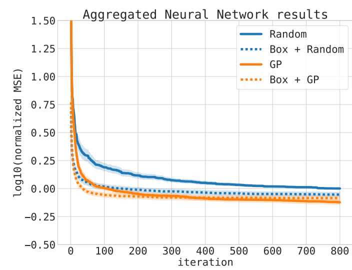
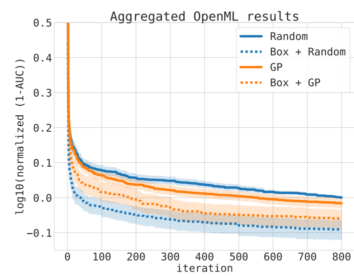
We thank the reviewers for the positive and constructive comments. Our detailed responses are below.

R2. We are grateful to the reviewer for the comments and suggested reference, which we will discuss in the revision. The idea of learning hyperparameter importance bears interesting connections to our work, and is an alternative viewpoint on transfer learning that could be combined with our novel representations of the search space. It would be insightful to apply a post-hoc functional ANOVA analysis to validate the way we prune the search space a priori.

We would like to clarify that, while the first step of our method only restricts the search space for numerical features, it still performs standard BO over the joint space, capturing the interactions between the two types of variables. The bounding box formulation naturally handles categorical parameters that are one-hot encoded: if none of the $\mathbf{x}_t^\star$'s uses a given category, say corresponding to the j-th dimension, we have $[\mathbf{x}_t^\star]_j = 0$ for all t, and the corresponding entries of $\mathbf{l}^*$ and $\mathbf{u}^*$ in (4) will satisfy $l_j^* = u_j^* = 0$; this means that a category not present in the $\mathbf{x}_t^\star$'s is automatically pruned out of the search space. We did not mention this aspect to keep the discussion simpler, but would be happy to clarify this in the revision. We would also like to clarify that there are no extra hyperparameters introduced with the ellipsoid (see equation (6) of the paper), similar to the bounding box case. Providing a parameter-free, off-the-shelf methodology was one of the main motivations of our approach.





Our goal was to lift the burden of choosing the search space. When running experiments for a fixed amount of resources as we do, this has the effect of exploring the (restricted) search space more densely, which we showed to be beneficial. Following on the reviewer's observations, we re-ran our experiments with 16 times as many iterations to assess how much accuracy can potentially be lost compared to methods that search the entire space (with more resources). The figures on the right show the neural network and OpenML results: restricting the search space leads to considerably faster convergence and the GP fails to find a better solution, only eventually catching up in the neural network case. This suggests that there is almost no loss of performance, but just a speed up effect of finding a very good solution in as few evaluations as possible. We thank the reviewer for this insightful comment and would like to add these new results to the revision.

Toy SGD experiment: We replicated the setup of Valkov et al., 2018, to provide insights into the proposed algorithm in a fully controlled setting, where resource-aware optimizers can be tested and we could build intuition by visualizing the evaluations together with the search space restriction in Fig. 2 of the paper and confirming a meaningful behavior. *Neural network experiment*: Following the setup of Klein and Hutter (2019), the parameters in the neural network experiment were discretized to allow for an exhaustive look-up table, eliminate the noise coming from interpolation and ease reproducibility. *Larger hyperparameter setting*: We agree that finding automatic search spaces in larger hyperparameter settings would be a valuable study to conduct, with the potential of further speeding up the BO by removing or restricting irrelevant dimensions. As we aimed to compare to previous transfer learning approaches developed in similar settings, we chose to benchmark the moderate-sized P (around 10) regime as in Snoek, et al. (2015), Springenberg, et al. (2016), and Perrone et al. (2018).

R3. Choosing the search space is difficult in practice and being conservative, by picking a large search space, can negatively impact the optimization performance. Our goal was to highlight this aspect as it is often overlooked in the literature, and to propose a simple, yet effective methodology to automate this critical step. When evaluating the methodology we asked ourselves the same question regarding the possibility of excluding the best solution in some situations. The experimental results show that focusing on the search space to induce transfer learning in BO is more effective than more complex and computationally intensive approaches. We also found that, when we were restricting the search space, good solutions were always found in the interior of the search space. We conducted an extra set of experiments to see how much accuracy could potentially be lost compared to methods that search the entire space (with more resources). The results indicate that results were robust and the difference was small, if any (see Figs on the side).

We confirm n in lines 256-257 corresponds to n_t in line 83, which we will fix in the revision. In all OpenML experiments, the n_t previous evaluations per task are drawn uniformly at random from the ones available, with $T = 30$ in all OpenML experiments except for Figure 4b.

R4. We would like to thank the reviewer for the positive feedback and suggestions. Several extensions are indeed possible, some of which are outlined in the discussion section. We thank the reviewer for the additional suggestion of combining our method with PESMOC and would be happy to further discuss this off-line. Striving for reproducibility, we made the algorithmic details as thorough as possible, and provided the pseudo-code of the proposed methodology.