Reviewers Comments & Authors Replies

Manuscript No.	Paper T-SP-18051-2014, submitted to *"IEEE Transactions on Signal Processing"*
Title	"Traffic Aware Resource Allocation Schemes for Multi-Cell MIMO-OFDM Systems"
Authors	Ganesh Venkatraman, Antti Tölli, Markku Juntti, and Le-Nam Tran

The authors would like to thank the associate editor and the reviewers for their valuable comments on the manuscript, which have been greatly helpful to improve the paper quality. Based on the comments, we have made several revisions to the paper, which are summarized below.

1. We have shortened the discussions on the existing *backpressure algorithm* in Section III-A.
2. Initialization of the successive convex approximation (SCA) operating point is included in the revised manuscript
3. Section III-D is updated to include the guidelines involved in selecting the sub-channel order.
4. Strong convexity of the objective function is emphasized in Appendix A-B.
5. Strict monotonicity of the algorithm is discussed in a detailed manner in Appendix A-C.
6. We have provided additional details on the distributed algorithm convergence as suggested by the reviewers

In what follows, the comments are listed, each followed immediately by the corresponding reply from the authors. The reviewers questions in the revised manuscript are highlighted in blue color and the authors responses are presented in black. Unless otherwise stated, all the numbered items (figures, equations, references, citations, etc) in this response letter refer to the revised manuscript. The revisions in the manuscript are highlighted in blue color.

Response to Reviewer - 1's Comments

Comments: The response to the reviewer's concerns are generally satisfying, except the convergence proof.

Reply: We thank the reviewer for providing valuable and insightful comments.

For the resubmitted manuscript, the reviewer still has the following concerns

1. Considering the length of the manuscript, it would be better to shorten some parts that are not new in this manuscript, e.g. III.A. More space can be left for convergence proof, which is very important.

Reply: We have removed a couple of paragraphs from Section III-A and shortened the discussions on Section V (*i.e.*, the simulation results) to provide additional details for the convergence proof. We believe the removed parts do not affect the readability of the paper and the added text regarding convergence proof certainly improves the quality.

2. In convergence proof (48), why does the 2nd inequality hold? In fact, to prove the feasibility of $m_{k+1}^{(i)}, w_*^{(i-1)}; m_k^{(i)}$, the part between 2nd and 3rd inequality is not necessary, ≤ 0 directly follows the 2nd inequality since the solution $m_{k+1}^{(i)}, \gamma_{k+1}^{(i)}$ is the optimal solution, and therefore feasible.

Reply: We have removed the additional inequality relating the previous operating point from (49) in the revised manuscript.

3. The solutions SCA iterations $\mathbf{m}_k^{(i)}$ does not necessarily converge. In fact $\mathbf{m}$ has compact feasible region, and thus $\mathbf{m}_k^{(i)}$ has limit points for any specific i. However $m_*^{(i)}$ does not necessarily exist (the whole sequence $\mathbf{m}_k^{(i)}$ may be not convergent). Similar problem happens to $\mathbf{w}_k^{(i)}$.

Reply: Even though the objective converges in each SCA update, it is not guaranteed for the iterates involved in the iterative algorithm, namely, $\mathbf{m}_k^{(i)}$ and $\mathbf{w}_k^{(i)}$ to converge. It follows from the convexity of the objective in problem (16), which can have multiple solutions in the feasible set. However, we have modified the convergence discussion using the strong convexity argument by regularizing the objective function with a quadratic term as discussed in [30] and [31]. This approach ensures the uniqueness of the solution in each SCA update.

Since our problem is (highly) nonconvex, there are possibly many stationary points. If we start from different initial points, we may end up in different stationary points. That is to say, the limit point of the SCA approach is not unique. However, by regularizing the objective with a small quadratic term, the iterate in each step of the SCA algorithm is unique. This and the fact that the set is compact ensure that the SCA converges to a limit point. We have updated the manuscript to include this information in the convergence proof of the centralized algorithm in Appendix A. The above modifications are included in Appendix A-B around (46) and in Appendix A-C after (50).

4. Strict monotonicity with respect to the objective function f should be rigorously proved. Note that to guarantee the uniqueness of the beamformer iterates, (52) instead of the objective function is used.

Reply: As suggested by the reviewer, we have included the strict monotonicity of the objective sequence in Appendix A-C following (52). Moreover, we have also mentioned that we in fact use the regularized objective in (46b) to analytically prove the convergence of the iterates, although we have numerically observed that the objective sequence of Algorithm 1 (with the original objective) always converges. Please refer to the paragraph before Section III-C.

5. Note that the conclusions [32, Thm 2] and [26, Thm 10] have lots of assumptions. To invoke these reference, explicit exposition should be provided to show that these conclusions can be applied to our problem. The same questions occur to the proof in Appendix B, where conclusions in [11] [36] and [37] are used. Too many details are omitted to make the proof convincing and clear.

Reply: We have updated the manuscript to include the details regarding the stationary point discussion in Appendix A-E. The convergence analysis for the primal and the alternating directions method of multipliers (ADMM) algorithms are presented in Appendix B. We have also mentioned the conditions required for showing the stationarity property of the limit point and included more mathematical steps to make the convergence proof of distributed approaches rigorous.

Response to Reviewer - 2's Comments

The authors have addressed many of my previous comments. However, there are still several major issues that need further clarification.

Reply: We thank the reviewer for providing further useful comments which help us to greatly improve the manuscript.

1. The revised paper did not address my previous comment about how to select the sub-channel ordering. I understand that finding the best sub-channel ordering requires exhaustive search which has extremely high complexity. But it is important to provide a guidance on what would be a good choice of sub-channel ordering. For example, can we achieve a good performance by using a low complexity ordering algorithm such as a greedy sub-channel ordering algorithm?

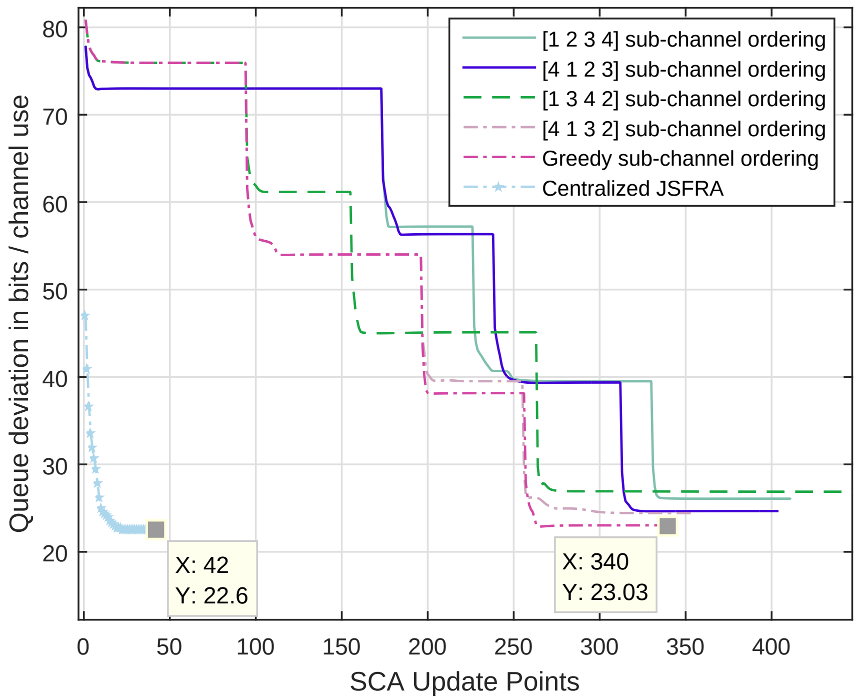
Reply: We apologize for not clarifying the sub-channel ordering in detail in our previous manuscript. For simplicity we use random sub-channel ordering in our paper. That is, after finding the precoders for a current sub-channel, we can choose any previously unselected sub-channels as the next candidate sub-channel for which the precoders are identified using the updated backlogged packets. As suggested by the reviewer, the greedy sub-channel ordering can also be considered while selecting the order. The greedy ordering can be based on sorting the best channel gain from each sub-channel, which is obtained by finding the highest channel norm seen between any user and the corresponding serving base station (BS).

We remark that the choice of a sub-channel ordering scheme should also consider the number of backlogged packets associated with each user. We have emphasized this issue by comparing various ordering schemes for the same system model with two different set of backlogged packets associated with each user. We do not include in the manuscript due to space limitations, we have included in the response letter to answer the reviewers' question. We considered a system with $N = 4$ sub-channels, $N_B = 2$ BSs with $N_T = 4$ transmit antennas and $K = 12$ single antenna users. The path loss (PL) is distributed uniformly over $[0, -3]$ dB. The number of backlogged packets assumed for each user is provided in the corresponding captions in Figure A.

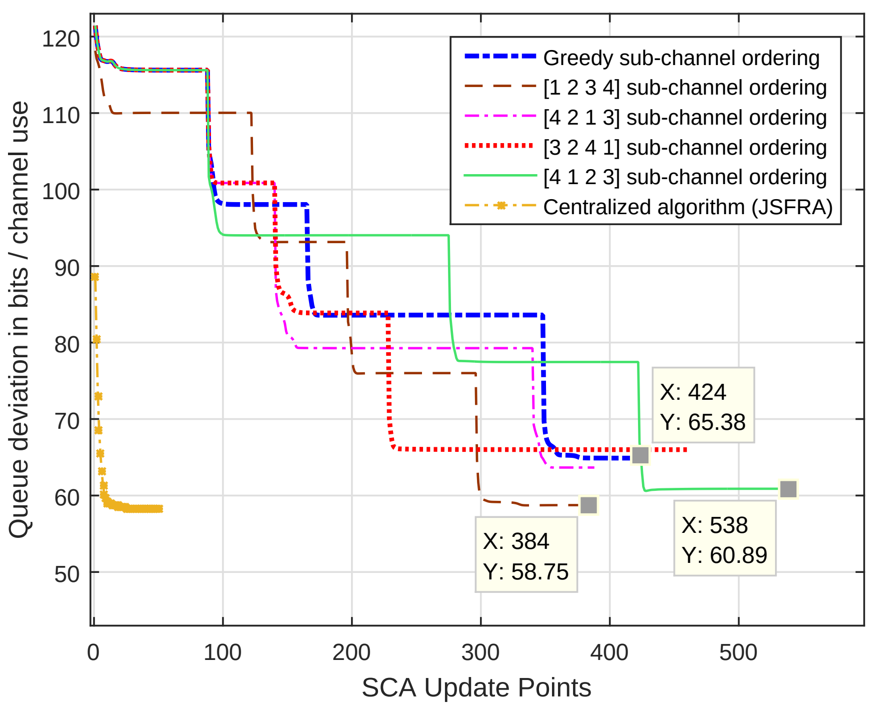
Figure A compares different sub-channel ordering schemes in terms of the total number of backlogged packets remaining in the system. As can be seen from Figure A(i), the greedy sub-channel ordering provides a favorable way of choosing sub-channels to minimize the total number of backlogged packets. However in Figure A(ii), it is evident that the greedy sub-channel ordering is no longer better in comparison to the few other random ordering schemes. We have included the discussions regarding the greedy sub-channel ordering in the revised manuscript as a heuristic method under Section III-D final paragraph. However, it is worth noting that as the number of users in the system increases, all sub-channel ordering schemes are similar in minimizing the total number of backlogged packets in the system without any noticeable gain.

2. The authors mentioned that the signaling overhead of the distributed algorithm can be reduced by using a smaller number of iterations $J_{\max}$. But still, you didn't answer my question about whether the signaling overhead of the distributed algorithm is smaller than the centralized algorithm. You should first analyze the signaling overhead of the distributed algorithm for fixed $J_{\max}$ and the signaling overhead of the centralized algorithm. Then you should point out under what $J_{\max}$ the distributed algorithm will have less signaling overhead than the centralized algorithm. Is it possible that the distributed algorithm always has more signaling overhead than the centralized algorithm even when $J_{\max} = 1$? Finally, there is a trade-off between performance and signaling overhead ($J_{\max}$) for the distributed algorithm. For the same signaling overhead (we can control $J_{\max}$ to make the signaling overhead of the distributed algorithm approximately equal to that of the centralized algorithm), does the distributed algorithm achieve better performance than the centralized algorithm?

Reply: We thank the reviewer for the insightful comment and we apologize for the lack of clarity in explaining this information in our earlier manuscript.



(i) Number of backlogged packets for each user in bits $Q_k = [11, 8, 14, 6, 6, 2, 10, 10, 5, 6, 9, 5]$



(ii) Number of backlogged packets for each user in bits $Q_k = [8, 9, 12, 8, 12, 5, 4, 10, 8, 5, 7, 9]$

Figure A: Convergence of the algorithms for $\{N, N_B, K, N_T, N_R\} = \{4, 2, 12, 4, 1\}$ using ℓ_1 norm

(a) The amount of signaling overhead of the distributed algorithm and the centralized one depends on the system model of consideration. For example, let us consider a model with $N = 1$ sub-channel, $K = 4$ users and $N_B = 2$ BSs, each serving 2 users. Let $N_T = 4$ be the number of transmit antennas and $N_R = 1$ be the number of receive antennas at each user. We note that a centralized algorithm requires the knowledge of all channels matrices in the system and thus the resulting amount of information exchange is proportional to the product of the numbers of users (K), BSs (N_B), and transmit and receive antennas (*i.e.*, N_T and N_R).

In order to quantify the total number of bits required in a centralized solution, let us assume that each complex channel for a single-input single-output (SISO) model requires 10 bits, *i.e*, 4 bits for amplitude and 6 bits for phase (assuming phase is more important) or it can be a equal share of 5 bits for both amplitude and phase. Using this assumption, the total number of bits for channel information to be exchanged via backhaul is $10 \times K \times N_B \times N_R \times N_T = 320$ bits. On the other hand, for the distributed algorithms, let us consider 6 bits for quantizing each element of the consensus vectors. Consequently, the proposed distributed solutions require $6 \times 2 \times 2$ bits to be exchanged in each iteration.

In the above example, for the same signaling overhead as in centralized method, we can perform only up to 6 SCA updates for $J_{\max} = 2$ (*i.e.*, two updates for ADMM part). This may not be sufficient for the distributed algorithms to attain the same performance as the centralized method. However, as the number of sub-channels, users and/or the antenna elements increases, it may not be a feasible option to send complete channel state informations (CSIs) across the coordinating BSs to the centralized controller in the current cellular systems. In addition to comparing the signaling overhead, we also need to consider the performance loss due to the CSI quantization in the centralized approaches, which is beyond the scope of our paper. Generally, the performance is significantly degraded if precoders are designed with the quantized CSI as discussed in [R1]. Moreover, in the centralized algorithm, resulting transmit precoders need to be exchanged with the corresponding BSs before the actual transmission, which involves significant overhead in the backhaul usage.

[R1] Muhammad-Fainan Hanif, Le-Nam Tran, Antti Tlli, Markku Juntti, and Savo Glisic, "Efficient Solutions for Weighted Sum Rate Maximization in Multicellular Networks With Channel Uncertainties", *IEEE Trans. Signal Process.*, vol 61, no. 22, pp. 5659-5674, Nov. 2013.

(b) Using the above discussion, we can say that when the system size is huge, it would be favorable to consider the distributed algorithm over the centralized approach due to the huge signaling overhead involved in exchanging the channels.

(c) From the above discussions, we can see that the ADMM and primal schemes require significant overhead compared to the centralized algorithms for a small system. However, for a system involving more coordinating BSs, users and antenna elements, it would be beneficial to use the proposed distributed algorithms with $J_{\max} > 1$ to have a strictly monotonic decrease in the objective. If we set $J_{\max} = 1$, the convergence of the proposed algorithms is not guaranteed as shown in Figure B. The system model considered is the same as the one used for Figure A(i). Note that Figure B plots the queue deviation at the SCA update points only. The number of backlogged packets in each ADMM iteration is not shown for clarity reason.

The above discussion is included in a condensed form in the first and the last paragraph of Section IV-C.

In reality, since the channel is time-correlated, it is enough to update the precoders once per radio frame. Thus, it is not necessary for the distributed algorithm to converge until the end. Instead, the decentralized parts only need to follow the fading process when $J_{\max} > 1$. The performance of the distributed algorithm based on the dual decomposition scheme was discussed for the time-correlated fading in Section C of [12], which showed that it is enough for the distributed precoder design to follow the fading process to provide the desired performance. Studying the performance of distributed algorithms for the time correlated case is beyond the scope of our paper and thus is not considered in

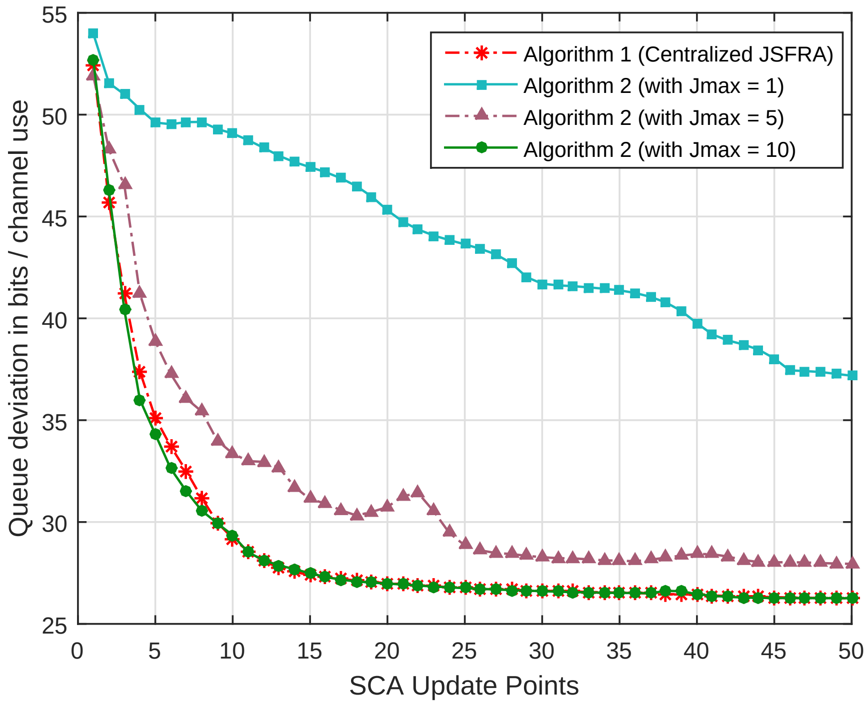


Figure B: Convergence of the centralized and distributed algorithms for $\{N, N_B, K, N_T, N_R\} = \{4, 2, 12, 4, 1\}$ using ℓ_1 norm

the current manuscript. However, we take this opportunity to show a plot demonstrating this behavior for the Karush-Kuhn-Tucker (KKT) based algorithm presented in Section IV-C of the manuscript.

Figure C compares the performance of the distributed KKT approach, *i.e.*, Algorithm 3 for different values of $I_{\max}$. The signaling requirements are outlined in Algorithm 3 and the overhead involved in the signaling is penalized in the achievable rate of the users.[1] We considered that the channel is coherent over $N_S = 100$ symbols and the precoder update is performed by exchanging the required information for $J_{\max} = 3, 5, 10$. The overhead is considered as $\tilde{t}_{l,k,n} = (1 - \frac{J_{\max}}{N_S}) \times t_{l,k,n}$, where $\tilde{t}_{l,k,n}$ is the rate seen by the user and the factor $(1 - \frac{J_{\max}}{N_S})$ is considered as a penalty involved due to the precoder exchange as discussed in [R2] and [R3]. The average number of backlogged packets after each transmission slot is evaluated as

$$\chi = \sum_{k=1}^{K} [Q_k - \tilde{t}_k]^+ \tag{i}$$

Unlike Algorithm 3, the centralized scheme presented in Figure C has no penalty term and it is used as a benchmark. In order to improve the performance of the distributed scheme, the operating point involved in the SCA algorithm is considered from the earlier frame instead of begin generated randomly.

[R2] Changxin Shi, Berry R.A., Honig M.L., "Bi-Directional Training for Adaptive Beamforming and Power Control in Interference Networks," *IEEE Transactions on Signal Processing*, vol.62, no.3, pp.607,618, Feb.1, 2014

[R3] P. Jayasinghe, A. Tölli, J. Kaleva, M. Latva-aho, "Bi-directional Signaling for Dynamic TDD with Decentralized Beamforming", in *Proceedings of IEEE ICC SmallNets Workshop*, London, UK, June, 2015

[1]Please refer to [27] for further details.

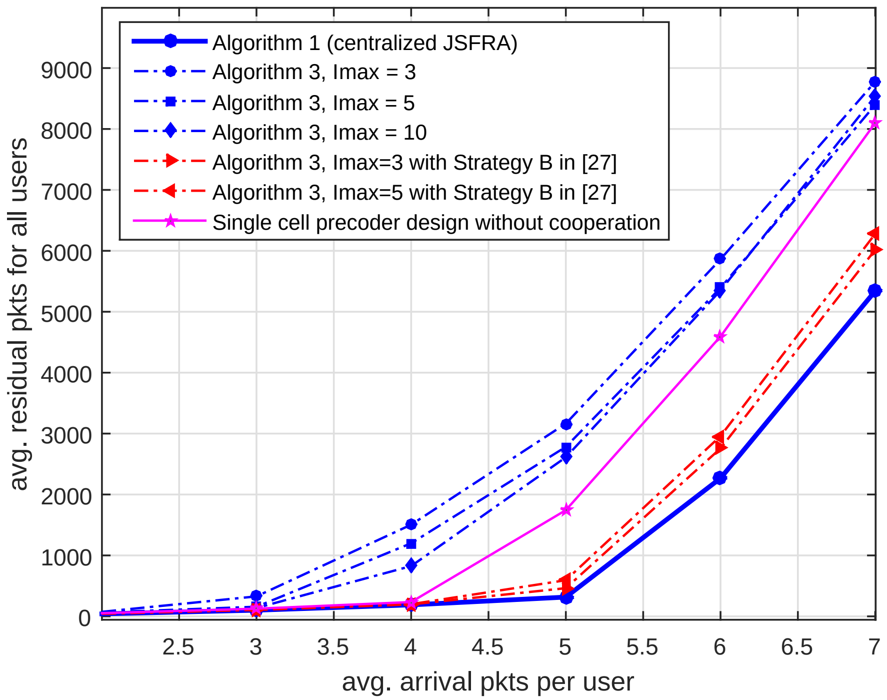


Figure C: Expected number backlogged packets after each transmission instant for a system $\{N, N_B, K, N_T, N_R\} = \{4, 2, 16, 4, 2\}$ evaluated for 500 slots

3. If the authors can't prove the convergence of the ADMM algorithm (or the decomposition approach via KKT conditions) in Section IV.B, then at least, you should discuss the property of the fixed point of the algorithm. For example, does there exist a fixed point of the algorithm? If so, is the fixed point of the algorithm unique? Is any fixed point of the algorithm also the optimal solution of the original problem in (20)? Assuming that the ADMM algorithm converges to a fixed point, will the interference vector in (39) converges to the actual interference in the network? These questions must be clarified in the paper. Otherwise, it is not clear how the ADMM algorithm is related to the original problem in (20). Similar questions should also be answered for the decomposition approach via KKT conditions.

Reply: We thank the reviewer for raising the concern regarding the convergence of the distributed algorithms. We have substantially improved the clarity of both primal and the ADMM based approach in the revised manuscript (see Appendix B). There, we have utilized the strong convexity of the objective function to ensure the uniqueness of the minimizer in each SCA step. Additionally, since each subproblem in (20) or (28) has a convex feasible set, the convergence of the ADMM scheme is guaranteed by using [10] and [36, Prop. 4.2] as discussed in Appendix B.

In our paper, the number of ADMM iterations is denoted by $J_{\max}$ in Algorithm 2. In order to obtain the centralized solution, we need to perform the distributed approaches until convergence or for very large value of $J_{\max}$. In practice, it may not be possible to perform the distributed methods until convergence. In such cases, the overall convergence of the objective sequence and beamformer iterates of Algorithm 2 is not guaranteed. The reason is that the strict monotonicity of the objective is not ensured in each SCA update.

However, if we perform the distributed algorithms for significantly enough to ensure strict monotonicity of the objective, then the overall of convergence of Algorithm 2 is guaranteed by following the arguments in [27]. Note that in addition, uniqueness of the limit point is also necessary for the sequence convergence. It can be assured by regularizing the original convex objective with a quadratic term as in (46b) in each SCA step to guarantee strong convexity, and thereby ensuring a unique minimizer and limit point.

Additionally, to reduce the signaling overhead involved with the information exchange due to multiple loops, we can also combine the minimum mean squared error (MMSE) receiver update in (23b) after each SCA step for the transmit precoders. If in addition $J_{\max}$ is sufficiently large, then the strict monotonicity of the objective sequence is still ensured by the combined update. It follows from the fact that the MMSE receivers are optimal for fixed transmit precoders obtained after each SCA update. The newly found solution is also feasible and the objective improves monotonically in strict sense. Therefore, if the above conditions are satisfied, then the convergence of the sequence of iterates is guaranteed.

For our simulation scenario, we have fixed the maximum number of ADMM iterations in each SCA step as $J_{\max} = 20$. Note that it is sufficient to ensure strict monotonic decrease of the objective value in each SCA step. However, the number of iterations to ensure strict monotonicity of the objective depends on the problem under consideration. The practical significance of limited number of ADMM iterations is discussed briefly in [10, Section 3.2.2].

Response for the reviewer questions related to the ADMM convergence for subproblem (20) is discussed below.

(a) Note that the objective is regularized with a quadratic term to ensure strong convexity as discussed in [30], [31]. Since the primal objective and the augmented Lagrangian are strongly convex, the optimal solution is unique due to the convexity of the feasible domain. Therefore, the limit point of the ADMM algorithm is unique, which guarantees the existence of fixed point.

(b) Note that if the objective function is convex, then it can have multiple optimal solutions. However, due to the additional quadratic term in the objective of (20) in each SCA step, uniqueness of the minimizer is ensured. Therefore, even though the original problem (20) with the convex objective has a set of fixed points, due to the strong convexity of the modified objective, the ADMM will find a fixed point upon convergence. However, note that the choice of fixed point depends on the SCA operating point, since it is also used to regularize the objective as in (46b).

(c) Using the earlier argument, any fixed point is a solution for the problem (20), since the original convex objective can have multiple solutions. The particular choice of the fixed point of the ADMM approach depends on the operating point used in the regularization term in the objective.

(d) Upon convergence of the ADMM approach, the interference vector is guaranteed to be equal to the actual interference seen in the network. This is the main principle of ADMM where the local variables (the vectors in (38)) converge to the global variables, which are the actual inference vectors.

However, to answer the reviewer questions in relation to the overall convergence of the distributed method presented in Algorithm 2, we consider the following conditions are satisfied in each SCA update step.

- The objective function is regularized by a quadratic term as discussed in Appendix A-B to ensure strict convexity of the objective function, which leads to the uniqueness of the solution.
- The distributed algorithms are performed until convergence or for sufficient number of iterations to ensure strict monotonic decrease of the objective sequence.

Upon satisfying the above conditions, even if we update the receive beamformers using the MMSE receiver in (23b) after each SCA step for solving the transmit precoders, strict monotonicity is still ensured. It follows from the fact that the MMSE receivers are optimal for the fixed transmit precoders, and therefore the objective sequence decreases monotonically after each update in strict sense.

(a) Since the above conditions are satisfied by Algorithm 2, then the existence of a fixed point for Algorithm 2 is guaranteed by following the arguments in [27].

(b) In general, fixed point of Algorithm 2 is not unique. Note that the iterative nature of the SCA approach is based on approximating the nonconvex constraint by a convex one around some fixed operating point. While initializing the iterative Algorithm 2, we need to find a fixed operating point upon which the convex function is approximated. Therefore, depending on the choice of initial feasible point, the algorithm terminates at different fixed point upon convergence. Hence, there exist a set of fixed points for Algorithm 2.

(c) The nonconvex problem (16) can have multiple stationary points and the iterative method proposed in Algorithm 2 finds a stationary point of the nonconvex problem. Therefore, all fixed points found by Algorithm 2 upon convergence are indeed the stationary points of the original nonconvex problem.

(d) Upon convergence of the ADMM approach, the interference vector is guaranteed to be equal to the actual interference seen in the network. This is the main principle of ADMM where the local variables (the vectors in (38)) converge to the global variables, which are the actual inference vectors.

On the other hand, the iterative method described in Algorithm 3 is based on a heuristic iterative approach to find a solution of KKT equations, we cannot prove its convergence. When Algorithm 3 terminates, we will use the transmit and receive beamforming vectors to calculate the original objective. Based on the reviewer comments, we have mentioned that the convergence of the Algorithm 3 cannot be proved theoretically (see Appendix B last paragraph).

We have updated the manuscript to clarify all the points mentioned above. Please refer to Section IV-B last paragraph after (40) in the revised manuscript. Additional information regarding the convergence of the distributed algorithms is provided in Appendix B second paragraph. The discussion on the convergence of the KKT based approach for MSE reformulation scheme is also presented in Appendix B last paragraph. For reference purpose, we have also referred the interested reader to [10], which discusses exclusively about the ADMM approach.

Response to Reviewer - 3's Comments

The authors have introduced changes in the manuscript that improved the paper's quality. Additionally, the authors have taken into account the reviewers' comments giving clarifications and modifying the content when required. More specifically, the following aspects have been treated:

Reply: We thank the reviewer for recognizing the changes made on our earlier manuscript. We also thank the reviewer for providing constructive comments to help us to improve our manuscript.

1. Convexity of problem (16). The paragraphs surrounding (16) allow a better understanding of the usage of the additional variables, i.e. gamma and beta, to remove the equality constraint in (2). For the reviewer remains however unclear, the procedure/criterion to determine the operating point for the parameter $\tilde{\beta}$ required in (19) and used in the convex subproblems (20) and (21).

Reply: We thank the reviewer for the comment. We have included additional material on how to choose the operating point while initializing the iterative algorithm. It is described briefly in the fourth paragraph following (23). We have also provided a detailed discussion on how to update the operating point after each iteration in the paragraphs following (23). Since the objective function is convex, we can update the current operating point with the solution obtained from the previous iteration as

$$\tilde{\mathbf{m}}_j^{(i)} = \mathbf{m}_j^{(i)} \tag{ii}$$

where $\mathbf{m} = \{\mathbf{m}_{l,k,n}\}_{\forall l, \forall k \in \mathcal{U}, \forall n}$ corresponds to the collection of all transmit precoders in the system. Note that $\tilde{\mathbf{m}}_j^{(i)}$ is the operating point for the $(j+1)$th SCA iteration in the ith alternating optimization (AO) step and $\mathbf{m}_j^{(i)}$ is the solution obtained by solving (20) in the jth SCA iteration.

2. Proof of convergence. The proof of convergence introduced by the authors in the Appendix seems correct and enhances the content of the manuscript.

Reply: We thank the reviewer for the comment. We have included several materials to make the proof of convergence more rigorous.

Additional Comments -

(a) - The reviewer considers that closing statements regarding the applicability of the proposed schemes are missing. Since the results are quite similar (when not identical), which formulation is preferable between the centralized schemes? Which one for the distributed solutions?

Reply: We have updated the manuscript to discuss possible options for selecting an algorithm for a specific implementation. The choice of considering a centralized approach is equally good when the number of receive antennas is larger than one, *i.e.*, $N_R > 1$. However, in case of a single antenna receiver, the joint space-frequency resource allocation (JSFRA) formulation outlined in Section III-B is more efficient than the mean squared error (MSE) reformulation, as there is no need for the receiver update, thereby reducing the complexity significantly.

As far as the distributed approaches are concerned, the primal and the ADMM schemes are equally favorable when $N_R = 1$ since the signaling required to be exchanged between the coordinating BSs involves only the scalar interference values. However, when $N_R > 1$, then the KKT based distributed approach is more efficient than the primal and the ADMM schemes, since it has less signaling overhead. The choice of selecting ℓ_q norm is discussed in Section II-B last paragraph. The above information is included in Section IV last paragraph, *i.e.*, before the simulations section.

(b) - The last discussion in section IV-C could benefit from restructuring. The information on how to obtain a practical distributed precoder design and to avoid backhaul exchange is too condensed and difficult to understand.

Reply: It has been rewritten to improve the readability as suggested by the reviewer. For additional information, the reviewer is requested to refer [R1].

[R1] Changxin Shi, Berry R.A., Honig M.L., "Bi-Directional Training for Adaptive Beamforming and Power Control in Interference Networks," *IEEE Transactions on Signal Processing*, vol.62, no.3, pp.607,618, Feb.1, 2014

(c) - For the simulation results, why not to unify configurations when possible? Having to read a different configuration for each graph is cumbersome and no additional comparisons are possible between figures. E.g. PL uniformly distributed between [0,-6] dB in Fig. 2 and [0,-3] dB in Fig. 3.

Reply: The main purpose of considering several path loss models is simply to show that the proposed algorithms works reasonably well on various scenarios mentioned in Section V (Simulation Section). For benchmarking, we have also included the centralized algorithm to draw the difference between the other schemes. We are conducting more numerical experiments to have a more complete comparison of all the schemes. Unfortunately those results cannot be included in the current manuscript due to space limitation. We aim at making them available online as a companion technical report whey they are ready.

(d) - In Fig. 1, the description of the system model does not agree with the statement of N = 3 subchannels.

Reply: The error has been fixed in the revised manuscript.

(e) - In Fig. 4(b), the performance of Q-WSRME seems to be (in average) worse than that of Q-WSRM. However, that should not be the case, since Q-WSRME is taking into account the over allocation. Any reason for this?

Reply: Indeed the performance of the Q-WSRME is better than Q-WSRM in Fig. 4(b). The misleading observation was probably due to the visualization effect where the colors representing the performance of sum arrivals and Q-WSRME are quite similar. We have updated Fig. 4 (and all other figures) in the revised manuscript to avoid such visualization problems.

(f) - p 6, col 2, row 49: it should be $t_{l,k,n}$ instead of $t_{l,n,k}$

Reply: This issue is corrected in the revised manuscript.

(g) - p 10, col 1, row 28: is it λ a dual variable?

Reply: In fact, it should be the dual variable $\sigma_{l,k,n}$ and not $\lambda_{l,k,n}$. We have fixed this problem in the revised manuscript.

(h) - p 10, col 2, row 59: typo wHith

Reply: We have fixed this error.

(i) - In general, a grammar check is recommended, several mistakes with respect to singular and plural nouns have been observed, e.g. -p 6, col 2, row 56 "... for A fixed receiverS" is not correct.

Reply: We thank the reviewer for the comment and we have proofread the manuscript carefully to eliminate all grammar errors.

Response to Reviewer - 4's Comments

This reviewer's concerns have been addressed, and this manuscript is now deemed fit for publication.

Reply: We thank the reviewer for the constructive comments and recommending the revised manuscript for the publication.