We thank the reviewers for their thorough and insightful comments. We emphasize that our contributions extend EHVI to the important case of parallel constrained optimization. Our approach is practical and highly competitive (including against a novel baseline algorithm, qPAREGO, that bests all publicly-available MOBO algorithms). Furthermore, our recent experience with attempting to obtain code from several MOBO authors highlight the value of our work in providing a replicable foundation for advancing cumulative research in this area.

Reviewers raised concerns about limited applicability of qEHVI to noisy settings. While EHVI does not account for the noise in the acquisition function (AF) itself, this does not necessarily imply that it would perform poorly on problems with noise. In fact, no previous work on EHVI explicitly accounts for observation noise. Among the many MOBO papers, to our knowledge only PESMO [Hernández-Lobato et al., 2015] and its extensions [Garrido-Merchán and Hernández-Lobato, 2019, 2020], scalarized TS [Paria et al., 2018], and ϵ-PAL [Zuluaga et al., 2016] explicitly handle noise. With the additional page, we plan to address these concerns by including additional evaluation on noisy problems, and we can include a simple extension of qEHVI to the noisy setting if the reviewers wish.

In our conclusion, we note that extending qEHVI to account for observation noise would be non-trivial, but upon further consideration, there is a straightforward extension inspired by Noisy EI [Letham et al., 2019, Balandat et al., 2019]. The idea behind the approach, which we call qNEHVI, is to integrate over the uncertainty in the Pareto frontier (PF) over the previously evaluated points X_{baseline} by drawing N joint samples from the posterior of $\boldsymbol{f}_t(X_{\text{baseline}}) \sim \mathbb{P}(\boldsymbol{f}(X_{\text{baseline}})|\mathcal{D}),\ t = 1, ..., N$. We prune X_{baseline} to remove points with zero probability of being Pareto optimal (estimated using MC). For each MC sample, we compute the PF and partition the non-dominated space into disjoint rectangles; this is only required once per BO iteration and can be easily parallelized across multiple processes. Computing the AF is the same as in qEHVI, except that we draw N joint samples from $\mathbb{P}(\boldsymbol{f}(X_{\text{baseline}}, \mathcal{X}_{\text{cand}})|\mathcal{D})$ where the t^{th} sample is conditioned on the original samples $\boldsymbol{f}_t(X_{\text{baseline}})$ to ensure the t^{th} cached partitioning uses the original samples: $\boldsymbol{f}_t(\mathcal{X}_{\text{cand}}) \sim \mathbb{P}(\boldsymbol{f}(X_{\text{baseline}}, \mathcal{X}_{\text{cand}})|\mathcal{D}, \boldsymbol{f}(X_{\text{baseline}}) = \boldsymbol{f}_t(X_{\text{baseline}}))$. Thus, the only distinction between qEHVI and qNEHVI is that qEHVI uses a partitioning on the PF over observations and qNEHVI uses a separate partitioning for the PF over the function values under each MC sample.

We empirically evaluate the performance of all algorithms on a BraninCurrin function where observations have additive, zero-mean, iid Gaussian noise; the unknown standard deviation of the noise is set to be 1% of the range of each objective. Fig 1 shows that qEHVI performs favorably in the presence of noise, besting PESMO and TS-TCH, which explicitly account for noise. qNEHVI dominates all algorithms, including Noisy qPAREGO (described in Appendix E) with respect to log hypervolume difference. Additional problems will be included in the final script, but we have seen similar evidence across a number of informal experiments during testing.

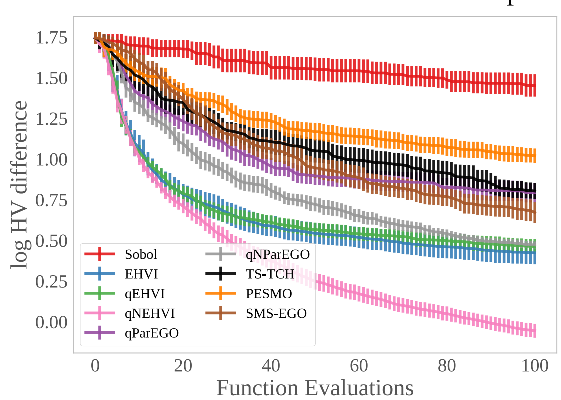

Figure 1: MOBO on the noisy BraninCurrin problem.

qNEHVI maintains acceptable optimization wall time. We report the mean and 2 std errors of the wall time per BO iteration in seconds on a CPU: qPAREGO 1.8 ($\pm$0.2); qNPAREGO 2.1 ($\pm$0.2); EHVI 2.1 ($\pm$0.2); qEHVI 2.7 ($\pm$0.3); TS-TCH 13.2 ($\pm$0.3); qNEHVI 52.9 ($\pm$4.4); SMS-EGO 87.2 ($\pm$5.0); PESMO 233.12 ($\pm$15.02). We also provide GPU wall times for parallel qNEHVI: (q=1) 44.9 ($\pm$3.4); (q=2) 47.6 ($\pm$5.3); (q=4) 82.7 ($\pm$8.2); (q=8) 197.5 ($\pm$26.4).

R4 requests additional comparisons. We have included TS-TCH from Paria et al. [2018] in Fig 1 and will include it in all experiments in the final version. We have would liked to include comparisons with other recent work, but Belakaria et al. [2019, 2020], Garrido-Merchán and Hernández-Lobato [2020], and Suzuki et al. [2019], have all graciously declined to share code in response to our request (R4 also mentions Abdolshah et al. [2019], but it is not applicable to the work here).

R1 brought up the handling of constraints. While qEHVI cannot use the Pareto dominance rule from Feliot et al. [2016] and maintain differentiability, qEHVI gives higher value to candidates that are more likely to be feasible given the candidates have the same MC samples of the objectives, even in the case of no feasible observations.

Regarding R3's comment on scalability with respect to the number of objectives M: any exact EHVI computation is exponential in M and therefore is only suitable for moderate M [Yang et al., 2019]. Since qEHVI uses the disjoint partitioning for piece-wise integration, it is agnostic to both the partitioning algorithm and the realized partitioning. Therefore, our approach is compatible with more efficient alternative partitioning methods including Yang et al. [2019] and even more scalable approximate partitioning methods such as Couckuyt et al. [2012] (although EHVI computation may no longer be exact with an approximate partitioning).