4
Mathematical and Computational Techniques

For investigating systems-level correlated of phenotypic side effects, we employed mathematical and computational approaches. Primarily, we used canonical correlation analysis (CCA) as a core method. Starting with the integrated data of drugs, targets and side effects (Chapter 3), we developed a generalized CCA method for pinning down drug features central for specifying side effects (Chapter 5). Further, we applied the CCA on the genomic space to find topological features that have bearing on adverse drug reactions (Chapter 6). We also used CCA for identification of subset of side effects, whose prior knowledge can be used for predicting remaining side effects, for seven different organs/systems (Chapter 7). Finally, we developed a partial CCA method that can be used to identify the most independent drug features, and demonstrated it for prediction of eye- and node-related side effects (Chapter 8). In this Chapter, we present technical details of canonical correlation analysis, partial canonical correlation, and their generalized versions.

4.1 CANONICAL CORRELATION ANALYSIS

CCA is a method employed for multivariate analysis aimed at finding information from cross-covariance matrices. This procedure involves projection of two datasets, between which correlation is sought for, into a common subspace. For example, two datasets A and B comprising of n entities each, such that each entity is represented by p and q features, can be written as,

$$A \rightarrow n \times p$$
$$B \rightarrow n \times q$$

One of the basic measures for finding similarity between two vectors, is that of cosine similarity. But it requires that these vectors/datasets to be same dimension. For vectors $X = A\alpha$ and $Y = B\beta$ (with projection parameters α, β) correlation measured with cosine similarity is as follows [M.E.J Newman, 2010]:

$$Corr(X, Y) = \frac{\alpha^T A^T B\beta}{\|A\alpha\|_2 \|B\beta\|_2}$$

It is clear that for maximizing the correlation, towards finding projection parameters, it is sufficient to maximize the numerator. Therefore, the objective function could be expressed as,

$$f = \alpha^T A^T B\beta.$$

The optimization could be achieved by using unconstrained as well as constrained methods. The prior method allows open exploration for parameters, whereas the latter curbs the search space.

4.1.1 Unconstrained optimization

Unconstrained optimization involves searching for parameters by minimizing the distance between two datasets A and B. This minimization of correlation could be expressed in terms of an objective function,

$$f = \min_{\alpha,\beta} (A\alpha - B\beta)^T (A\alpha - B\beta)$$

The parameters for which minimization is achieved are obtained for partial differentiation as follows:

$$\frac{df}{d\alpha} = 0, \; P_{AA}\alpha - 2P_{AB}\beta = 0 \tag{4.1}$$

$$\frac{df}{d\beta} = 0,\ -2P_{BA}\alpha + P_{BB}\beta = 0 \tag{4.2}$$

Here, $P_{AA} = A^T A,\ P_{BA} = B^T A,$ and $P_{BB} = B^T B.$
The value of β as obtained from Eq. (4.2) is,

$$\beta = 2P_{BA}^{-1}P_{BB}\alpha. \tag{4.3}$$

Substituting the value of β from Eq. (4.3) into Eq. (4.2) produces,

$$(P_{BB}P_{BA}^{-1}P_{BB} - P_{BA})\alpha = 0 \tag{4.4}$$

Accordingly, $\alpha \in Null\ (P_{BB}P_{BA}^{-1}P_{BB} - P_{BA})$, subsequently β could be computed using Eq. (4.3).

4.1.2 Constrained optimization

In contrast to unconstrained optimization, constrained optimization involves finding parameters within a restricted domain. In order to solve with such constrains, Lagrange multiplier method is used [Bertsekas, 1982]. It transforms objective function into unconstrained equation. This can be written as,

$$f = \alpha^T A^T B\beta + \lambda(1-\alpha^T\alpha) + \mu(1-\beta^T\beta)$$

Here the constraints are $\alpha^T\alpha = \beta^T\beta = 1.$ λ and μ are the Lagrange multipliers for constraint equations.

Differentiating f w.r.t α, β for maximizing correlation and substituting $A^T B = Z,$

$$\frac{\partial f}{\partial \alpha} = 0,\ Z\beta - \lambda\alpha = 0$$

$$\alpha = \frac{1}{\lambda}Z\beta \tag{4.5}$$

$$\frac{\partial f}{\partial \beta} = 0,\ Z^T\alpha - \mu\beta = 0$$

$$\beta = \frac{1}{\mu}Z^T\alpha \tag{4.6}$$

Using these values of α and β into Eqs. (4.5) and (4.4), we obtain

$$ZZ^T\alpha = \lambda\mu\alpha \tag{4.7}$$

$$Z^T Z\beta = \lambda\mu\beta \tag{4.8}$$

In general, Eq. (4.7) and Eq. (4.8) could be written as,

$$ZZ^T\alpha_i = \gamma_i\alpha_i \text{ and } Z^T Z\beta_i = \gamma_i\beta_i$$

$\alpha_{i=1,2,3\cdots p}$ and $\beta_{i=1,2,3\cdots q}$ are the eigenvectors of ZZ^T and $Z^T Z$ matrix, respectively. These vectors are also known as left and right singular vectors of Z. This canonical representations is found by single value decomposition (SVD) procedure, which is conventionally denoted as $Z = U\Sigma V^T$ [D.Meyer, 2000]. Here, U and V matrices consist of the eigenvectors of ZZ^T and $Z^T Z$ matrix, respectively, corresponding to eigenvalues (γ_i). The diagonal of Σ holds singular values ($\sigma_i = |\gamma_i|$). Importantly, Z matrix could have at most p and q number of left and right singular vectors. α_i and β_i vectors belongs to R^p and R^q vector space, respectively.

Finding optimum parameters

We assume α_i and β_i are the optimum parameters which maximize the objective function. With these parameters the objective function produces highest correlation value, which could be expressed as follows.

$$f = \alpha_i^T A^T B \beta_i = \alpha_i^T [\alpha_1 \alpha_2 ... \alpha_i ... \alpha_p] \begin{bmatrix} \sigma_1 & 0 & . & . & . & 0 \\ 0 & \sigma_2 & . & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & \sigma_l & 0 \end{bmatrix} [\beta_1 \beta_2 ... \beta_i ... \beta_q] \beta_i^T$$

Using the orthogonal nature of U and V and the following constraints, the objective function can be restated.

$$\alpha_i^T \alpha_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \quad \text{and} \quad \beta_i^T \beta_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Accordingly, the objective function takes the following form:

$$f = \sigma_i [00...1...0][00...1...0]^T = \sigma_i$$

In the above expressions, l denotes the number of nonzero singular values and σ_i is the respective singular value or correlation value. As a convention, all singular values are ordered in ascending order ($\sigma_1 > \sigma_2 > \cdots > \sigma_l$) according to SVD procedure. This implies that, for achieving highest correlation value optimum parameters (α_i and β_i) are required to be the first left and first right singular vectors, respectively.

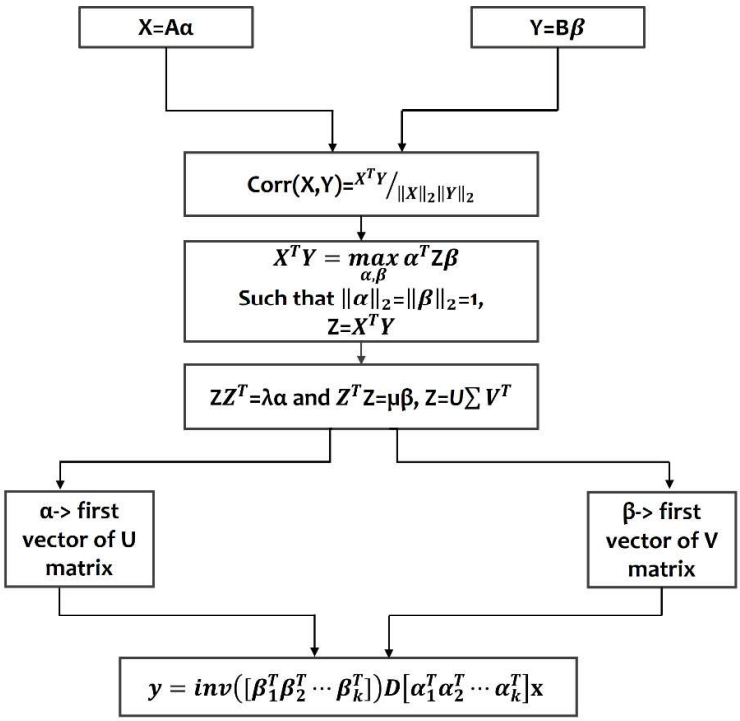

Figure 4.1: The flowchart depicting implementation of canonical correlation analysis.

Projection into common subspace

After finding parameters ($\alpha_{i=1,2,\cdots k}$ and $\beta_{i=1,2\cdots k}$) and correlation values ($\sigma_{i=1,2\cdots k}$), datasets are projected into a common subspace. The projection is represented as,

$$[\beta_1 \beta_2 ... \beta_k]^T Y = D\ [\alpha_1 \alpha_2 ... \alpha_k]^T X$$

Hence, the predicted profile is obtained as,

$$Y = ([\beta_1\beta_2...\beta_k]^T)^{-1} D\ [\alpha_1\alpha_2...\alpha_k]^T X. \quad (4.9)$$

Here, X is the known feature, Y is the predicted feature, $D = \begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_k \end{bmatrix}$, and k is a positive integer, $1 \le k \le n$. For $k = n$, orthogonal property of square matrix implies $inv([\beta_1\beta_2 \ldots\ldots \beta_n]^T) = [\beta_1\beta_2 \ldots\ldots \beta_n]$ We may normalize the correlation matrix (D) by defining, $D_{pp} = \frac{\sigma_p}{\sum_{j=1}^{k} \sigma_j}$ such that correlation value is bounded in the range of [-1,1].

The procedure described above is depicted below as a flowchart (Figure 4.1).

4.2 GENERALIZED CANONICAL CORRELATION ANALYSIS (GCCA)

CCA is of importance in situations when an independent variable needs to be correlated with a dependent variable. Going beyond this constraint, GCCA provides a framework for integrating multiple independent variables in a linear combination, so as to optimize the correlation with the dependent variable. Analogous to CCA, we derived implementation of unconstrained and constrained optimization as described below.

4.2.1 Unconstrained optimization

This procedure obtains the parameters by minimizing the distance between integrated dependent variables and the independent variable. Formally, the objective function minimization can be represented as,

$$f = \min_{\alpha_i,\beta} (\sum_{i=1}^{n} A_i\alpha_i - B\beta)^T (\sum_{i=1}^{n} A_i\alpha_i - B\beta)$$

Partial differentiation of the function with these parameters yields,

$$\frac{\partial f}{\partial \alpha_i} = 0,\ P_{ii}\alpha_i + 2\sum_{i=1, i\neq j}^{n} P_{ij}\alpha_j - 2P_{i\beta}\beta = 0 \quad (4.9)$$

$$\frac{\partial f}{\partial \beta} = 0,\ -2\sum_{i=1}^{n} P_{Bi}\alpha_i + 2P_{BB}\beta = 0 \quad (4.10)$$

Here, $P_{ii} = A_i^T A_i$, $P_{ij} = A_i^T A_j$, $P_{iB} = A_i^T B$, $P_{Bi} = B^T A_i$, $P_{BB} = B^T B$. Eq. (4.9) represents n equations, corresponding to $i = 1, \ldots, n$. These n equations, along with Eq. (4.10), could be presented in the form of a matrix of order $n + 1$ for finding the unknown parameters, as expressed in Eq. (4.11).

$$\begin{bmatrix} P_{11} & .. & P_{1j} & .. & P_{1n} & P_{1B} \\ P_{21} & P_{22} & .. & .. & P_{2n} & P_{2B} \\ .. & .. & .. & .. & .. & .. \\ P_{i1} & .. & P_{ii} & .. & P_{in} & P_{iB} \\ .. & .. & .. & .. & .. & .. \\ P_{B1} & .. & .. & .. & P_{Bn} & P_{BB} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ . \\ \alpha_i \\ . \\ \beta \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ . \\ . \\ 0 \\ 0 \end{bmatrix} \quad (4.11)$$

Prediction model

For an entity represented with n features $(x_{i=1,2,\cdots n})$, following formula was used for predicting its unknown profile (y).

$$y = (\sum_{i=1}^{n} x_i^T U_i) V^{T^{-1}}$$

Here, $U_i = [\alpha_i^1 \alpha_i^2 \cdots \alpha_i^l]$ and $V = [\beta^1 \beta^2 \cdots \beta^l]$; l represents the dimension of the null space. The parameters α_i and β_i in Eq. (4.11) could be obtained from the null space represented by the matrix depicted in this equation.

4.2.2 Constrained optimization

Analogous to the procedure implemented for constraint optimization described in Section 4.2.2, the objective function could be written in the following form:

$$f = \sum_{i=1}^{n} \max_{\alpha_i, \beta} \alpha_i^T Z_i \beta \tag{4.12}$$

Such that, $\|\alpha_i\|_2 = \|\beta\| = 1$ and $Z_i = A_i^T B$.

Objective function is rewritten according to Lagrange equation as,

$$f = \sum_{i=1}^{n} \max_{\alpha_i, \beta} \alpha_i^T Z_i \beta + \lambda_i (\sum_{i=1}^{n} (1 - \alpha_i^T \alpha_i) + \mu(1 - \beta^T \beta).$$

Here, λ_i and μ are Lagrange multiplier for constraints.

Differentiating f w.r.t $\alpha_{i=1,\ldots,n}$ and β^T yields,

$$\frac{\partial f}{\partial \alpha_i} = 0, \; Z_i \beta - \lambda_i \alpha_i = 0, \text{ and}$$

$$\alpha_i = \frac{1}{\lambda_i} Z_i \beta. \tag{4.13}$$

$$\frac{\partial f}{\partial \beta^T} = 0, \; \sum_{i=1}^{n} \alpha_i^T Z_i - \mu \beta^T = 0, \text{ and}$$

$$\beta = \frac{1}{\mu} \sum_{i=1}^{n} Z_i^T \alpha_i. \tag{4.14}$$

Substituting the value of α_i from Eq. (4.12) into Eq. (4.13) produces,

$$\sum_{i=1}^{n} (\frac{1}{\lambda_i} Z_i^T Z_i) \beta = \mu \beta. \tag{4.15}$$

By multiplying Eq. (4.13) with β^T and substituting the value of α_i from Eq. (4.13) into Eq. (4.14) produces

$$\mu = \sum_{i=1}^{n} \lambda_i. \tag{4.16}$$

Eq. (4.16) indicates that it is not possible to compute unique λ_i. Therefore, we follow hyper-sphere transformation to define unique λ_i. We demonstrate this with an example.

For $i = 2$ and $\mu = \mu(cos^2\theta + sin^2\theta)$, Eq. (4.16) yields $\lambda_1 = \mu cos^2\theta$ and $\lambda_2 = \mu sin^2\theta$.

For $i = 3$ and $\mu = \mu(cos^2\theta(cos^2\phi + sin^2\phi) + sin^2\theta)$, Eq. (4.16) yields $\lambda_1 = \mu cos^2\theta cos^2\phi$, and $\lambda_2 = \mu cos^2\theta sin^2\phi$ and $\lambda_3 = \mu sin^2\theta$.

Therefore, following the method of induction for $i \geq 3$, unique λ_i could be expressed as:

$$\lambda_{i \leq n-1} = \mu g_{i-1} \cos^2\theta \text{ and} \tag{4.17}$$

$$\lambda_{i=n} = \mu \sin^2\theta_1. \tag{4.18}$$

Here, $g_i = Cos^2\theta_1$ and $g_{i+1} = g_i Cos^2\theta_{i+1}$.

Eq. (4.17) and Eq. (4.18) suggest that the bound for hyper-sphere coordinates would be $[0, \frac{\pi}{2}]$. After finding λ_i, we use Eq. (4.13) and Eq. (4.14) to obtain α_i and β, respectively. This procedure will be referred to as 'hyper-sphere GCCA' in the rest of the text.

Complexity of hyper-sphere GCCA

Finding optimum parameters of objective function depends on finding hyper coordinates of each of the Lagrange multipliers. Out of n coordinates, $n-1$ hyper coordinates were fixed to a particular value, and rest of the hyper coordinates were varied in the range of $[0,\frac{\pi}{2}]$. Cardinality of this solution space is given by multiplication of number of intervals and number of hyper coordinates. Thus the complexity of computational procedure implemented for GCCA is $(0(2^n))$.

Choice of unknown parameters

Substituting β from Eq. (4.14) into Eq. (4.12), yields the objective function $f = \mu\beta^T\beta = \mu$ since $\beta^T\beta = \alpha_i^T\alpha_i = 1$. Therefore, μ should be the chosen as the highest eigenvalue to maximize the objective function. Thus one needs to choose the highest eigenvector as optimum parameter (β), corresponding to the dependent variable. Starting from β, one could find α_i from Eq. (4.13).

Prediction model

Analogous to Section 4.4.2, the procedure for the prediction model can be arrived at, with the following formal definition:

$$Y = ([\beta_1\beta_2...\beta_k]^T)^{-1}[\sum_{i=1}^{n} D^i A_i^T X_i]$$

Here, $D_{pp}^i = \frac{\lambda_p}{\sum_{j=1}^{k}|\lambda_j|}$ and $A_i = [\alpha_i^1\alpha_i^2 \dots \alpha_i^l]$.

4.3 GENERALIZED PARTIAL CANONICAL CORRELATION ANALYSIS (GPCCA)

In CCA and GCCA, we need to assume that dependent feature solely correlates with only one and n number of independent feature(s), respectively. In reality, any complex entity such as side effects profile would depend on more than one features, and those features could be interdependent. Hence, we developed a partial canonical correlation analysis model which measures the contribution of features, while subtracting dependencies. GPCCA is a generalized version of PCCA and deals with multiple features. Towards derivation of GPCCA, we start from the definition of partial correlation:

$$\rho_{xy,w} = \frac{\rho_{xy} - \rho_{xw}\rho_{yw}}{\sqrt{(1-\rho_{xw}^2)}\sqrt{(1-\rho_{yw}^2)}} \tag{4.18}$$

Thus, $\rho_{xy,w}$ defines partial correlation between x and y such that the effect of third controlling variable w is removed. So as to obtain partial correlation of X_1 matrix with Y matrix by eliminating the effect of third controlling variable X_2, these matrices are required to be of the same dimension. This could be achieved through transformation of these matrices: $U_1 = X_1\alpha_1$, $U_2 = X_2\alpha_2$, and $V = Y\beta$. These transformations implicitly entail following assumptions:

$$\rho_{xy} = \frac{U_1^T V}{\|U_1\|_2\|V\|_2}$$

$$\rho_{xw} = \frac{U_1^T U_2}{\|U_1\|_2\|U_2\|_2}$$

$$\rho_{yw} = \frac{U_2^T V}{\|U_2\|_2\|V\|_2}$$

In general, where independent variable would be dictated by a large number of drug features, the objective function would take a generalized form as derived from numerator of Eq. (4.18),

$$f = \alpha_1^T Z_1\beta - \sum_{i=2}^{n}\alpha_1^T Z_i\alpha_i\alpha_i^T Z_{i+1}\beta,$$

such that, $\alpha_1^T\alpha_1 = \alpha_i^T\alpha_i = \beta^T\beta = 1$. Here $Z_1 = X_1^T Y$, $Z_i = X_1^T X_i$ and $Z_{i+1} = X_i^T Y$.

To find unknown parameters α_1, α_i and, β, objective function (f) is written according to Lagrange form as,

$$f = \alpha_1^T Z_1 \beta - \sum_{i=2}^{n} \alpha_1^T Z_1 \alpha_i \alpha_i^T Z_{i+1} \beta + \sum_{i=2}^{n} \lambda_i (1 - \alpha_i^T \alpha_i) + \mu(1 - \beta^T \beta)$$

Here, λ_1, λ_i and μ are Lagrange multipliers for constraints equation associated with unknown parameters. Maximizing f w.r.t α_1, α_i and, β^T we obtain:

$$[Z_1 - \sum_{i=2}^{n} Z_i \alpha_i \alpha_i^T Z_{i+1}]\beta = \lambda_1 \alpha_1 \tag{4.19}$$

$$[Z_1 - \sum_{i=2}^{n} Z_i \alpha_i \alpha_i^T Z_{i+1}]^T \alpha_1 = \lambda_1 \beta \tag{4.20}$$

$$[Z_i^T \alpha_1 \beta^T Z_{i+1}^T - Z_{i+1} \beta \alpha_1^T Z_i]\alpha_i = \lambda_i \alpha_i \tag{4.21}$$

Next, substituting α_1 and β from Eq. (4.19) and Eq. (4.20) into Eq. (4.20) and Eq. (4.19), respectively, generates eigenvector equations as given below.

$$[Z_1 - \sum_{i=2}^{n} Z_i \alpha_i \alpha_i^T Z_{i+1}]^T [Z_1 - \sum_{i=2}^{n} Z_i \alpha_i \alpha_i^T Z_{i+1}]\beta = \lambda_1 \mu \beta \tag{4.22}$$

$$[Z_1 - \sum_{i=2}^{n} Z_i \alpha_i \alpha_i^T Z_{i+1}][Z_1 - \sum_{i=2}^{n} Z_i \alpha_i \alpha_i^T Z_{i+1}]^T \alpha = \lambda_1 \mu \alpha \tag{4.23}$$

Eqs. (4.21), (4.22) and (4.23) could not yield analytical solution for parameters as computation of α_i depends on the knowledge of α_1 and β , and since these parameters are in turn dependent on α_i . Therefore, an iterative approach was applied for solving these parameters. Figure 4.2 illustrates the logic of this iterative approach described below.

Pseudocode

Step 1: Initialize α_1 and β as all ones (1) vectors, and compute α_i using Eq. (4.21).
Step 2: Estimate α_1 and β by substituting α_i into (4.22) and (4.23).
Step 3: Update old values of α_1 and β with new values.
Step 4: Check correlation between new and old α_1 and β, with predefined threshold (0.95).
Step 5: If correlation value exceeds the threshold, then stop this process with the estimated values of α_1, β and α_i. Otherwise, go to Step 2.

Prediction model

α_i and β represent the most informative parameters of known features and unknown features respectively. As described in section 4.2.2, these parameters were used to generate the common subspace for predicting unknown feature (y) given the known features (x_j), using following equations.

$$y = (\beta^T)^{-1}[\alpha_1^T x_1 - \sum_{i=1, i\neq j}^{n} \frac{\alpha_j^T x_j x_i^T \alpha_i}{\sqrt{1 - (\alpha_j^T x_j x_i^T \alpha_i)^2}} \alpha_i^T x_i] \tag{4.24}$$

Eq. (4.24) can be generalized to include other eigenvectors from Eq. (4.20), Eq. (4.21) and Eq. (4.22).

$$y = (\beta^T)^{-1}[\alpha_1^T x_1 - \sum_{i=1, i\neq j}^{n} \sum_{k=1}^{m} \frac{\alpha_{jk}^T x_j x_i^T \alpha_{ik}}{\sqrt{1 - (\alpha_{jk}^T x_j x_i^T \alpha_{ik})^2}} \alpha_{ik}^T x_i] \tag{4.25}$$

Flowchart

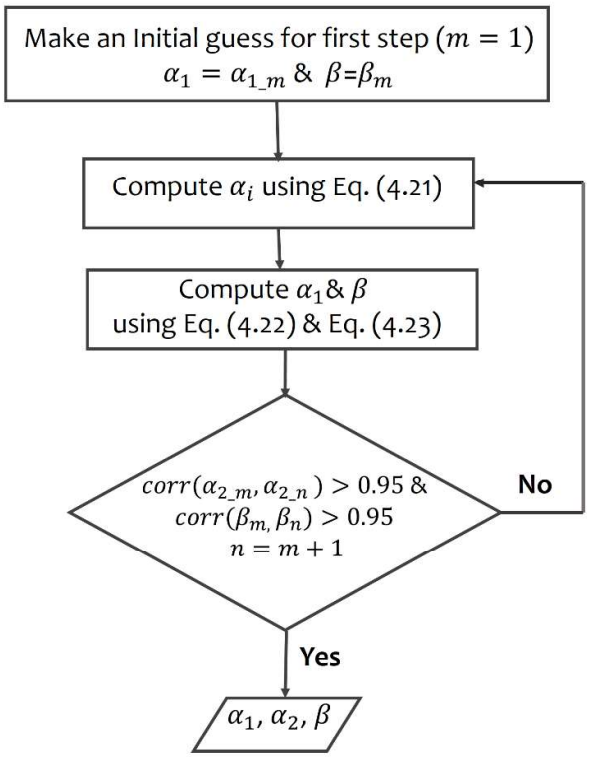

Figure 4.2 : Flowchart depicting the procedure implemented for the Generalized Partial Canonical Correlation Analysis.

Verification of model

We intend to maximize the objective function while also maximizing interrelationships among all feature vectors. Hence, we chose the highest eigenvectors of Eqs. (4.20), (4.21) and (4.22). α_1 and β would produce highest singular value of $Z_1 - \sum_{i=2}^{n} Z_i \alpha_i \alpha_i^T Z_{i+1}$matrix.

Since no other eigenvectors can exceed the highest singular value, the same parameters α_1 and β need to be used so as to maximize Z_1. α_i were computed by differentiating the objective function to maximize the interrelationship among feature vectors. For more details on computation of eigenvectors and highest singular value, please refer to discussions on singular value decomposition in [D.Meyer, 2000].

Formulae used for derivation of CCA and PCCA are provided (with numerator layout) in Annexure A.

…