Assessing Convergence in Gaussian Process Surrogate Model Optimization

Nicholas R. Grunloh and Herbert K. H. Lee

Abstract

Identifying convergence in numerical optimization is an ever-present, difficult, and often subjective task. The statistical framework provided by Gaussian Process surrogate model optimization provides useful secondary measures for tracking optimization progress; however the identification of convergence via these criteria has often provided only limited success and often requires a more subjective analysis. Here we use ideas originally introduced in the field of Statistical Process Control to define convergence in the context of an robust and objective convergence heuristic. The Exponentially Weighted Moving Average (EWMA) chart provides an ideal starting point for adaptation to track convergence via the EWMA convergence chart introduced here.

Keywords: Computer Model, Derivative-free Optimization, Emulator, Expected Improvement.

1 Introduction

Black-box derivative-free optimization is an important problem with a wide variety of applications, especially in the realm of computer simulations cite. When dealing with computationally expensive computer models, a key question is that of convergence of the optimization. Because each function evaluation is expensive, one wants to terminate the optimization as early as possible. However for complex simluators, 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. In this paper, we provide an automated method for assessing convergence of Gaussian process surrogate model optimization by bringing in elements of Statistical Process Control.

Our motivating example is a hydrology application, the Lockwood pump-and-treat problem **??**, discussed in more detail in Section XX, wherein contamination in the groundwater near the Yellowstone River is remediated via a set of treatment wells. The goal is to minimize the cost of running the wells while ensuring that no contamination enters the river. The contamination constraint results in a complicated boundary that is unknown in advance and requires evaluation of the simulator, and thus finding the global constrained minimum is a difficult problem and it is easy for optimization routines to at least temporarily get stuck in a local minimum. Without knowing the answer in advance, how do we know when to terminate the optimization routine?

The context of this paper is Gaussian Process surrogate model optimization, a statistical modeling approach to derivative-free numerical optimization routine that constructs a fast approximation to the expensive computer simulation using a statistical surrogate model [3]. Analysis of the surrogate model allows for efficient exploration the objective solution space. Typically a Gaussian Process (GP) surrogate model is chosen for its robustness, relative ease of computation, and its predictive framework [4]. Arising naturally from the GP predictive distribution [5, 2], the maximum Expected Improvement (EI) criterion has shown to be a valuable criterion for guiding the exploration of the objective function [6]Jones?; furthermore the EI shows promise for use as a convergence criterion cite.

Literature cite recommends considering the EI as a convergence criterion for GP surrogate model optimization; as of yet, little work has been done to describe what convergence of these algorithms actually looks like in the context of the EI criterion. However, the basic idea behind the use of the EI criterion, as a convergence criterion, is that convergence should occur when the GP surrogate model produces low expectations for discovering new optima; that is to say, small EI values should be associated with convergence of the algorithm. Thus, a common stopping rule, involving the EI criterion, first defines some lower EI threshold, then claims convergence upon first sight of an EI value falling below this threshold [1], cite. This use of EI as a convergence criteria falls well within the comfort zone of the numerical optimization literature as this is certainly a reasonable approach for monitoring the convergence criteria of other routines (ex. the vanishing step sizes of Newton-Raphson, Pattern Search (PS) methods, etc.). However, applying this same threshold strategy to the convergence of GP surrogate model optimization, in the context of the EI criterion, 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).

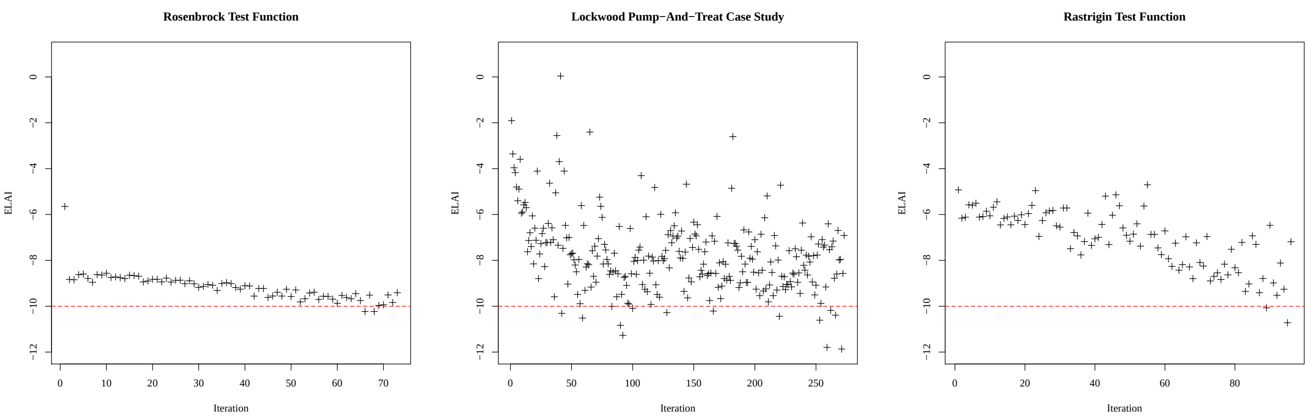

Figure 1: Three ELAI series plotted alongside an example convergence threshold value shown as a dashed line at -10.

Because Expected Improvement is non-negative but decreasingly small, we find it more productive to work on the log scale, using a lognormal approximation to the improvement distribution, as described in more detail in Section XX. Figure (1) represents three series of the Expected Lognormal Approximation to the Improvement (ELAI) values from three different optimization problems that will be demonstrated later in this paper, where it will be shown that convergence is established near the end of each of these series. These three series demonstrate 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 accuratly identify convergence. However the center panel (the Lockwood problem) 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 with occasional large values indicates that the optimization routine 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 (Rastigin 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.

As the Improvement function is itself a random variable, attempting to use a threshold test oversimplifies the dynamics of convergence. 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.

2 Gaussian Process Surrogate Model Optimization

3 Statistical Process Control

4 EWMA Convergence Chart

References

[1] Sanket Sanjay Diwale, Ioannis Lymperopoulos, and Colin Jones. Optimization of an airborne wind energy system using constrained gaussian processes. In *IEEE Multi-Conference on Systems and Control*, number EPFL-CONF-199717, 2014.

[2] David Ginsbourger, Rodolphe Le Riche, Laurent Carraro, et al. A multi-points criterion for deterministic parallel global optimization based on gaussian processes. 2008.

[3] Robert B. Gramacy and Matthew Taddy. Categorical inputs, sensitivity analysis, optimization and importance tempering with tgp version 2, an r package for treed gaussian process models. *Journal of Statistical Software*, 33(i06), 2012.

[4] Thomas J. Santner, Brian J. Williams, and William Notz. *The design and analysis of computer experiments*. Springer, 2003.

[5] Matthias Schonlau, William J. Welch, and Donald R. Jones. Global versus local search in constrained optimization of computer models. *Lecture Notes-Monograph Series*, pages 11–25, 1998.

[6] Matthew A. Taddy, Herbert H. K. Lee, Genetha A. Gray, and Joshua D. Griffin. Bayesian guided pattern search for robust local optimization. *Technometrics*, 51(4):389–401, 2009.