Christos Boutsidis

Contact Information

Work address
Yahoo Labs
229 West 43rd Street
New York, NY 10036

Office phone (USA): +1 (212) 381-6872
E-mail: christos.boutsidis@gmail.com
www: www.boutsidis.org

Biographical Information

Date of Birth: August 1983
Nationality: Greek
Citizenship: Greek
US Visa: H1-B

Education

Rensselaer Polytechnic Institute, Troy, NY — **Aug 2006 - May 2011**
Ph.D. in Computer Science
Dissertation: Topics in Matrix Sampling Algorithms
Advisor: Petros Drineas

University of Patras, GREECE — **Oct 2001 - July 2006**
BS in Computer Engineering
Thesis: SVD-based initialization: A head start
on nonnegative matrix factorization
Thesis Advisor: Efstratios Gallopoulos

Expertise

Machine learning; data analysis; numerical linear algebra; matrix theory. Specifically, approximation algorithms for linear algebraic and machine learning problems. Using "sketching" techniques to solve such problems in the streaming, online, and distributed model of computation.

Professional Experience

Yahoo Labs, New York — **March 2014 - present**
Research Scientist

(i) Design and implementation (Matlab, JAVA) of a recommendation system for Yahoo videos. (ii) Online/Distributed implementations (JAVA) of linear algebra algorithms, e.g., Principal Component Analysis, QR factorization, etc; (iii) Research on distributed matrix algorithms.

IBM T.J. Watson Research Center — **Aug 2011 - March 2014**
Business Analytics and Mathematical Sciences
Research Staff Member

(i) Design and implementation (SPSS) of machine learning algorithms for Workforce Optimization, e.g., retention modeling and prediction. (ii) Government (DARPA) sponsored project on large scale implementations (Python, C++) of Randomized Numerical Linear Algebra methods. (iii) Research on matrix algorithms and their applications to machine learning.

WorldQuant — **June 2010 - Aug 2010**
Hedge Fund located in Old Greenwich, CT
Summer intern

Worked under the supervision of a portfolio manager on designing and implementing (Matlab) alphas for modeling future contracts.

IBM Zurich Research Laboratory — **May 2009 - Oct 2009**
Mathematical and Computational Sciences
Summer intern

Designed and implemented (JAVA) algorithms for a recommendation system that was being developed for the sales department of IBM. Researched on co-clustering algorithms.

University of California, Los Angeles (UCLA) **Sep 2008 - Dec 2008**
Institute for Pure and Applied Mathematics (IPAM)
Research fellow

Participated in 3-month length Program on "Internet Multi-Resolution Analysis: Foundations, Applications and Practice".

IBM T.J. Watson Research Center **May 2008 - Aug 2008**
Service Engineering Department
Summer intern

Designed and implemented (Matlab) algorithms for clustering and classification of large scale data arising from Information Technology (IT) service providers.

Honors

1. "Near-Optimal Column-Based Matrix Reconstruction", which appeared in the 52nd IEEE Symposium on Foundations of Computer Science (FOCS), was invited to the special issue of the SIAM Journal on Computing for the top papers from FOCS 2011.
2. Awarded the 2011 "Robert McNaughton Prize", given to an outstanding student in the computer science department of Rensselaer Polytechnic Institute, yearly.

Representative Publications

1. Online Principal Components Analysis
   Christos Boutsidis, David Garber, Zohar Karnin, Edo Liberty
   ACM-SIAM Symposium on Discrete Algorithms (**SODA**)
   San Diego, California, USA, January 4-6, 2015.
2. Optimal CUR Matrix Decompositions
   Christos Boutsidis and David Woodruff
   ACM Symposium on Theory of Computing (**STOC**)
   New York, New York, May 31-Jun 3, 2014.
3. Efficient Dimensionality Reduction for Canonical Correlation Analysis
   Haim Avron, Christos Boutsidis, Sivan Toledo, and Anastasios Zouzias
   International Conference on Machine Learning (**ICML**)
   Atlanta, Georgia, USA, June 17-19, 2013
4. Near-Optimal Column-Based Matrix Reconstruction
   Christos Boutsidis, Petros Drineas, and Malik Magdon-Ismail
   Annual IEEE Symposium on Foundations of Computer Science (**FOCS**)
   Palm Springs, California, USA, October 23-25, 2011.
5. Random Projections for k-means Clustering
   Christos Boutsidis, Anastasios Zouzias, and Petros Drineas
   Annual Conference on Neural Information Processing Systems (**NIPS**)
   Vancouver, B.C., Canada, December 6-9, 2010.
6. Unsupervised Feature Selection for Principal Components Analysis
   Christos Boutsidis, Michael W. Mahoney, and Petros Drineas
   ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (**KDD**)
   Las Vegas, Nevada, USA, August 24-27, 2008.

Patents

1. *Method and Apparatus for optimally finding a CUR decomposition*,
   with David P. Woodruff (filed with IBM), 2014
2. *Matrix Reordering and Visualization Techniques in the Presence of Data Hierarchies*,
   with M. Vlachos and A. Labbi (filed with IBM), 2009

Computer skills

1. Java (Advanced)
2. Python (intermediate)
3. C/C++ (familiarity)
4. Matlab (advanced)

Publications full list (Journals)

(Accepted)

1. Near-Optimal Column-based Matrix Reconstruction
   Christos Boutsidis, Petros Drineas, and Malik Magdon-Ismail
   SIAM Journal on Computing, 43(2), 687-717, 2014
2. Efficient Dimensionality Reduction for Canonical Correlation Analysis
   Haim Avron, Christos Boutsidis, Sivan Toledo, and Anastasios Zouzias
   SIAM Journal on Scientific Computing (SISC), 36(5), 111-131, 2014
3. Faster Subset Selection for Matrices and Applications
   Haim Avron and Christos Boutsidis
   SIAM Journal on Matrix Analysis and Applications, 34(4), 1464-1499, 2013
4. Improved matrix algorithms via the Subsampled Randomized Hadamard Transform
   Christos Boutsidis and Alex Gittens
   SIAM Journal on Matrix Analysis and Applications, 34(2), 1301-1340, 2013
5. Near-optimal Coresets For Least-Squares Regression
   Christos Boutsidis, Petros Drineas, and Malik Magdon-Ismail
   IEEE Transactions on Information Theory, vol. 59, no. 10, pp. 6880-6892, Oct. 2013
6. Deterministic Feature Selection for K-means Clustering
   Christos Boutsidis and M. Magdon-Ismail
   IEEE Transactions on Information Theory, vol. 59, no. 9, pp. 6099-6110, Sept. 2013
7. Randomized Dimensionality Reduction for K-means Clustering
   Christos Boutsidis, Anastasios Zouzias, Michael W. Mahoney, and Petros Drineas
   IEEE Transactions on Information Theory, accepted, Nov 2014.
8. A note on sparse least-squares regression
   Christos Boutsidis and Malik Magdon-Ismail.
   Information Processing Letters 114 (5), 273-276, 2014.
9. Random Projections for Linear Support Vector Machines
   Saraubh Paul, Christos Boutsidis, Malik Magdon-Ismail, and Petros Drineas
   ACM Transactions on Knowledge Discovery from Data (TKDD), 8(4):22, 2014.
10. Spectral Clustering: An empirical study of Approximation Algorithms and its Application to the Attrition Problem
    B. Cung, T. Jin, J. Ramirez, A. Thompson, C. Boutsidis, and D. Needell
    SIAM Undergraduate Research Online, Volume 5, Dec. 2012
11. Atomic-level characterization of the ensemble of the Ab(1-42) monomer in water using unbiased molecular dynamics simulations and spectral algorithms
    N. Sgourakis, M. Serrano, C. Boutsidis, P. Drineas, Z. Du, C. Wang, and A. Garcia
    Journal of Molecular Biology, 405(2):570-83, 2011.
12. Random Projections for the Nonnegative Least Squares Problem
    Christos Boutsidis and Petros Drineas
    Linear Algebra and its Applications, Volume 431, Issues 5-7, 1 August 2009, pages 760-771.
13. SVD-based initialization: A head start on nonnegative matrix factorization
    Christos Boutsidis and Efstratios Gallopoulos
    Pattern Recognition, Volume 41, Issue 4, April 2008, pages 1350-1362.

(Submitted)

14. Optimal CUR Matrix Decompositions
    Christos Boutsidis and David Woodruff
    SIAM Journal on Computing, submitted, 2014

Publications full list (Conferences)

(Accepted)

15. Online Principal Components Analysis
Christos Boutsidis, David Garber, Zohar Karnin, Edo Liberty
ACM-SIAM Symposium on Discrete Algorithms (**SODA**)
San Diego, California, USA, January 4-6, 2015.

16. Provable Deterministic Leverage Scores Sampling
Dimitris Papailiopoulos, Anastasios Kyrillidis, and Christos Boutsidis
ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (**KDD**)
New York, NY, USA, August 24-27, 2014.

17. Faster SVD-truncated Regularized Least-squares
Christos Boutsidis and Malik Magdon-Ismail
IEEE International Symposium on Information Theory (**ISIT**)
Honolulu, HI, USA, June 29 - July 4, 2014.

18. Optimal CUR Matrix Decompositions
Christos Boutsidis and David Woodruff
ACM Symposium on Theory of Computing (**STOC**)
New York, New York, May 31-Jun 3, 2014.

19. Efficient Dimensionality Reduction for Canonical Correlation Analysis
Haim Avron, Christos Boutsidis, Sivan Toledo, and Anastasios Zouzias
International Conference on Machine Learning (**ICML**)
Atlanta, Georgia, USA, June 17-19, 2013

20. Random Projections for Support Vector Machines
Saraubh Paul, Christos Boutsidis, Malik Magdon-Ismail, Petros Drineas
International Conference on Artificial Intelligence and Statistics (**AISTATS**)
Scottsdale, Arizona, USA, April 29-May 1, 2013

21. Sparse Features for PCA-like Linear Regression
Christos Boutsidis, Petros Drineas, and Malik Magdon-Ismail
Annual Conference on Neural Information Processing Systems (**NIPS**)
Granada, Spain, December 12-15, 2011

22. Near-Optimal Column-Based Matrix Reconstruction
Christos Boutsidis, Petros Drineas, and Malik Magdon-Ismail
Annual IEEE Symposium on Foundations of Computer Science (**FOCS**)
Palm Springs, California, USA, October 23-25, 2011.

23. Random Projections for k-means Clustering
Christos Boutsidis, Anastasios Zouzias, and Petros Drineas
Annual Conference on Neural Information Processing Systems (**NIPS**)
Vancouver, B.C., Canada, December 6-9, 2010.

24. Unsupervised Feature Selection for the k-means Clustering Problem
Christos Boutsidis, Michael W. Mahoney, and Petros Drineas
Annual Conference on Neural Information Processing Systems (**NIPS**)
Vancouver, B.C., Canada, December 7-10, 2009.

25. An Improved Approximation Algorithm for the Column Subset Selection Problem
Christos Boutsidis, Michael W. Mahoney, and P. Drineas
ACM-SIAM Symposium on Discrete Algorithms (**SODA**)
New York, New York, USA, January 4-6, 2009.

26. Clustered Subset Selection and its Applications on IT Service Metrics
Christos Boutsidis, Jimeng Sun, and Nikos Anerousis
ACM Conference on Information and Knowledge Management (**CIKM**)
Napa Valley, California, USA, October 26-30, 2008.

27. Unsupervised Feature Selection for Principal Components Analysis
Christos Boutsidis, Michael W. Mahoney, and Petros Drineas
ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (**KDD**)
Las Vegas, Nevada, USA, August 24-27, 2008.

(Submitted)

28. Spectral Clustering via the Power Method - Provably
Christos Boutsidis, Alex Gittens, and Anju Kambadur
International Conference on Machine Learning (**ICML**)
29. Optimal Sparse Linear Auto-encoders and Sparse PCA
Malik Magdon-Ismail and Christos Boutsidis
International Conference on Machine Learning (**ICML**)
30. A Randomized Algorithm for Approximating the Log
Determinant of a Symmetric Positive Definite Matrix
Christos Boutsidis, Petros Drineas, Anju Kambadur, and Anastasios Zouzias
ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (**KDD**)
31. Optimal Distributed Principal Component Analysis
Christos Boutsidis, Maxim Sviridenko, and David Woodruff
Annual IEEE Symposium on Foundations of Computer Science (**FOCS**)

Invited Talks

(Invited presentations only; contributed conference presentations are not included.)

1. Optimal CUR Matrix Decompositions
Information Theory and Applications (ITA)
San Diego, California, February 2015
2. Principal Component Analysis: offline, online, and distributed
University of Illinois at Urbana-Champaign, Computer Science Dept. Seminar
Urbana-Champaign, Illinois, USA, October 2014
3. Principal Component Analysis: offline, online, and distributed
Digital Technology Center - University of Minnesota.
Twin Cities, Minneapolis, USA, October 2014
4. Optimal CUR Matrix Decompositions
Householder Symposium
Spa, Belgium, June 2014
5. Optimal CUR Matrix Decompositions
SIAM Conference on Optimization
San Diego, California, USA, May 2014
6. Approximate Spectral Clustering via Randomized Sketching
Electrical Flows, Graph Laplacians, and Algorithms: Spectral Graph Theory and Beyond
Institute for Computational and Experimental Research in Mathematics (ICERM)
Providence, RI, USA, April 2014
7. Sampling Algorithms for Matrix Computations
Yahoo Labs
New York, New York, USA, Dec 2013
8. Randomized Dimensionality Reduction in Machine Learning
SIAM Annual Meeting
San Diego, California, USA, July 2013
9. Near-optimal Column-based Matrix Reconstruction
SIAM Annual Meeting
San Diego, California, USA, July 2013
10. Near-optimal Column-based Matrix Reconstruction
Workshop on Randomized Numerical Linear Algebra (RandNLA): Theory and Practice
Under the auspices of the Annual IEEE Symposium on Foundations of Computer Science
New Brunswick NJ, USA October 20, 2012
11. Near-optimal Column-based Matrix Reconstruction
SIAM Conference on Applied Linear Algebra
Valencia, Spain July 2012

12. Randomized Dimensionality Reduction in Machine Learning
    Mysore Workshop on Machine Learning
    Mysore, India August 2012
13. Data Analytics Solutions through Stochastic, Sampling-based Matrix Algorithms
    IBM T.J. Watson Research Lab
    Yorktown Heights, NY, USA, September 2010
14. Randomized Matrix Algorithms and Applications to the Column Subset Selection Problem
    SIAM Conference on Parallel Processing and Scientific Computing
    Seattle, WA, USA, February 2010
15. Unsupervised Feature Selection for the K-means Clustering Problem
    GAMM Workshop Applied and Numerical Linear Algebra
    ETH Zurich, Switzerland, September 2009
16. Clustered Subset Selection and its Applications on IT Service Metrics
    Institute of Pure and Applied Mathematics (IPAM)
    University of California at Los Angeles (UCLA), Los Angeles, CA, USA, November 2008.
17. Effective Initializations for NMF Algorithms
    SIAM Conference on Optimization
    Boston, MA, USA, May 2008
18. A Randomized Algorithm for Rank-revealing QR Factorizations and Applications
    4th Montreal Scientific Computing Days
    Centre de Researches Mathematiques, Universite de Montreal
    Montreal, Canada, April 2007

Student Mentoring

- Alex Gittens
  Internship with IBM Research. Summer 2012.
- Industry Mentor
  Research in Industrial Projects for Students (RIPS) 2012
  Institute for Pure and Applied Mathematics (IPAM), UCLA
  Academic Mentor: D. Needell
  Students: B. Cung, T. Jin, J. Ramirez, A. Thompson

Teaching Experience

Fall 2007: Teaching Assistant, Computability and Complexity (CSCI-6050/4050)
Computer Science Department, Rensselaer Polytechnic Institute
Instructor: Prof. Mark Goldberg
Duties: Teaching (occasionally), proctoring exams, grading, holding office hours

Spring 2008: Teaching Assistant, Models of Computation (CSCI 2400)
Computer Science Department, Rensselaer Polytechnic Institute
Instructor: Prof. Petros Drineas
Duties: Proctoring exams, grading, holding office hours

Society Service

- Program Committee Member
  ACM Conference on Knowledge Discovery and Data Mining (KDD 2015)
  Research Track
- Program Committee Member
  ACM Conference on Knowledge Discovery and Data Mining (KDD 2015)
  Industry and Government Track
- Program Committee Member
  International Conference on Machine Learning (ICML 2015)
- Program Committee Member
  IEEE International Conference on Data Mining - demos session (ICDM 2014)
- Program Committee Member
  Neural Information Processing Systems (NIPS 2014)

- Program Committee Member
  ACM International Conference on Information and Knowledge Management (CIKM 2014)
- Program Committee Member
  2014 International Conference on Parallel Processing (ICPP-2014)
- Program Committee Member
  ACM Conference on Knowledge Discovery and Data Mining (KDD 2014)
  Research Track
- Program Committee Member
  ACM Conference on Knowledge Discovery and Data Mining (KDD 2014)
  Industry and Government Track
- Organizing Committee Member
  ICML 2013 Workshop on Numerical Linear Algebra in Machine Learning
  Under the auspices of the International Conference on Machine Learning (ICML)
  Atlanta, Georgia, USA June 17-19, 2013
- Program Committee Member
  2013 IEEE International Conference on Big Data (IEEE BigData 2013)
- Program Committee Member
  ACM International Conference on Information and Knowledge Management (CIKM 2013)
- Program Committee Member
  Neural Information Processing Systems (NIPS 2013)
- Organizing Committee Member
  Workshop on Randomized Numerical Linear Algebra (RandNLA): Theory and Practice
  Under the auspices of the Annual IEEE Symposium on Foundations of Computer Science
  New Brunswick NJ, USA October 20, 2012
- Program Committee Member
  ACM International Conference on Information and Knowledge Management (CIKM 2012)
- Program Committee Member
  Neural Information Processing Systems (NIPS 2012)
- Program Committee Member
  Low-rank Matrix Approximation for Large-scale Learning
  NIPS 2010 Workshop, Whistler, Canada, December 11, 2010
- Invited reviewer for the following conferences:
  - STOC 2009, 2014, 2015
  - SODA 2010, 2013, 2014.
  - ICDM 2014
  - ECML-PKDD 2009.
  - ICALP 2014.
  - ISIT 2014.
  - SDM 2008.
  - KDD 2006, 2007, 2009, 2012.
- Invited reviewer for the following journals:
  - Linear Algebra and its Applications,
  - SIAM Journal on Scientific Computing,
  - SIAM Journal on Matrix Analysis and Applications,
  - SIAM Journal on Computing,
  - Journal of Machine Learning Research,
  - ACM Transactions on Knowledge Discovery from Data,
  - IEEE Transactions on Neural Networks,
  - IEEE Transactions on Knowledge and Data Engineering,
  - IEEE Signal Processing Letters,
  - IEEE Journal of Selected Topics in Signal Processing,
  - Journal of Computational and Applied Mathematics,
  - Computational Intelligence and Neuroscience,
  - Machine Learning, Neurocomputing, Pattern Recognition, PLOS One
  - Pattern Analysis and Machine Intelligence, Theory of Computing Systems.