INDERJIT S. DHILLON

Associate Professor
Department of Computer Sciences, The University of Texas at Austin
1 University Station C0500, Austin, TX 78712-0500
Office: ACES 2.332
Phone: 512-471-9725
Fax: 512-471-8885

E-mail: inderjit@cs.utexas.edu
WWW: http://www.cs.utexas.edu/users/inderjit

EDUCATION

Ph.D. University of California at Berkeley - May 1997.

Major:	Computer Science
Thesis:	*A New $O(n^2)$ Algorithm for the Symmetric Tridiagonal Eigenvalue/Eigenvector Problem*
Advisors:	Profs. Beresford N. Parlett and James W. Demmel
Minors:	Mathematics and Theoretical Computer Science.

B. Tech. Indian Institute of Technology, Bombay, India - April 1989.

Major:	Computer Science and Engineering
Thesis:	*Parallel Architectures for Sparse Matrix Computations*
Advisors:	Prof. S. Biswas and Dr. N. K. Karmarkar

RESEARCH INTERESTS

Data mining & machine learning, numerical linear algebra, statistical pattern recognition, bioinformatics, scientific computing, numerical optimization.

RESEARCH EXPERIENCE

09/05-present: Associate Professor, Department of Computer Sciences, University of Texas, Austin.
09/07-12/07: Senior Research Fellow, Institute of Pure & Applied Mathematics (IPAM), UCLA.
09/99-08/05: Assistant Professor, Department of Computer Sciences, University of Texas, Austin.
11/97-08/99: Researcher, IBM Almaden Research Center, San Jose, CA.
05/97-10/97: Post-Doctoral Scholar, EECS Department, University of California at Berkeley, AND
08/91-04/97: Graduate Student Researcher, EECS Department, University of California at Berkeley.
9/89-8/91: Member of Technical Staff, Math Sciences Research Center, AT&T Bell Labs, Murray Hill, NJ.

HONORS & AWARDS

2002-present: Faculty Fellowship, Dept of Computer Sciences, The University of Texas at Austin.

2006: SIAG/LA Prize for the journal paper, "Orthogonal Eigenvectors and Relative Gaps". The award is for "the most outstanding paper on a topic in applicable linear algebra published in English in a peer-reviewed journal in the three calendar years preceding the year of the award."

Spring 2006: Dean's Fellowship, The University of Texas at Austin.

2005: University Cooperative Society's Research Excellence Award for Best Research Paper for "Clustering with Bregman Divergences" co-authored with A. Banerjee, S. Merugu & J. Ghosh.

2007 & 2005: Best Student Paper Awards at ICML (for J. Davis, B. Kulis, P. Jain & S. Sra) at the 24th Int'l Conference on Machine Learning for the paper "Information-Theoretic Metric Learning", and (for B. Kulis & S. Basu) at the 22nd Int'l Conference on Machine Learning for the paper "Semi-Supervised Graph-Based Clustering: A Kernel Approach".

2004: Best Paper Award at the Third SIAM Int'l Conference on Data Mining for the paper "Clustering with Bregman Divergences" co-authored with A. Banerjee, S. Merugu & J. Ghosh, April 2004.

2006: Two Plenary talks at the Ninth SIAM Conference on Applied Linear Algebra, Dusseldorf, Germany.

2008, 2005, 2002 & 1999: Plenary talks at the XVII, XVI, XV and XIV Householder Symposiums on Numerical Linear Algebra (Zeuthen in Germany, Pennsylvania in USA, Peebles in Scotland & Whistler in Canada).

2002: Semi-plenary talk at The Fourth Foundations of Computational Mathematics Conference (FoCM), Minneapolis.

2001: NSF CAREER Award for the period 2001-2006.

1999: Householder Award for the *Best Dissertation in Numerical Linear Algebra* for the period 1996-1998, Honorable Mention.

Fall 1996-Spring 1997: Graduate Research Fellowship from Pacific Northwest National Laboratory (PNNL).

1985-1989: Ranked 2nd (in a class of 300) at Indian Institute of Technology, Bombay.

PHD STUDENTS

Graduated: Joel Tropp in 2004 (Assistant Professor, Caltech, Pasadena), Yuqiang Guan in 2005 (Google, LA), Suvrit Sra in 2007 (Max Planck Institute, Germany), Jason Davis in 2008 (Startup).

Current: Hyuk Cho, Prateek Jain, Dongmin Kim, Brian Kulis, Mátyás Sustik and Wei Tang.

REPRESENTATIVE ACTIVITIES

- Associate Editor, Foundations and Trends in Machine Learning, 2007-present.
- Associate Editor, SIAM Journal for Matrix Analysis and Applications, 2002-present.
- Served on 2006 NSF panel in the Division of Information and Intelligent Systems (IIS), 2004 NSF panel in the Formal and Mathematical Foundations Cluster (FMF), 2001 NSF panel in the Division of Advanced Computational Research (ACR), and 1999 NSF panel in the Division of Information and Intelligent Systems (IIS).
- Area Chair for SIAM Int'l Conference on Data Mining (SDM): 2008 (Atlanta, GA).
- Program Committee Vice-Chair for IEEE Int'l Conference on Data Mining (ICDM), 2005 (New Orleans, LA).
- Senior Program Committee Member for ACM Int'l Conferences on Knowledge Discovery & Data Mining (KDD): 2007 (San Jose, CA).
- Program Committee Member for the SIAM Conference on Applied Linear Algebra: 2009 (Seaside, California).
- Program Committee Member for the World Wide Web Conference (WWW): 2008 (Beijing, China).
- Invited minisymposium on "Mathematical Methods in Data Mining", SIAM Annual Meeting, San Diego, CA, July 2008.
- Program Committee Member for Int'l Conference in Machine Learning (ICML): 2007 (Corvallis, Oregon).
- Reviewer for Neural Information Processing Systems Conference (NIPS): 2008, 2007 & 2006 (Vancouver, Canada).

- Program Committee Member for ACM Int'l Conferences on Knowledge Discovery & Data Mining (KDD): 2008 (Las Vegas, NV), 2006 (Philadelphia, PA), 2005 (Chicago, IL), 2004 (Seattle, WA), 2000 (Boston, MA).
- Program Committee Member for SIAM Int'l Conferences on Data Mining (SDM): 2007 (Minneapolis, MN), 2006 (Bethesda, MD), 2005 (Newport Beach, CA), 2004 (Orlando, FL), 2003 (San Francisco, CA), 2002 (Arlington, VA), 2001 (Chicago, IL).
- Program Committee Member for ACM Conference on Information & Knowledge Management (CIKM): 2006 (Arlington, VA).
- Program Committee Member for IEEE Int'l Conferences on Data Mining (ICDM): 2004 (Brighton, UK), 2003 (Melbourne, FL).
- Program Co-Chair for Workshops on "Clustering High-Dimensional Data and its Applications" at SIAM Int'l Conferences on Data Mining (SDM): 2005 (Newport Beach, CA), 2004 (Orlando, FL), 2003 (San Francisco, CA), 2002 (Arlington, VA).
- Invited minisymposium on "Linear Algebra in Data Mining and Information Retrieval", SIAM Conference on Applied Linear Algebra, Williamsburg, VA, July 2003.
- Program Co-Chair for Workshop on "Clustering High-Dimensional Data and its Applications" at IEEE Int'l Conference on Data Mining (ICDM): 2003 (Melbourne, FL).
- Program Co-Chair for Workshop on "Text Mining" at the Second SIAM Int'l Conference on Data Mining (SDM): 2002 (Arlington, VA).
- Referee for SIAM Review, SIAM Journal for Matrix Analysis and Applications, SIAM Journal on Scientific Computing, SIAM Journal on Numerical Analysis, Linear Algebra and its Applications, BIT, Journal of the ACM, Journal of Machine Learning Research (JMLR), Internet Mathematics, Data Mining and Knowledge Discovery Journal, AI Review, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), IEEE Transactions on Knowledge and Data Engineering (TKDE), IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB), IEEE Transactions on Signal Processing, IEEE Transactions on Image Processing, Information Processing Letters, Decision Support Systems, ACM Transactions on Internet Computing, International Journal of Neural Systems and various conference proceedings.

TEACHING

Spring 2008 & 2007: Instructor for the graduate course CS395T, *"Data Mining: A Statistical Learning Perspective"*.

Fall 2006: Instructor for the undergraduate course CS378, *"Introduction to Data Mining"*.

Fall 2004, 2003, 2001, Spring 2001 & 2000: Instructor for the graduate topics course CS395T, *"Large-Scale Data Mining"*.

Fall 2004, Spring 2004, 2003, 2002 & Fall 2000: Instructor for the undergraduate course CS323E, *"Elements of Scientific Computing"*.

Fall 2008, 2005, 2002 & 1999: Instructor for the graduate breadth course CS383C, *"Numerical Analysis: Linear Algebra"*.

Spring 1993: Teaching Assistant for CS170, *"Efficient Algorithms and Intractable Problems"*. Instructor - Prof. Manuel Blum.

PUBLICATIONS, TALKS, PATENTS & GRANTS

Publications in Progress

1. J. Davis, E. Witchel and I. S. Dhillon, "Unit Test Ordering: A Statistical Approach", working manuscript, 2008.

2. B. Kulis, M. Sustik and I. S. Dhillon, "Low-Rank Kernel Learning with Bregman Matrix Divergences", submitted for publication, 2008.

Journal Publications

1. P. Jain, R. Meka and I. S. Dhillon, "Unsupervised Learning of Disparate Clusterings", To Appear in *Statistical Analysis and Data Mining*, 2008.

2. B. Kulis, S. Basu, I. S. Dhillon and R. J. Mooney, "Semi-Supervised Graph-Based Clustering: A Kernel Approach", To Appear in *Machine Learning*, 2008.

3. I. S. Dhillon, R. Heath Jr., T. Strohmer and J. Tropp, "Constructing Packings in Grassmannian Manifolds via Alternating Projection", *Experimental Mathematics*, vol. 17:1, pages 9–35, 2008.

4. H. Cho and I. S. Dhillon, "Co-clustering of Human Cancer Microarrays using Minimum Sum-Squared Residue Co-clustering", *IEEE/ACM Transactions on Computational Biology and Bioinformatics(TCBB)*, 2008.

5. J. Brickell, I. S. Dhillon, S. Sra and J. Tropp, "The Metric Nearness Problem", *SIAM Journal of Matrix Analysis and Applications*, vol. 30:1, pages 375–396, April 2008.

6. D. Kim, S. Sra, and I. S. Dhillon, "Fast Projection-Based Methods for the Least Squares Nonnegative Matrix Approximation Problem", *Statistical Analysis and Data Mining*, vol. 1:1, pages 38–51, February 2008.

7. I. S. Dhillon and J. Tropp, "Matrix Nearness Problems using Bregman Divergences", *SIAM Journal of Matrix Analysis and Applications*, vol. 29:4, pages 1120–1146, November 2007.

8. I. S. Dhillon, Y. Guan, and B. Kulis, "Weighted Graph Cuts without Eigenvectors: A Multilevel Approach", *IEEE Transactions on Pattern Analysis and Machine Intelligence(PAMI)*, vol. 29:11, pages 1944–1957, November 2007.

9. M. Sustik, J. Tropp, I. S. Dhillon and R. Heath Jr., "On the existence of Equiangular Tight Frames", *Linear Algebra and its Applications*, vol. 426:2–3, pages 619–635, October 2007.

10. A. Banerjee, I. S. Dhillon, J. Ghosh, S. Merugu and D. S. Modha, "A Generalized Maximum Entropy Approach to Bregman Co-Clustering and Matrix Approximations", *Journal of Machine Learning Research(JMLR)*, vol. 8, pages 1919–1986, August 2007.

11. I. S. Dhillon, B. N. Parlett and C. Vömel, "The Design and Implementation of the MRRR Algorithm", *ACM Transactions on Mathematical Software*, vol. 32:4, pages 533–560, December 2006.

12. I. S. Dhillon, B. N. Parlett and C. Vömel, "Glued Matrices and the MRRR Algorithm", *SIAM Journal on Scientific Computing*, vol. 27:2, pages 496–510, October 2005.

13. A. Banerjee, S. Merugu, I. S. Dhillon and J. Ghosh, "Clustering with Bregman Divergences", *Journal of Machine Learning Research(JMLR)*, vol. 6, pages 1705–1749, October 2005.

14. A. Banerjee, I. S. Dhillon, J. Ghosh and S. Sra, "Clustering on the Unit Hypersphere using von Mises-Fisher distributions", *Journal of Machine Learning Research(JMLR)*, vol. 6, pages 1345–1382, September 2005.

15. P. Bientinesi, I. S. Dhillon and R. van de Geijn, "A Parallel Eigensolver for Dense Symmetric Matrices Based on Multiple Relatively Robust Representations", *SIAM Journal on Scientific Computing*, vol. 27:1, pages 43–66, September 2005.

16. J. Tropp, I. S. Dhillon, R. Heath Jr. and T. Strohmer, "Designing Structured Tight Frames via an Alternating Projection Method", *IEEE Transactions on Information Theory*, vol. 51:1, pages 188–209, January 2005.

17. I. S. Dhillon, R. Heath Jr., M. Sustik, J. Tropp, "Generalized finite algorithms for constructing Hermitian matrices with prescribed diagonal and spectrum", *SIAM Journal of Matrix Analysis and Applications*, vol. 27:1, pages 61–71, May 2005 (a longer version appears as UT CS Technical Report # TR-03-49, Nov 2003).

18. J. Tropp, I. S. Dhillon and R. Heath Jr., "Finite-step algorithms for constructing optimal CDMA signature sequences", *IEEE Transactions on Information Theory*, vol. 50:11, pages 2916–2921, November 2004.

19. I. S. Dhillon and B. N. Parlett, "Multiple Representations to Compute Orthogonal Eigenvectors of Symmetric Tridiagonal Matrices", *Linear Algebra and its Applications*, vol. 387, pages 1–28, August 2004.

20. I. S. Dhillon and B. N. Parlett, "Orthogonal Eigenvectors and Relative Gaps", *SIAM Journal of Matrix Analysis and Applications*, vol. 25:3, pages 858–899, March 2004 — **2006 SIAG/LA Prize for the best journal paper in applied linear algebra in the three year period from 2002–2005**.

21. I. S. Dhillon, E. M. Marcotte and U. Roshan, "Diametrical Clustering for identifying anti-correlated gene clusters", *Bioinformatics*, vol. 19:13, pages 1612–1619, September 2003.

22. I. S. Dhillon, S. Mallela and R. Kumar, "A Divisive Information-Theoretic Feature Clustering Algorithm for Text Classification", *Journal of Machine Learning Research(JMLR)*, vol. 3, pages 1265–1287, March 2003.

23. I. S. Dhillon and A. Malyshev, "Inner deflation for symmetric tridiagonal matrices", *Linear Algebra and its Applications*, vol. 358:1-3, pages 139–144, January 2003.

24. I. S. Dhillon, D. S. Modha and W. S. Spangler, "Class Visualization of High-Dimensional Data with Applications", *Computational Statistics & Data Analysis (special issue on Matrix Computations & Statistics)*, vol. 4:1, pages 59–90, November 2002.

25. I. S. Dhillon, and D. S. Modha, "Concept Decompositions for Large Sparse Text Data using Clustering", *Machine Learning*, 42:1, pages 143–175, January 2001.

26. B. N. Parlett and I. S. Dhillon, "Relatively Robust Representations of Symmetric Tridiagonals", *Linear Algebra and its Applications*, vol. 309, pages 121–151, April 2000.

27. I. S. Dhillon, "Current inverse iteration software can fail", *BIT Numerical Mathematics*, 38:4, pages 685–704, December 1998.

28. I. S. Dhillon, "Reliable computation of the condition number of a tridiagonal matrix in $O(n)$ time", *SIAM Journal of Matrix Analysis and Applications*, 19:3, pages 776–796, July 1998.

29. B. N. Parlett, and I. S. Dhillon, "Fernando's solution to Wilkinson's problem: an application of Double Factorization", *Linear Algebra and its Applications*, vol. 267, pages 247–279, November 1997.

30. L. Blackford, A. Cleary, J. Demmel, I. Dhillon, J. Dongarra, S. Hammarling, A. Petitet, H. Ren, K. Stanley and R. Whaley, "Practical Experience in the Numerical Dangers of Heterogeneous Computing", *ACM Transactions on Mathematical Software*, vol. 23, no. 2, pages 133–147, June 1997.

31. J. Choi, J. Demmel, I. Dhillon, J. Dongarra, S. Ostrouchov, A. Petitet, K. Stanley, D. Walker and R. Whaley, "ScaLAPACK: A Portable Linear Algebra Library for Distributed Memory Computers - Design Issues and Performance", *Computer Physics Communications*, vol. 97, pages 1–15, August 1996.

32. J. W. Demmel, I. S. Dhillon, and H. Ren, "On the correctness of some bisection-like eigenvalue algorithms in floating point arithmetic", *Electronic Transactions of Numerical Analysis*, vol. 3, pages 116–140, December 1995.

Conference Publications

1. J. Davis and I. S. Dhillon "Structured Metric Learning for High-Dimensional Problems", To Appear in *Proceedings of the Fourteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(KDD)*, August 2008.

2. P. Jain, R. Meka, C. Caramanis and I. S. Dhillon, "Rank Minimization via Online Learning", To Appear in *Proceedings of the 25th International Conference on Machine Learning(ICML)*, July 2008.

3. P. Jain, R. Meka and I. S. Dhillon, "Unsupervised Learning of Disparate Clusterings", *Proceedings of the Seventh SIAM International Conference on Data Mining*, pages 858–869, April 2008 — **Best of SDM'08 Award**.

4. J. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, "Information-Theoretic Metric Learning", *Proceedings of the 24th International Conference on Machine Learning(ICML)*, pages 209–216, June 2007 — **Best Student Paper Award**.

5. D. Kim, S. Sra, and I. S. Dhillon, "Fast Newton-type Methods for the Least Squares Nonnegative Matrix Approximation Problem", *Proceedings of the Sixth SIAM International Conference on Data Mining*, pages 343–354, April 2007 — **Best of SDM'07 Award**.

6. J. Davis and I. S. Dhillon, "Differential Entropic Clustering of Multivariate Gaussians", *Proceedings of the Neural Information Processing Systems Conference (NIPS)*, pages 337–344, December 2006.

7. J. Davis and I. S. Dhillon "Estimating the Global PageRank of Web Communities", *Proceedings of the Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(KDD)*, pages 116–125, August 2006.

8. B. Kulis, M. Sustik, and I. S. Dhillon, "Learning Low-Rank Kernel Matrices", *Proceedings of the 23rd International Conference on Machine Learning(ICML)*, pages 505–512, July 2006.

9. I. S. Dhillon and S. Sra, "Generalized Nonnegative Matrix Approximations with Bregman Divergences", *Proceedings of the Neural Information Processing Systems Conference (NIPS)*, pages 283–290, December 2005 (also UT CS Technical Report # TR-05-31, June 2005).

10. B. Kulis, S. Basu, I. S. Dhillon and R. J. Mooney, "Semi-Supervised Graph-Based Clustering: A Kernel Approach", *Proceedings of the 22nd International Conference on Machine Learning(ICML)*, pages 457–464, July 2005 — **Distinguished Student Paper Award**.

11. I. S. Dhillon, Y. Guan and B. Kulis, "A Fast Kernel-based Multilevel Algorithm for Graph Clustering", *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(KDD)*, pages 629–634, August 2005.

12. I. S. Dhillon, S. Sra and J. Tropp, "Triangle Fixing Algorithms for the Metric Nearness Problem", *Proceedings of the Neural Information Processing Systems Conference (NIPS)*, pages 361–368, December 2004 (also UT CS Technical Report # TR-04-22, June 2004).

13. A. Banerjee, I. S. Dhillon, J. Ghosh, S. Merugu and D. S. Modha, "A Generalized Maximum Entropy Approach to Bregman Co-Clustering and Matrix Approximations", *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(KDD)*, pages 509–514, August 2004 (a longer version appears as UT CS Technical Report # TR-04-24, June 2004).

14. I. S. Dhillon, Y. Guan, and B. Kulis, "Kernel k-means, Spectral Clustering and Normalized Cuts", *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(KDD)*, pages 551–556, August 2004.

15. A. Banerjee, I. S. Dhillon, J. Ghosh and S. Merugu, "An Information Theoretic Analysis of Maximum Likelihood Mixture Estimation for Exponential Families", *Proceedings of the Twenty-First International Conference on Machine Learning(ICML)*, pages 57–64, July 2004.

16. A. Banerjee, S. Merugu, I. S. Dhillon and J. Ghosh, "Clustering with Bregman Divergences", *Proceedings of the Third SIAM International Conference on Data Mining*, pages 234–245, April 2004 — **Best Paper Award**.

17. H. Cho, I. S. Dhillon, Y. Guan and S. Sra, "Minimum Sum-Squared Residue Co-clustering of Gene Expression Data", *Proceedings of the Third SIAM International Conference on Data Mining*, pages 114–125, April 2004.

18. R. Heath Jr., J. Tropp, I. S. Dhillon and T. Strohmer, "Construction of Equiangular Signatures for Synchronous CDMA Systems", *Proceedings of IEEE International Symposium on Spread Spectrum Techniques and Applications*, Sydney, Australia, August 2004.

19. J. Tropp, I. S. Dhillon and R. Heath Jr., "Optimal CDMA Signatures: A Finite-Step Approach", *Proceedings of IEEE International Symposium on Spread Spectrum Techniques and Applications*, Sydney, Australia, August 2004.

20. J. Tropp, I. S. Dhillon, R. Heath Jr. and T. Strohmer, "CDMA Signature Sequences with Low Peak-to-Average-Power Ratio via Alternating Projection", *Proceedings of the Thirty-Seventh IEEE Asilomar Conference on Signals, Systems, and Computers*, pages 475–479, November 2003.

21. I. S. Dhillon, S. Mallela and D. S. Modha, "Information-Theoretic Co-clustering", *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(KDD)*, pages 89–98, August 2003 (also UT CS Technical Report # TR-03-12, April 2003).

22. A. Banerjee, I. S. Dhillon, J. Ghosh and S. Sra, "Generative Model-Based Clustering of Directional Data", *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(KDD)*, pages 19–28, August 2003 (also UT CS Technical Report # TR-03-07, Feb 2003).

23. I. S. Dhillon and Y. Guan, "Information-Theoretic Clustering of Sparse Co-occurrence Data", *Proceedings of the 3rd IEEE International Conference on Data Mining*, pages 517–520, November 2003 (a longer version appears as UT CS Technical Report # TR-03-39, Sept 2003).

24. I. S. Dhillon, Y. Guan and J. Kogan, "Iterative Clustering of High Dimensional Text Data Augmented by Local Search", *Proceedings of the 2nd IEEE International Conference on Data Mining*, pages 131–138, December 2002.

25. I. S. Dhillon, S. Mallela and R. Kumar, "Enhanced Word Clustering for Hierarchical Text Classification", *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(KDD)*, July 2002 (also UT CS Technical Report # TR-02-17, March 2002).

26. I. S. Dhillon, "Co-Clustering Documents and Words Using Bipartite Spectral Graph Partitioning", *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(KDD)*, August 2001 (also UT CS Technical Report # TR-01-05, March 2001).

27. I. S. Dhillon, D. S. Modha, and W. S. Spangler, "Visualizing Class Structure of Multi-Dimensional Data", In *Proceedings of the 30th Symposium of the Interface: Computing Science and Statistics*, Interface Foundation of North America, vol. 30, pages 488–493, May 1998.

28. L. Blackford, J. Choi, A. Cleary, E. D'Azevedo, J. Demmel, I. Dhillon, J. Dongarra, S. Hammarling, G. Henry, A. Petitet, K. Stanley, D. Walker and R. Whaley, "ScaLAPACK: A Linear Algebra Library for Message-Passing Computers", *Proceedings of the Eighth SIAM Conference on Parallel Processing for Scientific Computing*, March 1997.

29. I. S. Dhillon, G. Fann and B. N. Parlett, "Application of a New Algorithm for the Symmetric Eigenproblem to Computational Quantum Chemistry", *Proceedings of the Eighth SIAM Conference on Parallel Processing for Scientific Computing*, March 1997.

30. L. Blackford, J. Choi, A. Cleary, J. Demmel, I. Dhillon, J. Dongarra, S. Hammarling, G. Henry, A. Petitet, K. Stanley, D. Walker and R. Whaley, "ScaLAPACK: A Portable Linear Algebra Library for Distributed Memory Computers - Design Issues and Performance", *Proceedings of Supercomputing '96*, 1996.

31. I. S. Dhillon, N. K. Karmarkar, and K. G. Ramakrishnan, "An Overview of the Compilation Process for a New Parallel Architecture", *Supercomputing Symposium '91*, pages 471–486, June 1991.

Workshop Papers

1. J. Brickell, I. S. Dhillon and D. Modha, "Adaptive Website Design using Caching Algorithms", Twelfth ACM International Conference on Knowledge Discovery and Data Mining (KDD) (Workshop on Web Mining and Web Usage Analysis (WebKDD-2006)), August 2006.

2. I. S. Dhillon and Y. Guan, "Clustering Large, Sparse, Co-occurrence Data", 3rd SIAM International Conference on Data Mining (Workshop on Clustering High-Dimensional Data and its Applications), May 2003.

3. I. S. Dhillon, S. Mallela and R. Kumar, "Information-Theoretic Feature Clustering for Text Classification", Nineteenth International Conference on Machine Learning (ICML) (Workshop on Text Learning (TextML-2002)), July 2002.

4. I. S. Dhillon, Y. Guan and J. Kogan, "Refining clusters in high-dimensional text data", 2nd SIAM International Conference on Data Mining (Workshop on Clustering High-Dimensional Data and its Applications), April 2002 (also UT CS Technical Report # TR-02-03, January 2002).

Book Chapters

1. J. Brickell, I. S. Dhillon and D. Modha, "Adaptive Website Design using Caching Algorithms", In: O. Nasraoui, M. Spiliopoulou, J. Srivastava, B. Mobasher, B. Masand(eds), *Advances in Web Mining and Web Usage Analysis*, Springer Lecture Notes in Computer Science (LNCS/LNAI), vol. 4811, pages 1–20, Sept 2007.

2. A. K. Cline and I. S. Dhillon, "Computation of the Singular Value Decomposition", In: L. Hogben, R. Brualdi, A. Greenbaum and R. Mathias(eds): *Handbook of Linear Algebra*, Invited Book Chapter, CRC Press, pages 45-1–45-13, 2006.

3. M. Teboulle, P. Berkhin, I. S. Dhillon, Y. Guan and J. Kogan, "Clustering with Entropy-like k-means Algorithms", invited book chapter, In: *Grouping Multidimensional Data – Recent Advances in Clustering*, Springer-Verlag, pages 127–160, 2005.

4. I. S. Dhillon, J. Kogan and C. Nicholas, "Feature Selection and Document Clustering", In: Michael Berry(ed): *A Comprehensive Survey of Text Mining*, Springer-Verlag, pages 73–100, 2003.

5. I. S. Dhillon, Y. Guan and J. Fan, "Efficient Clustering of Very Large Document Collections", invited book chapter, In: R. Grossman and C. Kamath and P. Kegelmeyer and V. Kumar and R. Namburu(eds): *Data Mining for Scientific and Engineering Applications*, In *Data Mining for Scientific and Engineering Applications*, Kluwer Academic Publishers, pages 357–381, 2001.

6. I. S. Dhillon, and D. S. Modha, "A Data Clustering Algorithm on Distributed Memory Multiprocessors", In: M.Zaki and C.T.Ho(eds): *Large-Scale Parallel Data Mining, Lecture Notes in Artificial Intelligence, vol. 1759*, Springer-Verlag, pages 245–260, March 2000 (also IBM Research Report RJ 10134).

Book

1. L. Blackford, J. Choi, A. Cleary, E. D'Azevedo, J. Demmel, I. Dhillon, J. Dongarra, S. Hammarling, G. Henry, A. Petitet, K. Stanley, D. Walker and R. Whaley, "ScaLAPACK Users' Guide", *SIAM*, 1997.

Technical Reports

1. B. Kulis, S. Sra, S. Jegelka, and I. S. Dhillon. "Scalable Semidefinite Programming using Convex Perturbations". UT CS Technical Report # TR-07-47, September 2007.

2. P. Jain, B. Kulis and I. S. Dhillon, "Online Linear Regression using Burg Entropy", UT CS Technical Report # TR-07-08, Feb 2007.

3. S. Sra and I. S. Dhillon, "Nonnegative Matrix Approximation: Algorithms and Applications", UT CS Technical Report # TR-06-27, June 2006.

4. I. S. Dhillon and S. Sra, "Generalized Nonnegative Matrix Approximations with Bregman Divergences", UT CS Technical Report # TR-05-31, June 2005.

5. I. S. Dhillon, Y. Guan, and B. Kulis, "A Unified View of Kernel k-means, Spectral Clustering and Graph Cuts", UT CS Technical Report # TR-04-25, June 2004.

6. I. S. Dhillon, S. Sra and J. Tropp, "Triangle Fixing Algorithms for the Metric Nearness Problem", UT CS Technical Report # TR-04-22, June 2004.

7. I. S. Dhillon, S. Sra and J. Tropp, "The Metric Nearness Problem with Applications", UT CS Technical Report # TR-03-23, July 2003.

8. A. Banerjee, I. S. Dhillon, J. Ghosh and S. Sra, "Clustering on Hyperspheres using Expectation Maximization", UT CS Technical Report # TR-03-07, February 2003.

9. I. S. Dhillon and S. Sra, "Modeling data using directional distributions", UT CS Technical Report # TR-03-06, January 2003.

10. I. S. Dhillon, "A New $O(n^2)$ Algorithm for the Symmetric Tridiagonal Eigenvalue/Eigenvector Problem", PhD Thesis, University of California, Berkeley, May 1997 (also available as UCB Tech. Report No. UCB//CSD-97-971).

11. M. Gu, J. W. Demmel and I. S. Dhillon, "Efficient Computation of the Singular Value Decomposition with Applications to Least Squares Problems", *Technical Report LBL-36201, Lawrence Berkeley National Laboratory*, 1994 (also available as LAPACK working note no. 88).

12. J. Choi, J. Demmel, I. Dhillon, J. Dongarra, S. Ostrouchov, A. Petitet, K. Stanley, D. Walker and R. Whaley, "Installation Guide for ScaLAPACK", University of Tennessee Computer Science Technical Report, UT-CS-95-280, March 1995 (version 1.0), updated August 31, 2001 (version 1.7) — also available as LAPACK working note no. 93.

13. I. S. Dhillon, N. K. Karmarkar and K. G. Ramakrishnan, "Performance Analysis of a Proposed Parallel Architecture on Matrix Vector Multiply Like Routines", *Technical Memorandum 11216-901004-13TM, AT&T Bell Laboratories*, Murray Hill, NJ, 1990.

14. I. S. Dhillon, "A Parallel Architecture for Sparse Matrix Computations", *B.Tech. Project Report, Indian Institute of Technology*, Bombay, 1989.

INVITED TALKS AT MAJOR CONFERENCES

July 2008: "The Log-Determinant Divergence and its Applications", Plenary talk, *Householder XVII Symposium*, Zeuthen, Germany.

July 2006: "Orthogonal Eigenvectors and Relative Gaps", SIAG/LA Prize talk, *Ninth SIAM Conference on Applied Linear Algebra*, Dusseldorf, Germany.

July 2006: "From Shannon to von Neumann: New Distance Measures for Matrix Nearness Problems", Plenary talk, *Ninth SIAM Conference on Applied Linear Algebra*, Dusseldorf, Germany.

May 2005: "Matrix Nearness Problems using Bregman Divergences", Plenary talk, *Householder XVI Symposium*, Seven Springs, Pennsylvania.

August 2002: "Fast and Accurate Eigenvector Computation in Finite Precision Arithmetic", Semi-plenary talk, *The Fourth Foundations of Computational Mathematics Conference (FoCM)*, Minneapolis, Minnesota.

June 2002: "Matrix Problems in Data Mining", Plenary talk, *Householder XV Symposium*, Peebles, Scotland.

June 1999: "Orthogonal Eigenvectors through Relatively Robust Representations", Plenary talk, *Householder XIV Symposium*, Whistler, Canada.

INVITED TALKS

June 2008: "Rank Minimization via Online Learning", Invited talk, *Workshop on Algorithms for Modern Massive Data Sets*, Stanford University, California.

May 2008: "Machine Learning with Bregman Divergences", Invited talk, *EurOPT-2008*, Neringa, Lithuania.

March 2008: "The Symmetric Tridiagonal Eigenproblem", Invited talk, Bay Area Scientific Computing Day, Mathematical Sciences Research Institute (MSRI), Berkeley, California.

February 2008: "Metric and Kernel Learning", Invited colloquium talk, ORFE (Operations Research & Financial Engineering) Department, Princeton University, Princeton, New Jersey.

November 2007: "Metric and Kernel Learning", Invited colloquium talk, Department of Computer Sciences, Cornell University, Ithaca, New York.

Sept, Oct & Nov 2007: "Clustering Tutorial", "Metric and Kernel Learning" & "Multilevel Graph Clustering", Special program on *Mathematics of Knowledge and Seach Engines*, Institute of Pure & Applied Mathematics (IPAM), UCLA, California.

June 2007: "Machine Learning and Optimization", Panel Speaker, *A-C-N-W Optimization Tutorials*, Chicago, Illinois.

Feb 2007: "Fast Newton-type Methods for Nonnegative Matrix Approximation", Invited talk, *NISS Workshop on Non-negative Matrix Factorization*, Raleigh, N. Carolina.

Jan 2007: "Machine Learning with Bregman Divergences", *BIRS Seminar on Mathematical Programming in Data Mining and Machine Learning*, organized by M.Jordan, J.Peng, T.Poggio, K.Scheinberg, D.Schuurmans and T.Terlaky, Banff, Canada.

June 2006: "Kernel Learning with Bregman Matrix Divergences" Invited talk, *Workshop on Algorithms for Modern Massive Data Sets*, Stanford University and Yahoo! Research, California.

May 2006: "Spectral Measures for Nearness Problems", Plenary talk, *Sixth International Workshop on Accurate Solution of Eigenvalue Problems*, Pennsylvania State University, University Park, Pennsylvania.

December 2005: "Co-Clustering, Matrix Approximations and Bregman Divergences", Invited colloquium talk, Department of Computer Sciences, Cornell University, Ithaca, New York.

November 2005: "Co-Clustering, Matrix Approximations and Bregman Divergences", Invited talk, McMaster University, Hamilton, Canada.

March 2004: "Information-Theoretic Clustering, Co-clustering and Matrix Approximations", Invited talk, IBM TJ Watson Research Center, New York.

Feb 2004: "Fast Eigenvalue/Eigenvector Computation for Dense Symmetric Matrices", Invited talk, University of Illinois, Urbana-Champaign.

Nov 2003: "Inverse Eigenvalue Problems in Wireless Communications", *BIRS Seminar on Theory and Numerics of Matrix Eigenvalue Problems*, organized by J.Demmel, N.Higham and P.Lancaster, Banff, Canada.

Oct 2003: "Inverse Eigenvalue Problems in Wireless Communications", *Dagstuhl Seminar on Theoretical and Computational Aspects of Matrix Algorithms*, organized by N.Higham, V.Mehrmann, S.Rump and D.Szyld, Wadern, Germany.

Aug 2003: "Accurate Computation of Eigenvalues and Eigenvectors of Dense Symmetric Matrices", Sandia National Laboratories, Albuquerque.

May 2003: "Information-Theoretic Clustering, Co-clustering and Matrix Approximations", *IMA Workshop on Data Analysis and Optimization*, organized by R.Kannan, J.Kleinberg, C.Papadimitriou and P.Ragahavan, Minneapolis, Minnesota.

December 2001: "Clustering High-Dimensional Data and Data Approximation", Invited colloquium talk, University of Minnesota, Minneapolis.

October 2000: "Concept Decompositions for Large-Scale Information Retrieval", Invited talk, *Computational Information Retrieval Workshop*, Raleigh, Carolina.

April 2000: "Matrix Approximations for Large Sparse Text Data using Clustering", Invited talk, *IMA Workshop on Text Mining*, Minneapolis, Minnesota.

September 1999: "Class Visualization of High-Dimensional Data", Invited talk, *Workshop on Mining Scientific Datasets*, Minneapolis, Minnesota.

March & April 1999: "Eigenvectors and Concept Vectors", Invited talks at UC Santa Barbara, Stanford, UT Austin, Yale, UW Madison and Caltech.

OTHER MAJOR TALKS

June 2008: Invited minisymposium talk, "On some modified root finding problems", *15th Conference of the International Linear Algebra Society (ILAS)*, Cancun, Mexico.

July 2007: Invited minisymposium talk, "Fast Newton-type Methods for Nonnegative Tensor Approximation", *Sixth Int'l Conference on Industrial and Applied Mathematics (ICIAM)*, Zurich, Switzerland.

July 2005: Invited minisymposium talk, "Glued Matrices and the MRRR Algorithm", Invited minisymposium talk, *SIAM Annual Meeting*, New Orleans, Louisiana.

Feb 2004: "A Parallel Eigensolver for Dense Symmetric Matrices based on Multiple Relatively Robust Representations", Invited minisymposium talk, *SIAM Conference on Parallel Processing for Scientific Computing*, San Francisco, California.

Aug 2003 – June 2004: "Information Theoretic Clustering, Co-Clustering and Matrix Approximations", PARC, Yahoo!, Verity, Univ. of Wisconsin-Madison.

August 2003: "Information-Theoretic Co-clustering", *The Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(KDD)*, Washington DC.

July 2003: "Data Clustering using Generalized Distortion Measures", Invited minisymposium talk, *SIAM Conference on Applied Linear Algebra*, Williamsburg, Virginia.

July 2003: "Matrix Nearness Problems in Data Mining", Invited minisymposium talk, *SIAM Conference on Applied Linear Algebra*, Williamsburg, Virginia.

July 2002: "Enhanced Word Clustering for Hierarchical Text Classification", *The Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(KDD)*, Edmonton, Canada.

June 2002: "Accurate Computation of Eigenvalues and Eigenvectors of Tridiagonal Matrices", *Fourth International Workshop on Accurate Eigensolving and Applications*, Split, Croatia.

April 2002: "Refining Clusters in High-Dimensional Text Data", *2nd SIAM International Conference on Data Mining*, Arlington, Virginia.

July 2001: "Large-Scale Data Mining", Invited minisymposium talk, *SIAM Annual Meeting*, San Diego, California.

October 2000: "Multiple Representations for Orthogonal Eigenvectors", Invited minisymposium talk, *SIAM Conference on Applied Linear Algebra*, Raleigh, Carolina.

December 1999: "Class Visualization of Multidimensional Data with Applications", Invited minisymposium on Data Mining, *6th International Conference on High Performance Computing (HiPC '99)*, Calcutta, India.

May 1999: "Concept-Revealing Subspaces for Large Text Collections", Invited minisymposium talk, *SIAM Annual Meeting*, Atlanta, Georgia.

August 1998: "Concept Identification in Large Text Collections", *5th International Symposium, IRREGULAR'98*, Berkeley, California.

April 1998: "Orthogonal Eigenvectors without Gram-Schmidt", *Ph.D. Dissertation Talk*, Berkeley, California.

October 1997: "When are Factors of Indefinite Matrices Relatively Robust?", *Sixth SIAM Conference on Applied Linear Algebra*, Snowbird, Utah.

July 1997: "Perfect Shifts and Twisted Q Factorizations", *SIAM 45th Anniversary and Annual Meeting*, Stanford, California.

March 1997: "A New Algorithm for the Symmetric Eigenproblem Applied to Computational Quantum Chemistry", *Eighth SIAM Conference on Parallel Processing for Scientific Computing*, Minneapolis, Minnesota.

August 1996: "Accuracy and Orthogonality", *First International Workshop on Accurate Eigensolving and Applications*, Split, Croatia.

July 1996: "A New Approach to the Symmetric Tridiagonal Eigenproblem", *Householder XIII Symposium*, Pontresina, Switzerland.

June 1991: "Compilation Process for a New Parallel Architecture based on Finite Geometries", *Supercomputing Symposium '91*, New Brunswick, Canada.

PATENTS

1. US6269376 awarded in 2001: "Method and system for clustering data in parallel on a distributed memory multiprocessor system", I.S.Dhillon and D.S.Modha.

2. US6560597 awarded in 2003: "Concept decomposition using clustering", I.S.Dhillon and D.S.Modha.

GRANTS

1. "Graph Data Mining", Sandia National Laboratories, $24,360, 06/01/08-08/31/08.

2. "Non-Negative Matrix and Tensor Approximations: Algorithms, Software and Applications", National Science Foundation, CCF-0728879, $250,000, 01/01/08-12/31/10.

3. "III-COR: Versatile Co-clustering Analysis for Bi-modal and Multi-modal Data", National Science Foundation, IIS-0713142, $430,000, 09/01/07-08/31/10.

4. "Sparse Data Estimation through Hierarchical Aggregation", Yahoo! Research, $25,000, 01/01/08-12/31/08.

5. Supplemental Award for National Science Foundation Grant "Novel Matrix Problems in Modern Applications", CCF-0431257, $30,000, 08/15/05-07/31/07.

6. REU Supplemental Award for National Science Foundation Grant "Scalable Algorithms for Large-Scale Data Mining", ACI-0093404, $6,000, 06/01/05-05/31/06.

7. "Novel Matrix Problems in Modern Applications", National Science Foundation, CCF-0431257, $200,000, 08/15/04-07/31/07.

8. "Web data mining and algorithms", Sabre Holdings, Inc., $72,000, 09/01/04-08/31/07.

9. "RI: Mastodon: A Large-Memory, High-Throughput Simulation Infrastructure", 18 co-PIs, National Science Foundation, CISE Research Infrastructure, extramural funding: $1,418,231, UT match: $508,000, total grant: $1,927,031, 2003-2008.

10. "ITR: Feedback from Multi-Source Data Mining to Experimentation for Gene Network Discovery", with R. Mooney, D. Miranker, V. Iyer and E. Marcotte, NSF Information Technology Research Award, IIS-0325116, $1,700,000, 11/03/03-10/31/07.

11. "Reconstructing Gene Networks by Mining Expression, Genomic and Literature Data", with E. Marcotte, Texas Higher Education Coordinating Board, Advanced Research Program (TARP), Award # 003658-0431-2001, $147,000, 01/01/02-08/31/04.

12. "Scalable Algorithms for Large-Scale Data Mining", NSF CAREER Award, ACI-0093404, $478,305, 06/01/01-05/31/07.

13. "Data Mining Seminar Series", with J. Ghosh and R. Mooney, Tivoli Inc., $5,000, 06/01/01-05/31/03.

14. "Linear Algebra for Text Classification", Lawrence Berkeley National Laboratories, $11,570.13, 08/01/00-05/31/01.

15. "An Interdisciplinary Practicum Course in Data Mining", with J. Ghosh, Tivoli Inc., $20,000, 06/01/00-05/31/03.

16. "Data Mining Seminar Series", with J. Ghosh, R. Mooney and L. Stokes, Tivoli Inc., $5,000, 06/01/00-05/31/01.

17. "Research in Web Mining", Neonyoyo Inc., $8,520, 06/01/00-08/31/00.

18. "Studies in Numerical Linear Algebra and Applications," University of Texas, Summer Research Award, $15,112, 06/01/00-07/31/00.

19. "SCOUT: Scientific Computing Cluster of UT (CISE Research Instrumentation)," with D. Burger, S. Keckler, H. Vin and T. Warnow, National Science Foundation, CISE-9985991, extramural funding: $139,481, UT match: $70,000, total grant: $209,481, 03/15/00-03/14/03.