmatrix} \tag{3.24}$$

The minimum value of the optimization criteria is

$$J_{MIN} = \frac{\sigma_1^2\cdot\sigma_2^2\cdot\left(1-\rho^2\right)}{\sigma_1^2+\sigma_2^2-2\rho\cdot\sigma_1\cdot\sigma_2} \tag{3.25}$$

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

$$\underline{k} = \frac{1}{\sigma_1^2+\sigma_2^2-2\rho\cdot\sigma_1\cdot\sigma_2}\cdot\begin{bmatrix} \sigma_2^2-\rho\cdot\sigma_1\cdot\sigma_2 \\ \sigma_1^2-\rho\cdot\sigma_1\cdot\sigma_2 \end{bmatrix}. \tag{3.26}$$

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

$$\underline{1}^T\cdot R_v^{-1}\cdot\underline{1} = \underline{1}^T\cdot\begin{bmatrix} \frac{1}{\sigma_1^2} & 0 & \cdots & \\ 0 & \frac{1}{\sigma_2^2} & & \\ \vdots & & \ddots & \\ & & & \frac{1}{\sigma_M^2} \end{bmatrix}\cdot\underline{1} = \sum_{j=1}^{M}\frac{1}{\sigma_j^2} \tag{3.27}$$

and

$$R_v^{-1}\cdot\underline{1} = \begin{bmatrix} \frac{1}{\sigma_1^2} \\ \frac{1}{\sigma_2^2} \\ \vdots \\ \frac{1}{\sigma_M^2} \end{bmatrix}. \tag{3.28}$$

4 Examples of Simple Tracker Methods

4.1.1 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 $\underline{x}$:

$$\underline{x} = \begin{bmatrix} \text{position} \\ \text{velocity} \end{bmatrix} \tag{4.1}$$

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

$$\hat{\underline{x}}_1 = \begin{bmatrix} y_1 \\ 0 \end{bmatrix} \tag{4.2}$$

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

$$\hat{\underline{x}}_2 = \begin{bmatrix} y_2 \\ \dfrac{y_2 - y_1}{t_2 - t_1} \end{bmatrix} \tag{4.3}$$

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

$$\begin{gathered} T_i = t_i - t_{i-1} \\ \Phi_i = \begin{bmatrix} 1 & T_i \\ 0 & 1 \end{bmatrix} \\ \tilde{\underline{x}}_i = \Phi_i \cdot \hat{\underline{x}}_{i-1} \end{gathered} \tag{4.4}$$

Update using new data y_i:

$$\begin{aligned} K_i &= \begin{bmatrix} \alpha & \dfrac{\beta}{T_i} \end{bmatrix} \\ h(\underline{x}) &= x_1 = \begin{bmatrix} 1 & 0 \end{bmatrix} \cdot \underline{x} \\ \hat{\underline{x}}_i &= \tilde{\underline{x}}_i + K_i \cdot \left(y_i - h(\tilde{\underline{x}}_i) \right) \end{aligned} \tag{4.5}$$

4.1.2 Relationship between α and β

The most commonly used value of β is

$$\beta = \frac{\alpha^2}{2 - \alpha} \tag{4.6}$$

The choice of α is determined by signal-to-noise ratio, time between updates, target behavior, etc.

Study problems:
Graph β as a function of α. Keep the graph for discussions in the next class.

Show that the estimate for time t_2 is obtained from the general equations by taking $\alpha = 1$ for this step.

5 High Level Language (HLL) Examples

We will use the laptop, and run the examples given last time.

6 The Z Transform of the Alpha-Beta Tracker

The extrapolated state vector for the alpha-beta tracker is

$$\Phi_i = \begin{bmatrix} 1 & T_i \\ 0 & 1 \end{bmatrix} \tag{6.1}$$
$$\underline{\tilde{x}}_i = \Phi_i \cdot \underline{\hat{x}}_{i-1}$$

The updated state vector of the alpha-beta tracker is

$$K_i = \begin{bmatrix} \alpha \\ \dfrac{\beta}{T_i} \end{bmatrix}, H = \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{6.2}$$
$$\underline{\hat{x}}_i = \underline{\tilde{x}}_i + K_i \cdot \left(y_i - H \cdot \underline{\tilde{x}}_i \right)$$

We can examine this equation under the assumption that the update time *T* is constant to get an idea of the behavior of the alpha-beta tracker, viewed as a one-pole digital filter.. Substituting the extrapolation equation into the update equation and collecting terms, we have

$$\underline{\hat{x}}_i = \left(I - K_i \cdot H \right) \cdot \Phi_i \cdot \underline{\hat{x}}_{i-1} + K_i \cdot y_i \tag{6.3}$$

This equation is of the general form

$$\underline{s}_i = A \cdot \underline{s}_{i-1} + B \cdot \underline{y}_i \tag{6.4}$$

or, a simple one-pole filter, posed in vector-matrix notation. The *A* matrix is

$$\begin{aligned} A &= \left(I - K \cdot H \right) \cdot \Phi \\ &= \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} \alpha \\ \dfrac{\beta}{T} \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \end{bmatrix} \right) \cdot \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1-\alpha & \left(1-\alpha\right) \cdot T \\ -\dfrac{\beta}{T} & 1-\beta \end{bmatrix} \end{aligned} \tag{6.5}$$

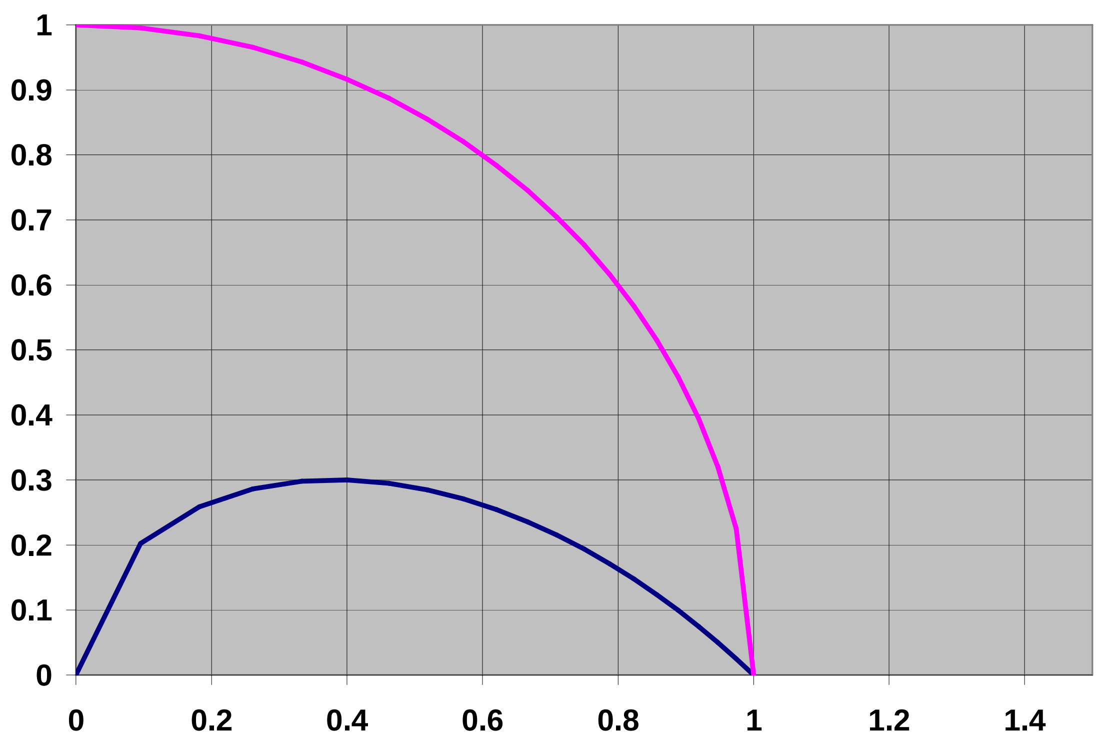
The characteristic values of the matrix *A* are

$$\lambda_A = 1 - \frac{\alpha+\beta}{2} \pm j \cdot \sqrt{\beta - \left(\frac{\alpha+\beta}{2} \right)^2} \tag{6.6}$$

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:

Alpha	Real{z}	Imag{z}
0	1	0
0.05	0.974359	0.024992
0.1	0.947368	0.049931
0.15	0.918919	0.074753
0.2	0.888889	0.099381
0.25	0.857143	0.123718
0.3	0.823529	0.147646
0.35	0.787879	0.171018
0.4	0.75	0.193649
0.45	0.709677	0.215309
0.5	0.666667	0.235702
0.55	0.62069	0.254449
0.6	0.571429	0.271052
0.65	0.518519	0.284848
0.7	0.461538	0.294928
0.75	0.4	0.3
0.8	0.333333	0.298142
0.85	0.26087	0.286264
0.9	0.181818	0.258732
0.95	0.095238	0.202311
1	0	0

A plot of the locus of the characteristic value in the first quadrant, as α varies from zero to one, is shown in the plot below.



Note that the locus begins at (1,0) for $\alpha = 0$ and stays well within the unit circle.

7 Reading

Read the rest of Chapter 2 and use what you learn to do the homework.

8 Homework

Gelb, problem 2-2 page 46.

9 Study Problems

Look at Gelb, problems on pages 46 and 47:
Problems 2-1, 2-3, 2-4, and 2-5.

Examine Problem 2-6, page 47.

Glance at Chapter 3, pages 51-56.

If you have studied the Z-transform in the past, review the basic concepts of the definition of the Z transform and the concept that stability of a transfer function is equivalent to the poles being in the unit circle. Next time we will look at the alpha-beta tracker for the case of uniform update times in terms of Z transforms.

10 One Final Note (Repeated from Last Week)

Approximate Amazon sales ranks of popular tracking references as of January 26, 2006:

Book	Approximate Sales Rank on Amazon.com
Gelb (1974)	78,000
Bar-Shalom (2002)	85,000
Blackman and Popoli (2000)	225,000
Blackman (1986)	690,000