Description of the implemented PCA method

Foundations and analytical derivation of the methods ????

Block diagram of the code (general)

1. Get\_mu – center the data
2. Svd
3. Enforce a sign convention

General description:

In general PCA methods include some optional steps and exist different versions based on singular value decomposition (SVD), eigenvalue decomposition or alternating least squares

The first step (but might be optional) in the PCA method is to center the data. Then the main part is coming which is

Enforce a sign convention on the coefficients – the largest element in each column will have a positive sign.

Multicore threads analysis – CUDA IMPLEMENTATION

“to conform CULA alignment requirements”

The implemented algorithm was optimized for specific dimensions of matrices (m >> n) of tested fMRI data. It might not be efficient on data with another ratio of dimensions.

The method starts with centering the data. It means calculate the average of each column, and subtract this average from each element of the column.

The simplest scheme to do it seems to do a sum reduction

Each column is processed by one block, (sum reduction) so that shared memory

A version using shuffle instruction (to compute sum reduction in the columns) was tested but did not result in better performance (there was no speed-up).

W sumie to można napisać o tych transpozycjach, że je robimy.

The most important (computational complexity and cost) part of the algorithm is singular value decomposition. A quick research of already implemented SVD method has shown that there are not many libraries for CUDA offering it. In fact CUDA API includes cuSOLVER library with SVD methods but they do not support “economic” version of the algorithm, so they are impractical for the large datasets (large matrices). In this work an implementation of SVD from CULA library was used. This is the library of linear algebra methods basing on LAPACK library implementation.

However CULA library has not been developed since 2013, so it does not take advantage of the features new CUDA release offers.

Przy podawaniu wyników czasowych, należy podać oczywiście dane techniczne procesorów na których były przeprowadzane testy.