最邻近相关矩阵新建功能

在 NAG 算法库第 23 版中，我们持续扩充最邻近相关矩阵函数的相关功能。在本篇短文中，我们先了解最邻近相关矩阵问题，接着提供一些背景介绍，并说明求解的函数。

简介

相关矩阵的特征是实数、对角线为 1 的对称矩阵，且具有非负数的特征值。具有非负数的特征值矩阵称为半正定。如果一个矩阵 *C* 是相关矩阵，则 *C* 的元素 *C(I,J)* 表示 *I* 与 *J* 成对的相关性，也就是两者之间的强度与方向的线性关系。

文献中有许多使用相关矩阵的例子，但我们最常看到的是在金融领域中用来表示两两股票之间的相关性，用来建构合适的投资组合。然而，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。

基本最邻近相关矩阵问题

NAG 的 g02aaX 函数以牛顿算法求解我们在简介中提到的基本问题。它会以矩阵范数 (Frobenius norm) 找出最接近输入矩阵 *G* 的关系矩阵 *X*，也就是找出其最小值：

$$||G - X||_F$$

在 Qi 与 Sun 的论文中提到优于前述方法的算法，具有更佳的收敛性。英国曼彻斯特大学的研究生 Rüdiger Borsdorf，在指导教授 Higham 的带领下，进一步仔细的研究相关细节，并提出改善的方法。其中包含了不同迭代求解器 (使用 Conjugate Gradient 方法的 MINRES) 以及预处理线性方程方法等。这个新的算法已经集成到我们新版的算法库中了。

我们在 22 版的算法库中透过对算法结构的理解，已经强化许多函数的执行性能。

权重范数问题与正定矩阵输出

在新的 NAG g02abX 函数中，我们加强了 g02aaX 函数的功能。我们对估计的相关矩阵有个合理的假设，并非所有矩阵中的值都是估计的，而是仅仅其中一部份。例如：我们也许知道某个我们量测的子集合中的相关性是已知的。

在此算法中，我们利用了原来 Qi 与 Sun 的方法，并加上权重范数 (weighted norm)。因此我们要找出最小值：

$$||W^{1/2}(G-X)W^{1/2}||_F$$

其中 *W* 是权重的对角矩阵。这意味着我们在寻求 *W(I,I)(G(I,J)-X(I,J))W(J,J)* 的最小值。因此通过选择 *W* 合适的元素值，我们可偏好 *G* 中某些的元素，以迫使 *X* 中对应的元素更接近他们。

此函数同时允许用户指定计算得到的相关矩阵是正定的，也就是其特征值必须大于 0。因为在某些应用中需要这样的特性改善矩阵条件而且增加稳定性。

最邻近相关矩阵与因素结构 (Factor Structure)

因素结构的的相关矩阵是非对角线元素有一些 *k* 阶的矩阵。也就是相关矩阵 *C* 可以改写为：

$$\mathrm{C}=\mathrm{diag}(\mathrm{I}-XX^T)+XX^T$$

其中 *X* 是 *n* x *k* 的矩阵，通常被称为因子负荷矩阵 (factor loading matrix)，而 *k* 远比 *n* 小的多。这些相关矩阵通常会在资产报酬因子模型、抵押借贷债务与多变量时间序列中出现。

g02aeX 函数用来计算上述所定义的最邻近相关矩阵 *G* 的最邻近因子负荷矩阵 *X*，求以下最小值：

$$||G-XX^T+diag(XX^T-I)||_F$$

我们采用 Borsdorf 与 Higham 建议的 Birgin, Martinez 与 Raydan 所提的光谱投影梯度 (spectral projected gradient) 法。