*P. Koumoutsakos*
*ETH Zentrum, CLT E 13*
*CH-8092 Zürich*

# HW 3 (Part 2 of 2): Applied Sampling with Korali

Issued: March 25, 2019
Due Date: April 1, 2019 10:00am

## Warm-Up & Information

This homework requires the use of *Korali*, our new Uncertainty Quantification and Optimization framework developed at ETH Zürich. Along with this pdf, you should download the provided code which contains the following folders:

- *korali/* contains a ready-to-use, pre-compiled version of Korali. The library and some of the task models will only work on Euler (or compatible ETH systems).
- *example/* contains examples based on the Ackley function. These examples contain all the Korali use cases required to solve this homework.
- *plotting/* Contains Python and MATLAB codes required to plot the distribution plots based on the output of Korali's TMCMC engine.
- *task1/* Contains the skeleton codes and model for Task 1.
- *task2/* Contains the skeleton codes and model for Task 2.

## Guidelines for Moodle submission:

- For HW3 Part 1: Two solution codes (recommended: in python).
- For HW3 Part 2: 5 solution codes: task1a.cpp task1c.cpp task1d.cpp task2a.cpp task2b.cpp, containing your Korali-based solutions.
- The final version of your report PDF which include the solutions to both parts of Homework 3. Do not forget to answer every question in the HWs.

**Example: The Ackley Function.**

The *Ackley* function is a well-known test case for optimizer algortithms. It contains many local minima in which an optimizer can get 'stuck', but only one global maximum. In the *ackley/* folder, we provide 6 use case for Korali. Feel free to take a look at them and use them as basis for your solution codes.

**Case 1**: We want to determine the values of the $(x_1, x_2)$ parameters that represent the global minimum for the 2D Ackley function. The code in *minimize_function.cpp*, uses Korali's CMAES engine to find these values. Since CMAES maximizes $f(x)$, we use *maximize_ackley()* as model for Korali, which returns *-Ackley()*, instead. Note: In this case, we are given the values for $(a, b, c)$ which define the 'shape' of the Ackley function, so these are not parameters to infer.

**Case 2**: The code in *sample_function.cpp*, uses Korali's TMCMC to generate tmcmc.txt, which will give a better idea of the distribution of Ackley around its global minimum, along with the marginal distributions of the parameters. Use **plotmatrix_hist** to plot the distribution graph.

**Case 3**: In this case, we do not know anything about the values of the $a, b$ and $c$ parameters that shape the Ackley. Instead, we are given a list of measurements with the $(x_1, x_2)$ position and their $f(x_1, x_2)$. We need to use these measurements to infer the values for $a, b$ and $c$. The code in *maximize_likelihood.cpp* will load this information from *data.in* and use it as reference data for Korali to use. The model, *likelihood_ackley*, will store the values of $f(x_1, x_2)$ for every point in the list, given $a, b, c$ samples. This code will produce the values of these parameters that maximize their likelihood given the data.

**Case 4**: The code in *sample_likelihood.cpp*, uses Korali's TMCMC to sample and give a better understanding of the distribution of the parameters that give the maximum likelihood.

**Case 5**: In this case, we calculate the posterior distribution of the $(a, b, c)$ parameters, given some prior information. We are told that parameter $a$ follows a Gamma distribution, while parameter *c* follows a Gaussian distribution. No prior information given for $b$. The code in *maximize_posterior.cpp* uses Korali CMAES to find the parameters that maximize the posterior for (*a,b,c*).

**Case 6**: The code in *sample_posterior.cpp*, uses Korali's TMCMC to sample and give a better understanding of the distribution of the parameters that give the maximum posterior.

## Task 1: Herr Küheli's Farm. (40 Points)

Herr Küheli's dairy farm is located deep within the hills of the Appenzell canton. His farm, famous for its delicious milk and cheeses, is home to hundreds of cows that freely graze around its grass fields and slopes. Once a year, during the month of September, the farm is visited by the Swiss Federal Food Safety and Veterinary authorities to ensure the good health and proper treatment of the cattle. Herr Küheli takes great pride in his cows, and eagerly awaits the opportunity for quality certification.

Figure 1: A painting of Appenzell by Ulrich Martinelli Druck

During the last few years, however, certification visits have been problematic. Given the vast square territory ($5\text{km} \times 5\text{km}$) enclosed by the Farm, the task of finding the herd has proven to be quite a challenge, taking hours to locate the majority of the cows. Authorities have warned Herr Küheli that such delays should not repeat during their next visit. This year, they have given him a maximum of one hour to locate at least 150 cows for examination. If he fails to achieve this goal, he may lose his certification.

To ensure a successful visit, Herr Küheli has hired you, an expert in optimization and sampling methods (and skilled Korali user) to help him find a way to quickly locate his cows. Your job is to determine the best spot in the farm to look for cows. He will personally go to each spot you suggest and let you know how many cows he has seen (given his old age, his sight is limited to a $200m^2$ radius). Each trip takes him 5 minutes. Knowing that you only have a limited number of trials, you inquire about cow grazing behavior and their possible location. He gives you a hunch: "Cows will always graze where grass grows tall. They must surely be all around the spot of the farm where grass grows taller!".

Upon consultation with local agricultural engineers, you find that they have a model for grass growth at the local area. The model considers the capricious topography of the farm and outputs the estimated height of grass given a (x,y) coordinate pair. To produce correct estimations, the model requires two input parameters: *ph* – the current pH acidity of the ground, and; *mm*, the amount of rain fallen in the last month, measured in millimeters. Fortunately, the exact values for this parameters are well known: "Last month we had 80mm of rain and the soil under the farm currently has a pH of 6.0"

a) (10 Points) Use Korali to find the X and Y coordinates of the spot where the grass grows taller. Send Herr Küheli to check for cows to that spot using the *$check_cows* command. Write down the result of this first attempt along Herr Küheli's feedback.

   **Syntax:** $check_cows X Y, the coordinates of the farm to check, in km, from 0.0 to 5.0.

b) (10 Points) Seems like you have found a bunch of cows in your first attempt. Confident you are on your way to success, you decide to follow Herr Küheli's suggestion as to where to look for the rest of the cows. Write in your report the location of these spots and try them. Answer in your report: why do you think they were suggested? Did you achieve the results you were expecting?

c) (10 Points) Use Korali's sampling engine to gain a better understanding of grass growth at the Farm. What new points would you instruct Herr Küheli to look for? Explain your decision. Were you able to find all 100 cows within an hour? Explain why the initial strategy was not effective, and why current your strategy provided better results.

d) (10 Points) Frau Kleineblume is planning to build a pumpkin garden right besides Herr Küheli's farm. The optimal range of soil pH for pumpkins to grow is within $5.5 - 7.5$. Additionally, these plants require constant irrigation, and therefore can only thrive under a rain volume of $> 100mm$ per month. Unfortunately, none of the locals know the value of these parameters for the current month of March, not even Herr Küheli.

   Local experts are able to give you some clues: "The pH of the soil never goes below 4 or above 9. Also, historical trends indicate that at this point of the year, rain volume averages 90mm with a 20mm standard deviation"

Once again, you are summoned to aid the locals with their endeavors. Armed with the knowledge you acquired on HPCSEII, you ask Herr Küheli to inform you about the height of the grass at many spots in his farm as possible. He assembles this information into a file called *grass.in*.

Use Korali to infer the likely values for the pH of the soil and rain volume. Write the rationale and steps you took to obtain these values. How did you integrate the expert's clues in your calculations? What can you say about the uncertainty of your results? Would you FULLY recommend (or not recommend) Frau Kleineblume to plant her pumpkins at this point of the year? Explain your decision.

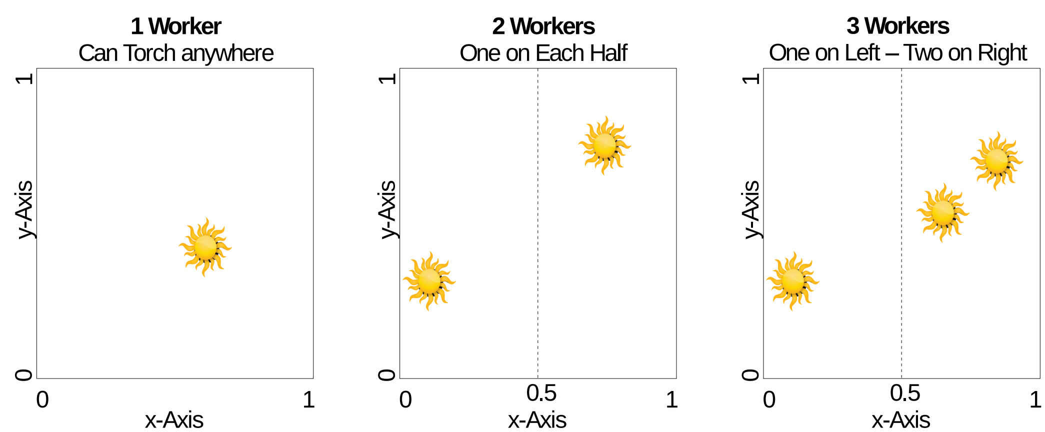
### Task 2: n-Candle Problem. (40 Points)

As the newly appointed head of operations in an up-and-coming electric car startup company, it is your job to ensure that all manufacturing processes are as efficient as possible. After all, the success of the company depends on the quality of their cars.

Not long after you take your post, you receive concerning news from the technical director of the chassis department. Apparently, cars are leaving the factory with inconsistent build quality. Some of them never report a failure, while others suffer from irreparable structural damage.

"The problem is the metal sheets we are using to build the chassis. Some of them seem to have a better heat treatment that make them resistant, while others turn out brittle". After a visit to the metal sheet manufacturing section, you find out that it is severely understaffed. Indeed, they have only 3 employees to apply heat treatment to all the metal sheets used in the factory.

Metal treatment consists of heating one or more sections of the 2D metal sheet using a torch (or candle). When demand is very high, each worker will treat different metal sheets by applying a torch anywhere in the plate. When there's average demand, two employees may collaborate on treating the same plate, by each working on their half of the plate (as divided by the $x$-axis). Finally, when demand is slow, all three employees will work together on each metal sheet, with one employee working on the first half, and two on the second half. As a result, there are 3 possible configurations by which a metal sheet could be treated:




Unfortunately, the current treatment process does not keep track of how many workers treated each sheet, nor the exact position where they applied their torch.

We have, however, an automatic machine that registers the temperature at multiple points of each sheet, just after they leave the treatment facility.

You receive a new report indicating that Sheet #004392 seems to have successfully passed all structural integrity tests, with outstanding results over all other sheets. Your technical director proposes: "We should replicate the same heat treatment that this sheet received to all the new sheets from now on. That would solve our problem.".

a) (20 Points) Armed with the temperature measurements for Sheet #004392 (file: *data.in*), use Korali to indicate:

   - Describe the model, solver, and parameters you will use to represent this problem in Korali. (Hint: the *p.nCandles* variable lets you choose which model to use, either 1-Candle, 2-Candles, or 3-Candles.)
   - What is the most likely number of workers employed in the treatment of this sheet?
   - What metrics did you use to verify which model fits the data better?
   - What is the most likely $(x, y)$ position each of them applied their torches.

b) (20 Points) Thanks to your assessment, treating metal sheets with the new procedure has the quality of the cars chassis has greatly improved. However, none of the new sheets seems to match the exceptional quality obtained by sheet #004392. It seems like there is some piece of information missing. Intrigued, you approach the metal workers in search for more clues.

   The workers report that: "We seldom torched the sheets close to the edge. Rather, we try to stay more or less close to the center of our section, horizontally speaking.". After analyzing their report, you realize that the x-axis position of their torches has a mean of 25cm (if working in the first half of the plate), and 75cm (if working on the second half), with a 5cm standard deviation in either case. No additional information about their position in the y-axis. Answer:

   - What would you change in your model to better reflect this new information?
   - What is the new recommended values for the position of the torches?
   - Use Korali to compare the *Evidence* with or without the prior information presented by the workers. What conclusion can you make from this comparison?