THE UNIVERSITY *of* EDINBURGH

Technical Report

Afonso Eduardo
`afonso.eduardo@ed.ac.uk`

August 2022

1 Collapsed Expected Improvement

Expected Improvement (`EI`) and Lower Confidence Bound (`LCB`) are susceptible to local minima. There is no intrinsic mechanism that prevents the rules from choosing points with already small predictive variances. In `LCB`, this is typically addressed by increasing β_n and decreasing lengthscales over time [1, 2], ensuring that the entire domain is eventually explored. Increasing β_n is functionally almost the same as increasing the prior variance. So, a similar strategy can carry over to `EI`. However, if the prior variance increases too fast, or the lengthscales decrease rapidly, BO becomes too explorative before it has the chance to exploit any information given by the posterior predictive mean. So, unconditional updates should be avoided.[1]

Wang et al. [4] provide a conditional rule for such update that seems reasonable: Lengthscales only decrease if the predictive variances of successive future acquisitions are less than a certain threshold, e.g., noise variance. The problem with this approach is that the amount of decrease that leads to mode collapse of `EI` is unknown, so it may take multiple updates and evaluations before this main acquisition mode vanishes. Here, we modify the acquisition function directly by removing this acquisition mode, and the procedure is repeated until the predictive variance of the acquisition is larger than the threshold. To this end, we use the Laplace method, estimating a log quadratic model centered at the location that maximizes the current acquisition function.

Collapsed EI (`CEI`) may be computationally more expensive than `EI`. Still, larger posterior predictive variances translate into more informative acquisitions, making better use of the limited evaluation budget. As a byproduct, note that repeated application of the Laplace method with mode collapse retrieves a Gaussian mixture search model $q(\boldsymbol{x}) = \sum_i w_i \mathcal{N}(\boldsymbol{x}; \boldsymbol{x}^{(i)}, \mathbf{H}^{-1(i)})$, with $w_i \propto \mathrm{EI}(\boldsymbol{x}^{(i)})$, that approximates the implicit posterior over promising regions $p(\boldsymbol{x}^\star|\mathcal{D}_n) \propto \mathrm{EI}(\boldsymbol{x}^\star)$. This information about promising regions can also later be fed to informative covariance functions [3] via the shaping function ϕ that induces the nonstationary effects.

Algorithm 1 Collapsed EI (`CEI`)

Input: posterior predictive variance v_n, threshold ϵ
Output: acquisition $\boldsymbol{x}^{(i)}$
$\boldsymbol{x}^{(0)} = \arg\max \mathrm{EI}(\boldsymbol{x})$
$i = 0$
$\mathrm{CEI}^{(i)} = \mathrm{EI}$
while $v_n(\boldsymbol{x}^{(i)}) \leq \epsilon$ **do**
  $\mathbf{H}^{(i)} = \nabla^2_{\boldsymbol{x}} - \log \mathrm{CEI}^{(i)}(\boldsymbol{x})|_{\boldsymbol{x}=\boldsymbol{x}^{(i)}}$
  $\mathrm{CEI}^{(i+1)}(\boldsymbol{x}) = \mathrm{CEI}^{(i)}(\boldsymbol{x}) - \mathrm{CEI}^{(i)}(\boldsymbol{x}^{(i)}) \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}^{(i)})^\intercal \mathbf{H}^{(i)} (\boldsymbol{x}-\boldsymbol{x}^{(i)})\right)$
  $\boldsymbol{x}^{(i+1)} = \arg\max \mathrm{CEI}^{(i+1)}(\boldsymbol{x})$
  $i = i + 1$
end while

[1] Interestingly, informative covariance functions [3] can too be of help in this case: Information from the posterior predictive mean can be directly incorporated in the proposed covariance functions through the shaping function ϕ.

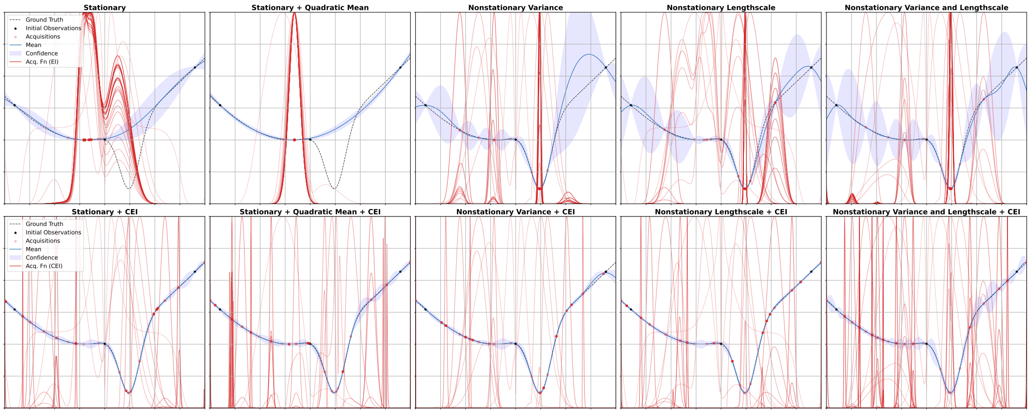

Figure 1. A comparison of `EI` and `CEI`. In this example, `CEI` can solve the problem that is due to overconfident stationary models. Overall, it leads to more informative acquisitions as it avoids locations with small predictive variances.

2 BOIC and Trust Regions

Bayesian Optimization (BO) with Trust Regions [e.g., `TuRBO` 5] has become popular for high-dimensional optimization due to its superior performance over standard BO. Importantly, informative covariances [3] and trust regions are complementary and not mutually exclusive. As shown in Figure 2, `I+XA+TR` outperforms `S+TR`. Fundamentally, `+TR` alone does not solve the problems that are due to stationary models. For instance, the boundary issue can still occur, but confined to each trust region. Informative models and trust regions prove to be an effective combination for high-dimensional optimization.

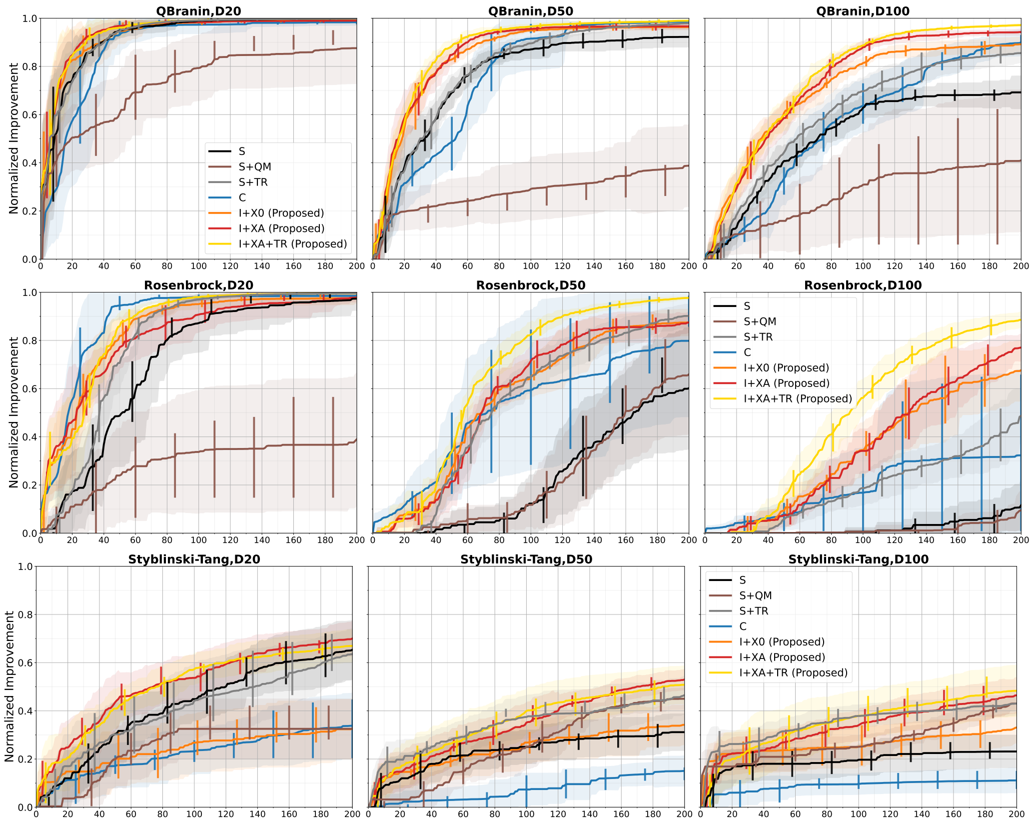

Figure 2. Performance on objectives with 20, 50 and 100 dimensions [Appendix B.3 3]. Budget of 200 acquisitions. Solid curves and shaded regions represent the normalized improvement mean and one standard deviation over 10 trials, each with different initial conditions. Solid vertical lines show the interquartile range. **Abbreviations:** Stationary (`S`), Cylindrical (`C`) and Informative (`I`) covariances; Quadratic Mean (`+QM`); Origin (`+X0`) and Adaptive greedy (`+XA`) anchors; acquisitions within Trust Region (`+TR`).

References

[1] N. Srinivas et al. "Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design". In: *Proceedings of the 27th International Conference on International Conference on Machine Learning*. 2010, pages 1015–1022.

[2] F. Berkenkamp et al. "No-regret Bayesian Optimization with Unknown Hyperparameters". In: *JMLR* 20.50 (2019), pages 1–24.

[3] A. Eduardo and M. U. Gutmann. "Bayesian Optimization with Informative Covariance". In: *arXiv preprint arXiv:2208.02704* (2022).

[4] Z. Wang and N. de Freitas. "Theoretical Analysis of Bayesian Optimisation with Unknown Gaussian Process Hyper-Parameters". In: *arXiv preprint arXiv:1406.7758* (2014).

[5] D. Eriksson et al. "Scalable Global Optimization via Local Bayesian Optimization". In: *Advances in Neural Information Processing Systems* 32 (2019).