This file provides a list of project topic examples. Both reference lists and descriptions of projects are not complete. Interested students are expected to clarify details with teachers (and assistants). We want to draw your attention to the fact that a student can suggest a project theme if her/his area of scientific interest nontrivially intersects with linear algebra.

1. **Singular decomposition for face recognition**

Write a program that performs a person recognition from a photograph using a singular value decomposition (SVD) (see ref.[1]). Write a program that constructs SVD for rectangular matrices of an arbitrary size and test this program on small matrices. Apply the program to the face recognition problem (see ref. [2]). One can convert a photo to a matrix using any standard application or package. What will change if one modifies individual entries/factors of a singular decomposition?

References:

1. Shevtsov G. S. "Linear algebra: Theory and Applied aspects": textbook, 2010 — 525 sec.

2. J.Demmel "Applied Numerical Linear Algebra"

3. http://link.springer.com/chapter/10.1007%2F978-1-4020-6264-3_26

4. TianY, TanT, WangY, FangY: Do singular values contain adequate information for face recognition? Pattern Recogn. 2003, 36: 649–655.

2. Linear Algebra and the Search Problem

Write a program that implements an official open version of the algorithm of the most famous search website. You can choose the degree of implementation complexity yourself. We hope that the dataset will be really non-trivial.

References:

1. *K Bryan, T Leise The $25,000,000,000 eigenvector: The linear algebra behind Google - Siam Review, 2006 - SIAM*

3. Quadratic optimization

The problem of quadratic optimization (or quadratic programming) is a generalization of the least squares method. It requires finding a solution to a system of linear equations that minimizes some quadratic function (the generalized version of the "squared error") under constraints defined in the form of linear inequalities. Several effective methods for solving this problem are known. We suggest to study these methods, implement them (or use already existing packages) and compare their effectiveness for the case of economic statistics - the reconstruction of tables of intersectoral transactions in the scale of national economy.

References:

1. For quadratic optimization algorithms, see Pshenichny BN, Danilin Yu.M. "Numerical methods in extreme problems", chap. III, paragraph 1, or

2. Hadley J; "Nonlinear and dynamic programming", chap. 7

3. For the more contemporary examples of general optimization methods, see: Y.E. Nesterov. "Convex optimization methods" (in Russian)

4. An example of economic challenges can be found here: http://www.wiod.org/publications/papers/wiod2.pdf, paragraphs 2.7 and 2.8

4. Multidimensional arrays, TT decomposition and big data analysis

Illustrate the use of the "tensor-train decompositions" or their analogues in problems of big data analysis.

References:

1. http://epubs.siam.org/doi/abs/10.1137/090752286

2. I. V. Oseledets and E. E. Tyrtyshnikov. “TT-Cross approximation for multidimensional arrays” INM RAS Preprint, 2009-05.

3. http://people.csail.mit.edu/moitra/docs/bookex.pdf

5. Solution of systems of algebraic equations and inequalities

The modern symbolic methods for solving systems of algebraic equations and inequalities are associated with the Gröbner bases and the Cylindrical Algebraic Decomposition concepts. We assume that the basic algorithms will be analyzed and illustrated. It is possible to analyze only one of the two theories in the project. See chapters 11 and 12 in [1], as well as [2] and [3] for information.

References:

1. Basu, Saugata; Pollack, Richard; Roy, Marie-Françoise “Algorithms in real algebraic geometry.” Second edition. Algorithms and Computation in Mathematics, 10. Springer-Verlag, Berlin, 2006. x+662 pp. ISBN 978-3-540-33098-1; 3-540-33098-4 ; author's edition — https://perso.univ-rennes1.fr/marie-francoise.roy/bpr-ed2-posted3.pdf

2. I.V. Arzhantsev. “Gröbner bases and systems of algebraic equations” — https://www.mccme.ru/free-books/dubna/arjantsev.pdf

3. Mats Jirstrand. “Cylindrical Algebraic Decomposition – an Introduction” — http://www.diva-portal.org/smash/get/diva2:315832/FULLTEXT02

6. Methods of the Krylov subspace for solving linear systems

It is proposed to numerically solve the equations of mathematical physics (for example, the Poisson equation) using one of the Krylov subspace methods.

References:

1. J.Demmel “Applied Numerical Linear Algebra”

7. Epsilon spectrum and the construction of spectral portraits

The calculation of the eigenvalues of a real matrix on a computer is always associated with errors. How can we visualize the actual results of computer calculations at a given accuracy?

References:

1. Godunov S.K. "Lectures on modern aspects of linear algebra" - Nauchnaya Kniga, Novosibirsk, 2002

8. Tropical linear algebra and scheduling

Illustrate the use of operations on a tropical semiring.

References:

1. P. Butkovic. Max-Linear Systems: Theory and Algorithms. Springer, 2010.

2. Francois Louis Baccelli, Guy Cohen, Geert Jan Olsder, Jean-Pierre Quadrat, "Synchronization and Linearity: An Algebra for Discrete Event Systems" John Wiley & Sons, 1993, pp. 514

9. Spectral clustering

Clustering - the task of dividing objects into groups of "similar" elements - is one of the most widely used approaches for data analysis, which has found its application in many tasks, ranging from statistics, computer science, biology to social sciences or psychology.

As a project, it is proposed to study the spectral clustering method and try to apply this algorithm to real data. Implementation of the algorithm can be found in the python package *sklearn*. As the data, one can use the most famous data set "Karate Club" from the python *networkx* package.

References:

1. http://people.csail.mit.edu/dsontag/courses/ml14/notes/Luxburg07_tutorial_spectral_clustering.pdf

2. https://scikit-learn.org/stable/modules/generated/sklearn.cluster.SpectralClustering.html

10. Inverse perspective transform (rectification)

Write a program (in Python with use of OpenCV, NumPy, SciPy, etc.) that would, given an image of the whiteboard in class with something written on it (or an A4 sheet of paper), taken at an non-right angle, remove the perspective from it – transform the part of the image that is the quadrilateral of the whiteboard (or a sheet of paper) to a proper perspective-free rectangle of it. This means implementing a certain linear transform.

You would have to detect the quadrilateral of the whiteboard on the image. If you're doing this project alone – you may use OpenCV's methods for it. If you're doing this project in a group – it would be good to separate it into two problems: 1) perspective removal for a detected quadrilateral 2) detection of the quadrilateral – and for the 2-nd problem, implement the detection yourselves – this can be done with the generalized Hough transform.

Extra points will be awarded if the detection and rectification are done in real-time, so the program can run on a video stream!

References:

1) *Klein, Philip N. "Coding the matrix: Linear algebra through applications to computer science." Newtonian Press, 2013. – Chapter 5.12, Lab: Perspective Rectification*
2) *https://github.com/luczeng/HoughRectangle – a C++ project on using the generalized Hough transform to detect a rectangle*

11. Topic modelling with TF-IDF and non-negative matrix factorisation

Singular Value Decomposition (SVD) is good for detecting "dominating" features in a dataset, but its drawback is that the matrix elements can be negative numbers – that is not always interpretable: if the dataset vectors are, for example, grayscale images – negative values of pixel intensity make no sense. In such situations, Non-negative Matrix Factorisation (NMF), can be used – it restricts matrix elements to non-negative numbers.

The aim of this project is to use NMF to detect the topics of some text. Use TF-IDF (term frequency inverse document frequency) encoding to convert texts to vectors.

Extra points will be awarded if you figure out how to use this approach on something interesting beyond texts, or use it on an interesting dataset. You should be able to explain all the undergoing linear algebra, etc.!

References:

1) *https://scikit-learn.org/0.19/datasets/twenty_newsgroups.html – The 20 Newsgroups Dataset that you can use to work out the approach*

12 Topological Data Analysis and hand-written digit recognition

The goal of this project is to get familiar with homology – a notion from algebraic topology, and to implement an algorithm to compute it – then applying it to do some simple topological analysis of the real data.

Consider a (finite) sequence of vector spaces: V_1, V_2 V_3, … You will be given a sequence of linear maps (just matrices of 0-s and ±1-s of appropriate dimensions): $D_i : V_{i+1} \to V_i$, such that $Im(D_{i+1}) \subset Ker(D_i)$. The problem is to find the basis (vectors) of the quotient vector space $Ker(D_i)/Im(D_{i+1})$. This will give us what's called the k-th homology group.

Once you figure out how to solve the above abstract problem (and implement an algorithm that does it), we will use it to find 1-th homologies of the MNIST dataset in the following way (note that MNIST is a collection of images of hand-written digits):

- Naturally turn MNIST images into vectors in R^n
- Properly define a metric on this vector space
- "Grow balls" around MNIST data-points, see how they intersect (you don't have to solve this problem, there is software for that) – construct the Čech complex for given ball size
- For this Čech complex, compute the dimension of H_1, 1-st homology
- Using this information, try to recognize

Reference:

1) *David Austin, "Finding holes in the data", AMS Feature Column – http://www.ams.org/publicoutreach/feature-column/fc-2016-12*