We would like to thank all the reviewers for their efforts in reading the manuscript and providing feedback. The reviewers have appreciated our central objectives such as bridging DRO to regularized machine learning in a "unified view" (R2), providing "foundation for lots of empirical work" (R1), and to "inspire future work connecting the plethora of DRO uncertainty sets" (R2) which are "highly relevant to NeurIPS" (R2). We particularly thank the reviewers for making us aware that we failed in explaining two aspects of our work with sufficient clarity, which we are happy to include in the updated version. We state these here and will reference them in our individual reviewer rebuttals below. We will be using citation references from the main submission.

1 **Concrete directions for practitioners:** Although this is deferred to future work, an immediate consequence in this direction would be robust certification, based on the black-box verification framework in [14], which is briefly mentioned at the end of Section 3. We outline how Corollary 1 directly implies a certificate for practitioners: Given a binary classifier and reference distribution ρ, one can compute $\mathbb{E}_{\rho(X)}[h(X)] - \epsilon\Theta_{\mathcal{F}}(-h)$ and check if this value is ≥ 0. Using Definition 2.2 of [14] and Corollary 1 of our work, if this value is ≥ 0 then this certifies that the classifier is robust to $\mathcal{F}$-IPM perturbations around ρ. This follows from the fact that Corollary 1 (using $-h$) implies $\mathbb{E}_{\rho(X)}[h(X)] - \epsilon\Theta_{\mathcal{F}}(-h) \leq \inf_{Q \in B_{\epsilon,\mathcal{F}}(\rho)} \mathbb{E}_{Q(X)}[h(X)]$ and positivity of the term on the right is precisely the condition laid out in Definition 2.2 of [14]. We will include this immediate consequence in the updated manuscript.

2 **Relevance of the GAN robustness results:** The main takeaway from the results presented in Section 4 is to advocate the use of regularized discriminators when training GANs. In particular, we show that the generative distribution learned using regularized discriminators gives guarantees on the worst-case perturbed distribution (robustness). This is particularly relevant for the robustness community since lines of work [55, 8, 60, 59, 28, 26, 41, 47, 48, 24, 57, 42] implement GANs as a robustifying mechanism by training a binary classifier on the learned GAN distribution. In light of our results, learning a binary classifier using a GAN (trained with regularized discriminators) as a downstream task implies this classifier will consequently be robust. For the GAN community, our finding complements existing empirical evidence that shows benefits of regularized discriminators such as the Wasserstein-, MMD-, and Sobelov-GAN and other discriminator regularizers outlined in [19]. Furthermore, another subtle benefit of Theorem 3 is it shows how a DRO result can be applied to the GAN objective. This paves the way for future developments of DRO to be applied to GANs and consequently helps bridge these two communities.

Reviewer 1: Thank you for your feedback and pointing out how our work provides foundation for many empirical work and the importance to be added to the literature. Regarding your points on the usefulness, we have made some headway on how a robust certificate is immediate in **1** and outline the relevance of Section 4 to both the GAN and robustness community in **2**. We really appreciated your feedback, which will be included in the updated version, and are confident that it will improve the paper by virtue of the changes outlined in **1+2**.

Reviewer 2: Thank you for your feedback and support of the unified view we present. Regarding the significance of the GAN section, we have outlined this in **2**. More specifically, the motivation for studying GAN robustness comes from lines of work that use the distribution learned by a GAN to train a classifier or to attack (cited above in **2**). In this context, our results allow us to understand how robust a GAN is and what makes them more robust. In particular, our results link this to regularizing the discriminator - validating methods that use GANs for these purposes. We appreciate your comment on pointing out the motivation and to hint on this earlier, which we believe will strengthen our paper. Thank you for the definitions we missed and related work regarding the DRO results, we will amend the statement and include these references. We will also include the paper you mentioned, which focuses on Wasserstein distances and supplements the use of restricted discriminators in Wasserstein GANs (WGAN). In contrast, we develop results for the IPMs (including Wasserstein) and supplement a large family of existing GANs that use restricted discriminator sets (including WGAN). Indeed, since well-known IPMs are of the form $\{h : \zeta(h) \leq 1\}$ (such as Wasserstein distance, Total Variation, MMD, Dudley, etc.), we will take your advice of delegating the general statement to the Appendix as this is only a slight change in notation yet retains the generality of the story and improves presentation - thank you.

Reviewer 3: Thank you for your feedback and support of the paper, including comments regarding the novelty and improved tightness of the results we present. Indeed, in terms of empirical connection, the main insight of our results is to supplement existing empirical work that use regularized discriminators in GANs (for example Wasserstein-, MMD- and Sobolev-GANs) , and contributing more largely to this narrative of robustness through regularization. A more direct practical ramification is outlined in **1** above and **2** for more detail regarding the takeaway from our GAN result, such as the robustness guarantees of a classifier trained using GANs.

Reviewer 4: Thank you for your feedback and support of the paper. The ‘untried robustness perspectives’ refers to linking regularization towards robustness. Indeed, the ‘positive results’ refer to the validity of the approaches. Regarding clarity of our GAN results, it is to provide theoretical support for many popular practices in GAN development as you mention in (1), which we will clarify in the updated version as outlined in **2**.