A

Appendix

In this appendix, we give the solution to the weighted BiNMF. Assume that each data point has the weight γ_i, the weighted sum of squared errors is:

$$\begin{aligned}
J &= \frac{1}{2}\sum_i \gamma_i (X_i - \mathbf{X}\mathbf{W}V_i^T)^T(X_i - \mathbf{X}\mathbf{W}V_i^T) \\
&= \frac{1}{2}\text{trace}((\mathbf{X} - \mathbf{X}\mathbf{W}\mathbf{V}^T)\mathbf{\Gamma}(\mathbf{X} - \mathbf{X}\mathbf{W}\mathbf{V}^T))^T \\
&= \frac{1}{2}\text{trace}((\mathbf{X}\mathbf{\Gamma}^{\mathbf{1/2}} - \mathbf{X}\mathbf{\Gamma}^{\mathbf{1/2}}\mathbf{W'}\mathbf{V'}^T)(\mathbf{X}\mathbf{\Gamma}^{\mathbf{1/2}} - \mathbf{X}\mathbf{\Gamma}^{\mathbf{1/2}}\mathbf{W'}\mathbf{V'}^T)^T) \\
&= \frac{1}{2}\text{trace}((\mathbf{X}\mathbf{\Gamma}^{\mathbf{1/2}} - \mathbf{X}\mathbf{\Gamma}^{\mathbf{1/2}}\mathbf{W'}\mathbf{V'}^T)^T(\mathbf{X}\mathbf{\Gamma}^{\mathbf{1/2}} - \mathbf{X}\mathbf{\Gamma}^{\mathbf{1/2}}\mathbf{W'}\mathbf{V'}^T)) \\
&= \frac{1}{2}\text{trace}((\mathbf{I} - \mathbf{W'}\mathbf{V'}^T)^T\mathbf{\Gamma}^{\mathbf{1/2}}\mathbf{K}\mathbf{\Gamma}^{\mathbf{1/2}}(\mathbf{I} - \mathbf{W'}\mathbf{V'}^T)) \\
&= \frac{1}{2}\text{trace}((\mathbf{I} - \mathbf{W'}\mathbf{V'}^T)^T\mathbf{K'}(\mathbf{I} - \mathbf{W'}\mathbf{V'}^T))\,, \qquad \text{(A.1)}
\end{aligned}$$

where $\mathbf{\Gamma}$ is the diagonal matrix with γ_i as its diagonal elements, $\mathbf{W'} = \mathbf{\Gamma}^{\mathbf{-1/2}}\mathbf{W}$, $\mathbf{V'} = \mathbf{\Gamma}^{\mathbf{1/2}}\mathbf{V}$ and $\mathbf{K'} = \mathbf{\Gamma}^{\mathbf{1/2}}\mathbf{K}\mathbf{\Gamma}^{\mathbf{1/2}}$.

Notice that the above equation has the same form as (3.40) in Section 3.3.2, so the same algorithm can be used to find the solution.

References

1. T. M. Mitchell, *Machine Learning*. McGraw Hill, 1997.
2. L. Wasserman, *All of Statistics – A Concise Course in Statistical Inference*. Springer, 1997.
3. F. Jelinek, *Statistical Methods for Speech Recognition*. The MIT Press, 1997.
4. L. R. Rabiner, "A tuorial on hidden markov models and selected applications in speech recognition," *Proceedings of IEEE*, vol. 77, no. 2, pp. 257–286, 1989.
5. C. R. Rao, "R. A. Fisher: The founder of modern statistics," *Statistical Science*, vol. 7, Feb. 1992.
6. V. Vapnik, *Estimation of Dependences Based on Empirical Data*. Berlin: Springer Verlag, 1982.
7. V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer Verlag, 1995.
8. Y. Gong, *Intelligent Image Databases – Towards Advanced Image Retrieval*. Boston: Kluwer Academic Publishers, 1997.
9. A. del Bimbo, *Visual Information Retrieval*. San Francisco: Morgan Kaufmann, 1999.
10. O. Margues and B. Furth, *Content-Based Image and Video Retrieval*. New York: Springer, 2002.
11. G. Golub and C. Loan, *Matrix Computations*. Baltimore: Johns-Hopkins, 2 ed., 1989.
12. A. Hyvarinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Networks*, vol. 13, pp. 411–430, 2000.
13. A. Hyvarinen, "New approximations of differential entropy for independent component analysis and projection pursuit," *Advances in Neural Information Processing Systems*, vol. 10, pp. 273–279, 1998.
14. S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
15. "Mnist database website." http://yann.lecun.com/exdb/mnist.

16. P. Willett, "Document clustering using an inverted file approach," *Journal of Information Science*, vol. 2, pp. 223–231, 1990.
17. W. Croft, "Clustering large files of documents using the single-link method," *Journal of the American Society of Information Science*, vol. 28, pp. 341–344, 1977.
18. P. Willett, "Recent trend in hierarchical document clustering: a critical review," *Information Processing and Management*, vol. 24, no. 5, pp. 577–597, 1988.
19. J. Y. Zien, M. D. F. Schlag, and P. K. Chan, "Multilevel spectral hypergraph partitioning with artibary vertex sizes," *IEEE Transactions on Computer-Aided Design*, vol. 18, pp. 1389–1399, sep 1999.
20. J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.
21. C. Ding, X. He, H. Zha, M. Gu, and H. D. Simon, "A min-max cut algorithm for graph partitioning and data clustering," in *Proceedings of IEEE ICDM 2001*, pp. 107–114, 2001.
22. H. Zha, C. Ding, M. Gu, X. He, and H. Simon, "Spectral relaxation for k-means clustering," in *Advances in Neural Information Processing Systems*, vol. 14, 2002.
23. A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Advances in Neural Information Processing Systems*, vol. 14, 2002.
24. W. Xu and Y. Gong, "Document clustering based on non-negative matrix factorization," in *Proceedings of ACM SIGIR 2003*, (Toronto, Canada), July 2003.
25. W. Xu and Y. Gong, "Document clustering by concept factorization," in *Proceedings of ACM SIGIR 2004*, (Sheffield, United Kingdom), July 2004.
26. D. P. Bertsekas, *Nonlinear Programming*. Belmont, Massachusetts: Athena Scientific, second ed., 1999.
27. D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in Neural Information Processing Systems*, vol. 13, pp. 556–562, 2001.
28. K.-R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf, "An introduction to kernel-based learning algorithms," *IEEE Transactions on Neural Networks*, vol. 12, no. 3, pp. 181–202, 2001.
29. F. Sha, L. K. Saul, and D. D. Lee, "Multiplicative updates for nonnegative quadratic programming in support vector machines," in *Advances in Neural Information Processing Systems*, vol. 14, 2002.
30. G. Salton and M. McGill, *Introduction to Modern Information Retrieval*. McGraw-Hill, 1983.
31. P. Bremaud, *Markov Chains, Gibbs Fields, Monte Carlo Simulation, and Queues*. Springer, 2001.

32. T. Lindvall, *lectures on the Coupling Method*. New York: Wiley, 1992.
33. F. R. Gantmacher, *Applications of the Theory of Matrices*. New York: Dover Publications, 2005.
34. R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.
35. N. Metropolis, M. Rosenbluth, A.W.Rosenbluth, A. Teller, and E. Teller, "Equations of state calculations by fast computing machines," *Journal of Chemistry and Physics*, vol. 21, pp. 1087–1091, 1953.
36. A. Barker, "Monte carlo calculations of the radial distribution functions for proton-electron plasma," *Australia Journal of Physics*, vol. 18, pp. 119–133, 1965.
37. G. Grimmett, "A theorem on random fields," *Bulletin of the London Mathematical Society*, pp. 81–84, 1973.
38. S. Kirkpatrick, C. Gelatt, and M. Vechhi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.
39. V. Cerny, "A thermodynamical approach to the travelling salesman problem: An efficient simulation algorithm," *Journal of Optimization Theory and Applications*, vol. 45, pp. 41–51, 1985.
40. M. Han, W. Xu, and Y. Gong, "Video object segmentation by motion-based sequential feature clustering," in *Proceedings of ACM Multimedia Conference*, (Santa Barbara, CA), pp. 773–781, Oct. 2006.
41. A. D. Jepson and M. Black, "Mixture models for optical flow computation," in *Proceedings of IEEE CVPR*, (New York, NY), June 1993.
42. S. Ayer and H. S. Sawhney, "Layered representation of motion video using robust maximum-likelihood estimation of mixture models and MDL encoding," in *Proceedings of ICCV*, (Washington, DC), June 1995.
43. T. Darrell and A. Pentland, "Cooperative robust estimation using layers of support," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 5, pp. 474–487, 1995.
44. P. H. Torr, R. Szeliski, and P. Anandan, "An integrated bayesian approach to layer extraction from image sequences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 3, pp. 297–303, 2001.
45. A. D. Jepson, D. J. Fleet, and M. J. Black, "A layered motion representation with occlusion and compact spatial support," in *Proceedings of ECCV*, (London, UK), pp. 692–706, May 2002.
46. E. Adelson and J. Wang, "Representing moving images with layers," in *IEEE Transactions on Image Processing*, vol. 3, pp. 625–638, Sept. 1994.
47. J. Shi and J. Malik, "Motion segmentation and tracking using normalized cuts," in *Proceedings of ICCV*, pp. 1154–1160, Jan. 1998.
48. N. Jojic and B. Frey, "Learning flexible sprites in video layers," in *Proceedings of IEEE CVPR*, (Maui, Hawaii), Dec. 2001.

49. J. Xiao and M. Shah, "Motion layer extraction in the presence of occlusion using graph cut," *IEEE Transactions on PAMI*, vol. 27, pp. 1644–1659, Oct. 2005.
50. S. Khan and M. Shah, "Object based segmentation of video using color motion and spatial information," in *Proceedings of IEEE CVPR*, (Maui, Hawaii), pp. 746–751, Dec. 2001.
51. Q. Ke and T. Kanade, "A subspace approach to layer extraction," in *Proceedings of IEEE CVPR*, (Maui, Hawaii), pp. 255–262, Dec. 2001.
52. Y. Wang and Q. Ji, "A dynamic conditional random field model for object segmentation in image sequences," in *Proceedings of IEEE CVPR*, (San Diego, CA), pp. 264–270, June 2005.
53. J. Wang and M. F. Cohen, "An iterative optimization approach for unified image segmentation and matting," in *Proceedings of ICCV*, (Beijing, China), pp. 936–943, 2005.
54. Y. Li, J. Sun, and H.-Y. Shum, "Video object cut and paste," *ACM Transactions on Graphics*, vol. 24, pp. 595–600, July 2005.
55. J. Sun, J. Jia, C.-K. Tang, and H.-Y. Sham, "Poisson matting," in *Proceedings of ACM SIGGRAPH*, (Los Angeles, CA), pp. 315–321, Aug. 2004.
56. N. Apostoloff and A. Fitzgibbon, "Bayesian video matting using learnt image priors," in *Proceedings of IEEE CVPR*, (Washington, DC), pp. 407–414, June 2004.
57. Y. Chuang, A. Agarwala, B. Curless, D. Salesin, and R. Szeliski, "Video matting of complex scenes," in *Proceedings of ACM SIGGRAPH*, (San Antonio, Texas), pp. 243–248, July 2002.
58. Y. Chuang, B. Curless, D. Salesin, and R. Szeliski, "A bayesian approach to digital matting," in *Proceedings of IEEE CVPR*, (Maui, Hawaii), pp. 264–271, Dec. 2001.
59. M. Ruzon and C. Tomasi, "Alpha estimation in natural images," in *Proceedings of IEEE CVPR*, (Hilton Head Island, South Carolina), pp. 18–25, June 2000.
60. H.-Y. Shum, J. Sun, S. Yamazaki, Y. Li, and C. keung Tang, "Pop-up light field: An interactive image-based modeling and rendering system," *ACM Transactions on Graphics*, vol. 23, pp. 143–162, Apr. 2004.
61. J. Canny, "A computational approach to edge detection," *IEEE Transacions on PAMI*, vol. 8, pp. 679–714, 1986.
62. B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *IJCAI81*, pp. 674–679, Apr. 1981.
63. P. Chang, M. Han, and Y. Gong, "Extract highlights from baseball game videos with hidden markov models," in *IEEE International Conference on Image processing (ICIP)*, (Rochester, NY), Sept. 2002.
64. F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 47, no. 2, 2002.

65. S. L. Lauritzen, *Graphical Models*. Oxford University Press, 1996.
66. M. I. J. (ed), *Learning in Graphical Models*. Kluwer Academic Publishers, 1998.
67. J. Yedidia, "An idiosyncratic journey beyond mean field theory," in *Advanced Mean Field Methods, Theory and Practice*, (MIT Press), pp. 21–36, 2001.
68. A. Berger, S. D. Pietra, and V. D. Pietra, "A maximum entropy approach to natural language processing," *Journal of Computational Linguistics*, vol. 22, 1996.
69. S. D. Pietra, V. D. Pietra, and J. Lafferty, "Inducing features of random fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, Apr. 1997.
70. D. Brown, "A note on approximations to discrete probability distributions," *Journal of Information and Control*, vol. 2, pp. 386–392, 1959.
71. J. N. Darroch and D. Ratcliff, "Generalized iterative scaling for log-linear models," *Annals of Mathematical Statistics*, vol. 43, pp. 1470–1480, 1972.
72. J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proceedings of International Conference on Machine Learning*, (Williams College, MA), June 2001.
73. C. Sutton, K. Rohanimanesh, and A. McCallum, "Dynamic conditional random fields: Factorized probabilistic models for labeling and segmenting sequence data," in *Proceedings of International Conference on Machine Learning*, (Banff, Canada), July 2004.
74. Y. Gong, M. Han, W. Hua, and W. Xu, "Maximum entropy model-based baseball highlight detection and classification," *International Journal on Computer Vision and Image Understanding*, vol. 96, pp. 181–199, 2004.
75. Y. T. Tse and R. L. Baker, "Camera zoom/pan estimation and compensation for video compression," *Image Processing Algorithms and Techniques*, vol. 1452, pp. 468–479, 1991.
76. R. Szeliski and H. Shum, "Creating full view panoramic image mosaics and texture-mapped models," in *SIGGRAPH97*, pp. 251–258, 1997.
77. M. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *CACM*, vol. 24, pp. 381–395, June 1981.
78. A. Dempster, N. Laird, and D. Rubin, "Maximum-likelihood from incomplete data via em algorithm," *Royal Statistics Society Series B*, vol. 39, 1977.
79. G. Schwarts, "Estimating the dimension of a model," *Annals of Statistics*, vol. 6, pp. 461–464, 1990.
80. C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning Journal*, vol. 20, no. 3, pp. 273–297, 1999.
81. B. Taskar, C. Guestrin, and D. Koller, "Max-margin markov networks," in *Neural Information Processing Systems Conference (NIPS)*, (Vancouver, Canada), Dec. 2003.

82. C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery Journal*, vol. 2, no. 2, pp. 121–167, 1998.
83. T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer, 2001.
84. V. Vapnik, *Estimation of Dependences Based on Empirical data*. Springer-Verlag, 1982.
85. E. Osuna, R. Freund, and F. Girosi, "Improved training algorithm for support vector machines," in *Proceedings of IEEE NNSP*, (Amelia Island, FL), Sept. 1997.
86. J. C. Platt, "Sequential minimal optimization: A fast algorithm for training support vector machines," in *Technical Report 98-14, Microsoft Research*, (Redmond, WA), 1998.

Index