

最邻近相关矩阵

在 NAG Fortran 算法库第 22 版、NAG MATLAB 工具箱与 NAG C 算法库第 9 版中，新建了最邻近相关矩阵函数。在此短文中，我们透过了解最邻近相关矩阵及其背景，可以清楚知道为什么我们需要增加这个函数。

简介

相关矩阵的特性是实数、对角线为 1、非负特征值的正方形矩阵。我们称非负特征值的矩阵为半正定 (positive semi-definite) 矩阵。

若矩阵 C 是一个相关矩阵，则 C 中的每个元素 $C(I,J)$ 代表的是每个成对实体 (I 与 J) 的相关性，表示他们两个之间的线性关系的强度与方向。

在文献中有许多例子说明相关矩阵的应用，然而在金融应用中，最常遇到的就是以用各个股票的相关性来建构合理的投资组合。

很不幸的，基于很多原因，输入的相关性矩阵可能不是半正定，例如：可能经过一段时间各股票间的资料已经遗漏，若不正确的去处理这些遗漏值 (missing data)，会导致此矩阵为非正定矩阵。另一个在金融上应用的例子是：研究人员也许想要透过对特定资产的历史资料，所计算得到的不同相关性中，去研究其对投资组合的影响。这同样也会破坏矩阵的半正定性。

在以上的情况下，使用者会有一个近似的相关矩阵，但它并不符合相关矩阵所定义的要件。由于后续的分析都必须根基于一个有效的相关矩阵，这样才能得到正确的结果。所以很自然的，我们可以去寻找一个与原始矩阵差异性最小的相接近的矩阵，用它来取代原始要用来分析的相关矩阵。

NAG 最邻近相关矩阵

Higham 教授发表了能解决此问题的可信赖方法 – 交替投影的算法，当时我们期望将其纳入 NAG 算法库中。(在此之前，大部分文献关注的都是不能保证能解决这个问题的各种方法)。随后 Qi 与 Sun 的论文中提到牛顿算法具有更高的收敛率。我们希望也能加入这个算法。很幸运的，曼彻斯特大学的研究生 Rudiger Borsdorf 及其指导教授 Higham，仔细的了解其中的细节并提出进一步的改进方法。方法包含了不同迭代求解器 – MINRES 比共轭梯度法要合适 – 以及预处理线性方程。此新改良的算法已经加入到我们最新的算法库版本中。

时间测量

这个算法的特征之一就是大量的矩阵需要进行处理。由于需广泛使用 BLAS 函数，我们可以

利用优化的 BLAS 算法库进行计算。因此，这个算法非常适合加入到我们的串行及并行的算法库版本。

在 10,000 X 10,000 的矩阵计算中，我们采用 AMD 四核的处理器，链接 NAG Fortran 及 NAG 多核算法库，及 AMD ACML 数学库，得到以下数据 (单位：秒)：

处理器个数	NAG 多核算法库	NAG Fortran 算法库
1	16,119	16,119
4	6,176	7,663