Daniel LeJeune

Dept. of Statistics
390 Jane Stanford Way
Stanford, CA 94305-4020
daniel@dlej.net
https://dlej.net/

CURRICULUM VITAE

RESEARCH INTERESTS

Statistical machine learning, randomized algorithms, high-dimensional statistics, theory of deep learning, optimization, signal processing

ACADEMICS

Jan 2023–Present **Postdoctoral Scholar, Statistics**, *Stanford University*, Stanford, CA.
Host: Emmanuel Candès

Aug 2017–Dec 2022 **Ph.D., Electrical and Computer Engineering**, *Rice University*, Houston, TX.
Thesis: Ridge Regularization by Randomization in Linear Ensembles
Advisor: Richard G. Baraniuk
Committee members: Genevera I. Allen, Anshumali Shrivastava, Francis Bach

Oct 2021–Feb 2022 **Guest Doctoral Candidate**, *Technical University of Munich*, Munich, Germany.
Host: Reinhard Heckel

Sep 2014–Apr 2016 **M.S., Electrical and Computer Engineering**, *University of Michigan*, Ann Arbor, MI, *4.0/4.0*.
Major: Signal Processing; *Minor:* Computers

Aug 2010–May 2014 **B.S., Engineering**, *McNeese State University*, Lake Charles, LA, *4.0/4.0*.
Major Concentration: Electrical Engineering; *Minors:* Computer Science, Mathematics

HONORS AND AWARDS

2024 **CPAL Rising Stars Award**, *Conference on Parsimony and Learning*, Hong Kong.

2023 **Ralph Budd Ph.D. Engineering Thesis Award Finalist**, *Rice University*, Houston, TX.

2017–2022 **Texas Instruments Distinguished Fellowship**, *Rice University*, Houston, TX.

2019–2020 **Data2Knowledge Lab Fellow**, *Rice University*, Houston, TX.

2014 **Honors College Graduate**, *McNeese State University*, Lake Charles, LA.

2010–2014 **President's List** (8 semesters), *McNeese State University*, Lake Charles, LA.

2013 **Engineering Student Leadership Award**, *McNeese State University*, Lake Charles, LA.

2012 **Outstanding IEEE Student Award**, *McNeese State University*, Lake Charles, LA.

PUBLICATIONS (*equal contribution)

PREPRINTS

[1] P. Patil* and **D. LeJeune*** (rev. 2023). Asymptotically free sketched ridge ensembles: Risks, cross-validation, and tuning. (First revision under review at) *International Conference on Learning Representations (ICLR)*. arXiv: 2310.04357.

[2] S. Alemohammad, J. Casco-Rodriguez, L. Luzi, A. I. Humayun, H. Babaei, **D. LeJeune**, A. Siahkoohi, and R. G. Baraniuk (rev. 2023). Self-consuming generative models go MAD. (First revision under review at) *International Conference on Learning Representations (ICLR)*. arXiv: 2307.01850.

[3] **D. LeJeune**, J. Liu, and R. Heckel (rev. 2023). Monotonic risk relationships under distribution shifts for regularized risk minimization. (First revision under review at) *Journal of Machine Learning Research (JMLR)*. arXiv: 2210.11589.

[4] Y. Dar, **D. LeJeune**, and R. G. Baraniuk (rev. 2023). The common intuition to transfer learning can win or lose: Case studies for linear regression. (Second revision under review at) *SIAM Journal on Mathematics of Data Science (SIMODS)*. arXiv: 2103.05621.

PEER-REVIEWED PUBLICATIONS

[5] **D. LeJeune***, P. Patil*, H. Javadi, R. G. Baraniuk, and R. J. Tibshirani (2024). Asymptotics of the sketched pseudoinverse. (To appear in) *SIAM Journal on Mathematics of Data Science (SIMODS)*. arXiv: 2211.03751.

[6] **D. LeJeune** and S. Alemohammad (2024). An adaptive tangent feature perspective of neural networks. (To appear in) *Conference on Parsimony and Learning (CPAL)*. arXiv: 2308.15478.

[7] L. Luzi, **D. LeJeune**, A. Siahkoohi, S. Alemohammad, V. Saragadam, H. Babaei, N. Liu, Z. Wang, and R. G. Baraniuk (2024). TITAN: Bringing the deep image prior to implicit representations. (To appear in) *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

[8] P. K. Kota, H.-A. Vu, **D. LeJeune**, M. Han, S. Syed, R. G. Baraniuk, and R. A. Drezek (2023). Expanded multiplexing on sensor-constrained microfluidic partitioning systems. *Analytical Chemistry* 95.48, pp. 17458–17466. DOI: 10.1021/acs.analchem.3c01176.

[9] V. Saragadam, **D. LeJeune**, J. Tan, G. Balakrishnan, A. Veeraraghavan, and R. G. Baraniuk (2023). WIRE: Wavelet implicit neural representations. *IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR)*. arXiv: 2301.05187.

[10] J. Tan*, **D. LeJeune***, B. Mason, H. Javadi, and R. G. Baraniuk (2023). A blessing of dimensionality in membership inference through regularization. *International Conference on Artificial Intelligence and Machine Learning (AISTATS)*. arXiv: 2205.14055.

[11] P. K. Kota, **D. LeJeune**, R. A. Drezek, and R. G. Baraniuk (2022). Extreme compressed sensing of Poisson rates from multiple measurements. *IEEE Transactions on Signal Processing* 70, pp. 2388–2401. DOI: 10.1109/TSP.2022.3172028.

[12] **D. LeJeune**, H. Javadi, and R. G. Baraniuk (2021). The flip side of the reweighted coin: Duality of adaptive dropout and regularization. *Advances in Neural Information Processing Systems (NeurIPS)*. arXiv: 2106.07769.

[13] S. Alemohammad, H. Babaei, R. Balestriero, M. Y. Cheung, A. I. Humayun, **D. LeJeune**, N. Liu, L. Luzi, J. Tan, Z. Wang, and R. G. Baraniuk (2021). Wearing a MASK: Compressed representations of variable-length sequences using recurrent neural tangent kernels. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. DOI: 10.1109/ICASSP39728.2021.9413450.

[14] T. Yao, **D. LeJeune**, H. Javadi, R. G. Baraniuk, and G. I. Allen (2021). Minipatch learning as implicit ridge-like regularization. *IEEE International Conference on Big Data and Smart Computing (BigComp)*. DOI: 10.1109/BigComp51126.2021.00021.

[15] **D. LeJeune**, H. Javadi, and R. G. Baraniuk (2020). The implicit regularization of ordinary least squares ensembles. *International Conference on Artificial Intelligence and Statistics (AISTATS)*. arXiv: 1910.04743.

[16] **D. LeJeune**, G. Dasarathy, and R. G. Baraniuk (2020). Thresholding graph bandits with GrAPL. *International Conference on Artificial Intelligence and Statistics (AISTATS)*. arXiv: 1905.09190.

[17] **D. LeJeune**, R. Heckel, and R. G. Baraniuk (2019). Adaptive estimation for approximate k-nearest-neighbor computations. *International Conference on Artificial Intelligence and Statistics (AISTATS)*. arXiv: 1902.09465.

[18] A. Aghazadeh, R. Spring, **D. LeJeune**, G. Dasarathy, A. Shrivastava, and R. G. Baraniuk (2018). MISSION: Ultra large-scale feature selection using count-sketches. *International Conference on Machine Learning (ICML)*. arXiv: 1806.04310.

[19] J. Matterer and **D. LeJeune** (2018). Peer group metadata-informed LSTM ensembles for insider threat detection. *Florida Artificial Intelligence Research Society (FLAIRS) Conference*.

PATENTS

[20] P. K. Kota, R. G. Baraniuk, **D. LeJeune**, R. A. Drezek, and H.-A. Vu (2023). Poisson signal recovery from multiple measurements. U.S. pat. 2023/0386603 A1. William Marsh Rice University. Pending.

TECHNICAL REPORTS

[21] **D. LeJeune**, R. Balestriero, H. Javadi, and R. G. Baraniuk (2019). Implicit rugosity regularization via data augmentation. arXiv: 1905.11639.

THESIS

[22] **D. LeJeune** (2022). Ridge regularization by randomization in linear ensembles. PhD thesis. Rice University.

TALKS

Jan 2024 **Emergent Properties of Heuristics in Machine Learning**.
Conference on Parsimony and Learning (CPAL), Hong Kong.

Jan 2024 **An Adaptive Tangent Feature Perspective of Neural Networks**.
Conference on Parsimony and Learning (CPAL), Hong Kong.

Aug 2023 **Equivalence of Sketching and Ridge Regularization**.
Department of Statistics Seminar at Stanford University.

Nov 2021– Jun 2023	**The Flip Side of the Reweighted Coin: Duality of Adaptive Dropout and Regularization**. SIAM Conference on Optimization, Seattle, WA (Jun 2023). [virtual] Bio-Data Science Group at LMU Munich, Germany (Feb 2022). Statistical Machine Learning group at ETH Zürich, Switzerland (Dec 2021). [recorded] Neural Information Processing Systems (NeurIPS), online (Dec 2021). [virtual] Scalable Parallel Computing Lab at ETH Zürich, Switzerland (Nov 2021).
Jan 2022	**Towards Understanding Adaptive Random Sampling in Ensemble Methods**. [virtual] SIERRA Project-team at INRIA, Paris, France.
Aug 2020	**The Implicit Regularization of Ordinary Least Squares Ensembles**. [recorded] International Conference on Artificial Intelligence and Statistics (AISTATS), online.
Aug 2020	**Thresholding Graph Bandits with GrAPL**. [recorded] International Conference on Artificial Intelligence and Statistics (AISTATS), online.

TEACHING

Spring 2021	**ELEC 631 Data Science for Education Seminar**, *Rice University*, Houston, TX. Coordinated student presentations and guest speakers.
Fall 2020, Fall 2018	**ELEC 478/578 Machine Learning**, *Rice University*, Houston, TX. Prepared tutorials, homework sets, and exams; held weekly office hours; managed grading.
Spring 2020, Fall 2019, Fall 2018	**STAT 435/535 Data Science Projects**, *Rice University*, Houston, TX. Mentored teams of undergraduate and early graduate students working to apply data science techniques to real-world data problems in the financial, emergency response, and energy domains.
Winter 2016	**EECS 545 Machine Learning**, *University of Michigan*, Ann Arbor, MI. Prepared tutorials, homework sets, and exams; held weekly office hours; managed grading.
Fall 2015, Winter 2015	**EECS 203 Discrete Mathematics**, *University of Michigan*, Ann Arbor, MI. Prepared homework sets, and exams; led weekly recitations; managed grading for 500+ students.
Spring 2014	**ELEN 410/510 Communication Theory**, *McNeese State University*, Lake Charles, LA. Prepared and graded quizzes and exams; gave lectures each week.

SELECTED STUDENT FEEDBACK FROM COURSE EVALUATIONS

"He answered every single little question I had even though he was in sort of a rush to get to every question which is valid considering that he wanted to help us out and not leave us at a disadvantage. He would even stay after even though he had a class right after in a different room. He is great."

"He was a really sweet guy, and made me feel accepted in a class where I felt like I was constantly filled with terror and failure."

"For office hours Daniel is my favorite GSI to go to. His comments are quick, clear, patient, and constructive, while still leading me to arrive at answers on my own."

"Really helpful when he showed his own tricks to solving different types of problems. Helped make the material more understandable. Felt very comfortable asking questions in discussion."

"Daniel has been doing a great job working through examples in discussion. When I have questions on a problem, Daniel is more than willing to work through the problem with me without giving me the answer."

"[Homework] difficulty was perfect; usually doable without being too hard or tedious. The TAs were very helpful as well (also helpful in their purposeful unhelpfulness, which forced me to answer the question for myself instead of them just giving me the answers)... [T]he TAs and instructor were very responsive on Piazza, often responding within minutes of questions being asked."

COMMUNITY SERVICE

2021–Present   **Conference and Workshop Organization**, *machine learning*.

Theory of Overparameterized Machine Learning (TOPML) Workshop
- Part of organizing committee (2024)
- Hosted lightning talk sessions (2021–2022)

2019–Present   **Reviewing**, *signal processing, statistics, and machine learning*.

Bernoulli (2023)
Neural Information Processing Systems (2021–2023)
International Conference on Artificial Intelligence and Statistics (2022)
International Conference on Machine Learning (2022)
Journal of Machine Learning Research (2020–2022)
Workshop on the Theory of Overparameterized Machine Learning (2021)
IEEE International Symposium on Information Theory (2021)
IEEE Transactions on Signal Processing (2019–2020)

2019–2020   **ECE GSA Treasurer**, *Rice University*, Houston, TX.

Planned social and professional development events for graduate students.

VOCATIONAL EXPERIENCE

2016–2017   **Associate Staff**, *MIT Lincoln Laboratory*, Lexington, MA.

Applied machine learning to cyber security as part of the Cyber Analytics and Decision Systems group.
- Introduced active learning for analyst feedback into a document ranking system, in addition to client- and server-side software development for a prototype of the same system.
- Collected datasets and evaluated baseline machine learning performances for DARPA's D3M program for automated machine learning, leveraging massive parallelism to perform hyperparameter search.
- On a multi-disciplinary team, incubated a method for passive detection of aircraft fuselage damage and won third place in an internal competition for this idea.

2015   **Summer Intern**, *MIT Lincoln Laboratory*, Lexington, MA.

- Developed a web-based visualization for a community detection analytic supporting large graphs.
- Investigated the performance of a community detection algorithm under partial observability conditions.

2013   **Automation Engineer Intern**, *Champion Technology Services*, Sulphur, LA.

- Programmed data acquisition, control, and visualization interfaces for industrial control systems.
- Developed an in-house application for streamlining project-specific document preparation.

2011–2013   **Tutor**, *McNeese State University*, Lake Charles, LA.

Tutored undergraduate students in math, computer science, engineering, physics, and chemistry.