

DMP title

Project Name FWO DMP Grant 12Y7622N - DMP title

Project Identifier 0000-0002-7969-8565

Grant Title 12Y7622N

Principal Investigator / Researcher Puya Latafat

Description Efficient algorithms for nonconvex minimax problems is the subject of investigation in this project. Such problems have recently emerged in a diverse range of applications related to robust learning and dynamical decision making in uncertain environments. For example, adversarial training that addresses the issue of sensitivity of machine learning models, or distributionally robust optimization that tackles decision making under uncertainty, and many other applications are all formulated as minimax problems. Until very recently, algorithms for solving such potentially nonconvex minimax problems have had little (if any) convergence guarantees, and have often been only addressed by means of heuristics. The envisioned datatypes consist of common electronic documents used for our manuscripts and reports (e.g. pdf), as well as open source software that would be publicly available and hosted on GitHub. The developed software would have specific format related to the programming language (Julia, Python and C code). Moreover, in our benchmarks we would use public datasets that are freely available online and have been used in other studies. The raw output of the simulation codes would come in variety of standard formats that can be imported by other software. All the research articles submitted to journals and conference proceedings will be made publicly available on ArXiv e- print archive.

Institution KU Leuven

1. General Information

Name applicant

Puya Latafat

FWO Project Number & Title

Project number 12Y7622N:

Efficient Algorithms for Large-Scale Minimax Problems

Affiliation

- KU Leuven

Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics

2. Data description

Will you generate/collect new data and/or make use of existing data?

- Generate new data
- Reuse existing data

Describe in detail the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a table (see example) or as a data flow and per WP or objective of the project. If you reuse existing data, specify the source of these data. Distinguish data types (the kind of content) from data formats (the technical format).

WP1 First-order methods for nonsmooth, nonconvex, structured minimax problems : To develop first-order proximal methods for nonsmooth nonconvex minimax problems with additive cost structures. To further expand the spectrum of both the covered algorithms and the tractable problems through extensions with Bregman distances. The numerical verification and publications would involve

Type of data	Format	Volume	How created
publication	.pdf	1-10MB	the pdf files generated by latex for submission to journals and conferences
Algorithm iterates, and all other data outputs used for plotting	.dat	1-10MB	Collected algorithm iterates on toy example problems using Python, Julia or C
plots	.png, .pdf	0-10 MB	plots generated from the iterates using latex (tikz)
codes and demos	.jl, .py, .mat	0-1 MB	Pytho, Julia, MATLAB code for solving minimax problems as well as demos
datasets used in simulations	.csv	1-5 GB	publicly available datasets used in experiments

WP2 Block-coordinate, stochastic and Newton-type algorithms: To develop distributed algorithms suitable for large-scale minimax problems. To develop stochastic and incremental aggregated methods for problems with finite sum structure. To accelerate and robustify our algorithms by developing Newton-type schemes.

The numerical packages and datasets are built upon the work in the previous workpackage and use public datasets.

Type of data	Format	Volume	How created
publication	.pdf	1-10MB	the pdf files generated by latex for submission to journals and conferences
Algorithm iterates, and all other data outputs used for plotting	.dat	1-10MB	Collected algorithm iterates on toy example problems using Python, Julia or C
plots	.png, .pdf	0-10 MB	plots generated from the iterates using latex (tikz)
codes and demos	.jl, .py, .mat	0-1 MB	Pytho, Julia, MATLAB code for solving minimax problems as well as demos
datasets used in simulations	.csv	1-5 GB	publicly available datasets used in experiments

WP3 Distributionally robust optimization for machine learning and control: To develop efficient algorithms for distributionally robust optimization, with applications to robust ML and MPC leading to learning paradigms with performance guarantees with respect to the (unknown) true probability distribution.

The nature of the codes are still very similar to the previous packages and builds upon them.

Type of data	Format	Volume	How created
publication	.pdf	1-10MB	the pdf files generated by latex for submission to journals and conferences
Algorithm iterates, and all other data outputs used for plotting	.dat	1-10MB	Collected algorithm iterates on toy example problems using Python, Julia or C
plots	.png, .pdf	0-10 MB	plots generated from the iterates using latex (tikz)
code and demos	.jl, .py, .mat	0-1 MB	Pytho, Julia, MATLAB code for solving minimax problems as well as demos
datasets used in simulations	.csv	1-5 GB	publicly available datasets used in experiments
videos for MPC	.mp4	1-5GB	videos from motion planning for MPC demos generated using ffmpeg

3. Legal and ethical issues

Will you use personal data? If so, shortly describe the kind of personal data you will

use. Add the reference to your file in KU Leuven's Register of Data Processing for Research and Public Service Purposes (PRET application). Be aware that registering the fact that you process personal data is a legal obligation.

- No

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, add the reference to the formal approval by the relevant ethical review committee(s)

- No

Does your work possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?

- No

Do existing 3rd party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?

- No

4. Documentation and metadata

What documentation will be provided to enable reuse of the data collected/generated in this project?

1- All the codes will involve standard naming convention in the relevant programming language (eg. Julia and Python). For example, there will be a readme.txt file, project.toml and licence files as required by standard package registration tools. All the packages will have an src folder (the main code), a test folder (unit tests), demos folder, and documentation.

2- The datasets used will consist of publicly available datasets that are typically in the .csv format. There are readme files describing the number of features and number of data points for each dataset. When a part of a given dataset is used, the truncated or preprocessed dataset will be accompanied by a readme.txt file explaining exactly where the data is taken from and how to load it.

Will a metadata standard be used? If so, describe in detail which standard will be used. If no, state in detail which metadata will be created to make the data easy/easier to find and reuse.

- Yes

As we are mainly dealing with data from numerical simulations, we will use a dedicated metadata standard for this purpose such as the EngMeta standard.

5. Data storage and backup during the FWO project

Where will the data be stored?

1. A time-stamped version of the data will be stored on the STADIUS DATASET Server, which is our research unit central storage server. Copies can be made and kept on personal devices.

2. To allow for efficient collaboration with researchers from other research groups, the data will additionally be stored on GitHub (subject to the Github regulations).

How is backup of the data provided?

All data on the STADIUS DATASET Server are backed up daily and replicated to an off-site storage system housed in the ICTS data center.

Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely. If no or insufficient storage or backup capacities are available then explain how this will be taken care of.

- Yes

The STADIUS DATASET Server has a total capacity of 14.88 TB. The capacity of the dataset server is monitored daily by the ESAT system admins.

What are the expected costs for data storage and back up during the project? How will these costs be covered?

All data storage and back up will be performed on the STADIUS DATASET Server in pre-existing storage facilities of the Department of Electrical Engineering (ESAT) without the need of purchasing new infrastructures.

Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

The access of the data on the STADIUS DATASET Server is regulated by an access control list (ACL)

that grants:

- read-write access to the project owner and the FWO fellow
- read-only access to specific users

The ACL is managed by the project owner (Panagiotis Patrinos). Client computers can access the data

using:

- SMB2 (or higher) from specific IP ranges
- NFSv4 from specific (IT managed) systems

6. Data preservation after the FWO project

Which data will be retained for the expected 5 year period after the end of the project? In case only a selection of the data can/will be preserved, clearly state the reasons for this (legal or contractual restrictions, physical preservation issues, ...).

All research data detailed previously will be retained for at least 5 years, conform the FWO data preservation policy.

Where will the data be archived (= stored for the longer term)?

The data will be stored on the university's central servers (with automatic back-up procedures) for at least 10 years, conform the KU Leuven RDM policy.

What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?

All data hosted on the STADIUS DATASET Server will be stored for a long term in pre-existing storage facilities of the Department of Electrical Engineering (ESAT) without the need of purchasing new infrastructures. Similarly, the data hosted on GitHub will be stored for a long term.

7. Data sharing and reuse

Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

- No

Which data will be made available after the end of the project?

All research articles submitted to journals and conference proceedings will be made publicly available.

Furthermore, all relevant source code and datasets will be made publicly available in previously mentioned formats. Software packages in Julia, Python along with demos and documentation will be made publicly available on github under the MIT License.

Where/how will the data be made available for reuse?

- In an Open Access repository

1. The research articles submitted to journals and conference proceedings will be made publicly available on ArXiv e-prints archive.

2. The source code will be released publicly on GitHub under the MIT license, along with the corresponding datasets.

When will the data be made available?

- Upon publication of the research results

Upon publication of new research results, the corresponding research article will be made publicly available on ArXiv and the corresponding source code and datasets will be released publicly on GitHub.

Who will be able to access the data and under what conditions?

1. The research articles submitted to journals and conference proceedings will be made publicly available on ArXiv e-prints archive under the perpetual, non-exclusive license. Therefore, it will be available to anyone for any purpose, provided that they give appropriate credit to the creators.
2. The source code will be released publicly on GitHub under the MIT license, along with the corresponding datasets. Hence, it will be available for everyone to use, change, and distribute the software, only requiring preservation of copyright and license notices.

What are the expected costs for data sharing? How will the costs be covered?

Since free services like GitHub and ArXiv are used to distribute the research data, there are no expected costs.

8. Responsibilities

Who will be responsible for data documentation & metadata?

The FWO Fellow will be responsible for the data documentation and the metadata.

Who will be responsible for data storage & back up during the project?

The FWO Fellow will be responsible for the data storage. The system administrators and data manager of the research division are responsible for the back up during and after the project.

Who will be responsible for ensuring data preservation and reuse ?

The FWO Fellow will be responsible for ensuring data preservation and reuse.

Who bears the end responsibility for updating & implementing this DMP?

The project owner (Panagiotis Patrinos) bears the end responsibility of updating & implementing this DMP