Supplementary Material: Bayesian Optimization with Inequality Constraints

Jacob R. Gardner[1] GARDNER.JAKE@WUSTL.EDU
Matt J. Kusner[1] MKUSNER@WUSTL.EDU
Zhixiang (Eddie) Xu[1] XUZX@CSE.WUSTL.EDU
Kilian Q. Weinberger[1] KILIAN@WUSTL.EDU
John P. Cunningham[2] JPC2181@COLUMBIA.EDU
[1]Washington University in St. Louis, 1 Brookings Dr., St. Louis, MO 63130

[2]Columbia University, 116th St and Broadway, New York, NY 10027

A: Different Acquisition Functions

In this section we evaluate using different acquisition functions for constrained Bayesian optimization. The constrained acquisiton function in our paper was ultimately the probability of feasibility times the original acquisition function. In principle any acquisition function can be modified for constrained optimization by similarly multiplying it by the probability of feasibility.

We compare three popular acquisition functions, Expected Improvement (EI), Probability of Improvement (PI), and the Lower Confidence Bound (LCB) measure. For a full description of these, see (Brochu et al., 2010). Table **??** displays the results of using these modified acquisition functions to tune the hyper-parameters of LSH. In terms of end result (speedup), EI outperforms the other two acquisition function on nearly all datasets (save letters). However, the more myopic PI acquisition function runs substantially more feasible points. This is likely because it explores less, and prefers to exploit by continuing to sample in a feasible region once it has found one.

B: Tree of Parzen Estimators

Another common global optimization scheme for hyperparameter selection is Tree of Parzen Estimators (TPE) (Bergstra & Bengio, 2012). For completeness, we demonstrate here that the TPE algorithm fares no better than standard Bayesian Optimization. Ultimately, TPE and BO fare worse than uniform sampling if the feasible regions are not near the global optima, as TPE and BO might actively *avoid* sampling in these regions.

Figure 1 displays the results of running TPE on our simulation functions. As with standard BO and uniform sampling, the small feasible region in simulation 2 prevents TPE from finding any feasible points at all.

References

Bergstra, J. and Bengio, Y. Random search for hyper-parameter optimization. *JMLR*, 13:281–305, 2012.

Brochu, E., Cora, V. M., and De Freitas, N. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv preprint arXiv:1012.2599*, 2010.

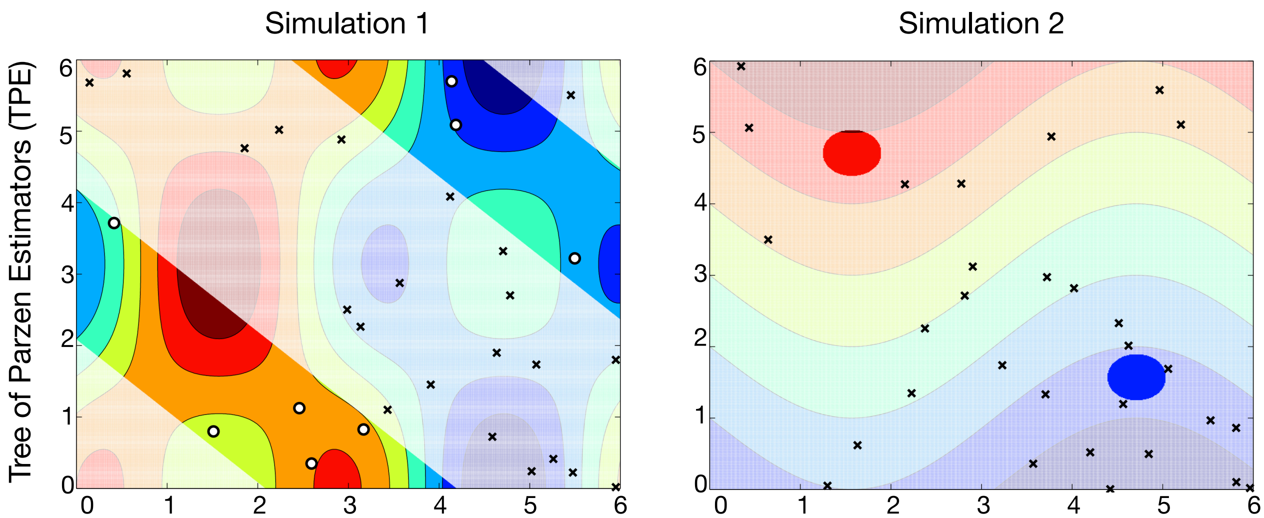

Figure 1. Evaluation of Tree of Parzen Estimators (TPE) on the simulation problems. Areas shaded in white are infeasible regions. White circle indicate feasible points, and black crosses indicate infeasible points.

Table 1. LSH results over 10 runs for selecting the number of hash tables and functions for approximate kNN search using different acquisition functions. We show speedup over kNN and the percentage of infeasible points sampled.

LSH						
	SPEEDUP (ℓ)			% INFEASIBLE		
DATASET	EI	PI	LCB	EI	PI	LCB
YALE-FACES	$\mathbf{3.33} \pm 1.53\times$	$1.54 \pm 0.39\times$	$1.04 \pm 0.71\times$	$89 \pm 7.0\%$	$\mathbf{3.8} \pm 2.2\%$	$99 \pm 0.5\%$
COIL	$\mathbf{18.6} \pm 13.6\times$	$2.30 \pm 1.98\times$	$3.20 \pm 4.17\times$	$84 \pm 8.4\%$	$\mathbf{1.9} \pm 0.3\%$	$99 \pm 0.6\%$
ISOLET	$\mathbf{6.97} \pm 1.21\times$	$4.35 \pm 0.71\times$	$4.44 \pm 1.04\times$	$67 \pm 16\%$	$\mathbf{1.3} \pm 0.8\%$	$97 \pm 1.0\%$
USPS	$\mathbf{3.58} \pm 0.89\times$	$2.45 \pm 1.24\times$	$2.71 \pm 1.42\times$	$81 \pm 14\%$	$\mathbf{2.5} \pm 1.2\%$	$99 \pm 1.1\%$
LETTERS	$1.64 \pm 0.70\times$	$1.56 \pm 0.92\times$	$\mathbf{1.92} \pm 0.42\times$	$70 \pm 14\%$	$\mathbf{29} \pm 4.9\%$	$96 \pm 4.6\%$
ADULT	$\mathbf{2.80} \pm 2.13\times$	$1.95 \pm 0.66\times$	$2.53 \pm 1.67\times$	$97 \pm 3.5\%$	$\mathbf{95} \pm 12\%$	$99 \pm 0.9\%$
W8A	$\mathbf{3.01} \pm 0.30\times$	$0.68 \pm 0.29\times$	$0.89 \pm 0.66\times$	$54 \pm 15\%$	$\mathbf{1.0} \pm 0.0\%$	$98 \pm 2.1\%$
MNIST	$\mathbf{1.69} \pm 0.59\times$	$0.73 \pm 0.60\times$	$0.73 \pm 0.67\times$	$71 \pm 16\%$	$\mathbf{2.4} \pm 1.3\%$	$96 \pm 2.3\%$