Gelb Example 1-1 Pages 5 and 6

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

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

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

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

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

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

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:

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

There is a very interesting matrix in this equation:

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

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 $\underline{k}$:

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

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 $\underline{v}$:

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

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

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

Since we have

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

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:

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

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

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

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:

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

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

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 $\underline{v}$:

$$R_v^{-1} = \begin{bmatrix} \sigma_1^2 & \rho\cdot\sigma_1\cdot\sigma_2 \\ \rho\cdot\sigma_1\cdot\sigma_2 & \sigma_2^2 \end{bmatrix}^{-1}$$

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

The minimum value of the optimization criteria is

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

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

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

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

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