We thank all the reviewers for their valuable comments and overall positive feedback. We will fix typos and improve our presentation in the next version.

Reviewer #1:

Thank you very much for your careful reading and recognition our work. We will make modifications to the paper as suggested. In particular, we will provide the precise statements of the local assumptions in the main text if space permits and replace the qualitative description in our proof (e.g., large enough, proper choice) by formal mathematical statement whenever possible. We will also explain that the so-called covariance learning problem here is just a very simple GAN model to avoid confusion. The other points will be addressed similarly. Concerning the quantity Γ: we mean that when Γ is small enough those inequalities are verified and therefore there exists such Γ satisfying the desired properties. Finally, the goal of Appendix B is to show that the same conclusion may hold for optimistic gradient as well by adequately choosing the output vector. We also attempt to give an explanation of some experimental results observed in the paper of Ryu et al. (2019) through the lens of DSEG.

Reviewer #2:

1. In terms of the relation of our counterexample to the one of Chavdarova et al. (2019): The purpose of this counterexample was to motivate the analysis to come, it was not intended as a separate contribution, nor labeled as such. We will discuss in more detail the exact differences between our setting and that of Chavdarova et al.
2. Regarding the paper of Loizou et al. (2020): We thank the reviewer for pointing us to this paper. At the same time, we would like to point out that this paper was accepted to ICML in June 2020, and it only appeared on arxiv on July 8, 2020. We will be happy to cite and discuss it, but mentioning it as a "weakness" is unfair: our work cannot be compared to a paper that was not even available on arxiv at the time of abstract registration at NeurIPS.
3. Strong monotonicity is a standard notion, but we will provide a reference to the standard textbook of Facchinei-Pang (as we would like to avoid introducing definition that are not explicitly used in the results). As for the definition of affine operator and its associated matrix, we will add a footnote about it since most readers are likely familiar with this notion.
4. About the notations: A capital X is used to mark a state of an algorithm (a stochastic process, so typically noted with uppercase letters) while a lower case x represents a point in the space.
5. Thank you for your suggestion of comparing against other methods. We will run the corresponding experiments and discuss the differences between different algorithms. We will also update the plot for colorblind readers.

Reviewer #3:

Hyperparameter tuning is a recurring problem in machine learning. Our goal was not to provide a parameter-free method, but a method with superior convergence guarantees. Developing an adaptive, parameter-free variant is an entire new paper by itself. Our paper provides the bedrock for this (and other) future considerations, a fact which highlights the potential impact of our work. As for the numerical experiments, since our paper is mainly theoretical, experimental details were relegated to the appendix, where we provide all the necessary reproducibility material. Concerning what we mean by a localized version of the assumptions: it could indeed be difficult to grasp for many readers. While in the submitted version the formal statements of these local assumptions appear in the appendix, we plan to move them to the main text in a newer version. Finally, the unbiasedness of the operator is a standard assumption in the theoretical analysis of stochastic optimization methods and is also verified in most machine learning applications (e.g., mini-batch setting where the estimator is unbiased by construction). Bias only appears to be an issue in very specific applications.

Reviewer #4:

Thank you very much for your careful reading, comments and appreciation of our work.

References

Chavdarova, T., Gidel, G., Fleuret, F., and Lacoste-Julien, S. Reducing noise in gan training with variance reduced extragradient. In *NeurIPS*, 2019.

Loizou, N., Berard, H., Jolicoeur-Martineau, A., Vincent, P., Lacoste-Julien, S., and Mitliagkas, I. Stochastic hamiltonian gradient methods for smooth games. In *ICML*, 2020.

Ryu, E. K., Yuan, K., and Yin, W. ODE analysis of stochastic gradient methods with optimism and anchoring for minimax problems and GANs. `https://arxiv.org/abs/1905.10899`, 2019.