KALMAN FILTERING WITH STATISTICAL STATE CONSTRAINTS

T. L. Chia
ControlSoft Inc.
5387 Avion Park Drive
Highland Heights, OH 44143
tchia@controlsoftinc.com

D. Simon
Cleveland State University
Electrical Engineering
Cleveland, OH 44115
d.j.simon@csuohio.edu

H. J. Chizeck
University of Washington
Electrical Engineering
Box 352500
Seattle, WA 98195
chizeck@ee.washington.edu

Abstract

For linear dynamic systems with white process and measurement noise, the Kalman filter is known to be the minimum variance linear state estimator. In the case that the random quantities are Gaussian, then the Kalman filter is the minimim variance state estimator. However, in the application of Kalman filters known signal information is often either ignored or dealt with heuristically. For instance, state variable constraints (which may be based on physical considerations) are often neglected because they do not fit easily into the structure of the optimal filter. Previous work by the authors demonstrated an analytic method of incorporating deterministic state equality constraints in the Kalman filter. This paper extends that work to develop the properties of Kalman filters in the presence of statistical state constraints. That is, given a linear system such that the expected values of the state variables satisfy some linear equality, we can constrain the Kalman filter estimates to satisfy those constraints. This results in a family of constrained filters with each member parameterized by a weighting matrix. This paper derives several interesting properties of the constrained Kalman filters.

Key Words – Kalman Filter, State Constraints, State Estimation.

1. Introduction

For linear dynamic systems with white process and measurement noise, the Kalman filter is known to be the minimum variance linear state estimator. In the case that the random quantities are Gaussian, then the Kalman filter is the minimim variance state estimator [1]. However, in the application of Kalman filters there is often known model or signal information that is either ignored or dealt with heuristically [2]. One method for incorporating state equality constraints includes reduction of the system model parameterization [3]. This approach can be used for deterministic state constraints but cannot be used for statistical constraints. Another disadvantage associated with model reduction is the loss of the natural form and structure of the state equations. Other researchers have treated state constraints as perfect measurements [4, 5]. However, this method has disadvantages. First, it increases the dimension of the problem, which results in increased computational effort. Second, it results in a singular covariance matrix in the Kalman filter, which causes numerical problems [6, p. 249],[7, p. 365].

There is often a need to impose state constraints in practical control and state estimation applications. For example, parameter equality and inequality constraints were applied in [8, 9] for electronic muscle stimulation for neuroprosthetic systems to restore motion to individuals with spinal cord injuries and other paralyzing disorders.

A method based on [10] for incorporating deterministic state constraints has recently been proposed by the authors for the Kalman filter [11]. This involves the projection of the unconstrained state estimate onto the constraint surface. In this paper we consider constraints on the expected values of the state variables. That is, the state constraints are statistical rather than deterministic. In this case we can project the unconstrained estimate onto the constraint surface as in [11]. The constrained filters in [11] are therefore identical to the constrained filters used in the present paper. However, the *properties* of the constrained estimate are different than those in [11] because the state constraints are statistical rather than deterministic.

In this paper we derive the properties of the constrained Kalman filter when the underlying system is subject to statistical state constraints. We show that each member of the family of constrained filters is unbiased. We also show that one particular constrained filter has a covariance that is smaller than the covariance of the unconstrained filter. However, the error covariance of the constrained filter is larger than the error covariance of the unconstrained filter. We also derive in this paper the particular constrained filter that has the smallest possible covariance.

Section 2 presents a review of the standard discrete time Kalman filter. Some important properties of the Kalman filter that will be used later in this paper are also reviewed. Section 3 discusses the extension of the Kalman filter to statistical state equality constraints. Section 4 derives the properties of the Kalman filter subject to statistical state constraints. Section 5 shows how the variance of the unconstrained filter can be computed recursively, which is an important practical aspect of the constrained filters. Section 6 presents a simple scalar example that illustrates the theory, and Section 7 provides some concluding remarks.

Since the filters subject to statistical constraints are identical to the filters subject to deterministic constraints, this paper does not present any simulation studies. Readers who are interested in simulation results are directed to earlier work on deterministic constraints [11].

2. Kalman Filtering

This section reviews standard (unconstrained) state estimation via the Kalman filter and some important properties of the filter that will be used later in this paper. The results and notation are taken from [1]. Consider the discrete linear time-invariant system given by:

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + w_k \\ y_k &= Cx_k + e_k \end{aligned} \qquad (1)$$

where k is the time index, x is the state vector, u is the known control input, y is the measurement, and $\{w_k\}$ and $\{e_k\}$ are noise input sequences. The problem is to find an estimate $\hat{x}_{k+1}$ of x_{k+1} given the measurements $\{y_0, y_1, \cdots, y_k\}$. We will use the symbol Y_k to denote the column vector that contains the measurements $\{y_0, y_1, \cdots, y_k\}$. We assume that the following standard conditions are satisifed:

$$\begin{aligned}
E[x_0] &= \bar{x}_0 \\
E[w_k] = E[e_k] &= 0 \\
E[(x_0 - \bar{x}_0)(x_0 - \bar{x}_0)^T] &= \Sigma_0 \\
E[w_k w_m^T] &= Q\delta_{km} \\
E[e_k e_m^T] &= R\delta_{km} \\
E[w_k e_m^T] = E[x_k e_m^T] &= 0 \\
E[x_k w_m^T] &= 0 \quad (k \leq m)
\end{aligned} \tag{2}$$

where $E[\cdot]$ is the expectation operator, $\bar{x}$ is the expected value of x, and δ_{km} is the Kronecker delta function ($\delta_{km} = 1$ if $k = m$, 0 otherwise). Q and R are positive semidefinite covariance matrices. In general, Q and R could be time-varying, but for notational convenience we will assume in this paper that they are constant. The Kalman filter equations are given by:

$$\begin{aligned}
K_k &= A\Sigma_k C^T (C\Sigma_k C^T + R)^{-1} \\
\hat{x}_{k+1} &= A\hat{x}_k + Bu_k + K_k(y_k - C\hat{x}_k) \\
\Sigma_{k+1} &= (A\Sigma_k - K_k C\Sigma_k)A^T + Q
\end{aligned} \tag{3}$$

where the filter is initialized with $\hat{x}_0 = \bar{x}_0$, and Σ_0 given above. It can be shown [1] that the Kalman filter has several attractive properties. For instance, if x_0, $\{w_k\}$, and $\{e_k\}$ are jointly Gaussian, the Kalman filter estimate $\hat{x}_{k+1}$ is the conditional mean of x_{k+1} given the measurements Y_k; i.e., $\hat{x}_{k+1} = E[x_{k+1}|Y_k]$. Even if x_0, $\{w_k\}$, and $\{e_k\}$ are not jointly Gaussian, the Kalman filter estimate is the best affine estimator given the measurements Y_k; i.e., of all estimates of x_{k+1} that are of the form $FY_k + g$ (where F is a constant matrix and g is a constant vector), the Kalman filter estimate is the one that minimizes the variance of the estimation error. It can be shown [1, pp. 92 ff.] that the Kalman filter estimate (i.e., the minimum variance estimate) can be given by:

$$\hat{x}_{k+1} = \bar{\bar{x}}_{k+1} \equiv \bar{x}_{k+1} + \Sigma_{xy}\Sigma_{yy}^{-1}(Y_k - \bar{Y}_k) \tag{4}$$

where $\bar{x}_{k+1}$ is the mean of x_{k+1}, Σ_{xy} is the covariance matrix of x_{k+1} and Y_k, Σ_{yy} is the covariance matrix of Y_k, and $\bar{\bar{x}}_{k+1}$ is the conditional mean of x_{k+1} given the measurements Y_k. In addition, from [1, p. 93] we know that if x_0, $\{w_k\}$, and $\{e_k\}$ are jointly Gaussian, then the Kalman filter estimate $\hat{x}_{k+1}$ and Y_k are jointly Gaussian, in which case $\hat{x}_{k+1}$ is conditionally Gaussian given Y_k. The conditional probability density function of x_{k+1} given Y_k is then given by:

$$P(x_{k+1}|Y_k) = \frac{\exp[-(x_{k+1} - \bar{\bar{x}}_{k+1})^T \Sigma_{k+1}^{-1}(x_{k+1} - \bar{\bar{x}}_{k+1})/2]}{(2\pi)^{n/2}|\Sigma_{k+1}|^{1/2}} \tag{5}$$

where n is the dimension of x, Σ_{k+1} is the covariance of the Kalman filter estimation error at time $k+1$, and the Kalman filter estimate is that value of x_{k+1} that maximizes the conditional probability density function $P(x_{k+1}|Y_k)$. Even if x_0, $\{w_k\}$, and $\{e_k\}$ are not jointly Gaussian, the covariance of the Kalman filter estimation error is given by:

$$\Sigma_{k+1} = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} \tag{6}$$

3. Constrained Kalman Filtering

This section reviews the extension of the well known results of the previous sections to cases where there are known linear equality constraints among the expected values of the state components. Consider the dynamic system of (1) where we are given the additional statistical state equality constraint:

$$D\bar{x}_k = d_k \tag{7}$$

This is a constraint on $\bar{x}_k = E(x_k)$, the *a priori* expected value of x_k. D is a known full rank $s \times n$ matrix, s is the number of constraints, n is the number of state variables, $s \leq n$, and d_k is a known vector. If D is not full rank that means we have redundant state constraints. In that case we can simply remove linearly dependent rows from D (i.e., remove redundant state constraints) until D is full rank. In general, D can be time-varying, but to simplify the notation we assume in this paper that D is constant.

We can incorporate the above constraint in the Kalman filter by requiring that the constrained state estimate, denoted as $\tilde{x}$, satisfy:

$$D\tilde{x}_k = d_k \tag{8}$$

That is, we require satisfaction of the constraint pointwise on the sample path of $\{\tilde{x}_k\}$. This is a sufficient but not necessary condition that the statistical state constraint specified in (7) be satisfied by the state estimate. This is exactly how we incorporate deterministic state constraints ($Dx_k = d_k$) in the Kalman filter [11]. Therefore, the optimal filter subject to statistical state constraints is exactly the same as the optimal filter subject to deterministic state constraints.

The constrained Kalman filter, as shown in [11], can be given by:

$$\tilde{x}_k(W) = \hat{x}_k - W^{-1}D^T(DW^{-1}D^T)^{-1}(D\hat{x}_k - d_k) \tag{9}$$

where $\hat{x}_k$ is the unconstrained Kalman filter estimate and W is a symmetric positive definite weighting matrix. A constrained state estimate $\tilde{x}_k(W)$ can be derived for any symmetric positive definite weighting matrix W. Several special cases of the weighting matrix W can be derived as shown in [11]. If the constrained Kalman filter is derived using a maximum probability approach, then $W = \Sigma^{-1}$. If the constrained filter is derived using a mean square minimization method, then $W = I$. If the constrained filter is derived using a projection method, then W can be any symmetric positive definite matrix, chosen to weight the difference between $\hat{x}$ and $\tilde{x}(W)$. Hence, (9) is a general form of a class of constrained Kalman filters. With the selection of different weighting matrices W, different statistical properties can be derived as shown in the next section.

Note that (3) and (9) can be combined to obtain:

$$\tilde{x}_{k+1}(W) \quad = \quad A\hat{x}_k + Bu_k + K_k(y_k - C\hat{x}_k) - W^{-1}D^T(DW^{-1}D^T)^{-1}(D\hat{x}_{k+1} - d_{k+1}) \tag{10}$$

However, this combination of the update equation and constraint equation cannot be used to directly update $\tilde{x}$, because the right hand side contains terms at time $k + 1$, as well as k.

4. Properties of the Statistically Constrained Filter

This section derives various properties of the constrained state estimate (9) when the system satisfies the constraint (7). In this section we drop the time subscript k for notational convenience.

Theorem 1 *Assuming that the weighting matrix W is independent of x and $\hat{x}$, each member of the class of constrained state estimates is unbiased. That is, $E[x - \tilde{x}(W)] = 0$.*

Proof: From (9) we obtain:

$$x - \tilde{x}(W) = x - \hat{x} + W^{-1}D^T(DW^{-1}D^T)^{-1}(D\hat{x} - d) \tag{11}$$

Taking the expected value of both sides gives:

$$\begin{aligned} E[x - \tilde{x}(W)] &= E[x - \hat{x}] + W^{-1}D^T(DW^{-1}D^T)^{-1}(DE[\hat{x}] - d) \\ &= 0 \end{aligned} \tag{12}$$

where the last equality comes from combining (7) with the fact that $\hat{x}$ is known to be unbiased. **QED**

This is the same as Theorem 1 in [11] but the proof is different.

At this point we will define $\hat{V}$ as the covariance of the unconstrained filter estimate. That is:

$$\begin{aligned} \hat{V} &\equiv E[(\hat{x} - \bar{\hat{x}})(\hat{x} - \bar{\hat{x}})^T] \\ &= E[(\hat{x} - \bar{x})(\hat{x} - \bar{x})^T] \end{aligned} \tag{13}$$

Note that $\hat{V}$ is not the covariance of the *error* of the estimate; rather, $\hat{V}$ is the covariance of the estimate. Since $\hat{V}$ is a covariance, we know that it is nonsingular. Similarly, we define $\tilde{V}$ as the covariance of the constrained filter estimate:

$$\begin{aligned} \tilde{V}(W) &\equiv E[(\tilde{x}(W) - \bar{\tilde{x}}(W))(\tilde{x}(W) - \bar{\tilde{x}}(W))^T] \\ &= E[(\tilde{x}(W) - \bar{x})(\tilde{x}(W) - \bar{x})^T] \end{aligned} \tag{14}$$

Note that $\tilde{V}(W)$ is a function of the weighting matrix W that is used in (9). The *error* covariances of the unconstrained and constrained state estimate is given as:

$$\begin{aligned} \Sigma &= E[(\hat{x}-x)(\hat{x}-x)^T] \\ \tilde{\Sigma}(W) &= E[(\tilde{x}(W)-x)(\tilde{x}(W)-x)^T] \end{aligned} \tag{15}$$

Note that a recursive formula for Σ is given in (3).

Theorem 2 *If we set $W = \hat{V}^{-1}$ in (9) to obtain the constrained state estimate then the covariance of $\tilde{x}(\hat{V}^{-1})$ is less than the covariance of $\hat{x}$. That is, $\tilde{V}(\hat{V}^{-1}) = \hat{V} - \hat{V}D^T(D\hat{V}D^T)^{-1}D\hat{V}$. However, the error covariance of $\tilde{x}(\hat{V}^{-1})$ is greater than the error covariance of $\hat{x}$ by the same amount. That is, $\tilde{\Sigma}(\hat{V}^{-1}) = \Sigma + \hat{V}D^T(D\hat{V}D^T)^{-1}D\hat{V}$.*

Proof: For ease of notation, we will use $\tilde{x}$ to indicate $\tilde{x}(\hat{V}^{-1})$, and $\tilde{\Sigma}$ to indicate $\tilde{\Sigma}(\hat{V}^{-1})$ in this proof. First we prove that the covariance of $\tilde{x}$ is less than the covariance of $\hat{x}$. If $W = \hat{V}^{-1}$, we see from (9) that:

$$\begin{aligned} \tilde{x} &= \hat{x} - \hat{V}D^T(D\hat{V}D^T)^{-1}(D\hat{x}-d) \\ &= \hat{x} - \hat{V}D^T(D\hat{V}D^T)^{-1}D\hat{x} + \hat{V}D^T(D\hat{V}D^T)^{-1}d \\ &= \hat{x} - M\hat{x} + Nd \end{aligned} \tag{16}$$

where $N = \hat{V}D^T(D\hat{V}D^T)^{-1}$ and $M = ND$. We next obtain:

$$\begin{aligned} \tilde{V} &= E[(\tilde{x}-\bar{\tilde{x}})(\tilde{x}-\bar{\tilde{x}})^T] \\ &= E[(\hat{x}-M\hat{x}+Nd-\bar{\tilde{x}})(\hat{x}-M\hat{x}+Nd-\bar{\tilde{x}})^T] \\ &= E[\hat{x}\hat{x}^T] - E[\hat{x}\hat{x}^T]M^T + E[\hat{x}]d^TN^T - E[\hat{x}]\bar{\tilde{x}}^T \\ &\quad -ME[\hat{x}\hat{x}^T] + ME[\hat{x}\hat{x}^T]M^T - ME[\hat{x}]d^TN^T + ME[\hat{x}]\bar{\tilde{x}}^T \\ &\quad +NdE[\hat{x}^T] - NdE[\hat{x}^T]M^T + Ndd^TN^T - Nd\bar{\tilde{x}}^T \\ &\quad -\bar{\tilde{x}}E[\hat{x}^T] + \bar{\tilde{x}}E[\hat{x}^T]M^T - \bar{\tilde{x}}d^TN^T + \bar{\tilde{x}}\bar{\tilde{x}}^T \\ &= T_1 + \cdots + T_{16} \end{aligned} \tag{17}$$

where the definitions of the 16 terms T_1 through T_{16} are apparent from the above equation. Recalling that $\bar{\hat{x}} = \bar{\tilde{x}} = \bar{x}$, we see that $T_9 + T_{12} = 0$. Recalling that $D\bar{x} = d$, we see that $D\bar{\tilde{x}} = d$. We therefore obtain:

$$T_{10} + T_{11} = 0 \tag{18}$$

Since $\bar{\hat{x}} = \bar{\tilde{x}}$ we can see that $T_{13} + T_{16} = 0$ and $T_3 + T_{15} = 0$. Next we compute:

$$\begin{aligned} T_1 + T_4 &= \hat{V} \\ T_5 + T_8 &= -M\hat{V} \\ T_2 + T_{14} &= -\hat{V}M^T \\ T_6 + T_7 &= M\hat{V}M^T \end{aligned} \tag{19}$$

Combining the above calculations with (17) gives:

$$\tilde{V} = \hat{V} - M\hat{V} - \hat{V}M^T + M\hat{V}M^T \tag{20}$$

Now note that the last two terms in the above equation can be written as:

$$\begin{aligned} -\hat{V}M^T + M\hat{V}M^T &= -\hat{V}D^T(D\hat{V}D^T)^{-1}D\hat{V} + \hat{V}D^T(D\hat{V}D^T)^{-1}D\hat{V}D^T(D\hat{V}D^T)^{-1}D\hat{V} \\ &= 0 \end{aligned} \tag{21}$$

Combining this with (20) gives:

$$\begin{aligned} \tilde{V} &= \hat{V} - M\hat{V} \\ &= \hat{V} - \hat{V}D^T(D\hat{V}D^T)^{-1}D\hat{V} \end{aligned} \tag{22}$$

But the rightmost term in the above equation is positive semidefinite since $\hat{V}$ is a covariance matrix and D is full rank.

Now we prove that the error covariance of $\tilde{x}$ is greater than the error covariance of $\hat{x}$. From (16) we see that:

$$\tilde{x} = \hat{x} - M\hat{x} + Nd \tag{23}$$

From this we can obtain:

$$\begin{aligned}
\tilde{\Sigma} &= E[(\tilde{x}-x)(\tilde{x}-x)^T] \qquad (24)\\
&= E[(\hat{x}-M\hat{x}+Nd-x)(\hat{x}-M\hat{x}+Nd-x)^T]\\
&= E[\hat{x}\hat{x}^T - \hat{x}\hat{x}^T M^T + \hat{x}d^T N^T - \hat{x}x^T\\
&\quad -M\hat{x}\hat{x}^T + M\hat{x}\hat{x}^T M^T - M\hat{x}d^T N^T + M\hat{x}x^T\\
&\quad +Nd\hat{x}^T - Nd\hat{x}^T M^T + Ndd^T N^T - Ndx^T\\
&\quad -x\hat{x}^T + x\hat{x}^T M^T - xd^T N^T + xx^T]\\
&= T_1 + T_2 + \cdots + T_{16}
\end{aligned}$$

where the T_i terms ($i = 1, \cdots, 16$) are apparent from the above equation. Since $E(\hat{x}) = E(x)$, we see that $T_3 + T_{15} = 0$. For the same reason, we see that $T_9 + T_{12} = 0$. We can also see that:

$$\begin{aligned}
T_{10} + T_{11} &= 0 \qquad (25)\\
T_1 + T_4 &= 0\\
T_2 + T_{14} &= 0\\
T_5 + T_8 &= 0\\
T_6 + T_7 &= M\hat{V}M^T
\end{aligned}$$

Now we define the estimation error as:

$$\epsilon \equiv x - \hat{x} \qquad (26)$$

With this definition we compute the sum:

$$\begin{aligned}
T_{13} + T_{16} &= -E[x\hat{x}^T] + E[xx^T] \qquad (27)\\
&= -E[(\hat{x}+\epsilon)\hat{x}^T] + E[(\hat{x}+\epsilon)(\hat{x}+\epsilon)^T]\\
&= -E[\hat{x}\hat{x}^T] - E[\epsilon\hat{x}^T] + E[\hat{x}\hat{x}^T] + E[\epsilon\hat{x}^T]\\
&\quad +E[\hat{x}\epsilon^T] + E[\epsilon\epsilon^T]\\
&= E[\hat{x}\epsilon^T] + E[\epsilon\epsilon^T]
\end{aligned}$$

$$
\begin{aligned}
&= 0 + E[\epsilon\epsilon^T] \\
&= \Sigma
\end{aligned}
$$

Combining these calculations with (24) gives us the result:

$$
\begin{aligned}
\tilde{\Sigma} &= \Sigma + M\hat{V}M^T \qquad (28) \\
&= \Sigma + \hat{V}D^T(D\hat{V}D^T)^{-1}D\hat{V}D^T(D\hat{V}D^T)^{-1}D\hat{V} \\
&= \Sigma + \hat{V}D^T(D\hat{V}D^T)^{-1}D\hat{V}
\end{aligned}
$$

But the rightmost term in the above equation is positive semidefinite since $\hat{V}$ is a covariance matrix and D is full rank.

QED

The previous theorem shows that for the constrained Kalman filter $\tilde{\Sigma}(\hat{V}^{-1})$ is greater than Σ by the same amount that $\tilde{V}(\hat{V}^{-1})$ is less than $\hat{V}$. The use of the constrained filter decreases the accuracy of the estimation but it reduces the variance of the estimation. The unbiased constrained filter may therefore be attractive in state feedback control applications where it is desirable to limit changes in the control signal. The constrained filter will also give better performance in cases where it is known *a priori* that the true state has deterministic constraints [11].

The previous theorem shows the constrast between the statistically constrained filter and the deterministically constrained filter [11]. For deterministic constraints the error covariance of the constrained filter is smaller than that of the unconstrained filter. However, for statistical constraints the error covariance of the constrained filter is larger than that of the unconstrained filter.

Theorem 3 *If we set $W = \hat{V}^{-1}$ in (9) to obtain the constrained state estimate $\tilde{x}(\hat{V}^{-1})$, then the trace of the covariance of $\tilde{x}(\hat{V}^{-1})$ is less than the trace of the covariance of $\hat{x}$. That is, trace($\tilde{V}(\hat{V}^{-1})$) $<$ trace($\hat{V}$). Similarly, if we set $W = I$ in (9) to obtain the constrained state estimate $\tilde{x}(I)$, then the trace of the covariance of $\tilde{x}(I)$ is less than the trace of the covariance of $\hat{x}$. That is, trace($\tilde{V}(I)$) $<$ trace($\hat{V}$).*

Proof: If $W = \hat{V}^{-1}$ the result is apparent from (22). If $W = I$, we see from (9) that:

$$\begin{aligned} \tilde{x}(I) &= \hat{x} - D^T(DD^T)^{-1}(D\hat{x} - d) \\ &= \hat{x} - D^T(DD^T)^{-1}D\hat{x} + D^T(DD^T)^{-1}d \end{aligned} \tag{29}$$

Following the proof of the first part of the previous theorem we can derive:

$$\tilde{V}(I) = \hat{V} - H\hat{V} - \hat{V}H^T + H\hat{V}H^T \tag{30}$$

where the symmetric matrix H is given by:

$$H = D^T(DD^T)^{-1}D \tag{31}$$

For two equally-dimensioned square matrices E and F we know that $\text{Trace}(E) = \text{Trace}(E^T)$, $\text{Trace}(E + F) = \text{Trace}(E) + \text{Trace}(F)$, and $\text{Trace}(EF) = \text{Trace}(FE)$. We therefore obtain:

$$\begin{aligned} \text{Trace}(\tilde{V}(I)) &= \text{Trace}(\hat{V} - H\hat{V} - \hat{V}H^T + H\hat{V}H^T) \\ &= \text{Trace}(\hat{V} - H\hat{V} - \hat{V}H^T + HH^T\hat{V}) \\ &= \text{Trace}(\hat{V} - H\hat{V} - H\hat{V} + H\hat{V}) \\ &= \text{Trace}(\hat{V} - H\hat{V}) \end{aligned} \tag{32}$$

Both H and $\hat{V}$ are positive definite, so $\text{Trace}(H\hat{V})$ is positive. We therefore conclude that

$$\text{Trace}(\tilde{V}(I)) < \text{Trace}(\hat{V}) \tag{33}$$

as the theorem states.

QED

This theorem shows a similarity between the constrained filters for deterministic and statistical state constraints [11]. Both types of constraints result in a constrained filter with a smaller estimate covariance than the unconstrained filter.

Theorem 4 *Setting $W = \hat{V}^{-1}$ in (9) results in the constrained state estimate $\tilde{x}(\hat{V}^{-1})$. This is the minimum variance constrained state estimate among all constrained state estimates $\tilde{x}(W)$. That is, $\tilde{V}(\hat{V}^{-1}) \leq \tilde{V}(W)$ for all positive definite symmetric W.*

Proof: First we note that with a general weighting matrix W the variance of the constrained state estimate can be written as:

$$\begin{aligned}\tilde{V}(W) &= E[(\tilde{x}(W) - \bar{\tilde{x}}(W))(\tilde{x}(W) - \bar{\tilde{x}}(W))^T] \\ &= E[(\hat{x} - W^{-1}D^T(DW^{-1}D^T)^{-1}D\hat{x} + W^{-1}D^T(DW^{-1}D^T)^{-1}d - \bar{\tilde{x}})(\cdots)^T]\end{aligned} \tag{34}$$

After several lines of algebraic manipulation we obtain:

$$\begin{aligned}\tilde{V}(W) &= \hat{V} - W^{-1}D^T(DW^{-1}D^T)^{-1}D\hat{V} - \hat{V}D^T(DW^{-1}D^T)^{-1}DW^{-1} \\ &\quad + W^{-1}D^T(DW^{-1}D^T)^{-1}D\hat{V}D^T(DW^{-1}D^T)^{-1}DW^{-1}\end{aligned} \tag{35}$$

Now replace W with $\hat{V}^{-1}$ in the above equation to obtain:

$$\begin{aligned}\tilde{V}(\hat{V}^{-1}) &= \hat{V} - \hat{V}D^T(D\hat{V}D^T)^{-1}D\hat{V} - \hat{V}D^T(D\hat{V}D^T)^{-1}D\hat{V} \\ &\quad + \hat{V}D^T(D\hat{V}D^T)^{-1}D\hat{V}D^T(D\hat{V}D^T)^{-1}D\hat{V} \\ &= \hat{V} - \hat{V}D^T(D\hat{V}D^T)^{-1}D\hat{V}\end{aligned} \tag{36}$$

Subtracting $\tilde{V}(\hat{V}^{-1})$ from $\tilde{V}(W)$ results in:

$$\begin{aligned}\tilde{V}(W) - \tilde{V}(\hat{V}^{-1}) &= W^{-1}D^T(DW^{-1}D^T)^{-1}D\hat{V}D^T(DW^{-1}D^T)^{-1}DW^{-1} \\ &\quad - W^{-1}D^T(DW^{-1}D^T)^{-1}D\hat{V} - \hat{V}D^T(DW^{-1}D^T)^{-1}DW^{-1} \\ &\quad + \hat{V}D^T(D\hat{V}D^T)^{-1}D\hat{V} \\ &= JJ^T\end{aligned} \tag{37}$$

where J is given by:

$$J = W^{-1}D^T(DW^{-1}D^T)^{-1}(D\hat{V}D^T)^{1/2} - \hat{V}D^T(D\hat{V}D^T)^{-1/2} \tag{38}$$

The square root of a matrix F is a matrix G such that $F = GG^T$. If F is positive definite, then G always exists and is invertible. If F is symmetric then a symmetric G exists. Since $\tilde{V}(W) - \tilde{V}(\hat{V}^{-1})$ is of the form JJ^T, we know that $\tilde{V}(W) - \tilde{V}(\hat{V}^{-1})$ is positive semidefinite, which proves that $\tilde{V}(\hat{V}^{-1}) \leq \tilde{V}(W)$.

QED

5. Recursive Computation of $\hat{V}$

The previous section shows that setting $W = \hat{V}^{-1}$ in (9) results in several attractive properties of the constrained Kalman filter. This section shows how $\hat{V}$ can be computed recursively. Note that:

$$\begin{aligned} \hat{V} &= E[(\hat{x} - \bar{\hat{x}})(\hat{x} - \bar{\hat{x}})^T] \\ &= E[\hat{x}\hat{x}^T] - \bar{\hat{x}}\bar{\hat{x}}^T \\ &= E[(x - \epsilon)(x - \epsilon)^T] - E(x - \epsilon)E(x - \epsilon)^T \\ &= E(xx^T) - E(\epsilon x^T) - E(x\epsilon^T) + E(\epsilon\epsilon^T) - \bar{x}\bar{x}^T + \bar{x}\bar{\epsilon}^T + \bar{\epsilon}\bar{x}^T - \bar{\epsilon}\bar{\epsilon}^T \\ &= V + \Sigma - \alpha - \alpha^T \end{aligned} \tag{39}$$

where we have used the fact that $\bar{\epsilon} = 0$, and α and V are defined as:

$$\begin{aligned} \alpha &\equiv E(\epsilon x^T) \\ V &\equiv E[(x - \bar{x})(x - \bar{x})^T] \end{aligned} \tag{40}$$

Note that α can be written as:

$$\begin{aligned} \alpha &= E(\epsilon x^T) \\ &= E[\epsilon(\hat{x} + \epsilon)^T] \\ &= E(\epsilon\epsilon^T) \\ &= \Sigma \end{aligned} \tag{41}$$

where we have used the projection theorem. $\hat{V}$ is therefore given as:

$$\hat{V} = V - \Sigma \tag{42}$$

We can obtain a recursive formula for V as follows:

$$\begin{aligned} V_{k+1} &= E[(x_{k+1} - \bar{x}_{k+1})(x_{k+1} - \bar{x}_{k+1})^T] \\ &= E[(A(x_k - \hat{x}_k) + B(u_k - \bar{u}_k) + w_k - \bar{w}_k)(A(x_k - \hat{x}_k) + B(u_k - \bar{u}_k) + w_k - \bar{w}_k)^T] \\ &= AE[(x_k - \bar{x}_k)(x_k - \bar{x}_k)^T]A^T + E[(w_k - \bar{w}_k)(w_k - \bar{w}_k)^T] \\ &= AV_kA^T + Q \end{aligned} \tag{43}$$

We combine (42) with the recursive formula for V and the recursive formula for Σ in (3) to obtain a recursive computation for $\hat{V}$.

6. Example

This section presents a simple example to illustrate the theory of this paper. Consider the scalar system:

$$\begin{aligned} x_{k+1} &= \frac{1}{2}x_k + w_k \\ y_k &= x_k + e_k \\ x_0 &= 0 \\ \Sigma_0 &= 0 \\ w_k &\sim N(0,Q), \qquad Q = \sqrt{2} - 5/4 \\ e_k &\sim N(0,R), \qquad R = 1 \end{aligned} \tag{44}$$

The steady state covariance of the unconstrained estimation error can be derived as:

$$\Sigma = \frac{\sqrt{2}-1}{2} \tag{45}$$

The steady state covariance of the state can be derived as:

$$V = 4/3 \tag{46}$$

The steady state covariance of the unconstrained filter estimate is derived from (42) as:

$$\begin{aligned} \hat{V} &= V - \Sigma \\ &= \frac{4}{3} - \frac{\sqrt{2}-1}{2} \end{aligned} \tag{47}$$

It can be seen that the expected value of the state is zero for all time. Therefore, in the constraint $D\bar{x}_k = d_k$, the D matrix is equal to one, and the d_k vector is equal to zero. Therefore the constrained state estimate is computed as:

$$\tilde{x}_k = \hat{x}_k - W^{-1}D^T(DW^{-1}D^T)^{-1}(D\hat{x}_k - d_k) \tag{48}$$

$$
\begin{aligned}
&= \hat{x}_k - W^{-1}(W^{-1})^{-1}\hat{x}_k \\
&= 0
\end{aligned}
$$

Note that in this simple example the constrained estimate is independent of the weighting matrix W. The covariance of the constrained state estimate is derived from (14) as:

$$
\begin{aligned}
\tilde{V} &= E[(\tilde{x} - \bar{x})^2] \\
&= E[(0 - 0)^2] \\
&= 0
\end{aligned}
\tag{49}
$$

This is consistent with Theorem 2, which states that:

$$
\begin{aligned}
\tilde{V} &= \hat{V} - \hat{V}D^T(D\hat{V}D^T)^{-1}D\hat{V} \\
&= 0
\end{aligned}
\tag{50}
$$

Clearly the trace of $\tilde{V}$ is less than the trace of $\hat{V}$, which illustrates Theorem 3. Since $\tilde{x} = 0$, the error covariance of the constrained estimate is equal to the covariance of the state:

$$
\begin{aligned}
\tilde{\Sigma} &= E[(\tilde{x} - x)^2] \\
&= E[x^2] \\
&= 4/3
\end{aligned}
\tag{51}
$$

This is consistent with Theorem 2, which states that:

$$
\begin{aligned}
\tilde{\Sigma} &= \Sigma + \hat{V}D^T(D\hat{V}D^T)^{-1}D\hat{V} \\
&= \Sigma + \hat{V} \\
&= V \\
&= 4/3
\end{aligned}
\tag{52}
$$

7. Conclusion

We have presented a method for incorporating linear statistical state equality constraints in the Kalman filter. We obtained a family of constrained filters with each member parameterized by a weighting matrix. We showed that each constrained filter in this family is unbiased. We also showed that, for a particular member of the family of constrained filters, the covariance of the constrained filter is smaller than the covariance of the unconstrained filter. However, the error covariance of the constrained filter is larger than the error covariance of the unconstrained filter. Finally, we derived the particular constrained filter that has the smallest possible covariance.

If we have nonlinear constraints they can easily be linearized in a manner similar to [11]. If we have inequality constraints the methods of this paper can be extended in a manner similar to [12]. Since the statistically constrained filter is identical in its formulation to the deterministically constrained filter we do not show simulation results here. Readers who are interested in simulation results and Matlab source code are referred to [11, 12].

References

[1] B. Anderson & J. Moore, *Optimal filtering* (Englewood Cliffs, New Jersey: Prentice Hall, 1979).

[2] D. Massicotte, R. Morawski, & A. Barwicz, Incorporation of a positivity constraint into a Kalman-filter-based algorithm for correction of spectrometric data, *IEEE Transactions on Instrumentation and Measurement 44*(1), February 1995, 2-7.

[3] W. Wen & H. Durrant-Whyte, Model-based multi-sensor data fusion, *IEEE International Conference on Robotics and Automation*, Nice, France, 1992, 1720-1726.

[4] S. Hayward, Constrained Kalman filter for least-squares estimation of time-varying beamforming weights, in J. McWhirter & I. Proudler (Eds.), *Mathematics in Signal Processing*, IV (New York, NY: Oxford University Press, 1998) 113-125.

[5] J. Porrill, Optimal combination and constraints for geometrical sensor data, *International Journal of Robotics Research 7*(6), December 1988, 66-77.

[6] P. Maybeck, *Stochastic models, estimation, and control – volume 1* (New York, NY: Academic Press, 1979).

[7] R. Stengel, *Optimal control and estimation* (New York, NY: Dover Publications, 1994).

[8] T. Chia, P. Chow, & H. Chizeck, Recursive Parameter identification of constrained systems: an application to electrically stimulated muscle, *IEEE Transactions on Biomedical Engineering 38*(5), May 1991, 429-442.

[9] W. Timmons, P. Katona, V. Chankong, & H. Chizeck, Parameter constrained adaptive control, *Industrial and Engineering Chemistry Research 36*(11), November 1997, 4894-4905.

[10] T. Chia, *Parameter identification and state estimation of constrained systems*, doctoral diss., Case Western Reserve University, Cleveland, OH, 1985.

[11] D. Simon & T. Chia, Kalman filtering with state equality constraints, *IEEE Transactions on Aerospace and Electronic Systems, 39*(1), January 2002, 128-136.

[12] D. Simon & D.L. Simon, Aircraft turbofan engine health estimation using constrained Kalman filtering, *Journal of Engineering for Gas Turbines and Power*, in print.