(4)

FINAL TECHNICAL REPORT

GRANT N00014-90-J-1712

THEORY AND APPLICATIONS OF RANDOM FIELDS

Office of Naval Research
Arlington, Virginia

Gary L. Wise
Department of Electrical and Computer Engineering
and
Department of Mathematics
The University of Texas at Austin
Austin, Texas 78712

March 31, 1993

INTRODUCTION

This Final Technical Report constitutes a summary of the research performed under Grant N00014-90-J-1712 during the period of April 1, 1990 through September 30, 1992. First we present a list of the personnel involved in the research effort. Then in the following section we present a brief summary of the research results that have been achieved. Each of these results is well documented in technical articles, and references to these articles are made in the summary of the research results. We hope you find these interesting.

PERSONNEL

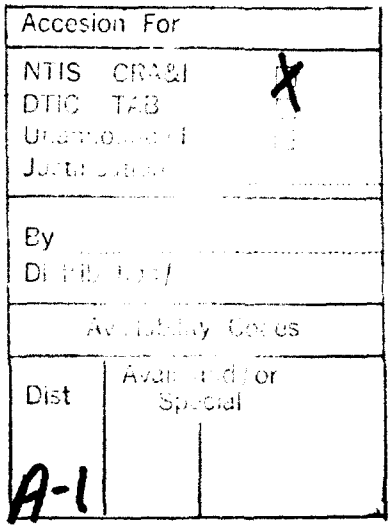
Principal Investigator:

Gary L. Wise

Post-Doctoral Research Associates:

Eric B. Hall

Alan E. Wessel



Statement A per telecon Julia Abrahams
ONCR/Code 1111
Arlington, VA 22217-5000

NWW 4/8/93

A SURVEY OF RESULTS

In this final technical report we will briefly comment upon our research accomplishments sponsored by Grant N00014-90-J-1712. Much of our work during this period was concerned with various aspects of random fields. The principal subareas of research activity may be characterized by the following:

Research in distributed estimation which might readily arise in a typical estimation problem in the context of random fields,

Optimal estimation with respect to a large family of cost functions,

Decentralized estimation with nontraditional fidelity criteria,

Multidimensional quantization which could arise in the effort to quantize a random field,

Multidimensional convolution,

The concept of finite memory of a stochastic process or a random field,

Distribution of the determinant of a random matrix,

Stationary random processes,

Zero-crossing rates for Gaussian processes,

Martingale characteristics of a Weiner process,

Estimation of a random variable based on multidimensional data,

Detection Theory versus Hypothesis Testing,

Importance Sampling

, and

Mutual Independence

A. Distributed Estimation

The results achieved in the area of distributed estimation are found in Appendix A in the paper entitled, "Some Aspects of Fusion in Estimation Theory," which appeared in the the March 1991 issue of *IEEE Transactions on Information Theory*. In this paper we considered the problem of fusion in estimation theory. We presented several examples, using common distributions, in which virtually any method of fusion would be useless in approximating the random variable of interest. Further, we presented a theorem which for a very general situation shows that fusion resulting in an almost surely exact approximation is always possible. In particular, this result addressed the situation in which the data consisted of the random variable of interest corrupted by additive Gaussian noise and the random variable of interest could be any second order random variable. An example was presented which illustrates the utility of this result.

Some Aspects of Fusion in Estimation Theory

Eric B. Hall
Alan E. Wessel
Gary L. Wise

Reprinted from
IEEE TRANSACTIONS ON INFORMATION THEORY
Vol. 37, No. 2, March 1991

Some Aspects of Fusion in Estimation Theory

Eric B. Hall, Alan E. Wessel, and Gary L. Wise

Abstract—**The problem of fusing or combining various estimates to obtain a single good estimate is investigated. Several examples are presented in which virtually any method of fusion fails. Finally, a very general situation is considered and an example is presented in which almost surely exact fusion is always possible.**

Index Terms—**Fusion, distributed estimation, conditional expectation.**

I. Introduction

In this correspondence we consider the problem of fusion in estimation theory. Our primary concern is directed toward finding a method of fusing or combining a finite number of estimates of a fixed second order random variable X in order to achieve a single "best" estimate of X. Our concern throughout this paper is directed toward minimizing the mean-square error (mse). In this context, for an arbitrary probability space $(\Omega, \mathcal{F}, P)$, we recall [1] that it is necessary to take versions of conditional expectations which are Borel measurable functions of the conditioned random variables, and we do so throughout the correspondence.

As an example of fusion in estimation, if Y_1 and Y_2 are random variables, how might $E[X|Y_1]$ and $E[X|Y_2]$ be fused to obtain a good approximation to $E[X|Y_1, Y_2]$. Notice that the statistical knowledge required to calculate $E[X|Y_1]$ and $E[X|Y_2]$ is less than that required to calculate $E[X|Y_1, Y_2]$, since $E[X|Y_1]$ and $E[X|Y_2]$ can be obtained from the appropriate bivariate distributions, whereas $E[X|Y_1, Y_2]$ in general cannot. Also, Y_1 and Y_2 might not be simultaneously available to the person desiring $E[X|Y_1, Y_2]$, a situation occurring in the usual context of distributed estimation in which a central location desires to construct a good estimate based on the estimates obtained by a set of spatially distinct field observers. Although other authors have attempted to address the problem of fusion in estimation theory (see for example [2]–[5]), many important questions in this area have not been resolved. In this correspondence we place the problem of fusion in estimation theory in a rigorous setting and address and answer several crucial questions. For a treatment of fusion via linear combinations of best linear estimates, of fusion via linear combinations of best Borel measurable estimates, and for other comments on this general problem, we refer the reader to [6].

Manuscript received October 2, 1989; revised August 29, 1990. This work was supported in part by the Office of Naval Research under Grant N00014-90-J-1712 and in part by the Air Force Office of Scientific Research under Grant AFOSR-86-0026. This work was presented in part at the 26th Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, September 28–30, 1988.

E. B. Hall is with the Department of Electrical Engineering, Southern Methodist University, Dallas, TX 75275.

A. E. Wessel is with the Department of Mathematics, Santa Clara University, Santa Clara, CA 95053.

G. L. Wise is with the Department of Statistics, University of California–Berkeley, and is also with the Department of Electrical and Computer Engineering and the Department of Mathematics, University of Texas at Austin, Austin, TX 78712.

IEEE Log Number 9040910.

II. Some Difficulties and Curiosities

We will now consider the general problem of fusion in which a Borel measurable transformation of best L_2 Borel measurable estimates is sought. In particular, if X is a second-order random variable, if n is a positive integer, and if $Y_1, Y_2, \cdots, Y_n$ are n random variables, how may $E[X|Y_1]$, $E[X|Y_2]$, $\cdots$, and $E[X|Y_n]$ be fused to approximate $E[X|Y_1, Y_2, \cdots, Y_n]$? In other words, under what conditions is $E(X|E[X|Y_1], E[X|Y_2], \cdots, E[X|Y_n])$ (the best L_2 estimate of X based on a Borel measurable function of $E[X|Y_1]$, $E[X|Y_2]$, $\cdots$, and $E[X|Y_n]$) a good approximation of $E[X|Y_1, Y_2, \cdots, Y_n]$? As the following examples indicate, there are numerous subtleties that arise in this context. For example, as will be seen, even if

$$E[X|Y_1] = E[X|Y_2] = \cdots = E[X|Y_n] \text{ a.s.},$$

$E[X|Y_1, Y_2, \cdots, Y_n]$ could be wildly different from $E[X|Y_1]$.

Example 1: Let $\Omega = [0, 1]$, $\mathcal{F}$ denote the Borel subsets of Ω, and P denote Lebesgue measure on $\mathcal{F}$. Let A be a positive real number, $\sigma(Y_1) = \sigma([0, 1/2))$, $\sigma(Y_2) = \sigma([1/4, 3/4))$, and $X(\omega) = A$ for $\omega \in [0, 1/4) \cup [1/2, 3/4)$ and $X(\omega) = -A$ for $\omega \in [1/4, 1/2) \cup [3/4, 1]$. Then it straightforwardly follows that $E[X|Y_1] = E[X|Y_2] = 0$ a.s., but $E[X|Y_1, Y_2] = X$, a.s. Notice that in this special case, any linear combination of $E[X|Y_1]$ and $E[X|Y_2]$ yields an estimate equal to 0 a.s., resulting in a mse in approximating X of A^2, which can exceed any preassigned real number. Recalling that $E[X|Y_1]$ and $E[X|Y_2]$, respectively, are $\sigma(Y_1)$-measurable and $\sigma(Y_2)$-measurable, we see that $E[X|Y_1] = E[X|Y_2] = 0$ pointwise in ω; similarly, we see that $E[X|Y_1, Y_2] = X$ pointwise in ω. Thus, in this situation, it is fruitless to attempt to approximate X based on *any* function of $E[X|Y_1]$ and $E[X|Y_2]$.

Note that in Example 1, Y_1 and Y_2 are independent, X and Y_1 are independent, and X and Y_2 are independent. This might have led an unwary investigator to perhaps assert that $E[X|Y_1, Y_2] = E[X]$ a.s., or perhaps that $E[X|Y_1, Y_2] = E[X|Y_i]$ a.s. for $i = 1$ or 2. Each of these assertions, which happen to be equivalent in the setting of Example 1, is incorrect.

Note, also, that Example 1 concerned simple random variables. The phenomenon exhibited in Example 1, however, can hold for nonsimple random variables as shown in the following three examples which involve more commonplace distributions of random variables.

Example 2: Let Y_1 and Y_2 be independent Gaussian random variables defined on the same probability space, each having

0018-9448/91/0300-0420$01.00 © 1991 IEEE

zero-mean and unit variance. Let $X = Y_1Y_2$. Then $E[X|Y_1] = E[X|Y_2] = 0$ a.s. whereas $E[X|Y_1, Y_2] = X$ a.s.

Example 3: Let Y_1 and Y_2 be independent random variables defined on the same probability space, such that Y_1 is uniformly distributed on $[-1, 1]$ and Y_2 has a probability density function given by $f(x) = x^2/\sqrt{2\pi}\exp(-x^2/2)$, and let $X = Y_1Y_2$. It can be shown that X has a Gaussian distribution with zero-mean and unit variance (cf. [7, pp. 172, 176]). Then again $E[X|Y_1] = E[X|Y_2] = 0$ a.s., whereas $E[X|Y_1, Y_2] = X$ a.s.

Example 4: For an integer $n > 1$, consider a set of random variables $\{X, Y_1, \cdots, Y_n\}$ with a joint probability density function given, as in [8], by

$$f(x, y_1, \cdots, y_n) = \left(\frac{1}{\sqrt{2\pi}}\right)^{n+1} \exp\left[\frac{-1}{2}\left(x^2 + \sum_{i=1}^{n} y_i^2\right)\right] \cdot \left[1 + x\exp\left(\frac{-x^2}{2}\right)\prod_{i=1}^{n}\left(y_i \exp\left(\frac{-y_i^2}{2}\right)\right)\right].$$

It follows straightforwardly that the set $\{X, Y_1, \cdots, Y_n\}$ is not mutually Gaussian and not mutually independent, yet any proper subset of $\{X, Y_1, \cdots, Y_n\}$ containing at least two random variables is mutually independent, mutually Gaussian, and identically distributed with each random variable having zero-mean and unit variance. For any nonempty proper subset $\mathcal{S}$ of $\{Y_1, \cdots, Y_n\}$, we note that $E[X|\mathcal{S}] = 0$ a.s. since X is independent of $\mathcal{S}$. However, it follows quickly that

$$E[X|Y_1, \cdots, Y_n] = \frac{1}{2\sqrt{2}} Y_1 \cdots Y_n \exp\left[\frac{-1}{2}\left(Y_1^2 + Y_2^2 + \cdots + Y_n^2\right)\right] \text{ a.s.}$$

Thus, since any Borel measurable function of the estimates $E[X|\mathcal{S}]$ where $\mathcal{S}$ ranges over all nonempty proper subsets of $(Y_1, \cdots, Y_n)$ would be constant almost surely, it would not be reasonable to attempt to estimate $E[X|Y_1, \cdots, Y_n]$ based on a combination of these estimates.

Notice that in Example 2 the observations are Gaussian, and in Example 3 the signal of interest is Gaussian. Further, in Example 4 the signal of interest is Gaussian, the observations are mutually Gaussian, and the problem under consideration is expanded to include fusion of estimates of the form $E[X|\mathcal{S}]$, where $\mathcal{S}$ is any nonempty proper subset of the observations. In each case, estimation of X via fusion is hopeless and even the ubiquitous Gaussian assumption does not alleviate this difficulty. However, as will be shown next, with an appropriate restriction on the observations, almost surely exact fusion is possible.

III. Fusion in a Practical Setting

Let $(\Omega, \mathcal{F}, P)$ be a fixed probability space on which all of the following random variables will be defined, and let n be a positive integer. We will now consider an approach motivated by more practical concerns. The following notation will be used throughout the remainder of this section. Let $Y_i = X + N_i$, where $X, N_1, N_2, \cdots, N_n$ are mutually independent, N_i represents additive noise, and X is a second-order random variable representing the signal of interest. As before, we consider the problem of estimating X via some combination of the $E[X|Y_i]$'s.

Assume that $N_1, N_2, \cdots, N_n$ each possesses a probability density function. Let F_X denote the distribution function of X, and f_{N_i} denote a Borel measurable density function of N_i. Further, notice that via a straightforward change of variable and using the independence assumption, we have that

$$E[X|Y_1, \cdots, Y_n] = \frac{\int_R x \prod_{i=1}^{n} f_{N_i}(Y_i - x)\, dF_X(x)}{\int_R \prod_{i=1}^{n} f_{N_i}(Y_i - x)\, dF_X(x)} \text{ a.s.}$$

Now, in addition to these assumptions, assume that for each positive integer $i \le n$, N_i has a zero-mean Gaussian distribution with a positive variance denoted by σ_i^2. Notice that in this case, choosing continuous versions of the density functions, we have

$$\begin{aligned}\prod_{i=1}^{n} f_{N_i}(Y_i - x) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(\frac{-(Y_i - x)^2}{2\sigma_i^2}\right) \\ &= \left[\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_i^2}}\right] \exp\left(\frac{-\lambda}{2}\left[x^2 - \frac{2Sx}{\lambda} + \frac{V}{\lambda}\right]\right) \\ &= K \exp\left(\frac{-\lambda}{2}\left[x - \frac{S}{\lambda}\right]^2\right),\end{aligned}$$

where we define

$$S = \sum_{i=1}^{n} \frac{Y_i}{\sigma_i^2}, \qquad V = \sum_{i=1}^{n} \frac{Y_i^2}{\sigma_i^2}, \qquad \lambda = \sum_{i=1}^{n} \frac{1}{\sigma_i^2},$$

and

$$K = \left[\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_i^2}}\right] \exp\left(\frac{-1}{2}\left[V - \frac{S^2}{\lambda}\right]\right).$$

Substitution into the previous expression for $E[X|Y_1, \cdots, Y_n]$ now implies that

$$E[X|Y_1, \cdots, Y_n] = \frac{\int_R x \exp\left[\frac{-\lambda}{2}\left(x - \frac{S}{\lambda}\right)^2\right] dF_X(x)}{\int_R \exp\left[\frac{-\lambda}{2}\left(x - \frac{S}{\lambda}\right)^2\right] dF_X(x)} \text{ a.s.} \quad (1)$$

For the special case where X has a density given by $1/\sqrt{2\pi\sigma^2}\exp(-x^2/2\sigma^2)$, we see that $E[X|Y_1, \cdots, Y_n] = S\sigma^2/(1 + \lambda\sigma^2)$ a.s. Furthermore, $E[X|Y_i] = Y_i\sigma^2/(\sigma_i^2 + \sigma^2)$ a.s. Thus, we see that in this situation,

$$E[X|Y_1, \cdots, Y_n] = \left(\frac{1}{1 + \lambda\sigma^2}\right) \sum_{i=1}^{n} \left(\frac{\sigma_i^2 + \sigma^2}{\sigma_i^2}\right) E[X|Y_i] \text{ a.s.}$$

In other words, $E[X|Y_1, \cdots, Y_n]$ is equal a.s. to a Borel measurable function of the $E[X|Y_i]$'s for this case when X is Gaussian. One might ask if such a result holds for any other distributions on X. The following theorem addresses this question and shows that, in the context of the previous assumptions, almost surely exact fusion is always possible for *any* second-order random variable X.

Theorem 1: Consider a probability space $(\Omega, \mathcal{F}, P)$ and random variables $X, N_1, \cdots, N_n$ defined on $(\Omega, \mathcal{F}, P)$ where n is a positive integer and X is a second-order random variable. Further, assume that for each positive integer $i \le n$, N_i has a zero-mean Gaussian distribution with positive variance given by σ_i^2, and that $X, N_1, \cdots, N_n$ are mutually independent. Define $Y_i = X + N_i$ for $i = 1, \cdots, n$. Then there exists a Borel measurable function $g: \boldsymbol{R}^n \to \boldsymbol{R}$ such that $E[X|Y_1, \cdots, Y_n] = g(E[X|Y_1], \cdots, E[X|Y_n])$ a.s.

Proof: If X is a.s. equal to a constant, the result is obvious. Assume that X is not almost surely equal to a constant. Using (1) it immediately follows that a version of the regression func-

tion $E[X|Y_i = y]$ is given by

$$\frac{\int_R x \exp\left(\frac{-x^2}{2\sigma_i^2} + \frac{yx}{\sigma_i^2}\right) dF_X(x)}{\int_R \exp\left(\frac{-x^2}{2\sigma_i^2} + \frac{yx}{\sigma_i^2}\right) dF_X(x)}.$$

This version will be used throughout the remainder of the proof. It now follows that

$$\frac{d}{dy} E[X|Y_i = y] \qquad \frac{\left[\frac{-x^2+2xy}{2\sigma_i^2}\right] dF_X(x) \int_R \frac{x^2}{\sigma_i^2} \exp\left[\frac{-x^2+2xy}{2\sigma_i^2}\right] dF_X(x) - \left[\int_R \frac{x}{\sigma_i} \exp\left[\frac{-x^2+2xy}{2\sigma_i^2}\right] dF_X(x)\right]^2}{\left[\int_R \exp\left[\frac{-x^2+2xy}{2\sigma_i^2}\right] dF_X(x)\right]^2}.$$

Notice that the denominator of this expression is positive. Further, the Schwarz inequality, which is a strict inequality since X is not a.s. equal to a constant, implies that the numerator is also positive. Thus, since $d/dy\, E[X|Y_i = y] > 0$, we see that $E[X|Y_i = y]$ is a strictly increasing function of y. Hence, there exists a Borel measurable function g_i so that $g_i(E[X|Y_i]) = Y_i$ a.s. Notice that

$$S = \sum_{i=1}^{n} \frac{1}{\sigma_i^2} g_i(E[X|Y_i]) \text{ a.s.}$$

Thus, substitution of this expression for S into (1) provides a Borel measurable function $g: \boldsymbol{R}^n \to \boldsymbol{R}$ such that

$$E[X|Y_1, \cdots, Y_n] = g(E[X|Y_1], E[X|Y_2], \cdots, E[X|Y_n]) \text{ a.s.}$$

□

Hence, Theorem 1 shows that almost surely exact fusion in the setting under consideration is always possible. Notice again that this result holds for *any* second-order random variable of interest. We next present an example which serves to illustrate the utility of Theorem 1.

Example 5: In the context of Theorem 1, let $X = 1$ with probability one half and let $X = -1$ with probability one half. Then it straightforwardly follows that a version of $E[X|Y_i = y]$ is given by $\tanh(y/\sigma_i^2)$. Now, fixing this version and adopting the notation used earlier in this section, we have that $S = \sum_{i=1}^{n} \tanh^{-1}(E[X|Y_i])$ a.s. Further, (1) simplifies to $E[X|Y_1, \cdots, Y_n] = \tanh(S)$ a.s. Hence, we see that

$$E[X|Y_1, \cdots, Y_n] = \tanh\left[\sum_{i=1}^{n} \tanh^{-1}(E[X|Y_i])\right] \text{ a.s.}$$

Thus, in this case we have, as guaranteed by Theorem 1, a precise expression where $E[X|Y_1, \cdots, Y_n]$ is equal a.s. to a specific Borel measurable transformation of the $E[X|Y_i]$'s.

IV. Conclusion

We considered the problem of fusion in estimation theory. We presented several examples, using common distributions, in which virtually any method of fusion would be useless in approximating the random variable of interest. Further, we presented a theorem which, for a very general situation, shows that fusion resulting in an almost surely exact approximation is always possible. In particular, this result addressed the situation in which the data consisted of the random variable of interest corrupted by additive Gaussian noise and the random variable of interest could be any second-order random variable. Finally, an example was presented that illustrates the utility of this result.

Acknowledgment

The authors wish to acknowledge helpful comments on the subject of this correspondence from Prof. Bruce Hajek at the University of Illinois at Urbana-Champaign.

References

[1] G. L. Wise, "A note on a common misconception in estimation," *Syst. Contr. Lett.*, vol. 5, pp. 355–356, Apr. 1985.

[2] A. T. Alouani and J. D. Birdwell, "Distributed estimation: Constraints on the choice of the local models," *IEEE Trans. Automatic Contr.*, vol. AC-33, pp. 503–506, May 1988.

[3] D. A. Castanon and K. Teneketzis, "Distributed estimation algorithms for nonlinear systems," *IEEE Trans. Automatic Contr.*, vol. AC-20, pp. 418–425, May 1985.

[4] J. L. Speyer, "Computation and transmission requirements for a decentralized linear-quadratic-Gaussian control problem," *IEEE Trans. Automatic Contr.*, vol. AC-24, pp. 266–269, Apr. 1979.

[5] A. S. Willsky, M. G. Bello, D. A. Castanon, B. C. Levy, and G. C. Verghese, "Combining and updating of local estimates and regional maps along sets of one-dimensional tracks," *IEEE Trans. Automatic Contr.*, vol. AC-27, pp. 799–813, Aug. 1982.

[6] E. B. Hall, A. E. Wessel, and G. L. Wise, "On fusion in estimation theory," in *Proc. Twenty-Sixth Ann. Allerton Conf. Commun., Contr., Comput.*, Monticello, IL, Sept. 28–30, 1988, pp. 599–608.

[7] L. Devroye, *Non-Uniform Random Variate Generation*. New York: Springer-Verlag, 1986.

[8] D. A. Pierce and R. L. Dykstra, "Independence and the normal distribution," *The American Statistician*, vol. 23, p. 39, Oct. 1969.

B. Optimal Estimation with Respect to a Large Family of Cost Functions

Our results pertaining to optimal estimation are found in Appendix B in the article entitled, "On optimal estimation with respect to a large family of cost functions," which appeared in the May 1991 issue of *IEEE Transactions on Information Theory*. In this article we considered the problem of optimal estimation of a random variable X based on an observation denoted by a random vector Y. We gave a mild restriction on the regular conditional distribution function of X given $\sigma(Y)$ that ensures that $E[\Phi(X-g(Y))]$ is minimized for any cost function Φ that is nonnegative, even, and convex. We showed that given any real valued Borel measurable function there exist random variables X and Y, possessing a joint density function, so that the chosen function is an optimal estimator, with respect to any of the cost functions previously described, of the random variable X based on the random variable Y. The results were then extended to estimation of X based upon a random variable that is measurable with respect to any given σ-subalgebra.

On Optimal Estimation with Respect to a Large Family of Cost Functions

Eric B. Hall and Gary L. Wise

Abstract—**Consider two random variables X and Y. A commonly encountered problem involves estimating X via $h(Y)$ so as to minimize $E[\Phi(X-h(Y))]$ where h is Borel measurable and Φ is a Borel measurable cost function chosen to adequately reflect the fidelity demands of the problem under consideration. This correspondence places a mild condition on the regular conditional distribution of X given $\sigma(Y)$ that ensures that $E[\Phi(X-h(Y))]$ is minimized for any cost function Φ that is nonnegative, even, and convex. In addition, it is shown that given any Borel measurable function $g: R \to R$, there exist random variables X and Y possessing a joint density function such that $E[X|Y=y]=g(y)$ a.e., with respect to Lebesgue measure.**

Index Terms—**Optimal nonlinear estimation, non-mean-square-error fidelity criteria, regression functions.**

I. Introduction

In this correspondence we consider the problem of estimation with respect to nontraditional cost functions. In an estimation problem one is often confronted with two concerns in choosing a cost function: the concern that the cost function adequately reflects the cost one wishes to associate with an error, and the concern that the cost function results in a problem which one finds to be mathematically tractable. Traditional cost functions, such as the quadratic cost function that is associated with the extremely popular mean-square-error criterion, are usually chosen solely on the basis of the second of these two concerns. As a result, the fidelity demands of the specific problem under consideration are rarely relied upon, and, in fact, are often not even considered, when determining the cost function which will be used. This sacrifice of suitability for mathematical ease in the choice of a cost function should be the cause of some concern since the traditional choices are unsuitable for many problems in estimation. This correspondence lessens this problem by extending the domain of mathematical tractability to include many cost functions that, even though pertinent to the subjective demands of many problems, have in the past been excluded from consideration.

Manuscript received August 27, 1990; revised October 25, 1990. This work was supported in part by the Office of Naval Research under Grant N00014-90-J-1712. This work was presented at the 27th Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, September 27–29, 1989.

E. B. Hall is with the Department of Electrical Engineering, Southern Methodist University, Dallas, TX 75275.

G. L. Wise is with the Department of Electrical and Computer Engineering and the Department of Mathematics, The University of Texas at Austin, Austin, TX 78712.

IEEE Log Number 9142965.

II. Development

In 1958, Seymour Sherman published a paper entitled "Non-Mean-Square Error Criteria" [1] in which he proposed conditions on a conditional distribution that would allow for the simultaneous minimization of a large family of cost functions. In [2] we provided a proof of Sherman's proposal and explored several extensions and practical consequences. Although Sherman's result had been widely quoted prior to [2], a correct proof seems to have been elusive. For example, several proofs [3, pp. 308–310]; [4, pp. 10–12]; [5, p. 61] using integration by parts were attempted even though the conditions placed on the cost function were insufficient to allow such a method to be used. (In particular, the proofs referenced above may fail for X and Y mutually Gaussian if a cost function $\Phi: R \to [0,\infty)$ is chosen which is even, strictly increasing on $[0,\infty)$, and singular. For an explicit counterexample the interested reader is referred to [6].)

For a topological space T, we will let $\mathcal{B}(T)$ denote the Borel subsets of T; and $I_S(\cdot)$ will denote the indicator function of the set S. Let N denote the set of positive integers. Also, recall that a probability distribution function $F: R \to [0,1]$ is said to be unimodal about $\gamma \in R$ if F is convex on $(-\infty,\gamma)$ and concave on (γ,∞), and a probability distribution function $F: R \to [0,1]$ is said to be symmetric if for all real x, $F(x)=1-\lim_{h \downarrow 0} F(-x-h)$. If k is a positive integer and $Y_1,\cdots,Y_k$ are k random variables defined on a common probability space, the random vector $Y=[Y_1,\cdots,Y_k]^T$ induces a probability measure on $\mathcal{B}(R^k)$; we will denote this resulting measure, conventionally known as the distribution of Y, by the notation μ_Y. Finally, we recall that for a random variable X and a σ-subalgebra $\mathcal{A}$, a regular conditional distribution function $F: R \times \Omega \to [0,1]$ always exists [7, pp. 263–264]; such a function is characterized by the following two conditions: for each $\omega \in \Omega$, $F(\cdot,\omega)$ is a probability distribution function, and, for each $x \in R$, $F(x,\omega)=P(X \le x|\mathcal{A})(\omega)$ a.s.

Sherman's original proposal (generalized and proved in [2]) required a regular conditional distribution function that, when properly shifted, is symmetric and unimodal about the origin and a cost function that is nonnegative, even, and nondecreasing to the right of the origin. For mutually Gaussian random variables X and Y it follows easily that the resulting regular conditional distribution function is symmetric and unimodal about $E[X|Y](\omega)$ for any fixed ω. This special case explains why Sherman's result is often invoked to add a flavor of generality to papers that consider Gaussian distributions. When one attempts to venture outside this somewhat limited arena, however, the conditions which Sherman placed on a regular conditional distribution function immediately begin to feel overly restrictive. The conditions on the cost function, however, are extremely nonrestrictive and, in fact, allow for many interesting, albeit impractical, choices. This imbalance suggests the possibility of lessening the restrictions on the regular conditional distribution function by perhaps slightly increasing the restrictions

0018-9448/91/0500-0691$01.00

imposed on the cost function. The following lemma will allow us to present such a result.

Lemma 1: Assume that F is a symmetric probability distribution function and that Φ: $\boldsymbol{R} \to [0,\infty)$ is even and convex. Then

$$\int_{\boldsymbol{R}} \Phi(x)\,dF(x) \le \int_{\boldsymbol{R}} \Phi(x-\alpha)\,dF(x), \qquad \text{for all } \alpha \in \boldsymbol{R}.$$

Proof: Since Φ is convex we see that $\Phi(x) \le (1/2)\Phi(x-\alpha) + (1/2)\Phi(x+\alpha)$. Further, since F is symmetric and Φ is even, we see that

$$\int_{\boldsymbol{R}} \Phi(x+\alpha)\,dF(x) = \int_{\boldsymbol{R}} \Phi(x-\alpha)\,dF(x).$$

Thus, we see that

$$\int_{\boldsymbol{R}} \Phi(x)\,dF(x) \le \frac{1}{2}\int_{\boldsymbol{R}} \Phi(x-\alpha)\,dF(x) + \frac{1}{2}\int_{\boldsymbol{R}} \Phi(x+\alpha)\,dF(x)$$

$$= \int_{\boldsymbol{R}} \Phi(x-\alpha)\,dF(x). \qquad \square$$

Lemma 1 will allow a result similar to that given in [1] to be stated for a much less restrictive family of regular conditional distribution functions by slightly restricting the family of allowable cost functions. In particular, we will be able to drop the restriction that the conditional distribution function be unimodal by requiring that the cost function, in addition to the previous restrictions, also be convex. Notice that requiring the cost function to be even and convex implies that it is also nondecreasing to the right of the origin.

We are now in a position to state and prove the following result.

Theorem 2: Let $k \in N$, $(\Omega, \mathcal{S}, P)$ be a probability space, and $X, Y_1, \cdots, Y_k$ be random variables defined on $(\Omega, \mathcal{S}, P)$, with X integrable. Let M: $\boldsymbol{R}^k \to \boldsymbol{R}$ be a Borel measurable function such that $M(Y_1(\omega), \cdots, Y_k(\omega)) = E[X|Y_1, \cdots, Y_k](\omega)$ a.s., and let F: $\boldsymbol{R} \times \Omega \to [0,1]$ be a regular conditional distribution function of X conditioned on $\sigma(Y_1, \cdots, Y_k)$ such that

$$F(x + M(Y_1(\omega), \cdots, Y_k(\omega)), \omega),$$

as a function of x with ω fixed, is symmetric. Then M minimizes the quantity $E[\Phi(X - f(Y_1, \cdots, Y_k))]$ over all Borel measurable functions f: $\boldsymbol{R}^k \to \boldsymbol{R}$ where Φ: $\boldsymbol{R} \to [0,\infty)$ is even and convex.

Proof: Lemma 1 and a change of variables imply that for each fixed ω and for $\alpha \in \boldsymbol{R}$,

$$\int_{\boldsymbol{R}} \Phi(x - M(Y_1(\omega), \cdots, Y_k(\omega)))\,dF(x,\omega)$$

$$\le \int_{\boldsymbol{R}} \Phi(x - \alpha - M(Y_1(\omega), \cdots, Y_k(\omega)))\,dF(x,\omega).$$

Let g: $\boldsymbol{R}^k \to \boldsymbol{R}$ be a Borel measurable function by which X is to be estimated and $E[\Phi(X - g(Y_1, \cdots, Y_k))]$ minimized. Note that

$$E[\Phi(X - g(Y_1, \cdots, Y_k))]$$

$$= E[E[\Phi(X - g(Y_1, \cdots, Y_k))|\sigma(Y_1, \cdots, Y_k)]].$$

From the preceding inequality and [8, p. 79], the inner expectation, and thus this expression, is minimized when $g(Y_1, \cdots, Y_k) = M(Y_1, \cdots, Y_k)$. $\square$

We will next present a useful corollary to Theorem 2. Let $k \in N$, $(\Omega, \mathcal{S}, P)$ be a probability space, X be a random variable defined on $(\Omega, \mathcal{S}, P)$, and Y be a random vector defined on $(\Omega, \mathcal{S}, P)$ taking values in $\boldsymbol{R}^k$. Recall that $F(x|Y = y)$ is said to be a regular conditional distribution function for X given $Y = y$ if for each fixed $y \in \boldsymbol{R}^k$, $F(x|Y = y)$ is a probability distribution function as a function of x, and for each fixed $x \in \boldsymbol{R}$, $F(x|Y = y)$ is a version of the regression function $E[I_{(-\infty,x]}(X)|Y = y]$. Further, recall that a regular conditional distribution function for X given $Y = y$ always exists [cf. 9, pp. 372–376]. The next corollary, which follows straightforwardly from Theorem 2, removes the need to work on the underlying probability space.

Corollary 3: Let $k \in N$, $(\Omega, \mathcal{S}, P)$ be a probability space, X be an integrable random variable defined on $(\Omega, \mathcal{S}, P)$, Y be a random vector defined on $(\Omega, \mathcal{S}, P)$ taking values in $\boldsymbol{R}^k$, and M: $\boldsymbol{R}^k \to \boldsymbol{R}$ be any Borel measurable function equal a.e. $[\mu_Y]$ to $E[X|Y = y]$. Further, assume that, as a function of x with y fixed, a regular conditional distribution function of X given $Y = y$, denoted by $F(x|Y = y)$, is such that

$$F(x + M(y)|Y = y)$$

is symmetric. Then $g = M$ minimizes

$$E[\Phi(X - g(Y))]$$

over all Borel measurable functions g: $\boldsymbol{R}^k \to \boldsymbol{R}$ where Φ: $\boldsymbol{R} \to [0,\infty)$ is even and convex.

III. A Non-Gaussian Application

The following example illustrates the usefulness of Theorem 2 and Corollary 3 and shows how these results may be applied to non-Gaussian distributions. In particular, given *any* real valued Borel measurable function $g(\cdot)$, we show that there exists a random variable X and a random variable Y, possessing a joint density function, so that $E[\Phi(X - h(Y))]$ is minimized when $h(\cdot) = g(\cdot)$ for any cost function Φ that is nonnegative, even, and convex.

Example: Let g: $\boldsymbol{R} \to \boldsymbol{R}$ be Borel measurable and define

$$f(x,y) = \frac{1}{8}\exp(-\exp(|y|)|x - g(y) + K|) + \frac{1}{8}\exp(-\exp(|y|)|x - g(y) - K|),$$

where K is some fixed real number. Note that $f(x,y)$ is a joint probability density function since

$$\int_{\boldsymbol{R}}\int_{\boldsymbol{R}} f(x,y)\,dx\,dy = \int_{\boldsymbol{R}}\int_{\boldsymbol{R}} \frac{1}{8}\exp(-\exp(|y|)|x - g(y) + K|) + \frac{1}{8}\exp(-\exp(|y|)|x - g(y) - K|)\,dx\,dy$$

$$= \int_{\boldsymbol{R}}\int_{\boldsymbol{R}} \frac{1}{4}\exp(-\exp(|y|)|z|)\,dz\,dy$$

$$= \int_{\boldsymbol{R}} \frac{1}{2}\exp(-|y|)\,dy = 1.$$

Let X and Y be random variables such that the pair (X,Y) has a joint density function given by $f(x,y)$. Notice from the above calculation that a second marginal density of $f(x,y)$ is given by $f_Y(y)=\frac{1}{2}\exp(-|y|)$. Also, notice that $f(x+g(y),y)$, as a function of x with y fixed, is even. Recall that a version of $E[X|Y=y]$ is given by $\int_R xf(x,y)/f_Y(y)\,dx$. Fixing this version and using this expression for $f_Y(y)$ implies that

$$
\begin{aligned}
&E[X|Y=y]\\
&\quad=2\exp(|y|)\int_R \frac{x}{8}\big(\exp(-\exp(|y|)|x-g(y)+K|)\\
&\qquad+\exp(-\exp(|y|)|x-g(y)-K|)\big)\,dx\\
&\quad=2\exp(|y|)\int_R\big((z+g(y)-K)\\
&\qquad+(z+g(y)+K)\big)\tfrac{1}{8}\exp(-\exp(|y|)|z|)\,dz\\
&\quad=2\exp(|y|)(2(g(y))+K-K)\frac{1}{4\exp(|y|)}=g(y).
\end{aligned}
$$

Since $f(x,y)$, as a function of x for y fixed, is even about $g(y)$, it follows that the conditional density function $f(x,y)/f_Y(y)$ shares this same property. Thus, it is easy to see that the associated regular conditional distribution function, when properly shifted, is symmetric (and not unimodal if $K\neq 0$).

Corollary 3 may thus be applied to see that $h(\cdot)=g(\cdot)$, which we recall was an arbitrary Borel measurable function, minimizes $E[\Phi(X-h(Y))]$ over all Borel measurable functions h: $R\to R$ where Φ: $R\to[0,\infty)$ is even and convex. Notice that this example illustrates the applicability of Theorem 2 in a situation where Sherman's result would not apply, and, in addition, it demonstrates the applicability of these results to non-Gaussian distributions. Notice further that this example also points out that the existence of a joint density function in no way guarantees that a regression function will obey any regularity property other than Borel measurability.

IV. The General Case

The preceding development was concerned with an attempt to estimate the integrable random variable X based on a Borel measurable function of the random variables $Y_1,\cdots,Y_k$, where k is a positive integer. In this case, our estimate was a $\sigma(Y_1,\cdots,Y_k)$-measurable random variable. Notice that it straightforwardly follows that $\sigma(Y_1,\cdots,Y_k)$ is countably generated since $\mathcal{B}(R^k)$ is countably generated. In many cases, we might wish to estimate X by a random variable that is measurable with respect to some other σ-algebra. For example, consider the case where $\{Y_t\colon t\in R^k\}$ is a random field and we wish to estimate X via a random variable that is $\sigma(\{Y_t\colon t\in R^k\})$-measurable. Note that this σ-algebra need not be countably generated. Also, consider the case where H is a real, separable Hilbert space and Z is an H-valued random variable; here we might wish to estimate X via a $\sigma(Z)$-measurable random variable. In the general case, Z could be a random object; that is, a random variable taking values in a measurable space $(G,\mathcal{G})$, and we would be interested in estimating X via a random variable which is measurable with respect to $\sigma(Z)=Z^{-1}(\mathcal{G})$.

The following theorem addresses the estimation of X via a random variable which is measurable with respect to any σ-subalgebra of $\mathcal{F}$.

Theorem 4: Let $(\Omega,\mathcal{F},P)$ be a probability space, $\mathcal{A}$ be a σ-subalgebra of $\mathcal{F}$, and X be a random variable defined on $(\Omega,\mathcal{F},P)$ such that X is integrable. For each $\omega\in\Omega$, let $M(\omega)=E[X|\mathcal{A}](\omega)$, and assume that there exists a regular conditional distribution function of X conditioned on $\mathcal{A}$, F: $R\times\Omega\to[0,1]$, such that $F(x+M(\omega),\omega)$, as a function of x with ω fixed, is symmetric. Then M minimizes the quantity $E[\Phi(X-\hat{X})]$ over all $\mathcal{A}$-measurable random variables $\hat{X}$, where Φ: $R\to[0,\infty)$ is even and convex.

Proof: Lemma 1 and a change of variables imply that for each fixed ω and for $\alpha\in R$,

$$\int_R \Phi(x-M(\omega))\,dF(x,\omega)\le\int_R\Phi(x-\alpha-M(\omega))\,dF(x,\omega).$$

Let $\hat{X}$ be an $\mathcal{A}$-measurable random variable by which X is to be estimated and $E[\Phi(X-\hat{X})]$ minimized. Note that

$$E\big[\Phi(X-\hat{X})\big]=E\big[E\big[\Phi(X-\hat{X})|\mathcal{A}\big]\big].$$

From the preceding inequality and [8, p. 79], the inner expectation, and thus the above expression, is minimized when $\hat{X}=M$. □

V. Conclusion

In this correspondence we have considered the problem of optimal estimation of a random variable X based on an observation denoted by a random vector Y. We have given a mild restriction on the regular conditional distribution function of X given $\sigma(Y)$ that ensures that $E[\Phi(X-g(Y))]$ is minimized for any cost function Φ that is nonnegative, even, and convex. Further, we have shown that given *any* real valued Borel measurable function there exist random variables X and Y, possessing a joint density function, so that the chosen function is an optimal estimator, with respect to any of the cost functions previously described, of the random variable X based on the random variable Y. The results were then extended to estimation of X based upon a random variable that is measurable with respect to any given σ-subalgebra.

References

[1] S. Sherman, "Non-mean-square error criteria," *IRE Trans. Inform. Theory*, vol. IT-4, pp. 125–126, Sept. 1958.

[2] E. B. Hall and G. L. Wise, "Simultaneous optimal estimation over a family of fidelity criteria," in *Proc. Twenty-First Ann. Conf. Inform. Sci. Syst.*, Baltimore, MD, Mar. 25–27, 1987, pp. 65–70.

[3] A. J. Viterbi, *Principles of Coherent Communication*. New York: McGraw-Hill, 1966.

[4] D. J. Sakrison, *Notes on Analog Communication*. New York: Van Nostrand Reinhold Company, 1970.

[5] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*. New York: Wiley, 1968.

[6] E. B. Hall and G. L. Wise, "On an aspect of optimal nonlinear estimation," in *Proc. Twenty-Seventh Ann. Allerton Conf. Commun., Contr. Computing*, Monticello, IL, Sept. 27–29, 1989, pp. 692–701.

[7] R. Ash, *Real Analysis and Probability*. New York: Academic Press, 1972.

[8] L. Breiman, *Probability*. Reading, MA: Addison-Wesley, 1968.

[9] R. G. Laha and V. K. Rohatgi, *Probability Theory*. New York: Wiley, 1979.

C. Decentralized Estimation

With a reasonable effort one may combine the results in Appendices A and B to result in methods of fusing best estimates where the field observers estimate under different fidelity criteria. For example, we presented results of this nature in Appendix C which appeared in the paper entitled, "Decentralized estimation with nontraditional fidelity criteria and corrupted estimates," which appeared in the *Proceedings of the Twenty-Sixth Annual Conference on Information Sciences and Systems*.

In the context of decentralized estimation, there is a need to efficiently and effectively process the estimates provided by multiple sensors. It is this problem of how best to fuse the separate estimates that is the essence of decentralized estimation. This paper was concerned with decentralized estimates when the estimates provided by the various sensors were corrupted by noise during transmission to the central processor and when different fidelity criteria were used by the different sensors. Decentralized techniques arise naturally in a number of diverse applications such as radar tracking, fault tolerance, two-way communications, highly redundant sensor systems, image processing, impact point prediction, moving source location, map updating in oceanography or meteorology, multiple sensor navigation systems, surveillance and search systems, underwater acoustic telemetry, power systems, object recognition, and communications between subsystems along unreliable or limited channels. Decentralized procedures promise many advantages over their centralized counterparts. For example, they may offer increased system reliability and fault tolerance, increased immunity and resistance to noise and jamming, increased accuracy, increased data compression and rate reduction, increased isolation and recovery capability, a parallel structure useful when processing a large volume of information, increased processing speed, increased computational efficiency, increased coverage, and an increase in the overall robustness of the system. In this paper it is shown, in contrast to previous results, that in a general additive Gaussian noise setting a decentralized procedure may produce the same estimate as a centralized procedure without the need for any intersensor communication.

DECENTRALIZED ESTIMATION WITH NONTRADITIONAL FIDELITY CRITERIA AND CORRUPTED ESTIMATES

Eric B. Hall
Department of Electrical Engineering
Southern Methodist University
Dallas, Texas 75275

Gary L. Wise
Department of Electrical and Computer Engineering
and
Department of Mathematics
The University of Texas at Austin
Austin, Texas 78712

Abstract

Decentralized estimation techniques are proposed based upon models that allow different nontraditional cost functions to be employed by each sensor and allow for noise to exist between the sensors and the central processor.

Introduction

In the context of decentralized estimation, there is a need to efficiently and effectively process the estimates provided by multiple sensors. Indeed, it is this problem of how best to combine or fuse the separate estimates that is the essence of decentralized estimation. This paper is concerned with decentralized estimation when the estimates provided by the various sensors are corrupted by noise during transmission to the central processor and when different fidelity criteria are used by the different sensors.

Decentralized techniques arise naturally in a number of diverse applications. For example, decentralized techniques have been proposed in the areas of radar tracking, fault tolerance, two-way communications, highly redundant sensor systems, image processing, impact point prediction, moving source location, map updating in oceanography or meteorology, multiple sensor navigation systems, surveillance and search systems, underwater acoustic telemetry, power systems, object recognition, and communications between subsystems along unreliable or limited channels.

Decentralized procedures promise many advantages over their centralized counterparts. For example, decentralized procedures may offer increased system reliability and fault tolerance, increased immunity and resistance to noise and jamming, increased accuracy, increased data compression and rate reduction, increased isolation and recovery capability, a parallel structure useful when processing a large volume of information, increased processing speed, increased computational efficiency, increased coverage, and an increase in the overall robustness of the system.

Although much has been written on the subject of decentralized estimation and data fusion, many important questions remain unanswered. Further, when answers have appeared they have often been incorrect or misleading. For example, is it true (as intuition might suggest) that a decentralized procedure can produce the same optimal estimate produced by a centralized procedure if and only if the sensors are allowed to communicate with each other? In [1] it is flatly stated that for a decentralized estimation structure to be effective the local sensors must communicate with each other, and in [2] a decentralized procedure has been proposed based upon additive noise and intersensor communication that provided the same estimate as a cen-

tralized procedure. In [3], however, it is shown, in contrast to the previous results, that in a general additive Gaussian noise setting a decentralized procedure may produce the same estimate as a centralized procedure without the need for any intersensor communication.

Data Fusion

In [3] the problem of decentralized estimation was considered in a general setting. In particular, it considered the problem of estimating a fixed second order random variable X defined on a probability space $(\Omega, \mathcal{F}, P)$ via a combination of estimates of the form $E[X \mid Y_i]$ where $1 \le i \le n$ and $Y_1, \ldots, Y_n$ are random variables also defined on $(\Omega, \mathcal{F}, P)$. That is, [3] considered the case in which the central processor was provided by each sensor with a best mean square estimate of X as a Borel measurable transformation of the observation Y_i. Focusing attention on Borel measurable transformations of the data it follows that the central processor must find a way of approximating $E[X \mid Y_1, \ldots, Y_n]$ based on random variables from the set $\{E[X \mid Y_1], \ldots, E[X \mid Y_n]\}$. (That is, it must approximate the orthogonal projection of X onto $L_2(\Omega, \sigma(Y_1, \ldots, Y_n), P)$ using the orthogonal projections of X onto $L_2(\Omega, \sigma(Y_1), P), \ldots, L_2(\Omega, \sigma(Y_n), P)$.) Thus, from a theoretical perspective, one seeks conditions under which $E[X \mid E[X \mid Y_1], \ldots, E[X \mid Y_n]]$ may provide a good approximation to $E[X \mid Y_1, \ldots, Y_n]$. Unfortunately, positive results to this question are elusive in many common settings. For example, (as shown in [3]) even if X is a Gaussian random variable, the observations $\{Y_1, \ldots, Y_n\}$ are mutually Gaussian random variables, and the problem expands to include estimates from the sensors of the form $E[X \mid \mathcal{D}]$ where $\mathcal{D}$ is a σ-algebra generated by any nonempty proper subset of the observations (that is, the problem expands to allow estimates from the sensors of the form $E[X \mid Y_{j_1}, \ldots, Y_{j_k}]$ for $k < n$), it still might not be possible to provide a reasonable estimate for X based on the data from the sensors. Thus, even invoking the ubiquitous Gaussian assumption and allowing any proper subset of the sensors to communicate may not be enough to establish reasonable rules for data fusion in such a general setting. In the next section we extend this example to include many non-Gaussian distributions.

A Counterexample

As in [4], let $n > 2$ be an integer and consider random variables $\{X_1, \ldots, X_n\}$ that have a joint probability density function $f : \mathbf{R}^n \to \mathbf{R}$ of the form

$$f(x_1, \ldots, x_n) = \left[\prod_{i=1}^{n} \hat{f}(x_i)\right]\left[1 + \prod_{i=1}^{n} x_i \hat{f}(x_i)\right] ; x_i \in \mathbf{R}$$

where $\hat{f}:\mathbf{R} \to \mathbf{R}$ is a standard Gaussian density function. It follows that the random variables in $\{X_1, \ldots, X_n\}$ each possess a standard Gaussian distribution. Further, it follows that, whereas the random variables in $\{X_1, \ldots, X_n\}$ are not mutually independent, the random variables in any proper subset of $\{X_1, \ldots, X_n\}$ are mutually independent.

Next, as in [5], consider a density function $\hat{g}$ for which the function $g : \mathbf{R}^n \to \mathbf{R}$ defined by

$$g(y_1, \ldots, y_n) = \left[\prod_{i=1}^{n} \hat{g}(y_i)\right]\left[1 + \prod_{i=1}^{n} y_i \hat{g}(y_i)\right] ; y_i \in \mathbf{R}$$

is nonnegative and integrates to 1. Let $n > 2$ be an integer. If $\{Y_1, \ldots, Y_n\}$ are random variables with a joint density function given by g then, paralleling the work in [4], it follows that each random variable in $\{Y_1, \ldots, Y_n\}$ possesses a density given by $\hat{g}$, and further that, although the random variables in $\{Y_1, \ldots, Y_n\}$ are not mutually independent, the random variables in any proper subset of $\{Y_1, \ldots, Y_n\}$ are mutually independent. Note that $g(y_1, \ldots, y_n)$ is nonnegative and integrates to 1 if $|x\hat{g}(x)| \le 1$ for all $x \in \mathbf{R}$ and if $\int_{\mathbf{R}} x\hat{g}^2(x)\,dx = 0$.

Now, let $\hat{f}$ be any probability density function such that $|x\hat{f}(x)| \le 1$ for all $x \in \mathbf{R}$ and such that $\int_{\mathbf{R}} x\hat{f}^2(x)\,dx = 0$, let $n > 1$ be an integer, and let $\{X, Y_1, \ldots, Y_n\}$ be random variables possessing a joint density function $f : \mathbf{R}^{n+1} \to \mathbf{R}$ of the form

$$f(x, y_1, \ldots, y_n) =$$

$$\left[\hat{f}(x)\prod_{i=1}^{n} \hat{f}(y_i)\right]\left[1 + x\hat{f}(x)\prod_{i=1}^{n} y_i \hat{f}(y_i)\right] ; x, y_i \in \mathbf{R}.$$

Since the conditional expectation of X given any proper subset of $\{Y_1, \ldots, Y_n\}$ is almost surely zero, it follows that any attempt to estimate X via best Borel measurable transformations of random variables from proper subsets of the data is hopeless.

Non-Mean-Square Cost Functions

Now we direct our attention toward non-mean-square cost functions. Given two random variables X and Y defined on a common probability space, a frequently encountered problem in estimation theory involves finding a function $g:\mathbf{R}\rightarrow\mathbf{R}$ that minimizes $E[\Phi(X - g(Y))]$ for some cost function $\Phi:\mathbf{R} \rightarrow [0, \infty)$. Generally, one is confronted with two concerns in choosing an appropriate cost function. First, one is concerned that the cost function should adequately reflect the cost one wishes to associate with an error, and second, one is concerned that the cost function should result in a problem that is mathematically tractable. Traditional cost functions, such as the popular mean-square error cost function, are usually chosen solely on the basis of the second of the above two concerns. As a result, the fidelity demands of the specific problem under consideration are rarely relied upon, and, in fact, are often not even considered when determining the cost function that will be used.

This sacrifice of suitability for mathematical ease in the choice of a cost function should be the cause of some concern since the traditional choices are unsuitable for many problems in estimation. As an example, consider the problem of estimating the position of a projectile. If one is interested in shooting down the projectile then a small error in the estimate of its position may not result in a penalty. If, however, the error in the estimate is such that the projectile is missed then the penalty might suddenly become enormous. Further, this penalty should not increase if the error in the estimate becomes even larger since the result of any two such errors is the same. The mean-square error cost function applied to this situation penalizes small inconsequential errors, assigns almost identical costs to barely hitting and barely missing the target, and assigns a larger cost to a far-miss than to a near-miss even though the result in each case is the same. Clearly, the mean-square cost function is not a very good choice in this commonly encountered situation. Such an example demonstrates the great need to extend the domain of mathematical tractability to include many cost functions that, though pertinent to the subjective demands of many problems, have often been excluded from consideration.

In [6], general conditions were given allowing for simultaneous use of any cost function that is nonnegative, even, and nondecreasing on $[0, \infty)$ or any cost function that is nonnegative, even, and convex. Further, these results were applied to non-Gaussian situations and extended to cover estimation based on random variables measurable with respect to a σ-algebra generated by a random object. In the next section we apply these results to the area of decentralized estimation.

Non-Mean-Square Fusion

Recall that a probability distribution function F is symmetric if for all real x, $F(x) = 1 - \lim_{h\downarrow 0} F(-x - h)$. The following result follows directly from Theorem 1 in [3] and Theorem 2 in [6].

Theorem 1: Consider a probability space $(\Omega, \mathcal{F}, P)$ and random variables $X, N_1, \ldots, N_n$ defined on $(\Omega, \mathcal{F}, P)$ where n is a positive integer and X is a second order random variable. Further, assume that for each positive integer $i \leq n$, N_i has a zero-mean Gaussian distribution with positive variance given by σ_i^2, and that $X, N_1, \ldots, N_n$ are mutually independent. Let $Y_i = X + N_i$ for $i = 1, \ldots, n$ and assume that a regular conditional distribution function $F:\mathbf{R} \times \Omega \rightarrow [0, 1]$ of X conditioned on $\sigma(Y_i)$ exists such that $F(x + E[X \mid Y_i](\omega), \omega)$, as a function of x with ω fixed, is symmetric. For each positive integer $i \leq n$, let $\Phi_i:\mathbf{R} \rightarrow [0, \infty)$ be even and convex. For each positive integer $i \leq n$, there exists a Borel measurable function $h_i:\mathbf{R} \rightarrow \mathbf{R}$ that minimizes $E[\Phi_i(X - h(Y_i))]$ over all Borel measurable functions h. Further, there exists a Borel mea-

surable function $g:\mathbf{R}^n \to \mathbf{R}$ such that $E[X \mid Y_1, \ldots, Y_n] = g(h_1(Y_1), \ldots, h_n(Y_n))$ a.s.

Fusion with Inter-Sensor Noise

The following example was shown in [3].

Example: Assume that all of the random variables in this example are defined on the same probability space denoted by $(\Omega, \mathcal{F}, P)$. For each positive integer $i \leq n$, let N_i be a zero-mean Gaussian distribution with positive variance given by σ_i^2, let X be a random variable such that $P(X = 1) = P(X = -1) = \frac{1}{2}$, assume that $X, N_1, \ldots, N_n$ are mutually independent, and let $Y_i = X + N_i$. It follows from [3] that a version of $E[X \mid Y_i = y]$ is given by $\tanh\left(\frac{1}{\sigma_i^2} y\right)$ and that

$$E[X \mid Y_1, \ldots, Y_n] = \tanh\left(\sum_{i=1}^{n} \tanh^{-1}(E[X \mid Y_i])\right) \text{ a.s.}$$

To begin, in the context of the previous example, note that $E[X \mid Y_i]$ possesses a density function $c_i:\mathbf{R} \to \mathbf{R}$ of the form

$$c_i(t) = \frac{\sigma_i^2}{1 - t^2}\, d_i(\sigma_i^2 \tanh^{-1}(t));\ |t| < 1$$

where

$$d_i(y) = \frac{1}{2\sigma_i\sqrt{2\pi}}\left[\exp\left(-\frac{(y+1)^2}{2\sigma_i^2}\right) + \exp\left(-\frac{(y-1)^2}{2\sigma_i^2}\right)\right].$$

Also, note that $|\tanh(x_1) - \tanh(x_2)| \leq |x_1 - x_2|$ for all $x_1, x_2 \in \mathbf{R}$ and, for $0 < \alpha < 1$, that $|\tanh^{-1}(x_1) - \tanh^{-1}(x_2)| \leq \frac{1}{1-\alpha^2}|x_1 - x_2|$ for all $x_1, x_2 \in [-\alpha, \alpha]$.

Consider now random variables $Z_1, \ldots, Z_n$ defined on the same probability space as above such that, for each positive integer $i \leq n$, Z_i is Gaussian with zero mean and positive variance s_i^2 and such that $X, N_1, \ldots, N_n, Z_1, \ldots, Z_n$ are mutually independent. Let α be an element from $(0, 1)$. Further, for each positive integer $i \leq n$, let β_i be a positive number, let $W_i = \beta_i E[X \mid Y_i] + Z_i$, let $A_i = \{\omega \in \Omega : \frac{1}{\beta_i}|W_i| \leq \alpha\}$, and let $V_i = \frac{1}{\beta_i} W_i I_{A_i}$. Let $C_i = \{\omega \in \Omega : |E[X \mid Y_i]| \leq \alpha\}$ for each positive integer $i \leq n$ and note that $P(A_i^c)$ and $P(C_i^c)$ may be made arbitrarily small by choice of α and β_i. Next, note that for each fixed positive integer $i \leq n$, it follows that for $\omega \in C_i \cap A_i$ we have

$$\left|\tanh\left(\sum_{i=1}^{n} \tanh^{-1}(E[X \mid Y_i])\right) - \tanh\left(\sum_{i=1}^{n} \tanh^{-1}(V_i)\right)\right|$$

$$\leq \left|\sum_{i=1}^{n} \tanh^{-1}(E[X \mid Y_i]) - \sum_{i=1}^{n} \tanh^{-1}(V_i)\right|$$

$$\leq \sum_{i=1}^{n} |\tanh^{-1}(E[X \mid Y_i]) - \tanh^{-1}(V_i)|$$

$$\leq \frac{1}{1-\alpha^2} \sum_{i=1}^{n} [|E[X \mid Y_i] - V_i|]$$

$$= \frac{1}{1-\alpha^2}\left[\sum_{i=1}^{n}\left(\frac{|Z_i|}{\beta_i}\right)\right].$$

Next, note that, for any positive integer $i \leq n$ and any positive number $\epsilon < 1$,

$$P\left(\frac{1}{\beta_i}|Z_i| \geq \epsilon\right) \leq \frac{s_i^2}{\epsilon^2 \beta_i^2}.$$

Thus, given any positive $p < 1$ and any positive $\epsilon < 1$, the upper bound above may be made less than ϵ with probability p by choice of α and β_i for $i = 1, \ldots, n$.

Finally, for $r \geq 1$, note from above that

$$E\left[\left|\tanh\left(\sum_{i=1}^{n} \tanh^{-1}(E[X \mid Y_i])\right) - \tanh\left(\sum_{i=1}^{n} \tanh^{-1}(V_i)\right)\right|^r I_{C_i \cap A_i}\right]$$

$$\leq \left(\frac{1}{1-\alpha^2}\right)^r E\left[\left(\sum_{i=1}^{n}\left(\frac{|Z_i|}{\beta_i}\right)\right)^r\right]$$

$$\leq \left(\frac{1}{1-\alpha^2}\right)^r 2^{r-1}\left(\sum_{i=1}^{n} E\left[\left(\frac{|Z_i|}{\beta_i}\right)^r\right]\right).$$

Also, note that

$$E\left[\left|\tanh\left(\sum_{i=1}^{n} \tanh^{-1}(E[X \mid Y_i])\right) - \tanh\left(\sum_{i=1}^{n} \tanh^{-1}(V_i)\right)\right|^r I_{C_i^c \cup A_i^c}\right]$$

$$\leq 2^r P(C_i^c \cup A_i^c).$$

Thus, it follows that the rth mean of the error between our estimate and a best Borel measurable mean square estimate of X without inter-sensory noise may be made arbitrarily small by choice of α and β_i for $i = 1, \ldots, n$.

Acknowledgement

This research was partially supported by the Office of Naval Research under Grant N00014-90-J-1712.

References

[1] Levy, B. C., D. A. Castañon, G. C. Verghese, and A. S. Willsky, "A Scattering Framework for Decentralized Estimation Problems," *Automatica*, Vol. 19, No. 4, 1983, pp. 373–384.

[2] Speyer, J. L., "Computation and Transmission Requirements for a Decentralized Linear-Quadratic-Gaussian Control Problem," *IEEE Transactions on Automatic Control*, Vol. 24, No. 2, April 1979, pp. 266–269.

[3] Hall, E. B., A. E. Wessel, and G. L. Wise, "Some Aspects of Fusion in Estimation Theory," *IEEE Transactions on Information Theory*, Vol. 37, No. 2, March 1991, pp. 420–422.

[4] Pierce, D. A., and R. L. Dykstra, "Independence and the Normal Distribution," *The American Statistician*, Vol. 23, No. 4, p. 39.

[5] Han, C.-P., "Dependence of Random Variables," *The American Statistician*, Vol. 25, No. 4, p. 35.

[6] Hall, E. B. and G. L. Wise, "On Optimal Estimation with Respect to a Large Family of Cost Functions," *IEEE Transactions on Information Theory*, Vol. 37, No. 3, May 1991, pp. 691–693.

D. Multidimensional Quantization

Results on multidimensional quantization are presented in the paper entitled, "A result on multidimensional quantization," which will appear in *Proceedings of the American Mathematical Society* and is given in Appendix D. Multidimensional quantization often arises in an effort to use digital processing techniques on data, since a quantizer is literally at the heart of analog to digital conversion. In this appendix we show that a popularly used technique for designing multidimensional quantizers fails spectacularly.

PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 000, Number 0, 0000

PROOF
RETURN THIS SET
OF PROOF TO
PROVIDENCE

A RESULT ON MULTIDIMENSIONAL QUANTIZATION

ERIC B. HALL AND GARY L. WISE

(Communicated by George C. Papanicolaou)

ABSTRACT. For any integer $N > 1$, a probability space, a Gaussian random vector X defined on the space with a positive definite covariance matrix, and an N-level quantizer Q are presented such that the random vector $Q(X)$ takes on each of the N values in its range with equal probability and such that X and $Q(X)$ are independent.

INTRODUCTION

Quantization, the process by which a set is mapped into a finite subset of a given cardinality, plays a pivotal role in virtually any application that requires analog to digital conversion; indeed, it is at the heart of much of modern digital technology. In such applications, a quantizer is often taken to be a function mapping $\mathbf{R}^k$ into a subset of $\mathbf{R}^k$ of cardinality N, where k is a positive integer and N is an integer greater than one (see, e.g., [1, 5, 3, 6, 2, 8, 10]). In this paper we present what might be a surprising consequence of such a general approach to quantization.

9

7

DEVELOPMENT

For a topological space T, we will let $\mathscr{B}(T)$ denote the family of Borel subsets of T. For a set S, we will let $\mathbf{P}(S)$ denote the power set of S and I_S denote the indicator function of S. By a standard Gaussian measure we will mean a Gaussian measure whose first moment is zero and whose second moment is one. Let k be a positive integer. For any measure m on $(\mathbf{R}^k, \mathscr{B}(\mathbf{R}^k))$ we will let m_* denote the inner measure on $(\mathbf{R}^k, \mathbf{P}(\mathbf{R}^k))$ induced by m and we will let m^* denote the outer measure on $(\mathbf{R}^k, \mathbf{P}(\mathbf{R}^k))$ induced by m. Recall from [4, p. 61] that if $B \in \mathscr{B}(\mathbf{R}^k)$ and $A \in \mathbf{P}(\mathbf{R}^k)$, then $m_*(B \cap A) + m^*(B \cap A^c) = m(B)$. We will let λ denote Lebesgue measure on $(\mathbf{R}, \mathscr{B}(\mathbf{R}))$ and, for integers $k > 1$, we will let Λ denote Lebesgue measure on $(\mathbf{R}^k, \mathscr{B}(\mathbf{R}^k))$, where k will be determined from the context. Recall that for a measure space $(\mathbf{R}^k, \mathscr{B}(\mathbf{R}^k), m)$, a subset S of $\mathbf{R}^k$ is said to be a saturated non-m-measurable

insert left paranthesis as shown

Aq.

Received by the editors April 20, 1991 and, in revised form, September 17, 1991.
1991 *Mathematics Subject Classification.* Primary 60A10; Secondary 62H12.
Key words and phrases. Multidimensional quantization, saturated nonmeasurable sets.
The first author was supported by the University Research Council of Southern Methodist University. The second author was supported by the Office of Naval Research under Grant N00014-90-J-1712.

Author Please verify. (they are correct)

Author: A preliminary version of this paper was presented at the 863rd meeting of the American Mathematical Society.

If the contents of this paper have been presented to the American Mathematical Society, or to a meeting of another scientific group, it is suggested you add this information to the first footnote. Include the date of presentation, the meeting, and the name of the sponsoring group.

Only one (1) set of galley proof will be sent to double authors. Please contact your co-author and make all necessary corrections.

set if $m_*(S) = m_*(S^c) = 0$. Finally, a k-dimensional quantizer of a random variable X defined on a probability space $(\Omega, \mathcal{F}, P)$ is any function $Q: \mathbf{R}^k \to F$ such that F is a finite subset of $\mathbf{R}^k$, such that $Q(x) = x$ for all x in F (i.e., such that Q restricted to F is the identity map on F) and such that $Q(X)$ is itself a random variable defined on $(\Omega, \mathcal{F}, P)$. If F is a finite subset of $\mathbf{R}^k$ with cardinality N then a quantizer $Q: \mathbf{R}^k \to F$ of a random variable X is said to be an N-level quantizer.

The following lemma is proved in [9, pp. 381–382].

Lemma 1. *For any positive integer M there exist M disjoint subsets $Z_1, Z_2, \ldots, Z_M$ of the real line such that $Z_1, Z_2, \ldots, Z_M$ and such that $Z = Z_1 \cup \cdots \cup Z_M$ are saturated non-λ-measurable sets.*

The next result is an immediate consequence of Lemma 1.

Corollary 1. *For any integer $N > 1$ there exist N subsets $T_1, T_2, \ldots, T_N$ of the real line that partition the real line and are such that for each positive integer $j \le N$, T_j is a saturated non-λ-measurable set.*

For our purposes the following corollary will prove useful.

Corollary 2. *For any positive integer k and any integer $N > 1$, there exist N subsets $S_1, S_2, \ldots, S_N$ of $\mathbf{R}^k$ that partition $\mathbf{R}^k$ and are such that, for each positive integer $j \le N$, S_j is a saturated non-Λ-measurable set.*

Proof. For $k = 1$, the result follows from Corollary 1. Assume $k > 1$. Let $T_1, \ldots, T_N$ be a partition of the real line as given by Corollary 1. For positive integers $j \le N$, let $S_j = T_j \times \mathbf{R} \times \cdots \times \mathbf{R} \subset \mathbf{R}^k$. Fix a positive integer $j \le N$ and assume that there exists a set $B \in \mathcal{B}(\mathbf{R}^k)$ such that $B \subset S_j$ and $\Lambda(B) > 0$. Define a subset $\hat{B}$ of $\mathbf{R}$ as follows:

$$\hat{B} = \{b_1 \in \mathbf{R} : (b_1, b_2, \ldots, b_k) \in B \text{ for some } (b_2, \ldots, b_k) \in \mathbf{R}^{k-1}\}.$$

Recall from [7, p. 161] that $\hat{B} \in \mathcal{B}(\mathbf{R})$. Further, notice that $\lambda(\hat{B}) > 0$ since $B \subset \hat{B} \times \mathbf{R} \times \cdots \times \mathbf{R} \subset \mathbf{R}^k$ and $\Lambda(B) > 0$. But, $\lambda(\hat{B}) = 0$ since $\hat{B} \subset T_j$ and $\lambda_*(T_j) = 0$. This contradiction implies that $\Lambda(B) = 0$ and hence that $\Lambda_*(S_j) = 0$. It follows similarly that $\Lambda_*(S_j^c) = 0$ also. Q.E.D.

Lemma 2. *For a positive integer k and an integer $N > 1$, let $S_1, S_2, \ldots, S_N$ comprise a partition of $\mathbf{R}^k$ such that for each positive integer $J \le N$, S_j is a saturated non-Λ-measurable set. The set*

$$\mathcal{F} = \{(S_1 \cap A_1) \cup \cdots \cup (S_N \cap A_N) : A_i \in \mathcal{B}(\mathbf{R}^k) \text{ for } 1 \le i \le N\}$$

is a σ-algebra on $\mathbf{R}^k$.

Proof. Choosing $A_1 = \cdots = A_N = \varnothing$ implies that $\varnothing \in \mathcal{F}$. Let A be an element of $\mathcal{F}$. Then $A = (S_1 \cap A_1) \cup \cdots \cup (S_N \cap A_N)$ for some choice of the A_i's from $\mathcal{B}(\mathbf{R}^k)$. Further, $A^c = (S_1 \cap A_1)^c \cap \cdots \cap (S_N \cap A_N)^c$. Since

$$S_i^c = \bigcup_{j=1\ (i \ne j)}^{N} S_j,$$

it follows that

$$A^c = \bigcap_{i=1}^{N} \bigcup_{j=1\ (i \ne j)}^{N} S_j \cup A_i^c.$$

insert extra space here

Hence A^c is a finite union of sets, each of which is of one of the following three forms:

(i) $S_{n_1} \cap \cdots \cap S_{n_k} \cap B$ where $1 \le n_1 < \cdots < n_k \le N$, $k > 1$, and $B \in \mathcal{B}(\mathbf{R}^k)$;

(ii) $S_j \cap B$ for $1 \le j \le N$ and $B \in \mathcal{B}(\mathbf{R}^k)$;

(iii) $B \in \mathcal{B}(\mathbf{R}^k)$.

Every set of the form given by (i) is empty since the S_i's are disjoint. Further, any set $B \in \mathcal{B}(\mathbf{R}^k)$ may be expressed as $B = (S_1 \cap B) \cup \cdots \cup (S_N \cap B)$. Hence, A^c is an element of $\mathcal{F}$.

Finally, if $B_1, B_2, \ldots$ are in $\mathcal{F}$, then for some choice of the $A_{i,j}$'s from $\mathcal{B}(\mathbf{R}^k)$,

$$\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} \bigcap_{j=1}^{N} (S_j \cap A_{i,j}) = \bigcap_{j=1}^{N} S_j \cap \left(\bigcup_{i=1}^{\infty} A_{i,j} \right) \in \mathcal{F}. \quad \square$$

Q.E.D.

Recall that two measures P_1 and P_2 on a given measurable space $(\Omega, \mathcal{S})$ are said to be equivalent if $\{A \in \mathcal{S} : P_1(A) = 0\} = \{A \in \mathcal{S} : P_2(A) = 0\}$. Notice that for sets $S_1, S_2, \ldots, S_N$ as above, it follows that, for any positive integer $i \le N$ and any $\mathcal{B}(\mathbf{R}^k)$-measurable set H, $P_*(S_i \cap H) = 0$, $P_*(S_i^c \cap H) = 0$, $P^*(S_i \cap H) = P(H)$, and $P^*(S_i^c \cap H) = P(H)$ for any probability measure P on $(\mathbf{R}^k, \mathcal{B}(\mathbf{R}^k))$ that is equivalent to Lebesgue measure on $(\mathbf{R}^k, \mathcal{B}(\mathbf{R}^k))$. The following lemma will be used in the proof of a subsequent theorem.

Lemma 3. *For a positive integer k and an integer $N > 1$, let $S_1, S_2, \ldots, S_N$ comprise a partition of $\mathbf{R}^k$ such that for each positive integer $j \le N$, S_j is a saturated non-Λ-measurable set. Let P be a probability measure on $(\mathbf{R}^k, \mathcal{B}(\mathbf{R}^k))$ that is equivalent to Lebesgue measure on $(\mathbf{R}^k, \mathcal{B}(\mathbf{R}^k))$. Let $A_1, \ldots, A_N$ and $B_1, \ldots, B_N$ be sets from $\mathcal{B}(\mathbf{R}^k)$ such that*

$$(S_1 \cap A_1) \cup \cdots \cup (S_N \cap A_N) = (S_1 \cap B_1) \cup \cdots \cup (S_N \cap B_N).$$

Then $P(A_i \Delta B_i) = 0$ for any positive integer $i \le N$ where for any two subsets A and B of $\mathbf{R}^k$, $A \Delta B$ denotes the symmetric difference of A and B.

Proof. Fix a positive integer $i \le N$. By assumption,

$$(S_1 \cap A_1) \cup \cdots \cup (S_N \cap A_N) = (S_1 \cap B_1) \cup \cdots \cup (S_N \cap B_N).$$

Intersecting each side with S_i implies that $(S_i \cap A_i) = (S_i \cap B_i)$, which implies that $(S_i \cap A_i) \cap (S_i \cap B_i)^c = (S_i \cap A_i) \cap (S_i^c \cup B_i^c) = (S_i \cap A_i \cap S_i^c) \cup (S_i \cap A_i \cap B_i^c) = (S_i \cap A_i \cap B_i^c) = \varnothing$ and, similarly, that $(S_i \cap B_i \cap A_i^c) = \varnothing$. Thus, we see that $(S_i \cap A_i \cap B_i^c) \cup (S_i \cap B_i \cap A_i^c) = S_i \cap (A_i \Delta B_i) = \varnothing$. Since $(A_i \Delta B_i) \in \mathcal{B}(\mathbf{R}^k)$, it follows that $P(A_i \Delta B_i) = P^*(S_i \cap (A_i \Delta B_i)) = P^*(\varnothing) = 0$. Q.E.D.

The following theorem provides a probability space upon which the principal result of this paper will be based.

Theorem 1. *For a positive integer k and an integer $N > 1$, let $S_1, S_2, \ldots, S_N$ comprise a partition of $\mathbf{R}^k$ such that for each positive integer $j \le N$, S_j is a saturated non-Λ-measurable set. Let P be a probability measure on $(\mathbf{R}^k, \mathcal{B}(\mathbf{R}^k))$ that is equivalent to Lebesgue measure on $(\mathbf{R}^k, \mathcal{B}(\mathbf{R}^k))$. There exists a probability space $(\mathbf{R}^k, \mathcal{S}, \mu)$ such that $\mathcal{S}$ includes $\mathcal{B}(\mathbf{R}^k)$, such that $\mathcal{S}$ contains*

$S_1, \ldots, S_N$, such that the measure μ agrees with P on $\mathscr{B}(\mathbf{R}^k)$, and such that $\mathscr{B}(\mathbf{R}^k)$ is independent of $\sigma(S_1, \ldots, S_N)$.

Proof. Let $\mathscr{G}$ be the σ-algebra $\mathscr{F}$ provided by Lemma 2. Recall that $\mathscr{F}$ contains all sets of the form $(S_1 \cap A_1) \cup \cdots \cup (S_n \cap A_n)$ where $A_i \in \mathscr{B}(\mathbf{R}^k)$ for each positive integer $i \leq N$. (if) $A \in \mathscr{B}(\mathbf{R}^k)$ then choosing $A_1 = \cdots = A_N = A$ implies that $A \in \mathscr{G}$. Similarly, for any positive integer $i \leq N$, setting $A_i = \mathbf{R}^k$ and all other A_j's equal to the empty set implies that $S_i \in \mathscr{G}$. Define a measure μ on the measurable space $(\mathbf{R}^k, \mathscr{G})$ via

$$\mu((S_1 \cap A_1) \cup \cdots \cup (S_N \cap A_N)) = \frac{1}{N}(P(A_1) + \cdots + P(A_N))$$

for $(S_1 \cap A_1) \cup \cdots \cup (S_N \cap A_N) \in \mathscr{G}$. That μ is well defined follows from Lemma 3 and that μ is in fact a probability measure that agrees with P on $\mathscr{B}(\mathbf{R}^k)$ is then straightforward. Further notice that $\mu(S_i) = \frac{1}{N}$ for each positive integer $i \leq N$ and that, for any set $B \in \mathscr{B}(\mathbf{R}^k)$ and any positive integer $i \leq N$, $\mu(S_i \cap B) = \frac{1}{N}P(B) = \mu(S_i)\mu(B)$. Thus S_i is independent of $\mathscr{B}(\mathbf{R}^k)$ for each positive integer $i \leq N$. Finally, notice that $\mathscr{B}(\mathbf{R}^k)$ is in fact independent of $\sigma(S_1, \ldots, S_N)$ since $\{\varnothing, S_1, \ldots, S_N\}$ is a π-system. Q.E.D.

We are now in a position to state and prove the principal result of this paper.

Theorem 2. *Let k be a positive integer and let N be an integer greater than one. There exists a probability space $(\Omega, \mathscr{S}, \nu)$, a Gaussian random vector X defined on $(\Omega, \mathscr{S}, \nu)$ taking values in $\mathbf{R}^k$ with a positive definite covariance matrix, and an N-level k-dimensional quantizer $Q: \mathbf{R}^k \to F$ such that $\nu(Q(X) = x) = 1/N$ for each x in F and such that X and $Q(X)$ are independent.*

Proof. Let $S_1, \ldots, S_N$ be sets as provided by Corollary 2. For these N subsets of $\mathbf{R}^k$, let $(\Omega, \mathscr{S}, \nu)$ be a probability space as provided by Theorem 1 where P is chosen to be the product measure induced by placing standard Gaussian measure on each factor of $(\mathbf{R}^k, \mathscr{B}(\mathbf{R}^k))$. For each positive integer $i \leq N$, let α_i be an element from S_i. Let F denote the set $\{\alpha_1, \ldots, \alpha_N\}$. Define an N-level k-dimensional quantizer $Q: \mathbf{R}^k \to F$ via $Q(x) = \sum_{i=1}^{N} \alpha_i I_{S_i}(x)$. Further, notice that the random vector $X(\omega) = \omega$; $\omega \in \Omega$, is a zero mean Gaussian random vector defined on $(\Omega, \mathscr{S}, \nu)$ whose covariance matrix is the $k \times k$ identity matrix. Also, notice that for $1 \leq i \leq N$, $\nu(Q(X(\omega)) = \alpha_i) = \nu(\omega \in S_i) = 1/N$. Finally, notice that X and $Q(X)$ are independent via Theorem 1. Q.E.D.

References

1. A. Gersho, *Asymptotically optimal block quantization*, IEEE Trans. Inform. Theory IT-25 (1979), 373–380.
2. R. M. Gray and E. D. Karnin, *Multiple local optimal in vector quantizers*, IEEE Trans. Inform. Theory IT-28 (1982), 256–261.
3. R. M. Gray, J. C. Kieffer, and Y. Linde, *Locally optimal block quantizer design*, Inform. and Control 45 (1980), 178–198.
4. P. R. Halmos, *Measure theory*, Springer-Verlag, New York, 1974.
5. (L.) Linde, A. Buzo, and R. M. Gray, *An algorithm for vector quantizer design*, IEEE Trans. Comm. COM-28 (1980), 84–95.

Author:

In order to make the address listing useful to the reader, we wish to include your address at the time of publication. If it is different from the one given above, please indicate it below. It will be added to the end of the paper as a current address

Authors are encouraged to supply electronic addresses when available. These will be printed after the postal address at the end of the article. Please supply them for this paper if available.

6. D. Pollard, *Quantization and the method of k means*, IEEE Trans. Inform. Theory IT-28 (1982), 199–205.

~~7. W. Rudin, *Real and complex analysis*, third edition, McGraw-Hill, New York, 1987.~~

8. M. J. Sabin and R. M. Gray, *Global convergence and empirical consistency of the generalized Lloyd algorithm*, IT-32 (1986), 148–155.

9. M. Shiffman, *Measure-theoretic properties of non-measurable sets*, Pacific J. Math. 138 (1989), 357–389.

10. P. F. Swaszek, *A vector quantizer for the Laplacian source*, IEEE Trans. Inform. Theory 37 (1991), 1355–1365.

Department of Electrical Engineering, Southern Methodist University, Dallas, Texas 75275

AND

Department of Electrical & Computer Engineering, and Department of Mathematics, University of Texas at Austin, Austin, Texas 78712

K-means

Delete reference 7 entirely.

IEEE Trans. Inform. Theo

and Author is this correct as set

No. see indicated changes.

The

email: ebh@smunews.smu.edu

email: gwise@ccwf.cc.utexas.edu

Summary of Corrections to PROC1451 (Hall and Wise)

1. In the sixth line of the introduction on page 1, references 8 and 10 should be changed to become references 7 and 9.

2. In the sixth line of the development on page 1, a left parenthesis should be inserted before $\mathbf{R}^k$.

3. The following footnote should be added to the bottom of page 1:
"A preliminary version of this paper was presented at the 863rd meeting of the American Mathematical Society."

4. Reference number 9 in the line above Lemma 1 on page 2 should be changed to become Reference number 8.

5. The lowercase m in Z_m in the statement of Lemma 1 on page 2 should be an uppercase M as you have already indicated.

6. The second phrase "such that" in the statement of Lemma 1 on page 2 should be deleted as you have already indicated.

* 7. The phrase "a set $B \in \mathcal{B}(\mathbf{R}^k)$" in the fourth line of the proof of Corollary 2 on page 2 should be deleted and replaced instead by the phrase "an $\mathcal{F}_\sigma$ subset B of $\mathbf{R}^k$". Note: $\mathcal{F}_\sigma$ is an uppercase script F followed by a subscripted lowercase Greek letter sigma.

* 8. The phrase "Recall from [7, p. 161]" in the 7th line of the proof of Corollary 2 on page 2 should be replaced by the phrase "Note".

9. The uppercase J in the second line of the statement of Lemma 2 on page 2 should be replaced by a lowercase j.

---

* Items 7 and 8 correct a small but critical oversight in the proof of Corollary 2 that slipped by both the authors and the reviewers and was not noticed until the proofreading stage. If you have any questions regarding these two changes please don't hesitate to contact either author. (Eric Hall at (214)-692-4367 or Gary Wise at (512)-471-3356.)

10. Extra space should be inserted between the $\bigcap_{i=1}^{N}$ and the $\bigcup_{j=1\,(i\neq j)}^{N}$ in the sixth line of the proof of Lemma 2 at the bottom of page 2.

11. The sentence beginning with the word "Every" near the top of page 3 should not be indented.

12. The hyphen following the word "non" in the third line of Lemma 3 should be moved over slightly as you have already indicated.

13. If word "if" in the third line of the proof of Theorem 1 on page 4 should be capitalized.

14. The word "optimal" in reference 2 on page 4 should be "optima".

15. The name "L. Linde" in reference 5 on page 4 should be "Y. Linde".

16. The phrase "k means" in reference 6 on page 5 should be "k-means".

17. Reference 7 on page 5 should be deleted entirely.

18. Reference 8 on page 5 should be renumbered to become reference 7 and the journal name "IEEE Trans. Inform. Theory" should be added as you have already indicated.

19. Reference 9 on page 5 should be renumbered to become reference 8.

20. Reference 10 on page 5 should be renumbered to become reference 9.

21. The phrase "Electrical Computer" should be "Electrical and Computer" in the second address on page 5 as you have already indicated.

22. The phrase "University of Texas" should be "The University of Texas" in the second address on page 5.

23. The following email addresses should be included in the addresses:
author 1: ebh@smunews.smu.edu
author 2: gwise@ccwf.cc.utexas.edu

E. Multidimensional Convolution

Results on multidimensional convolution are presented in the paper entitled "Some Aspects of Multidimensional Convolution" which appeared in *Proceedings of the 1991 IEEE International Conference on Acoustics, Speech, and Signal Processing* and is given in Appendix E. It is shown that multidimensional convolution need not be associative . Further, for any positive integer k , it is shown that the multidimensional convolution of two real valued, bounded integrable nowhere zero functions defined on $\mathbf{R}^k$ can be identically equal to zero. These results are discusses in an algebraic setting, and a consequence involving random fields is briefly discussed.

SOME ASPECTS OF MULTIDIMENSIONAL CONVOLUTION

Eric B. Hall
Department of Electrical Engineering
Southern Methodist University
Dallas, Texas 75275

Gary L. Wise
Department of Electrical and Computer Engineering
and
Department of Mathematics
The University of Texas at Austin
Austin, Texas 78712

Abstract

It is shown that multidimensional convolution need not be associative. Further, for any positive integer k, it is shown that the multidimensional convolution of two real valued, bounded, integrable, nowhere zero functions defined on $\mathbf{R}^k$ can be identically equal to zero. These results are discussed in an algebraic setting, and a consequence involving random fields is briefly considered.

Introduction

Real valued functions of several variables frequently occur in such areas of signal processing as image processing, optics, and oceanography. In these areas, as well as in many others, convolution plays a major role. This paper treats several aspects of multidimensional convolution which should be of interest to the signal processing community.

In applications, multidimensional convolution often arises when considering multidimensional linear systems. In this context, the linear system is characterized via convolution with an integrable function g and the input to the linear system is denoted by an integrable function f. The function h given by f convolved with g then denotes the resulting output. A problem which frequently arises in system identification is that of deconvolution which is concerned with approximating or identifying the function g from a knowledge of the pair of functions f and h.

In [1] it was shown that in a one-dimensional setting there exist integrable, bounded, nowhere zero functions f and g such that f convolved with g is identically equal to zero. That is, in the context of linear systems, there exists a linear system described via convolution with a fixed, bounded, nowhere zero function g which may be a nopass filter to an input which is nowhere zero. Clearly, such a phenomenon should be a cause of some concern to one who is attempting to derive a general method of deconvolution. In this paper, we extend this result to the case of multidimensional convolution.

Development

Let k be a positive integer. For a set $S \subset \mathbf{R}^k$, we will let I_S denote the indicator function of the set S. We will denote by $L_1(\mathbf{R}^k)$ the set of all extended real valued Lebesgue integrable functions modulo almost everywhere equivalence defined on $\mathbf{R}^k$ equipped with the norm given by the integral of the absolute value of an element of $L_1(\mathbf{R}^k)$. By a k-sequence of real numbers we will mean any function mapping $\mathbf{Z}^k$ into $\mathbf{R}$ where $\mathbf{Z}$ denotes the integers, and we will denote the value of such a function α at the point x via α_x. A k-sequence will be called absolutely summable if it is integrable with respect to counting measure on the power set of $\mathbf{Z}^k$. Further, we will occasionally denote points x in $\mathbf{R}^k$ as $x = (x_1, x_2, \ldots, x_k)$ where the x_i's are real numbers. Finally, for two points x and y in $\mathbf{R}^k$ we will denote the Euclidean inner product via

$$\langle x, y \rangle = \sum_{j=1}^{k} x_j y_j.$$

Recall that the convolution of two functions f and g in $L_1(\mathbf{R}^k)$, denoted by $f * g$, is defined via

$$(f * g)(x) = \int_{\mathbf{R}^k} f(x - y)\, g(y)\, dy$$

provided that this integral exists for all $x \in \mathbf{R}^k$. Further, we recall [2, pp. 247-248] that if f and g are in $L_1(\mathbf{R}^k)$ then $f * g$ is also in $L_1(\mathbf{R}^k)$ and satisfies $\|f * g\|_{L_1} \le \|f\|_{L_1} \|g\|_{L_1}$.

The following lemma shows that multidimensional convolution need not be associative.

Presented at the 1991 IEEE International Conference on Acoustics, Speech, and Signal Processing, May 14-17, 1991; to be published in the Proceedings of the Conference.

Lemma 1: Let k be a positive integer. There exist three bounded real valued Lebesgue measurable functions f, g, and h defined on $\mathbf{R}^k$ such that, even though the convolutions are each defined, $f * (g * h) \neq (f * g) * h$, i.e. such that convolution is not associative.

Proof: Consider first the special case when k = 1. As in [3, p. 177], for t and x real, define $p(t) = (1 - \cos(t))\, I_{[0, 2\pi]}(t)$ and let $f(x) = 1$, $g(x) = p'(x)$, and $h(x) = \int_{-\infty}^{x} p(t)\, dt$. Note that $(f * g)(x) = \int_{\mathbf{R}} f(x - t)\, g(t)\, dt$

$= \int_{\mathbf{R}} p'(t)\, dt = p(2\pi) - p(0) = 0$. Further, $(g * h)(x)$

$= \int_{\mathbf{R}} g(x - t)\, h(t)\, dt = \int_{\mathbf{R}} p'(x - t) \int_{-\infty}^{t} p(s)\, ds\, dt = (p * p)(x)$ via integration by parts. Note that

$$(p * p)(x) = \int_{0}^{2\pi} (1 - \cos(x - y))\, (1 - \cos(y))\, I_{[x - 2\pi,\, x]}(y)\, dy.$$

Hence, $(g * h)(x)$ is positive on $(0, 4\pi)$ and zero elsewhere. Finally, even though $(f * g) * h = 0$, we see that $f * (g * h)$ is a positive constant.

Now, let k be an integer greater than 1. With f, g, and h defined as in the preceding paragraph, let $\tilde{f}$, $\tilde{g}$, and $\tilde{h}$ map $\mathbf{R}^k$ into $\mathbf{R}$ via $\tilde{f}(x) = 1$, $\tilde{g}(x) = g(x_1)\, g(x_2) \ldots g(x_k)$, and $\tilde{h}(x) = h(x_1)\, h(x_2) \ldots h(x_k)$. It follows immediately that $(\tilde{f} * \tilde{g}) = 0$ and $(\tilde{g} * \tilde{h})$ is positive on $(0, 4\pi)^k$ and zero elsewhere. Hence, $(\tilde{f} * \tilde{g}) * \tilde{h} = 0$ but $\tilde{f} * (\tilde{g} * \tilde{h})$ is a positive constant.

Q.E.D.

Next, consider two bounded, real valued, Lebesgue integrable functions f and g defined on $\mathbf{R}^k$. Further, assume that f and g are *nowhere* zero. Does it follow that $f * g$ is nowhere zero? Does it follow that $f * g$ is nonzero on some nonempty set? From a linear systems viewpoint, does a nowhere zero input to a linear time-invariant system described via convolution with a fixed nowhere zero function result in an output which is nonzero somewhere?

To begin, we will need the following notation. For an absolutely summable k-sequence of real numbers α, define a bounded linear operator on $L_1(\mathbf{R}^k)$ to $L_1(\mathbf{R}^k)$ via

$$(T_\alpha(f))(x) = \int_{\mathbf{Z}^k} \alpha_y\, f(x - y)\, dC(y),$$

where C denotes counting measure on the power set of $\mathbf{Z}^k$, for any element f from $L_1(\mathbf{R}^k)$. For any two absolutely summable k-sequences of real numbers α and β, it follows that

$$((T_\alpha \circ T_\beta)(f))(x) = T_\alpha \left(\int_{\mathbf{Z}^k} \beta_y\, f(x - y)\, dC(y) \right)$$

$$= \int_{\mathbf{Z}^k} \int_{\mathbf{Z}^k} \alpha_z\, \beta_y\, f(x - y - z)\, dC(y)\, dC(z)$$

$$= \int_{\mathbf{Z}^k} \lambda_y\, f(x - y)\, dC(y)$$

where we define

$$\lambda_y = \int_{\mathbf{Z}^k} \int_{\mathbf{Z}^k} \alpha_p\, \beta_q\, g_y(p, q)\, dC(p)\, dC(q),$$

where $g_y(p, q)$ equals one if $p + q = y$ and equals zero otherwise. Finally, note that for any two elements f and g from $L_1(\mathbf{R}^k)$ it follows via Fubini's theorem that $(T_\alpha(f)) * (T_\beta(g)) = (T_\alpha \circ T_\beta)(f * g)$.

Theorem 1: Letting the above set notation, there exist two non-identically zero absolutely summable k-sequences of real numbers α and β such that for any f and g from $L_1(\mathbf{R}^k)$, $(T_\alpha(f)) * (T_\beta(g)) = 0$.

Proof: Recall that the function $|\cos(x_1) \cos(x_2) \cdots \cos(x_k)|$ is expressible as a multiple Fourier series given by

$$\int_{\mathbf{Z}^k} c_y \exp(i \langle x, y \rangle)\, dC(y)$$

where it follows easily that $c_y = a_{y_1}\, a_{y_2} \ldots a_{y_k}$ where $a_n = 0$ if n is odd and

$$a_n = \frac{2}{\pi} \left[\frac{(-1)^{n/2}}{1 - n^2} \right]$$

if n is even. Further, if we define

$f_1(x) = \frac{1}{2} (|\cos(x_1) \cos(x_2) \cdots \cos(x_k)| +$
$(\cos(x_1) \cos(x_2) \cdots \cos(x_k)))$ and
$f_2(x) = \frac{1}{2} (|\cos(x_1) \cos(x_2) \cdots \cos(x_k)| -$
$(\cos(x_1) \cos(x_2) \cdots \cos(x_k)))$, then $f_1(x)\, f_2(x) = 0$,

$$f_1(x) = \int_{\mathbf{Z}^k} \alpha_y \exp(i \langle x, y \rangle)\, dC(y),$$

and

$$f_2(x) = \int_{Z^k} \beta_y \exp(i \langle x, y \rangle)\, dC(y)$$

where

$$\alpha_y = \begin{cases} \frac{c_y}{2} & \text{if } y \notin \{-1, 1\}^k \\ \left(\frac{1}{4}\right)^k & \text{if } y \in \{-1, 1\}^k \end{cases}$$

and

$$\beta_y = \begin{cases} \frac{c_y}{2} & \text{if } y \notin \{-1, 1\}^k \\ -\left(\frac{1}{4}\right)^k & \text{if } y \in \{-1, 1\}^k. \end{cases}$$

But, $f_1(x)\, f_2(x) = 0$

$$= \int_{Z^k} \alpha_y \exp(i \langle x, y \rangle)\, dC(y) \int_{Z^k} \beta_y \exp(i \langle x, y \rangle)\, dC(y)$$

$$= \int_{Z^k} \int_{Z^k} \alpha_y \beta_z \exp(i (\langle x, y \rangle + \langle x, z \rangle))\, dC(y)\, dC(z)$$

$$= \int_{Z^k} \lambda_y \exp(i \langle x, y \rangle)\, dC(y) \text{ where, as before,}$$

$$\lambda_y = \int_{Z^k} \int_{Z^k} \alpha_p \beta_q g_y(p, q)\, dC(p)\, dC(q)$$

and $g_y(p, q)$ equals one if $p + q = y$ and equals zero otherwise. Note that, via Fubini's theorem, $\lambda_y = 0$ for every y. Thus, it follows that $(T_\alpha(f)) * (T_\beta(g)) = (T_\alpha \circ T_\beta)(f * g) = 0$ for any integrable f and g. **Q.E.D.**

Scholium 1: Let k be a positive integer. There exist two real valued, bounded, nowhere zero, Lebesgue integrable functions defined on R^k such that their convolution is identically equal to zero.

Proof: In the proof of Theorem 1, choose $f(x) = g(x) = I_S(x)$ where $S = (-1, 1]^k$. **Q.E.D.**

Before commenting further upon this result, we shall detour for a moment to review a few algebraic concepts. Recall that a nonempty set G and two operations + and · form an associative ring, hereafter referred to as a ring, if G is an Abelian group with respect to the + operation (denoting the identity element by 0 and $(a)^{-1}$ by –a), G is closed and associative with respect to ·, and finally if, for any a, b, and c in G, $a \cdot (b + c) = a \cdot b + a \cdot c$ and $(b + c) \cdot a = b \cdot a + c \cdot a$.

Further, if there exists an element u in G such that $a \cdot u = u \cdot a = a$ for every a in G, then G is said to possess a unit element. Also, if $a \cdot b = b \cdot a$ for every a and b in G then G is said to be a commutative ring. Recall that for a commutative ring G an element $a \neq 0$ in G is said to be a zero-divisor if there exists an element $b \neq 0$ in G such that $a \cdot b = 0$. Further, recall that a commutative ring is said to be an integral domain if it possesses no zero-divisors. For a more complete discussion of rings, the interested reader is referred to [4].

It follows easily that $L_1(R^k)$ equipped with the operations of pointwise addition and convolution is a commutative ring $\mathcal{R}$; in fact, $\mathcal{R}$ is a commutative Banach algebra. Even though it can be seen that this ring possesses no unit element [2, p. 248], it does possess so-called "approximate units" which often serve just as well for many purposes.

The previous results can now be viewed in a different setting. Recall that Lemma 1 showed that multidimensional convolution need not be associative. Although this result may seem surprising to some, notice that the function $\tilde{f}$ given in the proof of Lemma 1 is not an element of $\mathcal{R}$. Further, from an algebraic standpoint, the perhaps disturbing result of Scholium 1 yields the following corollary as a direct consequence.

Corollary 1: Let k be a positive integer. The commutative ring given by $L_1(R^k)$ equipped with the operations of pointwise addition and convolution is not an integral domain.

Hence, Corollary 1 implies that $(f * g) = 0$ can occur even when neither f nor g is equal to zero. In fact, we have actually shown something stronger via Scholium 1 since it exhibits bounded integrable functions f and g defined on R^k which are *nowhere* equal to zero and yet for which $(f * g)$ is identically equal to zero.

Finally, again let k be a positive integer. It follows from Theorem 1 and Scholium 1 that there exists a random field $\{X(p): p \in R^k\}$ with integrable sample paths and a function $f: R^k \rightarrow R$ which is Lebesgue integrable and nowhere zero such that $f * X$ is identically zero. Such a result should be of interest to those in areas such as seismology, radio astronomy, underwater acoustics, and channel equalization where blind deconvolution techniques are frequently employed.

Conclusion

In this paper we have considered multidimensional convolution from an algebraic standpoint and presented a result which may be of interest to the engineering community. In particular, we showed that, for any positive integer k, the convolution of two nowhere zero, bounded, integrable, real valued functions defined on R^k may be everywhere zero. This result should be of interest to those attempting to identify the input to a linear time-invariant system via some operations on the output, such as in deconvolution problems.

Acknowledgement

This research was partially supported by the Office of Naval Research under Grant N00014-90-J-1712.

References

1. Hall, E. B., and G. L. Wise, "An Algebraic Aspect of Linear System Theory," *IEEE Trans. on Circuits and Systems*, Vol. 37, No. 5, May, 1990, pp. 651-653.

2. Torchinsky, A., *Real Variables*. Redwood City, California: Addison–Wesley, 1988.

3. Rudin, W., *Real and Complex Analysis*. 3rd ed., New York: McGraw-Hill, 1988.

4. Herstein, I. N., *Topics in Algebra*. 2nd ed., New York: John Wiley, 1975.

F. The Concept of Finite Memory of a Stochastic Process or of a Random Field.

Results on the concept of finite memory of a stochastic process or of a random field are presented in the paper entitled "A comment on finite memory of stochastic processes" which appeared in the September 1992 issue of the *IEEE Transactions on Signal Processing* and is given in Appendix F. It is shown that a recently proposed concept of finite memory for a zero mean strictly stationary stochastic process results in a stochastic of random variables each of which is almost surely equal to zero. We eagerly note that the earlier work about which this paper comments was work supported by the Office of Naval Research.

A COMMENT ON FINITE MEMORY OF STOCHASTIC PROCESSES

Gary L. Wise
Department of Electrical and Computer Engineering
and
Department of Mathematics
The University of Texas at Austin
Austin, Texas 78712
(512) 471-3356

Eric B. Hall
Department of Electrical Engineering
Southern Methodist University
Dallas, Texas 75275
(214) 692-4367

EDICS number: 5.1.5

ABSTRACT

It is shown that a recently proposed concept of finite memory for a zero mean strictly stationary stochastic process results in a stochastic process of random variables each of which is almost surely equal to zero.

DEVELOPMENT

Let k be a positive integer and let $\{X(t): t \in \mathbf{R}\}$ be a zero mean, strictly stationary stochastic process defined on some probability space and taking values in $\mathbf{R}^k$. In [1] such a stochastic process is said to have finite memory if there exists a positive real number D such that for any positive integer n and for any n times $t_1, t_2, \ldots, t_n$, the two sets of random variables $\{X(t_1), X(t_2), \ldots, X(t_n)\}$ and $\{X(t_1 + d), X(t_2 + d), \ldots, X(t_n + d)\}$ are statistically independent for any real number d such that $d > D$. Here we note that such a stochastic process is degenerate in the sense that any random variable in the stochastic process is almost surely equal to zero.

First, consider the situation of a finite memory, zero mean, strictly stationary stochastic process as above. Let θ denote the origin of $\mathbf{R}^k$; let $n = 2$; let $t_1 = -2D$; and let $t_2 = 0$. In this case note that the set of random variables $\{X(-2D)$ and $X(0)\}$ and the set of random variables $\{X(-2D + d)$ and $X(d)\}$ must be statistically independent for any $d > D$. If we choose $d = 2D$, then we see that $X(0)$ must be independent from itself, and hence, since $E[X(0)] = \theta$, we see that $X(0) = \theta$ a.s. Now, since the stochastic process is strictly stationary, it follows that for any real number t, $X(t) = \theta$ a.s. Hence, for each real number t, each component of the random vector $X(t)$ is almost surely zero.

ACKNOWLEDGEMENT

This research was supported by the Office of Naval Research under Grant N00014-90-J-1712.

REFERENCE

[1] M. J. Hinich and G. R. Wilson, "Time delay estimation using the cross bispectrum," *IEEE Transactions on Signal Processing*, Vol. 40, January 1992, pp. 106-113.

G. Distribution of the Determinant of a Random Matrix

Results on the distribution of the determinant of a random matrix are presented in the paper entitled "A note on the distribution of the determinant of a random matrix" which appeared in the February 1991 issue of *Statistics and Probability Letters* and is given in Appendix G. An analysis of the tail behavior of a probability density function of the determinant of a random matrix is presented, and an oversight in an earlier paper on this subject is noted.

Statistics & Probability Letters 11 (1991) 147–148
North-Holland

February 1991

A note on the distribution of the determinant of a random matrix

Gary L. Wise

Department of Statistics, University of California, Berkeley, CA 94720, USA

Eric B. Hall

Department of Electrical Engineering, Southern Methodist University, Dallas, TX 75275, USA

Received December 1989
Revised February 1990

Abstract: An analysis of a probability density function of the determinant of a random matrix is presented, and an oversight in an earlier paper on this subject is noted.

Keywords: Determent, random matrix.

Let A_1, A_2, A_3, and A_4 denote four mutually independent identically distributed random variables defined on the same probability space and uniformly distributed over $[0, 1]$. Denote by M the following matrix:

$$M = \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}.$$

In Williamson and Downs (1988), a graph was presented for a probability density function (pdf) for the determinant of M. In this paper we show that this graph provides a misleading representation for such a pdf.

It follows straightforwardly that the random variable A_1A_4 has a pdf given by $-\chi_{(0,1)}(x)\log(x)$ (all logarithms in our paper are Naperian logarithms). Let X and Y be independent random variables defined on a common probability space, each having pdf $-\chi_{(0,1)}(x)\log(x)$. Further, let $W = X - Y$. Notice that the distribution of W is the same as the distribution of the determinant of M. Also, notice that there exists a pdf for W which is even and which is supported on $[-1, 1]$. For $x \in (0, 1)$ we have that a pdf of W at x, say $p(x)$, is given by

$$p(x) = \int_0^{1-x} \log(w + x)\log(w)\,dw.$$

Using integration by parts, we get

$$p(x) = \int_0^{1-x} \left[\frac{w}{w+x} - \left(\frac{w}{w+x}\right)\log(w)\right] dw.$$

Now, upon simplification we get

$$p(x) = (1-x)[2 - \log(1-x)] + x\log(x) + x\int_0^{1-x} \frac{\log(w)}{w+x}\,dw.$$

For a fixed positive number α, let $h: (0, 1) \to \mathbb{R}$ via

$$h(w) = \log(w)\log(1 + w/\alpha) + \int_{-w/\alpha}^{0} \frac{\log(1-t)}{t}\,dt.$$

This research was partially supported by the Office of Naval Research under Grant No. N00014-90-J-1712.

0167-7152/91/$03.50

Note that $h'(w) = \log(w)/(w+\alpha)$. Thus we see that

$$\int_0^{1-x} \frac{\log(w)}{w+x}\,dw = -\log(1-x)\log(x) + \int_{1-1/x}^{0} \frac{\log(1-t)}{t}\,dt,$$

and therefore we get that

$$p(x) = (1-x)[2-\log(1-x)] + x\log(x) - x\log(1-x)\cdot(\quad) + x\int_{1-1/x}^{0} \frac{\log(1-t)}{t}\,dt.$$

It follows straightforwardly that

$$p''(x) = \frac{-\log(1-x)}{x} - \frac{\log(x)}{x(1-x)} + \frac{\log(x)}{1-x},$$

and hence we see that $p''(x) > 0$. Thus, we note that, restricted to $(0, 1)$, $p(\cdot)$ is convex. This shows that the graph given in Fig. 1 of Williamson and Downs (1988) is misleading as a representation for a cdf of the determinant of M, and it points out an important, yet often unheeded, caveat associated with truncation effects in numerical schemes.

Reference

Williamson, R.C. and T. Downs (1988). The inverse and determinant of a 2×2 uniformly distributed matrix, *Statist. Probab. Lett.* 7, 167–170.

Statistics & Pro
North-Holland

Loca
equal
with

Manzoor
University of Q

Yogendra
Department of

Received April
Revised Februa

Abstract: In th
parameters is
transformations
studied by Cha

Keywords: Hete

1. Introducti

Consider tw

Y_i

where Y_i is
$(n_i \geq k)$, β i
with zero me
models have
Chap. 9)). W
who propose
This test was
cases; (i) the
by Geisser
alternative te

No optim
arising out o

This work was p

0167-7152/91/$

H. Stationary Random Processes

Results on stationary random processes are presented in the paper entitled "A Cautionary Aspect of Stationary Random Processes" which appeared in the November 1991 issue of *IEEE Transactions on Circuits and Systems*and is given in Appendix H. A problem associated with determining the stationarity of a random process from discrete time samples is noted. In particular, a nonstationary Gaussian random process $\{X(t): t \in \mathbf{R}\}$is given such that for any positive real number Δ, the discrete time random process $\{X(n\Delta): n \in \mathbf{Z}\}$ is strictly stationary, where $\mathbf{Z}$ denotes the set of all integers.

A Cautionary Aspect of Stationary Random Processes

Gary L. Wise

Abstract—**A problem associated with determining the stationarity of a random process from discrete time samples is noted.**

Development

Let $\{X(t)\colon t \in \boldsymbol{R}\}$ be some random process. Let Δ be a positive real number, and consider the random process $\{X(n\Delta)\colon n \in \boldsymbol{Z}\}$, where $\boldsymbol{Z}$ denotes the set of integers. In many practical problems, one would be interested in knowing whether or not the random process $\{X(t)\colon t \in \boldsymbol{R}\}$ is stationary. However, due to the current digital trend in signal processing, one might attempt to determine the stationarity of $\{X(n\Delta)\colon n \in \boldsymbol{Z}\}$. What if $\{X(n\Delta)\colon n \in \boldsymbol{Z}\}$ were stationary for any positive real number Δ? Would this imply stationarity of $\{X(t)\colon t \in \boldsymbol{R}\}$? We show by an example that the answer to this second question is no.

Let $\{Y(t)\colon t \in \boldsymbol{R}\}$ be a stationary zero mean Gaussian random process defined on some underlying probability space, such that $E[Y(t)\ Y(t+\tau)] = e^{-|\tau|}$. Define the stationary zero mean Gaussian random process $\{Z(t)\colon t \in \boldsymbol{R}\}$ via $Z(t) = Y(2t)$. Now, define a zero mean Gaussian random process $\{X(t)\colon t \in \boldsymbol{R}\}$ via $X(t) = Y(t)$ if t is rational, and $X(t) = Z(t)$ if t is irrational. Observe that $\{X(t)\colon t \in \boldsymbol{R}\}$ is not stationary since if t and τ are rational, then $E[X(t)\ X(t+\tau)] = e^{-|\tau|}$, yet if t is irrational and τ is rational, then $E[X(t)\ X(t+\tau)] = e^{-2|\tau|}$. Next, pick any positive real number Δ. Note that if Δ is rational, then $n\Delta$ is rational for all integers n. Also, if Δ is irrational, then $n\Delta$ is irrational for all nonzero integers n. Further, $Y(0) = Z(0) = X(0)$. Hence, for all integers n, if Δ is rational, then $X(n\Delta) = Y(n\Delta)$, and if Δ is irrational, then $X(n\Delta) = Z(n\Delta)$. Thus for any positive real number Δ, $\{X(n\Delta)\colon n \in \boldsymbol{Z}\}$ is a stationary Gaussian random process, yet $\{X(t)\colon t \in \boldsymbol{R}\}$ is a Gaussian random process that is not stationary.

Manuscript received July 12, 1991. This work was supported by the Office of Naval Research under Grant N00014-90-J-1712. This paper was recommended by Editor R. Liu.

The author is with the Department of Statistics, University of California, Berkeley, CA 94720.

IEEE Log Number 9102901.

J. Martingale characteristics of a Weiner process

Results on a martingale characterization of a Weiner process are presented in the paper entitled "A counterexample to a martingale characterization of a Wiener process" which was planned to have been given in given in Appendix J. We hasten to note that this investigator has recently gone through the trauma of having experienced a stroke. Unfortunately, he lost all of his documentation of this paper at some time during this experience. However, the paper should appear in the journal *Statistics and Probability Letters*, and in it we show that a recently proposed scheme for characterizing a Wiener process was incorrect. We regret this omission in this report.

K. Estimation of a random variable based on multidimensional data

Results on estimation of a random variable based on multidimensional data are presented in the paper entitled "Estimation of a random variable based on multidimensional data" which appeared in the *Proceedings of the 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing* and is presented in Appendix K. Several aspects associated with the mean square estimation of a second order random variable based upon elements from a random field are considered. Throughout the paper, the oft-neglected role of the underlying probability space is stressed. Numerous examples are presented that point out many of the subtleties associated with this endeavor.

ESTIMATION OF A RANDOM VARIABLE BASED ON MULTIDIMENSIONAL DATA

Gary L. Wise
Department of Electrical and Computer Engineering
and
Department of Mathematics
The University of Texas at Austin
Austin, Texas 78712

Eric B. Hall
Department of Electrical Engineering
Southern Methodist University
Dallas, Texas 75275

ABSTRACT

Several aspects associated with the mean square estimation of a second order random variable based upon elements from a random field are considered. Throughout the paper, the oft-neglected role of the underlying probability space is stressed. Numerous examples are presented that point out many of the subtleties associated with this endeavor.

INTRODUCTION

Let $\mathbf{N}$ denote the set of positive integers and let $\aleph_0$ denote the cardinality of $\mathbf{N}$. Let $k \in \mathbf{N}$ and let ε be a positive number. Let $S = \{k\varepsilon: k \in \mathbf{N}\}$, and let $G = S^k$. We may view G as a grid of points. For points $t = (t_1, t_2, \ldots, t_k) \in G$, and $s = (s_1, s_2, \ldots, s_k) \in G$, we write $s \le t$ to mean that $s_i \le t_i$ for $i = 1, 2, \ldots, k$. Note that this relation is a partial order on G. This paper will be concerned with attempts to estimate a second order random variable X via estimates of the form $E[X \mid Y_p: p \le n]$, where the Y_p's are random variables indexed by G. Notice that $\{Y_p: p \in G\}$ is a random field. Also, note that, in this situation, $\{E[X \mid Y_p: p \le n]: n \in G\}$ is a second order random field.

Let $(\Omega, \mathcal{F}, P)$ be a probability space, let X be a second order random variable defined on $(\Omega, \mathcal{F}, P)$, and let $\{Y_p: p \in G\}$ be a random field defined on $(\Omega, \mathcal{F}, P)$. For each $n \in G$, pick and fix a version of $E[X \mid Y_p: p \le n]$. For $n \in G$, let $\mathcal{F}_n = \sigma(Y_p: p \le n)$ and let $M_n = E[X \mid \mathcal{F}_n]$. As noted above, M_n is a second order random variable. Further, M_n is $\mathcal{F}_n$-measurable. Finally, note that for n_1 and n_2 in G with $n_1 \le n_2$, $E[M_{n_2} \mid \mathcal{F}_{n_1}] = E[E[X \mid Y_p: p \le n_2] \mid Y_p: p \le n_1] = E[X \mid Y_p: p \le n_1] = M_{n_1}$ a.s. via standard properties of conditional expectation. Thus, we see that $\{M_n: n \in G\}$ is a second order multiparameter martingale with respect to the filtration $\{\mathcal{F}_n: n \in G\}$. We remark that the above comments hold where $(\Omega, \mathcal{F}, P)$ is *any* probability space. Now we pose the question: how might one estimate the random variable X from the data $\{Y_p: p \in G\}$ so as to minimize the mean square error?

PRELIMINARIES

Before proceeding, we will review some definitions and introduce some conventions and notations which will prove useful. We will let $\mathcal{B}(\mathbf{R})$ denote the family of Borel subsets of R. For a set S, we will let $\mathbf{P}(S)$ denote the power set of S and I_S denote the indicator function of S. We will let $\mathbf{R}$ denote the set of real numbers. If A is a subset of $\mathbf{R}$, $-A$ will denote the set $\{x \in \mathbf{R}: -x \in A\}$. By a standard Gaussian measure we will mean a Gaussian measure whose first moment is zero and whose second moment is one. For any measure m on $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ we will let m_* denote the inner measure on $(\mathbf{R}, \mathbf{P}(\mathbf{R}))$ induced by m and we will let m^* denote the outer measure on $(\mathbf{R}, \mathbf{P}(\mathbf{R}))$ induced by m. Recall that for any subset A of R, m_* and m^* are defined via $m_*(A) = \sup\{ m(B) : A \supset B \in \mathcal{B}(\mathbf{R}) \}$ and $m^*(A) = \inf\{ m(B) : A \subset B \in \mathcal{B}(\mathbf{R}) \}$. We will say that a subset S of the reals is saturated non-m-measurable if $m_*(S) = m_*(S^c) = 0$. We will let λ denote Lebesgue measure on $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$.

Throughout this paper, we will let n and p, with or without subscripts, denote elements of G; we will let m, ν, P, and μ denote measures; we will let i, j, and k denote positive integers; and we will let N denote an integer greater than one.

Presented at the 1992 International Conference on Acoustics, Speech, and Signal Processing, March 23-26, 1992, San Francisco, California; to be published in the Proceedings of the Conference.

The following result is developed in [1].

Theorem 1: Let N be an integer greater than 1. There exist N subsets $S_1, S_2, \ldots, S_N$ of the real line that partition the reals and are each saturated non-Lebesgue measurable. Letting $S_1, \ldots, S_N$ be as above and letting μ be a probability measure on $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ such that μ is equivalent to Lebesgue measure on $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$, there exists a probability space $(\mathbf{R}, \mathcal{G}, P)$ where $\mathcal{G}$ is given by $\{(S_1 \cap B_1) \cup \ldots \cup (S_N \cap B_N)\colon B_i \in \mathcal{B}(\mathbf{R}) \text{ for } i=1, \ldots, N\}$ and where $P((S_1 \cap B_1) \cup \ldots \cup (S_N \cap B_N)) = \frac{1}{N}[\mu(B_1) + \ldots + \mu(B_N)]$.

The following corollary is an immediate result of Theorem 1.

Corollary 1: Let Theorem 1 set notation. The σ-algebra $\mathcal{G}$ includes $\mathcal{B}(\mathbf{R})$ and contains $S_1, \ldots, S_N$. Further, the probability measure P agrees with μ on $\mathcal{B}(\mathbf{R})$ and, for the probability space $(\mathbf{R}, \mathcal{G}, P)$, $\mathcal{B}(\mathbf{R})$ is *independent* of $\sigma(S_1, \ldots, S_N)$.

Further, we recall the following result from [2], which calls into question the validity of many claims in mean square estimation theory.

Theorem 2: For any real number B, there exists a probability space $(\Omega, \mathcal{F}, P)$, two bounded random variables X and Y defined on $(\Omega, \mathcal{F}, P)$, and a function $f\colon \mathbf{R} \to \mathbf{R}$ such that $E[(Y - E[Y \mid X])^2] > B$ and yet $f(X) = Y$ pointwise on Ω.

Now, we present an observation which will be of use to us.

Lemma 1: Let the introduction set notation. If $n_1 \le n_2 \le \ldots$ is any nondecreasing infinite sequence of elements from G, then the sequence of random variables $\{E[X \mid Y_{n_i}\colon i = 1, 2, \ldots, j]\colon j \in \mathbf{N}\}$ is a second order martingale with respect to the filtration $\{\mathcal{F}_{n_i}\colon i \in \mathbf{N}\}$.

Proof: First, it follows from Jensen's inequality for conditional expectations that $\{E[X \mid Y_{n_i}\colon i = 1, 2, \ldots, j]\colon j \in \mathbf{N}\}$ is a sequence of second order random variables. Also, it follows from the definition of conditional expectation that $E[X \mid Y_{n_i}\colon i = 1, 2, \ldots, j]$ is $\mathcal{F}_{n_j}$-measurable. Finally, note that for positive integers $j_1 < j_2$, it follows from standard properties of conditional expectation that $E[E[X \mid Y_{n_i}\colon i = 1, 2, \ldots, j_1] \mid \mathcal{F}_{n_{j_2}}] = E[X \mid Y_{n_i}\colon i = 1, 2, \ldots, j_1]$ a.s. **Q.E.D.**

DEVELOPMENT

The development will be a set of examples which will serve to indicate some problems which may await the unwary investigator. In particular, these examples suggest the importance of a careful consideration of the underlying probability space.

EXAMPLE A: For an integer $N > 1$, let $S_1, \ldots, S_N$ be subsets of the real line that partition the reals and are each saturated non-Lebesgue measurable. Let $(\mathbf{R}, \mathcal{G}, P)$ be the probability space provided by Theorem 1 for these sets where the measure μ in Theorem 1 is taken to be standard Gaussian measure on $\mathcal{B}(\mathbf{R})$. All random variables in this example will be defined on the probability space $(\mathbf{R}, \mathcal{G}, P)$. Let $X(\omega) = I_{S_1}(\omega) - I_{S_1^c}(\omega)$. Note that X is a Bernoulli random variable, and $P(X = -1) = 1 - \frac{1}{N}$ and $P(X = 1) = \frac{1}{N}$. For some $p_0 \in G$, let $Y_{p_0}(\omega) = \sigma\omega$, where σ is a positive real number, and for all other p's in G, let $Y_p = 0$. Note that $\{Y_p\colon p \in G\}$ is a Gaussian random field. Further, note that $\sigma\{Y_p\colon p \in G\} = \mathcal{B}(\mathbf{R})$ and $\sigma(X) = \{\Omega, \emptyset, S_1, S_1^c\}$. Recalling Corollary 1, we see that X is independent of the data $\{Y_p\colon p \in G\}$. Thus, we see that $M_n = \frac{2-N}{N}$ a.s. for all $n \in G$. Further, we see that $P(M_n = X) = 0$ for all $n \in G$. However, if one knew p_0, one could reconstruct X precisely from Y_{p_0} via $X = g(Y_{p_0})$ pointwise on $\mathbf{R}$, where $g\colon \mathbf{R} \to \mathbf{R}$ via $g(x) = I_{S_1}\left(\frac{x}{\sigma}\right) - I_{S_1^c}\left(\frac{x}{\sigma}\right)$. Note that X can be precisely written as a function of an independent random variable. Further, note that this can be done regardless of how small or large the positive variance of Y_{p_0} is, and $P(X = 1)$ can be arbitrarily small by choice of N. Of course, knowledge of p_0 is crucial.

EXAMPLE B: In this example, we let the random variable X be Gaussian, and we get a result similar to that in Example A. For an integer $N > 1$, let the probability space be the same as in Example A, and let $X(\omega) = \omega[I_{S_1} - I_{S_1^c}]$. For any real Borel set B, note that $P(X \in B) = P((S_1 \cap B) \cup (S_1^c \cap (-B))) = \mu(B)$, where μ as in Example A is standard Gaussian measure on $\mathcal{B}(\mathbf{R})$. Thus, we see that X is a standard Gaussian random variable. Let the random field $\{Y_p\colon p \in G\}$ be as in Example A. Note that $E[X \mid Y_{p_0}] = E[\omega[I_{S_1} - I_{S_1^c}] \mid \mathcal{B}(\mathbf{R})] = \omega E[I_{S_1} - I_{S_1^c} \mid \mathcal{B}(\mathbf{R})] = \omega\left[\frac{2-N}{N}\right]$ a.s., since the identity map is Borel measurable, since S_1 is independent of

$\mathcal{B}(\mathbf{R})$, and since $P(S_1) = \frac{1}{N}$. In this case, we see that for $p_0 \le n$, $M_n = \omega\left[\frac{2-N}{N}\right]$ a.s. Thus, when $N = 2$, $M_p = 0$ a.s. for all $p \in G$. On the other hand, for large N, M_n is close to $-\omega$ for $p_0 \le n$, and $P(X = -\omega) = 1 - \frac{1}{N}$. Nevertheless, we have that $P(M_n = X) = 0$ for any $n \in G$. However, we can once again write X precisely as a function of Y_{p_0}. That is, let $h:\mathbf{R}\to\mathbf{R}$ via $h(x) = \left(\frac{x}{\sigma}\right) g(x)$, where g is as in Example A, and note that $X = h(Y_{p_0})$ pointwise on R.

EXAMPLE C: For an integer $N > 1$, let the probability space be the same as in Example A. Let $X(\omega) = \omega[N I_{S_1}(\omega) - \frac{N}{N-1} I_{S_1^c}(\omega)]$. Notice that $E[X] = 0$. Also, notice that if $N = 2$, X is a Gaussian random variable. Let the random field $\{Y_p: p \in G\}$ be defined via $Y_p(\omega) = s_p\omega$ for each $p \in G$, where $\{s_p: p \in G\}$ is a set of nonzero real numbers. Notice that in this case, $\{Y_p: p \in G\}$ is a Gaussian random field, and each random variable in this random field has zero mean and a positive variance. Now, what if we tried to estimate X from elements of the random field $\{Y_p: p \in G\}$? Notice that $\sigma(Y_p: p \in G) = \mathcal{B}(\mathbf{R})$. Further, notice that $E[X \mid \mathcal{B}(\mathbf{R})] = E[\omega[N I_{S_1}(\omega) - \frac{N}{N-1} I_{S_1^c}(\omega)] \mid \mathcal{B}(\mathbf{R})] = 0$ a.s., since the identity map is $\mathcal{B}(\mathbf{R})$-measurable, since S_1 is independent of $\mathcal{B}(\mathbf{R})$, and since $E[X] = 0$. Thus, for any $n \in G$, $M_n = 0$ a.s. However, notice that for any $p \in G$, X can be written precisely as a function of Y_p. That is, $X(\omega) = r_p(Y_p(\omega))$ where $r_p:\mathbf{R}\to\mathbf{R}$ via

$$r_p(x) = \left(\frac{x}{s_p}\right)\left[N I_{S_1}\left(\frac{x}{s_p}\right) - \frac{N}{N-1} I_{S_1^c}\left(\frac{x}{s_p}\right)\right].$$

EXAMPLE D: In this example, assume $k > 1$, and let the probability space $(\Omega, \mathcal{F}, P)$ be given by R, $\mathcal{B}(\mathbf{R})$, and standard Gaussian measure on $\mathcal{B}(\mathbf{R})$. Let $X(\omega) = \omega$. Let d be the element of G given by $d = (\varepsilon, \varepsilon, \ldots, \varepsilon)$. Now, let $Y_d = \omega I_{(-\infty, 1]}(\omega)$, and for integers $j > 1$, let $Y_{(jd)} = \omega I_{(j-1, j]}(\omega)$. For n in G but not equal to positive integral multiples of d, let $Y_n = 0$. Now how might we estimate X? Fix any point n_0 in G, and for positive integers $i \le k$ and positive integers j, let p_j be the point whose coordinates are the same as those of n_0 except that the i-th coordinate is the i-th coordinate of n_0 plus $j\varepsilon$. Then $\{E[X \mid Y_{p_1}, \ldots, Y_{p_j}]: j \in \mathbf{N}\}$ is an ordinary martingale (see Lemma 1); indeed, in the context of random fields, it is called an i-martingale. It follows immediately that this martingale is equal to a fixed random variable for all j greater than some positive integer J. Further, note that the martingale does not converge to X. Thus, for no value of $i = 1, 2, \ldots, k$, will this i-martingale converge to X. However, for positive integers j, if we let $q_j = jd$, the martingale $\{E[X \mid Y_{q_1}, \ldots, Y_{q_j}]: j \in \mathbf{N}\}$ is an ordinary martingale (see Lemma 1); and it follows from elementary martingale theory that this martingale converges in $L_p(\Omega, \mathcal{F}, P)$, for any $p \in (1, \infty)$, and a.s. as $m\to\infty$ to X; indeed, here it will converge pointwise.

EXAMPLE E: For an integer $N > 1$, let the probability space be the same as in Example A. Let $X(\omega) = \omega[I_{S_1} - I_{S_1^c}]$. As in Example B, note that X is a standard Gaussian random variable. Now, for each $n \in G$, let s_n denote the sum of the components of n, and let $Y_n(\omega) = \omega s_n$. Notice that $\{Y_p: p \in G\}$ is a Gaussian random field, and Y_p has zero mean and a variance of $(s_p)^2$. Now, for any $p \in G$, as in Example B, $M_p(\omega) = \omega\left[\frac{2-N}{N}\right]$ a.s. If $N = 2$, we see that $M_p = 0$ a.s. for all $p \in G$. In any case, $M_p = E[X \mid Y_q \le p]$ must be a Borel measurable function of ω, since $\sigma(Y_q: q \le p) = \mathcal{B}(\mathbf{R})$, and X is not a Borel measurable function of ω. Thus, once again, conditional expectation is of no help to us here. However, for any $p \in G$, we can write X precisely as a function of Y_p; that is, $X = \frac{Y_p}{s_p} g\left(\frac{Y_p}{s_p}\right)$ pointwise on R, where g is the function given in Example A.

EXAMPLE F: For an integer $N > 1$, let the probability space be the same as in Example A. Let $X(\omega) = \omega$, and note that X is a standard Gaussian random variable. For some N distinct points $p_1, p_2, \ldots, p_N$ in G, let $Y_{p_i}(\omega) = \omega I_{S_i}(\omega)$, and for all other points $p \in G$, let $Y_p = 0$. Note that $X(\omega) = \sum_{i=1}^{N} Y_{p_i}(\omega)$, pointwise on R. Further, for certain points $n \in G$, $M_n(\omega) = X(\omega)$ pointwise on R, and the cardinality of such points n is $\aleph_0$. Note that if $k > 1$, depending on the location of the N points $p_1, p_2, \ldots, p_N$ in G, there could exist a subset H of G having cardinality $\aleph_0$ such that $M_n = 0$ for all $n \in H$.

EXAMPLE G: For an integer $N > 1$, let the probability space be the same as in Example A. Let $X(\omega) = \omega$, and note that X is a standard Gaussian random variable. For some N

points $p_1, p_2, \ldots, p_N$ in G, let $Y_{p_i}(\omega) = \omega I_{S_i}(\omega) + 1$.

Note that $X(\omega) = \left(\sum_{i=1}^{N} Y_{p_i}(\omega)\right) - N = \left(\prod_{i=1}^{N} Y_{p_i}(\omega)\right) - 1$

pointwise on R. Further, for certain points $n \in G$, $M_n(\omega) = X(\omega)$ pointwise on R, and

$\text{card}(\{n \in G: M_n(\omega) = X(\omega) \text{ for all } \omega \in R\}) = \aleph_0$.

EXAMPLE H: Let k be greater than one, and let $(\Omega, \mathcal{F}, P)$ be a probability space on which can be defined a random field $\{Z_p: p \in G\}$ of identically distributed, mutually independent random variables each having a probability density function given by

$$f(x) = \begin{cases} \dfrac{C}{x^2(\log|x|)^2} & \text{if } |x| > 2 \\ 0 & \text{if } |x| \le 2 \end{cases}$$

where C is the normalizing constant, and a zero mean unit variance Gaussian random variable X independent of $\{Z_p: p \in G\}$. Define the random field $\{Y_p: p \in G\}$ via $Y_p = Z_p + X$. Notice that the mean of Z_p exists and is zero for each $p \in G$. How might one attempt to estimate X from the data $\{Y_p: p \in G\}$? For $t \in G$, let $|t| = e^{-k}\prod_{i=1}^{k} t_i$. With an eye on Kolmogorov's strong law of large numbers, one might be tempted to try to estimate X via an estimate of the form $\hat{X}_n = \frac{1}{|n|} \sum_{\{p \in G:\, p \le n\}} Y_p$. Note that

$\hat{X}_n = X + \frac{1}{|n|} \sum_{\{p \in G:\, p \le n\}} Z_p$, and that the Z_p's have zero mean. Might we guess that $\hat{X}_n$ should then converge to X, in some meaningful sense? If so, we would be well advised to guess again, since it follows from [3] (see also [4, pp. 369-370] for the k=2 case) that

$$\limsup_{p \in G} \frac{1}{|n|} \left| \sum_{\{p \in G:\, p \le n\}} Z_p \right| = \infty \text{ a.s.}$$

EXAMPLE I: Consider the situation depicted in Example H. Recall that G is a countable set, and let $\{t_n: n \in N\}$ be an enumeration of G. Now, recalling Kolmogorov's strong law of large numbers, note that $\frac{1}{k}\sum_{i=1}^{k} Y_{t_i}$ converges to almost surely to X as $k \to \infty$.

CONCLUSION

We have developed a set of examples pointing out some caveats in the use of multiparameter martingales in estimation theory. In particular, we noted some instances in which estimators existed which yielded superior performance than estimators based on conditional expectation. We hope these results will be of use to those concerned with such endeavors.

ACKNOWLEDGEMENT

This research was partially supported by the Office of Naval Research under Grant N00014-90-J-1712.

REFERENCES

[1] E. B. Hall and G. L. Wise, "A result on multidimensional quantization," to appear in *Proceedings of the American Mathematical Society*.

[2] G. L. Wise, "A note on a common misconception in estimation," *Systems and Control Letters*, Vol. 5, pp. 355-356, April 1985.

[3] R. T. Smythe, "Strong laws of large numbers for r-dimensional arrays of random variables," *Annals of Probability*, Vol. 1, February, 1973.

[4] J. B. Walsh, "Martingales with a multidimensional parameter and stochastic integrals in the plane," in *Lectures in Probability and Statistics*, Lecture Notes in Mathematics number 1215, edited by G. Del Pino and R. Rebolledo, Springer-Verlag, 1986.