Appendix A
Convergence Proof for Centralized Algorithm

To prove the convergence of the transmit, receive beamformers, and the objective values of the centralized algorithms in (16) and (26), we need to show that the following conditions are satisfied by the centralized formulations.

(a) The function should be bounded below
(b) The feasible set should be a compact set
(c) The sequence of the objective values should be decreasing in each iteration
(d) Uniqueness of the minimizer, *i.e*, the transmit and the receive beamformers should be unique in each iteration.

The conditions (a) (b) and (c) are required for the convergence of the objective values generated by the iterative algorithm. Since the feasible set is not fixed in each iteration, we require the condition (d) to prove the convergence of the objective function and the corresponding arguments namely, the transmit and the receive precoders to a stationary point [30]. Assuming the conditions (a), (b) and (c) are satisfied, using [31, Th. 3.14], we can show that the bounded monotonically decreasing objective value sequence has a unique minimum. Finally, we use the discussions in [22], [26], [32] to show that the limit point of the iterative algorithm is indeed a stationary point of the nonconvex problem in (16) and (26).

A. Bounded Objective Function and Compact Feasible Set

The feasible sets of the problems (16) and (26) are bounded and closed, which is verified by the total power constraint on the transmit precoders (16d), therefore, the sets are compact.

The minimum value of the norm operator in the objective is zero, *i.e*, $\|x\|_q > -\infty$, therefore it is bounded below. The objective function is Lipschitz continuous over the feasible set, and therefore, it is bounded from above as well, since the feasible set is bounded.

B. Monotonicity of the Objective Sequence

In this section, we show the nonincreasing behavior of the objective values in each SCA and AO update. Let us consider a general formulation for the problems in (16) and (26) as

$$\begin{aligned} \underset{\mathbf{m},\mathbf{w},\boldsymbol{\gamma}}{\text{minimize}} \quad & f(\mathbf{m},\mathbf{w},\boldsymbol{\gamma}) && \text{(44a)} \\ \text{subject to} \quad & h(\boldsymbol{\gamma}) - g_0(\mathbf{m},\mathbf{w}) \le 0 && \text{(44b)} \\ & g_1(\mathbf{m},\mathbf{w}) \le 0 && \text{(44c)} \\ & g_2(\mathbf{m}) \le 0 && \text{(44d)} \end{aligned}$$

where g_2, f are convex and h is a linear function. Let g_0, g_1 be convex in either $\mathbf{m}$ or $\mathbf{w}$ but not on both. The constraint in (44b) corresponds to (16b) or (26b) and the constraint (44c) correspond to (16c) or (26c). Other convex constraints are addressed by (44d) and the feasible set of (44) is given by

$$\begin{aligned} \mathcal{F} = \{\, \mathbf{m},\mathbf{w},\boldsymbol{\gamma} \mid \quad & h(\boldsymbol{\gamma}) - g_0(\mathbf{m},\mathbf{w}) \le 0, \\ & g_1(\mathbf{m},\mathbf{w}) \le 0, g_2(\mathbf{m}) \le 0\}. \end{aligned} \quad (45)$$

To solve (44), we adopt AO by fixing a block of variables and optimize for others [33]. In (44), even after fixing the variable $\mathbf{w}$, the problem is nonconvex due to the DC constraint (44b). We adopt SCA presented in [25], [26], [32] by relaxing the nonconvex set by a sequence of convex subsets. Since it involves two nested iterations, we denote the AO iteration index by a superscript (i) and the DC constraint relaxations by a subscript k. Let $\mathbf{m}_*^{(i)}$ be the value of $\mathbf{m}$ when the SCA iteration converges in the i^{th} AO iteration and let $\boldsymbol{\gamma}_{*|\mathbf{w}}^{(i)}$ be the solution for $\boldsymbol{\gamma}$ obtained in the i^{th} AO iteration for a fixed $\mathbf{w}$.

To begin with, let us consider the variable $\mathbf{w}$ is fixed for the i^{th} AO with the optimal value achieved from the previous iteration $i-1$ as $\mathbf{w}_*^{(i-1)}$. In order to solve for $\mathbf{m}$ in the SCA iteration k, we linearize the nonconvex function g_0 using previous SCA iterate $\mathbf{m}_k^{(i)}$ of $\mathbf{m}$ as

$$\begin{aligned} \hat{g}_o(\mathbf{m},\mathbf{w}_*^{(i-1)};\mathbf{m}_k^{(i)}) &= g_0(\mathbf{m}_k^{(i)},\mathbf{w}_*^{(i-1)}) \\ &\quad + \nabla g_0(\mathbf{m}_k^{(i)},\mathbf{w}_*^{(i-1)})^{\mathrm{T}}(\mathbf{m}-\mathbf{m}_k^{(i)}). \end{aligned} \quad (46)$$

Let $\mathcal{X}_k^{(i)}$ be the feasible set for the i^{th} AO iteration and the k^{th} SCA point for a fixed $\mathbf{w}_*^{(i-1)}$ and $\mathbf{m}_k^{(i)}$. Similarly, $\mathcal{Y}_k^{(i)}$ denotes the feasible set for a fixed $\mathbf{m}_*^{(i)}$ and $\mathbf{w}_k^{(i)}$. Using (46), the convex subproblem for the i^{th} AO iteration and the k^{th} SCA point for the variable $\mathbf{m}$ and $\boldsymbol{\gamma}$ is given by

$$\begin{aligned} \underset{\mathbf{m},\boldsymbol{\gamma}}{\text{minimize}} \quad & f(\mathbf{m},\mathbf{w}_*^{(i-1)},\boldsymbol{\gamma}) && \text{(47a)} \\ \text{subject to} \quad & h(\boldsymbol{\gamma}) - \hat{g}_0(\mathbf{m},\mathbf{w}_*^{(i-1)};\mathbf{m}_k^{(i)}) \le 0 && \text{(47b)} \\ & g_1(\mathbf{m},\mathbf{w}_*^{(i-1)}) \le 0, \quad g_2(\mathbf{m}) \le 0 && \text{(47c)} \end{aligned}$$

The feasible set defined by the problem in (47) is denoted as $\mathcal{X}_k^{(i)} \subset \mathcal{F}$. To prove the convergence of the SCA updates in the i^{th} AO iteration, let us consider that (47) yields $\mathbf{m}_{k+1}^{(i)}$ and $\boldsymbol{\gamma}_{k+1}^{(i)}$ as the solution in the k^{th} iteration. The point $\mathbf{m}_{k+1}^{(i)}$ and $\boldsymbol{\gamma}_{k+1}^{(i)}$, which minimizes the objective function satisfies

$$\begin{aligned} & h(\boldsymbol{\gamma}_{k+1}^{(i)}) - g_0(\mathbf{m}_{k+1}^{(i)},\mathbf{w}_*^{(i-1)}) \le -\hat{g}_0(\mathbf{m}_{k+1}^{(i)},\mathbf{w}_*^{(i-1)};\mathbf{m}_k^{(i)}) \\ & + h(\boldsymbol{\gamma}_{k+1}^{(i)}) \le h(\boldsymbol{\gamma}_k^{(i)}) - \hat{g}_0(\mathbf{m}_k^{(i)},\mathbf{w}_*^{(i-1)};\mathbf{m}_k^{(i)}) \le 0. \end{aligned} \quad (48)$$

Using (48), we can show that $\{\mathbf{m}_{k+1}^{(i)},\mathbf{w}_*^{(i-1)},\boldsymbol{\gamma}_{k+1}^{(i)}\}$ is feasible, since the initial SCA operating point $\mathbf{m}_*^{(i-1)}$ was chosen to be feasible from the $i-1^{\text{th}}$ AO iteration. At each SCA update, the feasible set includes the solution from the previous iteration as $\{\mathbf{m}_{k+1}^{(i)},\mathbf{w}_*^{(i-1)},\boldsymbol{\gamma}_{k+1}^{(i)}\} \in \mathcal{X}_{k+1}^{(i)} \subset \mathcal{F}$, therefore, it decreases the objective as [26], [32], [34]

$$\begin{aligned} & f(\mathbf{m}_0^{(i)},\mathbf{w}_*^{(i-1)},\boldsymbol{\gamma}_0^{(i)}) \ge f(\mathbf{m}_k^{(i)},\mathbf{w}_*^{(i-1)},\boldsymbol{\gamma}_k^{(i)}) \\ & \ge f(\mathbf{m}_{k+1}^{(i)},\mathbf{w}_*^{(i-1)},\boldsymbol{\gamma}_{k+1}^{(i)}) \ge f(\mathbf{m}_*^{(i)},\mathbf{w}_*^{(i-1)},\boldsymbol{\gamma}_{*|\mathbf{w}}^{(i)}). \end{aligned} \quad (49)$$

Thus the sequence $f(\mathbf{m}_k^{(i)},\mathbf{w}_*^{(i-1)},\boldsymbol{\gamma}_k^{(i)})$ is nonincreasing and approaches a limit point of the SCA as $k \to \infty$. The equality in (49) is achieved upon the SCA convergence. The feasible point $\{\mathbf{m}_*^{(i)},\mathbf{w}_*^{(i-1)},\boldsymbol{\gamma}_{*|\mathbf{w}}^{(i)}\}$ need not to be a stationary point of (44), since it is a minimizer over the set $\mathcal{X}_*^{(i)} \subset \mathcal{F}$.

Once the solution is found for a fixed $\mathbf{w}$, it is then optimized by fixing $\mathbf{m}$ as $\mathbf{m}_*^{(i)}$. Even after treating $\mathbf{m}$ as a constant, the problem is nonconvex due to the DC constraint. Following similar approach as above, we can find a minimizer $\{\mathbf{m}_*^{(i)},\mathbf{w}_{k+1}^{(i)},\boldsymbol{\gamma}_{k+1}^{(i)}\}$ for a similar convex subproblem (47) at each iteration k. Note that $\gamma_{k+1}^{(i)}$ is reused since the variable $\mathbf{m}$ is fixed in the i^{th} AO iteration. The convergence and the non-increasing behavior of the objective follows similar arguments as above.[5] The limit point of the converged SCA subproblems with $\mathbf{w}$ as a variable is given by $\{\mathbf{m}_*^{(i)}, \mathbf{w}_*^{(i)}, \boldsymbol{\gamma}_{*|\mathbf{m}}^{(i)}\}$, which is a minimizer in the set $\mathcal{Y}_*^{(i)} \subset \mathcal{F}$.

Finally, to prove the global convergence of the iterative algorithm, we need to show the nonincreasing behavior of the objective function in between the AO updates, *i.e*,

$$\begin{aligned} f(\mathbf{m}_*^{(i)}, \mathbf{w}_*^{(i)}, \boldsymbol{\gamma}_{*|\mathbf{m}}^{(i)}) &\leq f(\mathbf{m}_*^{(i)}, \mathbf{w}_0^{(i)}, \boldsymbol{\gamma}_0^{(i)}) \\ &\leq f(\mathbf{m}_*^{(i)}, \mathbf{w}_*^{(i-1)}, \boldsymbol{\gamma}_{*|\mathbf{w}}^{(i)}). \end{aligned} \tag{50}$$

Let $\mathbf{m}_*^{(i)}$ and $\boldsymbol{\gamma}_{*|\mathbf{w}}^{(i)}$ be the solution obtained by solving (47) iteratively until SCA convergence in the i^{th} AO iteration for $\mathbf{m}$ and $\boldsymbol{\gamma}$ with fixed $\mathbf{w} = \mathbf{w}_*^{(i-1)}$. To find $\mathbf{w}_0^{(i)}$, we fix $\mathbf{m}$ as $\mathbf{m}_*^{(i)}$ and optimize for $\mathbf{w}$. Since the convex function is linearized in (44b), the fixed operating point is also included in the feasible set $\{\mathbf{m}_*^{(i)}, \mathbf{w}_*^{(i-1)}, \boldsymbol{\gamma}_{*|\mathbf{w}}^{(i)}\} \in \mathcal{Y}_0^{(i)}$ by following (48). Using this, we can show the monotonicity of the objective sequence as

$$f(\mathbf{m}_*^{(i)}, \mathbf{w}_0^{(i)}, \boldsymbol{\gamma}_0^{(i)}) \leq f(\mathbf{m}_*^{(i)}, \mathbf{w}_*^{(i-1)}, \boldsymbol{\gamma}_{*|\mathbf{m}}^{(i)}). \tag{51}$$

The AO update follows $\{\mathbf{m}_*^{(i)}, \mathbf{w}_*^{(i-1)}, \boldsymbol{\gamma}_{*|\mathbf{w}}^{(i)}\} \in \{\mathcal{X}_*^{(i)} \cap \mathcal{Y}_0^{(i)}\}$. Using (a),(b) and (c), we can guarantee the convergence of the objective value sequence generated by the iterative algorithm.

C. Uniqueness of the Beamformer Iterates

The uniqueness of the iterates $\{\mathbf{m}_k^{(i)}, \mathbf{w}_*^{(i-1)}, \boldsymbol{\gamma}_{k|\mathbf{w}}^{(i)}\}$ are guaranteed for the MSE reformulated problem in (28) when all the constraints are active. However when $f(\mathbf{m}_k^{(i)}, \mathbf{w}_*^{(i-1)}, \boldsymbol{\gamma}_{k|\mathbf{w}}^{(i)}) = 0$ after some iteration, say k and i, the uniqueness of the iterates are not guaranteed since there are many solutions satisfying the constraints without affecting the objective value. Similarly, when the objective is non-zero, problem defined in (20) has a unique minimizer if the receivers are matched to the corresponding transmit beamformers, *i.e*, $q_{l,k,n} = 0$ in (18), while choosing a feasible operating point $\mathbf{m}_0^{(0)}$ and $\mathbf{w}_0^{(0)}$ to initialize the iterative procedure.

In general, to ensure the uniqueness of the transmit precoders and the receive beamformers in all iterations, *i.e*, even when the objective value is zero, the convex subproblems in (20) and (28) can be regularized with a strongly convex function. The objective function for the $i+1^{\text{th}}$ iteration can be regularized as

$$\|\tilde{\mathbf{v}}\|_q + c\,\|\mathbf{x} - \mathbf{x}^{(i)}\|_2^2 \tag{52}$$

where $\mathbf{x}$ is the vector containing all the optimization variables and c is a positive constant to make the objective strongly convex. The uniqueness of the minimizer is guaranteed by the strongly convex objective function [32]. The vector $\mathbf{x}^{(i)}$ is the value of $\mathbf{x}$ obtained from the i^{th} iteration.

[5]Note that we can also use the MMSE receiver in (23b) instead of performing the SCA updates until convergence for the optimal receiver.

D. Convergence of the Beamformer Iterates

So far, we have considered the convergence of the objective sequence. In this section, we discuss the convergence of the beamformer iterates generated by the iterative algorithm for (44) using [30], [35]. Let $\mathbf{x} \triangleq [\mathbf{m}_*^{(i)}, \mathbf{w}_*^{(i)}, \boldsymbol{\gamma}_{*|\mathbf{m}}^{(i)}]$ be the stacked vector of the solution point in the i^{th} AO iteration and let $\mathcal{A}$ be the point-to-set mapping algorithm defined as

$$\mathbf{x}^{(i+1)} \in \mathcal{A}(\mathbf{x}^{(i)}) : \underset{\mathbf{x}}{\operatorname{argmin}}\ f(\mathbf{x}; \mathbf{x}^{(i)})\ \forall \mathbf{x} \in \mathcal{Y}_*^{(i)} \subset \mathcal{F}. \tag{53}$$

Since the algorithm finds a unique solution in each iteration,[6] the mapping denoted by $\mathbf{x}^{(i+1)} = \mathcal{A}(\mathbf{x}^{(i)})$. Let $\mathcal{F}^* \subset \mathcal{F}$ be the set of fixed points, *i.e* $\forall \mathbf{x}^* \in \mathcal{F}^*$, $\mathbf{x}^* = \mathcal{A}(\mathbf{x}^*)$, identified by the algorithm $\mathcal{A}$ for different initialization points.

Using [30, Th. 3.1], convergence of the iterates to a fixed point can be shown, if the following conditions are satisfied.

- Objective function f should be bounded and continuous.
- The feasible set $\mathcal{Y}_*^{(i)}$ in each iteration should be compact.
- The sequence $\{\mathbf{x}^{(i)}\}$ generated by the algorithm $\mathcal{A}$ should be strictly monotonic with respect to the objective function f, *i.e*, $\mathbf{x}' = \mathcal{A}(\mathbf{x})$ implies $f(\mathbf{x}') < f(\mathbf{x})$ whenever $\mathbf{x}$ is not a fixed point.

Note that the above conditions are satisfied by the proposed algorithm $\mathcal{A}$ in (53) and the strict monotonicity can also be verified from (51). Using [35] and [30, Th. 3.1], we can show that (i) all limit points will be a fixed point, (ii) $f(\mathbf{x}^{(i)}) \to f(\mathbf{x}^*)$, where $\mathbf{x}^*$ is a fixed point, and (iii) $\mathbf{x}^*$ is a regular point, *i.e*, $\|\mathbf{x}^{(i)} - \mathcal{A}(\mathbf{x}^{(i)})\| \to 0$. Even though the fixed points in the set $\mathcal{F}^*$ can be achieved by the algorithm $\mathcal{A}$ for different initial points, the algorithm $\mathcal{A}$ achieves a unique limit point, *i.e*, $\mathcal{F}^* = \{\mathbf{x}^*\}$, once a feasible point is chosen to be the operating point while initializing $\mathcal{A}$.

E. Stationary Points of the Nonconvex Problem

To show the limit point of the sequence of iterates $\{\mathbf{x}^{(i)}\}$ is a stationary point, it must satisfy the KKT conditions of the problem in (44). As $i \to \infty$, the objective converges as

$$\begin{aligned} f(\mathbf{m}_*^{(i)}, \mathbf{w}_*^{(i)}, \boldsymbol{\gamma}_{*|x}^{(i)}) &= f(\mathbf{m}_*^{(i+1)}, \mathbf{w}_*^{(i)}, \boldsymbol{\gamma}_{*|y}^{(i+1)}) \\ &= f(\mathbf{m}_*^{(i+1)}, \mathbf{w}_*^{(i+1)}, \boldsymbol{\gamma}_{*|x}^{(i+1)}) \end{aligned} \tag{54}$$

where $\{\mathbf{m}_*^{(i+1)}, \mathbf{w}_*^{(i+1)}, \boldsymbol{\gamma}_{*|x}^{(i+1)}\} \in \mathcal{F}^*$ is the limit point of $\mathcal{A}$ for a given initial point. As discussed in Appendix A-D, the unique limit point of the sequence of iterates $\{\mathbf{x}^{(i)}\}$ is a fixed point of $\mathcal{A}$. Using [26, Th. 10] and [32, Th. 2 and 11], we can show that the fixed points obtained by the algorithm $\mathcal{A}$ under the mapping f is a stationary point of the nonconvex problem (44). It can be verified by writing the KKT expressions for the original problem in (44) and the relaxed one in (47) at the limit point (54), using the relation in (46) between the nonconvex constraint and the corresponding convex approximation [22].

[6]Index i denotes the AO iteration to update the vector $\mathbf{x}$, which involves two SCA iterations performed until convergence, *i.e*, one for the variable $\mathbf{m}$ by keeping $\mathbf{w}$ fixed and another for the variable $\mathbf{w}$ by fixing $\mathbf{m}$ as constant.

Appendix B
Convergence Proof for Distributed Algorithms

The convergence of the distributed algorithm in Algorithm 2 follows the same discussion as the one in Appendix A, if the subproblem in (31) converges to the centralized solution. The feasible set of the problem in (31) satisfies the Slater's constraint by having a non-empty interior and is also compact. Since the subproblem in (31) is convex, the convergence of the primal and the ADMM scheme can be guaranteed by using the discussions in [11], [36] and [37, Prop. 4.2].

Unlike the primal or the ADMM methods, the decomposition approach via KKT conditions, presented in Section IV-C, updates all the optimization variables at once, *i.e*, the SCA update of $\epsilon^{(i-1)}$, the AO update of $\mathbf{w}_{l,k,n}$ and the dual variable update of α using subgradient method. Therefore, it is difficult to prove theoretically the convergence of the algorithm to a stationary point of the nonconvex problem in (16).

The algorithm in (42) is identical to (28), if the receivers $\mathbf{w}_{l,k,n}$ and the MSE operating point $\epsilon_{l,k,n}^{(i-1)}$ are fixed to find the optimal transmit precoders $\mathbf{m}_{l,k,n}$ and the dual variable $\alpha_{l,k,n}$. Note that it requires four nested loops to obtain the centralized solution, namely, the receive beamformer loop, MSE operating point loop, the dual variable update loop and the bisection method to find the transmit precoders.

However, to avoid the nested iterations, the proposed method performs group update of all the variables at once to obtain the transmit and the receive beamformers with the limited number of iterations, thus it achieves improved speed of convergence. Since the optimization variables are updated together, it is theoretically difficult to prove the monotonicity of the objective in each block update. Therefore, it is difficult to prove the convergence of the Algorithm 3 to a stationary point of the original nonconvex problem in (26). Even though the convergence is not theoretically guaranteed, it converges in all numerical experiments considered in Section V-B.

Appendix C
KKT conditions for MSE approach

In order to solve for an iterative precoder design algorithm, the KKT expressions for the problem in (41) are obtained by differentiating the Lagrangian by assuming the equality constraint for (41b) and (41c). At the stationary points, following conditions are satisfied.

$$\nabla_{\epsilon_{l,k,n}} : -\alpha_{l,k,n} + \frac{\sigma_{l,k,n}}{\tilde{\epsilon}_{l,k,n}} = 0 \tag{55a}$$

$$\nabla_{t_{l,k,n}} : -q\, a_k \Big(Q_k - \sum_{n=1}^{N} \sum_{l=1}^{L} t_{l,k,n} \Big)^{(q-1)} + \frac{\sigma_{l,k,n}}{\log_2(e)} = 0 \tag{55b}$$

$$\begin{aligned} \nabla_{\mathbf{m}_{l,k,n}} : & \sum_{y \in \mathcal{U}} \sum_{x=1}^{L} \alpha_{x,y,n} \mathbf{H}_{b_k,y,n}^{\mathrm{H}} \mathbf{w}_{x,y,n} \mathbf{w}_{x,y,n}^{\mathrm{H}} \mathbf{H}_{b_k,y,n} \mathbf{m}_{l,k,n} \\ & + \delta_b \mathbf{m}_{l,k,n} = \alpha_{l,k,n} \mathbf{H}_{b_k,k,n}^{\mathrm{H}} \mathbf{w}_{l,k,n} \end{aligned} \tag{55c}$$

$$\begin{aligned} \nabla_{\mathbf{w}_{l,k,n}} : & \sum_{(x,y) \neq (l,k)} \mathbf{H}_{b_y,k,n} \mathbf{m}_{x,y,n} \mathbf{m}_{x,y,n}^{\mathrm{H}} \mathbf{H}_{b_y,k,n}^{\mathrm{H}} \mathbf{w}_{l,k,n} \\ & + \mathbf{I}_{N_R} \mathbf{w}_{l,k,n} = \mathbf{H}_{b_k,k,n}\, \mathbf{m}_{l,k,n}. \end{aligned} \tag{55d}$$

In addition to the primal constraints given in (41b), (41c) and (41d), the complementary slackness criterion must also be satisfied at the stationary point. Upon solving the above expressions in (55) with the complementary slackness conditions, we obtain the iterative algorithm to determine the transmit and the receive beamformers as shown in (42).