Fabian Pedregosa

Professional experience

Google DeepMind

2023.05 – present, Senior Research Scientist, Google DeepMind

I currently lead an optimization team of 10 members within Google DeepMind. Some of my latest projects include the organization of a machine unlearning competition, the development of the Optax library, as well as research into deep learning theory optimization [1, 2]



2018.10 – **2023.05, Senior Research Scientist**, *Google Research, Brain team* During this time I developed a novel analysis of optimization methods based on an average-case analysis [8, 5], and co-developed the JAXopt optimization library [6]. Since 2020 I also lead an optimization team within Google Brain.



2017.09–2018.08, Postdoctoral researcher, UC Berkeley / ETH Zürich

Marie Skłodowska-Curie postdoctoral fellow between UC Berkeley and ETH Zürich. During this period, I developed automatic step-size tuning methods for the Frank-Wolfe algorithm [9], as well as scalable methods for complex non-smooth penalties [10] **Supervisors**: Laurent El Ghaoui (UC Berkeley) and Thomas Hofmann (ETH Zürich).



Ínnía

2015.05 – 2017.05, Postdoctoral researcher, École Normale Supérieure and INRIA-Paris

Postdoctoral researcher at the SIERRA team, a joint INRIA-ENS team, working on fundamental aspects of optimization and machine learning. During this period, among other things, I developed a novel framework to analyze parallel asynchronous optimization methods [11] and hyperparameter optimization methods [14] that are significantly more scalable than previous state of the art.

Supervisor: Alexandre D'Aspremont.

2009.12 - 2012.10, Engineer, INRIA Saclay, France

Lead developer and maintainer of the highly successful open source library scikit-learn. I made the first public release of the project in February 2010 and in the following years my work attracted a large amount of contributors, counting more than 30 active contributors in 2012. This library is nowadays considered the reference machine learning library in Python, with an associated paper [20] counting over 48k citations.

Supervisors: Gael Varoquaux and Bertrand Thirion.



Education

10/2011 – 02/2015 PhD in Computer Science, Université Pierre et Marie Curie (Paris VI) and INRIA Saclay, Paris, France

My PhD thesis, entitled *Feature extraction and supervised learning on fMRI: from practice to theory* [16], focuses on the application of statistical learning techniques to functional neuroimaging datasets. It includes prediction of a cognitive state from brain activation images [15] and on the estimation of brain activation images from raw fMRI signal [17]. **Advisors**: Alexandre Gramfort and Francis Bach.



07/2003–10/2010, Degree in Pure and applied Mathematics, *Universidad de Granada*, Granada, Spain

I studied Mathematics at University of Granada. The University of Granada consistently ranks as the top Spanish university. Being founded in 1531, it is also one of the oldest universities in the world. This is was a monolythical degree of 5 years, equivalent to a Bachelors + Masters degree.

Selected publications

According to Google Scholar, my work has been cited 115071 times and my current h-index is 26.

Below is a list with the most representative publications for my two main themes: "optimization and machine learning" and "scientific software". A full list of publications can be found in my Google Scholar profile and in my website.

Optimization and machine learning

- [1] Vincent Roulet, Atish Agarwala, and **P.** "On the Interplay Between Stepsize Tuning and Progressive Sharpening". *arXiv preprint arXiv:2312.00209* (2023).
- [2] Atish Agarwala, Jeffrey Pennington, and Fabian Pedregosa. "A second order regression model shows edge of stability behavior". *ICML 2023* (2022).
- [3] Baptiste Goujaud, Damien Scieur, Aymeric Dieuleveut, Adrien B Taylor, and P. "Super-acceleration with cyclical step-sizes". International Conference on Artificial Intelligence and Statistics. PMLR [blog post]. 2022.
- [4] Damien Scieur, Quentin Bertrand, Gauthier Gidel, and **P.** "The Curse of Unrolling: Rate of Differentiating Through Optimization". *NeurIPS 2022* (2022).
- [5] Courtney Paquette, Bart van Merriënboer, Elliot Paquette, and **P.** "Halting time is predictable for large models: A universality property and average-case analysis". *Foundations of Computational Mathematics* (2021).
- [8] **P.** and Damien Scieur. "Average-case acceleration through spectral density estimation". *Proceedings* of the 37rd International Conference on Machine Learning (2020).
- [9] **P.**, Geoffrey Negiar, Armin Askari, and Martin Jaggi. "Linearly convergent Frank-Wolfe with backtracking line-search". *AISTATS*. PMLR. 2020.
- [10] P., Kilian Fatras, and Mattia Casotto. "Proximal splitting meets variance reduction". AISTATS. PMLR. 2019, pp. 1–10.
- [11] P., Rémi Leblond, and Simon Lacoste-Julien. "Breaking the nonsmooth barrier: a scalable parallel method for composite optimization". Advances in Neural Information Processing Systems 30 (NIPS) [paper selected for spotlight presentation, top 3% of submitted papers]. 2017.
- [12] **P.**, Francis Bach, and Alexandre Gramfort. "On the consistency of ordinal regression methods". *Journal of Machine Learning Research* (2017).
- [14] P. "Hyperparameter optimization with approximate gradient". Proceedings of the 33rd International Conference on Machine Learning (2016).

Scientific software

- [6] Mathieu Blondel, ..., **P.**, and Jean-Philippe Vert. "Efficient and Modular Implicit Differentiation". *arXiv preprint arXiv:2105.15183* (2021).
- [13] Aaron Meurer, ..., P., et al. "SymPy: Symbolic computing in Python". PeerJ Computer Science (2017).
- [18] Lars Buitinck, Gilles Louppe, Mathieu Blondel, P., Andreas Mueller, et al. "API design for machine learning software: experiences from the scikit-learn project". European Conference on Machine Learning and Principles and Practices of Knowledge Discovery in Databases. 2013.
- [20] **P.**, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, et al. "Scikit-learn: machine learning in Python". *Journal of Machine Learning Research* (2011).



PhD Thesis

[16] Fabian Pedregosa-Izquierdo. "Feature extraction and supervised learning on fMRI: from practice to theory". PhD thesis. Université Pierre et Marie Curie-Paris VI, 2015.



Grants & Awards

2017– 2018 (initial funding until 2020), *Marie Skłodowska-Curie Global Fellowship*, Research fellowship

Grant to lead a 3-year project on optimization, machine learning and natural language processing. The Marie Skłodowska-Curie are "Europe's most competitive and prestigious award" (Source: Wikipedia). **Total amount**: \in 359k \approx \$384k

Teaching and mentoring

2018–2021 Google Internship program

I co-mentored 5 intern at Google: Fartash Fahgri (co-hosted with Nicolas Le Roux), Konstantin Mischenko (with Courtney Paquette), Geoffrey Negiar (with Danny Tarlow), Teodora Baluta (with Danny Tarlow) and Tomas Gonzalez Lara (with Cristobal Guzman).

- 2021 ICML 2021, *Tutorial*, Random Matrix Theory and Machine Learning Together with Courtney Paquette, Jeffrey Pennington and Thomas Trogdon, I gave a tutorial entitled "Random Matrix Theory and Machine Learning" during the ICML conference. Course website: https://random-matrix-learning.github.io/.
- 2018 **McGill university**, *Invited Lecture*, Graduate level Course: COMP-652, Lecture title: "A birds-eye view of optimization", Slides: http://fa.bianp.net/teaching/2018/COMP-652/.
- Fall 2017 UC Berkeley, Professor in charge, Master level Course title: "Data Science". Website: https://dsl2l2017.github.io Responsabilities: 12 Lectures (36h course) and supervising projects.
- 2016 2017 **ENSAE (French national statistics school)**, *Teaching assistant*, Master Level Course title: "Distributed and Stochastic Optimization for Machine Learning". Responsabilities: Theoretical and applied projects (10h). Professor in charge: Marco Cuturi.
- 2016 2017 **Université Paris-Dauphine**, *Professor in charge*, Master level Course title: "Introduction to machine learning with scikit-learn". Responsabilities: 10 Lectures (20h) and supervise student projects. Teaching material: https://github.com/fabianp/mash_2016_sklearn_intro

Selected Presentations

Communication is an integral part of research, and I strive to give talks that are clear and engaging.

- 2021 ICML 2021, *Tutorial*, Random Matrix Theory and Machine Learning Together with Courtney Paquette, Jeffrey Pennington and Thomas Trogdon, I gave a tutorial entitled "Random Matrix Theory and Machine Learning" during the ICML conference. Videos: https://slideslive.com/icml-2021/tutorial-random-matrix-theory-and-ml-rmtml. Tutorial website: https://random-matrix-learning.github.io/.
- 2017 NIPS 2017, spotlight presentation (top 3% of submitted papers), Long Beach Convention Center, United States Breaking the Nonsmooth Barrier: A Scalable Parallel Method for Composite Optimization [11]. Video recording.
- 2016 ICML 2016, New York City, United States Presentation at the International Conference for Machine Learning. Talk title: "Hyperparameter optimization with approximate gradients" [14]. Video recording.

Editorial Services

2023 ICLR post track, Core organizer

As core organizer of the ICLR blog post track, I led all technical aspects of this call: defined the work pipeline for submitting posts, implemented checks on blog posts using github actions, implemented tweaks and fixed bugs on the template, etc.

2021-present	Journal of Machine Learning Research, Editorial Board, Webmaster
	Define and create content for the journal website (http://jmlr.org/), publish accepted
	papers.
2021–2023	Transactions on Machine Learning Research, Fouding member, Managing editor
	and Webmaster
	Responsabilities: define scope of journal, define and implement user interface for reviewing system, define and implement journal website. Journal launched in February 2022 (http://tmlr.org).
2018–2021	Journal of Machine Learning Research, Editorial Board, Managing Editor
	Responsabilities: main contact point for authors, reviewers and action editors, define and implement improvements on the submission system.
2017, 2018	International Conference on Machine Learning, Program committee, Reviewer
2017, 2018	Neural Information Processing Systems (NIPS), Program committee, Reviewer
2017-present	SIAM Journal in Optimization, Reviewer
2016–2017	NeuroImage, Reviewer
2014–present	Journal of Machine Learning Research, Reviewer

Open Source Contributions

I believe that advancement of science goes hand in hand with advancement of the software tools we use. Over the years, I have developed many open source packages, mostly of the Python language, from scientific packages to profiling tools:

2023–Present	Python, Optax, Core contributor
	Optimization library for JAX. I created a gallery of examples for this library, and contributed bugfixes.
2023–Present	Python, JAXOpt, Core contributor
	Hardware accelerated, batchable and differentiable optimizers in JAX. As a founding member of this library, my contributions have been to define the scope and implement core functionality, as well as contribute large parts of the documentation. Project website: https://jaxopt.github.io.
2015–2018	Python, lightning, Core contributor
	lightning is a library for large-scale linear classification, regression and ranking in Python.
2011–Present	Python, memory_profiler, Author
	Profile memory consumption of Python programs
2010–2012	Python/Cython/C , <i>scikit-learn</i> , Lead developer
	Maintainer and lead developer of the scikit-learn machine learning library.
2010–2012	Python/Fortran, scipy.linalg, Maintainer
	Maintainer of the linear algebra module in SciPy (Python's most popular scientific library).
2007–2009	Python, SymPy
	Active contributor to project SymPy (design, patches, etc.), including a Summer of Code scholarship in 2009 $$

Languages

Spanish	Mother language
French	Fluent in spoken and written (12 years in french-speaking contries)
English	Fluent in spoken and written
German	Understanding & writing
Italian	Understanding