**Solutions – Assignment 4**

1 a) of the given network

(i) Condition 1: For the network to be NCA compliant, the number of zeros in each column has to be . Here m =3 and each column has at least 2 zeros

(ii) Condition 2: Deleting each column and corresponding rows containing non-zero elements, we check whether resulting sub-matrix has full structural column rank

All sub-matrices have full column rank. Therefore network is NCA compliant

b) Assuming the number of species is known to be 3, the rotation matrix can be calculated by solving system of linear equations

Thus the calculated rotation matrix is given by

The correlation matrix is given in Figure 3



Figure 3: Correlation matrix for part (b)

C) Solution using the NCA toolbox



Figure 4: Correlation matrix from NCA algorithm



**2. Apply NCA directly to the given microarraydata.** You can either use the toolbox or the code from question 4. If Y=A\*P and the structure of A is given. Once you use NCA, use the matrix P to calculate the standard deviation of the 33 tfas ( not the standard deviation of matrix A, which was a mistake most of the students did). From the top 11 tfas with maximum variance we normally find around 8 out of the 11 tfas that regulate cell cycle

**3. Non-negative matrix factorization** (NMF) to extract pure component spectra

The absorbance data considered from **‘Inorfull.mat’.** In Part (a), the absorbance data of the first replicate of the mixtures collectively forms the data matrix. The number of species is known a priori to be 3. The NMF code provided by *Prof. Haesun Park,* *GATech* is used for solving the problem. The input to the *NMF* function **(‘nmf.m’)** are the data matrix, number of species and the initial guesses are the absolute values of the loadings and the scores matrix calculated from *svd(Z)*.

In part (b), the average of the 5 replicates of every mixture is considered for the data set *Z*. The correlation matrices computed between the actual pure component spectra and the evaluated one is presented below.



Figure 1: Correlation matrix for part (a)



Figure 2: Correlation matrix for part (b)

Applying NMF on individual samples is not able to separate all pure species spectra (third species not separated properly). However, NMF on averaged data is able to separate the sources reasonably well. The average of the replicates have less noise and therefore NMF is able to separate sources better.

**4. FASTNCA code using helper functions rearrange and reconstitute**

function [A, P] = fastNCA(Z, Astruct, p);

[nsamples, nvar] = size(Z);

[u s v] = svd(Z,'econ');

W = u(:,1:p);

Amix = zeros(nsamples,p); // Initialize

for k = 1:p

[Zc Zr] = rearrange(W,Astruct,k); // Step 1

[u s v] = svd(Zr); // Step 2

S = v(:,p);

Zcp= Zc\*S; // Step 3

[u s v] = svd(Zcp,'econ'); // Step 4

ak = u(:,1);

Amix(1:size(Zc,1),k) = ak; // Store elements of A for column k

end

A = reconstitute(Amix,Astruct); // Rearrange to match A with Astruct

P = inv(A'\*A)\*A'\*Z; // Compute P using OLS