Sensitivity of Channel Capacity

Mark S. Pinsker, *Member, IEEE*, Vyacheslav V. Prelov, and Sergio Verdú, *Fellow, IEEE*

***Abstract*—In some channels subject to crosstalk or other types of additive interference, the noise is the sum of a dominant Gaussian noise and a relatively weak non-Gaussian contaminating noise. Although the capacity of such channels cannot be evaluated in general, we analyze the decrease in capacity, or sensitivity of the channel capacity to the weak contaminating noise. The main result of this paper is that for a very large class of contaminating noise processes, explicit expressions for the sensitivity of a discrete-time channel capacity do exist. Moreover, in those cases the sensitivity depends on the contaminating process distribution only through its autocorrelation function and so it coincides with the sensitivity with respect to a Gaussian contaminating noise with the same autocorrelation function.**

***Index Terms*— Non-Gaussian channels, channel capacity, water-filling formula, regular process, entropy-singular process, channels with memory.**

I. Introduction

Additive Gaussian noise channels subject to power constraints are one of the few examples in information theory where closed-form formulas for capacity exist. In some applications, such as channels subject to crosstalk or other types of additive interference, it is of interest to evaluate the capacity of channels where the noise consists of the sum of a dominant (so-called nominal) Gaussian noise $\{N_i\}$ and a relatively weak contaminating noise

$$Y_i = X_i + N_i + \theta Z_i \tag{1}$$

where for convenience, we have normalized the power of $\{Z_i\}$ to $\boldsymbol{E}[Z_i^2] = 1$.

Unfortunately, unless $\{Z_i\}$ is also a Gaussian process, no closed-form expressions are feasible for the capacity $C(\theta)$ of channel (1) if $\theta \neq 0$. However, since an explicit expression for $C(0)$ is indeed available, it makes sense to consider the sensitivity of the channel capacity to the weak contaminating noise defined as

$$\boldsymbol{S}_Z = \boldsymbol{S} = \lim_{\theta\to 0} \frac{C(0) - C(\theta)}{\theta^2}. \tag{2}$$

The main result of this paper is that for a very large class of contaminating noise processes, explicit expressions for the sensitivity of channel capacity do exist. Moreover, in those cases, the sensitivity depends on the contaminating process distribution only through its autocorrelation function.

Let us first consider the case where $\{Z_i\}$ is Gaussian, and moreover, let us find the sensitivity of channel capacity in the special case when both $\{N_i\}$ and $\{Z_i\}$ are independent, identically distributed. If the input power is constrained to P, a simple derivative of the Gaussian capacity formula

$$C(\theta) = \frac{1}{2} \ln\left(1 + \frac{P}{\sigma^2 + \theta^2}\right) \tag{3}$$

yields

$$\boldsymbol{S} = \frac{P}{2\sigma^2} \frac{1}{P + \sigma^2} \tag{4}$$

which means that

$$C(\theta) = C(0) - \frac{\theta^2}{2\sigma^2} \frac{P/\sigma^2}{P/\sigma^2 + 1} + o(\theta^2/\sigma^2). \tag{5}$$

The worst case degradation occurs for high signal-to-noise ratio; in which case we lose 1/2 nat per symbol times the relative power of the contaminating noise. Conversely, for vanishing signal-to-noise ratio, the degradation in capacity decreases faster than θ^2.

A nontrivial generalization of the simple formula in (4) is obtained in this paper by dropping the assumption that $\{N_i\}$ and $\{Z_i\}$ are white. In that case, the channel capacity admits the well-known water-filling solution

$$C(\theta) = \frac{1}{2} \int_{-1/2}^{1/2} \ln\left(1 + \frac{[K_\theta - N_0(f) - \theta^2 Z(f)]^+}{N_0(f) + \theta^2 Z(f)}\right) df \tag{6}$$

where $N_0(f)$ and $Z(f)$ are the power spectral densities of the nominal and contaminating noises, respectively, and the water level K_θ is adjusted so that the integral of the optimum input power spectral density $S_\theta(f)$ is equal to P, where

$$S_\theta(f) = [K_\theta - N_0(f) - \theta^2 Z(f)]^+. \tag{7}$$

We show in this paper that the sensitivity of the water-filling channel capacity formula admits the following simple expression:

$$\boldsymbol{S} = \frac{1}{2K_0} \int_{-1/2}^{1/2} Z(f) \frac{S_0(f)}{N_0(f)} df \tag{8}$$

where K_0 is the nominal water level. Thus the sensitivity of Gaussian channel capacity is equal to $(2K_0)^{-1}$ nat times the inner product of the contaminating spectral density and the nominal optimal signal-to-noise spectral density.

It follows that the sensitivity is maximized by a contaminating random process which concentrates its power at those

Manuscript received April 5, 1994; revised May 20, 1995. The material in this paper was presented in part at the 1995 IEEE International Symposium on Information Theory, Whistler, Canada, September 1995. This work was supported in part by the U.S. National Science Foundation under PYI Award ECSE-8857689 and in part by the Russian Fundamental Research Foundation under Grant 93-012-458.

M. S. Pinsker and V. V. Prelov are with the Institute for Problems of Information Transmission of the Russian Academy of Sciences, Moscow 101447, Russia.

S. Verdú is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA.

IEEE Log Number 9414770.

0018-9448/95$04.00

frequencies where the optimal nominal signal-to-noise spectral density is maximum, which corresponds to the case where the nominal noise spectral density is minimum

$$\begin{aligned} \boldsymbol{S} &\leq \frac{1}{2}\left[\frac{1}{\min_f N_0(f)} - \frac{1}{K_0}\right] \\ &\leq \frac{1}{2}\frac{1}{\min_f N_0(f)}. \end{aligned} \tag{9}$$

Note that the latter upper bound is independent of the input power. The worst case sensitivity is minimized over the nominal noise spectral density by white noise, in which case the sensitivity is equal to (4), regardless of the power spectral density of the contaminating process.

Let us now return to the general case of non-Gaussian contaminating noise. Since for a given power-spectral density, Gaussian noise minimizes capacity [1] it follows immediately that (8) is an upper bound to sensitivity for non-Gaussian contamination

$$\boldsymbol{S} \leq \frac{1}{2K_0}\int_{-1/2}^{1/2} Z(f)\frac{S_0(f)}{N_0(f)}\,df. \tag{10}$$

In fact, in the applications that motivate our model (such as twisted pairs channels in subscriber loops), it is customary to evaluate lower bounds on capacity by replacing the additive interference by a Gaussian process with the same spectrum [2] even though the interference is far from Gaussian.

The results of this paper establish that (10) *holds with equality* subject to a mild regularity condition on the contaminating noise, which is satisfied in most situations of practical interest.

It is interesting to point out that the results of [3] imply that the sensitivity of channel capacity does not increase if the decoder assumes that the noise statistics are Gaussian; in fact, the decoder can even assume that the noise is white without decreasing capacity as long as the encoder is allowed to account for such a mismatch.

The rest of the paper is structured as follows. It is useful in our development to use the measure of non-Gaussianness discussed in Section II along with its relationship to mutual information. Section III proves our main result for regular contaminating noise. Section IV shows that the sensitivity is equal to 0 if the contaminating noise is entropy singular. Section V gives a proof of (8) as the derivative of the water-filling capacity.

II. Non-Gaussianness

In this section we find a general decomposition of mutual information that will be useful throughout the paper. We apply this decomposition to the sensitivity problem in its most simple setting which avoids the major technical hurdles we will face in Section III and enables the use of existing results.

Throughout the paper, we will employ the notation $\overline{X}$ to denote a Gaussian random variable with the same mean and variance as the random variable X. The distribution of $\overline{X}$ will be denoted by Φ_X.

We define the non-Gaussianness $D(X)$ of a random variable X with finite variance as its distance (in the sense of divergence) from $\overline{X}$

$$D(X) = D(P_X \| \Phi_X). \tag{11}$$

Obviously, $D(X) \geq 0$ with equality if and only if X is Gaussian. If X has point masses, then $D(X) = +\infty$.

We can express non-Gaussianness as the difference between the differential entropies of $\overline{X}$ and X

$$\begin{aligned} D(X) &= h(\overline{X}) - h(X) \\ &= \frac{1}{2}\ln 2\pi e\sigma_X^2 - h(X) \\ &= \frac{1}{2}\ln\frac{\sigma_X^2}{\eta_X^2} \end{aligned} \tag{12}$$

where η_X^2 is the entropy power of X , i.e., the variance of a Gaussian random variable with the same differential entropy as X .

It is immediate to generalize this concept and define the (joint) non-Gaussianness of several random variables

$$D(X_1, \cdots, X_K) = D(P_{X_1\cdots X_K} \| \Phi_{X_1\cdots X_K}) \tag{13}$$

where $\Phi_{X_1\cdots X_K}$ is the Gaussian density with the same mean and covariance matrix as $(X_1, \cdots, X_K)$. Furthermore, we can define the conditional non-Gaussianness of X given Y using the conditional divergence [4]

$$D(X|Y) = D(P_{X|Y} \| \Phi_{X|Y} | P_Y) \tag{14}$$

where $\Phi_{X|Y}(\cdot|y)$ denotes the conditional density of $\overline{X}$ given $\overline{Y} = y$, where $(\overline{X}, \overline{Y})$ are jointly Gaussian random variables with the same means, variances, and correlation coefficient as (X, Y).

It is easy to see that non-Gaussianness satisfies the telescoping property

$$D(X_1, \cdots, X_K) = \sum_{i=1}^{K} D(X_i | X_{i-1}, \cdots, X_1). \tag{15}$$

Note that

$$\int \Phi_{X|Y}(\cdot|y)\,dP_Y(y)$$

does not necessarily equal $\Phi_X(\cdot)$, and, consequently, we cannot claim that $D(X|Y) \geq D(X)$. However, the difference between conditional and unconditional non-Gaussianness does admit an interesting representation:

$$\begin{aligned} D(X|Y) - D(X) &= \boldsymbol{E}\left[\ln\frac{P_{X|Y}(X|Y)}{\Phi_{X|Y}(X|Y)}\frac{\Phi_X(X)}{P_X(X)}\right] \\ &= I(X;Y) - \boldsymbol{E}\left[\ln\frac{\Phi_{X|Y}(X|Y)}{\Phi_X(X)}\right] \end{aligned} \tag{16}$$

where the expectations are with respect to the joint distribution of X and Y. The second term in the right-hand side of (16) would coincide with the mutual information if X and Y were jointly Gaussian; in fact, it depends on the joint distribution of X and Y only through their correlation coefficient. We will refer to this quantity as the *second-order information* between X and Y

$$I_2(X;Y) = I(\overline{X};\overline{Y}) = \boldsymbol{E}\left[\ln \frac{\Phi_{X|Y}(X|Y)}{\Phi_X(X)}\right] = \frac{1}{2}\ln\frac{1}{1-\varrho_{XY}^2} \tag{17}$$

where ϱ_{XY} is the correlation coefficient between X and Y

$$\varrho_{XY} = \frac{\boldsymbol{E}[XY] - \boldsymbol{E}[X]\boldsymbol{E}[Y]}{\sigma_X \sigma_Y}. \tag{18}$$

Thus we have the decomposition of mutual information as

$$I(X;Y) = I_2(X;Y) + D(X|Y) - D(X) = I_2(X;Y) + D(Y|X) - D(Y) \tag{19}$$

where we have used the symmetry of $I(X;Y)$ and $I_2(X;Y)$ to write the last equation.

The second-order information is obviously nonnegative; it may be greater or smaller than mutual information. For example, if X is not Gaussian and N is Gaussian and independent of X, then

$$I(N;N+X) > I_2(N;N+X) \tag{20}$$

and

$$I(X;N+X) < I_2(X;N+X). \tag{21}$$

Note also that the decomposition in (19) is not useful for discrete random variables as in that case $D(X|Y) - D(X) = \infty - \infty$.

The second-order information can be generalized to any number of random variables

$$I_2(X_1,\cdots,X_n;Y_1,\cdots,Y_m)$$

using the formula in [5, Th. 9.2.1]. For the purposes of our results, it is important to use (19) in the special case where $Y = X + V$ and X and V are independent. Since $D(P_{V+a}\|\Phi_{V+a})$ is constant in a

$$D(X+V|X) = D(V).$$

Therefore, it follows from decomposition (19) that

$$I(X;X+V) \le I(\overline{X};\overline{X}+\overline{V}) + D(V) \tag{22}$$

a result anticipated in a continuous-time context by Ihara [6] and which actually goes back to Shannon's 1948 paper [7] in a different guise.

It is well known [8] that among all V with the same variance, $I(\overline{X};\overline{X}+V)$ is minimized by $\overline{V}$

$$I(\overline{X};\overline{X}+\overline{V}) \le I(\overline{X};\overline{X}+V). \tag{23}$$

This result follows immediately from decomposition (19) upon noting the inequality (from the divergence data-processing theorem)

$$D(\overline{X}+V) \le D(V) = D(\overline{X}+V|\overline{X}). \tag{24}$$

In our setting, we are interested in the case where V is almost Gaussian: $V = N + \theta Z$, with N a Gaussian random variable. The bounds in (22) and (23) immediately provide the following bounds for the sensitivity of the memoryless channel (1):

$$\frac{P}{2\sigma^2}\frac{1}{P+\sigma^2} - \lim_{\theta\to 0}\frac{D(N+\theta Z)}{\theta^2} \le \boldsymbol{S} \le \frac{P}{2\sigma^2}\frac{1}{P+\sigma^2}. \tag{25}$$

The following result proves that the nuisance term in (25) is zero, and, thus the exact formula (4) for the sensitivity of the capacity of the memoryless nominally Gaussian channel is established.

Theorem 1: If N is Gaussian, Z has finite variance and N and Z are independent, then

$$D(N+\theta Z) = o(\theta^2). \tag{26}$$

Proof: Define the auxiliary random variable

$$S = \begin{cases} 1, & \text{if } |Z| \le M \\ 0, & \text{otherwise} \end{cases} \tag{27}$$

for some constant M. Consider the following lower bound on differential entropy:

$$\begin{aligned} h(N+\theta Z) &\ge h(N+\theta Z|S) \\ &= \boldsymbol{P}[S=1]h(N+\theta Z|S=1) \\ &\quad + \boldsymbol{P}[S=0]h(N+\theta Z|S=0) \\ &\ge \boldsymbol{P}[S=1]h(N+\theta Z|S=1) \\ &\quad + \boldsymbol{P}[S=0]h(N). \end{aligned} \tag{28}$$

Thus (13) yields (assuming without loss of generality that Z has unit variance)

$$\begin{aligned} D(N+\theta Z) &\le \frac{1}{2}\ln 2\pi e(\sigma^2+\theta^2) \\ &\quad - \boldsymbol{P}[S=1]h(N+\theta Z|S=1) \\ &\quad - \boldsymbol{P}[S=0]\frac{1}{2}\ln 2\pi e\sigma^2. \end{aligned} \tag{29}$$

The remaining differential entropy is taken care of using the following result.

Lemma 1 [9]: If N is Gaussian and independent of Y such that

$$\boldsymbol{E}\left[|Y|^{2+\alpha}\right] \le K < \infty$$

for some $\alpha > 0$, then

$$h(N+\theta Y) = \frac{1}{2}\ln 2\pi e\sigma_N^2 + \frac{1}{2}[\theta^2 + o(\theta^2)]\frac{\sigma_Y^2}{\sigma_N^2}\theta \to 0 \tag{30}$$

where $o(\theta^2)$ depends only on K.

Using Lemma 1 in order to evaluate $h(N+\theta Z|S=1)$ in (29) yields

$$\begin{aligned} D(N+\theta Z) &\le \frac{1}{2}\ln\left(1+\frac{\theta^2}{\sigma^2}\right) \\ &\quad - \frac{1}{2}[\theta^2 + o(\theta^2)]\frac{\text{var}\,(Z|S=1)}{\sigma^2}\boldsymbol{P}[S=1]. \end{aligned} \tag{31}$$

But because of the assumption that Z has finite (unit) variance, we can choose M so large that for any given $\delta > 0$

$$P[S = 1] \geq 1 - \delta$$
$$\text{var}\,(Z|S = 1) \geq 1 - \delta$$

which implies that

$$D(N + \theta Z) \leq \frac{\theta^2}{\sigma^2}\delta + o(\theta^2)$$

and the result follows because δ is arbitrarily small. □

Some intuition about the nature of Theorem 1 can be gained as follows. It is well known [4] that divergence between neighboring distributions can be expanded as

$$D(Q_\theta \| Q_o) = \frac{1}{2} I(0)\theta^2 + o(\theta^2)$$

where $I(\theta)$ is the Fisher information

$$I(\theta) = \boldsymbol{E}\left[\left(\frac{\partial}{\partial\theta} \ln Q_\theta(X)\right)^2\right]$$

where the expectation is with respect to Q_θ. If $Q_\theta = P_{N+\theta Z}$, it turns out that $I(0) = 0$.

We have made use of the decomposition of mutual information in (19) in order to derive (25), which as we have seen leads to sensitivity formulas in simple cases. Other applications of (19) will be fruitfully used in the next section.

III. Regular Contaminating Noise

The capacity of the channel with colored noise can be written as the limit of the capacities of memoryless vector channels

$$C(\theta) = \lim_{n\to\infty} \frac{1}{n} C_n(\theta) \tag{32}$$

where $C_n(\theta)$ is the capacity of the memoryless channel

$$\boldsymbol{Y}_i^n = \boldsymbol{X}_i^n + \boldsymbol{N}_i^n + \theta \boldsymbol{Z}_i^n \tag{33}$$

where the average power of the input codewords is constrained to be less than or equal to P, and $\{\boldsymbol{Z}_i^n\}$ and $\{\boldsymbol{N}_i^n\}$ have the distributions of the consecutive noise samples $(N_1, \cdots, N_n)$ and $(Z_1, \cdots, Z_n)$. If $\overline{C}_n(\theta)$ denotes the capacity of (33) where P_{Z^n} is substituted by Φ_{Z^n}, then (22) and (23) yield the well-known result [6]

$$\overline{C}_n(\theta) \leq C_n(\theta) \leq \overline{C}_n(\theta) + D(\boldsymbol{N}^n + \theta \boldsymbol{Z}^n). \tag{34}$$

This implies that we could conclude that the sensitivity is equal to the sensitivity of the channel with contaminating Gaussian noise (i.e., the sensitivity of the water-filling formula which is found in Section V), if we could show that

$$\lim_{\theta\to 0} \lim_{n\to\infty} \frac{1}{\theta^2 n} D(\boldsymbol{N}^n + \theta \boldsymbol{Z}^n) = 0. \tag{35}$$

Unfortunately, it is not possible to directly interchange the limits of (35) and the sought-after result is far from a direct consequence of Lemma 1.

The key sufficient condition on the contaminating noise in our proof of (35) is that it is *regular*:

Definition 1 (cf. [10]): A random sequence $\{Z_i\}$ is called *regular*, if the σ-algebra

$$\bigcap_t \mathcal{A}_Z(-\infty, t)$$

is trivial, i.e., if it contains only events of probability 0 or 1, where $\mathcal{A}_Z(-\infty, t)$ is the minimal σ-algebra containing all events

$$\{Z_{i_1} \in E_1, \cdots, Z_{i_s} \in E_s\}$$

for any integers $i_1, \cdots, i_s \in (-\infty, t)$ and any Borel sets E_j of real line.

Regularity is one of the weakest conditions in the literature for capturing the weakening of dependence between past and present events as their time separation increases.

It is well known [11, ch. IV, sec. 1] that a stationary Gaussian random sequence $\{N_i\}$ is regular if and only if it has a spectral density $N_0(f)$ and

$$\int_{-1/2}^{1/2} \ln N_0(f) df > -\infty. \tag{36}$$

Moreover, condition (36) means that $C(0)$ is finite.

The purpose of this section is to show the following general result on the invariance of sensitivity to non-Gaussian noise.

Theorem 2: Assume that the contaminating noise is a regular, stationary, second-order process. Then (35) holds under either of the following conditions:

i) $\{N_i\}$ is i.i.d. Gaussian.

ii) $\{N_i\}$ is a regular Gaussian stationary process and the ratio of spectral densities of contaminating to nominal noises: $Z(f)/N_0(f)$ is bounded on $\left[0, \frac{1}{2}\right]$.

Proof: We shall denote

$$V_i = N_i + \theta Z_i. \tag{37}$$

Accordingly, the goal is to show

$$\lim_{\theta\to 0} \lim_{n\to\infty} \frac{1}{n\theta^2} D(V_1, \cdots, V_n) = 0. \tag{38}$$

Using property (15), we can write

$$\frac{1}{n} D(V_1, \cdots, V_n) = \frac{1}{n} \sum_{j=1}^{n} D(V_j | V_{j-1}, \cdots, V_1) \tag{39}$$

and each of those terms can be decomposed via (19) into

$$\begin{aligned} D(V_j|V_{j-1}, \cdots, V_1) = {} & I(V_1, \cdots, V_{j-1}; V_j) \\ & - I_2(V_1, \cdots, V_{j-1}; V_j) + D(V_j). \end{aligned} \tag{40}$$

We will deal first with case i) of Theorem 2 where $\{N_i\}$ is i.i.d. Using the nonnegativity of the second-order information, and the stationarity of $\{V_j\}$ it will be enough for our purposes to show

$$D(N_1 + \theta Z_1) = o(\theta^2) \tag{41}$$

and

$$\lim_{j\to\infty} I(N_1 + \theta Z_1; N_0 + \theta Z_0, \cdots, N_{-j} + \theta Z_{-j}) = o(\theta^2) \tag{42}$$

as $\theta \to 0$.

Since (41) has been shown in Theorem 1, the additional complexity over the memoryless case is brought about by the need to show (42). A main difficulty in the proof of (42) is that θ appears in both of the arguments of mutual information.

We shall prove a statement stronger than (42)

$$I(N_1+\theta Z_1; N_0+\theta Z_0,\cdots,N_{-j}+\theta Z_{-j})=o(\theta^2) \quad (43)$$

where $o(\theta^2)/\theta^2 \to 0$ uniformly with respect to all integers j as $\theta \to 0$.

Let us denote

$$U_j(\theta)=(N_0+\theta Z_0,\cdots,N_{-j}+\theta Z_{-j}). \quad (44)$$

It follows from (12) that

$$\begin{aligned}I(N_1+\theta Z_1;U_j(\theta)) &= h(N_1+\theta\overline{Z}_1)-D(N_1+\theta Z_1)\\&\quad -h(N_1+\theta Z_1|U_j(\theta))\\&=\frac{1}{2}\ln 2\pi e\sigma^2+\frac{\theta^2}{2\sigma^2}\\&\quad -h(N_1+\theta Z_1|U_j(\theta))+o(\theta^2) \quad (45)\end{aligned}$$

using the result of Theorem 1. In order to lower-bound $h(N_1+\theta Z_1|U_j(\theta))$ we introduce again the auxiliary random variable (cf. (27))

$$S=\begin{cases}1, & \text{if } |Z_1|\le M\\ 0, & \text{otherwise}\end{cases} \quad (46)$$

for some constant M and will proceed analogously to the proof of Theorem 1.

$$\begin{aligned}h(N_1+\theta Z_1|U_j(\theta)) &\ge h(N_1+\theta Z_1|U_j(\theta),S)\\&= \boldsymbol{P}[S=1]h(N_1+\theta Z_1|U_j(\theta),S=1)\\&\quad +\boldsymbol{P}[S=0]h(N_1+\theta Z_1|U_j(\theta),S=0)\\&\ge \boldsymbol{P}[S=1]h(N_1+\theta Z_1|U_j(\theta),S=1)\\&\quad +\boldsymbol{P}[S=0]h(N_1) \quad (47)\end{aligned}$$

because

$$\begin{aligned}&h(N_1+\theta Z_1|U_j(\theta),S=0)\\&\qquad \ge h(N_1|U_j(\theta),S=0)=h(N_1). \quad (48)\end{aligned}$$

According to Lemma 1 for every realization of $U_j(\theta)=u$

$$\begin{aligned}h(N_1+\theta Z_1|U_j(\theta)=u,S=1)&=\frac{1}{2}\ln 2\pi e\sigma^2\\&\quad+\frac{\theta^2}{2\sigma^2}\text{var }(Z_1|U_j(\theta)=u,S=1)\\&\quad+o(\theta^2) \quad (49)\end{aligned}$$

where $o(\theta^2)$ depens only on M, because for any $\alpha>0$

$$\boldsymbol{E}[Z_1^{2+\alpha}|U_j(\theta)=u,S=1]$$

is uniformly bounded with respect to all u and j.

Putting together (45), (47), and (49) we obtain

$$\begin{aligned}&I(N_1+\theta Z_1;U_j(\theta))\\&\le\frac{\theta^2}{2\sigma^2}[1-\boldsymbol{P}[S=1]\text{var }(Z_1|U_j(\theta),S=1)]+o(\theta^2). \quad (50)\end{aligned}$$

It is evident that

$$\lim_{M\to\infty}\boldsymbol{P}[S=1]=1. \quad (51)$$

Thus in order to show that the right-hand side of (50) is $o(\theta^2)$, all we need to show is that the conditional variance of Z_1 given $U_j(\theta)$ and $S=1$ converges to the unconditional variance, which is equal to unity, uniformly with respect to all integers j as $\theta\to 0$ and $M\to\infty$. This is where regularity of $\{Z_i\}$ comes into play.

Lemma 2: If $\{Z_i\}$ is a regular, stationary, second-order process, and $\{N_i\}$ is i.i.d. Gaussian, then

$$\text{var }(Z_1|U_j(\theta),S=1)\to\text{var } Z_1 \quad (52)$$

uniformly with respect to all integers j as $\theta\to 0$ and $M\to\infty$.

Proof: The main idea is the following. First, using the regularity of the sequence $\{Z_i\}$ we prove that

$$I(Z_1+W_1;U_j(\theta))\to 0$$

uniformly in j as $\theta\to 0$, where W_1 is a Gaussian random variable independent of $\{N_i\}$ and $\{Z_i\}$ (unfortunately, we cannot prove directly that $I(Z_1;U_j(\theta))\to 0$ uniformly in j as $\theta\to 0$). Then, using this fact we show that

$$I(Z_1+W_1;U_j(\theta)|S=1)$$

also converges to zero uniformly with respect to all j as $\theta\to 0$ and $M\to\infty$. It means that Z_1 and $\{U_j(\theta)\}$ become "almost" independent under condition $S=1$ as $\theta\to 0$ and $M\to\infty$ and so the conditional variance of Z_1 given $U_j(\theta)$ and $S=1$ converges to the unconditional one.

Let $\delta>0$ be an arbitrarily small positive number. To prove that

$$I(Z_1+W_1;U_j(\theta))\to 0$$

uniformly with respect to all j as $\theta\to 0$ we shall show that there exists $\theta_0=\theta_0(\delta)>0$ such that for all θ, $0<\theta<\theta_0$

$$I(Z_1+W_1;U_j(\theta))<\delta \quad (53)$$

simultaneously for all j.

For some integer $m=m(\delta)$, which will be chosen later, and for any $j, j>m$ we have

$$\begin{aligned}&I(Z_1+W_1;U_j(\theta))\\&\quad\le I(Z_1+W_1;V_0,\cdots,V_{-m},N_{-m-1},\cdots,N_{-j},\\&\qquad Z_{-m-1},\cdots,Z_{-j})\\&\quad=I(Z_1+W_1;Z_{-m-1},\cdots,Z_{-j})\\&\qquad+I(Z_1+W_1;V_0,\cdots,V_{-m}|Z_{-m-1},\cdots,Z_{-j}) \quad (54)\end{aligned}$$

because $(N_{-m-1},\cdots,N_{-j})$ is independent on all other random variables.

We estimate first the second term of the right-hand side of (54). It is readily shown that

$$\begin{aligned}&I(Z_1+W_1;V_0,\cdots,V_{-m}|Z_{-m-1},\cdots,Z_{-j})\\&\quad\le I((Z_1,W_1);V_0,\cdots,V_{-m}|Z_{-m-1},\cdots,Z_{-j})\\&\quad=I(Z_1;V_0,\cdots,V_{-m}|Z_{-m-1},\cdots,Z_{-j})\\&\quad\le I(Z_1,\cdots,Z_{-m};V_1,\cdots,V_{-m}|Z_{-m-1},\cdots,Z_{-j})\\&\quad\le I(Z_1,\cdots,Z_{-m};V_1,\cdots,V_{-m})\\&\quad\le\frac{m+2}{2}\ln\left(1+\frac{\theta^2}{\sigma^2}\right)<\frac{\delta}{3} \quad (55)\end{aligned}$$

if

$$\theta < \theta_0 = \sigma\sqrt{\frac{2\delta}{3(m+2)}}. \tag{56}$$

Indeed, the third inequality in (55) follows from the fact that

$$(Z_{-m-1},\cdots,Z_{-j}) \to (Z_1,\cdots,Z_{-m}) \to (V_1,\cdots,V_{-m})$$

forms a Markov triple of random vectors, and the fourth one means that $I(Z_1,\cdots,Z_{-m};V_1,\cdots,V_{-m})$ is less than the capacity of a parallel white Gaussian channel in which the sum of input signal powers is not more than $(m+2)\theta^2$.

To estimate the first term of the right-hand side of (54) let us introduce $\{Z_i(M)\}$, a stationary "quantized" version of $\{Z_i\}$, with a finite number M of values such that

$$\kappa^2 = \boldsymbol{E}\Big[\widetilde{Z}_1^2(M)\Big] < \frac{2}{3}\delta\lambda^2 \tag{57}$$

where

$$\widetilde{Z}_1(M) = Z_1 - Z_1(M)$$

and

$$\lambda^2 = \text{var } W_1.$$

Now we have

$$\begin{aligned} I(Z_1+W_1;&Z_{-m-1},\cdots,Z_{-j}) \\ &\le I(Z_1(M), W_1+\widetilde{Z}_1(M);Z_{-m-1},\cdots,Z_{-j}) \\ &= I(Z_1(M);Z_{-m-1},\cdots,Z_{-j}) + I(W_1+\widetilde{Z}_1(M); \\ &\qquad Z_{-m-1},\cdots,Z_{-j}|Z_1(M)). \end{aligned} \tag{58}$$

It is easy to see that the parameter m can be chosen such that for all $j > m$

$$I(Z_1(M);Z_{-m-1},\cdots,Z_{-j}) < \delta/3 \tag{59}$$

because we assume that the sequence $\{Z_i\}$ is regular and $Z_1(M)$ has a finite number of values.

Indeed, it follows from regularity of $\{Z_i\}$ that for $m \to \infty$ [11, ch. IV, sec. 1] (see (60) at the bottom of this page) where $d(\cdot,\cdot)$ is the variational distance between corresponding probability measures.

It is shown in [12] that

$$I(Z_1(M);Z_{-m-1},\cdots,Z_{-j}) \le \mu_m \ln\frac{M}{\mu_m}. \tag{61}$$

Thus choosing m large enough so that

$$\mu_m \ln(M/\mu_m) < \delta/3$$

and using (61) we obtain (59). Then, we have

$$\begin{aligned} I(W_1&+\widetilde{Z}_1(M);Z_{-m-1},\cdots,Z_{-j}|Z_1(M)) \\ &= h(W_1+\widetilde{Z}_1(M)|Z_1(M)) \\ &\quad - h(W_1+\widetilde{Z}_1(M)|Z_1(M),Z_{-m-1},\cdots,Z_{-j}) \\ &\le h(W_1+\widetilde{Z}_1(M)) - h(W_1) \\ &\le \frac{1}{2}\ln\left(1+\frac{\kappa^2}{\lambda^2}\right) < \frac{\delta}{3} \end{aligned} \tag{62}$$

by virtue of (57).

The inequality (53) for all $j > m$ follows from (54), (55), (58), (59), and (62) if $\theta < \theta_0$, where θ_0 was defined in (56) and $m = m(M(\delta))$ and $M = M(\delta)$ are chosen such that the inequalities (57) and (59) are satisfied.

In the case $j \le m$ we obtain

$$\begin{aligned} I(Z_1+W_1;U_j(\theta)) &\le I(Z_1;V_0,\cdots,V_{-m}) \\ &\le I(Z_1,\cdots,Z_{-m};V_1,\cdots,V_{-m}) \\ &\le \frac{m+2}{2}\ln\left(1+\frac{\theta^2}{\sigma^2}\right) < \frac{\delta}{3} \end{aligned}$$

if $\theta < \theta_0$. Thus (53) is proved.

Let us show that $I(Z_1+W_1;U_j(\theta)|S=1)$ also converges to zero uniformly with respect to all j as $\theta \to 0$ and $M \to \infty$. This statement immediately follows from the relations

$$\begin{aligned} I(Z_1+W_1;&U_j(\theta),S) \\ &= I(Z_1+W_1;U_j(\theta)) + I(Z_1+W_1;S|U_j(\theta)) \\ &\le I(Z_1+W_1;U_j(\theta)) + H(S) \end{aligned} \tag{63}$$

and

$$\begin{aligned} I(Z_1+W_1;&U_j(\theta),S) \\ &= I(Z_1+W_1;S) + I(Z_1+W_1;U_j(\theta)|S) \\ &\ge I(Z_1+W_1;U_j(\theta)|S) \\ &\ge \boldsymbol{P}[S=1]I(Z_1+W_1;U_j(\theta)|S=1) \end{aligned} \tag{64}$$

because

$$\lim_{M\to\infty} \boldsymbol{P}[S=1] = 1$$

$$\lim_{M\to\infty} H(S) = 0$$

and

$$I(Z_1+W_1;U_j(\theta)) \to 0$$

uniformly with respect to all j as $\theta \to 0$.

Now Lemma 2 can be easily proven. Indeed, let $\widehat{Z}_1, \widehat{W}_1$, and $\widehat{U}_j(\theta)$ be the random variables whose distribution coincides with the conditional joint distribution of Z_1, W_1, and $U_j(\theta)$ given the event $\{S=1\}$. Then we have

$$\begin{aligned} I(Z_1+W_1;U_j(\theta)|S=1) &= I(\widehat{Z}_1+\widehat{W}_1;\widehat{U}_j(\theta)) \\ &\ge I\Big(\widehat{Z}_1+\widehat{W}_1;\boldsymbol{E}\Big[\widehat{Z}_1+\widehat{W}_1|\widehat{U}_j(\theta)\Big]\Big) \\ &\ge H_{\varepsilon_j}(\widehat{Z}_1+\widehat{W}_1) \end{aligned} \tag{65}$$

$$d(\boldsymbol{P}_{Z_1(M)(Z_{-m-1},\cdots,Z_{-j})}, \boldsymbol{P}_{Z_1(M)\times(Z_{-m-1},\cdots,Z_{-j})}) \le \mu_m = d(\boldsymbol{P}_{Z_1(M)(Z_{-m-1},Z_{-m_2},\cdots)}, \boldsymbol{P}_{Z_1(M)\times(Z_{-m-1},Z_{-m-2},\cdots)}) \to 0 \tag{60}$$

where

$$\varepsilon_j^2 = \text{var}\,(\widehat{Z}_1 + \widehat{W}_1 | \widehat{U}_j(\theta)) = \text{var}\,(Z_1 + W_1 | U_j(\theta), S = 1)$$

and

$$H_\varepsilon(\xi) = \inf_{X: \boldsymbol{E}[(\xi - X)^2] \le \varepsilon^2} I(\xi; X)$$

is the ε-entropy (rate-distortion function) of random variable ξ.

The inequality (65) shows that $H_{\varepsilon_j}(\widehat{Z}_1 + \widehat{W}_1) \to 0$ uniformly with respect to all j as $\theta \to 0$ and $M \to \infty$. In a view of the well-known property of ε-entropy [1, ch. IX] it means that

$$\begin{aligned}\text{var}\,(Z_1 + W_1 | U_j(\theta), S = 1) &= \text{var}\,(Z_1 | U_j(\theta), S = 1) + \lambda^2 \\ &\to \text{var}\,(Z_1 + W_1) = \text{var}\, Z_1 + \lambda^2\end{aligned}$$

uniformly in all j as $\theta \to 0$ and $M \to \infty$ from which the statement of Lemma 2 follows. □

Let us proceed now to the case ii) of Theorem 2, where $\{N_i\}$ is a regular Gaussian stationary process. First of all we note that, by regularity of $\{N_i\}$, the following Wold representation holds for all i:

$$N_i = \sum_{j=-\infty}^{i} a_{i-j}\eta_j \qquad \left(\sum_{j=0}^{\infty} a_j^2 < \infty\right) \tag{66}$$

where $\{\eta_i\}$ is a sequence of orthonormal random variables. Moreover, since $\{N_i\}$ is Gaussian, the sequence $\{\eta_i\}$ is a sequence of i.i.d. Gaussian random variables (see, for example, [11, ch. IV, sec. 1]). It is also well known [10, ch. II, sec. 3] that for the Wold representation (66), we can write for all i

$$\eta_i = \lim_{i \to \infty} \sum_{j: j \le i} c_{i-j}^{(l)} N_j \tag{67}$$

i.e., η_i is the limit in mean-square sense of a sequence of finite sums

$$\sum_{j: j \le i} c_{i-j}^{(l)} N_j.$$

Applying the linear transform (67) to the sequence $\{Z_i\}$ we obtain

$$\zeta_i = \lim_{i \to \infty} \sum_{j: j \le i} c_{i-j}^{(l)} Z_j, \quad \text{for all } i. \tag{68}$$

At first, we note that the limit on the right-hand side of (68) exists, so that ζ_i is defined. To see this, denote by $\varphi(f), -\frac{1}{2} \le f \le \frac{1}{2}$ the spectrum of the linear transform (67). Then, we have

$$\begin{aligned}\int_{-1/2}^{1/2} \varphi^2(f) Z(f)\, df &= \int_{-1/2}^{1/2} \varphi^2(f) \frac{Z(f)}{N_0(f)} N_0(f)\, df \\ &\le L \int_{-1/2}^{1/2} \varphi^2(f) N_0(f)\, df < \infty \end{aligned} \tag{69}$$

where

$$L = \sup_f \frac{Z(f)}{N_0(f)} < \infty$$

by condition ii), and the integral

$$\int_{-1/2}^{1/2} \varphi^2(f) N_0(f)\, df$$

converges by virtue of (68). The existence of the limit in mean square on the right-hand side of (68) follows from (69) [13, ch. X, sec. 9].

Next, it immediately follows from (68) and regularity of $\{Z_i\}$ that the stationary sequence $\{\zeta_i\}$ is regular, because it is evident that $\mathcal{A}_Z(-\infty, t) = \mathcal{A}_\zeta(-\infty, t)$ for every t (see Definition 1).

Now, to finish the proof of Theorem 2, we shall show that

$$\lim_{n \to \infty} \frac{1}{n} D(N_1 + \theta Z_1, \cdots, N_N + \theta Z_n) = \lim_{n \to \infty} \frac{1}{n} D(\eta_1 + \theta\zeta_1, \cdots, \eta_n + \theta\zeta_n). \tag{70}$$

This will suffice because (70) yields

$$\lim_{\theta \to 0} \lim_{n \to \infty} \frac{1}{n\theta^2} D(N_1 + \theta Z_1, \cdots, N_N + \theta Z_n) = 0$$

since the equality

$$\lim_{\theta \to 0} \lim_{n \to \infty} \frac{1}{n\theta^2} D(\eta_1 + \theta\zeta_1, \cdots, \eta_n + \theta\zeta_n) = 0$$

was already proved in statement i) of Theorem 2. To prove (70) it is enough to show that

$$\begin{aligned}&h(\eta_1 + \theta\zeta_1 | \eta_0 + \theta\zeta_0, \eta_{-1} + \theta\zeta_{-1}, \cdots) \\ &\quad = h(N_1 + \theta Z_1 | N_0 + \theta Z_0, N_{-1} + \theta Z_{-1}, \cdots) + c_0 \end{aligned} \tag{71}$$

where c_0 is some constant, by stationarity of the sequences $\{N_i\}$, $\{\eta_i\}$, $\{Z_i\}$, and $\{\zeta_i\}$. But (71) trivially follows from (66)–(68), because $(\eta_0 + \theta\zeta_0, \eta_{-1} + \theta\zeta_{-1}, \cdots)$ is a function of $(N_0 + \theta Z_0, N_{-1} + \theta Z_{-1}, \cdots)$ and *vice versa*. □

Remark 1: It should be noted that in the condition ii) of Theorem 2 we can assume that the function $[Z(f)/N_0(f)]$ is essentially bounded (rather than strictly bounded) on $[0, \frac{1}{2}]$, i.e., that $\operatorname{ess\,sup}[Z(f)/N_0(f)] < \infty$ on $[0, \frac{1}{2}]$. The proof of the theorem under this weaker condition is basically the same.

IV. Entropy-Singular Contaminating Noise

Let us consider now the case when the contaminating noise $\{Z_i\}$ is a singular random sequence.

Definition 2: A stationary random sequence $\{Z_i\}$ is called *entropy-singular* if the entropy of any subordinate process $\{\widetilde{Z}_i\}$ with a finite number of states equals zero, i.e.

$$\overline{H}\left(\left\{\widetilde{Z}_i\right\}\right) \stackrel{\text{def}}{=} \lim_{n \to \infty} \frac{1}{n} H\left(\widetilde{Z}^n\right) = 0$$

(for the definition of subordinate process see, for example, [5, sec. 5]).

Remark 2: In the theory of dynamical systems, entropy-singular sequences are usually referred to as zero-entropy sequences. It should be noted that, normally, a stationary random sequence $\{Z_i\}$ is called singular if for any integer n the random object $(\cdots, Z_{n-2}, Z_{n-1}, Z_n)$ is anywhere tight in the random variable $Z = \{Z_i\}$, (cf. [5, sec. 5]). It is known that any singular sequence $\{Z_i\}$ with a finite number of states is an entropy-singular sequence (see [5, Th. 6.2.1]).

Theorem 3: If $\{N_i\}$ is a regular Gaussian stationary sequence and $\{Z_i\}$ is a second-order entropy-singular stationary sequence, then $C(\theta) = C(0)$ for any $\theta \geq 0$ and, in particular, $\boldsymbol{S}_Z = 0$.

Proof: We start by writing the obvious equality

$$I(\boldsymbol{X}^n; \boldsymbol{Y}^n) = I(\boldsymbol{X}^n; \boldsymbol{X}^n + \boldsymbol{N}^n) + I(\theta \boldsymbol{Z}^n; \boldsymbol{X}^n + \boldsymbol{N}^n + \theta \boldsymbol{Z}^n) - I(\theta \boldsymbol{Z}^n; \boldsymbol{N}^n + \theta \boldsymbol{Z}^n). \tag{72}$$

Let us show now that the entropy-singularity of $\{Z_i\}$ yields the relations

$$\lim_{n\to\infty} \frac{1}{n} I(\theta \boldsymbol{Z}^n; \boldsymbol{X}^n + \boldsymbol{N}^n + \theta \boldsymbol{Z}^n) = \lim_{n\to\infty} \frac{1}{n} I(\theta \boldsymbol{Z}^n; \boldsymbol{N}^n + \theta \boldsymbol{Z}^n) = 0. \tag{73}$$

We need to prove only the second of the equalities in (73), because

$$0 \leq I(\theta \boldsymbol{Z}^n; \boldsymbol{X}^n + \boldsymbol{N}^n + \theta \boldsymbol{Z}^n) \leq I(\theta \boldsymbol{Z}^n; \boldsymbol{N}^n + \theta \boldsymbol{Z}^n).$$

Let $\delta > 0$ be an arbitrarily small positive number. Now, given this δ, we choose $\varepsilon > 0$ such that

$$C_Z^{(\varepsilon\theta)}(0) < \delta \tag{74}$$

where $C_Z^{(\varepsilon\theta)}(0)$ is a capacity of the channel (1) with $\theta = 0$ under input constraint $P = \varepsilon^2\theta^2$.

We introduce again a stationary sequence $\{Z_i(M)\}$ subordinate to $\{Z_i\}$ with finite number M of values (see also proof of Lemma 2) such that

$$\boldsymbol{E}\left[\widetilde{Z}_i^2(M)\right] \leq \varepsilon^2 \tag{75}$$

where $\widetilde{Z}_i(M) = Z_i - Z_i(M)$ for all i, and ε was defined in (74). Then, we have

$$\begin{aligned} I(\theta \boldsymbol{Z}^n; \theta \boldsymbol{Z}^n + \boldsymbol{N}^n) &\leq I\left(\theta \boldsymbol{Z}^n(M), \theta \widetilde{\boldsymbol{Z}}^n(M); \theta \boldsymbol{Z}^n + \boldsymbol{N}^n\right) \\ &= I(\theta \boldsymbol{Z}^n(M); \theta \boldsymbol{Z}^n + \boldsymbol{N}^n) \\ &\quad + I\left(\theta \widetilde{\boldsymbol{Z}}^n(M); \theta \boldsymbol{Z}^n + \boldsymbol{N}^n | \theta \boldsymbol{Z}^n(M)\right). \end{aligned} \tag{76}$$

Since

$$I(\theta \boldsymbol{Z}^n(M); \theta \boldsymbol{Z}^n + \boldsymbol{N}^n) \leq H(\boldsymbol{Z}^n(M)) \tag{77}$$

and

$$I\left(\theta \widetilde{\boldsymbol{Z}}^n(M); \theta \boldsymbol{Z}^n + \boldsymbol{N}^n | \theta \boldsymbol{Z}^n(M)\right) \leq I\left(\theta \widetilde{\boldsymbol{Z}}^n(M); \theta \widetilde{\boldsymbol{Z}}^n(M) + \boldsymbol{N}^n\right) \tag{78}$$

we obtain using (76)–(78) that

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} I(\theta \boldsymbol{Z}^n; \theta \boldsymbol{Z}^n + \boldsymbol{N}^n) &\leq \lim_{n\to\infty} \frac{1}{n} H(\boldsymbol{Z}^n(M)) \\ &\quad + \lim_{n\to\infty} \frac{1}{n} I\left(\theta \widetilde{\boldsymbol{Z}}^n(M); \theta \widetilde{\boldsymbol{Z}}^n(M) + \boldsymbol{N}^n\right) \\ &\leq \delta \end{aligned} \tag{79}$$

because the first term on the right-hand side is equal to zero by entropy-singularity of $\{Z_i\}$ and

$$\lim_{n\to\infty} \frac{1}{n} I\left(\theta \widetilde{\boldsymbol{Z}}^n(M); \theta \widetilde{\boldsymbol{Z}}^n(M) + \boldsymbol{N}^n\right) \leq C_Z^{(\varepsilon\theta)}(0) < \delta$$

by virtue of (75) and (74).

Since $\delta > 0$ can be arbitrarily small it follows from (79) that (73) is valid, which proves Theorem 3, because (73) and (72) lead to the equality $C_Z(\theta) = C_Z(0)$ for all $\theta > 0$. □

Corollary: Let $\{Z_i\}$ be a sum

$$Z_i = \sqrt{\alpha} U_i + \sqrt{1-\alpha} V_i, \quad 0 \leq \alpha \leq 1$$

where $\{U_i\}$ and $\{V_i\}$ are two independent stationary random sequences, such that

$$\boldsymbol{E}[U_i^2] = \boldsymbol{E}[V_i^2] = 1.$$

If $\{N_i\}$ is a regular Gaussian stationary sequence and $\{V_i\}$ is entropy-singular, then the sensitivity $\boldsymbol{S}_Z$ of the channel capacity (1) satisfies

$$\boldsymbol{S}_Z = \alpha \boldsymbol{S}_U.$$

Proof: Analogously to (72) we have

$$\begin{aligned} I(\boldsymbol{X}^n; \boldsymbol{Y}^n) &= I\left(\boldsymbol{X}^n; \boldsymbol{X}^n + \boldsymbol{N}^n + \sqrt{\alpha}\theta \boldsymbol{U}^n\right) \\ &\quad + I\left(\sqrt{1-\alpha}\theta \boldsymbol{V}^n; \boldsymbol{X}^n + \boldsymbol{N}^n + \boldsymbol{Z}^n\right) \\ &\quad - I\left(\sqrt{1-\alpha}\theta \boldsymbol{V}^n; \boldsymbol{N}^n + \boldsymbol{Z}^n\right). \end{aligned} \tag{80}$$

In a similar manner as (73) was proved, we can show that

$$\lim_{n\to\infty} \frac{1}{n} I\left(\sqrt{1-\alpha}\theta \boldsymbol{V}^n; \boldsymbol{X}^n + \boldsymbol{N}^n + \boldsymbol{Z}^n\right) = \lim_{n\to\infty} \frac{1}{n} I\left(\sqrt{1-\alpha}\theta \boldsymbol{V}^n; \boldsymbol{N}^n + \boldsymbol{Z}^n\right) = 0 \tag{81}$$

because $\{V_i\}$ is an entropy-singular sequence.

Now from (80) and (81) it follows that

$$C_Z(\theta) = C_U(\sqrt{\alpha}\theta). \tag{82}$$

Thus taking account (82) we conclude

$$\boldsymbol{S}_Z = \lim_{\theta\to 0} \alpha \frac{C_U(0) - C_U(\sqrt{\alpha}\theta)}{\alpha\theta^2} = \alpha \boldsymbol{S}_U. \qquad \square$$

V. Derivative of Water-Filling Formula

In this section we shall prove formula (8) for the sensitivity. According to Theorem 2 we can assume that the contaminating noise $\{Z_i\}$ is regular Gaussian satisfying condition ii) of Theorem 2. But for Gaussian contaminating noise we shall prove (8) without any additional conditions.

Theorem 4: Let $\{N_i\}$ be a regular Gaussian stationary process. Then the sensitivity of the channel capacity (1) with Gaussian contaminating noise $\{Z_i\}$ having a spectral density $Z(f)$ is given by (8). This formula holds even if the integral on its right-hand side diverges.

Proof: The proof of this theorem is divided into two parts, depending on whether the power spectral density $Z(f)$ is bounded.

Case A: $Z(f)$ is a bounded function:

It is well known [1] that for any $\theta \geq 0$

$$C(\theta) = \frac{1}{2}\int_{-1/2}^{1/2} \ln\left(1 + \frac{S_\theta(f)}{N_\theta(f)}\right) df = \int_{0}^{1/2} \ln\left(1 + \frac{S_\theta(f)}{N_\theta(f)}\right) df \tag{83}$$

where

$$N_\theta(f) = N_0(f) + \theta^2 Z(f)$$

and

$$S_\theta(f) = [K_\theta - N_\theta(f)]^+$$

with K_θ chosen so that

$$2\int_0^{1/2} S_\theta(f)\, df = P$$

because we suppose now that the contaminating noise $\{Z_i\}$ is Gaussian.

Now we need the following lemma proven in the Appendix.

Lemma 3: Let us introduce for any $\theta \geq 0$ the sets

$$A_\theta = \left\{f \in \left[0, \frac{1}{2}\right] : K_0 > N_\theta(f)\right\} \tag{84}$$

$$B = \left\{f \in \left[0, \frac{1}{2}\right] : K_0 = N_0(f)\right\} \tag{85}$$

$$G_\theta = \left\{f \in \left[0, \frac{1}{2}\right] : K_0 \leq N_\theta(f)\right\}. \tag{86}$$

Then, if $Z(f)$ is bounded function on $\left[0, \frac{1}{2}\right]$, the following relations hold:

a) $\lim_{\theta\to 0}[\text{mes } (A_0 \cap G_\theta)] = 0$

b) $\lim_{\theta\to 0}[\text{mes } (A_0 \setminus A_\theta)] = 0$

c) $\lim_{\theta\to 0}[\text{mes } \{(A_\theta \setminus A_0) \setminus B\}] = 0$

d) $K_\theta = K_0 + \frac{\theta^2}{\text{mes } A_\theta}\int_{A_\theta} Z(f)df + o(\theta^2), \quad \theta \to 0.$

Let us return to the proof of the theorem. From (83) we obtain

$$\begin{aligned}\frac{1}{\theta^2}[C(0) - C(\theta)] = &\frac{1}{\theta^2}\int_{A_0\setminus A_\theta} \ln\frac{K_0}{N_0(f)}\, df \\ &- \frac{1}{\theta^2}\int_{A_\theta\setminus A_0} \ln\frac{K_\theta}{N_\theta(f)}\, df \\ &+ \frac{1}{\theta^2}\int_{A_\theta\cap A_0} \ln\left(1 + \theta^2\frac{Z(f)}{N_0(f)}\right) df \\ &- \frac{1}{\theta^2}\int_{A_\theta\cap A_0} \ln\frac{K_\theta}{K_0}\, df.\end{aligned} \tag{87}$$

We evaluate now every integral in the right-hand side of (87). At first we observe that

$$J_1(\theta) \overset{\text{def}}{=} \frac{1}{\theta^2}\int_{A_0\setminus A_\theta} \ln\frac{K_0}{N_0(f)} df = o(1), \quad \theta \to 0. \tag{88}$$

To see this we note that for $f \in A_0 \setminus A_\theta$ we have

$$K_0 > N_0(f) \geq K_0 - Z(f)\theta^2 \geq K_0 - M\theta^2$$

where

$$M = \sup Z(f), \qquad f \in \left[0, \frac{1}{2}\right].$$

Therefore

$$0 \leq J_1(\theta) \leq \frac{1}{\theta^2}\int_{A_0\setminus A_\theta} \ln\frac{K_0}{K_0 - M\theta^2} df = o(1), \quad \theta \to 0$$

since mes $(A_0 \setminus A_\theta) \to 0$ as $\theta \to 0$ by b) of Lemma 3.

The second integral is divided into two parts as

$$J_2(\theta) \overset{\text{def}}{=} \frac{1}{\theta^2}\int_{A_\theta\setminus A_0} \ln\frac{K_\theta}{N_\theta(f)}\, df = J_2'(\theta) + J_2''(\theta) \tag{89}$$

where

$$J_2'(\theta) \overset{\text{def}}{=} \frac{1}{\theta^2}\int_{(A_\theta\setminus A_0)\setminus B} \ln\frac{K_\theta}{N_\theta(f)} df$$

and

$$J_2''(\theta) \overset{\text{def}}{=} \frac{1}{\theta^2}\int_{(A_\theta\setminus A_0)\cap B} \ln\frac{K_\theta}{N_\theta(f)} df.$$

It is readily seen that

$$J_2'(\theta) = o(1), \quad \theta \to 0. \tag{90}$$

Indeed, we have

$$\begin{aligned}0 &\leq J_2'(\theta) \\ &\leq \frac{1}{\theta^2}\int_{(A_\theta\setminus A_0)\setminus B} \ln\frac{K_\theta}{K_0 + \theta^2 Z(f)}\, df \\ &= \int_{(A_\theta\setminus A_0)\setminus B}\left[\frac{1}{K_0 \text{mes } A_\theta}\int_{A_\theta} Z(f)\, df - \frac{Z(f)}{K_0} + o(1)\right] df \\ &= o(1), \quad \theta \to 0.\end{aligned}$$

Here we used the inequality $N_0(f) \geq K_0$ for $f \in A_\theta \setminus A_0$ and the statements c) and d) of Lemma 3.

To evaluate $J_2''(\theta)$ we note that $N_0(f) = K_0$ for $f \in B$, so that

$$\begin{aligned}J_2''(\theta) &= \frac{1}{\theta^2}\int_{(A_\theta\setminus A_0)\cap B} \ln\frac{K_0 + \frac{\theta^2}{\text{mes } A_\theta}\int_{A_\theta} Z(f)df + o(\theta^2)}{K_0 + \theta^2 Z(f)}\, df \\ &= \frac{\text{mes } (A_\theta \setminus A_0)}{K_0 \text{mes } A_\theta}\int_{A_\theta} Z(f)df \\ &\quad - \frac{1}{K_0}\int_{A_\theta\setminus A_0} Z(f)df + o(1), \quad \theta \to 0.\end{aligned} \tag{91}$$

Here we used again statement d) of Lemma 3 and the relation

$$\text{mes}\,[(A_\theta \setminus A_0) \cap B] = \text{mes}\,(A_\theta \setminus A_0) + o(1), \quad \theta \to 0$$

by c) of Lemma 3.

Assume for now that the integral on the right-hand side of (8) converges and consider the integral

$$J_3(\theta) \stackrel{\text{def}}{=} \frac{1}{\theta^2} \int_{A_\theta \cap A_0} \ln\left(1 + \theta^2 \frac{Z(f)}{N_0(f)}\right) df. \tag{92}$$

First of all note that convergence of (8) implies that the integral

$$\int_0^{1/2} \frac{Z(f)}{N_0(f)}\, df \tag{93}$$

also converges (recall that now we assume the boundedness of $Z(f)$). So, using the inequality $\ln(1+x) \leq x$ we obtain from (92)

$$J_3(\theta) \leq \int_{A_0} \frac{Z(f)}{N_0(f)}\, df + o(1), \quad \theta \to 0 \tag{94}$$

because

$$\int_{A_0 \setminus A_\theta} \frac{Z(f)}{N_0(f)}\, df = o(1), \quad \theta \to 0$$

by statement b) of Lemma 3 and convergence of (93).

To obtain a lower bound for $J_3(\theta)$ let us introduce momentarily an additional condition

$$\int_0^{1/2} \left[\frac{Z(f)}{N_0(f)}\right]^2 df < \infty. \tag{95}$$

Then, using (95) and the inequality $\ln(1+x) \geq x - x^2$ for $x \geq 0$ we also verify that

$$J_3(\theta) \geq \int_{A_0} \frac{Z(f)}{N_0(f)}\, df + o(1), \quad \theta \to 0 \tag{96}$$

and so from (94) and (96) we conclude that

$$J_3(\theta) = \int_{A_0} \frac{Z(f)}{N_0(f)}\, df + o(1), \quad \theta \to 0. \tag{97}$$

Finally, using again statements b) and d) of Lemma 3, we obtain

$$\begin{aligned} J_4(\theta) &\stackrel{\text{def}}{=} \frac{1}{\theta^2} \int_{A_\theta \cap A_0} \ln \frac{K_\theta}{K_0}\, df \\ &= \frac{\text{mes}\, A_0}{K_0 \text{mes}\, A_\theta} \int_{A_\theta} Z(f)\, df + o(1), \quad \theta \to 0. \end{aligned} \tag{98}$$

Substituting (88)–(91), (97), and (98) into the right-hand side of (87) and noting that

$$\text{mes}\,(A_\theta \setminus A_0) + \text{mes}\, A_0 = \text{mes}\, A_\theta + o(1), \quad \theta \to 0$$

by b) of Lemma 3, we derive

$$\begin{aligned} \frac{1}{\theta^2}[C(0) - C(\theta)] &= \int_{A_0} \frac{Z(f)}{N_0(f)}\, df + \frac{1}{K_0} \int_{A_\theta \setminus A_0} Z(f)\, df \\ &\quad - \frac{1}{K_0} \int_{A_\theta} Z(f)\, df + o(1) \\ &= \int_{A_0} \frac{Z(f)}{N_0(f)}\, df - \frac{1}{K_0} \int_{A_0} Z(f)\, df + o(1) \\ &= \frac{1}{K_0} \int_0^{1/2} Z(f) \frac{S_0(f)}{N_0(f)}\, df + o(1), \quad \theta \to 0. \end{aligned}$$

This completes the proof of the theorem in Case A under the additional condition (95) and the assumption that the integral on the right-hand side of (8) converges.

Let us dispense with the condition (95). It should be noted that the right-hand side of (8) is an upper bound on the sensitivity $\boldsymbol{S}_Z$ without the assumption (95) because it was used only for deriving a lower bound on $\boldsymbol{S}_Z$ (see (96)). So, we need only prove that the right-hand side of (8) is also a lower bound for $\boldsymbol{S}_Z$ without the condition (95). To this end we introduce a sequence of stationary Gaussian contaminating noises

$$Z^{(m)} = \left\{Z_i^{(m)}\right\}, \quad m = 1, 2, \cdots$$

independent on $\{N_i\}$ with power spectral densities given by

$$Z^{(m)}(f) \stackrel{\text{def}}{=} \begin{cases} Z(f), & \text{if } f \in Q_m \\ 0, & \text{otherwise} \end{cases} \tag{99}$$

where

$$Q_m = \left\{ f \in \left[0, \frac{1}{2}\right] : \frac{1}{N_0(f)} < m \right\}. \tag{100}$$

At first we note that for any m

$$\boldsymbol{S}_Z \geq \boldsymbol{S}_{Z^{(m)}}$$

because it follows from (99) that

$$C_Z(0) = C_{Z^{(m)}}(0), \quad C_Z(\theta) \leq C_{Z^{(m)}}, \quad (\theta)\theta > 0.$$

Moreover, for any sequence $Z^{(m)}$, an analog of the condition (95) is fulfilled, i.e.

$$\int_0^{1/2} \left[\frac{Z^{(m)}(f)}{N_0(f)}\right]^2 df < \infty$$

and, therefore

$$\boldsymbol{S}_{Z^{(m)}} = \frac{1}{K_0} \int_0^{1/2} Z^{(m)}(f) \frac{S_0(f)}{N_0(f)}\, df. \tag{101}$$

It is readily noted that the condition (36) of nominal noise regularity implies that $\text{mes}\, \overline{Q}_m \to 0$ as $m \to \infty$, where $\overline{Q}_m = \left[0, \frac{1}{2}\right] \setminus Q_m$, and consequently for $f \in \left[0, \frac{1}{2}\right]$

$$Z^{(m)}(f) \uparrow Z(f) \text{ as } m \to \infty \tag{102}$$

by the definitions (99) and (100).

Finally, it follows from (102) that the integral on the right-hand side of (101) goes to the integral in (8) or goes to infinity, if the integral in (8) diverges, as $m \to \infty$. This completes the proof of Theorem 4 in Case A.

Case B: $Z(f)$ is an unbounded function:

Let us introduce two sequences of independent stationary Gaussian contaminating noises

$$\widehat{Z}^{(m)} = \left\{\widehat{Z}_i^{(m)}\right\}$$

and

$$\widetilde{Z}^{(m)} = \left\{\widetilde{Z}_i^{(m)}\right\}, \qquad m = 1, 2, \cdots$$

independent also of $\{N_i\}$ with power spectral densities given, respectively, by

$$\widehat{Z}^{(m)}(f) \overset{\text{def}}{=} \begin{cases} Z(f), & \text{if } Z(f) < m \\ 0, & \text{otherwise} \end{cases} \tag{103}$$

and

$$\widetilde{Z}^{(m)}(f) \overset{\text{def}}{=} \begin{cases} Z(f), & \text{if } Z(f) \geq m \\ 0, & \text{otherwise} \end{cases} \tag{104}$$

such that for any $m = 1, 2, \cdots$ we have

$$Z = \widehat{Z}^{(m)} + \widetilde{Z}^{(m)}. \tag{105}$$

It is easily verified now that

$$\boldsymbol{S}_Z \geq \frac{1}{K_0} \int_0^{1/2} Z(f) \frac{S_0(f)}{N_0(f)}\, df. \tag{106}$$

Indeed, (105) shows that for any integer m

$$\boldsymbol{S}_Z \geq \boldsymbol{S}_{\widehat{Z}^{(m)}} \tag{107}$$

and since $\widehat{Z}^{(m)}(f)$ is a bounded function by definition (103), we can apply our theorem in Case A, such that

$$\boldsymbol{S}_{\widehat{Z}^{(m)}} = \frac{1}{K_0} \int_0^{1/2} \widehat{Z}^{(m)}(f) \frac{S_0(f)}{N_0(f)}\, df. \tag{108}$$

Noting that

$$\widehat{Z}^{(m)}(f) \uparrow Z(f) \text{ as } m \to \infty$$

by (103), we immediately obtain (106) from (107) and (108). In particular, (106) implies that $\boldsymbol{S}_Z = \infty$ if the integral on the right-hand side of (106) diverges.

To end the proof of the theorem we need only show that the right-hand side of (106) is also an upper bound for $\boldsymbol{S}_Z$, if the integral converges.

Observe that for any integers n and m we have

$$\begin{aligned} &\frac{1}{n} I(\boldsymbol{X}^n; \boldsymbol{X}^n + \boldsymbol{N}^n + \theta \boldsymbol{Z}^n) \\ &\quad \geq \frac{1}{n} I\left(\boldsymbol{X}^n; \boldsymbol{X}^n + \boldsymbol{N}^n + \theta\left(\widehat{\boldsymbol{Z}}^{(m)}\right)^n\right) \\ &\qquad - \frac{1}{n} I\left(\theta\left(\widetilde{\boldsymbol{Z}}^{(m)}\right)^n; \theta\left(\widetilde{\boldsymbol{Z}}^{(m)}\right)^n + \boldsymbol{N}^n + \theta\left(\widehat{\boldsymbol{Z}}^{(m)}\right)^n\right) \\ &\quad \geq \frac{1}{n} I\left(\boldsymbol{X}^n; \boldsymbol{X}^n + \boldsymbol{N}^n + \theta\left(\widehat{\boldsymbol{Z}}^{(m)}\right)^n\right) \\ &\qquad - \frac{1}{n} I\left(\theta\left(\widetilde{\boldsymbol{Z}}^{(m)}\right)^n; \theta\left(\widetilde{\boldsymbol{Z}}^{(m)}\right)^n + \boldsymbol{N}^n\right). \end{aligned} \tag{109}$$

Since

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} I\left(\theta\left(\widetilde{\boldsymbol{Z}}^{(m)}\right)^n; \theta\left(\widetilde{\boldsymbol{Z}}^{(m)}\right)^n + \boldsymbol{N}^n\right) &\leq \int_0^{1/2} \ln\left(1 + \frac{\theta^2 \widetilde{Z}^{(m)}(f)}{N_0(f)}\right) df \\ &\leq \theta^2 \int_0^{1/2} \frac{\widetilde{Z}^{(m)}(f)}{N_0(f)}\, df \end{aligned} \tag{110}$$

and

$$\lim_{m\to\infty} \int_0^{1/2} \frac{\widetilde{Z}^{(m)}(f)}{N_0(f)}\, df = 0 \tag{111}$$

which follows from definition (104) and convergence of the integral on the right-hand side of (106). Putting together (109)–(111) we get

$$\boldsymbol{S}_Z \leq \frac{1}{K_0} \int_0^{1/2} Z(f) \frac{S_0(f)}{N_0(f)}\, df.$$

This completes the proof of the theorem. □

Appendix
Proof of Lemma 3

a) By the definitions (84) and (86) we have

$$\begin{aligned} A_0 \cap G_\theta &= \{f : K_0 - \theta^2 Z(f) \leq N_0(f) < K_0\} \\ &\subseteq \{f : K_0 - M\theta^2 \leq N_0(f) < K_0\} \downarrow \emptyset, \quad \theta \to 0 \end{aligned}$$

where

$$M = \sup_{f \in [0, \frac{1}{2}]} Z(f)$$

and so that mes $(A_0 \cap G_\theta) \to 0$ as $\theta \to 0$.

b) This follows immediately from a)

$$A_0 \setminus A_\theta \subseteq A_0 \cap G_\theta.$$

c) To prove c) let us note that

$$\begin{aligned} &(A_\theta \setminus A_0) \setminus B \\ &\quad = \{f : K_0 + \theta^2 Z(f) \leq N_0(f) + \theta^2 Z(f) < K_\theta\} \setminus B \\ &\quad \subseteq \{f : K_0 \leq N_0(f) < K_0 + c\theta^2\} \setminus \\ &\qquad \{f : K_0 = N_0(f)\} \downarrow \emptyset, \theta \to 0 \end{aligned}$$

where c is some positive constant. Here we used the obvious inequality $K_\theta \leq K_0 + (\theta^2/2 \,\text{mes}\, A_0)$. Thus we conclude that mes $\{(A_\theta \setminus A_0) \setminus B\} \to 0$ as $\theta \to 0$ and c) is proved.

d) We start by writing the "balance" equation:

$$\begin{aligned} \int_{A_0} \theta^2 Z(f)\, df - \int_{A_0 \cap G_\theta} (N_\theta(f) - K_0)\, df &= (K_\theta - K_0) \text{mes } A_\theta \\ &\quad - \int_{A_\theta \cap A_0 \cap G_\theta} (N_\theta(f) - K_0)\, df \\ &\quad - \int_{(A_\theta \setminus A_0) \cap B} \theta^2 Z(f)\, df \\ &\quad - \int_{(A_\theta \setminus A_0) \setminus B} (N_\theta(f) - K_0)\, df. \end{aligned} \tag{112}$$

From a physical point of view, the left-hand side of (112) is the volume of ''ballast'' $\theta^2 Z(f)$ under the old water level K_0, and the right-hand side of (112) is the water volume above the level K_0 after we dropped this ''ballast'' into the water.

Now observe that

$$\int_{A_0 \cap G_\theta} (N_\theta(f) - K_0)\, df = o(\theta^2), \quad \theta \to 0 \qquad (113)$$

$$\int_{A_\theta \cap A_0 \cap G_\theta} (N_\theta(f) - K_0)\, df = o(\theta^2), \quad \theta \to 0 \qquad (114)$$

$$\int_{(A_\theta \setminus A_0) \setminus B} (N_\theta(f) - K_0)\, df = o(\theta^2), \quad \theta \to 0 \qquad (115)$$

because

$$N_\theta(f) - K_0 = O(\theta^2), \quad \theta \to 0$$

for any

$$f \in (A_0 \cap G_\theta) f \in (A_\theta \cap A_0 \cap G_\theta)$$

or

$$f \in [(A_\theta \setminus A_0) \setminus B]$$

and

$$\begin{aligned} &\text{mes}\ (A_0 \cap G_\theta) \to 0 \\ &\text{mes}\ (A_\theta \cap A_0 \cap G_\theta) \to 0 \\ &\text{mes}\ [(A_\theta \setminus A_0) \setminus B] \to 0 \end{aligned}$$

as $\theta \to 0$ by a) and c) of the Lemma.

Moreover,

$$\int_{(A_\theta \setminus A_0) \cap B} \theta^2 Z(f)\, df = \theta^2 \int_{A_\theta \setminus A_0} Z(f)\, df + o(\theta^2), \quad \theta \to 0 \qquad (116)$$

since

$$\theta^2 \int_{(A_\theta \setminus A_0) \setminus B} Z(f)\, df = o(\theta^2), \quad \theta \to 0$$

according to c). Now from (112)—(116) the statement d) follows directly. □

Acknowledgment

The authors wish to thank Prof. M. L. Honig, Prof. R. Z. Khas'minskii, and Prof. S. Shamai (Shitz) for useful discussions and helpful comments.

References

[1] R. G. Gallager, *Information Theory and Reliable Communication.* New York: Wiley, 1968.

[2] S. Shamai (Shitz) and I. Kalet, "On the capacity of a twisted-wire pair: Gaussian model," *IEEE Trans. Commun.*, vol. 38, pp. 379–383, Mar. 1990.

[3] N. Merhav, G. Kaplan, A. Lapidoth, S. Shamai, "On information rates for mismatched decoders," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1953–1967, Nov. 1994.

[4] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Sources.* Budapest, Hungary: Akadémiai Kiadó, 1981.

[5] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes.* San Francisco, CA: Holden–Day, 1964.

[6] S. Ihara, "On the capacity of channels with additive non-Gaussian noise," *Informat. Contr.*, vol. 37, pp. 34–39, 1978.

[7] C. E. Shannon, "A mathematical theory of communication," *Bell. Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948.

[8] M. S. Pinsker, "Calculation of the rate of information production by means of stationary random processes and the capacity of stationary channel," *Dokl. Akad. Nauk USSR*, vol. 111, pp. 753–756, 1956.

[9] V. V. Prelov, "The capacity of the channels with almost Gaussian noise," *Teoria Verojatnostei i ee Primenenia,* vol. 33, no. 3, pp. 433–452, 1988.

[10] Yu. A. Rozanov, *Stationary Random Processes.* Moscow, USSR: Fizmatgiz, 1963.

[11] I. A. Ibragimov and Yu. A. Rozanov, *Gaussian Random Processes.* Moscow, USSR: Nauka, 1970.

[12] M. S. Pinsker, to be published.

[13] J. L. Doob, *Stochastic Processes.* New York: Wiley, 1953.