Finite Horizon Optimization for Large-Scale MIMO

Yi Feng, *Graduate Student Member, IEEE* and Kaiming Shen, *Senior Member, IEEE*

Abstract—Large-scale multiple-input multiple-output (MIMO) is an emerging wireless technology that deploys thousands of transmit antennas at the base-station to boost spectral efficiency. The classic weighted minimum mean-square-error (WMMSE) algorithm for beamforming is no suited for the large-scale MIMO because each iteration of the algorithm then requires inverting a matrix whose size equals the number of transmit antennas. While the existing methods such as the reduced WMMSE algorithm seek to decrease the size of matrix to invert, this work proposes to eliminate this large matrix inversion completely by applying gradient descent method in conjunction with fractional programming. Furthermore, we optimize the step sizes for gradient descent from a finite horizon optimization perspective, aiming to maximize the performance after a limited number of iterations of gradient descent. Simulations show that the proposed algorithm is much more efficient than the WMMSE algorithm in optimizing the large-scale MIMO precoders.

Index Terms—Beamforming, finite horizon optimization, large-scale multiple-input multiple-output (MIMO), large matrix inversion, gradient descent, step size, fractional programming.
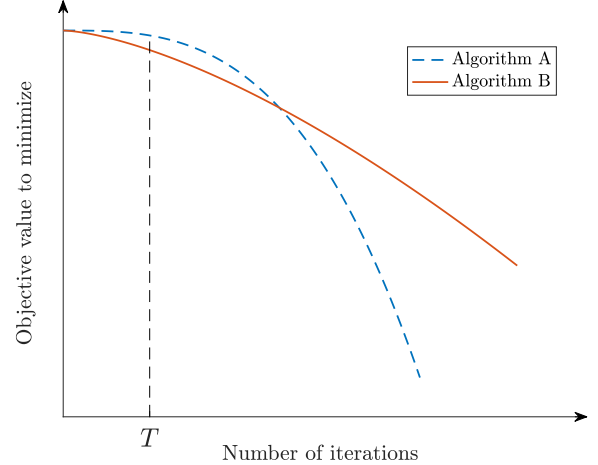

Fig. 1. According to the conventional optimization theory, algorithm A is better because its asymptotic convergence rate is higher when the number of iterations is sufficiently large. However, in the context of finite horizon optimization, algorithm B can outperform algorithm A when the attention is restricted to a finite range up to T iterations.

I. INTRODUCTION

LARGE-SCALE multiple-input multiple-output (MIMO) has imposed a huge challenge to the wireless network optimization because the corresponding beamforming problem is not only nonconvex but also involves high-dimensional variables [1]. Unfortunately, the well-known weighted minimum mean-square-error (WMMSE) algorithm [2], [3] is not suited for the large-scale MIMO case, because it entails inverting a large matrix whose size equals the number of transmit antennas for each iteration. To remedy the situation, this work examines the large-scale MIMO beamforming problem from a novel viewpoint of *finite horizon optimization*: how to maximize the performance given a limited number of iterations as illustrated in Fig. 1.

For the beamforming problem, many early works [4]–[8] pursue the minimum total power of precoders under the given signal-to-interference-plus-noise ratio (SINR) constraints. Despite the nonconvexity, this min-power beamforming problem can be solved globally in an efficient way based on the uplink-downlink duality [4]. Another beamforming problem formulation in the literature [9]–[11] is to maximize the minimum SINR across receivers under the given power constraint; the resulting max-min problem is amenable to the semidefinite programming (SDP) approach. In particular, the max-min SINR problem can be approximately solved to within a constant factor for the single-cell network [9]. The modern works in the area mostly study the weighted sum rates (WSR) maximization

for the beamforming problem. Since the WSR problem is NP-hard [12], the previous works that aim at global optimum must resort to those exponential-time optimization methods, e.g., the exhaustive search [13], the branch-and-bound algorithm [14], and the polyblock optimization [15]. Some other existing works have suggested various approximations of the WSR problem to make the solution practical, e.g., the signomial programming approximation [16], the geometric programming approximation [17], the successive convex approximation [18], and the minorization-maximization (MM) approximation [19]. In particular, as a classic and simple approximation, the zero-forcing (ZF) beamforming method [20] just ignores the background noise and then nullifies the interfering signals. The authors of [21] improve the ZF beamforming method to reduce the approximation error, by dropping a set of weakest eigen-channels. Moreover, it is worth mentioning the information theoretic approach to the beamforming problem based on interference alignment [22]–[25], wherein the data rate in the WSR problem is approximated as the degrees-of-freedom (DoF).

The WMMSE algorithm [2], [3], first proposed about 20 years ago, still constitutes one of the most popular beamforming methods nowadays. It derives from the following insight: the data rate maximization is equivalent to the received signal mean-square-error (MSE) minimization. Furthermore, [26] shows that the WMMSE algorithm is intimately related to fractional programming (FP). From an FP perspective, the WMMSE algorithm can be interpreted as a two-stage method:

first, it moves the SINR terms to the outside of $\log(1+\mathrm{SINR})$ to obtain a sum-of-ratios problem; second, it addresses the sum-of-ratios problem by the quadratic transform (QT) [26]. We remark that, at the second stage, the QT can be performed in various ways, one particular way leading to the WMMSE algorithm. Thus, the WMMSE algorithm turns out to be a special case of FP, and it may not be the best case of FP for beamforming as shown in [27], [28].

The WMMSE algorithm [2], [3] is based on an iterative paradigm. As a desirable feature, WMMSE provides closed-form solution for each iteration. Note that such closed-form solution can be interpreted as the projection onto the surface of an M-dimensional ellipsoid [29], where M is the number of transmit antennas, so it requires inverting an $M \times M$ matrix. This matrix inverse operation becomes costly when M is large, so the WMMSE algorithm is not suited for the large-scale MIMO beamforming (e.g., with thousands of antennas deployed [30], [31]). To address this issue, a recent work [32] proposes the so-called *reduced WMMSE* algorithm that maps the original beamforming variable from the $M \times M$ space into the $NK \times NK$ space, where N is the number of receive antennas at each user terminal and K is the number of user terminals per cell. As a consequence, the new algorithm requires inverting an $NK \times NK$ matrix instead for each iteration. Yet there are two subtle issues with the reduced WMMSE algorithm [32]. First, it is possible that $NK > M$ in practice, and then the $NK \times NK$ matrix inversion can be even more complex than the $M \times M$ one. Second, it can be difficult to recover the original solution for the multi-cell network, because the beamforming variable mapping is shown to be invertible only for the single-cell case.

Alternatively, a more recent work [33] proposes to eliminate the matrix inversion by the *nonhomogeneous quadratic transform (NQT)*—an improved version of the QT method for FP [26]. However, the NQT would slow down the convergence of the iterative optimization. As shown in [29], both NQT and WMMSE can be interpreted as constructing a concave surrogate function for the original objective function of the beamforming problem, but the surrogate function constructed by WMMSE is tighter, so the WMMSE algorithm [2], [3] converges faster than the NQT algorithm [33]. To be more specific, letting f^t be the value of the WSR objective after t iterations, and $f^\star$ the value of convergence, [29] shows that WMMSE achieves

$$|f^t - f^\star| = \frac{\alpha}{t} \quad \text{as} \quad t \to \infty, \tag{1}$$

while the QT method achieves

$$|f^t - f^\star| = \frac{\beta}{t} \quad \text{as} \quad t \to \infty, \tag{2}$$

where $\beta > \alpha > 0$. Furthermore, [29] shows that the NQT method can be interpreted as a gradient projection, so Nesterov's extrapolation scheme [34] can be incorporated into NQT to accelerate its convergence. Eventually, [29] shows that the extrapolated NQT leads to the following convergence rate:

$$|f^t - f^\star| = \frac{\gamma}{t^2} \quad \text{as} \quad t \to \infty, \tag{3}$$

where $\gamma > 0$. Thus, the extrapolated NQT has much faster convergence than WMMSE and NQT.

Nevertheless, the above convergence rate analyses from [29] only reflect the asymptotic speed as $t \to \infty$. But for a real-world network it is the finite or even small t that truly matters. Toward this end, this work considers the *finite horizon optimization* of the large-scale MIMO precoders. Specifically, given a finite positive integer T, we aim to maximize the performance after T iterations, i.e.,

$$\text{maximize} \quad f^T \quad \text{for some given} \quad T < \infty. \tag{4}$$

The finite horizon optimization is a newly emerging notion initiated in the machine learning field [35] to facilitate the finite-depth neural network training [36], [37]. To the best of our knowledge, the finite horizon optimization has not yet been explored in the area of wireless communications and signal processing. We wish to clarify the distinction between the finite horizon optimization and the optimization with finite convergence (e.g., the conjugate gradient descent [38]). The finite horizon optimization does not guarantee convergence to a global optimum or stationary point after T iterations. Rather, it aims at what one can expect at best after T iterations. The main contributions of this paper are summarized below:

- We recast the constrained log-det beamforming problem into an unconstrained quadratic program by means of the matrix FP, extending the analogous result by Zhao, Lu, Shi, and Luo [32].
- We propose solving the unconstrained quadratic program by the finite-horizon gradient descent, and thereby avoid inverting a large matrix whose size equals the number of transmit antennas of the large-scale MIMO.
- We further optimize the step sizes of the finite-horizon gradient descent by the Chebyshev polynomial theory. To the best of our knowledge, this is the first time that the finite horizon optimization is applied to the communication system design.

Notation: For a matrix $\boldsymbol{A}$, $\boldsymbol{A}^{\mathrm{H}}$ denotes its conjugate transpose, $\|\boldsymbol{A}\|_2$ denotes its ℓ_2 norm that equals the largest singular value of $\boldsymbol{A}$, $\|\boldsymbol{A}\|_F$ denotes its Frobenius norm. For a square matrix $\boldsymbol{A}$, $\mathrm{tr}(\boldsymbol{A})$ denotes its trace, and $|\boldsymbol{A}|$ denotes its determinate. For a nonsingular matrix $\boldsymbol{A}$, $\boldsymbol{A}^{-1}$ denotes its inverse. Moreover, $\boldsymbol{I}$ denotes the identity matrix, $\mathbb{C}^d$ the set of $d \times 1$ vectors, $\mathbb{C}^{d \times m}$ the set of $d \times m$ matrices, and $\mathbb{S}_+$ the set of positive semidefinite matrices. For a complex number $a \in \mathbb{C}$, $\Re\{a\}$ is its real part, $|a|$ is its absolute value. For a differentiable function $g(x)$, $\nabla g(x)$ denotes its gradient. Let $\mathbb{P}_T = \{c_T x^T + c_{T-1} x^{T-1} + \ldots + c_1 x + c_0 : c_T \neq 0\}$ be the set of degree-T polynomials given a positive integer T.

II. System Model

Consider a single-cell downlink large-scale MIMO network; the multi-cell case is discussed later in Section V. Assume that the base-station (BS) has M transmit antennas and each user terminal has N antennas; it shall be understood that $M \gg N$ in the large-scale MIMO system. The BS aims to send independent messages to a total of K user terminals in the cell simultaneously via spatial multiplexing. Assume that at

most $d \leq N$ data streams can be transmitted to the same user terminal. We use $k, j = 1, \ldots, K$ to index the users. Denote by $\boldsymbol{H}_k \in \mathbb{C}^{N \times M}$ the downlink channel from BS to the kth user, $\boldsymbol{V}_k \in \mathbb{C}^{M \times d}$ the transmit precoder intended for the kth user, and σ^2 the background noise power. Thus, an achievable rate R_k of user terminal k can be computed as

$$R_k = \log \left| \boldsymbol{I} + \boldsymbol{V}_k^{\mathrm{H}} \boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{F}_k^{-1} \boldsymbol{H}_k \boldsymbol{V}_k \right|, \tag{5}$$

where

$$\boldsymbol{F}_k = \sigma^2 \boldsymbol{I} + \sum_{j \neq k} \boldsymbol{H}_k \boldsymbol{V}_j \boldsymbol{V}_j^{\mathrm{H}} \boldsymbol{H}_k^{\mathrm{H}}. \tag{6}$$

Each user terminal k is assigned a positive weight $\omega_k > 0$ in accordance with its priority. We seek the optimal precoding variable $\boldsymbol{V} = \{\boldsymbol{V}_k\}$ to maximize the WSR:

$$\underset{\boldsymbol{V}}{\text{maximize}} \quad \sum_{k=1}^{K} \omega_k R_k \tag{7a}$$

$$\text{subject to} \quad \sum_{k=1}^{K} \|\boldsymbol{V}_k\|_F^2 \leq P, \tag{7b}$$

where P is the power constraint. We remark that the number of transmit antennas M in (7) is much larger than the conventional MIMO case, e.g., it is of the order of thousands.

III. CONVERSION TO QUADRATIC PROGRAM

It is difficult to tackle the problem (7) directly because of its nonconvexity. This section shows that the problem can be rewritten as an unconstrained quadratic program. For the conventional MIMO network, the new problem is favorable in that each $\boldsymbol{V}_k$ can then be solved in closed form. Nevertheless, for the large-scale MIMO case with M being large, the closed-form solution of $\boldsymbol{V}_k$ entails an $M \times M$ matrix inversion which is computationally formidable in practice. A finite horizon optimization treatment of this new quadratic program is discussed in Section IV.

First of all, it is easy to notice that the power constraint (7b) must be tight at any *nontrivial*[1] stationary-point solution, for otherwise we can further increase the value of $\sum_{k=1}^{K} \omega_k R_k$ by scaling up all the $\boldsymbol{V}_k$'s simultaneously. Since the global optimum must be nontrivial, problem (7) is equivalent to

$$\underset{\boldsymbol{V}}{\text{maximize}} \quad \sum_{k=1}^{K} \omega_k R_k \tag{8a}$$

$$\text{subject to} \quad \sum_{k=1}^{K} \|\boldsymbol{V}_k\|_F^2 = P, \tag{8b}$$

in the sense that the two problems have the same global optimum (and also the same set of nontrivial stationary points).

Furthermore, problem (8) can be immediately rewritten as

$$\underset{\boldsymbol{V}}{\text{maximize}} \quad \sum_{k=1}^{K} \omega_k \log \left| \boldsymbol{I} + \boldsymbol{V}_k^{\mathrm{H}} \boldsymbol{H}_k^{\mathrm{H}} \widetilde{\boldsymbol{F}}_k^{-1} \boldsymbol{H}_k \boldsymbol{V}_k \right| \tag{9a}$$

$$\text{subject to} \quad \sum_{k=1}^{K} \|\boldsymbol{V}_k\|_F^2 = P, \tag{9b}$$

with $\boldsymbol{F}_k$ in (6) replaced by

$$\widetilde{\boldsymbol{F}}_k = \left(\frac{\sigma^2}{P} \sum_{k=1}^{K} \|\boldsymbol{V}_k\|_F^2 \right) \boldsymbol{I} + \sum_{j \neq k} \boldsymbol{H}_k \boldsymbol{V}_j \boldsymbol{V}_j^{\mathrm{H}} \boldsymbol{H}_k^{\mathrm{H}}. \tag{10}$$

We now remove the power equality constraint (9b) and obtain an unconstrained reformulation of problem (9) as

$$\underset{\boldsymbol{V}}{\text{maximize}} \quad \sum_{k=1}^{K} \omega_k \log \left| \boldsymbol{I} + \boldsymbol{V}_k^{\mathrm{H}} \boldsymbol{H}_k^{\mathrm{H}} \widetilde{\boldsymbol{F}}_k^{-1} \boldsymbol{H}_k \boldsymbol{V}_k \right|. \tag{11}$$

A crucial trait of the above new problem is that its objective value does not alter when all the $\boldsymbol{V}_k$ are multiplied by the same nonzero factor, as stated in the following lemma.

Lemma 1: Denote by $g(\boldsymbol{V})$ the objective function in (11). For any nonzero $\boldsymbol{V}$, we have $g(\boldsymbol{V}) = g(c\boldsymbol{V})$ given any nonzero factor c, and thus

$$\nabla g(\boldsymbol{V}) \perp \boldsymbol{V} \tag{12}$$

for any $\boldsymbol{V}$.

In light of the above lemma, we can show that[2]:

Theorem 1: For any $\boldsymbol{V}$ with $\sum_{k=1}^{K} \|\boldsymbol{V}_k\|_F^2 > 0$, it is a stationary point of problem (11) if and only if $\rho \boldsymbol{V}$ is a stationary point of problem (7), where $\rho = (P / \sum_{k=1}^{K} \|\boldsymbol{V}_k\|_F^2)^{1/2}$.

Proof: Denote by $f(\boldsymbol{V})$ the objective function in (7). If $\boldsymbol{V}$ is a trivial solution so that $\boldsymbol{H}_k \boldsymbol{V}_k = \boldsymbol{0}$ for every $k = 1, 2, \ldots, K$, then we have $\nabla f(\boldsymbol{V}) = \nabla g(\boldsymbol{V}) = \boldsymbol{0}$, so the theorem holds in this case. The rest of the proof assumes nontrivial $\boldsymbol{V}$.

We start with the necessity. If $\boldsymbol{V}$ is a stationary point of (11), then $\rho \boldsymbol{V}$ is a stationary point of (11) as well because $\nabla g(\rho \boldsymbol{V}) = \nabla g(\boldsymbol{V}) = \boldsymbol{0}$ according to Lemma 1. Clearly, $\rho \boldsymbol{V}$ is feasible for the power constraint and continues to be a stationary point in (9). Since (9) is equivalent to (7) in the nontrivial regime, we have $\rho \boldsymbol{V}$ be a stationary point of (7).

We then show the sufficiency. If $\rho \boldsymbol{V}$ is a stationary point of (7), then it is also a stationary point of (9) since the two problems are equivalent in the nontrivial regime. Thus, $\rho \boldsymbol{V}$ must satisfy the KKT condition in (9), i.e., $\nabla g(\rho \boldsymbol{V}) + 2\mu^{\star}(\rho \boldsymbol{V}) = \boldsymbol{0}$, where $\mu^{\star} \in \mathbb{R}$ is the optimal Lagrange multiplier for the constraint $\sum_{k=1}^{K} \|\boldsymbol{V}_k\|_F^2 = P$. We rewrite this KKT condition as $\nabla g(\boldsymbol{V}') = -2\mu^{\star} \boldsymbol{V}'$ where $\boldsymbol{V}' = \rho \boldsymbol{V}$. Combining the above result with Lemma 1, i.e., $\nabla g(\boldsymbol{V}') \perp \boldsymbol{V}'$, we obtain $\mu^{\star} = 0$ and $\nabla g(\boldsymbol{V}') = \boldsymbol{0}$. Again, by Lemma 1, we have $\nabla g(\boldsymbol{V}) = \nabla g(\rho^{-1} \boldsymbol{V}') = \nabla g(\boldsymbol{V}') = \boldsymbol{0}$. The proof is then completed. ∎

We remark that problems (11) and (7) are not guaranteed to be equivalent since their global optimum may not be

---

[1]As defined in [32], the solution $\boldsymbol{V}$ is said to be nontrivial if $\boldsymbol{H}_k \boldsymbol{V}_k \neq \boldsymbol{0}$ for at least one $k = 1, 2, \ldots, K$.

[2]A weak version of Theorem 1 has been shown in [32] assuming that $\boldsymbol{V}$ is nontrivial. The proof technique in [32] is different and is more computationally intensive: it explicitly compares the KKT conditions of (7) and (11).

identical. Yet the two problems have the same set of stationary points excluding the zero solution. Because the beamforming problem is NP-hard [12], it is adequate in most cases to reach a nonzero stationary point. As such, it suffices to consider the new problem (11) in the rest of the paper.

Next, we deal with problem (11) from an FP perspective. First, apply the Lagrangian dual transform [26] so as to move the matrix ratio $V_k^{\mathrm{H}} H_k^{\mathrm{H}} \widetilde{F}_k^{-1} H_k V_k$ to the outside of log-det. As a result, problem (11) is converted to

$$\underset{V, \Gamma}{\text{maximize}} \quad f_r(V, \Gamma) \tag{13a}$$

$$\text{subject to} \quad \Gamma_k \in \mathbb{S}_+, \ k = 1, \dots, K, \tag{13b}$$

where $\Gamma = \{\Gamma_k\}$ is a set of auxiliary variables, and the new objective function is given by

$$f_r(V, \Gamma) = \sum_{k=1}^{K} \omega_k \Big[\log |I + \Gamma_k| - \text{tr}(\Gamma_k)$$
$$+ \text{tr}\big((I + \Gamma_k) V_k^{\mathrm{H}} H_k^{\mathrm{H}} J_k^{-1} H_k V_k\big) \Big] \tag{14}$$

with

$$J_k = H_k V_k V_k^{\mathrm{H}} H_k^{\mathrm{H}} + \widetilde{F}_k. \tag{15}$$

When V is held fixed, Γ in problem (13) can be optimally determined as

$$\Gamma_k^{\star} = V_k^{\mathrm{H}} H_k^{\mathrm{H}} \widetilde{F}_k^{-1} H_k V_k. \tag{16}$$

It remains to optimize V when Γ is fixed. Note that problem (13) is a sum-of-matrix-ratios problem of V, for which the quadratic transform [26] can be readily applied. As a result, we further convert problem (13) to

$$\underset{V, \Gamma, Y}{\text{maximize}} \quad f_q(V, \Gamma, Y) \tag{17a}$$

$$\text{subject to} \quad \Gamma_k \in \mathbb{S}_+, \ k = 1, \dots, K \tag{17b}$$

$$\qquad\qquad Y_k \in \mathbb{C}^{N \times d}, \ k = 1, \dots, K, \tag{17c}$$

where $Y = \{Y_k\}$ is a set of auxiliary variables, and the new objective function is

$$f_q(V, \Gamma, Y) = \sum_{k=1}^{K} \omega_k \Big[\log |I + \Gamma_k| - \text{tr}(\Gamma_k) +$$
$$\text{tr}\big(2\Re\{V_k^{\mathrm{H}} H_k^{\mathrm{H}} Y_k (I + \Gamma_k)\} - Y_k^{\mathrm{H}} J_k Y_k (I + \Gamma_k)\big) \Big]. \tag{18}$$

When Γ and V are both held fixed, each Y_k can be optimally determined as

$$Y_k^{\star} = J_k^{-1} H_k V_k. \tag{19}$$

We remark that both (16) and (19) require inverting an $N \times N$ matrix. This shall not incur any computational difficulty since N is a small number (e.g., $N = 1$ or 2 in a typical massive MIMO system [39]).

The key step is to optimize the precoding variable V in (17) with Γ and Y both held fixed. Note that $f_q(V, \Gamma, Y)$ is a concave quadratic function in V, so the solution of each V_k can be obtained in closed-form by completing the square, i.e.,

$$V_k^{\star} = D^{-1} Q_k, \tag{20}$$

Algorithm 1 Unconstrained WMMSE Beamforming
1: Initialize V to any nonzero beamforming matrix.
2: **repeat**
3:      Update each Γ_k by (16).
4:      Update each Y_k by (19).
5:      Update each V_k by (20).
6: **until** convergence
7: $V \leftarrow \rho V$, where $\rho = (P / \sum_{k=1}^{K} \|V_k\|_F^2)^{1/2}$.

where

$$D = \sum_{j=1}^{K} \omega_j \Big[H_j^{\mathrm{H}} Y_j (I + \Gamma_j) Y_j^{\mathrm{H}} H_j$$
$$+ \frac{\sigma^2}{P} \text{tr}(Y_j^{\mathrm{H}} Y_j (I + \Gamma_j)) I \Big], \tag{21}$$

and

$$Q_k = \omega_k H_k^{\mathrm{H}} Y_k (I + \Gamma_k), \tag{22}$$

As shown in [26], updating Γ, Y, and V iteratively guarantees convergence to a stationary point of (17). We can subsequently recover a stationary point of the original problem (13) according to Theorem 1. Algorithm 1 summarizes the above steps; observe that it resembles the conventional WMMSE algorithm [3].

Despite the closed-form steps in Algorithm 1, the practical implementation of this algorithm can still be difficult because the optimal update of V_k in (20) requires computing the $M \times M$ matrix inversion D^{-1}, where M is a large number in our problem case. Thus, the real challenge is to optimize V in (17) for fixed (Γ, Y), which can be divided into a set of unconstrained quadratic program on a per V_k basis:

$$\underset{V_k}{\text{minimize}} \quad \text{tr}\Big(\frac{1}{2} V_k^{\mathrm{H}} D V_k - \Re\{V_k^{\mathrm{H}} Q_k\}\Big), \tag{23}$$

for $k = 1, 2, \dots, K$. The next section proposes a finite horizon optimization approach to the above quadratic program.

IV. FINITE HORIZON OPTIMIZATION

Denote by $g_k(V_k)$ the objective function in (23). To avoid the matrix inversion D^{-1}, we solve problem (23) by the gradient descent as

$$V_k^{t+1} = V_k^t - \eta_t \nabla g_k(V_k^t), \tag{24}$$

where V_k^t is the beamforming variable decision in the tth iteration, $\eta_t \geq 0$ is the step size, and the gradient can be computed as

$$\nabla g_k(V_k^t) = D V_k^t - Q_k. \tag{25}$$

In particular, to cut down the computation cost, we only allow the gradient update in (24) to run T times, i.e., $t = 0, 1, \dots, T - 1$ in (24). For the finite horizon optimization purpose, we seek the optimal step sizes η_t to maximize the performance of the ultimate solution V_k^T after T iterations:

$$\underset{\eta_0, \cdots, \eta_{T-1}}{\text{minimize}} \quad g_k(V_k^T) \tag{26a}$$

$$\text{subject to} \quad V_k^{t+1} = V_k^t - \eta_t \nabla g_k(V_k^t). \tag{26b}$$

Denote by $\boldsymbol{V}_k^\star$ the optimal solution to (23). The objective function in (26) can be rewritten as

$$g_k(\boldsymbol{V}_k) = \operatorname{tr}\Big(\frac{1}{2}(\boldsymbol{V}_k - \boldsymbol{V}_k^\star)^{\mathrm{H}} \boldsymbol{D}(\boldsymbol{V}_k - \boldsymbol{V}_k^\star)\Big) + C, \qquad (27)$$

where C is a constant term depending on $\boldsymbol{V}_k^\star$. Moreover, let λ_M be the largest eigenvalue of $\boldsymbol{D}$ and λ_1 the smallest eigenvalue. Since $\boldsymbol{D}$ is positive definite, we have $\lambda_M \geq \lambda_1 > 0$. It can be shown that

$$\frac{\lambda_1}{2}\|\boldsymbol{V}_k - \boldsymbol{V}_k^*\|_F^2 \leq g_k(\boldsymbol{V}_k) - C \leq \frac{\lambda_M}{2}\|\boldsymbol{V}_k - \boldsymbol{V}_k^*\|_F^2. \quad (28)$$

Based on the above bounds, problem (26) can be approximated as

$$\underset{\eta_0,\cdots,\eta_{T-1}}{\operatorname{minimize}} \quad \|\boldsymbol{V}_k^T - \boldsymbol{V}_k^*\|_F \qquad (29\mathrm{a})$$

$$\text{subject to} \quad \boldsymbol{V}_k^{t+1} = \boldsymbol{V}_k^t - \eta_t \nabla g_k(\boldsymbol{V}_k^t). \qquad (29\mathrm{b})$$

Intuitively, we aim to minimize the (Euclidean) distance between the ultimate solution $\boldsymbol{V}_k^T$ and the optimal solution $\boldsymbol{V}_k^\star$ after T iterations.

Furthermore, we bound the distance $\|\boldsymbol{V}_k^T - \boldsymbol{V}_k^*\|_F$ from above as

$$\begin{aligned}
&\|\boldsymbol{V}_k^T - \boldsymbol{V}_k^\star\|_F \\
&= \left\| (\boldsymbol{V}_k^0 - \boldsymbol{V}_k^\star) \prod_{t=0}^{T-1} (\boldsymbol{I} - \eta_t \boldsymbol{D}) \right\|_F \\
&\leq \left\| \prod_{t=0}^{T-1} (\boldsymbol{I} - \eta_t \boldsymbol{D}) \right\|_2 \times \|\boldsymbol{V}_k^0 - \boldsymbol{V}_k^*\|_F \\
&\leq \sup_{\lambda_1 \leq \lambda \leq \lambda_M} \left| \prod_{t=0}^{T-1} (1 - \eta_t \lambda) \right| \times \|\boldsymbol{V}_k^0 - \boldsymbol{V}_k^*\|_F. \qquad (30)
\end{aligned}$$

Substituting this upper bound in problem (29), we arrive at

$$\underset{\eta_0,\cdots,\eta_{T-1}}{\operatorname{minimize}} \quad \sup_{\lambda_1 \leq \lambda \leq \lambda_M} \left| \prod_{t=0}^{T-1} (1 - \eta_t \lambda) \right|. \qquad (31)$$

We shall decide the step sizes for the finite horizon optimization by solving the above problem.

Proposition 1: If we require all the step sizes to be identical, i.e., $\eta_0 = \eta_1 = \ldots = \eta_{T-1} = \eta$, then the optimal solution is

$$\eta = \frac{2}{\lambda_1 + \lambda_M}. \qquad (32)$$

As a result, we obtain from (30) that

$$\|\boldsymbol{V}_k^T - \boldsymbol{V}_k^\star\|_F \leq \left(1 - \frac{2}{\kappa+1}\right)^T \|\boldsymbol{V}_k^0 - \boldsymbol{V}_k^\star\|_F, \qquad (33)$$

where $\kappa = \lambda_M/\lambda_1$ is the condition number of matrix $\boldsymbol{D}$.

Next, we consider (31) with $(\eta_0,\cdots,\eta_{T-1})$ allowed to be different. It turns out that the problem is closely related to the degree-T *Chebyshev polynomial*:

$$\begin{aligned}
S(x) &= \cos(T \arccos x) \\
&= 2^{T-1}(x - \xi_0)(x - \xi_1)\cdots(x - \xi_{T-1}), \qquad (34)
\end{aligned}$$

where

$$\xi_t = \cos\left(\left(t + \frac{1}{2}\right)\frac{\pi}{T}\right), \ t = 0, 1, \ldots, T-1. \qquad (35)$$

For a constant $\gamma > 1$, we normalize $S(x)$ at the position γ as

$$\widetilde{S}(x;\gamma) = \frac{S(x)}{S(\gamma)}. \qquad (36)$$

Clearly, we have $\widetilde{S}(\gamma;\gamma) = 1$.

Now, an important result from [40]–[42] is that $\widetilde{S}(x;\gamma)$ has "smaller" fluctuations on the interval $[-1, 1]$ than any other degree-T polynomials that are normalized at γ, i.e.,

$$\sup_{-1 \leq x \leq 1} |\widetilde{S}(x;\gamma)| \leq \sup_{-1 \leq x \leq 1} |p(x)|, \quad \forall p(x) \in \mathcal{P}_\gamma, \qquad (37)$$

where

$$\mathcal{P}_\gamma = \{p(x) \in \mathbb{P}_T : p(\gamma) = 1\}. \qquad (38)$$

To utilize the above result, we need to rewrite the objective function of (31) in the form of the Chebyshev polynomial. Denote by $S_0(\lambda)$ the objective function in (31) as a polynomial of λ given $(\eta_0,\cdots,\eta_{T-1})$:

$$S_0(\lambda) = \prod_{t=0}^{T-1}(1 - \eta_t \lambda). \qquad (39)$$

Letting

$$x = \frac{\lambda_M + \lambda_1}{\lambda_M - \lambda_1} - \frac{2}{\lambda_M - \lambda_1}\lambda \qquad (40)$$

so that $-1 \leq x \leq 1$ when $\lambda_1 \leq \lambda \leq \lambda_M$, we rewrite $S_0(\lambda)$ as a polynomial of x:

$$S_1(x) = \prod_{t=0}^{T-1}\left(1 - \eta_t\left(\frac{\lambda_M + \lambda_1}{2} - \frac{\lambda_M - \lambda_1}{2}x\right)\right). \qquad (41)$$

Because of the one-to-one correspondence between x and λ, we have the problem equivalence:

$$\underset{\eta_0,\cdots,\eta_{T-1}}{\operatorname{minimize}} \sup_{\lambda_1 \leq \lambda \leq \lambda_M} |S_0(\lambda)| \iff \underset{\eta_0,\cdots,\eta_{T-1}}{\operatorname{minimize}} \sup_{-1 \leq x \leq 1} |S_1(x)|.$$

Observing that every possible $S_1(x)$ equals 1 when $x = \frac{\lambda_M + \lambda_1}{\lambda_M - \lambda_1}$, we further have

$$\underset{\eta_0,\cdots,\eta_{T-1}}{\operatorname{minimize}} \sup_{-1 \leq x \leq 1} |S_1(x)| \iff \underset{p(x) \in \mathcal{P}_\gamma}{\operatorname{minimize}} \sup_{-1 \leq x \leq 1} |p(x)|,$$

where

$$\gamma = \frac{\lambda_M + \lambda_1}{\lambda_M - \lambda_1}. \qquad (42)$$

According to (37), the optimal solution of $(\eta_0,\ldots,\eta_{T-1})$ is the one that makes

$$S_1(x) = \widetilde{S}\left(x; \frac{\lambda_M + \lambda_1}{\lambda_M - \lambda_1}\right). \qquad (43)$$

In other words, we should choose $(\eta_0,\ldots,\eta_{T-1})$ to let $S_1(x)S(\frac{\lambda_M + \lambda_1}{\lambda_M - \lambda_1})$ be the degree-T Chebyshev polynomial. This optimal solution is stated in the following proposition.

Proposition 2: An optimal solution to problem (31) is

$$\eta_t^\star = \left(\frac{\lambda_M + \lambda_1}{2} + \frac{\lambda_M - \lambda_1}{2}\xi_t\right)^{-1}, \qquad (44)$$

for $t = 0, 1, \ldots, T-1$. This optimal solution leads to

$$\|\boldsymbol{V}_k^T - \boldsymbol{V}_k^\star\|_F \leq 2\left(1 - \frac{2}{\sqrt{\kappa}+1}\right)^T \|\boldsymbol{V}_k^0 - \boldsymbol{V}_k^\star\|_F, \qquad (45)$$

Algorithm 2 Finite Horizon Beamforming for a Single Cell

1: Initialize $\boldsymbol{V}$ to any nonzero beamforming matrix.
2: **repeat**
3:   Update each $\boldsymbol{\Gamma}_k$ by (16).
4:   Update each $\boldsymbol{Y}_k$ by (19).
5:   Update each $\boldsymbol{V}_k$ by (24) for T iterations, with the step sizes determined as in (44).
6: **until** convergence
7: $\boldsymbol{V} \leftarrow \rho\boldsymbol{V}$, where $\rho = (P/\sum_{k=1}^{K}\|\boldsymbol{V}_k\|_F^2)^{1/2}$.

Algorithm 3 Finite Horizon Beamforming for Multiple Cells

1: Initialize $\boldsymbol{V}$ to any nonzero beamforming matrix.
2: **repeat**
3:   Update each $\boldsymbol{\Gamma}_{\ell k}$ by (55).
4:   Update each $\boldsymbol{Y}_{\ell k}$ by (56).
5:   Update each $\boldsymbol{V}_k$ by (57) for T iterations, with the step sizes determined as in (59).
6: **until** convergence
7: $\boldsymbol{V}_{\ell k} \leftarrow \rho_\ell\boldsymbol{V}_{\ell k}$, where $\rho_\ell = (P_\ell/\sum_{k=1}^{K}\|\boldsymbol{V}_{\ell k}\|_F^2)^{1/2}$.

where $\kappa = \lambda_M/\lambda_1$ is the condition number of matrix $\boldsymbol{D}$.

Remark 1: The convergence bound in (33) continues to hold when T is replaced by any other positive integer $t = 1, 2, \ldots$, whereas the convergence bound in (45) holds only for this particular T.

Remark 2: Problem (31) has multiple optimal solutions. Actually, since the variables $\eta_0, \eta_1, \ldots, \eta_{T-1}$ are symmetric in the objective function of (31), the reordering of $(\eta_0^\star, \eta_1^\star, \ldots, \eta_{T-1}^\star)$ in (44) still gives an optimal solution.

V. MULTI-CELL CASE

Thus far our attention is limited to the single-cell downlink network with only one BS. This section aims at a multi-cell extension. Assume that there are L BSs, each equipped with M transmit antennas. The rest setting follows that in the single-cell case, e.g., each cell has K user terminals, and each user terminal has N receive antennas. We use $\ell, i = 1, \ldots, L$ to index the cells and the corresponding BSs, and still use $k, j = 1, \ldots, K$ to index the users within each cell. Denote by $\boldsymbol{H}_{\ell k,i} \in \mathbb{C}^{N\times M}$ the channel from BS i to the kth user in cell ℓ, and $\boldsymbol{V}_{\ell k} \in \mathbb{C}^{M\times d}$ the transmit precoder of BS ℓ for its kth associated user. As a result, the achievable rate of the kth user in cell ℓ, written $R_{\ell k}$, is given by

$$R_{\ell k} = \log|\boldsymbol{I} + \boldsymbol{V}_{\ell k}^{\mathrm{H}}\boldsymbol{H}_{\ell k,\ell}^{\mathrm{H}}\boldsymbol{F}_{\ell k}^{-1}\boldsymbol{H}_{\ell k,\ell}\boldsymbol{V}_{\ell k}|, \quad (46)$$

where

$$\boldsymbol{F}_{\ell k} = \sigma^2\boldsymbol{I} + \sum_{(i,j)\neq(\ell,k)}\boldsymbol{H}_{\ell k,i}\boldsymbol{V}_{ij}\boldsymbol{V}_{ij}^{\mathrm{H}}\boldsymbol{H}_{\ell k,i}^{\mathrm{H}}. \quad (47)$$

The weighted sum-of-rates maximization problem across multiple cells is:

$$\underset{\boldsymbol{V}}{\text{maximize}} \quad \sum_{\ell=1}^{L}\sum_{k=1}^{K}\omega_{\ell k}R_{\ell k} \quad (48\text{a})$$

$$\text{subject to} \quad \sum_{k=1}^{K}\|\boldsymbol{V}_{\ell k}\|_F^2 \leq P, \ \ell = 1, \ldots, L. \quad (48\text{b})$$

Unlike the single-cell case, the power constraint in the above problem need not be tight at the optimum point. We propose

rewriting (48) as

$$\underset{\boldsymbol{V}}{\text{maximize}} \quad \sum_{\ell=1}^{L}\sum_{k=1}^{K}\omega_{\ell k}R_{\ell k} \quad (49\text{a})$$

$$\text{subject to} \quad \sum_{k=1}^{K}\|\boldsymbol{V}_{\ell k}\|_F^2 = P_\ell, \ \ell = 1, \ldots, L, \quad (49\text{b})$$

with the power variables $P_\ell \leq P$. In practice, we may optimize P_ℓ separately via the existing power control method for fixed $\boldsymbol{V}$.

Repeating the procedure from (8) to (11) in Section III, we recast problem (49) into an unconstrained problem:

$$\underset{\boldsymbol{V}}{\text{maximize}} \quad \sum_{\ell,k}\omega_{\ell k}\log\left|\boldsymbol{I} + \boldsymbol{V}_{\ell k}^{\mathrm{H}}\boldsymbol{H}_{\ell k,\ell}^{\mathrm{H}}\widetilde{\boldsymbol{F}}_{\ell k}^{-1}\boldsymbol{H}_{\ell k,\ell}\boldsymbol{V}_{\ell k}\right| \quad (50)$$

where

$$\widetilde{\boldsymbol{F}}_{\ell k} = \left(\frac{\sigma^2}{P_\ell}\sum_{k=1}^{K}\|\boldsymbol{V}_{\ell k}\|_F^2\right)\boldsymbol{I} + \sum_{(i,j)\neq(\ell,k)}\boldsymbol{H}_{\ell k,i}\boldsymbol{V}_{ij}\boldsymbol{V}_{ij}^{\mathrm{H}}\boldsymbol{H}_{\ell k,i}^{\mathrm{H}}. \quad (51)$$

Again, by the matrix FP technique, we obtain the further reformulation as an unconstrained quadratic program:

$$\underset{\boldsymbol{V},\boldsymbol{\Gamma},\boldsymbol{Y}}{\text{maximize}} \quad f_q^{\mathrm{MC}}(\boldsymbol{V},\boldsymbol{\Gamma},\boldsymbol{Y}) \quad (52\text{a})$$

$$\text{subject to} \quad \boldsymbol{\Gamma}_{\ell k} \in \mathbb{S}_+, \ \forall(\ell,k) \quad (52\text{b})$$

$$\boldsymbol{Y}_{\ell k} \in \mathbb{C}^{N\times d}, \ \forall(\ell,k), \quad (52\text{c})$$

where the new objective function is shown in (53) along with

$$\boldsymbol{J}_{\ell k} = \boldsymbol{H}_{\ell k,\ell}\boldsymbol{V}_{\ell k}\boldsymbol{V}_{\ell k}^{\mathrm{H}}\boldsymbol{H}_{\ell k,\ell}^{\mathrm{H}} + \widetilde{\boldsymbol{F}}_{\ell k}. \quad (54)$$

Similar to the single-cell case, the two auxiliary variables are optimally determined as

$$\boldsymbol{\Gamma}_{\ell k}^\star = \boldsymbol{V}_{\ell k}^{\mathrm{H}}\boldsymbol{H}_{\ell k,\ell}^{\mathrm{H}}\widetilde{\boldsymbol{F}}_{\ell k}^{-1}\boldsymbol{H}_{\ell k,\ell}\boldsymbol{V}_{\ell k}, \quad (55)$$

$$\boldsymbol{Y}_{\ell k}^\star = \boldsymbol{J}_{\ell k}^{-1}\boldsymbol{H}_{\ell k,\ell}\boldsymbol{V}_{\ell k}. \quad (56)$$

For fixed auxiliary variables, we update the beamforming variable by gradient descent up to T iterations:

$$\boldsymbol{V}_{\ell k}^{t+1} = \boldsymbol{V}_{\ell k}^t - \eta_{\ell,t}\nabla g_{\ell k}(\boldsymbol{V}_k^t), \quad (57)$$

$$f_q^{\mathrm{MC}}(\boldsymbol{V},\boldsymbol{\Gamma},\boldsymbol{Y}) = \sum_{\ell=1}^{L}\sum_{k=1}^{K}\omega_{\ell k}\left[\log|\boldsymbol{I}+\boldsymbol{\Gamma}_{\ell k}| - \mathrm{tr}(\boldsymbol{\Gamma}_{\ell k}) + \mathrm{tr}\left(2\Re\{\boldsymbol{V}_{\ell k}^{\mathrm{H}}\boldsymbol{H}_{\ell k,\ell}^{\mathrm{H}}\boldsymbol{Y}_{\ell k}(\boldsymbol{I}+\boldsymbol{\Gamma}_{\ell k})\} - \boldsymbol{Y}_{\ell k}^{\mathrm{H}}\boldsymbol{J}_{\ell k}\boldsymbol{Y}_{\ell k}(\boldsymbol{I}+\boldsymbol{\Gamma}_{\ell k})\right)\right] \quad (53)$$

where

$$\nabla g_{\ell k}(\boldsymbol{V}_{\ell k}) = \boldsymbol{D}_\ell \boldsymbol{V}_{\ell k} - \omega_{\ell k} \boldsymbol{H}_{\ell k,\ell}^{\mathrm{H}} \boldsymbol{Y}_{\ell k}(\boldsymbol{I} + \boldsymbol{\Gamma}_{\ell k}) \qquad (58)$$

Now, the step sizes $(\eta_{\ell,0}, \eta_{\ell,1}, \ldots, \eta_{\ell,T-1})$ are determined for each cell ℓ individually as in Proposition 2:

$$\eta_{\ell,t} = \left(\frac{\lambda_{\ell M} + \lambda_{\ell 1}}{2} + \frac{\lambda_{\ell M} - \lambda_{\ell 1}}{2} \xi_t \right)^{-1}, \qquad (59)$$

only that the largest eigenvalue $\lambda_{\ell M}$ and the smallest eigenvalue $\lambda_{\ell 1}$ are now associated with the cell-specific matrix

$$\boldsymbol{D}_\ell = \sum_{i=1}^{L} \sum_{j=1}^{K} \omega_{ij} \boldsymbol{H}_{ij,\ell}^{\mathrm{H}} \boldsymbol{Y}_{ij}(\boldsymbol{I} + \boldsymbol{\Gamma}_{ij}) \boldsymbol{Y}_{ij}^{\mathrm{H}} \boldsymbol{H}_{ij,\ell}$$

$$+ \sum_{j=1}^{K} \frac{\omega_{\ell j} \sigma^2}{P_\ell} \mathrm{tr}\left(\boldsymbol{Y}_{\ell j}^{\mathrm{H}} \boldsymbol{Y}_{\ell j}(\boldsymbol{I} + \boldsymbol{\Gamma}_{\ell j}) \right) \boldsymbol{I}. \quad (60)$$

Algorithm 3 summarizes the above steps for the extended finite horizon beamforming for multiple cells.

VI. SIMULATION RESULTS

We first validate the performance of the proposed algorithms numerically in a single-cell large-scale MIMO network. Within the cell, the BS is located at the center, and 6 downlink users are randomly distributed. Let the BS have 2048 transmit antennas, and let each downlink user have 8 receive antennas, 8 independent data streams being intended for each downlink receiver. The maximum transmit power P equals 20 dBm. The background noise level σ^2 equals -80 dBm. The rate weights ω_k are all set to 1. The downlink distance-dependent path-loss is modeled as $15.3 + 37.6 \log_{10}(l) + \Delta$ (in dB), where l represents the distance in meters, and Δ is a zero-mean Gaussian random variable with a standard variance of 8 dB—which models the shadowing effect. The parameter T is fixed at 5 for Algorithm 2 by default. Aside from Algorithm 1 and Algorithm 2, we consider the conventional gradient descent with equal step size as in (32); we let the conventional gradient descent run T iterations as well for fairness.

Fig. 2 and Fig. 3 compare the convergences of the different beamforming algorithms in iterations and in CPU time, respectively. Observe from Fig. 2 that Algorithm 1, i.e., the modified WMMSE algorithm without need to tune the Lagrangian multiplier for the power constraint, converges faster than the proposed Algorithm 2 and conventional gradient descent algorithm in iterations. This is because Algorithm 1 finds the exact global optimum of the quadratic program in (23) every after the updates of the auxiliary variables. From an MM perspective, it gives a tighter approximation of the original objective function (7). However, this is at a high cost of computing a 2048×2048 matrix inversion for each iteration. As a result, it can be seen from Fig. 3 that the proposed Algorithm 2 converges faster than Algorithm 1 in CPU time, since the per-iteration complexity of the former is much lower than that of the latter. For instance, Algorithm 1 requires about 2.6 seconds to reach 100 bps/Hz, while the proposed Algorithm 2 merely requires 1.6 seconds to attain the same performance.
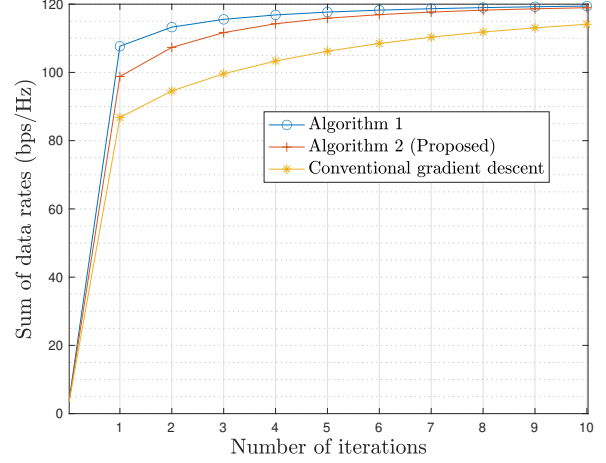

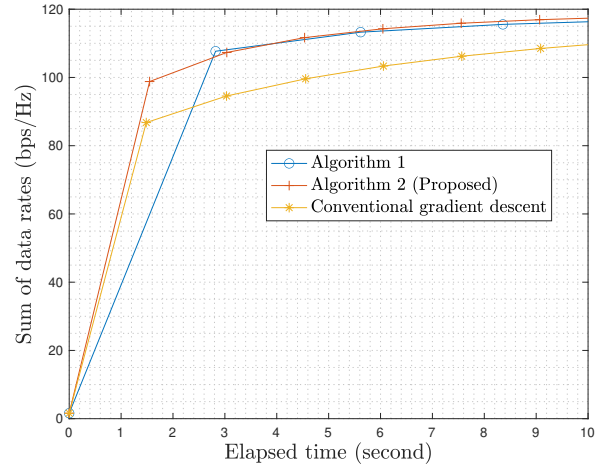Fig. 2. Sum rates vs. number of iterations in the single-cell case.

Fig. 3. Sum rates vs. CPU time in the single-cell case.

Furthermore, Fig. 4 focuses on how the different algorithms tackle the quadratic program in (23) iteratively when the auxiliary variables $\boldsymbol{\Gamma}_k$ and $\boldsymbol{Y}_k$ are both held fixed. Note that Algorithm 1 is not included in the figure because it solves the quadratic program (23) directly. We try out different values of T for Algorithm 2, i.e., $T \in \{3, 5, 7\}$. According to Fig. 4, Algorithm 2 converges more quickly when T is smaller, but its ultimate performance after T iterations becomes worse. Observe also that Algorithm 2 need not be always faster than the conventional gradient descent, but when it terminates at the Tth iteration, it must outperform the conventional gradient descent. Further, we remark that Algorithm 2 is already quite close to the global optimum when $T = 7$, so the quadratic program in (23) can be well addressed without computing the large matrix inversion.

Next, Fig. 5 shows how the different algorithms behave when the number of transmit antennas M is changed. For fair comparison, the same amount of CPU time is given to all the algorithms. Clearly, Algorithm 1, which is the modified WMMSE algorithm, is sensitive to the increase of the number
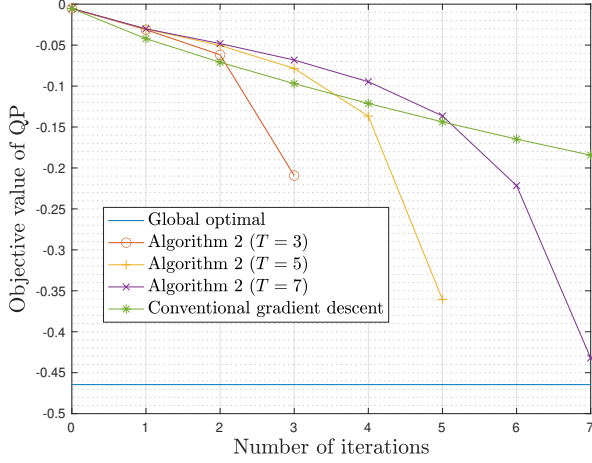
Fig. 4. Objective value in (23) vs. number of iterations in the single-cell case.
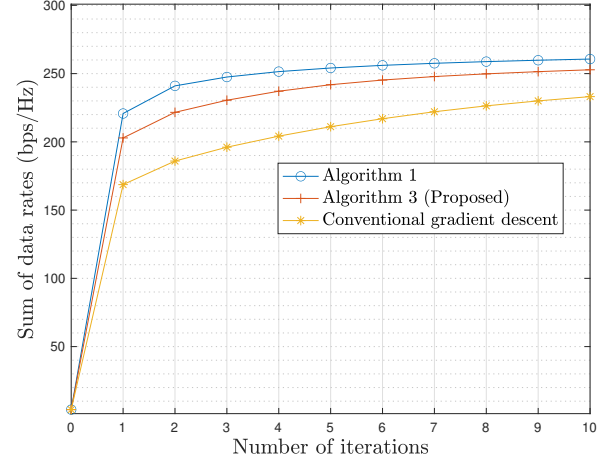

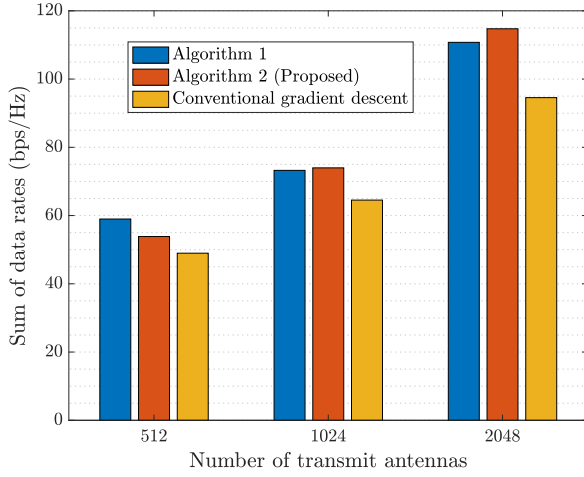Fig. 6. Sum rates vs. number of iterations in the multi-cell case.

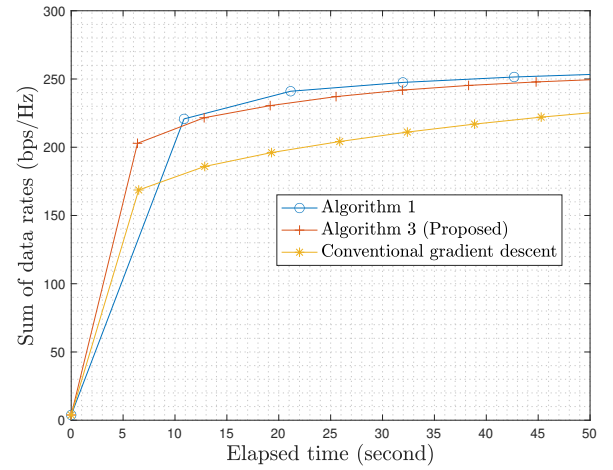Fig. 5. Sum rates vs. number of transmit antennas in the single-cell case.

Fig. 7. Sum rates vs. CPU time in the multi-cell case.

of antennas. While Algorithm 1 outperforms Algorithm 2 when $M = 512$, it is surpassed when T increases to 1024. The benefit of eliminating the large matrix inversion now starts to pay off. The advantage of Algorithm 2 over Algorithm 1 becomes even larger when M further increases to 2048.

We now consider the multi-cell case. The different algorithms are tested in a 3-cell large-scale MIMO network, with the BS-to-BS distance set to 800 meters. The rest settings (e.g., the number of transmit antennas of each BS) follows the previous single-cell case. We do not consider power control in this case, simply letting each $P_\ell = P$.

Fig. 6 and Fig. 7 repeat the convergence tests for the multi-cell network. The results do not differ much from the previous single-cell case. Again, the modified WMMSE of Algorithm 1 requires the fewest iterations to converge according to Fig. 6, and yet is much slower than the proposed Algorithm 2 according to Fig. 7 when it comes to CPU time, because of its high computation cost in large matrix inversion. Actually, Algorithm 1 is even slower than the conventional gradient descent as shown in Fig. 7. For instance, Algorithm 1 requires

almost twice as much time as the other two algorithms to finish the first iteration.

VII. CONCLUSION

This paper proposes a novel finite-horizon approach to the beamforming problem for the large-scale MIMO. We first convert the constrained log-det problem to an unconstrained quadratic program by the matrix FP. Optimizing the beamforming variables in the new problem iteratively by the gradient descent can avoid the costly matrix inversion. Our key contribution is to incorporate finite horizon optimization into the gradient descent to enhance efficiency. This is the very first application case of finite horizon optimization in the area of signal processing and digital communications.

REFERENCES

[1] Y.-F. Liu, T.-H. Chang, M. Hong, Z. Wu, A. M.-C. So, E. A. Jorswieck, and W. Yu, "A survey of recent advances in optimization methods for wireless communications," *IEEE J. Sel. Areas Commun.*, Sep. 2024.

[2] S. S. Christensen, R. Agarwal, E. De Carvalho, and J. 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.

[3] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. 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, Apr. 2011.

[4] W. Yu and T. Lan, "Transmitter optimization for the multi-antenna downlink with per-antenna power constraints," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2646–2660, May 2007.

[5] B. L. Ng, J. S. Evans, S. V. Hanly, and D. Aktas, "Distributed downlink beamforming with cooperative base stations," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5491–5499, Nov. 2008.

[6] F. Rashid-Farrokhi, K. R. Liu, and L. Tassiulas, "Transmit beamforming and power control for cellular wireless systems," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1437–1450, Aug. 1998.

[7] R. Zakhour and S. V. Hanly, "Min-max power allocation in cellular networks with coordinated beamforming," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 287–302, Jan. 2013.

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

[9] T.-H. Chang, Z.-Q. Luo, and C.-Y. Chi, "Approximation bounds for semidefinite relaxation of max-min-fair multicast transmit beamforming problem," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3932–3943, Jul. 2008.

[10] Z. Zhang, K. C. Teh, and K. H. Li, "A semidefinite relaxation approach for beamforming in cooperative clustered multicell systems with novel limited feedback scheme," *IEEE Trans. Veh. Technol.*, vol. 63, no. 4, pp. 1740–1748, Oct. 2013.

[11] M. Ashraphijuo, X. Wang, and M. Tao, "Multicast beamforming design in multicell networks with successive group decoding," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 3492–3506, Mar. 2017.

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

[13] R. Cendrillon, W. Yu, M. Moonen, J. Verlinden, and T. Bostoen, "Optimal multiuser spectrum balancing for digital subscriber lines," *IEEE Trans. Commun.*, vol. 54, no. 5, pp. 922–933, May 2006.

[14] K. Cumanan, R. Krishna, L. Musavian, and S. Lambotharan, "Joint beamforming and user maximization techniques for cognitive radio networks based on branch and bound method," *IEEE Trans. Wireless Commun.*, vol. 9, no. 10, pp. 3082–3092, Aug. 2010.

[15] L. P. Qian, Y. J. Zhang, and J. Huang, "MAPEL: Achieving global optimality for a non-convex wireless power control problem," *IEEE Trans. Wireless Commun.*, vol. 8, no. 3, pp. 1553–1563, Mar. 2009.

[16] S. Boyd, S.-J. Kim, L. Vandenberghe, and A. Hassibi, "A tutorial on geometric programming," *Optim. Eng.*, vol. 8, pp. 67–127, Apr. 2007.

[17] M. Chiang, "Geometric programming for communication systems," *Found. Trends Commun. Inf. Theory*, vol. 2, no. 1–2, pp. 1–154, Jul. 2005.

[18] S.-J. Kim and G. B. Giannakis, "Optimal resource allocation for MIMO ad hoc cognitive radio networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3117–3131, Apr. 2011.

[19] S. Gong, C. Xing, V. K. Lau, S. Chen, and L. Hanzo, "Majorization-minimization aided hybrid transceivers for MIMO interference channels," *IEEE Trans. Signal Process.*, vol. 68, pp. 4903–4918, Aug. 2020.

[20] A. Goldsmith, *Wireless communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[21] L. Sun and M. R. McKay, "Eigen-based transceivers for the MIMO broadcast channel with semi-orthogonal user selection," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5246–5261, Jun. 2010.

[22] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the k-user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Jul. 2008.

[23] M. A. Maddah-Ali, A. S. Motahari, and A. K. Khandani, "Communication over MIMO X channels: Interference alignment, decomposition, and performance analysis," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3457–3470, Aug. 2008.

[24] M. Razaviyayn, G. Lyubeznik, and Z.-Q. Luo, "On the degrees of freedom achievable through interference alignment in a MIMO interference channel," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 812–821, Oct. 2011.

[25] G. Sridharan, S. Liu, and W. Yu, "Role of interference alignment in wireless cellular network optimization," *IEEE Trans. Commun.*, vol. 64, no. 12, pp. 5319–5333, Sep. 2016.

[26] K. Shen and W. Yu, "Fractional programming for communication systems—Part I: Power control and beamforming," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2616–2630, Mar. 2018.

[27] K. Shen and W. Yu, "Fractional programming for communication systems—Part II: Uplink scheduling via matching," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2631–2644, Mar. 2018.

[28] K. Shen, W. Yu, L. Zhao, and D. P. Palomar, "Optimization of MIMO device-to-device networks via matrix fractional programming: A minorization–maximization approach," *IEEE/ACM Trans. Netw.*, vol. 27, no. 5, pp. 2164–2177, Oct. 2019.

[29] K. Shen, Z. Zhao, Y. Chen, Z. Zhang, and H. V. Cheng, "Accelerating quadratic transform and WMMSE," *IEEE J. Sel. Areas Commun.*, Mar. 2024.

[30] E. De Carvalho, A. Ali, A. Amiri, M. Angjelichinoski, and R. W. Heath, "Non-stationarities in extra-large-scale massive MIMO," *IEEE Wireless Commun.*, vol. 27, no. 4, pp. 74–80, Aug. 2020.

[31] J. H. I. de Souza, A. Amiri, T. Abrão, E. De Carvalho, and P. Popovski, "Quasi-distributed antenna selection for spectral efficiency maximization in subarray switching XL-MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 70, no. 7, pp. 6713–6725, May 2021.

[32] X. Zhao, S. Lu, Q. Shi, and Z.-Q. 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.

[33] Z. Zhang, Z. Zhao, and K. Shen, "Enhancing the efficiency of WMMSE and FP for beamforming by Minorization-Maximization," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Jun. 2023, pp. 1–5.

[34] Y. Nesterov, "A method for solving the convex programming problem with convergence rate o(1/k2)," in *Dokl akad nauk Sssr*, vol. 269, Feb. 1983, p. 543.

[35] Y. Zhang, D. Rybin, and Z.-Q. Luo, "Finite horizon optimization: Framework and applications," 2024, [Online]. Available: https://arxiv.org/abs/2412.21068.

[36] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proc. Int. Conf. Mach. Learn. (ICLR)*, Jun. 2010, pp. 399–406.

[37] J. Sun, H. Li, Z. Xu *et al.*, "Deep ADMM-Net for compressive sensing MRI," *Adv. Neural Inf. Process. Syst.*, vol. 29, Dec. 2016.

[38] D. F. Shanno, "On the convergence of a new conjugate gradient algorithm," *SIAM J. Numer. Anal.*, vol. 15, no. 6, pp. 1247–1257, Dec. 1978.

[39] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Dec. 2012.

[40] W. Markoff and J. Grossmann, "Über polynome, die in einem gegebenen intervalle möglichst wenig von null abweichen," *Math. Ann.*, vol. 77, no. 2, pp. 213–258, Jun. 1916.

[41] D. A. Flanders and G. Shortley, "Numerical determination of fundamental modes," *J. Appl. Phys.*, vol. 21, no. 12, pp. 1326–1332, Dec. 1950.

[42] D. Young, "On Richardson's method for solving linear systems with positive definite matrices," *J. Math. Phys.*, vol. 32, no. 1-4, pp. 243–255, Apr. 1953.