Problem Set 3

Issued: Monday, October 8, 2018 **Due:** Sunday, October 21, 2018

Part I - Theory

Problem 4.1: [10pts] Gaussian processes.

(a) Prediction with Gaussian processes can become computationally very expensive when we have lots of observations. Why?

(b) (*optional for undergraduates*) How could we make this computation more efficient? Suggest a solution and an algorithm (including pseudocode).

Hint: there are many possibilities, e.g., shrinking the data or using the Sherman-Morrison-Woodbury formula. If you follow any of these or your own idea, say how exactly you would do it and why, and how you would ensure to not lose too much prediction quality. Your predictions will be approximations to predictions using the full, expensive Gaussian process. You may use the form of prediction with noisy observations.

As a further hint, if you want to use the Sherman-Morrison-Woodbury matrix inversion lemma, here is a form that is useful for the problem:

$$(Z + UWV^\top)^{-1} = Z^{-1} - Z^{-1}U(W^{-1} + V^\top Z^{-1}U)V^\top Z^{-1},$$

where $Z \in \mathbb{R}^{n\times n}$, and $U, V \in \mathbb{R}^{n\times m}$.

Part II - Ocean Flow

The Philippine Archipelago is a fascinating multiscale ocean region. Its geometry is very complex, with multiple straits, islands, steep shelf-breaks, and coastal features, leading to partially interconnected seas and basins. In this part, we will be studying, understanding and navigating through the ocean current flows.

The data set may be found in `OceanFlow.zip`. It consists of the ocean flow vectors for time T from 1 to 100. The flow in the data set is an averaged flow from the surface to either near the bottom or 400m of depth, whichever is shallower. It is thus a 2D vector field. The files `*u.csv` contain the horizontal components of the vectors, while the files `*v.csv` contain the vertical component. The numbers in the file names indicate the time. For instance, files `24u.csv` and `24v.csv` contain the information of the flow at time 24. The file `mask.csv`, if needed, contains a a 0-1 matrix identifying land and water.

Additional info and units: The data has been collected in January 2009. Multiply the flow values by 25/0.9 to get a unit of cm/second (cmps). The time interval between the data snapshots is 3hrs. The grid spacing used is 3km. The matrix index $(0, 0)$ will correspond in this problem to the coordinate (0km,0km), or the ***bottom, left*** of the plot. For simplicity, we will not be using longitudes and latitudes in this problem.

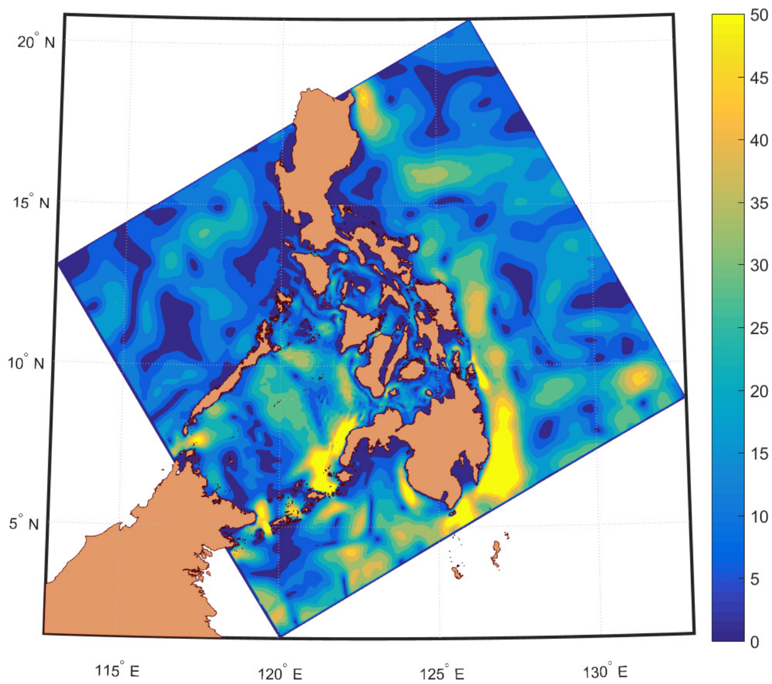

Figure 1: Snapshot of the ocean flow speed in the Philippine Archipelago.

The data has been provided by the MSEAS research group at MIT (`http://mseas.mit.edu/`). The flow field is from a data-assimilative multiresolution simulation obtained using their MSEAS primitive-equation ocean modeling system. It simulates tidal flows to larger-scale dynamics in the region, assimilating a varied set of gappy observations.

Problem 4.2: [30pts] Flows and correlation.

We first study spacial correlations in the ocean flow.

(a) Describe the average flow (averaged over all times, and not location). Are there any constant flow currents that run in the archipelago? Compute the needed characteristics to explain your description. Again, remember, the matrix index $(0,0)$ will correspond in this problem to the coordinate (0km,0km), or the ***bottom, left*** of the plot.

(b) Describe the speed of the average flow (again averaged over all times). Compute the needed characteristics to explain your description.

(c) Visualize the evolution of the flow and its speed over time. Do you observe any spatial correlation?

Problem 4.3: [30pts] Predicting trajectories.

The goal of this problem is to simulate the trajectory of a particle moving in the flow.

(a) We assume that a particle in the ocean, with certain coordinates, will inherit the velocity correponding to the flow at those coordinates. Implement a procedure to track its position and movement caused by the time-varying flow. Explain the procedure, and show that it works by providing examples and plots.

(**Hint / suggested approach:** The data provides a discretization of the ocean flow. The particles will however be moving on a continuous surface. For simplicity, let us assume that the surface is the plane $\mathbb{R}^2$. The data can be seen to provide flow information at integer points, namely at (m, n) for m and n integers. Divide the continuous surface into squares, in such a way that each square contains a unique data point. One way to achieve this is to assign to every point in the surface the closest data point. For instance, given $(x, y) \in \mathbb{R}^2$, this consist of rounding both x and y to the closest integer. You may then suppose that each square has the same flow information as the data point it contains.

Now take a particle at (x, y) in a certain square. The flow in the square will displace it at the registered velocity. Once the particle moves out of this square, it is then governed by the new squares' flow information.)

(b) A (toy) plane has crashed in the Sulu Sea at $T = 0$. The exact location is unknown, but data suggests that the location of the crash follows a Gaussian distribution with mean $(100, 350)$ (namely $(300km, 1050km)$) with variance σ^2. The debris from the plane have been carried away by the ocean flow. You are about to lead a search expedition for the debris. Where would you expect the parts to be at 48hrs, 72hrs, 120hrs? Study the problem varying the variance of the Gaussian distribution. Either pick a few variance samples or sweep through the variances if desired. (**Hint:** Sample particles and track their evolution.)

Problem 4.4: [30pts] Path planning

The goal of this problem is to route a boat through the ocean water. We will consider a fixed current flow in this problem, specifically that of `40u.csv` and `40v.csv`.

The vehicle is initially positionned at (x_0, y_0). The goal is to plan a travel route for the vehicle to arrive at (x_f, y_f). The vehicle is equipped with an engine, that can be turned on and off. When on, the engine provides the vehicle with a velocity vector of magnitude V in any direction. When off the velocity vector is 0. The velocity V is fixed, and cannot vary. The current flow of the ocean however affects the velocity of the vehicle. If vel_{flow} and vel_{engine} are the velocity vector of the flow and the engine respectively, then the velocity vector of the vehicle is $vel_{flow} + vel_{engine}$. Of course, the vehicle can only travel by water, and is not allowed to cross over land.

We will set $(x_0, y_0) = (70, 400)$ and $(x_f, y_f) = (360, 170)$.

(a) Devise and implement a scheme to plan a route that minimizes travel time. The vehicle consumes 1 unit of fuel for every 1 unit of time the engine is on. Plan a route from $(x_0, y_0) = (70, 400)$ to $(x_f, y_f) = (360, 170)$. How does the route vary for different values of V? You are not required to find the optimal solution, but a very good solution.

(**Hint / suggested approach:** Approximate the continuous surface with a collection of points. If the surface is $\mathbb{R}^2$, then consider for example the grid $\mathbb{Z}^2$ of integer points. Between any two integer points find the shortest route in the appropriate sense, and add an edge with weight equal to the length of the route. One then gets a graph with edges whose weights correspond to the routes lengths. You may then run a shortest path algorithm. The closer the grid points are, the better estimates you obtain.)

(b) (**Optional for undergraduates**) Describe a potential scheme that would compute the shortest path in a time varying flow. Specifically, reconsider (a) while working with the whole data set. Explain the different pieces of the algorithm.