We can now proceed with the construction of a code whose expected coding length is too small. To guarantee invertibility we assume that all our codes, $\hat{c}_1, \tilde{c}_m, c_n, M \le n \le N$, were defined so that any concatenation of images of any of these codes is uniquely decodable. This can be done by adding a short prefix to each code specifying whether blocks of length 1, m, or n, with $M \le n \le N$, were used. (There are many ways to do this; one particular way is described in [3].) The code $\bar{c}_K$ is defined as follows. If $x_1^K \notin A_K$ then each letter is encoded separately, using the code $\hat{c}_1$. Otherwise, $x_1^K \in A_K$, and the partition $\{J_t\}$ is constructed using Lemma 1. The coding, $\bar{c}_K(x_1^K)$, is the concatenation of the binary sequences $\bar{J}_t$, defined as follows. Suppose $J_t = [u, v]$.

1) If $v = u$, then $\bar{J}_t = \hat{c}_1(x_u)$.
2) If $v = u + m - 1$ then $\bar{J}_t = \tilde{c}_m(x_u^v)$.
3) If $v = u + L - 1$, with $L > m$, then $\bar{J}_t = c_L(x_u^v)$.

Let $\bar{l}(x_1^K)$ be the length function for the code $\bar{c}_K$. If $x_1^K \in A_K$ then Lemma 1 gives the upper bound

$$\bar{l}(x_1^K) \le d(3\delta K) + (H - \epsilon)\gamma\frac{K}{2} + \left(1 - \frac{\gamma}{2} - 3\delta\right)K(H + \delta).$$

The ergodic theorem can now be applied to guarantee that $\mathrm{Prob}(A_K)$ goes to 1. Thus we can choose δ small enough and K large enough to contradict the Shannon lower bound, (5). This completes the proof of Theorem 3. □

Proof of the Entropy Theorem: Let $\{c_n\}$ be a sequence of noiseless codes such that $\limsup_n l(x_1^n)/n \le H$, a.s. Fix $\epsilon > 0$ and define

$$D_n = \{x : l(x_1^n)/n < H + \epsilon\}, \qquad \tilde{D}_n = \{x_1^n : x \in D_n\}.$$

Note that $x \in D_n$, eventually almost surely, in the sense that for almost all x there is an $N(x)$ such that $x \in D_n, n \ge N(x)$. Note also that $|\tilde{D}_n| \le 2^{n(H+\epsilon)}$, since c_n is invertible. Thus, if

$$C_n = \{x : p(x_1^n) \le 2^{-n(H+2\epsilon)}\},$$

then

$$\mathrm{Prob}\,(C_n \cap D_n) \le |\tilde{D}_n| 2^{-n(H+2\epsilon)} \le 2^{-n\epsilon}.$$

Therefore, $x \notin C_n \cap D_n$, eventually almost surely, and so $x \notin C_n$, eventually almost surely; hence, we have established the upper bound result

$$\limsup_n \frac{1}{n} \log \frac{1}{p(x_1^n)} \le H, \text{ a.s.}$$

To obtain the analogous lower bound result define

$$U_n = \{x : p(x_1^n) \ge 2^{-n(H-\epsilon)}\}, \qquad \tilde{U}_n = \{x_1^n : x \in U_n\},$$

and note that $|\tilde{U}_n| \le 2^{n(H-\epsilon)}$. Thus there is a one-to-one function ϕ from $\tilde{U}_n$ into binary sequences of length 2 more than $n(H - \epsilon)$, such that the first symbol in $\phi(x_1^n)$ is always a 0. Let $\tilde{c}_n$ be the code obtained by adding the prefix 1 to the code c_n. Define $\bar{c}_n(x_1^n) = \phi(x_1^n)$, $x_1^n \in \tilde{U}_n$ and $\bar{c}_n(x_1^n) = \tilde{c}_n(x_1^n), x_1^n \notin \tilde{U}_n$. The resulting code $\bar{c}_n$ is invertible, so that Theorem 3 guarantees that $x \notin U_n$, eventually almost surely. This proves that

$$\liminf_n \frac{1}{n} \log \frac{1}{p(x_1^n)} \ge H, \text{ a.s.},$$

and completes the proof of the entropy theorem. □

References

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

[2] J. Kieffer, "Sample converses in source coding theory," *IEEE Trans. Inform. Theory*, vol. 37, pp. 263-268, 1991.

[3] D. Ornstein and P. Shields, "Universal almost sure data compression," *Annals of Prob.*, vol. 18, pp. 441-452, 1990.

[4] D. Ornstein and B. Weiss, "The Shannon-McMillan-Breiman theorem for a class of amenable groups," *Israel J. of Math.*, vol. 44, pp. 53-60, 1983.

[5] P. Shields, "Universal almost sure data compression using Markov types," *Probl. Contr. Inform. Theory*, vol. 19, pp. 269-277, 1990.

[6] ____, "The ergodic and entropy theorems revisited," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 263-266, 1987.

[7] J. Ziv and A. Lempel, "A universal algorithm for sequential data compression," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 337-343, 1977.

Two Proofs of Pinsker's Conjecture Concerning Arbitrarily Varying Channels

Rudolf Ahlswede and Ning Cai

***Abstract*—At the 1990 IEEE Information Theory workshop, M. S. Pinsker conjectured the following theorem: For an arbitrary varying channel (AVC) every rate below the random code capacity is achievable with deterministic list codes of *constant* list size, if the average error criterion is used. Two proofs of this theorem are given.**

***Index Terms*—Arbitrarily varying channels, list codes.**

I. Proof of Theorem by the Elimination Technique of [1]

An AVC is defined here by a sequence $\mathscr{A} = (\{P(\cdot|\cdot|s^n): s^n \in \mathscr{S}^n\})_{n=1}^{\infty}$ of sets of transmission probabilities, where for a finite input alphabet $\mathscr{X}$, a finite output alphabet $\mathscr{Y}$, and a finite set $\{w(\cdot|\cdot|s): s \in \mathscr{S}\}$ of stochastic $|\mathscr{X}| \times |\mathscr{Y}|$-matrices,

$$P(y^n|x^n|s^n) = \prod_{t=1}^{n} w(y_t|x_t|s_t), \tag{1}$$

for all $x^n = (x_1, x_2, \cdots, x_n) \in \mathscr{X}^n = \prod_1^n \mathscr{X}$, for all $y^n \in \mathscr{Y}^n$, and for all $s^n \in \mathscr{S}^n$.

From a random code (n, N) of the AVC with average error less than $e^{-\epsilon n}$ one can produce for any $\lambda \in (0,1)$ by random selection of L (used here instead of the letter R in [1]) deterministic codes in the original ensemble and by randomization with the uniform distribution μ^* over those codes a new random code, say

$$\left(\left\{(u_i^l, D_i^l): 1 \le i \le N;\ l = 1, \cdots, L\right\}, \mu^*\right).$$

It was shown in [1], (inequality (4.3)), that this random selection fails to lead for the new code and a fixed $s^n \in \mathscr{S}^n$ to an average error probability less than λ with a probability smaller

Manuscript received September 19, 1990; revised April 5, 1991. This work was presented at the IEEE International Symposium on Information Theory, Budapest, Hungary, June 24-29, 1991.

The authors are with the Fakultät für Mathematik, Universität Bielefeld, Postfach 8640, D-4800 Bielefeld 1, Germany.

IEEE Log Number 9102263.

than

$$e^{-\alpha\lambda L}(1+e^{\alpha}e^{-\epsilon n})^{L}, \qquad \text{for any } \alpha > 0.$$

It suffices now to choose α and L in such a way that

$$|\mathscr{S}|^{n}e^{-\alpha\lambda L}(1+e^{\alpha}e^{-\epsilon n})^{L} < 1. \tag{2}$$

The original choice in [1] was $\alpha = 2$ and $L = n^2$. We show now how to keep L constant.

Clearly, for $\alpha = \epsilon n$ the left side in (2) is bounded from above by

$$\exp\{n\log|\mathscr{S}| - n\epsilon\lambda L + L\}$$

and this is smaller than 1 for

$$L > \left(\epsilon\lambda - \frac{1}{n}\right)^{-1}\log|\mathscr{S}|, \tag{3}$$

and therefore, any constant

$$L > (\epsilon\lambda)^{-1}\log|\mathscr{S}| \tag{4}$$

is good for n sufficiently large.

Now our random code can be modified to the deterministic list code

$$\{(u_i^l, D_i^l) : (i,l) \in \{1,2,\cdots,N\}\times\{1,\cdots,L\}\}.$$

It has list size L, a length $N\cdot L$ even greater than the original code, and an average error probability less than λ, because

$$\sum_{l=1}^{L}\frac{1}{L}\left(\frac{1}{N}\sum_{i=1}^{N}P(D_i^l|u_i^l|s^n)\right) \geq 1-\lambda, \qquad \text{for } s^n \in \mathscr{S}^n.$$

Remark 1: We are curious to know whether the random code capacity can be obtained with a universal list length L^* that is a function of $|\mathscr{X}||\mathscr{Y}|$ alone, that is, it does not depend on the error exponent ϵ (which is a function of the rate), λ, and on the class of matrices, in particular on $|\mathscr{S}|$.

Remark 2: The present and also the following approach yield the bound in (4). Both approaches are based on modifications of certain classical ($L = 1$) codes. By a clever direct selection of the list codes one should do better. The ideas of [3] are promising.

II. Proof of Theorem by a Refinement of the Robustification Technique of [2]

As in [2, Section 5] we start with a code for an *associated compound* channel and build from these a random code for the AVC *via permutations*. Here this is done so that L, defined as in (4), permutations suffice. Then we pass over to the list code as before. For the compound channel with class of matrices

$$\overline{\mathscr{W}} = \left\{\sum_{s} w(\cdot|\cdot|s)Q(s) : Q \in \mathscr{P}(\mathscr{S})\right\},$$

the coding theorem says that any rate R below the random code capacity for the AVC, that is,

$$\overline{C} = \max_{P\in\mathscr{P}(\mathscr{X})}\min_{\overline{w}\in\overline{\mathscr{W}}} I(P,\overline{w}),$$

is achievable with deterministic codes of an error probability less than $\exp\{-n\epsilon(R)\}$, where $\epsilon(R) > 0$. Clearly, this code meets the same error bound for the compound channel with class of matrices $\mathscr{W}_n \subset \overline{\mathscr{W}}$, where

$$\mathscr{W}_n = \left\{\sum_{s} w(\cdot|\cdot|s)Q(s) : Q \in \mathscr{P}_n\right\}, \tag{5}$$

$$\mathscr{P}_n = \left\{Q \in \mathscr{P}(\mathscr{S}) : Q(s) \in \left\{0, \frac{1}{n}, \frac{2}{n}, \cdots, 1\right\} \text{ for all } s \in \mathscr{S}\right\}. \tag{6}$$

With Chebyshev's inequality we derive, from the inequality

$$\sum_{s^n\in\mathscr{S}^n}\frac{1}{N}\sum_{i=1}^{N}P(D_i^c|u_i|s^n)\prod_{t=1}^{n}Q(s_t) \leq e^{-\epsilon n}, \qquad \text{for } Q \in \mathscr{P}_n,$$

that for any $\lambda \in (0,1)$ and $\gamma \in (0,1)$ there are subsets $A_Q^n \subset T_Q^n$, the set of strict Q-typical sequences in $\mathscr{S}^n$, with the properties

$$\frac{1}{n}\sum_{i=1}^{N}P(D_i^c|u_i|s^n) \leq \lambda\gamma, \qquad \text{for } s^n \in A_Q^n, \tag{7}$$

$$|A_Q^n| \geq \left(1 - \frac{1}{\lambda\gamma}e^{-\epsilon n}\right)|T_Q^n|. \tag{8}$$

Consider now the symmetric group (the set of all permutations) Σ_n acting on $\{1,2,\cdots,n\}$. We then define for $s^n \in \mathscr{S}^n$, $A \subset \mathscr{S}^n$ and $\pi \in \Sigma_n$

$$\pi(s^n) = \pi(s_1,\cdots,s_n) = (s_{\pi(1)},\cdots,s_{\pi(n)}), \tag{9}$$

$$\pi(A) = \{\pi(s^n) : s^n \in A\}. \tag{10}$$

The desired result is an immediate consequence of the following fact.

Lemma: Suppose that for a family of sets $\{B_P : P \in \mathscr{P}_n\}$ with $B_P \subset T_P^u$ and for some $\eta > 0$

$$|B_P| \geq |T_P^n|(1 - e^{-\eta n}),$$

then for every $\delta > 0$ and integer

$$L > \frac{\log|\mathscr{S}|}{\delta\eta}, \tag{11}$$

there are permutations $\{\pi_l\}_{l=1}^{L}$ with

$$|\{\pi_l(s^n) : l = 1,2,\cdots,L\} \cap B_P| \geq L(1-\delta),$$
$$\text{for all } P \in \mathscr{P}_n \text{ and all } s^n \in T_P^n, \tag{12}$$

if n is larger than a suitable $n_0(\delta,\eta)$. (If we choose $L > 2(\log|\mathscr{S}|/\delta\eta)$, then $n_0(\delta,\eta) = \lceil 2/\delta\eta \rceil$ does it.)

Proof: Let $\{\tilde{\pi}_l\}_{l=1}^{L}$ be random permutations taking values in Σ_n according to the uniform distribution on Σ_n. Then $\Pr(\tilde{\pi} = \tau) = 1/n!$ for $\tau \in \Sigma_n$ and for all $s^n, s'^n \in T_P^n$ $\Pr(\tilde{\pi}_t(s^n) = s'^n) = \prod_s(nP(s))!/n! = 1/|T_P^n|$. □

Consider now for $s^n \in T_P^n$ the event $B(s^n,\delta)$: "there are at least $\lfloor\delta L\rfloor$ many l's with $\tilde{\pi}_l(s^n) \in B_P^c$." Its probability is given by

$$\Pr(B(s^n,\delta)) = \sum_{l=\lfloor\delta L\rfloor}^{L}\binom{L}{l}\left(\frac{|B_P^c|}{|T_P^n|}\right)^{l}\left(\frac{|B_P|}{|T_P^n|}\right)^{L-l}$$
$$\leq 2^{L}\left(\frac{|B_P^c|}{|T_P^n|}\right)^{\lfloor\delta L\rfloor} \leq e^{L-n\eta\lfloor\delta L\rfloor}.$$

Therefore, the probability that for *all* $s^n \in \mathscr{S}^n$ $B(s^n,\delta)^c$ occurs exceeds $1 - e^{n\log|\mathscr{S}|}e^{L-n\eta\lfloor\delta L\rfloor}$. Clearly, this quantity is positive for L as specified in (11) and n large.

We now continue with the final steps in the second proof. Choose $B_P = A_P^n$, $\delta \in \{0,1\}$ and η such that $e^{-\eta n} \geq 1/(\lambda\gamma)e^{-\epsilon n}$. For large n, η can be made arbitrarily close to ϵ, because λ is constant. By the Lemma, from (12) every $s^n \in T_P^n$ is contained in

$L(1-\delta)$ of the sets $\{\pi_l^{-1}(A_P^n)\}_{l=1}^{L}$. By (7) therefore

$$\frac{1}{L}\sum_{l=1}^{L}\frac{1}{N}\sum_{i=1}^{N}P\left(\pi_l^{-1}(D_i^c)\middle|\pi_l^{-1}(u_i)\middle|s^n\right)$$

$$\le \lambda\gamma(1-\delta)+\delta.$$

By choosing $\delta=\lambda$ and by letting γ tend to zero from the bound in (11), we get again our previous bound on L in (4).

References

[1] R. Ahlswede, "Elimination of correlation in random codes for arbitrarily varying channels," *Z. Wahrscheinlichkeitstheorie u. verw. Geb.*, vol. 44, pp. 159–175, 1978.

[2] ____, "Coloring hypergraphs: A new approach to multi-user source coding," *J. Combin. Inform. Syst. Sci.*, pt. II, vol. 5, pp. 220–268, 1980.

[3] ____, "A method of coding and an application to arbitrarily varying channels," *J. Combin. Inform. Syst. Sci.*, vol. 5, no. 1, pp. 10–35, 1980.

Sequence Estimation and Synchronization from Nonsynchronized Samples

Costas N. Georghiades, *Senior Member, IEEE* and
Marc Moeneclaey

Abstract—**Recent advances in digital signal processing chips have motivated the study of algorithms for data detection and timing extraction that are easily digitally implemented. These algorithms operate on samples of the output of a matched-filter obtained asynchronously with the actual symbol timing in order to extract their timing and data estimates. A general analysis is presented of sampled receivers that handle arbitrary baseband pulse-shapes and arbitrary sampling rates. It is observed that the optimal processing of the matched-filter samples consists of digital interpolation, followed by symbol-by-symbol decoding when sampling is at (or above) the Nyqnist rate, or Viterbi (sequence) decoding when sampling is below the Nyquist rate. Performance is stndied throngh the Cramer–Rao bound on mean-square estimation error and a lower-bound on error-probability.**

Index Terms—**Sequence estimation, synchronization, likelihood-function, sampling, Nyquist pulses.**

I. Introduction

The advent of high-speed digital signal processing (DSP) chips in recent years has generated a need for receiver algorithms that operate on sampled as opposed to analog data to produce their timing and symbol estimates. A number of such algorithms that treat the problem of timing-recovery from sampled data has already appeared in the literature over the past few years [1]–[4]. These algorithms have a common theme in that they operate on samples of the matched filter output taken at some integer multiple of the symbol rate and at some arbitrary phase which may or may not be updated in time. This updating can be done by adjusting the phase of an analog sampler or by digitally interpolating between samples taken by a free-running clock.

In their classic paper, Mueller and Muller [1] study the problem of timing recovery from samples taken at the rate of one sample per symbol. The timing-recovery algorithms introduced are decision-directed and operate on baseband signals. The overall approach is to use the samples in order to estimate the timing-error, which is further used to correct the phase of the analog sampler in order to reduce the error. In a subsequent paper, Agazzi *et al.* [2] study timing-recovery in digital subscriber loops, and show that the wave difference method (WDM), originally introduced in [5], can be implemented with two samples per symbol. According to their proposed algorithm, even at phase lock, neither of the samples coincides with the actual decision timing and an interpolator is used to provide the decision data. The authors note that to avoid the need for an interpolator, sampling at the rate of four samples per symbol may be necessary.

Another timing-error detector of interest which operates on two samples per symbol was recently introduced by Gardner [3]. In contrast to [1], this algorithm is not decision-directed and does not assume carrier phase lock. Although similar to the algorithm introduced in [2], it has the advantage that one of the samples is taken at the data strobe and handles both baseband, as well as carrier signals. As with the other two algorithms in [1] and [2], Gardner's algorithm generates an error-signal that may subsequently be used to correct the sampling phase in the direction of reducing the timing-error, or to control a digital interpolator. The latter approach has the advantage of a fully digital implementation, and may well be the preferred way, especially at sufficiently high sampling rates when interpolation improves.

A further algorithm that does not update the sampling phase was introduced by Oerder and Meyr [4]. Their algorithm is effectively open-loop and operates on samples obtained under the command of a free-running oscillator. Although not explicitly studied, the authors assume that symbol detection is done by operating on data obtained through interpolation of the samples.

All of the previously described algorithms deal almost exclusively with timing-recovery, and assume that once timing is established, data detection can be accomplished using a slicer operating on the samples (or an interpolated value) obtained at the timing estimate. In contrast to [1]–[5], in recent work [6], [7], the problem of optimum *joint* timing and sequence estimation from samples obtained at the rate of one sample per symbol is investigated. The jointly optimal algorithm is based on maximizing a likelihood function for the sampled data, obtained by a free-running oscillator, over both timing and modulation sequences. The joint optimization is done naturally using the Viterbi algorithm in parallel for a number of discrete timing delays. An iterative algorithm for evaluating the likelihood function based on the expectation-maximization (EM) algorithm [8], [9] is also introduced in [10]. Results obtained indicate that such joint, open-loop, algorithms are quite robust to timing errors and can yield completely digital receiver realizations.

In another relevant paper, Falconer and Salz [11] treat the problem of joint data, delay and phase estimation for a Gauss-

Manuscript received November 28, 1989; revised April 26, 1991. This work was presented at the IEEE International Symposium on Information Theory, San Diego, CA, January 14–19, 1990.

C. N. Georghiades is with the Electrical Engineering Department, Texas A&M University, College Station, TX 77843-3128.

M. Moeneclaey is with the Communications Engineering Laboratory, University of Ghent, B-9000 Ghent, Belgium.

IEEE Log Number 9102264.

0018-9448/91 $01.00 ©1991 IEEE