OPTIMUM FILTERS (5)

DISCRETE KALMAN FILTER

Kalman filter:
- Kalman filtering problem
- Kalman prediction problem
- Summary of the implementation of discrete Kalman filter
- Steady state properties of Kalman filters

Appendix 7B. Derivation of matrix differentiation formulas
Kalman filters are optimum filters that are different from Wiener filters and have the following distinctive features:

(i) the problems are formulated in terms of state-space model (or state-space form);
(ii) the solution of the problems is computed recursively;
(iii) they can be used to treat stationary random processes, as well as non-stationary processes.

As we learnt from the previous lectures, an optimum Wiener filter is designed using the Wiener-Hopf equations in which the autocorrelation $r_s(k)$ of the noise observation $x(n)$ (that contains the desired signal $d(n)$ with $r_d(k)$) and the cross-correlation $r_{ds}(k)$ are used. Use of $r_d(k)$, $r_s(k)$ and $r_{ds}(k)$ indicates that $d(n)$ and $x(n)$ are wide-sense stationary as well as jointly wide-sense stationary random processes. Whereas, for a non-stationary random process, its statistical properties are not stationary; in other words, they are time varying. For example, a deterministic sinusoidal signal $d(n) = A \sin(\omega n + \varphi)$ interfered with noise $v(n)$ is a non-stationary random process, $x(n) = d(n) + v(n)$, because the mean of $x(n)$, $m_x(n) = A \sin(\omega n + \varphi)$, depends on time $n$, and the autocorrelation, $r_x(n+k,n) = r_x(k) + (A^2/2)\left[\cos(\alpha k) - \cos[\omega(2n+k)+2\varphi]\right]$, depends not only on time lag $k$ but also on time $n$. The cross-correlation $r_{dx}(n+k,n) = (A^2/2)\left[\cos(\alpha k) - \cos[\omega(2n+k)+2\varphi]\right]$ also depends on time $n$.

Therefore, in the optimum sense a Wiener filter can only be used to treat stationary random processes. As we will see below, Wiener filtering problems are just a special case, the steady-state case, of Kalman filtering problems.

**Kalman filter:**

**Kalman filtering problem**

Kalman filtering addresses the general problem of trying to get the best estimate of the state $x(n)$ of a process governed by the state equation (linear stochastic difference equation)

$$ x(n) = A(n-1)x(n-1) + w(n) \quad (217) $$

from measurements given by the observation equation

$$ y(n) = C(n)x(n) + v(n) \quad . \quad (218) $$

![Fig. 19. The diagram for constructing a Kalman filter.](image)

Specifically, when the measurements $y(n)$ (measurement vector) are input to the Kalman filter $h(n)$, the filter outputs $\hat{x}(n|n)$, which is the best linear estimate vector of the state vector $x(n)$ of length $p$ based on the observations $y(l)$ for $0 \leq l \leq n$ (Fig. 19).

The best estimate of $x(n)$, $\hat{x}(n|n)$, from the Kalman filter is realized by minimizing the mean-square error
The initial estimation error covariance matrix (Eq. (220)) for this estimate is
\[
P(n | n) = E\{e(n | n)e^H(n | n)\}
\]
is called estimation error covariance matrix (a \(p \times p\) matrix), and \(e(n | n)\) (a vector of length \(p\)) is the estimate error, given by
\[
e(n | n) = \begin{bmatrix} x(n) - \hat{x}(n | n) \\ e(n) - \hat{e}(n | n) \\ \vdots \\ e(n-p) - \hat{e}(n-p | n-p) \end{bmatrix}
\]
\[
= \begin{bmatrix} x(n) - \hat{x}(n | n) \\ x(n-1) - \hat{x}(n-1 | n-1) \\ \vdots \\ x(n-p) - \hat{x}(n-p | n-p) \end{bmatrix}
\]
\[
\text{tr}\{A\} = \sum_{n=1}^{p} a_{nm}
\]
\[
\text{Thus}, \text{ } \text{tr}\{P(n | n)\} = \sum_{k=0}^{n} E\{e(n-k | n-k)^2\}.
\]

Another error covariance matrix that we need in deriving the Kalman filter is prediction error covariance matrix, defined as
\[
P(n-1 | n) = E\{e(n-1 | n)e^H(n-1 | n)\}
\]
where
\[
e(n-1 | n) = x(n) - \hat{x}(n | n-1)
\]
is the one-step prediction error (a vector of length \(p\)), in which \(\hat{x}(n | n-1)\) is the prediction vector of the state vector \(x(n)\) of length \(p\), based on the observations \(y(l)\) for \(0 \leq l \leq n-1\).

To derive the Kalman filter we shall carry out a recursive procedure of two steps: prediction using previous measurements, and estimation using new measurement (refer to Fig. 20). In the first step, the state \(x(n)\) is predicted using the previous observations \(y(l)\) up to time \(l=n-1\) so as to obtain \(\hat{x}(n | n-1)\). In the second step, the new information \(\varepsilon(n) = y(n) - \hat{y}(n | n-1)\) (the innovation process of \(y(n)\)) from new measurement \(y(n)\) is added to the estimate \(\hat{x}(n | n-1)\) so as to get the best estimate of \(x(n)\), \(\hat{x}(n | n)\).

**Initialization** \(\hat{x}(0 | 0) = E\{x(0)\}\)

Step 1: predicting \(x(1)\) from \(\hat{x}(0 | 0)\) \(\Rightarrow \hat{x}(1 | 0) = A(0)\hat{x}(0 | 0)\);

Step 2: estimating \(x(1)\) by adding new info \(\varepsilon(1) \Rightarrow \hat{x}(1 | 1) = \hat{x}(1 | 0) + K(1)\varepsilon(1)\)

Repeat 1: predicting \(x(2)\) from \(\hat{x}(1 | 1) \Rightarrow \hat{x}(2 | 1) = A(1)\hat{x}(1 | 1)\);

Repeat 2: estimating \(x(2)\) by adding new info \(\varepsilon(2) \Rightarrow \hat{x}(2 | 2) = \hat{x}(2 | 0) + K(2)\varepsilon(2)\)

\[\ldots\]

Repeat 1: predicting \(x(n) \Rightarrow \hat{x}(n | n-1) = A(n-1)\hat{x}(n-1 | n-1)\);

Repeat 2: estimating \(x(n)\) by adding new info \(\varepsilon(n) \Rightarrow \hat{x}(n | n) = \hat{x}(n | n-1) + K(n)\varepsilon(n)\)

**Fig. 20** The two-step recursive procedure of deriving Kalman filter.

Suppose that a given random process \(x(n)\) has an initial state \(x(0)\) with a given estimate \(\hat{x}(0 | 0) = E\{x(0)\}\). The initial estimation error covariance matrix (Eq. (220)) for this estimate is
\[
P(0 | 0) = E\{e(0 | 0)e^H(0 | 0)\}
\]
With the given estimate \(\hat{x}(0 | 0) = E\{x(0)\}\), we first predict \(x(1)\) using the relation \(\hat{x}(1 | 0) = A(0)\hat{x}(0 | 0)\) and
find $\hat{y}(1|0) = C(1) \hat{x}(1|0)$ as well. Secondly, we make optimal use of the available measurement $y(1)$ by adding to $\hat{x}(1|0)$ the new information $\varepsilon(1) = y(1) - \hat{y}(1|0)$ (the innovation process of $y(1)$), and find the best estimate of state $x(1)$, $\hat{x}(1|1)$. Optimal use of $\varepsilon(1)$ is made by a weighting factor $K(1)$ that minimizes the mean-square error $\xi(n)$ at time $n=1$, $\xi(1) = \text{tr}\{P(1|1)\} = \sum_{k=0}^{p-1} E\{e(1-k | 1-k)^T e(1-k | 1-k)\}$.

Repeating the above two steps for the next time instant, we use $\hat{x}(1|1)$ to predict $x(2)$ using the relation $\hat{x}(2|1) = A(1) \hat{x}(1|1)$ and find $\hat{y}(2|1) = C(2) \hat{x}(2|1)$. Then we update $\hat{x}(2|1)$ by adding new information $\varepsilon(2) = y(2) - \hat{y}(2|1)$ from the next available measurement $y(2)$, and find the best estimate of $x(2)$, $\hat{x}(2|2)$, by optimizing $K(2)$, the weight on $\varepsilon(2)$, that minimizes $\xi(2) = \text{tr}\{P(2|2)\} = \sum_{k=0}^{p-1} E\{e(2-k | 2-k)^T e(2-k | 2-k)\}$.

We carry on this until time $n$, and then we can formulate a recursive procedure for time $n$; that is for each $n>0$, given $\hat{x}(n-1|n-1)$ and $P(n-1|n-1)$, we can predict $x(n)$ and obtain $\hat{x}(n|n-1)$, and then we make optimal use of the new information $\varepsilon(n) = y(n) - \hat{y}(n|n-1)$ from a new available measurement $y(n)$ to find the best estimate $\hat{x}(n|n)$ of the state $x(n)$ by finding optimal $K(n)$ that minimizes the mean-square error $\xi(n) = \text{tr}\{P(n|n)\} = \sum_{k=0}^{p-1} E\{e(n-k | n-k)^T e(n-k | n-k)\}$.

The Kalman filter is established based on the above recursive procedure, as follows. In the first step, $x(n)$ is projected ahead (predicted) via the state transition matrix in the following manner

$$\hat{x}(n|n-1) = A(n-1) \hat{x}(n-1|n-1)$$

which is obtained by ignoring the contribution of $w(n)$ in the state equation (Eq. (217)) because $w(n)$ is a white noise process with zero mean and is not correlated with its previous values, $w(l)$ for $l<n$.

The prediction error in this case is

$$e(n|n-1) = x(n) - \hat{x}(n|n-1) = A(n-1)x(n-1) + w(n) - A(n-1)\hat{x}(n-1|n-1)$$

$$= A(n-1)e(n-1|n-1) + w(n)$$

(266)

Since $e(n-1|n-1)$ is uncorrelated with $w(n)$ (due to the fact that $w(n)$ is white noise and thus uncorrelated with $w(l)$ for $n \neq l$) so that $E\{e(n-1|n-1)w^H(n)\} = 0$, the prediction error covariance matrix becomes

$$P(n|n-1) = E\{e(n|n-1)e^H(n|n-1)\} = A(n-1)P(n-1|n-1)A^H(n-1) + Q_w(n)$$

(277)

This completes the first step of deriving the Kalman filter.

In the second step, the optimal estimation of $x(n)$ is conducted by adding the new measurement $y(n)$ to the estimate $\hat{x}(n|n-1)$ in terms of the innovation process of $y(n)$,

$$\varepsilon(n) = y(n) - \hat{y}(n|n-1)$$

(288)

which is the error between $y(n)$ and its prediction $\hat{y}(n|n-1)$ and represents the new information that can not be predicted by $\hat{y}(n|n-1)$. The predict $\hat{y}(n|n-1)$ is given by

$$\hat{y}(n|n-1) = C(n)\hat{x}(n|n-1)$$

(299)

which is derived by ignoring contribution of $v(n)$ in the observation equation (Eq. (218)) because $v(n)$ is assumed to be zero-mean white noise and is not correlated with its previous values, $v(l)$ for $l<n$.

Therefore, the optimum estimate $\hat{x}(n|n)$ of $x(n)$ turns out to be

$$\hat{x}(n|n) = \hat{x}(n|n-1) + K(n)\{y(n) - \hat{y}(n|n-1)\}$$

(300)

in which the innovations process is scaled by an optimum gain $K(n)$ that is determined by minimizing mean-square error $\xi(n)$. Inserting Eq. (299) in Eq. (300), we have...
A feature in this equation is that the estimate $\hat{x}(n|n)$ is related with $\hat{x}(n|n-1)$, the predict of $x(n)$ using previous values of $x(l)$ up to time instant $l=n-1$. Applying Eq. (225) to Eq. (231) yields

$$\hat{x}(n|n) = \Lambda(n-1)\hat{x}(n-1|n-1) + K(n)[y(n) - C(n)\hat{x}(n|n-1)]$$

(232)

which is the recursive relation between the estimate $\hat{x}(n|n)$ and the previous estimate $\hat{x}(n-1|n-1)$. The gain $K(n)$ is known as the Kalman gain. The recursive relation in Eq. (232) corresponds to the discrete Kalman filter, which is a recursive filter and illustrated in Fig. 21.

As we will see below, the recursive optimum estimation of $x(n)$ results from the optimization of Kalman gain $K(n)$ by minimizing MSE $\xi(n) = \text{tr}\{P(n|n)\}$, like we did for optimum Wiener filters. It is because the Kalman gain $K(n)$ is optimum that the Kalman filter is an optimum filter! In general, $\Lambda(n-1)$, $C(n)$ and $K(n)$ are time varying, and thus the Kalman filters are applicable to non-stationary random processes as well.

The estimate error (see Eq. (221)) is

$$e(n|n) = x(n) - \hat{x}(n|n) = x(n) - \hat{x}(n|n-1) - K(n)[C(n)x(n) + v(n) - C(n)\hat{x}(n|n-1)]$$

$$= e(n|n-1) - K(n)C(n)e(n|n-1) - K(n)v(n)$$

(237)

in which Eqs. (231) and (218) are used. Since $w(n)$ and $v(n)$ are uncorrelated, then $x(n)$ and $e(n|n-1)$ both are uncorrelated with $v(n)$. Thus, the prediction error covariance matrix becomes

$$P(n|n) = E\{e(n|n)e^H(n|n)\} = [I - K(n)C(n)]P(n|n-1)[I - K(n)C(n)]^H + K(n)Q_s(n)K^H(n)$$

(238)

With Eq. (238), we minimize the mean square error $\xi(n) = \text{tr}\{P(n|n)\}$ to get the optimum Kalman gain $K(n)$. To do so we need to use the following matrix differentiation formulas,

$$\frac{d}{dK}\text{tr}\{KC\} = C^H$$

(239)

$$\frac{d}{dK}\text{tr}\{KCK^H\} = 2KC$$

(240)

where the derivative with respect to the matrix $K$ means

$$\frac{ds}{dK} = \begin{bmatrix}
\frac{ds}{dk_{11}} & \frac{ds}{dk_{12}} & \cdots \\
\frac{ds}{dk_{21}} & \frac{ds}{dk_{22}} & \cdots \\
& & \ddots \\
& & & \ddots
\end{bmatrix};$$

(241)
The derivations of Eqs. (239) and (240) are given in Appendix 7B. With the help of Eqs. (239) and (240) and setting \( d\xi(n)/dK(n) = 0 \), we have

\[
\frac{d\xi(n)}{dK(n)} = \frac{d}{dK(n)} \text{tr}(P(n|n)) = -2[I - K(n)C(n)]P(n|n-1)C^H(n) + 2K(n)Q_x(n) = 0
\]  

(242)

Solving Eq. (242) for \( K(n) \) we get the optimum Kalman gain,

\[
K(n) = P(n|n-1)C(n)[P(n|n-1)C^H(n) + Q_x(n)]^{-1}
\]  

(243)

It is interesting to look into how \( K(n) \) is adapted to measurement noise. When \( Q_x(n) = 0 \) (namely, \( v(n) = 0 \) then \( K(n) = C^{-1}(n) \) so that \( \hat{x}(n|n) = C^{-1}(n)y(n) = x(n) \), that is, the estimate is identical to the state \( x(n) \). When \( Q_x(n) \to \infty \), then \( K(n) \to 0 \). In this case, Eq. (232) becomes \( \hat{x}(n|n) = A(n-1)\hat{x}(n-1|n-1) \), which shows that the measurements \( y(n) \) become completely unreliable and are not used in the estimation.

Like we had a minimum mean square error for an optimum Wiener filter due to the orthogonality principle, we can find the optimum mean square error by using Eq. (242).

Re-arranging the error covariance matrix \( P(n|n) \) in Eq. (238) in the following manner,

\[
P(n|n) = [I - K(n)C(n)]P(n|n-1) + \left\{ [I - K(n)C(n)]P(n|n-1)C^H(n) + K(n)Q_x(n) \right\}K^H(n)
\]  

(244)

and it follows from Eq. (242) that the second term in the braces \{ \} in Eq. (244) is zero. Thus, we obtain the minimized error covariance matrix

\[
P(n|n) = [I - K(n)C(n)]P(n|n-1)
\]  

(245)

Since the Kalman gain \( K(n) \) and the error covariance matrix \( P(n|n) \) are independent of the data \( x(n) \), they can be computed off-line.

It should be pointed out that one of the very appealing features of the Kalman filter is the recursive nature (shown in Fig. 20), which makes practical implementations much more feasible than an implementation of a Wiener filter which is designed to operate on all of the data directly for each estimate. The Kalman filter instead recursively conditions the current estimate on all of the past measurement.

**Kalman prediction problem**

A (linear) prediction of signal \( x(n) \) is concerned with the estimation (prediction) of \( x(n+m) \) for \( m > 0 \) using the measurements \( y(l) \) up to time \( l=n \). It can be a one-step or a multiple-step prediction.

**A. One-step Kalman prediction**

For a one-step prediction problem, we perform a one-step forward time shift on Eq. (225), i.e., \( n \) replaced with \( n+1 \), and then we have

\[
\hat{x}(n+1|n) = A(n)\hat{x}(n|n)
\]  

(246)

Inserting Eq. (231) into Eq. (246), we obtain the one-step discrete Kalman filter

\[
[\hat{x}(n+1|n) = A(n)\hat{x}(n|n-1) + A(n)K(n)[y(n) - C(n)\hat{x}(n|n-1)]
\]  

(247)

which is the recursive relation between the predict \( \hat{x}(n+1|n) \) and the previous predict \( \hat{x}(n|n-1) \). For the recursive prediction relation, we introduce Kalman *prediction* gain \( K_p(n) \)

\[
K_p(n) = A(n)K(n) = A(n)P(n|n-1)C^H(n)\left[ C(n)P(n|n-1)C^H(n) + Q_x(n) \right]^{-1}
\]  

(248)

With Kalman *prediction* gain \( K_p(n) \), Eq. (247) becomes

\[
\hat{x}(n+1|n) = A(n)\hat{x}(n|n-1) + K_p(n)[y(n) - C(n)\hat{x}(n|n-1)]
\]  

(249)
In relation to the recursive prediction, it is interesting to study the prediction error covariance matrix,
\[ P(n+1 \mid n) = E\{ e(n+1 \mid n)e^H(n+1 \mid n) \} \quad (250) \]
Since
\[
e(n+1 \mid n) = x(n+1) - \hat{x}(n+1 \mid n)
\]
\[ = A(n)x(n) + w(n+1) - A(n)\hat{x}(n \mid n-1) - K_p(n)\{ C(n)x(n) + v(n) - C(n)\hat{x}(n \mid n-1) \}
\]
\[ = A(n)e(n \mid n-1) + w(n+1) - K_p(n)\{ C(n)e(n \mid n-1) + v(n) \}
\]
\[ = [A(n) - K_p(n)C(n)]E(n \mid n-1) + w(n+1) - K_p(n)v(n) \quad (251) \]
and \( w(n+1) \) and \( e(n \mid n-1) \) both are uncorrelated with \( v(n) \), then
\[
P(n+1 \mid n) = A(n)P(n \mid n-1)A^H(n) + Q_w(n) - K_p(n)C(n)P(n \mid n-1)A^H(n)
\]
Using Eq. (248) in the above equation, we may have
\[
P(n+1 \mid n) = A(n)P(n \mid n-1)A^H(n) + Q_w(n)
\]
\[ - A(n)P(n \mid n-1)C^H(n)\{ C(n)P(n \mid n-1)C^H(n) + Q_v(n) \}^{-1}C(n)P(n \mid n-1)A^H(n) \quad (252) \]
which is known as Riccati equation that gives the recursive relation between \( P(n+1 \mid n) \) and \( P(n \mid n-1) \).
The best one-step prediction of \( y(n) \) is easily obtained from Eq. (229)
\[
\hat{y}(n+1 \mid n) = C(n+1)\hat{x}(n+1 \mid n)
\]
where \( \hat{x}(n+1 \mid n) \) is given in Eq. (247) or Eq. (249).

**B. m-step Kalman prediction**

An \( m \)-step prediction of signal \( x(n) \), denoted by \( \hat{x}(n+m \mid n) \) for \( m>0 \), means that we predict the signal \( x(l) \) at \( l=n+m \) using the values of measurement \( y(n) \) up to time \( n \), and it can be obtained by extending the relation in Eq. (225) and noting that only the values of \( y(n) \) up to time \( n \) are used, in the following way
\[ \hat{x}(n+m \mid n) = A(n+m-1)\hat{x}(n+m-1 \mid n) \quad (254) \]
Performing the recursive computation on Eq. (254),
\[ \hat{x}(n+m \mid n) = A(n+m-1)\hat{x}(n+m-1 \mid n)
\]
\[ = A(n+m-1)A(n+m-2)\hat{x}(n+m-2 \mid n)
\]
\[ = \ldots \quad (255) \]
we get the \( m \)-step prediction of \( x(n) \)
\[
\hat{x}(n+m \mid n) = \prod_{k=0}^{m-1}A(n+k) \hat{x}(n \mid n)
\]
\[ \quad (256) \]
which is determined by the estimate \( \hat{x}(n \mid n) \).
Similarly, by extending the relation in Eq. (229) and noting that only the signal values \( y(n) \) up to time \( n \) are used, we can find the best \( m \)-step prediction of \( y(n) \)
\[
\hat{y}(n+m \mid n) = C(n+m)\hat{x}(n+m \mid n) = \prod_{k=0}^{m-1}A(n+k) \hat{x}(n \mid n)
\]
\[ \quad (257) \]

**Summary of the implementation of discrete Kalman filter**

State equation: \( x(n) = A(n-1)x(n-1) + w(n) \)

Observation (or measurement) equation: \( y(n) = C(n)x(n) + v(n) \)

Initial value: \( \hat{x}(0 \mid 0) = E\{ x(0) \} \), a \( p \times 1 \) vector (due to \( a(k) \) for \( k = 1, 2, \ldots, p \)
\[ P(0|0) = E[x(0)x^H(0)] \], a \( p \times p \) initial error covariance matrix

Recursive computation: For \( n = 1, 2, \ldots \)
\[ P(n|n-1) = \Lambda(n-1)P(n-1|n-1)\Lambda^H(n-1) + \mathbf{Q}_w(n) \] (227)
\[ K(n) = P(n|n-1)C^H(n)[C(n)P(n|n-1)C^H(n) + Q_v(n)]^{-1} \] (243)
\[ P(n|n) = [I - K(n)C(n)]P(n|n-1) \] (245)
\[ \hat{x}(n|n-1) = \mathbf{A}(n-1)\hat{x}(n-1|n-1) \] (225)
\[ \hat{x}(n|n) = \hat{x}(n|n-1) + K(n)[y(n) - C(n)\hat{x}(n|n-1)] \] (231)

alternatively, Eq. (231) with Eq. (225) inserted in,
\[ \hat{x}(n|n) = \mathbf{A}(n-1)\hat{x}(n-1|n-1) + K(n)[y(n) - C(n)\hat{x}(n|n-1)] \] (232)

The one-step prediction of \( x(n) \) is
\[ \hat{x}(n+1|n) = \mathbf{A}(n)\hat{x}(n|n-1) + \mathbf{A}(n)K(n)[y(n) - C(n)\hat{x}(n|n-1)] \] (247)

The best \( m \)-step prediction of \( x(n) \) is
\[ \hat{x}(n+m|n) = \prod_{k=0}^{m-1} \mathbf{A}(n+k) \hat{x}(n|n) \] (256)

The best \( m \)-step prediction of \( y(n) \) is
\[ \hat{y}(n+m|n) = C(n+m)\hat{x}(n+m|n) = \prod_{k=0}^{m-1} \mathbf{A}(n+k) \hat{x}(n|n) \] (257)

Riccati equation is
\[ P(n+1|n) = \mathbf{A}(n)P(n|n-1)\mathbf{A}^H(n) + \mathbf{Q}_w(n) \]
\[ - \mathbf{A}(n)P(n|n-1)C^H(n)[C(n)P(n|n-1)C^H(n) + Q_v(n)]^{-1} C(n)P(n|n-1)\mathbf{A}^H(n) \]

\( P(n|n-1) \) and \( P(n|n) \) are \( p \times p \) error covariance matrices.
\( \mathbf{K}(n) \) is a \( p \times r \) matrix. For single observation, \( \mathbf{K}(n) \) is a \( p \times 1 \) vector.
The process white noise variance matrix \( \mathbf{Q}_w(n) = E[w(n)w^H(n)] = \mathbf{B}(n)\mathbf{B}^H(n)\mathbf{\sigma}_w^2 \) is a \( p \times p \) matrix.
The measurement noise variance matrix \( \mathbf{Q}_v(n) = E[v(n)v^H(n)] \) is a \( r \times r \) matrix. For single observation, \( \mathbf{Q}_v(n) = \mathbf{\sigma}_v^2 \) is a scalar.

**Example 14. Kalman filter for estimating an unknown DC voltage**

Suppose that the measurement of an unknown DC voltage \( x(n) \) is corrupted by noise \( v(n) \) that is characterized by a zero-mean white noise process with a variance \( \mathbf{\sigma}_v^2 \). Find the Kalman filter of the form
\[ \hat{x}(n|n) = \mathbf{A}(n-1)\hat{x}(n-1|n-1) + K(n)[y(n) - C(n)\hat{x}(n|n-1)] \]

for the estimation of the unknown DC voltage.

**Solution.**

Since \( x(n) \) is a DC voltage, then the state equation can be expressed as
\[ x(n) = x(n-1) \]
The measurement equation is
\[ y(n) = x(n) + v(n) \]
From the state space form (the above equations), we may find that \( \mathbf{A}(n) = 1, \mathbf{C}(n) = 1, \mathbf{Q}_w(n) = 0, \) and \( \mathbf{Q}_v(n) = \mathbf{\sigma}_v^2 \). Since \( x(n) \) is scalar, then we may have
\[ P(n|n-1) = P(n-1|n-1) \]
\[ K(n) = P(n | n-1) \left[ P(n | n-1) + \sigma_v^2 \right]^{-1} \]

\[ P(n | n) = [1 - K(n)] P(n | n-1) = \left[ 1 - \frac{P(n | n-1)}{P(n | n-1) + \sigma_v^2} \right] P(n | n-1) = \frac{P(n | n-1)\sigma_v^2}{P(n | n-1) + \sigma_v^2} = \frac{P(n-1 | n-1)\sigma_v^2}{P(n-1 | n-1) + \sigma_v^2} \]

and

\[ P(n | n) = \frac{P(n | n-1)\sigma_v^2}{P(n | n-1) + \sigma_v^2} = \frac{P(n | n-1)\sigma_v^2}{K(n)\sigma_v^2} \]

Writing \( P(n | n) \) recursively as follows

\[ P(1 | 1) = \frac{P(0 | 0)\sigma_v^2}{P(0 | 0) + \sigma_v^2} \]

\[ P(2 | 2) = \frac{P(1 | 1)\sigma_v^2}{P(1 | 1) + \sigma_v^2} = \frac{P(0 | 0)\sigma_v^2}{2P(0 | 0) + \sigma_v^2} \]

\[ P(3 | 3) = \frac{P(2 | 2)\sigma_v^2}{P(2 | 2) + \sigma_v^2} = \frac{P(0 | 0)\sigma_v^2}{3P(0 | 0) + \sigma_v^2} \]

\[ \vdots \]

we may infer a general expression for \( P(n | n) \) as follows

\[ P(n | n) = \frac{P(0 | 0)\sigma_v^2}{nP(0 | 0) + \sigma_v^2} \]

Since

\[ K(n) = \frac{P(n | n)}{\sigma_v^2} = \frac{P(0 | 0)}{nP(0 | 0) + \sigma_v^2} \]

then the Kalman filter for estimating the unknown DC voltage becomes

\[ \hat{x}(n | n) = \hat{x}(n-1 | n-1) + \frac{P(0)}{nP(0) + \sigma_v^2} \left[ y(n) - \hat{x}(n-1 | n-1) \right] \]

Note that \( K(n) \to 0 \) as \( n \to \infty \), and then \( \hat{x}(n | n) = \hat{x}(n-1 | n-1) \), which means that \( \hat{x}(n | n) \) approaches a steady state value.

If \( \sigma_v^2 \to \infty \) (which implies that the measurements are completely unreliable), then \( K(n) \to 0 \), and the estimate is

\[ \hat{x}(n | n) = \hat{x}(n-1 | n-1) \]

In this case, the measurements are ignored and \( \hat{x}(n | n) = \hat{x}(0 | 0) \), the initial estimate, which has an error variance \( P(0 | 0) \).

If \( \hat{x}(0 | 0) = 0 \) and \( P(0 | 0) \to \infty \) (which corresponds to the case of no a priori information about \( x(n) \)), then \( K(n) = 1 / n \), and the estimate becomes

\[ \hat{x}(n | n) = \hat{x}(n-1 | n-1) + \frac{1}{n} \left[ y(n) - \hat{x}(n-1 | n-1) \right] \]

\[ = \frac{n-1}{n} \hat{x}(n-1 | n-1) + \frac{1}{n} y(n) \]

which is simply a recursive implementation of the sample mean

\[ \hat{x}(n | n) = \frac{1}{n} \sum_{k=1}^{n} y(k) \]

Example 15. Kalman filter for estimating an AR(1) process

Consider that an AR(1) process

\[ x(n) = 0.8 x(n-1) + w(n) \]
(where \( w(n) \) is white noise with a variance \( \sigma_w^2 = 0.36 \) is measured in a noisy environment, as follows

\[
y(n) = x(n) + v(n)
\]

where \( v(n) \) is the measurement noise, a white noise process with zero mean and unit variance, and is uncorrelated with \( w(n) \). Find a Kalman filter that gives a best estimate of \( x(n) \) in the following manner,

\[
\hat{x}(n | n) = A(n - 1)\hat{x}(n - 1 | n - 1) + K(n)[y(n) - C(n)A(n - 1)\hat{x}(n - 1 | n - 1)]
\]

Solution.

To find the Kalman filter, we shall determine the Kalman gain \( K(n) \).

Form the given equations, it follows that \( A(n-1) = 0.8 \), \( B(n) = 1 \) and \( C(n) = 1 \), \( Q_v = \sigma_v^2 = 1 \), and \( Q_w = BB^T \sigma_w^2 = \sigma_w^2 = 0.36 \). Since the state vector is scalar, then the estimation equation becomes

\[
\hat{x}(n | n) = 0.8\hat{x}(n - 1 | n - 1) + K(n)[y(n) - 0.8\hat{x}(n - 1 | n - 1)]
\]

and we also have

\[
P(n | n - 1) = 0.8^2 P(n - 1 | n - 1) + 0.36
\]

\[
K(n) = P(n | n - 1)[P(n | n - 1) + 1]^{-1}
\]

\[
P(n | n) = [I - K(n)]P(n | n - 1)
\]

Inserting \( K(n) = P(n | n - 1)[P(n | n - 1) + 1]^{-1} \) into \( P(n | n) = [I - K(n)]P(n | n - 1) \) yields

\[
P(n | n) = \left[1 - \frac{P(n | n - 1)}{P(n | n - 1) + 1}\right]P(n | n - 1) = \frac{P(n | n - 1)}{P(n | n - 1) + 1} = K(n)
\]

Thus, \( P(n | n) = K(n) \). Noting that \( x(n) \) is a zero-mean process with \( r_x(k) = 0.8^{|k|} \), and assuming that \( \hat{x}(0 | 0) = E[x(0)] = 0 \) and \( P(0 | 0) = E[\hat{x}(0)] = r_x(0) = 1 \), the recursive calculations of the Kalman gain \( K(n) \) and the error covariances \( P(n | n - 1) \) and \( P(n | n) \) for the first few values are shown in the following table.

| \( n \) | \( K(n) \) | \( P(n | n - 1) \) | \( P(n | n) \) |
|---|---|---|---|
| 1 | 0.5000 | 1.0000 | 0.5000 |
| 2 | 0.4048 | 0.6800 | 0.4048 |
| 3 | 0.3824 | 0.6190 | 0.3824 |
| 4 | 0.3768 | 0.6047 | 0.3768 |
| 5 | 0.3755 | 0.6012 | 0.3755 |
| 6 | 0.3751 | 0.6003 | 0.3751 |
| ... | ... | ... | ... |
| \( \infty \) | 0.3750 | 0.6000 | 0.3750 |

From the table we can see that after a few iterations \( K(n) \), \( P(n | n - 1) \) and \( P(n | n) \) approach their steady-state values. Thus, the Kalman filter tends to its steady-state solution

\[
\hat{x}(n | n) = 0.8\hat{x}(n - 1 | n - 1) + 0.375[y(n) - 0.8\hat{x}(n - 1 | n - 1)]
\]

Re-writing the above equation in the following way,

\[
\hat{x}(n | n) = 0.5\hat{x}(n - 1 | n - 1) + 0.375y(n), \text{ or } \hat{x}(n | n) = \frac{0.375}{1 - 0.5q^{-1}} y(n) = H(q^{-1})y(n)
\]

and comparing it with the IIR causal Wiener filter in Example 5, we can see that both \( H(q^{-1})'s \) are identical. Substituting \( P(n | n - 1) = 0.8^2 P(n - 1 | n - 1) + 0.36 \) into \( K(n) = P(n | n - 1)[P(n | n - 1) + 1]^{-1} \), we have

\[
K(n) = \frac{0.8^2 P(n - 1 | n - 1) + 0.36}{0.8^2 P(n - 1 | n - 1) + 1} = \frac{0.64K(n-1) + 0.36}{0.64K(n-1) + 1.36}
\]

where \( P(n | n) = K(n) \) is used. Since \( K(n) \) approaches constant as \( n \to \infty \), then we may assume that \( \lim_{n \to \infty} K(n) = \lim_{n \to \infty} K(n - 1) = K \), which is called steady-state solution. The above equation becomes

\[
K = \frac{0.64K + 0.36}{0.64K + 1.36}, \text{ or } K^2 + 1.125K - 0.5625 = 0
\]

59
Solving the above equation for $K$, we may have $K = 0.375$ (another solution, $K = -1.5$, is unreasonable and thus, discarded); and $\lim_{n \to \infty} P(n \mid n-1) = 0.8^2 K + 0.36 = 0.8^2 \times 0.375 + 0.36 = 0.6000$. Because the steady-state solution of $K(n)$ for a stationary process is constant, then it is independent of initial value $P(0\mid0)$ ($=K(0)$). In other words, the initial value $P(0\mid0)$ does not affect the steady-state solution.

Example 15 shows that in the case of steady state ($n \to \infty$) the error covariance matrices $P(n \mid n-1)$ and $P(n \mid n)$ are different, that is, $\lim_{n \to \infty} P(n \mid n-1) \neq \lim_{n \to \infty} P(n \mid n)$. Therefore, if we assume $\lim_{n \to \infty} P(n \mid n-1) = P$, then we should assume $\lim_{n \to \infty} P(n \mid n) = P'$, because $P \neq P'$ in most cases.

**Steady state properties of Kalman filters**

As we have seen in Example 15, the steady state solution of the Kalman filter is identical to the Wiener filter. In other words, the Wiener filter is a special case of the Kalman filter in the steady state case. Here we are going to deal with the steady state properties of Kalman filters.

For a WSS process, e.g., an ARMA process $d(n)$ of the form $d(n) = H_\delta(q^{-1}) w(n)$, the state-space form is of the form

$$
x(n) = A x(n-1) + B w(n)
$$

$$
y(n) = C x(n) + v(n)
$$

where $A$, $B$, and $C$ are all constant matrices, and $w(n)$ and $v(n)$ have the variances of $\sigma_w^2$ and $\sigma_v^2$ with corresponding variance matrices $Q_w = E\{w(n)w^H(n)\} = BB^H \sigma_w^2$ and $Q_v = \sigma_v^2$, respectively.

**A. Steady-state Kalman gain, estimation error, and prediction error**

In this case, $K(n)$, $P(n \mid n)$ and $P(n \mid n-1)$ approach their steady state values for $n \to \infty$, namely, we may have $\lim_{n \to \infty} K(n) = K$, $\lim_{n \to \infty} P(n \mid n) = \lim_{n \to \infty} P(n \mid n-1) = P$, $\lim_{n \to \infty} P(n \mid n-1) = \lim_{n \to \infty} P(n-1 \mid n-2) = P'$. In general, $\lim_{n \to \infty} P(n \mid n) \neq \lim_{n \to \infty} P(n \mid n-1)$.

For given $A$, $B$, $C$, $Q_w$ and $Q_v$, the steady-state prediction error $\lim_{n \to \infty} P(n+1 \mid n) = \lim_{n \to \infty} P(n \mid n-1) = P'$ can be determined using the Riccati equation in the following manner,

$$
P' = AP'A^H + Q_w - AP'C^H \left[ C'P'C^H + Q_v \right]^{-1} C'P'A^H. \tag{259}
$$

When $A = I$, Eq. (259) is of a simpler form

$$
P'C^H \left[ C'P'C^H + Q_v \right]^{-1} C'P' + Q_w = 0. \tag{259'}
$$

When the prediction error $P'$ is found, the steady-state Kalman gain $\lim_{n \to \infty} K(n) = K$ can be determined by

$$
K = P'C^H \left[ C'P'C^H + Q_v \right]^{-1}. \tag{260}
$$

and the steady-state estimation error $\lim_{n \to \infty} P(n \mid n) = \lim_{n \to \infty} P(n \mid n-1) = P$ can be obtained with

$$
P = [I - KC]P'. \tag{261}
$$

**B. Steady-state system function for estimation**

From Eq. (232), it follows that

$$
\hat{x}(n \mid n) = A\hat{x}(n-1 \mid n-1) + K[y(n) - CA\hat{x}(n-1 \mid n-1)] = A\hat{x}(n-1 \mid n-1) - KCA\hat{x}(n-1 \mid n-1) + Ky(n),
$$

i.e.,
\[ \hat{x}(n \mid n) = [I - KC] \hat{x}(n-1 \mid n-1) + Ky(n) \]

The \( q \)-transforming of the above equation yields
\[ \{I - [I - KC]Aq^{-1}\} \hat{x}(n \mid n) = Ky(n) \]

and the system function for estimation is, thus, determined in the following way
\[ \hat{x}(n \mid n) = \{I - [I - KC]Aq^{-1}\}^{-1} Ky(n) = H(q^{-1})y(n) \]

where
\[ H(q^{-1}) = \{I - [I - KC]Aq^{-1}\}^{-1} K. \] (263)

C. Steady-state system function for prediction of \( x(n) \) and \( y(n) \)

One step forward shifting \( \hat{x}(n \mid n-1) = A\hat{x}(n-1 \mid n-1) \) yields \( \hat{x}(n+1 \mid n) = A\hat{x}(n \mid n) \), and from Eqs. (232) and (263), we may have the one-step prediction of \( x(n) \) as follows
\[ \hat{x}(n+1 \mid n) = A\hat{x}(n \mid n) = AH(q^{-1})y(n) \]

where
\[ H_{s}(q^{-1}) = AH(q^{-1}) = A\{I - [I - KC]Aq^{-1}\}^{-1} K. \] (265)

Alternatively, since \( \hat{x}(n+1 \mid n) = A\hat{x}(n \mid n-1) + AK[y(n) - C\hat{x}(n \mid n-1)] \) we have
\[ \hat{x}(n+1 \mid n) = A\hat{x}(n \mid n-1) - AKC\hat{x}(n \mid n-1) + AKy(n) \]
i.e.,
\[ \hat{x}(n+1 \mid n) = AK\hat{x}(n \mid n-1) = AKy(n) \]

The \( q \)-transform of the above equation is
\[ \{I - A[I - KC]q^{-1}\} \hat{x}(n+1 \mid n) = AKy(n) \]

and the system function becomes
\[ \hat{x}(n+1 \mid n) = \{I - A[I - KC]q^{-1}\}^{-1} AKy(n) = H_{s}(q^{-1})y(n) \]

where
\[ H_{s}(q^{-1}) = \{I - A[I - KC]q^{-1}\}^{-1} AK. \] (267)

The best \( m \)-step prediction of \( x(n) \) is
\[ \hat{x}(n+m \mid n) = \left[ \prod_{k=0}^{m-1} A \right] \hat{x}(n \mid n) \]

(268)

The best \( m \)-step prediction of \( y(n) \) is
\[ \hat{y}(n+m \mid n) = C\hat{x}(n+m \mid n) = C \left[ \prod_{k=0}^{m-1} A \right] \hat{x}(n \mid n) \]

(269)


Consider a random process \( x(n) \) of the form,
\[ x(n) = ax(n-1) + w(n) \]

where \( w(n) \) is zero-mean white noise with a variance \( \sigma^2_w = 1 \) and \( \mid a \mid < 1 \), and assume
\[ y(n) = x(n) + v(n) \]

to be noisy measurement of \( x(n) \) where \( v(n) \) is zero-mean white noise with a unit variance and uncorrelated with \( w(n) \).
(a) Find the steady-state Kalman filter gain $K$, i.e., $K=\lim_{n\to\infty}K(n)$, in terms of $a$;

(b) Show that the causal IIR Wiener filter for one-step prediction satisfies the steady-state solution of the Kalman filter as follows

$$\hat{x}(n+1\mid n) = a\hat{x}(n\mid n-1) + aK[y(n) - \hat{x}(n\mid n-1)].$$

Note that $\hat{y}(n+1\mid n) = C(n+1)\hat{x}(n+1\mid n)$.

Solution.

(a) From given state-space form, it follows that $A=a$, $C=1$, $Q_w=\sigma_w^2=1$, and $Q_v=\sigma_v^2=1$.

The Riccati equation

$$P = AP^T A + Q_w - AP^T C R\left(CP^T C + Q_v\right)^{-1} CP^T A$$

becomes the scalar form,

$$P = a^2 P + 1 - a^2 \frac{P^2}{P+1}$$

$$P^2 - a^2 P - 1 = 0$$

Solving the above equation for $P$ and considering $P\geq0$, we take the positive solution as follows

$$P = \frac{a^2 + \sqrt{a^4 + 4}}{2}$$

The steady-state Kalman gain can be found by

$$K = \frac{P^T}{P+1} = \frac{a^2 + \sqrt{a^4 + 4}}{a^2 + \sqrt{a^4 + 4} + 2}$$

(b) Inserting $x(n) = ax(n-1) + w(n)$ into $y(n) = x(n) + v(n)$ and taking the $q$-transform, we have

$$y(n) = \frac{w(n)}{1-aq^{-1}} + v(n)$$

The power spectrum of $y(n)$ is

$$P_y(z) = \frac{1}{(1-aq^{-1})(1-az)} + 1 = \frac{2 + a^2 - a(z^{-1} + z)}{(1-aq^{-1})(1-az)}$$

Spectrally factoring $P_y(z)$ yields

$$P_y(z) = \sigma_y^2 \frac{(1-dz^{-1})(1-dz)}{(1-az^{-1})(1-az)}$$

where

$$d = 2 + a^2 - \sqrt{a^4 + 4} \quad \text{(taking into account $d<1$), and} \quad \sigma_y^2 = \frac{a}{d} = \frac{2}{2 + a^2 - \sqrt{a^4 + 4}}$$

Thus, we may write

$$y(n) = \frac{1 - dq^{-1}}{1-aq^{-1}} \epsilon(n)$$

From Eq. (166) in Optimum Filters (4) (or Example 6 in Optimum Filters (2)) we can find the one-step prediction of $y(n)$,

$$\hat{y}(n+1\mid n) = \frac{d + a}{1-aq^{-1}} y(n)$$

From Eq. (269), it follows that $\hat{y}(n+1\mid n) = C\hat{x}(n+1\mid n) = \hat{x}(n+1\mid n)$. Thus, the above equation can be expressed as
\[
\hat{x}(n+1 | n) - a \hat{x}(n | n - 1) = (a - d) y(n) \\
\hat{x}(n+1 | n) = a \hat{x}(n | n - 1) + (a - d) [y(n) - \hat{x}(n | n - 1)]
\]

Since
\[
a - d = \frac{a^2 - 2 + \sqrt{a^4 + 4}}{2a^2} = \frac{a^2 + \sqrt{a^4 + 4} - 2}{2a^2} = \frac{a^2 + \sqrt{a^4 + 4} - 2}{2a^2 (a^2 + \sqrt{a^4 + 4} + 2)} = \frac{a^2 + \sqrt{a^4 + 4} - 2}{a^2 + \sqrt{a^4 + 4} + 2} = K
\]
then we have
\[
\hat{x}(n+1 | n) = a \hat{x}(n | n - 1) + aK [y(n) - \hat{x}(n | n - 1)]
\]

**Example 17. Steady-state solution of a Kalman filter.**

Consider a random process \(x(n)\) of the form,
\[
x(n) = ax(n-1) + w(n)
\]
where \(w(n)\) is zero-mean white noise with a variance \(\sigma^2_w = 1\) and \(|a| < 1\), and assume
\[
y(n) = x(n) + v(n)
\]
to be noisy measurement of \(x(n)\) where \(v(n)\) is zero-mean white noise with a unit variance. Assuming that the steady-state Kalman gain \(\lim_{n \to \infty} K(n) = K\) is a known constant,

(a) find the steady-state system function \(H(q^{-1})\) of the Kalman filter for the optimal two-step prediction, \(\hat{x}(n+2 | n)\), of \(x(n)\).

(b) Show that the steady-state Kalman gain \(K\) satisfies \(0 \leq K < 1\) and the Kalman filter \(H(q^{-1})\) in (a) is stable.

**Solution.**

(a) From given state-space form, it follows that \(A = a\), \(C = 1\), \(B = 1\), \(Q_w = BB^H \sigma^2_v = 1\), and \(Q_v = \sigma^2_v = 1\).

Since the steady-state Kalman gain \(\lim_{n \to \infty} K(n) = K\) is known and \(x(n)\) is scalar, the one-step prediction
\[
\hat{x}(n+1 | n) = A \hat{x}(n | n - 1) + AK [y(n) - C \hat{x}(n | n - 1)]
\]
in this case becomes
\[
\hat{x}(n+1 | n) = a \hat{x}(n | n - 1) + aK [y(n) - \hat{x}(n | n - 1)].
\]
The \(q\)-transform of the above equation gives
\[
\hat{x}(n+1 | n) = \frac{aK}{1 - a(1 - K)q^{-1}} y(n)
\]
One-step forward shifting of \(\hat{x}(n | n - 1) = A(n-1) \hat{x}(n-1 | n - 1)\) yields \(\hat{x}(n+1 | n) = A(n) \hat{x}(n | n)\). Moreover, since the \(m\)-step prediction of \(x(n)\) is \(\hat{x}(n+m | n) = \prod_{k=0}^{m-1} A \hat{x}(n | n)\), we have the two-step prediction of \(x(n)\)
\[
\hat{x}(n+2 | n) = AA \hat{x}(n | n) = A \hat{x}(n+1 | n) = \frac{a^2 K}{1 - a(1 - K)q^{-1}} y(n) = H(q^{-1}) y(n)
\]
where \(H(q^{-1}) = \frac{a^2 K}{1 - a(1 - K)q^{-1}}\) is the desired system function for the optimal two-step prediction.

(b) Since \(K = P' c^H [C P' c^H + Q_v]^{-1} = P' / (P' + 1)\) and \(P' > 0\), then \(0 \leq K < 1\). As \(|a| < 1\) and \(0 \leq K < 1\), we have \(|a(1 - K)| \leq |a| < 1\) so that the pole \(q = a(1 - K)\) is inside the unit circle, and thus \(H(q^{-1})\) is stable.
Appendix 7B. Derivation of matrix differentiation formulas

In this appendix we will derive the matrix differentiation formulas in Eqs. (239) and (240). Here we consider the real valued matrices so that
\[
\frac{d}{dK} \text{tr}\{KC\} = C^T \quad \text{and} \quad \frac{d}{dK} \text{tr}\{KCK^T\} = 2KC
\]

For the complex-valued matrices we may replace the transpose with the conjugate transpose.

Assuming a 3x2 valued matrix \( K = \begin{bmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \\ k_{31} & k_{32} \end{bmatrix} \) and a 2x3 matrix \( C = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \end{bmatrix} \), then we have
\[
KC = \begin{bmatrix} k_{11}c_{11} + k_{12}c_{21} & k_{21}c_{12} + k_{22}c_{22} \\ k_{31}c_{11} + k_{32}c_{21} \\ k_{31}c_{12} + k_{32}c_{22} \end{bmatrix}, \quad \text{and} \quad \text{tr}\{KC\} = k_{11}c_{11} + k_{12}c_{21} + k_{21}c_{12} + k_{22}c_{22} + k_{31}c_{13} + k_{32}c_{23} = s
\]

The derivative of \( KC \) with respect to \( K \) is
\[
\frac{d}{dK} \text{tr}\{KC\} = \frac{d}{dK} \begin{bmatrix} ds \\ ds \\ ds \end{bmatrix} = \begin{bmatrix} \frac{dC_{11}}{dk_{11}} & \frac{dC_{12}}{dk_{12}} \\ \frac{dC_{21}}{dk_{21}} & \frac{dC_{22}}{dk_{22}} \\ \frac{dC_{31}}{dk_{31}} & \frac{dC_{32}}{dk_{32}} \end{bmatrix} = C^T \Rightarrow \frac{d}{dK} \text{tr}\{KC\} = C^T
\]

Assuming a 3x2 matrix \( K = \begin{bmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \\ k_{31} & k_{32} \end{bmatrix} \) and a 2x2 matrix \( C = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \), and \( K^T = \begin{bmatrix} k_{11} & k_{12} & k_{31} \\ k_{12} & k_{22} & k_{32} \end{bmatrix} \), then we have
\[
KCK^T = \begin{bmatrix} k_{11}c_{11} + k_{12}c_{12} & k_{11}c_{12} + k_{12}c_{22} \\ k_{21}c_{11} + k_{22}c_{12} & k_{21}c_{12} + k_{22}c_{22} \\ k_{31}c_{11} + k_{32}c_{12} & k_{31}c_{12} + k_{32}c_{22} \end{bmatrix} K^T
\]

\[
= \begin{bmatrix} k_{11}c_{11} + k_{12}c_{21} \\ k_{12}c_{11} + k_{12}c_{22} \\ k_{31}c_{11} + k_{32}c_{21} + k_{21}c_{12} + k_{22}c_{22} \\ k_{21}c_{12} + k_{22}c_{21} + k_{31}c_{12} + k_{32}c_{22} + k_{32}c_{22} \end{bmatrix}
\]

and
\[
\text{tr}\{KCK^T\} = s = k_{11}c_{11} + k_{12}c_{21} + k_{12}c_{12} + k_{12}c_{22} + k_{21}c_{12} + k_{21}c_{22} + k_{22}c_{21} + k_{22}c_{22} + k_{31}c_{11} + k_{31}c_{12} + k_{32}c_{12} + k_{32}c_{22} + k_{32}c_{22}
\]

Thus \( \frac{d}{dK} \text{tr}\{KCK^T\} = \frac{d}{dK} \begin{bmatrix} ds \\ ds \\ ds \end{bmatrix} = 2 \begin{bmatrix} k_{11}c_{11} + k_{12}c_{21} & k_{11}c_{12} + k_{12}c_{22} \\ k_{21}c_{11} + k_{22}c_{21} & k_{21}c_{12} + k_{22}c_{22} \\ k_{31}c_{11} + k_{32}c_{22} & k_{31}c_{12} + k_{32}c_{22} \end{bmatrix} = 2KC
\]