Soledad Villar

Assistant Professor

Applied Mathematics and Statistics Mathematical Institute for Data Science Johns Hopkins University

⊠ soledad.villar@jhu.edu ™ www.ams.jhu.edu/villar/ ♥ solevillar

Education

- 2017 **PhD in Mathematics**, University of Texas at Austin, United States. Supervisor: Rachel Ward Dissertation: Relax, descend and certify: optimization techniques for typically tractable data problems
- 2012 **Master of Science in Mathematics**, Universidad de la República, Uruguay. Supervisor: Gonzalo Tornaría Dissertation: Gross formula on heights and special values of *L*-series.
- 2010 **Bachelor in Mathematics**, Universidad de la República, Uruguay. Supervisor: Gonzalo Tornaría Project: Pell curves cryptography and generalizations.
- 2010 Bachelor in Software Engineering, Universidad Católica del Uruguay, Uruguay.

Academic positions

- 2020-now Assistant Professor, Department of Applied Mathematics and Statistics. Mathematical Institute for Data Science, Johns Hopkins University, Baltimore, USA.
- 2017-2020 Moore-Sloan Research Fellow, Center for Data Science and Courant Institute of Mathematical Sciences, New York University, New York City, USA.
- 2017-2020 Collaboration Scientist, Algorithms and Geometry Simons Collaboration, Simons Foundation, New York City, USA.
- Fall 2017 Research Fellow, Bridging Continuous and Discrete Optimization, Simons Institute, University of California, Berkeley, USA. Mentor: Benjamin Recht

- Grants and awards

- 2022 NSF Collaborative Research: CIF: Medium., \$1.2M grant for 3-year project (\$900,000 to JHU). Understanding Robustness via Parsimonious Structures.
 (PI, co-PIs: Rene Vidal (JHU), Jeremias Sulam (JHU), Soheil Feizi (UMD))
- 2022 Amazon AI2AI Faculty Research Award, \$80,000 grant for 1-year project. Green AI: Powerful and Lightweight Machine Learning via Exploiting Symmetries.
 (PI)
- 2022 ONR, \$394,995 grant for 3-year project. Geometric methods for optimal matching and feature identification in data sets.
 (PI)
- 2021 **JHU Covid Bridge grant**, \$50,000 grant for one year. (PI)
- 2020 NSF-Simons Research Collaborations on the Mathematical and Scientific Foundations of Deep Learning, 5-year \$10M grant for large collaborative project (16 PIs at Johns Hopkins, Duke, Stanford, Berkeley and UPenn). Collaborative Research: Transferable, Hierarchical, Expressive, Optimal, Robust, Interpretable NETworks (THEORINET). (co-PI)
- 2019 NSF-DMS Computational Mathematics, \$58k grant for project Optimization techniques for geometrizing real-world data. (PI)
- 2019 Rising star in Computational and Data Sciences, University of Texas, Austin, USA.

- 2018 European Office of Aerospace Research and Development (EOARD), \$50k grant for project Error quantification and complexity limits in deep learning.
 Co-PI with Augustin Cosse (then Postdoc at Ecole Normale Superieure, France)
 - Co-r1 with Augustin Cosse (then rostdoc at Ecole Normale Superiore, rrance)
- 2017 **Speaker at UT Commencement Ceremony**, presented remarks representing the graduating class of PhD students at University of Texas at Austin.
- 2014 Frank Gerth III Graduate Excellence Award, Department of Mathematics, University of Texas at Austin.

2012-2013 Fulbright Fellow.

Publications

 \dagger Equal contribution. $\hfill *$ Student coauthor .

Journal publications

Dimensionless machine learning: Imposing exact units equivariance, S. Villar, W. Yao^{*}, D. W. Hogg, B. Blum-Smith, B. Dumitrascu, Journal of Machine Learning Research (JMLR), to appear.

Dimensionality reduction, regularization, and generalization in overparameterized regressions, *N. Huang*^{*}, *D. W. Hogg and S. Villar*, SIAM Journal on Mathematics of Data Science (SIMODS) 4 (1), 126-152, 2022.

Optimal marker gene selection for cell type discrimination in single cell analyses, *B. Dumitrascu*[†], *S. Villar*[†], *D. G. Mixon and B. E. Engelhardt*, Nature Communications 12 (1): 1-8. 2021.

Fitting very flexible models: Linear regression with large numbers of parameters, D. W. Hogg and S. Villar, Publications of the Astronomical Society of the Pacific, 133:093001 (18pp), 2021.

Seabed classification using physics-based modeling and machine learning, C. Frederick, S. Villar, Z. H. Michalopoulou, The Journal of the Acoustical Society of America 148, 859, 2020.

SqueezeFit: label-aware dimensionality reduction, C. McWhirter^{*}, D. G. Mixon and S. Villar, IEEE Transactions on Information Theory 66 (6), 3878-3892, 2019.

Fair redistricting is hard, R. Kueng, D. G. Mixon and S. Villar, Theoretical Computer Science 791, 28-35, 2019.

Clustering subgaussian mixtures by semidefinite programming, D. G. Mixon, S. Villar and R. Ward, Information and Inference: A Journal of the IMA, 6(4):389–415, 2017.

Probably certifiably correct k-means clustering, *T. Iguchi*, *D. G. Mixon*, *J. Peterson and S. Villar*, Mathematical Programming, 165(2):605–642, 2017.

Computer science and machine learning conferences

Scalars are universal: Equivariant machine learning, structured like classical physics, S. Villar, D. W. Hogg, K. Storey-Fisher^{*}, W. Yao^{*}, B. Blum-Smith, Advances in Neural Information Processing Systems (NeurIPS) 34, 28848-28863, 2021.

Can graph neural networks count substructures?, Z. Chen^{*}, L. Chen^{*}, S. Villar and J. Bruna, Advances in Neural Information Processing Systems (NeurIPS), 10383–10395, vol 33. 2020.

On the equivalence between graph isomorphism testing and function approximation with GNNs, Z. Chen^{*}, S. Villar, L. Chen^{*} and J. Bruna, Advances in Neural Information Processing Systems (NeurIPS), 15894-15902, 2019.

Relax, no need to round: Integrality of clustering formulations, *P. Awasthi, A. S. Bandeira, M. Charikar, R. Krishnaswamy, S. Villar and R. Ward*, In Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science, pages 191–200. ACM, 2015.

Conference proceedings and workshops (peer-reviewed)

From local to global: Spectral-inspired graph neural networks., *N. Huang*^{*}, *S. Villar, C.E. Priebe, D. Zheng, C. Huang, L. Yang, V. Braverman*, NeurIPS 2022 Workshop: New Frontiers in Graph Learning. 2022.

SE(3)-equivariant self-attention via invariant features., N. Chen^{*} and S. Villar, NeurIPS workshop Machine learning for the Physical Sciences, 2022.

A simple equivariant machine learning method for dynamics based on scalars, W. Yao^{*}, K. Storey-Fisher^{*}, W. Hogg, S. Villar, NeurIPS workshop Machine learning for the Physical Sciences, 2021.

A short tutorial on the Weisfeiler-Lehman test and its variants, *N. Huang*^{*}, *S. Villar*, 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021, pp. 8533-8537.

Efficient belief propagation for graph matching, E. Onaran^{*}, S. Villar, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 9060-9064, 2020.

Experimental performance of graph neural networks on random instances of max-cut, W. Yao^{*}, A. S. Bandeira and S. Villar, Wavelets and Sparsity XVIII, vol. 11138, p. 111380S. International Society for Optics and Photonics, 2019.

Revised note on learning algorithms for quadratic assignment with graph neural networks, A. Nowak^{*†}, S. Villar[†], A. S. Bandeira and J. Bruna, In 2018 IEEE Data Science Workshop, DSW 2018, Lausanne, Switzerland,: 229-233, 2018.

Projected power iteration for network alignment, *E. Onaran and S. Villar*, In Wavelets and Sparsity XVII, volume 10394, pages 103941C. International Society for Optics and Photonics, 2017.

Manifold optimization for k-means clustering, T. Carson, D. Mixon, S. Villar and R. Ward, In 2017 International Conference on Sampling Theory and Applications (SampTA), pages 73–77. IEEE, 2017.

Book chapters (peer-reviewed)

Three proofs of the Benedetto-Fickus theorem, *D. G Mixon*, *T. Needham*, *C. Shonkwiler and S. Villar*, To appear as a chapter in: Sampling, Approximation, and Signal Analysis, 2023.

Other (peer-reviewed)

Cibercoloquio latinoamericano de matemáticas, D. Campos, M. Rivera, M. A. Salazar, J. A. Samper, J. Simental, S. Villar, Notices of the American Mathematical Society, 68(5):793-797, 2021.

Preprints and working papers

The passive symmetries of machine learning, S. Villar[†], D.W. Hogg[†], W. Yao^{*}, G.A. Kevrekidis^{*} and B. Schölkopf, arXiv:2301.13724.

Sketch-and-solve approaches to k-means clustering by semidefinite programming, C. Clum^{*}, D.G. Mixon, S. Villar and K. Xie^{*}, arXiv:2211.15744.

Graph Neural Networks for Community Detection on Sparse Graphs, L. Ruiz, N. Huang^{*}, S. Villar, arXiv:2211.03231.

Deep Learning is Provably Robust to Symmetric Label Noise, C.E. Priebe, N Huang^{*}, S. Villar, C. Mu, L. Chen, arXiv:2210.15083.

Shuffled linear regression through graduated convex relaxation, E. Onaran, S. Villar, arXiv:2209.15608.

Equivariant maps from invariant functions., B. Blum-Smith, S. Villar, arXiv:2209.14991.

MarkerMap: nonlinear marker selection for single-cell studies., N. Sarwar^{*}, W. Gregory^{*}, G.A. Kevrekidis^{*}, S. Villar^{*†}, B. Dumitrascu^{*†}, arXiv:2207.14106.

MREC: a fast and versatile framework for aligning and matching point clouds with applications to single cell molecular data, A. J. Blumberg, M. Carriere, M. A. Mandell, R. Rabadan and S. Villar, arXiv:2001.01666.

Utility Ghost: Gamified redistricting with partian symmetry, D. G. Mixon and S. Villar, arXiv:1812.07377.

SUNLayer: Stable denoising with generative networks, D. G. Mixon and S. Villar, arXiv:1803.09319.

Monte Carlo approximation certificates for k-means clustering, D. G. Mixon, S. Villar, arXiv:1710.00956.

A polynomial-time relaxation of the Gromov-Hausdorff distance, S. Villar, A. S. Bandeira, A. J. Blumberg and R. Ward, arXiv:1610.05214.

Student mentees and group members

Ben-Blum Smith, Postdoctoral Fellow (JHU).

Teresa Huang, Third year PhD student in Applied Mathematics and Statistics (JHU), (co-advised with Carey Priebe).

George Kevrekidis, Second year PhD student in Applied Mathematics and Statistics (JHU).

Wilson Gregory, Second year PhD student in Applied Mathematics and Statistics (JHU).

Evan Mata, Master student in Applied Mathematics and Statistics (JHU).

Weichi Yao, Fifth year PhD student in Statistics (NYU), (currently collaborating in GNN projects). Santiago Robaina, Undergraduate student in Mathematics (Universidad de la República, Uruguay), (co-advised with Gabriel Illanes).

Professional service

Educational activities, conferences and seminar organization

- Mar 2023 Khipu, Latin American Meeting In Artificial Intelligence, Co-organizer, https://khipu.ai/.
 - 2022 Mathematics Research Communities (MRC), Program: Data Science at the crossroads of analysis, geometry, and topology, Co-organizer.
 - 2022 Whiting Internships in Science & Engineering (WISE) Program, Hosted a student from a Baltimore Public School to work on a computational redistricting problem (using MCMC techniques to assess whether the 2022 Maryland Congressional map is gerrymandered).
- Dec 2021 Out of distribution generalization Workshop at NeurIPS 2021, Co-organizer.
- May 2021 Geometrical and Topological Representation Learning Workshop at ICLR 2021, Coorganizer, https://gt-rl.github.io/.
- Mar 2021 AMS Spring Southeastern Sectional Meeting, Co-organizer of session: Graphs in Data Science.
- 2021-now One world MINDS seminar, Co-organizer, https://sites.google.com/view/minds-seminar/.
- 2021-now Mathematics and Democracy Institute at Wellesley College, Affiliated scholar, https://mathematics-democracy-institute.org/.
- 2020-now DeepMath Conference, Co-organizer, https://deepmath-conference.com/.
- 2020-now Cibercoloquio Latinoamericano de Matemáticas, Co-organizer of a weekly virtual math colloquium in Spanish directed to the global Spanish-speaking mathematical community, http: //www.cibercoloquio.com/.
- 2020-now Mathematical and Scientific Machine Learning (MSML), Program chair, https://msml21.github.io/.
- Mar 2020 Centre International de Rencontres Mathématiques, Co-organizer of program: Optimization for Machine Learning, Luminy, France.
- Jan 2019 Joint Math Meetings, Co-organizer of session: Low Complexity Models in Data Analysis and Machine Learning, Baltimore, USA.

JHU service

- 2022-now AI2AI Amazon-JHU initiative, Advisory board member.
- 2022 MINDS director search committee member.
- May 2022 Lead a workshop on getting NSF funding directed to AMS postdocs.
- 2022 AMS faculty search committee member.
- 2021-now Mentor of 25 students between AMS undergraduates, MSE and data science.
- 2021-now MINDS/CIS Seminar organizer.
- 2021-now AMS Department Seminar organizer.
- 2021 MINDS faculty search committee member and diversity advocate.
- Jan 2021 TRIPODS Winter School & Workshop on Graph Learning and Deep Learning, Coorganizer.
- Oct 2020 MINDS IDES Symposium, Co-organizer.
- Sep 2020 JHU Fulbright panel member.

Reviewer

- NSF CISE panels (2022).
- NSF DMS panels (2020, 2021 and 2022).
- SIAM Journal on Mathematics of Data Science (SIMODS).
- SIAM Journal on Optimization (SIOPT).
- SIAM Journal on Applied Algebra and Geometry (SIAGA).
- SIAM Journal on Scientific Computing (SISC).
- IEEE Transactions on Information Theory.
- IEEE Transactions on Signal Processing.
- Statistics and Public Policy Journal.
- Computational Learning Theory Conference (COLT).
- Information Theory Workshop (ITW).
- ACM Symposium on the Theory of Computing (STOC).
- Conference on Neural Information Processing Systems (NeurIPS).
- International Conference on Machine Learning (ICML).
- Journal on Machine Learning research (JLMR).

Teaching

Instructor

2020-now Johns Hopkins University.

- Equivariant machine learning
- Topics on trustworthy machine learning
- Non-linear optimization II
- Introduction to convexity
- 2019 NYU Center for Data Science.
 - Inference and representation
- 2012 Universidad de la República, Engineering School, Uruguay.
 o Calculus I
- 2011 Universidad de la Católica del Uruguay, Electrical Engineering, Uruguay.
 o Linear Algebra and Discrete Mathematics

Teaching assistant

2012-2017 University of Texas at Austin, Department of Mathematics.

- Differential equations and linear algebra.
- From numbers to chaos.
- Introduction to mathematics.
- Calculus of complex variables.
- Integral calculus.
- Differential calculus.
- Differential equations and linear algebra.

2008-2012 Universidad de la República, Department of Mathematics, Uruguay.

- Mathematics for life sciences.
- Linear algebra for Mathematics majors.
- Introduction to programming in Haskell.
- Introduction to programming in Python.
- General topology.

Selected talks and presentations

- Feb 2023 Data-driven physical simulation webinar at Lawrence Livermore National Laboratory, virtual talk, Livermore, USA.
- Jan 2023 Plenary speaker at LACIAM, Rio de Janeiro, Brazil.
- Dec 2022 Keynote talk at Learning on Graphs Conference, virtual talk.

Keynote talk at LatinX in AI workshop at NeurIPS, virtual talk. IMA Seminar at University of Minnesota, MN, USA.

NYU Center for Data Science Math and Data Seminar, NY, USA.

- Nov 2022 Seminar at Imperial College, virtual talk, London, UK.
 MATH + X Symposium on Matter under Extreme Conditions in Solar System Giant Planets and Exoplanets, Inverse Problems and Deep Learning, Las Catalinas, Costa Rica.
 Colóquio de Matemática Aplicada (IM - UFRJ), virtual talk, Rio de Janeiro, Brazil.
- Oct 2022 Mathematical Advances for Multi-Dimensional Microscopy, virtual talk, IPAM, UCLA, USA. MGGG Redistricting seminar at Tufts, virtual talk, Tufts, USA.
- Sep 2022 Mathematical and Scientific Foundations of Deep Learning Annual Meeting, virtual talk, Simons Foundation, NY, USA.

Dagstuhl Workshop: Machine Learning for Science: Bridging Data-Driven and Mechanistic Modeling, Warden, Germany.

Center for Mathematics and Artificial Intelligence Colloquium at George Mason University, Fairfax, VA, USA.

Jul 2022 Keynote talk at ICML Workshop: Topology, Algebra, and Geometry in Machine Learning, Baltimore, USA.

Keynote talk at ICML Workshop: Machine Learning for Astrophysics, Baltimore, USA.

BIRS workshop, Deep Exploration of non-Euclidean Data with Geometric and Topological Representation Learning, Kelowna, Canada.

LatinX in the Mathematical Sciences, IPAM workshop, University of California Los Angeles, USA.

June 2022 Institut de recherche en mathématique et physique seminar, UCLouvain, Belgium. Friedrich-Alexander-Universität Erlangen-Nürnberg, virtual research talk.

Lecturer at Summer School in Machine Learning Theory, Princeton University, USA.

Prospects and Challenges of Machine Learning for the Physical Sciences Conference, Flatiron Institute, New York, USA.

May 2022 Data Science, Approximation Theory, and Harmonic Analysis workshop, Fields Institute, Toronto, Canada.

Fast Faraway Talks (University of Maryland, College Park), virtual research seminar.

Apr 2022 CRM Applied Math Seminar, Montreal, Canada.

Mathematics in Imaging, Data and Optimization at Rensselaer Polytechnic Institute, virtual research seminar.

Panel moderator at ICLR Workshop on Geometric and Topological Representation Learning, virtual panelist.

Mar 2022 University of Chicago IMSI Workshop: The Mathematics of Soft Matter, virtual research talk.

AMS Spring Sectional at Tufts, virtual research talk.

Data-oriented Mathematical & Statistical Sciences Seminar at Arizona State University, virtual research talk.

Codex Seminar, virtual research talk.

Statistics Seminar at George Mason University, Fairfax, VA, USA.

AI + Math Colloquia at Institute of Natural Sciences, Shanghai Jiao Tong University, virtual research seminar.

Dec 2021 Simons Institute Optimization Under Symmetry Workshop, virtual research talk.

Oberwolfach workshop: Applied Harmonic Analysis and Data Science, virtual research talk.

Lisbon webinar Mathematics, Physics and Machine Learning, virtual research talk.

Mathematical Foundations of Machine Learning at the 2021 Canadian Mathematical Society Winter Meeting, virtual research talk.

- Oct 2021 Princeton Day of Statistics, Princeton, USA. BIRS-CMO Workshop: Geometry & Learning from Data, virtual research talk. The Ohio State Mathematics Colloquium, Columbus, USA.
- Sep 2021 University of Florida data science seminar, virtual research talk. University of Houston Data-Enabled Science Seminar, virtual research talk.
- Jun 2021 Fields Institute workshop: Low-Rank Models and Applications, virtual research talk. ICTP workshop: Youth in high dimensions, virtual research talk.
- May 2021 ICLR 2021 LatinX in AI workshop, Invited keynote speaker.
- Mar 2021 University of Maryland College Park Statistics Seminar, virtual seminar.
- Feb 2021 UCLA Statistics Seminar, virtual seminar.
- Jan 2021 UCLA Applied Mathematics Seminar, virtual seminar.
- Dec 2020 NeurIPS Women in Machine learning mentoring session, participant in virtual panel. CERN String data workshop, virtual research talk.

ASA Acoustics Virtually Everywhere, virtual research talk.

Nov 2020 **RWTH Aachen University**, virtual seminar. **Duke Probability Seminar**, virtual seminar. **INFORMS 2020 annual meeting**, virtual research talk.

- Oct 2020 C3.ai DTI Workshop on The Analytical Foundations of Deep Learning, virtual research talk.
- Aug 2020 One World ML, virtual research talk.
- Apr 2020 ICERM Workshop on Computational Statistics and Data-Driven Models, virtual research talk.
- Jan 2020 Joint Math Meetings, Denver, USA.
- Dec 2019 Machine Learning Tools for Research in Astronomy, Ringberg, Germany.
- Nov 2019 Using Physical Insights for Machine Learning, University of California, Los Angeles, USA.
- Oct 2019 Computational Harmonic Analysis and Data Science, Oaxaca, Mexico.
- Aug 2019 Microsoft Research AI Institute workshop Geometry of Deep Learning, Redmond, USA.
- May 2019 Oberwolfach workshop Statistical and computational aspects of learning with complex structure, Oberwolfach, Germany.

TU Berlin Theoretical Computer Science seminar, Berlin, Germany.

- Apr 2019 Rising stars in Computational and Data Sciences, University of Texas, Austin, USA.
- Mar 2019 Halıcıoğlu Data Science Institute Seminar, University of California, San Diego, USA.
 Data Institute SF Annual Conference, University of San Francisco, San Francisco, USA.
 Mathematical Institute for Data Science Seminar, Johns Hopkins University, Baltimore, USA.
- Jan 2019 NJIT Applied Mathematics and Statistics Seminar, New Jersey Institute of Technology, Newark, USA.

Joint Math Meetings, Session organizer: Low Complexity Models in Data Analysis and Machine Learning, Baltimore, USA.

Math + X Symposium on inverse problems and deep learning in space exploration, Rice University, Houston, USA.

- Nov 2018 Sublinear Algorithms and Nearest-Neighbor Search, Simons Institute for the Theory of Computing, Berkeley, USA.
- Oct 2018 Young Researchers Workshop: Kinetic descriptions in theory and applications, University of Maryland, College Park, USA.

Quantitative Redistricting, Duke University, Durham, USA.

- Sep 2018 NJIT Mechanical Engineering Colloquium, New Jersey Institute of Technology, Newark, USA.
- Aug 2018 Statistical physics and machine learning back together, Cargese, France.
- Apr 2018 SILO Seminar, University of Wisconsin, Madison, USA.
 Norbert Wiener Center Seminar, University of Maryland, College Park, USA.
 DIMACS Theory Seminar, Rutgers University, New Brunswick, USA.

Tèlecom ParisTech, Paris, France.

INRIA, Paris, France.

Mar 2018 **Oberwolfach workshop Applied Harmonic Analysis and Data Processing**, Oberwolfach, Germany.

NYU Mathematics Colloquium, New York University, USA.

- Feb 2018 Center for Data Science lunch seminar, New York University, USA.
- Jan 2018 Microsoft Research, Redmond, Washington, USA.
- Dec 2017 Uruguayan Colloquium of Mathematics, Universidad de la República, Uruguay.

Young Researchers Workshop: new trends in Computational and Applied Mathematics, Peking University, Beijing, China.

Minisymposium on Spectral Graph Theory and Optimization, University of California, Berkeley, USA.

- Nov 2017 Mathematical Data Science Seminar, Department of Mathematics, University of Tennessee, Knoxville, USA.
- Sep 2017 **Topology, Geometry and Data Seminar**, Department of Mathematics, Ohio State University, Columbus, Ohio, USA.

Machine Learning Lunch Seminar, Electrical and Computer Engineering Department, Ohio State University, Columbus, Ohio, USA.

IMA Data Science Seminar, Institute of Mathematics and its Applications, University of Minnesota Twin Cities, Minneapolis, USA.

- Ago 2017 SPIE Wavelets and Sparsity, San Diego, USA.
- Jul 2017 Foundation of Computational Mathematics, Barcelona, Spain.

Approximation Theory and Function Spaces Worshop, Centre de Recerca Matematica, Barcelona, Spain.

- Oct 2016 Applied Harmonic Analysis, Massive Data Sets, Machine Learning, and Signal Processing, Oaxaca, Mexico.
- Sep 2016 Information Theory Workshop, Cambridge University, UK.
- May 2016 SIAM Imaging, Albuquerque, New Mexico, USA.
- Apr 2016 MIT Applied Mathematics Seminar, MIT, Cambridge, USA.
- Mar 2015 AMS Sectional Meeting, Michigan State University, East Lansing, USA.
- Sep 2014 IDeAS seminar, Princeton University, Princeton, USA.