Low-Complexity MIMO Precoding with Discrete Signals and Statistical CSI

Yongpeng Wu, Chao-Kai Wen, Derrick Wing Kwan Ng, Robert Schober, and Angel Lozano

Abstract—In this paper, we investigate the design of multiple-input multiple-output single-user precoders for finite-alphabet signals under the premise of statistical channel-state information at the transmitter. Based on an asymptotic expression for the mutual information of channels exhibiting antenna correlations, we propose a low-complexity iterative algorithm that radically reduces the computational load of existing approaches by orders of magnitude with only minimal losses in performance. The savings increase with the number of transmit antennas and with the cardinality of the signal alphabet, making it possible to support values thereof that were unwieldy in existing solutions.

I. INTRODUCTION

Although Gaussian signals are capacity-achieving in a multiple-input multiple-output MIMO channel under perfect channel-state information (CSI) at the receiver, signals conforming to discrete constellations are transmitted in practice, and the design of precoders optimized for such signal formats is a topic that has gathered momentum in recent years [1–9]. The works in [3–9] consider the problem under the assumption of perfect CSI at the transmitter, which is a reasonable premise in reciprocal or slow fading channels. Often though, perfect CSI at the transmitter is an impossibility and only statistical CSI can be available therein; these are the conditions on which we concentrate here. For Gaussian signals, the design of single-user precoders with statistical CSI has been addressed in [10–15]. For discrete signals, an iterative precoding algorithm was proposed in [16] and shown to achieve a high ergodic spectral efficiency in simulations.

The complexity of this complete-search algorithm is exponential in the number of transmit antennas and, even with modest numbers thereof (say, eight), it becomes unwieldy. The alternative algorithm proposed in this paper drastically reduces the search space, and with it the complexity, but in such a way that the loss in performance—established by means of the 3GPP spatial channel model (SCM) [21]—is minimal.

The following notations are adopted throughout the paper: diag{b} denotes a diagonal matrix containing in its main diagonal the entries of vector b, diag{B} denotes a diagonal matrix containing the diagonal of matrix B, vec(A) is a column vector containing the stacked columns of matrix A, [A]_{m,n} denotes the (m,n)th entry of matrix A, and [a]_{m} denotes the mth entry of vector a.

II. SIGNAL MODEL

Consider a single-user MIMO channel where transmitter and receiver are equipped with N_t and N_r antennas, respectively. The received signal \( y \in \mathbb{C}^{N_r} \) can be written as

\[
y = Hx + n
\]

where \( H \in \mathbb{C}^{N_r \times N_t} \) is a random channel matrix whose \( (i,j) \)th entry denotes the complex fading coefficient between the \( j \)th transmit and the \( i \)th receive antenna, \( x \in \mathbb{C}^{N_t} \) denotes the zero-mean transmitted vector with covariance \( \Sigma_x \), \( n \in \mathbb{C}^{N_r} \) is the zero-mean complex Gaussian noise vector with covariance \( I_{N_r} \). The transmit vector \( x \) satisfies the power constraint

\[
x^H \Sigma_x x \leq P.
\]

Based on the statistical CSI, and subject to this constraint, the transmitter needs to optimize \( \Sigma_x \) to maximize the ergodic spectral efficiency.

With \( H \) known at the receiver, the ergodic mutual information between \( x \) and \( y \) is given by [17]

\[
I(x; y) = E \left[ \log \frac{p(y|x,H)}{p(y|H)} \right].
\]

where the outer expectation is over \( H \) and the inner expectation is over \( p(x,y|H) \).

III. COMPLETE-SEARCH PRECODER DESIGN

In this section, we review the complete-search approach and introduce an idea to reduce its computational load.

Let \( x = B d \), where \( d \in \mathbb{C}^{N_t \times 1} \) is the signal vector drawn from an equiprobable constellation of size \( M^{N_t} \) whereas \( B \in \mathbb{C}^{N_r \times N_t} \) is the precoder. Let \( d_m \) denote the \( m \)th element in the constellation. Consider the singular value decomposition (SVD) \( B = U_B \Lambda_B V_B^H \) where \( \Lambda_B \in \mathbb{C}^{N_r \times N_t} \) is diagonal while \( U_B \in \mathbb{C}^{N_r \times N_t} \) and \( V_B \in \mathbb{C}^{N_t \times N_t} \) are unitary.

When Gaussian-signal precoding solutions are applied to discrete constellations, the performance suffers because, in the face of major power variations between MIMO subchannels, these solutions insist on beamforming over an extensive range of SNRs, well beyond the point where beamforming is appropriate for a discrete constellation. With beamforming, signalling is only possible over the dominant subchannel.
which causes a performance loss with discrete signals (cf. [4, 16]). By properly designing $U_B$, $A_B$, and $V_B$, the complete-search precoder design minimizes this loss [4, 16]. The matrix $V_B$ mixes the $N_t$ original signals into $N_t$ beams, then $A_B$ allocates power to those beams, and finally $U_B$ aligns them spatially as they are launched onto the channel. With a proper choice of $V_B$, all the $N_t$ signals can be effectively transmitted even if a single beam is active.

The next example illustrates the role of $U_B$, $A_B$, and $V_B$.

**Example 1:** Consider a $4 \times 4$ deterministic channel $A$ with $\text{SVD } A = U_A A_B V_A$. The received signal is given by

$$y = A U_B A_B V_B d + n$$  \hspace{1cm} (4)

where $d = [d_1, d_2, d_3, d_4]^T$. From [4, Prop. 2], the optimal design satisfies $U_B = V_B^H$. Then, based on [4, Eq. (8)], (4) can be rewritten as

$$\bar{y} = \begin{bmatrix} a_1 \lambda_1 \\ \vdots \\ a_4 \lambda_1 \end{bmatrix} \begin{bmatrix} V_{11} & \ldots & V_{14} \\ \vdots & \ddots & \vdots \\ V_{41} & \ldots & V_{44} \end{bmatrix} d + n$$  \hspace{1cm} (5)

where $\bar{y} = U_B^H y$ while $a_i$, and $\lambda_i$ are the diagonal entries of $A_A$ and $A_B$, respectively, and $V_{ij} = [V]_{ij}$.

Assume one of the subchannel gains, say $a_2$, is very weak. Then, with a Gaussian-signal precoder, the power allocated to the corresponding subchannel will be very small even at moderate SNRs. Since, with Gaussian signals, $V_B$ is immaterial, $d_2$ then cannot be transmitted. With a proper $V_B$, in contrast, the received signal equals

$$[\bar{y}]_i = a_i \lambda_i \sum_{j=1}^4 V_{ij} d_j \quad i = 1, 2, 3, 4$$  \hspace{1cm} (6)

and now, even if $a_2 \lambda_2 \approx 0$, $d_2$ can still be effectively transmitted along other subchannels.

As indicated by (6), an adequate design for discrete constellations in general mixes all the signals $(d_1, d_2, d_3, d_4)$ and transmit the ensuing beams on different subchannels. As a result, the search space grows exponentially with $N_t$.

Intuitively though, if there is only one weak subchannel, say $a_2$ in Example 1, it is not necessary to mix all the signals. It suffices to mix $d_2$ for instance with $d_1$ and transmit the ensuing beam on the stronger subchannel $a_1$. This corresponds to

$$V = \begin{bmatrix} V_{11} & V_{12} & 0 & 0 \\ V_{21} & V_{22} & 0 & 0 \\ 0 & 0 & V_{33} & V_{34} \\ 0 & 0 & V_{43} & V_{44} \end{bmatrix}$$  \hspace{1cm} (7)

which, plugged into (5), gives

$$[\bar{y}]_i = a_i \lambda_i \sum_{j=1}^2 V_{ij} d_j \quad i = 1, 2$$  \hspace{1cm} (8)

$$[\bar{y}]_i = a_i \lambda_i \sum_{j=3}^4 V_{ij} d_j \quad i = 3, 4.$$  \hspace{1cm} (9)

Observe from (8) and (9) that $(d_1, d_2)$ and $(d_2, d_3)$ are decoupled. If $d$ is drawn from QPSK distributions, then the search space for (8) and (9) is of dimension $2 \times 4^2 = 512$. In contrast, for the complete search in (6), it is of dimension $4^2 \times 4 = 65536$. Since $d_2$ is transmitted all the same, the structure in (7) may perform close to the complete-search design, but with a substantially lower computational complexity.

**IV. LOW-COMPLEXITY PRECODER DESIGN**

**A. Channel Model**

Inspired by (8) and (9), we propose a low-complexity design to maximize the ergodic spectral efficiency in (3). We consider the popular Kronecker channel model [18]

$$H = A_R^{1/2} W A_T^{1/2}$$  \hspace{1cm} (10)

where $A_R \in \mathbb{C}^{N_r \times N_t}$, and $A_T \in \mathbb{C}^{N_t \times N_t}$ are transmit and receive correlation matrices while $W \in \mathbb{C}^{N_r \times N_t}$ is a random matrix whose entries are independent and identically distributed (IID) complex Gaussians. The eigenvalue decompositions of $A_R$ and $A_T$ are

$$A_R = U_R A_R U_R^H$$  \hspace{1cm} (11)

$$A_T = U_T A_T U_T^H.$$  \hspace{1cm} (12)

For this channel model, the optimal left singular matrix $U_B$ of the precoder $B$ equals $U_T$ [16]. From this, using [16, Eq. (5)], and recalling (1) and (10), we can rewrite (1) as

$$y_{eq} = H_{eq} x_{eq} + \tilde{n}$$  \hspace{1cm} (13)

where

$$x_{eq} = A_B V_B d$$  \hspace{1cm} (14)

$$H_{eq} = A_R^{1/2} W A_T^{1/2}$$  \hspace{1cm} (15)

and where $\tilde{n}$ and $W$ have the same distributions as $n$ in (1) and $W$ in (10), respectively.

**B. Mutual Information in the Large-Dimensional Regime**

In order to obtain counterparts to (8) and (9) for this setting, we move into the large-dimensional regime [19]. When both $N_r$ and $N_t$ grow large with ratio $c = N_t / N_r$, the mutual information in (3) satisfies [19]

$$I(x; y) \approx I_{Asy}(x; y)$$  \hspace{1cm} (16)

where

$$I_{Asy}(x; y) = I(x_{eq}; z_{eq}) + \log_2 \det (I_{N_r} + R_{eq}) - \gamma_{eq} \psi_{eq} \log_2 e.$$  \hspace{1cm} (17)

given the diagonal MIMO relationship

$$z_{eq} = \Xi_{eq}^{1/2} x_{eq} + \tilde{n}$$  \hspace{1cm} (18)
where \( \tilde{n} \in \mathbb{C}^{N_t} \) is a standard complex Gaussian random vector while

\[
\begin{align*}
\Xi_{eq} &= \gamma_{eq} \Lambda_T \\
\mathbf{R}_{eq} &= \psi_{eq} \Lambda_R \\
\gamma_{eq} &= \text{tr} \left( (\mathbf{I}_{N_t} + \mathbf{R}_{eq})^{-1} \Lambda_R \right) \\
\psi_{eq} &= \text{tr} \left( \Omega_{eq} \Lambda_T \right)
\end{align*}
\]

The diagonal relationship in (18) does not relate to any physical channel, but it is merely an instrument to obtain an asymptotic expression for the mutual information. We shall take advantage of this relationship.

Also necessary for later derivations is the MMSE estimate of \( \mathbf{x}_{eq} \) based on (18), which is given by

\[
\hat{x}_{eq} = E \left[ x_{eq} | z_{eq} \right].
\]

It will be convenient to define the following MMSE matrix as the covariance of the error vector between the transmitted signal and its estimate,

\[
\Omega_{eq} = E \left[ (x_{eq} - \hat{x}_{eq})(x_{eq} - \hat{x}_{eq})^H \right].
\]

\[C. Precoder Structure\]

Let us divide the transmit signal \( \mathbf{d} \) into \( S \) streams. Each stream \( \mathbf{d}_s \in \mathbb{C}^{N_r \times 1} \) is to be conveyed over \( N_s = N_t / S \) diagonal entries of \( \Xi_{eq} \). Let the set \( \{1, \ldots, N_t\} \) denote a permutation of \( \{1, \ldots, N_t\} \) and let \( \Lambda_s \in \mathbb{C}^{N_t \times N_t} \) and \( \mathbf{V}_s \in \mathbb{C}^{N_r \times N_r} \) denote a diagonal matrix and a unitary matrix, respectively, for \( s = 1, \ldots, S \). The goal of arranging these \( S \) streams as in (8) and (9) prompts the following design steps:

1) Structure of \( \Lambda_B \): We define

\[\Lambda_{B} |_{\ell_s, \ell_j} = [\Lambda_s]_{ij}\] (25)

where \( i = 1, \ldots, N_s, \; s = 1, \ldots, S \) and \( j = (N_s - 1)s + i \). Under this structure, the \( s \)th stream is transmitted along the \( \ell_{(N_s - 1)s + 1}, \ldots, \ell_{(N_s - 1)s + N_r} \) diagonal entries of \( \Xi_{eq} \).

2) Structure of \( \mathbf{V}_B \): We define

\[\mathbf{V}_{B} |_{\ell_s, \ell_j} = \begin{cases} 
[\mathbf{V}_{s}]_{mn} & \text{if } i = (N_s - 1)s + m, \; j = (N_s - 1)s + n \\
0 & \text{otherwise}
\end{cases}\] (26)

where \( m = 1, \ldots, N_r, \; n = 1, \ldots, N_s, \; s = 1, \ldots, S, \; i = 1, \ldots, N_t \) and \( j = 1, \ldots, N_t \). Under this structure, for the \( s \)th stream the entries of \( \mathbf{V}_s \) map only to rows \( \ell_{(N_s - 1)s + 1}, \ldots, \ell_{(N_s - 1)s + N_r} \) and columns \( \ell_{(N_s - 1)s + 1}, \ldots, \ell_{(N_s - 1)s + N_r} \) of \( \mathbf{V}_B \). This yields \( S \) decoupled groups of streams at the receiver.

The design in (7) is a specific instance of (26) with \( \{\ell_1, \ldots, \ell_{N_t}\} = \{1, 2, 3, 4\} \) and \( S = 2 \). Recall how \( (d_1, d_2) \) and \( (\hat{d}_3, \hat{d}_4) \) are indeed decoupled in (8) and (9).

3) Structure of \( \mathbf{d}_s \): Finally, we let

\[
[d_s]_s = [d]_{\ell_s}.
\]

\[D. Precoder Optimization\]

Based on (25)–(27), the relationship in (14) can be rewritten as

\[
[x_{eq}]_{\ell_{j_s}} = [\Lambda_s \mathbf{V}_s \mathbf{d}_s]_{\ell_{j_s}}
\]

for \( i = 1, \ldots, N_s, \; s = 1, \ldots, S \) and \( j = (N_s - 1)s + i \). Recalling that \( \Xi_{eq} \) is diagonal, (18) then reduces to

\[
[z_{eq}]_{\ell_{j_s}} = [\Xi_{eq}]_{\ell_{j_s}} [x_{eq}]_{\ell_{j_s}} + [\tilde{n}]_{\ell_{j_s}}.
\]

Eqs. (28) and (29) indicate that each independent data stream \( \mathbf{d}_s \) is transmitted along its own \( N_s \) separate subchannels without interfering with other streams. Furthermore, the MMSE matrix in (24) then equals

\[
\Omega_{eq} |_{\ell_{j_s}, \ell_{j_s}} = \begin{cases} 
[\Xi_{eq}]_{\ell_{j_s}, \ell_{j_s}} & \text{if } i = (N_s - 1)s + m, \; j = (N_s - 1)s + n \\
0 & \text{otherwise}
\end{cases}
\]

(30)

where \( \Xi_s = \Lambda_s \mathbf{V}_s \mathbf{V}_s^H \Lambda_s^H \)

\[\mathbf{d}_s = E \left[ d_s | z_s \right].
\]

and \( [z_s]_{\ell_{j_s}} = [z_{eq}]_{\ell_{j_s}}, \) and further defining diagonal matrices \( \Xi_s \), for \( s = 1, \ldots, S \), with entries \( [\Xi_s]_{\ell_{j_s}} = [\Xi_{eq}]_{\ell_{j_s}} \).

The main term in the mutual information in (16) is

\[
I(x_{eq}; z_{eq}) = \sum_{s=1}^{S} I(d_s; z_s)
\]

(34)

based on which the gradients of \( I_{a,v_s}(x; y) \) with respect to \( \Lambda_s^2 \) and \( \mathbf{V}_s \) are given by \( [20, \text{Eq. (22)}] \)

\[
\nabla_{\Lambda_s^2} I_{a,v_s}(x; y) = \text{diag} \left( \mathbf{V}_s^H \mathbf{E}_s \mathbf{V}_s \xi_s \right)
\]

(35)

\[
\nabla_{\mathbf{V}_s} I_{a,v_s}(x; y) = \xi_s \Lambda_s^2 \mathbf{V}_s \mathbf{E}_s
\]

(36)

where

\[
\mathbf{E}_s = E \left[ (d_s - \hat{d}_s)(d_s - \hat{d}_s)^H \right].
\]

(37)

From (34), and from the relationship between \( \Lambda_1, \ldots, \Lambda_S \) and \( \mathbf{A}_B \) in (25) as well as the relationship between \( \mathbf{V}_1, \ldots, \mathbf{V}_S \) and \( \mathbf{V}_B \) in (26), we propose Algorithm 1 to optimize \( \Lambda_B \) and \( \mathbf{V}_B \).

Remark 1: For the complete-search design algorithm [4, 16], the complexity is dominated by the computation of the mutual information and the MMSE matrix, which grows exponentially with \( 2N_t \). For Algorithm 1, alternatively, the complexity of computing the mutual information and the MMSE matrix in Algorithm 1 grows exponentially with \( 2N_t \). Thus, by choosing proper values of \( S \) and \( N_s \), Algorithm 1 offers a tradeoff between performance and complexity. At one end, when \( S = 1 \) and \( N_s = N_t \), Algorithm 1 searches the entire space, while at the other end, when \( S = N_t \) and \( N_s = 1 \), Algorithm 1 merely allocates power among the \( N_t \) parallel subchannels. The selec
TABLE I: Running time (sec.) per iteration with BPSK.

<table>
<thead>
<tr>
<th>$N_t$</th>
<th>$N = 2$</th>
<th>$N_r = 4$</th>
<th>$N_s = N_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.0051</td>
<td>0.0190</td>
<td>0.0190</td>
</tr>
<tr>
<td>8</td>
<td>0.0112</td>
<td>0.0473</td>
<td>11.6209</td>
</tr>
<tr>
<td>16</td>
<td>0.0210</td>
<td>0.1939</td>
<td>×</td>
</tr>
<tr>
<td>32</td>
<td>0.0570</td>
<td>0.4111</td>
<td>×</td>
</tr>
</tbody>
</table>

Algorithm 1: Maximization of $I(x; y)$ with respect to $B$.

1) Initialize $\Lambda_s^{(0)}$, $V_s^{(0)}$ for $s = 1, \ldots, S$. Fix a maximum number of iterations, $N_{\text{iter}}$, and a threshold $\varepsilon$.
2) Initialize $\Xi_{\text{eq}}$, $R_{\text{eq}}$, $\gamma_{\text{eq}}$ and $\psi_{\text{eq}}$ based on (19)–(22), with $\Omega_{\text{eq}}$ based on (30). Then, initialize $I^{(1)}(x; y)$ based on (16) with $I(x; y_{\text{eq}})$ as per (34). Set counter $n = 1$.
3) Update $\Lambda_s^{(n)}$ for $s = 1, \ldots, S$ along the gradient descent direction given by (35).
4) Normalize $\sum_{s=1}^{S} [\Lambda_s^{(n)}]^2 = P$.
5) Update $V_s^{(n)}$ for $s = 1, \ldots, S$ along the gradient descent direction in (36).
6) Update $\Xi_{\text{eq}}$, $R_{\text{eq}}$, $\gamma_{\text{eq}}$ and $\psi_{\text{eq}}$ based on (19)–(22), (30).
7) Compute $I^{(n+1)}(x; y)$ based on (16) and (34). If $I^{(n+1)}(x; y) - I^{(n)}(x; y) > \varepsilon$ and $n \leq N_{\text{iter}}$, set $n = n + 1$ and repeat Steps 3–7;
8) Compute $\Lambda_B$ and $V_B$ based on (25) and (26). Set $B = U_T \Lambda_B V_B$.

- tion of $N_s$ from 1 to $N_t$ bridges the gap between separate and fully joint transmission of the $N_t$ original signals.

Remark 2: An adequate choice of $\ell_1, \ldots, \ell_{N_t}$ is important for Algorithm 1 to performance satisfactorily. As discussed in Sec. III, the $N_s/2$ largest diagonal entries of $\Xi_{\text{eq}}$ are paired with the $N_s/2$ smallest diagonal entries. Then, the remaining $N_s/2$ largest diagonal entries of $\Xi_{\text{eq}}$ are paired with the remaining $N_s/2$ smallest ones, and so on. This generalizes the two-antenna scheme in [5].

Remark 3: In Steps 3 and 5, $\Lambda_s^{(n)}$ and $V_s^{(n)}$ are updated along the gradient descent direction, with the backtracking line search method used to determine the step size. Hence, in Step 7 the mutual information $I^{(n)}(x; y)$ is nondecreasing. Since Algorithm 1 generates sequences that are nondecreasing and upper-bounded, it is convergent. However, due to the nonconvexity of $I^{(n)}(x; y)$ in $\Lambda_s^{(n)}$ and $V_s^{(n)}$, Algorithm 1 may only find local optima. As a result, the algorithm is run several times with different random initializations of $\Lambda_s^{(n)}$ and $V_s^{(n)}$ and the final precoder that provides the highest mutual information is retained.

V. PERFORMANCE EVALUATION

First, let us evaluate the complexity of Algorithm 1 for different values of $N_s$. Matlab is used on an Intel Core i7-4510U 2.6GHz processor. Tables I–III provide the running time per iteration, for various numbers if antennas and constellations, with $\times$ indicating that the time exceeds one hour. See how, for $N_s = N_t$, the computational complexity grows exponentially with $N_t$ and quickly becomes unwieldy.

Figure 1 depicts the spectral efficiency in the 3GPP SCM (urban scenario, half-wavelength antenna spacing, velocity 36 km/h) for different precoder designs with $N_t = N_r = 4$ and QPSK. A Gauss-Seidel algorithm using stochastic programming is employed to obtain the capacity-achieving precoder [13]. For Algorithm 1, both $N_s = 4$ and $N_s = 2$ are considered, and despite their enormous computational gap (cf. Table II) the difference in performance is minor. Both these reduced-complexity precoders hug the capacity up to the point where the QPSK cardinality becomes insufficient, gaining many dB over an unprecoded transmitter and also over a capacity-achieving precoder applied with QPSK.

Figure 2 contrasts the spectral efficiency given by the asymptotic expression in (16) with the exact form in (3) for the precoders obtained by Algorithm 1 with $N_s = 2$. The channel model is as in Fig. 1. The perfect match between the two confirms that (16) is a very good proxy for (3), and hence that Algorithm 1 is indeed effective even for small numbers of antennas.

Figures 3 and 4 present further results for $N_t = N_r = 32$ and for 16-QAM, respectively. We set $N_s = 4$ for the former and $N_s = 2$ for the latter. (Precoder design for such large arrays were, to best of the authors’ knowledge, not available.
henceforth for discrete signals.) As the numbers of antennas grow, the conditioning of the transmit correlation matrix $A_T$ becomes progressively poorer [22] and the performance of the capacity-achieving precoder applied to discrete signals degrades, even failing to achieve the saturation spectral efficiency of $N_t \log_2 M \text{ b/s/Hz}$ at relevant SNRs; some subchannels are simply never activated by a precoder intended for Gaussian signals. Algorithm 1, in contrast, is tailored to finite-cardinality constellations.

VI. CONCLUSION

With a proper design of $V_B$ (right unitary matrix in the SVD decomposition of the precoder), it is possible to achieve a satisfactory tradeoff between the need to feed into the channel mixings of multiple finite-cardinality signals and the computational complexity of exploring all possible such mixings. Building on this idea, an algorithm has been proposed that—under the 3GPP SCM channel model—exhibits very good performance with orders-of-magnitude less complexity than complete-search solutions. More refined versions of this algorithm, equipped with alternative subchannel pairing schemes, may perform even better. Additional extensions include the applicability to settings with CSIT, or to multiuser contexts, as well as the performance under other channel models.

REFERENCES


