Chapter 7

Conclusion

The three pillars of the thesis are the theoretical results described in Chapter 4, the polygonal line algorithm proposed in Chapter 5, and the experimental results with the principal graph algorithm presented in Chapter 6. Below we summarize our main results in these three areas, and briefly discuss some of the possible areas of future research.

Theory

We proposed a new definition of principal curves with a length constraint. Based on the new definition, we proved the following two results.

- **Existence of principal curves.** Principal curves in the new sense exist for all distributions with final second moments.

- **Consistency and rate of convergence.** For distributions concentrated on a bounded and closed convex set, an estimator of the principal curve can be constructed based on a data set of n i.i.d. sample points such that the expected loss of the estimator converges to the loss of the principal curve at a rate of $n^{-1/3}$.

Two interesting open problems are the following.

- **Concrete principal curves.** It would be of both theoretical and practical interest if concrete examples of principal curves of basic multivariate densities could be found.

- **A more practical constraint.** It would be convenient to replace the length constraint with a practically more suitable restriction, such as a limit on the maximum curvature of the curve, that is more closely related to the curvature penalty applied in the practical algorithm.

Algorithm

Our main result here is a practical algorithm to estimate principal curves based on data. Experimental results on simulated data demonstrate that the polygonal line algorithm compares favorably to previous methods both in terms of performance and computational complexity.

A possible area of further research is to extend the polygonal line algorithm to find multi-dimensional manifolds. There are two fundamentally different approaches to extend principal curves to principal surfaces or to arbitrary-dimensional principal manifolds. In the first approach, the theoretical model and the algorithm are either extended to include smooth non-parametric surfaces [Has84, HS89], or they can be used, without modification, to find arbitrary-dimensional principal manifolds [SMS98, SWS98]. The second approach follows the strategy of an iterative PCA algorithm which finds the ith largest principal component by finding the first principal component in the linear subspace orthogonal to the first $i-1$ principal components. In the second approach, therefore, the one-dimensional principal curve routine is called iteratively so that in ith iteration, we compute the principal curve of the data set obtained by subtracting from the data points their projections to the principal curve computed in the $(i-1)$th iteration [Del98, CG98b, DM95].

Theoretically, it is not impossible to use the first approach, i.e., to extend the polygonal line algorithm to find arbitrary-dimensional piecewise linear manifolds. Technically, however, it is not clear at this point how this extension could be done. The second approach, on the other hand, seems feasible to be implemented with the polygonal line algorithm. The exact design of the algorithm is subject of future research.

Applications

We proposed an extended version of the polygonal line algorithm to find principal graphs of data sets obtained from binary templates of black-and-white images. Test results indicate that the principal graph algorithm can be used to find a smooth medial axis of a wide variety of character templates, and to represent hand-written text efficiently.

Here, the main objective of future research is to improve the computational complexity of the method. At this point the “general purpose” polygonal line algorithm is used as the “main engine” for the principal graph algorithm. We expect that by incorporating the special features of the highly structured data obtained from binary templates into the algorithm, the efficiency of the algorithm can be increased substantially.

Bibliography

[AH69] T. M. Alcorn and C. W. Hoggar. Preprocessing of data for character recognition. *Marconi Review*, 32:61–81, 1969.

[Ale84] K. Alexander. Probability inequalities for empirical processes and a law of the iterated logarithm. *Annals of Probability*, pages 1041–1067, 1984.

[Ash72] R. B. Ash. *Real Analysis and Probability*. Academic Press, New York, 1972.

[Bar87] D. J. Bartholomew. *Latent Variable Models and Factor Analysis*. Charles Griffin & Co. Ltd., London, 1987.

[BCJL94] P. V. Balakrishnan, M. C. Cooper, V. S. Jacob, and P. A. Lewis. A study of the classification capabilities of neural networks using unsupervised learning: a comparison with k-means clustering. *Psychometrika*, 59(4):509–525, 1994.

[BD75] J. Bezdek and J. Dunn. Optimal fuzzy partitions: A heuristic for estimating the parameters in a mixture of normal distributions. *IEEE Transactions on Computers*, 24(4):835–838, 1975.

[BDHW97] G. Balzuweit, R. Der, M. Herrmann, and M. Welk. An algorithm for generalized principal curves with adaptive topology in complex data sets. Technical Report 3/97, Institut für Informatik, Universität Leipzig, 1997.

[BLL98] P. Bartlett, T. Linder, and G. Lugosi. The minimax distortion redundancy in empirical quantizer design. *IEEE Transactions on Information Theory*, 44(5):1802–1813, 1998.

[BP95] J. C. Bezdek and N. R. Pal. An index of topological preservation for feature extraction. *Pattern Recognition*, 28(3):381–391, 1995.

[BR92] J. D. Banfield and A. E. Raftery. Ice floe identification in satellite images using mathematical morphology and clustering about principal curves. *Journal of the American Statistical Association*, 87:7–16, 1992.

[BSW96] C. M. Bishop, M. Svensén, and C. K. I. Williams. EM optimization of latent-variables density models. In David S. Touretzky, Michael C. Mozer, and Michael E. Hasselmo, editors, *Advances in Neural Information Processing Systems*, volume 8, pages 465–471. The MIT Press, 1996.

[BSW98] C. M. Bishop, M. Svensén, and C. K. I. Williams. GTM: The generative topographic mapping. *Neural Computation*, 10(1):215–235, 1998.

[BT98] C. M. Bishop and M. E. Tipping. A hierarchical latent variable model for data visualization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):281–293, 1998.

[CG92] G. Celeux and G. Govaert. A classification EM algorithm and two stochastic versions. *Computational Statistics and Data Analysis*, 14:315–332, 1992.

[CG98a] K. Chang and J. Ghosh. Principal curve classifier – a nonlinear approach to pattern classification. In *IEEE International Joint Conference on Neural Networks*, pages 695–700, Anchorage, AL, May 5–9 1998.

[CG98b] K. Chang and J. Ghosh. Principal curves for nonlinear feature extraction and classification. In *Applications of Artificial Neural Networks in Image Processing III*, volume 3307, pages 120–129, San Jose, CA, Jan 24–30 1998. SPIE Photonics West '98 Electronic Image Conference.

[Cho94] P. A. Chou. The distortion of vector quantizers trained on n vectors decreases to the optimum as $o_p(1/n)$. In *Proceedings of IEEE International Symposium on Information Theory*, Trondheim, Norway, 1994.

[Cle79] W. S. Cleveland. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74:829–835, 1979.

[DBH96] R. Der, G. Balzuweit, and M. Herrmann. Constructing principal manifolds in sparse data sets by self-organizing maps with self-regulating neighborhood width. In *Proceedings of the International Conference on Neural Networks*, pages 480–483, 1996.

[Del98] P. Delicado. Principal curves and principal oriented points. Technical Report 309, Department d'Economia i Empresa, Universitat Pompeu Fabra, 1998.

[Deu68] E. S. Deutsch. Preprocessing for character recognition. In *Proceedings of the IEE NPL Conference on Pattern Recognition*, pages 179–190, 1968.

[DGL96] L. Devroye, L. Györfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition*. Springer, New York, 1996.

[Din55] G. P. Dinnen. Programming pattern recognition. In *Proceedings of the Western Joint Computer Conference*, pages 94–100, New York, 1955.

[DK82] P. A. Devijver and J. Kittler. *Pattern Recognition: a Statistical Approach*. Prentice Hall, Englewood Cliffs, New Jersey, 1982.

[DLR77] A.P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B*, 39:1–38, 1977.

[DM95] D. Dong and T. J. McAvoy. Nonlinear principal component analysis – based on principal curves and neural networks. *Computers Chem. Engineering*, 20(1):65–78, 1995.

[DP97] A. Datta and S. K. Parui. Skeletons from dot patterns: A neural network approach. *Pattern Recognition Letters*, 18:335–342, 1997.

[DS96a] T. Duchamp and W. Stuetzle. Extremal properties of principal curves in the plane. *Annals of Statistics*, 24(4):1511–1520, 1996.

[DS96b] T. Duchamp and W. Stuetzle. Geometric properties of principal curves in the plane. In Helmut Rieder, editor, *Robust statistics, data analysis, and computer intensive methods: in honor of Peter Huber's 60th birthday*, volume 109 of *Lecture notes in statistics*, pages 135–152. Springer-Verlag, 1996.

[EOS92] E. Erwin, K. Obermayer, and K. Schulten. Self-organizing maps: ordering, convergence properties and energy functions. *Biological Cybernetics*, 67:47–55, 1992.

[Eve84] B. S. Everitt. *An Introduction to Latent Variable Models*. Chapman and Hall, London, 1984.

[Fle97] A. Flexer. Limitations of self-organizing maps for vector quantization and multidimensional scaling. In M. C. Mozer, M. I. Jordan, and T. Petsche, editors, *Advances in Neural Information Processing Systems: Proceedings of the 1996 Conference*, volume 9, pages 445–451. MIT Press, 1997.

[Fle99] A Flexer. On the use of self-organizing maps for clustering and visualization. In J.M. Zytkow and J. Rauch, editors, *Principles of Data Mining and Knowledge Discovery, Third European Conference, PKDD'99*, Lecture Notes in Artificial Intelligence 1704, pages 80–88, Prague, Czech Republic, 1999. Springer.

[Föl89] P. Földiák. Adaptive network for optimal linear feature extraction. In IEEE Press, editor, *Proceedings of the IEEE/INNS International Joint Conference on Neural Networks*, volume 1, pages 401–405, New York, 1989.

[GG92] A. Gersho and R. M. Gray. *Vector Quantization and Signal Compression*. Kluwer, Boston, 1992.

[GKL80] R. M. Gray, J. C. Kieffer, and Y. Linde. Locally optimal block quantizer design. *Information and Control*, 45:178–198, May 1980.

[Gro95] P. Grother. *NIST Special Database 19*. National Institute of Standards and Technology, Advanced Systems Division, 1995.

[Har75] J. A. Hartigan. *Clustering Algorithms*. Wiley, New York, 1975.

[Has84] T. Hastie. *Principal curves and surfaces*. PhD thesis, Stanford University, 1984.

[Hoe63] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.

[Hot33] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24:417–441, 498–520, 1933.

[HS89] T. Hastie and W. Stuetzle. Principal curves. *Journal of the American Statistical Association*, 84:502–516, 1989.

[JD88] A. K. Jain and R. C. Dubes. *Algorithms for Clustering Data*. Prentice Hall, Englewood Cliffs, New Jersey, 1988.

[JW92] R. A. Johnson and D. W. Wichern. *Applied Multivariate Statistical Analysis*. Prentice Hall, Englewood Cliffs, New Jersey, 1992.

[Kar47] K. Karhunen. Über lineare methoden in der wahrscheinlichkeitsrechnung. *Amer. Acad. Sci., Fennicade, Ser. A, I*, 37:3–79, 1947. (Translation: RAND Corporation, Santa Monica, California, Rep. T-131, Aug. 1960).

[KCRU57] R. A. Kirsh, L. Cahn, C. Ray, and G. J. Urban. Experiment in processing pictorial information with a digital computer. In *Proceedings of the Eastern Joint Computer Conference*, pages 221–229, New York, 1957.

[KKLZ] B. Kégl, A. Krzyżak, T. Linder, and K. Zeger. Principal curves: Learning and convergence. In *Proceedings of IEEE International Symposium on Information Theory*, page 387, Cambridge, Mass., Aug. 1998.

[KKLZ99] B. Kégl, A. Krzyżak, T. Linder, and K. Zeger. A polygonal line algorithm for constructing principal curves. In *Advances in Neural Information Processing Systems*, volume 11, pages 501–507, Denver, Col., Nov. 1998, 1999. The MIT Press.

[KKLZ00] B. Kégl, A. Krzyżak, T. Linder, and K. Zeger. Learning and design of principal curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(3):281–297, 2000.

[Koh82] T. Kohonen. Clustering, taxonomy, and topological maps of patterns. In *Proceedings of the 6th International Conference on Pattern Recognition*, pages 114–128, Munich, 1982.

[Koh97] T. Kohonen. *The Self-Organizing Map*. Springer-Verlag, 2nd edition, 1997.

[KR95] R. E. Kass and A. E. Raftery. Bayes factors. *Journal of the American Statistical Association*, pages 773–795, 1995.

[Kra91] M.A. Kramer. Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal*, 37(2):233–243, 1991.

[KT61] A. N. Kolmogorov and V. M. Tikhomirov. ε-entropy and ε-capacity of sets in function spaces. *Translations of the American Mathematical Society*, 17:277–364, 1961.

[KW78] J. B. Kruskal and M. Wish. Multidimensional scaling. In *Sage University Paper Series on Quantitative Applications in the Social Sciences*, number 07-011. Sage Publications, Beverly Hills, California, 1978.

[LBG80] Y. Linde, A. Buzo, and R. M. Gray. An algorithm for vector quantizer design. *IEEE Transactions on Communications*, COM-28(1):84–95, 1980.

[LLS93] S. W. Lee, L. Lam, and C.Y. Suen. A systematic evaluation of skeletonization algorithms. *International Journal of Pattern Recognition and Artificial Intelligence*, 7(5):1203–1225, 1993.

[LLZ94] T. Linder, G. Lugosi, and K. Zeger. Rates of convergence in the source coding theorem, in empirical quantizer design and in universal lossy source coding. *IEEE Transactions on Information Theory*, 40:1728–1740, 1994.

[Lut90] S. P. Luttrell. Derivation of a class of training algorithms. *IEEE Transactions on Neural Networks*, 1(2):229–232, 1990.

[Mac67] J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematics, Statistics and Probability*, pages 281–296, 1967.

[Mac95] D. J. C. MacKay. Bayesian neural networks and density networks. *Nuclear Instruments and Methods in Physics Research*, 354(1):73–80, 1995.

[MAG91] S. Mahmoud, I. Abuhaiba, and R. Green. Skeletonization of arabic characters using clustering based skeletonization algorithm (CBSA). *Pattern Recognition*, 24(5):453–464, 1991.

[MC95] F. Mulier and V. Cherkassky. Self-organization as an iterative kernel smoothing process. *Neural Computation*, 7:1165–1177, 1995.

[MMT95] E. C. Malthouse, R. H. S. Mah, and A. C. Tamhane. Some theoretical results on nonlinear principal component analysis. In *Proceedings of the American Control Conference*, pages 744–748, June 1995.

[MZ97] N. Merhav and J. Ziv. On the amount of side information required for lossy data compression. *IEEE Transactions on Information Theory*, 43:1112–1121, 1997.

[NS84] N. J. Naccache and R. Shingal. SPTA: A proposed algorithm for thinning binary patterns. *IEEE Transactions on Systems, Man and Cybernetics*, 14(3), 1984.

[Oja92] E. Oja. Principal components, minor components, and linear neural networks. *Neural Networks*, 5:927–935, 1992.

[O'N66] B. O'Neil. *Elementary Differential Geometry*. Academic Press, Inc., Orlando, Florida, 1966.

[Pav80] T. Pavlidis. A thinning algorithm for discrete binary images. *Computer Graphics and Image Processing*, 13(2):142–157, 1980.

[Pea01] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philos. Mag.*, 6(2):559–572, 1901.

[Pol81] D. Pollard. Strong consistency of k-means clustering. *Annals of Statistics*, 9:135–140, 1981.

[Pol82] D. Pollard. A central limit theorem for k-means clustering. *Annals of Probability*, 10:919–926, 1982.

[RMS92] H. Ritter, T. Martinetz, and K. Schulten. *Neural Computation and Self-Organizing Maps: An Introduction*. Addison-Wesley, Reading, Massachusetts, 1992.

[RN98] K. Reinhard and M. Niranjan. Subspace models for speech transitions using principal curves. *Proceedings of Institute of Acoustics*, 20(6):53–60, 1998.

[Row98] S. Roweis. EM algorithms for PCA and SPCA. In *Advances in Neural Information Processing Systems*, volume 10, pages 626–632. The MIT Press, 1998.

[RT89] J. Rubner and P. Tavan. A self-organizing network for principal component analysis. *Europhysics Letters*, 10:693–698, 1989.

[SA86] S. Suzuki and K. Abe. Sequential thinning of binary pictures using distance transformation. In *Proceedings of the 8th International Conference on Pattern Recognition*, pages 289–292, 1986.

[Sam69] J. W. Sammon. A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers*, C-18:401–409, 1969.

[Sil85] B. W. Silverman. Some aspects of spline smoothing approaches to non-parametric regression curve fitting. *Journal of the Royal Statistical Society*, Ser. B, 47:1–52, 1985.

[SMS98] A. J. Smola, S. Mika, and B. Schölkopf. Quantization functionals and regularized principal manifolds. Technical Report NC2-TR-1998-028, NeuroCOLT2 Technical Report Series, 1998.
`http://www.neurocolt.com/abstracts/contents_1998.html`.

[SR00] D. Stanford and A. E. Raftery. Principal curve clustering with noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000. to appear.

[SWP98] R. Singh, M. C. Wade, and N. P. Papanikolopoulos. Letter-level shape description by skeletonization in faded documents. In *Proceedings of the Fourth IEEE Workshop on Applications of Computer Vision*, pages 121–126. IEEE Comput. Soc. Press, 1998.

[SWS98] A. J. Smola, R. C. Williamson, and B. Schölkopf. Generalization bounds and learning rates for regularized principal manifolds. Technical Report NC2-TR-1998-027, NeuroCOLT2 Technical Report Series, 1998.
`http://www.neurocolt.com/abstracts/contents_1998.html`.

[Tal94] M. Talagrand. Sharper bounds for gaussian and empirical processes. *Annals of Probability*, 22:28–76, 1994.

[TB99] M. E. Tipping and C. M. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society, Series B*, 61(3):611–622, 1999.

[Tib92] R. Tibshirani. Principal curves revisited. *Statistics and Computation*, 2:183–190, 1992.

[TLF95] T. Tarpey, L. Li, and B. D. Flury. Principal points and self-consistent points of elliptical distributions. *Annals of Statistics*, 23(1):103–112, 1995.

[Vap98] V. N. Vapnik. *Statistical Learning Theory*. Wiley, New York, 1998.

[VT68] H. L. Van Trees. *Detection, Estimation, and Modulation Theory*. Wiley, New York, 1968.

[Wil65] J. H. Wilkinson. *The algebraic eigenvalue problem*. Claredon Press, Oxford, England, 1965.

[WKIM98] N. G. Waller, H. A. Kaiser, J. B. Illian, and M. Manry. A comparison of the classification capabilities of the 1-dimensional Kohonen neural network with two partitioning and three hierarchical cluster analysis algorithms. *Psychometrika*, 63(1):5–22, 1998.

[YMMS92] N. Yamashita, M. Minami, M. Mizuta, and Y. Sato. A refined algorithm of principal curves and the evaluation of its complexity. *Bulletin of the Computational Statistics of Japan*, 5(1):33–43, 1992.