The Modulo-Lattice Channel: The Key Feature in Precoding Schemes

Robert F.H. Fischer

Abstract Precoding for transmission over channels with interference known at the transmitter is reviewed. In contrast to what is usually discussed in literature, precoding employing higher-dimensional lattice quantization for generating the transmit signal is assessed. It is shown that thereby the gap between low-complexity schemes using scalar modulo reductions and theoretical asymptotic results assuming hypothetical optimum quantization operations can be bridged. In this paper, capacity curves obtained from numerical calculations, and approximations to the capacity are provided for various lattices. Moreover, the role of appropriate scaling at transmitter and receiver is discussed in detail.

Keywords Precoding schemes, modulo channel, interference channel, lattice quantization, channel capacity

1. Introduction

Precoding is an attractive strategy when transmitting over channels which suffer from additive noise and an additional interference term which is known at the transmitter side [9]. Over the last years, Tomlinson-Harashima precoding [21, 16], initially introduced for intersymbol-interference channels, has been extended to the equalization of MIMO channels, e.g., wireless multiple antenna systems [10, 11] or interference avoidance in digital subscriber lines transmission [14]. Key point in these precoding schemes is the use of modulo arithmetic. Classically, this modulo operation is done for one- or two-dimensional symbols.

Beside this practical implementation, the capacity of channels with interference known at the transmitter has been studied extensively, see e.g., [5, 8, 2, 24] and the references therein. It was shown that "writing-on-dirty-paper approaches" are able to approach the capacity of the AWGN channel, as if the interference were not present. However, these theoretical results assume operations on blocks whose lengths go to infinity and the use of optimum lattice quantizers with dimension approaching infinity.

In this paper we review the concept of precoding for transmission over channels with interference known at the transmitter in a non-causal fashion. In particular we discuss the role of the interference term and assess different optimization criteria (zero-forcing, minimum mean-squared error, and minimum error entropy approach; biased versus unbiased solution in case of finite interference variance), only partially known in literature, in a unified way. Shaping loss, modulo loss, and precoding loss of the strategies are considered. Moreover, contrary to scalar quantization usually discussed in literature, the potential gain due to the use of lattice quantizers whose dimensionality is manageable in practical schemes is determined. Thereby, the gap between low-complexity schemes and theoretical asymptotic results is bridged. The effectiveness of modulo-lattice reduction is studied by capacity curves obtained from numerical calculations, and approximations to the capacity are provided. The important role of appropriate scaling at transmitter and receiver is discussed in detail.

In Section 2 the system model is introduced. The end-to-end capacity when using precoding is studied in Section 3 and its potential increase due to an MMSE approach is discussed in Section 4. In Section 5 modifications in the case of finite interference variance are discussed and Section 6 draws some conclusions.

2. System Model

The System Model

The mod-$\Lambda$ channel considered in this paper is depicted in Fig. 1. For ease of exposition, without loss of generality, all signals are assumed to be real-valued. Complex symbols, e.g., signals in the equivalent complex baseband domain [17], are regarded as pairs of real-valued (one-dimensional) ones.

![Modulo-Lattice Channel](image)

Data symbols $a$ are transmitted over an additive white Gaussian noise (AWGN) channel. The variance (per dimension) of the zero-mean noise is denoted by $\sigma_n^2 \triangleq \mathbb{E}\{n^2\}$. At the receiver, (blocks of) symbols $y$ are modulo reduced with respect to a given lattice $\Lambda$ to produce (blocks of) symbols $u$. Given a block (vector) $x = [x_1, \ldots, x_N]$ of length $N$ ($N$: dimensionality of the lattice), and a quantization operation $Q(x)$, which results in the lattice point $\lambda$ closest to $x$ (with respect to squared Euclidean distance), the modulo operation is defined as

$$\text{mod}_\Lambda(x) \triangleq x - Q(x) \in \mathcal{R}_\Lambda.$$  

$$\mathcal{R}_\Lambda \triangleq \{ x | \|x\|^2 \leq \|x-\lambda\|^2, \forall \lambda \in \Lambda \} \{\|\cdot\|^2: \text{squared Euclidean norm}\} \text{denotes the Voronoi region of the lattice}$$

1 For an introduction to lattices, i.e., infinite sets of points which have the algebraic structure of a group under ordinary vector addition, see, e.g., [9, Appendix C].
addition to the channel noise $n$, an interference term $f$ is active, which is known (or can be calculated) at the transmitter. In transmission over ISI channels this term is the sum of post-cursors of prior transmitted symbols. In multi-antenna scenarios, this interference is due to data streams transmitted in parallel using antenna arrays.

A popular transmission strategy is to subtract the interference term $f$ from the data symbols $a$ and apply a modulo reduction to the difference signal in order to generate the channel symbols $x$. This precoding operation can be viewed as using effective data symbols taken from an extended signal obtained by shifting the initial signal set by lattice points $\Lambda$ [9]. Classically, modulo operations based on one- or two-dimensional lattices $\Lambda$ are used, i.e., they work either with scalar signals or on two consecutive symbols, usually real and imaginary part of signals in the equivalent complex baseband. Building blocks of length $N$, the generalization to arbitrary dimensions is straightforward. This, however, requires the interference term $f$ to be known in a non-causal fashion; the next $N$ interference samples $f$ have to be given in order to generate the next $N$ channel symbols $x$. In multi-antenna or multi-access schemes this can be achieved by successive off-line generation of the parallel transmit signals. For ISI channels, the use of higher-dimensional lattices is not as straightforward; here, clever interleaving techniques (cf., e.g., [24]) are a possible solution.

The channel symbols $x$ are transmitted over the channel with interference and noise. Using $x = \text{mod}_\Lambda(a - f)$ and because of the properties of the modulo operation, the received signal is given by

$$u = \text{mod}_\Lambda(x + f + n) = \text{mod}_\Lambda(\text{mod}_\Lambda(a - f) + f + n) = \text{mod}_\Lambda(a - f + f + n) = \text{mod}_\Lambda(a + n),$$

i.e., the interference term vanishes and the mod-$\Lambda$ channel of Fig. 1 is present.

Further examples, where mod-$\Lambda$ channels occur, are decoding of multilevel codes [22] or multilevel coset code constructions [12]. Here, the influence of other coding levels are eliminated by the modulo frontend, i.e., the term $f$ is some kind of self-interference. Moreover, in digital watermarking, an embedding scheme equivalent to that of Fig. 2 using the one-dimensional lattice $\mathbb{Z}$, called “scalar Costa scheme,” has become popular [7]. The interference seen by the watermark message is the host signal, into which the message is to be woven.

3. Capacity

In the first step we are interested in the end-to-end capacity of the communication scheme according to Fig. 2. In [12] it was proven that the capacity of the mod-$\Lambda$ channel is achieved if the vectors $a$ are uniformly distributed over the Voronoi region $\mathcal{R}_\Lambda$, i.e., $f_a(a) = \mathcal{U}_{\mathcal{R}_\Lambda}(a)$, with $\mathcal{U}_{\mathcal{R}_\Lambda}(x) \equiv 1/V(\Lambda), x \in \mathcal{R}_\Lambda$, zero else. It then calculates (in bits per dimension) to

$$C_\Lambda(\sigma_n^2) = \frac{1}{N}(\log_2(V(\Lambda)) - h(\tilde{n})), \quad (3)$$

where $V(\Lambda) = \int_{\mathcal{R}_\Lambda} d\mathbf{r}$ is the volume of the lattice [4] and $h(\cdot)$ denotes differential entropy. The active noise term $\tilde{n}$ is a $\Lambda$-aliased version of the initial noise vector $n$, and its pdf reads

$$f_{\tilde{n}}(\tilde{n}) = \begin{cases} \frac{2}{\left(2\pi\sigma_n^2\right)^{d/2}} e^{-\|\tilde{n}-\Lambda\|^2/(2\sigma_n^2)}, & \tilde{n} \in \mathcal{R}_\Lambda, \\ 0, & \tilde{n} \notin \mathcal{R}_\Lambda. \end{cases}$$

(4)

Since up to now such capacity curves are missing in literature (except the curves given in [12]), we now numerically evaluate the capacity of the mod-$\Lambda$ channel for various finite-dimensional lattices. Table 1 summarizes the considered lattices and their most important parameters (the values are taken from [4]).

<table>
<thead>
<tr>
<th>Name</th>
<th>$N$</th>
<th>$G(\Lambda)$</th>
<th>$G_s$ [dB]</th>
<th>$\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathbb{Z}$</td>
<td>1</td>
<td>0.0833</td>
<td>0.000</td>
<td>2</td>
</tr>
<tr>
<td>$A_2$</td>
<td>2</td>
<td>0.0802</td>
<td>0.167</td>
<td>6</td>
</tr>
<tr>
<td>$D_4$</td>
<td>4</td>
<td>0.0766</td>
<td>0.364</td>
<td>24</td>
</tr>
<tr>
<td>$E_6$</td>
<td>8</td>
<td>0.0717</td>
<td>0.652</td>
<td>240</td>
</tr>
<tr>
<td>$A_{16}$</td>
<td>16</td>
<td>0.0682</td>
<td>0.867</td>
<td>4320</td>
</tr>
<tr>
<td>$A_{24}$</td>
<td>24</td>
<td>0.0657</td>
<td>1.030</td>
<td>196560</td>
</tr>
</tbody>
</table>

Table 1. Lattices and their important parameters (the values are taken from [4]).

The capacity curves for the mod-$\Lambda$ channel are numerically evaluated using Monte-Carlo integration of the differential entropy in (3). For that, the closest-point search
algorithm of [1] is extended to provide all lattice points which have a relevant contribution to the aliased pdf for a given \( \tilde{n} \in \mathcal{R}_A \).

The numerical results are displayed in Fig. 3 over the signal-to-noise ratio (SNR) \( \sigma_r^2/\sigma_n^2 \) in dB. Additionally, the AWGN capacity curve and the asymptotic limit for the mod-A channel derived in [12] is plotted. As the normalized second moment of the lattice approaches \( 1/(2n^2) \), that of an hypersphere of infinity dimension (sphere-bound-achieving lattice), the capacity reads

\[
C_{\infty}(\sigma_r^2/\sigma_n^2) = \max \left\{ 0, \frac{1}{2} \log_2(\sigma_r^2/\sigma_n^2) \right\}.
\]

Using the \( \mathbb{Z} \) lattice, i.e., the simplest modulo-reduction strategy, at high SNR the well-known *shaping loss* of 1.53 dB [13, 9] compared to AWGN channel capacity is visible. This gap can be bridged using higher-dimensional lattices. In the high-SNR region, the normalized second moment of the lattice, which determines average transmit power \( \sigma_r^2 \), is decisive. As \( G(\Lambda) \) approaches its lower limit, the capacity curves converge to that of the AWGN channel. Noteworthy, the one-dimensional projection of a uniform distribution over the Voronoi region of “good” lattices tends to a Gaussian pdf as \( N \) increases [13]. Hence, the pdf of the transmit symbols \( x \) tends to a Gaussian one, as required for approaching the AWGN channel capacity.

However, in the low-SNR region no convergence towards the AWGN capacity is visible. This *modulation loss* is caused by the modulo frontend at the receiver. A zoom into the region around 0 dB SNR (\( \sigma_n^2 \approx \sigma_r^2 \)) shows that here going to higher dimensions does not provide additional gains but causes a loss in capacity. Asymptotically, in the region below \( 10 \log_{10}(\sigma_r^2/\sigma_n^2) = 0 \) dB no transmission is possible. This effect is caused by the exploding number of neighboring signal points which contribute significant portions to the mod-A-aliased noise pdf \( f_{a,n}(\tilde{n}) \).

Geometrically, at 0 dB SNR the noise space in high-dimensional space with squared radius \( N\sigma_r^2 \) equals the Voronoi region—hence the decision region of the points—with squared radius \( N\sigma_r^2 \). Increasing \( \sigma_r^2 \) leads to unavoidable decision errors; no reliable transmission is possible and the capacity tends to zero. In summary, in the low-SNR region, basing the mod-A operation in the transmission scheme of Fig. 2 on higher-dimensional lattices is not rewarding.

Finally, in the present situation we do not encounter a *precoding loss* since \( \alpha \) is already assumed to be continuously distributed over \( \mathcal{R}_A \). The precoding loss quantifies the increase in average transmit power of the continuous distributed channel symbols \( x \) compared to the power required for transmission of data symbols \( a \) drawn from a finite signal constellation [9].

4. Minimum Error Entropy Approach

In [8] it was shown that modifying the precoding scheme in Fig. 2 slightly, the AWGN capacity can be approached asymptotically; cf. also the similar approach in digital watermarking [19, 20, 6, 7] and the original work of Costa [5]. The scheme is a particular realization of the “*writing-on-dirty-paper* principle” dating back to Costa [5]. Main point is not to pre-subtract the known interference term completely, but to use it constructively.

The possible gain in capacity is achieved by observing the following property: Assuring the variance of \( f \) is very large, for arbitrary \( a \in \mathcal{R}_A \) the pdf of \( x \) will be almost uniform over the Voronoi region \( \mathcal{R}_A \), an effect which is well-known in lattice quantization [15]. Then, the joint pdf of \( a \) and \( x \) can be written using the chain rule as

\[
f_{a,x}(a, x) = f_a(a) \cdot f_{x|a}(x|a) = f_a(a) \cdot f_x(x).
\]

Hence, the channel symbols \( x \) are statistically independent of the data symbols \( a \) to be communicated.\(^2\) Only after addition of the interference at the channel, i.e., with the knowledge of \( x \) and \( f \), data can be recovered. Hence we can argue that part of the information is carried via the interference term leading to a potential gain in capacity.

It is worth noting that this situation is similar to encryptions schemes employing “one-time pads” [18]. Optimal, the mutual information between plain text and cipher text should be zero, which is the case for statistical independence. Only if cipher text and key are known, the message can be decrypted.

Fig. 4 shows the modified precoding transmission scheme. The interference term \( f \) is scaled by \( \alpha \in [0, 1] \)

![Image](attachment:image.png)

**Fig. 4. Transmission scheme using improved mod-Λ precoding.**

and pre-subtracted from the data symbols \( a \). Via the mod-Λ reduction the transmit symbols \( x \) are obtained. The receive signal \( y \) is first scaled by the same parameter \( \alpha \) and then modulo reduced to the decision symbols \( u \). Since \( \alpha \leq 1 \), \( y \) is effectively processed using an enlarged version of \( \Lambda \), hence this scheme is sometimes called *inflated lattice precoding* [8, 24]. In contrast to the preceding section, where the interference is completely eliminated by choosing \( \alpha = 1 \) (zero-forcing (ZF) approach), this scheme employs some kind of minimum mean-squared error (MMSE) approach.

However, this strategy is only possible if the interference term can explicitly be scaled at the transmitter. E.g., for multilevel codes, where the interference is due to higher coding levels which are fixed for a given mapping, no improvement of the capacity can be obtained.

\(^2\) If \( \alpha \) is uniformly distributed over the Voronoi region \( \mathcal{R}_A \), as it is required for approaching capacity, then \( x \) is uniform, too, regardless of the interference term \( f \). In this case, using the same arguments as in (6), \( f \) and \( x \) are statistically independent, however, \( \alpha \) and \( x \) are not independent. This is only the case for \( \sigma_r^2 \rightarrow \infty \) or if the (scaled) interference term which is subtracted is uniformly distributed over the Voronoi region. Then, \( \alpha \) and \( x \) are independent, even if \( \alpha \) is drawn from a finite constellation.
Furthermore, the results are only valid, if the above stated statistically independence between \( a \) and \( x \) holds. Using an additional (pseudo) random “dither sequence”, uniformly distributed over \( R_A \) and known to transmitter and receiver (“common randomness”), the interference term can be forced to guarantee (6), see [8, 24]. Here, additional side information which has to be transmitted, is used to increase capacity. In practice, the recovery of a pseudo random dither sequence at the receiver requires additional complexity and signaling overhead. This obstacle makes this theoretical concept questionable for practical schemes. Subsequently, we assume that the (known) interference \( f \) is such that (6) is (nearly) fulfilled.

The effect of the scaling factor can be seen by calculating the effective noise term. Considering \( x = \text{mod}_A (a - \alpha f) \), which gives \( a = \text{mod}_A (x + \alpha f) \), and \( y = x + f + \bar{n} \) we have [8]

\[
\bar{n} \equiv \text{mod}_A (a - \alpha) = \text{mod}_A (\text{mod}_A (\alpha y) - \text{mod}_A (x + \alpha f)) = \text{mod}_A (\alpha x + \alpha f + \alpha n - x - \alpha f) = \text{mod}_A (\alpha n - (1 - \alpha)x). \tag{7}
\]

The additive disturbance \( \bar{n} \) is hence composed of the channel noise \( \bar{n} \) and parts of the transmit signal \( x \). Since \( x \) is statistically independent of the data \( a \) to be detected at the receiver, this component accounts to the noise and does not lead to a biased solution, where parts of the desired signal is falsely apportioned to the noise, cf. [3].

Noteworthy, the effective noise is no longer Gaussian and isotropically distributed, but is given by the convolution of an \( N \)-dimensional Gaussian pdf and a pdf uniform over (a scaled version of) the Voronoi region \( R_A \). Hence, strictly speaking, detection or channel decoding based on squared Euclidean distances is no longer optimum. However, as the dimensionality of the lattice goes to infinity and its second moment approaches \( 1/(2\pi e) \), the low-dimensional projections tend to Gaussian pdfs, cf. [8].

For achieving capacity, \( \alpha \) uniformly distributed over the Voronoi region is again optimum (cf. (3)), but \( \alpha \) has to be optimized. Unfortunately, for finite \( N \), this has to be done numerically. Since

\[
C_{A,\text{MEE}}(\sigma_n^2) = \max_{\alpha} \frac{1}{N} \left( \log_2 (V(A)) - h(\bar{n}) \right), \tag{8}
\]

maximization of \( C_{A,\text{MEE}}(\sigma_n^2) \) is equivalent to a minimization of the differential entropy of \( n \). Hence we call this new optimization criterion minimum error entropy (MEE) approach.

An approximation \( C_{A,\text{MMSE}}(\sigma_n^2) \) to this solution can be obtained by minimizing the variance of \( \bar{n} = \alpha n - (1 - \alpha)x \), i.e., the unfolded version of the effective noise. The solution to this classical and well-known MMSE setting is given straightforwardly by \( \alpha_{\text{opt}} = \sigma_n^2 / (\sigma_n^2 + \sigma_m^2) \), for which the variance of \( \bar{n} \) calculates to

\[
\sigma_{\bar{n}}^2 = \frac{\sigma_n^2 \cdot \sigma_m^2}{\sigma_n^2 + \sigma_m^2}. \tag{9}
\]

Moreover, ignoring the fact that \( \bar{n} \) is not strictly Gaussian, the capacity curves derived in the last section (ZF solution) may simply be transformed to obtain a new and simple approximation of the capacity curves. Writing \( C_{A,\text{ZF}} = F(\sigma_n^2 / \sigma_m^2) \) as a function \( F(\cdot) \) of the signal-to-noise ratio, and replacing \( \sigma_n^2 \) by \( \sigma_m^2 \), an approximation for
the capacity is given by \( C_{\Lambda, \text{MMSE}} \approx F(\sigma_n^2/\sigma_\alpha^2 + 1) \). As the dimensionality of \( \Lambda \) approaches infinity this estimate becomes exact and, using (5), asymptotically approaches \( C_{\Lambda, \text{MMSE}} = 1/2 \log_2 \left( 1 + \sigma_\alpha^2/\sigma_n^2 \right) \), i.e., the capacity of the AWGN channel [8, 24].

Again, using Monte-Carlo methods for the integration of (8) and the convolution of the pdfs of \( \alpha n \) and \( (1-\alpha)x \), the capacity curves for the mod-\( \Lambda \) channel are generated. The results are plotted in Fig. 5. The solid lines correspond to the MEE approach using at each SNR an individually numerically optimized gain parameter \( \alpha \). For reference, the ZF capacity curves (dashed-dotted) and the AWGN capacity curve (dashed) are shown as well.

Replacing ZF precoding by the MEE approach, especially at low SNR significant gains are possible. Conversely, for high SNR ZF and MEE solutions converge; for rates above 2 bits/dimension no significant gains are observable. Moreover, it should be noted that in the low-SNR region the convergence towards the AWGN capacity by using higher-dimensional lattices is rather slow. At high SNR, basically the shaping gain of the lattice is utilisable.

In Fig. 6 the different approaches are compared. The solid lines correspond to the situation where \( \alpha \) was numerically optimized (MEE approach) and the dotted lines are valid for \( \alpha = 1 \) (ZF solution). Using the classical MMSE solution for the parameter \( \alpha \) the dashed curves result. The respective values of \( \alpha \) are compared in the bottom part of the figures. Only for the \( \mathbb{Z} \) and the \( \mathbb{Z}_2 \) lattice at low SNR a small loss in capacity and differences in the parameter \( \alpha \) are visible. Going to higher dimensional modulo operations, the MMSE setting for \( \alpha \) is almost optimum.

Finally, the approximation of the capacity by transforming the ZF curve is shown as well (dashed-dotted), which would be exact if the effective noise were Gaussian, i.e., as the dimensionality approaches infinity. At low SNR an overestimate is obtained, whereas at high SNR, the true capacity is slightly higher than the approximation.

5. Finite Interference Variance

The above derivations are only valid if the variance of the interference approaches infinity or the independence of \( \alpha \) and \( x \) is forced by introducing common randomness (dither sequence). If non of this is valid the precoding schemes can be improved as depicted in Fig. 7. Now, the scaling factors at transmitter (\( \alpha \)) and at receiver (\( \beta \)) differ, this offers a new degree of freedom results, not used up to now in precoding schemes. Since we have seen in the last section, that, especially for higher-dimensional lattices, MEE and MMSE approach lead to similar results, we now concentrate on the error variance of the (unfolded) additive noise term.

Before deriving the optimum parameters for given lattice \( \Lambda \) (and hence variance \( \sigma_n^2 \)), noise variance \( \sigma_\alpha^2 \), and interference variance \( \sigma_f^2 \), we first observe important properties of the respective signals: Obviously, the noise term \( n \) is statistically independent of the data symbols \( a \) and the transmit symbols \( x \), as well as independent of the interference term \( f \); we short write: \( n \perp a, n \perp x, n \perp f \) (orthogonality in Hilbert space). Additionally, \( f \perp a \) is valid. Finally, assuming again \( a \) to be uniformly distributed over the Voronoï region \( \mathcal{V}_\Lambda \), \( f \) and \( x \) are statistically independent: \( f \perp x \). However, for finite interference variance \( \sigma_f^2 \), data symbols \( a \) and channel symbols \( x \) are no longer independent (cf. footnote above).

For the further derivations, Fig. 8 gives a linearized description of the precoding scheme. Here, the effective data symbols \( v \) have been introduced, and it is assumed that the modulo operations at transmitter and receiver uses the same precoding symbol \( d \), which describes the action of the modulo operation. This assumption is equivalent to ignoring the modulo folding of the error symbols \( e \) as is already done in the derivation above.

![Fig. 7. Transmission scheme using modified mod-\( \Lambda \) precoding.](image)

\[ e = \beta y - v = \beta (x + f + n) - v = \beta x + \beta n + (\beta - \alpha) f + \alpha f - v = \beta x + \beta n + (\beta - \alpha) f - x = (\beta - 1) x + \beta n + (\beta - \alpha) f \] (10)
Fig. 5. Capacity curves according to the minimum error entropy approach of the modulo-Λ channel for various lattices over the signal-to-noise ratio $\sigma_x^2/\sigma_n^2$. Bottom to top: $E_4$, $A_2$, $D_4$, $E_8$ lattice. Dashed-dotted: Zero-forcing capacity curves. Dashed: AWGN channel.

Since $x$, $n$, and $f$ are pairwise orthogonal (see above), the error variance (per dimension) reads

$$\sigma_e^2 = (\beta - 1)^2 \sigma_x^2 + \beta^2 \sigma_n^2 + (\beta - \alpha)^2 \sigma_f^2. \quad (11)$$

5.1 MMSE Solution

A possible strategy is to choose $\alpha$ and $\beta$ such that the error variance $\sigma_e^2$ is minimized. Standard optimization gives

$$\frac{\partial \sigma_e^2}{\partial \alpha} = 2(\beta - \alpha)\sigma_f^2 = 0 \quad \Rightarrow \quad \beta = \alpha \quad (12)$$

which in turn leads to

$$\beta_{\text{MMSE}} = \alpha_{\text{MMSE}} = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_n^2}. \quad (13)$$

Hence, the optimization according to the MMSE criterion leads to identical scaling factors $\alpha$ and $\beta$ (as already derived above), which are moreover independent of the interference variance $\sigma_f^2$. However, since for finite variance $\sigma_f^2$ the channel symbols $x$ are not statistically independent of the data signal $v$, the MMSE solution has some bias, i.e., part of the desired signal is apportioned to the error signal.

Figure 9 visualizes the signals in three-dimensional (Hilbert) space (cf. the similar figures in [19, 20]). The squared length of the vectors corresponds to their variance, and perpendicular vectors imply orthogonality. The space is spanned by the pairwise orthogonal vectors $x$, $f$, and $n$. As can be seen from Fig. 6, $x = v - \alpha f$ holds. The error vector is given by the difference between the scaled receive vector $\alpha (x + f + n)$ and the data vector $v$. As usual in MMSE optimization, the error $e$ is perpendicular to the observation $y = x + f + n$, but not perpendicular to the desired signal $v$. 

Fig. 9. Visualization of the signals in three-dimensional space. Optimization according to the MMSE criterion.
5.2 Unbiased MMSE Solution

Compensation for the bias leads to a somewhat increased mean-squared error but at the same time to better error performance. In the optimum, the error is orthogonal to the desired signal. This situation is depicted in Fig. 10.

Since the lengths of the vectors \( f \), \( x \), and \( n \) correspond to the standard deviation of the respective signals, we have

\[
f = \begin{bmatrix} \sigma_f \\ 0 \end{bmatrix}, \quad x = \begin{bmatrix} 0 \\ \sigma_x \end{bmatrix}, \quad n = \begin{bmatrix} 0 \\ 0 \end{bmatrix},
\]

\[
y = \begin{bmatrix} \sigma_f \\ \sigma_x \\ \sigma_n \end{bmatrix}, \quad v = \begin{bmatrix} \alpha \sigma_f \\ \sigma_x \\ 0 \end{bmatrix},
\]

and the error vector reads

\[
e = \beta y - v = \beta \begin{bmatrix} \sigma_f \\ \sigma_x \\ \sigma_n \end{bmatrix} - \begin{bmatrix} \alpha \sigma_f \\ \sigma_x \\ 0 \end{bmatrix} = \begin{bmatrix} (\beta - \alpha) \sigma_f \\ (\beta - 1) \sigma_x \\ \beta \sigma_n \end{bmatrix}. \tag{15}
\]

For orthogonality between \( v \) and \( e \)

\[
e^T v = \begin{bmatrix} (\beta - \alpha) \sigma_f \\ (\beta - 1) \sigma_x \\ \beta \sigma_n \end{bmatrix}^T \begin{bmatrix} \alpha \sigma_f \\ \sigma_x \\ 0 \end{bmatrix} = (\beta - \alpha) \alpha \sigma_f^2 + (\beta - 1) \sigma_x^2 \equiv 0 \tag{16}
\]

has to hold. Solving for \( \beta \) results in \( \beta (\alpha \sigma_f^2 + \sigma_x^2) = \alpha^2 \sigma_f^2 + \sigma_x^2 \) or (cf. [19, 20])

\[
\beta_{\text{opt}} = \frac{\sigma_x^2 + \alpha^2 \sigma_f^2}{\sigma_x^2 + \alpha \sigma_f^2}. \tag{17}
\]

Using this, the error variance (equation (11)) reads

\[
\sigma_e^2 = \left( \alpha^4 (\sigma_x^2 + \sigma_n^2) \sigma_f^4 - 2 \alpha^3 \sigma_x^2 \sigma_f^4 \\
+ \alpha^2 (\sigma_x^2 + \sigma_n^2) \sigma_f^2 \sigma_x^4 \\
- 2 \alpha^2 \sigma_x^4 (\sigma_n^2 + \sigma_f^2) \sigma_x^2 \right) / (\sigma_x^2 + \alpha \sigma_f^2)^2. \tag{18}
\]
In the final step, this variance has to be minimized with respect to $\alpha$. Unfortunately, no analytic solution is possible. In particular, the above derived parameter $\alpha = \sigma_x^2/(\sigma_n^2 + \sigma_f^2)$ is not optimum in the unbiased MMSE case. However, as $\sigma_f^2$ approaches infinity, $\beta_{\text{opt}} = \alpha$ holds (cf. (17)). Then, the biased and the unbiased solution converge; asymptotically, for $\sigma_f^2 \to \infty$, the error is perpendicular to both $v$ and $f$, since in the limit they are parallel.

The (normalized) error variance $\sigma_x^2/\sigma_n^2$ for the biased and unbiased MMSE solution is shown in Figures 11 (a) through (d) over the parameter $\alpha$ and the interference variance $\sigma_f^2/\sigma_n^2$. The figures show the situation for $\sigma_x^2/\sigma_n^2 = 1, 2, 5$, and $10$. The parameters $\alpha$ leading to the minimum error variance are plotted beneath the surface (marker “$\times$” for the biased and marker “$\circ$” for the unbiased MMSE solution). Additionally, the parameter $\beta_{\text{opt}}$ corresponding to the optimal choice for $\alpha$ for the unbiased solution is plotted (marker “$\Box$”). Noteworthy, for the biased MMSE case, $\beta = \alpha$ holds.

Especially for low signal-to-noise ratios $\sigma_x^2/\sigma_n^2$, the optimum parameter $\alpha$ is significantly higher for the unbiased MMSE case compared to the biased situation, where $\sigma_{\text{MMSE}}^2 = \sigma_x^2/(\sigma_x^2 + \sigma_n^2)$ holds independent of the interference variance $\sigma_f^2$. Moreover, the receiver side scaling $\beta$ is (much) larger than the transmitter side scaling $\alpha$. This fact has to be taken into account in the design of future transmission schemes employing modulo-$\Lambda$ arithmetics. As the interference variance $\sigma_f^2$ and/or the signal-to-noise ratio tend to infinity, biased and unbiased solution converge, and $\sigma_{\text{MMSE}}^2$ is the optimum choice. Additionally, $\alpha$ and $\beta$ converge, too.

6. Conclusions

In this paper, the key elements of precoding for transmission over channels with interference known at the transmitter have been discussed. In particular, the choice of the lattice on which the modulo operations at transmitter and receiver are based and the effect of proper scaling have been assessed.

At high signal-to-noise ratios, going to higher-dimensional lattices provides additional shaping gain. At low SNR the situation depends on whether the interference can be scaled before the pre-subtraction and modulo reduction. If this is not the case (e.g., in multilevel codes), using higher-dimensional lattices is not rewarding; asymptotically, below an SNR of $0$ dB no transmission is possible at all. However, appropriate scaling of the interference term to be pre-subtracted, i.e., establishing an MMSE solution, leads to an increase in capacity.

Usually in literature the derivations assume that the interference is such that the data symbols $x$ and the generated channel symbols $\nu$ are statistically independent. This is achieved by very “strong” interference, or by using a dither sequence, known to transmitter and receiver. But in practical schemes, this assumption of having common randomness is questionable. It has been shown that for interferences with finite variance, which is the case in all precoding schemes for ISI or MIMO channels, scaling with different factors at transmitter and receiver is rewarding.

References


Fig. 11. (Normalized) error variance $\sigma_e^2/\sigma_n^2$ for biased (lower surface) and unbiased (upper surface) MMSE solution over the parameter $\alpha$ and the interference variance $\sigma_f^2/\sigma_n^2$. Optimum choice of $\alpha = \beta$ for the biased MMSE solution: "x". Optimum choice of $\alpha$ ("\(\sigma\)) and the corresponding parameter $\beta_{\text{opt}}$ ("\(\Box\)) for the unbiased MMSE solution. (a) $\sigma_f^2/\sigma_n^2 = 1$, (b) $\sigma_f^2/\sigma_n^2 = 2$, (c) $\sigma_f^2/\sigma_n^2 = 5$, (d) $\sigma_f^2/\sigma_n^2 = 10$.


