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

CHAPTER 6

Conclusion

In this thesis, several important aspects of the stochastic optimization are investigated in depth. In **Chapter 1**, the discussion begins with the fundamental definitions on single- and multi-objective optimality (Section 1.1). It is attempted to give rigorous definitions on the procedure of *stochastic optimization*. In addition, a mathematical prerequisite, called *Matrix Calculus* is also described. The research question in this chapter is:

> *Can we generalize the notion of local optimality in the single objective case to multi-objective problems?*

We give a generalization by revisiting the local efficient points and defining the local efficient set (Def. 1.8). Basically, a local efficient set should consist of local efficient points that collectively forms a connected component.

Chapter 2 discusses several issues on generating *stochastic variations* in $\mathbb{R}^d$. Herein, the major concerns are:

> *When using a small sample size in stochastic variation, what is the drawback of the so-called sampling error? How to mitigate such drawbacks? Is it possible to justify our approach both theoretically and empirically?*

Random sampling from the multivariate Gaussian distribution is the most commonly applied stochastic variation. It suffers from the well-known sampling error when the sample size is small. This issue is re-visited in Chapter 2 and an improved sampling method, called *mirrored orthogonal sampling* is proposed to relax the drawback. Both theoretical analysis and empirical study are conducted on the proposed sampling method. After plugged into evolution strategies, this sampling method is, in addition, tested on the well-known BBOB benchmark. The second research question is:

> *The well-known Efficient Global Optimization (EGO) algorithm differs largely from the canonical stochastic optimization algorithms (e.g., evolutionary algorithms). Is it possible to find a unified treatment to incorporate EGO into stochastic optimization algorithms?*

In EGO, new candidate locations are generated by optimizing the so-called infill criterion. This procedure is considered as a special stochastic variation method, called *mutation by optimization*.

Chapter 3 resorts to a different topic, the surrogate modeling. The Kriging/Gaussian Process Regression (GPR) is discussed in depth, where an unified treatment is presented from three different perspectives: 1) the theory on the best linear predictor, 2) reproducing kernel Hilbert Space and 3) Bayesian inference. Our major theoretical concerns are:

> *The most prominent feature of Kriging/GPR is the prediction uncertainty. What does this uncertainty truly characterize? How does the uncertainty related to the inaccuracy in the function approximation?*

Our consideration starts with a clear specification of the modeling assumption: the target function f is a *sample function* of a prescribed stochastic process Y. An optimal linear predictor $\widehat{Y}$ is derived for Y based on the partial information (finite locations) on Y. Then the function approximation is obtained by taking a realization $\hat{f}$ from $\widehat{Y}$. Based on this treatment, it is clearly seen that the Kriging uncertainty is the MSE of the predictor: $s^2 = \mathbb{E}\{Y - \widehat{Y}\}^2$. Moreover, the accuracy of function approximation is related to s^2 (theorem 3.1). When it comes to the application of the Kriging/GPR method, the following question is of our interest:

> *The Kriging/GPR method suffers from a cubic time complexity when dealing with large data sets. It is then crucial to propose methods that reduce the time complexity.*

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. Despite the successful experiments on Cluster Kriging, there is still a lack of theoretical treatment (using one of the three aforementioned perspectives) to illustrate its asymptotic convergence property to the target function. This part should be investigated in future work.

Chapter 4 focuses on the following problem: given a thorough study on the surrogate model, naturally the next step is to investigate how the surrogate model can be utilized in a reasonable manner. The utility of each location on a surrogate model is quantified by a well-defined function, called *infill criterion*. The research question in this chapter are:

> *Since some infill criteria are more explorative and some others are more exploitative, does it make sense to look for the optimal trade-off among infill criteria? Can we design an infill criterion that controls this trade-off explicitly and smoothly?*

Prior to the investigation on the exploration-exploitation trade-off, most of the improvement-based infill criteria is discussed. Among them, two infill criteria, Probability of Improvement (PI) and Expected Improvement (EI) are selected for the investigation, where PI and EI are treated as a bi-objective optimization task. This task is solved using a gradient-based multi-objective optimization algorithm that is proposed later in this thesis (Section 5.2). In addition, to explicitly control such trade-offs, a novel infill criterion, *Moment-Generate Function of Improvement* (MGFI) is proposed as the an extension of all improvement-based criteria. It is equipped with an continuous control parameter t (called temperature), which is enable to scale the exploration behavior smoothly. The other research question discussed is:

> *How to parallelize the EGO algorithm, namely proposing multiple candidate locations in each iteration?*

The goal of parallelizing infill criteria is formulated in the first place. Then several well-known parallelization methods are discussed: multi-point Expected Improvement and Multi-Instance of infill criteria. In addition, we proposed two nee approached: 1) packing several infill criteria as a multi-objective function and thus multiple point could be obtained via the multi-objective optimization. 2) when using a single infill criteria, multiple distinct locations can be found by applying the so-called *niching* technique. Although the proposed parallelization methods seem plausible, some of them remains untested as of the time of writing. It is important to perform systematic experiments on them in the future.

Chapter 5 discusses the numerical multi-objective optimization (MOO). The demands on this topic originate from many numerical multi-objective tasks that arise in the study of stochastic optimization, e.g., the multi-objective treatment of infill criteria in Chapter 4. The research question in this chapter is:

> *How to design a deterministic optimization algorithm for multi-objective problems using either the gradient or the Hessian matrix of the objective function, such that both the convergence rate to the Pareto front and the distribution of non-dominated points are considered simultaneously?*

The contribution in this chapter are three-fold: firstly, we mathematically analyze the so-called *Mixed-Peak* bi-objective test problem. Secondly, the gradient field and Hessian matrix of the hypervolume indicator are derived and studied in depth. Thirdly, two novel numerical MOO algorithms, namely the hypervolume-based first- (gradient) and second-order (Hessian) methods are proposed and tested. Differentiating the hypervolume indicator resolves our research question due to the fact that maximizing the hypervolume leads to the convergence to the Pareto front and a set of well-distributed efficient points. In parallel to the algorithmic development, we also investigate the condition on which the hypervolume indicator Hessian would be singular. The singularity would make the Hessian inapplicable for the optimization. The preliminary proof/results show that the subset (of the search space S) where the Hessian is singular is of measure zero in S, meaning that it is generally safe to apply the Hypervolume Hessian method. However, this theoretical study is conducted in an idealized setting: 1) two points in the approximation set and 2) the objective function is set to the simple sphere model. As the next step, it is necessary to investigate the singularity condition furthermore on more general functions with more approximation points.

Many research directions/proposals can be given by combining the techniques in multiple chapters. For example, the novel infill criteria MGFI (Section 4.3) can be combined with Kriging-based multi-objective optimization algorithm, e.g., SMS-EGO (Ponweiser et al., 2008), replacing the EI criteria. In addition, as the surrogate model is required in SMS-EGO, it is also recommended to use the proposed Cluster Kriging (Section 3.2) with this algorithm, aiming at improving the algorithm running time and convergence rate.

In addition, the theoretical study in this thesis can also be continued. In Chapter 2, although the mirrored orthogonal sampling is analyzed in depth, the distribution of the proposed *uniform random orthogonal vectors* (Def. 2.1) is yet to discover. In Chapter 3, despite the successful experimental result of Cluster Kriging, the theoretical aspects are not treated. Thus, it is quite important to check its modeling ability from the perspective of either Bayesian inference or RKHS. In Chapter 4, when combining the EGO algorithm and the proposed MGFI criteria, the convergence rate of the resulting optimizer is theoretically unknown, although the faster empirical convergence is validated from benchmarking.