We would first like to thank the reviewers for their insightful comments on our work.

Reviewer #1 and #3 suggested that our current experimental section was offering little added value on top of our theoretical analysis. Although our main motivations are theoretical (i.e. explain theoretically why zero-order algorithms can provide the same guarantees as the first-order counterparts in non-convex settings), we provide a new experimental section that we believe couples well both with our stated goals as well with the prior related work in the area. Given the reported empirical success of zero-order algorithms in many applications, the scope of our experiments was not to verify the success of AGD or PAGD in general practical settings but to demonstrate the effect of saddle points for the case of zero order algorithms (versus first order methods). To offer an even better illustration of the above point, we propose to replace the 2-d rastrigin function with a high dimensional version of octopus function as presented in [DLJ$^+$17]. This function is particularly relevant to our setting as it possesses a sequence of saddle points. The authors of [DLJ$^+$17] proved that gradient descent needs exponential time to avoid saddle points before converging to a local minimum. In contrast the perturbed version of gradient descent does not suffer from the same limitation. Figure 1 clearly shows that the zero-order versions have the same iteration performance with the first-order ones. In fact, AGD is shown to behave even better than GD in this example thanks to the noise induced by the gradient approximation. Thus our theoretical findings are verified even in this well-established and challenging benchmark.

Reviewer #1 asked us about the motivation behind our choice to study the case of an arbitrary h_0 selection in contrast to a random one. Our main motivation was to offer similar guarantees with the first order counterparts [LPP$^+$17] , i.e the avoidance of saddle point stems from their instability and not from an extra random dimension. Additionally, in our experience practitioners of zero-order methods tend to use some fixed values based on the machine precision and not generally some randomly sampled numbers. Thus, providing these stronger guarantees, our result reflects better what actually happens in practice.

Reviewer #1 asked for a clarification about the cost of a line search alternative. The main intuition is that one can try progressively smaller values of h until the value of f is decreased [Torczon, V.J. (1997)]. Using Lemma 3 and property iii) of Definition 4, one can see that the required number of trials actually depends on the norm of the gradient at the current iterate. For convergence to first-order stationary points ,i.e ($\|\nabla f(\mathbf{x}_k)\| \leq \epsilon$), this is hardly a problem, since $\|f(x)\| \geq \epsilon$ until termination and finding an appropriate h is trivial. However, for second-order stationary points, such trivial termination condition does not hold. Therefore $\|\nabla f(x)\|$ can be arbitrarily small and thus the sample complexity of this process may be unbounded in terms of $\frac{1}{\epsilon}$. One of the surprising contribution of our work is that there exist zero-order methods (like PAGD) that can escape saddle points in a bounded number of iterations with a fixed h.

Finally, all the minor typos spotted by the reviewers will be corrected for the camera ready version. Regarding the issues of presentation for both the main text and the appendix, we will be happy to follow the suggestions of Reviewer #1. Additionally, about the difference of the writing style between section 4 and 5, that Reviewer #4 mentioned, we plan to hide the details of Corollary 1, Lemma 2 and Theorem 3 expressing their main intuition in text.

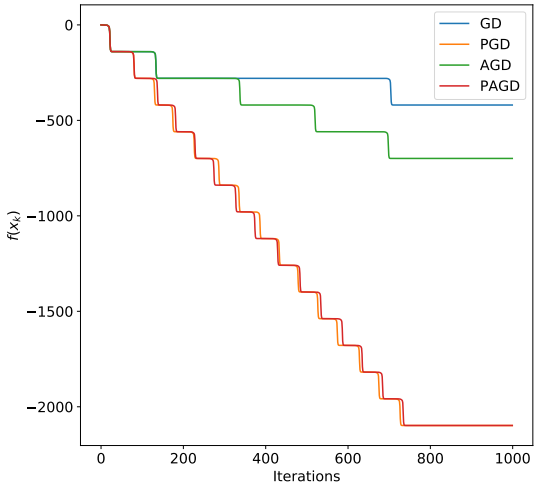

Figure 1: Octopus function of $d = 15$. Parameters of the function $\tau = e$, $L = e$, $\gamma = 1$. Parameters of first order methods taken from [DLJ$^+$17]. Zero order methods use symmetric differencing with $h = 0.01$

[Torczon, V.J. (1997)]. "On the convergence of pattern search algorithms" (PDF). SIAM Journal on Optimization.