6.454 Fall 2004
MMSE estimation and lattice encoding/decoding

Todd P. Coleman

September 22, 2004

1 Notations

Throughout this discussion we will abbreviate 'independent and identically distributed' as 'i.i.d.' and denote 'X is Gaussian with mean m and variance σ^2' by $X \sim \mathcal{N}(m, \sigma^2)$.

2 The AWGN channel

The AWGN channel and coding for it in an intuitive sense has been known since Shannon. The AWGN channel model is

$$Y = X + N$$

where $N \sim \mathcal{N}(0, \sigma_N^2)$ is independent of the transmitted signal and the transmitted codeword must satisfy the power constraint $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \leq P_X$. Shannon showed that the capacity of the Gaussian channel with power constraint P_X and noise variance σ_N^2 is

$$C = \frac{1}{2}\log_2(1 + SNR)$$

where $SNR = \frac{P_X}{\sigma_N^2}$.

Shannon showed that in the limit of using long block length n, generating 2^{nC} i.i.d. $\mathcal{N}(0, P_X)$ codewords and averaging across all codebooks is a capacity-achieving strategy: with high likelihood the power constraint will be satisfied and the average probability of error under ML decoding tends to 0 as $n \to \infty$.

This result can be derived geometrically as well. With very high likelihood, from the law of large numbers, a length-n i.i.d. $\mathcal{N}(0, \sigma^2)$ vector lies on the boundary shell of an n-sphere of radius $\sqrt{n\sigma^2}$. Since we've generated $X \sim \mathcal{N}(0, P_X)$ and $N \sim \mathcal{N}(0, \sigma_N^2)$, the received signal is $\mathcal{N}(0, P_X + \sigma_N^2)$ and thus with very high likelihood lies on the boundary shell of an n-sphere of radius $\sqrt{n(P_X + \sigma_N^2)}$. So we can choose codewords as the centers of spheres of radius $\sqrt{n\sigma_N^2}$, choose the decoding region of each codeword to be that particular sphere, and observe that if all the codeword spheres are non-overlapping, then the probability of error will tend to 0 as $n \to 0$. The question becomes, how many such disjoint spheres can we pack and meet our power constraint? See Figure 1.

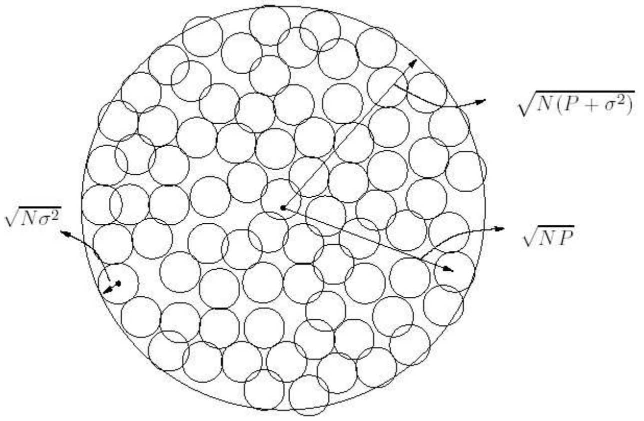

Figure 1: Sphere-packing for the AWGN channel.

The volume of an n-sphere with radius r is of the form $A_n r^n$. Since the decoding spheres are of radius $\sqrt{n\sigma_N^2}$ and the received vector lies in an n-sphere of radius $\sqrt{n(P_X+\sigma_N^2)}$, the maximum number of non-overlapping decoding spheres is given by

$$\frac{A_n\left[n(P_X+\sigma_N^2)\right]^{\frac{n}{2}}}{A_n\left[n\sigma_N^2\right]^{\frac{n}{2}}} = 2^{\frac{n}{2}\log_2\left(1+\frac{P_X}{\sigma_N^2}\right)} = 2^{nC}$$

Researchers for decades have been interested in constructing *structured* codes, encoding mechanisms, and decoding mechanisms that can achieve capacity on the AWGN channel for *arbitrary* SNRs. We will devote our attention in this discussion to codes that are based on lattices, which are algebraic in nature. Specifically, we will discuss how Uri Erez and Ram Zamir solved the decades-old problem of achieving the AWGN channel capacity at all SNRs, using lattice codes and lattice decoding. Surprisingly, (and non-intuitively at first glance), using a *biased* MMSE estimator is essential to achieve capacity. This is also related to the recently shown deep connection between mutual information and estimation theory (which will be discussed in more depth by Baris in a couple of weeks).

3 Introduction to Lattices

- A lattice Λ is a discrete group which is a subset of $\mathbb{R}^n$. As in the case of binary linear codes, a lattice can be described in terms of a generator matrix.

  $$\Lambda = \{\lambda = Gx : x \in \mathbb{Z}^n\}$$

  where G is an n by n real matrix. Note that for any lattice, $\underline{0} \in \Lambda$.

- A fundamental Voronoi region of Λ is denoted by $\mathcal{V}$ and is the set of all $x \in \mathbb{R}^n$ such that

  $$\|x-\underline{0}\| \leq \|x-\lambda\| \;\; \forall\, \lambda \in \Lambda,$$

where ties are broken consistently in such a way that every $x \in \mathbb{R}^n$ can be expressed as

$$x = \lambda + r$$

where $\lambda \in \Lambda$ and $r \in \mathcal{V}$. Thus, the Voronoi region can be thought of as the remainder in modular arithmetic. We denote

$$\lambda = Q_{\mathcal{V}}(x)$$

as the nearest neighbor of x in Λ and

$$r = x \bmod_{\mathcal{V}} \Lambda \ = x - Q_{\mathcal{V}}(x)$$

as the error. In some cases, when considering the modulo operation with the fundamental Voronoi region, we will drop the $\mathcal{V}$ subscript and simply write

$$r = x \bmod \Lambda \, .$$

- More generally, we can consider any fundamental region and quantizer. A fundamental region Ω of Λ has the property that every $x \in \mathbb{R}^n$ can be uniquely expressed as $x = \lambda + e$ where $\lambda \in \Lambda$ and $e \in \Omega$. The quantizer and modulo-Λ operations associated with Ω are given by

$$\begin{aligned} Q_{\Omega}(x) &= \lambda \text{ if } x \in \lambda + \Omega \\ x \bmod_{\Omega} \Lambda &= x - Q_{\Omega}(x) \, . \end{aligned}$$

- For any region $\mathcal{R}$, we denote its volume as $V(\mathcal{R})$.
- For any $\mathcal{R} \subset \mathbb{R}^n$, we define its second moment per dimension to be

$$P(\mathcal{R}) = \frac{1}{n} \frac{\int_{\mathcal{R}} \|x\|^2 dx}{V(\mathcal{R})},$$

which is the average energy per dimension of a uniform probability distribution over $\mathcal{R}$.

- The normalized second moment of a region $\mathcal{R}$ is given by

$$G(\mathcal{R}) = \frac{P(\mathcal{R})}{V(\mathcal{R})^{\frac{2}{n}}}$$

- Let us denote the n-sphere with radius $\sqrt{n\sigma^2}$ as S_{n,σ^2}. As $n \to \infty$ it is known that

$$\begin{aligned} V(S_{n,\sigma^2})^{\frac{2}{n}} &\to 2\pi e \sigma^2 \\ P(S_{n,\sigma^2}) &\to \sigma^2 \\ \Rightarrow G(S_{n,\sigma^2}) &\to \frac{1}{2\pi e} \end{aligned}$$

We denote a sphere of volume V as S_V. Since a sphere has the minimal moment of inertia of all shapes with equal volume, it follows that for any lattice,

$$G(\mathcal{V}) \geq G(S_{V(\mathcal{V})}) \to \frac{1}{2\pi e}$$

- We say that high-dimensional lattices have Voronoi regions that are quasi-spherical if $G(\mathcal{V}) \to \frac{1}{2\pi e}$. It is well-known that such lattices exist. They are termed 'good for quantization' or 'good for shaping', and we will denote those wich such a desired property as Λ_S.
- We also note for any $0 < \epsilon < \sigma^2$, the probability that an i.i.d. $\mathcal{N}(0, \sigma^2 - \epsilon)$ n-tuple falls outside S_{n,σ^2} can be made arbitrarily small.
- There also exist high-dimensional lattices whose Voronoi regions are quasi-sphereical in the above sense: For any $\epsilon > 0$, for all $n > n_0(\epsilon)$:

  $$S_n < \frac{V(\mathcal{V})^{\frac{2}{n}}}{2\pi e} \Rightarrow P([X \sim \mathcal{N}(0, S_n)] \notin \mathcal{V}) < \epsilon.$$

  Such lattices are termed '*sphere-bound achieving*' or '*good for AWGN channel coding*' and we will denote those with such a desired property as Λ_C.
- A *lattice code* $\mathcal{C}$ is simply the intersection of a lattice with a bounded shaping region $\mathcal{S}$:

  $$C = \Lambda_C \cap \mathcal{S}.$$

  The shaping region $\mathcal{S}$ helps to impose any signaling constraint on the communication problem (for instance the input power constraint for the AWGN channel).
- A lattice decoder is simply a quantizer $Q_{\Omega_C}(x)$ with respect to some fundamental region Ω_C of Λ_C. Namely, a lattice decoder for a lattice Λ_C takes as input $y \in \mathbb{R}^n$ and for some fundamental region Ω_C of Λ_C (not necessarily the fundamental Voronoi region $\mathcal{V}_C$ of Λ_C), it performs the operation

  $$\lambda = Q_{\Omega_C}(y) \in \Lambda_C.$$

  It is important to note that when applied to a received signal across a channel whose input was a lattice code symbol, a lattice decoder does not take into account the shaping region $\mathcal{S}$ associated with the lattice code, which simplifies the decoding process.

4 Spherical lattice codes

It has been known that a spherical lattice code (where $\mathcal{S}$ is a thin spherical shell) with a second moment P_X can achieve the AWGN channel capacity.

4.1 Spherical Lattice codes with ML decoding

It has been shown by numerous researchers that provided that Λ_C is 'good for channel coding', then with ML decoding, capacity can be achieved. Because of the intersection of Λ with the spherical shell, ML decoding requires finding the lattice point *inside the shell* that is closest to the received signal. This does not correspond to true lattice decoding and the decoding regions associated with any codeword lose structure. Furthermore, almost all of the signal points in a spherical lattice code lie near the boundary of the sphere, where Euclidean minimum-distance lattice decoding (or any form of lattice decoding for that matter) significantly differs from ML decoding.

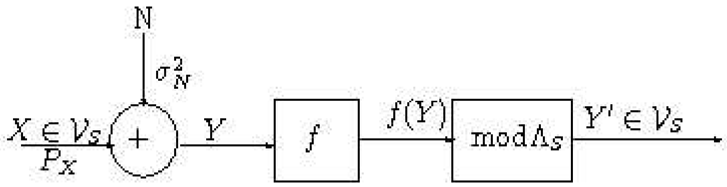

Figure 2: Canonical model of a $\bmod \Lambda_S$ channel.

4.2 Spherical lattice codes with lattice decoding

It has also been shown in the research literature that if we replace the ML-decoder, which differs significantly from a lattice decoder, with a more structured lattice decoder, then the maximum achievable rate is $\frac{1}{2}\log_2(SNR)$. So at high SNR, this essentially achieves capacity, but it exhibits significant performance loss at low SNR. We will discuss later why this decoding strategy cannot fully attain the $\frac{1}{2}\log_2(1+SNR)$.

5 Mod-Lattice Transmission and Lattice Decoding

We now continue to consider lattice codes Λ_C, but the shaping region $\mathcal{S}$ is not a spherical shell but rather is a fundamental Voronoi region $\mathcal{V}_S$ of another lattice Λ_S. We now show that for the canonical model given in Figure 2, the AWGN channel capacity can be achieved [1], [2], [3]:

Theorem 5.1. *Consider the encoding/decoding mechanism given in Figure 2. If Λ_S is 'good for shaping' ($G(\mathcal{V}_S) \to \frac{1}{2\pi e}$), the transmitted signal X is uniformly distributed over $\mathcal{V}_S$, and f is an MMSE estimator of X, then the rate $\frac{1}{2}\log_2(1+SNR)$ is achievable.*

Note that this differs from Section 4 in that i) the shaping region $\mathcal{S}$ is now the Voronoi region of a 'good for shaping' lattice Λ_S, and ii) there is no 'good for channel coding' lattice Λ_C; the channel inputs are simply uniformly distributed over Λ_S. We will resolve ii) shortly, but for now let us go ahead with the outline of the proof.

Proof:

Introduce a *dither* random variable U that is uniformly distributed over Λ_S and known to both the encoder and decoder. Given any data vector $C \in \mathcal{V}_S$, the channel input is

$$X = C + U \bmod_{\mathcal{V}_S} \Lambda_S \, .$$

This makes X uniformly distributed over $\mathcal{V}_S$, *and* independent of C. This is because $P_U(u)$ is constant over any $u \in \mathcal{V}_S$, and as x runs through $\mathcal{V}_S$, $x - c \bmod_{\mathcal{V}_S} \Lambda_S$ runs through $\mathcal{V}_S$, so $P_{X|C}(x|c) = P_U(x - c \bmod_{\mathcal{V}_S} \Lambda_S)$ is constant for any $x \in \mathcal{V}_S$ and $c \in \mathcal{V}_S$.

The dither contributes two nice things: first of all, it allows the transmitted vector to be uniformly distributed across $\mathcal{V}_S$ and thus the power constraint is met with equality. More importantly, it allows the transmitted vector (and thus the received vector after the mod operation) to be independent of the data vector $C \in \mathcal{V}_S$. See Figure 3.

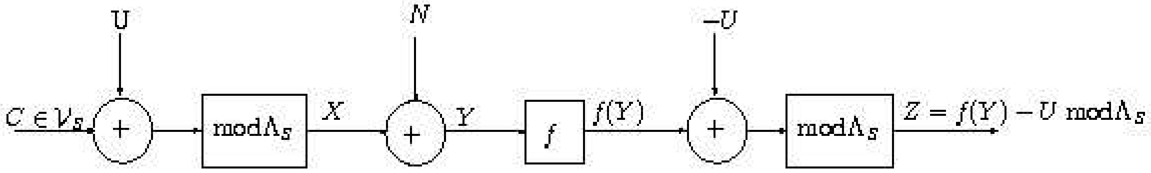

Figure 3: Using a mod Λ_S channel along with dither U.

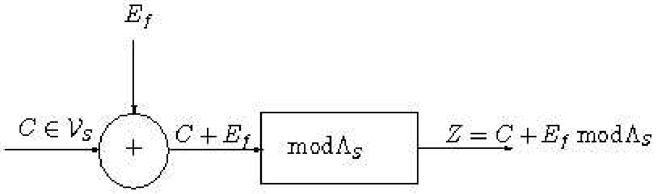

Figure 4: Equivalent mod Λ_S channel model.

Since X is independent of C, and since $Y = X + N$, it follows that the estimation error

$$E_f = f(Y) - X \tag{1}$$

is also independent of any codeword C. Thus we now have an additive noise channel (see Figure 4) whose input and output alphabets are a fundamental Voronoi region $\mathcal{V}_S$ of a lattice Λ_S, such that $Z = C + E_f \bmod \Lambda_S$ and E_f is independent of C. We note that a uniform input is capacity-achieving and in this case the output is also uniform.

Thus

$$\begin{aligned}
C \geq C(\Lambda_S, f) &= \frac{1}{N}\left[h(Z) - h(Z|C)\right] \\
&= \frac{1}{N}\left[\log_2 V(\Lambda_S) - h(E'_f)\right] && (2) \\
&= \frac{1}{2}\log_2 2\pi e P_X - \frac{1}{2}\log_2 2\pi e G(\mathcal{V}_S) - \frac{1}{N}h(E'_f) \\
&\geq \frac{1}{2}\log_2 2\pi e P_X - \frac{1}{2}\log_2 2\pi e G(\mathcal{V}_S) - \frac{1}{N}h(E_f) && (3) \\
&\geq \frac{1}{2}\log_2 2\pi e P_X - \frac{1}{2}\log_2 2\pi e G(\mathcal{V}_S) - \frac{1}{2}\log_2 2\pi e P_{E_f} && (4) \\
&= \frac{1}{2}\log_2 \frac{P_X}{P_{E_f}} - \frac{1}{2}\log_2 2\pi e G(\mathcal{V}_S).
\end{aligned}$$

where (2) follows by defining $E'_f = E_f \bmod \Lambda_S$, (3) follows from Lemma A.1, and (4) follows from the entropy-power inequality where P_{E_f} is the average energy per dimension of E_f. Now suppose that Λ_S is 'good for shaping'. Then we have that $\log_2 2\pi e G(\mathcal{V}_S) \to 0$ and thus

$$C \geq C(\Lambda_S, f) \;\geq\; \frac{1}{2}\log_2 \frac{P_X}{P_{E_f}}. \tag{5}$$

Finally, suppose we let $f(Y) = \alpha Y$. Then

$$\begin{aligned} E_f &= \alpha Y - X = \alpha N - (1-\alpha)X \\ \Rightarrow P_{E_f} &= \alpha^2 \sigma_N^2 + (1-\alpha)^2 P_X \end{aligned}$$

By choosing f to be a MMSE estimator, we minimize P_{E_f} and obtain

$$\begin{aligned} \alpha^* &= \frac{P_X}{P_X + \sigma_N^2} = \frac{SNR}{1+SNR} \\ \Rightarrow P_{E_f}^* &= \frac{P_X \sigma_N^2}{P_X + \sigma_N^2} \\ \Rightarrow C(\Lambda_S, f^*) &= \frac{1}{2}\log_2(1+SNR) \end{aligned}$$

5.1 Comments on dither and MMSE scaling

Note that the dither U is used in a non-symmetric way. At the encoder it is simply added to the codeword and then the $\bmod \Lambda_S$ operation is performed. However, at the decoder, a scalar multiplies the received signal and then the dither is subtracted out after the $\bmod \Lambda_S$ operation. This is somewhat un-intuitive at first glance, since the dither contributes to part of the equivalent noise E_f. Moreover, we note that when we think of α^* in terms of estimation, we note that since $E_f = \alpha Y - X$, and $\alpha^* \neq 1$, the optimal estimator is *biased.* Thus, the only way to achieve capacity using lattice decoding is to first perform biased estimation.

Traditional ways of using a mod-lattice system (without MMSE scaling) in the research literature have only been able to attain achievable rates of $\frac{1}{2}\log_2(SNR)$, which amounts to letting $\alpha = 1$ in the previous section. The MMSE scaling operation $f(Y) = \alpha^* Y = \frac{SNR}{SNR+1} Y$ minimizes the variance of E_f and increases the 'effective SNR' by a factor of $\frac{SNR+1}{SNR}$. For this reason, this is sometimes called an 'inflated lattice decoder' because relative to the noise, the decoding regions for the lattice are larger. This accounts for the jump from $\frac{1}{2}\log_2(SNR)$ to $\frac{1}{2}\log_2(1+SNR)$.

5.2 Nested lattices (Voronoi codes)

Note that in the previous section we discussed shaping lattices Λ_S and required them to be 'good for shaping'. We discussed maximizing mutual information and spoke of a data vector $C \in \mathcal{V}_S$ but did not go into more detail. Here Erez and Zamir consider using structured coding schemes that use the dither and MMSE estimation as in the previous section, along with more constructive ways to signal a codeword $C \in \mathcal{V}_S$. They choose the shaping lattice Λ_S to be 'good for shaping' as in the previous section and choose it to be a sublattice of a fine lattice Λ_C that is 'good for channel coding'. Define

$$\mathcal{C} = \{\Lambda_C \bmod \Lambda_S\} = \{\Lambda_C \cap \mathcal{V}_S\}$$

The rate is given by

$$R = \frac{1}{n}\log_2|\mathcal{C}| = \frac{1}{n}\log_2\frac{V(\mathcal{V}_S)}{V(\mathcal{V}_C)}$$

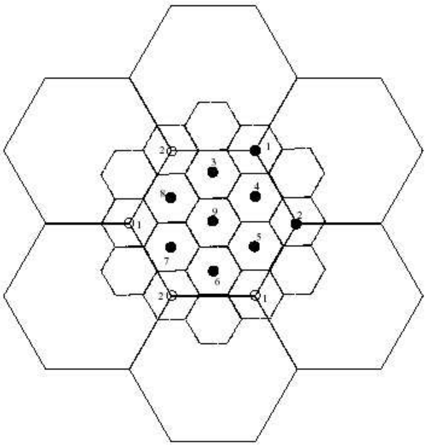

Figure 5: Example of a pair of nested lattices.

Figure 5 gives an example of what a pair of nested lattices looks like.

Erez and Zamir have shown using clever random coding techniques that *for all SNRs*, there exist nested lattices (Λ_C, Λ_S) where $\Lambda_S \subset \Lambda_C$, Λ_C is 'good for channel coding', and Λ_S is 'good for shaping'.

It is also important to note that under the mod-Λ_S transformation, on the equivalent channel, ML decoding coincides with lattice decoding. Thus once constrained to this type of signaling, there is no penalty paid by performing structured lattice decoding. However, the fundamental region associated with the ML decoder's quantizer

$$\Omega_C^* = \{e : f_{E_f}(e) \geq f_{E_f}(e - c \bmod \Lambda_S\ ) \;\forall\; c \in \mathcal{C}\}$$

is not that of a Euclidean minimum-distance lattice decoder.

Erez and Zamir also show that one can in fact use a Euclidean minimum-distance lattice decoder (i.e. replace $Q_{\Omega_C^*}(x)$ with $Q_{\mathcal{V}_C}(x)$) and still achieve capacity.

6 More on MMSE, bias, inflated lattice

One of the most interesting aspects of these results (in my opinion) is that if one is to use a lattice decoder for lattice transmission, then scaling the output by a factor $\alpha^* = \frac{SNR}{SNR+1}$ is necessary. Let us compare MMSE scaling and lattice decoding to simply non-scaled lattice decoding. Note that in high dimensions the Gaussian noise lies *on the surface* of a sphere around the transmitted point. If non-scaled lattice decoding works, then correct decoding only occurs when the noise falls within the Voronoi region. If Λ_C is 'good for channel coding', then this corresponds to the noise falling within the spherical shell corresponding to the power constraint. Thus $\frac{1}{2}\log_2 SNR$ can only be attained.

Using an inflated lattice decoder, correct decoding only occurs when the noise falls within the Voronoi region of the *inflated lattice* (see Figure 6),which corresponds to a larger spherical shell

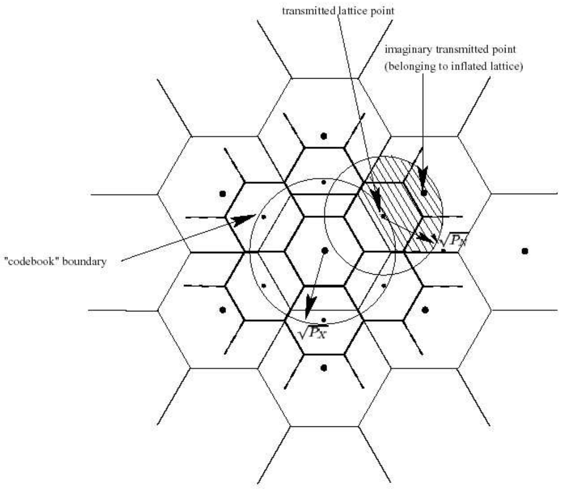
transmitted lattice point
imaginary transmitted point
(belonging to inflated lattice)
"codebook" boundary
$\sqrt{P_X}$
$\sqrt{P_X}$

Figure 6: Inflated lattice.

(if the code is 'good for channel coding') than the original one whose radius corresponds to the power constraint. Thus the spherical shell that the noise must lie in for correct decoding can in fact be larger than the original spherical shell of radius corresponding to the power constraint. It follows that we cannot inflate the lattice too much because as α tends to 0, the output signal αY will have very small variance and thus will lie in a very small sphere. If this is the case then with probability one it will lie inside the Voronoi region of Λ_C, which means that the 0 codeword is always the output of the decoder, and thus $P(\text{error}) \rightarrow 1$. So it makes sense that scaling Y by some α between 0 and 1 is necessary if a mod-Λ_S operation is to follow. At first glance, just based on geometry, it would make sense to at least force the scaled version of Y, $\tilde{Y} = \alpha Y$, to have the same volume as Λ_S. However, the correct scaling coefficient in this case would be $\tilde{\alpha} = \sqrt{\frac{SNR}{SNR+1}}$ rather than $\frac{SNR}{SNR+1}$. So this does not appear to be the correct way to look at the scaling operation. However, one thing we can gain from that geometric argument is that by scaling by $\alpha^* < \tilde{\alpha}$, we are guaranteeing that with high probability, from the law of large numbers, there is no loss of information from the $\alpha^* Y \rightarrow \alpha^* Y \bmod \Lambda_S$ transformation.

7 Extensions: Costa Precoding, Wyner-Ziv, etc.

The Erez/Zamir nested lattice construction can also address multiterminal information theory problems involving Gaussian distributions. For instance, consider the Costa 'Writing on Dirty Paper' problem, the AWGN channel w/ encoder side information. The channel model is

$$Y = S + X + N$$

where $N \sim \mathcal{N}(0, \sigma_N^2)$, X must satisfy a power constraint P_X, and S is an arbitrary additive interference signal known to the encoder. Costa showed that the capacity of this channel is the same as a channel where S is not present. The capacity of this channel can be shown in a *constructive* and *trivial* manner using the Erez/Zamir construction: apply all the arguments previously with the extra twist that the channel input becomes $X = C + U - \alpha^* S \bmod \Lambda_S$ instead. Let the receiver perform the same operation as in the previous section, and it will cancel out S entirely (without knowledge of it). The remaining signal is just as in the AWGN channel case and all the arguments still hold. Thus the capacity is $\frac{1}{2}\log_2(1 + SNR)$.

Also, Erez and Zamir show [2] how these same constructive nested codes with dithering and MMSE scaling achieve the rate-distortion bound for the Wyner-Ziv quantization problem.

Error exponent analysis for nested lattice codes with lattice decoding is discussed in [4]. It is shown that $\alpha = \alpha^*$ maximizes the error exponent only as $R \to C$. At lower rates, α^* is strictly suboptimal. By choosing α correctly, however, the random coding exponent can be achieved at all rates.

The lattice encoding/decoding with MMSE estimation has recently been applied to address coherent communication over MIMO flat fading channels [5]. It has been shown that a class of lattice codes generalized for the MIMO setting under lattice decoding achieve the optimal optimal diversity-vs-multiplexing tradeoff, defined by Zheng and Tse.

References

[1] U. Erez and R. Zamir, "Achieving 0.5 log(1+SNR) on the AWGN channel with lattice encoding and decoding," *IEEE Transactions on Information Theory*, Oct 2004.

[2] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured muliterminal binning," *IEEE Transactions on Information Theory*, vol. 48, no. 6, 2002.

[3] G. D. Forney, "On the role of MMSE estimation in approaching the information-theoretic limits of linear Gaussian channels: Shannon meets Wiener," *Allerton Conference on Communications, Control and Computing*, 2003.

[4] T. Liu, P. Moulin, and R. Koetter, "On error exponents of nested lattice codes for the AWGN channel," *IEEE Transactions on Information Theory*, 2004, Submitted for publication.

[5] H. El Gamal, G. Caire, and M. O. Damen, "Lattice coding and decoding achieve the optimal diversity-vs-multiplexing tradeoff of MIMO channels," *IEEE Transactions on Information Theory*, 2004, To appear.

A Appendix

Lemma A.1. *The random variables $E_f \in \mathbb{R}^n$ defined in (1) and $E'_f = E_f \bmod \Lambda_S$ satisfy $h(E'_f) \leq h(E_f)$.*

Proof. Denote the probability densities of E_f and E'_f as $f_{E_f}(e)$ and $f_{E'_f}(e)$, respectively. Let us denote G_S as the generator associated with Λ_S. By the definition of the $\bmod \Lambda_S$ operation, we have that for any $\underline{e} \in \mathcal{V}_S$,

$$f_{E'_f}(\underline{e}) = \sum_{\underline{x} \in \mathbb{Z}^n} f_{E_f}(\underline{e} + G_S \underline{x}).$$

Note that we can express $h(E_f)$ as

$$\begin{aligned} h(E_f) &= -\int_{\underline{e}\in\mathbb{R}^n} f_{E_f}\left(\underline{e}\right)\log_2 f_{E_f}\left(\underline{e}\right) \\ &= -\sum_{\underline{x}\in\mathbb{Z}^n}\int_{\underline{e}\in\mathcal{V}_S} f_{E_f}\left(\underline{e}+G_S\underline{x}\right)\log_2 f_{E_f}\left(\underline{e}+G_S\underline{x}\right). \end{aligned}$$

We now discuss $h(E'_f)$:

$$\begin{aligned} h(E'_f) &= -\int_{\underline{e}\in\mathcal{V}_S} f_{E'_f}\left(\underline{e}\right)\log_2 f_{E'_f}\left(\underline{e}\right) \\ &= -\int_{\underline{e}\in\mathcal{V}_S}\left[\sum_{\underline{x}\in\mathbb{Z}^n} f_{E_f}\left(\underline{e}+G_S\underline{x}\right)\right]\log_2\left[\sum_{\underline{y}\in\mathbb{Z}^n} f_{E_f}\left(\underline{e}+G_S\underline{y}\right)\right] \\ &= -\sum_{\underline{x}\in\mathbb{Z}^n}\int_{\underline{e}\in\mathcal{V}_S} f_{E_f}\left(\underline{e}+G_S\underline{x}\right)\log_2\left[\sum_{\underline{y}\in\mathbb{Z}^n} f_{E_f}\left(\underline{e}+G_S\underline{y}\right)\right] \\ &\le -\sum_{\underline{x}\in\mathbb{Z}^n}\int_{\underline{e}\in\mathcal{V}_S} f_{E_f}\left(\underline{e}+G_S\underline{x}\right)\log_2\left[f_{E_f}\left(\underline{e}+G_S\underline{x}\right)\right] \qquad (6) \\ &= h(E_f) \end{aligned}$$

where (6) follows from the monotonicity of the log function along with the non-negativity of pdfs. □