Mutual Information and MMSE in Gaussian Channels

Dongning Guo[2] Shlomo Shamai[3] Sergio Verdú [2]

[2] Dept. of Electrical Engineering
Princeton University
Princeton, NJ 08544, USA

[3] Dept. of Electrical Engineering
Technion-Israel Institute of Technology
32000 Haifa, Israel

Abstract — **Consider arbitrarily distributed input signals observed in additive Gaussian noise. A new fundamental relationship is found between the input-output mutual information and the minimum mean-square error (MMSE) of an estimate of the input given the output: The derivative of the mutual information (nats) with respect to the signal-to-noise ratio (SNR) is equal to half the MMSE. This identity holds for both scalar and vector signals, as well as for discrete- and continuous-time noncausal MMSE estimation (smoothing). A consequence of the result is a new relationship in continuous-time nonlinear filtering: Regardless of the input statistics, the causal MMSE achieved at snr is equal to the expected value of the noncausal MMSE achieved with a channel whose SNR is chosen uniformly distributed between 0 and snr.**

I. Introduction

Consider a scalar additive Gaussian noise model:

$$Y = \sqrt{\mathsf{snr}} \cdot X + W, \tag{1}$$

where $\mathsf{snr} > 0$ denotes the signal-to-noise ratio, and the noise $W \sim \mathcal{N}(0,1)$. Fix an arbitrary input distribution P_X. The minimum mean-square error (MMSE) in estimating the input given the output is achieved by the conditional mean estimator, and is a function of snr:

$$\mathsf{mmse}(\mathsf{snr}) = \mathsf{E}\left\{ (X - \mathsf{E}\{X \mid Y; \mathsf{snr}\})^2 \right\}. \tag{2}$$

The input-output mutual information (in nats) of the channel (1) is also a function of snr, denoted by $I(\mathsf{snr})$.

Theorem 1 *For every P_X with $\mathsf{E}X^2 < \infty$ and every snr,*

$$\frac{\mathrm{d}}{\mathrm{d}\mathsf{snr}} I(\mathsf{snr}) = \frac{1}{2}\mathsf{mmse}(\mathsf{snr}). \tag{3}$$

Surprisingly simple as it is, the identity was previously unknown. In case of standard Gaussian input, (3) is easy to check: $I(\mathsf{snr}) = \frac{1}{2}\log(1+\mathsf{snr})$ and $\mathsf{mmse}(\mathsf{snr}) = 1/(1+\mathsf{snr})$.

The mutual information-MMSE relationship (3) holds for the broadest settings of additive Gaussian noise models and allows the following generalization.

Theorem 2 *Consider an N-dimensional Gaussian channel:*

$$\boldsymbol{Y} = \sqrt{\mathsf{snr}} \cdot \boldsymbol{Z} + \boldsymbol{W} \tag{4}$$

where $\boldsymbol{W}$ is consists of independent standard Gaussian entries. Also fix a random transformation $P_{\boldsymbol{Z}|X}$ where X is an arbitrary random input. Then

$$\frac{\mathrm{d}}{\mathrm{d}\mathsf{snr}} I(X;\boldsymbol{Y}) = \frac{1}{2}\mathsf{E}\left\{ \|\mathsf{E}\{\boldsymbol{Z}|\boldsymbol{Y},X\}\|^2 - \|\mathsf{E}\{\boldsymbol{Z}|\boldsymbol{Y}\}\|^2 \right\} \tag{5}$$

as long as $\mathsf{E}\|\boldsymbol{Z}\|^2 < \infty$.

[1] This work was partially supported by NSF Grants NCR-0074277 and CCR-0312879.

In the special case where $X = \boldsymbol{X}$ is a vector and $\boldsymbol{Z} = \boldsymbol{H}\boldsymbol{X}$ for some fixed matrix $\boldsymbol{H}$, the right hand side of (5) is equal to half the MMSE in estimating $\boldsymbol{H}\boldsymbol{X}$ given $\boldsymbol{Y}$.

Consider a continuous-time channel:

$$Y_t = \sqrt{\mathsf{snr}} \cdot X_t + W_t, \quad t \in [0,T], \tag{6}$$

where W_t is white Gaussian noise with unit spectral density.

Theorem 3 *For every finite-power input process in (6), the identity (3) holds, where*

$$I(\mathsf{snr}) = \tfrac{1}{T} I\left(X_0^T; Y_0^T\right) \tag{7}$$

stands for the mutual information per unit time, and

$$\mathsf{mmse}(\mathsf{snr}) = \int_0^T \mathsf{E}\left\{ \left(X_t - \mathsf{E}\left\{ X_t \mid Y_0^T; \mathsf{snr} \right\} \right)^2 \right\} \frac{\mathrm{d}t}{T} \tag{8}$$

is the average noncausal MMSE per unit time.

Together with Duncan's relationship between the mutual information and the causal MMSE [1], Theorem 3 implies that the causal and non-causal MMSEs determine each other:

Theorem 4 *For every finite power input process in (6),*

$$\mathsf{cmmse}(\mathsf{snr}) = \frac{1}{\mathsf{snr}} \int_0^{\mathsf{snr}} \mathsf{mmse}(\gamma)\, \mathrm{d}\gamma, \tag{9}$$

where cmmse *is similarly defined as (8) with Y_0^t replacing Y_0^T.*

The discrete-time model:

$$Y_n = \sqrt{\mathsf{snr}} \cdot X_n + W_n, \quad n = 1,2,\dots,N, \tag{10}$$

where $W_n \sim \mathcal{N}(0,1)$ are independent, is equivalent to the vector channel (4). Hence the relationship (3) holds in an appropriate form. However, the counterpart to Duncan's relationship is given by inequalities:

Theorem 5 *The mutual information can be bounded:*

$$\mathsf{cmmse}(\mathsf{snr}) \le \frac{I(X^N;Y^N)}{N \cdot \mathsf{snr}} \le \mathsf{pmmse}(\mathsf{snr}) \tag{11}$$

where cmmse *and* pmmse *are the average causal MMSE and one-step prediction MMSE per sample respectively.*

Theorems 1–5 illuminate new connections between information theory and estimation theory. The new relationships as well as Duncan's [1] are proved using a novel idea of "incremental channels", which analyzes the increase in the mutual information due to an infinitesimal increase in the SNR or observation time. The white Gaussian nature of the noise is key to this approach. Applications as well as extensions of these results can be found in [2].

References

[1] T. E. Duncan, "On the calculation of mutual information," *SIAM Journal of Applied Mathematics*, vol. 19, pp. 215–220, July 1970.

[2] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *submitted to IEEE Trans. on Inform. Theory*, 2004.