Nearest Correlation Matrix

In the extensive new functionality included at Mark 22 of the NAG Fortran Library and the NAG Toolbox for MATLAB and Mark 9 of the NAG C Library is a Nearest Correlation Matrix routine. In this mini-article we take a closer look at Nearest Correlation Matrix and give background to how it came to be included in the NAG Libraries.

Introduction

A correlation matrix is characterised as being a real, square matrix with ones on the diagonal and with non-negative eigenvalues. A matrix with non-negative eigenvalues is called positive semi-definite.

If a matrix C is a correlation matrix then the elements of C, C(I,J) represent the pair-wise correlation of entity I with entity J, that is. the strength and direction of a linear relationship between two.

In the literature there are numerous examples illustrating the use of correlation matrices but the one we have encountered the most arises in finance where the correlation between various stocks is used to construct sensible portfolios. On our website Simon Acomb talks about general use of correlation in finance http://www.nag.co.uk/market/training/manchester_finance_feb09/.

Unfortunately, for a variety of reasons, an input matrix which is supposed to be a correlation matrix may fail to be semi-definite. The correlations may be between stocks measured over a period of time and missing data for example. Treated incorrectly the missing data problem can give rise to an indefinite matrix. Still drawing from finance, a practitioner may wish to explore the effect on a portfolio of assigning correlations between certain assets different from those computed from historical values. This can destroy the semi-definiteness of the matrix too.

In such situations the user has a matrix which approximates a correlation matrix but which fails to be so. Since subsequent analysis relies upon the matrix being a valid correlation matrix for the results to be valid, it is natural to seek a neighbouring matrix which differs least from the input matrix to act in its stead. This problem is discussed in greater detail by Professor Nick Higham http://www.nag.co.uk/market/training/manchester_finance_feb09/.

The NAG Nearest Correlation Matrix Algorithm

Professor Higham published an alternating projection method for the reliable solution to this problem and it was our intention to make this available in the NAG Libraries. (Prior to this much of the literature had been concerned with ad hoc methods that were not guaranteed to solve the problem.) However a subsequent paper by Qi & Sun appeared describing a Newton algorithm that had superior rate of convergence properties. We wanted to include this algorithm instead. Fortunately a research student at Manchester, Rudiger Borsdorf, with Professor Higham as supervisor, looked at this in greater detail and offered further improvements. These included a different iterative solver -MINRES was preferred to Conjugate Gradient - and a means of pre-conditioning the linear equations. It is this enhanced algorithm that has been incorporated into our latest library releases.

Timings

One of the features of this algorithm is that large matrices may be handled. As extensive use is made of BLAS routines we can exploit the optimised BLAS libraries available on many machines. The algorithm thus lends itself well to exploitation of the machine architecture in both our serial and SMP implementations:

On a 10,000 by 10,000 matrix we are able to obtain the following timings using an AMD quad-core system and linking to the NAG serial, NAG SMP Library and AMD Core Math Library (ACML):

Number of Processors	NAG SMP Library Time (secs)	NAG Fortran Library Time (secs)
1	16,119	16,119
4	6,176	7,663