ECE4094 Requirements Specification

Monash University

Large-Scale Matrix Computations for Statistical Signal Processing

Project Code	ECE4094
Document Number	0001
Document Title	ECE4094 Requirements Specification
Version	E
Status	[ ]Draft [ ]Review [ ]Approved
Date Issued	20/12/2017

1 Table of Contents

2 Document Control

2.1 Revision Control

Version	Date	Modified By	Details
A	12/12/2017	ST	Document Outline
B	14/12/2017	ST	Section 2,3
C	16/12/2017	ST	Section 4,5
D	18/12/2017	ST	Incorporated Feedback from JS in sections 3, 4 & 5
E	19/12/2017	ST	Final edits

2.2 Contributors

Name	Position	Company
Simon Teshuva	Student	Monash University
Dr James Sauderson	Technical Supervisor	Monash University
Dr Jonathan Li	Administrative Supervisor	Monash University

3 Introduction

3.1 Objectives

The Objective of this document is to record and store the requirements for the components of the "Large-Scale Matrix Computations for Statistical Signal Processing" final year project that Simon Teshuva is to complete.

3.2 Context

The Log Determinant of a Sparse Large Positive Semi-Definite Matrix is a function used in many applications in statistics, economics and engineering. When modelling the correlation between many variables, the covariance is a useful property. The logdet of a matrix is a useful function, closely related to the covariance. The logdet of a matrix is an analogue to the polynomial log function. It works by approximating the logarithm with a polynomial (usually a Chebyshev Polynomial) over a finite domain. Outside this domain, the accuracy of this approximation is much looser.

As matrices get very large (in the order of 10s or 100s of millions of variables) standard techniques, such as Cholesky Decomposition and Eigenvalue analysis, used to evaluate the logdet of a matrix become very time and space inefficient. As a result, techniques which approximate this function to a high degree of accuracy have been developed.

This project aims to:

- replicate the results of Han, Malioutov and Shin [2] and compare their solution with the standard Cholesky Decomposition.
- Develop a new method, which approximates the logdet as a rational function rather than as a polynomial will be developed, in the hopes of avoiding Han, Malioutov and Shin's reliance on the condition number.
- Test these methods on pre-generated data with controlled condition numbers, and subsequently use them to experiment on real world data.

3.3 Document Scope

The areas that requirements are specified include:

- High Level Requirements
- Software Requirements
- Communication Requirements

3.4 Types of requirements

There are several distinctions made between the different types of requirements:

1. **Requirements**. Standard requirements that are necessary to fulfill in order for the product to meet customer expectations. These are listed in the form "R.xxx"
2. **Optional**. Standard requirements that the customer has indicated are to be fulfilled if possible, with the expectation that they might not be possible or feasible to fulfill. These are listed in the form "OR.xxx".
3. **Caveats**. Requirements that have been put in place by Simon Teshuva in order to flag restrictions imposed on the system in order to meet the requirements. They are listed in the form "C.xxx".

All requirements, regardless of type, have a unique numerical value (xxx) and that value has no significance other than providing a unique ID for the requirement. [1]

3.5 References

Reference	Document
[1]	“Sample Requirements Analysis from Grey Innovation” [Supplied by Monash University]
[2]	“Large-scale Log-determinant Computation through Stochastic Chebyshev Expansions”, I. Han, D. Malioutov, J. Shin
[3]	“Functions of Matrices: Theory and Computation”, N.J. Higham, 2008

3.6 Glossary

Term	Definition
Matrix Logdet	A matrix analogue to the polynomial definition of the log function. It is approximated by a Chebyshev Polynomial over a bounded domain.
Large Matrix	In this context, a matrix with at least 1,000,000 elements
Sparse Matrix	A matrix in which the overwhelming majority elements have a value of 0
Positive Semi-Definite Matrix	An n-by-n A is positive semidefinite if, for any vector z of size n, $z^T Az$ returns either 0 or a positive number.
Cholesky Decomposition	A method of decomposing a Hermitian, Positive-Semi Definite Matrix into lower a lower triangular matrix and its conjugate transpose.
Condition Number	The ratio between the largest and smallest Eigenvalue/Singular value

4 Requirements

4.1 Project Overview

The project has several aims, which have been split in into High Level Requirements, Code Functionality Requirements and Communication Requirements, each of which is described below.

4.2 Technical Requirements

4.2.1 High Level Requirements

At a high level, the project should produce the following results:

Requirement ID	Requirement Description
[R.001]	To reproduce results from "Large-scale Log-determinant Computation through Stochastic Chebyshev Expansions" [2]
[R.002]	Implement the Sparse Cholesky and Eigenvalue analysis methods of computing the logdet of a matrix
[R.003]	To conduct a comparative analysis of the effectiveness of above methods in terms of their time complexity, space complexity and accuracy.
[R.004]	To Implement and evaluate a method based on rational approximation of the log function.

4.2.2 Code Functionality Requirements

Requirement ID	Requirement Description
[R.005]	Create a dataset in the form of a Large, Sparse Positive Semi-Definite Matrix with control over the condition number, without inherent biases.
[R.006]	Implement the method shown in [2] in MATLAB, on a matrix with 1e6 variables for a sparse input.
[R.007]	Implement Sparse Cholesky and Eigenvalue analysis methods for computing the logdet in MATLAB and determine the at what size matrix these methods stop being viable.
[R.008]	Implement method based on rational approximation of the log function in MATLAB on a matrix with 1e6 variables for a sparse input.
[R.009]	Implement the above method in parallel using MATLAB's parallel programming toolbox.
[R.010]	Compare the effectiveness of the above methods in terms of relative error, time complexity and space complexity.
[R.011]	Run applications on real world data to compare all implemented methods.

4.2.3 Communication Requirements

Requirement ID	Requirement Description
[R.012]	Write a report which clearly explains the problem, existing solutions, methodology, the new solution, and the effectiveness of the new solution.
[R.013]	Using the results obtained in [R.011], develop a set of graphs, charts or interactive GUIs to be used in presentations explaining this project.

4.3 Caveats

Requirement ID	Requirement Description
[C.014]	The effectiveness of the new method produced will be limited by the power of the computer used.
[C.015]	The solution developed may not out-perform existing solutions in all contexts.
[C.016]	The new solution will be designed to operate within a certain set of parameters and may not be effective outside of these parameters.

4.4 Optional

Requirement ID	Requirement Description
[O.017]	Improve the effectiveness of the method developed in [R.008] by implementing the algorithm in a lower level language such as Python or C.
[O.018]	Improve the effectiveness of the method developed in [R.008] and parallelised in [R.009] by implementing it using C's parallel computing libraries.

5 Non-Functional Requirements

5.1 Risk Analysis

Requirement ID	Requirement Description
[R.017]	To undertake a Risk Analysis
[R.018]	Proceed or alter the direction of the project as determined by the results of the risk analysis.

5.2 Design Specifications

Requirement ID	Requirement Description
[R.019]	By 12/2/2018, produce a Design Specification Report. The report will include a detailed description of; - How the project is to be completed, including a timeline - What experimental methodology is being used - Explanation on the mathematical and programming theory that motivates the experimental methodology - Samples of code used in the project so far

5.3 Progress Report

Requirement ID	Requirement Description
[R.020]	By 12/2/2018, produce a Progress Report. The report will include a detailed report on progress made in addressing the requirements listed in section 4. In addition, a detailed description of the remaining tasks will be included.

5.4 Final Report

Requirement ID	Requirement Description
[R.021]	By the end of Semester 1 2018, produce a Final Project Report which will include: - An executive summary - Introduction to the project - Overview of the mathematical and programming theory - Description of Methodology - Results from experiments which use developed code - A description of the project management - Key challenges encountered during the project - Appendices with the code developed, and result from experiments

5.5 Project Summary Video

Requirement ID	Requirement Description
[R.022]	By the end of Semester 1 2018, produce a Project Summary Video. The Video will include an executive summary; outline of the project theory, methods and results; and will showcase the potential of the project to future employers.

5.6 Poster

Requirement ID	Requirement Description
[R.023]	By the end of Semester 1 2018, produce a Project Presentation Poster. The poster will explain in a concise and interesting manner, the core aims and ideas that were explored during the project. The poster will be used at the end of Semester 2 2018 during the “Spark Night”
[R.024]	To assist viewers of the poster in getting an in depth understanding of the project, develop a set of interactive charts, graphs and GUIs, as described in [R.013].