A Variation on the Littlewood-Offord Theme with Applications to Phase Transitions in CDMA Detection

Santosh S. Venkatesh
Department of Electrical and Systems Engineering
University of Pennsylvania
Philadelphia, PA 19104, USA
Email: venkatesh@ee.upenn.edu

***Abstract*—Let $\{\sigma_i, 1 \le i \le m\}$ be a random constellation of vertices of the cube $\{-1.+1\}^n$, $\mathcal{F}$ the family of 2^m constellations corresponding to all possible reflections of the vertices σ_i about the origin. When does the centre of mass $\sum_i \epsilon_i \sigma_i$ of a reflected constellation $\{\epsilon_i \sigma_i\}$ in $\mathcal{F}$ (with $\epsilon_i \in \{-1, +1\}$) uniquely identify it? The constellation centres of mass are reminiscent of Littlewood-Offord sums. I will demonstrate that the critical range for a phase transition is $n/\log n \ll m \ll n \log n$. This question has a direct relevance to the number of users that can be supported in a direct sequence, code division multiple access system.**

Proc. ITA Workshop, San Diego, California, February 2011.

I. Problem Statement

Write $\mathbb{B} := \{-1, +1\}$. Suppose $\boldsymbol{\sigma}_i = (\sigma_1, \ldots, \sigma_n)$ $(1 \le i \le m)$ is a sequence of random points chosen by independent sampling from the law placing equal mass at each of the vertices of the n-dimensional cube, $\mathbb{B}^n$. We call the collection $\{\boldsymbol{\sigma}_i\}$, permitting (rare) replications, a *constellation*. For each $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_m) \in \mathbb{B}^m$, form the weighted sum of vertices

$$\mathbf{r}(\boldsymbol{\epsilon}) := \epsilon_1 \boldsymbol{\sigma}_1 + \cdots + \epsilon_m \boldsymbol{\sigma}_m.$$

Geometrically speaking, the lattice point $\mathbf{r}(\boldsymbol{\epsilon})$ lies on the ray passing through the centre of mass of the constellation $\{\epsilon_i \boldsymbol{\sigma}_i\}$ obtained by reflecting some of the vertices $\boldsymbol{\sigma}_i$ about the origin. Accordingly, we call $\boldsymbol{\epsilon}$ a *reflection vector*. Given the constellation $\{\boldsymbol{\sigma}_i\}$, when is the reflection vector uniquely determined by $\mathbf{r}(\boldsymbol{\epsilon})$, equivalently, the centre of mass of the collection $\{\epsilon_i \boldsymbol{\sigma}_i\}$?

As $\boldsymbol{\epsilon}$ varies over $\mathbb{B}^m$, $\mathbf{r}(\boldsymbol{\epsilon})$ takes values in the set of lattice points

$$\mathbb{L}_m^n := \{-m, -m+2, \ldots, m-2, m\}^n.$$

For each lattice point $\mathbf{k} = (k_1, \ldots, k_n)$ in this set, write $\mathcal{R}_{\mathbf{k}}$ for the equivalence class of those $\boldsymbol{\epsilon} \in \mathbb{B}^m$ for which $\sum_{i=1}^m \epsilon_i \boldsymbol{\sigma}_i = \mathbf{k}$.

Definition 1 We say that a reflection vector $\boldsymbol{\epsilon} \in \mathbb{B}^m$ is *identifiable* (with respect to a given constellation $\{\boldsymbol{\sigma}_i\}$) if $\operatorname{card} \mathcal{R}_{\mathbf{r}(\boldsymbol{\epsilon})} = 1$.

Write $\mathcal{F} = \mathcal{F}(\{\boldsymbol{\sigma}_i\})$ for the family of 2^m constellations of the form $\{\epsilon_i \boldsymbol{\sigma}_i, 1 \le i \le m\}$ with $\epsilon_i \in \mathbb{B}$. Then $\mathcal{F}$ is identically the family consisting of the constellation $\{\boldsymbol{\sigma}_i\}$ and all its reflected variants. Observe that as the joint distribution of the $\boldsymbol{\sigma}_i$ is invariant with respect to reflections, conditioned upon $\mathcal{F}$, each of the constellation variants $\{\epsilon_i \boldsymbol{\sigma}_i\}$ has equal probability.

Definition 2 Given a basic constellation $\{\boldsymbol{\sigma}_i\}$, we say that a reflected constellation $\{\epsilon_i \boldsymbol{\sigma}_i\} \in \mathcal{F}$ is *identifiable* if $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_m)$ is identifiable. We say that $\mathcal{F}\big(\{\boldsymbol{\sigma}_i\}\big)$ is identifiable if each of its 2^m constituent constellations $\{\epsilon_i \boldsymbol{\sigma}_i\}$ is identifiable.

II. An Upper Bound via the Pigeon-Hole Principle

As $\boldsymbol{\epsilon}$ varies across $\mathbb{B}^m$ the values of $\mathbf{r}(\boldsymbol{\epsilon})$ can range across at most 2^m distinct values. On the other hand, there are only $(m+1)^n$ values that $\mathbf{r}(\boldsymbol{\epsilon})$ can take in $\mathbb{L}_m^n$. By the pigeon-hole principle it follows that if $2^m > (m+1)^n$ then there is at least one reflection vector $\boldsymbol{\epsilon}$ that is not identifiable. A necessary condition for $\mathcal{F}$ to be identifiable (equivalently, all $\boldsymbol{\epsilon} \in \mathbb{B}^m$ to be identifiable) is hence that $2^m \le (m+1)^n$.

Let m_n^* be the largest integer m satisfying the inequality $2^m \le (m+1)^n$. As may now be readily verified by taking logarithms and substituting,

$$m_n^* = n \log_2 n + n \log_2 \log_2 n + \mathcal{O}\left(\frac{n \log_2 \log_2 n}{\log_2 n}\right)$$

as $n \to \infty$. A necessary condition for *all* reflection vectors $\boldsymbol{\epsilon} \in \mathbb{B}^m$ to be identifiable is hence that $m \le m_n^*$. But more can be said.

Theorem 1 *Suppose κ is any fixed positive integer. Let $\boldsymbol{\epsilon}^{(0)} \in \mathbb{B}^m$ be any reflection vector. If $m \ge m_n^* + \kappa$ then the probability that $\boldsymbol{\epsilon}^{(0)}$ is identifiable is bounded above by $2^{-\kappa+1}\big(1 + o(1)\big)$.*

Thus, if $m \ge m_n^* + 2$ (at least) half the reflection vectors are not identifiable. The proof is based on a refinement of a probabilistic pigeon-hole argument. I will reserve the details for elsewhere.

The notion of identifiability provides a basic characterisation of the capacity of CDMA systems in a precise combinatorial sense. I shall consider the following sanitised setting.

III. Direct Sequence Spread Spectrum

Suppose there are m users in a direct sequence spread spectrum setting. The kth user is assigned a signature waveform

$$s_k(t) = \sum_{j=1}^{n} \sigma_{jk} p(t - jT_c) \qquad (0 \le t \le T)$$

where $\boldsymbol{\sigma}_k = (\sigma_{1k}, \ldots, \sigma_{nk}) \in \mathbb{B}^n$ is a code sequence (the so-called *spreading sequence*) of n binary-valued chips, $p(t)$ is a pulse of energy $1/n$ with support in an interval $[-T_c, 0]$ where T_c is the chip interval, and $T = nT_c$ is the duration of the signature waveform, i.e., the bit transmission interval. Each signature waveform $s_k(t)$ is antipodally modulated by the bit $\epsilon_k \in \mathbb{B}$ transmitted by the kth user. If the users transmit in synchrony, channel noise is abeyant, and the received amplitudes of the individual waveforms are the same then the received waveform is of the form

$$r(t) = \sum_{k=1}^{m} A\epsilon_k s_k(t) = A \sum_{j=1}^{n} \sum_{k=1}^{m} \epsilon_k \sigma_{jk} p(t - jT_c)$$

for $0 \le t \le T$. This antiseptic setting provides the clearest venue in which to examine user interference shorn of external complexity; perturbants to the system such as asynchrony, the presence of channel noise, both additive and fading, and variations in received amplitudes (the "near-far problem") only serve to worsen the communication problem. Our results hence suggest fundamental limits on direct sequence spread spectrum systems.

Now, it is easy to see that the likelihood function depends on the observations only through the outputs of a bank of matched filters (cf. Verdú [1]),

$$Y_\ell := \int_0^T r(t) s_\ell(t)\, dt \qquad (1 \le \ell \le m).$$

It follows that $\boldsymbol{Y} = (Y_1, \ldots, Y_m)$ is a *sufficient statistic* for demodulating the bit sequence $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_m)$.

Let us rewrite these equations in a more compact vector form. The received signal is the vector

$$\mathbf{r} = \frac{A}{\sqrt{n}} \sum_{k=1}^{m} \epsilon_k \boldsymbol{\sigma}_k$$

from which we form the sufficient statistic $\boldsymbol{Y} = (Y_1, \ldots, Y_m)$ via the inner products

$$Y_\ell = \frac{1}{\sqrt{n}} \langle \mathbf{r}, \boldsymbol{\sigma}_\ell \rangle = \frac{A}{n} \sum_{k=1}^{m} \epsilon_k \langle \boldsymbol{\sigma}_k, \boldsymbol{\sigma}_\ell \rangle = \frac{A}{n} \sum_{j=1}^{n} \sum_{k=1}^{m} \epsilon_k \sigma_{jk} \sigma_{j\ell} \tag{1}$$

for $1 \le \ell \le m$.

From now on we suppose that the spreading sequences are randomly selected by independent sampling from the outcomes of symmetric Bernoulli trials, that is to say, the random variables $\{\sigma_{jk}, 1 \le j \le n, 1 \le k \le m\}$ are independent and have the common distribution

$$\mathbf{P}\{\sigma_{jk} = -1\} = \mathbf{P}\{\sigma_{jk} = +1\} = 1/2.$$

This is, of course, an idealisation of the specification of pseudonoise sequences in practice. We suppose further that the information bits $\epsilon_1, \ldots, \epsilon_m$ are fixed, but arbitrary. We are interested in determining how many users (as a function of the available bandwidth) can be supported in such a scheme. It is clear that the recovery of the bit sequence is equivalent to the identifiability of the reflection vector $\boldsymbol{\epsilon}$ with respect to the constellation $\{\boldsymbol{\sigma}_k, 1 \le k \le m\}$. We are interested in determining how the probability of bit error at the receiver depends on the number of users.

IV. Matched Filter Receiver

In conventional single user detection the receiver forms the estimate $\widehat{\epsilon}_\ell$ of the information bit ϵ_ℓ of the ℓth user according to[1]

$$\widehat{\epsilon}_\ell = \operatorname{sgn} Y_\ell = \operatorname{sgn}\left(\sum_{j=1}^{n} \sum_{k=1}^{m} \epsilon_k \sigma_{jk} \sigma_{j\ell} \right) \qquad (1 \le \ell \le m). \tag{2}$$

The procedure is intuitive and its simplicity has much to recommend it. Its complexity measured in terms of elementary operations is only mn multiplications, the same number of additions, and m comparisons to decode all the transmitted bits. As we see in the following, it's performance is also surprisingly robust.

The bit transmitted by the ℓth user is received in error if $\widehat{\epsilon}_\ell \ne \epsilon_\ell$. Write

$$W = \sum_{\ell=1}^{m} 1\big(\widehat{\epsilon}_\ell \ne \epsilon_\ell\big)$$

for the number of bit errors at the receiver. Perhaps counter-intuitively, our main result shows that, for large n, the number of errors W increases abruptly beyond a very narrow range of values of $m = m_n$.

Theorem 2 *For any fixed, real c, suppose the number of users $m = m_n$ grows with n such that*

$$\frac{m_n}{n/(2\log n)} = 1 + \frac{3\log\log n}{2\log n} + \frac{\log(16\pi)}{2\log n} - \frac{c}{\log n} + O\left(\frac{(\log\log n)^2}{(\log n)^2} \right). \tag{3}$$

Then the number of errors, $W = W_n$, converges in law to the Poisson distribution with mean e^{-c}. A fortiori, for every nonnegative integer m,

$$\mathbf{P}\{W_n = m\} \to \frac{(e^{-c})^m}{m!} e^{-e^{-c}}$$

as $n \to \infty$.

[1] Setting $\widehat{\epsilon}_\ell = 0$ when $Y_\ell = 0$ is merely an analytical convenience—any other convention may be chosen without affecting the asymptotic results as the event $Y_\ell = 0$ has asymptotically vanishing probability.

The proof requires a careful consideration of weak but pervasive statistical dependencies either via classical sieve methods or by means of an appropriately chosen coupling in the Stein-Chen method. I will reserve the details for elsewhere.

The fine order of infinity manifested in the expression (3) is worthy of remark: for any choice of the constant c, the dominant term in the expression (3) shows that $m_n \sim n/(2\log n)$; the nature of the emergent Poisson law is hidden within the second sub-dominant term in the expression.

Thus, asymptotically, the number of receiver errors is Poisson when the rate of growth of the number of users with the number of chips (that is to say, the available bandwidth) is in a critical, sub-linear range. In particular, for the rate of growth given in (3), the probability that there are no receiver errors tends to $e^{-e^{-c}}$. This double exponential law is familiar from other contexts, most notably in connectivity in random graphs. Thus, if c is chosen to be negative and large in absolute value there will be a fixed but large number of errors; if, on the other hand, c is positive and large there will be essentially no errors. We hence observe a phase transition (or threshold function) in performance when the number of users varies around $n/(2\log n)$.

Corollary 1 *Fix any* $\epsilon > 0$.

a) *If* $m_n \le (1-\epsilon)n/(2\log n)$ *for all sufficiently large* n *then* $\mathbf{P}\{W_n = 0\} \to 1$ *as* $n \to \infty$.
b) *If* $m_n \ge (1+\epsilon)n/(2\log n)$ *for all sufficiently large* n *then* $\mathbf{P}\{W_n = 0\} \to 0$ *as* $n \to \infty$.

Roughly speaking, *if* m_n *is less than* $n/(2\log n)$*, the matched filter decisions are guaranteed to be error-free; contrariwise, if* m_n *exceeds* $n/(2\log n)$ *there are guaranteed to be (many) errors*. Thus, no more than roughly $n/(2\log n)$ users can be supported if bit errors cannot be tolerated at the receiver.

What if a significant number of errors can be tolerated? The answer is provided in part below.

Corollary 2 *Fix any* $0 < \gamma < 1/2$*, and suppose that* $m = m_n$ *increases so that*

$$\frac{m_n}{n} \to \left[\Phi^{-1}(\gamma)\right]^{-2} \tag{4}$$

where Φ^{-1} *is the inverse of the Gaussian distribution function* Φ*. Then* $\frac{1}{m_n}\mathsf{E}(W_n) \to \gamma$ *as* $n \to \infty$.

If $m_n \sim n/4$ then about two percent of the transmitted bits are received in error; if $m_n \sim n$ then approximately sixteen percent of the bits are received in error; and if $m_n \sim 4n$ then approximately thirty percent of the bits are received in error. Thus, *even if a constant fraction of bit errors can be tolerated, the number of users cannot increase faster than linearly in* n.

It should be remarked that these results reflect the best case in some sense for matched filter detection of pseudonoise sequences. *The elimination of various perturbants such as noise, amplitude fluctuations, and asynchrony has allowed us to focus on the fundamental rôle played by user interference.*

The basic approach may be extended, at some notational and computational cost, to include some of the other perturbants to the system that we had excluded. Embedding the waveforms in additive, white Gaussian noise does not, in principle, create any new difficulties in analysis though the computational burden starts becoming significant—the additive noise terms at the output of the matched filters are i.i.d. Gaussian, and we can, in principle, run the analysis through by first conditioning on them and finally taking expectations to remove the conditioning. Different waveform amplitudes can likewise be handled if, for instance, the amplitudes are bounded between known limits or their distribution known. Again, one can proceed, in principle, by conditioning though the burgeoning complexity of the expressions will ultimately limit the utility of this approach.

V. Decorrelating Receiver

The matched filter receiver, while beguiling in its simplicity, is not optimal because, as is well known, although $\mathbf{Y}$ is a sufficient statistic for demodulating the bit sequence $\boldsymbol{\epsilon}$, the component Y_i is not in itself a sufficient statistic for demodulating the bit ϵ_i. The decorrelating detector (cf. Verdú [3]) is an alternative formulation which systematically attempts to take correlations into account in the detection process. While its implementation complexity is higher, it has some very attractive properties among which is the fact that it does not require knowledge of the energies of any of the active users.

Let us first introduce some new notation. Form the $n \times m$ *spreading matrix*

$$\boldsymbol{\Sigma} := \begin{bmatrix}\boldsymbol{\sigma}_1^{\mathsf{T}} & \boldsymbol{\sigma}_2^{\mathsf{T}} & \dots & \boldsymbol{\sigma}_m^{\mathsf{T}}\end{bmatrix} = \begin{bmatrix}\sigma_{11} & \sigma_{12} & \dots & \sigma_{1m} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2m} \\ \dots & \dots & \dots & \dots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_{nm}\end{bmatrix}$$

whose components are all ± 1 and whose columns are identically the random spreading sequences of each of the $m = m_n$ users. We can now succinctly rewrite the sufficient statistic $\mathbf{Y}$ whose components are given by (1) in matrix-vector form as

$$\mathbf{Y} = \tfrac{A}{n}\,\boldsymbol{\epsilon}\boldsymbol{\Sigma}^{\mathsf{T}}\boldsymbol{\Sigma}$$

where $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \dots, \epsilon_m)$ is the vector of data bits. The decorrelating detector forms the bit estimates $\mathbf{Y} \overset{\text{decor}}{\longmapsto} \widehat{\boldsymbol{\epsilon}} = \left(\widehat{\epsilon}_1, \widehat{\epsilon}_2, \dots, \widehat{\epsilon}_m\right)$ according to

$$\widehat{\boldsymbol{\epsilon}} = \operatorname{sgn}\left\{\mathbf{Y}\left(\boldsymbol{\Sigma}^{\mathsf{T}}\boldsymbol{\Sigma}\right)^{\dagger}\right\},$$

where $\left(\boldsymbol{\Sigma}^{\mathsf{T}}\boldsymbol{\Sigma}\right)^{\dagger}$ denotes the Moore-Penrose generalised inverse of $\boldsymbol{\Sigma}^{\mathsf{T}}\boldsymbol{\Sigma}$ and the signum operation applied to a vector is to be interpreted as applying component-wise. Observe that

$$\left(\boldsymbol{\Sigma}^{\mathsf{T}}\boldsymbol{\Sigma}\right)^{\dagger} = \left(\boldsymbol{\Sigma}^{\mathsf{T}}\boldsymbol{\Sigma}\right)^{-1}$$

if the spreading sequences are linearly independent, in which case, of course, $m \le n$ is requisite. If this is the case then $\widehat{\boldsymbol{\epsilon}} = \operatorname{sgn}\left(\frac{A}{n}\boldsymbol{\epsilon}\right) = \boldsymbol{\epsilon}$ and we automatically recover all data bits without error.

As before, suppose the number of users $m = m_n$ is explicitly allowed to vary with n and let $W_n^\dagger$ denote the number of bit errors made by the decorrelating detector.

Theorem 3 *If $m_n \le n$ then the decorrelating detector makes no errors with probability approaching one as $n \to \infty$ for any sequence of data bits. Conversely, if $m_n > n$, with probability bounded away from zero, there exist sequences of data bits which cannot all be recovered by the decorrelating detector.*

The key technical instrument needed in the proof is a refinement of a classical result on the singularity of random matrices due to J. Komlós which makes key use of the Littlewood-Offord lemma. I shall again reserve the details.

I have chosen to eschew making the strongest possible statement that can be made here in the interests of presenting the result as cleanly as possible.

Noisy channels are harder to analyse for the decorrelating detector as the generalised inverse causes correlations in the noise variables. See Verdú [3] for details.

VI. Maximum Likelihood Detector

The optimum detector maximises the likelihood function. This is unfortunately an NP-hard problem in general (cf. Verdú [1, Chapter 4]) so that this is not a very practical solution to the demodulation problem unless a special structure is built into the code sequences themselves. From a practical point of view, there may not be much incentive to seek the jointly optimum solution to the demodulation problem because the increase in capacity is modest at best.

Consider now the maximum likelihood solution to the detection problem, $\mathbf{Y} \overset{\text{opt}}{\longmapsto} \widehat{\boldsymbol{\epsilon}}$, and let W_n^* be the number of errors made by the optimal receiver. Then $W_n^* = 0$ if, and only if, the reflection vector $\boldsymbol{\epsilon}$ is identifiable with respect to the constellation $\{\boldsymbol{\sigma}_k, 1 \le k \le m\}$ and so, by a restatement of Theorem 1 we have:

Theorem 4 *Suppose κ is any fixed positive integer, ϵ any positive constant. If $m_n \ge m_n^* + \kappa$ then*

$$\mathbf{P}\{W_n^* = 0\} \le 2^{-\kappa+1}(1+\epsilon)$$

for all sufficiently large n.

Thus, the wheels fall off the bus for even a constant excursion in the number of users beyond m_n^*.

The result is not the best possible but suffices to show that, on the one hand, the phase transitions that are manifested in the number of users who can be supported are not artefacts of the algorithms employed at the receiver but are indeed emblematic of the system as a whole, and, on the other, there is not much gain or profit in trying for optimality as the gain in capacity is relatively modest.

Results of this nature may be of only combinatorial significance as the computational intractability of the decision problem renders suspect the practical utility of the jointly optimum receiver. I will not in consequence attempt to present the result in its strongest form here but leave details for elsewhere.

VII. Conclusion

User interference induces random graph-theoretic effects in direct sequence spread spectrum systems. These effects show up not gradually as one might perhaps have expected but abruptly in the form of threshold phenomena in the number of users that can be accommodated.

Acknowledgments

Aaron Wyner, Sergio Verdú, and Len Cimini at various times (chronologically) urged publication of the results. The deficiencies are all mine.

Acknowledgment

This work was supported in part by the NSF under grant 0915697.

References

[1] S. Verdú, *Multiuser Detection*. Cambridge, UK: Cambridge University Press, 1998.

[2] S. S. Venkatesh, "CDMA capacity," in *Proc. Conf. on Information Sciences and Systems*, Princeton University, March 2000.

[3] S. Verdú, "Optimum multi-user asymptotic efficiency," *IEEE Trans. Commun.*, vol. COM–34, no. 9, pp. 890–897, 1986.

[4] W. Hoeffding, Probability inequalities for sums of bounded random variables, *J. Amer. Stat. Assoc.*, **58**, 13–30, 1963.

[5] W. Feller, *An Introduction to Probability Theory and Its Applications, Volume I*. New York: Wiley, 1968.

[6] W. Richter, "Multidimensional local limit theorems for large deviations," *Theor. Prob. Appl.* (translation), vol. 3, pp. 100–106, 1958.

[7] S. C. Fang and S. S. Venkatesh, "The capacity of majority rule," *Random Structures and Algorithms*, vol. 12, pp. 83–109, 1998.

[8] H. Ruben, "An asymptotic expansion for a class of multivariate normal integrals," *J. Australian Math. Soc.*, vol. 2, pp. 253–264, 1962.

[9] J. Komlós, "On the determinant of $(0,1)$ matrices," *Studia Sci. Math. Hung.*, vol. 2, pp. 7–21, 1967.

[10] J. Kahn, J. Komlós and E. Szemerédi, "On the probability that a random ± 1 matrix is singular," *J. Amer. Math. Soc.*, vol. 8, pp. 223–240, 1995.

[11] R. J. McEliece, E. C. Posner, E. R. Rodemich, and S. S. Venkatesh, "The capacity of the Hopfield associative memory," *IEEE Trans. Inform. Theory*, vol. IT–33, pp. 461–482, 1987.

[12] S. S. Venkatesh, "Connectivity versus capacity in the Hebb rule," in *Theoretical Advances in Neural Computation and Learning* (eds. V. Roychowdhuri, K-Y. Siu, and A. Orlitsky). Norwell, Massachusetts: Kluwer Academic Publishers, 1994.