Robust Convergence in Gaussian Process Surrogate Model Optimization, via Statistical Process Control

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 (SPC) 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: Emulator, Expected Improvement, Exponentially Weighted Moving Average.

1 Introduction

Releasing numerical optimization from its dependence on derivative information, has made black-box optimization a powerful and flexible tool for solving a wide range of engineering problems. Of primary interest here, is the application of derivative-free optimization on computationally expensive black-box objective functions, such as computer simultions. Often computer simulations can function as fast, cost-effective, development test beds for processes that are otherwise too expensive, or dangerous, to test physically. For example, the Lockwood pump-and-treat simulator, discussed in more detail in section XX, computationally simulates the flow of contaminated groundwater alongside the Yellowstone River as numerical solutions to coupled systems cite. The focus of the Lockwood simulation is to optimize the management of this contaminated groundwater flow, so as to avoid contamination of the river. However, the space of possible management schemes is large, and it is hard to identify when derivative-free optimization has converged to an optimal solution. Due to the computational expense of such simulators, it is desirable to do effective optimization with as few simulation evaluations as possible. This is achieved by efficiently exploring the solution space and limiting the number of function evaluations by robustly recognizing convergence.

Gaussian Process surrogate model optimization is a derivative-free numerical optimization routine that reduces the number of neccissary objective function evalutations by using each evaluation as data to construct a statistical surrogate model of the objective function. The surrogate model is relatively fast to work with, as compared with a large computer simulation, and 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 the predictive framework that arises naturally from its Bayesian representation [?]cite. Arising naturally from the GP predictive distributioncite, the maximum Expected Improvement (EI) criterion has shown to be a valuable criterion for assessing future points of exploration on the objective function cite, and 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 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, this use of the EI criterion on GP surrogate model optimization has not yet been adequately justified. In fact, this use ignores the nature of the EI criterion as a random variable, and oversimplifies the stochastic dynamics of convergence in this setting. Thus it is no surprise that this treatment of the EI criterion results in an inconsistent stopping rule as demonstrated in Figure (**??**).

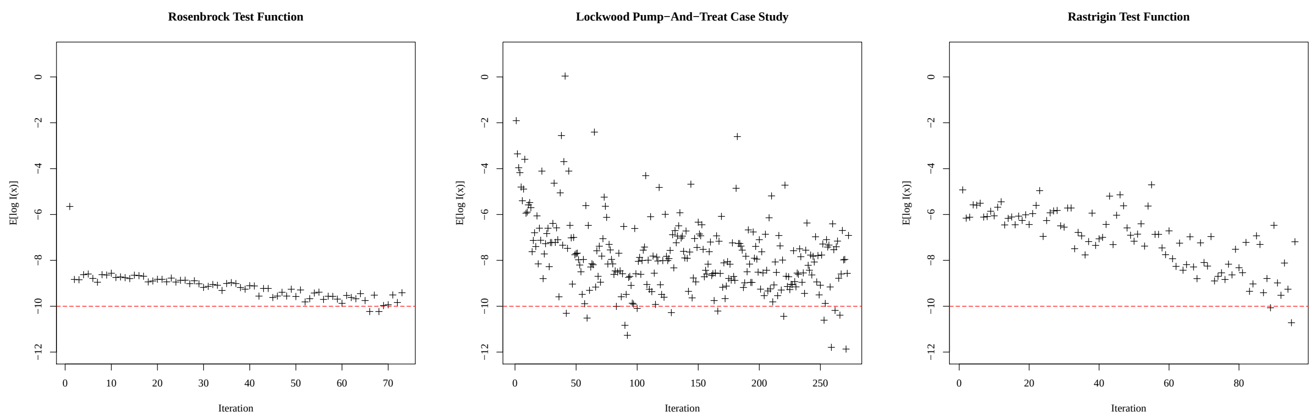

Figure 1: Three EI series shown as Expected Log-Improvement (ELI) plotted against iteration number. ELI values plotted alongside an example convergence threshold value shown as a dashed line at -10.

Figure (??) represents three series of EI values, all of which only manage to achieve convergence within the final quarter of values. These three series demonstrate various EI convergence behaviors, based on GP surrogate model optimization of three different objective functions. In each panel the y-axis represents a monotone transformation of the basic EI criterion (i.e. ELI) so as to benefit the asymptotic behavior of the improvement criterion's distribution, to be discussed further in Section XX. In the left-most panel, optimization of the Rosenbrock test function results in a nicely well-behaved series of EI values, demonstrating a case in which the simple threshold stopping rule can accuratly identify convergence. However the center panel (i.e. Lockwood) demonstrates a failure of the threshold stopping rule, as this EI series contains much more variance, and thus small EI values are observed quite regularly. In the Lockwood example a simple threshold stopping rule falsly claims convergence, for the first time, within the first 50 iterations of the algorithm, additionally as the algorithm proceeds, several improvements of the objective function are discovered while ELI values repeatedly fall below the convergence threshold.

- Thus it is common to observe stopping rules based on constant threshold values of the EI, with algorithms claiming convergence upon first sight of a maximum EI value below this threshold.
- Recall that the EI is a metric derived from the predictive distribution of the GP surrogate model; as such the EI improvement is itself a random variable.
- Thus such a threshold test of convergence over simplifies the dynamics of convergence in this setting, as quite small EI values should be expected with some regularity based on the particular

topology of the problem and the stochasticity of the criterion itself.

- A theoretically more reasonable approach might be to monitor the EI improvement as a series of values, then search for consistency among the distribution of the most recently observed values as iterations of the routine pass.

- Such a search for stationarity among the series of EI values is certainly a valuable endeavor; however Figure (**??**) shows that this is not enough to ensure robust identification of convergence among a broad range of potential objective applications.

- Figure (**??**) empirically demonstrates that simply employing either a simple threshold test, or a test of stationarity, always incorrectly identifies convergence in at least one of the shown cases.