Optimization and Modeling of Black Box Functions with Multiple Local Minima

Author: Tatsiana Ivonchyk
Supervisor: Prof. Dr. rer. nat. Marc Toussaint
Course of Study: Electrical Engineering and Computer Science
Home University: Massachusetts Institute of Technology

August 7, 2014
1 My Research Project

1.1 Motivation

Optimization is a field of mathematics present in many applications. In particular, the goal is often to find an extrema of a function for which a closed form expression is not known and no information about the gradient or Hessian is known. This problem is known as optimization of a black box function. Knowledge of the function is limited to function evaluations, which can be noisy. In addition, function calls can be expensive motivating the goal of using the minimal number of evaluations possible. Finding the global minimum of a function is a more complex problem than simply locating a local minimum as a balance between exploiting areas where a low function values are expected and exploration of areas of input with high uncertainty must be found.

Bayesian optimization uses a prior belief and the data collected to motivate a posterior belief of the function. Gaussian Processes have commonly been used as priors in Bayesian optimization. In a Gaussian Process, every combination of random variables has a normal Gaussian distribution. The motivation of this project was to explore whether Bayesian Optimization could be approached with a simpler prior. The goal was to develop a method which would perform optimization on a black box function as while creating a representation of the unknown function. The functions that the method focuses on working with are functions with several local minima. A quadratic model was chosen as the prior because it is simple, is guaranteed to have a minimum, and the interest is in representing areas of local and global minima of the function.

1.2 The Method

The optimization method is focused on using a quadratic prior to create local models which will allow for the global optimal to be found. A quadratic feature vector with \((n+1)(n+2)/2\) features is used for an n-dimensional function. The method begins by sampling \((n+1)(n+2)/2\) points from the function. \(2^n\) points are chosen to be sampled in the area of the outer corners of the bounding square, cube, or hypercube. The remaining \((n+1)(n+2)/2 - 2^n\) points are sampled randomly from the bounded input space.

An iterative loop is performed during which the next location to query the function is chosen. At the beginning of each iteration all of the sampled points are checked for being a local minimum. This check is done by first performing a Delaunay triangulation of the space and testing whether the point in question has a smaller function value than its neighbors. If the point is indeed a local minima, then Ridge Regression is utilized to compute the optimal parameter \(\beta\) to model each of the local minima and their neighbors.
with higher function values. A local model using those data points is created. Only the quadratic models with a positive definite Hessian are kept so that each of the local models have distinct minima.

In each iteration, the next point to be sampled is chosen by sampling at the location with the expected lowest function value based on the existing local models. If this point is within a threshold distance of an already sampled point, the algorithm moves on to the next best local model and attempts to sample at the minima of that model. If through this process, the local models are exhausted, or no local models existed from the beginning, then the method samples randomly.

The method converges when the next point to sample is within a threshold distance of all of its neighbors.

**Algorithm 1** Model and Optimize Black Box Function with Multiple Optima in $n$-dimensions

**Require:** $n \geq 0$ and search space is bounded

1: KeptModels is an empty set which will store Local Models
2: Sample $\binom{n+1}{2} \cdot \binom{n+2}{2}$ data points to span search space
3: while $\forall$ sampled $x_i, |x_i - x_j| < \epsilon, x_j$ in Delaunay neighbors of $x_i$ do
4:    Check each data point for being local minima using Delaunay neighbors
5:    $\forall$ minima, calculate $\beta$ using using Ridge Regression
6:    Add local model to KeptModels set
7:    if KeptModels is empty then
8:        $x^+ = \frac{x_a + x_b}{2}$ where $x_a, x_b$ are $\text{argmax} |x_a - x_b|$
9:    else
10:       Sort KeptModels by function value of minima of each Local Model
11:      $x^+ = \text{argmin}$ best Local Model
12:     while $|x^+ - x_i| < \epsilon, \forall x_i$ in sampled points do
13:        Move on to next best Local Model in KeptModels
14:     if KeptModels exhausted then
15:        $x^+ = \frac{x_a + x_b}{2}$ where $x_a, x_b$ are $\text{argmax} |x_a - x_b|$
16:     end if
17: end while
18: end if
19: Sample at $x^+$
20: end while
1.3 Results

1.3.1 Rastrigin Function

The method described was tested and graphed for both one-dimensional and two-dimensional input spaces for the Rastrigin function

\[ f(x) = 10n + \sum_{i=1}^{n} x_i^2 - 10\cos(2x_i\pi) \]

Figure 1: 1D Rastrigin Function. Each graph represents one iteration.
After 15 iterations, three local minima are modeled by the quadratic models. In the last frame, two other quadratic models (colored purple and cyan) can be seen to span several local minima. These would likely converge to closely fit a local minima in several more iterations of the method.

Figure 2: 2D Rastrigin Function. Each graph represents ten iterations.

Figure 2 shows the method on the 2D Rastrigin function after 80 iterations. In the last frame, three quadratic models exist, focusing on three different local minima of the 2D Rastrigin function.

1.3.2 Combination of Gaussians

In Figure 3, the real function is defined as a mixture of five Gaussians with means at -6.0, -2.0, 0.0, 8.0, and 11.0 with standard deviations of 1.0, 1.2, 1.5, 1.0, 0.9 respectively. After 15 iterations, the local models maintained by the method represent the three local minima of the function defined by the sum of Gaussian distributions.
Figure 3: Mixture of 1D Gaussian Distributions. Each graph represents one iteration.
1.4 Summary and Future Work

The method proposed uses the Bayesian Optimization method with a simpler prior of a quadratic function. On the tested functions, the method exploited areas of believed low function values until the local area was well-modeled, at which point the method explored the areas of the input space with the most uncertainty. On the functions tested, the method performed as desired, maintaining a representation of the true functions’ areas of minimal value through several local models. Future work could include developing a method of combining the local models into a model of the entire function. In addition, a stopping criterion could be specified which requires fewer function calls than one which is based on the distances between sampled points, as this is essentially a brute-force method of sampling the entire input space.

1.5 My Experience in the Lab

During my stay at Universität Stuttgart, I was given the amazing opportunity to experience what it is like to be a researcher. This summer I gained experience about how to work more independently. In my previous research experiences, I had always been given more specific instructions or tasks to complete. During this project, however, my supervisor and I would discuss generally the next steps to take in and the specifics would be up to me. I was responsible for deciding how to write the code, which packages or libraries were best to use, and how to tackle any emerging challenges in the developing method. During my summer, I further developed my programming skills, learned a lot about Optimization and Machine Learning, and have gained research skills which will be necessary for my future success.

I was the only undergraduate student in the Maschinelles Lernen und Robotik lab. In addition, I was only the second female who was a member of the laboratory group consisting of over ten people; Carola Stahl, Professor Toussaint’s administrative assistant is the only woman who is a permanent member of the lab group. Being a ‘minority’ in the laboratory group did not faze me, but rather motivated me to continue working hard in my field. I gained a lot of knowledge about research, conferences, and publication in the fields of Machine Learning, Robotics, an Optimization through conversations with the Postdoctoral researchers, PhD students in the group, and Professor Toussaint. I went out to dinner with the group several times after work and we had a goodbye barbecue in the forest. These events allowed us to talk about topics beyond the scope of our research and I got to know the members of the group on a more personal level.
2 Personal Impressions of My Stay

2.1 My Travels

During my two month stay, I travelled to cities in Germany, Austria, and Switzerland during the weekends. I visited Tübingen, München, Berlin, Heidelberg, Düsseldorf, Zürich, and Wien. During my travels, I experienced new customs, foods, and clothing styles that I had never seen or experienced before. I saw paintings, sights, and castles that I had written essays about in my German language course at MIT. I feel that I became a better traveler and have broadened my knowledge about the different cultures of the world. I feel that my experiences are invaluable and the impressions that these places have made on me have changed me for the better: I think that seeing different ways of approaching life and the problems that the modern world faces is vital if I am to one day try to help change the world for the better.

2.2 My Living Situation

I lived in a student dormitory on the Vaihingen campus of the university. During my two month long stay I became very close friends with my neighbors. The dormitory where I resided was a very international community. I shared a kitchen with students from Germany, Greece, Italy, Spain, Colom-
baja, Brazil, Mexico, and Turkey. It was an amazing experience to become a member of this community because I was able to closely experience and learn aspects from all of these different cultures. As a group, we cooked, explored the city of Stuttgart, and planned and hosted various events together. As several of my neighbors were participants in the Erasmus Programme, I met many other students who were participants of this program and lived in different dormitories on the campus. I had a great time talking and spending time with these students from all over the world since at my home university, international students do not comprise a large percentage of the student population.

Figure 5: Dinner with My Neighbors

2.3 Noted Differences between Germany and the United States

My impression of Germany is that it is a country which is better organized and environmentally conscious than the United States. The first and perhaps most obvious example of this is the public transportation system. In New York City, the subway system is completely unpredictable. There are no timetables for the trains and no screens announcing when the next train will be arriving. In Stuttgart, the trains and buses that I took were almost always on schedule. If there was a delay in service in Germany, I was notified in advance of the circumstances. In addition, the subway trains in Stuttgart
had special compartments for trash by the seats. This contributed to the trains being much cleaner than the ones I am used to taking at home.

The pfand system and presence of recycling bins and bottle deposit machines throughout the city made it extremely easy to be environmentally-friendly. In addition, paying for grocery bags in supermarkets is not a practice common in the United States. These differences made me more conscious of how much plastic I had been wasting in the United States and show practices that I feel the United States needs to implement if we are to attempt to slow down global warming and the destruction of the planet’s resources.

3 Conclusion

My experience conducting research at the Universität Stuttgart through the SUPER program is unforgettable. I learned so much about both Machine Learning and Optimization, as well as about myself as a traveller and person. I learned about how other countries and cultures approach the problems of the world and my previous beliefs were challenged through this new knowledge. I enjoyed my stay in Germany so much that I am seriously considering searching for employment in the country after I finish my studies at MIT. I would like to sincerely thank everyone who supported me and made this summer experience a reality!