Page 1

Hi everyone, my name is Muyi Song. Today I’d like to introduce my Master project to you which is [Machine learning assisted cyclic coordinate descent algorithm for scalable black-box optimization]

Page 2

I’m going to present my project in these five parts.

Page 3

At the beginning, let me introduce some basic background information to you.

So what is black-box optimization? For example, if a set of power plant layouts is given, how do we find out the best layout in terms of capital investment? Another example, if a series of design variables of a machine is given, such as reactors or containers, how do we know its life? How do we figure out the connection between those variables and the damage? Actually we can find there are some common problems in these situations:

1. Lack of physical understanding about the model.
2. The cost is unaffordable. We do know that layouts of industries are designed on purpose. Equipment tend to be built at downwind places to prevent the diffusion of dangerous gases for process safety. However, we have no idea if there is any direct relationship between layouts and benefits.
3. Sometimes noisy statistics is obtained from experiments. But we don’t know where is the noise coming from.

All of the situations I referred above lead to an unknown objective function during optimization. So we can only input a vector into black boxes and get an output value without any information of inner structure of the boxes.

Page 4

When we speak of traditional optimization problems, a mathematical expression of objective function is always available. Then what we need to do is to find the minimum or maximum under constraints. Several methods like random search method and a bunch of gradient-based search have been proved that they can solve traditional optimization problems efficiently and effectively.

But two major problems make black box optimization problems more complexed and unsolvable. Firstly, gradient information is not accessible which is usually used to decide the direction to go in the next iteration. Secondly, function evaluations are always extremely expensive. So how to find the optimal value within limited number of evaluations has become the primary goal for researchers.

Page 5

This the flow chart showing the logic of my proposed algorithm. After initialization of starting point and the range of search, the algorithm is sampling a group of data points to build a surrogate model. ALAMO is implemented here to build the model, which accepts input data from the algorithm to build a surrogate model by using adaptive sampling technique repeatedly until convergence. In this project, quadratic based model is used. A mathematical expression of the model should be returned which will then be sent to BARON software which is a mixed-integer nonlinear optimization solver to get a local optimal solution. This process is repeatedly applied until meeting the termination condition. Otherwise the range of search and number of sampling points should be modified to make a more accurate model in the next iteration.

Page 6

I used cyclic coordinate descent as searching algorithm. The main advantage is that it is very easily applied. And it is flexible to be integrated with other tools.

The simple process of cyclic coordinate search can be visualized like this figure. It changes a chosen coordinate at each iteration. The most natural approach to choose an index is to select components cyclically. Regarding the update schemes, I simply renew by minimizing the objective with respect to the specific variable while fixing the others.

Page 7

Another problem is if we can build an accurate surrogate model with limited sampling point, which requires each point is able to contribute to the final solution. Because we assume that each evaluation is expensive. Intuitively, I’d like to test the performances of different sampling methods, which is one of the most important factor in modeling. The first image is generated by python random library. You can see that points are actually not distributed evenly in local space. In comparison, four types of quasi-random sequences are selected including Halton, Van Der Corput, Hammersley and Sobol sequences. Another sampling method is Latin random square.

Page 8

100 black box models were chosen from multiple libraries to with different smoothness, convexity and number of variables. The distribution is shown on the plot. All of these information is available before solving the problems.

Page 9

A realistic industrial case is selected as well. We know that process system simulation has been applied in chemical engineering in terms of efficiency and economic feasibility. Software like Aspen has been widely used in companies and academic systems. But their inner structures were unavailable for users like black boxes. The development of simulation optimization is increasingly important recently. The single mixed refrigerant liquefaction process is one of the most famous technique for offshore natural gas fields development. In this project, the proposed algorithm was trying to optimize a base case of SMR aiming to find the best configuration in terms of the expense.

Page 10

These summary plots show how many test problems has been solved by the algorithm with a specific number of evaluations with different sampling methods.

From the figures we can find that the proposed algorithm shows poor ability of solving nonsmooth-convex problems. With increasing evaluations, better solutions were obtained for smooth-convex and some nonsmooth-nonconvex problems which are represented by yellow and green lines.

Although it seems like there is no big difference between them, we can still find that Sobol sequence has shown the fastest speed of convergence among all of the methods, and Latin random square proved that it can solve a part of the problems consistently given more evaluations.

Page 11

The plots on the left hand side show the whole process of convergence for some typical problems with different properties.

We can see that an obvious convergence process can be observed in smooth-nonconvex and smooth-convex problems. I set a limitation for the number of evaluations. But if more evaluations were given, the global optimal points can be expected. On the contrary, the algorithm tends to be stuck in local optimal points in nonsmooth models.

Difference between sampling methods can be found in the plots, which is not so obvious. Because we focused on single dimension modeling and some methods are designed for higher dimensions like Halton, Hammersley and Latin random square. Halton and Hammersley derived from Van Der Corput sequence which is one of the most classical low discrepancy sequences generated from radical inversion. In comparison, Sobol sequence not only has a good distribution over the sampling space, but also it shows high-qualified sampling performance when the number of points is integer power of two. The red line represents the Sobol sequence. It has the fastest convergence speed in these problems.

Page 12

Then I applied the proposed algorithm on the SMR simulation. However, I found the search progress was kept being stuck after hundreds of evaluations and cannot find the global optimal solution. I looked into the source code and some paper trying to find out the reasons.

There is a big difference between SMR simulation and previous black box models. Discontinuous hidden constraints and pre-defined constraints coming from physical configuration of the technique have been assumed. An extreme big number should be returned as long as those constraints are violated. In this situation, our algorithm did not perform well with the existence of those noisy data points. It cannot capture the pattern of the simulator with surrogate modeling. Consequently, the range of search was adjusted based on the accuracy of last iteration’s modeling. The algorithm tends to be stuck at the local optimal when the space has many infeasible solutions.

These problems may possibly be solved with the application of global search. We can sample data from the whole range of model and filter the infeasible points out. Block coordinate search method is another potential solution to correlated variables which update multiple components simultaneously in each iteration.

Page 13

This project is mainly focusing on implementing an existed algorithm and tools, and test their performances with different combinations. Although I did not get perfect results at the end, several conclusions can be drawn based on my experiments.

1. Surrogate modeling has been proved that it is a flexible and useful method in black-box optimization problems
2. Different sampling methods has shown different performances depending on the properties of the models
3. Sobol sequence has shown the best performance yet combing with cyclic coordinate descent and 1-d modeling
4. Proposed algorithm is not able to solve the problems with discontinuous hidden constraints and correlated variables. More techniques are expected to be applied.