We greatly appreciate the reviewers' effort and helpful comments. We will fix the typos and polish the writing by incorporating the reviewers' suggestions.

Response to Reviewer #1

Comment 1: "The significance of the proposed method is not very clear..."
Response 1: First, the question of solving saddle-point problems using only projection-free methods is interesting (Reviewer #3 also mentions this point). It also has great theoretical significance in the optimization area.

Secondly, though our analysis is specified for the convex-strongly-concave setting, there is a simple way to adopt our algorithms to solve the general convex-concave saddle point problems. For a convex-concave function $f(\mathbf{x}, \mathbf{y})$, we can construct a convex-strongly-concave function as $f_\epsilon(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}, \mathbf{y}) - \epsilon\|\mathbf{y} - \mathbf{y}_0\|^2$ and solve $f_\epsilon(\mathbf{x}, \mathbf{y})$ by our algorithms (such as MPCGS). Though the convergence rate of this method could be suboptimal, it's a practical way to deal with the general convex-concave situations.

In addition, [6] shows some examples of saddle point algorithms where projection onto the constrain sets is hard. These applications includes robust optimization, two-player games and sparse structured SVM.

Comment 2: "Why do we consider nuclear norm constraint for this classification problem?"
Response 2: Nuclear norm is a popular penalty in multi-class classification because datasets with many categories usually exhibit low-rank embedding of the classes behaviour (see [4]).

Comment 3: "(arXiv:1804.08554, section 5.4 and 5.6) can be added."
Response 3: We find that this paper does not have section 5.4 and 5.6. Also, it is irrelevant to our paper. Perhaps you give the wrong paper id.

Comment 4: "The presentation is mostly clear, but some parts are lacking important details."
Response 4: We will modify the confused sentences and clarify our results.

Comment 5: "Line 116: the linear optimization on Xc only needs to find the top singular vector of X, which only costs O(nnz(X)) time. This statement is inaccurate."
Response 5: You're right. The complexity should be $\tilde{O}(\frac{N}{\sqrt{\epsilon}})$, where N is the number of non-zero entries in the gradient.

Response to Reviewer #3

Comment 1: "It is not clear why the assumption that the objective is strongly concave is needed."
Response 1: Notice that we adopt CGS algorithm to approximately solve a concave problem in Alg 4 (line 3). When the objective is strongly concave, the CGS method only requires to call $\sqrt{\kappa}\log(1/\epsilon)$ SFO. When the objective is not strongly concave, the CGS method requires to call $1/\sqrt{\epsilon}$ SFO. The convergence rate of CGS will significantly influence the total number of iterations of our algorithm because CGS is performed in the inner loop.

Comment 2: "It seems that the bounds are loose at several points."
Response 2: For our algorithms, we think our bounds are almost tight. We think that there exists better algorithms which only requires to call $O(1/\epsilon)$ LO as the projection-free algorithms for the convex optimization, but finding such an algorithm is a big challenge because minimax problem is much more complicate than the minimization problem.

Comment 3: "Line 4 of Alg 3: not clear what we get v_k here as one of the outputs of prox-step if its updated in the following line via CndG"
Response 3: Actually, we do not compute v_k via CndG. We only update x_k,y_k and v_k by the prox-step. According to Alg 3 (the procedure of prox-step), the results of the prox-step guarantee that x_k,y_k and v_k satisfies the equations and inequality in the Line 4 of Alg 3.

Response to Reviewer #6

Comment 1: "L40 is a bit of an over-claim".
Response 1: We will modify the over-claim sentences and clarify our setting. On the other hand, there is a simple way to adapt our methods to the convex-concave setting (see the second paragraph of the Response 1 to Reviewer #1).

Comment 2: "I am a bit confused about Remark 2. since when ϵ is small we could have $\log(1/\epsilon) \gg \sqrt{\kappa}$. Moreover, isn't the condition you would like to require $\sqrt{\kappa}/\epsilon \gg \kappa^2$?"
Response 2: The condition should be $2^{-\sqrt{\kappa}} < \epsilon < \kappa^{-1.5}$. Then we can get $(\sqrt{\kappa}/\epsilon + \kappa^2)\log(1/\epsilon) < \kappa/\epsilon$.

Comment 3: "SVRE has no-guarantees in the convex-strongly-concave setting."
Response 3: To our knowledge, there is no stochastic projection algorithm has guarantees in the convex-strongly-concave setting. On the other hand, we have already took a nuclear norm regularization. Usually it does not need additional L2 regularization.