Mutual Information and MMSE in Gaussian Channels

Dongning Guo[a] Shlomo Shamai[b] Sergio Verdú [a]

[a] Dept. of Electrical Engineering, Princeton University

[b] Dept. of Electrical Engineering, Technion-Israel Institute of Technology

International Symposium on Information Theory
July 1, 2004

Mutual Information & MMSE

- A pair of random objects

$$(X, Y) \sim P_{XY}.$$

- Minimum mean-squared error:

$$\mathsf{mmse}(X|Y) = \min_f \mathsf{E}\,|X - f(Y)|^2 .$$

Achieved by $\mathsf{E}\{X \mid Y\}$.

- Mutual information:

$$I(X;Y) = \mathsf{E}\left\{\log \frac{p_{XY}(X,Y)}{p_X(X)p_Y(Y)}\right\}.$$

Gaussian Channel

$$Y = \alpha \cdot X + N.$$

$$I(X;Y) \longleftrightarrow \mathsf{mmse}(X|Y).$$

Implications & applications.

Gaussian Input

- Scalar channel:

$$Y = \sqrt{\mathsf{snr}} \cdot X + N, \quad N \sim \mathcal{N}(0,1).$$

- If $X \sim \mathcal{N}(0,1)$,

$$I(\mathsf{snr}) = I(X;Y) = \frac{1}{2}\log(1+\mathsf{snr}).$$

$$\mathsf{mmse}(\mathsf{snr}) = \mathsf{mmse}(X|Y) = \frac{1}{1+\mathsf{snr}}.$$

$$\frac{\mathsf{d}}{\mathsf{dsnr}} I(\mathsf{snr}) = \frac{1}{2}\mathsf{mmse}(\mathsf{snr}) \cdot \log e.$$

Binary Input

- Scalar channel:

$$Y = \sqrt{\mathsf{snr}} \cdot X + N.$$

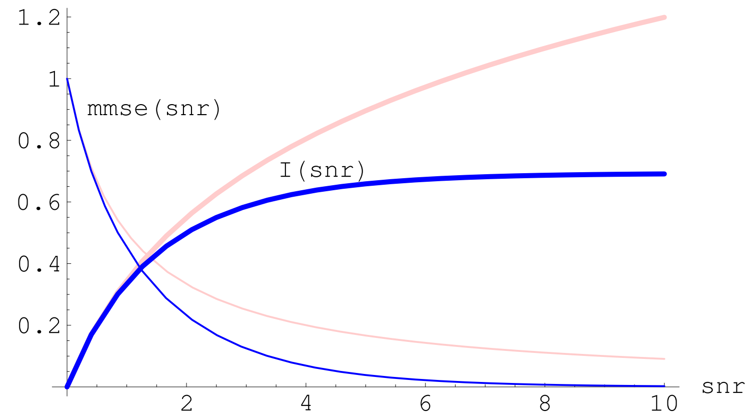
- Binary input: $X = \pm 1$ equally likely.

$$I(\mathsf{snr}) = \mathsf{snr} - \int_{-\infty}^{\infty} \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} \log\cosh(\mathsf{snr} - \sqrt{\mathsf{snr}}\, y)\, \mathrm{d}y,$$

$$\mathsf{mmse}(\mathsf{snr}) = 1 - \int_{-\infty}^{\infty} \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} \tanh(\mathsf{snr} - \sqrt{\mathsf{snr}}\, y)\, \mathrm{d}y.$$

$$\frac{\mathrm{d}}{\mathrm{d}\mathsf{snr}} I(\mathsf{snr}) = \frac{1}{2} \mathsf{mmse}(\mathsf{snr}).$$

$\frac{\mathrm{d}}{\mathrm{d}\mathsf{snr}} I(\mathsf{snr}) = \frac{1}{2} \mathsf{mmse}(\mathsf{snr})$



Gaussian input, binary input.

Central Theorem

- **Theorem 1** *Let* $Y = \sqrt{\mathsf{snr}} \cdot X + N$. $\forall\, P_X$ *with* $\mathsf{E} X^2 < \infty$,

$$\frac{\mathrm{d}}{\mathrm{d}\mathsf{snr}} I(\mathsf{snr}) = \frac{1}{2} \mathsf{mmse}(\mathsf{snr}).$$

- Also true for:
  - Vector channel:

$$\boldsymbol{Y} = \sqrt{\mathsf{snr}} \cdot \boldsymbol{H}\boldsymbol{X} + \boldsymbol{N}.$$

  - Continuous-time:

$$Y_t = \sqrt{\mathsf{snr}} \cdot X_t + N_t, \quad t \in [0, T].$$

  - Discrete-time:

$$Y_i = \sqrt{\mathsf{snr}} \cdot X_i + N_i, \quad i = 1, 2, \ldots$$

Information vs. Estimation

Information theory

$I(X;Y)$

likelihood ratio

detection theory | estimation theory

$\mathsf{mmse}(X|Y)$

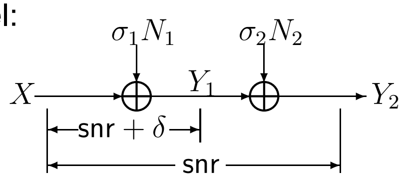
- Wiener, Shannon, Kolmogorov.
- Price, Kailath, '50–'60s: "Estimator-correlator" principle. Esposito '68, Jafar-Gupta '72: Geometry.
- Duncan '70: Mutual information vs. filtering MMSE.
- Kadota-Zakai-Ziv '71: With feedback.

Proof of $\frac{\mathsf{d}}{\mathsf{dsnr}} I(\mathsf{snr}) = \frac{1}{2}\mathsf{mmse}(\mathsf{snr})$

- Equivalent to:

$$I(\mathsf{snr}+\delta) - I(\mathsf{snr}) = \frac{\delta}{2}\cdot \mathsf{mmse}(\mathsf{snr}) + o(\delta).$$

- Incremental channel:



- Chain rule:

$$I(X;Y_1) - I(X;Y_2) = I(X;Y_1 \mid Y_2).$$

- Given X, (Y_1, Y_2) jointly Gaussian:

$$(\mathsf{snr}+\delta)\cdot Y_1 = \mathsf{snr}\cdot Y_2 + \delta \cdot X + \mathcal{N}(0,\delta).$$

Proof of $\frac{\mathsf{d}}{\mathsf{dsnr}} I(\mathsf{snr}) = \frac{1}{2}\mathsf{mmse}(\mathsf{snr})$

- **Lemma 1** *Let* $Y = \sqrt{\delta}\cdot Z + N$. *As* $\delta \to 0$,

$$I(Y;Z) = \frac{\delta}{2}\cdot \mathsf{var} Z + o(\delta).$$

  Verdú ’90, ’02, Lapidoth & Shamai ’02.

- Apply Lemma 1 to $X \to Y_1$ conditioned on Y_2:

$$I(X;Y_1|Y_2) = \frac{\delta}{2}\cdot \mathsf{var}\{X \mid Y_2\} + o(\delta).$$

- Thus,

$$I(\mathsf{snr}+\delta) - I(\mathsf{snr}) = I(X;Y_1|Y_2) = \frac{\delta}{2}\mathsf{mmse}(\mathsf{snr}) + o(\delta).$$

- Key property: $\mathcal{N}(0,\sigma_1^2) + \mathcal{N}(0,\sigma_2^2) \sim \mathcal{N}(0, \sigma_1^2+\sigma_2^2)$.

Continuous-time Channel

- Model:

$$Y_t = \sqrt{\mathsf{snr}}\cdot X_t + N_t, \quad t \in [0,T].$$

- Mutual information rate:

$$I(\mathsf{snr}) = \frac{1}{T} I(X_0^T; Y_0^T).$$

- MMSEs per unit time:

$$\mathsf{cmmse}(\mathsf{snr}) = \frac{1}{T}\int_0^T \mathsf{E}\left(X_t - \mathsf{E}\left\{X_t \mid Y_0^t\right\}\right)^2 \mathsf{d}t,$$

$= \mathsf{cmmse}(t, \mathsf{snr})$

$$\mathsf{mmse}(\mathsf{snr}) = \frac{1}{T}\int_0^T \mathsf{E}\left(X_t - \mathsf{E}\left\{X_t \mid Y_0^T\right\}\right)^2 \mathsf{d}t.$$

Special Input: Gaussian $S_X(\omega)$

- Shannon ’49

$$I(\mathsf{snr}) = \frac{1}{2}\int_{-\infty}^{\infty} \log[1+\mathsf{snr} S_X(\omega)] \frac{\mathsf{d}\omega}{2\pi}.$$

- Wiener ’49

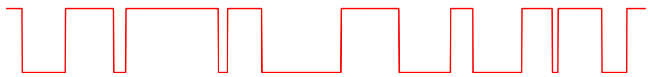
$$\mathsf{mmse}(\mathsf{snr}) = \int_{-\infty}^{\infty} \frac{S_X(\omega)}{1+\mathsf{snr}S_X(\omega)}\frac{\mathsf{d}\omega}{2\pi}.$$

- Yovits-Jackson ’55

$$\mathsf{cmmse}(\mathsf{snr}) = \frac{1}{\mathsf{snr}}\int_{-\infty}^{\infty} \log\left(1+\mathsf{snr}S_X(\omega)\right)\frac{\mathsf{d}\omega}{2\pi}.$$

$$\frac{\mathsf{snr}}{2}\mathsf{cmmse}(\mathsf{snr}) = I(\mathsf{snr}) = \frac{1}{2}\int_0^{\mathsf{snr}} \mathsf{mmse}(\gamma)\,\mathsf{d}\gamma.$$

Special Input: Random Telegraph



- Wonham '65

$$\mathsf{cmmse}(\mathsf{snr}) = \frac{\int_1^\infty u^{-\frac{1}{2}}(u-1)^{-\frac{1}{2}} e^{-\frac{2\nu u}{\mathsf{snr}}}\, \mathrm{d}u}{\int_1^\infty u^{\frac{1}{2}}(u-1)^{-\frac{1}{2}} e^{-\frac{2\nu u}{\mathsf{snr}}}\, \mathrm{d}u}.$$

- Yao '85

$$\mathsf{mmse}(\mathsf{snr}) = \frac{\int_{-1}^{1}\int_{-1}^{1} \frac{(1+xy)\exp\left[-\frac{2\nu}{\mathsf{snr}}\left(\frac{1}{1-x^2}+\frac{1}{1-y^2}\right)\right]}{-(1-x)^3(1-y)^3(1+x)(1+y)}\, \mathrm{d}x\, \mathrm{d}y}{\left[\int_1^\infty u^{\frac{1}{2}}(u-1)^{-\frac{1}{2}} e^{-\frac{2\nu u}{\mathsf{snr}}}\, \mathrm{d}u\right]^2}.$$

$$\mathsf{cmmse}(\mathsf{snr}) = \frac{1}{\mathsf{snr}}\int_0^{\mathsf{snr}} \mathsf{mmse}(\gamma)\, \mathrm{d}\gamma.$$

Triangle Relationship

- All $\{X_t\}$ that satisfies $\int_0^T \mathsf{E}X_t^2\, \mathrm{d}t < \infty$.
- **Theorem 2**

$$\frac{\mathrm{d}}{\mathrm{d}\mathsf{snr}} I(\mathsf{snr}) = \frac{1}{2}\mathsf{mmse}(\mathsf{snr}).$$

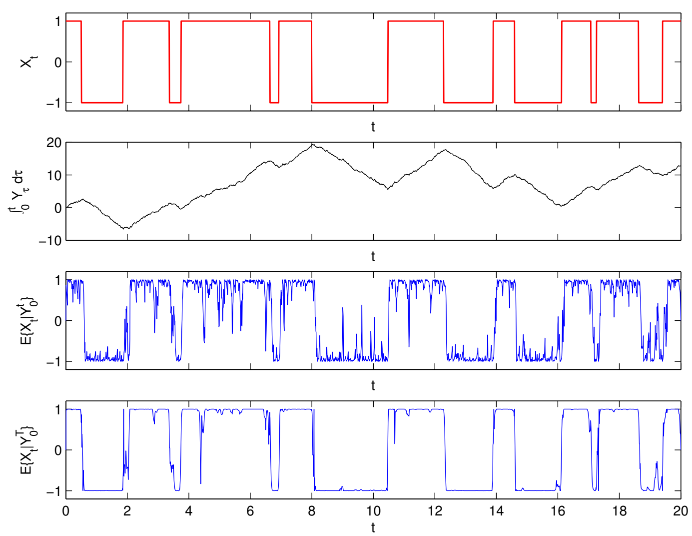
- **Theorem 3 (Duncan '70)**

$$I(\mathsf{snr}) = \frac{\mathsf{snr}}{2} \cdot \mathsf{cmmse}(\mathsf{snr}).$$

- **Theorem 4**

$$\mathsf{cmmse}(\mathsf{snr}) = \frac{1}{\mathsf{snr}}\int_0^{\mathsf{snr}} \mathsf{mmse}(\gamma)\, \mathrm{d}\gamma.$$

Nonlinear Filtering (Telegraph)



Proof: Incremental Channels

- **Lemma 2** *Let $Y_t' = \sqrt{\delta}\cdot X_t' + N_t$. As $\delta \to 0$,*

$$I(\delta) = \frac{\delta}{2T}\int_0^T \mathsf{var}\{X_t'\}\, \mathrm{d}t + o(\delta).$$

- SNR-incremental channel:

$$\begin{aligned} I(\mathsf{snr}+\delta) - I(\mathsf{snr}) &= \frac{1}{T} I\left(X_0^T; Y_{0,\mathsf{snr}+\delta}^T | Y_{0,\mathsf{snr}}^T\right) \\ &= (\delta/2)\cdot \mathsf{mmse}(\mathsf{snr}) + o(\delta). \end{aligned}$$

- Time-incremental channel:

$$\begin{aligned} I\left(X_0^{t+\delta}; Y_0^{t+\delta}\right) - I\left(X_0^t; Y_0^t\right) &= I\left(X_t^{t+\delta}; Y_t^{t+\delta} \,|\, Y_0^t\right) \\ &= (\delta/2)\cdot \mathsf{snr}\cdot \mathsf{cmmse}(t,\mathsf{snr}) + o(\delta). \end{aligned}$$

Representation of Info. Measures

- $\forall X$ with $\mathsf{E}X^2 < \infty$,

$$I\left(X; \sqrt{\mathsf{snr}}X + N\right) = \frac{1}{2}\int_0^{\mathsf{snr}} \mathsf{mmse}\left(X | \sqrt{\gamma}\, X + N\right)\, \mathsf{d}\gamma.$$

- **Theorem 5** $\forall$ *discrete r.v.* X. $\forall$ *1-to-1 mapping* g *to* $\mathbb{R}$,

$$H(X) = \frac{1}{2}\int_0^{\infty} \mathsf{mmse}\left(g(X) | \sqrt{\mathsf{snr}}\, g(X) + N\right)\, \mathsf{dsnr}.$$

- **Theorem 6** $\forall$ *continuous r.v.* X *in* $\mathbb{R}$.

$$\mathsf{D}\left(P_X \| \mathcal{N}(\mathsf{E}X, \mathsf{var}X)\right) = \frac{1}{2}\int_0^{\infty} \frac{\mathsf{var}X}{1 + \mathsf{snr}\cdot\mathsf{var}X} - \mathsf{mmse}(\mathsf{snr})\, \mathsf{dsnr}.$$

General Channel

$$I(X;Y) = I(Z;Y) - I(Z;Y \mid X).$$

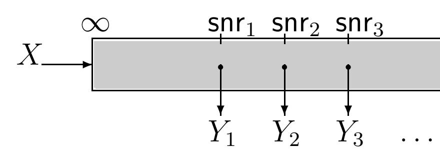
- **Theorem 7**

$$I(X;Y) = \frac{1}{2}\int_0^{\mathsf{snr}} \mathsf{mmse}(Z|Y,X;\gamma) - \mathsf{mmse}(Z|Y;\gamma)\, \mathsf{d}\gamma.$$

Applications & Extensions

- Representation of information measures in estimation errors.
- Key relationship in nonlinear filtering/smoothing.
- Upper bound on MMSE $\Rightarrow$ upper bound on mutual information.
- Large-population CDMA.
- ... ...
- Lévy processes (independent increments):
  - Gaussian channel — $\mathsf{E}\left\{X_t^2 - \widehat{X}_t^2\right\}$;
  - Poisson channel — $\mathsf{E}\left\{\log X_t - \log \widehat{X}_t\right\}$.

Why $\frac{\mathsf{d}}{\mathsf{dsnr}} I(\mathsf{snr}) = \frac{1}{2}\mathsf{mmse}(\mathsf{snr})$?



$$\begin{aligned} I(\mathsf{snr}_1) &= \sum_{n=1}^{\infty} \left[I(X;Y_n) - I(X;Y_{n+1})\right] \\ &= \sum_{n=1}^{\infty} I(X;Y_n | Y_{n+1}) \\ &= \frac{1}{2}\sum_{n=1}^{\infty} \Delta_n \cdot \mathsf{mmse}(\mathsf{snr}_n) + o(\Delta_n) \\ &= \frac{1}{2}\int_0^{\mathsf{snr}_1} \mathsf{mmse}(\gamma)\, \mathsf{d}\gamma. \end{aligned}$$

Vector/Discrete-time Channel

- **Theorem 8** $\boldsymbol{Y} = \sqrt{\mathsf{snr}} \cdot \boldsymbol{H}\boldsymbol{X} + \boldsymbol{N}$. *If* $\mathsf{E}\|\boldsymbol{X}\|^2 < \infty$*,*

$$\frac{\mathsf{d}}{\mathsf{dsnr}} I(\boldsymbol{X};\boldsymbol{Y}) = \frac{1}{2} \cdot \mathsf{E}\,\|\boldsymbol{H}\,\boldsymbol{X} - \boldsymbol{H}\,\mathsf{E}\,\{\,\boldsymbol{X} \mid \boldsymbol{Y}\}\,\|^2.$$

- Discrete-time:

$$Y_i = \sqrt{\mathsf{snr}} \cdot X_i + N_i, \quad i = 1, 2, \dots, n.$$

- **Theorem 9** *If* $\sum_{i=1}^{n} \mathsf{E} X_i^2 < \infty$*, then*

$$\frac{\mathsf{d}}{\mathsf{dsnr}} I(\mathsf{snr}) = \frac{1}{2}\mathsf{mmse}(\mathsf{snr}).$$

- **Theorem 10**

$$\frac{\mathsf{snr}}{2}\mathsf{cmmse}(\mathsf{snr}) \le I(\mathsf{snr}) \le \frac{\mathsf{snr}}{2}\mathsf{pmmse}(\mathsf{snr}).$$

Application: CDMA/MIMO

$$\boldsymbol{Y} = \boldsymbol{H}_{L\times K}\boldsymbol{\Gamma}\boldsymbol{X} + \boldsymbol{N}, \quad \lim K/L = \beta.$$

- Guo-Verdú '03

$$\begin{aligned} \mathsf{C}_{\mathsf{sep}}(\beta) &= \beta \cdot \mathsf{E}\,\{I(\eta\,\mathsf{snr})\}\,, \\ \mathsf{C}_{\mathsf{joint}}(\beta) &= \beta \cdot \mathsf{E}\,\{I(\eta\,\mathsf{snr})\} + (\eta - 1 - \log\eta)/2, \\ \eta^{-1} &= 1 \;+ \beta\,\mathsf{E}\,\{\mathsf{snr} \cdot \mathsf{mmse}(\eta\,\mathsf{snr})\}\,. \end{aligned}$$

- Guo-Verdú '03

$$\mathsf{C}_{\mathsf{joint}}(\beta) = \int_0^{\beta} \frac{1}{\beta'} \mathsf{C}_{\mathsf{sep}}(\beta')\; \mathsf{d}\beta'.$$

  Proof: By $\frac{\mathsf{d}}{\mathsf{dsnr}} I(\mathsf{snr}) = \frac{1}{2}\mathsf{mmse}(\mathsf{snr})$ and fixed-point eqn.
- Interpretation: chain rule. Successive cancellation.