

Introduction to the Math of Multiple Co-Inertia Analysis

Max Mattessich

Joaquin Reyna

Anna Konstorum

9/1/2022

Introduction

Multiple Co-Inertia Analysis (MCIA) is a joint dimensionality reduction method designed to work on data that can be organized into several blocks of variables measured on the same subjects. These blocks could indicate variables of different categories (e.g. genes vs. proteins) or different experiments conducted on the same subjects. MCIA creates a set of ‘factors’ or ‘scores’ vectors along with accompanying ‘loadings’ vectors that aim to highlight trends common across all datasets.

Table of Notation

Notation	Description
$\mathbf{X}_b \in \mathbb{R}^{n \times p_b}$	The matrix representing data block b , in samples-by-features format.
$\mathbf{X} \in \mathbb{R}^{n \times p}$	The global (concatenated) matrix of data blocks.
p_b	The number of variables/features in block b .
p	The total number of variables/features, $p = \sum p_b$.
n	The number of samples/rows in the data (constant across all blocks).
N	The number of blocks in the data.
$\mathbf{f}^{(j)} \in \mathbb{R}^{n \times 1}$	The j -th order global score/factor vector.
$\mathbf{a}^{(j)} \in \mathbb{R}^{p \times 1}$	The j -th order global loadings vector.
$\mathbf{f}_b^{(j)} \in \mathbb{R}^{n \times 1}$	The j -th order block score/factor vector for block b .
$\mathbf{a}_b^{(j)} \in \mathbb{R}^{p_b \times 1}$	The j -th order block loadings vector for block b .
r	The target dimension of the low-dimensional representation of the data.

Mathematical Theory

MCIA works on datasets that are made up of multiple blocks of variables measured on the same set of samples. Since each data block has the same number of rows, we define the global data matrix as

$$\mathbf{X} = [\mathbf{X}_1 | \mathbf{X}_2 | \dots | \mathbf{X}_N] \in \mathbb{R}^{n \times p}, \quad p = \sum_b p_b.$$

It is sometimes helpful to compare MCIA to the better-known Principal Component Analysis (PCA). The global scores/factors and loadings are broadly similar to the principal component scores and loadings respectively of matrix \mathbf{X} , although they are **not** identical.

The Objective Function

The ultimate goal of MCIA is to provide a set of scores and loadings vectors of decreasing optimality up to a user-chosen target dimension r . These vectors then can be used as a low-dimensional representation of the

data or to find associations between variables or blocks. MCIA defines optimality as maximizing the sum of squared covariance between a set of block scores \mathbf{f}_b and a global score \mathbf{f} , i.e.

$$\operatorname{argmax}_{\mathbf{f}_1, \dots, \mathbf{f}_N, \mathbf{f}} \sum_{b=1}^N \operatorname{cov}^2(\mathbf{f}_b, \mathbf{f}), \quad (1)$$

where $\mathbf{f}_b = \mathbf{X}_b \mathbf{a}_b$ for $b = 1, \dots, N$ and $\mathbf{f} = \mathbf{X} \mathbf{a}$ are linear combinations of the variables in the block and global matrices respectively.