We thank the reviewers for their insightful comments. Below please find our responses to the major points raised.

To Reviewers 1: We appreciate your very detailed and thoughtful comments.
Q1: Quantify the relative contributions of the golden section search and the LP-ADMM to the runtime speedup.
A1: The use of the golden section search alone will not lead to any substantial speedup, as the main computational burden lies in the β-subproblem. In general, standard solvers (such as interior-point methods) cannot exploit the structure of the β-subproblem. Thus, even when combined with the golden section search strategy, they still cannot achieve the speedup obtained by our proposed LP-ADMM. Furthermore, the golden section search enjoys the $\Theta(\log \frac{1}{\epsilon})$ complexity, which is already optimal in the information-theoretic sense. Therefore, any other univariate search methods can at best achieve similar complexity.

Q2: Strengthen Section 5.2 with further empirical evidence of faster convergence in a variety of settings.
A2: Thank you for your kind reminder. As mentioned in the paper (line 235-237), we did the comparison for the real datasets (i.e., UCI Adult) but found that all baseline methods cannot achieve the desired accuracy (i.e.,$||x_k - x^*|| \leq 10^{-6}$). Due to the ill-conditioned data matrix, all baseline methods require an extremely careful choice of hyper-parameters. That's why we did not include the details. Based on your advice, we will add back the comparison in the revision.

Q3: The change in notation from the DRLR sections (1, 3, and 5) to the generic LP-ADMM sections (2 and 4) is potentially confusing. Is there a way to make the notation consistent?
A3: Thank you for your suggestion. We will unify our notation for clarity in the revision.

Q4: Section 5.3 bears little relation to the rest of the paper ...... Should it be omitted entirely?
A4: Thank you for your comments. We want to further verify the power of DRO modeling for the large-scale datasets, which is different from Table 1 in [1]. That is why we include it.

Q5: The choice of inner solvers for β-update in LP-ADMM for different settings (illustrate the difference empirically?)
A5: Thank you for your advice. We will add them in the main text in the revision for clarity.

To Reviewers 2: Thanks for appreciating the contributions of our work. Thanks for the suggestion on the literature review. We have discussed [1] in our revised manuscript.

To Reviewers 3:
Q1: Can you prove results and conduct experiments for different norm constraint on beta, such as ℓ_1 or ℓ_2? It would benefit the reader if the result is stated for the general norm for beta.
A1: Thank you for your suggestion. Absolutely yes. Firstly, for an upper bound on optimal λ for ℓ_1 and ℓ_2 cases, we already have the same upper bound whose proof is just a slight modification of the current one in the appendix for the ℓ_∞ case. In details, we replace the ball constraint $||\beta||_1 \leq \lambda$ by the equivalent one $B\beta \leq \lambda e_{2^n}$ where B is the $2^n \times n$ matrix whose rows are all the possible arrangements of $+1's$ and $-1's$, and $||\beta||_2 \leq \lambda$ by $||\beta||_2^2 \leq \lambda^2$. Other steps are the same as the ℓ_∞ case. For the convergence analysis of our LP-ADMM, the convergence result has already covered the general norm setting. We will add it in the revision.

Q2: Test performance on smaller datasets, and demonstrate it worths solving $\kappa < \infty$, comparing to $\kappa = \infty$.
A2: The test performance on smaller datasets has already been done in the previous work, see Table 1 in the reference [21]. Thank you for your suggestions. We will better motivate the results and clarify their relationship with the literature in the revision.

Q3: What is the information-theoretic lower bound of the DRLR problem? This is related to the last sentence in the conclusion section. Discuss whether the algorithm achieves the optimal complexity.
A3: Thank you for pointing out the interesting research direction. We mentioned it in the paper, see Remark 6.8 in the appendix and reference [31]. We have proved that the β-subproblem (1.2) enjoys the Luo-Tseng error bound. Thus, the optimal local convergence rate is linear for first-order algorithms theoretically. However, only a sublinear rate (i.e., $\mathcal{O}(\frac{1}{K})$) has been established in our paper since it is open whether the primal-dual error bound holds for the problem (1.3). Under the ADMM framework, to the best of our knowledge, this is the best complexity bound to date. However, it remains open whether one can prove that our proposed LP-ADMM or some other algorithms can achieve the optimal complexity (i.e., linear rate) for the β-subproblem.

Q4: The relationship between the KKT point of (2.1) and the global solution of the original problem? In particular, if we get an eps-optimal solution of (2.1), what is the optimality gap of the original DRLR problem?
A4: Firstly, the KKT point of the problem (2.1) (i.e., x^*) is the corresponding optimal solution of the original problem. Furthermore, if $||(x_k, y_k, w_k) - (x^*, y^*, w^*)|| \leq \epsilon$, then we have $||x_k - x^*|| \leq ||(x_k, y_k, w_k) - (x^*, y^*, w^*)|| \leq \epsilon$. Then, we have the ϵ-optimal solution of the original problem.