References for Kyoto University CoE Lectures Oct. 25-26, 2010
Matrix Computations, Information Retrieval & Search Engines

Books

(1) Belew, R. (2000) Finding Out About, Cambridge Univ. Press, Cambridge, UK.

This book gives is an entertaining and friendly introduction to vector space modeling of text-based documents in very large data bases. It can become very tiresome to read because of the long prose. I recommend reading as much of it until you grow tired of having to read too much English text. I recommend it for reading during long commutes on a train.

(2) Berry, M. and Browne, M. (2005) Understanding Search Engines, Mathematical Modeling and Text Retrieval (2nd ed.), SIAM, Philadelphia, PA, USA.

Most of material can be found in Berry's review paper (see refs in papers section) and in Belew's book (above). Since it's a SIAM book, the price is relatively low.

(3) Golub, G. and Van Loan, C. (1996) Matrix Computations (3rd ed.), Johns Hopkins Univ. Press, Baltimore MD, USA.

All graduate students in my generation in applied mathematics used this text in the United States. Recent editions present all pseudo-code for algorithms in a format that facilitates implementation in MatLab. I personally prefer the editions that use regular pseudo-code to describe algorithms, but I have never used MatLab. Warning: This book is more useful as a reference and is rather difficult for self-study.

(4) Horn, R. and Johnson, C. (1985) Matrix Analysis, Cambridge Univ. Press, Cambridge, UK.

This book is a very friendly introduction to matrices. It gives easy-to-follow examples. Unfortunately, it is not a particularly good reference book so if you already know most of the material, it's not worth buying.

(5) Jennings, A. and McKeown, J. (1992) Matrix Computation (2nd ed.), John Wiley & Sons, NY.

I personally like this book because it gives a lot of details and concrete examples. It is a good manual for self-study, especially if you plan to implement the algorithms yourself.

(6) Jolliffe, I. (1986) Principal Component Analysis. Springer, NY.

This book is more of a reference book than a textbook. If you want to understand the basic idea behind PCA, most engineering books give a better introduction. Try a search on Google, Bing or Yahoo (input query: PCA Principal Component Analysis) Wikipedia has some pages on PCA.

(7) Langville, A. and Meyer, C. (2006) Google's PageRank and Beyond: the Science of Search Engine Rankings, Princeton Univ. Press, Princeton, NJ.

Most of the material presented on the second day (Tues. Oct. 26) was from this book. It is a very easy read, and is entertaining (with less tedious prose than Belew's book). Japanese edition available. Very reasonable price since it's a CS-Engineering book which has a wider distribution than math books.

(8) Parlett, B. (1998) The Symmetric Eigenvalue Problem, SIAM, Philadelphia, PA.

This book was very, very expensive until the copyright was sold to SIAM. It is now a very good deal. Parlett writes very well.

References for Kyoto University CoE Lectures Oct. 25-26, 2010
Matrix Computations, Information Retrieval & Search Engines

(9) Press, W. et al. (1986) Numerical recipes: The Art of Scientific Computing. Cambridge Univ. Press, Cambridge, UK.

This book was originally published with FORTRAN code to help experimental physicists and engineers who wanted computer programs to help process data. The code for C is a translation of FORTRAN into C so it may be inefficient. The book has very nice tutorials on algorithms so even if you do not intend to use the code, the tutorials before each program can be useful for self-study. In particular, the SVD tutorial is nice.

(10) Salton, G. (editor) (1971) The SMART Retrieval System: Experiments in Automatic Document Processing. Prentice-Hall, Englewood Cliffs, NJ.

This book is not worth buying. You might want to take a look at it in your library just to see what researchers did many decades ago in information retrieval and search.

(11) Stewart, G. (2001) Matrix Algorithms, Vols. 1&2, SIAM, Philadelphia, PA.

These 2 books are a really good deal for the price. They contain a lot of advanced material that is presented in a more readavle format than Golub and Van Loan. They can also serve as valuable references.

(12) Strang, G. (1998) Introduction to Linear Algebra, 3rd ed., Wellesley-Cambridge Press, Wellesley, MA.

This is one of the more popular linear algebra books for undergraduate students in the Unites States. If you know most of the material, it is not worth buying since it is not intended to be a reference book.

(13) Trefethen, N. and Bau, D. (1997) Numerical Linear Algebra, SIAM, Philadelphia, PA.

Some of my friends in the High Performance Computing Center love this book and keep telling me to read it. I must confess that haven't read it in detail. It's definitely a good buy for the money.

(14) Watkins, D. (2010) Fundamentals of Matrix Computations, 3rd ed., Wiley, New York, NY, USA.

I like this book and used the 2nd edition for a course at the University of Electro- Communications. It is intended for an upper division students (3rd-4th year undergraduates) and has many examples and figures to show what's going on geometrically. For graduate students, it is a very good text for self-study (or review of concepts). I haven't seen the 3rd edition, which has a solid 5-star rating @amazon.com. The only drawback is it is a bit pricey.

Remarks:

(1) There is big discount on SIAM Books for all SIAM members so you might consider joining SIAM (especially while the Japanese yen has a highly favorable exchange rate). SIAM has big summer sales on books for its members, and you can usually get better discounts (and sometimes a free T-shirt or mug) when buying 3 or more books.

(2) If you plan to attend in a large, international conference in the near future, you might consider waiting to purchase books at the conference since most publishers who have a display booth offer 20-30% discounts for conference attendees.

References for Kyoto University CoE Lectures Oct. 25-26, 2010
Matrix Computations, Information Retrieval & Search Engines

Papers

(1) Berry M., Dumais S. and O'Brien G. (1995) Using linear algebra for intelligent information-retrieval. SIAM Review 1995; 37:573-95.

This is a good review paper that covers the mathematics from the lecture on the first day. It does not cover math used in PageRank, HITS and Salsa.

(2) Deerwester et al. (1990) Indexing by latent semantic analysis. Journal of the American for Society Information Science, 41(6), 391-407.

This paper was co-authored by the team at Bell Labs that worked on building computerized information retrieval systems described in lecture 1.

(3) Golub, G. and Kahan, W. (1965) Calculating the singular values and pseudo-inverse of a matrix, Journal of SIAM, Series B, Numerical Analysis **2** (2): 205–224.

This paper gives the algorithm used for converting a bi-diagonal rectangular matrix to diagonal form for the final step of the SVD.

(3) Kleinberg, J. (1999) Authoritative sources in a hyperlinked environment, Journal of the ACM 46(5): 604-632.

This paper introduces the HITS algorithm that was introduced on the second day (after PageRank)

(4) Stewart, G. W. (1993) On the early history of the singular value decomposition, SIAM Review **35** (4): 551–566.

Software

(1) SVDLIBC: http://tedlab.mit.edu/~dr/SVDLIBC/ (search terms: svdlibc, berry)

(2) SVDPACK: http://www.netlib.org/svdpack/ (search terms: berry svd)

(3) Numerical Recipes – there is a CD-ROM available for purchase for code in the book.

Webpages

(1) Jon Kleinberg: http://www.cs.cornell.edu/home/kleinber/ (search terms: kleinberg, cornell, hits)

(2) Michael Berry: http://www.cs.utk.edu/~berry/ (search terms: michael berry tennessee)

(3) Susuan Dumais: http://research.microsoft.com/en-us/um/people/sdumais/
(search terms: susan dumais)

(4) Google Dance: http://www.google-dance-tool.com/ (search terms: google dance)

(5) SALSA: http://www9.org/w9cdrom/175/175.html (search terms: SALSA, lepel, moran)