Information Regular and ψ-Mixing Channels

Martin Mittelbach and Eduard A. Jorswieck
Communications Theory, Communications Laboratory
Department of Electrical Engineering and Information Technology
Dresden University of Technology, Germany
Email: {martin.mittelbach, eduard.jorswieck}@tu-dresden.de

***Abstract*— For channels with time structure we introduce a condition that characterizes infinite, asymptotically decreasing output memory called information regularity. We show that this condition is strictly weaker than asymptotic output-memorylessness, a property that was used to prove a coding theorem for continuous-time channels in case of infinite information capacity. We further show that the coding theorem still holds for information regular channels. We prove that Gaussian asymptotic output-memoryless channels have in fact finite output memory. This demonstrates that asymptotic output-memorylessness is quite restrictive. We consider discrete- and continuous-time channels with completely abstract alphabets in a unified framework and discuss practically relevant examples.**

I. Introduction

Background. For channels with time structure coding theorems were derived under finite (output) memory conditions in [1], [2], [3] and [4]. In [5] more general channels with infinite output memory were considered, which still satisfy the following essential property: Any probability measure induced by the channel and an ergodic input probability measure is ergodic as well. The conditions formulated in [5] are based on strong and weak mixing properties in the ergodic-theoretic sense. A more restrictive constraint on the infinite channel output memory was introduced in [6]. The so-called asymptotic output-memorylessness allowed to prove a coding theorem for continuous-time channels also for the case of infinite information capacity.

Finite channel output memory means, sufficiently time-separated output events are independent given a fixed input. The asymptotic independence of remote output events, in turn, is the defining property of the infinite output memory conditions in [5] and [6]. The measures of dependence listed in [7, Defs. 3.3, 5.1] can be used equivalently to characterize stochastic independence. However, if one defines asymptotic independence based on these coefficients, then the resulting conditions, called strong mixing conditions, are no longer equivalent [7, Sec. 5.22].

Now we observe that infinite output memory conditions of the sort as proposed in [6] can be formulated using conditional versions of the dependence coefficients considered in [7]. The asymptotic output-memorylessness, e. g., can be expressed with a conditional version of the so-called ψ-coefficient. An advantage of using this representation is that one can exploit more easily the connections to the rich field of strong mixing conditions, treated in [7] (and two more volumes).

Contribution and outline. In this paper we introduce in analogy to so-called information regular sequences [7, Sec. 5.1] a class of channels with infinite output memory, which we call information regular channels. The definition is based on the conditional mutual information, which seems very natural in an information-theoretic context. We show that the considered condition is strictly weaker than asymptotic output-memorylessness and that Gaussian asymptotic output-memoryless channels have in fact finite output memory, demonstrating that this condition is rather restrictive. We further show that the coding theorem in [6] is still valid for information regular channels. We consider discrete- and continuous-time channels with completely abstract alphabets in a unified framework, which requires some basics of measure theory. Some practically relevant examples are discussed.

Section II provides the basic material for the rest of the paper. We introduce the ψ- and the information coefficient (mutual information) in a conditional and unconditional form. We define corresponding strong mixing conditions and τ-dependence for random processes, collect some properties, and give a useful example of an information regular but not ψ-mixing process. Based on Section II we define in Section III information regular and ψ-mixing (asymptotically output-memoryless) channels and those with finite output memory. Relations between these conditions are derived. In Section IV we show that the coding theorem in [6] still holds for information regular channels. In Section V we consider a general example for which the strong mixing property of the noise source implies the corresponding mixing property of the channel. Additive noise and multipath fading channels are discussed as special cases. Relevant examples of those channels are shown to be information regular but not ψ-mixing.

Notation and preliminaries. Let T denote the set of time indices, which is either equal to the integers $\mathbb{Z}$ or to the real numbers $\mathbb{R}$, modeling discrete or continuous time. Occasionally, it is convenient to extend T and define $\overline{T} = T \cup \{-\infty, \infty\}$.

Let $(\Omega, \mathcal{F}, P)$ be an abstract probability space and $\{(X_t, \mathcal{X}_t)\}$, $t \in T$, be a family of arbitrary measurable spaces. For any $s \leq t \in \overline{T}$ we define the product space and σ-algebra

$$X_s^t = \times_{u \in J} X_u, \qquad \mathcal{X}_s^t = \bigotimes_{u \in J} \mathcal{X}_u, \tag{1}$$

where $J = (s, t] \cap T$ if $s < t < \infty$, $J = (s, \infty) \cap T$ if $s < t = \infty$, $J = \{t\}$ if $s = t \neq \pm\infty$, and $J = \varnothing$ otherwise. By $x_s^t = \{x_u\}$, $x_u \in X_u$, $u \in J$, we denote an element of X_s^t. For any $t \in T$ let α_t be a random variable on $(\Omega, \mathcal{F}, P)$ with values in $(X_t, \mathcal{X}_t)$. We write $\sigma(\alpha_t)$ for the inverse image $\alpha_t^{-1}(\mathcal{X}_t)$ of $\mathcal{X}_t$ with respect to (w. r. t.) α_t. Given $s \leq t \in \overline{T}$ we denote by α_s^t the random variable on $(\Omega, \mathcal{F}, P)$ with values in $(X_s^t, \mathcal{X}_s^t)$, which is defined for any $\omega \in \Omega$ by

$$\alpha_s^t(\omega) = \{\alpha_u(\omega)\},\ u \in J.$$

This research was supported in part by the Deutsche Forschungsgemeinschaft (DFG) under grant JO 801/3-1.

We identify α_s^t with the family $\{\alpha_u\}$, $u \in J$, of random variables, which is (a segment of) a random process either with discrete (random sequence) or continuous time. We further put $X = X_{-\infty}^{\infty}$, $\mathcal{X} = \mathcal{X}_{-\infty}^{\infty}$, $x = x_{-\infty}^{\infty}$, and $\alpha = \alpha_{-\infty}^{\infty}$.

From the family $\{(Y_t, \mathcal{Y}_t)\}$, $t \in T$, of measurable spaces and the random variables β_t on $(\Omega, \mathcal{F}, P)$ with values in $(Y_t, \mathcal{Y}_t)$ we derive similar quantities as in the previous paragraph using a corresponding notation. The same applies to the family $\{(Z_t, \mathcal{Z}_t)\}$, $t \in T$, of measurable spaces and the random variables φ_t on $(\Omega, \mathcal{F}, P)$ with values in $(Z_t, \mathcal{Z}_t)$. It is convenient to consider projections to sub-product spaces. Thus, for all $t \in T$ let ξ_t, η_t, and ζ_t denote the projection from X, Y, and Z to X_t, Y_t, and Z_t, respectively. In connection with channels ξ_t and η_t are defined on $X \times Y$. Corresponding to the notation of the previous paragraph ξ_s^t, η_s^t, and ζ_s^t denote the projections to X_s^t, Y_s^t, and Z_s^t for $s \leq t \in \overline{T}$.

If $(X_t, \mathcal{X}_t) = (X_0, \mathcal{X}_0)$ for all $t \in T$, then a probability measure μ on $\mathcal{X}$ is called stationary if

$$\mu(\theta_s(A)) = \mu(A)$$

holds for all $A \in \mathcal{X}$ and $s \in T$. By θ_s we denote the s-shift operator on X, which is defined for any $x = \{x_t\} \in X$ by

$$\theta_s(x) = \tilde{x} = \{\tilde{x}_t\}, \quad \tilde{x}_t = x_{t-s}, \quad t \in T.$$

The random process $\{\alpha_t\}$, $t \in T$, is called stationary if its distribution P_α is stationary.

II. Dependence coefficients, ψ-mixing, information regular, and τ-dependent random processes

We define special dependence coefficients as well as strong mixing conditions for random processes as introduced in [7, Chs. 3, 5], where the discrete-time case is considered. The collected properties and the given example are required later in connection with strong mixing channels. Finally, we define conditional versions of the dependence coefficients.

(II.1) Definition. (Dependence coefficients) For the two random variables α_0 and β_0 we define the ψ-coefficient as

$$\psi(\alpha_0;\beta_0) = \sup\left|\frac{P_{\alpha_0,\beta_0}(A \times B)}{P_{\alpha_0}(A)P_{\beta_0}(B)} - 1\right|,$$

where the supremum is taken w. r. t. all sets $A \in \mathcal{X}_0$ and $B \in \mathcal{Y}_0$ with $P_{\alpha_0}(A)P_{\beta_0}(B) > 0$. The information coefficient[1] (mutual information) is defined as

$$I(\alpha_0;\beta_0) = \sup\sum_{i=1}^{m}\sum_{j=1}^{n} P_{\alpha_0,\beta_0}(A_i \times B_j)\log\frac{P_{\alpha_0,\beta_0}(A_i \times B_j)}{P_{\alpha_0}(A_i)P_{\beta_0}(B_j)},$$

where the supremum is taken w. r. t. to all finite partitions $\{A_1, \ldots, A_m\}$ of X_0 and $\{B_1, \ldots, B_n\}$ of Y_0 with $A_i \in \mathcal{X}_0$, $B_j \in \mathcal{Y}_0$, and $P_{\alpha_0}(A_i)P_{\beta_0}(B_j) > 0$.

By P_{α_0,β_0} we denote the distribution of (α_0, β_0) and by P_{α_0} and P_{β_0} the corresponding marginal distributions.

(II.2) Remark. We also write $\psi\big(P_{\alpha_0,\beta_0}\big)$ and $I\big(P_{\alpha_0,\beta_0}\big)$ instead of $\psi(\alpha_0;\beta_0)$ and $I(\alpha_0;\beta_0)$.

Note that both dependence coefficients are nonnegative and that they are zero if and only if α_0 and β_0 are independent.

[1]We use the logarithm w. r. t. Euler's number and the convention $0 \log 0 = 0$ throughout the paper.

The two (in)equalities of the next proposition are simple measure-theoretic observations (see [7, p. 68]) and the implication follows from [7, Prop. 5.2(I)] and [7, Th. 5.3(II)]. Obviously, (3) is a special form of the well-known data-processing inequality.

(II.3) Proposition. *Let $\epsilon \geq 0$ and g be a measurable function on $(X_0, \mathcal{X}_0)$ with values in $(Z_0, \mathcal{Z}_0)$. Then we have for the random variables α_0 and β_0:*

$$\psi(g(\alpha_0);\beta_0) \leq \psi(\alpha_0;\beta_0), \tag{2}$$

$$I(g(\alpha_0);\beta_0) \leq I(\alpha_0;\beta_0), \tag{3}$$

$$\psi(\alpha_0;\beta_0) \leq \epsilon \Longrightarrow I(\alpha_0;\beta_0) \leq (1+\epsilon)\log(1+\epsilon). \tag{4}$$

If $g^{-1}(\mathcal{Z}_0) = \mathcal{X}_0$, then we have equality in (2) *and* (3)*.*

(II.4) Definition. (ψ-mixing, information regular, τ-dependent random process) The random process $\{\alpha_t\}$, $t \in T$, is called ψ-mixing if

$$\lim_{t\to\infty}\sup_{s\in T}\psi(\alpha_{-\infty}^{s};\alpha_{s+t}^{\infty}) = 0 \tag{5}$$

and information regular if

$$\lim_{t\to\infty}\sup_{s\in T} I(\alpha_{-\infty}^{s};\alpha_{s+t}^{\infty}) = 0. \tag{6}$$

It is called τ-dependent, if there exists a nonnegative $\tau \in T$ such that for all $s \in T$ the random variables $\alpha_{-\infty}^{s}$ and $\alpha_{s+\tau}^{\infty}$ are independent.

A probability measure μ on $\mathcal{X}$ is called ψ-mixing (information regular, τ-dependent) if the random process $\{\xi_t\}$, $t \in T$, of projections from X to X_t is ψ-mixing (information regular, τ-dependent).

(II.5) Remark. If the random process $\{\alpha_t\}$, $t \in T$, is stationary, then

$$\lim_{t\to\infty}\psi(\alpha_{-\infty}^{0};\alpha_{t}^{\infty}) = 0$$

is equivalent to (5). A similar statement is true for (6) and τ-dependence follows from the independence of $\alpha_{-\infty}^{0}$ and α_{τ}^{∞}.

The information regularity and ψ-mixing conditions are different versions of characterizing that random variables with remote time index are asymptotically independent. The next proposition relates τ-dependence and the strong mixing properties. Part (a) follows from Remark II.2 and from (4). The examples in [7, Expl. 7.9 and 7.11] together with [7, Sec. 5.22] show that the reversed implications are not true in general. Part (b) is a straightforward generalization of [7, Th. 9.7(II)].

(II.6) Proposition. *(a) A τ-dependent random process is ψ-mixing and a ψ-mixing random process is information regular.*

(b) A real-valued ψ-mixing Gaussian (vector) process is τ-dependent.

(II.7) Example. Consider the product measurable space $(Z, \mathcal{Z})$ with $(Z_t, \mathcal{Z}_t) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ for all $t \in T$, where $\mathcal{B}(\mathbb{R})$ denotes the Borel-σ-algebra on $\mathbb{R}$. Let λ be a probability measure on $\mathcal{Z}$ such that the family $\{\zeta_t\}$, $t \in T$, of projections from Z to Z_t is a stationary Gaussian process with finite second moments and rational spectral density $\hat{f}$. If $T = \mathbb{Z}$, then $\hat{f}$ is given as in [8, eq. (2.3.8)]. We assume [8, (i)–(iii), p. 71] is satisfied and the denominator polynomial has at least one root. If $T = \mathbb{R}$, then $\hat{f}$ is given as in [8, eq. (2.5.5)] and we assume the conditions listed thereafter are satisfied. Such processes result from passing stationary white Gaussian noise through a (well-behaved) linear filter.

Using [9, Ch. IV, Th. 8] we can show that ζ is information regular in the discrete-time case. From [10, Th. 10.1.1] or [9, Ch. IV, Th. 9] together with the remark in [9, p. 128] we obtain that ζ is information regular also in the continuous-time case. In both cases the covariance function of ζ is not concentrated in a finite interval. Thus, ζ is not τ-dependent and due to Proposition II.6.b it is not ψ-mixing.

Proposition II.3 together with the basic measure-theoretic Lemma A.1 in the Appendix yield the subsequent result. A special case is also considered in [7, Rmk. 1.8(b)].

(II.8) Proposition. *Let $0 \leq u, v \in T$ be fixed and for any $t \in T$ assume g_t is a measurable function on $(X_{t-u}^{t+v}, \mathcal{X}_{t-u}^{t+v})$ with values in $(Z_t, \mathcal{Z}_t)$. If the random process $\{\alpha_t\}$, $t \in T$, is ψ-mixing (information regular, τ-dependent), then the random process $\{g_t(\alpha_{t-u}^{t+v})\}$, $t \in T$, is ψ-mixing (information regular, τ-dependent). If $g_t^{-1}(\mathcal{Z}_t) = \mathcal{X}_{t-u}^{t+v}$ holds for all $t \in T$, then the reversed implication is also true.*

We complete the section by defining conditional versions of the ψ- and information coefficient.

(II.9) Definition. (Conditional dependence coefficients) Assume the regular conditional distribution $P_{\alpha_0,\beta_0|\varphi_0}$ of (α_0, β_0) given φ_0 exists. For any $z_0 \in Z_0$ we define

$$\begin{aligned} \psi(\alpha_0; \beta_0 \,|\, \varphi_0 = z_0) &= \psi\big(P_{\alpha_0,\beta_0|\varphi_0=z_0}\big) \\ I(\alpha_0; \beta_0 \,|\, \varphi_0 = z_0) &= I\big(P_{\alpha_0,\beta_0|\varphi_0=z_0}\big), \end{aligned}$$

where $P_{\alpha_0,\beta_0|\varphi_0=z_0}$ denotes the (conditional) distribution of (α_0, β_0) given $\varphi_0 = z_0$. We further define

$$\begin{aligned} \psi(\alpha_0; \beta_0 \,|\, \varphi_0) &= \int_{Z_0} \psi(\alpha_0; \beta_0 \,|\, \varphi_0 = z_0) \,\mathrm{d}P_{\varphi_0}(z_0) \\ I(\alpha_0; \beta_0 \,|\, \varphi_0) &= \int_{Z_0} I(\alpha_0; \beta_0 \,|\, \varphi_0 = z_0) \,\mathrm{d}P_{\varphi_0}(z_0), \end{aligned} \tag{7}$$

where P_{φ_0} denotes the distribution of φ_0.

(II.10) Remark. The conditional dependence coefficients are natural extensions of the unconditional versions, which require the existence of a regular conditional distribution. Formulations without this precondition are possible, but the given ones are sufficient subsequently. Note that the conditional information coefficient $I(\alpha_0; \beta_0 \,|\, \varphi_0)$ is the usual conditional mutual information between α_0 and β_0 given φ_0, as considered in the situation of [10, pp. 32–34].

Finally note, that Proposition II.3 can be directly extended to the conditional dependence coefficients by adding '$|\, \varphi_0 = z_0$' or '$|\, \varphi_0$' to all quantities in (2)–(4).

III. ψ-Mixing, Information Regular, and Finite Output Memory Channels

First, we define (information) channels with time structure and stationarity as well as Gaussian channels. Then finite and infinite output memory conditions are introduced and relations between them are derived.

The abstract definition of a channel (with time structure) as adopted here can be found, e. g., in [11] or [12, Sec. 3.1].

(III.1) Definition. (Channel with time structure, stationarity, Gaussian channel) A channel κ with time structure is a Markov-kernel from the input product measurable space $(X, \mathcal{X})$ to the output product measurable space $(Y, \mathcal{Y})$. If $T = \mathbb{Z}$, then κ is called a discrete-time channel and if $T = \mathbb{R}$, then a continuous-time channel.

Given $(X_t, \mathcal{X}_t) = (Y_t, \mathcal{Y}_t) = (X_0, \mathcal{X}_0)$ for all $t \in T$, then the channel κ is called stationary, if for any $s \in T$, $x \in X$, and $B \in \mathcal{Y}$ we have

$$\kappa(x, B) = \kappa(\theta_s(x), \theta_s(B)).$$

Given $(Y_t, \mathcal{Y}_t) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ for all $t \in T$, where $\mathcal{B}(\mathbb{R}^n)$ denotes the Borel-σ-algebra on $\mathbb{R}^n$, then κ is called a Gaussian channel, if for all fixed inputs $x \in X$ the family $\{\eta_t\}$, $t \in T$, of output projections (see end of Section I) is a Gaussian random (vector) process with finite second moments.

(III.2) Remark. The definition of a channel means $\kappa(x, \cdot)$ is a probability measure on $\mathcal{Y}$ for any $x \in X$ and $\kappa(\cdot, B)$ is an $\mathcal{X}/\mathcal{B}([0,1])$-measurable function on X for any $B \in \mathcal{Y}$.

Any probability measure μ on $\mathcal{X}$ induces together with the channel κ a probability measure $\mu\kappa$ on $\mathcal{X} \otimes \mathcal{Y}$ given by

$$\mu\kappa(C) = \int_X \kappa(x, C_x) \,\mathrm{d}\mu(x), \quad C \in \mathcal{X} \otimes \mathcal{Y}, \tag{8}$$

where C_x denotes the x-section of C. For the projections defined at the end of Section I we have that for any μ and $s, t \in T$ with $t \geq 0$ the regular conditional distribution of $(\eta_{-\infty}^s, \eta_{s+t}^\infty)$ given ξ always exists. The Markov-kernel K with

$$K(x, B) = \kappa(x, [B]), \quad x \in X,\ B \in \mathcal{Y}_{-\infty}^s \otimes \mathcal{Y}_{s+t}^\infty,$$

is a version of this conditional distribution, where $[B]$ denotes the inverse image of the set B w. r. t. to $(\eta_{-\infty}^s, \eta_{s+t}^\infty)$. Therefore, the following definition is always meaningful.

(III.3) Definition. (ψ-mixing, information regular, finite output memory channel) The channel κ is called ψ-mixing if for all $s \in T$

$$\lim_{t\to\infty} \sup_{x \in X} \psi(\eta_{-\infty}^s; \eta_{s+t}^\infty \,|\, \xi = x) = 0 \tag{9}$$

and information regular if for all $s \in T$

$$\lim_{t\to\infty} \sup_{x \in X} I(\eta_{-\infty}^s; \eta_{s+t}^\infty \,|\, \xi = x) = 0. \tag{10}$$

It has finite output memory, if for all $s \in T$ there exists a nonnegative $t_0(s) \in T$, such that for all $t \geq t_0(s)$ we have

$$\sup_{x \in X} \psi(\eta_{-\infty}^s; \eta_{s+t}^\infty \,|\, \xi = x) = 0. \tag{11}$$

(III.4) Remark. In [6] ψ-mixing channels have been introduced under the name asymptotically output-memoryless channels. We use the new name to emphasize the connection to the corresponding strong mixing condition.

If the channel κ is stationary, then (9) is true for all $s \in T$, if it is true for $s = 0$. The same holds for (10) and (11).

The definition of finite output memory is equivalent to the one given in [13] and note (11) is equivalent to the conditional independence of $\eta_{-\infty}^s$ and η_{s+t}^∞ given $\xi = x$ for all $x \in X$.

The mixing conditions formulated in the definition characterize channels with infinite output memory of the following type:

For given input, future and present outputs are asymptotically independent from outputs remote in the past. Both conditions imply the strong mixing condition given in [5]. The property defined there is inspired by the strong mixing condition in the ergodic-theoretic sense, whereas the conditions introduced here are related to the strong mixing concepts studied in [7].

The next result is the counterpart of Proposition II.6 for channels. The reversed implications of part (a) are not true in general as seen from Example V.3.a and the comment before Proposition II.6. Part (b) is remarkable and shows that for the important class of Gaussian channels the ψ-mixing condition is in fact a finite memory condition.

(III.5) Proposition. *(a) A channel with finite output memory is ψ-mixing and a ψ-mixing channel is information regular.*

(b) A ψ-mixing Gaussian channel has finite output memory.

Proof. Part (a) follows directly from Definition III.3 and from the conditional version of (4).

For part (b) assume the channel κ is ψ-mixing and Gaussian. From (9) we obtain that for any $s \in T$ there exists a nonnegative $t_0(s) \in T$, such that for all $t \geq t_0(s)$ we have

$$\sup_{x \in X} \psi(\eta_{-\infty}^{s}; \eta_{s+t}^{\infty} \,|\, \xi = x) < 2. \tag{12}$$

From [7, Prop. 3.11(a)] together with [7, Th. 9.7(I)] we conclude that if the ψ-coefficient between two jointly normal random variables is less than 2, then their correlation coefficient is zero. We use this result and the same argumentation as in the proof of [7, Th. 9.7(II)] to conclude from (12) that for any $x \in X$ the projections $\eta_{-\infty}^{s}$ and η_{s+t}^{∞} are (conditionally) independent. With the comment in the third paragraph of Remark III.4 it follows that a Gaussian ψ-mixing channel has finite output memory. □

IV. Application to a coding theorem

Assume κ is a stationary, continuous-time channel, μ is a probability measure on $\mathcal{X}$, and $\mu\kappa$ is the probability measure on $\mathcal{X} \otimes \mathcal{Y}$, generated according to (8). In [6] it has been shown that if κ is ψ-mixing and satisfies some further conditions, then

$$\lim_{t \to \infty} \psi(\xi_{-\infty}^{0}\eta_{-\infty}^{0}; \xi_{t}^{\infty}\eta_{t}^{\infty}) = 0 \tag{13}$$

holds, for the projections introduced at the end of Section I. From this convergence

$$\lim_{t \to \infty} I(\xi_{-\infty}^{0}\eta_{-\infty}^{0}; \xi_{t}^{\infty}\eta_{t}^{\infty}) = 0 \tag{14}$$

is concluded, which, in turn, is used to prove a coding theorem for continuous-time channels that also hold for infinite information capacity (see [6, Sec. 3(b), App. III]).

Subsequently, we will demonstrate that (14) holds already for information regular channels if we assume all the remaining constraints as in [6]. Thus, the considered coding theorem holds also under weaker conditions. Note, that the rest of the results in [6] are not affected by this modification. In particular, for the case of finite information capacity, the strong mixing condition of [5] is already sufficient to prove the theorem.

We need to introduce two more channel properties. The definitions are taken from [6], where the first is exactly adopted and the second is a modified version. However, note that this modified form is also necessary in [6] to obtain (13).

(IV.1) Definition. (Causality, asymptotic input-memorylessness) The channel κ is called causal if for any $t \in T$, $B \in \sigma(\eta_{-\infty}^{t})$, and $x, \tilde{x} \in X$ coinciding on $(-\infty, t]$

$$\kappa(x, B) = \kappa(\tilde{x}, B).$$

It is called asymptotically input-memoryless if for any $\epsilon > 0$ and $s \in T$ there exists a nonnegative $t(\epsilon, s) \in T$ such that for any $B \in \sigma(\eta_{s}^{\infty})$ and $x \in X$ with $\kappa(x, B) > 0$, and for any $\tilde{x} \in X$ coinciding on $(s - t(\epsilon, s), \infty)$ with x we have

$$\left|1 - \kappa(\tilde{x}, B)\big/\kappa(x, B)\right| < \epsilon.$$

(IV.2) Proposition. *If κ is a causal, asymptotically input-memoryless, information regular channel, and the channel input probability measure μ is such that the projections $\xi_{-\infty}^{0}$ and ξ_{0}^{∞} are independent, then* (14) *holds.*

Proof sketch. We apply (3) to upper-bound the mutual information in (14) by

$$\begin{aligned}
&I(\xi_{-\infty}^{0}\eta_{-\infty}^{0}; \xi_{0}^{\infty}\eta_{t}^{\infty}) \\
&= I(\xi_{-\infty}^{0}; \xi_{0}^{\infty}) \qquad + I(\eta_{-\infty}^{0}; \xi_{0}^{\infty} \,|\, \xi_{-\infty}^{0}) \qquad (15)\\
&\quad + I(\xi_{-\infty}^{0}; \eta_{t}^{\infty} \,|\, \xi_{0}^{\infty}) \quad + I(\eta_{-\infty}^{0}; \eta_{t}^{\infty} \,|\, \xi), \qquad (16)
\end{aligned}$$

where the equality results from using the chain rule of mutual information twice. Since $\xi_{-\infty}^{0}$ and ξ_{0}^{∞} are independent, the first term in (15) is zero. The second is also zero because the channel is causal, which implies the conditional independence of $\eta_{-\infty}^{0}$ and ξ_{0}^{∞} given $\xi_{-\infty}^{0}$. We can show that for asymptotically input-memoryless channels we have

$$\lim_{t \to \infty} \psi(\xi_{-\infty}^{0}; \eta_{t}^{\infty} \,|\, \xi_{0}^{\infty}) = 0. \tag{17}$$

This implies together with the conditional version of (4), that the first term in (16) converges to zero. Due to space limitations, we omit the proof of (17) but note that the derivations are similar to those in [6, App. I]. Finally, because the channel κ is information regular, the second term in (16) converges to zero, which follows from (7) and (10). □

As a result, the limit in (14) is zero under the conditions of Proposition IV.2, which are identical to those in [6], except that ψ-mixing is replaced by information regularity.

V. Examples

We consider examples, where the channel output is the result of a deterministic mapping f applied to the channel input and a random noise source λ. For a class of mappings we show that mixing properties of the noise source imply corresponding mixing properties of the channel, which provides a direct link to the field of mixing processes and the material in Section II. Finally, additive noise and fading channels with possibly infinite memory in the noise and fading process are discussed as special cases.

Consider the product measurable space $(Z, \mathcal{Z})$ and assume λ is a probability measure on $\mathcal{Z}$. Assume f is a measurable function on $(X \times Z, \mathcal{X} \otimes \mathcal{Z})$ with values in $(Y, \mathcal{Y})$. Then we define the channel κ with input $(X, \mathcal{X})$ and output $(Y, \mathcal{Y})$ by

$$\kappa(x, B) = \lambda(f(x, \cdot) \in B) = \int_{Z} \mathbb{1}_{B}\big(f(x, z)\big)\, \mathrm{d}\lambda(z), \tag{18}$$

for any $x \in X$, $B \in \mathcal{Y}$, where $\mathbb{1}_{B}(y)$ denotes the indicator function at point $y \in Y$, which is one if $y \in B$ and zero otherwise. Such channels are also considered in [14], [12, Ch. 4.1].

In the next special case the channel output at time t is the result of a deterministic mapping f_t applied to the channel input and the random noise in the time interval $(t-u, t+v]$.

(V.1) Example. Let $0 \le u, v \in T$ be fixed. For any $w \in T$ let f_w be a measurable function on $(X_{w-u}^{w+v} \times Z_{w-u}^{w+v}, \mathcal{X}_{w-u}^{w+v} \otimes \mathcal{Z}_{w-u}^{w+v})$ with values in $(Y_w, \mathcal{Y}_w)$. Further, for any $s \le t \in \overline{T}$ let f_s^t denote the function on $(X_{s-u}^{t+v} \times Z_{s-u}^{t+v}, \mathcal{X}_{s-u}^{t+v} \otimes \mathcal{Z}_{s-u}^{t+v})$ with values in $(Y_s^t, \mathcal{Y}_s^t)$, which is build from the functions f_w according to the construction in Lemma A.1. Due to the lemma the function f_s^t is measurable. The channel κ is defined by (18) with $f = f_{-\infty}^{\infty}$.

(V.2) Proposition. *Consider the channel in Example V.1.*

(a) If the noise source λ is ψ-mixing (information regular, τ-dependent), then the channel κ is ψ-mixing (information regular, has finite output memory).

(b) If λ is stationary and $f_t^{-1}(x_{t-u}^{t+v}, \mathcal{Y}_t) = \mathcal{Z}_{t-u}^{t+v}$ for all $t \in T$ and $x_{t-u}^{t+v} \in X_{t-u}^{t+v}$, then the converse of (a) *is also true.*

Proof. We adopt the notation of Example V.1 and denote by ζ_s^t the projection from Z to Z_s^t (see end of Section I). Since sections of measurable functions are measurable $f_s^t(x_{s-u}^{t+v}, \cdot)$ is $\mathcal{Z}_{s-u}^{t+v}/\mathcal{Y}_s^t$-measurable for all $x_{s-u}^{t+v} \in X_{s-u}^{t+v}$ according to Lemma A.1. Assume $\tilde{s} \in T$ and $u + v \le \tilde{t} \in T$, then

$$
\begin{aligned}
&\sup_{x\in X} \psi(\eta_{-\infty}^{\tilde{s}}; \eta_{\tilde{s}+\tilde{t}}^{\infty} \,|\, \xi = x) \\
=\ &\sup_{x\in X} \psi(f_{-\infty}^{\tilde{s}}(x_{-\infty}^{\tilde{s}+v}, \zeta_{-\infty}^{\tilde{s}+v})\,;\, f_{\tilde{s}+\tilde{t}}^{\infty}(x_{\tilde{s}+\tilde{t}-u}^{\infty}, \zeta_{\tilde{s}+\tilde{t}-u}^{\infty})) \quad (19)\\
\le\ &\sup_{x\in X} \psi(\zeta_{-\infty}^{\tilde{s}+v}; \zeta_{\tilde{s}+\tilde{t}-u}^{\infty}) \quad (20)\\
=\ &\psi(\zeta_{-\infty}^{\tilde{s}+v}; \zeta_{\tilde{s}+\tilde{t}-u}^{\infty}), \quad (21)
\end{aligned}
$$

where (19) follows from the definition of the channel according to (18), the definition of the mapping f, and from the remarks before Definition III.3. The inequality in (20) is due to (2) and we obtain part (a) for the ψ-mixing condition. The equality $f_t^{-1}(x_{t-u}^{t+v}, \mathcal{Y}_t) = \mathcal{Z}_{t-u}^{t+v}$ for all $t \in T$, $x_{t-u}^{t+v} \in X_{t-u}^{t+v}$ yields together with Lemma A.1 equality in (20). Moreover, if λ is stationary, then (21) is equal to $\psi(\zeta_{-\infty}^{0}; \zeta_{\tilde{t}-(u+v)}^{\infty})$ and with Remark II.5 we obtain part (b). The proof is identical for information regularity and τ-dependence / finite memory. □

We complete the section by discussing concrete examples including those for which the original form of the coding theorem in [6] does not apply but for which the generalized version considered in Section IV holds.

(V.3) Example. Consider the channel of Example V.1 either with discrete or continuous time, i. e., $T = \mathbb{Z}$ or $T = \mathbb{R}$. For the sake of simplicity we assume real scalar inputs and outputs, i. e., $X_t = Y_t = \mathbb{R}$ for all $t \in T$. Modifications with subsets of $\mathbb{R}$ or vectors are straightforward. The σ-algebras in the example are the usual Borel σ-algebras.

(a) *Additive noise channel*. We assume real-valued noise samples, i. e., $Z_t = \mathbb{R}$ for all $t \in T$. We have $u = v = 0$ and the channel is specified by

$$y_t = f_t(x_t, z_t) = x_t + z_t, \quad x_t \in X_t,\ z_t \in Z_t,\ t \in T.$$

The noise source λ is the distribution of the random process $\{\zeta_t\}$, $t \in T$, of projections, which models the additive noise.

Proposition V.2.a applies and if λ is stationary, then even the converse is true. If ζ is a Gaussian process, then the channel is Gaussian and Proposition III.5.b applies. In particular, if the additive noise results from passing stationary white Gaussian noise through a linear filter (see Example II.7), then the channel is information regular but not ψ-mixing.

(b) *Multipath fading channel*. Let $l \in \mathbb{N}$ be the number of signal path and $0 \le \tau_1 < \ldots < \tau_l \in T$ be the path delays. We assume real-valued fading coefficients and real-valued additive noise samples, i. e., $Z_t = Z_t^{\odot} \times Z_t^{\oplus} = \mathbb{R}^l \times \mathbb{R}$. We have $u = \tau_l + 1$, $v = 0$ and the channel is specified by

$$y_t = f_t(x_{t-\tau_l}, \ldots, x_{t-\tau_1}, z_t) = \textstyle\sum_{k=1}^{l} z_{t,k}^{\odot} x_{t-\tau_k} + z_t^{\oplus},$$

$$x_{t-\tau_k} \in X_{t-\tau_k},\ z_t = (z_t^{\odot}, z_t^{\oplus}) \in Z_t,\ t \in T.$$

The overall random noise process $\{\zeta_t\}$, $t \in T$, is composed of the fading (vector) process $\{\zeta_t^{\odot}\}$ and the additive noise process $\{\zeta_t^{\oplus}\}$, where $\zeta_t = (\zeta_t^{\odot}, \zeta_t^{\oplus})$ with $\zeta_t^{\odot}$ and $\zeta_t^{\oplus}$ being the projections from Z to $Z_t^{\odot}$ and $Z_t^{\oplus}$. The noise source λ is such that $\zeta^{\odot}$ and $\zeta^{\oplus}$ are independent. If $l = 1$ and $\tau_1 = 0$, then we have a flat fading channel.

Proposition V.2.a applies. If $\zeta^{\odot}$ and $\zeta^{\oplus}$ satisfy a certain mixing condition, then ζ (resp. the noise source λ) satisfies the same condition due to [7, Lem. 6.2]. If $\zeta^{\odot}$ and $\zeta^{\oplus}$ are Gaussian processes, then the channel is Gaussian and Proposition III.5.b applies. Based on Example II.7 we easily find multipath fading channels that are information regular but not ψ-mixing.

References

[1] A. I. Khinchin, *Mathematical Foundations of Information Theory*. Dover, 1957.

[2] K. Takano, “On the Basic Theorems of Information Theory,” *Ann. Inst. Stat. Math.*, vol. 9, no. 1, pp. 53–77, 1957.

[3] A. Feinstein, “On the Coding Theorem and Its Converse for Finite-Memory Channels,” *Inf. Control*, vol. 2, no. 1, pp. 25–44, 1959.

[4] J. Wolfowitz, *Coding Theorems of Information Theory*, 3rd ed. Springer, 1978.

[5] R. L. Adler, “Ergodic and Mixing Properties of Infinite Memory Channels,” *Proc. Amer. Math. Soc.*, vol. 12, no. 6, pp. 924–930, 1961.

[6] T. T. Kadota and A. D. Wyner, “Coding Theorem for Stationary, Asymptotically Memoryless, Continuous-Time Channels,” *Ann. Math. Statist.*, vol. 43, no. 5, pp. 1603–1611, 1972.

[7] R. C. Bradley, *Introduction to Strong Mixing Conditions*. Kendrick Press, 2007, vol. 1.

[8] S. Ihara, *Information Theory for Continuous Systems*. World Scientific, 1993.

[9] I. A. Ibragimov and Y. A. Rozanov, *Gaussian Random Processes*. Springer, 1978.

[10] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*. Holden-Day, 1964.

[11] A. N. Kolmogorov, “On the Shannon Theory of Information Transmission in the Case of Continuous Signals,” *IRE Trans. Inf. Theory*, vol. 2, no. 4, pp. 102–108, 1956.

[12] Y. Kakihara, *Abstract Methods in Information Theory*. World Scientific, 1999.

[13] T. T. Kadota, “Coding Theorem and its Converse for Continuous Incrementally Stationary Channels with Finite Incremental Memory,” *IEEE Trans. Inf. Theory*, vol. 18, no. 4, pp. 510–513, 1972.

[14] C. R. Baker, “Calculation of the Shannon Information,” *J. Math. Anal. Appl.*, vol. 69, pp. 115–123, 1979.

Appendix

(A.1) Lemma. *Let $0 \le u, v \in T$ and $s \le t \in \overline{T}$. Assume J is defined as below* (1)*. For any $w \in J$ let g_w be a measurable function on $(X_{w-u}^{w+v}, \mathcal{X}_{w-u}^{w+v})$ with values in $(Z_w, \mathcal{Z}_w)$. By g_s^t we denote the function on $(X_{s-u}^{t+v}, \mathcal{X}_{s-u}^{t+v})$ with values in $(Z_s^t, \mathcal{Z}_s^t)$, which is defined for any $x_{s-u}^{t+v} = \{x_w\} \in X_{s-u}^{t+v}$ by*

$$g_s^t(x_{s-u}^{t+v}) = z_s^t = \{g_w(x_{w-u}^{w+v})\},\ w \in J.$$

The function g_s^t is measurable and if $g_w^{-1}(\mathcal{Z}_w) = \mathcal{X}_{w-u}^{w+v}$ holds for all $w \in J$, then $(g_s^t)^{-1}(\mathcal{Z}_s^t) = \mathcal{X}_{s-u}^{t+v}$.