Optimization of Signal Sets for Partial-Response Channels—Part I: Numerical Techniques

Michael L. Honig, *Member, IEEE*, Kenneth Steiglitz, *Fellow, IEEE*, and Stephen A. Norman

***Abstract*—Given a linear, time-invariant, discrete-time channel, the problem of constructing N input signals of finite length K that maximize minimum l_2 distance between pairs of outputs is considered. Two constraints on the input signals are considered: a power constraint on each of the N inputs (*hard constraint*) and an average power constraint over the entire set of inputs (*soft constraint*). The hard constraint problem is equivalent to packing N points in an ellipsoid in $\min(K, N-1)$ dimensions to maximize the minimum Euclidean distance between pairs of points. Gradient-based numerical algorithms and a constructive technique based on dense lattices are used to find locally optimal solutions to the preceding signal design problems. Numerical results, consisting of minimum distance vs. input length for different information rates, are given for the soft constraint problem. The channels considered are the identity channel, the $1-D$ channel, and the $1-D^2$ channel. Signal constellations found via gradient search are superior to the multidimensional lattice constructions when the number of points per dimension is small (i.e., when the information rate is 1 bit/T or less, $1/T$ being the symbol rate). The average spectra of optimized signal sets is examined. It is shown that transmitted energy is concentrated into frequency bands where the channel attenuation is relatively small. The measure of this frequency band increases with information rate. It is observed that the average spectrum of a signal set is primarily determined by the shape, or boundary, of the signal constellation, assuming the points are uniformly distributed throughout this region. Two numerical examples are shown for which the average spectrum of an optimized signal set resembles the water pouring spectrum that achieves Shannon capacity, assuming additive white Gaussian noise.**

***Index Terms*—Coding, partial-response channels, intersymbol interference, multidimensional signal sets, lattices.**

I. Introduction

GIVEN A LINEAR, time-invariant, dispersive channel with additive white Gaussian noise, and a maximum-likelihood receiver, a classical problem is to encode N messages into energy limited signals over a finite-time interval so as to minimize the probability of making a detection error. Unfortunately, this problem is quite difficult in general, and proving the optimality of any given signal set remains elusive even when the channel is not dispersive. Rather than assume a complete statistical characterization of the channel, however, an alternative approach to signal design for digital communications is to assume that the receiver can distinguish between two channel outputs provided that they are separated by some minimum amount in a suitable metric space. Conversely, given a fixed number of messages, the channel inputs can be designed to maximize the minimum distance between any pair of outputs. Of course, if the noise in the system is known to be Gaussian, then the obvious choice of metric space is L_2 (l_2 for discrete-time channels).

Motivated by the application of high-speed data transmission over the telephone subscriber loop, we assume that it is very difficult, if not impossible, to characterize the statistical properties of the channel. This is due to the difficulty in building statistical models for impairments such as residual intersymbol interference (ISI), crosstalk, timing inaccuracy, VLSI nonlinearities, etc. We therefore model the discrete-time channel as a linear, shift-invariant system with given impulse response $h[k]$, or frequency response $H(f)$, and assume that the receiver can distinguish between any two channel outputs provided that they are separated by some constant $d > 0$ in a suitable metric space, say, l_p in general. One interpretation of this assumption is that the channel and receiver add noise to the received signal, where the noise has *unknown* statistics, but is bounded in l_p norm by $d/2$ almost surely. Assuming that the channel inputs are time limited to $[1, K]$, and that they are constrained in l_q norm, then this leads to the l_q/l_p *signal design problem*: Given $N \geq 2$ and K, find N inputs, bounded in l_q norm, to maximize the minimum l_p distance between channel outputs.

As an example, if it is assumed that the receiver can distinguish two channel outputs separated in amplitude by d at some time k, and that the inputs are constrained in amplitude, then this leads to the l_∞/l_∞ signal design problem, which is considered in [1]. The implicit assumption there is that random disturbances to the received signal are bounded in amplitude. In this paper, we consider the l_2/l_2 problem, namely, we insist that all pairs of outputs be separated by at least d in l_2 norm with an l_2 constraint on the inputs. Strictly speaking, the l_2/l_2 problem assumes that *each* input is bounded in l_2. A more common assumption, however, is that the *average* l_2 norm over the entire signal set is bounded. The former

Manuscript received December 10, 1989; revised April 17, 1991. This work was supported in part by NSF Grant MIP-8912100 and U.S. Army Research Office-Durham Contract DAAL03-89-K-0074. This work was presented in part at the IEEE International Symposium on Information Theory, San Diego, CA, January 14-19, 1990.

M. L. Honig is with Bellcore, Room MRE 2L-343, 445 South Street, Morristown, NJ, 07960-1910.

K. Steiglitz is with the Department of Computer Science, Princeton University, Princeton, NJ 08854.

S. A. Norman is with Information Systems Laboratory, Stanford University, Stanford, CA 94305.

IEEE Log Number 9101604.

0018-9448/91/0900-1327$01.00 ©1991 IEEE

constraint will be referred to as the *hard input constraint* (HIC), and the latter will be referred to as the *soft input constraint* (SIC). As previously mentioned, these problems are closely related to coding for dispersive channels with additive Gaussian noise, although our approach is simply to assume that any additive noise is bounded in l_2 norm by $d/2$ almost surely.

Designing power-limited discrete-time input signals to maximize minimum l_2 distance between pairs of channel outputs in the absence of noise, subject to various receiver constraints, has been the topic of intensive research in recent years. This research includes the work on multidimensional lattice codes [3]–[5], trellis coding, as proposed by Ungerboeck [2], and the combination of these two schemes [4]–[7]. Much of this work assumes a nondispersive (identity) channel; however, more recently there has been significant attention devoted to partial-response (PR) channels. References [8] and [9] discuss coding schemes in which nulls are created in the transmitted spectrum to more closely resemble the channel frequency response. In contrast, the problem of designing codes specifically for PR channels, or channels with ISI, is discussed in [10]–[12]. Block coding combined with decision feedback equalization for dispersive channels is studied in [13]. In all of these cases an important figure of merit is minimum l_2 distance between pairs of channel outputs.

For a given channel impulse response, $h[\cdot]$, if we fix the number of messages N, and restrict the inputs to the time interval $[1, K]$, an "optimal signal set," or "optimal code," is defined here to be a set of inputs $u_1[\cdot], \cdots, u_N[\cdot]$ that maximizes the minimum l_2 distance between pairs of channel outputs. This paper studies the construction of such optimal codes. It is shown in the next section that for fixed number of messages N, time interval $[1, K]$, and assuming the HIC, the l_2/l_2 signal design problem is equivalent to packing N points in an ellipsoid in $\min(K, N-1)$ dimensions to maximize the minimum Euclidean distance between pairs of points. Each point corresponds to a channel output in response to a codeword input, and the collection of points is referred to as the "output signal constellation." The SIC problem has a similar geometric interpretation in which points, corresponding to channel outputs, are to be packed in Euclidean space subject to an average power constraint, which is transformed by the linear channel operator. In both cases the basis vectors for the Euclidean space are taken to be the eigenvectors of the linear channel operator. These geometric interpretations are not entirely new. It has been shown in [14] and [15] that the analogous L_2/L_2 problem for continuous-time channels is equivalent to packing points in an ellipsoid in Hilbert space. The geometric interpretation assuming the SIC can be inferred from discussions in [12].

Section III describes numerical techniques that are used to find approximate solutions to the preceding packing problems assuming either the HIC or SIC. A heuristic construction technique based on cropping dense lattice packings is first described, followed by a description of gradient-based search techniques. Ascent directions in the latter case can be found by solving linear programs, assuming either the HIC or SIC. However, for the SIC problem we use an alternative algorithm that maximizes a continuous penalty function that approximates the minimum distance d. This penalty function is similar in form to the error criterion used in [16], where two-dimensional signal sets are optimized for the identity channel with additive white Gaussian noise.

In Section IV, numerical results, consisting of minimum distance vs. input length for different information rates, are given for the SIC problem. The channels considered are the identity channel, the $1-D$ channel, and the $1-D^2$ channel. When the number of points per dimension becomes large, nearly optimal solutions (i.e., output signal constellations) can be obtained by cropping dense lattice packings by ellipsoids. There are two sources of degrees of freedom in cropping lattice packings: namely, the orientation of the lattice axes and origin with respect to the coordinate axes and origin. A simple search algorithm is used to find good choices for these degrees of freedom, and the resulting codes are compared with those of the gradient search algorithm for the SIC problem. The densest lattice packings are known for dimensions up to eight [3]. When the number of points per dimension is large the cropped lattice codes are typically better than codes found via gradient search for two reasons: first, because the gradient search penalty function differs from the true minimum distance, and second, because the gradient method finds local, but not necessarily global, optima. When the number of points per dimension is small, however, the gradient method generally produces the better solutions.

The constructed codes have the effect of concentrating the transmitted signal energy into frequency bands where the channel has the least attenuation. It is shown in Section 5 that if the signal constellation at the channel output has uniform density, that is, the points are uniformly distributed within the output region determined by the input constraint and are equally likely, then for large information rates the average spectrum of the corresponding input signal set is approximately constant over a subset of the channel bandwidth. This type of transmitted spectrum has appeared in different, but related, contexts. Specifically, Price [17] has shown that if the channel is followed by additive white Gaussian noise, then the transmitted spectrum that maximizes the signal-to-noise ratio at the output of a decision-feedback equalizer is a constant over a subset of the channel bandwidth, assuming no error propagation. This subset is again where the channel has the least attenuation. The average spectra for two optimized constellations are shown in Section V. Similar transmitted spectra have also been observed in systems that use vector coding, as proposed in [12].

Part II of this paper [18] investigates the minimum distance between outputs, or coding gain, of optimal codes as $K \to \infty$.

II. Preliminaries

Let $h[k]$, $k = 0, 1, \cdots, \tau - 1$, be the discrete-time, real-valued impulse response of the channel, so that if $u[k]$ is an input, then the output at time m is

$$y[m] = \sum_{k=0}^{m} u[k]h[m-k]. \tag{2.1}$$

It is assumed throughout this paper that $h[k] = 0$ for $k < 0$ and $k > \tau - 1$, and that $|h[k]| < \infty$, $k = 0, \cdots, \tau - 1$. Given a set of inputs, or "codewords" $u_i[k]$, $i = 1, \cdots, N$, defined for $k = 1, \cdots, K$, then the minimum distance between pairs of outputs is defined as

$$d = \min_{i \neq j} \| y_i[\cdot] - y_j[\cdot] \|, \tag{2.2}$$

where the norm is the l_2 (Euclidean) norm evaluated over the time interval $[1, K + \tau - 1]$, and y_i is the output corresponding to the input u_i.

Two constraints on the input signals are considered: a power constraint on each input (*hard input constraint*, or HIC),

$$\|u_i[\cdot]\|^2 = \sum_{k=1}^{K} u_i^2[k] \leq PK, \qquad \text{for each } i = 1, \cdots, N \tag{2.3a}$$

and an average power constraint (*soft input constraint*, or SIC),

$$\frac{1}{N}\sum_{i=1}^{N} \|u_i[\cdot]\|^2 = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} u_i^2[k] \leq PK, \tag{2.3b}$$

where P is the transmitted power. Defining vectors of channel inputs and outputs as

$$\boldsymbol{u}_i' = [u_i[1], u_i[2], \cdots, u_i[K]] \tag{2.4a}$$

and

$$\boldsymbol{y}_i' = [y_i[1], y_i[2], \cdots, y_i[K + \tau - 1]], \tag{2.4b}$$

respectively, where "$'$" denotes transpose, then (2.1) can be rewritten as

$$\boldsymbol{y}_i = \boldsymbol{H}\boldsymbol{u}_i, \tag{2.5}$$

where $\boldsymbol{H}$ is the appropriate $(K + \tau - 1) \times K$ convolution matrix. The squared minimum distance is therefore

$$d^2 = \min_{i \neq j} \| \boldsymbol{H}(\boldsymbol{u}_i - \boldsymbol{u}_j) \|^2 = \min_{i \neq j} (\boldsymbol{u}_i - \boldsymbol{u}_j)'\boldsymbol{H}'\boldsymbol{H}(\boldsymbol{u}_i - \boldsymbol{u}_j). \tag{2.6}$$

Now $\boldsymbol{H}'\boldsymbol{H}$ is real, symmetric, and Toeplitz, and can therefore be factored as

$$\boldsymbol{H}'\boldsymbol{H} = \Phi\Lambda\Phi', \qquad \Phi' = \Phi^{-1}, \tag{2.7}$$

where Φ is the $K \times K$ orthonormal matrix whose columns are eigenvectors of $\boldsymbol{H}'\boldsymbol{H}$ (also referred to as the "channel eigenvectors"), and

$$\Lambda = \text{diag}[\lambda_1, \lambda_2, \cdots, \lambda_K], \tag{2.8}$$

where λ_i, $i = 1, \cdots, K$ are the real, nonnegative eigenvalues of $\boldsymbol{H}'\boldsymbol{H}$ arranged in nonincreasing order. Note that Φ and the λ_i's are functions of K, the time index. If $\boldsymbol{H}'\boldsymbol{H}$ is singular, so that some of the eigenvalues are zero, then we must project all signal vectors onto the subspace of $\mathbb{R}^K$ spanned by the eigenvectors of $\boldsymbol{H}'\boldsymbol{H}$ corresponding to nonzero eigenvalues. This is simply accomplished by discarding the eigenvectors corresponding to the zero eigenvalues. Consequently, throughout the rest of this paper we assume that $\boldsymbol{H}'\boldsymbol{H}$ is nonsingular with strictly positive eigenvalues. Letting $\tilde{\boldsymbol{u}}_i = \Phi'\boldsymbol{u}_i$, then

$$\|\tilde{\boldsymbol{u}}_i\|^2 = \boldsymbol{u}_i'\Phi\Phi'\boldsymbol{u}_i = \|\boldsymbol{u}_i\|^2, \tag{2.9}$$

since Φ is orthonormal. The signal set $\{\boldsymbol{u}_i\}$ therefore satisfies the HIC (SIC) constraint, if and only if the transformed signal set $\{\tilde{\boldsymbol{u}}_i\}$ satisfies the HIC (SIC) constraint.

Assuming the HIC, fixing the number of inputs N, and the time interval $[1, K]$, the l_2/l_2 signal design problem can be written as

$$\max_{\tilde{\boldsymbol{u}}_1, \cdots, \tilde{\boldsymbol{u}}_N} \left\{ d = \min_{i \neq j} \| \Lambda^{1/2}(\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{u}}_j) \| \right\}$$

$$\text{subject to } \|\tilde{\boldsymbol{u}}_i\|^2 \leq PK, \qquad i = 1, \cdots, N. \tag{P1}$$

The inputs $\tilde{\boldsymbol{u}}_i$ are therefore points in $\mathbb{R}^K$ that lie within the sphere with radius $\sqrt{PK}$, and the objective is to maximize the minimum distance between pairs of points with respect to the Euclidean metric "warped" by Λ. Letting $\tilde{\boldsymbol{y}}_i = \Lambda^{1/2}\tilde{\boldsymbol{u}}_i$, then we can rewrite the l_2/l_2 signal design problem as

$$\max_{\tilde{\boldsymbol{y}}_1, \cdots, \tilde{\boldsymbol{y}}_N} \min_{i \neq j} \|\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{y}}_j\|^2$$

$$\text{subject to } \|\Lambda^{-1/2}\tilde{\boldsymbol{y}}_i\|^2 \leq PK, \qquad i = 1, \cdots, N. \tag{P2}$$

The points $\tilde{\boldsymbol{y}}_i$ lie inside an ellipsoid in $\mathbb{R}^K$, and the objective is to maximize the minimum Euclidean distance between pairs of points. The axes of the ellipsoid are oriented along the eigenvectors of $\boldsymbol{H}'\boldsymbol{H}$. Any set of vectors $\tilde{\boldsymbol{y}}_1, \cdots, \tilde{\boldsymbol{y}}_N$ will be referred to as an "output signal constellation," and a set of vectors $\tilde{\boldsymbol{u}}_1, \cdots, \tilde{\boldsymbol{u}}_N$ will be referred to as an "input signal constellation."

Given an output signal constellation, the inputs to the channel are given by $\boldsymbol{u}_i = \Phi\Lambda^{-1/2}\tilde{\boldsymbol{y}}_i$. The constellation point $\tilde{\boldsymbol{y}}_i$ can then be recovered from the channel output $\boldsymbol{y}_i$ by

$$\tilde{\boldsymbol{y}}_i = \Lambda^{-1/2}\Phi'\boldsymbol{H}'\boldsymbol{y}_i, \tag{2.10}$$

as shown in Fig. 1. The type of signal representation described here, in which the basis vectors used to define the channel inputs are the eigenvectors of $\boldsymbol{H}'\boldsymbol{H}$, is a discrete-time version of one first used by Holsinger [19] and Gallager [20] for continuous-time channels. More recently, this signal representation has been proposed in [11], [12], and [21]. The preceding discussion shows that the l_2/l_2 signal design problem leads directly to this representation. Because the eigenvectors of $\boldsymbol{H}'\boldsymbol{H}$ become sinusoidal for large values of K (see [19] and the discus-

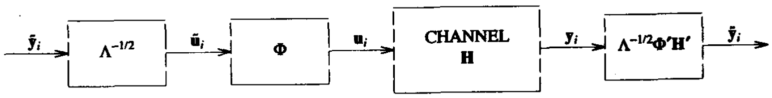


Fig. 1. Transmission of codeword.

sion in Section V), the signaling scheme shown in Fig. 1 is also closely related to multicarrier modulation [19], [22].

Consider now an output constellation $\tilde{y}_1, \cdots, \tilde{y}_n$ for which $N \leq K$, that is, the number of points is less than or equal to the number of dimensions. For discrete-time channels this assumption is rarely true for information rates of interest; however, for continuous-time channels the analogous signal points lie on the surface of an ellipsoid in Hilbert space [14], so that the number of dimensions is infinite. However, the output signal constellation spans an $N-1$ dimensional hyperplane, and the intersection of this hyperplane with the original ellipsoid is an ellipsoid in $\mathbb{R}^{N-1}$. Let E' denote this ellipsoid in $\mathbb{R}^{N-1}$, with axes having lengths $\mu_1^{1/2}, \cdots, \mu_{N-1}^{1/2}$, listed in nonincreasing order, and let E denote the ellipsoid in $\mathbb{R}^N$ with axes having lengths $\lambda_1^{1/2}, \cdots, \lambda_N^{1/2}$. (For the time being, we assume that $PK = 1$.)

Theorem:

$$\mu_i^{1/2} \leq \lambda_i^{1/2}, \; i = 1, \cdots, N-1,$$

which implies that if E' is rotated so that its axes corresponding to $\mu_1, \cdots, \mu_{N-1}$ are coincident with the axes of E corresponding to $\lambda_1, \cdots, \lambda_{N-1}$, respectively, then the resulting ellipsoid E'' satisfies $E'' \subset E$.

The proof is given in Appendix A. This theorem implies that any N-point constellation in the original ellipsoid in $\mathbb{R}^K$ can be rotated and translated to fit inside E, so that there always exists an optimal constellation that lies in E. If $N \leq K$, then the l_2/l_2 signal design problem is equivalent to packing N points in E to maximize the minimum Euclidean distance between pairs of points. Of course, $N = 2^{RK}$ for fixed information rate R, so that the number of dimensions spanned by an optimal signal set grows as $O(\log N)$. We shall see in Section IV that for most PR channels, and for fixed R, the number of dimensions spanned by an optimal signal set grows as βK, where $0 < \beta < 1$. This is also true of the continuous-time L_2/L_2 problem considered in [14]. That is, although the N points can span at most $N-1$ dimensions in this case, the number of dimensions spanned by an optimal signal set grows as $O(\beta \log_2 N)$.

If the SIC (2.3b) is assumed then the constraint in (P2) becomes

$$\frac{1}{N} \sum_{i=1}^{N} \|\Lambda^{-1/2}\tilde{y}_i\|^2 \leq PK, \tag{2.11}$$

so that some points in the output constellation may lie outside of the ellipsoid with axes having lengths $(PK\lambda_1)^{1/2}, \cdots, (PK\lambda_{N-1})^{1/2}$. Nevertheless, the results in Section IV indicate that when the density of points is relatively large (i.e., at least 2 bits/symbol), the optimal output signal constellation approaches a uniform dense lattice cropped by an ellipsoid. The corresponding rotated inputs $\tilde{u}_1, \cdots, \tilde{u}_n$ lie within a sphere in $\mathbb{R}^N$, but the density of points along the ith dimension is $(\lambda_i/\lambda_1)^{1/2}$ times the density of points along the first dimension (that is, the dimension spanned by the eigenvector corresponding to the largest eigenvalue). The fact that optimal signal sets lie inside a sphere has been observed for the identity channel [4], and is due to the fact that the sphere has the least average energy per unit volume of any region. The situation for PR channels is the same because for an optimal n-dimensional input signal constellation, the density of points in $\mathbb{R}^n$ can be assumed to be constant, even though the densities with respect to each axis are generally different.

III. Searching for Optimal Signal Sets

A. Construction of Signal Sets From Lattices

Consider first problem (P2) when the information rate, or number of points per dimension, is large ($R \geq 2$ bits/T where $1/T$ is the symbol rate). Since we will later show that the optimal signal set consisting of $N = 2^{RK}$ points may not span all K dimensions, we will denote the number of dimensions as n, where $n \leq K$. It is plausible that the optimal packing of a finite number of points must approach in some sense an optimally dense packing in $\mathbb{R}^n$. This suggests the idea of constructing a finite-length signal set by selecting a piece of a dense infinite lattice near the origin. In this section we describe an algorithm for doing this, assuming the SIC.

The densest lattice packings of infinite extent are known for dimensions up to dimension $n = 8$ [3]. For $n = 2$, the result is the familiar hexagonal lattice, denoted as A_2, and for $3 < n < 5$ the densest lattices are the *checkerboard lattice* defined by

$$D_n = \{(x_1, \cdots, x_n) \in Z^n | x_1 + \cdots + x_n \text{ is even}\}, \tag{3.1a}$$

where Z^n is the fundamental lattice of integers. We will be searching for good finite croppings, and for this purpose we may also want to allow a displacement of this lattice so that a point appears at the origin. We therefore introduce a parameter $\pi \in \{0, 1\}$ in the definition of D_n as follows:

$$D_n = \{(x_1, \cdots, x_n) \in Z^n | x_1 + \cdots + x_n + \pi \text{ is even}\}. \tag{3.1b}$$

We call π the *parity* of the lattice, and speak of the *even* (*odd*) checkerboard lattice when $\pi = 0(1)$. The only difference between A_2 and D_2 is a scale factor of $\sqrt{3}$ in one dimension.

The construction proceeds as follows. We first select a finite section of D_n (A_2 if $n = 2$), called the *field*, to represent points $\tilde{y}$ in the output space. These points are densely packed with respect to a minimum distance criterion, as desired in (P2). We then map these points back to points $\tilde{u}$ in the input space, where we are interested in minimizing average power. We sort those points with respect to distance from the origin, and take the first n of them for an N-point signal set. The algorithm is thus a "greedy" one. We can view this as minimizing input power for fixed minimum distance d.

The results of this strategy depend on the parity, the exact size of the field, and the orientation of the axes with respect to the fundamental lattice. In the final algorithm we therefore search over these parameters, to the extent allowed by our time budget. For a small number of dimensions n we can afford to do a much more thorough search than for larger n.

The final algorithm contains some heuristic features arrived at through trial and error. The parity, the size of the field k_0, and the angles θ_j, $j = 1, \cdots, n-1$, are parameters searched over in an outside loop.

a) Generate the field $F_{\tilde{y}}$ of points $\tilde{y}$ where each coordinate satisfies $-k_0 \le \tilde{y}[k] \le k_0$.
b) Move $F_{\tilde{y}}$ to its centroid.
c) Rotate $F_{\tilde{y}}$ by θ_1 in the $1-2$ plane, by θ_2 in the $2-3$ plane, and so on, up to θ_{n-1} in the $(n-1)-n$ plane.
d) Map the result back to a field $F_{\tilde{u}}$ in the input constellation space by the transformation $\tilde{u} = \Lambda^{-1/2}\tilde{y}$.
e) Select the N points closed to the origin.
f) Move the resulting input signal constellation to its centroid and normalize to unit average power.

Numerical results will be presented in Section IV, and compared with the results of a gradient search algorithm, described next.

B. Construction of Signal Sets by Gradient Search

The problem of finding an optimal signal set is a nonlinear optimization problem that we can attempt to solve numerically. We will view the problem in the space of possible input signal constellations, formulated as (P1) in the last section, but assuming the SIC. With N points in n dimensions, the unknown vector, denoted as U, is $N \cdot n$-dimensional, and the average power constraint means that allowable solutions must lie on the unit sphere in $N \cdot n$-space.

The major difficulty with this approach is caused by the fact that the cost function is of the form max-min, and is therefore not differentiable. A standard method which allows us to apply gradient search is to replace the true cost criterion

$$d^2 = \min_{i \ne j} \left\| \Lambda^{1/2}(\tilde{u}_i - \tilde{u}_j) \right\|^2$$

by another function, called the *potential function*, which is smooth, but which can be made to approximate the true cost function. This technique has been used in other applications, such as digital filter design [23]. In this case we can use the function

$$f = -\frac{1}{W} \ln \left(\sum_{i \ne k} e^{-W \| \Lambda^{1/2}(\tilde{u}_i - \tilde{u}_k) \|^2} \right), \tag{3.2}$$

where W is a *weight*, large enough so that f is an accurate approximation to d^2, but small enough so that f is sufficiently well behaved to allow a gradient search algorithm to converge. Note that $f \to d^2$ as $W \to \infty$. Good values of W must be found by trial and error; typically $W = 20$ was found to be effective in the smaller problems considered in this paper (i.e., $N \le 32$). If the same value of W is used in larger problems, however, the potential function becomes a less accurate approximation to d. This is because f depends on the number of nearest-neighbors to each point. As N and n increase, the number of nearest-neighbors increases, and the function f deviates more from d. Consequently, larger values of W, i.e., $W > 50$, were typically used in larger problems. This potential function is nearly identical to the probability of error criterion in [16], which was used to optimize two-dimensional signal sets in the case of a nondispersive channel with additive white Gaussian noise. The primary difference between f in (3.2) and the error criterion in [16] is the presence of the matrix Λ in (3.2). This matrix would also appear in the error criterion used in [16] if the noise were correlated with covariance matrix $(\boldsymbol{H}'\boldsymbol{H})^{-1}$.

One optimization algorithm, which was implemented, picked a random starting point U on the unit sphere, and then minimized the function

$$g(\alpha) = f\left(\frac{U + \alpha \nabla f}{\|U + \alpha \nabla f\|} \right) \tag{3.3}$$

with respect to the positive step-size α. This is the cost f of the point obtained by moving in the gradient direction and projecting on to the unit sphere. Such iterative one-dimensional optimization is encountered frequently in nonlinear optimization algorithms, and it is important to balance its precision: it should be done accurately enough to increase the cost effectively, but not so accurately as to waste time working on a cost that is only approximate. As in the construction algorithm described in the previous section, some trial and error was necessary to arrive at a useful heuristic. A one-dimensional search for α was used that shrunk the step-size until an improvement was found, and then expanded it (more slowly than it was shrunk) as much as possible.

Another, simpler, gradient search strategy is to take one step at each iteration, and then recompute the gradient. The step-size, α, is adjusted adaptively at each iteration, based on how much the potential function increases (or decreases). This approach is faster than optimizing over α at each iteration; however, convergence to a stationary point, as will be described shortly, is sometimes harder to achieve with this method.

Finally, another strategy for gradient search uses a linear programming subproblem to find an optimal ascent direction. This was implemented also, and worked well for small problems. The memory requirements for large problems were excessive, so this algorithm was used only for checking results on smaller problems. The idea is to move the vector $\boldsymbol{U}$ along a direction in the plane tangent to the unit sphere, so that the smallest rate of change of distance from each constellation point to its nearest neighbors is as large as possible. This rate of change is, for a constellation point $\tilde{\boldsymbol{u}}_k$, a nearest neighbor $\tilde{\boldsymbol{u}}_i$ of $\tilde{\boldsymbol{u}}_k$, and a direction vector $\boldsymbol{g}_k$ in the tangent plane,

$$\left.\frac{\partial\|\tilde{\boldsymbol{u}}_k+\alpha\boldsymbol{g}_k-\tilde{\boldsymbol{u}}_i\|^2}{\partial\alpha}\right|_{\alpha=0}=2(\tilde{\boldsymbol{u}}_k-\tilde{\boldsymbol{u}}_i)'\boldsymbol{g}_k. \tag{3.4}$$

Since $\boldsymbol{g}_k$ is in the tangent plane, $\boldsymbol{g}_k'\tilde{\boldsymbol{u}}_k=0$, so the conditions just described lead to the following linear program:

$$\max C \tag{LP}$$

subject to

$$\boldsymbol{g}_k'\tilde{\boldsymbol{u}}_k=0,$$
$$-\tilde{\boldsymbol{u}}_i'\boldsymbol{g}_k\ge C,\qquad i\in J_k,$$
$$|[\boldsymbol{g}_k]_m|\le 1,\qquad m=1,\ldots,K,$$

where $k=1,\cdots,K$, and J_k is the set of indices associated with the nearest neighbors of $\tilde{\boldsymbol{u}}_k$. The variables in (LP) are C and $\boldsymbol{g}_k$. The last constraint on the components of $\boldsymbol{g}_k$ is added to bound C. Without this constraint any $\boldsymbol{g}_k$ and C that satisfy the remaining constraints can be scaled by any positive constant. This is to be expected, since it is only the direction of $\boldsymbol{g}_k$ that is desired. The norm of $\boldsymbol{g}_k$ is irrelevant.

The convergence criterion for all gradient search algorithms described was to stop when $\sin^2\theta$ was sufficiently small, where θ is the angle between ∇f and $\boldsymbol{U}$. The angle θ is zero when $\boldsymbol{U}$ and ∇f are colinear; that is, when the gradient is normal to the spherical constraint surface. Explicitly, θ is determined by

$$\sin^2\theta=1-\cos^2\theta=1-(|\boldsymbol{U}'\nabla f|/(\|\boldsymbol{U}\|\cdot\|\nabla f\|))^2. \tag{3.5}$$

An alternative characterization of a good signal set is obtained by defining the Lagrangian

$$L=f+\mu'\left(\sum_{i=1}^{N}\|\tilde{\boldsymbol{u}}_i\|^2-1\right), \tag{3.6}$$

where μ' is the Lagrange multiplier, and setting $\nabla L=0$. Subsequently taking the limit as $W\to\infty$ gives the following condition,

$$\sum_{k\in J_i}(\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{u}}_k)=\mu\Lambda^{-1}\tilde{\boldsymbol{u}}_i,\qquad i=1,\cdots,N, \tag{3.7}$$

where μ is a constant, and J_i is the set of indexes corresponding to nearest-neighbors of $\tilde{\boldsymbol{u}}_i$. This condition has also been observed in [16] for the case $\Lambda=\boldsymbol{I}$, the identity matrix. The geometric interpretation is that for fixed $\tilde{\boldsymbol{u}}_i$, the sum of the vectors from nearest-neighbor points to $\tilde{\boldsymbol{u}}_i$ is colinear with $\Lambda^{-1}\tilde{\boldsymbol{u}}_i$. We point out that the condition (3.7) is *not* a necessary condition for local optimality of a signal set, in the sense of problems (P1) or (P2). This is because d is not a differentiable function of the $\tilde{\boldsymbol{u}}_i$'s. One can easily construct an example for which (3.7) is not satisfied, yet the points cannot be moved to increase d locally. (Take $\Lambda=\boldsymbol{I}$ and initially place six uniformly spaced points on a circle. Construct the constellation by moving one of the points to the center of the circle.) Nevertheless, because the potential function is a continuous approximation to d, the solutions found by the preceding gradient algorithms always satisfy (3.7) approximately.

In the preceding variations of gradient search, we can move all of the constellation points simultaneously, or one at a time. Generally, faster convergence was achieved by moving all of the points simultaneously. Although these methods can also be used to find signal sets assuming the HIC as well as the SIC, we will only present numerical results assuming the SIC.

Gradient search algorithms lead to different local optima, depending on the initial condition. We, therefore, tried several random starting points, typically 10 or 20 for relatively small N and n. The best solution from among those was then refined further by a random search algorithm using the max–min distance cost criterion, rather than the potential function. The random search strategy was similar to simulated annealing: A large step size was used initially to find improvements, and after a given number of consecutive failures, the step size was shrunk, and this process continued until the step size fell below a prescribed value. The motivation for this procedure is that an approximate cost function can be used to find approximate locations of many local optima, and then more computer time can be invested in refining the best of those.

As an example, consider packing 64 points in two dimensions for the $1-D$ channel. Ten different random initial conditions led to ten different local optima, with minimum distances ranging from 0.420 to 0.430, assuming the average energy $PK=1$. Random search applied to the solution with the latter value led to one with a minimum distance of 0.435.

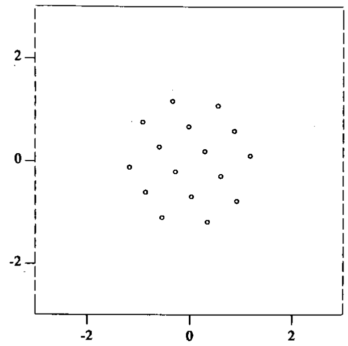
In the next section we present the results obtained when the algorithms in this and the preceding section were used to design sets. In cases where several variations of the gradient algorithms were used, we give the best result obtained.

IV. Search Results

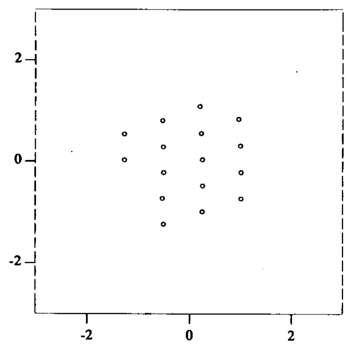
For a given constellation the figure of merit used here is coding gain, which is defined as

$$CG=10\log_{10}\left(\frac{d^2/P}{d_{ss}^2/P_{ss}}\right) \tag{4.1}$$

in dB where d_{ss} and P_{ss} are the minimum distance and average power assuming single-step detection. Single-step detection refers to one-dimensional, multilevel signaling in which the receiver makes a decision on a (scalar) transmitted symbol based on a single channel output. In



(a)



(b)

Fig. 2. (a) Best constructed input signal constellation, 16 points in two dimensions, $1-D$ channel, minimum distance = 0.903 (3.10 dB gain). (b) Locally optimal constellation from gradient search, 16 points in two dimensions, $1-D$ channel, minimum distance = 0.877 (2.84 dB gain). Another local optimum essentially coincides with the solution shown in (a).

this case the normalized distance between outputs is

$$\frac{d_{ss}^2}{P_{ss}} = \frac{12\alpha^2}{4^R - 1}, \tag{4.2}$$

where α is the ratio of the minimum distance between any two possible outputs at a particular time to the spacing between input levels. For the channels considered, $\alpha = 1$.

The optimization problems considered in this paper are similar to many difficult combinatorial optimization problems in that they typically have many different locally optimal solutions with associated cost very close to the global optimum. Fig. 2 illustrates this point for the case $n = 2$, $N = 16$, and the $1-D$ channel. Fig. 2(a) shows the result of the construction algorithm (verified to be locally optimal), with $d = 0.903$, and Fig. 2(b) shows a local optimum obtained from the gradient algorithm, with $d = 0.877$. Throughout this section, the numerical results for minimum distance assume that the average energy $PK = 1$.

Note that both the construction and the gradient methods produced signal sets with a lattice tilted with respect to the axes. The gradient method did so "automatically," while the tilt in the lattice construction was the result of searching over a grid of possible angles. The search over angles is more important for cases with a small number of points per dimension, where conformation to a spherical envelope in the input space is more difficult, and the discrete clipping of the lattice is more significant. To illustrate the possible improvement, consider the case of optimizing 16 points in three dimensions for the $1-D$ channel (this corresponds to an information rate of 1 bit/T with "padding", which will be explained shortly). Without allowing rotations of the lattice, the construction procedure yielded $d = 1.155$ (1.25 dB gain relative to single-step detection). The corresponding result searching over 21 possibilities for two angles was $d = 1.212$ (1.67 dB gain). Note that this is still not as good as the result of gradient search, which yielded $d = 1.235$ (1.83 dB gain). In such examples with few points per dimension, the construction method is very likely to be outperformed by the gradient method, and when there are many points per dimension angle search offers little advantage (and is very expensive). Thus, while angle search may improve some of our gains slightly in intermediate cases, it is not critical.

Tables I–III show minimum distance, assuming $PK = 1$, and coding gains, relative to single-step detection, for a range of cases for three different channels: the identity, $1-D$, and $1-D^2$ channels. Two sets of results are shown, one assuming the inputs are "padded" with $\tau - 1$ zeros, where τ is the length of the impulse response, and one assuming that no additional zeros are appended to the input vectors. Zero padding is used to eliminate intersymbol interference (ISI) between successive output vectors. That is, if the input vectors $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_N$ are of length K, then the corresponding outputs, $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_N$, are of length $K + \tau - 1$ where τ is the length of the impulse response. When this input signal set is used as a block code, then ISI occurs between successive output vectors when $\tau > 1$. This ISI is eliminated when each input vector is padded with $\tau - 1$ zeros. Assuming this is the case, then the receiver can detect a sequence of transmitted K-vectors by examining successive channel output blocks of length $K + \tau - 1$, and in each case selecting the corresponding input that produces an output closest to the received vector in l_2. If the input vectors are transmitted without padding, then a decision feedback equalizer (DFE) can be used at the receiver to remove the ISI between successive output blocks, as proposed in [11], [13], and [24]. Combining (4.1) and (4.2), the entries for coding gain are therefore given by

$$CG = \begin{cases} 10\log_{10}\left[d^2 K \cdot (4^R - 1)/12\right] & \text{without padding,} \\ 10\log_{10}\left[d^2 (K + \tau - 1) \cdot (4^R - 1)/12\right] & \text{with padding,} \end{cases} \tag{4.3}$$

where d is the corresponding entry for minimum distance in the tables.

Consider a block coding scheme with padding at the information rate R. If the inputs are of length K, then each input codeword uses $K + \tau - 1$ time slots, so that the

TABLE I
CODING GAIN (CG), IDENTITY CHANNEL

1 bit/T			1.5 bits/T			2 bits/T		
N/K	d	CG	N/K	d	CG	N/K	d	CG
*4/2	1.414	0.00	*8/2	0.963	0.34	*16/2	0.676	0.58
8/3	1.228	0.53	*23/3	0.821	0.72	*64/3	0.585	1.08
16/4	1.135	1.10	*64/4	0.766	1.36	*256/4	0.544	1.71
32/5	1.075	1.60	*181/5	0.719	1.78	#1024/5	0.507	2.06
†64/6	0.984	1.62						
†128/7	0.944	1.93						
†256/8	0.900	2.10						

Minimum distance d, assuming $PK = 1$, and corresponding coding gain in dB for the identity channel. The "*" indicates cases where the construction yielded results at least as good as the gradient search, "#" indicates that only a construction result is available, and "†" indicates that only a gradient result is available.

TABLE II
CODING GAIN (CG), $1 - D$ CHANNEL

Without Padding

1 bit/T			1.5 bits/T			2 bits/T		
N/K	d	CG	N/K	d	CG	N/K	d	CG
*4/2	2.000	3.01	*8/2	1.281	2.82	*16/2	0.903	3.10
8/3	1.647	3.08	23/3	1.073	3.04	*64/3	0.743	3.16
16/4	1.480	3.40	*64/4	0.952	3.25	*256/4	0.664	3.43
*32/5	1.389	3.82	*181/5	0.871	3.45	#1024/5	0.597	3.48
†64/6	1.292	3.98						
†128/7	1.227	4.21						
†256/8	1.171	4.38						

With Padding

1 bit/T			1.5 bits/T			2 bits/T		
N/K	d	CG	N/K	d	CG	N/k	d	CG
*8/2	1.281	0.90	*23/2	0.746	−0.11	*64/2	0.444	−1.31
16/3	1.235	1.83	*64/3	0.743	1.10	*256/3	0.465	0.34
32/4	1.179	2.40	*181/4	0.727	1.88	#1024/4	0.470	1.40
64/5	1.120	2.74	#512/5	0.702	2.36			
†128/6	1.092	3.19						
†256/7	1.064	3.55						

Minimum distance d, assuming $PK = 1$, and corresponding coding gain in dB for the 1-D channel. The "*" indicates cases where the construction yielded results at least as good as the gradient search, "#" indicates that only a construction result is available, and "†" indicates that only a gradient result is available.

TABLE III
CODING GAIN (CG), $1 - D^2$ CHANNEL

Without Padding

1 bit/T			1.5 bits/T			2 bits/T		
N/K	d	CG	N/K	d	CG	N/K	d	CG
*4/2	2.000	3.01	*8/2	1.362	3.35	*16/2	0.956	3.59
8/3	1.712	3.42	23/3	1.130	3.49	*64/3	0.794	3.73
16/4	1.516	3.61	*64/4	1.033	3.96	*256/4	0.716	4.09
32/5	1.376	3.74	*181/5	0.931	4.03	#1024/5	0.650	4.22
†64/6	1.291	3.98						
†128/7	1.224	4.19						
†256/8	1.156	4.27						

With Padding

1 bit/T			1.5 bits/T			2 bits/T		
N/K	d	CG	N/K	d	CG	N/K	d	CG
*16/2	0.956	−0.39	*64/2	0.476	−2.76	*256/2	0.238	−5.47
*32/3	1.013	1.08	*181/3	0.559	−0.40	#1024/4	0.312	−2.16
*64/4	1.033	2.04	#512/4	0.602	1.03	#4096/4	0.357	−0.19
*128/5	1.000	2.43						
†256/6	0.966	2.71						

Minimum distance d, assuming $PK = 1$, and corresponding gain in dB for the 1-D^2 channel. The "*" indicates cases where the construction yielded results at least as good as the gradient search, "#" indicates that only a construction result is available, and "†" indicates that only a gradient result is available.

number of points in the K-dimensional signal constellation is $N = 2^{R(K+\tau-1)}$. Since the number of points per dimension is larger with padding than without (by the factor $2^{R(\tau-1)}$), the coding gain of a K-dimensional signal set with padding is less than the corresponding coding gain without padding. This can be observed from the tables.

Each entry in the Tables is the best result available. Unmarked entries correspond to cases where the gradient method is better than the construction method; entries marked "*" correspond to the reverse. Entries marked "#" correspond to cases that were too large to run for the gradient method, so only the construction result is available. The construction method was run only for cases where the dimension $n \le 5$, so that entries marked "†" correspond to cases where only the gradient result is available. Because of the computer time involved, the random search procedure mentioned in the last section was used only for the cases $N \le 64$. For the same reason the entries obtained via the gradient algorithm for $N > 64$ are the best of only one or two runs. Generally, the gradient method out-performed the construction procedure only when the number of points per dimension was small. This is intuitively reasonable because it is in these cases that we would expect the optimal packing to deviate most from a lattice. We also note that more coding gain is obtained for the PR channels than for the identity channel, as will be discussed in [18].

Each coding gain shown for the identity channel in Table I at the rate of 2 bits/T is 0.2 to 0.3 dB less than the corresponding coding gain shown in [4], which results from adding the coding gain of the densest known lattice in that dimension, relative to the integer lattice, to the corresponding shaping gain obtained by selecting the constellation points from within a sphere rather than a cube. The estimates in [4] also assume a continuous distribution of points within the boundary region, which is the limiting situation as $R \to \infty$. For dimensions $n = 2$, 3, 4 and 5, the estimated coding gains shown in [4] in dB are 0.82, 1.35, 1.96, and 2.35, respectively.

A. Discussion

It is often the case that an optimized signal set does not use all dimensions available. That is, $\tilde{u}_i[k] = 0$ for each $1 \le i \le N$, and $k > n$, where $n < K$. For example, the 256/8 signal sets for the identity and $1 - D$ channels represented in Tables I and II use seven of eight dimensions, and the 256/8, $1 - D^2$ signal set uses six of eight dimensions. To see why this is the case, consider packing N points inside a K-dimensional ellipsoid with axes $\sqrt{PK}\lambda_k^{1/2}$, $k = 1, \cdots, K$, and suppose that the optimal packing results in a minimum distance d. Then we can surround each point by disjoint spheres of radius $d/2$. The volume of the set of disjoint spheres is approximately the volume of the ellipsoid, i.e.,

$$N\gamma_n\left(\frac{d}{2}\right)^n \approx \gamma_n \prod_{i=1}^{n}\left(\sqrt{PK}\lambda_i^{1/2}\right),$$

or

$$N \approx \prod_{i=1}^{n}\left(\frac{2\sqrt{PK}\lambda_i^{1/2}}{d}\right), \tag{4.4}$$

where $\gamma_n r^n$ is the volume of the n-dimensional sphere with radius r. Since $N = 2^{RK}$,

$$R \approx \frac{1}{K}\sum_{i=1}^{n}\log_2\left(\frac{2\sqrt{PK}\lambda_i^{1/2}}{d}\right). \tag{4.5}$$

For fixed d, (4.4) indicates that to maximize N, n should be the largest integer such that $\sqrt{PK}\lambda_n^{1/2} \ge d/2$. Now for a fixed rate R, as $K \to \infty$, d^2 for optimal signal sets increases linearly with K (see [14] and Section 4 of [18]). Consequently, as $K \to \infty$, the number of dimensions, n, is given approximately by the largest integer such that

$$P\lambda_n \ge \frac{d^2}{4K} \to \bar{d}^2, \tag{4.6}$$

where $\bar{d}$ is a constant that depends only on the channel and R.

According to the Szegö Theorem [25], as $K \to \infty$, the distribution of eigenvalues $\lambda_k(K)$, $k = 1, \cdots, K$, converges to the channel spectrum $|H(f)|^2$ (see (5.6)), where

$$H(f) = \sum_{k=0}^{\tau-1} h[k]e^{-j2\pi fk}. \tag{4.7}$$

This implies that as $K \to \infty$, the fraction of eigenvalues that satisfy (4.6), i.e., n/K, converges to a constant β, where $0 \le \beta \le 1$. Furthermore, if $|H(f)|^2 < \bar{d}^2/P$ for f in a set of positive measure, then $\beta < 1$, and n behaves as $\beta K < K$ for large K. (The preceding argument can be made rigorous by using upper and lower bounds on $\bar{d}$ given in Section 4 of [18]; however, the details are straightforward and are omitted.) Since the normalized distance, $\bar{d}$, decreases monotonically as R increases, β increases with R, and is approximately the fraction of channel bandwidth used by optimal signal sets as $K \to \infty$. This is investigated further in the next section.[1]

The numerical techniques discussed in Section III are concerned only with finding solutions to the signal design problem (P1), and ignore decoding complexity. Nevertheless, because output signal constellations obtained from the construction technique presented in Section III-A are taken from regular lattices, decoding these signal sets can proceed as in the case of the identity channel [3] after post-processing the channel outputs by $\Lambda^{-1/2}\Phi' \boldsymbol{H}'$, as shown in Fig. 1. It may be worthwhile, however, to modify these constellations so as to enforce symmetries that reduce decoding complexity. The signal sets found by gradient search are not necessarily selected from a dense lattice; however, it is likely that these constellations ex-

[1]A similar type of argument to the one just given is used in [12] to show that the optimal transmitted spectrum in vector coded systems generally occupies a proper subset of the available channel bandwidth. The subset of channel bandwidth used in vector coded systems will, however, be somewhat different than the subset of channel bandwidth used by optimal signal sets, as defined in the next section.

hibit sufficient structure to enable efficient decoding procedures.

As pointed out in [12], a trellis code can be defined with respect to an output signal constellation $\tilde{y}_1, \cdots, \tilde{y}_N$. The coding gain of the trellis code is then added to the coding gain of the signal constellation. A 16-state trellis code is proposed in [12] for the $1-D$ channel at a rate of 1 bit$/T$, and uses a padded input signal constellation corresponding to output points selected from the four-dimensional integer lattice. This code gives an overall gain of 5.86 dB relative to single-step detection, which is about 2.3 dB more gain than the 256/7 code shown in Table I. Of course, more coding gain than reported in the tables can be obtained by using longer block lengths. Results in [18] indicate that an optimal signal set of block length $K=10$ ($N=2048$ for $R=1$ bit$/T$ with padding) is needed to achieve a 6 dB coding gain. However, practical issues, such as decoding complexity, are likely to favor the trellis coding approach.

Of course, it is also possible to superimpose a trellis code on one of the optimized signal sets found in this section. Specifically, some of the smaller signal sets can be set-partitioned by inspection [2], and a trellis code constructed according to rules in [2]. This approach was tried for some examples, and was found to offer relatively little additional coding gain for the added complexity. This is consistent with previous results in the literature that state that trellis codes based on coset partitions of the multidimensional integer lattice often give more coding gain than trellis codes of the same complexity, but which are based on cosets of a denser lattice [5], [26].

V. Transmitted Spectra

As discussed in the last section, the optimized signal constellations have the property that more signal levels, or bits, are allocated to dimensions corresponding to larger eigenvalues. Furthermore, in many cases dimensions corresponding to the smallest eigenvalues are not used. This implies that as $K \to \infty$, the spectrum of the transmitted signal is concentrated in frequency bands where the channel transfer function has the least attenuation. In this section we show that for fixed minimum distance d, large information rate R (many bits$/T$), and large K, the transmitted spectrum associated with an optimized signal set is approximately given by

$$S(f) \approx \begin{cases} c, & \text{if } f \in F(f;\kappa), \\ 0, & \text{if } f \notin F(f;1), \end{cases} \tag{5.1}$$

where c is a constant determined by the input power constraint, and

$$F(f;\kappa) = \left\{ f\colon |H(f)| \geq \frac{\kappa d}{2} \right\}, \tag{5.2}$$

where κ is a constant, and the average energy per signal point, PK, is normalized to one. The numerical examples considered in this section suggest that $\kappa \approx 2$. For frequencies f such that $d/2 < |H(f)| < \kappa d/2$, $S(f)$ makes a smooth transition between 0 and c. For large K, solutions to the l_2/l_2 signal design problem therefore result in a transmitted spectrum which is approximately white over the frequency band $F(f;\kappa)$.

The input power spectral density given by (5.1) is similar to the input power spectral density obtained from "water pouring," which achieves Shannon capacity for dispersive channels with additive Gaussian noise. Specifically, assuming additive white Gaussian noise with spectral density N_0, the input power spectral density that achieves capacity is [20, Section 8.3]

$$S(f) = \begin{cases} M - \dfrac{N_0}{|H(f)|^2}, & |H(f)|^2 \geq \dfrac{N_0}{M}, \\ 0, & |H(f)|^2 \leq \dfrac{N_0}{M}, \end{cases} \tag{5.3}$$

where M is a constant chosen to satisfy the input average power constraint. As the information rate becomes large, the minimum distance d tends to zero, assuming fixed average energy. If the error exponent is a constant, then N_0 is proportional to d^2, and also tends to zero. Consequently, for large rates the water pouring spectrum (5.3) also becomes constant over the frequency band $\{f\colon |H(f)| \gg N_0\}$. This property has previously been noted by Price [17] (see also [27]).

Given the input signal set $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_N$ on the interval $[1, K]$, the Fourier series associated with the input $\boldsymbol{u}_i$ is

$$\hat{u}_i(f) = \sum_{k=1}^{K} u_i[k] e^{-j2\pi f(k-1)}, \tag{5.4}$$

and the average input spectrum is defined as

$$S(f) = \frac{1}{KN} \sum_{i=1}^{N} |\hat{u}_i(f)|^2, \tag{5.5}$$

where f denotes normalized frequency (that is, the analog frequency times the symbol interval T). The transmitted signal is then obtained by concatenating a sequence of vectors chosen from the preceding signal set. Each vector is selected from successive RK source bits. Strictly speaking, this transmitted signal is nonstationary for $K > 1$. However, the preceding definition of average spectrum is the "short-term" spectrum of this transmitted signal within coded blocks, ignoring block-to-block edge effects.

The average input spectrum for the signal sets obtained in Section IV must be computed numerically. However, the estimate (5.1) of average input spectrum for an optimized signal set is easily obtained as both the block length K and the information rate R become large. This estimate relies upon the following assumption, which is based on the results in Section IV. It is assumed that for moderate to large R (i.e., $R \geq 2$ bits$/T$), the input signal constellation $\{\tilde{\boldsymbol{u}}_i\}$ consists of points inside an n-sphere of radius r, where r is determined by the average power constraint, and are uniformly distributed in each dimension, but the density of points along dimension i is $(\lambda_i/\lambda_1)^{1/2}$ times the density of points along the dimension corresponding to λ_1. This implies that the output signal constellation $\tilde{y}_1, \cdots, \tilde{y}_N$ consists of points uniformly distributed inside an ellipsoid with axes $r\lambda_1^{1/2}, \cdots, r\lambda_n^{1/2}$. The volume argument at the end of Section IV implies that the dimension n is approximately the largest integer such that $r\lambda_n^{1/2} \geq d/2$. Note that the density of input points is a constant throughout the sphere of radius r even though the densities along each axis are different. Although we have not proved this assumption in general, we remark that a proof for the special case of $n = K = 2$ and the identity channel appears in [16]. This proof is easily generalized so as to apply to PR channels with $n = 2$, resulting in the preceding assumption.

Approximation (5.1) also depends upon the well-known fact that as the block length $K \to \infty$, the columns of Φ, defined by (2.7), which are eigenvectors of $\boldsymbol{H}'\boldsymbol{H}$, become sinusoidal functions. In other words, a sinusoidal input with frequency f_0 to a linear system with transfer function $H(f)$ produces a sinusoidal output with frequency f_0; the corresponding eigenvalue is $|H(f_0)|^2$. Furthermore, as $K \to \infty$, the frequencies of these sinusoids are uniformly distributed on the interval $[-0.5, 0.5]$, or equivalently, on the interval $[0, 0.5]$ since $|H(f)|$ is an even function. More precisely, let $N_k(\alpha, \beta)$ be the number of eigenvalues of $\boldsymbol{H}'\boldsymbol{H}$ between α and β for fixed block length K. The Szegö theorem on the asymptotic distribution of eigenvalues of a linear operator implies that [25]

$$\lim_{K \to \infty} \frac{1}{K} N_K(\alpha, \beta) = 2\,\text{meas}\{f: \alpha < |H(f)|^2 < \beta, 0 < f < 0.5\}, \quad (5.6)$$

where $H(f)$ is defined by (4.7). As $K \to \infty$, the λ_i's therefore behave as uniformly distributed samples of the channel spectrum, so that the corresponding set of eigenvectors can be approximated as sinusoids at uniformly distributed frequencies. (A definition of "approximate" eigenvalues and eigenvectors is given in [28].) It is shown in Appendix B that if the impulse response is of length two, then for any finite K it is precisely the case that $\lambda_k(K) = |H[\sigma_k/(2K+2)]|^2$, $k = 1, \cdots, K$, where $\sigma_k, k = 1, \cdots, K$, is the set of permuted indexes for which $|H[\sigma_k/(2K+2)]|$, $k = 1, \cdots, K$, is monotonically decreasing. The eigenvector corresponding to λ_k has components $\sin(\pi\sigma_k m/(K+1))$, $m = 1, \cdots, K$.

Letting $\phi_{lm} = [\Phi]_{lm}$, then

$$\hat{u}_i(f) = \sum_{l=1}^{K} \left(\sum_{m=1}^{K} \phi_{lm} \tilde{u}_i[m] \right) e^{-2\pi f(l-1)} = \sum_{m=1}^{K} \left(\sum_{l=1}^{K} \phi_{lm} e^{-j2\pi f(l-1)} \right) \tilde{u}_i[m]. \quad (5.7)$$

Now the inner sum is the Fourier transform of the mth eigenvector of $\boldsymbol{H}'\boldsymbol{H}$, which is approximately sinusoidal with frequency $f = \sigma_m/(2K+2)$. If we consider $\hat{u}_i(f)$ evaluated at the discrete frequencies $f = k/(2K+2)$, $k = 1, \cdots, K$, then for large K the inner sum can be approximated as the Kronecker delta δ_{k,σ_k}. Let σ_k^{-1}, $k = 1, \cdots, K$, denote the inverse permutation to σ_k defined previously, i.e., if $k' = \sigma_k$, then $k = \sigma_{k'}^{-1}$. The σ_k^{-1}th eigenvector is therefore approximately sinusoidal with frequency $k/(2K+2)$. For large K, we assume that

$$\hat{u}_i\left(\frac{k}{2(K+1)}\right) \approx \tilde{u}_i[\sigma_k^{-1}], \quad k = 1, \cdots, K, \quad i = 1, \cdots, N. \quad (5.8)$$

If the impulse response has length $\tau = 2$, then because $\phi_{km} = \sin[(\pi\sigma_k m)/(K+1)]$, (5.8) becomes exact as $K \to \infty$.

Given that the Szegö theorem specifies only the asymptotic distribution of eigenvalues, and not the behavior of individual eigenvalues, we cannot expect (5.8) to be exact as $K \to \infty$, except when $\tau = 2$. However, we do expect that the components of $\tilde{\boldsymbol{u}}$ and $\hat{u}[k/(2K+2)]$, $k = 1, \cdots, K$, have the same distribution of values asymptotically. Specifically, for given positive real numbers α and β, let $\tilde{L}_K(\alpha, \beta)$ be the number of components k such that $\alpha \leq |\tilde{u}[k]| < \beta$, and $\hat{L}_K(\alpha, \beta)$ be the number of indexes k such that $\alpha \leq |\hat{u}[k/(2K+2)]| < \beta$. A weaker version of the assumption (5.8) states that for $\boldsymbol{u}$ in an optimal signal set, $\hat{L}_K$ can be accurately approximated by $\tilde{L}_K$ for large enough K. This statement can be made more precise by using the notion of approximate eigenvalues and eigenvectors introduced in [28].

Combining (5.5) and (5.8), we can estimate the average input spectrum at the frequency $f = k/(2K+2)$ as

$$S\left(\frac{k}{2(K+1)}\right) \approx \frac{1}{KN} \sum_{i=1}^{N} \left|\tilde{u}_i[\sigma_k^{-1}]\right|^2, \quad k = 1, \cdots, K. \quad (5.9)$$

Since $\tilde{u}_i[k] = 0$ for $k > n$, this implies that $S[\sigma_k/(2K+2)] = 0$ for $k > n$, where n is approximately the largest integer such that $r\lambda_n^{1/2} \geq d/2$, and r is the radius of the sphere containing the input signal constellation. Assuming the constellation points are uniformly and continuously distributed throughout this sphere, then the average energy per point is $nr^2/(n+2)$. As $K \to \infty$, $n \to \infty$, so that the average energy per point becomes r^2, i.e., all points gravitate towards the surface of the sphere. For large K, we therefore conclude that r is accurately approximated as $\sqrt{PK} = 1$, which implies that $S[\sigma_k/(2K+2)] \approx 0$ where $\lambda_k^{1/2} \leq d/2$. As $K \to \infty$, the Szegö theorem therefore implies that $S(f) \approx 0$ if $|H(f)| < d/2$, as stated in (5.1).

Assume now that the $\tilde{\boldsymbol{u}}_i$'s are uniformly distributed in each dimension and are contained in a sphere. Choose now a dimension k for which λ_k is large enough so that the input constellation points can be assumed to be continuously distributed along this dimension (i.e., $2\lambda_k^{1/2} \gg d$). Then, (5.9) becomes

$$S\left(\frac{\sigma_k}{2(K+1)}\right) \approx \frac{\int_{-1}^{1} \rho x^2\, dV(x)}{K \int_{-1}^{1} \rho\, dV(x)}, \quad (5.10)$$

where ρ is the density of points in the n-sphere, $dV(x)$ is the differential volume of the region formed by the intersection of the interior of the sphere containing the input points with planes perpendicular to the kth axis at the points x and $x+dx$, and the denominator is $N=2^{RK}$. Since $S[\sigma_k/(2K+2)]$ in (5.10) is independent of k, this implies (5.1) where κ in (5.2) is chosen to satisfy the assumption that the number of points distributed along dimension k is sufficiently large, so that the preceding integral approximation is valid. Taking k to be any dimension such that $\lambda_k^{1/2} \geq d$ implies that $\kappa = 2$, and appears to be adequate for the examples shown in this section.

Characterization of $S(f)$ for frequencies f such that $d/2 < |H(f)| < \kappa d/2$ is relatively difficult. However, the following plausibility argument indicates that for uniformly distributed input signal constellations bounded by a sphere, if $|H(f)|$ is monotonically increasing (decreasing) where $d/2 < |H(f)| < \kappa d/2$, then $S(f)$ is monotonically increasing (decreasing) for f in this frequency range. Consider first a two-dimensional input signal constellation consisting of a regular lattice with minimum distance d stretched in the x-y plane by $\lambda_k^{1/2}$, $k=1,2$, respectively, and bounded by a circle of radius one. That is, distances in the $x(y)$ direction are multiplied by $\lambda_1^{-1/2}(\lambda_2^{-1/2})$. Assume also that $\lambda_1 \gg \lambda_2$ (the density of points along the x axis is much greater than along the y axis), and that $\lambda_2^{-1/2} \geq 2/d$. In this case all of the points will be clustered near the x axis, so that the average of the squared y components will be close to zero. As λ_2 increases, however, the number of points that can fit inside the circle increases, and more points will be farther from the x axis, so that the average energy of the constellation along the y axis increases. As $\lambda_2^{1/2}$ becomes much larger than $d/2$, the distribution of points along the y axis becomes uniform, and the energy along the y axis becomes the same as the energy along the x axis.

The preceding argument easily generalizes to $n > 2$ dimensions by projecting the n-dimensional signal constellation onto two dimensions corresponding to λ_k and λ_1. When $r\lambda_k^{1/2}$ is not much greater than $d/2$, most of the points will be clustered around the axis corresponding to λ_1; however, as k decreases ($\lambda_k^{1/2}$ increases) more points will be spread along the kth dimension, so that the average energy with respect to the kth dimension increases. Combining this with (5.10) indicates that $S(f)$ increases (decreases) if $|H(f)|$ is increasing (decreasing) in the region where $d/2 < |H(f)| < \kappa d/2$. Note that this argument does not assume any particular lattice, just that the signal constellation lies within a sphere. The spectrum of the constellation is therefore determined primarily by its *shape*, rather than by the lattice from which it is chosen.

The preceding arguments, and hence approximation (5.1), apply to both the HIC and SIC. To estimate c, note that

$$\int_0^1 S(f)\,df = \frac{1}{KN}\sum_{i=1}^{N}\int_0^1 |\hat{u}_i(f)|^2\,df = \frac{1}{KN}\sum_{i=1}^{N}\|\boldsymbol{u}_i\|^2 = P, \tag{5.11}$$

which is true assuming either the HIC or SIC. Consequently, if meas$\{f\colon d/2 \leq |H(f)| \leq \kappa d/2\}$ is small, then

$$c \approx \frac{P}{\text{meas}\,F(f;\kappa)}. \tag{5.12}$$

From the preceding discussion it is obvious that the shape of the spectrum given by (5.1) applies to any input signal set consisting of points uniformly distributed throughout a region that is invariant with respect to permutation of axes. If the signal set is not optimal, as defined here, then the type of volume argument given in the last section implies that the set $F(f;\kappa)$ should be replaced by the set $F'(f;\kappa) = \{f\colon |H(f)| \geq \kappa d'/2\}$, where d' is the minimum distance between points in the output constellation. Since by definition $d' \leq d$, this argument would imply that the bandwidth (i.e., measure of the set F') occupied by a "suboptimal" signal constellation satisfying the preceding conditions (such as one of the constellations described in [12]) is greater than the bandwidth occupied by an optimal signal set with the same number of points. However, the "volume" estimate for the number of dimensions spanned by an optimal signal set assuming $PK = 1$, $n = \max\{k\colon \lambda_k^{1/2} \geq d/2\}$, is typically less than the number of dimensions spanned by a particular PR signal set found in Section IV. Consequently, the preceding approximate comparison of bandwidth occupied by optimal signal sets as $K \to \infty$ with the bandwidth of other signaling schemes is inconclusive.

The behavior of $S(f)$ where $d'/2 < |H(f)| < \kappa d'/2$, depends a great deal on the shape, or boundary region, of the signal constellation, rather than on the underlying lattice from which the points are chosen. This is simply because the density of any lattice is uniform throughout n-space (even when scaled by the $\lambda_k^{-1/2}$'s in each dimension), and the relative energy of a constellation in each dimension, given by (5.10), is independent of this density. This will be further demonstrated in Part II [18], where the spectrum of an input constellation based on the integer lattice and bounded by a cube is computed.

We conclude this discussion by showing plots of average spectra, given by (5.5), for two signal sets represented by the numerical results in Section IV. Figs. 3(a) and 3(b) show average spectra for the 256/8 codes obtained for the $1-D$ and $1-D^2$ channels, respectively. Also shown in each figure are the water pouring spectrum given by (5.3) where $N_0 = d^2/4$ (i.e., the noise has standard deviation $d/2$), and the approximation to the spectrum at frequencies $f = k/(2K+2)$, $k = 1,\cdots,K$, given by (5.9). It is interesting that the input spectrum in each of these two cases resembles the corresponding water pouring spectrum. All spectra are scaled so that they integrate to one.

Although the information rate (as well as the input length) for the codes used to generate Figs. 3(a) and 3(b) is relatively small (1 bit/T), some characteristics of asymptotic spectra previously described can be observed. In both cases, the estimate (5.9) accurately approximates

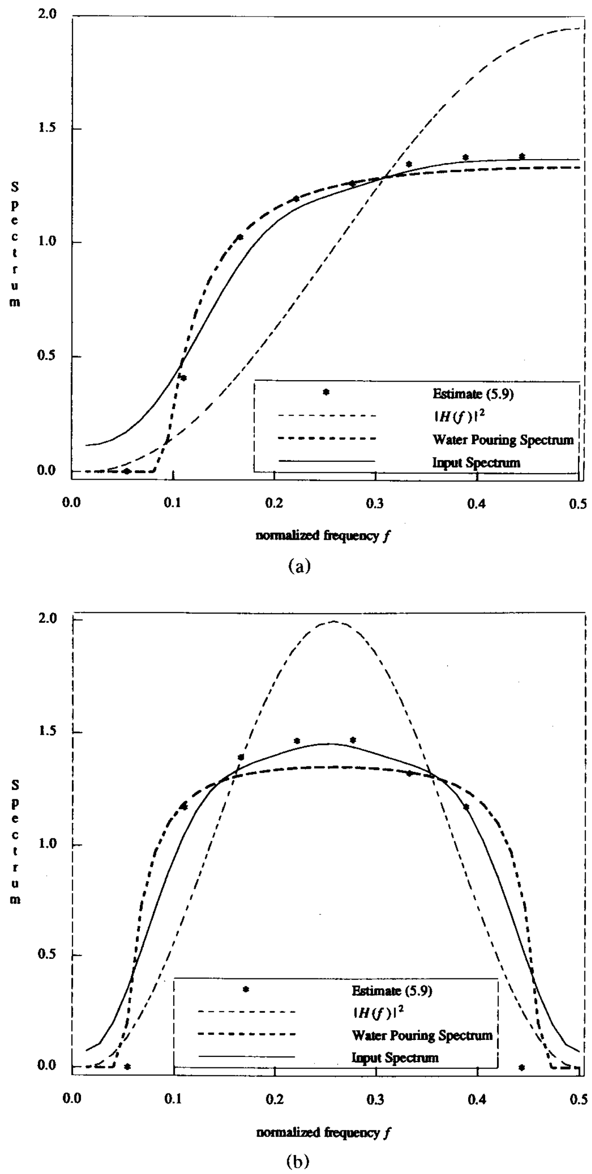


Fig. 3. (a) Input spectrum for the 256/8 code found by computer search for the $1-D$ channel. Also shown are the water pouring spectrum, assuming additive white noise, and the average squared signal components, given by (5.9). (b) Input spectrum for 256/8 code found by computer search for the $1-D^2$ channel. Also shown are the water pouring spectrum, assuming additive white noise, and the average squared signal components, given by (5.9).

the input spectrum at the discrete frequencies $f = k/18$, $k = 1,\cdots,8$. The $1-D$ code uses seven out of the available eight dimensions, so that the estimate (5.9) is zero at $f = 1/18$. The $1-D^2$ code uses six out of the available eight dimensions. The input spectra do not go to zero at the frequencies where the estimate (5.9) is zero, however, because in both cases the discrete Fourier transform of the mth eigenvector of $\boldsymbol{H}'\boldsymbol{H}$, evaluated at frequencies $f = k/(2K+2)$, $k = 1,\cdots,K$, is not exactly δ_{k,σ_k}. This, combined with (5.7), implies that the approximation (5.8) is not exact.

For both the $1-D$ and $1-D^2$ signal sets, $d^2/4 \approx 0.33$. Figs. 3(a) and 3(b) therefore agree with (5.1) where $\kappa = 2$. Specifically, in both cases the spectrum is relatively flat where $|H(f)|^2 > d^2$, and is close to zero where $|H(f)|^2 < d^2/4$. The input lengths for the signal sets corresponding to 2 bits/T shown in Tables I–III are too short to exhibit the asymptotic behavior predicted by (5.1). Specifically, for the 1024/5 codes, d is small enough so that there are many points with components in all five dimensions. The average spectra of these codes therefore do not decrease where the channel attenuation is small.

VI. Conclusion

We have shown that the l_2/l_2 signal design problem, assuming the HIC, is equivalent to packing N points in an ellipsoid in $\mathbb{R}^n$, where $n \le N-1$, so as to maximize the minimum Euclidean distance between points. A similar interpretation was shown to hold for the SIC problem. Numerical techniques were then proposed, and used to find locally optimal solutions to these packing problems. Proving that a locally optimal signal set is globally optimal appears to be quite difficult.

The optimization of signal sets in Sections III and IV paid no attention to important practical issues such as receiver complexity (i.e., decoding), and peak-to-average power. In addition, if passband transmission is assumed, then additional considerations such as phase symmetry and constituent two-dimensional constellations become important [29].

In practice, decoding may be only marginally more difficult for PR codes than for identity channel codes. This is because when the number of bits/T is two or greater, nearly optimal codes can be generated by cropping a dense lattice in $\mathbb{R}^n$. Assuming the eigenvectors of $\boldsymbol{H}'\boldsymbol{H}$ are the basis vectors for the input space, the lattice must be scaled along the kth dimension by $\lambda_k^{1/2}$. Assuming the channel impulse response is known *a priori*, after appropriately scaling and rotating the received signal points, decoding can proceed exactly as for the identity channel. However, if the rate is reduced to 1 bit/T, then the signal constellations found by computer search in Section IV are no longer based on lattices, although it is likely that these signal sets have considerable structure, which can be exploited to minimize decoding complexity. For any information rate, however, the fact that the output constellation is not necessarily symmetric about the origin can present problems with phase ambiguities. This may be a topic for future investigation.

An additional consideration that has been ignored here, but is important for channels with additive Gaussian noise, is the number of nearest neighbors corresponding to each codeword. This consideration is inherently ignored in l_q/l_p signal design, since the optimization criterion is simply minimum distance between outputs, and assumes no statistical model by which probability of error can be evaluated. Nevertheless, the numbers of nearest-neighbors for the A_2 and D_k lattices, $k = 3,\cdots,5$, on which many of the codes in Section IV are based, are relatively small [3], so that these codes might be useful for channels with additive Gaussian noise.

It has also been shown empirically that the spectra of two signal sets represented in Section IV with input length $K = 8$ are quite similar to the corresponding water pouring spectra that achieve Shannon capacity, assuming additive white Gaussian noise. As the information rate increases, it has also been shown that both spectra become constant over a frequency band whose measure increases with rate. It has been pointed out that the average spectrum of a signal constellation is primarily determined by its shape, i.e., the region in which it lies, assuming the signal points are uniformly distributed throughout this region.

The numerical techniques discussed in Section III are also applicable to other types of signal design, say l_q/l_p, in general. Specifically, maximization of minimum distance between outputs in the l_p sense, and given an l_q constraint on the inputs, is no longer an ellipsoid packing problem, but is another nonlinear optimization problem, and one can again find locally optimal solutions numerically. This approach has been used to design signal sets for the l_∞/l_∞ problem [30]. Application of these numerical techniques to other signal design problems remains to be explored; for example, a particularly interesting candidate is the l_∞/l_2 problem—maximization of minimum Euclidean distance between outputs subject to an *amplitude* constraint on the inputs.

ACKNOWLEDGMENT

The authors thank S. Boyd, D. J. Hajela, M. Fredman, H. J. Landau, P. Narayan, and V. Wei for helpful discussions concerning this work.

APPENDIX A
PROOF OF THEOREM

Since λ_k is the kth largest eigenvalue of $\boldsymbol{H}'\boldsymbol{H}$, we apply the Courant–Fischer theorem, which states that [31, Section 10-2]

$$\lambda_k = \max_{S_{N-k}} \min_{\boldsymbol{v} \perp S_{N-k}} \{\boldsymbol{v}'\boldsymbol{H}'\boldsymbol{H}\boldsymbol{v}: \|\boldsymbol{v}\| = 1\}, \qquad \text{(A.1)}$$

where the first maximum is over all $(N-k)$-dimensional spaces S_{N-k}, and the minimum is over all vectors orthogonal to S_{N-k}. Clearly, the S_{N-k} for which this maximum occurs is the space spanned by the $N-k$ eigenvectors of $\boldsymbol{H}'\boldsymbol{H}$ corresponding to $\lambda_{k+1},\cdots,\lambda_N$. Denoting this space as S'_{N-k}, then we have the orthogonal decomposition $S_N = S'_{N-k} \oplus S''_k$ where S''_k is the k-dimensional space orthogonal to S'_{N-k}. We can therefore rewrite (A.1) as

$$\lambda_k = \max_{S_k} \min_{\boldsymbol{v} \in S_k} \{\boldsymbol{v}'\boldsymbol{H}'\boldsymbol{H}\boldsymbol{v}: \|\boldsymbol{v}\| = 1\}. \qquad \text{(A.2)}$$

The S_k for which the maximum occurs is now the space spanned by the first k eigenvectors.

Assume initially that the center of the ellipsoid E' is at the origin. Letting Q denote the $(N-1)$-dimensional hyperplane in which E' lies, we again use the Courant–Fischer theorem to write

$$\mu_k = \max_{\substack{S_k \\ S_k \subset Q}} \min_{\boldsymbol{v} \in S_k} \{\boldsymbol{v}'\boldsymbol{H}'\boldsymbol{H}\boldsymbol{v}: \|\boldsymbol{v}\| = 1\}, \qquad \text{(A.3)}$$

that is, since E' lies in E, μ_k has the same form as λ_k with the additional restriction that the space S_k must lie in Q. Since the maximum in (A.3) is over a more restrictive set than in (A.1), it follows that $\mu_k \le \lambda_k$.

We add that the min-max form of the Courant–Fischer theorem states that

$$\lambda_k = \min_{S_{N-k+1}} \max_{\boldsymbol{v} \in S_{N-k+1}} \{\boldsymbol{v}'\boldsymbol{H}'\boldsymbol{H}\boldsymbol{v}: \|\boldsymbol{v}\| = 1\}$$
$$\le \min_{S_{N-k+1} \subset Q} \max_{\boldsymbol{v} \in S_{N-k+1}} \{\boldsymbol{v}'\boldsymbol{H}'\boldsymbol{H}\boldsymbol{v}: \|\boldsymbol{v}\| = 1\} = \mu_{k-1}, \qquad \text{(A.4)}$$

where the last equality follows from the fact that Q has dimension $N-1$, so that μ_k is found by minimizing over $N-k$ dimensional spaces. Consequently, if E' is centered at the origin, we have shown that $\lambda_{k+1} \le \mu_k \le \lambda_k$, $k = 1,\cdots,N-1$.

If E' is centered at $\boldsymbol{x}_0 \ne 0$, then we translate E by $\boldsymbol{x}_0$. This is accomplished with the operator $\boldsymbol{T} = \boldsymbol{T}_3\boldsymbol{T}_2\boldsymbol{T}_1$, where

$$\boldsymbol{T}_1: \boldsymbol{x} \to \Lambda^{-1/2}\boldsymbol{x}, \qquad \boldsymbol{T}_2: \boldsymbol{x} \to \boldsymbol{x} - \boldsymbol{x}_0, \qquad \boldsymbol{T}_3: \boldsymbol{x} \to \Lambda^{1/2}\boldsymbol{x}.$$

The transformation $\boldsymbol{T}_1$ maps both E and E' into spheres ($\boldsymbol{T}_1(E)$ has radius one and is centered at the origin), $\boldsymbol{T}_2$ translates $\boldsymbol{T}_1(E')$ so that its center is at its origin, and $\boldsymbol{T}_3$ restores the spheres to ellipsoids with the original dimensions.

Since $E = \{\boldsymbol{x}: \boldsymbol{x}'\Lambda^{-1}\boldsymbol{x} = 1\}$, we have that

$$\boldsymbol{T}(E) = \{\boldsymbol{x}: (\boldsymbol{x}-\boldsymbol{x}_0)'\Lambda^{-1}(\boldsymbol{x}-\boldsymbol{x}_0) = 1\}. \qquad \text{(A.5)}$$

Now $\boldsymbol{T}_1(E')$ is the intersection of a plane ($\boldsymbol{T}_1(Q)$) with the unit sphere, so that the vector from the origin to the center of $\boldsymbol{T}_1(E')$, i.e., $\boldsymbol{T}_1(\boldsymbol{x}_0)$, must be orthogonal to any vector in $\boldsymbol{T}_1(E')$. This implies that $\boldsymbol{x}'_0\Lambda^{-1}\boldsymbol{x} = 0$ for any $\boldsymbol{x} \in \boldsymbol{T}_1(E')$. Consequently,

$$\boldsymbol{T}(E') = \boldsymbol{T}(E) \cap \boldsymbol{T}(Q)$$
$$= \{\boldsymbol{x}: \boldsymbol{x}'\Lambda^{-1}\boldsymbol{x} = 1 - \boldsymbol{x}'_0\Lambda^{-1}\boldsymbol{x}_0,\ \boldsymbol{x} \in \boldsymbol{T}(Q)\}. \qquad \text{(A.6)}$$

Since $\boldsymbol{x}_0$ lies inside E, $0 \le 1 - \boldsymbol{x}'_0\Lambda^{-1}\boldsymbol{x}_0 \le 1$, so that μ_k is clearly maximized by taking $\boldsymbol{x}_0 = 0$, which reduces to the preceding case in which E' is centered at the origin. □

APPENDIX B
EIGENVECTORS OF $\boldsymbol{H}'\boldsymbol{H}$ FOR LENGTH-2 IMPULSE RESPONSE

Suppose that the impulse response satisfies $h[k] = 0$ for $k < 0$ and $k > 1$. With an input of length K, $\boldsymbol{H}$ is the $(K+1) \times K$ matrix with entries $[\boldsymbol{H}]_{km} = h[k-m]$, and

$$[\boldsymbol{H}'\boldsymbol{H}]_{kr} = \sum_{i=1}^{K+1} h[i-k]h[i-r], \qquad k,r = 1,\cdots,K. \qquad \text{(B.1)}$$

We will show that the K-vectors with components

$$\phi[k] = \sin k\omega, \qquad k = 1, \cdots, K \qquad \text{(B.2)}$$

are eigenvectors of $\boldsymbol{H'H}$ for the discrete values $\omega = \pi i/(K+1)$, $i = 1, \cdots, K$. Since the matrix $\boldsymbol{H'H}$ is $K \times K$, this is a complete set.

For convenience in algebraic manipulation, we will consider a complex equation, but enforce only the imaginary part. In particular, the imaginary part of the following equation must hold:

$$\sum_{r=1}^{K} [\boldsymbol{H'H}]_{kr} e^{jr\omega} = \lambda e^{jk\omega}, \qquad k = 1, \cdots, K. \quad \text{(B.3)}$$

Using (B.1) and rearranging sums, this becomes

$$\left[\sum_{s=1-k}^{K+1-k} h[s] e^{js\omega} \sum_{q=s+k-1}^{s+k-K} h[q] e^{-jq\omega} \right] e^{jk\omega} = \lambda e^{jk\omega}, \qquad k = 1, \cdots, K. \quad \text{(B.4)}$$

When $k = 2, \cdots, K-1$ we can write this as

$$\left| h[0] + h[1] e^{-j\omega} \right|^2 e^{jk\omega} = \lambda e^{jk\omega}, \qquad k = 2, \cdots, K-1, \quad \text{(B.5)}$$

so a candidate for the eigenvalue corresponding to ω is

$$\lambda = \left| h[0] + h[1] e^{-j\omega} \right|^2, \qquad \text{(B.6)}$$

which is, of course, the squared magnitude of the channel frequency response at frequency ω. When $k = 1$ or K, however, the left-hand side of (B.5) contains terms missing from the left-hand side of (B.4). These missing terms are

$$h(0)h(1), \qquad \text{for } k = 1,$$

and

$$h(0)h(1)e^{j\omega(K+1)}, \qquad \text{for } k = K.$$

The required eigenvalue conditions will be satisfied if the imaginary parts of these quantities are zero. The first quantity is real since the impulse response is real-valued; the second quantity is real when $\omega = i\pi/(K+1)$, which is what we wanted to show.

References

[1] M. L. Honig, K. Steiglitz, B. Gopinath, and S. Boyd, "Bounds on maximum throughput for digital communications with finite precision and amplitude constraints," *IEEE Trans. Inform. Theory*, vol. 36, no. 3, pp. 472–484, May 1990.

[2] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory*, vol. IT-28, no. 1, pp. 55–67, Jan. 1982.

[3] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices, and Groups.* New York: Springer-Verlag, 1988.

[4] G. D. Forney *et al.*, "Efficient modulation for band-limited channels," *IEEE J. Select. Areas Commun.*, vol. SAC-2, no. 5, pp. 632–647, Sept. 1984.

[5] G. D. Forney, "Coset codes—Part I: Introduction and geometrical classification," *IEEE Trans. Inform. Theory*, vol. 34, no. 5, pp. 1123–1151, Sept. 1988.

[6] A. R. Calderbank and N. J. A. Sloane, "New trellis codes based on lattices and cosets," *IEEE Trans. Inform. Theory*, vol. IT-33, no. 2, pp. 177–195, Mar. 1987.

[7] L. F. Wei, "Trellis-coded modulation with multidimensional constellations," *IEEE Trans. Inform. Theory*, vol. IT-33, no. 4, pp. 483–501, July 1987.

[8] G. D. Forney, Jr. and A. R. Calderbank, "Coset codes for partial response channels, or, coset codes with spectral nulls," *IEEE Trans. Inform. Theory*, vol. 35, no. 5, pp. 925–943, Sept. 1989.

[9] A. R. Calderbank and J. E. Mazo, "Baseband line codes via spectral factorization," *IEEE J. Select. Areas Commun.*, vol. 7, no. 6, pp. 914–928, Aug. 1989.

[10] J. K. Wolf and G. Ungerboeck, "Trellis coding for partial-response channels," *IEEE Trans. Commun.*, vol. COM-34, pp. 765–773, Aug. 1986.

[11] J. Lechleider, "The optimum combination of block codes and receivers for arbitrary channels," *IEEE Trans. Commun.*, vol. 38, no. 5, pp. 615–621, May 1990.

[12] S. Kasturia, J. T. Aslanis, and J. M. Cioffi, "Vector coding for partial-response channels," *IEEE Trans. Inform. Theory*, vol. 36, no. 4, pp. 741–762, July 1990.

[13] D. W. Lin, "Wideband digital subscriber access with multi-dimensional block modulation and decision-feedback equalization," *IEEE J. Select. Areas Commun.*, vol. 7, no. 6, pp. 996–1005, Aug. 1989.

[14] W. L. Root, "Estimates of ϵ-capacity for certain linear communication channels," *IEEE Trans. Inform. Theory*, vol. IT-14, no. 3, pp. 361–369, May 1968.

[15] R. T. Prosser and W. L. Root, "The ϵ-entropy and ϵ-capacity of certain time-invariant channels," *J. Math. Anal. Applicat.*, vol. 21, pp. 233–241, 1968.

[16] G. J. Foschini, R. D. Gitlin, and S. B. Weinstein, "Optimization of two-dimensional signal constellations in the presence of Gaussian noise," *IEEE Trans. Commun.*, vol. COM-22, pp. 28–38, 1974.

[17] R. Price, "Nonlinearly feedback-equalized PAM vs. capacity for noisy filter channels, in *Int. Conf. Commun.*, Philadelphia, PA, June 1972, pp. 22-12–22-17.

[18] M. L. Honig, "Optimization of signal sets for partial-response channels—Part II: Asymptotic coding gain," *IEEE Trans. Inform. Theory*, vol. 37, no. 5, pp. 1342–1354, Sept. 1991.

[19] J. L. Holsinger, "Digital communication over fixed time-continuous channels with memory—With special application to telephone channels," *M.I.T. Res. Lab. Electron. Rep.*, Tech. Rep. 430, 1964.

[20] R. G. Gallager, *Information Theory and Reliable Communication.* New York: John Wiley, 1968.

[21] K. Sistanizadeh, "Block coding capacity of digital subscriber lines by the structured-channel signalling technique," in *Proc. 1989 IEEE Globecom Conf.*, Dallas, TX, Nov. 1989.

[22] J. A. C. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Commun. Mag.*, vol. 28, no. 5, pp. 5–14, May 1990.

[23] A. G. Deczky, "Synthesis of recursive digital filters using the minimum p-error criterion," *IEEE Trans. Audio Electroacoust.*, vol. AU-20, pp. 257–263, Oct. 1972.

[24] S. Kasturia and J. M. Cioffi, "Vector coding with decision feedback equalization for partial response channels," in *Proc. 1988 Globecom Conf.*, Paper 27.3, Hollywood, FL, Nov. 1988.

[25] R. M. Gray, "On the asymptotic eigenvalue distribution of Toeplitz matrices," *IEEE Trans. Inform. Theory*, vol. IT-18, no. 6, pp. 725–730, Nov. 1972.

[26] G. J. Pottie and D. P. Taylor, "Sphere-packing upper bounds on the free distance of trellis codes," *IEEE Trans. Inform. Theory*, vol. 34, no. 3, pp. 435–447, May 1988.

[27] M. V. Eyuboglu, "Detection of coded modulation signals on linear, severely distorted channels using decision-feedback noise prediction with interleaving," *IEEE Trans. Commun.*, vol. 36, no. 4, pp. 401–409, Apr. 1988.

[28] H. J. Landau, "The notion of approximate eigenvalues applied to an integral equation of laser theory," *Quart. Appl. Math.*, pp. 165–172, Apr. 1977.

[29] G. D. Forney and L. F. Wei, "Multidimensional constellations—Part I: Introduction, figures of merit, and generalized cross constellations," *IEEE J. Select. Areas Commun.*, vol. 7, no. 6, pp. 877–892, Aug. 1989.

[30] M. L. Honig, K. Steiglitz, E. Rantapaa, and V. Balakrishnan, "l_∞/l_∞ signal design," presented at the *IEEE Int. Symp. Inform. Theory*, Budapest, Hungary, June 24–28, 1991.

[31] B. N. Parlett, *The Symmetric Eigenvalue Problem.* Englewood Cliffs, NJ: Prentice-Hall, 1980.