A LOW-COMPLEXITY ZERO-FORCING BEAMFORMER DESIGN FOR MULTIUSER MIMO SYSTEMS VIA A DUAL GRADIENT METHOD

BIN LI AND HAI HUYEN DAM
Department of Mathematics and Statistics
Curtin University, GPO Box U1987
Perth, WA 6845, Australia

ANTONIO CANTONI
School of Electrical, Electronic and Computer Engineering
The University of Western Australia
35 Stirling Highway, Crawley, WA 6009, Australia

ABSTRACT. In this paper, we consider the zero-forcing beamforming (ZFBF) under the per-antenna power constraints (PAPC). Our objective is to maximize the minimum user information rate. Traditionally, ZFBF under PAPC with a max-min performance measure can be transformed into a second order cone problem and then solved by applying the interior point method. However, it is expensive to realize this design in practice due to high computational complexity per iteration. An alternative low complexity zero-forcing beamformer design is proposed for MU-MIMO systems by applying a dual gradient method. Different from the step size rule in the literature, a backtracking line search is adopted. A numerical example is provided to show the effectiveness of the proposed method.

1. Introduction. Transmitter design for the multi-user multiple-input multiple-output (MU-MIMO) systems has been studied intensively in the literature (see, for example, [3]-[14]). The dirty paper coding (DPC) [3] is known as the capacity-achieving scheme. Due to high computational complexity, however, it is difficult to be implemented in practice. Consequently, emphasis has been shifted to finding suboptimal strategies (see, for example, [14, 15]). Zero-forcing beamforming (ZFBF) is one of the most commonly used linear pre-coding methods in MU-MIMO broadcast channel since it provides a good trade-off between the complexity and the performance. Traditionally, the beamformer design is investigated under the total power constraint [15]. However, for real world applications, each antenna of the transmitter has its own amplifier and its linear operation is important for modern efficient modulation, e.g. Orthogonal Frequency Division Multiplexing (OFDM). Hence, the per-antenna power constraints (PAPC) (see, for example, [15, 5, 6]) are more relevant for real world applications than the total power constraint.

2010 Mathematics Subject Classification. Primary: 90B18; Secondary: 90C25.

Key words and phrases. Zero-forcing beamforming, MIMO systems, a gradient method, a backtracking line search.

This work was supported by a Discovery Grant from the Australia Research Council (No. DP120103859).
ZFBF under the PAPC is a nontrivial problem. In [15], it is shown that ZF pre-coding is closely related to generalized inverses and it is a difficult optimization problem, which depends heavily on the performance criterion, to find the optimal generalized inverse under PAPC. Two specific performance criteria (i.e., the fairness and throughput) have been investigated. For the fairness performance, the problem is transformed into a convex second order cone program. For the throughput performance, the problem is transformed into a standard determinant maximization program subject to matrix linear inequalities, which is a convex optimization problem. Both of the transformed problems can be solved by using standard optimization packages. In [6], the pseudo-inverse is scaled to satisfy the PAPC with a scale factor. However, as we mentioned, this approach does not lead to optimal solution.

In real world applications, it is expensive for realizing the algorithms in [15] in hardware [4, 8, 9, 10]. This is due to the fact that many complex computations are involved in the algorithm, such as the calculations of a Hessian matrix and taking the inverse of a matrix in each iteration [2]. More specifically, it has a complexity of $O\left((M + N)^3 M^3\right)$ in each iteration [1].

In this paper, we consider a zero-forcing beamformer design problem under PAPC, and the minimum user information rate is taken as the performance measure. A dual gradient method method is proposed. By introducing the Lagrangian, the dual problem of the original problem is formed. Since the strong duality holds, the problem can be solved through solving its dual problem [2]. However, since the objective is non-differentiable, a Tikhonov regularization is applied to smooth the dual objective. The regularized dual problem can be solved by using the gradient method. The computational complexity is only $O(M^3N + N^2)$ in each iteration.

In the literature, a constant number is chosen as the step size, and the Lipschitz constant is always chosen [11]. However, in this case the convergence is slow and sometimes it does not even converge. To overcome this difficulty, we use a backtracking line search instead of a constant step size in each iteration.

2. Problem Formulation. Consider the standard MISO multiuser broadcast channel

\[ y_m = h_m^H x + n_m, \quad m = 1, 2, \ldots, M, \]  

where $y_m$ is the received signal of the $m^{th}$ user, $h_m$ is the channel vector of length $N$ of the $m^{th}$ user, $x$ is the transmitted vector of length $N$, and $n_m$ is the complex Gaussian noise with mean 0 and variance $\sigma^2$. It can be written in a compact form given below:

\[ y = Hx + n, \]

where $y = [y_1, y_2, \ldots, y_M]^\top$, $H = [h_1, h_2, \ldots, h_M]^H$, $n = [n_1, n_2, \ldots, n_M]^\top$, $(\cdot)^\top$ denotes the transpose, and $(\cdot)^H$ denotes the conjugate transpose.

Here, the linear zero-forcing pre-coding transmitter is applied, i.e.,

\[ x = Ws \]

\[ HW = \sqrt{\Lambda} \]

where $s$ is the information vector of length $M$ such that $E\{ss^H\} = I$, $I$ denotes the identity matrix of appropriate dimension, $W$ is an $M \times N$ complex matrix, and $\Lambda$ denotes a real and positive diagonal matrix.
The information rate for each user is denoted by \( r(m) \) and is given by
\[
r(m) = \log_2 (1 + \text{SINR}(m)), \quad j = 1, 2, \ldots, M, \tag{5}\]
where \( \text{SINR}(m) \) is the signal-to-interference-plus-noise ratio (SINR) for each user, which is given by
\[
\text{SINR}(m) = \frac{\left| (HW)_{m,m} \right|^2}{\sum_{j,j \neq m} \left| (HW)_{j,m} \right|^2 + \sigma_m^2}, \tag{6}\]
m, j = 1, 2, \ldots, M. From (4), we have \( \sum_{j,j \neq m} \left| (HW)_{j,m} \right|^2 = 0 \). In this paper, we take minimum user information rate, i.e., \( \min_{1 \leq m \leq M} r(m) \), as the performance measure.

To limit the power on the amplifier of each antenna, the per-antenna power constraints are imposed as follows:
\[
\sum_{m=1}^{M} |e_n^T w_m|^2 \leq P, \quad n = 1, 2, \ldots, N, \tag{7}\]
where \( w_m \) is the \( m \)th column vector of \( W \), \( e_n \) is a vector of length \( N \) with an 1 in the \( n \)th element while 0 in the other elements, and \( P \) is the maximum allowable power on each antenna.

The problem under consideration may now be formally stated below:

\begin{equation}
\textbf{Problem.} \quad \max_{w_m, r_0} \quad r_0 \quad \quad m = 1, \ldots, M \tag{8a}
\end{equation}

\begin{equation}
\text{s. t.} \quad \log_2 \left( 1 + \frac{|h_m^H w_m|^2}{\sigma^2} \right) \geq r_0, \quad m = 1, \ldots, M \tag{8b}
\end{equation}

\begin{equation}
\sum_{m=1}^{M} |e_n^T w_m|^2 \leq P, \quad n = 1, \ldots, N \tag{8c}
\end{equation}

\begin{equation}
h_j^H w_m = 0, \quad \forall j \neq m, \quad 1 \leq j, m \leq M \tag{8d}
\end{equation}

From [10], it follows from letting
\[
\mathbf{x} = [\mathbf{w}_{1\text{Re}}^\top \mathbf{w}_{1\text{Im}}^\top \ldots \mathbf{w}_{M\text{Re}}^\top \mathbf{w}_{M\text{Im}}^\top]^\top
\]
where \( \mathbf{x} \in \mathbb{R}^{2NM} \), that Problem 2 can be written as the following optimization problem.

\begin{equation}
\textbf{Problem.} \quad \min_{\mathbf{x}, t} \quad -t \tag{9}
\end{equation}

\begin{equation}
\text{s. t.} \quad \mathbf{H}_1 \mathbf{x} \leq -t \mathbf{1} \tag{10}
\end{equation}

\begin{equation}
\mathbf{x}^\top \mathbf{A}_n \mathbf{x} \leq P \quad n = 1, 2, \ldots, N \tag{11}
\end{equation}

\begin{equation}
\mathbf{H}_2 \mathbf{x} = 0 \tag{12}
\end{equation}

where \( \mathbf{1} \) is a vector of ones with appropriate dimension, \( \mathbf{A}_n = \text{diag} \{ \mathbf{B}_n \mathbf{B}_n \ldots \mathbf{B}_n \} \) \( \in \mathbb{R}^{2N^2 \times 2NM} \), \( \mathbf{B}_n \in \mathbb{R}^{2N^2 \times 2N} \) is a diagonal matrix with 1 appearing in the \( (n, n) \)th and \( (n + N, n + N) \)th positions and 0 elsewhere, \( \mathbf{H}_1 \in \mathbb{R}^{M \times 2NM} \), and \( \mathbf{H}_2 \in \mathbb{R}^{N \times 2NM} \).
$\mathbb{R}^{2M(M-1)+M\times 2NM}$. For the structures of $H_1$ and $H_2$, we refer the readers to [10].

3. A Dual Gradient Method. In this section, we develop a dual gradient method to solve Problem 2. To achieve this goal, we introduce the following Lagrangian of Problem 2 as

$$L(t, x, \lambda, v, \mu) = -t + \lambda^\top (H_1x + t1) + v^\top H_2x + \sum_{n=1}^{N} \mu_n (x^\top A_nx - P).$$

The corresponding dual function is denoted as

$$d(\lambda, v, \mu) = \min_{x,t} L(t, x, \lambda, v, \mu)$$

$$= \min_{x,t} \left\{ -t + \lambda^\top (H_1x + t1) + v^\top H_2x + \sum_{n=1}^{N} \mu_n (x^\top A_nx - P) \right\}. \quad (13)$$

Then, the dual problem of Problem 2 can be written as

Problem.

$$\max_{\lambda, v, \mu} \quad d(\lambda, v, \mu) \quad (14a)$$

$$s. \ t. \quad \lambda \succeq 0 \quad (14b)$$

$$\mu \succeq 0 \quad (14c)$$

where $a \succeq 0$ means each element of $a$ is greater than or equal to 0.

Problem 2 can be solved through solving its dual problem, i.e., Problem 3, since the strong duality holds. This can be verified by the fact that Problem 2 is convex and the following Slater’s condition holds [2].

3.1. Regularization of the Lagrangian. Note that the Lagrangian (13) is not strictly convex and the the dual function (13) is not differentiable. Thus, a Tikhonov regularization to Lagrangian (13) is applied to smooth the regularized dual function (13).

To regularize (13), we introduce two prox-functions $d_t(t) = \rho t^2$ and $d_x(x) = \rho \|x\|^2$, where $\rho$ is a smoothing parameter and $\| \cdot \|$ denotes Euclidian norm. By appending $d_t(t)$ and $d_x(x)$ into (13), we obtain the regularized Lagrangian as

$$L_\rho(t, x, \lambda, v, \mu) = L(t, x, \lambda, v, \mu) + \rho t^2 + \rho \|x\|^2 \quad (15)$$

Note that the optimal solution of

$$\min_{x,t} L_\rho(t, x, \lambda, v, \mu)$$

can be written in closed form as

$$t^* = \frac{1}{2\rho} \left( 1 - \sum_{m=1}^{M} \lambda_m \right), \quad x^* = -\frac{1}{2} V_\rho(\mu)s$$

where $V_\rho(\mu) = I_{2M} \otimes C \in \mathbb{R}^{2NM}$, $C = \text{diag} \left[ \frac{1}{\rho+\mu_1}, \frac{1}{\rho+\mu_2}, \ldots, \frac{1}{\rho+\mu_N} \right] \in \mathbb{R}^{N\times N}$, $s = H_1^\top \lambda + H_2^\top v$. 


By considering (13), (15) and (16), we obtain the regularized dual function as follows

\[ d_{\rho}(\lambda, v, \mu) = -\frac{1}{4} s^T V_{\rho}(\mu)s - \frac{1}{4\rho} \left( \sum_{m=1}^{M} \lambda_m - 1 \right)^2 - P \sum_{n=1}^{N} \mu_n \]

The regularized dual problem with (14a) replaced by (17) is referred to as Problem (3(\rho)) as follows.

\[
\begin{align*}
\max_{\lambda, v, \mu} & \quad d_{\rho}(\lambda, v, \mu) \\
\text{s. t.} & \quad \lambda \succeq 0 \\
& \quad \mu \succeq 0.
\end{align*}
\]

3.2. A Gradient Algorithm. In this section, we present a gradient method to solve Problem 3(\rho), which is given as below.

For the kth iteration, we update the dual variables \( \lambda, v \) and \( \mu \) as follows:

\[
\begin{align*}
\lambda^{k+1} &= \max \left\{ \lambda^k + t^k \frac{\partial d_{\rho}}{\partial \lambda}(\lambda^k, v^k, \mu^k), 0 \right\} \quad (18) \\
v^{k+1} &= v^k + t^k \frac{\partial d_{\rho}}{\partial v}(\lambda^k, v^k, \mu^k) \quad (19) \\
\mu^{k+1} &= \max \left\{ \mu^k + t^k \frac{\partial d_{\rho}}{\partial \mu}(\lambda^k, v^k, \mu^k), 0 \right\} \quad (20)
\end{align*}
\]

where

\[
\begin{align*}
\frac{\partial d_{\rho}}{\partial \lambda}(\lambda, v, \mu) &= \frac{1}{2\rho^2} \left( 1 - \sum_{m=1}^{M} \lambda_m \right) - \frac{1}{2} \mathbf{H}_1 V_{\rho}(\mu)s \\
\frac{\partial d_{\rho}}{\partial v}(\lambda, v, \mu) &= -\frac{1}{2} \mathbf{H}_2 V_{\rho}(\mu)s \\
\frac{\partial d_{\rho}}{\partial \mu}(\lambda, v, \mu) &= \frac{1}{4} (\mu_n + \rho)^2 s^T A_n s
\end{align*}
\]

In [11], the constant step size rule is adopted, where the estimated Lipschitz constant \( t \) is chosen for the step size in each iteration. Here, we apply a backtracking line search to determine the step size \( t^k \) in the kth iteration. Thus, the step size is varying in each iteration to guarantee a certain reduction of the objective value, which is described as blow.

**Line Search Rule**

Set \( t_{\text{max}}, t_{\text{min}}, l \) and \( \omega \), and denote

\[ d^k = \left[ \frac{\partial d_{\rho}}{\partial \lambda}(\lambda^k, v^k, \mu^k), \frac{\partial d_{\rho}}{\partial v}(\lambda^k, v^k, \mu^k), \frac{\partial d_{\rho}}{\partial \mu}(\lambda^k, v^k, \mu^k) \right] . \]

For the kth iteration, set \( t_k = t_{\text{max}} \) and goto **Step 1**.

**Step 1** Generate \( \lambda, \tilde{v}, \tilde{\mu} \) with (21)-(23).

**Step 2** Compute \( d_{\rho}(\lambda, \tilde{v}, \tilde{\mu}) \). If \( d_{\rho}(\lambda, \tilde{v}, \tilde{\mu}) \geq d_{\rho}(\lambda, v, \mu) + \omega t_k \| d^k \| \), stop and set \( \lambda^{k+1} = \lambda, v^{k+1} = \tilde{v}, \mu^{k+1} = \tilde{\mu} \). Else, set \( t_k = lt_k \) and goto **Step 3**.

**Step 3** If \( t_k \leq t_{\text{min}} \), stop and report error. Else, goto **Step 1**.

Now we state the proposed algorithm as follows.

**Algorithm 1**

Given an initial point \( \lambda^0, v^0 \) and \( \mu^0 \) and \( \epsilon \).
For the $k$th iteration:

**Step 1** Compute the search direction according to (21)-(23).

**Step 2** Apply the Line Search Rule and generate $\lambda^{k+1}, v^{k+1}$ and $\mu^{k+1}$ and the dual function value $d^{k+1}$.

**Step 3** Compute the primal function value $f^{k+1} = -t^{k+1} + \rho(t^{k+1})^2 + \rho\|x^{k+1}\|^2$ according to (16).

**Step 4** If $|d^{k+1} - f^{k+1}| \leq \epsilon$, exit. Else goto Step 1.

**Theorem 3.1.** Let $y^k = \{\lambda^k, v^k, \mu^k\}$ be the vector generated by Algorithm 1 in the $k$th iteration. Then

$$d_{\rho} (y^k) - d^*_{\rho} \leq \frac{1}{2kt_{\min}} ||y^0 - y^*||^2 = O \left( \frac{1}{k} \right)$$

where

$$d^*_{\rho} = \max_y d_{\rho} (y) = d_{\rho} (y^*) ,$$

**Proof.** The proof follows from the result on page 25 of Chapter 1 in [13].

4. **Numerical Results.** In this section, we provide a numerical example to justify the results in this paper. The base station array considered in the numerical examples is a uniform planar circular array. It consists of $N$ isotropic elements and the inter-element spacing is equal to $0.5\lambda$. $M$ users are spread uniformly around the base station. We have set $\sigma^2 = 0.001$ and denote $P = 0.01$. The computer used is a Dell desktop. Its CPU is i5-2500, 3.30G CPU, and it has a 8G RAM. The simulation is implemented in the Matlab environment. We implement Algorithm 1 within a channel that was tested by The Commonwealth Scientific and Industrial Research Organization (CSIRO) in rural Australia [12].

We compare Algorithm 1 with the gradient method by using a constant line search, where the estimated Lipschitz constant is adopted as the step size. In Algorithm 1, $\epsilon$ is set as $10^{-4}$, $\lambda^0 = 1$, $v^0 = 1$ and $\mu^0 = 1$. The residual errors of the two methods are plotted in Figure 1 by considering different scenarios. From
Figure 1, it shows that for each scenario the convergence by using the backtracking line search is faster than the one by using the constant step size rule. As expected, for each line search rule the scenario with less antennas \((N)\) and less users \((M)\) converges faster since there are less variables to optimize in the problem.

5. Conclusion. A dual gradient method is adopted in this paper. By applying this method, \(O\left(\frac{1}{k}\right)\) convergence rate is achieved with the computational complexity of \(O\left(M^4N + N^2\right)\) in each iteration, which is less than \(O\left((M + N)^3M^3\right)\) by using the interior point method. The numerical example shows that by using the backtracking line search the convergence of the gradient method is faster than that of using the constant step size rule. For future study, the sum rate performance may be considered.

Acknowledgments. This work was supported by a Discovery Grant from the Australia Research Council (No. DP120103859).

REFERENCES

[1] S. J. Benson, Y. Ye and X. Zhang, Solving large-scale sparse semidefinite programs for combinatorial optimization, SIAM J. Optim., 10 (2000), 443–461.

[2] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge, UK: Cambridge University Press, 2004.

[3] G. Caire and S. Shamai (Shitz), On the achievable throughput of multiantenna Gaussian broadcast channel, IEEE Trans. Inf. Theory., 49 (2003), 1691–1706.

[4] 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), 10–14.

[5] K. Karakayali, R. Yates, G. Foschini and R. Valenzuela, Optimal zero-forcing beamforming with per-antenna power constraints, IEEE International Symposium on Information Theory, Nice, France, (2007), 101–105.

[6] S. R. Lee, J. S. Kim, S. H. Moon, H. B. Kong and I. Lee, Zero-forcing beamforming in multiuser MISO downlink systems under per-antenna power constraint and equal-rate metric, IEEE Trans. Wireless Commun., 12 (2013), 228–236.

[7] B. Li, H. H. Dam, A. Cantoni and K. L. Teo, A global optimal zero-forcing beamformer design with signed Power-of-Two coefficients, Journal of Industrial and Management Optimization, 12 (2016), 625–636.

[8] B. Li, H. H. Dam, A. Cantoni and K. L. Teo, A first-order optimal zero-forcing beamformer design for multiuser MIMO systems via a regularized dual accelerated gradient method, IEEE Commun. Lett., 19 (2015), 195–198.

[9] B. Li, H. H. Dam, A. Cantoni and K. L. Teo, Some interesting properties for zero-forcing beamforming under per-antenna power constraints in rural areas, J. Glob. Optim., 62 (2015), 877–886.

[10] B. Li, H. H. Dam, A. Cantoni and K. L. Teo, A primal-dual interior point method for optimal zero-forcing beamformer design under per-antenna power constraints, Optim. Lett., 8 (2014), 1829–1843.

[11] B. Li, C. Z. Wu, H. H. Dam, A. Cantoni and K. L. Teo, A parallel low complexity zero-forcing beamformer design for multiuser MIMO systems via a regularized dual decomposition method, IEEE Trans. Signal Process., 63 (2015), 4179–4190.

[12] H. Suzuki, D. Robertson, N. L. Ratnayake and K. Ziri-Castro, Prediction and measurement of multiuser MIMO-OFDM channel in rural Australia, IEEE 75th Vehicular Technology Conference, (2012), 1–5.

[13] L. Vandenberghe, *Lecture Notes: Optimization Methods for Large-Scale Systems*, UCLA, Spring 2013-2014.
[14] A. Wiesel, Y. C. Eldar and S. Shamai (Shitz), Linear precoding via conic optimization for fixed MIMO receivers, *IEEE Trans. Signal Process.*, **54** (2006), 161–176.

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

Received March 2015; 1st revision December 2015; final revision September 2016.

E-mail address: bin.li@curtin.edu.au
E-mail address: h.dam@curtin.edu.au
E-mail address: antonio.cantoni@uwa.edu.au