Study on precoding optimization algorithms in massive MIMO system with multi-antenna users

Evgeny Bobrov a,b, Dmitry Kropotov b,c, Sergey Troshin c and Danila Zaeva a

aMoscow Research Center, Huawei Technologies, Moscow, Russia; bM.V. Lomonosov Moscow State University, Moscow, Russia; cNational Research University Higher School of Economics, Moscow, Russia

ABSTRACT
The paper studies the multi-user precoding problem as a non-convex optimization problem for wireless multiple inputs and multiple outputs (MIMO) systems. In our work, we approximate the target Spectral Efficiency function with a novel computationally simpler function. Then, we reduce the precoding problem to an unconstrained optimization task using a special differential projection method and solve it by the Quasi-Newton L-BFGS iterative procedure to achieve gains in capacity. We are testing the proposed approach in several scenarios generated using Quadriga-open-source software for generating realistic radio channel impulse response. Our method shows monotonic improvement over heuristic methods with reasonable computation time. The proposed L-BFGS optimization scheme is novel in this area and shows a significant advantage over the standard approaches. The proposed method has simple implementation and can be a good reference for other heuristic algorithms in this field.

1. Introduction

Wireless channels with multiple inputs and multiple outputs (MIMO) provide significantly more capacity than their counterparts with one input and one output. Thus, the MIMO system is an essential technology in modern wireless telecommunications, including Wi-Fi and 5G systems [1]. It allows using antenna array clusters to send multiple signal beams for multiple user devices simultaneously. The proper construction of such beams is called the beamforming or precoding procedure [8,11]. In the linear channel assumptions, the correct precoding is a complex matrix of the general form with given constraints, which corresponds to the physical limitations of the system.

We measure the quality of the precoding using the well-known functions such as Signal-to-Interference-and-Noise Ratio (SINR) and Spectral Efficiency (SE) [16]. In addition to these functions, the literature considers minimum MSE (MMSE) [10,19]) function, and the discrete bit-error-rate (BER) [6] quality measure. On the basis of MMSE function the most precoding algorithms in the field are derived [2].
The standard precoding algorithms, which are well known in the literature, are maximum-ratio transmission (MRT) [9], Zero-Forcing (ZF) [20] and regularized ZF (RZF) [17] (including its recent variant – adaptive RZF (ARZF) [2]). All these algorithms have analytical formulas, without taking into account the target function of SE or do it implicitly, maximizing the numerator of SINR using the MRT algorithm, or reducing the denominator of SINR using the ZF algorithm. This leads to simple but non-optimal precoding solutions. Therefore, one of our goals is to study the potential improvement of precoding methods.

While during downlink (DL) on the transmitter side, a base station applies precoding matrices, symmetrically, on the receiver side users apply detection matrices. The knowledge of the detection of the user allows to construct more efficient precoding for the base station, see e.g. [14]. In general, we do not have exact information about the user detection, but we can mean a type of it. The most common detections are well described in the literature, such as minimum MSE (MMSE) [10,19] and Interference Rejection Combiner (MMSE-IRC) [13], which can be applied in order to eliminate the multiuser interference. We also utilize the recently proposed theoretical conjugated detection (CD) [2].

We consider the most practical formulation of the problem with multiple-antenna users [15] and per-antenna power constraints at the base station [18]. Alternatively, simpler and mathematically attractive statements include single-antenna users and single total power constraint at the base station. Thus, the considered problem is constrained optimization programming with the non-convex function and multiple constraints. It can be solved by using suboptimal heuristic methods [9,20], interior point methods [5], projection-based methods [4], or any of them with reparametrization (variables transform). We find a convenient reparametrization that reduces the problem to an unconstrained one. This reduction allows to use advanced unconstrained gradient-based methods. The proper reparametrization is one of the main contributions of the article. Even with the right reparametrization, the target function is computationally complex and must be simplified to accelerate convergence. Proper simplification is the second and new approach presented in this article.

In this study we, firstly, approximate the SE [16] objective function using the CD matrix [2] to simplify our algorithm as much as possible. As the second step, we reduce the precoding problem to an unconstrained optimization task and solve it by the quasi-Newton L-BFGS iterative procedure [22] to achieve the maximum possible quality. We write an end-to-end ‘differentiable projection’-based method, where gradients are taken both for the functional and the projection, which tends to very fast convergence of the proposed method. The problem to be solved is not convex, and so we find the local maximum of the SE.

The novelty of the proposed method consists in simplifying the objective function SE, which makes the task more attractive from a computational point of view. Secondly, the reparametrization of precoding for an unlimited task was invented. We called this parametrization Differentiable projection. In addition, classical parameterization using the Softmax function is also described.

Our parametrization method is a know-how and is designed to solve the problem of precoding. In fact, it can be generalized to other topics of constrained optimization. In combination with the simplified function, we proposed the target problem of unlimited
optimization as a reference algorithm for evaluating the upper bound of quality. Finally, all our research is based on realistic power limits per antenna and multiple antenna users, which in itself is a separate problem for finding the right approach to solving.

All investigated algorithms were tested in several scenarios of Quadriga [7] – an open-source software for generating realistic radio channel impulse response. Our method shows monotonic improvement over heuristic methods with reasonable computational time.

We adopt commonly used notational conventions throughout the paper. Matrices and vectors will be denoted by bold-face upper and lower case letters, respectively. Furthermore, $(\cdot)^H$ denotes Hermitian transpose.

The paper is organized as follows. In Section 2 we describe the system model of the studied massive MIMO network, introduce quality measures, types of detection matrices, approximated quality function, power constraints and problem statement. In Section 3 we study solutions for the precoding problem, reference methods, proposed optimization solution and its relative computational complexity. In Section 4 we provide numerical results in the Quadriga radio simulator. Section 5 contains the conclusion.

2. System model

In 5G cellular networks, we look at a precoding problem in multi-user massive MIMO communication. A single base station with multiple transmit antennas is used. These antennas send signals to multiple users at the same time. Each user also has multiple antennas for receiving signals. The transmission quality between the base station and the users is measured by the base station. The precoding problem entails determining an appropriate weighting (phase and gain) for transmitting the signal in order to maximize the signal power at the receivers’ output.

In our system we consider $K$ users and for the $k$th user we would like to transmit $L_k$ symbols. In total, we would like to transmit a vector $x \in \mathbb{C}^L$, where $L = L_1 + \cdots + L_K$. We multiply the transmitted vector by a precoding matrix $W \in \mathbb{C}^{T \times L}$, where $T$ is a total number of transmit antennas on a base station. Then we transmit the precoded symbols for the all users. Suppose that the $k$th user has $R_k$ receive antennas and $H_k \in \mathbb{C}^{R_k \times T}$ is a channel between the $k$th user and the base station. Then the $k$th user receives $H_k W x + n_k$, where $n_k$ is a Gaussian noise. The noise appears as a result of thermal distortions of the system. Finally, the $k$th user applies a detection of transmitted symbols by multiplying the received vector by a detection matrix $G_k \in \mathbb{C}^{L_k \times R_k}$. The final vector of detected symbols of the all users is denoted by $r \in \mathbb{C}^L$. The whole process of transmitting symbols is presented in Figure 1. Note that the linear precoding and detection can be implemented by simple matrix multiplications.

Finally, the multi-user MIMO model is described using the following linear system:

$$r = G(HWx + n),$$

where $r \in \mathbb{C}^L$ is a received vector, $x \in \mathbb{C}^L$ is a sent vector, $H \in \mathbb{C}^{R \times T}$ is a channel matrix, $W \in \mathbb{C}^{T \times L}$ is a precoding matrix, $G \in \mathbb{C}^{L \times R}$ is a block-diagonal detection matrix and $n \sim \mathcal{CN}(0, I_L)$ is a noise-vector. The constant $T$ is the number of transmit antennas, $R$ is the total number of receive antennas and $L$ is the total number of transmitted symbols in the system. Usually they are related as $L \leq R \leq T$. Each of the matrices $G, H, W$ decomposes
**Figure 1.** Multi-user precoding allows to transmit different information to different users simultaneously. The problem is to find an optimal precoding matrix \( W \) of the system given the target SE function and the per-antenna power constraints (16).

by \( K \) users. The numbers \( K, L, R, T \) and channel matrix \( H \in \mathbb{C}^{R \times T} \) are predefined in the system.

It is convenient [15] to represent the user channel matrix \( k \) via its reduced singular value decomposition (SVD):

\[
H_k = U_k^H S_k V_k, \quad U_k U_k^H = I_{R_k}, \quad S_k = \text{diag}(s_1, \ldots, s_{R_k}), \quad V_k V_k^H = I_{R_k},
\]

where the channel matrix for user \( k, H_k \in \mathbb{C}^{R_k \times T} \) contains channel vectors \( h_i \in \mathbb{C}^T \) by rows, the singular values \( S_k \in \mathbb{C}^{R_k \times R_k} \) are sorted by descending, \( U_k \in \mathbb{C}^{R_k \times R_k} \) is a unitary matrix of left singular vectors, and matrix \( V_k \in \mathbb{C}^{R_k \times T} \) consists of right singular vectors.

Collecting all users together, we can write the following decomposition:

**Lemma 2.1 ([2]):** Denote \( H = [H_1, \ldots, H_K] \in \mathbb{C}^{R \times T} \) the concatenation of individual channel rows \( H_k \). The following representation holds:

\[
H = U^H S V,
\]

where the \( H \in \mathbb{C}^{R \times T} \), and \( S = \text{diag}(S_k) \in \mathbb{C}^{R \times R} \), and \( U = \text{bdig}(U_k) \in \mathbb{C}^{R \times R} \) is block-diagonal unitary matrix, but \( V = [V_1, \ldots, V_K] \in \mathbb{C}^{R \times T} \) has a general form.

**2.1. Quality measures**

The quality functions are based not on the actual sending symbols \( x \in \mathbb{C}^L \), but on some distribution of them [1]. Thus, we get the common functions for all the assumed symbols, which can be sent using the specified precoding. The Signal-to-Interference-and-Noise
(SINR) functional of the $l = l(k)$th symbol and user $k$, is defined as

$$\text{SINR}_l(W, H_k, g_l, \sigma, P) = \frac{|g_l H_k w_l|^2}{\sum_{i \neq l} |g_l H_k w_l|^2 + \|g_l\|^2 \sigma^2}, \quad \forall l \in L_k, \quad (4)$$

where $w_l \in \mathbb{C}^T$ in the $l$th column of the precoding matrix.

To get the formula for the $k$th user SINR, namely the effective SINR, that shows the signal quality of the user, we average the $L_k$ per-symbol SINRs (4) by the geometric mean:

$$\text{SINR}_k^{\text{eff}}(W, H_k, g_k, \sigma, P) = \left( \prod_{l \in L_k} \text{SINR}_l(W, H_k, g_l, \sigma, P) \right)^\frac{1}{L_k}, \quad (5)$$

where the detection matrix $G_k \in \mathbb{C}^{L_k \times R_k}$ contains vector of the $k$th users symbols $g_l$. To get the final functional of Spectral Efficiency we apply Shannon’s formula over all effective user SINRs (5):

$$\text{SE}(W, H, G, \sigma, P) = \sum_{k=1}^{K} L_k \log_2 (1 + \text{SINR}_k^{\text{eff}}(W, H_k, g_k, \sigma, P)) \to \max_W. \quad (6)$$

Formula (6) shows the throughput capacity of the system as a whole. Now it depends on the whole channel and detection matrices $H \in \mathbb{C}^{R \times T}$ and $G \in \mathbb{C}^{L \times R}$. We will use it as our final score function, compare algorithms and find precoding matrices $W \in \mathbb{C}^{T \times L}$, solving the problem of maximizing this function.

Additionally, we consider the function of Single-User SINR:

$$\text{SUSINR}(\tilde{S}, \sigma^2, P) = \frac{P}{\sigma^2} \left( \frac{1}{L_k} \left( \prod_{l \in L_k} \frac{1}{S_k^2} \right) \right)^\frac{1}{L_k}. \quad (7)$$

Formula (7) reflects the quality of the channel for the specified user without taking into account the others. It depends on the greatest $L_k$ singular values $\tilde{S}_k \in \mathbb{R}^{L_k \times L_k}$ of the $k$th user channel matrix $H_k \in \mathbb{C}^{R_k \times T}$. Finally, the matrix $\tilde{S} \in \mathbb{C}^{L \times L}$ contains all singular values on the main diagonal, such as $\tilde{S} = \text{diag}(\tilde{S}_1 \ldots \tilde{S}_K) \in \mathbb{C}^{L \times L}$. We will use this function in our experiments as a universal channel characteristic in our experiments.

### 2.2. Detection matrices

Detection aims to reverse the channel with precoding to an identity matrix. We study the most common forms of the detection matrix $G$, which is called Minimum MSE (MMSE) [10], Interference Rejection Combine (MMSE-IRC) [13] and Conjugate Detection (Conjugate) (CD). In our work we will focus our attention on the MMSE-IRC and CD methods.

Firstly, MMSE detection realizes the following rule:

$$G_k^{\text{MMSE}}(A_k) = A_k^H (A_k A_k^H + \lambda I)^{-1}, \quad A_k = H_k W_k. \quad (8)$$

The scalar $\lambda = \frac{\sigma^2}{P}$ is the system noise-to-signal ratio. The constant $P$ refers to the base station power and $\sigma^2$ refers to the system noise. Thus, the method implicitly considers the
noise equal for all symbols: \( n = \frac{\sigma^2}{P} \mathbf{1} \in \mathbb{C}^L \), and detection \( G \) which tends to eliminate it. One may notice that the use of the method requires solving a system of linear equations of the size of layers \( L \).

Secondly, \textit{MMSE-IRC} detection uses additional information about noise-covariance matrix \( R_{uu}^k \) as follows:

\[
G_{IRC}^k(A_k, R_{uu}^k) = A_k^H (A_k A_k^H + R_{uu}^k + \lambda I)^{-1}, \quad A_k = H_k W_k,
\]

where the covariance matrix has the form as

\[
R_{uu}^k = H_k (WW^H - W_k W_k^H) H_k^H = H_k \left( \sum_{u=1, u\neq k}^K W_u W_u^H \right) H_k^H.
\]

**Lemma 2.2:** Using the linear property, you can get the following representation of the detection matrix \textit{MMSE-IRC}:

\[
G_{IRC}^k(H_k, R_{uu}^k) = (H_k W_k)^H (H_k W_k (H_k W_k)^H + R_{uu}^k + \lambda I)^{-1}
\]

\[
= (H_k W_k)^H (H_k W_k (H_k W_k)^H + H_k \left( \sum_{u=1, u\neq k}^K W_u W_u^H \right) H_k^H + \lambda I)^{-1}
\]

\[
= (H_k W_k)^H (H_k W_k H_k W_k^H + \lambda I)^{-1}.
\]

Thirdly, \textit{Conjugate Detection} can be written in the following form:

\[
G^C := \tilde{S}^{-1} \tilde{U} \in \mathbb{C}^{L_k \times R}, \quad G_k^C = \tilde{S}_k^{-1} \tilde{U}_k \in \mathbb{C}^{L_k \times R_k},
\]

where \( \tilde{S}_k \in \mathbb{C}^{L_k \times L_k} \) contains the first \( L_k \) largest singular values and \( \tilde{U}_k \in \mathbb{C}^{L_k \times R_k} \) contains the first \( L_k \) corresponding singular vectors. The \textit{Conjugate} and \textit{MMSE-IRC} detection matrices are very closely related to each other, assuming low noise and ZF or RZF Precoding (the proof of this fact we put elsewhere). We can use the \textit{Conjugate} detection at the base station to configure the precoding matrix assuming \textit{MMSE-IRC} for user equipments.

### 2.3. Approximated quality function

Assuming the conjugated detection \( G^C \) (12) we have come to the following approximated function of the \textit{SINR} [2]:

\[
SINR_i^C(W, \tilde{v}_l, s_l, \sigma^2, P) = \frac{|\tilde{v}_l w_l|^2}{\sum_{l \neq i} |\tilde{v}_l w_l|^2 + s_l^{-2} \sigma^2 \frac{\sigma^2}{P}}.
\]

The spectral efficiency function can be simplified in the following way, where \( \tilde{v}_l \in \mathbb{C}^T \) is the singular vector of the \( l \)th symbol, and \( s_l \in \mathbb{R} \) is the singular value of the \( l \)th symbol:

\[
SE^C(W, \tilde{V}, \tilde{S}, \sigma, P) = \sum_{l=1}^L \log_2 (1 + SINR_i^C(W, \tilde{v}_l, s_l, \sigma, P))
\]
\[
\sum_{l=1}^{L} \log_2 \left( \sum_{i=1}^{L} |\tilde{v}_l w_i|^2 + s_l^{-2} \sigma^2 \right)
\]
\[
- \sum_{l=1}^{L} \log_2 \left( \sum_{i \neq l}^{L} |\tilde{v}_l w_i|^2 + s_l^{-2} \sigma^2 \right) \rightarrow \max_W .
\]

One may notice that we have completely moved away from user antennas of shapes \(R_k\) and \(R\) and work only with user layers of shapes \(L_k\) and \(L\). The formula now does not depend on channel matrix \(H_k \in \mathbb{C}^{R_k \times T}\) but on the eigenvectors of the \(l\)th layer \(v_l \in \mathbb{C}^T\), which has length \(T\) by the number of antennas. The inversed squared singular values \(s_l^{-2} \in \mathbb{R}\) scale the noise power.

### 2.4. Power constraints

We formulate the realistic per-antenna power constraints [21]. Since we have \(T\) equal transmitter antennas, the power limitation applied to each of them is \(P_T\). The antenna power can be described in terms of the row norms of the precoding matrix: \(\|w_m\|^2 \leq \frac{P}{T} \forall m\).

It is clear that per-antenna constraints satisfy the total power: \(\|w_1\|^2 + \cdots + \|w_T\|^2 \leq P\), which is the sum-power constraint across all transmit antennas. While analytically attractive, such a sum-power constraint is often unrealistic in practice. In physical implementation of a multi-antenna base station, each antenna has its own power amplifier in its analog front-end and is limited individually by the linearity of the power amplifier. Thus, a power constraint imposed on a per-antenna basis is more realistic.

### 2.5. Problem statement

We formulate the constrained smooth optimization problem:

\[
\begin{align*}
\text{maximize} & \quad SE(W, H, G, \sigma, P) \\
\text{subject to} & \quad \|w_m\|^2 \leq \frac{P}{T}, \quad m = 1 \ldots T,
\end{align*}
\]

where \(w_m \in \mathbb{C}^L\) in the \(m\)th row of the precoding matrix.

### 3. Solutions for precoding

#### 3.1. Reference methods

The precoding matrix \(W\) is responsible for the beamforming from the base station to the users. The linear methods for precoding do the following. Firstly, the linear solutions obtain singular value decomposition for each user \(H_k = U_k^H S_k V_k \in \mathbb{C}^{R_k \times T}\) (Lemma 2.1) and take the first \(L_k\) singular vectors \(\tilde{V}_k \in \mathbb{C}^{L_k \times T}\) which attend to the first \(L_k\) greatest singular values [15]. All these matrices are concatenated to the one matrix \(\tilde{V} \in \mathbb{C}^{L \times T}\), which is used as the main building block of these precoding construction.

Finally, the precoding matrix is constructed from the obtained singular vectors. In the next sections, we describe the precoding matrix construction in the linear methods. We
meet the power constraints using the scalar post-adjustment. We divide the precoding matrix on its maximal row norm \( \text{max}_t \{ \| w^t \| \} \) and scale them on \( \sqrt{\frac{P}{T}} \). Thus, the per-antenna power constraints can satisfy \( \| w^m \|^2 \leq \frac{P}{T} \ \forall m \).

Firstly, **Maximum-Ratio Transmission** is the simplest way of computing the precoding matrix. It takes the Hermitian conjugate of the channel singular vectors [9]:

\[
W_{\text{MRT}} = \mu \tilde{V}^H P \in \mathbb{C}^{T \times L},
\]

where matrix \( P \in \mathbb{R}^{L \times L} \) denotes power allocation between symbols and constant \( \mu > 0 \) utilizes the power constraints. In such a way, the signal beams can be sent directly to users without considering their interaction. The maximum-ratio approach is preferred in noisy systems where the noise power is higher than inter-user interference.

Secondly, **Zero-Forcing** is the next modification of the precoding algorithm, which performs decorrelation of the symbols using inverse correlation matrix of the channel vectors [20]:

\[
W_{\text{ZF}} = \mu \tilde{V}^H (\tilde{V} \tilde{V}^H)^{-1} P \in \mathbb{C}^{T \times L}.
\]

With matrix \( P \in \mathbb{R}^{L \times L} \) and constant \( \mu > 0 \), which have the same meaning as in the previous section. Such precoding construction sends the signal beams to the users without creating any interference between them. Different from the previous method, the Zero-Forcing approach is preferred when the potential inter-user interference is higher than the noise power. There is a huge performance gain by eliminating the interference.

In the previous method, beams are sent not directly to the users but with some deviation, which actually reduces the payload. Thirdly, the following modification, **Regularized Zero-Forcing**, corrects the beams, which allows some inter-user interference and significantly increases the payload [17]:

\[
W_{\text{RZF}} = \mu \tilde{V}^H (\tilde{V} \tilde{V}^H + R)^{-1} P \in \mathbb{C}^{T \times L},
\]

where diagonal matrix \( R \in \mathbb{R}^{L \times L} \) denotes symbol regularization and matrix \( P \in \mathbb{R}^{L \times L} \) and constant \( \mu > 0 \) as the previous ones. As the baseline, we use a special form [12] of the regularization matrix \( R = \frac{\sigma^2}{L} I \). It is the most common precoding in real practice.

Algorithm \( W_{\text{RZF}} \) is a precoding with scalar regularization. Finally, **Adaptive Regularized Zero-Forcing** takes into account effective noise of the system. It was the proposed **Adaptive RZF (ARZF) algorithm** [2] with diagonal matrix regularization:

\[
W_{\text{ARZF}} = \mu \tilde{V}^H (\tilde{V} \tilde{V}^H + \lambda S^{-2})^{-1}, \quad \lambda = \frac{\sigma^2}{P}.
\]

### 3.2. Proposed optimization solution

Let us introduce the following parametrization. This method explicitly constrains antenna rows using (sub-)differentiable projection on a ball as a part of the Conjugate Spectral Efficiency function:

\[
\maximize_{W \in \mathbb{C}^{T \times L}} S(\text{proj}_{P,T}(W)),
\]

where...
where $S(W)$ is defined assuming $CD$ or assuming $MMSE-IRC$ detection. $S(W) = SE(W, H, G^{{IRC}}(W), \sigma, P)$ and

\[
\text{proj}_{P,T}(W) = \begin{cases} \frac{w^m}{\|w^m\|}, & \|w^m\|^2 \leq \frac{P}{T} \\ \sqrt{\frac{P}{T}} \frac{w^m}{\|w^m\|}, & \|w^m\|^2 > \frac{P}{T}, \forall m = 1 \ldots T, \end{cases}
\]

with $w^m \in \mathbb{C}^L$ in the $m$th row of the precoding matrix.

**Remark 3.1:** We also add maximization of the target $SE$ function for reference.

**Remark 3.2:** As the starting point, we are using regularized $W_{RZF}$ or adaptive $W_{ARZF}$.

The key feature of the proposed method is that the optimization process, on the one hand, always remains within the feasible range and at the same time can be solved using unconstrained optimization methods, such as quasi-Newton methods. In contrast to the projection gradient method, the iteration solution always remains in a feasible area and does not need an explicit projection on the boundary. Moreover, in the particular case the projection-based method does not converge appropriately. And the same for interior point method: its basic realization does not give an appropriate solution for our task. An advantage of the proposed method is that it seeks the solution on the boundary (and we know that the optimum is on the boundary). This search within the such complicated edge is possible using the Quasi-Newton L-BFGS method [22].

**Algorithm 1:** On the optimal precoding matrix and Quasi-Newton L-BFGS [22]

Input: Channel singular vectors $\tilde{V}$, channel singular values $S$, station power $P$, noise $\sigma^2$, iterations $N$, Tolerance grad $\varepsilon_g$, termination tolerance on first order optimality (default: 1e-5), Tolerance change $\varepsilon_c$, termination tolerance on function value and parameter changes (default: 1e-9).

Initialize precoding matrix $W \leftarrow W_0$;

for $t = 1$ to $N$ do

\hspace{1cm} if True conditions on $\varepsilon_g$ or $\varepsilon_c$ then

\hspace{2cm} return proj$_{P,T}(W)$

\hspace{1cm} end

\hspace{1cm} Calculate the gradient: $\nabla_W S(\text{proj}_{P,T}(W))$;

\hspace{1cm} Find the optimal direction recursively: $D = D(\nabla_W S(\text{proj}_{P,T}(W)))$;

\hspace{1cm} Find the optimal step length $\alpha = \arg \max_{\alpha} S(\text{proj}_{P,T}(W + \alpha D))$;

\hspace{1cm} Make the optimization step: $W \leftarrow W + \alpha D$;

end

return proj$_{P,T}(W)$

We consider $S(W)$ assuming $CD$ or assuming $MMSE-IRC$ detection, and also the initial precoding matrix in the form of RZF or ARZF. By this way, we compare four options:

1. QN CD RZF
2. QN CD ARZF
3. QN IRC RZF
4. QN IRC ARZF
3.3. Relative computational complexity

The complexity of our algorithm is no more than $\sim 3Nx$ times higher than the baseline RZF. On each iteration of QNS we have to compute gradient $\nabla SE^C(W)$ of $x$ complexity and adjust step length $\alpha$, which takes roughly two calls of the function $SE^C(W)$ of $2x$ total complexity (see Table 1).

3.4. Alternative optimization solution

For comparison, let us consider the classic softmax parametrization of the precoding matrix (this is an option of interior point approach). It turned out that such parametrization leads to the same solution as the differentiable projection method, but the latter converges at least twice as fast as softmax. We believe that this is due to the fact that most constraints should be up to the boundary, and softmax parameterization arguments in this sense should have the value of infinity. So, we marked the softmax method as an alternative method and brought it here for a better understanding of the solvable problem.

\[
\maximize_{W \in \mathbb{C}^{T \times L}} SE^C(W, \tilde{V}, S, \sigma, P) \tag{22}
\]
\[
W = (w_{ij})_{i,j=1}^{T,L}, \quad w_{ij} = \rho_{ij} \exp(i\phi_{ij}), \quad \rho_{ij} \in \mathbb{R}^+, \quad \phi_{ij} \in [0, 2\pi) \tag{23}
\]
\[
\rho_{ij}^2 = \frac{\exp(\theta_{ij}) \sigma(\alpha_i)P}{\sum_{k=1}^{L} \exp(\theta_{ik})}, \quad \phi_{ij} = 2\pi \sigma(\eta_{ij}), \tag{24}
\]

where $\sigma(\cdot)$ means a sigmoid function and $\{\theta_{ij}, \eta_{ij}, \alpha_i\}$ are free parameters that can take any real values. Thus, in another way we can reduce the problem of optimizing precoding with constraints to the problem without constraints.

4. Numerical experiments

The main scenario is Urban Non-Line-of-Sight 3GPP_38.901_RMa_NLOS [3]. We are testing the proposed approach in several scenarios generated using Quadriga [7] – open-source software for generating realistic radio channel impulse response. For each scenario we generate 40 different channel matrices $H \in \mathbb{C}^{n_{users} \times 4 \times 64}$. The carrier frequency for each channel matrix is selected randomly over the bandwidth. User selection is described in the next section. The base station antenna array forms a grid with 8 placeholders along the $y$ axis and 4 placeholders along the $y$ axis. The receiver antenna array consists of two placeholders along the $x$ axis. Each placeholder contains two cross-polarized antennas. An interested reader can find detailed hyperparameters for antenna models and generation processes in Table 2. All unlisted Quadriga parameters are those set by default. This set-up is similar to the one used in the [2] work.
Table 2. Quadriga generation parameters.

| Parameter                                      | Value               |
|------------------------------------------------|---------------------|
| **Base station parameters**                   |                     |
| Number of base stations                       | 1                   |
| Position, m: (x, y, z) axes                   | (0, 0, 25)          |
| Number of antenna placeholders (y axis)       | 8                   |
| Number of antenna placeholders (z axis)       | 4                   |
| Distance between placeholders (y axis)         | 0.5 wavelength      |
| Distance between placeholders (z axis)         | 1.7 wavelength      |
| Antenna model                                 | 3gpp-macro          |
| Half-Power in azimuth direction, deg          | 60                  |
| Half-Power in elevation direction, deg        | 10                  |
| Front-to-back ratio, dB                       | 20                  |
| Total number of antennas                      | 64                  |
| **Receiver parameters**                       |                     |
| Number of placeholders at the receiver (x axis)| 2                   |
| Distance between placeholders (x axis)         | 0.5 wavelength      |
| Antenna model                                 | Half-wave-dipole    |
| **Quadriga simulation parameters**            |                     |
| Central band frequency                        | 3.5 GHz             |
| 1 sample per meter (default value)            | 1                   |
| Include delay of the LOS path                 | 1                   |
| Disable spherical waves (use_3GPP_baseline)   | 1                   |
| **Quadriga channel builders parameters**      |                     |
| Shadow fading sigma                           | 0                   |
| Cluster splitting                             | False               |
| Bandwidth                                     | 100 MHz             |
| Number of subcarriers                         | 42                  |

For each scenario we generate 40 different channels: $H \in \mathbb{C}^{K \times R_k \times T}$:

1. $T = 64$ base station antennas,
2. $K = 8$ users,
3. $R_k = 4$ user antennas,
4. $L_k = 2$ user layers.

4.1. Selecting user positions

We model users in the urban landscape (Figure 2). We locate users on the ground, setting the height to 1.5 m, and generate x, y positions uniformly in the 120° ray within 250 m from the base station. For the NLOS scenario, we locate users in the urban landscape. Firstly, we sample up to 8 cluster centers $x_c, y_c$ in the 120° ray from the base station within 2000 m from the base station. Each cluster represents a part of a city building. We assign a random cluster height $z_c = 1.5 m + (3 \cdot U([1, \ldots, 10]) - 1)$, selecting the cluster floor in a building from the uniform distribution $U$. Secondly, for each user, we assign a cluster id $c(u)$ and sample $x_u, y_u$ position randomly over the 60 m circle around the cluster centre. Thirdly, we sample the height of the user, given the height of the cluster, 80% of users we place at the floors near the cluster floor $z_u = z_{c(u)} + 3 \cdot U([-1, 0, 1])$ m, and 20% of users we place outdoor $z_u = 1.5$ m.

After generating channel matrices for a fairly large number of users (64), we select a subset of users, which are not too correlated, since too correlated users can be suited at
Figure 2. The graph shows IRC spectral efficiency of the different precoding algorithms.

The correlation between users $i, j$ is measured as squared cosine between the main singular vector: $\cos^2(v_i, v_j) = |\langle \tilde{v}_i, \tilde{v}_j \rangle|^2$. The statistics of mean correlations between all pairs of users served together (averaged over random seeds) is equal to 7.8%. We believe, our user generation produces realistic setups for both urban and rural cases and is simple to implement.

4.2. Qualitative results

For each generated channel seed, we fix station power $P = 1$ and select system noise $\sigma^2$ for the specified $SUSINR$ (7) average value. We use binary search to select the noise $\sigma^2$ since $SUSINR$ depends on it monotonically. We report the hyper-parameters for Quadriga in the Table 2.

The results of the comparison are in Figure 3 and Table 3. In all cases the final $SE$ function is calculated using MMSE-IRC detection (9), geometric mean effective SINR (5) and Shannon’s capacity (6).

To begin with, the experimental results in Figure 3 and Table 3 show that all gradient-based methods starting with ‘QN,’ which stands for ‘Quasi-Newton’ (Algorithm (1)), outperform heuristic algorithms such as MRT (16), ZF (17), RZF (18), and ARZF (19).

The second keyword ‘CD’ and ‘IRC’ stands for the assumed detection algorithm: Conjugate Detection (14) and MMSE-IRC (6) functions, respectively, and their gradients. It is important to observe that the performance of the obtained method is influenced by the target optimization function. Remember that the MMSE-IRC function was used to calculate the final quality measure. As a result, the ‘IRC’ method, which optimizes the target function directly, has a higher final quality measure.
Figure 3. Example of users for urban two building set-up [2].

Table 3. Urban NLOS 8 Users. The table shows IRC spectral efficiency of the different precoding algorithms.

| Precoding SUsINR | MRT | ZF | RZF | ARZF | QN CD RZF | QN CD ARZF | QN IRC RZF | QN IRC ARZF |
|------------------|-----|----|-----|------|-----------|------------|------------|------------|
| −4               | 0.60| 0.26| 0.59| 1.58 | 1.92      | 1.92       | 2.13       | 2.10       |
| 0                | 0.79| 0.52| 0.85| 2.25 | 2.70      | 2.69       | 2.95       | 2.95       |
| 4                | 0.98| 0.95| 1.24| 3.02 | 3.59      | 3.64       | 3.91       | 3.94       |
| 8                | 1.16| 1.56| 1.78| 3.89 | 4.59      | 4.70       | 4.99       | 5.00       |
| 12               | 1.33| 2.38| 2.52| 4.86 | 5.73      | 5.85       | 6.17       | 6.20       |
| 16               | 1.51| 3.39| 3.46| 5.90 | 6.92      | 7.11       | 7.42       | 7.47       |
| 20               | 1.70| 4.57| 4.61| 7.04 | 8.17      | 8.42       | 8.71       | 8.81       |
| 24               | 1.91| 5.93| 5.95| 8.29 | 9.43      | 9.83       | 10.01      | 10.25      |
| 28               | 2.11| 7.41| 7.42| 9.59 | 10.67     | 11.28      | 11.25      | 11.68      |
| 32               | 2.29| 9.15| 9.15| 11.06| 11.99     | 12.80      | 12.57      | 13.18      |
| 36               | 2.46| 11.06| 11.06| 12.63| 13.53     | 14.35      | 13.97      | 14.64      |
| 40               | 2.56| 13.14| 13.14| 14.31| 15.28     | 15.90      | 15.45      | 16.08      |

On the other hand, the ‘CD’ method, on the other hand, converges faster through iterations and each iteration is computationally simpler than IRC. On Figure 4 one can find the quality of the proposed gradient methods by iterations. We noticed that the modification of algorithm that uses the approximated ‘CD’ optimization function gives better results than the ‘IRC’ target function in the first iterations. At higher iterations the ‘IRC’ gradient method is superior to the ‘CD’ method.

The third keyword in the title of gradient methods, ‘RZF’ (18) and ‘ARZF’ (19), refers to the method’s starting point. The method will be more convergent if the starting point
is better. Methods based on the ‘ARZF’ starting point outperform methods based on the ‘RZF’ point in this way. Remember that our issue is not concave. As a result, we only find the local maximum, and a good starting point is critical for our method to work.

5. Conclusion

The paper studies the multi-user precoding problem as a non-convex optimization problem for wireless MIMO systems. In this study we approximate the spectral efficiency objective function using a differentiable projection-based method, where gradients are taken both for the function and the projection, which tends to very fast convergence of the proposed method. We then reduce the precoding problem to an unconstrained optimization task and solve it by the quasi-Newton L-BFGS iterative procedure. Finally, all our research is based on realistic power limits per antenna and multiple antenna users, which in itself is a standalone problem for finding the right approach to solving. All investigated algorithms were studied in massive experiments using Quadriga software. The proposed method shows monotonic improvement over heuristic methods with reasonable computation time. It has a simple implementation and can be a good reference for other heuristic algorithms in the MIMO field.

Acknowledgments

Authors are grateful for D. Minenkov for the fruitful discussions.

Disclosure statement

No potential conflict of interest was reported by the authors.
Funding

This work was supported by Huawei Technologies.

Notes on contributors

Evgeny Bobrov is a PhD student of the Department of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, Russian Federation. Graduated from Lomonosov Moscow State University with Bachelor's and Master’s degrees in the same university. The topic of my scientific qualification work is: “Machine Learning and Optimization Algorithms in Wireless Communications Problems”. Experience of teaching 4 years. Area of interest includes machine learning, optimization and modern wireless communication problems.

Dmitry Kropotov is a specialist: Lomonosov Moscow State University. M.V. Lomonosov Moscow State University, specialty “Applied mathematics and informatics”, qualification “mathematician, system programmer”. Professional interests: optimization, Bayesian methods, neural network models, applied data analysis. Work experience: February 2007 - December 2013, Junior Researcher, A.A. Dorodnitsyn Computer Center RAS. February 2014 - present, researcher, Lomonosov Moscow State University, Department of Computational Mathematics and Cybernetics, Department of Mathematical Methods of Forecasting. June 2014-present, Scientific Secretary, Lomonosov Moscow State University, Department of Computational Mathematics and Cybernetics, Department of Mathematical Methods of Forecasting. February 2018 - present, Senior Researcher, Higher School of Economics-Samsung Joint Laboratory.

Sergey Troshin is a Research Intern: Faculty of Computer Science / Department of Big Data and Information Retrieval / Center for Deep Learning and Bayesian Methods. Started working at NRU HSE in 2017. Research and teaching experience: 1 year, 6 months. Responsibilities: to conduct research for Centre of Deep Learning and Bayesian Methods. Professional Interests: Implicit models, automatic source code analysis. Employment history: Research Assistant at Advanced Research NLP Group ABBYY, June-August 2019. Development of a system for active learning of neural networks to search for relationships between entities in a text.

Danila Zaev received Ph.D degree in mathematics in 2017 at National Research University Higher School of Economics (SU HSE), Department of Mathematics. Title of the PhD thesis: Monge-Kantorovich problem with linear constraints. His professional Interests: wireless communication, massive MIMO, stochastic processes, mathematical optimization.

ORCID

Evgeny Bobrov http://orcid.org/0000-0002-2584-6649

References

[1] E. Björnson, J. Hoydis, and L. Sanguinetti, Massive mimo networks: Spectral, energy, and hardware efficiency, Found. Trends Signal Process. 11 (2017), pp. 154–655.
[2] E. Bobrov, B. Chinyaev, V. Kuznetsov, H. Lu, D. Minenkov, S. Troshin, D. Yudakov, and D. Zaev, Adaptive regularized zero-forcing beamforming in massive mimo with multi-antenna users, arXiv preprint arXiv:2107.00853 (2021).
[3] F. Bohagen, P. Orten, and G. Oien, Construction and capacity analysis of high-rank line-of-sight MIMO channels. IEEE Wireless Communications and Networking Conference, Vol. 1, 1995, pp. 432–437.
[4] J.C. Chen, Gradient projection-based alternating minimization algorithm for designing hybrid beamforming in millimeter-wave MIMO systems, IEEE Commun. Lett. 23 (2018), pp. 112–115.
[5] H.H. Dam and A. Cantoni, Interior point method for optimum zero-forcing beamforming with per-antenna power constraints and optimal step size, Signal Process. 106 (2015), pp. 10–14.
[6] L. Garcia-Ordonez, D.P. Palomar, A. Pages-Zamora, and J.R. Fonollosa, *Analytical BER performance in spatial multiplexing MIMO systems*. IEEE 6th Workshop on Signal Processing Advances in Wireless Communications, 2005, pp. 460–464.

[7] S. Jaeckel, L. Raschkowski, K. Börner, and L. Thiele, *Quadriga: A 3-D multi-cell channel model with time evolution for enabling virtual field trials*, IEEE Trans. Antennas Propag. 62 (2014), pp. 3242–3256.

[8] M. Joham, W. Utschick, and J.A. Nossek, *Linear transmit processing in MIMO communications systems*, IEEE Trans. Signal Process. 53 (2005), pp. 2700–2712.

[9] T.K. Lo, *Maximum ratio transmission*. 1999 IEEE International Conference on Communications (Cat. No. 99CH36311), Vol. 2, 1999, pp. 1310–1314.

[10] A.H. Mehana and A. Nosratinia, *Diversity of MMSE MIMO receivers*, IEEE Trans. Inf. Theory 58 (2012), pp. 6788–6805.

[11] L.D. Nguyen, H.D. Tuan, T.Q. Duong, and H.V. Poor, *Multi-user regularized zero-forcing beamforming*, IEEE Trans. Signal Process. 67 (2019), pp. 2839–2853.

[12] C.B. Peel, B.M. Hochwald, and A.L. Swindlehurst, *A vector-perturbation technique for near-capacity multi-antenna multiuser communication-part I: Channel inversion and regularization*, IEEE Trans. Commun. 53 (2005), pp. 195–202.

[13] B. Ren, Y. Wang, S. Sun, Y. Zhang, X. Dai, and K. Niu, *Low-complexity MMSE-IRC algorithm for uplink massive MIMO systems*, Electron. Lett. 53 (2017), pp. 972–974.

[14] S. Shi, M. Schubert, and H. Boche, *Downlink MMSE transceiver optimization for multi-user MIMO systems: Duality and sum-MSE minimization*, IEEE Trans. Signal Process. 55 (2007), pp. 5436–5446.

[15] L. Sun and M.R. McKay, *Eigen-based transceivers for the MIMO broadcast channel with semi-orthogonal user selection*, IEEE Trans. Signal Process. 58 (2010), pp. 5246–5261.

[16] S. Verdú, *Spectral efficiency in the wideband regime*, IEEE Trans. Inf. Theory 48 (2002), pp. 1319–1343.

[17] Z. Wang and W. Chen, *Regularized zero-forcing for multi-antenna broadcast channels with user selection*, IEEE Wirel. Commun. Lett. 1 (2012), pp. 129–132.

[18] A. Wiesel, Y.C. Eldar, and S. Shamai, *Zero-forcing precoding and generalized inverses*, IEEE Trans. Signal Process. 56 (2008), pp. 4409–4418.

[19] D. Wubben, R. Bohnke, V. Kuhn, and K.D. Kammeyer, *Near-maximum-likelihood detection of MIMO systems using MMSE-based lattice-reduction*. 2004 IEEE International Conference on Communications (IEEE Cat. No. 04CH37577), Vol. 2, 2004, pp. 798–802.

[20] T. Yoo and A. Goldsmith, *On the optimality of multi-antenna broadcast scheduling using zero-forcing beamforming*, IEEE J. Sel. Areas Commun. 24 (2006), pp. 528–541.

[21] W. Yu and T. Lan, *Transmitter optimization for the multi-antenna downlink with per-antenna power constraints*, IEEE Trans. Signal Process. 55 (2007), pp. 2646–2660.

[22] C. Zhu, R.H. Byrd, P. Lu, and J. Nocedal, *Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization*, ACM Trans. Math. Softw. 23 (1997), pp. 550–560.