APPENDIX A
PROOF OF THEOREM 1

Let the noise variance be non-zero, i.e., $\sigma^2 > 0$, the transmit power be finite, i.e., $P_n^{\max} < \infty$, and the channel realizations be bounded. Then, the SINR expressions (13) and (14) are finite, i.e., $\gamma_{g,k}^p < \infty$, $\forall g \in \mathcal{G}$, and $\gamma_{g,k}^c < \infty$, $\forall k \in \mathcal{M}_g, \forall g \in \mathcal{G}$. The limit of average spectral efficiency exists, when the sample size M tends to infinity. Let $\mathcal{W}$ be the feasible set of beamforming vectors determined by constraint (19c) . All previous assumptions make sure, that $\mathcal{W}$ is compact and not empty. Now, considering the ergodicity assumption and the law of large numbers, we make the following statement [1, Theorem 7.48]

$$\sup_{\mathbf{w}\in\mathcal{W}} \left| \frac{1}{M} \log_2(1+\gamma_{g,k}^o) - \mathbb{E}_{\mathbf{h}}\{\log_2(1+\gamma_{g,k}^o)\} \right| \to 0, \quad \text{as } M \to \infty, \quad o \in \{p, c\}. \tag{43}$$

Thus, given unlimited sample size, the SAA estimate of the rates converges to the ergodic rate uniformly on the compact set $\mathcal{W}$ with probability one. Therefore, the set of optimal solutions of problem (23) converges uniformly to the optimal solution set of problem (22) [1, Theorem 5.3]. This completes the proof.

APPENDIX B
PROOF OF THEOREM 1

The steps in this proof can be seen analog to [2, Theorem 2]. Let $G_g(\mathbf{w}_g^p, \mathbf{w}_g^c, \boldsymbol{\rho}_g^p, \boldsymbol{\rho}_g^c, \mathbf{u}_g^p, \mathbf{u}_g^c) = G_g^p(\mathbf{w}_g^p, \boldsymbol{\rho}_g^p, \mathbf{u}_g^p) + G_g^c(\mathbf{w}_g^c, \boldsymbol{\rho}_g^c, \mathbf{u}_g^c)$, where we define the following two functions

$$G_g^p(\mathbf{w}_g^p, \boldsymbol{\rho}_g^p, \mathbf{u}_g^p) = \frac{B}{M\log(2)} \sum_{m=1}^{M} \max_{u_{g,k}^p(m), \rho_{g,k}^p(m)} \left(\log(\rho_{g,k}^p(m)) - \rho_{g,k}^p(m) e_{g,k}^p(m) + 1 \right), \tag{44}$$

$$G_g^c(\mathbf{w}_g^c, \boldsymbol{\rho}_g^c, \mathbf{u}_g^c) = \frac{B}{M\log(2)} \sum_{m=1}^{M} \min_{k\in\mathcal{M}_g} \left(\max_{u_{g,k}^c(m), \rho_{g,k}^c(m)} \left(\log(\rho_{g,k}^c(m)) - \rho_{g,k}^c(m) e_{g,k}^c(m) + 1 \right) \right). \tag{45}$$

Using these definitions, we formulate following problem

$$\max_{\mathcal{V}_3} \quad \sum_{g\in\mathcal{G}} G_g(\mathbf{w}_g^p, \mathbf{w}_g^c, \boldsymbol{\rho}_g^p, \boldsymbol{\rho}_g^c, \mathbf{u}_g^p, \mathbf{u}_g^c) \tag{46a}$$

$$\text{s.t.} \quad (19\text{c}),$$

$$\sum_{g\in\mathcal{G}_n^p} G_g^p(\mathbf{w}_g^p, \mathbf{u}_g^p, \boldsymbol{\rho}_g^p) + \sum_{g\in\mathcal{G}_n^c} G_g^c(\mathbf{w}_g^c, \mathbf{u}_g^c, \boldsymbol{\rho}_g^c) \le F_n, \qquad \forall n \in \mathcal{N}. \tag{46b}$$

Here $\mathcal{V}_3 \triangleq \left\{\mathbf{w}_g^p, \mathbf{w}_g^c, \boldsymbol{\rho}_g^p, \boldsymbol{\rho}_g^c, \mathbf{u}_g^p, \mathbf{u}_g^c | \ \forall g \in \mathcal{G}\right\}$ is the set of optimization variables. Note, that problem (34) is the epigraph form of problem (46). According to [3, Chapter 4], problems (34) and (46) and their respective optimal solutions are equivalent. Therefore, we are able to use problem (46) as an equivalent formulation of problem (34) throughout this proof for simplicity reasons.

As Algorithm 2 is a block coordinate ascent algorithm operating iteratively, in iteration ν, we solve the following convex optimization problem

$$\max_{\mathcal{V}_4} \quad \sum_{g\in\mathcal{G}} G_g(\mathbf{w}_g^p, \mathbf{w}_g^c, (\boldsymbol{\rho}_g^p)^\nu, (\boldsymbol{\rho}_g^c)^\nu, (\mathbf{u}_g^p)^\nu, (\mathbf{u}_g^c)^\nu) \tag{47a}$$

$$\text{s.t.} \quad \textbf{(??)},$$

$$\sum_{g\in\mathcal{G}_n^p} G_g^p(\mathbf{w}_g^p, (\mathbf{u}_g^p)^\nu, (\boldsymbol{\rho}_g^p)^\nu) + \sum_{g\in\mathcal{G}_n^c} G_g^c(\mathbf{w}_g^c, (\mathbf{u}_g^c)^\nu, (\boldsymbol{\rho}_g^c)^\nu) \leq F_n, \qquad \forall n \in \mathcal{N}, \tag{47b}$$

where $\mathcal{V}_4 \triangleq \left\{\mathbf{w}_g^p, \mathbf{w}_g^c | \ \forall g \in \mathcal{G}\right\}$. Note that the fixed values, i.e., $(\boldsymbol{\rho}_g^p)^\nu, (\boldsymbol{\rho}_g^c)^\nu, (\mathbf{u}_g^p)^\nu, (\mathbf{u}_g^c)^\nu$, are computed by $(\rho_{g,k}^p)^\nu = 1/e_{g,k,\text{mmse}}^p$, $(\rho_{i,k}^c)^\nu = 1/e_{i,k,\text{mmse}}^c$, (28), and (29), using the optimal beamforming vectors computed in iteration $(\nu-1)$. We define the objective function of problem (46) as $Q(\mathbf{w}_g^p, \mathbf{w}_g^c, \boldsymbol{\rho}_g^p, \boldsymbol{\rho}_g^c, \mathbf{u}_g^p, \mathbf{u}_g^c)$. Since Q is a concave function and the achievable ergodic rates are bounded, the sequence $\{Q((\mathbf{w}_g^p)^\nu, (\mathbf{w}_g^c)^\nu, (\boldsymbol{\rho}_g^p)^\nu, (\boldsymbol{\rho}_g^c)^\nu, (\mathbf{u}_g^p)^\nu, (\mathbf{u}_g^c)^\nu)\}_{\nu=0}^\infty$ increases monotonically after each iteration and converges to the limit point $\bar{Q}$. Since the feasible set defined by constraints (19c) and (47b) is compact, $\{(\mathbf{w}_g^p)^\nu, (\mathbf{w}_g^c)^\nu\}_{\nu=0}^\infty$ must have a cluster point $\{\bar{\mathbf{w}}_g^p, \bar{\mathbf{w}}_g^c\}$. Meaning, it exists a subsequence $\{(\mathbf{w}_g^p)^{\nu_1}, (\mathbf{w}_g^c)^{\nu_1}\}_{\nu_1=\Lambda}^\infty$ for $\Lambda > 0$ that converges to $\{\bar{\mathbf{w}}_g^p, \bar{\mathbf{w}}_g^c\}$. Thus, the following statement holds

$$\lim_{\nu_1\to\infty}\{(\mathbf{w}_g^p)^{\nu_1}, (\mathbf{w}_g^c)^{\nu_1}, (\boldsymbol{\rho}_g^p)^{\nu_1}, (\boldsymbol{\rho}_g^c)^{\nu_1}, (\mathbf{u}_g^p)^{\nu_1}, (\mathbf{u}_g^c)^{\nu_1}\} = \{\bar{\mathbf{w}}_g^p, \bar{\mathbf{w}}_g^c, \bar{\boldsymbol{\rho}}_g^p, \bar{\boldsymbol{\rho}}_g^c, \bar{\mathbf{u}}_g^p, \bar{\mathbf{u}}_g^c\}. \tag{48}$$

Note that $\bar{\boldsymbol{\rho}}_g^p, \bar{\boldsymbol{\rho}}_g^c, \bar{\mathbf{u}}_g^p$, and $\bar{\mathbf{u}}_g^c$ are computed based on the beamforming vectors using the continuous functions (28), (29), $\rho_{g,k}^p = 1/e_{g,k,\text{mmse}}^p$, and $\rho_{i,k}^c = 1/e_{i,k,\text{mmse}}^c$. At this point, we have shown that $\{\bar{\boldsymbol{\rho}}_g^p, \bar{\boldsymbol{\rho}}_g^c, \bar{\mathbf{u}}_g^p, \bar{\mathbf{u}}_g^c\}$ are optimal when $\{\mathbf{w}_g^p, \mathbf{w}_g^c\} = \{\bar{\mathbf{w}}_g^p, \bar{\mathbf{w}}_g^c\}$. Following up, we show that the same applies vice versa, i.e., $\{\bar{\mathbf{w}}_g^p, \bar{\mathbf{w}}_g^c\}$ is optimal when $\{\boldsymbol{\rho}_g^p, \boldsymbol{\rho}_g^c, \mathbf{u}_g^p, \mathbf{u}_g^c\} = \{\bar{\boldsymbol{\rho}}_g^p, \bar{\boldsymbol{\rho}}_g^c, \bar{\mathbf{u}}_g^p, \bar{\mathbf{u}}_g^c\}$. With the optimal beamforming vectors from the previous iteration, and monotonicity of the objective function we write

$$
\begin{aligned}
Q((\mathbf{w}_g^p)^{\nu_1+1}, &(\mathbf{w}_g^c)^{\nu_1+1}, (\boldsymbol{\rho}_g^p)^{\nu_1+1}, (\boldsymbol{\rho}_g^c)^{\nu_1+1}, (\mathbf{u}_g^p)^{\nu_1+1}, (\mathbf{u}_g^c)^{\nu_1+1}) \\
&\geq Q((\mathbf{w}_g^p)^{\nu_1+1}, (\mathbf{w}_g^c)^{\nu_1+1}, (\boldsymbol{\rho}_g^p)^{\nu_1}, (\boldsymbol{\rho}_g^c)^{\nu_1}, (\mathbf{u}_g^p)^{\nu_1}, (\mathbf{u}_g^c)^{\nu_1}) \\
&\geq Q(\mathbf{w}_g^p, \mathbf{w}_g^c, (\boldsymbol{\rho}_g^p)^{\nu_1}, (\boldsymbol{\rho}_g^c)^{\nu_1}, (\mathbf{u}_g^p)^{\nu_1}, (\mathbf{u}_g^c)^{\nu_1}), \qquad \forall \mathbf{w}_g^p, \forall \mathbf{w}_g^c. \qquad (49)
\end{aligned}
$$

Taking the limit in this equation, we obtain the following relation

$$
\bar{Q} = Q(\bar{\mathbf{w}}_g^p, \bar{\mathbf{w}}_g^c, \bar{\boldsymbol{\rho}}_g^p, \bar{\boldsymbol{\rho}}_g^c, \bar{\mathbf{u}}_g^p, \bar{\mathbf{u}}_g^c) \geq Q(\mathbf{w}_g^p, \mathbf{w}_g^c, \bar{\boldsymbol{\rho}}_g^p, \bar{\boldsymbol{\rho}}_g^c, \bar{\mathbf{u}}_g^p, \bar{\mathbf{u}}_g^c), \qquad \forall \mathbf{w}_g^p, \forall \mathbf{w}_g^c. \qquad (50)
$$

Therefore, $\{\bar{\mathbf{w}}_g^p, \bar{\mathbf{w}}_g^c\}$ are the optimal beamforming vectors of problem (46). We have now shown that $\{\bar{\mathbf{w}}_g^p, \bar{\mathbf{w}}_g^c\}$ is optimal when $\{\boldsymbol{\rho}_g^p, \boldsymbol{\rho}_g^c, \mathbf{u}_g^p, \mathbf{u}_g^c\} = \{\bar{\boldsymbol{\rho}}_g^p, \bar{\boldsymbol{\rho}}_g^c, \bar{\mathbf{u}}_g^p, \bar{\mathbf{u}}_g^c\}$. At last, it can be shown that $\{\bar{\mathbf{w}}_g^p, \bar{\mathbf{w}}_g^c, \bar{\boldsymbol{\rho}}_g^p, \bar{\boldsymbol{\rho}}_g^c, \bar{\mathbf{u}}_g^p, \bar{\mathbf{u}}_g^c\}$ is a KKT solution to problem (46) by checking the KKT conditions. At this point, we have shown the solution set generated by Algorithm 2 converges to a KKT solution of problem (46). KKT points are not necessarily unique, however, any sequence $\{(\mathbf{w}_g^p)^{\nu}, (\mathbf{w}_g^c)^{\nu}, (\boldsymbol{\rho}_g^p)^{\nu}, (\boldsymbol{\rho}_g^c)^{\nu}, (\mathbf{u}_g^p)^{\nu}, (\mathbf{u}_g^c)^{\nu}\}_{\nu=0}^{\infty}$ converges to the KKT solution in the limit. This proof relates on the equivalence of problem (34) and problem (46). Therefore, we conclude that $\{\bar{\mathbf{w}}_g^p, \bar{\mathbf{w}}_g^c, \bar{\boldsymbol{\rho}}_g^p, \bar{\boldsymbol{\rho}}_g^c, \bar{\mathbf{u}}_g^p, \bar{\mathbf{u}}_g^c, \bar{R}_g^p, \bar{R}_g^c\}$ is also a KKT solution to problem (46), where we have $Q(\mathbf{w}_g^p, \mathbf{w}_g^c, \boldsymbol{\rho}_g^p, \boldsymbol{\rho}_g^c, \mathbf{u}_g^p, \mathbf{u}_g^c) = \sum_{g \in \mathcal{G}} (\bar{R}_g^p + \bar{R}_g^c)$. This completes the proof.

References

[1] D. D. A. Shapiro and A. P. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory.* PA, USA: SIAM, 2009.

[2] M. Razaviyayn, M. Hong, and Z. Luo, "Linear transceiver design for a MIMO interfering broadcast channel achieving max-min fairness," in *2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, 2011, pp. 1309–1313.

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