Yuan Yao

Current Address: Department of Mathematics, Stanford University, Stanford, CA 94305-2125.
Homepage: http://math.stanford.edu/~yuany, Email: yuany@stanford.edu, Phone: (510) 390 5030

LANGUAGES Native Chinese, Fluent English, Mathematics reading in French and German

CITIZENSHIP P. R. China

EDUCATION

Department of Mathematics, University of California at Berkeley
Ph.D. in Mathematics, December 2006
Dissertation Title: A Dynamical Theory for Learning
Advisor: Professor Steve Smale

Department of Mathematics, City University of Hong Kong
M.Phil. in Mathematics, June 2002
Thesis Title: On Some Problems in the Mathematical Foundation of Learning
Advisor: Professor Steve Smale and Professor Felipe Cucker

Department of Control Engineering, Harbin Institute of Technology
M.Sc. in Engineering, June 1998
Thesis Title: Linear Matrix Inequalities in Robust $\mathscr{H}_\infty$ Control
Advisor: Professor Guangxiong Wang

Department of Control Engineering, Harbin Institute of Technology
B.Sc. in Engineering, June 1996
Thesis Title: Finite Element Modeling in Control Systems
Advisor: Professor Guangxiong Wang and Dr. Jing Luo

RESEARCH INTERESTS

Topological and geometric methods in high dimensional data analysis
Online learning and stochastic optimization
Computational molecular biology and protein folding
Statistical machine leaning, *esp.* mathematical methods
Computer vision and pattern recognition
Convex optimization, System and control

RESEARCH EXPERIENCE

Department of Mathematics *Postdoctoral Fellow*
Stanford University **August 2006-current**
Working on topological and geometric methods for high dimensional data analysis. In partiluar, it includes the development of topological methods in exploring biomolecular folding pathways, which in the first time provides structural evidence from computer simulations that RNA hairpin folding has multiple pathways in support of a hypothesis from biological experiments; the development of combinatorial Hodge theoretic approach in rank learning or preference aggregation problems, which provides a novel framework to cope with modern ranking data which are cardinal, imbalanced and incomplete, beyond the classical social choice theory; and the application of metric learning based on semi-definite programming for selecting algebraic invariants in the topological reconstruction of phylogenetric trees.

Department of Mathematics, U.C. Berkeley *Graduate Student Researcher*
& Toyota Technological Institute at Chicago **August 2002-July 2006**
Studied online learning algorithms as stochastic optimizations in reproducing kernel Hilbert spaces. With S. Smale, proposed an online learning algorithm as stochastic gradient method for a regularized least square problem and showed some probabilistic upper bounds of convergence by introducing novel concentration inequalities in Hilbert spaces from Russian school.

Then improved these bounds to the same rates as batch learning in some settings, exploiting techniques of martingale and reverse martingale decomposition. Currently with P. Tarres, improving such results to reach minimax and individual optimality by following the entire regularization path. Furthermore investigated early stopping regularization, with L. Rosasco and A. Caponnetto, which is connected to Boosting in a gradient descent view and Landwebter iterations in classical inverse problems.

Department of Mathematics, City University of Hong Kong *Research Assistant*
Hong Kong SAR, P.R. China **June 1999 – June 2002**
Guided by S. Smale and F. Cucker, approached learning theory via Vapnik's framework and approximation theory. Also with M. Pontil, studied the applications of support vector machines for multi-category fingerprint classification problems, which gives some competitive or best results on NIST database 4. In the master thesis work, studied the application of Kullback-Sanov Theorem in learning, which leads to the well-known Hoeffding and Bernstein inequalities and some complexity upper bounds on averaging performance for mixed strategies in learning, in a spirit of Game theory and PAC-Bayesian learning developed by D. McAllester.

Department of Control Engineering, Harbin Institute of Technology *Research Assistant*
Harbin, P.R. China **June 1999 – June 2002**
Two aspects from control theory and computer vision in this period: 1. studied the applications of linear matrix inequalities in robust control, developed by S. Boyd *et al*. based on the interior-point method for semi-definite programming; 2. studied the image reconstruction via mathematical morphology implemented in Cellular Neural Networks, which are analog computing devices as universal as Turing machine.

PUBLICATIONS AND PREPRINTS

Book

1. Yao, Y., *A Dynamic Theory of Learning – Online Learning and Stochastic Algorithms in Reproducing Kernel Hilbert Spaces*, Verlag Dr. Müller, ISBN: 978-3-639-09390-2. 2008.

Recent Papers

1. Sun, J., Y. Yao, X. Huang, V. Pande, G. Carlsson, and L. Guibas, A Fast Geometric Clustering Method on Conformation Space of Biomolecules. *Journal of Physical Chemistry*, submitted.
2. Tarres, P. and Y. Yao. Online Learning as Stochastic Approximations of Regularization Paths. preprint. Short-report appeared in *Mathematisches Forschungsinstitut Oberwolfach, Report No.30/2008, Learning Theory and Approximation*.
3. Jiang, X., L.-H. Lim, Y. Ye and Y. Yao. Learning to Rank with Combinatorial Hodge Theory. *Mathematical Programming*, submitted.
4. Yao, Y., J. Sun, X. Huang, G. Bowman, V. Pande, L. Guibas and G. Carlsson, Topological Methods for Exploring Low-density States in Biomolecular Folding Pathways. *PLoS Computational Biology*, submitted.
5. Yao, Y. On Complexity Issue of Online Learning Algorithms. *IEEE Transactions on Information Theory*. Accepted.

6. Bowman, G., X. Huang, Y. Yao, J. Sun, G. Carlsson, L. Guibas and V. Pande, Structural insight into RNA hairpin folding intermediates. *Journal of American Chemistry Society*, 2008, 130 (30): 9676-8.
7. Eriksson, N. and Y. Yao. Metric Learning for Phylogenetic Invariants, 2007. *preprint*. `arxiv.org/abs/q-bio/0703034`.
8. Yao, Y., L. Rosasco and A. Caponnetto. On Early Stopping in Gradient Descent Learning. *Constructive Approximation*, Special Issue: Learning Theory, 2007, 26 (2): 289-315.
9. Minh, H.Q., P. Niyogi and Y. Yao. Mercer's Theorem, Feature Maps, and Smoothing. In: *Proc. of Computational Learning Theory (COLT)*, 2006.
10. Caponnetto, A. and Y. Yao. Adaption for Regularization Operators in Learning Theory, *CBCL Paper #265/AI Technical Report #063*, Massachusetts Institute of Technology, Cambridge, MA, September, 2006.
11. Smale, S. and Y. Yao. Online Learning Algorithms. *Foundation of Computational Mathematics*, 6(2): 145-170, 2006.

Papers on Computer Vision, Machine Learning and Pattern Recognition

1. Yao, Y., G. L. Marcialis, M. Pontil, P. Frasconi, and F. Roli. Combining Flat and Structured Representations for Fingerprint Classification with Recursive Neural Networks and Support Vector Machines. *Pattern Recognition*, 36(2): 397-406, 2003.
2. Yao, Y., G. L. Marcialis, M. Pontil, P. Frasconi, and F. Roli. A New Machine Learning Approach to Fingerprint Classification. *AI*IA 2001: Advances in Artificial Intelligence*, LNCS, vol.2175: pp. 57-63.
3. Yao, Y., M. Pontil and F. Roli. Fingerprint Classification with Combinations of Support Vector Machines. In: *Proceedings of Audio- and Video-Based Biometric Person Authentication, Third International Conference, AVBPA 2001*, Halmstad, Sweden.
4. Yao, Y., X.-F. Zhang, T.-W. Zhang and G.-X. Wang. Multiscale Morphology for Color Images Implemented by Fuzzy Cellular Neural Network. In: *Proceedings of IEEE Hong Kong Symposium on Robotics and Control*, July 1999, Hong Kong, pp. 459-462.
5. Yao, Y., G.-X. Wang and T.-W. Zhang. Morphological Reconstruction for Color Images Implemented by Fuzzy Cellular Neural Networks. *Chinese Journal of Computers* (in Chinese), 22(7): 727-732, 1999.
6. Yao, Y., G.-X. Wang and T.-W. Zhang. Application of Fuzzy Cellular Neural Networks to Stone Inscription Reconstruction in Chinese Calligraphy. *Journal of Computer Research and Development* (in Chinese), 36(3): 282-286, 1999.

Papers on System and Control

1. L.-F. Li, G.-X. Wang and Y. Yao. Global Optimal Robust Controller Design. *Journal of Control Theory and Applications* (in Chinese), 18(2): 266-269, 2001.
2. Yao, Y., L.-F. Li and G.-J. Bao. On the Application Problem of the Gap Metric for SISO Systems. *Journal of Harbin Institute of Technology* (in Chinese), 31(6): 19-21, 1999.
3. Yao, Y., J.-B. Wang, L.-F. Li and G.-X. Wang. Optimal Robust Performance in Constantly Scaled $\mathscr{H}_\infty$ control. *Korea-China Process System Engineering Workshop*. August 1999, Korea.

4. Yao, Y., L.-F. Li, G.-X. Wang and J.-B. Wang. Robust Gain-scheduled $\mathscr{H}_\infty$ control with Constant Diagonal Scaling. In: *Proceedings of IEEE Hong Kong Symposium on Robotics and Control*, July 1999, Hong Kong, pp. 628-632.
5. X.-F. Wang, Y. Yao, G.-X. Wang and J.-B. Wang. Application of Quadratic Stabilization, Constantly Scaled $\mathscr{H}_\infty$ control and μ-Synthesis. In: *Proceedings of IEEE Hong Kong Symposium on Robotics and Control*, July 1999, Hong Kong, pp. 633-637.
6. J.-B. Wang, J. Jiya, T.-Y. Chai, Y. Yao, G.-X. Wang and S.-J. Xu. Identifying Noise Model in Closed-Loop Using Subspace Method. In: *Proceedings of the IEEE International Vehicle Electronics Conference (IVEC '99)*, pp. 349-351, Changchun, China, September 6-9, 1999,
7. Yao, Y. and J. Luo. FEM-Based Modeling in Servo Design. *Electric Machine and Control* (in Chinese), 2(2): 108-111, 1998.

Poster Presentations

1. Yao, Y., J. Sun, X. Huang, V. Pande, L. Guibas and G. Carlsson, Topological Methods for Exploring Biomolecular Folding Pathways. *Biomedical Computation at Stanford (BCATS)*, spotlight poster presentation, Stanford, October 26, 2008.
2. Sun, J., X. Huang, Y. Yao, G. Carlsson, V. Pande and L. Guibas, A Well-controlled Fast Clustering Method on Conformation Space of Biomolecules. *Biomedical Computation at Stanford (BCATS)*, poster presentation, Stanford, October 26, 2008.
3. Bowman, G. R., X. Huang, Y. Yao, J. Sun and V. Pande, Adaptive Seeding: A New Method for Simulating Biologically Relevant Timescales. *Biomedical Computation at Stanford (BCATS)*, poster presentation, Stanford, October 26, 2008.
4. Scheler, G. and Y. Yao. *Equilibria in neuroadaptive pathways*. RECOMB Systems Biology, poster presentation, Dec 1-2, La Jolla, CA, 2007.
5. Yao, Y. and J.-Y. Ke. Exploring Semantic Complexity by Computational Learning Theory. *Evolutionary Computation and Cognitive Science Workshop (ECCS'2000)*, poster presentation, January, 2000, Melbourne, Australia.

Papers in preparation

1. *A Competitive Equilibrium Approach to Rank Aggregation*, with Yinyu Ye, et al.
2. *From PageRank to HodgeRank: a Perspective from Voting Theory and Algebraic Topology*, with Gunnar Carlsson, et al.
3. *Multiresolution Transition Networks for Biomolecular Folding*, with Xuhui Huang, et al.

TALKS

Topological Methods for Exploring Biomolecular Folding Pathways, spotlight presentation, Biomedical Computation at Stanford (BCATS), Stanford, October 26, 2008.

Combinatorial Hodge Theory and A Geometric Approach to Ranking, SIAM Annual Meeting, minisymposium: Mathematical Methods in Data Mining, San Diego, July 7-11, 2008.

Topological Methods for Exploring Low-density States in Biomolecular Folding Pathways, Modern Massive Data Sets (MMDS), Stanford, June 25-39, 2008.

Hodge Decomposition, Spectral Embedding, and the Netflix Dataset, Bay Area Scientific Computing Day: honoring Professors Kahan and Parlett, MSRI, Berkeley, March 29-30, 2008.

Hodge Theory and Rank Aggregation, Computational Applications of Algebraic Topology, I, AMS sectional meeting, Albuquerque, New Mexico, October 13, 2007.

Combinatorial Hodge Theory and Societal Rank Aggregation, Chern Institute of Mathematics, Nankai Univeristy, Tianjin, China, August 20-25, 2007.

Combinatorial Laplacians and Rank Aggregation, the 6^{th} International Congress of Industrial and Applied Mathematics (ICIAM), minisymposium: Novel Matrix Methods for Internet Data Mining. Zurich, Switzerland, July 16-20, 2007.

Topology Learning of High Dimensional Probability Density Functions, the 1^{st} International Congress of IPIA, Conference on Applied Inverse Problems, minisymposium: Learning from Examples as an Inverse Problem. PIMS, University of British Columbia, Vancouver, Canada, June 25-29, 2007.

Metric Learning for Phylogenetic Invariants, Student Algebraic Statistics Seminar, University of California at Berkeley, Feb. 5, 2007.

MEMBERSHIPS

American Mathematical Society (AMS)
Institute of Mathematical Statistics (IMS)
Society for Industrial and Applied Mathematics (SIAM)

JOURNAL AND CONFERENCE REFEREE

Foundations of Computational Mathematics
IEEE Transactions on Information Theory
Journal of Approximation Theory
Journal of Machine Learning Research
Neural Computation
NIPS'2003, 2005

TEACHING EXPERIENCE

Department of Mathematics *Graduate Student Instructor*
University of California at Berkeley August 2002 - December 2003

Fall 2002	Math 1A	Calculus
Spring 2003	Math 54	Linear Algebra and Differential Equations
Fall 2003	Math 54	Linear Algebra and Differential Equations

Department of Mathematics *Teaching Assistant*
City University of Hong Kong 2000

Spring 2000 Dynamical Systems Lecturer: Prof. Steve Smale and Prof. Lan Wen

SPECIAL SKILLS

Programming ability with Matlab, C^{++}, R and Python.
Playing music instruments such as Chinese long flute (Xiao) and Classical Guitar
Chinese Calligraphy

REFERENCES

1. *Professor Gunnar Carlsson*, Department of Mathematics, Stanford University, Stanford, CA 94305-2125. Email: gunnar@math.stanford.edu. Office phone: (650) 723-2224.
2. *Professor Leonidas Guibas*, Department of Computer Science and Electrical Engineering, Stanford University, Stanford, CA 94305. Email: guibas@cs.stanford.edu. Office phone: (650) 723-0304.
3. *Professor Vijay Pande*, Department of Chemistry and Structural Biology, Stanford University, Stanford, CA 94305. Email: pande@stanford.edu.
4. *Professor Steve Smale*, Toyota Technological Institute at Chicago, 1427 East 60^{th} Street, Chicago, IL 60637. Department of Mathematics, University of California, Berkeley, CA 94720. Email: smale@math.berkeley.edu. Office phone: (773) 834-2510 (TTI-C) and (510) 642-4367 (UCB).
5. *Professor Wing-Hung Wong*, Department of Statistics, Stanford University, Stanford, CA 94305-4065. Email: whwong@stanford.edu. Office phone: (650) 725-2915.
6. *Professor Yinyu Ye*, Department of Management Science and Engineering and Electrical Engineering, Stanford University, Stanford, CA 94305. Email: yyye@stanford.edu. Office phone: (650) 723-7262.
7. [Teaching Reference] *Barbara Peavy*, Director of Student Services, Department of Mathematics, University of California at Berkeley, 967 Evans Hall, Berkeley, CA 94720-3840. Email: peavy@math.berkeley.edu. Office phone: (510) 642-2479.
8. [For soliciting reference letters above] *Rose Stauder*, Office Assistant, Department of Mathematics, Building 380, Sloan Hall, Stanford University, Stanford, CA 94305. Email: rose@math.stanford.edu. Office phone: (650) 725-6284.