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 of the paper, which have been greatly helpful to improve the paper quality. Based on the comments, we have made several major revisions to the paper, following the suggestions of the reviewers. In what follows, the comments are listed, each followed immediately by the corresponding reply from the authors. The revisions in the revised manuscript are highlighted using blue color and the authors responses are also presented in blue color text. Following is the summary of the revision made on the manuscript in accordance with the reviewers comments.

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

In what follows, the comments are listed, each followed by the corresponding reply from the authors. Unless otherwise stated, all the numbered items (figures, equations, references, citations, etc) in this response letter refer to the revised manuscript.

Response to Reviewer - 1's Comments

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

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 understand the reviewer concern. Based on the suggestion, we have removed couple of paragraphs from Section III.A and shortened the discussions on the simulation section to provide additional details for the convergence proof.

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 thank the reviewer for the critical comment. It is not possible to define the inequality with the previous optimal point. Since it is not possible to comment on the inclusion of the previous constraint set in the current update (except the earlier optimal point), the inequality is not guaranteed. Based on the comment from the reviewer, we have removed the inequality specifying the previous operating point from (48), since the newly found optimal point is inside the feasible set.

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: We understand the reviewer concern. Even though the SCA algorithm converges, it is not guaranteed that the iterates involved in the iterative algorithm, namely, $\mathbf{m}_k^{(i)}$ and $\mathbf{w}_k^{(i)}$ to converge. It is true that the iterates need not converge when the objective function is convex. We have modified the discussion on the strong convexity of the objective function to impose the uniqueness of the iterates in each SCA update as well. By regularizing the objective function with a strongly convex term like $\|\mathbf{m} - \mathbf{m}_k^{(i)}\|^2$, we can guarantee the uniqueness of the iterates upon the SCA convergence. We thank the reviewer for citing this issue on the sequence convergence and the uniqueness of the minimizer. We have included this information in the centralized convergence proof on Appendix A-B after (44), after (49) and on Appendix A-C to describe the strong convexity.

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: We thank the reviewer for the pointing out the issue involving the strict monotonicity. As suggested by the reviewer, we have included Appendix A-D to discuss the strict monotonicity of the objective function in each update point of the algorithm.

Let $\mathbf{z}$ be the stacked vector of optimization variables that can either be $[\mathbf{m}, \mathbf{w}_*^{(i-1)}, \boldsymbol{\gamma}]$ or $[\mathbf{m}_*^{(i)}, \mathbf{w}, \boldsymbol{\gamma}]$ depending on the optimization problem that we are trying to solve, *i.e*, either solving for $\mathbf{m}$ and $\boldsymbol{\gamma}$ by fixing $\mathbf{w}_*^{(i-1)}$ or solving for $\mathbf{w}$ and $\boldsymbol{\gamma}$ by keeping $\mathbf{m}_*^{(i)}$ fixed.

Due to the strong convexity of the modified objective $f(\mathbf{z})$, there exist an unique minimizer in each SCA step k. Moreover, in each SCA step k, the feasible set $\mathcal{X}_k^{(i)}$ includes the earlier optimal point $\mathbf{z}_k^{(i)}$, therefore, the objective decreases strictly unless $\mathbf{z}_k^{(i)}$ is a fixed point or a limit point of the SCA iterations [22]. Following the strong convexity of $f(\mathbf{z})$, we have

$$f(\mathbf{z}) \geq f(\mathbf{z}_*^{(i)}) + \nabla f(\mathbf{z}_*^{(i)})(\mathbf{z} - \mathbf{z}_*^{(i)}) + c\|\mathbf{z} - \mathbf{z}_*^{(i)}\|^2, \forall \mathbf{z} \in \mathcal{X}_*^{(i)} \quad (1)$$

where $\mathbf{z}_*^{(i)}$ is a limit point as $k \to \infty$ for the i^{th} alternating optimization (AO) iteration. The equality in (1) is valid iff $\mathbf{z} = \mathbf{z}_*^{(i)}$. Strict monotonicity of $f(\mathbf{z}_k^{(i)})$ is ensured in the SCA steps by the additional quadratic term, which holds with equality at the limit point $\mathbf{z}_*^{(i)} \in \mathcal{X}_*^{(i)}$.

To ensure strict monotonicity of the objective while swapping the optimization variables from $\mathbf{m}, \boldsymbol{\gamma}$ to $\mathbf{w}, \boldsymbol{\gamma}$ in the i^{th} AO iteration, we rely on the discussions to show the inequality (51). Note that the modified objective f is strongly convex, and therefore has a unique minimizer in each SCA step. Additionally, the feasible set $\mathcal{Y}_0^{(i)}$ for $k = 0$ SCA step with the variables $\mathbf{w}, \boldsymbol{\gamma}$ includes the previous minimizer $\{\mathbf{m}_*^{(i)}, \mathbf{w}_*^{(i-1)}, \boldsymbol{\gamma}_{*|\mathbf{w}}^{(i)}\}$, thereby having strictly decreasing objective in (51). The equality in (51) holds for the limit point $\{\mathbf{m}_*^{(i)}, \mathbf{w}_*^{(i)}, \boldsymbol{\gamma}_{*|\mathbf{m}}^{(i)}\}$. Using the above arguments, we can ensure the strict monotonicity of the overall objective sequence.

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 thank the reviewer for raising the concern. We have updated the manuscript to include the details regarding the stationary point discussion in Appendix A-F and the convergence proof analysis for the primal and the alternating directions method of multipliers (ADMM) algorithms in Appendix B. Additional detail includes the conditions required for showing the stationarity of the limit point and we have included more details to utilize the conclusions from [37] to show the convergence of the ADMM scheme.

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.

We thank the reviewer for providing useful comments. The comments are constructive and helped us to improve the manuscript better.

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 discussing the sub-channel ordering guidelines in a detailed manner. As suggested by the reviewer, we have included the guidelines on selecting the sub-channel ordering based on the greedy selection, which could be an educated guess over the random sub-channel ordering. Even though the number of backlogged packets must also be involved in the sub-channel ordering, greedy selection provides an effective ordering when the queues are fairly equal among the serving users or when the number of backlogged packets are significantly large. This information is now included in the revised manuscript under Section III-D final paragraph. We have included a plot comparing different ordering schemes to justify the above statement. Even though we are not including it in the paper, we have included in the response letter to justify the reviewers claim. We considered a system model with $N = 6$ sub-channels, $N_B = 2$ base stations (BSs) with $N_T = 4$ transmit antennas and $K = 12$ single antenna users. The path loss (PL) is distributed uniformly over $[0, -3]$ dB.

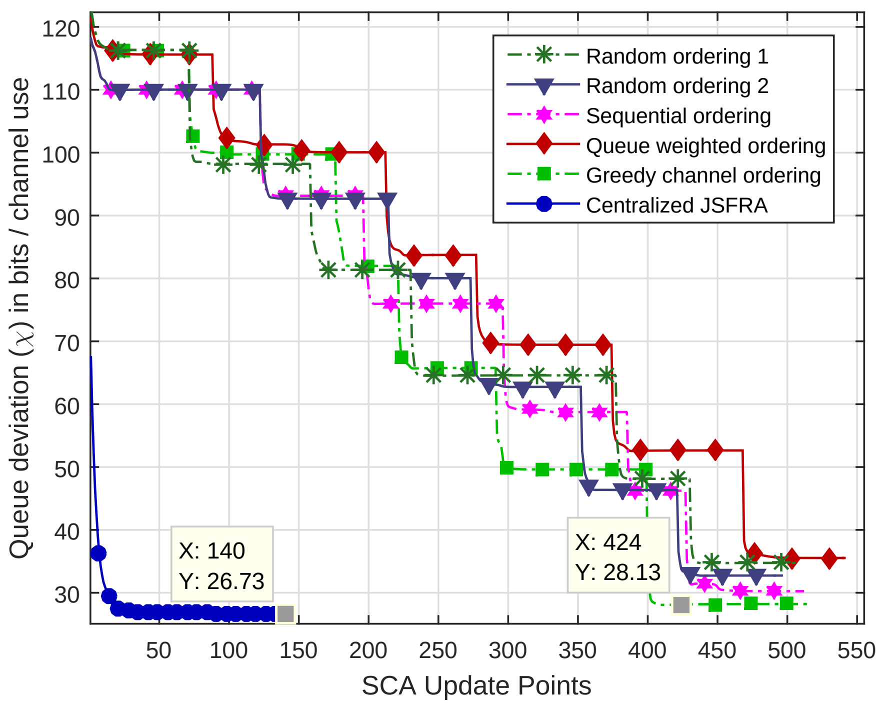

Figure 1: Convergence of the algorithms for $\{N, N_B, K, N_T, N_R\} = \{6, 2, 12, 4, 1\}$ using ℓ_1 norm

Fig. 1 compares different ordering of sub-channels with the total number of backlogged packets remaining in the system as the metric. The random ordering are based on permutation of the sub-channel indices and sequential ordering is the natural way of selection. The greedy sub-channel ordering is based on sorting the best channel from each sub-channel, which is obtained by finding the highest channel norm between the users from the respective serving BS. As pointed out by the reviewer, greedy scheduling performed much better when compared to the other random ordering schemes.

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.

(a) Quantifying the signaling overhead of the distributed algorithm over the centralized one depends on the system model under consideration. For example, let us consider a model with $N = 1, N_B = 2, K = 4, N_T = 4, N_R = 1$, where each BS has two users. In this scenario, the amount of information exchange to perform a centralized algorithm by a common controller requires the knowledge of complete channel matrices interlinked in the system, *i.e*, the number of users times the BSs. In order to quantify the total number of bits required to be exchanged, let us assume that each complex channel for a single-input single-output requires 10 bits, *i.e*, 4 bits for amplitude and 6 bits for phase (assuming phase is important). Using this assumption, the total number of channel information in bits to be exchanged via backhaul requires $10 \times K \times N_B \times N_R \times N_T \approx 320$ bits. On the other hand, for the distributed case, let us consider 6 bits are required to quantize the scalar interference in the consensus vector. Therefore, it requires $6 \times 2 \times 2$ bits of information to be exchanged for each iteration. With this overhead, we can perform only 6 SCA updates with $J_{\max} = 2$ internal iterations in each to reach the centralized overhead. Note that the number iteration includes the SCA update as well. However, by knowing the complete channel of the users in the system, the centralized controller can perform the precoder design until convergence. However, as the number of sub-channels, users or the antenna element increases, it may not be a feasible option to feedback the user channels across the coordinating BSs to the centralized controller. The quantizing the channel information degrades the achievable performance of the centralized solution that also needs to be considered while comparing the signaling overhead. Moreover, in the centralized algorithm, resulting transmit precoders need to be exchanged with the corresponding BSs before the actual transmission, involving huge overhead in the backhaul.

(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) In particular, ADMM and primal approach requires significant overhead compared to the centralized algorithm when $J_{\max} = 1$ for a small system, since by exchanging the quantized channels, each BS can perform the centralized algorithm independently until convergence. However, it depends on the channel quantizer, which is likely to be based on the channel density function (eg. Lloyd quantizer). For a system involving more number of coordinating BSs, users and number of antenna elements, it would be beneficial to use distributed algorithm with $J_{\max} > 1$ to have a strictly monotonic decrease in the objective.

(d) Even with the same signaling overhead, it is not possible to prove the benefits of the distributed algorithm like ADMM and primal approach over the independent centralized scheme performed by exchanging the channel state information across the BSs. It is dependent on the system model under consideration. As explained earlier, for the considered system model, the independent centralized algorithm performs better in terms of the over head involved.

We have included the above information briefly in Section IV-C first paragraph and in the last paragraph. We have also included this information in the paragraph before the Conclusions section. Having said that, note that in reality, the channel is time-correlated, and it is often enough to update the precoders once per radio frame or certain number of symbols later. It is not necessary for the decentralized algorithm to converge until the end, it is only important to follow the fading process when $J_{\max} > 1$. The performance of the distributed algorithm based on dual decomposition scheme is discussed for the time-correlated fading in Section C of [13], which shows that it is enough for the distributed precoder design to follow the fading process to provide desired performance. The distributed algorithm for the time correlated case is not provided in the current manuscript, since it is not in the scope of the precoder design algorithms considered in this manuscript. However, for the clarity purpose, we have provided a plot demonstrating this behavior for the Karush-Kuhn-Tucker (KKT) based algorithm in Section IV-C presented in the manuscript.

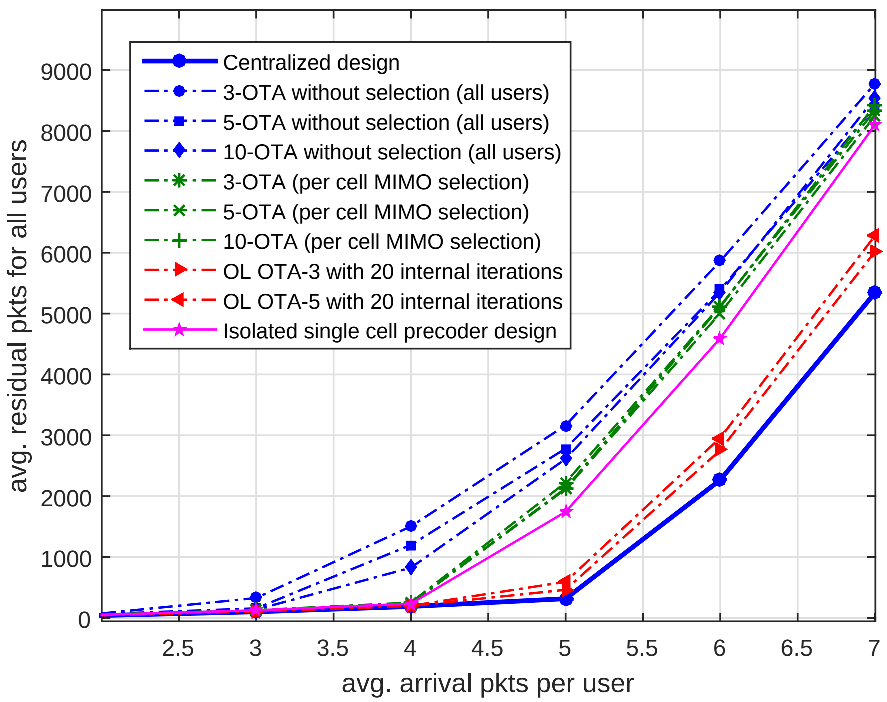

Figure 2: Average number backlogged packets after each transmission for system $\{N, N_B, K, N_T, N_R\} = \{4, 2, 16, 4, 2\}$ evaluated for 500 slots

Fig. 2 compares the performance of the distributed KKT approach presented in Section IV-C for different iteration. The signaling requirements are outlined in Algorithm 3 and the overhead involved in the signaling is penalized in the achievable rate of the users. We considered that the channel is coherent over $N_S = 100$ symbols or frames and the precoder update is performed by exchanging the equivalent channel for $J_{\max} = 3, 5, 10$ number of iterations. 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. The average number of backlogged packets after each transmission slot is evaluated as

$$\chi = \sum_{k=1}^{K} [Q_k - \tilde{t}_k]^+ \tag{2}$$

Unlike the distributed algorithm, the centralized scheme presented in Fig. 2 has no penalty term and it is used as a benchmark for the other figures. 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 starting initializing randomly. Since we use KKT approach, we can either use all users in the system for the precoder design or we can utilize single-cell MU-MIMO user selection presented in the literature to limit the number of users for which the precoders are designed, which leads to the faster convergence. As we can see from Fig. 2, as the arrival rate per user increases, the performance of KKT schemes with $J_{\max} = 3, 5, 10$ converges since the number of backlogged packets are significantly large, therefore, the same set of users will be served by the algorithm with better precoders by utilizing the memory.

In spite of using memory and prior scheduling in the KKT approach, isolated single BS processing performs much better than the distributed scheme due to the limited number of iterations allowed in the algorithm. Note that the precoders are not updated for the desired users until convergence, after the limited number of iterations. However, if we perform the single cell precoder design by considering the neighboring precoders as fixed after the recent exchange as discussed in [28], we can improve the performance significantly for Algorithm 3 as shown by red curves in Fig. 2. In this approach, in between each exchange across the coordinating BSs, each BS will perform $J_{\max} = 20$ with the neighboring precoders as fixed. Once the iterations are performed to update the precoders, it is then exchanged across the coordinating BSs to perform the same procedure as mentioned earlier.

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 important concern regarding the convergence issue of the distributed algorithm with limited number of iterations. In view of this, we have updated the manuscript to include the discussion on the convergence of the distributed algorithms with limited number of iterations in Appendix B third paragraph after (57). Since the distributed algorithm cannot be guaranteed to be monotonic in each iteration, it is not possible to prove the convergence of the algorithm. However, if the algorithm is allowed to converge or iterated to guarantee the monotonicity of the objective, it is possible to prove the convergence based on the discussions provided in the manuscript.

- The fixed points for each convex subproblem exist, but may not be unique unless the objective function is regularized with a strongly convex term as discussed in Appendix A-C.
- If the distributed algorithm is iterated for limited number of iterations, it is not guaranteed to achieve the fixed point even if the outer SCA update is performed for large number of iterations. In this scenario, it is not guaranteed to converge. In all our simulations on the primal and the ADMM approach, we have iterated $J_{\max} = 20$ in order to guarantee the monotonicity of the objective.
- Unless the objective function is regularized with a strongly convex term as in Appendix A-C, the uniqueness of the iterates is not guaranteed.
- In each SCA update, if the distributed algorithm converges to a fixed point, then the overall convergence will be a stationary point of the original nonconvex problem. It may not be an optimal solution to the original nonconvex problem. It can be proved based on the strict monotonicity of the objective, since the distributed algorithm converges to a centralized solution.
- Since the coupling between the distributed precoder designs are the interference between the BSs and the users, in the ADMM approach, the interference is treated as a local variable, which is then included in the precoder design problem for each coordinating BS. This is treated as a local variable for a specific BS. Note that the local variable is an assumption made by the BS on the actual interference caused by the neighboring BSs. Since the actual interference caused is different, the consensus has to be made between the local interference variable maintained at each BS with the global consensus interference variable, which is nothing but the average between

the corresponding BSs interference. These discussions have been made in the revised manuscript in Section IV-B. For further details we have also referred the interested reader to [11], which discusses exclusively about the ADMM approach. Upon convergence of the ADMM approach, the interference vector can be considered as an upper bound of the actual interference seen in the network. It can be equal to the actual interference but not less.

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:

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 understand the reviewer concern. The initial operating point for the SCA approach can be identified by initializing the transmit precoders $\mathbf{m}_{l,k,n}$ with the single user transmit beamformers satisfying the total power constraint at each BS. Once the transmit precoders are initialized, the receive beamformers $\mathbf{w}_{l,k,n}$ can be initialized with the minimum mean squared error (MMSE) receiver based on (23). Upon initializing both transmit and the receive beamformers, the interference term $\beta_{l,k,n}$ can be initialized from (20b). This information is included in the revised manuscript in the paragraph on Section III-B following after (23).

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 updated the manuscript further for better clarity and understanding based on the other reviewers.

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 understand the reviewer concern. We have updated the manuscript to include a discussion on the choice of selecting an algorithm for the practical purposes. The choice of selecting a centralized algorithm is equally fair when $N_R > 1$. For a single antenna receiver, the joint space-frequency resource allocation (JSFRA) formulation in Section III-B is efficient, as there is no receiver update, and therefore has less complexity. The distributed algorithms based on primal and ADMM schemes are favorable when $N_R = 1$ due to the limited signaling between the coordinating BSs, involving the scalar interference values. However, when $N_R > 1$, the KKT based distributed approach is much more efficient, since it has less signaling overhead for a given throughput. The choice of the ℓ_q norm is as discussed in Section II-B. These lines are included in the paragraph before the Conclusions section. If the reviewer is interested on the practical usage of the algorithms, please refer to the response for the reviewer 2's question 2 in this response document for more details involving the performance of the KKT based distributed algorithm.

(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 is rewritten to provide additional clarity as suggested by the reviewer. We have included the reference [28] for more illustrative discussions on the backhaul signaling.

(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: We understand the reviewer concern. We used different path loss models for improved clarity and to show that the proposed algorithms works reasonably well on all scenarios. To provide a fair comparison, we have also included the centralized algorithm to draw the difference between the other schemes. We will consider the reviewer comment in our future work. However, if the reviewer insists on the unanimous configuration across the figures, we will revise the manuscript accordingly.

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

Reply: We apologize for the error. We have modified the text to agree with the simulation Fig. 1.

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

Reply: The performance of the Q-WSRME is better than Q-WSRM in Fig. 4(b). Since we have plotted the instantaneous arrival rates also in the same figure, it is difficult to discriminate between the colors. We have updated the manuscript to provide better clarity in the figure.

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

Reply: We have update the manuscript with the proper subscript.

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

Reply: We thank the reviewer for pointing out the mistake. It is the dual variable $\sigma_{l,k,n}$ and not $\lambda_{l,k,n}$. We have updated the manuscript accordingly.

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

Reply: We have updated the manuscript with the proper word.

(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 have updated the manuscript with the singular tense and we have also verified the grammatical validity of the manuscript as per the reviewer suggestion.

Response to Reviewer - 4's Comments

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

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