Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

Determining Convergence for Bayesian Optimization

Nicholas Grunloh | **Herbert Lee**

[1]Department of Statistical Sciences, University of California, Santa Cruz, CA, USA

Correspondence
Corresponding author Nicholas Grunloh. 110 McAllister Way, Santa Cruz, CA 95060
Email: grunloh@soe.ucsc.edu

Abstract

Bayesian optimization routines may have theoretical convergence results, but determining whether a run has converged in practice can be a subjective task. This paper provides a framework inspired by statistical process control for monitoring an optimization run for convergence. An Exponentially Weighted Moving Average chart is adapted for automated convergence analysis.

KEYWORDS
Derivative-free Optimization, Computer Simulation, Emulator, Expected Improvement

1 | INTRODUCTION

Bayesian optimization aims to find a global optimum of a complex function that may not be analytically tractable, and where derivative information may not be readily available[1,2]. A common application is for computer simulation experiments[3]. Because each function evaluation may be expensive, one wants to terminate the optimization algorithm as early as possible. However for complex simulators, the response surface may be ill-behaved and optimization routines can easily become trapped in a local mode, so one needs to run the optimization sufficiently long to achieve a robust solution. So far there has been little work on assessing convergence for Bayesian optimization. In this paper, we provide an automated method for determining convergence of surrogate model-based optimization by bringing in elements of statistical process control.

Among the wide variety of Bayesian optimization approaches, we focus on those that are based on a statistical surrogate model, such as a Gaussian process[4]. We further focus on approaches based on Expected Improvement (EI)[5], although our methods are generalizable for other acquisition functions.

There have been a few hints in the literature that monitoring EI directly could be used to assess convergence[6].[7] considers the use of the improvement distribution for identifying global convergence. The basic idea is that convergence should occur when the surrogate model produces low expectations for discovering a new optimum; that is to say, globally small EI values should be associated with convergence of the algorithm. Thus a simplistic stopping rule might first define some lower EI threshold, then claim convergence upon the first instance of an EI value falling below this threshold, as seen in[8]. This use of EI as a convergence criterion is analogous to other standard convergence identification methods in numerical optimization (e.g., the vanishing step sizes of a Newton-Raphson algorithm). However, applying this same threshold strategy to the convergence of Bayesian optimization has not yet been adequately justified. In fact, this use of EI ignores the nature of the EI criterion as a random variable, and oversimplifies the stochastic nature of convergence in this setting. Thus it is no surprise that this treatment of the EI criterion can result in an inconsistent stopping rule as demonstrated in Figure (1).

Because EI is strictly positive but decreasingly small, we find it more productive to work on the log scale, using a log-normal approximation to the improvement distribution to generate a more appropriate convergence criterion, as described in Section 3.2. Figure (1) represents three series of the Expected Log-normal Approximation to the Improvement (ELAI) values from three different optimization problems. We will demonstrate later in this paper that convergence is established near the end of each of these series. These three series demonstrate the kind of diversity observed among various ELAI convergence behaviors, and illustrate the difficulty in assessing convergence. In the left-most panel, optimization of the Rosenbrock test function results in a well-behaved series of ELAI values, demonstrating a case in which the simple threshold stopping rule can accurately identify

Abbreviations: ELAI, Expected Log-Normal Approximation of Improvement; EWMA, Exponentially Weighted Moving Average.

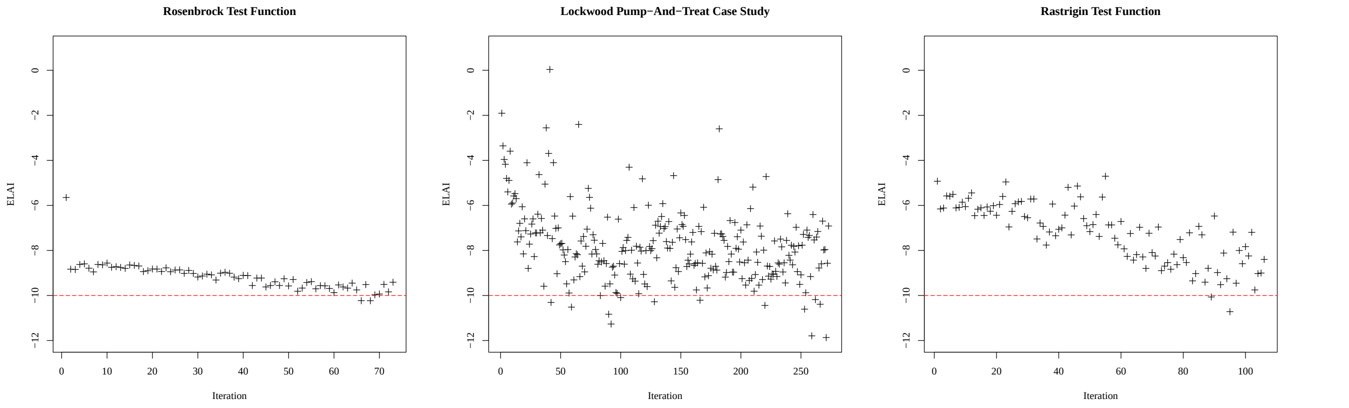

FIGURE 1 Three Expected Log-normal Approximation to the Improvement series (more details in Section **??**) plotted alongside an example convergence threshold value shown as a dashed line at -10.

convergence. However the center panel (the Lockwood problem described in Section 4.3) demonstrates a failure of the threshold stopping rule, as this ELAI series contains much more variance, and thus small ELAI values are observed quite regularly. In the Lockwood example a simple threshold stopping rule could falsely claim convergence within the first 50 iterations of the algorithm. The large variability in ELAI values with occasional large values indicates that the optimization routine sometimes briefly settles into a local minimum but is still exploring and is not yet convinced that it has found a global minimum. This optimization run appears to have converged only after the larger ELAI values stop appearing and the variability has decreased. Thus one might ask if a decrease in variability, or small variability, is a necessary condition for convergence. The right-most panel (the Rastrigin test function) shows a case where convergence occurs by meeting the threshold level, but where variability has increased, demonstrating that a decrease in variability is not a necessary condition.

Since the Improvement function is itself random, attempting to set a lower threshold bound on the EI, without consideration of the underlying EI distribution through time, over-simplifies the dynamics of convergence in this setting. Instead, we propose taking the perspective of Statistical Process Control (SPC), where a stochastic series is monitored for consistency of the distribution of the most recently observed values. In the next section, we review the surrogate model approach and the use of EI for optimization. In Section **??**, we discuss our inspiration from SPC and how we construct our convergence chart. Section **??** provides synthetic and real examples, and then we provide some conclusions in the final section.

2 | BAYESIAN OPTIMIZATION VIA EXPECTED IMPROVEMENT

REFERENCES

1. Mockus J. *Bayesian Approach to Global Optimization*. Dordrecht: Kluwer, 1989.
2. Brochu E, Cora VM, Freitas dN. A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning. Tech. Rep. 1012.2599, arXiv; 2010.
3. Gramacy RB. *Surrogates: Gaussian Process Modeling, Design, and Optimization for the Applied Sciences*. Chapman & Hall/CRC, 2020.
4. Santner TJ, Williams BJ, Notz W. *The design and analysis of computer experiments*. New York: Springer, 2003.
5. Schonlau M, Jones D, Welch W. Global versus local search in constrained optimization of computer models. In: , , No. 34 in IMS Lecture Notes - Monograph Series. JSTOR, 1998:11–25.
6. Jones DR, Schonlau M, Welch WJ. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization.* 1998;13(4):455–492.
7. Taddy MA, Lee HKH, Gray GA, Griffin JD. Bayesian guided pattern search for robust local optimization. *Technometrics.* 2009;51(4):389–401.
8. Diwale SS, Lymperopoulos I, Jones C. Optimization of an airborne wind energy system using constrained Gaussian processes with transient measurements. In: No. EPFL-CONF-199719. 2015.