Journal

Fan Zhuo

Spring 2008

Week 1 (03/31/08-04/06/08)

This is the first week for the spring quarter. My first task this quarter is to continue my study on the 'regression' paper and to understand the programs for the examples in this paper. The second task this quarter is to apply the regression analysis to the GRACE system. I plan to finish the first task in first 3 weeks and the second task in 6 weeks.

This week, I also read some materials about the Regression Analysis, including the linear and non-linear regression. This helps me to understand the 'regression' paper.

Week 2 (04/07/08-04/13/08)

This week, I am looking for the data that can be used in our regression analysis. Actually, there are three types of data from JPL. The third level data are used in some specific areas, such as Oceanography, Glaciology and Hydrology. The second level data refer to monthly estimates of spherical harmonic coefficients of the Earth gravity field. Since these data are estimations, we can only use first level data.

The first level data are complicated and need a Linux system to learn. I might need some time to get used to the new system in order to understand and use the first level data.

Week 3 (04/14/08-04/20/08)

The level 1B data are the only data available to apply the regression analysis. However, after reading the user handbook for these data, I find I can not directly use

them and a huge project is needed to process these data and translate them into the forms I can use. Hence, this difficult stops me from doing further research.

Next week, I will change subjects.

Week 4 (04/21/08-04/27/08)

Since proper data are not available for my study on Grace system. I have to change subjects. However, I found the GOCE project, which may provide more detailed data for Earth gravitational field next year. I hope the data from this project can be used in my study.

Dr. Martin gave me some keywords last week. These keywords were matrix, infilling, 3D, and seismic. I was trying to have a general understanding of these words, especially when people tried to put them together.

Week 5 (04/28/08-05/04/08)

This week, I learnt how to use Python and Latex in Linux system. It was difficult for me to write a program without any understanding of the Python. However, Ryan helped me out. We met almost everyday. He helped me to solve the system problems and the problems in understanding the Python program. I tried to use the vector function to generate a simple 2D matrix and use some of the data in this matrix to train a function and finally the function can fill in the missing data very well.

Next week, I will continue to generate more complicate matrices and use some other separable functions to learn the data.

Week 6 (05/5/08-05/11/08)

This week, I tried to randomly delete some data in a matrix and use other data in that matrix as the training data to train the new function. The new function gave 'good' approximations to all the data in the matrix. Of course, this depends on how do we define 'good'. So, next week, I will improve my coding and give the standard deviation and standard error for all the data. And, one more thing I need to consider is how many data do I need to train the function. This week, I just got a general picture. For example, if I randomly delete two data points from a matrix with 20 data points, the error at each point is about e^{-7}, and if I delete six data, the error at each point is e^{-5}. Next week, I will try to give a more clear relation between the number of deleted data and error.

Week 7 (05/12/08-05/18/08)

This week, I made some changes to the 2D infilling program in order to provide a clear error analysis. The fitfunction gave very good approximation to the original matrix when enough traindata are provided. So, I need to answer the question: how many data points do I need to achieve a good approximation of the original matrix.

I tried several kinds of combinations of data. I found if we provide all the data on four edges of the matrix, the fitfunction performed well. Then I tried to remove more data with the constrain that each row and column should have at least 2 data points. Then, for a matrix $A_{m\times n}$, we need $2 \times \max(m, n)$ points. however, the fitfunction failed. So, I tried to add one more data point. I found that the error depended not only on the number of the data points, but also their positions. I hope I can find the explanation for this.

Week 8 (05/19/08-05/25/08)

This week, I basically finished the 2D-infilling program(rank 1). The question I was supposed to answer was solved after I made some changes to the ALS function. First, I changed the conjugate gradient epsilon, so that, the ALS function can give better approximations to both training data and missing data. Second, I changed the times of iteration, according to the shape of the training data.

The main results were: 1. If we provided one complete row and one complete column data of a matrix, the approximation was very good after about 70 iterations. 2. If we provided the data that made sure there were at least two data in each row and column, the approximation was very good after 300 iterations.

Week 9 (05/26/08-06/01/08)

This week, I finished the 2D(rank2) program. There are some problems with rank 2 program in the beginning. I can only find good approximations for both the training data and the testing data occasionally. And, I found if the approximations for the training data were acceptable, the approximations for the testing data were also acceptable, and vise versa. Prof. Martin told me the reason might be that I found the local optimal solution. So, I tried to generate and randomize the fit function before getting an acceptable approximation. Finally, I found we could solve rank 2 problem if we provided three columns and three rows data for training. The approximations for both training data and testing data are very good.

Next week, I will start to consider the 3D problem.