Innovations, Ricatti and Kullback-Leibler Divergence

Youngchul Sung
Dept. of Electrical Engineering
KAIST
Daejeon, Korea

Presented at Niigata University, Japan, Oct 22, 2010

This visit and talk is supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2010-0021269).
Research Background

- Signal Processing
  - Statistical Signal Processing (Detection & Estimation)
  - SPTM
  - SAM
  - Adaptive
  - Machine Learning
  - Audio
  - Video
  - Multirate
  - Biomedical
  - ...

- Information Theory

- Networking

- Communications
  - Signal Processing for Communications
Outline

- Introduction
- Neyman-Pearson detection of hidden Gauss-Markov signals
- Optimal sampling for detection
- Application to wireless sensor networks
- Conclusion
Mathematical Engineering

Physical reality
Engineering system
(curiosity or necessity)

Modeling

Mathematical Model
(capturing the essence)

Solve

Equations explaining or predicting system behavior

Occam’s Razor

Occam’s Razor: Pluralitas non est ponenda sine necessitas.
Modern Probability and Statistics

- Lebesgue: Measure theory
- Kolmogorov: Measure-theoretic probability
- Wiener: Probabilistic approach to system theory
- Shannon: Probabilistic approach to communication theory
The Key Problem in Statistics

X \rightarrow \text{Blackbox} \rightarrow Y
The Key Problem in Statistics

P(Y|X)

The Wiener Model
Areas in Statistical Signal Processing

- **Detection theory**
  - Radar setup
  - Digital communication setup: Shannon formulation and following 60,000 Ph.D.’s
  - Applications: Radar, sonar, distributed (team) detection, digital communications

- **Point estimation theory**
  - Applications: Frequency/spectral estimation, DOA estimation, Beamforming, (Blind) parametric system identification

- **Signal tracking**
  - Optimal, adaptive, particle filtering
  - Applications: Radar, sonar, control, digital communications, equalization, negative feedback recovery

- **Time series analysis**
  - Applications: Finance, prediction theory

- **Machine learning**
  - Applications: Pattern recognition, voice recognition, face recognition, financial modeling
Statistical Inference: Detection & Estimation

\[ \Pr\{ X \neq X_{\text{guess}}(Y) \} \]

Deterministic Least Squares

\[ \mathbb{E} \{ \| X - X_{\text{guess}}(Y) \|^2 \} \]

Stochastic Least Squares or Minimum Mean Square Error

<table>
<thead>
<tr>
<th>Neyman</th>
<th>Pearson</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thomas Bayes</td>
<td>(1702 ~ 1762)</td>
</tr>
<tr>
<td>H. Chernoff</td>
<td></td>
</tr>
<tr>
<td>Kullback</td>
<td>Leibler</td>
</tr>
<tr>
<td>Carl Friedrich Gauß</td>
<td>(1777 ~ 1855)</td>
</tr>
<tr>
<td>Claude Shannon</td>
<td></td>
</tr>
<tr>
<td>R. Fisher</td>
<td>(FIM, 1925)</td>
</tr>
<tr>
<td>C. R. Rao</td>
<td>H. Cramer</td>
</tr>
<tr>
<td>Lucien Le Cam</td>
<td></td>
</tr>
<tr>
<td>Norbert Wiener</td>
<td></td>
</tr>
<tr>
<td>Rudolf Kalman</td>
<td></td>
</tr>
<tr>
<td>Bernard Widrow</td>
<td></td>
</tr>
</tbody>
</table>
Statistical Inference: Detection & Estimation

Pr\{ X \neq X_{\text{guess}}(Y) \}

\| X - X_{\text{guess}}(Y) \|^2

Deterministic Least Squares

E \{ || X - X_{\text{guess}}(Y) ||^2 \}

Stochastic Least Squares or Minimum Mean Square Error

State-space model
Introduction

- Neyman-Pearson detection of hidden Gauss-Markov signals
- Optimal sampling for detection
- Application to wireless sensor networks
- Conclusion
Related Work

Detection of correlated random fields
Detection of correlated random fields

Spatial correlation

Measurement noise

Sensor field

Data collection

Sensor deployment ↔ Signal sampling
Correlation and Noise

Different sample correlation

Different SNR

SNR=10dB

SNR= -3dB

SNR=10dB

SNR=-3dB

Design parameter: Number of samples and spacing
Problem Formulation: The **Gauss-Gauss** Problem

**Underlying diffusion phenomenon** (Ornstein-Uhlenbeck process)

\[
\frac{ds(x)}{dx} = -As(x) + Bu(x)
\]

Sampled signal

\[
S_{i+1} = aS_i + u_i
\]

\[
a = e^{-A\Delta}
\]

**Hypotheses**

\[
H_0 : y_i = w_i \quad \text{vs.}
\]

\[
H_1 : \begin{cases}
    s_{i+1} = aS_i + u_i \\
    y_i = s_i + w_i
\end{cases}
\]

**SNR**

\[
\text{SNR} \triangleq \frac{E s_i^2}{E w_i^2}
\]
Hidden Gauss-Markov Model: The State-Space Model

Hidden Gauss-Markov model

\[
\begin{align*}
S_{i+1} &= aS_i + u_i \\
Y_i &= S_i + w_i
\end{align*}
\]

Burg’s Theorem

The maximum-entropy-rate stochastic process \(\{S_i\}\) satisfying the constraints

\[ES_iS_{i+k} = a_k, \quad k = 0, 1, ..., p\]

is the p-th order Gauss-Markov process.

Kalman filtering

Optimal estimation of state using observations is based on the state-space model.
Optimal Detection: Neyman & Pearson

- Likelihood ratio detector

\[ \log \frac{p_{1,n}(y_1, \cdots, y_n)}{p_{0,n}(y_1, \cdots, y_n)} \begin{cases} \geq \tau_n & H_1 \\ < \tau_n & H_0 \end{cases} \]

- Threshold design

Minimize miss probability

\[ P_M = \Pr\{\text{Decision } n = H_0 \mid H_1\} \]

satisfying false alarm probability level

\[ P_F = \Pr\{\text{Decision } n = H_1 \mid H_0\} \leq \alpha \]

(Neyman-Pearson formulation)
Structure for Optimal Detection

- Extension by Middleton, Shepp

\[
\begin{align*}
\mathcal{H}_0 : \quad & y = w \sim \mathcal{N}_c(0, \Sigma_w) \\
\mathcal{H}_1 : \quad & y = s + w \sim \mathcal{N}_c(0, \Sigma_s + \Sigma_w) \\
T(y) &= y^H (\Sigma_w^{-1} - (\Sigma_w + \Sigma_s)^{-1}) y \\
&= y^H (\Sigma_w + \Sigma_s)^{-1} \Sigma_s \Sigma_w^{-1} y = \hat{s}^H \Sigma_w^{-1} y \\
\text{where} \quad & \hat{s} = \Sigma_s (\Sigma_w + \Sigma_s)^{-1} y = Wy
\end{align*}
\]
Schweppe’s Recursion

\[ l_{i-1} \triangleq \log p_1(y_0, \cdots, y_{i-1}) \]

\[ l_i = \log p_1(y_0, \cdots, y_i) \]

\[ l_i = l_{i-1} + \log p_1(y_i | y_0, \cdots, y_{i-1}) \]

\[ p_1(y_i | y_1, \cdots, y_{i-1}) = \frac{1}{\sqrt{2\pi R_{e,i}}} \exp\left(-\frac{1}{2} \frac{(y_i - \hat{y}_{i|i-1})^2}{R_{e,i}}\right) \]

\[ \hat{y}_{i|i-1} \triangleq \mathbb{E}_1(y_i | y_0^{i-1}), \quad \text{linear MMSE estimate of } y_i \]

\[ e_i = y_i - \hat{y}_{i|i-1}, \quad \text{innovation of } y_i \]

\[ R_{e,i} = \mathbb{E}e_i^2, \quad \text{innovation variance.} \]

(Schweppe, 1965)
Performance Analysis:
Known Results - I.I.D. Case

- **Energy detector**

\[ \sum_{i=1}^{n} y_i^2 \begin{cases} \geq \tau_n & H_1 \\ < \tau_n & H_0 \end{cases} \]

- **Performance**

\[ P_M = \Gamma \left[ \frac{n}{2}; \frac{1}{1+\text{SNR}} \right] \Gamma^{-1} \left( \frac{n}{2}; 1 - P_F \right) \]

![Graph showing performance analysis results](image)

- \( P_F = 0.1 \% \)
- \( \text{SNR}=10 \text{ dB} \)
- **Slope** = \(-K\)

\[ \log P_M \approx -nK \]
General Case

- **Challenge**: Exact error probability is **not available**!

\[
P_M \approx e^{-nK}
\]

\[
\log P_M \approx -nK
\]
\[ K \triangleq - \lim_{n \to \infty} \frac{\log P_M}{n} \]

\[ \log P_M \approx -nK \]

**Performance metric for large samples!**
Previous Results on Error Exponents

- I.i.d. observation
  - Stein’s lemma

\[
K = D(p_0 \parallel p_1) = E_0 \log \frac{p_0(Y)}{p_1(Y)} = \int \log \frac{p_0(y)}{p_1(y)} p_0(y)dy
\]

\[
D(N(0, \sigma_0^2) \parallel N(0, \sigma_1^2)) = E_{\sigma_0} \log \frac{1}{\sqrt{2\pi \sigma_0^2}} e^{-\frac{1}{2\sigma_0^2}x^2} = E_{\sigma_0} \left[ \log \frac{\sigma_1^2}{\sigma_0^2} + 1 \left( \frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2} \right)x^2 \right] = -\frac{1}{2} \log \frac{\sigma_0^2}{\sigma_1^2} + \frac{1}{2} \left( \frac{\sigma_0^2}{\sigma_1^2} - 1 \right)
\]
Interpretation via Sanov’s Theorem

\[ \Pr(P_Y \in E) = \min_{P \in E} \exp(-nD(P \| Q)) = \exp(-nD(P^* \| Q)) \]

\[ P_M = \Pr\{\text{Decision } n = H_0 \mid H_1\} \]
Previous Results on Error Exponents

- Asymptotic Kullback-Leibler rate

\[ K = \lim_{n \to \infty} \log \frac{p_{0,n}(y_1, y_2, \ldots, y_n)}{p_{1,n}(y_1, y_2, \ldots, y_n)} \quad \text{under H0} \]
Previous Results on Error Exponents

- Between two Markov or Gauss-Markov observations
  - Koopmans (1960), Hoeffding (1965), Boza (1971), Natarajan (1985)
  - Luschgy (1994)
  - Neyman-Pearson detection between two Gauss-Markov signals

\[ p(y_1, \ldots, y_n) = p(y_1)p(y_2 | y_1)p(y_3 | y_2) \cdots p(y_n | y_{n-1}) \]
Previous Results on Error Exponents

- Markov Gauss-Markov observation
  - Koopmans (1960), Hoeffding (1965), Boza (1971), Natarajan (1985)
  - Luschgy (1994)

- Neyman-Pearson detection between two Gauss-Markov signals

- State-space model: Hidden Markov
  \[ p(y_1, \ldots, y_n) = p(y_1)p(y_2 | y_1)p(y_3 | y_2) \cdots p(y_n | y_{n-1}) \]

**Observation process is not Markov!**
Spectral Domain Approach

Theorem (Itakura-Saito, 1970’s)

\[ K(a, \text{SNR}) = \frac{1}{2\pi} \int_{0}^{2\pi} D(\mathcal{N}(0, S_y^0(\omega)) \parallel \mathcal{N}(0, S_y^1(\omega))) d\omega \]

\[ D(\mathcal{N}(0, S_y^0(\omega)) \parallel \mathcal{N}(0, S_y^1(\omega))) = -\frac{1}{2} \log \frac{S_y^0(\omega)}{S_y^1(\omega)} + \frac{1}{2} \left( \frac{S_y^0(\omega)}{S_y^1(\omega)} - 1 \right) \]
Innovations Process

\[ \{y_1, y_2, y_3\} \iff \{e_1, e_2, e_3\}, \quad e_i \perp e_j, \ i \neq j \]

\[ e_i \triangleq y_i - \hat{y}_{i|i-1,...,1} \]

\[ R_{e,i} \triangleq \mathbb{E}\{e_i^2\} \]

\[ e_i \sim \mathcal{N}(0, R_{e,i}) \]
Innovations Approach to Asymptotic KL rate for the Hidden Gauss-Markov Model

\[
\log p_0(y_0, \cdots, y_{n-1}) = -\frac{1}{2} \sum_{i=0}^{n-1} \left( \log(2\pi\sigma^2) - \frac{y_i^2}{\sigma^2} \right),
\]

\[
\log p_1(y_0, \cdots, y_{n-1}) = -\frac{1}{2} \sum_{i=0}^{n-1} \left( \log(2\pi R_{e,i}) - \frac{e_i^2}{R_{e,i}} \right)
\]
Innovations Approach to Asymptotic KL rate for Hidden Gauss-Markov Model

\[- \log \frac{p_0(y_0, y_1, \cdots, y_{n-1})}{p_1(y_0, y_1, \cdots, y_{n-1})} = \]

\[- \frac{1}{2} \sum_{i=0}^{n-1} \log(2\pi R_{e,i}) - \frac{1}{2} \sum_{i=0}^{n-1} \frac{e_{i}^2}{R_{e,i}} + \frac{1}{2} n \log(2\pi \sigma^2) + \frac{1}{2} \sum_{i=0}^{n-1} \frac{y_{i}^2}{\sigma^2} \]
Error Exponent: Innovations Approach

Theorem (Sung et al. 2004)

\[
K(a, \text{SNR}) \triangleq - \lim_{n \to \infty} \frac{1}{n} \log P_M
\]

\[
= - \frac{1}{2} \log \frac{\sigma^2}{R_e} + \frac{1}{2} \tilde{R}_e - \frac{1}{2}
\]

\[
R_e = \lim_{n \to \infty} \mathbb{E}\{e_i^2 \mid H_1\} = P + \sigma^2
\]

\[
\tilde{R}_e = \lim_{n \to \infty} \mathbb{E}\{e_i^2 \mid H_0\} = \sigma^2 \left(1 + \frac{a^2 P^2}{P^2 + 2\sigma^2 P + (1-a^2)\sigma^4}\right)
\]

\[
P = \frac{1}{2} \sqrt{\sigma^2(1-a^2) - Q}^2 + 4\sigma^2 Q - \frac{1}{2} \sigma^2(1-a^2) + \frac{Q}{2}
\]
Extreme Correlations

**Corollary**

**I.i.d. case** \((a = 0) \iff \text{Stein's lemma} \)

\[
R_e = \Pi_0 + \sigma^2, \quad \tilde{R}_e = \sigma^2
\]

\[
K = D(p_0 \parallel p_1)
\]

\[
p_0 = N(0, \sigma^2), \quad p_1 = N(0, \Pi_0 + \sigma^2)
\]

**Perfectly correlated case** \((a = 1)\)

\[
R_e = \tilde{R}_e = \sigma^2
\]

\[
K = 0
\]

\[
P_M \sim \Theta\left(\frac{1}{\sqrt{n}}\right)
\]
K vs. Correlation Strength

**Theorem (Sung et al.)**

- For SNR \( \geq 1 \), K decreases monotonically as \( a \uparrow 1 \)
- For SNR < 1, there exists a non-zero optimal correlation \( a^* \)

\[
[1 + a^2 + \Gamma(1-a^2)]^2 - 2 \left( r_e + \frac{\Delta}{r_e} \right) = 0
\]

**Optimal Sampling**

- Maximum spacing
- Optimal finite spacing

\[
\Delta^* = -\frac{\log a^*}{A}
\]

\[
a = e^{\frac{-A \Delta^*}{\Delta^*}}
\]
Optimal Correlation vs. SNR

Transition is very sharp!
Intuition: Coherency vs. Diversity

SNR=-3 dB

SNR=10 dB

Error exponent, K

Correlation coefficient, a

Signal notch
K vs. Signal-to-Noise Ratio

Theorem 5:

- K is **monotone increasing** as SNR increases
- K is proportional to $(1/2) \log(1 + \text{SNR})$ at high SNR

![Graph showing K vs. Signal-to-Noise Ratio with a=exp(-1)]
Simulation Results

SNR = 10 dB

SNR = -3 dB
How Many Samples?

\[ P_M \approx P_1 e^{-K(n-1)} \]

\[ \hat{n} = -\frac{1}{K} \log \frac{P_M}{P_1 e^K} \]

Target \( P_M \): \(10^{-4} \)
Target \( P_F \): \(10^{-3} \)
Diffusion rate \( A = 1 \)

![Graph showing required number of sensors vs. sensor spacing with SNR levels of -3 dB, 0 dB, and 3 dB, and optimal and reasonable spacing indicated.]
Vector State Space Model

\[ H_0 : \vec{y}_i = \vec{w}_i, \quad i = 1, 2, \ldots, N \]
\[ H_1 : \vec{y}_i = \vec{s}_i + \vec{w}_i, \quad i = 1, 2, \ldots, N \]
\[ \vec{s}_{i+1} = A \vec{s}_i + B \vec{u}_i \]
\[ \vec{u}_i \overset{i.i.d.}{\sim} \mathcal{N}(0, Q), \quad Q \succeq 0 \]
\[ \vec{s}_1 \sim \mathcal{N}(0, C_0) \]
\[ C_0 = AC_0A^T + BQB^T. \]

\[ \vec{s}_i \triangleq [s_{1i}, s_{2i}, \ldots, s_{Mi}]^T, \quad i = 1, \ldots, N \]
\[ \vec{y}_i \triangleq [y_{1i}, y_{2i}, \ldots, y_{Mi}]^T, \quad i = 1, \ldots, N \]

Vector State Space Model

Theorem (Sung et al.)

\[ K_v = -\frac{1}{2} \log \frac{\sigma^{2M}}{\det(R_e)} + \frac{1}{2} \text{tr} \left( R_e^{-1} \tilde{R}_e \right) - \frac{M}{2} \]

\[
\begin{align*}
R_e & = \sigma^2 I_M + P \\
P & = APA^T + BQB^T - APR_e^{-1} PA^T \quad \text{(Riccati)} \\
K_p & = APR_e^{-1} \\
\tilde{R}_e & = \sigma^2 (I_M + \tilde{P}) \\
\tilde{P} & = (A - K_p)\tilde{P}(A - K_p)^T + K_pK_p^T \quad \text{(Lyapunov)}
\end{align*}
\]

Known in Kalman filter theory
Newly defined quantities

Numerical Results

Extension to d-Dimensional Gauss-Markov Random Fields

Hidden Signal Field on Lattice $\mathcal{I}_n$:

$$Y_{ij} = X_{ij} + W_{ij} \quad i,j \in \mathcal{I}_n \equiv \{i,j : 0 \leq i \leq n, \quad 0 \leq j \leq n\}$$

\[
\begin{align*}
\{ & X_{ij} : \text{a signal field on 2D lattice}, \\
& Y_{ij} : \text{observations.} \}
\end{align*}
\]

Extension to d-Dimensional Gauss-Markov Random Fields

**Theorem (Strong Convergence in 1-D):** If \( S_{y}^{(0)}(\omega) \) and \( S_{y}^{(1)}(\omega) \) have finite lower and upper bounds, and are continuous and strictly positive, we have

\[
\mathcal{K} = \lim_{n \to \infty} \frac{1}{n} \log \left( \frac{p_{0,n}(y)_{n}}{p_{1,n}} \right) \text{ a.s.} \, [p_{0,n}],
\]

\[
= \frac{1}{2\pi} \int_{0}^{2\pi} \left( \frac{1}{2} \log \frac{S_{y}^{(1)}(\omega)}{S_{y}^{(0)}(\omega)} + \frac{S_{y}^{(0)}(\omega)}{2S_{y}^{(1)}(\omega)} - \frac{1}{2} \right) d\omega,
\]

\[
= \frac{1}{2\pi} \int_{0}^{2\pi} D(\mathcal{N}(0, S_{y}^{(0)}(\omega)) \| \mathcal{N}(0, S_{y}^{(1)}(\omega))) d\omega.
\]

**Itakura-Saito (1970’s)**

**Hannon (1973)**

**Theorem (Asymptotic KLI rate in d-D):** Suppose that

A.1 the alternative spectrum \( f_{1}(\omega) \) has a positive lower bound, and

A.2 \( f_{1}(\omega) \) is more than twice differentiable.

Then, the asymptotic KLI rate \( \mathcal{K} \) for (24) is given by

\[
\mathcal{K} = \frac{1}{(2\pi)^{d}} \int_{[-\pi, \pi]^{d}} \left[ \frac{1}{2} \log \frac{(2\pi)^{d} f_{1}(\omega)}{\sigma^{2}} - \frac{1}{2} \left( 1 - \frac{\sigma^{2}}{(2\pi)^{d} f_{1}(\omega)} \right) \right] d\omega,
\]

\[
= \frac{1}{(2\pi)^{d}} \int_{[-\pi, \pi]^{d}} D(\mathcal{N}(0, \sigma^{2}) \| \mathcal{N}(0, (2\pi)^{d} f_{1}(\omega))) d\omega,
\]

where \( D(\cdot \| \cdot) \) denotes the Kullback-Leibler distance.

**Sung et al. (2009)**

Y. Sung, H. V. Poor and H. Yu, "How much information can one get from a wireless ad hoc sensor network over a correlated random field?" IEEE Trans. Infor. Theory, 2009
Conclusion

- The performance of Neyman-Pearson detection of hidden Gauss-Markov signals was analyzed using error exponent.
- Innovations approach to asymptotic Kullback-Leibler rate.
- Sharp transition behavior of error exponent as a function of correlation depending on SNR was proved.
- Connection of Kalman filtering quantities with asymptotic KL rate.
- Extension to vector state-space model.
- Extension to multi-dimensional case: Sufficient condition for strong convergence established.