Using Real Constellations in Fully- and Over-loaded Large MU-MIMO Systems with Simple Detection

J.C. De Luna Ducoing, Na Yi, Member, IEEE, Yi Ma, Senior Member, IEEE, and Rahim Tafazolli

Abstract—The aim of this letter is to exhibit some advantages of using real constellations in large multi-user (MU) MIMO systems. It is shown that a widely linear zero-forcing (WLZF) receiver with M-ASK modulation enjoys a spatial-domain diversity gain, which linearly increases with the MIMO size even in fully- and over-loaded systems. Using the decision of WLZF as the initial state, the likelihood ascent search (LAS) achieves near-optimal BER performance in fully-loaded large MIMO systems. Interestingly, for coded systems, WLZF shows a much closer optimal BER performance in fully-loaded large MIMO systems.

Index Terms—Multiple-input multiple-output (MIMO), over-loaded, M-ASK, widely linear receiver.

I. INTRODUCTION

EMPLOYING a large number of receive antennas at the network side for signal reception is a prominent uplink approach for the next generation of wireless communication systems. Practical implementation of large multi-user multiple-input multiple-output (MU-MIMO) faces several signal processing challenges, including synchronization, channel state information (CSI) estimation, and signal detection. In the scope of signal detection, one major problem is finding the best performance-complexity trade-off. MIMO receivers have been extensively investigated and many linear and non-linear approaches can be found in the literature [1].

In densely populated areas, the number of independent transmit antennas can be equal to or higher than the number of receive antennas. These systems are referred to as fully-loaded or over-loaded large MU-MIMO systems, for which linear receivers such as zero-forcing (ZF) and linear minimum mean-square error (LMMSE) yield very poor performance. In contrast, non-linear receivers [1] lead to high computational complexity, especially in over-loaded scenarios [2]. Therefore, their implementation becomes challenging in large MIMO systems, notably for certain emerging wireless applications such as large-scale Internet of Things (IoT) networks [3], which are very demanding in terms of both the large number of connected IoT devices and high link reliability.

Buzzi et al. applied a widely linear (WL) filter together with M-ASK to improve signal detection in fully- and over-loaded (small) MIMO systems [4]. WL filters typically exhibit superior performance compared with strictly linear filters when the transmitted signals (such as M-ASK), interference, or noise are improper [5], [6], i.e., their pseudo-covariance matrix is not 0. WL processing has been proposed for applications in many contexts, such as CDMA [5], [7], single antenna interference cancellation [8], [9], OFDM [10], space-time codes [11], SISO systems with adaptive constellations [12], and MIMO [4], [13].

In this work, we apply the WL filter derived in [4] (termed WL zero-forcing (WLZF)) into fully- and over-loaded large MU-MIMO systems with M-ASK. The main motivation is that the WLZF receiver has low complexity with respect to the MIMO system size, even in very large over-loaded scenarios. Specifically, our contributions include: 1) we find that the spatial diversity of WLZF under Rayleigh fading is \( M - K/2 + 1/2 \), where \( M \) and \( K \) are the number of receive and transmit antennas, respectively (cf. diversity of ZF is \( M - K + 1 \)); the diversity expression also allows us to find the exact SER of WLZF, and 2) we present some interesting results upon applying WLZF as the starting point for the likelihood ascent search (LAS) receiver, because due to the large spatial diversity gain, the algorithm can quickly converge to a near-optimal decision without the knowledge of noise variance. Moreover, it is found that a 1/3-rate coded system can further improve the link reliability by 3 dB or more in SNR. Interestingly, the SNR gap between WLZF and WLZF-LAS is largely mitigated (only 0.9 dB difference in some cases). This renders WLZF to be a more viable approach in complexity-constrained applications.

II. MU-MIMO WITH REAL CONSTELLATIONS

The uplink of a wireless large MU-MIMO communication system is considered, subject to \( K \leq 2M \), where the system loading factor is defined as \( \alpha = K/M \). Independent streams are sent through a flat block fading MIMO channel. At the base station, the received signal \( y \in \mathbb{C}^M \) can be expressed as

\[
y = \sqrt{\mathcal{E}} \mathbf{H} \mathbf{D} x + z,
\]

where \( \mathbf{H} \) is the channel matrix, and each of its elements \( h_{ij} \) represents the i.i.d. small-scale fading coefficients with unitary variance from the \( j \)-th transmit to the \( i \)-th receive antenna. \( \mathbf{D} \) is a diagonal matrix where each diagonal element \( d_{jj} \) represents the large-scale fading coefficient of the \( j \)-th transmit antenna. Without loss of generality, \( \mathbf{D} \) is normalized to the identity matrix \( \mathbf{I}_K \), in this letter. The transmitted symbols \( x \in \mathbb{R}^K \) are selected from an M-ASK constellation, where \( \mathcal{E} \) is the energy per transmitted symbol. The vector \( z \in \mathbb{C}^M \) denotes the AWGN noise, which is distributed as \( \mathcal{CN}(0, \sigma^2 \mathbf{I}_M) \). Additionally, the normalized SNR and \( \mathcal{E}_b/N_0 \),
per transmit antenna, are \( \gamma = \mathcal{E}_s/\sigma^2 \) and \( \gamma_0 = \gamma/(R_c \log_2 m) \), respectively, where \( R_c \) is the code rate and \( m \) is the modulation order. It is assumed that perfect CSI at the receiver (CSIR) is available, but none at the transmitters.

A. Widely Linear Zero-Forcing (WLZF)

In the following paragraphs, a simple characterization of the filter proposed in [4] (termed WLZF in this letter) is presented, and we obtain the expression for diversity that it achieves and the exact SER.

Equation (1) can be written in a real representation as [14]

\[
\begin{bmatrix} \Re(y) \\ \Im(y) \end{bmatrix} = \sqrt{\mathcal{E}_s} \begin{bmatrix} \Re(H) \\ \Im(H) \end{bmatrix} \begin{bmatrix} \Re(x) \\ \Im(x) \end{bmatrix} + \begin{bmatrix} \Re(z) \\ \Im(z) \end{bmatrix},
\]

where \( \Re(\cdot) \) and \( \Im(\cdot) \) are the real and imaginary components, respectively. However, because the transmit symbols are selected from a real constellation, \( \Im(x) = 0 \), therefore (2) is equivalent to

\[
\begin{bmatrix} \Re(y) \\ \Im(y) \end{bmatrix} = \sqrt{\mathcal{E}_s} \begin{bmatrix} \Re(H) \\ \Im(H) \end{bmatrix} \begin{bmatrix} \Re(x) \\ \Im(x) \end{bmatrix} + \begin{bmatrix} \Re(z) \\ \Im(z) \end{bmatrix},
\]

which can be written in compact form as

\[
\bar{y} = \sqrt{\mathcal{E}_s} \bar{H} \bar{x} + \bar{z},
\]

where \( \bar{y}, \bar{H}, \bar{x}, \) and \( \bar{z} \) are real-valued and have dimensions \( 2M \times 1, 2M \times K, K \times 1, \) and \( 2M \times 1 \), respectively. The ZF estimate of the transmitted signal using (1) is given by \( \hat{x} = \frac{1}{\sqrt{\mathcal{E}_s}} H^+ \bar{y} \), where \( (\cdot)^+ \) denotes the pseudoinverse. However, it should be noted that it is not necessary to perform a pseudoinversion on the full channel matrix; therefore, a better approach is to use (4)

\[
\hat{x} = \frac{1}{\sqrt{\mathcal{E}_s}} \bar{H}^+ \bar{y} = \frac{1}{\sqrt{\mathcal{E}_s}} \begin{bmatrix} \Re(H) \\ \Im(H) \end{bmatrix}^+ \bar{y},
\]

which is the WLZF approach. The benefits are due to the fact that the equivalent channel matrix \( \bar{H} \) has half the number of columns that the real representation of the original channel matrix \( H \) has, thus making it better conditioned. Furthermore, because half the column vectors are removed, twice as many transmit antennas can be used; this would make the system over-loaded but the channel matrix would not be underdetermined. It should be noted that these terms are not the same: ‘over-loaded’ refers to the number of antennas, and ‘underdetermined’ is an attribute of the channel matrix.

It is known that for i.i.d. Rayleigh channels, the diversity of the ZF receiver is \( L(ZF) = M - K + 1 \) [15], [16]. The WLZF approach described in this letter achieves a diversity of (see Appendix for the derivation)

\[
L(WLZF) = M - \frac{K}{2} + \frac{1}{2}.
\]

For Rayleigh fading, the exact average BER in terms of the diversity \( L \), where \( L \) is allowed to be fractional, for a MIMO ZF receiver using BPSK or QPSK, also applicable to WLZF with BPSK, can be obtained from [17, eq. (7)]

\[
P_{b}^{(B)} = \frac{1}{2} - \sqrt{\frac{\gamma_0}{\pi}} \frac{\Gamma(L + \frac{3}{2})}{\Gamma(L)} \times \frac{2 F_1 \left( \frac{1}{2}; L + \frac{1}{2}, \frac{3}{2}; -\gamma_0 \right)},
\]

where \( \Gamma(L) \) is the Gamma function and \( _2F_1(a, b; c; z) \) is the hypergeometric function.

The SER for uncoded ZF and WLZF MIMO receivers with general M-ASK in terms of \( L \) is determined by following the steps in [17] and noting that the SER for a SISO system with M-ASK in an AWGN channel is given by [18, eq. (8.3)]

\[
P_s^{(AWGN)} = \frac{2(m - 1)}{m} Q \left( \frac{3\gamma}{\sigma^2(m^2 - 1)} \right),
\]

where \( Q(x) \) is the Q-function defined as \( \frac{1}{\sqrt{\pi}} \int_0^x e^{-t^2/2} dt \). The SER for M-ASK for the ZF and WLZF detectors is then

\[
P_s = \frac{m - 1}{m} \frac{2(m - 1)}{m} \sqrt{\frac{3\gamma}{\pi(m^2 - 1)}} \frac{\Gamma(L + \frac{1}{2})}{\Gamma(L)} \times _2F_1 \left( \frac{1}{2}; L + \frac{1}{2}, \frac{3}{2}; -\frac{3\gamma}{m^2 - 1} \right).
\]

The average BER can then be approximated by [18, eq. (8.7)]

\[
P_b \approx P_s / \log_2 m.
\]

WLZF can be seen as a way of splitting the in-phase and quadrature components of a complex stream into two separate real streams; in doing so, twice the number of transmit antennas are supported. In exchange, the orthogonality that existed in the components of the complex stream disappears, which results in a power efficiency loss (as seen in Section III) and a slight decrease of 0.5 in diversity. This is in agreement with the analysis presented in [19], where it was observed that a complex stream occupies 1 degree of freedom (DoF) of the total \( M \) available in the system, but a real stream uses only 0.5 DoF; thus twice as many real streams can be accommodated. WLZF is also consistent with [8], where it is shown that a receiver array with \( M \) elements can suppress \( 2M - 1 \) rectilinear interferences, effectively supporting \( 2M \) real-valued streams.

Throughout this letter, it is assumed that noise variance knowledge \( \sigma^2 \) is not required. It is possible to estimate \( \sigma^2 \) by the use of pilots or training sequences, thus, it would be straightforward to use WL-MMSE detection, although the diversity and SER equations are not trivially determined.

B. WLZF-LAS

LAS is a heuristic algorithm and one of the simplest for large MIMO detection that can approach near-optimal performance [20]. Given an initial candidate vector, LAS searches among its neighbor vectors for those whose Euclidean distance to the received vector are shorter. If at least one is found, the closest vector is considered the new candidate and another iteration is performed, otherwise, the algorithm stops, and the candidate is chosen as the output vector. In this letter, the neighbor vectors are defined as the set of vectors that differ one bit from the candidate vector in its binary representation, and all the neighbor vectors are searched in parallel. Aside from the complexity of finding the initial vector, the complexity per bit is \( \mathcal{O}(MK) \) for the LAS algorithm [20]. The complexity and error performance of LAS are very sensitive to the initial vector. For \( m > 4 \) in fully-loaded scenario, ZF-LAS (LAS
with a ZF initial solution) does not converge to the optimal solution (MMSE-LAS does approach optimal performance for \( m = 4 \)). However, by using WLZF as the starting point, WLZF-LAS is able to find a near-optimal solution without the need for noise variance knowledge.

III. PERFORMANCE EVALUATION AND DISCUSSION

Computer simulations were performed to evaluate WLZF and WLZF-LAS in large MU-MIMO systems with an i.i.d. Rayleigh fading channel. Results are presented for a 128×128 fully-loaded case, and for up to 128×64 in over-loaded scenarios, for uncoded and coded systems. The proposed techniques allow even larger MIMO sizes due to their low complexity. However, the number of streams for which CSIR estimation can be performed in one channel coherence period limits the number of antennas (although the channel coherence time could be relatively long for some static or low-mobility IoT devices). Because the network density is expected to keep increasing [21], CSIR estimation for a large number of antennas is an active research area and some promising approaches to this problem have been presented, e.g., pilot reuse [22]. Due to the limited space, the case with imperfect CSIR is not considered, and is a subject for future research.

The performance of WLZF has been evaluated for small MIMO systems in [4]. This letter focuses on large MIMO systems, mainly because of two reasons: 1) the performance of WLZF improves as the MIMO size increases, given a fixed \( \alpha \) (recall that \( \alpha = K/\sqrt{M} \)), due to the increased diversity order; and 2) no other approaches (e.g., [21]) seem to be feasible (as far as the authors are aware) for very large over-loaded MIMO systems, due to their high complexity.

**Fully-loaded uncoded large MU-MIMO:** The BER vs. \( \varepsilon_b/N_0 \) results are exhibited in Fig. 1 for an uncoded fully-loaded scenario with 128×128 antennas and other supporting cases. The following plots are shown: a) ZF and ZF-LAS with BPSK and QPSK, b) WLZF with BPSK, 4-ASK, 8-ASK, and 16-ASK (both simulated and theoretical), c) WLZF-LAS with BPSK and 4-ASK, d) the lower bound 1×1 case with an AWGN channel, and finally, e) ZF with QPSK (which is the same as for BPSK) for a half-loaded 64×128 system. Several observations can be obtained from Fig. 1: 1) WLZF drastically outperforms ZF, as the former achieves a very high diversity (which is 64.5 for 128×128), but the latter obtains only unitary diversity; ii) for WLZF, a gap of approximately 3 dB in \( \varepsilon_b/N_0 \) remains in terms of achieving the AWGN bound. However, through the use of WLZF-LAS, near-optimal performance can be achieved; iii) WLZF-LAS outperforms ZF-LAS, especially when \( m > 2 \), because the latter does not approach optimal detection; iv) for WLZF, each increase in modulation order requires an additional 4 dB approximately in \( \varepsilon_b/N_0 \) to obtain the same BER, which is inherent to the use of M-ASK; v) the theoretical equations (8) and (9) for WLZF are seen to fit the simulation results closely; a noticeable mismatch can be seen in low \( \varepsilon_b/N_0 \) values for 8-ASK and 16-ASK, which is due to the approximation in (9); and vi) the channel matrices have the same dimensions (in their real representation) for the half-loaded 64×128 ZF with QPSK and the fully-loaded 128×128 WLZF with 4-ASK; however, there is a gap of approximately 7 dB (4 dB if BPSK is used) in \( \varepsilon_b/N_0 \); this means that twice the number of transmit antennas can be supported using WLZF at the expense of additional power.

**Over-loaded uncoded large MU-MIMO:** The performance evaluation of WLZF and WLZF-LAS in over-loaded scenarios in the range 1 ≤ \( \alpha \) ≤ 2 can be observed in Fig. 2. Results are shown for system settings using 4-ASK modulation and with increasing values of \( \alpha = \{1, 1.5, 1.875, 2\} \), with \( M = 64 \). Also included is the 4-ASK 1×1 AWGN lower bound that is valid only when \( \alpha = 1 \), because lower bounds for the other values of \( \alpha \) are not known. From Fig. 2, it can be seen that as \( \alpha \) increases, additional power is required for successful detection, and the diversity gradually decreases to the point where it becomes 0.5 for \( \alpha = 2 \), because the equivalent real channel matrix \( \tilde{H} \) becomes square. The results demonstrate that simple linear detection is capable of obtaining good performance for over-loaded systems with an unprecedented number of antennas.

1 It is worth noting that in this case, the corresponding curves used for comparison have different data rates. The same case applies to Fig. 2.
Coded large MU-MIMO: Further simulations were run for coded systems using WLZF and WLZF-LAS detection with 4-ASK modulation. Each transmitted stream was independently coded using a turbo code with a code rate $R_c = 1/3$, 6 iterations, and a frame size corresponding to 100 uncoded symbols. Each detection method was tested in a fully-loaded 64×64 and an over-loaded 96×64 scenario. Hard decision was used as the output of the WLZF detector. The results are shown in Fig. 2, including a coded SISO AWGN curve as reference.

It can be noted that for the fully-loaded case, a gap of 3 dB exists between uncoded WLZF and WLZF-LAS, but for the coded case, the gap is reduced to approximately 0.9 dB at a BER of $10^{-4}$. Similarly, for the over-loaded scenario the gap is reduced from 5 to 2 dB from the uncoded to the coded scenarios. These results suggest that in coded systems where computational complexity is an issue, high reliability can still be obtained using simple WLZF.

IV. CONCLUSION

This letter presented some advantages of using M-ASK in large MU-MIMO systems. Using WLZF, linear detection with low complexity is enabled in both fully-loaded and over-loaded systems. WLZF was found to offer a large spatial diversity gain in exchange for power efficiency. Additionally, using WLZF as an initial solution to the LAS algorithm, near-optimal BER can be achieved without the need for noise variance knowledge. Finally, it was observed that in coded systems, the gap in error performance between WLZF and WLZF-LAS is reduced. The results suggest use of M-ASK can lead to efficient receivers in highly dense networks.

APPENDIX – DIVERSITY OF WLZF

The estimated signal after applying ZF is given by

$$\hat{x} = x + (H^*H)^{-1}H^*z = x + \tilde{z},$$

where $(\cdot)^*$ denotes the conjugate transpose operation. Signal $\hat{x}$ can be seen as the original signal $x$ corrupted by colored noise $\tilde{z}$. Because the covariance matrix of $\tilde{z}$ is $\sigma^2(H^*H)^{-1}$, the SINR of the $k$-th transmitted symbol is given by

$$\gamma_k = \gamma \left[(H^*H)^{-1}\right]_{kk},$$

where $[A]_{kk}$ refers to the $k$-th diagonal element of $A$.

Following the derivation in [16], (11) can be written in the form

$$\gamma_k = \gamma h_k^*QAQh_k,$$

where $h_k$ denotes the $k$-th column of $H$, $Q$ is a unitary matrix, and $\Lambda$ is a diagonal matrix. When $H$ is complex, $\Lambda$ contains $M - K + 1$ diagonal elements equal to one and the rest are zero; hence, $\gamma_k^{(\text{ZF})}/\gamma$ follows a chi-squared distribution with $2M - 2K + 2$ degrees of freedom (DoF) [16]. It can be observed that because $H$ is complex, $h_k$ corresponds to 2 column vectors of $H^R$ in (2), as the second half of its column vectors are formed by the elements of the first half of the column vectors.

The above description is applicable for conventional ZF. For WLZF, the channel matrix is $\tilde{H}$ in (4); it is real-valued and is formed from only the first half of the columns of $H^R$, and therefore, $h_k$ consists of only one column in $\tilde{H}$. Because of this, for WLZF, $\Lambda$ has $2M - K + 1$ diagonal elements equal to one, and the rest are zero. Thus, $\gamma_k^{(\text{WLZF})}/\gamma$ is $\chi^2$-distributed with $2L = 2M - K + 1$ DoF, as each increase in 1 DoF accounts for a 0.5 increase in diversity [16]; it follows that the diversity for WLZF is $L^{(\text{WLZF})} = M - K/2 + 1/2$.

REFERENCES