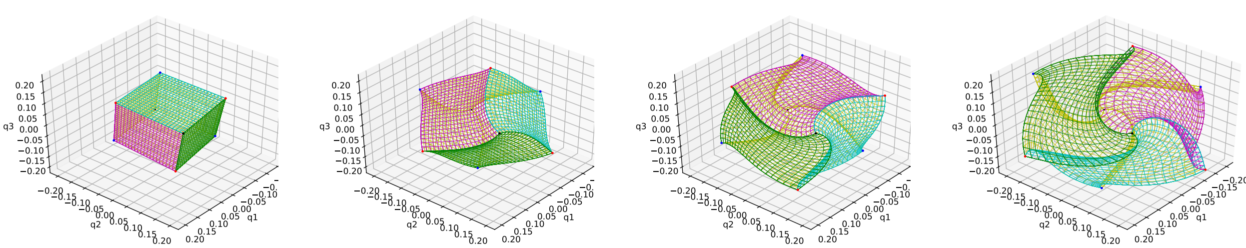
First of all, thanks all reviewers for your positive comments on this paper. We will fix the typos and minor errors you pointed out. We address your questions/suggestions below.

(1) Empirical Studies. If the paper is accepted to NeurIPS 2020, we will have one extra page. We will use it to include more empirical studies beyond the current Figures 1, 3 and 4 (Figures 3, 4 are in appendix). One example is the following volume-expansion figures in 3D. There are three players, each with two strategies Head and Tail. The underlying game is a graphical zero-sum game, where each edge-game is a Matching Pennies game: Player i wants to match with Player $(i+1)$ but wants to mis-match with Player $(i-1)$. We start with a small rectangular box around a Nash equilibrium (leftmost) in the payoff space, using Multiplicative Weights Update (MWU) for each player, and runs the algorithms for 4500, 9000 and 13500 steps. *The figures are plotted with high resolution; you may magnify them and read them more clearly.*

Indeed, we create more figures and combine them as a video. It demonstrates the evolution more clearly, regarding volume-expansion, edge-curving and edge-twisting. But we are not allowed to include a link to the video here; it will be in the final version of this paper. We also run with Optimistic MWU. Since the volume shrinks to a single point (boringly) as expected, we choose not to include the figures in this short response. They will appear in the final version.



(2) How may the results, particularly the volume analyses, be translated to the case with "diminishing step-size"? What about other learning dynamics, e.g., those in [1][5][17]? It is fairly straight-forward to generalize volume analyses to diminishing step-size; indeed, in all calculations we simply need to change ϵ to ϵ_t. About the results:

- The unavoidability theorem can be generalized to diminishing step-size when $\epsilon_t = \Omega(1/\sqrt{t})$, since this can guarantee that the volume expands quickly enough, so that the proof of the theorem can carry through.
- For the extremism theorem, we need unavoidability theorem and a technical lemma (Lemma 10 in Appendix C). As said, the unavoidability theorem component can carry through with $\epsilon_t = \Omega(1/\sqrt{t})$. For the lemma, we believe it holds with step-size which is diminishing but not too quickly. But we currently do not see a quick hack of the current argument to generalize. We believe it can be done with more careful argument.

Regarding applicability of volume analyses to other learning dynamics (e.g. those in [1,5,17]): yes it can be applied to those dynamics, as all of them belong to the family of *gradual* update rules which we formulate in this paper. Indeed we are planning to work on some of them in future work.

(3) The main results, the two negative properties, do not undermine the practicalities of the algorithms at all... MWU and OMWU are commonly applied mainly due to the convergence of their time average... We do not have any intent to undermine the practicalities of these algorithms. Rather, our key message is they are not practical in *every* scenario, we need to choose the appropriate ones depending on the applications. On the other hand, we note that there are simple games for which MWU and OMWU both do *not* converge, even in time-averages. One example is the game studied in Kleinberg et al. *Beyond the Nash Equilibrium Barrier* in ICS 2011. They showed instability of simple learning algorithms. We have run simulations to verify that both MWU and OMWU diverge in time-average.

(4) It would be better to compare with the most relevant a prior work [9] with more details. [9] is the first work in ML venue to *prove* that chaos exists using volume-expansion argument. Our focus in this paper is the new types of negative consequences of volume expansion, namely unavoidability and extremism. We also investigate how volume analyses can be applicable to other learning methods. As we see it, the generalization to OMWU is not straight-forward, and its analysis is interesting from a technical perspective.

(5) The discussion on RPS in Appendix C.1 is indeed complicated. I wonder whether it is possible to explore a more general condition for the non-trivial matrix, including both conditions in Theorem 5 and RPS. We will improve the discussion in Appendix C.1. We also thought about how to generalize to trivial matrices, but we did not see a simple way, unless we introduce further (heavy) mathematical notations. However, this appears to reduce the readability of this paper with relatively little gain, so we opted not to pursue.