Title
Real-time Estimation of a Markov Process Over a Noisy Digital Communication Channel

Authors
Xu, Qing
Sengupta, Raja

Publication Date
2005-11-01

CALIFORNIA PATH PROGRAM
INSTITUTE OF TRANSPORTATION STUDIES
UNIVERSITY OF CALIFORNIA, BERKELEY

Real-time Estimation of a Markov Process Over a Noisy Digital Communication Channel

Qing Xu
Raja Sengupta
University of California, Berkeley

California PATH Working Paper
UCB-ITS-PWP-2005-3

This work was performed as part of the California PATH Program of the University of California, in cooperation with the State of California Business, Transportation, and Housing Agency, Department of Transportation, and the United States Department Transportation, Federal Highway Administration.

Report for Task Order 4224

November 2005

ISSN 1055-1417

Real-time Estimation of a Markov Process Over a Noisy Digital Communication Channel

Qing Xu and Raja Sengupta

Abstract

We study the real-time estimation of a Markov process over a memoryless noisy digital communication channel. The goal of system design is to minimize the mean squared estimation error. We first show the optimal encoder and decoder can be memoryless in terms of the source symbols. We then prove the optimal encoder separates the real space with hyperplanes. In the case of the binary symmetric channel and scalar source, the optimal encoder can be a threshold. A recursive algorithm is given to jointly find a locally optimal encoder and decoder for the binary symmetric channel. For a memoryless Gaussian vector source and a binary symmetric channel, we show the optimal policy is to encode the principal component. We derive the minimum mean squared error as a function of the variance of source and the channel noise.

I. Introduction

This paper is about the design of encoders and decoders optimized to estimate the state of a stochastic dynamical system across a digital but noisy communication channel. Control and estimation over communication networks is attracting increasing attention. For example see the recent special issues of the IEEE Transactions on Automatic Control [3] and Control Systems Magazine [2] on networked control system. This class of problems is also given considerable weight in [17] in its evaluation of future directions in control, dynamics, and systems. They see control over communication networks as the natural next phase of the information revolution. It would transform current communication networks, now mainly concerned with the transmission of information, to have more interaction with the physical world. We ourselves have built control and estimation systems over digital communication networks for cars and airplanes [7][23][29][10][16]. For an audio-visual description of one of our systems see [1].

Here we present results on real-time estimation of the state of a Markov process over a noisy communication channel. Figure 1 shows the system schematically. A discrete-time continuous-valued Markov source is passed through an encoder at each discrete time-step. The encoder produces the input to the communication channel. The communication channel is assumed to have a finite, discrete alphabet. Thus we consider digital communications. The input and output alphabets are the same. In general the channel may output a symbol different from the one that is input, i.e., the channel is noisy. The channel is also memoryless. The output of the channel is fed to the decoder. The decoder is permitted to have memory. Its job is to output an estimate of the state of the Markov process. There are no communication delays.

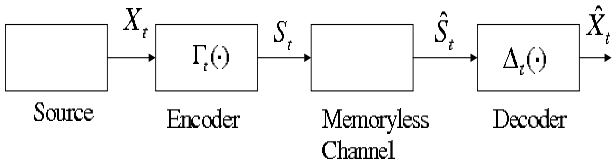

Fig. 1. State Estimation over Memoryless Channel

Our aim is to choose the encoder and decoder at each time-step to minimize the mean squared difference between the state of the Markov process and its estimate at the output of the decoder at the same time step. In other words, the encoder and decoder are to be designed for real-time minimum mean-square error estimation (MMSE). The Real-time has to do with the emphasis on choosing the encoder and decoder at time t to minimize the estimation error at time t. This distinguishes our formulation from the rate distortion, source and channel coding problems in information theory. Our objective function is the same as that in Kalman filtering [12]. However, the emphasis on the digital communication channel distinguishes this problem from Kalman's.

We review relevant previous works in section II. The problem statement is in section III. The rest of the paper is composed of two parts. The first part presents the structural results. There are three theorems and an algorithm in this part. The encoder at time t is permitted to be any function of the states up to time t. In section IV, Theorem 4.1 shows the encoder may be restricted, without loss of optimality, to a function of the current state of the Markov process and the probability mass function of the

This work was supported by California PATH program under project TO4224 and in part by General Motors R&D Center through contract TCS70709 and TCS53798 to U.C. Berkeley. The views expressed here are those of the authors and not of the research sponsors.

Q. Xu is with the Department of Mechanical Engineering, University of California, 1995 University Avenue, Suite 386, Berkeley, CA 94720, U.S.A. qingxu@me.berkeley.edu

R. Sengutpa is with the Department of Civil and Environmental Engineering, University of California, Berkeley, CA 94720, U.S.A. raja@path.berkeley.edu

memory of the decoder conditioned on the current state. This says the encoder can be causal and memoryless. Theorem 4.3 merely asserts that since ours is an MMSE problem, for any given encoder the optimal decoder is the conditional expectation of the state of the Markov process given all past channel outputs. Theorem 4.4 shows the encoder may be restricted, without loss of optimality, to a threshold type. In section V we present an iterative algorithm that converges to a locally optimal encoder and decoder for the binary symmetric channel. The algorithm synthesizes these results from control and information theory to derive a computational scheme to get optimal encoders and decoders for minimum mean-square error estimation. The algorithm itself turns out to be related to others well established in quantization and rate distortion theory. [4][5][15] The second part of the paper studies the special case of memoryless Gaussian vector source over binary symmetric channel. We show that the globally optimal encoding is to do a threshold encoding of the principal component. We also derive expressions for the minimum mean square error.

II. Previous Work

We situate our problem in a literature situated partly in control and partly in information theory. The problem of optimal estimation of a linear Gaussian Markov process, when the measurement is contaminated by an independent white Gaussian process, was studied by Kalman in [12] and [13]. However when the state is transmitted over a digital communication channel the state measurement, which is a real vector, has to be quantized into bits. Then the bits are transmitted over the noisy channel and are decoded on the other side. Thus the optimality of the orthogonal projection of the state onto the manifold generated by the observations no longer holds.

The problem of estimating state over a digital communication channel was first introduced in [28]. Nair and Evens extend this work in [19] [20] [21] and [18]. All these references consider a noise-less though bit-rate constrained channel, The system output at each time-step can be quantized into R bits, which are then transmitted over the channel without error. We consider both the bit rate limit and the channel noise, i.e., the bits received may not be the same as those transmitted.

In [24] Tatikonda derived the bit rates necessary for the controllability, observability, and stability of a dynamical system. Once again, the communication channel was assumed to be error free.

Walrand and Varaiya [26] studied the optimal coding-decoding problem. They consider a discrete alphabet source, and the Hamming distance as a measure of distortion. Our source is continuous valued. They also allow the encoder to have noiseless feedback from the channel which we do not assume.

Sahai studied the estimation problem of an unstable process over noisy channel in [**?**] and [**?**]. He considers the stability of the estimation. We on the other side study stable process, but concern more with the optimal performance in estimation.

Şimşek and Varaiya [6] extended the work of Sahai and studied the estimation over a binary symmetric channel. They derive conditions for stability. Once again, they assume channel feedback. We on the other hand find the optimal design to minimize the mean square estimation error without channel feedback.

Neuhoff and Gilbert [22] studied causal source codes, and show that the performance of memoryless coding is as good as any other causal coding at the minimum bit rate required to achieve a given distortion. We show similar results but in the presence of a noisy channel. They solve a pure source coding problem.

Quantization over a noisy channel problem was first introduced in [14]. They studied scalar quantization. Farvardin [8] extended the result to provide an iterative algorithm which converges to a locally optimal encoder for a given channel and distortion measure. Vector quantization is studied in [9]. The authors show the geometric structure of channel-optimized vector encoders and the implications on the complexity of encoding. We extend their results from the memoryless process to the Markov process. They provide an iterative algorithm to get a local optimum for any stationary source and discrete memoryless channel. We on the other hand present an algorithm for a Markov source, and binary symmetric channel.

Teneketzis [25] studied the real-time estimation of a discrete-time Markov process. They present a structural results similar to our first theorem. They assume a discrete-valued Markov source. We generalize their result to a continuous valued Markov source. Their cost function is also slightly different. They optimize the sum of all errors from the beginning to the current time, and design all the encoders and decoders at one time to minimize this sum. We on the other hand optimize the encoder and decoder at the current instant to minimize the distortion at the current instant, assuming the prior encoders and decoders are already fixed.

Xu and Hespanha study optimal communication logics for networked control systems in [30]. They derive communication policies for the optimal control of an estimator-based networked control system architecture to reduce communication load. Unlike us, they do not consider quantization of the communication signals. The channel in their problem is also error free. In [31] the authors study the minimal rate requirements for state estimation in linear time-invariant systems. For different estimation distortion criterion, they find the minimum data rate required from the channel. The channel they consider again has a constraint on data rate, but is noiseless. We consider noisy channels.

III. Problem Statement

The system is shown in Figure 1. We describe each part of it below.

1) Source:

$X_t \in \mathbb{R}^n$ is a Markov process.

2) Encoder: Define $X_1^t \triangleq \{X_1, X_2, \cdots, X_t\}$.

$$S_t = \Gamma_t(X_1^t) \tag{1}$$

where $S_t \in \mathrm{L} = \{1, 2, \cdots, K\}$ and $K \in \mathbb{N}^+$.

3) Channel: Memoryless

$$\hat{S}_t = H_t(S_t, N_t) \tag{2}$$

where $N_t \in \{1, 2, \cdots, \gamma\}$ and $\gamma \in \mathbb{N}^+$. N_t is independent for different t and independent of S_t. $\hat{S}_t \in \mathrm{L}$.

4) Receiver memory update:

a) At $t = 1$, $M_1 = l_1(\hat{S}_1)$.

b) At $t > 1$, $M_t = l_t(\hat{S}_t, M_{t-1})$.

where $M_t \in \mathrm{W}_t = \{1, 2, \cdots, \kappa_t\}$ and $\kappa_t \in \mathbb{N}^+$. Denote the space of probability mass functions in W_t as $\mathbb{P}^{\mathrm{W}_t}$, and the probability mass function of M_t as P_{M_t}. Define the probability mass function of M_t conditioned on $X_t = x_t$ as $P_{M_t(x_t)} \triangleq P(M_t \mid X_t = x_t)$.

5) Decoder:

$$\hat{X}_t = \Delta_t(\hat{S}_t, M_{t-1}) \tag{3}$$

6) Cost Function:

$$E\{\|X_t - \hat{X}_t\|^2\} \tag{4}$$

The Real-time Estimation Problem: At each time $t = \tau$, given $\Gamma_1^{\tau-1} \triangleq \{\Gamma_1, \Gamma_2, \cdots, \Gamma_{\tau-1}\}$, $\Delta_1^{\tau-1} \triangleq \{\Delta_1, \Delta_2, \cdots, \Delta_{\tau-1}\}$, and $l_1^\tau \triangleq \{l_1, l_2, \cdots, l_\tau\}$, find the encoder $\Gamma_\tau(\cdot)$, and decoder $\Delta_\tau(\cdot)$, such that $E\{\|X_\tau - \hat{X}_\tau\|^2\}$ is minimized.

IV. The structure of the optimal encoder and decoder

In this section we prove three structural results about the real-time estimation problem. Firstly, by Theorem 4.1 we prove that for a given decoder, the optimal encoder for real-time estimation of the state of a Markov process is separable, i.e., it need not depend on the previous states. This is an extension of the result in [25] to a continuous-valued source. Our proof is also similar although our cost function is a bit different, as discussed in section II. Lemma 4.2 is an intermediate result used to prove Theorem 4.1. Then Theorem 4.3 asserts the optimal decoder for any given encoder is the expected value of the state conditioned on the previously channel outputs. Finally we prove in Theorem 4.4 the optimal encoder is a hyper-plane encoder. It partitions the real space with hyper-planes, and maps the X_t values in each subspace to a distinct symbol. This result is based on Theorem 4.1. The proof technique is similar to [8].

A. The optimal encoder

This subsection is about the structure of the optimal encoder for the real-time Markov process estimation problem. The result is in Theorem 4.1.

Theorem 4.1: For any t, one can replace Γ_t with some Γ_t^*

$$\Gamma_t^* : \mathbb{R}^n \times \mathbb{P}^{\mathrm{W}_t} \to \mathrm{L}$$

so that $s_t = \Gamma_t^*(x_t, P_{M_{t-1}(x_{t-1})})$ without loss of optimality.

Like [25], we prove the theorem with a two-stage lemma, i.e., Lemma 4.2. This approach first appeared in [27].

Below in Lemma 4.2 we consider a vector Markov process. The states in the first two time instants are $X_1 \in \mathbb{R}^n$ and $X_2 \in \mathbb{R}^n$. The encoder at stage 1 is $\Gamma_1 : \mathbb{R}^n \to \mathrm{L}$ with $S_1 = \Gamma_1(X_1)$. The encoder at stage 2 is $\Gamma_2 : \mathbb{R}^{n \times n} \to \mathrm{L}$ with $S_2 = \Gamma_1(X_1, X_2)$. $S_t \in \mathrm{L} = \{1, 2, \cdots, K\}$ for $t = 1, 2$. Then we have the following lemma.

Lemma 4.2: Two-stage lemma:

Consider a two-stage system where

$$\Gamma_2 : \mathbb{R}^{n \times n} \to \mathrm{L}$$

so that $S_2 = \Gamma_2(X_1, X_2)$, then one can replace Γ_2 with Γ_2^*,

$$\Gamma_2 : \mathbb{R}^n \times \mathbb{P}^{\mathrm{W}_1} \to \mathrm{L}$$

so that $S_2 = \Gamma_2^*(X_2, P_{M_1}(x_1))$ without loss of optimality.

Proof: With a given design $d \triangleq (\Gamma_1, \Gamma_2, l_1, l_2, \Delta_1, \Delta_2)$, define $\hat{\rho}_2(X_2, M_1, S_2, N_2) \triangleq \|X_2 - \Delta_2(M_1, H_2(S_2, N_2))\|^2$. Define $P^d_{M_1(x_1)} \triangleq P^d(M_1 = m_1 \mid X_1 = x_1)$ to be the probability mass function of M_1 conditioned on $X_1 = x_1$, under design d. It depends on Γ_1 and l_1 but not Γ_2. We then have for any $X_1 = x_1$, $X_1 = x_2$

$$
\begin{aligned}
& E^d\{\|X_2-\hat{X}_2\|^2 \mid X_1=x_1, X_2=x_2\} \\
= \;& E^d\{\|X_2-\hat{X}_2\|^2 \mid X_1=x_1, X_2=x_2, P^d_{M_1(x_1)}\} \\
= \;& E^d\{\|X_2-\Delta_2(M_1,H_2(S_2,N_2))\|^2 \mid X_1=x_1, X_2=x_2, P^d_{M_1(x_1)}\} \\
= \;& E^d\{\hat{\rho}_2(X_2,M_1,S_2,N_2) \mid X_1=x_1, X_2=x_2, P^d_{M_1(x_1)}\} \\
= \;& \sum_{m_1}\sum_{s_2}\sum_{n_2} P^d(M_1=m_1, S_2=s_2, N_2=n_2 \mid X_1=x_1, X_2=x_2, P^d_{M_1(x_1)})\cdot\hat{\rho}_2(x_2,m_2,s_2,n_2) \\
= \;& \sum_{s_2} P^d(S_2=s_2 \mid X_1=x_1, X_2=x_2)\left[\sum_{n_2} P(N_2=n_2)\cdot[\sum_{m_1} P^d_{M_1(x_1)}(m_1)\hat{\rho}(x_2,m_1,s_2,n_2)]\right]
\end{aligned}
$$

Now consider a new design $\hat{d}$ where $\Gamma_2^*:\mathbb{R}^n\times\mathbb{P}^{\mathrm{W}_1}\to \mathrm{L}$ is chosen as follows: For any given $x_2\in\mathbb{R}^n$ and any given $P_{M_1}\in\mathbb{P}^{\mathrm{W}_1}$

$$
\Gamma_2^*(x_2,P_{M_1}(x_1)) = \arg\min_{s_2\in\mathrm{L}}\left\{\sum_{n_2} P(N_2=n_2)\cdot\left[\sum_{n_2} P(N_2=n_2)\left[\sum_{m_1} P^d_{M_1(x_1)}(m_1)\hat{\rho}(x_2,m_1,s_2,n_2)\right]\right]\right\}
$$

Keep the decoders the same in the new design. Then, under the new design $\hat{d}=(\Gamma_1,\Gamma_2^*,l_1,l_2,\Delta_1,\Delta_2)$, for all x_1,

$$
P^d_{M_1(x_1)} = P^{\hat{d}}_{M_1(x_1)}
$$

and

$$
\begin{aligned}
& E^{\hat{d}}\{\|X_2-\hat{X}_2\|^2 \mid X_2=x_2, P^d_{M_1(x_1)}\} \\
= \;& E^{\hat{d}}\{\|X_2-\hat{X}_2\|^2 \mid X_1=x_1, X_2=x_2, P^d_{M_1(x_1)}\} \\
\le \;& E^{d}\{\|X_2-\hat{X}_2\|^2 \mid X_1=x_1, X_2=x_2, P^d_{M_1(x_1)}\} \\
= \;& E^{d}\{\|X_2-\hat{X}_2\|^2 \mid X_2=x_2, P^d_{M_1(x_1)}\} \qquad (5)
\end{aligned}
$$

Therefore

$$
E^{\hat{d}}\{\|X_2-\hat{X}_2\|^2\} \le E^d\{\|X_2-\hat{X}_2\|^2\}
$$

■

Using Lemma 4.2, we can prove Theorem 4.1. The basic idea is to aggregate the system state from time 1 to $t-1$ into one "super-state" at the first stage, and view the state at t as the second stage so that the two-stage lemma can be applied.

Proof of Theorem 4.1

Proof: The given t-stage system can be considered as a two-stage system by setting

$$
\begin{aligned}
\bar{X_1} &\triangleq (X_1,X_2,\cdots,X_{t-1}) \\
\bar{X_2} &\triangleq X_t \\
\bar{N_1} &\triangleq (N_1,N_2,\cdots,N_{t-1}) \\
\bar{N_2} &\triangleq N_t \\
\bar{S_1} &\triangleq (S_1,S_2,\cdots,S_{t-1}) \\
\bar{S_2} &\triangleq S_t \\
\bar{\hat{S}_1} &\triangleq (\hat{S}_1,\hat{S}_2,\cdots,\hat{S}_{t-1}) \\
\bar{\hat{S}_2} &\triangleq \hat{S}_t \\
\bar{M_1} &\triangleq M_{t-1}=\phi(\bar{X_1},\bar{N_1}) \\
\bar{M_2} &\triangleq M_t \\
\bar{\hat{X}_1} &\triangleq (\hat{X}_1,\hat{X}_2,\cdots,\hat{X}_{t-1}) \\
\bar{\hat{X}_2} &\triangleq \hat{X}_t \\
\hat{\Gamma}_2(\bar{X_1},\bar{X_2}) &\triangleq \Gamma_t(X_1,X_2,\cdots,X_t) \\
\bar{l_2}(\bar{M_1},\bar{\hat{S}_2}) &\triangleq l_t(M_{t-1},\hat{S}_t)
\end{aligned}
$$

Then by the two-stage lemma there is an encoder Γ_2^* that has the structure

$$\bar{s_2} = \Gamma_2^*(\bar{x_2}, P_{\bar{M_1}}(x_1))$$

which does not increase the cost. This corresponds to

$$s_t = \Gamma_t^*(x_t, P_{M_{t-1}}(x_{t-1}))$$

■

B. The optimal decoder

The following theorem characterized the optimal decoder for any given encoder.

Theorem 4.3: For any encoder Γ_t the optimal decoder Δ_t is

$$\hat{X}_t = \Delta_t(\hat{S}_t, M_{t-1}) = E_{\Gamma_1^t, I_1^{t-1}}\{X_t \mid \hat{S}_t, M_{t-1}\}$$

Proof: This is a well-known result in estimation theory. It appears in, for example, Theorem 1 of [12]. The proof is omitted here. ■

C. Optimality of the hyper-plane encoder

Theorem 4.1 opens a way to use quantization techniques for noisy channels. The following theorem is similar to the result in [8]. Unlike the quantization problem, in our problem the receiver memory needs to be considered.

Theorem 4.4: For any given decoder, the optimal encoder for the real-time estimation problem is a hyperplane encoder. In particular, let the reconstruction points be $\{c_1, c_2, \cdots, c_{\hat{K}}\}$, where and $1 \le i \le \hat{K}$ and $c_i = \Delta(\hat{S}_t = i, M_{t-1})$ depends on the memory of the receiver. Define $A_i \triangleq \{x_t : \Gamma_t^*(x_t, P_{M_{t-1}}) = i\}$, then A_i and A_l are separated by the hyper-plane

$$\left\{ x_t \in \mathbb{R}^n : 2\sum_{j=1}^{\hat{K}} \left[P(\hat{S}_t = j \mid S_t = l) - P(\hat{S}_t = j \mid S_t = i)\right] \cdot \langle x_t, c_j \rangle = \sum_{j=1}^{\hat{K}} \left[P(\hat{S}_t = j \mid S_t = l) - P(\hat{S}_t = j \mid S_t = i)\right] \cdot \|c_j\|^2 \right\}$$

$A_i \cap A_l = \emptyset, \forall 1 \le i, l \le K, i \ne l$, and $\cup_{i=1}^K A_i = \mathbb{R}^n$.

Proof: An optimal encoder should map all the vectors in a way such that all the x_t mapped to the i-th region, i.e. $S_t = i$, produce smaller mean squared error than if they are mapped to any other, say, l-th region. Denote the set of vectors mapped to the i-th region by an optimal encoder as A_i^* Then all the vectors in A_i^* should satisfy the following equation for any given l other than i.

$$\begin{aligned}
& E\{\|x_t - \hat{x}_t\|^2 \mid S_t = i\} - E\{\|x_t - \hat{x}_t\|^2 \mid S_t = l\} \\
= \;& x_t^2 - 2E\{\langle x_t, \hat{x}_t \rangle \mid S_t = i\} + E\{\hat{x}_t^2 \mid S_t = i\} - x_t^2 + 2E\{\langle x_t, \hat{x}_t \rangle \mid S_t = l\} - E\{\hat{x}_t^2 \mid S_t = l\} \\
= \;& 2\left(E\{\langle x_t, \hat{x}_t \rangle \mid S_t = l\} - E\{\langle x_t, \hat{x}_t \rangle \mid S_t = i\}\right) + \left(E\{\hat{x}_t^2 \mid S_t = i\} - E\{\hat{x}_t^2 \mid S_t = l\}\right) \\
= \;& \sum_{j=1}^{\hat{K}} \left[P(\hat{S}_t = j \mid S_t = l) - P(\hat{S}_t = j \mid S_t = i)\right] \langle x_t, c_j \rangle - \sum_{j=1}^{\hat{K}} \left[P(\hat{S}_t = j \mid S_t = l) - P(\hat{S}_t = j \mid S_t = i)\right] \cdot \|c_j\|^2 \\
\le \;& 0
\end{aligned}$$

where $\langle a, b \rangle$ denotes the inner product of a and b.

For any given $1 \le l \le \hat{K}$, consider the following sets

$$A_{il} \triangleq \left\{ x_t \in \mathbb{R}^n : 2\sum_{j=1}^{\hat{K}} \left[P(\hat{S}_t = j \mid S_t = l) - P(\hat{S}_t = j \mid S_t = i)\right] \cdot \langle x_t, c_j \rangle \le \sum_{j=1}^{\hat{K}} \left[P(\hat{S}_t = j \mid S_t = l) - P(\hat{S}_t = j \mid S_t = i)\right] \cdot \|c_j\|^2 \right\}$$

For any l, the vectors in A_i^* are in the set A_{il}, hence

$$A_i^* = \bigcap_{l \ne i} A_{il} \tag{6}$$

The regions A_i^* and A_l^* are separated by the hyperplane

$$\left\{ x_t \in \mathbb{R}^n : 2\sum_{j=1}^{\hat{K}} \left[P(\hat{S}_t = j \mid S_t = l) - P(\hat{S}_t = j \mid S_t = i)\right] \cdot \langle x_t, c_j \rangle = \sum_{j=1}^{\hat{K}} \left[P(\hat{S}_t = j \mid S_t = l) - P(\hat{S}_t = j \mid S_t = i)\right] \cdot \|c_j\|^2 \right\}$$

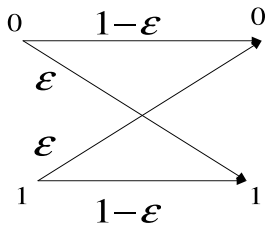

Fig. 2. A Binary Symmetric Channel

which is a hyper-plane in R^n. ■

Remark 4.5: Note when the source is scalar, the optimal encoder separates the range of the source into continuous intervals, and maps the points in each interval into a different symbol. That is, the optimal encoder for the scalar source is a threshold encoder.

V. Algorithm to find the optimal threshold of a scalar encoder for a binary symmetric channel

In this section we focus on the special case of the binary symmetric channel and the scalar Markov source.

$$\begin{cases} P(\hat{S}_t = S_t) = 1-\varepsilon \\ P(\hat{S}_t = 1-S_t) = \varepsilon \end{cases} \tag{7}$$

where $S_t, \hat{S}_t \in \{0,1\}$.

The channel is shown in Fig 2.

In this case the optimization problem reduces to that of finding an optimal threshold T such that

$$S_t = \begin{cases} 0 & X_t \le T \\ 1 & X_t > T \end{cases}$$

For this special case of the original real-time estimation problem, we state a recursive algorithm to find a locally optimal solution. The algorithm is based on the one presented in [8]. But unlike in their problem, we have to consider the receiver memory in our calculation. Our approach is to recursively find the optimal encoder for a given decoder and then find the optimal decoder for a given encoder. Since each iteration reduces the mean squared error, the algorithm converges. Recursive algorithms are used in information theory to find the rate-distortion function and channel capacity [4] [5]. There the optimization is performed over convex sets, so the solution obtained is globally optimal. We on the other hand only know the algorithm converges to a locally optimal solution.

Let the reconstruction points be R_0 and R_1, both in $\mathbb{R}$, such that

$$\hat{X}_t = \begin{cases} R_0 = \Delta(\hat{S}_t = 0, M_{t-1}) \\ R_1 = \Delta(\hat{S}_t = 1, M_{t-1}) \end{cases}$$

Then the mean squared error is

$$\begin{aligned} D &= E\{|X_t - \hat{X}_t|^2\} \\ &= P(\hat{S}=0 \mid S=0)\int_{-\infty}^{T} |x-R_0|^2 p_X(x)dx + P(\hat{S}=1 \mid S=0)\int_{-\infty}^{T} |x-R_1|^2 p_X(x)dx \\ &\quad + P(\hat{S}=0 \mid S=1)\int_{T}^{\infty} |x-R_0|^2 p_X(x)dx + P(\hat{S}=1 \mid S=1)\int_{T}^{\infty} |x-R_1|^2 p_X(x)dx \end{aligned}$$

A. The optimal encoder for a fixed decoder

For fixed R_0 and R_1, we can find the optimal threshold T^* by differentiating D with respect to T. In the following equation, let $P(A|B) \triangleq P(\hat{S}=A|S=B)$, where $A,B \in \{0,1\}$, then we have

$$\begin{aligned} \frac{dD}{dT} &= 0 \Rightarrow \\ T &= \frac{1}{2}\frac{(P(0|1)-P(0|0))R_0^2 + (P(1|1)-P(1|0))R_1^2}{(P(0|1)-P(0|0))R_0 + (P(1|1)-P(1|0))R_1} \\ &= \frac{1}{2}\cdot\frac{(2\varepsilon-1)(R_0^2-R_1^2)}{(2\varepsilon-1)(R_0-R_1)} \\ &= \frac{1}{2}(R_0+R_1) \end{aligned} \tag{8}$$

To minimize mean squared error for fixed R_0 and R_1 we also need

$$\frac{d^2D}{dT^2} = 2(2\varepsilon - 1)(R_0 - R_1) > 0 \tag{9}$$

Therefore

$$\frac{d^2D}{dT^2} > 0 \Leftrightarrow \begin{cases} R_0 < R_1 & \text{when } \varepsilon < \frac{1}{2} \\ R_0 \geq R_1 & \text{when } \varepsilon \geq \frac{1}{2} \end{cases} \tag{10}$$

Therefore for given R_0 and R_1 (hence a given decoder), the optimal encoder puts the threshold at the mid-point of the two reconstruction points. In addition, equation (10) must be satisfied. We will further discuss this point in the next subsection.

B. The optimal decoder for fixed encoder

For fixed encoder, the optimal decoder is the conditional expectation.

$$\hat{X}_t = \begin{cases} R_0 = E\{X_t \mid \hat{S}_t = 0, M_{t-1}\} \\ R_1 = E\{X_t \mid \hat{S}_t = 1, M_{t-1}\} \end{cases} \tag{11}$$

where M_{t-1} is the known receiver memory from the last step.

Now once the optimal decoder for a fixed encoder is given by (11), we can go back to check the optimality condition given by (10).

We notice first, for the threshold given in (8) to be optimal, (10) must be true, but this is not guaranteed by (11), i.e., there may be solutions of (11) that violate (10).

Secondly, if we flip the areas encoded to 0 and 1, R_0 and R_1 will also flip since

$$\begin{aligned} R_0 &= E\{X_t \mid \hat{S}_t = 0, M_{t-1}\} \\ &= E\{X_t \mid S_t = 0, M_{t-1}\} \cdot \frac{(1-\varepsilon)P(S_t=0)}{(1-\varepsilon)P(S_t=0) + \varepsilon P(S_t=1)} + \\ &\quad E\{X_t \mid S_t = 1, M_{t-1}\} \cdot \frac{\varepsilon P(S_t=1)}{(1-\varepsilon)P(S_t=0) + \varepsilon P(S_t=1)} \end{aligned}$$

and

$$\begin{aligned} R_1 &= E\{X_t \mid \hat{S}_t = 0, M_{t-1}\} \\ &= E\{X_t \mid S_t = 0, M_{t-1}\} \cdot \frac{\varepsilon P(S_t=0)}{\varepsilon P(S_t=0) + (1-\varepsilon)P(S_t=1)} + \\ &\quad E\{X_t \mid S_t = 1, M_{t-1}\} \cdot \frac{(1-\varepsilon)P(S_t=1)}{\varepsilon P(S_t=0) + (1-\varepsilon)P(S_t=1)} \end{aligned}$$

Hence if we have $S'_t = \begin{cases} 0 & X_t > T \\ 1 & X_t \leq T \end{cases}$, then $R'_0 = R_1$ and $R'_1 = R_0$. But the derivation of (8) is not affected by this flip. Therefore by simply exchanging the areas coded to 1 and 0 we can always make (10) true and thus make the threshold given by (8) optimal.

C. The algorithm to find the optimal encoder and decoder

In summary, we obtain the following recursive algorithm to find an encoder and decoder for transmission of a Markov process over a binary symmetric channel:

- Step 1: Set $\{R_0, R_1\} = \{R_0^{(0)}, R_1^{(0)}\}$, the initial reconstruction levels. They must satisfy (10) but are otherwise arbitrary.
- Step 2: Set $k = 0$ (the iteration index), and $D^{(0)} = \infty$.
- Step 3: Use (8) to determine the best threshold $T^{(k)}$.
- Step 4: Set $k = k+1$. Use (11) to find the best reconstruction levels $R_0^{(k)}$ and $R_1^{(k)}$.
- Step 5: Check if (10) is satisfied. If not, flip the areas encoded to 0 and 1, and therefore flip $R_0^{(k)}$ and $R_1^{(k)}$.
- Step 6: Compute the MSE $D^{(k)}$. If $\frac{D^{(k-1)} - D^{(k)}}{D^{(k)}} < \delta$, where δ is a preset positive fraction, go to step 7, otherwise go to step 3.
- Step 7: End the algorithm.

Remark 5.1: 1) Since with each iteration the MSE always decreases, the algorithm converges.

2) The role played by memory in the system is in (11), which further affects the solution of (8).

VI. Optimal estimation of memoryless Gaussian random vector source over binary symmetric channel

In this section we discuss the special case of the memoryless source. We are able to analytically characterize the optimal encoder and the minimum mean square error in this case. The optimal encoding strategy is to encode the principal component of the source. Lemma 6.2 asserts this for a random vector with independent components. Theorem 6.11 asserts the same for a source vector with correlated components.

A. System Description

Let $\mathbf{X} \in \mathbb{R}^n$ and $\mathbf{X} \sim \mathrm{N}(\mathbf{0}_n, \mathbf{K_x})$, where $\mathbf{K_x} \in \mathbb{R}^{n\times n}$. $\mathbf{X}$ is encoded with

$$S = G(\mathbf{X}) \tag{12}$$

where $S \in \{0,1\}$. K_x is a symmetric positive definite matrix.

S is transmitted through a memoryless channel. From now on assume the channel is binary symmetric, i.e.

$$P(\hat{S}=0|S=0) = P(\hat{S}=1|S=1) = 1-\varepsilon$$
$$P(\hat{S}=0|S=1) = P(\hat{S}=1|S=0) = \varepsilon$$

The decoder is

$$\hat{\mathbf{X}} \triangleq [\hat{X}_1\ \hat{X}_2 \ldots \hat{X}_n]^T = \Delta(\hat{S}) \tag{13}$$

The objective is to estimate $\mathbf{X}$ with minimum mean squared error, i.e. to design $G^*(\cdot)$ and $\Delta^*(\cdot)$ to minimize $E\{(\mathbf{X}-\hat{\mathbf{X}})^T(\mathbf{X}-\hat{\mathbf{X}})\}$.

For any given $G(\cdot)$, the optimal decoder is the conditional expectation, i.e. $\hat{\mathbf{X}} = E\{\mathbf{X}|\hat{S}\} = \left[E\{X_1|\hat{S}\}\ E\{X_2|\hat{S}\} \ldots E\{X_n|\hat{S}\}\right]^T$.

B. The Optimal Vector Encoder for Binary Symmetric Channel: Independent Gaussian Noise Case

Since in section IV it is shown that the optimal vector encoder over noisy channel partitions the vector space with hyperplanes, we search for our optimal design within this class of encoders. The following two lemmas provide the optimal encoder among all the encoders that partition the $\mathbb{R}^n$ space with a plane. In this subsection we derive the optimal encoder design when the components of the Gaussian random vector are mutually independent. We discuss the case of correlated components in the next subsection.

Lemma 6.1: Let $\mathbf{X} \in \mathbb{R}^n$ and $\mathbf{X} \sim \mathrm{N}(\mathbf{0}_n, \mathbf{K_x})$, and $\mathbf{K_x} = diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2)$. Let $\mathbf{w} \in \mathbb{R}^n$, $\|\mathbf{w}\| = 1$ and $b \in (-\infty, \infty)$. Define a encoder such that

$$S = G(\mathbf{X}) = \begin{cases} Y & \text{if } \mathbf{w^T X} \geq b \\ 1-Y & \text{otherwise} \end{cases}$$

where $Y \in \{0,1\}$. Then for any binary symmetric channel, among all encoders with the same $\mathbf{w}$, the encoder with $b=0$, i.e. when the plane passes through the origin, is optimal.

Lemma 6.2: Let $\mathbf{X} \in \mathbb{R}^n$ and $\mathbf{X} \sim \mathrm{N}(\mathbf{0}_n, \mathbf{K_x})$ and $\mathbf{K_x} = diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2)$. Consider the encoder $G^*(\cdot)$ defined as below

$$S = G^*(\mathbf{X}) = \begin{cases} Y & \text{if } \mathbf{w^{*T} X} \geq 0 \\ 1-Y & \text{otherwise} \end{cases}$$

where $Y \in \{0,1\}$, $\mathbf{w}^* \in \mathbb{R}^n$, $\|\mathbf{w}^*\| = 1$ is chosen as below.

1) If $\sigma_1 = \sigma_2 = \cdots = \sigma_n = \sigma$, i.e. all the n directions are equally noisy, let $\mathbf{w}^*$ be any vector in $\mathbb{R}^n$.
2) Let $\sigma_m = \max\{\sigma_1, \sigma_2, \ldots, \sigma_n\}$, with $m \in \{1,2,\ldots,n\}$, let $\mathbf{w}^*$ be the unit vector in the m-th direction.

Then $G^*(\cdot)$ is optimal, i.e., the optimal encoder only encodes the most noisy direction with one bit.

To prove Lemmas 6.1 and 6.2 we prove lemmas 6.3 to 6.8.

Lemma 6.3: Minimizing the mean squared error $E\{(\mathbf{X}-\hat{\mathbf{X}})^T(\mathbf{X}-\hat{\mathbf{X}})\}$ is equivalent to maximizing

$$\left\|E\{\mathbf{X}|\hat{S}=0\}\right\|^2 P(\hat{S}=0) + \left\|E\{\mathbf{X}|\hat{S}=1\}\right\|^2 P(\hat{S}=1) \tag{14}$$

Proof: Since $\hat{\mathbf{X}} = E\{\mathbf{X}|\hat{S}\}$ we have

$$
\begin{aligned}
& E\{(\mathbf{X}-\hat{\mathbf{X}})^T(\mathbf{X}-\hat{\mathbf{X}})\} \\
= \;& E\{\left\|\mathbf{X}-E\{\mathbf{X}|\hat{S}\}\right\|^2\} \\
= \;& E\left\{\left\|\mathbf{X}-E\{\mathbf{X}|\hat{S}=0\}\right\|^2|\hat{S}=0\right\}P(\hat{S}=0)+E\left\{\left\|\mathbf{X}-E\{\mathbf{X}|\hat{S}=1\}\right\|^2|\hat{S}=1\right\}P(\hat{S}=1) \\
= \;& \left(E\{\mathbf{X}^T\mathbf{X}|\hat{S}=0\}-2E\{\mathbf{X}^T|\hat{S}=0\}E\{\mathbf{X}|\hat{S}=0\}+\left\|E\{\mathbf{X}|\hat{S}=0\}\right\|^2\right)\cdot P(\hat{S}=0\Big) \\
& +\left(E\{\mathbf{X}^T\mathbf{X}|\hat{S}=1\}-2E\{\mathbf{X}^T|\hat{S}=1\}E\{\mathbf{X}|\hat{S}=1\}+\left\|E\{\mathbf{X}|\hat{S}=1\}\right\|^2\right)\cdot P(\hat{S}=1\Big) \\
= \;& E\{\mathbf{X}^T\mathbf{X}|\hat{S}=0\}P(\hat{S}=0)-\left\|E\{\mathbf{X}|\hat{S}=0\}\right\|^2P(\hat{S}=0)+E\{\mathbf{X}^T\mathbf{X}|\hat{S}=1\}P(\hat{S}=1)-\left\|E\{\mathbf{X}|\hat{S}=1\}\right\|^2P(\hat{S}=1) \\
= \;& E\{\mathbf{X}^T\mathbf{X}\}-\left\|E\{\mathbf{X}|\hat{S}=0\}\right\|^2P(\hat{S}=0)-\left\|E\{\mathbf{X}|\hat{S}=1\}\right\|^2P(\hat{S}=1) \\
= \;& \sum_{i=1}^{n}\sigma_i^2-\left\|E\{\mathbf{X}|\hat{S}=0\}\right\|^2P(\hat{S}=0)-\left\|E\{\mathbf{X}|\hat{S}=1\}\right\|^2P(\hat{S}=1)
\end{aligned}
$$

Thus the lemma is true. ■

Below we define the probability mass function of the section of $\mathbb{R}^n$ mapped to 0 and 1 respectively.

$$
\begin{aligned}
P_0(\mathbf{w},b) \triangleq P(S=0) &= \frac{1}{\sqrt{(2\pi)^n det(\mathbf{K_x})}}\int_{\mathbf{w^T x}<b} e^{-\frac{1}{2}\mathbf{x^T K_x^{-1} x}}d\mathbf{x} \\
&= \frac{1}{\sqrt{(2\pi)^n}\prod_{i=1}^n\sigma_i}\int_{\mathbf{w^T x}<b} e^{-\frac{1}{2}\mathbf{x^T K_x^{-1} x}}d\mathbf{x}
\end{aligned}
$$

and

$$
\begin{aligned}
P_1(\mathbf{w},b) &\triangleq P(S=1) \\
&= 1-P_0(\mathbf{w},b) \\
&= \frac{1}{\sqrt{(2\pi)^n det(\mathbf{K_x})}}\int_{\mathbf{w^T x}<b} e^{-\frac{1}{2}\mathbf{x^T K_x^{-1} x}}d\mathbf{x} \\
&= \frac{1}{\sqrt{(2\pi)^n}\prod_{i=1}^n\sigma_i}\int_{\mathbf{w^T x}<b} e^{-\frac{1}{2}\mathbf{x^T K_x^{-1} x}}d\mathbf{x}
\end{aligned}
$$

In above equations, the special case where $b=0$ is included. Obviously, when $b=0$, $P_0(\mathbf{w},0)=P_1(\mathbf{w},0)=\frac{1}{2}$. Hereafter we drop the arguments of $P_0(\mathbf{w},b)$ and $P_1(\mathbf{w},b)$ and simply write them as P_0 and P_1.

Lemma 6.4: For any $i\in\{1,2,\ldots,n\}$, define

$$
\begin{aligned}
\bar{X}_i(0) &\triangleq E\{X_i|S=0\} \\
&= \frac{1}{P_0\sqrt{(2\pi)^n det(\mathbf{K_x})}}\int_{\mathbf{w^T x}<b} x_i e^{-\frac{1}{2}\mathbf{x^T K_x^{-1} x}}d\mathbf{x}
\end{aligned}
$$

$$
\begin{aligned}
\bar{X}_i(1) &\triangleq E\{X_i|S=1\} \\
&= \frac{1}{P_1\sqrt{(2\pi)^n det(\mathbf{K_x})}}\int_{\mathbf{w^T x}\geq b} x_i e^{-\frac{1}{2}\mathbf{x^T K_x^{-1} x}}d\mathbf{x}
\end{aligned}
$$

and

$$
\bar{\mathbf{X}}(j)=[\bar{X}_1(j)\ \bar{X}_2(j)\ldots\bar{X}_n(j)]^T \tag{15}
$$

for $j\in\{0,1\}$.

Then we have

$$
E\{X_i|\hat{S}=0\}=\frac{(1-\varepsilon)P_0\bar{X}_i(0)+\varepsilon P_1\bar{X}_i(1)}{(1-\varepsilon)P_0+\varepsilon P_1} \tag{16}
$$

$$
E\{X_i|\hat{S}=1\}=\frac{\varepsilon P_0\bar{X}_i(0)+(1-\varepsilon)P_1\bar{X}_i(1)}{\varepsilon P_0+(1-\varepsilon)P_1} \tag{17}
$$

Proof:

$$\begin{aligned} E\{X_i|\hat{S}=0\} &= E\{X_i|S=0,\hat{S}=0\}P(S=0|\hat{S}=0)+E\{X_i|S=1,\hat{S}=0\}P(S=1|\hat{S}=0) \\ &= E\{X_i|S=0\}P(S=0|\hat{S}=0)+E\{X_i|S=1\}P(S=1|\hat{S}=0) \\ &= \frac{(1-\varepsilon)P_0}{(1-\varepsilon)P_0+\varepsilon P_1}\bar{X_i}(0)+\frac{\varepsilon P_1}{(1-\varepsilon)P_0+\varepsilon P_1}\bar{X_i}(1) \end{aligned}$$

The second equation above is because that $X_i \to S \to \hat{S}$ is a Markov chain. The third equation is because of Bayes Rule. Changing $\hat{S}=0$ to $\hat{S}=1$, we can prove equation (17) in the same way. ■

Lemma 6.5:

$$P_0\bar{X_i}(0)+P_1\bar{X_i}(1)=E\{X_i\}=0, \forall i \in \{1,2,\cdots,n\} \tag{18}$$

Proof: The proof is straightforward and omitted. ■

Lemma 6.6: Define for all $j \in \{1,2,\cdots,n\}$

$$\begin{aligned} I_j(b) &\triangleq \int_{\mathbf{w^T x}<b} x_j p_{\mathbf{X}}(\mathbf{x})d\mathbf{x}, \forall j \in \{1,2,\cdots,n\} \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}\prod_{i=1}^n \sigma_i}\int_{\mathbf{w^T x}<b} x_j e^{-\frac{1}{2}\mathbf{x^T K_x^{-1} x}}d\mathbf{x} \end{aligned} \tag{19}$$

where $p_{\mathbf{X}}(\cdot)$ is the probability density function of random vector $\mathbf{X}$. Then

$$\begin{aligned} &\left\|E\{\mathbf{X}|\hat{S}=0\}\right\|^2 P(\hat{S}=0)+\left\|E\{\mathbf{X}|\hat{S}=1\}\right\|^2 P(\hat{S}=1) \\ = &\frac{(2\varepsilon-1)^2\sum_{j=1}^n I_j(b)^2}{-(2\varepsilon-1)^2P_0^2+(2\varepsilon-1)P_0+\varepsilon(1-\varepsilon)} \end{aligned} \tag{20}$$

Proof:

From Lemma 6.4, we know that

$$\begin{aligned} &\left\|E\{\mathbf{X}|\hat{S}=0\}\right\|^2 P(\hat{S}=0)+\left\|E\{\mathbf{X}|\hat{S}=1\}\right\|^2 P(\hat{S}=1) \\ = &\sum_{j=1}^n \left(E\{X_j|\hat{S}=0\}\right)^2 P(\hat{S}=0)+\sum_{j=1}^n \left(E\{X_j|\hat{S}=1\}\right)^2 P(\hat{S}=1) \\ = &\sum_{j=1}^n \frac{\left((1-\varepsilon)P_0\bar{X_j}(0)+\varepsilon P_1\bar{X_j}(1)\right)^2}{(1-\varepsilon)P_0+\varepsilon P_1}+\sum_{j=1}^n \frac{\left(\varepsilon P_0\bar{X_j}(0)+(1-\varepsilon)P_1\bar{X_j}(1)\right)^2}{\varepsilon P_0+(1-\varepsilon)P_1} \\ = &(2\varepsilon-1)^2\sum_{j=1}^n \left(P_0\bar{X_j}(0)\right)^2\left(\frac{1}{(1-\varepsilon)P_0+\varepsilon P_1}+\frac{1}{\varepsilon P_0+(1-\varepsilon)P_1}\right) \\ = &(2\varepsilon-1)^2\sum_{j=1}^n \left(P_0\bar{X_j}(0)\right)^2\left(\frac{1}{(1-2\varepsilon)P_0+\varepsilon}+\frac{1}{(2\varepsilon-1)P_0+1-\varepsilon}\right) \\ = &\frac{(2\varepsilon-1)^2}{-(2\varepsilon-1)^2P_0^2+(2\varepsilon-1)P_0+\varepsilon(1-\varepsilon)}\sum_{j=1}^n \left(P_0\bar{X_j}(0)\right)^2 \\ = &\frac{(2\varepsilon-1)^2}{-(2\varepsilon-1)^2P_0^2+(2\varepsilon-1)P_0+\varepsilon(1-\varepsilon)}\sum_{j=1}^n \left(\frac{1}{\sqrt{(2\pi)^n det(\mathbf{K_x})}}\int_{\mathbf{w^T x}<b} x_j e^{-\frac{1}{2}\mathbf{x^T K_x^{-1} x}}d\mathbf{x}\right)^2 \\ = &\frac{(2\varepsilon-1)^2\sum_{j=1}^n I_j(b)^2}{-(2\varepsilon-1)^2P_0^2+(2\varepsilon-1)P_0+\varepsilon(1-\varepsilon)} \end{aligned} \tag{21}$$

The second equation comes from Lemmas 6.4, the third equation from (18), the fourth equation is true because $P_0+P_1=1$, and the sixth equation is due to the definition of $\bar{X_i}(0)$.

■

Since $P_0=\frac{1}{2}$ when $b=0$, by Lemma 6.6, to prove Lemma 6.1 we need to show

$$\begin{aligned} & \frac{\sum_{j=1}^{n} I_j^2(b)}{-(2\varepsilon-1)^2 P_0^2 + (2\varepsilon-1)P_0 + \varepsilon(1-\varepsilon)} \\ \leq & \frac{\sum_{j=1}^{n} I_j^2(0)}{-(2\varepsilon-1)^2 \left(\frac{1}{2}\right)^2 + (2\varepsilon-1)\left(\frac{1}{2}\right) + \varepsilon(1-\varepsilon)} \\ = & \frac{\sum_{j=1}^{n} I_j^2(0)}{\frac{1}{4}} \end{aligned} \tag{22}$$

We have the following lemma regarding $I_j(b)$.

Lemma 6.7: Let $\mathbf{w} = \begin{bmatrix} w_1 & w_2 & \cdots & w_n \end{bmatrix}^T$, $\Omega_{\mathbf{w}}^2 = \sum_{j=1}^{n} w_i^2 \sigma_i^2$

$$I_j(b) = R_j(\mathbf{w}) e^{-\frac{b^2}{2\Omega_{\mathbf{w}}^2}}, \forall j \in \{1, 2, \cdots, n\} \tag{23}$$

where $R_j(\mathbf{w})$ is independent of b.

Proof: Let

$$\mathbf{P}_1 = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1n} \\ P_{21} & P_{22} & \cdots & P_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{n1} & P_{n2} & \cdots & P_{nn} \end{bmatrix}$$

be a rotation matrix such that

$$\mathbf{P}_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix}$$

Then the original random vector $\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix}$ is rotated to $\mathbf{Z} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$. Since x_i's are orthogonal, the variance of z_1, which is in the direction of $\mathbf{w}$ in the original coordinate system and the first basic direction of the transformed coordinate, is the following:

$$\begin{aligned} \sigma_{z_1}^2 &= \Omega_{\mathbf{w}}^2 \\ &= \sum_{j=1}^{n} w_i^2 \sigma_i^2 \end{aligned} \tag{24}$$

The probability density function of random vector $\mathbf{Z}$ now is

$$p_{\mathbf{Z}}(\mathbf{z}) = \frac{1}{(2\pi)^{\frac{n}{2}} det(\mathbf{K_x})} e^{-\frac{1}{2}\mathbf{z^T P_1^T K_x^{-1} P_1 z}}$$

Notice since $\mathbf{P}_1$ is orthogonal, the determinant of the variance matrix is not changed.

Now we look at $I_1(b)$,

$$
\begin{aligned}
I_1(b) &\triangleq \int_{\mathbf{w}^T\mathbf{x}<b} x_1 p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\
&= \int_{\mathbf{w}^T\mathbf{x}<b} \sum_{j=1}^{n} P_{1j} z_j p_{\mathbf{Z}}(\mathbf{z}) d\mathbf{z} \\
&= P_{11} \int_{-\infty}^{b} z_1 dz_1 \int_{-\infty}^{\infty} p_{\mathbf{Z}}(\mathbf{z}) dz_2 \cdots dz_n + \sum_{j=2}^{n} P_{1j} \int_{-\infty}^{b} dz_1 \int_{-\infty}^{\infty} z_j p_{\mathbf{Z}}(\mathbf{z}) dz_2 \cdots dz_n \\
&= \frac{P_{11}}{\sqrt{2\pi}\sigma_{z_1}} \int_{-\infty}^{b} z_1 e^{-\frac{z_1^2}{2\sigma_1^2}} dz_1 + \sum_{j=2}^{n} \frac{P_{1j}}{2\pi\sqrt{\sigma_{z_1}^2\sigma_{z_j}^2 - \sigma_{z_1 z_j}^2}} \int_{-\infty}^{b}\int_{-\infty}^{\infty} z_j e^{-\frac{1}{2}\begin{pmatrix} z_1 & z_j \end{pmatrix}\begin{pmatrix} \sigma_{z_1}^2 & \sigma_{z_1 z_j} \\ \sigma_{z_1 z_j} & \sigma_{z_j}^2 \end{pmatrix}^{-1}\begin{pmatrix} z_1 \\ z_j \end{pmatrix}} dz_j dz_1 \\
&= -P_{11}\sigma_{z_1} e^{-\frac{b^2}{2\sigma_{z_1}^2}} + \sum_{j=2}^{n} \frac{P_{1j}}{2\pi\sqrt{\sigma_{z_1}^2\sigma_{z_j}^2 - \sigma_{z_1 z_j}^2}} \int_{-\infty}^{b}\int_{-\infty}^{\infty} z_j e^{-\frac{1}{2}\begin{pmatrix} z_1 & z_j \end{pmatrix}\begin{pmatrix} \sigma_{z_1}^2 & \sigma_{z_1 z_j} \\ \sigma_{z_1 z_j} & \sigma_{z_j}^2 \end{pmatrix}^{-1}\begin{pmatrix} z_1 \\ z_j \end{pmatrix}} dz_j dz_1
\end{aligned}
$$

Now we analyze the second term. Without loss of generality, let $j = 2$. Define

$$a \triangleq \frac{\sigma_{z_2}}{\sqrt{\sigma_{z_1}^2\sigma_{z_2}^2 - \sigma_{z_1 z_2}^2}} \tag{25}$$

$$f \triangleq \frac{\sigma_{z_1}}{\sqrt{\sigma_{z_1}^2\sigma_{z_2}^2 - \sigma_{z_1 z_2}^2}} \tag{26}$$

$$c \triangleq \frac{-\sigma_{z_1 z_2}}{\sigma_{z_1}^2\sigma_{z_2}^2 - \sigma_{z_1 z_2}^2} \tag{27}$$

and

$$g \triangleq \frac{c}{f} = \frac{-\sigma_{z_1 z_2}}{\sigma_{z_1}\sqrt{\sigma_{z_1}^2\sigma_{z_2}^2 - \sigma_{z_1 z_2}^2}} \tag{28}$$

Notice

$$a^2 - g^2 = \frac{1}{\sigma_{z_1}^2}$$

Then we have

$$
\begin{aligned}
&\int_{-\infty}^{b}\int_{-\infty}^{\infty} z_2 e^{-\frac{1}{2}\begin{pmatrix} z_1 & z_2 \end{pmatrix}\begin{pmatrix} \sigma_{z_1}^2 & \sigma_{z_1 z_2} \\ \sigma_{z_1 z_2} & \sigma_{z_2}^2 \end{pmatrix}^{-1}\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}} dz_2 dz_1 \\
&= \int_{-\infty}^{b}\int_{-\infty}^{\infty} z_2 e^{-\frac{1}{2}\left(a^2 z_1^2 + f^2 z_2^2 + 2c z_1 z_2\right)} dz_2 dz_1 \\
&= \int_{-\infty}^{b} e^{-\frac{a^2 z_1^2}{2}} \int_{-\infty}^{\infty} z_2 e^{-\frac{1}{2}\left(f^2 z_2^2 + 2c z_1 z_2\right)} dz_2 dz_1 \\
&= \int_{-\infty}^{b} e^{-\frac{1}{2}(a^2 - g^2) z_1^2} \int_{-\infty}^{\infty} z_2 e^{-\frac{1}{2}\left(f^2 z_2^2 + 2c z_1 z_2 + g^2 z_1^2\right) + \frac{1}{2} g^2 z_1^2} dz_2 dz_1 \\
&= \int_{-\infty}^{b} e^{-\frac{1}{2}(a^2 - g^2) z_1^2} \int_{-\infty}^{\infty} z_2 e^{-\frac{1}{2}(f z_2 + g z_1)^2} dz_2 dz_1 \\
&= -\frac{\sqrt{2\pi} g}{f^2} \int_{-\infty}^{b} z_1 e^{-\frac{1}{2}(a^2 - g^2) z_1^2} dz_1 \\
&= -\frac{\sqrt{2\pi} g}{f^2} \int_{-\infty}^{b} z_1 e^{-\frac{z_1^2}{2\sigma_{z_1}^2}} dz_1 \\
&= \frac{\sqrt{2\pi} g \sigma_{z_1}}{f^2} e^{-\frac{b^2}{2\sigma_{z_1}^2}} \\
&= r_{12} e^{-\frac{b^2}{2\sigma_{z_1}^2}}
\end{aligned} \tag{29}
$$

where $r_{12} = \frac{\sqrt{2\pi}g\sigma_{z_1}}{f^2}$ is independent of b. Similarly it can be shown that the j-th term in the summation in equation 25 can be written as $r_{1j}e^{-\frac{b^2}{2\sigma_{z_1}^2}}$ where r_{1j} is independent of b. It follows that

$$\begin{aligned} I_1(b) &= \sum_{j=1}^{n} r_{1j} e^{-\frac{b^2}{2\sigma_{z_1}^2}} \\ &\triangleq R_1 e^{-\frac{b^2}{2\Omega_{\mathbf{w}}^2}} \end{aligned} \tag{30}$$

where R_1 is independent of b and can be obtained from equations (25) and (29).

To show $I_j(b) = R_j e^{-\frac{b^2}{2\Omega_{\mathbf{w}}^2}}$, repeat the analysis with

$$\mathbf{P}_1 \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ \vdots \\ \vdots \\ w_n \end{pmatrix}$$

where $\begin{pmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{pmatrix}^T$ is the j-th basis vector. ■

Lemma 6.8:

$$\frac{\sqrt{1-e^{-\varphi^2}}}{2} \geq \frac{1}{\sqrt{2\pi}} \int_0^{\varphi} e^{-\frac{y^2}{2}} dy, \ \forall \varphi \geq 0 \tag{31}$$

where the equality is true when $\varphi = 0$ or $\varphi \to \infty$.

Proof:

See Appendix I. ■

With lemmas 6.3— 6.8, we prove Lemma 6.1 as follows.

- *Proof of Lemma 6.1*

  Lemma 6.7 tells us that

$$\begin{aligned} \sum_{j=1}^{n} I_j^2(b) &= \sum_{j=1}^{n} R_j^2 e^{-\frac{b^2}{\Omega_{\mathbf{w}}^2}} \\ &= F(\mathbf{w}) e^{-\frac{b^2}{\Omega_{\mathbf{w}}^2}} \end{aligned} \tag{32}$$

  where $F(\mathbf{w})$ does not depend on b.

  Hence (22) becomes

$$\frac{\sum_{j=1}^{n} F(\mathbf{w}) e^{-\frac{b^2}{\Omega_{\mathbf{w}}^2}}}{-(2\varepsilon-1)^2 P_0^2 + (2\varepsilon-1)P_0 + \varepsilon(1-\varepsilon)} \leq \frac{\sum_{j=1}^{n} F(\mathbf{w})}{\frac{1}{4}} \tag{33}$$

  Define

$$\alpha \triangleq \frac{1}{\sqrt{2\pi}\Omega_{\mathbf{w}}} \int_0^{b} e^{-\frac{z_1^2}{2\Omega_{\mathbf{w}}^2}} dz_1 \tag{34}$$

  Then

$$\begin{aligned} P_0 &= \frac{1}{\sqrt{2\pi}\Omega_{\mathbf{w}}} \int_{-\infty}^{b} e^{-\frac{z_1^2}{2\Omega_{\mathbf{w}}^2}} dz_1 \\ &= \frac{1}{2} + \alpha \end{aligned}$$

Continuing with (33), since $P_0 = \frac{1}{2} + \alpha$, we get

$$\begin{aligned} & \frac{\sum_{j=1}^{n} F(\mathbf{w}) e^{-\frac{b^2}{\Omega_{\mathbf{w}}^2}}}{-(2\varepsilon-1)^2 P_0^2 + (2\varepsilon-1)P_0 + \varepsilon(1-\varepsilon)} \\ \leq & \frac{\sum_{j=1}^{n} F(\mathbf{w})}{\frac{1}{4}} \\ \Leftrightarrow & -(2\varepsilon-1)^2 P_0^2 + (2\varepsilon-1)P_0 + \varepsilon(1-\varepsilon) \geq \frac{e^{-\frac{b^2}{\Omega_{\mathbf{w}}^2}}}{4} \\ \Leftrightarrow & \frac{1}{4} - (2\varepsilon-1)^2\alpha^2 \geq \frac{e^{-\frac{b^2}{\Omega_{\mathbf{w}}^2}}}{4} \end{aligned}$$

Since $(2\varepsilon-1)^2 \leq 1$ it suffices to prove

$$\frac{1}{4} - \alpha^2 \geq \frac{e^{-\frac{b^2}{\Omega_{\mathbf{w}}^2}}}{4}$$

i.e.

$$\begin{aligned} \alpha &= \frac{1}{\sqrt{2\pi}\Omega_{\mathbf{w}}} \int_0^b e^{-\frac{z_1^2}{2\Omega_{\mathbf{w}}^2}} dz_1 \\ &\leq \frac{\sqrt{1 - e^{-\frac{b^2}{\Omega_{\mathbf{w}}^2}}}}{2} \end{aligned} \tag{35}$$

By Lemma 6.8,

$$\frac{1}{\sqrt{2\pi}} \int_0^{\varphi} e^{-\frac{y^2}{2}} dy \leq \frac{\sqrt{1 - e^{-\varphi^2}}}{2}, \ \forall \varphi \geq 0$$

Putting $y = \frac{z_1}{\Omega_{\mathbf{w}}}$ in (35), the lemma follows.
We prove one more lemma before using Lemma 6.1 to prove Lemma 6.2.
Lemma 6.9: For $w_1 \neq 0$ and $\mathbf{K}$ non-singular,

$$\begin{aligned} & det\left[\frac{1}{\sigma_1^2 w_1^2}\begin{pmatrix} w_2^2 & w_2 w_3 & \cdots & w_2 w_n \\ w_3 w_2 & w_3^2 & \cdots & w_3 w_n \\ \vdots & \vdots & \ddots & \vdots \\ w_n w_2 & w_n w_3 & \cdots & w_n^2 \end{pmatrix} + diag\left(\frac{1}{\sigma_2^2}, \frac{1}{\sigma_3^2}, \cdots, \frac{1}{\sigma_n^2}\right)\right] \\ = & \frac{\sigma_1^2 w_1^2 + \sigma_2^2 w_2^2 + \cdots + \sigma_n^2 w_n^2}{w_1^2 \sigma_1^2 \sigma_2^2 \cdots \sigma_n^2} \\ = & \frac{\sum_{i=1}^{n} \sigma_i^2 w_i^2}{w_1^2 \prod_{i=1}^{n} \sigma_i^2} \end{aligned} \tag{36}$$

Proof:
See Appendix II

■

With Lemma 6.1 and Lemma 6.9, we prove Lemma 6.2 as follows.

- *Proof of Lemma 6.2*
From Lemmas 6.3, 6.6 and Lemma 6.1, we want to maximize the following expression when $b = 0$.

$$
\begin{aligned}
&\frac{\sum_{i=1}^{n}\left(\frac{1}{(2\pi)^{\frac{n}{2}}\prod_{j=1}^{n}\sigma_j}\int_{\mathbf{w^T x}<b} x_i e^{-\frac{1}{2}\sum_{j=1}^{n}\frac{x_j^2}{\sigma_j^2}}d\mathbf{x}\right)^2}{-(2\varepsilon-1)^2P_0^2+(2\varepsilon-1)P_0+\varepsilon(1-\varepsilon)}\\
= &\left(\frac{4}{(2\pi)^{\frac{n}{2}}\prod_{j=1}^{n}\sigma_j}\right)^2\cdot\\
&\sum_{i=1}^{n}\left(\int_{\mathbf{w^T x}<0} x_i e^{-\frac{1}{2}\sum_{j=1}^{n}\frac{x_j^2}{\sigma_j^2}}d\mathbf{x}\right)^2
\end{aligned}
\tag{37}
$$

Look at the term with $i=1$,

$$
\begin{aligned}
&\frac{1}{(2\pi)^{\frac{n}{2}}\prod_{j=1}^{n}\sigma_j}\int_{\mathbf{w^T x}<0} x_1 e^{-\frac{1}{2}\sum_{j=1}^{n}\frac{x_j^2}{\sigma_j^2}}d\mathbf{x}\\
= &\frac{1}{(2\pi)^{\frac{n}{2}}\prod_{j=1}^{n}\sigma_j}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\int_{-\infty}^{-\frac{\sum_{k=2}^{n}w_k x_k}{w_1}} x_1 e^{-\frac{1}{2}\sum_{j=1}^{n}\frac{x_j^2}{\sigma_j^2}}dx_1dx_2\cdots dx_n\\
= &\frac{1}{(2\pi)^{\frac{n}{2}}\prod_{j=1}^{n}\sigma_j}\int_{-\infty}^{\infty}e^{-\frac{x_n^2}{2\sigma_n^2}}\int_{-\infty}^{\infty}e^{-\frac{x_{n-1}^2}{2\sigma_{n-1}^2}}\int_{-\infty}^{\infty}e^{-\frac{x_2^2}{2\sigma_2^2}}\int_{-\infty}^{-\frac{\sum_{k=2}^{n}w_k x_k}{w_1}} x_1 e^{-\frac{x_1^2}{2\sigma_1^2}}dx_1dx_2\cdots dx_n\\
= &-\frac{\sigma_1}{(2\pi)^{\frac{n}{2}}\prod_{j=2}^{n}\sigma_j}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}e^{-\frac{1}{2}\sum_{j=2}^{n}\frac{x_j^2}{\sigma_j^2}}e^{-\frac{\left(\sum_{k=2}^{n}w_k x_k\right)^2}{2\sigma_1^2w_1^2}}dx_2dx_3\cdots dx_n
\end{aligned}
$$

If $w_1=0$, the innermost integration has limits from $-\infty$ to $+\infty$. Since random vector $\mathbf{X}$ is zero mean, we know the integral is 0.

Now assuming $w_1\neq 0$, continuing from above equation we have

$$
\begin{aligned}
&-\frac{\sigma_1}{(2\pi)^{\frac{n}{2}}\prod_{j=2}^{n}\sigma_j}\cdot\\
&\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}e^{-\frac{1}{2}\sum_{j=2}^{n}\frac{x_j^2}{\sigma_j^2}}e^{-\frac{\left(\sum_{k=2}^{n}w_k x_k\right)^2}{2\sigma_1^2w_1^2}}dx_2dx_3\cdots dx_n\\
= &-\frac{(2\pi)^{\frac{n-1}{2}}\sigma_1}{(2\pi)^{\frac{n}{2}}\prod_{j=2}^{n}\sigma_j}\left(det\left[\frac{1}{\sigma_1^2w_1^2}\begin{pmatrix}w_2^2 & w_2w_3 & \cdots & w_2w_n\\ w_3w_2 & w_3^2 & \cdots & w_3w_n\\ \vdots & \vdots & \ddots & \vdots\\ w_nw_2 & w_nw_3 & \cdots & w_n^2\end{pmatrix}+diag\left(\frac{1}{\sigma_2^2},\frac{1}{\sigma_3^2},\cdots,\frac{1}{\sigma_n^2}\right)\right]\right)^{-\frac{1}{2}}\\
= &-\frac{\sigma_1}{(2\pi)^{\frac{1}{2}}\prod_{j=2}^{n}\sigma_j}\frac{w_1\prod_{i=1}^{n}\sigma_i}{\sqrt{\sum_{i=1}^{n}\sigma_i^2w_i^2}} \qquad (38)\\
= &-\frac{\sigma_1^2w_1}{\sqrt{2\pi\sum_{i=1}^{n}\sigma_i^2w_i^2}} \qquad (39)
\end{aligned}
$$

where (38) comes from lemma 6.9.

Similarly we can show the j-th term is $\frac{-\sigma_j^2w_j}{\sqrt{2\pi\sum_{i=1}^{n}\sigma_i^2w_i^2}}$.

Thus (37) becomes

$$
(2\varepsilon-1)^2\frac{2}{\pi}\cdot\frac{\sum_{j=1}^{n}\sigma_j^4w_j^2}{\sum_{j=1}^{n}\sigma_j^2w_j^2} \tag{40}
$$

Without loss of generality, let $\sigma_k=\max\{\sigma_i : w_i\neq 0\}$, then

$$
\frac{\sum_{j=1}^{n}\sigma_j^4w_j^2}{\sum_{j=1}^{n}\sigma_j^2w_j^2}=\sigma_k^2-\frac{\sum_{j=1,j\neq k}^{n}(\sigma_1^2-\sigma_j^2)\sigma_j^2w_j^2}{\sum_{j=2}^{n}\sigma_j^2w_j^2} \tag{41}
$$

The second term in (41) is non-negative. There are two possible cases.

1) $\sigma_1 = \sigma_2 = \cdots = \sigma_n = \sigma$, i.e. the n directions are equally noisy. In this case the mean squared error is constant for all $\mathbf{w}$, thus any $\mathbf{w}$ is optimal.
2) Otherwise, (37) is maximized, i.e. the mean squared estimation error is minimized, if and only if $w_j = 0, \forall j \in \{2,3,\cdots,n\}$. In this case, since $\|\mathbf{w}\| = 1$, we know that $w_1 = 1$, i.e. $\mathbf{w}$ is in the direction with the maximum noise variance.

The lemma is proved.

Remark 6.10: When the encoding is performed according to Lemma 6.2, the minimum MSE for any given binary symmetric channel can be obtained:

1) $\sigma_1 = \sigma_2 = \cdots = \sigma_n = \sigma$, i.e. the n directions are equally noisy.
   The MSE is
   $$\left(n - \frac{2(2\varepsilon-1)^2}{\pi}\right)\sigma^2$$
   When the channel has maximum entropy, i.e. $\varepsilon = \frac{1}{2}$, the minimum MSE is $n\sigma^2$. Therefore no information is transmitted over the channel.
   When the channel is perfect, i.e. $\varepsilon = 0$, the minimum MSE is
   $$\left(n - \frac{2}{\pi}\right)\sigma^2$$
   The reduction in MSE is due to quantization.
2) When not all variances are the same, and $\sigma_i = \max(\sigma_1, \sigma_2, \cdots, \sigma_n)$, i.e. direction i is the most noisy.
   The MSE is
   $$\sum_{j\neq i}\sigma_j^2 + \left(1 - \frac{2(2\varepsilon-1)^2}{\pi}\right)\sigma_i^2$$
   When the channel has maximum entropy, i.e. $\varepsilon = \frac{1}{2}$, the minimum MSE is $\sum_{j=1}^n \sigma_j^2$. Therefore no information is transmitted over the channel.
   When the channel is perfect, i.e. $\varepsilon = 0$, the minimum MSE is
   $$\sum_{j\neq i}\sigma_j^2 + \left(1 - \frac{2}{\pi}\right)\sigma_i^2$$
   The reduction in MSE is due to quantization in the X_i direction.

C. The Optimal Vector Encoder for Binary Symmetric Channel: Correlated Gaussian Noise Case

For an n dimensional zero-mean Gaussian random vector with density

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{det(\mathbf{K_X})}} e^{-\frac{1}{2}\mathbf{x}^T\mathbf{K_X}^{-1}\mathbf{x}}$$

where $\mathbf{K_X}$ is the covariance matrix defined by

$$\mathbf{K_X} \triangleq E\{\mathbf{X}\mathbf{X}^T\} = E\begin{bmatrix} X_1^2 & X_1X_2 & \cdots & X_1X_n \\ X_2X_1 & X_2^2 & \cdots & X_2X_n \\ \vdots & \vdots & \ddots & \vdots \\ X_nX_1 & X_nX_2 & \cdots & X_n^2 \end{bmatrix}$$

If K_x is positive-definite then $\mathbf{K_X}$ can be diagonalized to be

$$\mathbf{M} = diag\{\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2\} = \mathbf{Q^T K_x Q}$$

where $\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2$ are the eigen-values of $\mathbf{K_X}$, with corresponding orthonormal eigen-vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ and $\mathbf{Q} \triangleq \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}$ (See e.g. [11]). $\mathbf{Q}$ is an orthonormal matrix with $\mathbf{Q}^T\mathbf{Q} = \mathbf{Q}\mathbf{Q}^T = \mathbf{I}$, where $\mathbf{I}$ is the unit matrix. We also have $det[\mathbf{K_X}] = det[\mathbf{M}] = \prod_{j=1}^n \sigma_j^2$, $\mathbf{K_X}^{-1} = \mathbf{Q}\mathbf{M}^{-1}\mathbf{Q}^T$, and $det[\mathbf{Q}] = 1$..

Then we have the following theorem.

Theorem 6.11: Let $\mathbf{X} \in \mathbb{R}^n$ and $\mathbf{X} \sim \mathrm{N}(\mathbf{0}_n, \mathbf{K_x})$ and hence $\mathbf{K_X}$ has orthonormal eigen-vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ corresponding to eigen-values $\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2$. Consider encoder $G^*(\cdot)$ defined as below

$$S = G^*(\mathbf{X}) = \begin{cases} Y & \text{if } \mathbf{w}^{*\mathbf{T}}\mathbf{X} \geq 0 \\ 1-Y & \text{otherwise} \end{cases}$$

where $Y \in \{0,1\}$, $\mathbf{w}^* \in \mathbb{R}^n$, $\|\mathbf{w}^*\| = 1$ and w^* is chosen as below. If

1) $\sigma_1 = \sigma_2 = \cdots = \sigma_n = \sigma$, i.e. all the n directions are equally noisy, $\mathbf{w}^*$ is any vector in $\mathbb{R}^n$.
2) Otherwise, let $\sigma_m = \max\{\sigma_1, \sigma_2, \ldots, \sigma_n\}$, with $m \in \{1, 2, \ldots, n\}$. Then $\mathbf{w}^* = \mathbf{v}_m$ is the unit vector in the direction of the m-th eigen-vector.

Then $G^*(\cdot)$ is optimal, i.e., the optimal encoder only encodes the most noisy direction with one bit.

Proof:

By Theorem 4.4, there is an optimal encoder of $\mathbf{X}$ within the class of encoders separating $\mathbb{R}^n$ by a hyperplane through the origin.

Consider random vector $\mathbf{Z} = \mathbf{Q}^T\mathbf{X}$, where $\mathbf{Q} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}$. Then the covariance of $\mathbf{Z}$ satisfies

$$\mathbf{K_Z} \triangleq E\{\mathbf{Z}\mathbf{Z}^T\} = \mathbf{Q}^T\mathbf{K_X}\mathbf{Q} = \mathbf{M} = diag\{\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2\} \tag{42}$$

and the density of $\mathbf{Z}$ is

$$\begin{aligned} f_{\mathbf{Z}}(\mathbf{z}) &= \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{det(\mathbf{K_Z})}} e^{-\frac{1}{2}\mathbf{z}^T\mathbf{K_Z}^{-1}\mathbf{z}} \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}\prod_{j=1}^n \sigma_j} e^{-\frac{1}{2}\mathbf{z}^T\mathbf{M}^{-1}\mathbf{z}} \end{aligned}$$

All the n components of random vector $\mathbf{Z}$ are independent. Lemma 6.2 gives an optimal encoder for such a random vector. Below we will prove there is a 1-to-1 map between the hyperplane encoders of $\mathbf{Z}$ and $\mathbf{X}$ producing the same mean squared error.

Consider the mean squared estimation error caused by the following two encoders

$$S_x = G_x(\mathbf{X}) = \begin{cases} Y & \text{if } \mathbf{w}^{\mathbf{T}}\mathbf{X} \geq 0 \\ 1-Y & \text{otherwise} \end{cases} \tag{43}$$

with $Y \in \{0,1\}$.

and

$$S_z = G_z(\mathbf{Z}) = \begin{cases} Y & \text{if } (\mathbf{Q}^{\mathbf{T}}\mathbf{w})^{\mathbf{T}}\mathbf{Z} = \mathbf{w}^{\mathbf{T}}\mathbf{Q}\mathbf{Z} \geq 0 \\ 1-Y & \text{otherwise} \end{cases} \tag{44}$$

with $Y \in \{0,1\}$.

The encoder $G_z(\cdot)$ encodes the random vector $\mathbf{Z}$, which is rotated from $\mathbf{X}$ by $\mathbf{Q}^{\mathbf{T}}$, with hyper-plane $\mathbf{Q}^{\mathbf{T}}\mathbf{w}$, which is rotated by $\mathbf{Q}^{\mathbf{T}}$ from the hyper-plane $\mathbf{w}$ used by $G_x(\cdot)$.

Let $P_{x0} \triangleq P(S_x = 0)$ and $P_{x1} \triangleq P(S_x = 1)$. From Lemma 6.4, the outputs of the decoder in estimating X when the encoder is $G_x(\cdot)$ are

$$E\{\mathbf{X}|\hat{S}_x = 0\} = \frac{(1-\varepsilon)P_{x0}E\{\mathbf{X}|S_x = 0\} + \varepsilon P_{x1}E\{\mathbf{X}|S_x = 1\}}{(1-\varepsilon)P_{x0} + \varepsilon P_{x1}} \tag{45}$$

and

$$E\{\mathbf{X}|\hat{S}_x = 1\} = \frac{\varepsilon P_{x0}E\{\mathbf{X}|S_x = 0\} + (1-\varepsilon)P_{x1}E\{\mathbf{X}|S_x = 1\}}{\varepsilon P_{x0} + (1-\varepsilon)P_{x1}} \tag{46}$$

where $\varepsilon = P(\hat{S} = 0 \mid S = 1) = P(\hat{S} = 1 \mid S = 0)$. Define $P_{z0} \triangleq P(S_z = 0)$ and $P_{z1} \triangleq P(S_z = 1)$. Again by Lemma 6.4, the outputs of the decoder in estimating $\mathbf{Z}$ when the encoder is $G_z(\cdot)$ are

$$E\{\mathbf{Z}|\hat{S}_z = 0\} = \frac{(1-\varepsilon)P_{z0}E\{\mathbf{X}|S_z = 0\} + \varepsilon P_{z1}E\{\mathbf{Z}|S_z = 1\}}{(1-\varepsilon)P_{z0} + \varepsilon P_{z1}} \tag{47}$$

and

$$E\{\mathbf{Z}|\hat{S}_z = 1\} = \frac{\varepsilon P_{z0}E\{\mathbf{Z}|S_z = 0\} + (1-\varepsilon)P_{z1}E\{\mathbf{Z}|S_z = 1\}}{\varepsilon P_{z0} + (1-\varepsilon)P_{z1}} \tag{48}$$

With encoder as $G_x(\cdot)$, the expectation of $\mathbf{X}$ conditioned on $S_x = 0$ is

$$\begin{aligned} & E\{\mathbf{X}|S_x = 0\} \\ = & \frac{1}{P_{x0}\sqrt{(2\pi)^n det(\mathbf{K_x})}} \int_{\mathbf{w^T x}<0} \mathbf{x} e^{-\frac{1}{2}\mathbf{x^T K_x^{-1} x}} d\mathbf{x} \\ = & \frac{1}{P_{x0}\sqrt{(2\pi)^n}\prod_{i=1}^n \sigma_i} \int_{\mathbf{w^T x}<0} \mathbf{x} e^{-\frac{1}{2}\mathbf{x^T K_x^{-1} x}} d\mathbf{x} \end{aligned}$$

and

$$\begin{aligned} P_{x0} &= \frac{1}{\sqrt{(2\pi)^n det(\mathbf{K_x})}} \int_{\mathbf{w^T x}<0} e^{-\frac{1}{2}\mathbf{x^T K_x^{-1} x}} d\mathbf{x} \\ &= \frac{1}{\sqrt{(2\pi)^n}\prod_{i=1}^n \sigma_i} \int_{\mathbf{w^T x}<0} e^{-\frac{1}{2}\mathbf{x^T K_x^{-1} x}} d\mathbf{x} \end{aligned}$$

On the other hand, when the encoder is $G_z(\cdot)$ the expectation of $\mathbf{Z}$ conditioned on $S_z = 0$ is

$$\begin{aligned} & E\{\mathbf{Z}|S_z = 0\} \\ = & \frac{1}{P_{z0}\sqrt{(2\pi)^n det(\mathbf{M})}} \int_{\mathbf{w^T Q z}<0} \mathbf{z} e^{-\frac{1}{2}\mathbf{z^T M^{-1} z}} d\mathbf{z} \\ = & \frac{1}{P_{z0}\sqrt{(2\pi)^n}\prod_{i=1}^n \sigma_i} \int_{\mathbf{w^T Q z}<0} \mathbf{z} e^{-\frac{1}{2}\mathbf{z^T M^{-1} z}} d\mathbf{z} \end{aligned}$$

and

$$\begin{aligned} P_{z0} &= \frac{1}{\sqrt{(2\pi)^n det(\mathbf{M})}} \int_{\mathbf{w^T Q z}<0} e^{-\frac{1}{2}\mathbf{z^T M^{-1} z}} d\mathbf{z} \\ &= \frac{1}{\sqrt{(2\pi)^n}\prod_{i=1}^n \sigma_i} \int_{\mathbf{w^T Q z}<0} e^{-\frac{1}{2}\mathbf{z^T M^{-1} z}} d\mathbf{z} \end{aligned}$$

Since $\mathbf{Z} = \mathbf{Q^T X}$ and $\mathbf{M} = \mathbf{Q^T K_x Q}$, we know

$$\begin{aligned} P_{z0} &= \frac{1}{\sqrt{(2\pi)^n}\prod_{i=1}^n \sigma_i} \int_{\mathbf{w^T Q z}<0} e^{-\frac{1}{2}\mathbf{z^T M^{-1} z}} d\mathbf{z} \\ &= \frac{1}{\sqrt{(2\pi)^n}\prod_{i=1}^n \sigma_i} \int_{\mathbf{w^T x}<0} e^{-\frac{1}{2}\mathbf{x^T Q M^{-1} Q^T x}} det[Q] d\mathbf{x} \\ &= \frac{1}{\sqrt{(2\pi)^n}\prod_{i=1}^n \sigma_i} \int_{\mathbf{w^T x}<0} e^{-\frac{1}{2}\mathbf{x^T K_x^{-1} x}} d\mathbf{x} \\ &= P_{x0} \end{aligned}$$

since $det[Q] = 1$. Also

$$\begin{aligned} & E\{\mathbf{Z}|S_z = 0\} \\ = & \frac{1}{P_{z0}\sqrt{(2\pi)^n}\prod_{i=1}^n \sigma_i} \int_{\mathbf{w^T Q z}<0} \mathbf{z} e^{-\frac{1}{2}\mathbf{z^T M^{-1} z}} d\mathbf{z} \\ = & \frac{1}{P_{x0}\sqrt{(2\pi)^n}\prod_{i=1}^n \sigma_i} \int_{\mathbf{w^T x}<0} \mathbf{Q^T x} e^{-\frac{1}{2}\mathbf{x^T Q M^{-1} Q^T x}} det[\mathbf{Q}] d\mathbf{x} \\ = & \frac{1}{P_{x0}\sqrt{(2\pi)^n}\prod_{i=1}^n \sigma_i} \int_{\mathbf{w^T x}<0} \mathbf{Q^T x} e^{-\frac{1}{2}\mathbf{x^T K_x^{-1} x}} d\mathbf{x} \\ = & \mathbf{Q^T} E\{\mathbf{X}|S_x = 0\} \end{aligned}$$

Similarly we can prove $E\{\mathbf{Z}|S_z = 1\} = \mathbf{Q^T} E\{\mathbf{X}|S_x = 1\}$ and therefore by (47) and (48)

$$E\{\mathbf{Z}|\hat{S}_z = 0\} = \mathbf{Q^T} E\{\mathbf{X}|\hat{S}_x = 0\}$$

and

$$E\{\mathbf{Z}|\hat{S}_z = 1\} = \mathbf{Q^T} E\{\mathbf{X}|\hat{S}_x = 1\}$$

Moreover,

$$\begin{aligned} & E_{G_z}\left\{ \left(\mathbf{Z}-\hat{\mathbf{Z}}\right)^T \left(\mathbf{Z}-\hat{\mathbf{Z}}\right) \right\} \\ = \; & E_{G_x}\left\{ \left(\mathbf{Q^T X}-\mathbf{Q^T}\hat{\mathbf{X}}\right)^T \left(\mathbf{Q^T X}-\mathbf{Q^T}\hat{\mathbf{X}}\right) \right\} \\ = \; & E_{G_x}\left\{ \left(\mathbf{X}-\hat{\mathbf{X}}\right)^T \mathbf{Q}\mathbf{Q}^T \left(\mathbf{X}-\hat{\mathbf{X}}\right) \right\} \\ = \; & E_{G_x}\left\{ \left(\mathbf{X}-\hat{\mathbf{X}}\right)^T \left(\mathbf{X}-\hat{\mathbf{X}}\right) \right\} \end{aligned}$$

The mean squared error achieved by $G_x(\cdot)$ in estimating $\mathbf{X}$ and the mean squared error achieved by $G_z(\cdot)$ in estimating $\mathbf{Z}$ are the same.

Thus for every encoder of $\mathbf{X}$ of the form (43), we can find an encoder of $\mathbf{Z}$ of the form (44), such that the error in estimating $\mathbf{X}$ and $\mathbf{Z}$ are the same, and vice versa.

Therefore if $\mathbf{w}^*$ in (43) is optimal, $\mathbf{Q^T w}$ in (44) is also optimal, and vice versa.

From Lemma 6.2, if $\sigma_m = \max\{\sigma_1, \sigma_2, \cdots, \sigma_n\}$, an optimal encoder for $\mathbf{Z}$ is in the form (44) with $\mathbf{Q^T w} = \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{bmatrix}^T$, where the 1 is the m-th component. Then an optimal encoder for $\mathbf{X}$ is in the form (43) with

$$\begin{aligned} \mathbf{w} &= Q \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\ &= \mathbf{v}_m \end{aligned}$$

■

VII. Conclusion

We study the real-time estimation of a Markov process over a memoryless noisy digital communication channel to minimize the mean squared estimation error. We first show the optimal encoder can be a function of the current state of the Markov process and the probability mass function of the state of the memory of the receiver given the current state. We then prove the optimal encoder separates the state space with hyper-planes. A recursive algorithm is then given to jointly find the locally optimal encoder and decoder for the special case of the binary symmetric channel and scalar source. For memoryless Gaussian vector source and binary symmetric channel we analytically derive the global joint optimal encoder and decoder. This turns out to be an encoding of the principal component of the source vector. We derive the minimum mean squared error as a function of the variance of source and the channel noise.

Many problems remain open. The recursive relation between the optimal design across time steps needs to be found. The recursive algorithm to find optimal encoder and decoder needs to be generalized to channels other than the binary symmetric channel. For the memoryless Gaussian vector case, we also need to find the optimal designs for more practical channels. The memory state update is given in our problem. The optimal joint design of encoder, decode, and memory update is another interesting problem.

ACKNOWLEDGMENTS

The authors greatly appreciate the valuable suggestions from professor Demosthenis Teneketzis of the University of Michigan, Professor Anant Sahai of the University of California at Berkeley, and Dr. Girish Nair of the University of Melbourne. Dr. Peter Seiler's contribution in the early discussion of the problem was very helpful.

REFERENCES

[1] http://path.berkeley.edu/∼raja/CCWVers4.mpeg.
[2] *IEEE Control Systems Magazine*, 21(1), February 2001.
[3] *IEEE Transactions on Automatic Control*, 49(9), September 2004.
[4] S. Arimoto. An algorithm for calculating the capacity of an arbitrary discrete memoryless channel. *IEEE Trans. Inform Theory*, IT-18:14–20, 1972.
[5] R. Blahut. Computation of channel capacity and rate-distortion function. *IEEE Trans. Information Theory*, IT-18:460–473, 1972.
[6] T. Şimşek and P. Varaiya. Noisy data-rate limited estimation: Renewal codes. *IEEE Conference on Decision and Control*, December 2003.
[7] M. Ergen, D. Lee, , R. Sengupta, and P. Varaiya. Wireless token ring protocol. *IEEE transactions on Vehicular Technology*, 53(6):1863–1881, November 2004.
[8] N. Farvardin and V. Vaishampayan. Optimal quantizer design for noisy channels: An approach to combined source-channel coding. *IEEE Trans. Information Theory*, 33(6):827–838, November 1987.
[9] N. Farvardin and V. Vaishampayan. On the performance and complexity of channel-optimized vector quantizers. *IEEE Trans. Information Theory*, 37(1):155–160, January 1991.
[10] E. Frew, T. McGee, Z. Kim, X. Xiao, S. Jackson, M. Morimoto, R. Rathinam, M. Zennaro, J. Padial, and R. Sengupta. Vision based road-following using a small autonomous aircraft. *IEEE Aerospace Conference*, March 2004.
[11] S. Friedberg, A. Insel, and L. Spence. *Linear Algebra*. Prentice Hall, 4th edition, 2002.
[12] R. Kalman. A new approach to liner filtering and prediction problems. *Transactions of the ASME, Journal of Basic Engineering*, 82D:34–45, March 1960.
[13] R. Kalman and R. Bucy. New results in linear filtering and prediction theory. *Transactions of the ASME, Journal of Basic Engineering*, 83D(1):95–108, March 1961.
[14] A. Kurtenbach and P. Wintz. Quantizing for noisy channels. *IEEE Trans. Communication Technology*, COM-17(2):291–302, April 1969.
[15] S. Lloyd. Least squares quantization in pcm. *IEEE Transactions on Information Theory*, IT-28(2):129–137, March 1982.
[16] A. Mahajan, J. Ko, and R. Sengupta. Distributed probabilistic map service. *Proc. of the 41st IEEE Conference on Decision and Control*, December 2002.
[17] R. Murray, K. Astrom, S. Boyd, R. Brockett, and G. Stein. Future directions in control in an information-rich world. *IEEE Control Systems Magazine*, 23(2):20–33, April 2003.
[18] G. Nair. *State Estimation Under Communication Constraints*. PhD thesis, University of Melbourne, 1999.
[19] G. Nair and R. Evens. State estimation via a capacity-limited communication channel. *Proc. IEEE 36th Conference on Decision and Control*, 1997.
[20] G. Nair and R. Evens. State estimation under bit-rate constraints. *Proc. IEEE 37th Conference on Decision and Control*, 1998.
[21] G. Nair and R. Evens. Structural results for finite bit-rate state estimation. *Proc. 1999 Information, Decision, and Control, Adelaide, Australia*, 1999.
[22] D. Neuhoff and R. Gilbert. Causal source codes. *IEEE Trans. on Information Theory*, IT-28(5):701–713, September 1982.
[23] P. Seiler and R. Sengupta. Analysis of communication losses in vehicle control problems. *American Control Conference*, June 2001.
[24] S. Tatikonda. *Control Under Communication Constraints*. PhD thesis, Massachusetts Institute of Technology, 2000.
[25] D. Teneketzis. On the optimal structure of real-time encoders and decoders in noisy communication. Technical Report Report No. CGR-04-03, Control Group, Department of Electrical Engineering and Computer Science,University of Michigan, Ann Arbor, March 2004.
[26] J. Walrand and P. Varaiya. Optiaml causal coding-decoding problems. *IEEE Trans. Inform. Theory*, IT-29(6):814–820, November 1983.
[27] H. S. Witsenhausen. On the structure of real-time source coders. *The Bell System Technical Journal*, 58(6):1437–1451, July-August 1978.
[28] W. Wong and R. Brockett. Systems with finite communication bandwidth constraints I: State estimation problems. *IEEE Trans. on Automatic Control*, 42(9):1294–1299, September 1997.
[29] Q. Xu, T. Mak, J. Ko, and R. Sengupta. Vehicle-vehicle safety messaging in dsrc. *Proc. of the 1st ACM Workshop on Vehicular Ad-hoc Networks*, October 2004.
[30] Y. Xu and J. Hespanha. Optimal commuication logics in networked control systems. *The 43rd IEEE Conference on Decision and Control*, December 2004.
[31] S. Yuksel and T. Başar. Minimum rate coding for state estiamtion over noiseless channels. *The 43rd IEEE Conference on Decision and Control*, December 2004.

APPENDIX I
PROOF OF LEMMA 6.8

Proof: When $\varphi = 0$, both sides of (31) equal to 0. When $\varphi \to +\infty$, both sides are equal to $\frac{1}{2}$. Hence inequality holds at both 0 and ∞.

We will show for all $\varphi \in (0,\infty)$, the left side of (31) is greater than the right side. Since both sides are positive, (31) is equivalent to

$$V(\varphi) \triangleq \frac{1-e^{-\varphi^2}}{4} - \frac{1}{2\pi}\left(\int_0^{\varphi} e^{-\frac{y^2}{2}} dy\right)^2 \geq 0 \tag{49}$$

We differentiate $V(\varphi)$ with φ and get

$$\frac{dV}{d\varphi} = \frac{\varphi e^{-\varphi^2}}{2} - \frac{1}{\pi} e^{-\frac{\varphi^2}{2}} \int_0^{\varphi} e^{-\frac{y^2}{2}} dy \tag{50}$$

Clearly, $\frac{dV}{d\varphi}$ is zero when either $\varphi = 0$ or $\varphi \to +\infty$. We study its sign in $(0,+\infty)$

For all $\varphi \in (0,\infty)$, define $W(\varphi) \triangleq e^{\frac{\varphi^2}{2}}\left(\frac{dV}{d\varphi}\right)$. Therefore

$$W(\varphi) = \frac{\varphi e^{-\frac{\varphi^2}{2}}}{2} - \frac{1}{\pi}\int_0^{\varphi} e^{-\frac{y^2}{2}} dy \tag{51}$$

$W(\varphi)$ and $\frac{dV}{d\varphi}$ have the same sign when $\varphi > 0$ and is finite. We can study the sign of $\frac{dV}{d\varphi}$ using $W(\varphi)$.

$W(0)=0$ and $W(+\infty)=-\frac{1}{\sqrt{2\pi}}<0$. The first term in (51) first increases from 0 with φ, then decreases until it converges to 0. The second term on the other hand keeps increasing from 0 with φ, so the sign of $W(\varphi)$, therefore the sign of $\frac{dV}{c\varphi}$, must be negative for large φ. Now we analyze the change trend of the sign of $W(\varphi)$ with φ by differentiating it.

$$\frac{dW}{d\varphi}=e^{-\frac{\varphi^2}{2}}\left(\frac{1}{2}-\frac{1}{\pi}-\frac{\varphi^2}{2}\right) \tag{52}$$

Clearly, $\frac{dW}{d\varphi}\geq 0$, if $\varphi\leq\sqrt{1-\frac{2}{\pi}}$. Since $W(0)=0$, we have $W(\varphi)>0$ in $\left(0,\sqrt{1-\frac{2}{\pi}}\right]$. On the other hand, $\frac{dW}{d\varphi}<0$, if $\varphi>\sqrt{1-\frac{2}{\pi}}$. Therefore in $(\sqrt{1-\frac{2}{\pi}},\infty)$ $W(\varphi)$ keeps decreasing, going from positive to negative. Since $W(\varphi)$ and $\frac{dV}{d\varphi}$ have the same sign except for $\varphi\to+\infty$, we know the following about $\frac{dV}{d\varphi}$:

$$\frac{dV}{d\varphi}\begin{cases}=0 & \varphi=0\\ >0 \text{ and increasing} & \varphi\in(0,\sqrt{1-\frac{2}{\pi}}]\\ >0 \text{ and decreasing} & \varphi\in(\sqrt{1-\frac{2}{\pi}},\varphi^*]\\ <0 \text{ and decreasing} & \varphi>\varphi^* \text{ and finite}\\ =0 & \varphi\to+\infty\end{cases}$$

where $\varphi*$ is a finite positive number in $(1-\frac{2}{\pi},+\infty)$ whose exact value is not important to us.

From the sign of $\frac{dV}{d\varphi}$ we can see that $V(\varphi)$ starts from 0 when $\varphi=0$. It first increases then decreases monotonically with φ, and converges to zero when $\varphi\to+\infty$. Hence it can never be negative. The only possibility is $V(\varphi)$ first increases with φ from 0 to be positive , then decreases while still being positive, and eventually goes back to 0. Figure 3 confirms our analysis.

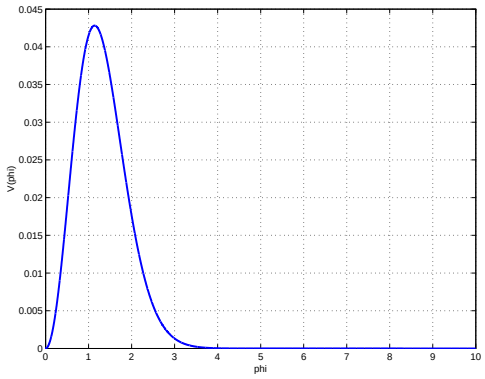

Fig. 3. V as a function of φ

Therefore we have proved that $V(\varphi)\geq 0,\ \forall\varphi\in(0,+\infty)$, and the lemma is proved. ■

APPENDIX II
PROOF OF LEMMA 6.9

Proof: We prove by induction.

When $n=2$, $det\left[\frac{w_2^2}{\sigma_1^2w_1^2}+\frac{1}{\sigma_2^2}\right]=\frac{\sigma_1^2w_1^2+\sigma_2^2w_2^2}{w_1^2\sigma_1^2\sigma_2^2}$. This establishes a base case.

Suppose the lemma holds for all matrices of the above form of size $(k-1)\times(k-1)$, then observe that

$$
\begin{aligned}
& det\left[\frac{1}{\sigma_1^2 w_1^2}\begin{pmatrix} w_2^2 & w_2w_3 & \cdots & w_2w_k \\ w_3w_2 & w_3^2 & \cdots & w_3w_k \\ \vdots & \vdots & \ddots & \vdots \\ w_kw_2 & w_kw_3 & \cdots & w_k^2 \end{pmatrix} + diag\left(\frac{1}{\sigma_2^2},\frac{1}{\sigma_3^2},\cdots,\frac{1}{\sigma_k^2}\right)\right] \\
= & \begin{vmatrix} \frac{w_2^2}{\sigma_1^2w_1^2}+\frac{1}{\sigma_2^2} & \frac{w_2w_3}{\sigma_1^2w_1^2} & \cdots & \frac{w_2w_k}{\sigma_1^2w_1^2} \\ \frac{w_3w_2}{\sigma_1^2w_1^2} & \frac{w_3^2}{\sigma_1^2w_1^2}+\frac{1}{\sigma_3^2} & \cdots & \frac{w_3w_k}{\sigma_1^2w_1^2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{w_kw_2}{\sigma_1^2w_1^2} & \frac{w_kw_3}{\sigma_1^2w_1^2} & \cdots & \frac{w_k^2}{\sigma_1^2w_1^2}+\frac{1}{\sigma_k^2} \end{vmatrix} \\
\triangleq & |A(k)|
\end{aligned}
$$

where $|M|$ stands for the determinant of matrix M.

Then

$$
\begin{aligned}
|A(k)| &= \frac{\sigma_1^2w_1^2+\sigma_2^2w_2^2+\cdots+\sigma_k^2w_k^2}{w_1^2\sigma_1^2\sigma_2^2\cdots\sigma_k^2} \\
&= \frac{\sum_{i=1}^k \sigma_i^2w_i^2}{w_1^2\prod_{i=1}^k\sigma_i^2}
\end{aligned}
$$

Now,

$$
|A(k+1)| = \begin{vmatrix} \frac{w_2^2}{\sigma_1^2w_1^2}+\frac{1}{\sigma_2^2} & \frac{w_2w_3}{\sigma_1^2w_1^2} & \cdots & \frac{w_2w_k}{\sigma_1^2w_1^2} & \frac{w_2w_{k+1}}{\sigma_1^2w_1^2} \\ \frac{w_3w_2}{\sigma_1^2w_1^2} & \frac{w_3^2}{\sigma_1^2w_1^2}+\frac{1}{\sigma_3^2} & \cdots & \frac{w_3w_k}{\sigma_1^2w_1^2} & \frac{w_3w_{k+1}}{\sigma_1^2w_1^2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{w_kw_2}{\sigma_1^2w_1^2} & \frac{w_kw_3}{\sigma_1^2w_1^2} & \cdots & \frac{w_k^2}{\sigma_1^2w_1^2}+\frac{1}{\sigma_k^2} & \frac{w_kw_{k+1}}{\sigma_1^2w_1^2} \\ \frac{w_{k+1}w_2}{\sigma_1^2w_1^2} & \frac{w_{k+1}w_3}{\sigma_1^2w_1^2} & \cdots & \frac{w_{k+1}w_k}{\sigma_1^2w_1^2} & \frac{w_{k+1}^2}{\sigma_1^2w_1^2}+\frac{1}{\sigma_{k+1}^2} \end{vmatrix}
$$

We prove in two cases depending on the value of w_k, i.e. $w_k \neq 0$ and $w_k = 0$.

1) Assume $w_k \neq 0$

Then

$$
\begin{aligned}
& |A(k+1)| \\
= & \begin{vmatrix}
\frac{w_2^2}{\sigma_1^2 w_1^2}+\frac{1}{\sigma_2^2} & \frac{w_2 w_3}{\sigma_1^2 w_1^2} & \cdots & \frac{w_2 w_{k-1}}{\sigma_1^2 w_1^2} & \frac{w_2 w_k}{\sigma_1^2 w_1^2} & \frac{w_2 w_{k+1}}{\sigma_1^2 w_1^2} \\
\frac{w_3 w_2}{\sigma_1^2 w_1^2} & \frac{w_3^2}{\sigma_1^2 w_1^2}+\frac{1}{\sigma_3^2} & \cdots & \frac{w_2 w_{k-1}}{\sigma_1^2 w_1^2} & \frac{w_3 w_k}{\sigma_1^2 w_1^2} & \frac{w_3 w_{k+1}}{\sigma_1^2 w_1^2} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
\frac{w_k w_2}{\sigma_1^2 w_1^2} & \frac{w_k w_3}{\sigma_1^2 w_1^2} & \cdots & \frac{w_k w_{k-1}}{\sigma_1^2 w_1^2} & \frac{w_k^2}{\sigma_1^2 w_1^2}+\frac{1}{\sigma_k^2} & \frac{w_k w_{k+1}}{\sigma_1^2 w_1^2} \\
0 & 0 & \cdots & 0 & -\frac{w_{k+1}}{\sigma_k^2 w_k} & \frac{1}{\sigma_{k+1}^2}
\end{vmatrix} \\
= & \begin{vmatrix}
\frac{w_2^2}{\sigma_1^2 w_1^2}+\frac{1}{\sigma_2^2} & \frac{w_2 w_3}{\sigma_1^2 w_1^2} & \cdots & \frac{w_2 w_{k-1}}{\sigma_1^2 w_1^2} & \frac{w_2 w_k}{\sigma_1^2 w_1^2} & 0 \\
\frac{w_3 w_2}{\sigma_1^2 w_1^2} & \frac{w_3^2}{\sigma_1^2 w_1^2}+\frac{1}{\sigma_3^2} & \cdots & \frac{w_2 w_{k-1}}{\sigma_1^2 w_1^2} & \frac{w_3 w_k}{\sigma_1^2 w_1^2} & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
\frac{w_{k-1} w_2}{\sigma_1^2 w_1^2} & \frac{w_{k-1} w_3}{\sigma_1^2 w_1^2} & \cdots & \frac{w_{k-1}^2}{\sigma_1^2 w_1^2}+\frac{1}{\sigma_{k-1}^2} & \frac{w_{k-1} w_k}{\sigma_1^2 w_1^2} & 0 \\
\frac{w_k w_2}{\sigma_1^2 w_1^2} & \frac{w_k w_3}{\sigma_1^2 w_1^2} & \cdots & \frac{w_k w_{k-1}}{\sigma_1^2 w_1^2} & \frac{w_k^2}{\sigma_1^2 w_1^2}+\frac{1}{\sigma_k^2} & -\frac{w_{k+1}}{\sigma_k^2 w_k} \\
0 & 0 & \cdots & 0 & -\frac{w_{k+1}}{\sigma_k^2 w_k} & \frac{1}{\sigma_{k+1}^2}+\frac{w_{k+1}^2}{\sigma_k^2 w_k^2}
\end{vmatrix} \\
= & \left|\begin{array}{ccc|c}
 & & & 0 \\
 & A(k) & & \vdots \\
 & & & 0 \\
 & & & -\frac{w_{k+1}}{\sigma_k^2 w_k} \\
\hline
0 & \cdots & 0 & \frac{1}{\sigma_{k+1}^2}+\frac{w_{k+1}^2}{\sigma_k^2 w_k^2}
\end{array}\right| \\
= & \left(\frac{1}{\sigma_{k+1}^2}+\frac{w_{k+1}^2}{\sigma_k^2 w_k^2}\right)|A(k)| \\
& -\frac{w_{k+1}^2}{\sigma_k^4 w_k^2}\left|\begin{array}{ccc|c}
 & & & 0 \\
 & A(k-1) & & \vdots \\
 & & & 0 \\
\hline
0 & \cdots & 0 & -\frac{w_{k+1}}{\sigma_k^2 w_k}
\end{array}\right| \\
= & \left(\frac{1}{\sigma_{k+1}^2}+\frac{w_{k+1}^2}{\sigma_k^2 w_k^2}\right)|A(k)|-\frac{w_{k+1}^2}{\sigma_k^4 w_k^2}|A(k-1)| \\
= & \left(\frac{1}{\sigma_{k+1}^2}+\frac{w_{k+1}^2}{\sigma_k^2 w_k^2}\right)\cdot\frac{\sum_{i=1}^{k}\sigma_i^2 w_i^2}{w_1^2\prod_{i=1}^{k}\sigma_i^2}-\frac{w_{k+1}^2}{\sigma_k^4 w_k^2}\cdot\frac{\sum_{i=1}^{k-1}\sigma_i^2 w_i^2}{w_1^2\prod_{i=1}^{k-1}\sigma_i^2} \\
= & \frac{\sum_{i=1}^{k}\sigma_i^2 w_i^2}{w_1^2\prod_{i=1}^{k+1}\sigma_i^2}+\frac{w_{k+1}^2}{\sigma_k^2 w_k^2}\left(\frac{w_k^2\sigma_k^2}{w_1^2\prod_{i=1}^{k}\sigma_i^2}\right) \\
= & \frac{\sum_{i=1}^{k}\sigma_i^2 w_i^2}{w_1^2\prod_{i=1}^{k+1}\sigma_i^2}+\frac{w_{k+1}^2\sigma_{k+1}^2}{w_1^2\prod_{i=1}^{k+1}\sigma_i^2} \\
= & \frac{\sum_{i=1}^{k+1}\sigma_i^2 w_i^2}{w_1^2\prod_{i=1}^{k+1}\sigma_i^2} \qquad (53)
\end{aligned}
$$

2) Assume $w_k = 0$.

Then,

$$
\begin{aligned}
& |A(k+1)| \\
= & \begin{vmatrix}
\frac{w_2^2}{\sigma_1^2 w_1^2}+\frac{1}{\sigma_2^2} & \frac{w_2 w_3}{\sigma_1^2 w_1^2} & \cdots & \frac{w_2 w_k}{\sigma_1^2 w_1^2} & \frac{w_2 w_{k+1}}{\sigma_1^2 w_1^2} \\
\frac{w_3 w_2}{\sigma_1^2 w_1^2} & \frac{w_3^2}{\sigma_1^2 w_1^2}+\frac{1}{\sigma_3^2} & \cdots & \frac{w_3 w_k}{\sigma_1^2 w_1^2} & \frac{w_3 w_{k+1}}{\sigma_1^2 w_1^2} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\frac{w_k w_2}{\sigma_1^2 w_1^2} & \frac{w_k w_3}{\sigma_1^2 w_1^2} & \cdots & \frac{w_k^2}{\sigma_1^2 w_1^2}+\frac{1}{\sigma_k^2} & \frac{w_k w_{k+1}}{\sigma_1^2 w_1^2} \\
\frac{w_{k+1} w_2}{\sigma_1^2 w_1^2} & \frac{w_{k+1} w_3}{\sigma_1^2 w_1^2} & \cdots & \frac{w_{k+1} w_k}{\sigma_1^2 w_1^2} & \frac{w_{k+1}^2}{\sigma_1^2 w_1^2}+\frac{1}{\sigma_{k+1}^2}
\end{vmatrix} \\
= & \begin{vmatrix}
\frac{w_2^2}{\sigma_1^2 w_1^2}+\frac{1}{\sigma_2^2} & \frac{w_2 w_3}{\sigma_1^2 w_1^2} & \cdots & \frac{w_2 w_{k-1}}{\sigma_1^2 w_1^2} & 0 & \frac{w_2 w_{k+1}}{\sigma_1^2 w_1^2} \\
\frac{w_3 w_2}{\sigma_1^2 w_1^2} & \frac{w_3^2}{\sigma_1^2 w_1^2}+\frac{1}{\sigma_3^2} & \cdots & \frac{w_3 w_{k-1}}{\sigma_1^2 w_1^2} & \vdots & \frac{w_3 w_{k+1}}{\sigma_1^2 w_1^2} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \\
\frac{w_{k-1} w_2}{\sigma_1^2 w_1^2} & \frac{w_{k-1} w_3}{\sigma_1^2 w_1^2} & \cdots & \frac{w_{k-1}^2}{\sigma_1^2 w_1^2}+\frac{1}{\sigma_k^2} & 0 & \frac{w_{k-1} w_{k+1}}{\sigma_1^2 w_1^2} \\
0 & \cdots & \cdots & 0 & \frac{1}{\sigma_k^2} & 0 \\
\frac{w_{k+1} w_2}{\sigma_1^2 w_1^2} & \frac{w_{k+1} w_3}{\sigma_1^2 w_1^2} & \cdots & \frac{w_{k+1} w_{k-1}}{\sigma_1^2 w_1^2} & 0 & \frac{w_{k+1}^2}{\sigma_1^2 w_1^2}+\frac{1}{\sigma_{k+1}^2}
\end{vmatrix} \\
= & \frac{1}{\sigma_k^2} \cdot \begin{vmatrix}
\frac{w_2^2}{\sigma_1^2 w_1^2}+\frac{1}{\sigma_2^2} & \frac{w_2 w_3}{\sigma_1^2 w_1^2} & \cdots & \frac{w_2 w_{k-1}}{\sigma_1^2 w_1^2} & \frac{w_2 w_{k+1}}{\sigma_1^2 w_1^2} \\
\frac{w_3 w_2}{\sigma_1^2 w_1^2} & \frac{w_3^2}{\sigma_1^2 w_1^2}+\frac{1}{\sigma_3^2} & \cdots & \frac{w_3 w_{k-1}}{\sigma_1^2 w_1^2} & \frac{w_3 w_{k+1}}{\sigma_1^2 w_1^2} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\frac{w_{k-1} w_2}{\sigma_1^2 w_1^2} & \frac{w_{k-1} w_3}{\sigma_1^2 w_1^2} & \cdots & \frac{w_{k-1}^2}{\sigma_1^2 w_1^2}+\frac{1}{\sigma_k^2} & \frac{w_{k-1} w_{k+1}}{\sigma_1^2 w_1^2} \\
\frac{w_{k+1} w_2}{\sigma_1^2 w_1^2} & \frac{w_{k+1} w_3}{\sigma_1^2 w_1^2} & \cdots & \frac{w_{k+1} w_{k-1}}{\sigma_1^2 w_1^2} & \frac{w_{k+1}^2}{\sigma_1^2 w_1^2}+\frac{1}{\sigma_{k+1}^2}
\end{vmatrix} \\
= & \frac{1}{\sigma_k^2} \cdot \frac{\sigma_1^2 w_1^2+\sigma_2^2 w_2^2+\cdots+\sigma_{k-1}^2 w_{k-1}^2+\sigma_{k+1}^2 w_{k+1}^2}{w_1^2 \sigma_1^2 \sigma_2^2 \cdots \sigma_{k-1}^2 \sigma_{k+1}^2} \\
= & \frac{\sum_{i=1}^{k+1} \sigma_i^2 w_i^2}{w_1^2 \prod_{i=1}^{k+1} \sigma_i^2} \qquad (54)
\end{aligned}
$$

The second last equation is because of the induction assumption, and the last equation is true considering $w_k = 0$. Therefore the Lemma is proved for both cases. ■