We thank all three reviewers for their constructive comments. We address them below one by one.

To Reviewer 1 (R1).

Q1: *what makes it nontrivial to extend the regularity condition and proof technique in [11] to Riemmanian optimization.*

A1: The key contribution of our work is not to extend sharpness and weak convexity to the Riemannian regularity condition (RRC), but rather to use the RRC to analyze the convergence of the projected Riemannian subgradient method (PRSM) and to demonstrate the applicability of the RRC to problems in robust subspace and dictionary learning. The result in [11] only works for convex sets that have a nice property such as their orthogonal projector being non-expansive. The Grassmannian manifold is nonconvex, making the analysis more complex. We have exploited specific structure in the Grassmannian and used a Riemannian subgradient instead of subgradient (which is used in [11]) to get our results.

Q2: *For DPCP, the contribution can be bigger if a global convergence guarantee is provided using some initialization (e.g. the spectral initialization analyzed in the appendix)*

A2: We do have a deterministic and a statistical analysis for the spectral initialization in the supplementary material; see Lemma 3 and Corollary 3, which are not mentioned in the manuscript. In particular, Corollary 3 implies that in a random model, the spectral method provides a valid initialization with high probability when the number of outliers is smaller than the square of the number of inliers. We will incorporate these into a revised version of the manuscript.

To Reviewer 3 (R3).

Q1: *The focus on the grassmannian limits the scope to some extent. The regularity condition seems like it can be extended to arbitrary submanifolds of Euclidean space. Is there anywhere using the special structure of the grassmannian?*

A1: We agree the definition of the RRC can be extended to arbitrary submanifolds of a Euclidean space with an appropriate definition of the Riemanniann metric and distance. However, using the RRC to analyze convergence of optimization methods requires exploiting the specific properties of the submanifold. In our paper the Grassmannian structure is utilized together with the RRC to analyze the convergence of the projected Riemannian subgradient method. Specifically, an important property of the Grassmannian manifold is that if a point $\boldsymbol{B} \in \mathbb{R}^{D\times c}$ is outside the Grassmannian (i.e., $\sigma_c(\boldsymbol{B}) \geq 1$), then projecting it onto the Grassmannian will not increase its distance to any other point in the Grassmannian with the distance defined in (2). If other manifolds also have such property, then the current convergence analysis can also be applied. We will incorporate this discussion in the final version.

Q2: *It would be helpful to clarify the assumptions that are needed to ensure that (4) is the Riemannian subdifferential.*

A2: We really appreciate this comment. The reviewer is correct that for a general nonsmooth function, the current (4) may not be the Riemannian subdifferential. To be more rigorous, in the revision we will directly define the Riemannian subdifferential using the Clarke subdifferential (which is more general than the Fréchet subdifferential) for locally Lipschitz functions on Riemannian manifolds [2]. As pointed out by the reviewer, the analysis still holds as long as the Riemannian regularity condition holds. According to [A], if a function is **regular**, then the Riemannian subdifferential based on the Clark subdifferential is equal to the projection (onto the tangent space) of the Clark subdifferential. Since both the robust subspace learning and dictionary learning problems are regular, their Riemannian subdifferentials computed in Section 4 are correct. We will incorporate this discussion in the revision.

[A] Yang, Zhang, Song. "Optimality conditions for the nonlinear programming problems on Riemannian manifolds." Pacific Journal of Optimization, 2014.

Q3: *The RRC here is not a generalization of sharpness and weak convexity, but rather of the consequence (9).*

A3: The referee is correct, and we will rephrase this sentence in the final version, if accepted.

Q4: *It is not clear that this condition has been phrased in an intrinsic manner; calling it RRC may introduce confusion.*

A4: We called our condition a Riemannian Regularity Condition because it involves the Riemannian subgradient. However, we agree that this definition is extrinsic. We will rename our definition as Extrinsic RRC to avoid confusion.

To Reviewer 4 (R4).

Q1: *Line 84: we usually think of the projection of $\boldsymbol{B}$ onto $\boldsymbol{A}$ as an operation on $\boldsymbol{B}$.*

A1: Since we project $\boldsymbol{B}$ onto $[\boldsymbol{A}]$, here $\boldsymbol{A}\boldsymbol{Q}^\star$ represents a point in $[\boldsymbol{A}]$. Of course, the reviewer is correct that $\boldsymbol{Q}^\star$ contains a nonlinear transformation of $\boldsymbol{A}^\top\boldsymbol{B}$. We will incorporate this discussion into the final version, if accepted.

Q2: *Line 118: Is it more appropriate to say this gives a bound on the sum of the cosines of the principal angles?*

A2: This is a great suggestion and we will incorporate it into a revision.

Q3: *Section 3.2: it maybe better to explicitly state that by projection onto Grassmannian you mean orthonormalization.*

A3: This is a great suggestion. We will make this statement in a revision of the paper.

Q4: *It seems to me that the convergence results should hold even if the steps are taken along a geodesic.*

A4: In this paper we take an extrinsic approach because extrinsic methods are typically easier to implement, e.g. when the projection map is easier to compute than the geodesic distance. Extending the current analysis to an intrinsic optimization method is definitely worth exploring and will be the subject of future research.