Bibliography

. Spiros Papadimitriou, . Jimeng Sun, and . Christos Faloutsos. Streaming pattern discovery in multiple time-series. In *VLDB '05: Proceedings of the 31st international conference on Very large data bases*, pages 697–708. VLDB Endowment, 2005. ISBN 1-59593-154-6. Cited on pages 5, 16, 20, 35, 37, 46, 56, and 70.

. Evan Hoke, . Jimeng Sun, . John D. Strunk, . Gregory R. Ganger, and . Christos Faloutsos. Intemon: continuous mining of sensor data in large-scale self-infrastructures. *SIGOPS Oper. Syst. Rev.*, 40(3):38–44, 2006. ISSN 0163-5980. doi: http://doi.acm.org/10.1145/1151374.1151384. Cited on pages 8, 16, 20, 21, 37, 56, 70, and 78.

Jake D. Brutlag. Aberrant behavior detection in time series for network monitoring. In *LISA '00: Proceedings of the 14th USENIX conference on System administration*, pages 139–146, Berkeley, CA, USA, 2000. USENIX Association. Cited on pages 8, 23, and 24.

Charu Aggarwal. *Data Streams: Models and Algorithms*. Advances in Database Systems. Springer, 2007. URL `http://www.springer.com/west/home/default?SGWID=4-40356-22-107949228-0`. Cited on page 13.

João Gama and Mohamed Medhat Gaber. *Learning From Data Streams : Processing Techniques in Sensor Networks*. Springer, September 2007. URL `http://www.liaad.up.pt/pub/2007/GG07`. Cited on page 13.

S. Muthukrishnan. Data streams: algorithms and applications. *Found. Trends Theor. Comput. Sci.*, 1(2):117–236, 2005. ISSN 1551-305X. doi: http://dx.doi.org/10.1561/0400000002. Cited on page 13.

Daniel J. Abadi, Don Carney, Ugur Cetintemel, Mitch Cherniack, Christian Convey, Sangdon Lee, Michael Stonebraker, Nesime Tatbul, and Stan Zdonik. Aurora: a new model and architecture for data stream management. *The VLDB Journal*, 12(2):120–139, 2003. ISSN 1066-8888. doi: http://dx.doi.org/10.1007/s00778-003-0095-z. Cited on page 13.

A. Arasu, B. Babcock, S. Babu, J. Cieslewicz, M. Datar, K. Ito, R. Motwani, U. Srivastava, and J. Widom. Stream: The stanford data stream management system. Technical Report 2004-20, Stanford, 2004. URL `http://ilpubs.stanford.edu:8090/641/`. Cited on page 13.

Sirish Chandrasekaran, Owen Cooper, Amol Deshpande, Michael J. Franklin, Joseph M. Hellerstein, Wei Hong, Sailesh Krishnamurthy, Samuel R. Madden, Fred Reiss, and Mehul A. Shah. Telegraphcq: continuous dataflow processing. In *SIGMOD '03: Proceedings of the 2003 ACM SIGMOD international conference on Management of data*, pages 668–668, New York, NY, USA, 2003. ACM. ISBN 1-58113-634-X. doi: http://doi.acm.org/10.1145/872757.872857. Cited on page 13.

Chuck Cranor, Theodore Johnson, Oliver Spataschek, and Vladislav Shkapenyuk. Gigascope: a stream database for network applications. In *SIGMOD '03: Proceedings of the 2003 ACM SIGMOD international conference on Management of data*, pages 647–651, New York, NY, USA, 2003. ACM. ISBN 1-58113-634-X. doi: http://doi.acm.org/10.1145/872757.872838. Cited on page 13.

Victoria Hodge and Jim Austin. A survey of outlier detection methodologies. *Artif. Intell. Rev.*, 22(2):85–126, 2004. ISSN 0269-2821. doi: http://dx.doi.org/10.1023/B:AIRE.0000045502.10941.a9. Cited on page 14.

Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection : A survey. *ACM Computing Surveys*, 2009. URL `http://www-users.cs.umn.edu/~kumar/papers/anomaly-survey.php`. Cited on page 14.

Animesh Patcha and Jung-Min Park. An overview of anomaly detection techniques: Existing solutions and latest technological trends. *Comput. Netw.*, 51(12):3448–3470, 2007. ISSN 1389-1286. doi: http://dx.doi.org/10.1016/j.comnet.2007.02.001. Cited on page 14.

. Dragomir Yankov, . Eamonn Keogh, and . Umaa Rebbapragada. Disk aware discord discovery: finding unusual time series in terabyte sized datasets. *Knowl. Inf. Syst.*, 17(2):241–262, 2008. ISSN 0219-1377. doi: http://dx.doi.org/10.1007/s10115-008-0131-9. Cited on page 14.

Yu Gu, Andrew McCallum, and Don Towsley. Detecting anomalies in network traffic using maximum entropy estimation. In *IMC '05: Proceedings of the 5th ACM SIGCOMM conference on Internet Measurement*, pages 32–32, Berkeley, CA, USA, 2005. USENIX Association. Cited on page 15.

Anukool Lakhina, Mark Crovella, and Christophe Diot. Mining anomalies using traffic feature distributions. In *SIGCOMM '05: Proceedings of the 2005 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 217–228, New York, NY, USA, 2005. ACM. ISBN 1-59593-009-4. doi: http://doi.acm.org/10.1145/1080091.1080118. Cited on pages 15 and 20.

Anukool Lakhina, Mark Crovella, and Christophe Diot. Diagnosing network-wide traffic anomalies. *SIGCOMM Comput. Commun. Rev.*, 34(4):219–230, 2004a. ISSN 0146-4833. doi: http://doi.acm.org/10.1145/1030194.1015492. Cited on pages 15, 20, and 54.

Paul Barford, Jeffery Kline, David Plonka, and Amos Ron. A signal analysis of network traffic anomalies. In *IMW '02: Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurment*, pages 71–82, New York, NY, USA, 2002. ACM. ISBN 1-58113-603-X. doi: http://doi.acm.org/10.1145/637201.637210. Cited on page 15.

Liwei Kuang and Mohammad Zulkernine. An anomaly intrusion detection method using the csi-knn algorithm. In *SAC '08: Proceedings of the 2008 ACM symposium on Applied computing*, pages 921–926, New York, NY, USA, 2008. ACM. ISBN 978-1-59593-753-7. doi: http://doi.acm.org/10.1145/1363686.1363897. Cited on page 15.

Kuai Xu, Zhi-Li Zhang, and Supratik Bhattacharyya. Profiling internet backbone traffic: behavior models and applications. *SIGCOMM Comput. Commun. Rev.*, 35(4):169–180, 2005. ISSN 0146-4833. doi: http://doi.acm.org/10.1145/1090191.1080112. Cited on page 15.

Wenke Lee and Salvatore J. Stolfo. A framework for constructing features and models for intrusion detection systems. *ACM Trans. Inf. Syst. Secur.*, 3(4):227–261, 2000. ISSN 1094-9224. doi: http://doi.acm.org/10.1145/382912.382914. Cited on page 15.

Robert Schweller, Ashish Gupta, Elliot Parsons, and Yan Chen. Reversible sketches for efficient and accurate change detection over network data streams. In *IMC '04: Proceedings of the 4th ACM SIGCOMM conference on Internet measurement*, pages 207–212, New York, NY, USA, 2004. ACM. ISBN 1-58113-821-0. doi: http://doi.acm.org/10.1145/1028788.1028814. Cited on page 15.

Varun Chandola, Varun Mithal, and Vipin Kumar. Comparative evaluation of anomaly detection techniques for sequence data. In *ICDM '08: Proceedings of the 2008 Eighth IEEE International Conference on Data Mining*, pages 743–748, Washington, DC, USA, 2008. IEEE Computer Society. ISBN 978-0-7695-3502-9. doi: http://dx.doi.org/10.1109/ICDM.2008.151. Cited on page 15.

Jessica Lin, Eamonn Keogh, Li Wei, and Stefano Lonardi. Experiencing sax: a novel symbolic representation of time series. *Data Min. Knowl. Discov.*, 15(2):107–144, 2007. ISSN 1384-5810. doi: http://dx.doi.org/10.1007/s10618-007-0064-z. Cited on page 15.

C. Bandt and B. Pompe. Permutation entropy: A natural complexity measure for time series. *Physical Review Letters*, 88(17):174102, apr 2002. doi: 10.1103/PhysRevLett.88.174102. Cited on page 15.

. Eamonn Keogh, . Stefano Lonardi, and . Chotirat Ann Ratanamahatana. Towards parameter-free data mining. In *KDD '04: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 206–215, New York, NY, USA, 2004. ACM. ISBN 1-58113-888-1. doi: http://doi.acm.org/10.1145/1014052.1014077. Cited on page 15.

. Li Wei, . Nitin Kumar, . Venkata Lolla, . Eamonn J. Keogh, . Stefano Lonardi, and . Chotirat Ratanamahatana. Assumption-free anomaly detection in time series. In *SSDBM'2005: Proceedings of the 17th international conference on Scientific and statistical database management*, pages 237–240, Berkeley, CA, US, 2005. Lawrence Berkeley Laboratory. ISBN 1-88888-111-X. Cited on page 15.

F. Y. Edgeworth. On discordant observations. *Philosophical Magazine*, 23(5): 364–375, 1887. Cited on page 15.

T. Ahmed, M. Coates, and A. Lakhina. Multivariate online anomaly detection using kernel recursive least squares. In *INFOCOM 2007. 26th IEEE International Conference on Computer Communications.*, pages 625–633. IEEE, 2007a. ISBN 1-4244-1047-9. Cited on pages 16, 22, 23, 52, 54, 64, 65, and 70.

Bin Yang. Projection approximation subspace tracking. *Signal Processing, IEEE Transactions on*, 43(1):95–107, Jan 1995a. ISSN 1053-587X. doi: 10.1109/78.365290. Cited on pages 16, 29, 30, and 33.

Peter Strobach. The fast recursive row-householder subspace tracking algorithm. *Signal Processing*, In Press, Corrected Proof, 2009a. ISSN 0165-1684. doi: DOI:10.1016/j.sigpro.2009.04.012. URL `http://www.sciencedirect.com/science/article/B6V18-4W4JDK7-6/2/e188abbf108d636c843932abf8300e75`. Cited on pages 17, 26, 28, 29, 31, 32, 33, 34, 42, 44, 52, 55, 56, 78, and 93.

Keinosuke Fukunaga. *Introduction to statistical pattern recognition (2nd ed.)*. Academic Press Professional, Inc., San Diego, CA, USA, 1990. ISBN 0-12-269851-7. Cited on page 20.

Jimeng Sun, Spiros Papadimitriou, and Christos Faloutsos. Online latent variable detection in sensor networks. In *ICDE '05: Proceedings of the 21st International Conference on Data Engineering*, pages 1126–1127, Washington, DC, USA, 2005. IEEE Computer Society. ISBN 0-7695-2285-8. doi: http://dx.doi.org/10.1109/ICDE.2005.100. Cited on page 20.

A. Lakhina, K. Papagiannaki, M. Crovella, C. Diot, E. Kolaczyk, and N. Taft. Structural analysis of network traffic flows, 2004b. URL `citeseer.ist.psu.edu/article/lakhina03structural.html`. Cited on pages 20 and 22.

Wei Wang and Roberto Battiti. Identifying intrusions in computer networks with principal component analysis. In *ARES '06: Proceedings of the First International Conference on Availability, Reliability and Security*, pages 270–279, Washington, DC, USA, 2006. IEEE Computer Society. ISBN 0-7695-2567-9. doi: http://dx.doi.org/10.1109/ARES.2006.73. Cited on page 20.

Mei-ling Shyu, Shu-ching Chen, Kanoksri Sarinnapakorn, and Liwu Chang. A novel anomaly detection scheme based on principal component classifier. In *in Proceedings of the IEEE Foundations and New Directions of Data Mining Workshop, in conjunction with the Third IEEE International Conference on Data Mining (ICDM)*, pages 172–179, 2003. Cited on page 20.

M. Thottan and C. Ji. Anomaly detection in ip networks. *IEEE Transactions on Signal Processing*, 51:2191–2204, aug 2003. doi: 10.1109/TSP.2003.814797. Cited on page 20.

Haakon Ringberg, Augustin Soule, Jennifer Rexford, and Christophe Diot. Sensitivity of pca for traffic anomaly detection. *SIGMETRICS Perform. Eval. Rev.*, 35(1): 109–120, 2007. ISSN 0163-5999. doi: http://doi.acm.org/10.1145/1269899.1254895. Cited on page 22.

Yaakov Engel, Shie Mannor, and Ron Meir. The kernel recursive least squares algorithm. *IEEE Transactions on Signal Processing*, 52:2275–2285, 2003. Cited on page 22.

Tarem Ahmed, Boris Oreshkin, and Mark Coates. Machine learning approaches to network anomaly detection. In *in Proceedings of the Second Workshop on Tackling Computer Systems Problems with Machine Learning (SysML*, 2007b. Cited on page 23.

Alberto Munoz and Javier M. Moguerza. Estimation of high-density regions using one-class neighbor machines. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(3): 476–480, 2006. ISSN 0162-8828. doi: http://dx.doi.org/10.1109/TPAMI.2006.52. Cited on page 23.

I. T. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. Springer, 2002. Cited on pages 26, 35, and 37.

P. Comon and G. H. Golub. Tracking a few extreme singular values and vectors in signal processing. *Proceedings of the IEEE*, 78(8):1327–1343, Aug 1990. ISSN 0018-9219. doi: 10.1109/5.58320. Cited on page 26.

X. G. Doukopoulos and G. V. Moustakides. Fast and stable subspace tracking. *IEEE Transactions on Signal Processing*, 56:1452–1465, apr 2008. doi: 10.1109/TSP.2007.909335. Cited on pages 26, 33, 36, and 52.

Xenofon Doukopoulos. *Power techniques for blind channel estimation in wireless communication systems*. PhD thesis, University of Rennes, 2004. Cited on page 26.

Magnus Borga and C. Magnus Borga. Learning multidimensional signal processing. Technical report, 1998. Cited on page 27.

N. Owsley. Adaptive data orthogonalization. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '78.*, volume 3, pages 109–112, Apr 1978. Cited on pages 28 and 33.

Gene H. Golub and Charles F. Van Loan. *Matrix computations (3rd ed.)*. Johns Hopkins University Press, Baltimore, MD, USA, 1996. ISBN 0-8018-5414-8. Cited on pages 28, 42, and 43.

P. Strobach. Low-rank adaptive filters. *Signal Processing, IEEE Transactions on*, 44(12):2932–2947, Dec 1996. ISSN 1053-587X. doi: 10.1109/78.553469. Cited on pages 28 and 33.

Peter Strobach. The fast householder bi-svd subspace tracking algorithm. *Signal Processing*, 88(11):2651–2661, 2008. ISSN 0165-1684. doi: DOI:10.1016/j.sigpro.2008.05.004. URL `http://www.sciencedirect.com/science/article/B6V18-4SJG65H-1/2/0555dc65a518eeda8ed4a067441d7422`. Cited on page 29.

Abbas Mohammed Ronnie Landqvist. The projection approximation subspace tracking algorithm applied to whitening and independent component analysis in wireless communications. Research Report 1, Blekinge Institute of Technology, 2005. Cited on page 30.

E. Oja, J. Karhunen, L. Wang, and R. Vigario. Principal and independent components in neural networks - recent developments. In *In Proc. VII Italian Workshop on Neural Nets*, pages 16–35, 1995. Cited on page 30.

Juyang Weng, Yilu Zhang, and Wey-shiuan Hwang. Candid covariance-free incremental principal component analysis. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 25:1034–1040, 2003. Cited on pages 30 and 70.

R. Badeau, G. Richard, and B. David. Fast and stable yast algorithm for principal and minor subspace tracking. *IEEE Transactions on Signal Processing*, 56:3437–3446, aug 2008. doi: 10.1109/TSP.2008.925924. Cited on pages 33 and 36.

R. Badeau, B. David, and G. Richard. Fast approximated power iteration subspace tracking. *Signal Processing, IEEE Transactions on*, 53(8):2931–2941, Aug. 2005. ISSN 1053-587X. doi: 10.1109/TSP.2005.850378. Cited on page 33.

Y. Miao and Y. Hua. Fast subspace tracking and neural network learning by a novel information criterion. *Signal Processing, IEEE Transactions on*, 46(7): 1967–1979, Jul 1998. ISSN 1053-587X. doi: 10.1109/78.700968. Cited on page 33.

J. F. Yang and M. Kaveh. Adaptive eigensubspace algorithms for direction or frequency estimation and tracking. *IEEE Transactions on Acoustics Speech and Signal Processing*, 36:241–251, feb 1988. Cited on page 33.

Simon Haykin. *Adaptive filter theory (3rd ed.)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1996. ISBN 0-13-322760-X. Cited on pages 33 and 64.

Lu Yang, S. Attallah, G. Mathew, and K. Abed-Meraim. Analysis of orthogonality error propagation for frans and hfrans algorithms. *Signal Processing, IEEE Transactions on*, 56(9):4515–4521, Sept. 2008. ISSN 1053-587X. doi: 10.1109/TSP.2008.923199. Cited on page 33.

S. Attallah. Adaptive algorithms for noise subspace estimation with application to mc-cdma. In *Signal Processing and Its Applications, 2003. Proceedings. Seventh International Symposium on*, volume 1, pages 501–504, July 2003. doi: 10.1109/ISSPA.2003.1224749. Cited on page 33.

C. E. Davila. Efficient, high performance, subspace tracking for time-domain data. *Signal Processing, IEEE Transactions on*, 48(12):3307–3315, Dec 2000. ISSN 1053-587X. doi: 10.1109/78.886994. Cited on page 34.

S. C. Douglas. Numerically-robust adaptive subspace tracking using householder transformations. In *Sensor Array and Multichannel Signal Processing Workshop. 2000. Proceedings of the 2000 IEEE*, pages 499–503, 2000. doi: 10.1109/SAM.2000.878059. Cited on pages 34 and 93.

Bin Yang. An extension of the pastd algorithm to both rank and subspace tracking. *Signal Processing Letters, IEEE*, 2(9):179–182, Sep 1995b. ISSN 1070-9908. doi: 10.1109/97.410547. Cited on pages 35 and 40.

A. Kavcic and Bin Yang. Adaptive rank estimation for spherical subspace trackers. *Signal Processing, IEEE Transactions on*, 44(6):1573–1579, Jun 1996. ISSN 1053-587X. doi: 10.1109/78.506625. Cited on page 35.

B. Champagne, W. Kang, and H. C. Tam. Improved strategy for adaptive rank estimation with spherical subspace trackers. In *Electrical and Computer Engineering, 2003. IEEE CCECE 2003. Canadian Conference on*, volume 3, pages 2029–2034, May 2003. Cited on page 35.

Patrick O. Perry and Patrick J. Wolfe. Minimax rank estimation for subspace tracking. Submitted, January 2009. URL `http://www.citebase.org/abstract?id=oai:arXiv.org:0906.3090`. Cited on pages 35 and 80.

Art B. Owen and Patrick O. Perry. Bi-cross-validation of the svd and the nonnegative matrix factorization. *Annals of Applied Statistics*, 3(2):564–594, 2009. Cited on page 35.

Shira Kritchman and Boaz Nadler. Determining the number of components in a factor model from limited noisy data. *Chemometrics and Intelligent Laboratory Systems*, 94(1):19–32, 2008. ISSN 0169-7439. doi: DOI:10.1016/j.chemolab.2008.06.002. URL `http://www.sciencedirect.com/science/article/B6TFP-4SS1HWM-1/2/69f77105a564dcf7231452f50bd020e4`. Cited on page 35.

. Spiros Papadimitriou. *Parameter-free spatial and stream mining*. PhD thesis, Carnegie Mellon University, Pittsburgh, PA, USA, 2005. Adviser-Faloutsos,, Christos. Cited on pages 37, 51, 56, 58, 61, and 64.

. Jimeng Sun. *Incremental pattern discovery on streams, graphs and tensors*. PhD thesis, Carnegie Mellon University, Pittsburgh, PA, USA, 2007. Adviser-Faloutsos,, Christos. Cited on pages 37, 51, 56, and 58.

Don Coppersmith and Shmuel Winograd. Matrix multiplication via arithmetic progressions. *J. Symb. Comput.*, 9(3):251–280, 1990. ISSN 0747-7171. doi: http://dx.doi.org/10.1016/S0747-7171(08)80013-2. Cited on page 42.

Sara Robinson. Toward an optimal algorithm for matrix multiplication. *Society for Industrial and Applied Mathematics News*, 38(9), 2005. Cited on page 42.

David H. Bailey, King Lee, and Horst D. Simon. Using strassen's algorithm to accelerate the solution of linear systems. *J. Supercomput.*, 4(4):357–371, 1990. ISSN 0920-8542. doi: http://dx.doi.org/10.1007/BF00129836. Cited on page 42.

Phillip E. Gill, Gene H. Golub, Walter A. Murray, and Michael A. Saunders. Methods for modifying matrix factorizations. Technical report, Stanford University, Stanford, CA, USA, 1972. Cited on page 42.

Jaroslav Hajek. qrupdate: a fortran library for fast updates of qr and cholesky decompositions. http://sourceforge.net/projects/qrupdate/, 2009. Cited on pages 43, 44, and 51.

John M. Bennett. Triangular factors of modified matrices. *Numerische Mathematik*, 7(3):217–221, 1965. Cited on page 44.

Peter Stange, Andreas Griewank, and Matthias Bollhofer. On the efficient update of rectangular lu factorizations subject to low rank modifications. *Eletronic Transactions on Numerical Analysis*, 26:161–177, 2007. Cited on pages 44, 45, 46, and 57.

Gregor Hohpe and Bobby Woolf. *Enterprise Integration Patterns: Designing, Building, and Deploying Messaging Solutions*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2003. ISBN 0321200683. Cited on page 47.

Ricardo Clemente and Denis Vieira. Inteligent monitoring. `http://dist.codehaus.org/esper/FISL_IntelligentMonitoring.pdf`, July 2009. 10th International Free Software Forum (FISL). Cited on pages 50 and 79.

Ross Ihaka and Robert Gentleman. R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5(3):299–314, 1996. URL `http://www.amstat.org/publications/jcgs/`. Cited on page 51.

Clint Whaley Antoine, Antoine Petitet, and Jack J. Dongarra. Automated empirical optimization of software and the atlas project. *Parallel Computing*, 27:2001, 2000. Cited on page 51.

T Oetiker. Round robin database tool. http://oss.oetiker.ch/rrdtool/. Cited on page 55.

Peter Strobach. The householder compressor theorem and its application in subspace tracking. *Signal Processing*, 89(5):857–875, 2009b. ISSN 0165-1684. doi: DOI:10.1016/j.sigpro.2008.11.001. URL `http://www.sciencedirect.com/science/article/B6V18-4TW1130-6/2/ad6a65d4d8a33baebe403a6eb4bf6aaf`. Cited on pages 57 and 79.

Jesse Davis and Mark Goadrich. The relationship between precision-recall and roc curves. In *ICML '06: Proceedings of the 23rd international conference on Machine learning*, pages 233–240, New York, NY, USA, 2006. ACM. ISBN 1-59593-383-2. doi: http://doi.acm.org/10.1145/1143844.1143874. Cited on page 64.

Pedro Henriques Santos Teixeira and Ruy Luiz Milidiú. Data stream anomaly detection through principal subspace tracking. In *SAC 2010: Proceedings of the 2010 ACM symposium on Applied Computing*. ACM, To Appear To Appear. Cited on page 79.

Haifeng Chen, Guofei Jiang, and Kenji Yoshihira. Failure detection in large-scale internet services by principal subspace mapping. *IEEE Trans. on Knowl. and Data Eng.*, 19(10):1308–1320, 2007. ISSN 1041-4347. doi: http://dx.doi.org/10.1109/TKDE.2007.190633. Cited on page 80.

R. L. Milidiú, Machado R.J., and Rentería R.P. Time-series forecasting through wavelets transformation and a mixture of expert models. *Neurocomputing*, 28 (1):145–156, 1999. Cited on page 80.

Xiaohui Gu, Spiros Papadimitriou, Philip S. Yu, and Shu-Ping Chang. Toward predictive failure management for distributed stream processing systems. In *ICDCS '08: Proceedings of the 2008 The 28th International Conference on Distributed Computing Systems*, pages 825–832, Washington, DC, USA, 2008a. IEEE Computer Society. ISBN 978-0-7695-3172-4. doi: http://dx.doi.org/10.1109/ICDCS.2008.34. Cited on page 80.

Xiaohui Gu and Haixun Wang. Online anomaly prediction for robust cluster systems. In *ICDE '09: Proceedings of the 2009 IEEE International Conference on Data Engineering*, pages 1000–1011, Washington, DC, USA, 2009. IEEE Computer Society. ISBN 978-0-7695-3545-6. doi: http://dx.doi.org/10.1109/ICDE.2009.128. Cited on page 80.

Xiaohui Gu, Spiros Papadimitriou, Philip S. Yu, and Shu-Ping Chang. Online failure forecast for fault-tolerant data stream processing. In *ICDE '08: Proceedings of* *the 2008 IEEE 24th International Conference on Data Engineering*, pages 1388–1390, Washington, DC, USA, 2008b. IEEE Computer Society. ISBN 978-1-4244-1836-7. doi: http://dx.doi.org/10.1109/ICDE.2008.4497565. Cited on page 80.

Per Brinch Hansen. Householder reduction of linear equations. *ACM Comput. Surv.*, 24(2):185–194, 1992. ISSN 0360-0300. doi: http://doi.acm.org/10.1145/130844.130851. Cited on page 92.

A
Householder Reflection

For completeness, we provide additional details on how the parameters of the desired Householder transformation used in Section 3.3.4 are obtained. Alston S. Householder introduced the transformation in the mid 1950s as part of an improved way to solve *least-squares* problems. The basic problem is to design a Householder matrix $\boldsymbol{H}$ determined by the vector $\underline{\boldsymbol{v}}$ orthogonal to the reflection plane with following properties:

$$\boldsymbol{H} = \boldsymbol{H}^T \qquad \textit{symmetric} \qquad \text{(A-1)}$$

$$\boldsymbol{H}^{-1} = \boldsymbol{H} \qquad \textit{unitary} \qquad \text{(A-2)}$$

$$\boldsymbol{H}\boldsymbol{H} = \boldsymbol{I} \qquad \textit{involutary.} \qquad \text{(A-3)}$$

This is accomplished with the following expression:

$$\boldsymbol{H} = \boldsymbol{I} - 2\underline{\boldsymbol{v}}\underline{\boldsymbol{v}}^T \qquad \text{(A-4)}$$

where we consider the decomposition

$$\underline{\boldsymbol{v}} = \begin{bmatrix} \boldsymbol{v} \\ \varphi \end{bmatrix}, \qquad \text{(A-5)}$$

with $\|\underline{\boldsymbol{v}}\| = 1$. The essential property is that reflection of an n-dimensional vector preserves the norm $\|\boldsymbol{H}a\| = \|\boldsymbol{a}\|$. Hansen [1992] offers a geometric derivation for Equation (A-4) and we present it here to explain why it has this particular form. The reflection plane P is determined by the vector $\underline{\boldsymbol{v}}$ as shown in Figure A.1. The difference between a vector $\boldsymbol{a}$ and its reflection $\boldsymbol{b}$ is a vector $f\underline{\boldsymbol{v}}$

$$f\underline{\boldsymbol{v}} = \boldsymbol{a} - \boldsymbol{b} \qquad \text{(A-6)}$$

The reflection of the arbitrary vector $\boldsymbol{a}$ through a plane P satifies:

$$\begin{aligned} \|\boldsymbol{a}\|^2 &= \|\boldsymbol{b}\|^2 && \text{(A-7)} \\ &= (\boldsymbol{a} - f\underline{\boldsymbol{v}})^T(\boldsymbol{a} - f\underline{\boldsymbol{v}}) && \text{(A-8)} \\ &= \|\boldsymbol{a}\|^2 - 2f\underline{\boldsymbol{v}}^T\boldsymbol{a} + f^2 && \text{(A-9)} \end{aligned}$$

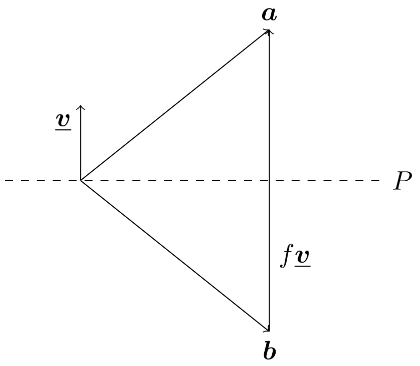

Figure A.1: Householder reflection in two-dimensional space

which determines the distance f between vector $\boldsymbol{a}$ and its image $\boldsymbol{b}$:

$$f = 2\underline{\boldsymbol{v}}^T\boldsymbol{a}. \tag{A-10}$$

The reflection of $\boldsymbol{b}$ into $\boldsymbol{a}$ displaces $\boldsymbol{b}$ by the same distance f in the opposite direction. So, the distance is also expressed as $f = -2\underline{\boldsymbol{v}}^T\boldsymbol{b}$. Therefore, we can define $\boldsymbol{b}$ in terms of $\boldsymbol{a}$:

$$\begin{aligned}\boldsymbol{b} &= \boldsymbol{a} - \underline{\boldsymbol{v}}f && \text{(A-11)}\\ &= \boldsymbol{I}\boldsymbol{a} - \underline{\boldsymbol{v}}(2\underline{\boldsymbol{v}}^T\boldsymbol{a}) && \text{(A-12)}\\ &= (\boldsymbol{I} - 2\underline{\boldsymbol{v}}\underline{\boldsymbol{v}}^T)\boldsymbol{a} && \text{(A-13)}\end{aligned}$$

hence, the expression for the Householder matrix:

$$b = \boldsymbol{H}a. \tag{A-14}$$

We reference [Douglas, 2000, Strobach, 2009a] for the applications of the row-Householder transformation. We are interested in choosing the plane that annihilates the bottom-row vector $\boldsymbol{x}^T$ in the problem of the kind

$$\begin{bmatrix}\boldsymbol{Y}\\ 0\dots0\end{bmatrix} = \boldsymbol{H}\begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{x}^T\end{bmatrix} \tag{A-15}$$

where both $\boldsymbol{X}$ and $\boldsymbol{Y}$ are square $r \times r$ matrices. The solution to this problem is $\boldsymbol{H}$ with a bottom-row vector that belongs to the nullspace of the $(r+1) \times r$ appended X-matrix. Therefore, the conditional equation that determines $\boldsymbol{v}$ is given by considering the bottom row of (A-15) and substituting $\boldsymbol{H}$ with (A-4):

$$\begin{aligned}\begin{bmatrix}0\dots0\end{bmatrix} &= (\begin{bmatrix}0\dots01\end{bmatrix} - 2\varphi\begin{bmatrix}\boldsymbol{v}^T\varphi\end{bmatrix})\begin{bmatrix}\boldsymbol{X}\\ \boldsymbol{x}^T\end{bmatrix} && \text{(A-16)}\\ &= \boldsymbol{x}^T - 2\varphi\boldsymbol{v}^T\boldsymbol{X} - 2\varphi^2\boldsymbol{x}^T. && \text{(A-17)}\end{aligned}$$

Rearranging the terms yields the following linear system of equations:

$$\boldsymbol{X}^T\boldsymbol{b} = \boldsymbol{x} \tag{A-18}$$

where

$$\boldsymbol{b} = \frac{2\varphi}{1-2\varphi^2}\boldsymbol{v}. \tag{A-19}$$

A remaining point is to determine φ, which follows from the unit-norm constraint of $\underline{\boldsymbol{v}}$:

$$\boldsymbol{v}^T\boldsymbol{v} + \varphi^2 = 1 \tag{A-20}$$

We can substitute (A-19) to write

$$\boldsymbol{v}^T\boldsymbol{v} = \frac{(1-2\varphi^2)^2}{4\varphi^2}\boldsymbol{b}^T\boldsymbol{b}. \tag{A-21}$$

Combining (A-20) and (A-21), we can eliminate $\boldsymbol{v}^T\boldsymbol{v}$:

$$\frac{(1-2\varphi^2)^2}{4\varphi^2}\boldsymbol{b}^T\boldsymbol{b} + \varphi^2 = 1 \tag{A-22}$$

or equivalently

$$(1-2\varphi^2)^2\boldsymbol{b}^T\boldsymbol{b} + 4\varphi^4 = 4\varphi^2. \tag{A-23}$$

The conditional equation has two solutions:

$$\varphi^2 = \frac{1}{2}\left(1 \pm \frac{1}{\sqrt{\boldsymbol{b}^T\boldsymbol{b}+1}}\right) \tag{A-24}$$

The solution with positive sign is arbitraly chosen, and this concludes the overall derivation for determining the parameters $\boldsymbol{v}$ and φ of the desired row-Householder reflector.

B

Annotated Anomalies

B.1

Abilene

The groud truth anomalies used for evaluation of the anomaly detection algorithms of both public datasets are summarized in Tables B.1 and B.1. These were extracted from the Matlab Abilene.mat file publicly available[1].

Start τ_i	**Duration ℓ_i**
647	2
708	3
777	4
1138	1
1100	2
1270	2
1456	6
1615	5
1723	4
1907	3
1952	3

Table B.1: 11 anomalies in the Abilene Packet dataset.

Start τ_i	**Duration ℓ_i**
708	2
862	6
1100	2
1160	2
1181	1
1384	1
1434	6
1546	3
1618	3
1625	1
1723	3
1869	1
1906	4
1968	1

Table B.2: 15 anomalies in the Abilene Flow dataset.

---

[1] From `http://www.tsp.ece.mcgill.ca/Networks/projects/monit-tarem-results.html`.