Numerical Methods for PCA: Jacobi Transformation

Jacobi Transformation is a *tractable* numerical method of matrix diagonalization (e.g., obtaining a diagonal matrix of eigenvalues). The method is based on eliminating the largest off-diagonal element by rotating the matrix. 'Rotation' is implemented by pre-multiplying matrix $\mathbf{A}$, which we ultimately want to decompose, by matrix $\mathbf{P_{p,q}}$ that is specially constructed in order to cancel an off-diagonal element $a_{p,q}$ so that $a'_{p,q} = 0$.

$$P_{p,q} = \begin{bmatrix} 1 & & & & & & & & 0 \\ & \ddots\, 1 & & & \vdots & & & \iddots & \\ & & \cos\phi & \cdots & 0 & \cdots & \sin\phi & & \\ & \cdots & 0 & \cdots & 1 & \cdots & 0 & \cdots & \\ & & -\sin\phi & \cdots & 0 & \cdots & \cos\phi & & \\ & \iddots & & & \vdots & & & 1\,\ddots & \\ 0 & & & & & & & & 1 \end{bmatrix}$$

For each rotation, we multiply

$$\mathbf{A'} = \mathbf{P_{p,q}}^{\mathbf{T}}\mathbf{A}\mathbf{P_{p,q}}$$

For a covariance matrix, the rotation occurs within the unit circle, and therefore, properties of trigonometric functions can be efficiently used. Key to implementation is calculation of the angle of rotation ϕ.

1. Describe the purpose of applying Jacobi Transformation to a covariance matrix.
   Solution: The method is part of the specific class of *spectral decomposition* that factorizes a matrix into eigenvalues and corresponding eigenvectors. Spectral decomposition is used to identify main uncorrelated (orthogonal) factors that determine the most variance of a system, usually expressed with a co-variance matrix.

2. Deduce why in order to eliminate the matrix element $a'_{p,q} = 0$ it is necessary that $\tan(2\phi) = \dfrac{2a_{p,q}}{a_{q,q} - a_{p,p}}$. **Hint:** consider multiplication of individual matrix elements.
   Solution: Consider the result of rotation matrix multiplication on the individual element with row p and column q

$$a'_{p,q} = \frac{1}{2}(a_{p,p} - a_{q,q})\sin(2\phi) + a_{p,q}\cos(2\phi) = 0$$

$$\frac{1}{2}(a_{q,q} - a_{p,p})\sin(2\phi) = a_{p,q}\cos(2\phi)$$

$$\frac{\sin(2\phi)}{\cos(2\phi)} = \frac{2a_{p,q}}{a_{q,q} - a_{p,p}}$$

$$\tan(2\phi) = \frac{2a_{p,q}}{a_{q,q} - a_{p,p}}$$

   Very close eigenvalues $a_{p,p} = a_{q,q}$ will make $\tan(2\phi) \to \infty$ implying that *stability* of the method improves with $\phi \ll \frac{\pi}{4}$.

3. Jacobi method is not the most computationally efficient because each new rotation destroys zero result obtained on the previous step. Nonetheless, *convergence* of the sum of the off-diagonal elements to zero occurs. Given that Jacobi method chooses $a_{p,q}$ to be greater than other off-diagonal elements on average

$$a_{p,q}^2 \geq \frac{\sum_{i\neq j} a_{i,j}^2}{n^2-n}, \tag{4}$$

show that for a matrix $n \times n$ convergence occurs with the factor of $1 - \dfrac{2}{n^2-n}$.

Solution: Each rotation reduces the sum of squares of the off-diagonal elements by the amount $2a_{p,q}^2$

$$\sum_{i\neq j} a_{i,j}'^2 = \sum_{i\neq j} a_{i,j}^2 - 2a_{p,q}^2. \tag{5}$$

This is possible to demonstrate with a case of symmetric 2×2 matrix $\mathbf{A} = \begin{bmatrix} a_{p,p} & a_{p,q} \\ a_{p,q} & a_{q,q} \end{bmatrix}$.

Then $\mathbf{A}' = \mathbf{P^T A P}$ implies $a_{p,p}'^2 + a_{q,q}'^2 = a_{p,p}^2 + 2a_{p,q}^2 + a_{q,q}^2$, where the sum of squares of diagonal elements increased by $2a_{p,q}^2$ (remember $2a_{p,q}'^2 = 0$ after a rotation).

The rotation deducts the same amount from off-diagonal elements as it adds to diagonal elements, i.e., the rotation does not change L^2 norms of column vectors constituting the matrix. Substituting (4) into (5) gives

$$\sum_{i\neq j} a_{i,j}'^2 \leq \sum_{i\neq j} a_{i,j}^2 - 2\frac{\sum_{i\neq j} a_{i,j}^2}{n^2-n}$$

$$\sum_{i\neq j} a_{i,j}'^2 \leq \left(1 - \frac{2}{n^2-n}\right) \sum_{i\neq j} a_{i,j}^2$$

The closer convergence factor is to 1 the slower is the numerical method because of the small reduction in the sum of squares occurring on each rotation.

4. Explore VBA code that implements Jacobi Transformation in Excel PCA file. Names of variables are self-explanatory and linked to the mathematical model, for example, *Athis(i,j)* for $\mathbf{A}$ and *Awork(i,j)* for $\mathbf{A}'$.

Solution: For the *spectral* decomposition of the covariance matrix

$$\mathbf{\Sigma} = \mathbf{V \Lambda V^T}$$

On convergence, matrix $\mathbf{A}'$ becomes a diagonal matrix with eigenvalues, so $\mathbf{\Lambda} = \mathbf{A}'$. In order to recover eigenvectors, matrices $\mathbf{P_{p,q}}$ from each transformation (rotation) should be multiplied, so $\mathbf{V} = \mathbf{P_0} \times \mathbf{P_1} \times \ldots \times \mathbf{P_m}$.

Note: Jacobi Transformation represents a balance between being tractable and computationally efficient. Power method to calculate eigenvalues by one, starting with the largest, is also simple to present (see Chapter 37.13 in Volume 2 of PWOQF). Other matrix decomposition methods (including non-spectral) can suit the task and work much faster, in particular, see Cholesky decomposition applicable if the matrix is positive definite.