Stochastic and deterministic algorithms for continuous black-box optimization

Wang, H.

Author: Wang, H.
Title: Stochastic and deterministic algorithms for continuous black-box optimization
Issue Date: 2018-11-01

Summary

The black-box optimization problem is frequently encountered in many applications. For example, the tuning task of a machine learning algorithm or fitting a curve to some experimental data. In the PROMIMOOC project (**PRO**cess **MI**ning for **M**ulti-**O**bjective **O**nline **C**ontrol with industrial partners Tata Steel and BMW group), the optimization problem we are facing is to search for proper control parameters of production processes (for both partners), such that the number of defects generated during the production would be largely reduced. Such a problem is typically referred as an "black-box", as we don't directly model the physical mechanism behind the production process and there is no additional information about its mathematical characteristics (e.g., convexity and continuity) that would be very useful for the optimization. Therefore, the black-box problem is also considered very challenging. Another difficulty arises in the extremely high cost of making trials on the production line: suppose a candidate setting of control parameters (or candidate solution) is proposed by an optimization algorithm. The quality of this setting can only be assessed by applying it to the actual production line and then measuring defect rate in the output. This is typically very costly and risky: when the candidate setting doesn't actually perform well, many defects will be generated, resulting in extra production costs for industrial partners. To solve this problem efficiently and carefully, several fundamental optimization techniques have to be combined in a reasonable way.

First of all, as there is not much mathematical assumptions on the problem, we have to resort to the so-called *stochastic optimization algorithm*, instead of using the traditional optimization techniques from mathematics/operational research. The stochastic optimization algorithm is a class of methods that directly optimize the objective function by solely using the assessment (evaluation) of the candidate solution. Stochastic optimization algorithms are underpinned by the so-called

stochastic variation, which generates (local) random perturbations to modify the current search point. In evolutionary computation, this is typically called the mutation operator. Intuitively, the efficiency of a stochastic variation method greatly affects the performance of the corresponding optimization algorithm. This is the reason why we investigate the efficiency issue of such methods in depth (Chapter 2). As a result of the investigation, we propose a novel stochastic variation method, called *mirrored orthogonal sampling*, which aims at generating random perturbations that cover the search space (subset of $\mathbb{R}^d$) evenly. Both theoretical analysis and empirical study are conducted on the proposed method.

Secondly, because it is very costly to assess candidate solutions, it is common to replace an actual expensive assessment by a machine learning model, which is trained on the historical assessments. Then an optimization algorithm can query the quality of a candidate solution from the model, instead of running the real production process with this solution. Such a technique is called *surrogate modeling*. One big challenge in surrogate modeling is to give a reliable quantification about the uncertainty in model prediction due to the fact that data-driven models usually yield significant errors in prediction. In Chapter 3, we study the well-known *Kriging/Gaussian Process Regression* (GPR) model, that is capable of quantifying the uncertainty. The quantification approach in Kriging/GPR is discussed in detail. When it comes to the application of the Kriging/GPR method to real-world data, we are confronted with the following obstacle: The Kriging/GPR method suffers from a cubic time complexity when dealing with large data sets, limiting its applicability for big data sets. In the reminder of this chapter, a novel algorithmic framework, called *Cluster Kriging* is proposed to tackle this issue. Cluster Kriging is tested on some selected functions and data sets, exhibiting an acceleration of the modeling speed as well as an improved modeling precision.

Naturally, once a good surrogate model is obtained from the previous discussion, the next question is how to use such a model in a reasonable manner such that the uncertainty quantification is taken into account. It is possible to select the most trustworthy solution based on the surrogate model, or alternatively the point that possesses the highest potential to help the optimization procedure if the actual assessment were conducted on it. Such decisions are usually determined through an utility function on the surrogate model, called *infill criterion*. This is the topic of Chapter 4. The difficulty in designing the infill criterion is how to balance the trade-off between the model prediction (exploitation) and the model uncertainty (exploration). In this chapter, we summarize the existing infill criterion and propose a novel infill criteria, called *Moment-Generating Function of Improvement* that allows for controlling this trade-off explicitly and smoothly. Furthermore, the parallelization issue of infill criteria is also considered thoroughly and several new parallelization methods are proposed and tested.

Lastly, we discuss the so-called multi-objective optimization problem: suppose we want to minimize the number of defects generated in the production and maximize the throughout simultaneously. In this case, it is typical not possible to find a setting of control parameters that satisfies both objectives in the same time and thus we have to adopt multi-objective optimization algorithms. In Chapter 5, we aims at designing a multi-objective optimization algorithm that is able to use either the gradient or the Hessian matrix of the objective function. To achieve this goal, the gradient field and Hessian matrix of the so-called hypervolume indicator are derived and studied in depth. As a result, two novel algorithms, namely the hypervolume-based first- (gradient) and second-order (Hessian) methods are proposed and tested.