A LATTICE PRECODING SCHEME FOR FLAT-FADING MIMO CHANNELS

Jin He  Masoud Salehi

Department of Electrical and Computer Engineering
Northeastern University
Email: {jhe, salehi}@ece.neu.edu

ABSTRACT

This paper provides some details of a novel vector perturbation based lattice precoding algorithm based on Eigenvalue decomposition, named EDLP, for single user multiple-input multiple-output (MIMO) systems operating over slowly-varying flat-fading channels. Linear precoding is used in this technique to orthogonalize the fading channel, thus nullifying the interference among spatial channels. A perturbation vector is employed as well to reduce the transmitted power. Joint optimization of the linear precoder and the perturbation vector enhances the performance of the algorithm. As a result, the EDLP features full diversity order and full spatial multiplexing order. Furthermore, compared with other precoding techniques, EDLP requires less computational complexity. In appendixes, a proof validating the transformation from the original optimization metric to the minimization of a quadratic form is provided, as well as an elaborated algorithm to generate perturbation vector.

1. INTRODUCTION

In rich scattering environments, MIMO systems can provide large data rates by employing spatial multiplexing [1]. In addition to channel noise and fading, MIMO systems suffer from interference among spatial channels as well. Linear precoding techniques such as singular value decomposition (SVD) or zero-forcing (ZF) techniques orthogonalize the spatial channels and hence, cancel the interference from data streams transmitted on different spatial channels require large transmitted power, especially when the communication channel is ill-conditioned. The diversity order obtained by using linear precoding techniques is limited and improved performance can be achieved by using the Tomlinson-Harashima precoding (THP) [2]. In order to exploit full diversity order, more advanced techniques such as lattice precoding can be used to significantly improve the performance of the uncoded error probability at moderate to high signal to noise ratio (SNR) values.

Lattice precoding [3], [4] [5] [6] [7] and lattice detection [8] [9] use lattice reduction (LR) techniques to implement dirty paper coding (DPC) [10] [11]. These techniques can be interpreted as closest lattice point search problems, which can be exactly solved by sphere decoding techniques with relatively higher computational complexity [12] [13] [14]. LR aided precoding techniques solve the closest lattice point search problem by employing rounding off approximation (using LR only) or nearest plane approximation (using LR plus sorted QR decomposition). However, the LR techniques in high dimensions is NP-hard [12] and the proposed suboptimal polynomial time LLL LR algorithm in the worst case has unbounded complexity, even for low dimensional cases [8]. This becomes particularly a problem since in implementation, before each frame transmission, the LR algorithm needs to be performed.

In this paper, a novel Lattice Precoding technique based on Eigenvalue decomposition (EDLP) is proposed which not only exploits full diversity order with spatial multiplexing, but also provides power gain over the performance of traditional LR precoding. The computational complexity of the proposed scheme is similar to THP with linear precoding. This precoding technique is based on the assumption of the availability of channel state information (CSI) at the transmitter and uses a ZF linear precoding matrix computed from the CSI. The well-known deficiency of using a ZF spatial equalizer at the receiver is the noise enhancement effect. Similarly, the ZF precoding suffers from transmitted power amplification. Despite power amplification at the transmitter introduced by ZF precoding, the received signal can be detected by simple per-component quantizers with limited loss when compared with maximum likelihood (ML) detection. In order to reduce the transmitted power, the proposed technique employs a vector perturbation technique through extension of the signal constellation at the transmitter and solving a closest lattice point search problem. In addition, coordinated spatial rotations at both transmitter and receiver sides are introduced to reduce the transmitted power without affecting the statistics of the noise.

In Section II, we present the structure of the transceiver with nonlinear processing at the transmitter and linear processing at the receiver. Based on the proposed structure, the optimization metric that minimizing the transmitted
power is proposed. This optimization problem is solved in Section III. We study the error rate performance of the system using simulation in Section IV, as well as the computational complexity of the proposed scheme. Finally, Section V concludes the paper with some remarks. Throughout this paper, vectors and matrices are denoted in bold letters. Transpose and Hermitian transpose are denoted by $(\cdot)^T$ and $(\cdot)^H$, respectively. $\det(A)$ denotes the determinant of the matrix $A$ and $\|a\|$ denotes the norm of vector $a$. $I_n$ denotes the $n$-dimensional identity matrix, and $\bar{a}$ denotes element-wise conjugate of vector $a$.

II. SYSTEM MODEL

Common techniques for linear coherent reception when CSI is available at the receiver include canceling interference (ZF), maximizing the signal-to-noise ratio (matched filtering), or minimizing the mean square error. Nonlinear techniques can be regarded as additional blocks that use decision feedback to suppress the interference (successive interference cancelation), or transform the problem into an equivalent domain where the signals on different dimensions become approximately orthogonal (lattice reduction). When CSI is available at the transmitter as well, precoding techniques can be utilized. In this case the computational burden to perform the aforementioned tasks can be put at the transmitter side, resulting a less complex receiver which is of extreme significance in cases of infrastructure and high data rate networks.

Precoding techniques are capable of managing the transmitted power by shaping the transmit signal such that the received signal does not suffer from noise enhancement and spatial interference. This operation usually enhances the power and hence some linear and nonlinear processing at the transmitter is necessary to reduce the transmitted power.

We consider a slowly-varying, flat-fading, $N_t$-input and $N_r$-output channel, where the complex channel matrix is assumed to remain constant during one transmission frame and changes independently from one frame to another. The SVD of the complex channel matrix is denoted by $H = U\Lambda V^H$, which is assumed to be known at both the transmitter and the receiver. The transmitted signal for the proposed EDLP can be expressed as

$$x = H^{-1}UB(s + p) = V\Lambda^{-1}B(s + p)$$

where $B$ is a unitary matrix. Each element of the transmit symbol vector $s$ takes value in an $M$-QAM signal constellation with equal probability. The perturbation vector $p$ belongs to a lattice structure which extends the original signal constellation periodically. The choice of this design parameter depends on the realization of the channel, $H$, and symbol vector $s$ in order to reduce the transmitted power. The dimensions of $B$, $s$ and $p$ are the same and denoted by $L$ where $L \leq \min(N_t,N_r)$ is the number of independent data streams that are transmitted over the spatial channels. The received signal, $y = Hx + n$, is linearly equalized in spatial domain by $R = UB$, as

$$r = R^Hy = R^H Hx + R^H n = s + p + n'$$

where $n \sim CN(0,\sigma_n^2I)$. Since $R$ is orthonormal, the noise after linear equalization, i.e., $n'$, has the same statistical properties as the channel noise $n$. The receiver also uses a quantizer to subtract the perturbation vector $p$, which was intentionally added at the transmitter for the sake of reducing the transmitted power. To constrain the implementation of the receiver, this quantizer is set to be a simple per-component complex quantizer, denoted by $Q_{\alpha G^L}$, which takes modulo operations on each of the real and imaginary parts of every element of the vector separately. Here $\alpha$ is the divisor of the modulo operation and $G$ is the set of the Gaussian integers. This constraint indicates that the lattice $p$ belongs to the scaled Gaussian integer lattice. The symbol decision statistic $\hat{s}$ is the output of the per-component complex quantizer,

$$\hat{s} = Q_{\alpha G^L}(r) = s + Q_{\alpha G^L}(n').$$

In this way, the estimated signal is free of interference and corrupted only by a quantized Gaussian noise $n'' \sim CN(0,\sigma_n^2I)$.

The power of the transmit signal, $\|x\|^2$ is given by

$$\|x\|^2 = \|\Lambda^{-1}B(s + p)\|^2,$$

which is a function of the linear precoding matrix $B$, the perturbation vector $p$, and the channel. As noted earlier, our design goal is to select the unitary matrix $B$ and $p$ to minimize the average transmitted power. This optimization is taken over the choices of $B$ and $p$,

$$\text{minimize } \mathbb{E}[\|\Lambda^{-1}B(s + p)\|^2 | H]$$

(1)

$$B \in U(L)$$

$$p \in \alpha G^L$$

where in (1) the expectation is over equiprobable symbol vectors. The symbol $U(L)$ denotes the unitary group with dimension $L$. The linear part of the precoding operation, i.e., the unitary matrix $B$ depends only on the channel and once is determined, will remain the same for the entire transmission frame. The perturbation vector $p$ is a point in the scaled $L$ dimensional Gaussian integer lattice. The scaling scalar $\alpha$ is selected such that the Voronoi regions of the signal constellation, which is periodically extended by the lattice through $s + p$, do not overlap.
III. AN EIGENVALUE DECOMPOSITION BASED LATTICE PRECODING

To be specific, we consider a system using an M-QAM constellation in each independent data stream and spatial multiplexing order of \( L \). Define the signal space as \( \tilde{\mathbf{S}} = [\tilde{s}_1 \ldots \tilde{s}_N] \) and \( \tilde{s}_i = s_i + p_i, 1 \leq i \leq N \) with cardinality \( N = M^L \). The set \( \{s_i\}_{i=1}^N \) is the set of all uniquely ordered combinations of \( L \) M-QAM signals. We can rewrite Equation (1) as

\[
\mathbb{E}[\|x\|^2 | H] = \mathbb{E}[\tilde{s}^H B^H \Lambda^{-2} B \tilde{s} | H]
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} \tilde{s}_i^H \sum_{j=1}^{L} \gamma_j^2 b_j b_j^H \tilde{s}_i
\]

\[
= \frac{1}{N} \sum_{j=1}^{L} \gamma_j^2 b_j^H R_a b_j
\]

(2)

where \( \gamma_j \) is the \( j \)th singular value on the diagonal of \( \Lambda^{-1} \), \( b_j \) is the \( j \)th column of \( B^H \), and \( R_a = SS^H \). For \( L = 2 \) case, Equation (2) can be written in the following form, as derived in Appendix I,

\[
\mathbb{E}[\|x\|^2 | H] = \frac{1}{N} b_1^H \Phi b_1
\]

(3)

where

\[\Phi = \gamma_1^2 R_a + \gamma_2^2 \det(R_a) R_a^{-1}.\]

Therefore, the optimization problem in (1) reduces to the minimization of the quadratic form given in Equation (3), which is minimized when: 1) the minimum eigenvalue of the Hermitian matrix \( \Phi \), denoted by \( \eta_{\text{min}} \), is minimized and 2) when \( b_1 \) is selected to be the eigenvector corresponding to \( \eta_{\text{min}} \).

Since \( \Phi \) is Hermitian, using the eigenvalue decomposition \( R_a = U_a D_a U_a^H \), we have

\[
\Phi = \ln(U_a e^{\gamma_1^2 D_a} U_a^H) U_a e^{\gamma_2^2 \det(R_a) D_a^{-1}} U_a^H
\]

\[
= U_a (\gamma_1^2 D_a + \gamma_2^2 \det(R_a) D_a^{-1}) U_a^H
\]

It is easy to see that the eigenvalues of \( \Phi \), denoted by \( \eta_1 \) and \( \eta_2 \), can be expressed as

\[
\eta_1 = \gamma_1^2 a_1 + \gamma_2^2 a_2
\]

\[
\eta_2 = \gamma_1^2 a_2 + \gamma_2^2 a_1
\]

where \( a_1 \) and \( a_2 \) are the eigenvalues of \( R_a \). Therefore, the minimum eigenvalue of \( \Phi \) is given by

\[
\eta_{\text{min}} = \gamma_1^2 a_{\text{min}} + \gamma_2^2 a_{\text{max}}.
\]

(4)

From the definition of \( R_a \), we have

\[
R_a = \frac{N}{2} I_2 + \sum_{j=1}^{N} (p_j s_j^H + s_j p_j^H + p_j p_j^H).
\]

Hence, \( R_a - \frac{N}{2} I_2 \) is positive semidefinite and

\[
\det(R_a) = a_{\min} a_{\max} \geq \frac{N^2}{4}.
\]

(5)

This result can be used to bound the minimum transmitted power through (4),

\[
\mathbb{E}[\|x\|^2 | H] \geq \frac{2}{N} \gamma_{\max} \gamma_{\min} \sqrt{a_{\min} a_{\max}}
\]

\[
= \frac{2}{N} \det(\Lambda^{-1}) \sqrt{\det(R_a)}
\]

\[
\geq \det(\Lambda^{-1})
\]

(6)

where first equality holds for \( \gamma_{\max} a_{\min} = \gamma_{\min} a_{\max} \).

One approach to find the optimizing set \( \{p_i\}_{i=1}^N \) for a given channel is to consider a graphical representation of Equation (4) as shown in Figure 1. In this figure the

![Fig. 1. Sketch for finding solutions of minimizing transmitted power](image)

eigenvalues of \( R_a \), \( (a_{\max}, a_{\min}) \), are represented as points in the plane, where each point is determined by the choice of \( \{p_i\}_{i=1}^N \). Since \( R_a \) is Hermitian, these discrete points are all in the first quadrant and are left and lower bounded by the hyperbola given by Equation (5), shown in Figure 1 as the shaded area. The channel determines the slope of the straight line

\[
a_{\min} = -\frac{\gamma_2^2}{\gamma_{\max}} a_{\max}
\]

(7)

going through the origin. One of the discrete points on the lower bound hyperbola that is at minimum distance to this straight line represents the set of perturbation vectors that minimize the transmitted power. We denote the minimum distance between them by \( \eta_{\text{min}} \). The points on the boundary hyperbola can be computed in a systematic way, as explained in Appendix II. If we denote the 2-norm condition number of matrix \( H \) by \( \kappa_2(H) \) with
\( \kappa_2(H) = \kappa_2(H^{-1}) = \frac{\gamma_{\text{max}}}{\gamma_{\text{min}}} \). Equation (7) can be rewritten as \( a_{\text{min}} = -\kappa_2^2(H)a_{\text{max}} \). Also denote the slope of the line segment connecting two consecutive points with index \( n-1 \) and \( n \) on the lower boundary as \( \delta_n \) and define \( \delta_0 = -1 \). The sequence \( -\delta_n^{-1} \) for \( 0 \leq n < \infty \) partitions the range \([1, \infty)\) and such partition is denoted by \( D \). It is obvious that the index of the partition that \( \kappa_2^2(H) \) falls in is the index of the closest point on the lower hyperbola to the line in Equation (7), as shown in Figure 1, and the corresponding set of \( \{p_i\}_{i=1}^N \) is the optimum perturbation vector set. The channel characteristics, which are known at the transmitter, result in a particular \( \kappa_2(H) \). Using this value, the transmitter looks for the corresponding partition and uses the corresponding set of \( \{p_i\}_{i=1}^N \) for precoding. The selected perturbation vector \( p \) is then used in Equation (3) to compute the linear precoding matrix \( B \).

**IV. PERFORMANCE AND COMPLEXITY ANALYSIS**

In Section III a transceiver structure for single user MIMO system is proposed. The parameters in the system are designed to optimizing the performance as given by Equation (1). In the following, we discuss the error probability performance and the computational complexity of the system.

**A. SIMULATION RESULTS**

In this section we also compare the performance and complexity of the proposed EDLP scheme with a number of other recently proposed techniques. In [15] a new technique for MIMO broadcast channels is proposed. The single antenna receivers use per-component complex scalar quantizers to determine symbol estimates. At the transmitter, the amplified transmitted power due to using ZF linear precoding is reduced by introducing a perturbation vector. This power reduction is carried out by performing a closest lattice point search, the exact solution of which is NP-hard. Nearest plane and rounding off approximation techniques are used to solve this problem and find the perturbation vector. The EDLP transceiver structure introduced above differs from [15] in the additional coordinated linear precoding and equalization represented by the orthonormal transformation matrix \( R \). By doing this, the transmit signal can be shaped to match the channel through a unitary transformation in order to achieve a larger channel gain. Furthermore, the original lattice basis, for example in Equation (5) in [15], are not optimized for LR with receive antenna joint processing. Different original lattice basis, when fed into the LR algorithm, require different numbers of iterations in the lattice reduction process and result in different reduced lattice basis. The transmitter rotation matrix \( B \) is introduced to provide the LR process an additional degree of freedom such that the resulting reduced basis is shorter and more orthogonal.

A single user LR aided precoding and reception technique is proposed in [3] and is referred to as LRAPR. This technique uses a ZF linear precoding matrix whose bases are reduced by LR from the channel inverse precoding matrix. Hence, the condition number of the precoding matrix becomes smaller and the transmitted power is reduced. However, at the receiver the unimodular lattice transformation matrix needs to be computed from the channel, or be sent by the transmitter to transform back the original transmit constellation.

In Figure 2, BER performance of EDLP is evaluated through simulation and is compared with the techniques described above. It can be seen from this figure that all these techniques achieve full diversity order, thanks to the nonlinear perturbation. However, EDLP displays a clear power gain over all other precoding techniques. Comparing to the nearest plane and rounding off approximation techniques, the receive antenna collaboration and the coordination between the transmitter and the receiver in the EDLP provides a significant 2.5 dB power gain. Although the single user LR aided precoding also provides a power gain over the two broadcast channel techniques mentioned above through the receive antenna collaboration, it still cannot outperform the EDLP in the normal operating BER-SNR range since its linear precoding part of it is not optimized for LR. Notice that in order to reduce the complexity at the receiver, the EDLP uses linear equalization and per-component quantizers, therefore, there remains a gap between its performance and the performance of ML detection.

Fig. 2. Error rate performance comparison of different techniques
B. COMPLEXITY

We use the required number of operations to evaluate the overall complexity of different transmission and reception techniques. The computation operations required for ML, LRAP, LRAPR, and EDLP are listed in Table I, where the operations are categorized based on location and stage of the computation. The stage in which the transceiver processes the CSI to prepare for the succeeding operations at the transmitter (Tx) and the receiver (Rx) is referred to as “frame overhead” at Tx and Rx. The operations needed for transmission and reception of each symbol is referred to as “per channel use”.

V. CONCLUSIONS

In this paper an eigenvalue decomposition based lattice precoding technique, EDLP, is proposed for use in MIMO single user flat-fading channels. This technique requires CSI at both transmitter and receiver sides. In return, the resulting computation at the receiver side is limited to linear processing and simple per-component quantization. Different from other lattice precoding techniques, although a nonlinear precoding scheme is used at the transmitter side, the computation does not require complex LR algorithms.

APPENDIX I

Define a unitary matrix as

$$B^H = [b_1, b_2] = \begin{bmatrix} x & z \\ y & w \end{bmatrix}. $$

Let us define $z = -\bar{y}\rho$. Based on the condition $-\bar{x}z + \bar{y}w = 0$, we simply have $w = \bar{x}\rho$ and

$$B^H = \begin{bmatrix} x & \bar{y}\rho \\ y & \bar{x}\rho \end{bmatrix}. $$

Since $|x|^2 + |y|^2 = 1$ and $|x|^2 + |\bar{y}\rho|^2 = 1$, the magnitude of $\rho$ is $|\rho|^2 = 1$. Now define an operator $\Upsilon$ as

$$\Upsilon = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}. $$

and write $b_2 = \rho \Upsilon b_1$. For the Hermitian matrix $R_a$,

$$b_2^H R_a b_2 = |\rho|^2 \bar{b}_1^H \Upsilon^H R_a \Upsilon b_1.$$

Since the cross terms of a quadratic form sums up to real, we have

$$b_2^H R_a b_2 = \bar{b}_1^H \Upsilon^H R_a \Upsilon b_1.$$

Applying the operator $\Upsilon$ on both sides results

$$\Upsilon^H R_a \Upsilon = \begin{bmatrix} R_a(2,2) & -R_a(2,1) \\ -R_a(1,2) & R_a(1,1) \end{bmatrix} = \det(R_a) R_a^{-1}$$

where $R_a(i,j)$ is the element on the $i$th row and $j$th column of $R_a$, this results in Equation (3).

APPENDIX II

Consider a system using a 4-QAM scheme, i.e., $M = 4$, as an example. Elements of $s_i$ take values of $\pm 1/2 \pm j/2$ and therefore $\alpha = 2$. The value of the signal on the constellation $s_i$ is defined in the following way: convert decimal integer $(i - 1)_d$ into binary $(i - 1)_b$; then map $+1 \rightarrow 1/2$ and $0 \rightarrow -1/2$ to construct the sequence $c_i = c_1^i c_2^i c_3^i c_4^i$ where $c_4^i$ is the most significant bit; finally define $s_i = [c_i^i + j c_2^i, c_3^i + j c_4^i]^T$. A series of points on the lower part of the hyperbola in the first quadrant are constructed below. These points are indexed by $n$, $0 \leq n \leq \infty$ in increasing order, as shown in Figure 1. In the rest of this appendix superscript $n$ indexes the den. Define two $L$-by-$N$ matrices

$$P_{b1} = \begin{bmatrix} 0 & -1 & 0 & 0 & 0 & -1 & -i & -i & i \ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \ \end{bmatrix};$$

$$P_{b2} = \begin{bmatrix} -1 & 0 & -i & -i & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \ \end{bmatrix}. $$

If $n$ is an even number, $p_{0i}^n$ is the $i$th column of the matrix $\alpha n (P_{b1} + P_{b2})$. If $n$ is odd $p_{0i}^n$ is the $i$th column of the matrix $\alpha (n + 1) P_{b1} + \alpha (n - 1) P_{b2}$. It can be easily verified that all points generated have $\det(R_a^0) = N^2/4$, indicating that all of these points lie on the hyperbola.
Every corresponding $R_n^m$ satisfies $\text{tr}(R_n^m) = N(n^2 + 1)$, therefore, the coordinates of point $n$ can be found as

$$\gamma_1, \gamma_2 = N\left(n^2 + 1 \pm \sqrt{(n^2 + 1)^2 - 1}\right).$$

However, this algorithm does not guarantee to generate all possible eigenpair points.

REFERENCES


