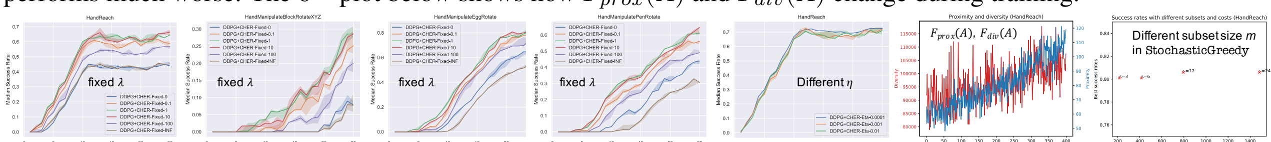
We appreciate the reviewers' efforts and suggestions (in blue)! We will address them all in the next version, and cite/discuss all the papers mentioned by reviewers. We will answer the shared question and then reply to each reviewer.
Shared: • (Improvements) How to choose similarity metric (or proximity) for different tasks? 1) We focus on multi-goal ML tasks that HER and many works aim to address, on which CHER using proximity in Eq.1 with Euclidean distance can achieve compelling results. Although Euclidean distance is not universal for arbitrary tasks, it works generally well over many tasks important to RL community. 2) Every RL system more or less relies on domain knowledge such as some physical laws. Tasks may prefer different distance metrics, but most physical systems have their own predefined ones, e.g., for environment with obstacles or a maze, it is natural to have proximity as the length of the shortest legal path to the goal; for a smooth surface, geodesic distance is a better choice. 3) Our key idea, a curriculum with increasing proximity and decreasing diversity, is not limited to Euclidean distance and can be applied with any predefined proximity.
Reviewer 1: • A theoretical motivation for having the curriculum on the λ value in the first place was not given. A critical theoretical motivation of having λ is to balance exploration-exploitation trade-off as in online learning problems.
• (Improvements) Comment on some of the related work on hindsight goal sampling. We will add discussion of those works. They are not directly comparable due to different task settings: their observations are raw pixels, while HER and CHER use physical positions. For fair comparison, we need to modify either HER/CHER or baselines, like [Nair et. al, 2018] modifying HER to minimize pixel-MSE. But this makes the original task harder with extra cost of training a VAE handling raw pixels. [Warde-Farley et.al, 2019] are different to CHER in: 1) it learns a reward function to address the sparse reward problem, and 2) it samples the goal buffer as uniform as possible.
• (Improvements) Ablations on fixed λ values. Plotting the value of $F_{prox}(A)$, $F_{div}(A)$ over the course of training. The left 4 plots below report the performance of 6 different fixed λ values on the four tasks, where $\lambda =$INF refers to proximity(goal-similarity)-only. Compared to Fig.3-4 using λ-curriculum (Fig.4 shows CHER initialized with different λ_0), fixed λ performs much worse. The 6^{th} plot below shows how $F_{prox}(A)$ and $F_{div}(A)$ change during training.



• What is the η and λ_0 value? How sensitive is the performance to this parameter? Fig.4 shows that CHER's performance remains stably good for $\lambda_0 \in [0.1, 10]$. The 5^{th} plot above (using $\lambda_0 = 1$) shows that CHER is also robust to different η. They indicates that CHER is robust to the choices of η and λ_0. We use $\eta = 0.0001$ and $\lambda_0 = 1$ for all the tasks in experiments.
Reviewer 2: • (Improvements) Ablation study of different σ in the RBF kernel. How η and τ in Eq.7 affect performance? We adopt an adaptive σ widely used in kernel methods: the average distance over all (x, y) pairs. It practically has promising performance while saves tuning cost. Result of different η is shown in the 5^{th} plot above. τ is the episode number.
• (Improvements) The diversity term shouldn't be called "curiosity" We will change it to "diversity".
• (Improvements) The curves in Fig.3(a) are suspiciously cut at Epoch=50, after which the baseline methods seem to catch up and perhaps surpass CHER. They saturate and won't surpass CHER later. Zooming in Fig.3(a) at Epoch=50 also shows that the orange curve (our method) increases the fastest, while the baselines are either decreasing or increasing slower.
Reviewer 3: • (Improvements) An empirical evaluation what effect sub optimality in sampling actually has on agent performance. In the rightest plot above, we report the time costs and success rates for StochasticGreedy with different sub-sampling sizes. It shows that small sub-sampling size improves efficiency but does not harm the optimality.
• (Improvements) Compare with simpler prioritized-replay mechanisms, e.g. directly using the goal-similarity metric in a priority queue. In above 4 plots of fixed λ, the success rate of solely using goal-similarity/proximity($\lambda =$INF) is lower than smaller fixed-λ and λ-curriculum in Fig.3-4. This indicates the importance of diversity for HER.
• In Fig.3, CHER offers no benefit over HER and HEREBP baselines for tasks b & c, but is significant on tasks a & d. To what do the authors attribute this difference? Task b, c and d are all rotating tasks. Comparing to d, b & c have shapes (block and egg) easier to handle, and the proximity and diversity of different achieved goals are more similar to each other since the shapes are more rotate-invariant. Hence, CHER makes less difference on b & c. Nevertheless, it still achieves the best performance on b & c and significantly outperforms the best baseline by 4% and 1.5% (note the baseline already has $> 95\%$).
• (Improvements) Inconsistency of c & d between Fig.3 and Fig.4: the effect of λ is quite small in Fig.4 for both c & d. In Fig.4, on tasks a & b, $\lambda = 1$ is optimal, but doesn't seem to hold in c & d. Sorry for the typo in Fig.4's caption: it shows CHER initialized with different λ_0. It shows that the effect of changing λ_0 in $[0.1, 10]$ is small, and $\lambda_0 \simeq 1$ performs the best, same as Fig.4(a)-(b). It actually exhibits the robustness of CHER to λ_0: its remarkable performance is mainly resulted from the dynamic curriculum rather than careful tuning of λ_0. We tuned all the baselines for their best performance, which are consistent with previous papers about HER on the same tasks. In Fig.3(a), CHER reaches a success rate of 78% only after 10 epochs while DDPG-HER spent 50 epochs. CHER is much more efficient for its careful selection of curriculum. Although a & d are similar tasks, their robots have different mechanical structures, which results in different performance.
• Plot titles and axis legends difficult to read. We will try our best to improve their readability.
• What is the compute cost of the proposed method? Running an iterative optimization inside the batch-sampling step of a Deep-RL algorithm sounds expensive. The only extra computation of CHER (comparing to HER) is to run StochasticGreedy in Line 12 of Algorithm 2, which only needs mk evaluations of Eq.5 (note the similarities in Eq.5 are invariant and pre-computed). In our experiments, $mk = 192$ and the extra computation is ignorable ($< 5\%$ of the total training time).