Reviewer 1: Q1: This paper should provide experimental results?
A: We believe it is important to deliver the message that our proof is novel that addresses the open problem for strongly-convex-strongly-concave minimization. Hence, we emphasize on theoretical analysis in this paper. On the other hand, previous studies have provided the numerical experiments on the state-of-the-art algorithms that are highly related to our Epoch-GDA, e.g., [36,32]. There is also a following-up work that uses a similar idea and has promising experimental results [ref1]. We will consider adding some experiments in the long version.

[ref1] Guo et al. "Fast Objective and Duality Gap Convergence for Non-convex Strongly-concave Min-max Problems". arXiv 2020.

Q2: Key difference of algorithm and analysis between the proposed Epoch-GDA and Epoch-GD?
A: The update of Epoch-GDA can be seen as a primal-dual variant of Epoch-GD. In terms of analysis, there is key difference between Epoch-GDA and Epoch-GD. In particular, Epoch-GD bounds the primal gap, while Epoch-GDA bounds the duality gap. Note that bounding the duality gap of a min-max problem is fundamentally more difficult than bounding the primal gap of a minimization problem. The difference between our analysis and that of Hazan & Kale is very subtle. Particularly, in Hazan & Kale, they used the fact that the primal gap at a solution x_{k+1} from stage $k+1$ can be bounded by the distance between a solution x_k from stage k and the optimal solution i.e., $\|x_k - x_*\|$, which can be further bounded by the primal gap at x_k using strong convexity. However, in our case, the duality gap at a solution (x_{k+1}, y_{k+1}) from stage $k+1$ cannot be bounded by the distance between (x_k, y_k) from stage k and the optimal solution. Instead, they are bounded by the distance from (x_k, y_k) to the corresponding optimal solutions to the minimization and maximization defined at (x_{k+1}, y_{k+1}) , i.e., $\|x_k - \hat{x}_R(y_{k+1})\|$ and $\|y_k - \hat{y}_R(x_{k+1})\|$ (cf. the key Lemma 3). More importantly, we have to show that this distance is strictly less than the imposed radius R such that adding the bounded ball preserves the duality gap of the original problem. Please note that such interior-point argument is very important and is not necessary in Hazan & Kale.

Q3: Why avoiding deterministic updates as in [36,32]?
A: The reason is two-fold: (i) the deterministic updates in [36, 32] require a specific form of objective function; hence by avoiding deterministic update we are able to handle more general problems without scarifying the complexity; (ii) the deterministic updates in [36, 32] have additional computional overhead, which usually needs to pass all data in machine learning applications. A key difference from [36, 32] is that we use the recursion on the duality gap as for convergence analysis, while [36,32] use the recursion on the primal objective gap for convergence analysis.

Reviewer 2: Q1: Key technical contribution (Lemma 1) is simple to prove, so unclear why it is important to extend Epoch-GD for SC min problem to Epoch-GDA for SCSC min-max problem.
A: We agree Lemma 1 is simple to prove. But **the key for proving the fast rate of duality gap for SCSC problems lies at Lemma 3**, which proves that the duality gap of the problem defined with the ball constraint is equal to the original duality gap. This proof is subtly different from that of Epoch-GD. Please refer to response to Q2 of reviewer 1.

Reviewer 3: Thanks for pointing out the relevant reference. We will add it in the revision.
Q1: Why we always have dist$(0, \partial P(\hat{x}_\tau^*)) \leq \gamma \|\hat{x}_\tau^* - x_0^\tau\|$ in Theorem 2?
A: Thanks for pointing this out. Indeed, we should use $\hat{P} = P + \mathbb{I}_X$ in place of P in the above inequality, where $\mathbb{I}_X$ is the indicator function of the set X, which gives us the desired result. To prove this, let us consider an unconstrained ρ-weakly convex function $\psi(x)$ and a reference point $\tilde{x}$, $f(x) = \psi(x) + \frac{\gamma}{2}\|x - \tilde{x}\|^2$ is $(\gamma - \rho)$-strongly convex. Hence, we have the unique optimal solution of $\min_x f(x)$, say $\hat{x}$, then the optimality condition gives that $0 \in \partial\psi(\hat{x}) + \gamma(\hat{x} - \tilde{x})$, i.e., $\gamma(\tilde{x} - \hat{x}) \in \partial\psi(\hat{x})$, which means dist$(0, \partial\psi(\hat{x})) \leq \gamma\|\hat{x} - \tilde{x}\|$. Applying this argument to $P(x) + \mathbb{I}_X$ leads to the corrected inequality.

Q2: Theorem 1 provides a high probability result, while Theorem 2 proves a bound in expectation?
A: Thanks for noticing this difference. We prove the expectation result for WCSC in Theorem 2 for consistency with previous results [32]. Indeed, we followed your suggestion and found that Thm. 2 can be also extended to high-probability result. The key idea is similar to that for proving Thm. 1. In particular, we can prove a high-probability result of Lemma 4 similar to Lemma 2. Then by appropriately setting the radius R_k according to η_k and T_k we can able to prove a similar result as in Lemma 3, which leads to a high-probability upper bound for the duality gap of $f_k(x, y)$. From this point, we can prove the high-prob convergence for the WCSC similar to the existing proof of Thm. 2 except replacing expectation result with high-probability result. We will discuss this in the revision.