RECENT ADVANCES IN COMMUNICATION AND CONTROL THEORY

EDITED BY

R.E. KALMAN
G.I. MARCHUK
A.E. RUBERTI
A.J. VITERBI

RECENT ADVANCES IN COMMUNICATION AND CONTROL THEORY

EDITED BY
R.E. KALMAN
G.I. MARCHUK
A.E. RUBERTI
A.J. VITERBI

OPTIMIZATION SOFTWARE, INC.
PUBLICATIONS DIVISION, NEW YORK

Editors

Rudolph E. Kalman
Mathematische Systemtheorie
ETH-Zentrum, Hauptgebaude
CH-8092 Zurich
Switzerland

Gurij I. Marchuk
Department of Numerical Mathematics
The Academy of Sciences of the U.S.S.R.
11 Gorky Street
Moscow 103009, U.S.S.R.

Antonio Ruberti
President
Università degli Studi di Roma
"La Sapienza"
Piazzale Aldo Moro, 2
1-0085 Roma, Italy

Andrew J. Viterbi
Qualcomm, Inc.
10555 Sorrento Valley Road
San Diego, California 92121
U.S.A.

Library of Congress-in-Publication Data

Recent advances in communication and control systems.

"Volume honoring A.V. Balakrishnan on his 60th birthday"—Half t.p.
Bibliography: p.
1. Control theory. 2. Statistical communication theory. 3. Balakrishnan, A.V. I. Kalman, R.E. (Rudolph Emil), 1930- . II. Balakrishnan, A.V.
QA402.3.R4 1987 629.8'312 87-18604
ISBN 0-911575-46-4

SERIES IN COMMUNICATION AND CONTROL SYSTEMS

Translations Series in Mathematics and Engineering

M. I. Yadrenko
Spectral Theory of Random Fields
1983, 267 pp., ISBN 0-911575-00-6

G. I. Marchuk
Mathematical Models in Immunology
1983, 378 pp., ISBN 0-911575-01-4

A. A. Borovkov, ed.
Advances in Probability Theory: Limit Theorems and Related Problems
1984, 392 pp., ISBN 0-911575-03-0

V. A. Dubovitskij
The Ulam Problem of Optimal Motion of Line Segments
1985, 128 pp., ISBN 0-911575-04-9

N. V. Krylov, R. Sh. Liptser, and A. A. Novikov, eds.
Statistics and Control of Stochastic Processes
1985, 521 pp., ISBN 0-911575-18-9

Yu. G. Evtushenko
Numerical Optimization Techniques
1985, 575 pp., ISBN 0-911575-07-3

V. F. Dem'yanov and L. V. Vasil'ev
Nondifferentiable Optimization
1985, 472 pp., ISBN 0-911575-09-X

A. A. Borovkov, ed.
Advances in Probability Theory: Limit Theorems for Sums of Random Variables
1985, 313 pp., ISBN 0-911575-17-0

A. V. Balakrishnan, A. A. Dorodnitsyn, and J. L. Lions, eds.
Vistas in Applied Mathematics: Numerical Analysis, Atmospheric Sciences, Immunology
1986, 396 pp., ISBN 0-911575-38-3

V. F. Kolchin
Random Mappings
1986, 224 pp., ISBN 0-911575-16-2

L. Telksnys, ed.
Detection of Changes in Random Processes
1986, 240 pp., ISBN 0-911575-20-0

V. F. Dem'yanov and A. M. Rubinov
Quasidifferential Calculus
1986, 301 pp., ISBN 0-911575-35-9

V. P. Chistyakov, B. A. Sevast'yanov, and V. K. Zakharov
Probability Theory for Engineers
1987, 175 pp., ISBN 0-911575-13-8

B. T. Polyak
Introduction to Optimization
1987, 464 pp., ISBN 0-911575-14-6

V. V. Ivanishchev and A. D. Krasnoshchekov
Control of Variable Structure Networks
1987, 154 pp., ISBN 0-911575-05-7

Comcon Conferences Proceedings

Proceedings of the First International Conference in Communication and Control Systems
Washington, D.C., June 18-20, 1987
Edited by N. DeClaris
Dec. 1987, 450 pp., ISBN 0-911575-47-2

Scientific Applications Software

Kalman Filtering Software
Diskette & User's Guide
1985, ISBN 0-911575-31-6
Updated version in Quick-Basic 1987

Volume honoring
A.V. Balakrishnan
on his 60th birthday

For his pioneering contributions to Communication and Control Theory

TABLE OF CONTENTS

PART I COMMUNICATION SYSTEMS

Information Theory

Signal Processing

Stochastic Processes

Stochastic Equations

Optimal Control

PART III OPTIMIZATION

PART IV MODELLING

PART I

COMMUNICATION SYSTEMS

INFORMATION THEORY

INFORMATION CAPACITY OF GAUSSIAN CHANNELS†

Charles R. Baker

Department of Statistics
University of North Carolina
Chapel Hill, North Carolina

Abstract. Information capacity of Gaussian channels is one of the basic problems of information theory. Shannon's results for white Gaussian channels and Fano's "waterfilling" analysis of stationary Gaussian channels are two of the best-known works of early information theory. Results are given here which extend to a general framework, these results and others due to Gallager and to Kadota, Zakai, and Ziv. The development applies to arbitrary Gaussian channels when the channel noise has sample paths in a separable Banach space, and to a large class of Gaussian channels when the noise has sample paths in a linear topological vector space. Solutions for the capacity are given for both matched and mismatched channels.

Introduction

The modern theory of information is largely based on the pioneering work of C.E. Shannon [1]. The contributions and importance of information theory to the advancement of technology are very well known, and need not be summarized here. However new applications of a different nature seem likely to arise in the not too far distant future. Some of these potential applications would require a much deeper development of the theory than has been needed heretofore. This is in part because of rapid advances in technology in areas such as computers and communications. Thus one may envision computers of such high capability that their optimum use will require mathe-

---

† Research supported by ONR Contract N00014-86-K-0039

matical models using infinite-dimensional methods. Fiber optics is already leading to communication channels of extremely high bandwidth. Also to be considered is the need to develop information-theoretic models and methods for applications which do not fit into the classical mold of a communications channel with stationary Gaussian noise or a discrete memoryless channel. On the one hand, some communication channels contain nonstationary noise as a major source of interference. In another direction, information theory is viewed as a means of evaluating and designing systems in areas such as image processing, artificial intelligence, and surveillance.

Thus, the scope of information theory as presently applied may require considerable expansion in order to meet the needs of the future. In particular, mathematical models may be needed for problems of a very general nature, including channels with memory which may be infinite-dimensional, nonstationary, and possibly non-Gaussian.

The present article gives a treatment of capacity for Gaussian channels in a very general setting: when the stochastic processes of interest induce measures on a linear topological vector space. The work is an extension of previous results for induced measures on a separable Hilbert space [2], [3]. Although the latter model will be sufficiently general for most applications, it is not likely to be adequate for a treatment of nonstandard applications such as random fields, artificial intelligence, and surveillance.

In the case of stochastic processes with sample functions belonging to a separable Hilbert space, the results given in [2] and [3] represent a substantial generalization of previous work. This previous work includes Shannon's original white noise channel [1], Gallager's further work on this model [4], Kadota, Zakai, and Ziv's work on the Wiener channel [5], and the results of Fano [6] and Gallager [4] for stationary Gaussian channels. All of this prior work makes various assumptions on the channel noise.

Of course, in practical applications the coding capacity is most important. Partial results in this area for these more general models have been obtained [7], [8]. It can be expected that more complete solutions of the coding capacity problem will require the availability of general results on information capacity such as those summarized here, since proofs of coding capacity typically involve use of the information capacity.

This paper discusses the general framework in which these problems have been solved, and summarizes solutions. Proofs will not be included; it will be seen that one can modify the proofs of the Hilbert space solutions given in [2] and [3]. This has already been done in [7] for the case of the "matched" channel analyzed in [2], and similar methods can be used for the "mismatched" channel considered in [3]. Thus, the development here will be limited to defining the framework of the problem, providing the supplementary details needed to adapt the Hilbert-space solutions and proofs of [2] and [3] to the present more general setup, and then stating the results.

Mutual Information and Channel Capacity

Let (X, β) and $(Y, \mathfrak{F})$ be two measurable spaces, with μ_{XY} a probability on $(X \times Y, \beta \times \mathfrak{F})$. For the sake of clarity, μ_{XY} is called a *joint* measure. Denote by μ_X and μ_Y the projections of μ_{XY} on (X, β) and $(Y, \mathfrak{F})$, $\mu_X \otimes \mu_Y$ the product measure on $(X \times Y, \beta \times \mathfrak{F})$. The (average) mutual information of μ_{XY} is defined to be

$$I(\mu_{XY}) = \sup_{N; C_1, \ldots, C_N} \sum_{n=1}^{N} \mu_{XY}(C_N) \log \frac{\mu_{XY}(C_N)}{\mu_X \otimes \mu_Y(C_N)}, \qquad (1)$$

where the supremum is over all $N \geq 1$ and all measurable partitions $C_1, \ldots, C_N$ of μ_{XY}. It follows immediately that $I(\mu_{XY}) = \infty$ when it is false that μ_{XY} is absolutely continuous with respect to $\mu_X \otimes \mu_Y$ ($\mu_{XY} \ll \mu_X \otimes \mu_Y$). When $\mu_{XY} \ll \mu_X \otimes \mu_Y$, then [9]

$$I(\mu_{XY}) = \int_{X \times Y} \left[\log \frac{d\mu_{XY}}{d\mu_X \otimes d\mu_Y}(x, y) \right] d\mu_{XY}(x, y) . \qquad (2)$$

Channel information capacity is defined as the supremum of $I(\mu_{XY})$ over all μ_{XY} in a suitable set. In the framework of most communication channels, to be used here, the channel model is defined as follows. A measure μ_{XN} on $(X \times Y, \beta \times \mathfrak{F})$ describes the statistical relationship between the message X and the channel noise N; usually, as we shall assume, $\mu_{XN} = \mu_X \otimes \mu_N$. The channel output Y is described by measure $\mu_Y = \mu_X \otimes \mu_N \circ g^{-1}$, where g is $(X \times Y, \beta \times \mathfrak{F})/(Y, \mathfrak{F})$-measurable. The joint measure μ_{XY} is then $\mu_X \otimes \mu_N \circ f^{-1}$, where $f(x, y) = (x, g(x, y))$. The most typical situation in engineering applications is for $g(x, n) = A(x) + n$, where A is an $(X, \mathfrak{F})/(Y, \beta)$-measurable coding function. In general, the capacity is then defined as $\sup_Q I(\mu_{XY})$, where Q is a set of constraints on all admissible pairs (μ_X, A) of message measures μ_X and coding functions A. However, if A is 1:1 and bimeasurable, then no information is lost due to A. That is, let $\mu_{SY} = [\mu_X \circ A^{-1} \otimes \mu_N] \circ h^{-1}$, where $h(x, y) = (x, x + y)$. If A is 1:1 and $\overline{(X, \beta)}^{\mu_X} / \overline{(Y, \mathfrak{F})}^{\mu_S}$ bimeasurable, then (1) shows that $I(\mu_{SY}) = I(\mu_{XY})$. If X and Y are Polish (complete, separable, metrizable), then by Kuratowski's Borel mapping theorem [10] any 1:1 Borel-measurable map $A: X \to Y$ is Borel-bimeasurable.

We shall assume here that $X = Y$, $\beta = \mathfrak{F}$, $A = I$ (identity), so that $g(x, y) = x + y$. The extension to the more general case can be obtained by either restricting attention to coding functions A which are 1:1 and bimeasurable, or else by computing the information lost due to a coding function which does not have these properties.

Mathematical Structure

The following assumptions will be made henceforth. E is a locally convex Hausdorff linear topological vector space over the real numbers, with topological dual E'. It will also be assumed that E is quasi-complete: every closed and bounded subset is complete. E must then be sequentially complete. $\sigma(E')$ will denote the cylindrical σ-field, generated by the elements of E', $\overline{\sigma(E')}^{\mu}$ the completion under the measure μ. For x in E and y in E', the value of y at the point x will be denoted by $\langle y, x\rangle$.

The noise measure μ_N will be defined on $(E, \sigma(E'))$. μ_N will be assumed to be Gaussian and zero-mean: $\mu_N \circ \ell^{-1}$ is a zero-mean Gaussian distribution on $\mathbf{R}$ for each ℓ in E'. μ_N will be assumed to have a covariance operator R_N: $E' \to E$. R_N is linear, self-adjoint and nonnegative: $\langle x, R_N y\rangle = \langle y, R_N x\rangle$ and $\langle x, R_N x\rangle \geq 0$ for all x, y in E'. μ_N has the characteristic function given by

$$\hat{\mu}_N(x) = \int_E e^{i\langle x,y\rangle}\, d\mu_N(y) = e^{-(1/2)\langle x,R_N x\rangle},$$

and $\langle x, R_N y\rangle = \int_E \langle x, u\rangle\langle y, u\rangle\, d\mu_N(u)$.

Under these assumptions, it is known [11] that there exists a unique Hilbert space H_N contained in E, such that the natural (canonical) injection j_N: $H_N \to E$ is continuous, $R_N = j_N j_N^*$, and H_N is the closure of range (R) under the inner product $\langle Ru, Rv\rangle_N = \langle u, Rv\rangle$. Here, H_N' is always identified with H_N. H_N is termed the reproducing kernel Hilbert space (RKHS) of R_N (or μ_N); it is actually the RKHS for the covariance function R_0: $E' \times E' \to \mathbf{R}$, $R_0(u, v) = \langle u, Rv\rangle$. It will be further assumed that H_N is separable; instances where this assumption is not necessary will be noted. If μ_N is Radon, then H_N is necessarily separable [12].

The message measure μ_X is a probability on $(E, \sigma(E'))$. The constraints to be imposed will ensure that μ_X has a covariance operator R_X: $E' \to E$; it can be assumed (WLOG) that μ_X has zero mean. As in the previous section, the measure of interest is μ_{XY}, defined by $\mu_{XY} = \mu_X \otimes \mu_N \circ f^{-1}$, where $f(x, y) = (x, x+y)$.

A basic result in the Shannon theory is that if the supports of μ_N and μ_X are restricted to be of finite dimension and the covariance of μ_X is fixed, then $I(\mu_{XY})$ is maximized when μ_X is Gaussian. From this one obtains the result that the channel capacity problem can be solved by assuming μ_X to be Gaussian (see [2, Lemma 6]). This assumption will be made henceforth.

The *observation* measure $\mu_Y = \mu_X \otimes \mu_N \circ g^{-1}$, where $g(x, y) = x + y$, is thus Gaussian, with covariance operator R_Y: $E' \to E$, $R_Y = R_X + R_N$. Of course, R_Y has a RKHS H_Y contained in E and $R_Y = j_Y j_Y^*$, where j_Y: $H_Y \to E$ is the natural injection and is continuous.

The joint Gaussian measure μ_{XY} has a joint covariance operator $\mathcal{R}_{XY}: E' \times E' \to E \times E$ [13], [7]. This operator and its properties are characterized by the following result. It does not require that H_N be separable. Moreover, the result holds for any joint Gaussian measure on $(E \times E, \sigma(E') \times \sigma(E'))$ having a covariance operator $\mathcal{R}_{XY}: E' \times E' \to E \times E$.

LEMMA 1 [13], [7].

(1) $\mathcal{R}_{XY} = \mathcal{J}(\mathcal{I}+\mathcal{V})\mathcal{J}^*$, where $\mathcal{J}: H_X \times H_Y \to E \times E$ is the natural injection, $\mathcal{I}$ is the identity in $E \times E$, and $\mathcal{V}$ is a self-adjoint bounded linear operator in $H_X \times H_Y$ with $\|\mathcal{V}\| \leq 1$.

(2) $\mathcal{V}(x,y) = (V_{XY}y, V_{XY}^* x)$, where $V_{XY}: H_Y \to H_X$ is a bounded linear operator with $\|V_{XY}\| \leq 1$. The operator V_{XY} is uniquely defined by $\int_E \langle u, x\rangle\langle v, y\rangle \, d\mu_{XY}(x, y) = \langle u, j_X V_{XY} j_Y^* v\rangle$ for all u, v in E'.

(3) $I(\mu_{XY}) < \infty$ if and only if V_{XY} is Hilbert-Schmidt with $\|V_{XY}\| < 1$.

(4) When V_{XY} is Hilbert-Schmidt with $\|V_{XY}\| < 1$, then $I(\mu_{XY}) = -\frac{1}{2}\Sigma_n \log(1-\gamma_n)$, where (γ_n) are the eigenvalues of $V_{XY}V_{XY}^*$.

Lemma 1 is fundamental to the solution of the channel capacity problem. It enables one to calculate the mutual information, yielding the following result.

LEMMA 2 [2], [7]. Suppose that μ_X is Gaussian. Then:

(1) $I(\mu_{XY}) < \infty$ if and only if $\overline{\mu}_X[\text{range}(j_N)] = 1$, where $\overline{\mu}_X$ is the extension of μ_X to $\overline{\sigma(E')}^{\mu_X}$;

(2) $I(\mu_{XY}) < \infty$ if and only if $R_X = j_N T j_N^*$, where $T: H_N \to H_N$ is trace-class. When this is satisfied, $I(\mu_{XY}) = \frac{1}{2}\Sigma_n \log(1+\tau_n)$, where (τ_n) are the eigenvalues of T.

If the RKHS H_N is not separable, then part (1) of Lemma 2 holds with the condition $\overline{\mu}_X[\text{range}(j_N)] = 1$ replaced by $\mu_X^*[\text{range}(j_N)] = 1$, where μ_X^* is the outer measure obtained from μ_X [7]. The following result is then useful.

LEMMA 3 [7]. Suppose that E is a locally convex l.t.v.s., μ a probability measure on $(E, \sigma(E'))$. Suppose that B is a separable or reflexive Banach space and that $j: B \to E$ is a continuous linear injection. Then, the following are equivalent:

(1) $\mu^*[j(B)] = 1$;

(2) $\mu = \nu \circ j^{-1}$, where ν is a unique probability measure on $(B, \sigma(B'))$.

If (1) or (2) holds, then μ is Gaussian if and only if ν is Gaussian. If B is both separable and reflexive, then $j[B] \in \overline{\sigma(E')}^{\mu}$ so that (1) is equivalent to $\overline{\mu}[j(B)] = 1$.

In the mismatched channel to be considered subsequently, the constraints are given in terms of the norm for another Hilbert subspace of E. The following result is then useful. It does not require that H_N be separable.

PROPOSITION 1. Suppose H_W is a Hilbert subspace of E. Let $j_W: H_W \to E$ be the natural injection map, and suppose that $\mu_X^*[\text{range } (j_W)] = 1$. Then $I(\mu_{XY}) < \infty$ if and only if H_W is a vector subspace of H_N. If $H_W \subset H_N$, then j_W is continuous, the natural injection $J: H_W \to H_N$ is continuous, and H_W is the RKHS for the covariance operator $j_W j_W^*$; if H_N is separable, then H_W is also separable.

Proof. Suppose that $H_W \subset H_N$. Since H_W is a Hilbert space contained in the RKHS H_N, H_W is also a RKHS of functions on E' and $\|Jx\|_N^2 \leq k\|x\|_W^2$ for all x in H_W, some $k < \infty$ [12], so that the natural injection $J: H_W \to H_N$ is continuous. Since $j_W = j_N J$, j_W must also be continuous, so that $j_W j_W^*$ is a covariance operator mapping $E' \to E$. By definition, H_W is the (unique) RKHS for $j_W j_W^*$. To see that H_W is separable (assuming that H_N is separable), one notes that the linear map $L: H_N \to H_W$, $L j_N^* u = j_W^* u$, is continuous and has dense range in H_W, so that L^* has only $\{0\}$ in its null space. Thus, if $\{u_n, n \geq 1\}$ is such that $\{j_N^* u_n, n \geq 1\}$ is dense in H_N, then $\{L j_N^* u_n, n \geq 1\}$ must be dense in H_W. $I(\mu_{XY}) < \infty$, by Lemma 2, since $\bar{\mu}_X[\text{range } (j_N)] = 1$.

If H_W is not contained in H_N, then there exists z in range (j_W), $z \notin$ range (j_N). The Gaussian measure μ_X with covariance $z \otimes z$ has $\mu_X^*[\text{range } (j_N)] = 0$; by Lemma 2, $I(\mu_{XY}) = \infty$. □

Constraints

The constraints that will be used to define the admissible set Q of message measures μ_X are the following:

$$\mu_X^*[\text{range } (j_W)] = 1, \tag{A.1}$$

$$\int_{H_W} \|x\|_W^2 \, d\nu_X(x) \leq P, \tag{A.2}$$

where H_W is a Hilbert space contained in H_N, with norm $\|\cdot\|_W$, $j_W: H_W \to E$ is the natural injection, and ν_X is the Borel measure on H_W satisfying $\mu_X = \nu_X \circ j_W^{-1}$.

Since we wish to have the constraint (A.2) apply a.e. $d\mu_X$, it is first necessary to require (A.1). The existence of the measure ν_X such that $\mu_X = \nu_X \circ j_W^{-1}$ follows from Lemma 3; H_W is separable, from Proposition 1. Also by Proposition 1, the capacity will be infinite if H_W is not a vector subspace of H_N.

The constraint (A.2) is motivated by the typical application when E is $L_2[0, T]$. In the case of formal white noise, the constraint is usually $E \int_0^T X_t^2(\omega)\, dt \le P$. This can be viewed as a constraint on $E\|X\|_W^2$, where W is the RKHS of the identity operator; this is the covariance of formal white noise. When white noise is viewed as the formal derivative of the Wiener process, then the "integrated" channel is analyzed [5]. In that case, the transmitted signal X is defined by $X_t = \int_0^t u(s)\, ds$, u in $L_2[0, T]$ and the constraint is typically $E\|U\|_{L_2}^2 \le PT$. $\|\dot{x}\|_{L_2}^2$ is the norm of x in the RKHS of Wiener measure. Finally, one may note that in his treatment of stationary power-and-frequency-limited Gaussian channels when the noise has integrable spectral density [4], Gallager first assumes a constraint on the message of the form $E\|X\|_{L_2[T]}^2 \le PT$. However, the transmitted signal is obtained by passing the message through a linear filter whose transfer function G satisfies

$$\int_{-\infty}^{\infty} \frac{|G(\lambda)|^2}{\phi_N(\lambda)}\, d\lambda < \infty,$$

where ϕ_N is the noise spectral density. Such a transmitted signal satisfies both (A.1) and a constraint of the type (A.2), with (assuming that $|G|^2/\phi_n$ is bounded) an upper bound of

$$E_X \|X\|_{N,T}^2 \le \frac{PT}{2\pi} \sup_{\lambda} \frac{|G(\lambda)|^2}{\Phi_N(\lambda)}$$

for any $T > 0$, where now X refers to the filtered message and $\|\cdot\|_{N,T}$ is the RKHS of the noise covariance for the interval $[0, T]$. Of course, the constraint (A.2) is not placed explicitly on the transmitted signal in [4]; instead it appears in the solution for the capacity. Gallager's analysis is for the water-filling model treated by Fano [6]. Fano's treatment does not yield finite capacity, precisely because the constraints (A.1) and (A.2) are not imposed.

In addition to its use in previous more specialized analyses, the use of a Hilbert space norm is plausible in light of two other considerations. First, as can be seen from Lemma 2, the capacity will be infinite unless the constraint used implies $E_X \|X\|_N^2 \le P'$ for some $P' < \infty$. Proposition 1 shows that H_W must be a RKHS of functions on E' if the capacity is to be finite. Second, a RKHS norm actually places a dual constraint on the signal; this corresponds to limitations on the amount and frequency distribution of the signal energy in typical applications.

The capacity subject to the constraints (A.1) and (A.2) will be denoted by $\mathscr{C}_W(P)$. If $H_W = H_N$ (consisting of the same elements and the identical inner product), then the capacity will be denoted by $\mathscr{C}_N(P)$ and the channel is said to be *matched* (to the constraint). If $H_W \ne H_N$ as Hilbert spaces,