# Concept paper (proposal) for the master´s project

## Template

### Note:

*A concept paper is a document in which you define the scholarly framework for your thesis at an early stage and develop a realistic working plan. Since it is very hard to predict the framework conditions as well as your own research in the context of scientific work, the concept paper will need to be modified on a regular basis. Ideally, the concept paper provides the basis for the thesis. Depending on the topic, style and preliminary work, a concept paper can be approximately 6 to 10 pages long.*

*The italicized sections are intended merely to provide explanations and can be deleted.*

Name:

Matriculation number:

(Project-) Title:

Institute:

Name of supervisor:

## Topic

*Concisely describe (no more than half a page) the issue that you want to work on. The relevance of the topic should be apparent.*

The Federated SVD (FedSVD) is an algorithm that allows for multiple actors to collectively compute the SVD from combined data sets, without revealing their source data to one another. This allows for modification of algorithms that use SVD as a subroutine, so that they may satisfy privacy requirements, by using FedSVD instead of SVD.

FedSVD is a relatively new algorithm, and while there exists some papers describing the usage of FedSVD as a subroutine, there is still some unexplored territory regarding how FedSVD might be used in the future. This paper therefore wishes to analyze how FedSVD might be used as a subroutine in numerical algorithms with respect to parameters such as running times, accuracy, numerical stability, usefulness, etc.

Therefore, a potential name for the paper could be “An Analysis of the Usefulness of Federated SVD as a Subroutine in Numerical Algorithms”. The research should be based on the FedSVD algorithm from Cai et al., and how it may be implemented as a subroutine in new algorithms, as well as a comparison of how these algorithms function with FedSVD as a subroutine instead of SVD.

It should be noted that this paper is solely interested in the numerical properties of FedSVD, and not of the privacy conserving properties.

## State of science and technology

*Try to gain as comprehensive an overview of the state of science and technology for your topic as possible and provide sources as evidence.*

*For the literature research, use your colleagues’ citation lists in the PhD dissertation (read the texts!) as well as performing your own free database searches. Correctly and fully understanding how your own work is positioned within the international state of science and technology as well as the institute’s research traditions is an important element of high-quality academic work.*

Federated SVD is a new field of research, with the first paper mentioning it being published in 2019, namely "Federated Singular Value Decomposition" by Li et al. This paper introduced a federated learning algorithm for SVD, which allows multiple parties to collaboratively compute the SVD of a matrix without sharing their data. Since then, more papers have emerged, including “Practical Lossless Federated Singular Vector Decomposition over Billion-Scale Data” by Cai et al. which claims its algorithm to be near lossless, and with minimal extra computational costs compared to the ordinary SVD.

Cai et al. proposes some algorithms where FedSVD could be used as a subroutine, e.g. Principal Component Analysis, Linear Regression and Latent Semantic Analysis.

## Objective of your project

*Demonstrate how you plan to proceed with your scientific project to achieve the goal. Show the chosen procedure and the selected examination methods. Describe the aspects/areas in which your contribution will go beyond the state of technology described above.*

1. Read the articles describing the different FedSVD methods, and how they may be incorporated as subroutines. Which algorithms have already been explored, and which have not? Most likely decide to implement the FedSVD described by Cai et al. as the subroutine, but there might be better alternatives?
2. Implement FedSVD in Matlab. Take inspiration from the open source code on Github.
3. Implement FedSVD as a subroutine in different algorithms.
4. Analyze the results of step 3.
5. Write the article.

The contributions here will go beyond the state of the state of technology by implementing FedSVD as a subroutine in algorithms that it has not been implemented in yet.

## Work plan, necessary resources

*Try to create a timeframe for your thesis now. Consider by what time frame and conditions you are bound and where you are dependent on your supervisor. To demonstrate your working plan create a “Gantt chart”.  
Describe which methods you will use to test your statements and why these methods are especially suitable. Describe which resources (e.g. experimental resources) you will need for your task.*

Week 1: Literature research. Deep dive into existing papers describing FedSVD as a subroutine. Which algorithms are already modified, and which are not? How does FedSVD really function as a subroutine?

Week 2: Understanding the code implementation of FedSVD and implementing it in Matlab. It exists as a Github repos (https://github.com/Di-Chai/FedEval/tree/master/research/FedSVD), but how can it be implemented in Matlab? How the privacy aspect of the algorithm is implemented is not of interest. Solely interested in the numerical properties of FedSVD, and not of the privacy conserving properties. But it should be implemented exactly as described in the paper, but without addressing how the different matrices are distributed among the participants.

Week 3: Implementation of FedSVD as a subroutine in different algorithms in Matlab.

Week 4: Implementation of FedSVD as a subroutine in different algorithms in Matlab.

Week 5: Testing and analysis of the FedSVD as a subroutine, wrt. quantities like running times, accuracy, general usefulness (can’t get the entire V matrix, how does this affect the algorithms?), etc., etc.

Week 6: Testing and analysis of the FedSVD as a subroutine, wrt. quantities like running times, accuracy, general usefulness (can’t get the entire V matrix, how does this affect the algorithms?), etc., etc.

Week 7: Writing paper

Week 8: Writing paper

Week 9: Slingringsmonn

Week 10: Slingringsmonn

Week 11: Slingringsmonn

## Literature

*List of used literature.*

1. Structure

*The outline should provide an overview of the topics covered, the focus and the guiding thread of the work.*

|  |  |  |  |
| --- | --- | --- | --- |
| *Content presentation created by: \_\_\_\_\_\_\_\_\_\_\_\_\_\_,\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Place, date* | *\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_,\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Place, date* | | *\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_,\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Place, date* |
| Student | | First examiner (supervisor) | Second examiner (Center of Key Competencies) |

*A formal review will be done by the Center of Key Competencies.*

*After control by your supervisor (first examiner) please hand a written version to the Center of Key Competencies.*