Proc. Conf. Information Sciences, Systems (Princeton, NJ), pp. 294-299, March 1992

Spectral Shaping with Unequal Power Distribution

A. K. Khandani[1] and P. Kabal[1,2]

[1]McGill University, 3480 University, Montreal, Canada, H3A 2A7
[2]INRS–Telecommunications, 3 Place du Commerce, Verdun, Canada, H3E 1H6

Abstract: We are going to maximize the entropy of a line code subject to some constraints on the power spectrum. The general tools are the selection of the constellation basis (modulating waveforms) and the power allocated to each constellation dimension. In our analysis, the basis is fixed and is selected to reduce the computational complexity of the modulation. The following constraints on the power spectrum are considered in detail: (i) A fraction of the total power equal to F_p is located in the frequency band $[0, \omega_c]$, and/or (ii) the spectrum has spectral nulls at the zero and/or at the Nyquist frequency. To realize a power spectrum with spectral null(s), we need a set of dimensions with the same set of nulls. We discuss the general structure of such basis. In specific, we give analytical expression for the basis providing spectral null at zero frequency and/or at the Nyquist frequency. These are either sine basis or closely related to it. This property reduces the computational complexity of the modulation by allowing for the use of the fast sine transform algorithms. The energy allocation is computed by an optimization procedure. We also propose a method to match the spectrum of the line code to a partial response channel. This procedure maximizes the entropy of the code subject to having equal minimum distance to noise ratio along all the dimensions at the channel output. The noise is the sum of the additive Gaussian noise and the intersymbol interference.

1 Introduction

Consider a signal constellation composed of a set of equiprobable points. In continuous approximation, the distribution of the constellation points is approximated by a continuous *uniform* density within the shaping region. Assuming continuous approximation, the selection of the constellation is composed of selecting a basis for the space and a boundary (shaping region) for the points. This selection can be formulated in terms of an optimization procedure. The objective function (to be maximized) is the rate of the constellation. There is always a constraint on the total energy. According to the application, we impose some additional constraints.

Due to the continuous approximation, the structure of the shaping region appears as an independent factor in the objective function. This reduces the complexity of the optimization procedure. Without loss of generality, we can assume that the shaping region is obtained by the scaling of a baseline region $\mathcal{B}$, [1]. For a fixed $\mathcal{B}$, the shaping problem reduces to the selection of an appropriate set of the scale factors. For a fixed total power, this is equivalent to a power allocation problem.

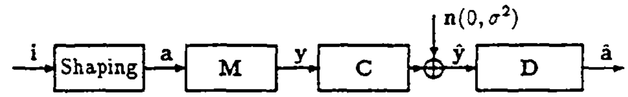

Fig. 1 System block diagram.

In this work, we impose some constraints on the power spectrum of the constellation. We assume that, the basis is fixed and the power allocation is optimized to maximize the rate. The basis is selected to reduce the computational complexity of the modulation by the use of the fast transform algorithms.

For a given cutoff frequency $\omega_c = 2\pi f_c$, define the power-ratio of a spectrum as the fraction of the total power in the frequency band $[0, \omega_c]$. The F_p-constraint is the constraint of having a power-ratio less than or equal to F_p. We study the following constraints in detail: (i) the F_p-constraint, and/or (ii) the spectrum has some spectral nulls.

We also study the selection of a line code for the signaling over a Partial-Response channel. In this case, the entropy is maximized subject to the constraint of having equal minimum distance to noise ratio along all the dimensions. The noise is composed of the additive white Gaussian noise and the intersymbol interference. This results in a spectrum which is matched to the channel. The optimization procedure also concerns the selection of the constellation basis.

2 System block diagram

Figure 1 shows the block diagram of the system under consideration. We use discrete time model and block based processing. In each signaling interval, which is composed of M channel use, a binary data vector $\mathbf{i}$ is transmitted. The shaping block maps the vector $\mathbf{i}$ to the N-dimensional point $\mathbf{a}$ in the baseband constellation. This is a finite portion of the coding lattice bounded within the shaping region $\mathcal{R}_a$. The second moment along the i'th dimension of $\mathcal{R}_a$ is equal to λ_i. The diagonal matrix Λ_a is defined as, $\Lambda_a = \text{diag}[\lambda_0, \ldots, \lambda_{N-1}]$. Normalizing the volume of the Voronoi region around each constellation point to unity, the entropy of $\mathbf{a}$ is equal to.

$$H(\mathbf{a}) = \log[V(\mathcal{R}_a)]. \quad (1)$$

This is based on the assumption that the points $\mathbf{a}$ are used with equal probability.

The modulator, $M \times N$, $M \geq N$, matrix $\mathbf{M}$ maps the point $\mathbf{a}$ to the point $\mathbf{y}$ in the modulated constellation. This is a set of points in an N-dimensional subspace of the M-dimensional channel space. The dimensions of the modulated constellation are along the columns of $\mathbf{M}$. In general, there exists $M-N$ linear relationships between the components of $\mathbf{y}$. As an example, for $M \geq N+1$, we can select $\mathbf{M}$ such the sum of the components in each of its columns is equal to zero. In this case, $\mathbf{y}$ will have a spectral null at zero frequency. For a fixed rate per each channel dimension, decreasing N beyond M results in a higher rate per each constellation dimension. For a fixed minimum distance, such a constellation requires a higher energy. This can be considered as the price of introducing $M-N$ linear relationships between the components of $\mathbf{y}$.

It can be shown that the entropy of $\mathbf{y}$ in Fig. 1 is equal to,

$$H(\mathbf{y}) = H(\mathbf{a}) + \log\left(|\mathbf{M}^t\mathbf{M}|\right). \quad (2)$$

Define,

$$H(\mathbf{M}) = \log\left(|\mathbf{M}^t\mathbf{M}|\right), \quad (3)$$

as the entropy of the system $\mathbf{M}$. For an orthonormal $\mathbf{M}$, $H(\mathbf{M})$ is zero and we have, $H(\mathbf{a}) = H(\mathbf{y})$. Under the constraint of, $\mathrm{Trace}(\mathbf{M}^t\mathbf{M}) = N$, $H(\mathbf{M})$ is always negative unless $\mathbf{M}$ is orthonormal.

Channel is linear and has a memory length of M_0 symbols (impulse response of length M_0+1). Without loss of generality, we assume that $M_0 \leq M$. The noise is additive white Gaussian with zero mean and power σ^2, $\mathbf{n}(0,\sigma^2)$. The output of the channel (M-vector $\hat{\mathbf{y}}$) is related by the following relation to the channel input (M-vector $\mathbf{y}$),

$$\hat{\mathbf{y}} = \mathbf{C}\mathbf{y} + \mathbf{C}_0\mathbf{y}_0 + \mathbf{n}(0,\sigma^2), \quad (4)$$

where the $M \times M$ matrix $\mathbf{C}$ is the transfer matrix of the channel, the $M \times M_0$ matrix $\mathbf{C}_0$ reflects the effect of the channel memory, the M_0-vector $\mathbf{y}_0$ is the set of M_0 symbols prior to $\mathbf{y}$, and $\mathbf{n}(0,\sigma^2)$ is the additive noise. The term $\mathbf{C}_0\mathbf{y}_0$ in (4) is the interference from the previous block. This is denoted as the interblock interference. The interference from the symbols within a block, denoted as intrablock interference, is due to the nondiagonal elements of $\mathbf{C}$.

The $N \times M$ demodulator matrix $\mathbf{D}$ is selected such that $\mathbf{DCM} = \mathbf{I}$ where $\mathbf{I}$ is the $N \times N$ identity matrix. The autocorrelation of the noise at the demodulator output is equal to,

$$\hat{\mathbf{R}}_n = \sigma^2\mathbf{D}\mathbf{D}^t + \mathbf{D}\mathbf{C}_0\mathbf{M}_0\mathbf{R}_a\mathbf{M}_0^t\mathbf{C}_0^t\mathbf{D}^t, \quad (5)$$

where the $M_0 \times N$ matrix $\mathbf{M}_0$ is composed of the last M_0 rows of the $\mathbf{M}$ and $\mathbf{R}_a$ is the autocorrelation of $\mathbf{a}$. The whole system is equivalent to an N-dimensional channel with an additive Gaussian noise of autocorrelation $\sigma^2\mathbf{D}\mathbf{D}^t$ and an intersymbol interference (ISI) term of autocorrelation $\mathbf{D}\mathbf{C}_0\mathbf{M}_0\mathbf{R}_a\mathbf{M}_0^t\mathbf{C}_0^t\mathbf{D}^t$.

We assume that the shaping region is obtained by the scaling of a baseline region B which is closed under the sign changings and the permutations of the coordinates. In this case, it is easy to show that $\mathbf{R}_a = \Lambda_a$.

We assume that the decision is made independently along each dimension of the output space. This is equivalent to maximum likelihood decision if $\hat{\mathbf{R}}_n$ given in (5) is diagonal. The equivalent noise power along the i'th dimension, denoted by σ_i^2, $i \in [0, N-1]$, is equal to the i'th diagonal element of $\hat{\mathbf{R}}_n$. To have equal minimum distance to noise ratio along all the dimensions, the coding lattice should be scaled with a factor proportional to σ_i along the i'th dimension.

In a class of the schemes, the last M_0 transmissions of each block are zero. This brings the channel to zero state at the beginning of each block and omits the interblock interference. This is obtained at the price of loosing M_0 dimensions per each M-dimensional block. In this case, the last M_0 rows of $\mathbf{M}$ are equal to zero and the last M_0 columns of $\mathbf{C}$ do not result in any output. As a result, $\mathbf{M}$ and $\mathbf{C}$ are written in $N \times N$ and $M \times N$-dimensional forms, respectively. We refer to this scheme as the zero state block based signaling. In this case, the ISI term in (5) is equal to zero. The optimum modulator for this scheme, which minimizes the product of the noise powers (diagonal elements of $\hat{\mathbf{R}}_n$), is the input eigenvectors of the channel, [2]. This results in a diagonal $\hat{\mathbf{R}}_n$, i.e., uncorrelated (independent) noise along different dimensions.

In this paper, we discuss the selection of the matrices $\mathbf{M}$ and Λ_a. In spectral shaping, the objective is to shape the power spectrum of $\mathbf{y}$. In signaling over a partial response channel, the objective is to maximize the distance to noise ratio of $\hat{\mathbf{a}}$.

3 Spectral shaping

The autocorrelation matrix of a sequence of M-dimensional blockwise uncorrelated vectors $\mathbf{y}^k$ is equal to,

$$\mathbf{R}_y^k = \mathbf{E}\left[\mathbf{y}^0(\mathbf{y}^k)^*\right], \quad (6)$$

where $\mathbf{E}[\,.\,]$ denotes the expectation over the set of vectors $\mathbf{y}^k$, k is the block index and $*$ denotes the Hermitian transpose. The power spectrum of $\mathbf{y}$ is equal to, [3],

$$S_y(\omega) = \frac{1}{M}\sum_{k=-\infty}^{\infty} e^{-jk\omega M}\,\mathbf{v}^*(\omega)\,\mathbf{R}_y^k\,\mathbf{v}(\omega), \quad |\omega| \leq \pi, \quad (7)$$

where,

$$\mathbf{v}(\omega) = \left[e^{-j\omega m},\ m = 0,\ldots,M-1\right]^t. \quad (8)$$

If vectors in different blocks are uncorrelated, we have $\mathbf{R}_y^k = 0$, $k \neq 0$. In this case, we use the notation $\mathbf{R}_y = [R_y(i,j),\ i,j = 0,\ldots,M-1]$ instead of $\mathbf{R}_y^0$. Any autocorrelation matrix is positive–semi–definite (has nonnegative eigenvalues). A positive–definite $\mathbf{R}_y$ (with all the eigenvalues strictly positive) results in a strictly positive $S_y(\omega)$. A spectral null results in a zero eigenvalue. For a real process, the autocorrelation matrix is symmetrical. From now on, we assume real processes.

Define $d_y(k)$ as,

$$d_y(k) = \sum_{|i-j|=k} R_y(i,j). \quad (9)$$

Using this notation for a real, blockwise uncorrelated $\mathbf{y}$, Eq. (7) reduces to,

$$S_y(\omega) = \frac{1}{M}\sum_{k=0}^{M-1} d_y(k)\cos(\omega k). \quad (10)$$

It is seen that the power spectrum of the scheme is completely determined by the autocorrelation matrix but the reverse is not necessarily true.

From (10), it is seen that different autocorrelation matrices with equal sum of the diagonal elements, d_k's, result in the same spectrum. The selection among the available choices is based on maximizing the entropy of the code.

To realize a given autocorrelation matrix, it is enough to use its eigenvectors as the space dimensions and allocate a power equal to the corresponding eigenvalue to each dimension.

4 Spectrum of a modulated process

The autocorrelation matrices of $\mathbf{a}$ and $\mathbf{y}$ in Fig. 1 are related by,

$$\mathbf{R}_y = \mathbf{M}\mathbf{R}_a\mathbf{M}^t . \tag{11}$$

The autocorrelation matrix and the power spectrum of the k'th column of $\mathbf{M}$, $\mathbf{m}_k$, are equal to,

$$\mathbf{R}_k = \mathbf{m}_k\mathbf{m}_k^t , \tag{12}$$

and,

$$S_k(\omega) = \frac{1}{M}\sum_{i=0}^{M-1}\sum_{j=0}^{M-1} m_k(i)m_k(j)\cos[\omega(i-j)] , \tag{13}$$

respectively, where $m_k(i)$ denotes the elements of $\mathbf{m}_k$.

A shaping region which is not a hypercube results in dependency between the distributions along different dimensions. In spite of this, under a set of mild conditions, the distributions are uncorrelated (autocorrelation matrix is diagonal) and the corresponding spectrum is white. Assuming that $\mathbf{M}$ is an orthonormal matrix ($\mathbf{M}\mathbf{M}^t = \mathbf{I}$), $\mathbf{R}_y$ will be diagonal if: (i) Matrices $\mathbf{R}_a$ and $\mathbf{M}$ are both diagonal. (ii) Matrix $\mathbf{R}_a$ is diagonal and has equal diagonal elements. On the other hand: (i) If the channel dimensions are used directly as the constellation dimensions, $\mathbf{M}$ will be diagonal. (ii) If the region $\mathcal{B}$ is closed under sign changings of the coordinates, $\mathbf{R}_a$ will be diagonal. As already mentioned, if $\mathcal{B}$ is closed under the sign changings and the permutations of the coordinates, we have $\mathbf{R}_a = \mathbf{\Lambda}_a$.

For $\mathbf{R}_a = \mathbf{\Lambda}_a$, it is easy to show that,

$$\mathbf{R}_y = \sum_{k=0}^{N-1} \lambda_k \mathbf{R}_k , \tag{14}$$

and,

$$S_y(\omega) = \sum_{k=0}^{N-1} \lambda_k S_k(\omega) . \tag{15}$$

Considering (13) and (15), the spectrum of $\mathbf{y}$ depends only on the space dimensions,, $\mathbf{M}$, and to the power allocated to each dimension, $\mathbf{\Lambda}_a$. The important issue is that the power spectrum is independent of the exact structure of the shaping region.

5 Performance loss of a nonflat spectrum

The price to be paid for a nonflat spectrum is a reduction in the signal space volume. This is measured in terms of the power loss with respect to a reference scheme with a white spectrum. In this case, equating the entropies, the power loss, P_l, is defined as the ratio of the second moments. The reference scheme for the system in Fig. 1 is M-dimensional. For a fixed entropy, the general shape of the spectrum is independent of the structure of $\mathcal{B}$. The structure of $\mathcal{B}$ changes the spectrum (the total power) by a multiplicative factor which is minimum for a sphere and maximum for a hypercube, [1]. Considering this fact, it is equally reasonable to select the shaping region of the reference scheme as spherical or cubic. In the following, we compute the P_l of a nonflat spectrum based on an N-dimensional elliptical/cubic shaping region with respect to an M-dimensional spherical/cubic reference scheme. Normalizing the energy per dimension to one, for the elliptical region we obtain,

$$P_l = \frac{\pi(M+2)\,[\Gamma(0.5N+1)]^{\frac{2}{M}}}{[\pi(N+2)]^{\frac{N}{M}}\,[\Gamma(0.5M+1)]^{\frac{2}{M}}} \times \frac{\sum_{i=0}^{N-1}\lambda_i}{M}\left(\prod_{i=0}^{N-1}\lambda_i\right)^{-\frac{1}{M}} , \tag{16}$$

and for the cubic region, we obtain,

$$P_l = (12)^{1-\frac{N}{M}} \times \frac{\sum_{i=0}^{N-1}\lambda_i}{M}\left(\prod_{i=0}^{N-1}\lambda_i\right)^{-\frac{1}{M}} . \tag{17}$$

For $N = M$, or $N \simeq M$ when N and M are large, and for $\sum_i \lambda_i = M$, (16) and (17) reduce to,

$$P_l = \left(\prod_{i=0}^{N-1}\lambda_i\right)^{-\frac{1}{N}} . \tag{18}$$

For $N \to \infty$, the eigenvectors tend to complex exponentials, $\exp(-j\omega)$, and the eigenvalues tend to the power spectrum, $S_y(\omega)$. In this case, it can be shown that assuming an elliptical region, the average entropy per channel dimension of $\mathbf{y}$ tends to,

$$H_0(\mathbf{y}) = \frac{1}{4\pi}\int_{-\pi}^{\pi} \log\,[2\pi e S_y(\omega)]\;d\omega = \frac{1}{2}\log(2\pi e\lambda) , \tag{19}$$

where,

$$\lambda = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\,[S_y(\omega)]\;d\omega\right\} . \tag{20}$$

To have the same volume with an infinite dimensional spherical region, the required energy per dimension is equal to λ given in (20). Using (18), the asymptotic value of P_l is found as,

$$P_l = \exp\left\{-\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\,[S_y(\omega)]\;d\omega\right\} . \tag{21}$$

This is the reciprocal of the output power of the linear minimum mean square predictor for $\mathbf{y}$. As the distribution along the dimensions are independent Gaussian, the linear minimum mean square predictor is the optimum predictor and (21) is the reciprocal of the innovation power of $\mathbf{y}$. Using (20) and considering that the process $\mathbf{y}$ has the same entropy as given in (19) but with unit energy per dimension results in (21).

6 Spectral shaping using fixed basis

This concerns selecting a fixed $\mathbf{M}$ and using only $\mathbf{\Lambda}_a$ to maximize the entropy of the code. This method is suboptimum. However, by the appropriate selection of $\mathbf{M}$, one can decrease the computational complexity. For a spectrum with spectral nulls, $\mathbf{M}$ is selected as an orthonormal basis with the same nulls. For the case of no spectral null, sine basis is used. First, we discuss the basis with spectral nulls.

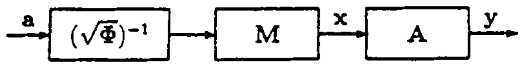

Fig. 2 Modulation with the output eigenvectors using the input eigenvectors.

6.1 Bases with spectral nulls

Consider an $M \times N$ matrix $\mathbf{A}$. The input eigenvectors/eigenvalues of $\mathbf{A}$ are the eigenvectors/eigenvalues of $\mathbf{A}^t\mathbf{A}$. The output eigenvectors/eigenvalues of $\mathbf{A}$ are the eigenvectors/eigenvalues of the $\mathbf{AA}^t$. From the M output eigenvalues, $M-N$ are equal to zero. The N nonzero output eigenvalues are equal to the input eigenvalues. These are denoted by ϕ_i, $i \in [0, N-1]$. The input eigenvector, $\mathbf{m}_i$, and the output eigenvector, $\hat{\mathbf{m}}_i$, corresponding to the same eigenvalue, ϕ_i, satisfy,

$$\begin{aligned} \mathbf{A}\mathbf{m}_i &= \sqrt{\phi_i}\,\hat{\mathbf{m}}_i\,, \\ \mathbf{A}^t\hat{\mathbf{m}}_i &= \sqrt{\phi_i}\,\mathbf{m}_i\,. \end{aligned} \tag{22}$$

As $\mathbf{A}^t\mathbf{A}$ and $\mathbf{AA}^t$ are both symmetrical, the input and the output eigenvectors form an orthonormal basis.

If the system $\mathbf{A}$ has spectral null at certain frequencies, its output eigenvectors form an orthonormal basis with the same nulls.

Considering Eq. (22), to modulate a signal with the output eigenvectors, we can use the cascade combination shown in Fig. 2 where $(\sqrt{\Phi})^{-1} = \mathrm{diag}\,[1/\sqrt{\phi_0}, \dots, 1/\sqrt{\phi_{N-1}}]$. This results in,

$$\mathbf{R}_y = \mathbf{AM}\Phi^{-1}\Lambda_a\mathbf{MA}^t = \hat{\mathbf{M}}\Lambda_a\hat{\mathbf{M}}^t, \tag{23}$$

where $\Phi = \mathrm{diag}\,[\phi_0, \dots, \phi_{N-1}]$.

The $1 \pm D$ and $1 - D^2$ systems are three important examples of the partial response channels, [4]. For a null at zero/Nyquist frequency, $\mathbf{A}$ is taken as $1-D/1+D$ system. The $1 \pm D$ systems has an $(N+1)\times N$ transfer matrix with the i'th column equal to, $[(0)^i, \sqrt{2}/2, \pm\sqrt{2}/2, (0)^{N-1-i}]$. For a null at both zero and Nyquist frequency, $\mathbf{A}$ is taken as $1-D^2$ system. This has an $(N+2)\times N$-dimensional transfer matrix with the i'th column equal to, $[(0)^i, \sqrt{2}/2, 0, -\sqrt{2}/2, (0)^{N-1-i}]$.

6.2 Block-based Eigensystem of the $1 \pm D$ and $1 - D^2$ systems

For the $1-D$ channel, the input/output eigenvectors are equal to, [5],

$$m_k(n) = \sqrt{\frac{2}{N+1}}\,\sin\frac{\pi(k+1)(n+1)}{N+1}, \tag{24}$$

where, $k, n = 0, \dots, N-1$, and,

$$\hat{m}(n) = \sqrt{\frac{2}{N+1}}\,\cos\frac{\pi(k+1)(n+0.5)}{N+1}, \tag{25}$$

where, $n = 0, \dots, N$ and $k = 0, \dots, N-1$, respectively. The corresponding eigenvalues are equal to,

$$\phi_k = 1 - \cos\frac{\pi(k+1)}{(N+1)}. \tag{26}$$

The input and output eigenvectors of $1+D$ channel are obtained by multiplying (24) and (25) with $(-1)^n$. The eigenvalues are the same as the $1-D$ channel given in Eq. (26).

For the $1 \pm D$ channels, the product of the nonzero eigenvalues is equal to, [5],

$$\prod_{k=0}^{N-1}\phi_k = |\mathbf{A}^t\mathbf{A}| = 2^{-N}(N+1)\,. \tag{27}$$

An N-dimensional $1-D^2$ channel, N even, can be considered as two time multiplexed $N/2$-dimensional $1-D$ channels. Consequently, the eigenvalues are in pair equal to,

$$\phi_k = 1 - \cos\frac{\pi(k+1)}{0.5N+1},\ k = 0, \dots, 0.5N-1. \tag{28}$$

The two eigenvectors corresponding to a pair of eigenvalues are of the general form $\alpha_1 m_k(2n) + \alpha_2 m_k(2n+1)$ where $\alpha_1^2 + \alpha_2^2 = 1$ and $m_k(n)$ is the eigenvector of the $1-D$ channel given in (24). For the $1-D^2$ channel, we have,

$$\prod_{k=0}^{N-1}\lambda_k = 2^{-N}(0.5N+1)^2\,. \tag{29}$$

In all the three cases, the matrix $\mathbf{M}$, in Fig. 2 is either sine matrix or closely related to it. This reduces the computational complexity of the modulation by using fast sine transform algorithms.

6.3 Optimization procedure

For a fixed basis, the F_p-constraint is formulated as,

$$\sum_{k=0}^{N-1}\lambda_k B_k(\omega_c) \le F_p\,, \tag{30}$$

where,

$$B_k(\omega_c) = 2\int_0^{\omega_c} S_k(\omega)d\omega\,, \tag{31}$$

and $S_k(\omega)$ is the spectrum of the k'th dimension of $\mathbf{M}$. For spectral nulls and/or F_p-constraint, the energy constraint is always active. This can be verified by considering that increasing the total energy increases the entropy while not affecting the constraints. The optimization procedure is formulated as,

$$\begin{cases} \text{Maximize} & \displaystyle\sum_{k=0}^{N-1}\log(\lambda_k)\,, \\ \text{Subject to:} & \displaystyle\sum_{k=0}^{N-1}\lambda_k B_k(\omega_c) \le F_p\,, \\ & \displaystyle\sum_{k=0}^{N-1}\lambda_k = M\,, \quad \lambda_k \ge 0\,. \end{cases} \tag{32}$$

This is a convex optimization problem.

For $F_p \in [F_{\min}, F_{\max}]$, the F_p-constraint is active. For $F_p < F_{\min}$, the optimization problem has no answer. For $F_p > F_{\max}$, the F_p-constraint is not active and the power-ratio is equal to $F_{\max}$. The $F_{\max}$ can be calculated by relaxing the F_p-constraint and finding the power-ratio of the answer. Without spectral null constraint, this results in a white spectrum and $F_{\max} = \omega_c/\pi$.

Assuming that the F_p-constraint is active and using the Lagrange method, we obtain,

$$\lambda_k = \frac{1}{\psi_1\,B_k(\omega_c) + \psi_2}\,, \tag{33}$$

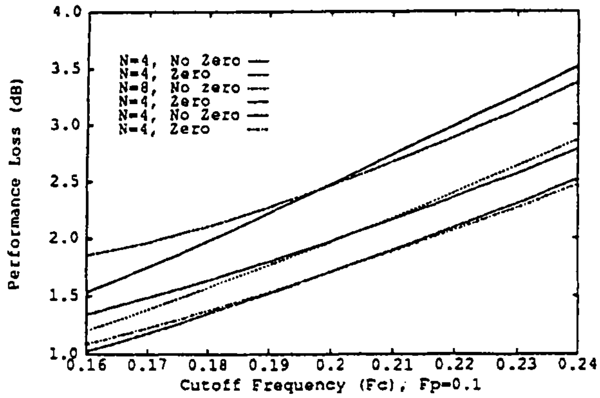

Fig. 3 Performance loss (in dB) as a function of the cutoff frequency, fixed basis, $F_p = 0.1$, with and without spectral null at zero frequency, cubic shaping region, $M = N + 1 = 4, 8, 16$.

where ψ_1 and ψ_2 are determined by solving,

$$\sum_{k=0}^{N-1} \frac{B_k(\omega_c)}{\psi_1 B_k(\omega_c) + \psi_2} = F_p, \quad \text{and} \quad \sum_{k=0}^{N-1} \frac{1}{\psi_1 B_k(\omega_c) + \psi_2} = M. \tag{34}$$

The solution with $\psi_1, \psi_2 \geq 0$ is unique and corresponds to the global optimum point.

Figure 3 shows the P_l as a function of the cutoff frequency using fixed basis, $F_p = 0.1$, with and without spectral null at zero frequency, cubic shaping region, $M = N + 1 = 4, 8, 16$.

In the case that the F_p-constraint is not active, the optimum answer is obtained by allocating equal energy to all the dimensions.

6.4 Spectral null

In the following, we consider the case of a null at zero frequency in more detail. In this case, $\mathbf{A}$ is the $1 - D$ system. In general, by changing the λ_k's while keeping $\sum_k \lambda_k = N+1$, one can tradeoff the width of the null and the entropy of the code. One interesting case in this tradeoff corresponds to,

$$\lambda_k = \frac{N+1}{N} \times \phi_k, \tag{35}$$

where ϕ_k's are the eigenvalues of the $1 - D$ system. This results in,

$$\mathbf{R}_y = \frac{N+1}{N} \times \mathbf{A}\mathbf{A}^t, \tag{36}$$

and,

$$S_y(\omega) = 1 - \cos(\omega). \tag{37}$$

In this case, it can be shown that,

$$\prod_{k=0}^{N-1} \lambda_k = \left(\frac{N+1}{N}\right)^N \times |\mathbf{A}^t\mathbf{A}| = \frac{(N+1)^{N+1}}{N^N} \times 2^{-N}. \tag{38}$$

Using Eqs. (21) and (37) or Eqs. (18) and (38) results in an asymptotic value of 3 dB for P_l.

Another interesting case corresponds to the spectrum with the maximum entropy. This is obtained by equal allocation of energy. In this case,

$$\lambda_k = \frac{N+1}{N}, \quad k \in [0, N-1]. \tag{39}$$

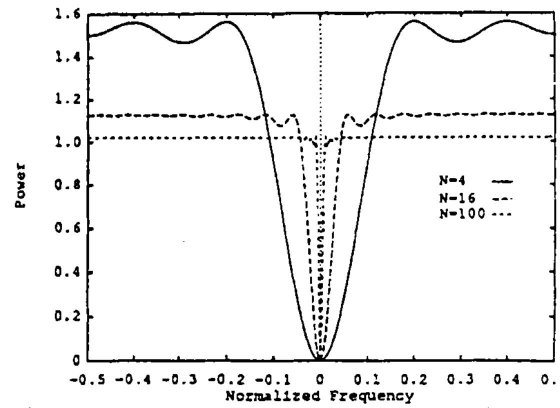

Fig. 4 Spectrum of the maximum entropy (narrowest null width) with spectral null.

Using Eq. (18), the asymptotic value for P_l is zero dB. The power spectrum is equal to,

$$S_y(\omega) = \frac{N+1}{N} \sum_{k=0}^{N-1} S_k(\omega), \tag{40}$$

where $S_k(\omega)$'s are the spectrum of the output eigenvectors of $1 - D$ system. Figure 4 shows the corresponding spectrum for different values of N.

It can be shown that this is the maximum entropy spectrum with a spectral null at zero over dimensionality $N + 1$. It can be also shown that the $F_{\max}$ with spectral null is equal to,

$$F_{\max} = \frac{\omega_c}{\pi} - \frac{2}{\pi} \sum_{i=1}^{N} \frac{N+1-i}{N(N+1)} \sin(\omega_c i)/i. \tag{41}$$

6.5 Example

We consider fixed basis analysis with a spectral null at zero frequency. For $N = 2$, the basis (output eigenvectors of $1 - D$ system) is equal to,

$$\mathbf{M} = \begin{bmatrix} \sqrt{2}/2 & \sqrt{6}/6 \\ 0 & -\sqrt{6}/3 \\ -\sqrt{2}/2 & \sqrt{6}/6 \end{bmatrix}. \tag{42}$$

Within a scale factor, the shape of the power spectrum is determined by the ratio of S_0 and S_1. By changing this ratio, we can tradeoff the width of the null and the entropy of the code. Figure 5 shows the corresponding spectrums. For $S_0 = S_1$, the volume of the region is maximum. From Fig. 5, it is seen that this case is closer to a white spectrum than the other cases.

7 Signaling over partial response channels

We put the restriction of having equal minimum distance to noise ratio along all the dimensions where the noise is the sum of the Gaussian noise and ISI. In general, the statistics of the ISI depends on the source and is not Gaussian. This is the major problem associated with the proposed analysis.

The decision is made independently along each dimension. This is equivalent to maximum likelihood decision if for $\mathbf{R}_a = \mathbf{\Lambda}_a$, the $\hat{\mathbf{R}}_n$ in (5) is diagonal. This will be the

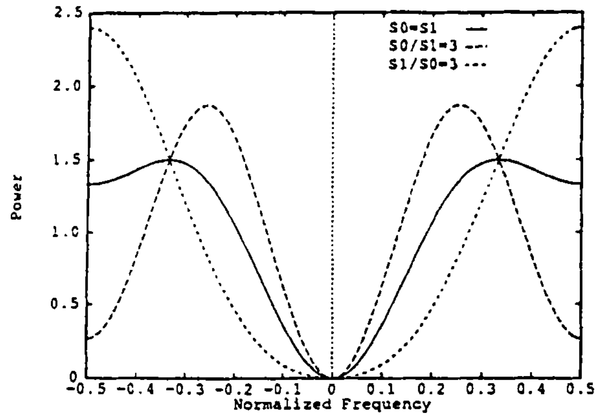

Fig. 5 Power spectrum of the example.

case if $\sigma^2 \mathbf{D}\mathbf{D}^t$ is diagonal and the ISI term in zero (zero state signaling). The equivalent noise power along the i'th dimension, denoted by σ_i^2, $i \in [0, N-1]$, is equal to the i'th diagonal element of $\hat{\mathbf{R}}_n$. To have equal minimum distance to noise ratio along all the dimensions, volume of the Voronoi region around each point of the input constellation should be proportional to $\prod_i \sigma_i$. Define the entropy of the channel as,

$$H(\mathbf{C}) = \sum_{i=0}^{N-1} \log[\hat{R}_n(i,i)]. \tag{43}$$

The entropy of a flat channel is equal to zero. Using this notation, the entropy of $\hat{\mathbf{a}}$ is equal to,

$$H(\hat{\mathbf{a}}) = H(\mathbf{a}) + H(\mathbf{M}) - H(\mathbf{C}), \tag{44}$$

where $H(\mathbf{a})$, $H(\mathbf{M})$ are given in Eqs. (1) and (3), respectively. The optimum N, $\mathbf{M}$ and Λ_a are found by solving the following optimization problem,

$$\begin{cases} \text{Maximize} & H(\hat{\mathbf{a}}), \\ \text{Subject to:} & \text{Trace}(\mathbf{M}\Lambda_a\mathbf{M}^t) = M, \end{cases} \tag{45}$$

where, $\text{Trace}(\mathbf{M}\Lambda_a\mathbf{M}^t)$ is the total energy of $\mathbf{y}$. This results in a spectrum for $\mathbf{y}$ which is matched to the channel characteristics. In this case, unlike to the case of the spectral shaping, $\mathbf{M}$ is not necessarily orthonormal. Optimization over N can be achieved by decreasing N (starting from M) until the optimum value is found.

In the zero state signaling, the σ_i^2's are equal to the diagonal elements of $\sigma^2\mathbf{D}\mathbf{D}^t$. Under the constraint of $\text{Trace}(\mathbf{M}\Lambda_a\mathbf{M}^t) = M$, $H(\mathbf{a}) + H(\mathbf{M})$ is maximized when the λ_i's are all equal and $\mathbf{M}$ is orthonormal. It can be shown that $H(\mathbf{C})$ is minimized when $\mathbf{M}$ is the set of the N input eigenvectors of the channel corresponding to the largest eigenvalues. Similar to the previous case, optimization over N can be achieved by a search. This omits the intrablock interference resulting in uncorrelated (independent) noise along different dimensions. By allocating equal energy to the input eigenvectors of the channel corresponding to the largest eigenvalues, both of these terms are simultaneously optimized. These results are consistent with those obtained in [2]. In this case, the dimensions are used in an on-off manner in the sense that a dimension has either zero energy or an amount of energy equal to other dimensions. If the rate is high enough such that all the dimensions are all nonempty, i.e., $N = M$, this scheme, independent of the channel structure, results in a white spectrum for $\mathbf{y}$.

Remarks: This work was mainly based on using the continuous approximation. Using this approximation, rate and energy distributions can be treated independently. In practice, rate is a discrete quantity. More importantly, in most of the practically interesting boundaries, the rate per two dimensional subspaces is restricted to be an integer. This imposes more restrictions on the optimization problem. In this case, the rate and the energy distribution should be jointly selected to maximize the total rate. The procedure will be quite similar to the one given in [6].

8 Summary and conclusions

We have studied the selection of a constellation for spectral shaping. This is achieved by the use of an optimization procedure which maximizes the rate of the constellation subject to some constraints on its power spectrum. The constellation basis is fixed and is selected to reduce the computational complexity of the modulation. The power allocated to each dimension is optimized. We also discussed how to match the power spectrum of a line code to a specific channel characteristics. The procedure ws based on assuming an effective additive noise composed of the sum of the Gaussian noise and the intersymbol interference. In this case, the optimization procedure maximizes the rate of the constellation subject to having equal minimum distance to noise ratio along all the dimensions.

References

[1] A. K. Khandani and P. Kabal, "Unsymmetrical boundary shaping in multi-dimensional spaces," *Proc. Twenty-ninth Annual Allerton Conference on Communications, Control, and Computing,* Allerton, IL, pp. 798–808, Oct 1991.

[2] S. Kasturia, J. T. Aslanis, and J. M. Cioffi, "Vector Coding for Partial Response Channels," *IEEE Trans. Inform. Theory,* vol. IT-36, pp. 741–762, July 1989.

[3] G. L. Cariolaro and G. P. Tronca, "Spectra of block coded digital signals,". *IEEE Trans. on Commun.,* vol. COM-22, pp. 1555–1563, Oct 1974

[4] P. Kabal and S. Pasupathy, "Partial-Response signaling," *IEEE Trans. Commun.,* vol. 23, pp. 921–934, Sept 1975.

[5] A. K. Khandani and P. Kabal, "Optimum block-based signaling over $1 \pm D$ and $1 - D^2$ partial-response channels," *Proc. Twenty-ninth Annual Allerton Conference on Communications, Control, and Computing,* Allerton, IL, pp. 779–789, Oct 1991.

[6] A. K. Khandani, P. Kabal and H. Leib, "Combined coding and shaping over a multitone channel," *Proc. IEEE Global Telecommunications Conference,* Phoenix, Arizona, pp. 1182–1186, Dec 1991,