Optimization of Linear Observations for the Stationary Kalman Filter based on a Generalized Water Filling Theorem

Yoshiki Takeuchi
Dept. of Information Science, Osaka Univ. of Education
4-698-1 Asahigaoka, Kashiwara City, Osaka 582-8582, Japan
Tel. +81-729-78-3669, Fax. +81-729-78-3554
e-mail: takeuti@cc.osaka-kyoiku.ac.jp

Abstract

We are concerned with a problem of the optimal selection of the gain matrix of a linear observation for the Kalman filter. The innovations process included in the Kalman filter has the same structure as the model of a set of parallel transmission channels with the optimal output feedback. In the linear coding problem for this set of channels, it is well-known that the optimal output feedback which minimizes the power of the encoded signal is given by the least-squares estimate of the linear term and that the channel output becomes the innovations process. By applying a solution of the optimal transmission problem for this model, we obtain a set of gains which maximizes the mutual information between the observation and the signal under a constraint on the power of the innovations process.

1 Introduction

Starting from 1970's, the problem of optimization of observations associated with the Kalman filter [1] has been studied in many literatures [2]-[4], [6] , [7]. Most of them formulated the problem as a kind of optimal control problem with a quadratic performance criterion. However, because of the existence of the nonlinearity in the Riccati equation, there are few result which is applicable in the real design of the Kalman filter. Logothetis et. al [3], discussed a sensor scheduling problem for linear observations via information theoretic criteria. They treated a discrete-time linear system, and under the constraint that only one sensor can be used at any time, they derived equations for the optimal sensor gains which maximize the mutual information between the signal and the observation. A similar kind of problem is that of sensor allocation. Namely, for a number of sensors, which are simultaneously available at one time, we want to determine a set of gains for the sensors in such a way that the estimation error is minimized. According to the Shannon's information theory, the observation is considered to be better when the mutual information between the signal and the observation takes a larger value.

In this paper, we are concerned with discrete-time linear systems and observations for the Kalman filter. Based on a generalized Water Filling Theorem [10] which is the solution of the optimal transmission problem through parallel channels with feedback, we obtain a set of gains which maximize the mutual information under a constraint on the power of the innovations process. We will show that the gain which maximizes the mutual information is not unique. The optimal gain is selected from this set of gains in such a way that the steady-state estimation error is minimized.

Mathematical symbols, in this paper, are used in the following way. \( R \) is the space of all real numbers, i.e., \( R = (-\infty, \infty) \). For positive integers \( m \) and \( n \), \( R^m \) and \( R^{n \times n} \) denote the spaces of \( n \)-dimensional vectors and \( m \times n \)-dimensional matrices whose components take values in \( R \). The prime denotes the transpose of a vector or a matrix and the Euclidean norm is \( \| \cdot \| \). Thus, for \( x \in R^n \), \( |x| = \sqrt{x^T x} \). The identity matrix of any dimension is denoted by \( I \) and \( \text{rank}[A] \) is the rank of matrix \( A \) of any dimension. The components of a matrix are denoted by using subscripts. Thus, \( [A]_{ij} \) is the \((i,j)\)-component of \( A \). In the case where no confusion may arise, we denote \( [A]_{ij} \) simply by \( a_{ij} \). If \( A \) is a square matrix, \( \det[A] \) and \( \text{tr}[A] \) respectively denote the determinant and the trace of \( A \). By \( \text{diag}(a_1, a_2, \ldots, a_n) \), we denote a diagonal matrix with the diagonal components \( a_1, a_2, \ldots, a_n \). We use \( A > 0 \) and \( A \geq 0 \) to denote that \( A \) is positive definite and non-negative definite, respectively. The triplet \( (\Omega, \mathcal{F}, P) \) is a complete probability space where \( \Omega \) is a sample space with elementary events \( \omega \), \( \mathcal{F} \) is a \( \sigma \)-field of subsets of \( \Omega \), and \( P \) is a probability measure. \( E\{ \cdot \} \) denotes the expectation and \( E\{\cdot|\mathcal{G}\} \) the conditional expectation, given \( \mathcal{G} \), with respect to \( P \). \( \sigma\{ \cdot \} \) is the minimal sub-\( \sigma \)-field of \( \mathcal{F} \) with respect to which the family of \( \mathcal{F} \)-measurable sets or random variables \( \{ \cdot \} \) is measurable. It is assumed that all random variables and stochastic processes are \( \mathcal{F} \)-measurable. Unless otherwise stated, stochastic properties are that with respect to \( P \).
2 Problem Formulation and Preliminaries

2.1 Kalman Filter and the Basic Equations

Let \( x(t) = \{x(t); t = 0, 1, \cdots \} \) denote an \( n \)-dimensional Gaussian stochastic process described by

\[
x_{t+1}(\omega) = A x_t(\omega) + w_t(\omega), \quad x_0(\omega) = x^0(\omega),
\]

where \( A \in \mathbb{R}^{n \times n} \), \( w_t(\omega) \) is a Gaussian random vector with mean \( 0 \) and covariance \( Q_t \), and \( w = \{w_t(\omega); t = 0, 1, \cdots \} \) is a \( d_1 \)-dimensional standard white Gaussian noise sequence. Suppose that the value of \( x \) is not available but we have an \( m \)-dimensional observation described by

\[
y_t(\omega) = H x_t(\omega) + v_t(\omega),
\]

where \( y = \{y_t(\omega); t = 1, 2, \cdots \} \) is an \( m \)-dimensional observation process, \( H \in \mathbb{R}^{m \times n} \), \( R_t \) is a \( d_2 \)-dimensional standard white Gaussian noise sequence. We will assume that \( x^0(\omega), v_t(\omega) \) and \( v_t(\omega) \) are mutually independent.

It is well-known that the least-squares estimate

\[
\hat{x}_{\text{ls}}(\omega) \triangleq \mathbb{E}\{x_t(\omega)|\mathcal{F}_t\}
\]

of \( x_t(\omega) \) based on \( \mathcal{F}_t = \sigma(y_s(\omega); s = 1, 2, \cdots, t) \) is given by the Kalman filter:

\[
\begin{align*}
\dot{x}_{\text{ls}}(\omega) &= A \hat{x}_{\text{ls}}(\omega) + Q_{\text{ls}} H' R_t^{-1} y_t(\omega), \\
Q_{\text{ls}}(t) &= Q_{\text{ls}}(t-1) - Q_{\text{ls}}(t-1) H'(H Q_{\text{ls}}(t-1) H' + R_t)^{-1} H Q_{\text{ls}}(t-1),
\end{align*}
\]

and

\[
\hat{x}_{\text{ls}}(\omega) \triangleq \mathbb{E}\{x_t(\omega)|\mathcal{F}_t\},
\]

shown in Fig. 1. Here, we will also assume that \( x = \{x_t(\omega); t = 0, 1, \cdots \} \) which denotes the signal is an \( n \)-dimensional Gaussian stochastic process given by (1).

2.2 The Innovations Process and its Relation to the Optimal Transmission Problem

From (8), we notice that the innovations process \( \bar{y} \) is described by the same equation which was used in the discussion of the optimal transmission problem of Gaussian signals through parallel channels with feedback [8, 9].

Now, let us consider the optimal transmission problem for a set of transmission channels with the construction

\[
\begin{align*}
\text{Additive} \\
\text{Noise} \quad R_v \\
\text{Output}
\end{align*}
\]

Fig. 1 Transmission channels with feedback.

When we send \( x \) through the set of \( m \)-parallel channels, we employ an encoding scheme. Namely, we send an encoded signal \( \beta = \{\beta_s(x, y); s = 1, 2, \cdots \} \) which is a function of both signal \( x \) and output \( y = \{y_t(\omega); t = 1, 2, \cdots \} \). Thus, the output of the transmission channels in Fig. 1 is described by

\[
\beta_t(\omega) = \beta_s(x, y) + R v_t(\omega).
\]

An optimal transmission problem, or an optimal coding problem, is the one to determine \( \beta \) optimally. In Takeuchi [8, 9], we were concerned with a linear coding problem where \( \beta \) is given by

\[
\beta_t(x, y) = H x_t(\omega) - \phi(t, y).
\]

Since \( \phi(t, y) \) must be a non-anticipative function of \( y \), we have

\[
\phi(t, y) = \phi(t, y_1, \cdots, y_{t-1}),
\]

namely, \( \phi(t, y) \) should be a \( \mathcal{F}_{t-1} \)-measurable random vector, where

\[
\mathcal{F}_t = \sigma(y_s(\omega); s = 1, 2, \cdots, t).
\]

As it is well-known, \( \phi(t, y) \) does not affect the information given by \( y \).

**Theorem 1** Assume that \( R_0 R' > 0 \). Then, we have the informational equivalence between \( y \) and \( y \), i.e., we have

\[
\mathcal{F}_t = \mathcal{G}_t, \quad t = 1, 2, \cdots.
\]

Because of (13), \( \phi(t, y) \) is usually taken to minimize the transmission power.

**Theorem 2** The optimal feedback signal which minimizes

\[
\Delta P_t \triangleq \mathbb{E}\{[H x_t(\omega) - \phi(t, y)]^2\},
\]

shown in Fig. 1. Here, we will also assume that \( x = \{x_t(\omega); t = 0, 1, \cdots \} \) which denotes the signal is an \( n \)-dimensional Gaussian stochastic process given by (1).
is given by
\[ \phi(t, \tilde{y}) = H \tilde{x}_{E_{t+1}}(\omega), \quad t = 1, 2, \ldots \] (15)

The proofs of the above theorems are trivial and can be found in [9].

Now, substituting (10) and (15) into (9), we have (8) as the equation for the output of the parallel channels. Thus, if we determine \( \phi(t, \tilde{y}) \) by (15), then we have the innovations process as the output of the parallel channels.

### 2.3 Optimization of \( H \) from an Information Theoretic Point of View

In what follows, we are concerned with the stationary case where the error covariance matrices have the properties: \( Q_0 \rightarrow Q \) and \( Q_{E_{t+1}} \rightarrow Q \) as \( t \rightarrow \infty \). As it is well-known, the steady-state covariance matrices \( Q \) and \( Q^- \) are determined by
\[ Q = Q^- - Q^- H'(HQ^- H' + R_0)^{-1}HQ^-, \] (16)
and
\[ Q^- = AQA' + GG'. \] (17)

Therefore, when (15) is used, we have
\[ \Delta P \rightarrow \bar{P} \triangleq \text{tr}[HQ'H'], \quad t \rightarrow \infty. \] (18)

In this paper, the optimization of \( H \) is done by the following two steps.

(i) Find a set \( \mathcal{H} \) of the values of \( H \) each of which maximizes the mutual information between \( x \) and \( y \) subject to the constraint: \( \bar{P} \leq p \).

(ii) Find \( H \in \mathcal{H} \) which minimizes \( \text{tr}[Q] \).

As for (i), the problem within the optimal transmission framework is described as follows:

**Problem 1 (Optimal Selection of \( H \))**

Find \( H \) such that
\[ \bar{T}(x, y) = \frac{1}{2} \log \{|R_0^{-1/2}HQ'H'R_0^{-1/2} + I|\} \rightarrow \max \] subject to
\[ \bar{P} \triangleq \text{tr}[HQ'H'] \leq p, \] (19)
(16) and (17). \( \square \)

In (19) and (20), \( \bar{P} \) and \( \bar{T}(x, y) \) are the transmission power and the mutual information between \( x \) and \( y \) (or \( \tilde{y} \)) per unit time, respectively, and \( p \) is the maximum admissible value of the former.

### 3 Optimal Selection of \( H \) by way of a Solution of the Optimal Transmission Problem

As we explained in the previous section, the optimization of \( H \) consists of two steps. For the first step, i.e., for Problem 1, the condition of the optimality is given by the following theorem.

**Theorem 3 (A generalized Water Filling Theorem)**

Assume that \( GG' > 0 \) and \( R_0 > 0 \). Let \( (\psi_1, \psi_2, \ldots, \psi_m) \) and \( (\gamma_1, \gamma_2, \ldots, \gamma_m) \) denote the set of eigen-values of \( R_0 \) and the corresponding set of eigen-vectors, respectively. Also, let \( \Psi \triangleq \text{diag}(\psi_1, \psi_2, \ldots, \psi_m) \) and \( \Gamma \triangleq [\gamma_1, \gamma_2, \ldots, \gamma_m] \), i.e.,
\[ R_0 = \Gamma \Psi \Gamma^T. \] (21)

Without loss of generality, we can assume that
\[ 0 < \psi_1 \leq \psi_2 \leq \cdots \leq \psi_m. \] (22)

Then, the condition of optimality of \( H \in \mathbb{R}^{m \times n} \) is given by
\[ HQ'H' = \Gamma \Xi \Gamma^T, \] (23)
where
\[ \Xi \triangleq \text{diag}(\xi_1, \xi_2, \ldots, \xi_m), \] (24)
\[ \xi_i \triangleq \max \{0, \alpha - \psi_i\}, \quad i \in \{1, 2, \ldots, m\} \] (25)
\[ \alpha \triangleq \frac{p + \sum_{i=1}^{m} \psi_i}{\tilde{m}}, \] (26)
and \( \tilde{m} \triangleq \text{rank}[\Xi] = \{\text{Number of positive } \xi_i's\}. \) (27)

Furthermore, for the optimal value of \( H \), we have
\[ HQ'H' = \Gamma \Xi(\Xi + \Psi)^{-1}\Psi \Gamma^T, \] (28)
and
\[ \bar{T}(x, y) = \frac{1}{2} \log \{|\Xi(\Xi + \Psi)^{-1}\Psi + I|\}. \] (29)

**Remark 1** It should be noted that \( HQ'H' \) and \( R_0 \) have the same set of eigen-vectors and the sum of the pair of eigen-values \( \xi_i \) and \( \psi_i \) corresponding to each eigen-vector \( \gamma_i \) is equal to \( \alpha \) if \( \psi_i < \alpha \). Otherwise, i.e., if \( \psi_i \geq \alpha \), we have \( \xi_i = 0 \). This is the same result as the well-known Water Filling Theorem[10] except for the point that our result is concerned with the eigen-values of the signal and noise covariance matrices whereas the usual case is with the signal and noise powers for channels.

**Remark 2** Due to (22), (25) implies that
\[ \xi_1 \geq \xi_2 \geq \cdots \geq \xi_m > \xi_{m+1} = \cdots = \xi_{2m} = 0. \] \( \square \)

**Remark 3** The maximum value of the mutual information given by (29) only depends on \( \Psi \) and \( p \) because \( \alpha \) and \( \Xi \) are determined by them. Thus, the mutual information is independent of the stochastic property of the signal except for the condition that the mean squares power of \( H^T(\tilde{x}_E(o) - \tilde{x}_{E_{t+1}}(o)) \) is \( p \).

**Remark 4** In order to determine the values of \( \alpha \) and
\( m \), we can use a simple recursive algorithm.

**Remark 5** It should be noted that \( \Xi \), \( \Psi \) and \( \{ \Psi + \Xi \} \) in (28) are all diagonal matrices, and that they are commutative in matrix multiplications.

By Theorem 3, we have the following solution for Problem 1.

**Theorem 4** Assume that \( GG' > 0 \) and \( R_0 > 0 \). Let
\[ \Xi = \text{diag}(\Xi, 0), \quad \tilde{\Xi} = \text{diag}(\xi_1, \xi_2, \ldots, \xi_n) > 0, \quad (30) \]
and
\[ \Gamma = \left[ \begin{array}{c} \tilde{\Gamma} \\ \Gamma \end{array} \right], \quad \tilde{\Gamma} \in R^{n \times \tilde{n}}, \quad \Gamma \in R^{n \times (n - \tilde{n})}. \quad (31) \]
Let \( S' \) denote a solution of the quadratic matrix equation:
\[ S^2 = AKS A' + GG', \quad (32) \]
where
\[ \tilde{K} = I - \frac{1}{\alpha} \tilde{U} \tilde{\Xi} \tilde{U}', \quad (33) \]
and where \( \tilde{U} \in R^{n \times \tilde{n}} \) is any matrix with the property: \( \tilde{U}' \tilde{U} = I \). Then, the solution of Problem 1 is given by the following set of equations:
\[ Q = S^2, \quad (34) \]
\[ Q = SKS, \quad (35) \]
\[ H = \tilde{\Xi}^{1/2} \tilde{U} S^{-1}. \quad (36) \]

**Remark 6** From (34)-(36), it is clear that any transmission schemes with \( H \) given by (36) and the same value of \( \tilde{K} \) produce the same filtering performance. Thus, the performance of the Kalman filter under (36) depends on \( \Psi' \), \( \beta \) and \( \tilde{U} \) other than \( A \) and \( G \), and is independent of \( \Gamma \).

**Remark 7** It should be noted that if \( A \in R^{m \times m} \) is a stable matrix, then we always have a positive definite solution of (32).

By (36), it is seen that \( H \) is not unique but we can take any \( \tilde{U} \) with property: \( \tilde{U}' \tilde{U} = I \). Namely,
\[ \mathcal{H} \triangleq \{ H \in R^{m \times m}; H = \tilde{\Xi}^{1/2} \tilde{U} S^{-1}, \tilde{U}' \tilde{U} = I \}, \]
is the set of solutions of Problem 1. From the view point of the performance of the Kalman filter, it is desirable to select \( \tilde{U} \) in the way that \( \text{tr}(Q) \) is minimized. The solution of this step, (ii), of the problem is given by the following theorem.

**Theorem 5** Assume that \( A \) is a stable matrix. The optimal selection of \( H \in \mathcal{H} \) for (ii) is given by a solution \((S, \tilde{U}, K, X, \Lambda)\) of the set of equations: (32), (33),
\[ S X S \tilde{U} = \tilde{U} \Lambda \tilde{\Xi} S^{-1}, \quad (37) \]
and
\[ -X + A' X A = -\{I - S^{-1} \tilde{U} \Lambda \tilde{U}' S^{-1}\}, \quad (38) \]
where \( X \in R^{n \times n}, \quad \Psi = \text{diag}(\tilde{\Psi}, \tilde{\Psi}), \quad \tilde{\Psi} \in R^{(m - \tilde{n}) \times (m - \tilde{n})} \), and \( \Lambda \in R^{(m - \tilde{n}) \times (m - \tilde{n})} \), and \( \Lambda \) is a diagonal matrix.

### 4 Proofs of Theorems

In this section, we will give proofs of the theorems presented in the previous section. For the proof of Theorem 3, we use the following lemma by which we can reduce the problem to the simpler one whose solution is referred to as the Water Filling Theorem [10].

**Lemma 1** Let \( F \in R^{m \times m} \) denote an orthogonal matrix whose columns are eigenvectors of \( HQH' \), i.e.,
\[ HQH' = F \Xi F', \quad FF' = FF' = I. \quad (39) \]
Then, the mutual information
\[ I(x, y) = \frac{1}{2} \log \left\{ \frac{\det |F\Xi F' + R_0|}{\det |R_0|} \right\} \quad (40) \]
takes the maximum value when \( F \) is the set of eigenvectors of \( R_0 = \Gamma \Psi' \Gamma' \), i.e.,
\[ F = \Gamma. \quad (41) \]

**(Proof)** Since
\[ \text{tr}[HQH'] = \text{tr}[F\Xi F'] = \text{tr}[\Xi FF'] = \text{tr}[\Xi], \quad (42) \]
the transmission power does not depend on \( F \). Hence, according to (39) and (40), let us define Lagrangean by
\[ L(F, \Lambda') = \log \left\{ \det |F\Xi F' + R_0| \right\} - \text{tr}[\Lambda'(FF' - I)], \quad (43) \]
where \( \Lambda' \in R^{m \times m} \) denotes a set of Lagrange multipliers for the components of \( (FF' - I) \), and we have used the relation
\[ \text{tr}[\Lambda'(FF' - I)] = \sum_{i=1}^{m} \sum_{j=1}^{m} \lambda_i' [FF' - I]_{ij}. \quad (44) \]
Note here that since \( (FF' - I) \) is symmetric, \( \Lambda' \) must be also symmetric.

It can be easily seen that the condition of the optimality is given by
\[ \frac{\partial L(F, \Lambda')}{\partial F} = 2(F\Xi F' + R_0)^{-1} F\Xi - 2FA' = 0. \quad (45) \]
Note that we can take \( \Lambda' \) as a diagonal matrix because (45) implies
\[ \Lambda' = F'(F\Xi F' + R_0)^{-1} F\Xi, \quad (46) \]
and the right-hand side is a product of symmetric and diagonal matrices whereas the left-hand side is a symmetric matrix. Now, according to (31), let
\[ F = \tilde{F}, \quad \tilde{F} \in R^{m \times \tilde{n}}, \quad \tilde{F} \in R^{(m - \tilde{n}) \times (m - \tilde{n})}. \quad (47) \]
We assume that \( \tilde{m} \leq m \) including the case \( \tilde{m} = m \).
Then, (45) implies
\[
(F \Xi F' + R_0)^{-1} \begin{bmatrix} \tilde{F} \tilde{X} \\ 0 \end{bmatrix} = \begin{bmatrix} \tilde{F} \\ 0 \end{bmatrix}
\]
from which we have
\[
(F \Xi F' + R_0)^{-1} \tilde{F} \tilde{X} = \tilde{F} \Lambda^*,
\] (49)
and \(\Lambda^* = 0\). Since \(\tilde{X} > 0\) and \(\text{rank}[\tilde{F}] = \tilde{m}\), we have \(\text{rank}[\Lambda^*] = \tilde{m}\). Hence, from (46), we have \(\Lambda^* > 0\).

Also, we have
\[
I - \Lambda^* = I - (\Xi + F'R_0F)^{-1} \Xi = (\Xi + F'R_0F)^{-1} F'R_0F > 0.
\] (50)
It can be easily seen that (49) implies
\[
R_0 \tilde{F} = \tilde{F} \Xi (I - \Lambda^*) \Lambda^{*-1}.
\] (51)
Thus, all column vectors of \(\tilde{F}\) are the eigenvectors of \(R_0\) because \(\tilde{X}(I - \Lambda^*) \Lambda^{*-1}\) is a positive definite diagonal matrix. That is, these are the \(\tilde{F}\) component vectors of \(F\). We can take the rest of components of \(F\), i.e., \(\tilde{F}\) to be an arbitrary matrix with properties: \(\tilde{F} \tilde{F}' = I\) and \(\tilde{F} \tilde{F}' = 0\).

(Proof of Theorem 3) By (39), (40) and (41) in Lemma 1, the problem is now reduced to be finding \(\Xi = \text{diag}(\xi_1, \xi_2, \cdots, \xi_m)\) such that
\[
\tilde{I}(x, y) = \frac{1}{2} \log \left( \frac{\det[\Xi + \Psi]}{\det[\Psi]} \right) \rightarrow \max.,
\] (52)
subject to
\[
\tilde{F} = \text{tr}[\Xi] \leq \rho,
\] (53)
and
\[
\Xi \geq 0, \quad \text{i.e.,} \quad \xi_i \geq 0, \quad i = 1, 2, \cdots, m,
\] (54)
where (54) must be satisfied since \(HQ' H' \geq 0\). This is the problem whose solution is known to be given by (25)-(27) and referred to as the Water Filling Theorem[10]. Thus, we have shown that (23)-(27) hold. Then, we have (29) from (52). Finally, (28) follows from (23) and (16) as
\[
HQ' = HQ' H' - HQ' H' (HQ' H' + R_0)^{-1} HQ' H'
\]
\[
= \Gamma \Xi \Gamma' - \Gamma \Xi (\Gamma \Xi + \Gamma \Psi')^{-1} \Gamma \Xi 
\]
\[
= \Gamma [\Xi - \Xi (\Xi + \Psi')^{-1} \Xi] \Gamma'
\]
\[
= \Gamma [\Xi (\Xi + \Psi')^{-1} (\Xi + \Psi') - \Xi (\Xi + \Psi')^{-1} \Xi] \Gamma'
\]
\[
= \Gamma (\Xi + \Psi')^{-1} \Xi 
\].
(55)
This completes the proof.

Now, let us proceed to the proof of Theorem 4. The proof is based on the following lemma.

[Lemma 2] For \(S'\) given by (34), let
\[
\tilde{U} \triangleq SH' \tilde{\Xi}^{-1/2},
\] (56)
and
\[
Y \triangleq SH' \tilde{\Gamma}.
\] (57)
Then, we have
(i) \(\tilde{U}' \tilde{U} = I\),
and
(ii) \(Y = 0\).

(Proof) First, (i) is easily seen from (23) and \(\tilde{F} \tilde{F}' = I\) as
\[
\tilde{U}' \tilde{U} = \tilde{X}^{-1/2} \tilde{F}' HS \cdot SH' \tilde{\Xi}^{-1/2}
\]
\[
= \tilde{X}^{-1/2} \tilde{F}' H \tilde{\Xi}^{-1/2} \tilde{F} \tilde{\Xi}^{-1/2}
\]
\[
= \tilde{X}^{-1/2} \tilde{X} \tilde{F} \tilde{\Xi}^{-1/2} = I.
\] (58)
Also, (ii) is easily seen by the fact
\[
Y' Y = \tilde{F}' HS \cdot SH' \tilde{\Gamma} = \tilde{F}' H \tilde{\Xi}^{-1/2} \tilde{F} \tilde{\Xi}^{-1/2}
\]
\[
= \tilde{F}' \tilde{\Xi}^{-1/2} \tilde{F} = 0,
\] (59)
where the last equality follows from \(\tilde{F} \tilde{F}' = 0\). This completes the proof.

(Proof of Theorem 4) First, let us show (36). By (56), we have
\[
\tilde{U} \tilde{\Xi}^{1/2} = SH' \tilde{\Gamma}.
\] (60)
Hence, by Lemma 2, we have
\[
\left[ \begin{array}{c} \tilde{U} \tilde{\Xi}^{1/2} \\ 0 \end{array} \right] = \left[ \begin{array}{c} SH' \tilde{\Gamma} \\ Y \end{array} \right] = SH' \tilde{\Gamma},
\] (61)
and which implies
\[
H' = S^{-1}[\tilde{U} \tilde{\Xi}^{1/2} \ 0] \Gamma' = S^{-1} \tilde{U} \tilde{\Xi}^{1/2} \tilde{F}.'
\] (62)
Thus, we have (36). Next, let us show (35) for \(S\) given by (34). Using the well-known formula of matrix to (16), we have
\[
(Q^{-1})^{-1} + H'R_0^{-1} H = Q^{-1}.
\] (63)
Note that by (62) and (21), we have
\[
H'R_0^{-1} H = S^{-1}[\tilde{U} \tilde{\Xi} \tilde{F}' \tilde{\Xi} S^{-1},
\] (64)
where \(\tilde{F}'\) is the set of components in \(\Psi\) corresponding to \(\tilde{F}\) in \(\Gamma\) defined by (31). Substituting (64) into (63) and noting (34), we have
\[
S^{-1}(I + \tilde{U} \tilde{\Xi} \tilde{F}' \tilde{\Xi} S^{-1}) = Q^{-1}.
\] (65)
Now, we have (35) from (65) if we show that
\[
(I + \tilde{U} \tilde{\Xi} \tilde{F}' \tilde{\Xi} S^{-1})^{-1} = K = \left( I + \frac{1}{\alpha} \tilde{U} \tilde{\Xi} \tilde{F}' \right).
\] (66)
Again by using the matrix inversion formula to the left-hand side of (66), we have
\[ (I + \ddot{U}\dddot{Z}^{-1}\dddot{U}^{-1})^{-1} = I - \ddot{U}\left(\dddot{Z}^{-1} + \dddot{U}^{-1}\right)^{-1}\dddot{U} \]
\[ = I - \frac{1}{\alpha} \ddot{U}\dddot{Z}^{-1}\dddot{U}, \tag{67} \]

where we have used the relation: \( \dddot{Z} + \dddot{U} = \alpha I \). Thus, we have shown (35) and (33). Finally, substituting (34) and (35) into (17), we have (32). This completes the proof. \( \blacksquare \)

(Proof of Theorem 5) Let \( Z \doteq (Q')^{-1} \ddot{U} = S\dddot{U} \). Then, note that the present problem is converted to the equivalent problem of the following form.

**Problem 2** Find \( Z \in R^{n \times k} \) such that

\[
\text{tr}[Q] \to \text{min.},
\tag{69}
\]

subject to

\[
Z'(Q')^{-1}Z = I,
\tag{70}
\]

and (17) and

\[
Q = Q^{-1} - \frac{1}{\alpha} Z\dddot{Z}Z'. \tag{71} \]

For this problem, let us define Lagrangean by

\[
L(Z, \Lambda) \doteq \text{tr}[Q] + \text{tr}[\Lambda(Z'(Q')^{-1}Z - I)],
\tag{72} \]

where \( \Lambda \in R^{n \times k} \) denotes the set of Lagrange multipliers for the matrix-valued constraint (70).

Then, from (72), we have

\[
\frac{\partial}{\partial \dddot{Z}_{ij}} L(Z, \Lambda) = \text{tr} \left[ \frac{\partial Q}{\partial \dddot{Z}_{ij}} + \text{tr} \left[ \Lambda [Z'(Q')^{-1}Z + Z'(Q')^{-1}E_{ij}] \right] \right] - \text{tr} \left[ AZ'(Q')^{-1} \frac{\partial Q}{\partial \dddot{Z}_{ij}} (Q')^{-1}Z \right], \tag{73} \]

where \( E_{ij} \doteq E_{ij}' \) and \( E_{ij} \in R^{n \times k} \) is defined by

\[
[E_{ij}]_{kl} = \delta_{ik}\delta_{j\ell}, \quad k = 1, 2, \ldots, n, \quad \ell = 1, 2, \ldots, m. \tag{74} \]

Note that from (17), we have

\[
\frac{\partial Q}{\partial \dddot{Z}_{ij}} = A \frac{\partial Q}{\partial \dddot{Z}_{ij}} A'. \tag{75} \]

Then, from (71) and (75), we have

\[
\frac{\partial Q}{\partial \dddot{Z}_{ij}} + A \frac{\partial Q}{\partial \dddot{Z}_{ij}} A' = \frac{1}{\alpha} \left[ E_{ij} \dddot{Z}Z' + Z\dddot{Z}E_{ij} \right]. \tag{76} \]

Hence, we have

\[
\frac{\partial Q}{\partial \dddot{Z}_{ij}} = -\frac{1}{\alpha} \sum_{k=1}^{\infty} A_k^*(E_{ij} \dddot{Z}Z' + Z\dddot{Z}E_{ij}) A_k^*. \tag{77} \]

Now, note the fact that because of (68), the solution of (38) has the expression

\[
X = \sum_{k=1}^{\infty} A_k^* \left( I - (Q')^{-1}ZA(Z')^{-1} \right) A_k. \tag{78} \]

Then, substitution of (75) and (77) into (73) with simple arrangements using (78) yields

\[
\frac{\partial}{\partial \dddot{Z}_{ij}} L(Z, \lambda) = - \frac{2}{\alpha} \left[ \Lambda Z\dddot{Z} - (Q')^{-1}Z\dddot{Z}(Q')^{-1} \right] \dddot{Z}_{ij}, \tag{79} \]

where we have used the relation: \( \alpha I - \dddot{Z} = \dddot{U} \). From (79) and (68), we have (37). Note that from (70), we have

\[
Z\dddot{Z} = (Q')^{-1} - (Q')^{-1}AZ = \Lambda \dddot{Z}^{-1}. \tag{80} \]

Since \( \dddot{U} \) and \( \dddot{Z} \) are diagonal matrices, \( \dddot{Z}^{-1} \) is also a diagonal matrix. Because \( \Lambda \) is a symmetric matrix, \( \Lambda \) must be a diagonal matrix. This completes the proof. \( \blacksquare \)

### 5 Concluding Remarks

In this paper, we did not demonstrate numerical examples of the proposed method of the optimization of \( H \). However, it is rather easy to construct a recursive algorithm to solve the set of equations given by Theorem 5 since our task is to determine \( \dddot{U} \) over a relatively small degree of freedom described by \( \dddot{U} \dddot{U} = I \). For example, starting with any \( \dddot{U} \) which satisfies \( \dddot{U} \dddot{U} = I \), we can update the value to meet the condition (37). The result of numerical studies using this type of algorithm will be reported in the near future.

### References


