ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-4485-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICASSP48485.2024.10446220

A STOCHASTIC PROXIMAL WMMSE FOR ERGODIC SUM RATE MAXIMIZATION

Xiaotong Zhao[1], *Xi Wang*[1], *Juncheng Wang*[2], *and Qingjiang Shi*[1,3]

[1]School of Software Engineering, Tongji University, Shanghai, China
[2]Department of Computer Science, Hong Kong Baptist University, Hong Kong, China
[3]Shenzhen Research Institute of Big Data, Shenzhen, China

ABSTRACT

We consider ergodic weighted sum rate (WSR) maximization in a massive multi-user multiple-input multiple-output system. Existing solutions iteratively minimize the average WSR based on all the historical information, and use bisection search to satisfy the power constraint at each iteration, resulting in both high storage burden and high computational complexity. In contrast, we propose an efficient stochastic proximal weighted minimum mean-square error (SP-WMMSE) algorithm, which updates the precoder only based on the current single channel realization, without checking the power constraint at each iteration. Furthermore, we propose a novel proximal term to incorporate all the previous channel and surrogate function information in precoder updates. Our analysis shows that SPWMMSE converges to the stationary point of the original ergodic WSR maximization problem almost surely. Simulation results demonstrate the effectiveness of SPWMMSE over the current best alternatives.

Index Terms— Massive MU-MIMO, precoding, imperfect CSI, stochastic WMMSE, sum-rate maximization

1. INTRODUCTION

Massive multi-user (MU) multiple-input multiple-output (MIMO) is an enabling technique for wireless communication systems [1, 2]. A fundamental problem in MU-MIMO systems is to design transmit precoders that maximize the weighted sum rate (WSR) subject to power constraints. The weighted minimum mean-square error (WMMSE) iterative algorithm [3, 4] has been recognized as the benchmark solution to WSR maximization. WMMSE reformulates the WSR objective function into a variational form and then employs analytic block coordinate descent (BCD) updates. Additionally, the WMMSE method guarantees convergence to stationary points in a non-decreasing manner and has been shown to outperform other optimization techniques numerically [4]. However, the standard WMMSE relies on perfect channel state information (CSI). In a practical system, perfect channel statistics may not be available due to channel estimation errors and aging [5].

In this work, we focus on ergodic WSR maximization in MU-MIMO systems with imperfect CSI, i.e., stochastic CSI with unknown distributions. Existing works have utilized the WMMSE framework to solve the ergodic WSR maximization problem [5–9]. In particular, a robust WMMSE scheme was proposed in [5] to maximize a lower bound of the ergodic WSR. However, the lower bound can be loose and thus maximizing it inevitably leads to performance degradation. Other approaches [6–9] utilized a stochastic optimization framework [9–11] to directly optimize the ergodic WSR. These methods are well-represented by the stochastic WMMSE (SWMMSE) algorithm proposed in [6], which iteratively constructs a surrogate function using the WMMSE approach [4] and then updates the precoder by optimizing the average of all the historical surrogate functions.

However, the SWMMSE algorithm has several limitations. First, it requires storing all the historical surrogate functions with numerous variables, resulting in a significant storage burden. Second, optimizing the average function can lead to inaccurate surrogate, since the previous iteration points can be far from the optimal solution. It should be noted that the deterministic WMMSE method only optimizes the current rather than the average surrogate. Finally, the bisection search used in SWMMSE to satisfy the power constraint in each precoder update is of high computational complexity.

In this paper, we propose an effective stochastic proximal WMMSE (SPWMMSE) algorithm for maximizing the ergodic WSR. SPWMMSE is built upon the WMMSE framework with a novel proximal term to incorporate all the historical precoder and surrogate function information. Unlike SWMMSE which relies on the historical channel realizations, SPWMMSE iteratively updates the precoder solely based on the current instantaneous channel realization, which significantly reduces the storage burden. Furthermore, we transform the ergodic WSR problem into an unconstrained problem in the SPWMMSE to completely remove the bisection search in each precoder update. Our precoder solutions are in closed form with lower computational complexity than SWMMSE. Our analysis shows that SPWMMSE converges almost surely to the stationary point of the original ergodic WSR maximization problem. Numerical results demonstrate that SP-WMMSE substantially outperforms the current best alterna-

This work was supported by the NSFC under Grant 62231019. Corresponding author: *Qingjiang Shi*.

tives in both the WSR and run time.

2. ERGODIC SUM RATE MAXIMIZATION

We consider a downlink massive MU-MIMO system, where a BS is equipped with M transmit antennas. The BS serves K users, and each user k has N_k receive antennas. Let $\mathbf{H}_k \in \mathbb{C}^{N_k \times M}$ denote the MIMO channel state between the BS and user k. Using $\mathbf{H} \triangleq \{\mathbf{H}_k\}_{k=1}^K$, the BS designs a precoding matrix $\mathbf{P}_k \in \mathbb{C}^{M \times D_k}$ to transmit D_k symbols $\mathbf{s}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ to each user k. The received signal vector $\mathbf{y}_k \in \mathbb{C}^{N_k \times 1}$ at each user k is $\mathbf{y}_k = \mathbf{H}_k\mathbf{P}_k\mathbf{s}_k + \sum_{j=1,j\neq k}^K \mathbf{H}_k\mathbf{P}_j\mathbf{s}_j + \mathbf{n}_k$, where $\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}, \sigma_k^2\mathbf{I})$ is the additive Gaussian noise vector with σ_k^2 being the noise power.

The instantaneous achievable rate of each user k is

$$R_k^{\text{inst}} \triangleq \log\det\Bigg(\mathbf{I} + \mathbf{H}_k\mathbf{P}_k\mathbf{P}_k^H\mathbf{H}_k^H \times \Big(\sum_{j\neq k}\mathbf{H}_k\mathbf{P}_j\mathbf{P}_j^H\mathbf{H}_k^H + \sigma_k^2\mathbf{I}\Big)^{-1}\Bigg). \tag{1}$$

If the precise channel information is unavailable, $\mathbf{H}$ can be treated as random variables. The ergodic achievable rate of user k is given by $R_k = \mathbb{E}_{\mathbf{H}}\left[R_k^{\text{inst}}\right]$. We aim at optimizing precoding matrices $\mathbf{P} \triangleq \{\mathbf{P}_k\}_{k=1}^K$ to maximize the ergodic WSR, subject to the maximum transmit power constraint:

$$\max_{\mathbf{P}} \ \sum_{k=1}^K \alpha_k R_k \tag{2a}$$

$$\text{s.t.} \ \sum_{k=1}^K \operatorname{Tr}\left(\mathbf{P}_k\mathbf{P}_k^H\right) \leq P_{\max}, \tag{2b}$$

where α_k is the weight of user k and $P_{\max}$ is the maximum transmit power limit. Without loss of generality, we set equal weights to the users and omit α_k in the rest of the paper.

Problem (2) is a stochastic non-convex optimization problem. It is challenging to solve, even its deterministic version has been shown to be NP-hard [12]. The WMMSE framework [4] is one of the most popular methods for addressing the deterministic version of problem (2). A stochastic WMMSE (SWMMSE) algorithm was proposed in [6] for solving problem (2). Some subsequent studies have followed this framework [7–9]. However, as will be discussed in Section 3.3, the SWMMSE framework has some drawbacks, such as high storage requirement, high computational complexity, and performance degradation especially when the BS is equipped with a large number of antennas. In this work, we aim to develop an efficient algorithm that utilizes the WMMSE framework [4] and is guaranteed to converge to a stationary point of problem (2).

3. STOCHASTIC PROXIMAL WMMSE (SPWMMSE)

3.1. Equivalent Unconstrained Reformulation

Conventional SWMMSE algorithms require bisection search to satisfy the power constraint (2b), resulting in a significant computational burden. In the following lemma, we show that the stationary points of problem (2) must consume the maximum power. We can then use this unique property to transform problem (2) into an equivalent unconstrained optimization problem, to avoid checking the power constraint (2b). The proof follows from analyzing the Karush-Kuhn-Tucker (KKT) conditions of problem (2) and is omitted due to space constraints [13].

Lemma 1 *Any stationary point of problem* (2) *must satisfy the power constraint* (2b) *with equality.*

Using Lemma 1 and the fractional structure of signal-to-interference-plus-noise ratio (SINR), we can simplify problem (2) to an unconstrained optimization problem as

$$\max_{\mathbf{P}} \sum_{k=1}^K \mathbb{E}_{\mathbf{H}}\left[\log\det\left(\mathbf{I} + \mathbf{H}_k\mathbf{P}_k\mathbf{P}_k^H\mathbf{H}_k^H\mathbf{N}_k^{-1}\right)\right], \tag{3}$$

where $\mathbf{N}_k \triangleq \sum_{j\neq k}\mathbf{H}_k\mathbf{P}_j\mathbf{P}_j^H\mathbf{H}_k^H + \frac{\sigma_k^2}{P_{\max}}\sum_{i=1}^K \operatorname{Tr}(\mathbf{P}_i\mathbf{P}_i^H)\mathbf{I}$. The following lemma shows that for any stationary point of problem (2), we can linearly scale it to obtain a stationary point of problem (3), and vice versa. The proof follows from analyzing and comparing the KKT conditions of problems (2) and (3), and is omitted due to space limitations [13].

Lemma 2 *For any stationary point* $\{\mathbf{P}_k^\star\}$ *of the constrained problem* (2)*, there exists a stationary point* $\{\mathbf{P}_k^\ddagger\}$ *of the unconstrained problem* (3) *such that* $\mathbf{P}_k^\star = \sqrt{\omega}\mathbf{P}_k^\ddagger, \forall k$*, where* $\omega \triangleq \frac{P_{\max}}{\sum_{k=1}^K \operatorname{Tr}\left(\mathbf{P}_k^\ddagger(\mathbf{P}_k^\ddagger)^H\right)}$ *is a scaling factor, and vice versa.*

Therefore, the original constrained problem (2) can be equivalently solved by the unconstrained problem (3).

3.2. SPWMMSE Algorithm

Now we show how to solve problem (3). The major challenge of solving the stochastic problem (3) lies in the expectation of the channel state $\mathbf{H}$. Stochastic gradient descent (SGD) [14] is one of the most commonly used methods for solving stochastic optimization problems such as problem (3). In each iteration, the SGD method first draws one realization of $\mathbf{H}$ and then takes a step in the negative gradient direction. Essentially, SGD minimizes the linear expansion of the objective function of problem (3), by adding a proximal term that is related to the gradient descent step size.

Our solution method also has a proximal term. However, different from the SGD method for general stochastic optimization problems, we utilize the majorization minimization (MM) method [15] to fully exploit the unique structure of problem (3), by minimizing the upper bound of the negative objective function instead of its linear expansion for faster convergence speed.

The following lemma will help us to construct an upper bound of the negative objective function of problem (3). The proof follows from checking the first-order optimality conditions of problem (3), and is omitted due to space limits.

Lemma 3 *The objective function of problem* (3) *satisfies*

$$\log\det(\mathbf{I}+\mathbf{H}_k\mathbf{P}_k\mathbf{P}_k\mathbf{P}_k^H\mathbf{N}_k^{-1}) = \max_{\mathbf{W}_k\succ\mathbf{0},\mathbf{U}_k}\log\det\mathbf{W}_k-\mathrm{Tr}(\mathbf{W}_k\mathbf{E}_k(\mathbf{U}_k,\mathbf{P},\mathbf{H}_k))+D_k,\forall k,$$

where $\{\mathbf{U}_k\}$, $\{\mathbf{W}_k\}$ *are auxiliary variables, and*

$$\mathbf{E}_k\left(\mathbf{U}_k,\mathbf{P},\mathbf{H}_k\right) \triangleq (\mathbf{I}-\mathbf{U}_k^H\mathbf{H}_k\mathbf{P}_k)(\mathbf{I}-\mathbf{U}_k^H\mathbf{H}_k\mathbf{P}_k)^H+\mathbf{U}_k^H\mathbf{N}_k\mathbf{U}_k.$$

Based on Lemma 3, problem (3) can be equivalently reformulated as the following problem:

$$\min_{\mathbf{P}}\mathbb{E}_{\mathbf{H}}\left[\min_{\mathbf{W},\mathbf{U}}\sum_{k=1}^{K}\left(\mathrm{Tr}\left(\mathbf{W}_k\mathbf{E}_k(\mathbf{U}_k,\mathbf{P},\mathbf{H}_k)\right)-\log\det\mathbf{W}_k\right)\right], \tag{4}$$

where $\mathbf{W}\triangleq\{\mathbf{W}_k\succ\mathbf{0}\}_{k=1}^K$, $\mathbf{U}\triangleq\{\mathbf{U}_k\}_{k=1}^K$. Define $h\left(\mathbf{P},\mathbf{U},\mathbf{W},\mathbf{H}\right)\triangleq\sum_{k=1}^K(\mathrm{Tr}(\mathbf{W}_k\mathbf{E}_k\left(\mathbf{U}_k,\mathbf{P},\mathbf{H}_k\right))-\log\det\mathbf{W}_k)$, whose minimal solution is denoted as $(\mathbf{U}^\star,\mathbf{W}^\star)$. Note that $h(\mathbf{P},\mathbf{U},\mathbf{W},\mathbf{H})$ is the objective function of the inner problem in (4). From Lemma 3, $h\left(\mathbf{P},\mathbf{U}^\star,\mathbf{W}^\star,\mathbf{H}\right)$ provides a upper bound of the negative objective function in (3). Note that problem (3) is a maximization problem. Here we equivalently consider it as a minimization problem for convenience and thus find the upper bound for the negative objective function.

We now provide a solution to problem (4), which iteratively minimizes the upper bound of the negative objective function of problem (3). Specifically, after observing one channel realization $\mathbf{H}^r\triangleq\{\mathbf{H}_k^r\}_{k=1}^K$ at iteration r, we first update $\mathbf{W}^r$ and $\mathbf{U}^r$ by minimizing $h\left(\mathbf{P}^{r-1},\mathbf{U},\mathbf{W},\mathbf{H}^r\right)$ via

$$(\mathbf{U}^r,\mathbf{W}^r)=\arg\min_{\mathbf{U},\mathbf{W}}\ h\left(\mathbf{P}^{r-1},\mathbf{U},\mathbf{W},\mathbf{H}^r\right). \tag{5}$$

We then update precoder $\mathbf{P}^r$ by minimizing the upper bound of the negative objective function in (3), plus a proximal term $\frac{1}{2\rho_r}\|\mathbf{P}-\mathbf{P}^{r-1}\|_F^2$, where $\rho_r>0$ is a proximal parameter that tunes the distance between $\mathbf{P}^r$ and $\mathbf{P}^{r-1}$. Specifically, we update $\mathbf{P}^r$ by solving

$$\mathbf{P}^r=\arg\min_{\mathbf{P}}\sum_{k=1}^{K}\mathrm{Tr}\left(\mathbf{W}_k^r\mathbf{E}_k(\mathbf{U}_k^r,\mathbf{P},\mathbf{H}_k^r)\right)+\frac{1}{2\rho_r}\|\mathbf{P}-\mathbf{P}^{r-1}\|_F^2. \tag{6}$$

We now provide solutions $\mathbf{U}_k^r$, $\mathbf{W}_k^r$, $\mathbf{P}_k^r$ to (5) and (6) for each user k at each iteration r. Our solutions are in closed-form, by checking the first-order optimality conditions of (5) and (6). Specifically, we update $\mathbf{U}_k^r$, $\mathbf{W}_k^r$, $\mathbf{P}_k^r$ as

- $$\mathbf{U}_k^r=\left(\sum_{i=1}^{K}\frac{\sigma_k^2}{P_{\max}}\mathrm{Tr}\left(\mathbf{P}_i^{r-1}(\mathbf{P}_i^{r-1})^H\right)\mathbf{I} +\sum_{j=1}^{K}\mathbf{H}_k^r\mathbf{P}_j^{r-1}(\mathbf{P}_j^{r-1})^H\left(\mathbf{H}_k^r\right)^H\right)^{-1}\mathbf{H}_k^r\mathbf{P}_k^{r-1},\ \forall\, k,$$
- $$\mathbf{W}_k^r\leftarrow\left(\mathbf{I}-(\mathbf{U}_k^r)^H\mathbf{H}_k^r\mathbf{P}_k^{r-1}\right)^{-1},\ \forall\, k,$$
- $$\mathbf{P}_k^r=\left(\left(\beta+\frac{1}{2\rho_r}\right)\mathbf{I}+\sum_{j=1}^{K}\left(\mathbf{H}_j^r\right)^H\mathbf{M}_j^r\mathbf{H}_j^r\right)^{-1} \times\left(\left(\mathbf{H}_k^r\right)^H\mathbf{U}_k^r\mathbf{W}_k^r+\frac{1}{2\rho_r}\mathbf{P}_k^{r-1}\right),\ \forall\, k,$$

where $\beta\triangleq\sum_{i=1}^K\frac{\sigma_i^2}{P_{\max}}\mathrm{Tr}\left(\mathbf{M}_i^r\right)$ and $\mathbf{M}_k^r\triangleq\mathbf{U}_k^r\mathbf{W}_k^r(\mathbf{U}_k^r)^H$.

Upon reaching convergence, the final precoders can be obtained through $\mathbf{P}_k^\star=\sqrt{\omega}\mathbf{P}_k,\forall k$, where ω is set such that $\mathbf{P}^\star$ consumes full transmit power as in Lemma 2. Our precoder solution approach is referred to as stochastic proximal WMMSE (SPWMMSE).

3.3. Performance and Complexity Analysis

The per-iteration complexities of the proposed SPWMMSE algorithm and the SWMMSE algorithms [6–9] are respectively $\mathcal{O}\left(M^3\right)$ and $\mathcal{O}\left(TM^3\right)$, where T is the number of bisections searched in SWMMSE at each iteration. Notably, the proposed SPWMMSE algorithm avoids bisection search to satisfy (2b) at each precoder updating iteration, which substantially reduces the computational complexity. Furthermore, compared with the SWMMSE algorithms, the proposed SPWMMSE algorithm significantly reduces the storage requirement by M^2, since SPWMMSE does not require the historical surrogate functions to run. This is particularly advantageous for practical massive MIMO systems with a large number of transmit antennas M.

The following theorem shows that our proposed SPWMMSE algorithm is guaranteed to converge to a stationary point of problem (3) almost surely.

Theorem 4 *Suppose* $\sum_{r=1}^\infty\rho_r=\infty$ *and* $\sum_{r=1}^\infty\rho_r^2<\infty$, *e.g.,* $\rho_r=1/r$, *the SPWMMSE algorithm converges to a stationary point of problem* (3) *almost surely.*

Proof: Firstly, we define $\mathbf{v}\triangleq(\mathbf{U},\mathbf{W})$ and $g\left(\mathbf{P},\mathbf{v},\mathbf{H}\right)\triangleq\sum_{k=1}^K\left(-\log\det\left(\mathbf{W}_k\right)+\mathrm{Tr}\left(\mathbf{W}_k\mathbf{E}_k\left(\mathbf{U}_k,\mathbf{P},\mathbf{H}_k\right)\right)\right)$. Let us further define $f\left(\mathbf{P},\mathbf{H}\right)\triangleq\min_{\mathbf{v}}g\left(\mathbf{P},\mathbf{v},\mathbf{H}\right)$, and $F\left(\mathbf{P}\right)\triangleq\mathbb{E}_{\mathbf{H}}\left[f\left(\mathbf{P},\mathbf{H}\right)\right]$, which is exactly the negative of the objective function in problem (3). Finally, we define $G\left(\mathbf{P},\mathbf{v}\right)\triangleq\mathbb{E}_{\mathbf{H}}\left[g\left(\mathbf{P},\mathbf{v},\mathbf{H}\right)\right]$. In the following, we provide a sketch of the proof in three steps.

Step 1: Using the optimality of $\mathbf{P}^r$ in (6), we find that $g\left(\mathbf{P}^{r-1},\mathbf{v}^r,\mathbf{H}^r\right)-g(\mathbf{P}^r,\mathbf{v}^r,\mathbf{H}^r)\geq\frac{\rho_r}{2}\|\nabla_{\mathbf{P}}g\left(\mathbf{P}^r,\mathbf{v}^r,\mathbf{H}^r\right)\|_F^2$. Note that $g\left(\mathbf{P},\mathbf{v}^r,\mathbf{H}^r\right)\geq f\left(\mathbf{P},\mathbf{H}^r\right)$, it can be shown that $f\left(\mathbf{P}^{r-1},\mathbf{H}^r\right)-f\left(\mathbf{P}^r,\mathbf{H}^r\right)\geq\frac{\rho_r}{2}\|\nabla_{\mathbf{P}}g\left(\mathbf{P}^r,\mathbf{v}^r,\mathbf{H}^r\right)\|_F^2$.

Step 2: By taking expectation over $\mathbf{H}^r$ given $\mathbf{P}^{r-1}$. We have $F\left(\mathbf{P}^{r-1}\right)-\mathbb{E}_{\mathbf{H}^r}\left[F\left(\mathbf{P}^r\right)\right]\geq\frac{\rho_r}{2}\mathbb{E}_{\mathbf{H}^r}\left[\|\nabla_{\mathbf{P}}G\left(\mathbf{P}^r,\mathbf{v}^r\right)\|_F^2\right]$. Then, by taking expectation with respect to $\{\mathbf{H}^1,\ldots,\mathbf{H}^{r-1}\}$, we have $\mathbb{E}\left[F\left(\mathbf{P}^0\right)\right]-F_{\min}\geq\sum_{r=1}^L\frac{\rho_r}{2}\mathbb{E}\left[\|\nabla_{\mathbf{P}}G\left(\mathbf{P}^r,\mathbf{v}^r\right)\|_F^2\right]$.

Step 3: Since we suppose $\sum_{r=1}^\infty\rho_r=\infty$, we have $\lim_{r\to\infty}\mathbb{E}\left[\|\nabla_{\mathbf{P}}G\left(\mathbf{P}^r,\mathbf{v}^r\right)\|_F^2\right]=0$. Further using the optimality condition, we have $\sum_{r=1}^L\frac{1}{\rho_r}\mathbb{E}\left[\|\mathbf{P}^r-\mathbf{P}^{r-1}\|_F^2\right]\leq\infty$. Since $\sum_{r=1}^\infty\rho_r^2<\infty$, implying $\rho_r\to0$, we have $\lim_{r\to\infty}\mathbb{E}\left[\|\mathbf{P}^r-\mathbf{P}^{r-1}\|_F^2\right]=0$. Noting that $\nabla_{\mathbf{P}}F\left(\mathbf{P}^{r-1}\right)=$

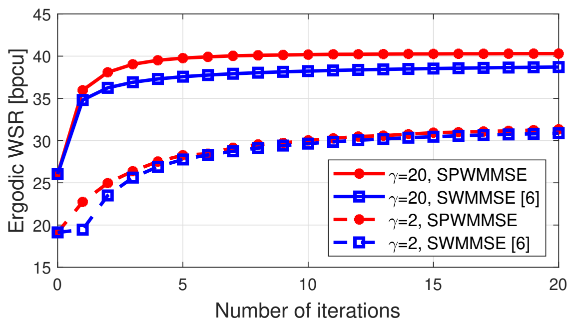

Fig. 1. Ergodic WSR versus iteration number.

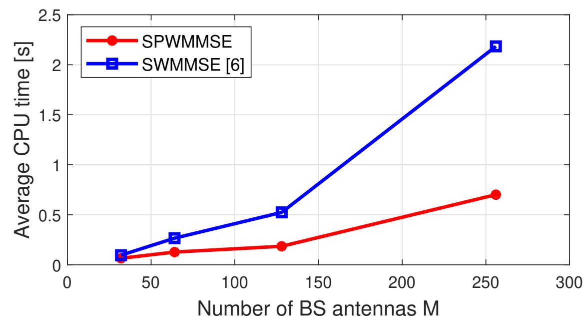

Fig. 2. Average CPU time versus M.

$\nabla_{\mathbf{P}} G\left(\mathbf{P}^{r-1}, \mathbf{v}^r\right)$, we have $\lim_{r\to\infty} \mathbb{E}\left[\|\nabla_{\mathbf{P}} F\left(\mathbf{P}^{r-1}\right)\|_F^2\right] = 0$. This proves that the precoder sequence $\{\mathbf{P}^r\}$ converges to the stationary point of problem (3) almost surely. ■

Note that the SWMMSE framework [6–9] iteratively optimizes the ensemble average of *all* the historical surrogate functions. The ensemble average tends to converge to a deterministic surrogate function of the WSR objective. It can then be shown that any limiting point of the SWMMSE is a stationary point. However, since our SPWMMSE method utilizes only the *current* channel realization in each iteration, proving Theorem 4 requires a completely different analysis from SWMMSE. Specifically, we leverage the optimality condition and the property of the surrogate function together with the proposed *proximal* term in (6), to show that the instantaneous WSR is sufficiently descent for each channel realization. Then, by handling the expectation over random channel realizations and summing all the historical descent values, we show that the gradient of the WSR objective approaches zero almost surely.

Furthermore, the SWMMSE framework [6–9] needs to store all the historical objective functions, which leads to a high storage burden, especially when the number of BS antennas is large. Also, optimizing the average surrogate function in SWMMSE may not be effective since the previous iteration point is generally far from the optimal solution, leading to an inaccurate estimation of the surrogate function. In contrast, the proposed SPWMMSE algorithm incorporates information about past iterates and surrogate functions within the proximal term, and uses a tunable parameter to tune the distance between the current precoder and the previous precoder.

4. NUMERICAL RESULTS

We consider a massive MU-MIMO system, where a BS is equipped with $M = 64$ antennas and serves $K = 8$ users. Each user k is equipped with $N_k = 4$ antennas to multiplex $D_k = 4$ data streams. We set the power budget of the BS to $P_{\text{max}} = 10$ [W]. The channel matrix $\mathbf{H}$ is generated from the circularly-symmetric standard complex normal distribution with path loss between the users and the BS, set to $\sigma_l^2 = 128.1 + 37.6\log_{10}(d)$[dB] [16], where d represents the distance between the user and the BS, ranging from $0.1 \sim 0.3$ km. The noise power is set as $\sigma_k^2 = 10^{\frac{1}{K}\sum_k \log_{10}\frac{1}{N_k}\|\mathbf{H}_k\|_F^2} \times 10^{-\frac{\text{SNR}}{10}}$. We consider inaccurate channel state by adding channel estimation error $\mathcal{CN}\left(0, \frac{\sigma_l^2}{1+\gamma^2}\right)$ to each channel state element of $\mathbf{H}$, where γ^2 is the effective SNR coefficient. In our simulations, the value of ρ_r is set to $\gamma/4r$ in the order of $1/r$ as in Theorem 4. Moreover, the ergodic WSR is calculated via 1000 Monte Carlo runs. We choose the algorithm in [6] as a baseline, which represents the SWMMSE framework [6–9].

In Fig. 1, we compare the WSR and the convergence performance of SPWMMSE and SWMMSE under different values of γ. We can see that the ergodic rate of the SPWMMSE algorithm is higher than that of the SWMMSE algorithm, particularly when γ is large and channel randomness is small. Furthermore, SPWMMSE converges faster than SWMMSE.

In Fig. 2, we compare the average CPU run time of the proposed SPWMMSE algorithm with that of the SWMMSE algorithm under different numbers of BS antennas M. We observe that the average execution time of the proposed SPWMMSE algorithm is significantly lower than that of the SWMMSE algorithm, especially when M is large. This indicates the performance advantage of the proposed SPWMMSE algorithm over the SWMMSE algorithm in massive MIMO systems.

5. CONCLUSIONS

We consider ergodic WSR maximization in MU-MIMO systems. We propose an efficient SPWMMSE algorithm, which iteratively minimizes the upper bound of the negative WSR objective with a novel proximal term that incorporates all the historical channel and surrogate function information. SPWMMSE updates the precoder only based on the current single channel realization without the need to check the power constraint at each iteration, leading to low storage burden and computational complexity. By equivalently transforming the original WSR maximization problem to an unconstrained optimization, we further prove that SPWMMSE converges almost surely to the stationary points of the original ergodic WSR maximization problem. Our numerical results demonstrate substantial performance advantage of SPWMMSE over the current best alternatives, in terms of both the WSR performance and run time.

6. REFERENCES

[1] Thomas L Marzetta and Hien Quoc Ngo, *Fundamentals of massive MIMO*, Cambridge Univ. Press, Cambridge, U.K., 2016.

[2] Mingjin Wang, Feifei Gao, Shi Jin, and Hai Lin, "An overview of enhanced massive MIMO with array signal processing techniques," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 5, pp. 886–901, Sep. 2019.

[3] Søren Skovgaard Christensen, Rajiv Agarwal, Elisabeth De Carvalho, and John M Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, Dec. 2008.

[4] Qingjiang Shi, Meisam Razaviyayn, Zhi-Quan Luo, and Chen He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.

[5] Junchao Shi, An-An Lu, Wen Zhong, Xiqi Gao, and Geoffrey Ye Li, "Robust WMMSE precoder with deep learning design for massive MIMO," *IEEE Trans. Commun.*, pp. 3963 – 3976, Apr. 2023.

[6] Meisam Razaviyayn, Maziar Sanjabi Boroujeni, and Zhi-Quan Luo, "A stochastic weighted mmse approach to sum rate maximization for a MIMO interference channel," in *Proc. IEEE Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, 2013.

[7] Hamdi Joudeh and Bruno Clerckx, "Sum-rate maximization for linearly precoded downlink multiuser MISO systems with partial CSIT: A rate-splitting approach," *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4847–4861, Nov. 2016.

[8] Alaa Alameer Ahmad, Yijie Mao, Aydin Sezgin, and Bruno Clerckx, "Rate splitting multiple access in C-RAN: A scalable and robust design," *IEEE Trans. Commun.*, vol. 69, no. 9, pp. 5727–5743, Jun. 2021.

[9] Meisam Razaviyayn, Maziar Sanjabi, and Zhi-Quan Luo, "A stochastic successive minimization method for nonsmooth nonconvex optimization with applications to transceiver design in wireless communication networks," *Math. Program.*, vol. 157, pp. 515–545, Jun. 2016.

[10] Julien Mairal, "Stochastic majorization-minimization algorithms for large-scale optimization," *Proc. Adv. Neural Info. Proc. Sys. (NeurIPS)*, 2013.

[11] Julien Mairal, "Incremental majorization-minimization optimization with application to large-scale machine learning," *SIAM J. Optim.*, vol. 25, no. 2, pp. 829–855, 2015.

[12] Zhi-Quan Luo and Shuzhong Zhang, "Dynamic spectrum management: Complexity and duality," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 57–73, Feb. 2008.

[13] Xiaotong Zhao, Siyuan Lu, Qingjiang Shi, and Zhi-Quan Luo, "Rethinking WMMSE: Can its complexity scale linearly with the number of BS antennas?," *IEEE Trans. Signal Process.*, vol. 71, pp. 433–446, Feb. 2023.

[14] Léon Bottou, Frank E Curtis, and Jorge Nocedal, "Optimization methods for large-scale machine learning," *SIAM review*, vol. 60, no. 2, pp. 223–311, 2018.

[15] Ying Sun, Prabhu Babu, and Daniel P Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, Feb. 2016.

[16] Hayssam Dahrouj and Wei Yu, "Coordinated beamforming for the multicell multi-antenna wireless system," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1748–1759, May 2010.