HAL open science

A Kashin Approach to the Capacity of the Discrete Amplitude Constrained Gaussian Channel

Brendan Farrell, Peter Jung

▶ To cite this version:

Brendan Farrell, Peter Jung. A Kashin Approach to the Capacity of the Discrete Amplitude Constrained Gaussian Channel. SAMPTA'09, May 2009, Marseille, France. Special Session on Sampling and Communication. hal-00451439

HAL Id: hal-00451439
https://hal.science/hal-00451439

Submitted on 29 Jan 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Kashin Approach to the Capacity of the Discrete Amplitude Constrained Gaussian Channel

Brendan Farrell [(1)] and Peter Jung [(2)]

(1) Heinrich-Hertz Lehrstuhl, Technische Universität Berlin, Einsteinufer 25, 10587 Berlin, Germany.
(2) Fraunhofer German-Sino Lab for Mobile Communications - MCI, Einsteinufer 37, 10587 Berlin, Germany.
brendan.farrell@mk.tu-berlin.de, peter.jung@hhi.fhg.de

Abstract:

We derive an explicit lower bound on the capacity of the discrete amplitude–constrained Gaussian channel by proving the existence of tight frames that permit redundant vector representations with small coefficients. Our method encodes the information in subspaces that are optimal in terms of the power to amplitude ratio. In a recent paper, Lyubarskii and Vershynin discuss how the work of Kashin (1977) implies the existence of such representations, and they term them Kashin respresentations. We use this work from frame theory to address the relationship between signal redundancy, peak–to–average power ratio and achievable data rates.

1. Introduction

Communication at high data rates and with moderate cost on hardware and complexity provide challenging topics in engineering and applied mathematics. An important problem in this direction is efficient signaling and coding under an amplitude constraint. In general, the cost for high data rate is related to a power budget. However, in practical communication systems, there sometimes exist disruptive or non-linear effects that only occur at high signal amplitudes. The information–theoretic treatment of amplitude–constrained channel is completely different from the power–constrained channel. On the other hand, coding for power–constrained Gaussian channels is well understood. Clearly, if a loss in data rate is accepted, signals can be constructed with lower maximum amplitude. The optimal scaling between power and amplitude and an explicit relation to achievable rates will be given in this paper. In this case, the data-rate loss is caused by considering redundant representations. Here, the original vectors are expanded with respect to a particular frame and the coefficients are then transmitted.

We show that there exist frames which allow the standard coding approach to be used for the amplitude-constrained channel. Our result is Theorem 2, which comes at the end of the paper. This theorem states that for the amplitude constrained, Gaussian channel the rate

$$\frac{1}{2\lambda_{\min}} \log\left(1+\lambda_{\min}\frac{\text{Signal Power}}{\text{Noise Power}}\right) \tag{1}$$

is achievable for a redundancy $\lambda_{\min}$ that is an *explicit function* of the peak-to-average power ratio. We note that by making the amplitude constraint compatible with Gaussian codebooks, we make the developed tools and understanding of Gaussian codebooks applicable to the amplitude–constrained channel. Results from frame theory, thus, allow us to address a question in information theory. While the results used from functional analysis are well known there, we show a new application.

1.1 The Information–Theoretic Problem

The capacity of a communication channel is the maximum amount of information per unit of time that can be sent from a sender through the channel to the receiver. Shannon made this operational concept mathematically rigorous by formulating it in terms of entropy [7]. In [7] Shannon addressed the discrete–time model:

$$Y = X + Z, \tag{2}$$

for the noisy channel, where X and Y denote the (real) channel input and output, and the additive noise Z is a Gaussian random variable with variance σ^2. Let X^n be a random vector in $\mathbb{R}^n$ according to a distribution to be determined and Z^n the random vector having n identical independent distributed (iid) copies of Z. Shannon introduced two concepts of a capacity for this model. The *information capacity* $C^{(i)}$ is the supremum of the information rates:

$$C^{(i)} = \lim_{n\to\infty}\frac{1}{n}\sup_{\mu^n\in\mathcal{F}^n}\mathcal{I}(X^n;Y^n) \tag{3}$$

taken over all distributions μ^n of X^n from a particular subset $\mathcal{F}^n \subset \mathcal{P}^n$ of probability distributions $\mathcal{P}^n$. $\mathcal{I}(X^n;Y^n)$ denotes the mutual information between the random variables X^n and Y^n and is equal to the entropy of Y^n minus the entropy of Y^n given X^n, $\mathcal{I}(X^n;Y^n) = h(Y^n) - h(Y^n|X^n)$. From its concavity in μ^n it follows that the optimum μ^n_{opt} is at least achieved for a product distribution, i.e. single letter coding with a measure $\mu = \mu^1$ is optimal in this sense. Shannon considered an averaged power constraint P which corresponds to the set $\mathcal{F} = \mathcal{F}^1$ of single–letter distributions:

$$\mathcal{F} = \{\mu\in\mathcal{P}\mid \int |x|^2 d\mu(x) \le P\} \tag{4}$$

or equivalently

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}|x_i|^2 \le P. \tag{5}$$

He found that the optimum μ_{opt} is attained for a Gaussian distribution with variance P and that

$$C^{(i)} = \frac{1}{2}\log(1+\frac{P}{\sigma^2}). \quad (6)$$

Shannon further showed with a so called *coding theorem* that it is even possible to get arbitrary close to that value justifying the term *channel capacity*. That is, for each rate $R < C^{(i)}$ there exist 2^{nR} codewords $\{X(\omega)\}_{\omega=1}^{\omega=2^{nR}}$ in $\mathbb{R}^n$ (called a $(2^{nR}, n)$ code) such that $X(\omega) + Z^n$ can be distinguished at the receiver with error probability going to zero as n increases. (X will now denote codewords and be indexed by ω.) Each admissible codeword satisfies the average power constraint $\frac{1}{n}\sum_{i=1}^n |X_i(\omega)|^2 \leq P$; however, to achieve the capacity it may be necessary to use codewords having maximum amplitudes which scale with $\sqrt{n}$.

We address an additive, white Gaussian noise (AWGN) channel under the assumption that there is both a power constraint,

$$\frac{1}{n}\sum_{i=1}^{n} |X_i(\omega)|^2 \leq P, \quad (7)$$

and a strict amplitude constraint:

$$\max_{i=1,\dots,n} |X_i(\omega)| \leq A, \quad (8)$$

for two positive, real numbers P and A and for all $\omega = 1, ..., 2^{nR}$.

The information capacity under a constraint A on the amplitudes of the signals was solved by Smith [8]. Similar to the Gaussian channel with power constraint only, Smith showed that the capacity of the amplitude–constrained channel is attained when the entries x_i are independent.

The set of (single–letter) input distributions is in this case:

$$\mathcal{F} = \{\mu \in \mathcal{P} \,|\, \mu(\{|x| > A\}) = 0\}. \quad (9)$$

Smith found that the optimum measure μ_{opt} has discrete and finite support. Similar results are known for other noise densities (see for example [6]). A characterization of the number of mass points in the Gaussian case is unknown. For a given assumption on this number the values and the positions can be computed. From this Smith gave an algorithm which numerically computes $C^{(i)}$. Smith establishes an algorithm to determine the optimal input probability measure given the constraints A, P and σ^2. However, to date there is not a general strategy applicable for a practical range of these parameters.

1.2 Frames and Banach Geometry

We will work strictly with real numbers. We have the following norms for $\mathbb{R}^n$: $\|x\|_{l_p^n} = (\sum_{i=1}^n |x_i|^p)^{1/p}$ and $\|x\|_{l_\infty^n} = \max_{i=1,\dots,n} |x_i|$. B_p^n will denote the unit ball in $\mathbb{R}^n$ with respect to the ℓ_p-norm. We denote by U_n^N an n-dimensional subspace of $\mathbb{R}^N$, $N \geq n$. We will often speak of a matrix $U \in \mathbb{R}^{n\times N}$ whose rows are orthonormal and span U_n^N or whose columns constitute a tight frame for $\mathbb{R}^n$.

Definition 1. A set of vectors $\{u_i\}_{i=1}^N \subset \mathbb{R}^n$ is a *tight frame* for $\mathbb{R}^n$ if

$$\|x\|_2^2 = \sum_{i=1}^{N} = |\langle x, u_i\rangle|^2 \quad (10)$$

for all $x \in \mathbb{R}^n$.

It follows that the columns of an $n\times N$ matrix U constitute a tight frame for $\mathbb{R}^n$ if and only if $UU^* = I_n$, where I_n denotes the identity matrix of size n. In the proof of the coding theorem (see, for example, [1]) for the Gaussian channel with average power constraint P, the constructed codewords $X \in \mathbb{R}^n$ satisfy the constraint $\|X\|_{\ell_2^n} \leq \sqrt{nP}$. Similarly, in the amplitude constrained channel codewords must satisfy $\|X\|_{\ell_\infty^n} \leq A$. In other words, admissible signals X for the amplitude constraint channel lie in a scaled cube, i.e. $X \in A \cdot B_\infty^n$. And for a power constrained channel the signals are contained in an increasing ball $X \in \sqrt{nP} \cdot B_2^n$.

Of course the difficult aspect of this channel is the amplitude constraint. We do not require that the random input variables $\{x_i\}_{i=1}^n$ be independent, which allows us to use redundant representations.

The basic idea for our approach is the following: given N vectors $\{u_i\}_{i=1}^N$ spanning $\mathbb{R}^n$, $N > n$, a vector $x \in \mathbb{R}^n$ may be expressed, in general, in multiple ways as a linear combination of the vectors $\{u_i\}_{i=1}^N$:

$$x = \sum_{i=1}^{N} b_i u_i. \quad (11)$$

In light of the amplitude constraint, the question is whether one of the possible expressions (10) satisfies $\|b\|_{l_\infty^N} \leq A$. If this is possible, then we may transmit the vector b and suffer an efficiency loss of $N - n$ symbols.

The representation (10) is called a *Kashin representation* [5] of the vector x if $\|b\|_\infty \leq C\|x\|_2$. We first address a general frame setting and then focus on the Kashin representations in Section 3.

2. General Frame Setting

As we have seen, the capacity of the discrete Gaussian channel with average power constraint P and noise variance σ^2 is $\frac{1}{2}\log(1+\frac{P}{\sigma^2})$. This means, If $R < \frac{1}{2}\log(1+\frac{P}{\sigma^2})$ is the rate, then there are 2^{nR} codewords, and all admissible codewords for this channel satisfy the power constraint $\|X(\omega)\|_2 \leq \sqrt{nP}$, $\omega = 1, ..., 2^{nR}$. If one has a tight frame $\{u_i\}_{i=1}^N$ for $\mathbb{R}^n$, $N = [\lambda n]$, then one can also achieve the rate:

$$\frac{1}{2\lambda}\log(1+\frac{\lambda P}{\sigma^2}) \quad (12)$$

by transmitting codewords $\{Y(\omega)\}_{\omega=1}^{2^{nR}} \subset \mathbb{R}^N$ satisfying $UY(\omega) = X(\omega)$ for $\omega = 1, ..., 2^{nR}$.

Since columns of $U \in \mathbb{C}^{n\times N}$ form a tight frame for $\mathbb{R}^n$, $\|UX(\omega)\|_{l_N^2} = \|Y(\omega)\|_{l_n^2}$, and thus:

$$\frac{1}{N}\sum_{i=1}^{N} |Y_i(\omega)|^2 = \frac{1}{\lambda n}\sum_{i=1}^{n} |X_i(\omega)|^2 \leq P. \quad (13)$$

The key point is that a vector $Y(\omega)$ that satisfies $UY(\omega) = X(\omega)$ is, in general, not unique. For a given additional constraint, one may ask if there exists a set $\mathcal{Y} \subset \mathbb{R}^N$ satisfying the additional constraint and a tight frame with matrix U such that:

$$U\mathcal{Y} = \{x | x \in \mathbb{R}^n,\ \|x\|_2 = 1\}. \tag{14}$$

The existence of such a set and a corresponding tight frame is sufficient to imply that $\frac{1}{2\lambda}\log(1 + \frac{\lambda P}{\sigma^2})$ is an achievable rate for the discrete Gaussian channel with the additional constraint.
The additional constraint of interest here is the amplitude constraint; that is, it is required that $\|Y(\omega)\|_{l_\infty^N} \leq A$ for all codewords $Y(\omega)$. Thus, for a given codebook $\{X(\omega)\}_{\omega=1}^{2^{nR}}$ satisfying $\|X(\omega)\|_{l_2^n} \leq \sqrt{nP}$ for all ω, we would like to determine a second codebook $\{Y(\omega)\}_{\omega=1}^{2^{nR}} \subset \mathbb{R}^N$ satisfying $\|Y(\omega)\|_{l_\infty^N} \leq A$ and a tight frame so that $UY(\omega) = X(\omega)$ for all ω. For completeness and clarity, we include the communication strategy. The next section will show that Step 2 is possible for an appropriate λ.

Communication Strategy:

1. The set of vectors $\{u_i\}_{i=1}^N$ form a tight frame for $\mathbb{R}^n$ and are known to both transmitter and receiver.
2. Each codeword $X(\omega)$ satisfies the power constraint , and its Kashin representation $Y(\omega) \in \mathbb{R}^N$ satisfying $\|Y(\omega)\|_{l_\infty^N} \leq A$ is determined.
3. To transmit the message ω, the transmitter sends $Y(\omega)$.
4. $Y(\omega) + Z^N \in \mathbb{R}^N$ is received.
5. Receiver multiplies $Y(\omega) + Z^N$ by U to obtain $X(\omega) + UZ^N \in \mathbb{R}^n$.
6. Receiver decodes $X(\omega) + UZ^N \in \mathbb{R}^n$.

We note that, in contrast to the approach of Smith [8], this approach is still based on Gaussian codebooks, and, therefore, the extensive tools developed for Gaussian codebooks are still applicable.

3. Kashin Representations or Optimal Subspaces

Definition 2 (Kashin Representations)**.** For a set of vectors $\{u_i\}_{i=1}^N \subset \mathbb{R}^n$, $N > n$, the expansion

$$x = \sum_{i=1}^{N} a_i u_i \tag{15}$$

is a *Kashin representation with level* K of the vector $x \in \mathbb{R}^n$ if

$$\|a\|_{l_\infty^N} \leq \frac{K\|x\|_{l_2^n}}{\sqrt{N}}, \quad i = 1, ..., N. \tag{16}$$

See [3, 4, 5]. We denote by U the $n \times N$ dimensional matrix with columns $\{u_i\}_{i=1}^N$. If these vectors constitute a tight frame, then $UU^* = I_n$, where I_n denotes the identity matrix on $\mathbb{R}^n$. One possible coefficient vector for equation (14) is $a = U^*x$. For this vector, we note

$$\begin{aligned} \|a\|_{l_\infty^N} &\leq \|a\|_{l_2^N} = \sqrt{\langle U^*x, U^*x\rangle} \quad (17)\\ &= \sqrt{\langle I_n x, x\rangle} = \|x\|_{l_2^n}. \quad (18)\end{aligned}$$

Consequently, for a tight frame, it is always possible to find a vector a satisfying $\|a\|_{l_\infty^N} \leq \|x\|_{l_2^n}$, and thus equation (15) can be satisfied for every tight frame with Kashin level $K = \sqrt{N}$.
Of course the study of Kashin representations is concerned with optimally small constants and their relation to the redundancy $\lambda = N/n$. We will be interested in the dependence of $K = K(\lambda)$ on λ, but we postpone the discussion of the constant $K(\lambda)$ until the next section. Now, we show a lower bound on the achievable capacity when the amplitude constraint is $K(\lambda)\sqrt{P}$ (or greater).
If we set any n orthonormal vectors in $\mathbb{R}^N$ to be the rows of a matrix U, then $UU^* = I_n$, and the columns of U constitute a tight frame for $\mathbb{R}^n$. Thus, a tight frame for $\mathbb{R}^n$ can be constructed from any n-dimensional subspace of $\mathbb{R}^N$. For $U \in \mathbb{C}^{n\times N}$, let U_n^N denote the subspace of $\mathbb{R}^N$ spanned by its rows. Then $U(B_2^N \cap U_n^N) = B_2^n$. Therefore, for any $x \in B_2^n$, as long as the rows of U are linearly independent there exists a $y \in (B_2^N \cap U_n^N)$ such that $x = Uy$. In the higher dimensional space, we have an $\|\cdot\|_{l_\infty^N}$-norm constraint. We thus want to find an n-dimensional subspace of $\mathbb{R}^N$ that can be mapped isometrically with respect to the $\|\cdot\|_{l_2^n}$-norm to $\mathbb{R}^n$, and we must be able to cover B_2^n in this way.
First results on the smallest constant C, such that a projection of the ball $C \cdot B_\infty^N$ covers B_2^n was given by Kashin in [3]. There he showed that the scaling is $\mathcal{O}(n^{-1/2})$, and the exact optimal scaling was then determined in [2]. Since the $\|\cdot\|_2$-isometric projection is equivalent to the existence of a tight frame, we formulate their result in terms of frames.

Theorem 1 ([3, 2])**.** *For all positive integers N and n, $N > n$, there exists a tight frame for $\mathbb{R}^n$ consisting of N vectors such that every vector in $\mathbb{R}^n$ has a Kashin representation of level:*

$$K(\lambda) := C\left(\frac{\lambda}{\lambda-1}\log\left(1+\frac{\lambda}{\lambda-1}\right)\right)^{1/2}, \tag{19}$$

where $\lambda = N/n$ with respect to this frame.

See also [4, 5] for further discussion of this result. In [5] Lyubarskii and Vershynin have recently given an algorithm for determining a Kashin representation. In the same paper they discuss various ways to generate the required frames and determine their Kashin constants.

Theorem 2. *For a given amplitude constraint A, there exists a constant $\lambda_{\min}$ such that the capacity $\mathcal{C}_{P,A}$ of the discrete Gaussian channel with average power constraint P, amplitude constraint A and noise variance σ^2 is lower bounded by*

$$\mathcal{C}_{P,A} \geq \frac{1}{2\lambda_{\min}}\log\left(1+\frac{\lambda_{\min}P}{\sigma^2}\right). \tag{20}$$

Proof Theorem 1 shows the existence of a frame with the necessary properties, as discussed in the communication strategy in Section 2. Denoting the matrix corresponding to this frame by U, for each codeword $X(\omega) \in \mathbb{R}^n$, there exists a codeword $Y(\omega) \in \mathbb{R}^N$ such that $X(\omega) = UY(\omega)$, and

$$\begin{aligned} \|Y(\omega)\|_{l_\infty^N} &\leq \frac{K(\lambda_{\min})}{\sqrt{N}} \|X(\omega)\|_{l_2^n} && (21) \\ &\leq K(\lambda_{\min})\sqrt{P}. && (22) \end{aligned}$$

Lastly, $\lambda_{\min}$ is the solution to

$$C\left(\frac{\lambda}{\lambda-1}\log(1+\frac{\lambda}{\lambda-1})\right)^{1/2} = \frac{A}{\sqrt{P}}, \quad (23)$$

which exists and is unique since $\left(\frac{\lambda}{\lambda-1}\log(1+\frac{\lambda}{\lambda-1})\right)^{1/2}$ is monotone increasing. ❑

4. Conclusion

We have considered an application of the redundant representations found in frame theory and geometric functional analysis to a fundamental question in information theory.

References:

[1] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, New York, 1991.

[2] A. Garnaev and E. D. Gluskin. The widths of euclidean balls. *Doklady An. SSSR.*, 277:1048–1052, 1984.

[3] B. S. Kashin. Diameters of some finite-dimensional sets and classes of smooth functions. *Izv. Akad. Nauk SSSR Ser. Mat.*, 41(2):334–351, 478, 1977. English transl. in Math. USSR IZV. 11 (1978), 317-333.

[4] B. S. Kashin and V. N. Temlyakov. A remark on compressed sensing. *Mathematical Notes*, 82(5):748–755, Nov 2007.

[5] Y Lyubarskii and R. Vershynin. Uncertainty principles and vector quantization. preprint.

[6] W. Oettli. Capacity-achieving input distributions for some amplitude-limited channels with additive noise (corresp.). *IEEE Transactions on Information Theory*, 20(3):372–374, May 1974.

[7] C.E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423,623–656, 1948.

[8] Joel G. Smith. The Information Capacity of Amplitude and Variance Constrained Scalar Gaussian Channels. *Information and Control*, 18:203–219, 1971.