

## CHAPTER 6

# CONCLUSIONS AND FUTURE WORK

In the introduction, I expressed the hope that the work in this thesis could be a ‘first step’ towards a general purpose inference system. In this final chapter, I will conclude by describing the progress made towards this goal in terms of my development of a variational inference framework and its application to problems in a range of domains. I will also suggest some future research directions that could provide the next steps along the path to a practical and widely applicable inference system.

### 6.1 Conclusions

The aim of this thesis has been to develop a framework for automatically performing variational inference. In Chapter 1, I explained that such automatic inference systems will meet the increasing need for machine-based reasoning in a range of applications. I demonstrated the importance of modelling uncertainty in such systems and showed how probabilistic models and Bayesian inference provide both the representation and the method for a powerful approach to automated learning. Belief propagation, an algorithm for performing exact Bayesian inference, was used as an example of how message passing in a graph allows the use of general-purpose inference algorithms. Unfortunately, exact inference is intractable in many models, leading to the requirement for approximate inference methods. I focussed on one such technique, variational inference, as a deterministic method which gives an analytic approximation to the posterior and has been shown to be successful in a wide range of models. In variational inference, an approximating distribution  $Q$  is selected which has a simpler form than the true posterior  $P$ . This variational distribution is then recursively adapted to be as close to the true posterior as possible, in terms of KL divergence. On convergence,  $Q$  provides a functional approximation to  $P$ , whose quality depends on the data set and initialisation.

One problem with variational inference is that implementing it for new models is extremely time consuming and prone to errors. In Chapter 2, I described a general purpose framework which can apply variational inference automatically to a large class of probabilistic models and so avoid these implementation problems. The framework uses message-passing in a graphical model, leading to the algorithm being called Variational Message Passing (VMP). In VMP,

the  $Q$  distribution is fully-factorised and hence the approximate posterior is separable with respect to all variables. I showed that the class of models that VMP supports includes all conjugate-exponential models (discrete, continuous and mixed) and it allows for both mixtures and deterministic relationships between variables. I also investigated some extensions to VMP that allow non-conjugate hyper-priors and the option of finding Maximum A Posteriori values for some variables.

The VMP framework has been implemented in a software package called VIBES (Variational Inference in BayESian networks). This Java software allows the model to be specified graphically, by simply drawing the Bayesian network and assigning conditional probabilities through a graphical user interface. Variational inference can then be applied automatically. During inference, individual variables can be monitored by a range of methods and the lower bound on the evidence automatically calculated and displayed. Experience with VIBES has shown that it dramatically simplifies the construction and testing of new variational models and readily allows a range of alternative models to be evaluated on a given problem. This has been illustrated by applying VIBES to problems in the domains of machine vision and of DNA microarrays.

In Chapter 3, the problem of modelling nonlinear image subspaces was addressed using the VMP framework and VIBES. A mixture of Bayesian Principal Component Analysis models was developed and shown to be able to model the nonlinear manifolds of sets of digit and face images, as well as segments of a natural image. By also implementing variational inference by hand, the significant time and effort saving of using VIBES was demonstrated. The resultant system was shown to be able to determine manifold shape and dimensionality automatically from the data, avoiding the need for cross-validation.

The new technology of DNA microarrays has raised a range of inference problems, two of which were addressed in Chapter 4. The main problem investigated was that of analysing the scanned images of DNA microarrays in order to determine the corresponding gene expression levels. A model of the image process was developed with a latent variable for the intensity distribution of each spot (corresponding to the expression of a single gene) which, combined with the spot locations and shapes, led to the observed image. As this model was not conjugate-exponential, the VMP algorithm was extended to allow importance sampling in the non-exponential parts of the graph. I showed that this extended algorithm allowed the model to be used to analyse images successfully. I also introduced models for missing and obscured spots which allowed these to be identified and handled separately. The second problem of how to analyse the resultant gene expression data was investigated and VMP was used to perform independent component analysis on a small gene expression data set. The aim of this experiment was to motivate the development of complex models for gene expression data (for example, based on knowledge of the latent biological processes) by using automatic inference systems, such as VIBES, to construct and evaluate each model.

Finally, in Chapter 5, I investigated the use of  $Q$  distributions which are not fully factorised and hence retain some dependencies between variables in the posterior approximation.

The use of such a structured  $Q$  distribution guarantees at least as good an approximation as a fully factorised distribution and, in general, a better one as it can capture posterior correlations between variables. The particular form of  $Q$  distribution investigated was a junction tree (or set of trees) because it allows local computation of marginals – essential for the development of local, message passing algorithms. I demonstrated how inference can be performed efficiently using such variational junction trees by creating an extended form of VMP called Structured VMP (SVMP). VIBES was extended to include a partial implementation of SVMP and applied to show the advantages of SVMP over VMP in a small Hidden Markov Model.

The use of inference frameworks, such as VIBES, separates the design of a probabilistic model from the process of performing inference within the model. This separation permits the construction and comparison of domain models to be left to the domain expert without their needing to understand the learning and inference process in detail. Equally, it lets the machine learning researcher concentrate on improving inference methods, since the comparative performance of different methods can be readily assessed using not only the same data set but also the same probabilistic model.

## 6.2 Suggestions for Future Work

A number of open problems must be solved to allow the development of a truly general purpose learning and inference system. These problems suggest a variety of research directions that need to be pursued to make such a system feasible.

One such direction would be to investigate allowing automatic learning of the structure of the probabilistic model. The current framework requires that the model be specified explicitly. It would be preferable that an initial model be suggested and the framework allowed to adapt or extend it so as to best fit the data. Bayesian model selection provides a mechanism for choosing between models and so learning one for a particular data set or domain.

Another possibility would be to create hybrid sampling/variational inference systems which would allow models outside the conjugate-exponential family. This idea was touched upon in Chapter 2 and led to the use of VMP with Importance Sampling in Chapter 4, and is one that would certainly merit further investigation. It would also be very valuable to perform a comparison of sampling methods (such as Markov chain Monte Carlo) and variational methods (with structured and unstructured variational distributions). The aim would be to compare speed and accuracy of approximation for identical models on real and toy data sets.

Expectation propagation provides an alternative to variational methods for Bayesian inference. As there are many similarities between the two, it seems likely that VIBES could be modified to use expectation propagation instead of, or in combination with, VMP. The resulting framework would allow comparison of the two methods and hence give an understanding of the strengths and weaknesses of each. The extended framework would then allow the appropriate method to be applied automatically to novel models, providing a more general

purpose inference tool.

Finally, in terms of applications of the variational inference framework, there are a plethora of possible areas in which it can be applied, from social sciences to cosmology. My particular interest would be to use the framework to learn more complex models of natural image formation, as a means of performing the more general image analysis required for machine vision applications.

### **6.3 Summary**

In summary, I have created a framework which allows variational inference to be performed automatically in a wide range of probabilistic models. The framework has been used to solve difficult problems in the real world domains of machine vision and DNA microarray analysis. I have also demonstrated that inference with structured variational distributions is both tractable and capable of being performed automatically. The use of such structured distributions both improves the quality of the approximation and extends the applicability of the original framework.

# BIBLIOGRAPHY

- D. Ackley, G. Hinton, and T. Sejnowski. A learning algorithm for Boltzmann machines. *Cognitive Science*, 9:147–169, 1985.
- H. Attias. A variational Bayesian framework for graphical models. In S. Solla, T. K. Leen, and K-L Muller, editors, *Advances in Neural Information Processing Systems*, volume 12, pages 209–215, Cambridge MA, 2000. MIT Press.
- Z. Bar-Joseph, D. Gifford, and T. Jaakkola. Fast optimal leaf ordering for hierarchical clustering. *Bioinformatics*, 17:S22–29, 2001.
- D. Barber and C. M. Bishop. Variational learning in Bayesian neural networks. In C. M. Bishop, editor, *Generalization in Neural Networks and Machine Learning*. Springer Verlag, 1998.
- K. J. Bathe. *Finite Element Procedures*. Prentice-Hall, Englewood Cliffs, NJ, 1996.
- Rev. T. Bayes. An essay towards solving a problem in the doctrine of chances. In *Philosophical Transactions of the Royal Society*, volume 53, pages 370–418, 1763.
- A. Ben-Dor, R. Shamir, and Z. Yakhini. Clustering gene expression patterns. *Journal of Computational Biology*, 6(3/4):281–297, 1999.
- J. M. Bernardo and A.F.M. Smith. *Bayesian Theory*. John Wiley and Sons, New York, 1994.
- C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.
- C. M. Bishop. Bayesian PCA. In S. A. Solla M. S. Kearns and D. A. Cohn, editors, *Advances in Neural Information Processing Systems*, volume 11, pages 382–388. MIT Press, 1999a.
- C. M. Bishop. Variational principal components. In *Proceedings Ninth International Conference on Artificial Neural Networks, ICANN'99*, volume 1, pages 509–514. IEE, 1999b.
- 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.
- C. M. Bishop and M. E. Tipping. Variational Relevance Vector Machines. In *Proceedings of 16th Conference in Uncertainty in Artificial Intelligence*, pages 46–53. Morgan Kaufmann, 2000.

- C. M. Bishop and J. M. Winn. Non-linear Bayesian image modelling. In *Proceedings Sixth European Conference on Computer Vision*, volume 1, pages 3–17. Springer-Verlag, 2000.
- C. M. Bishop and J. M. Winn. Structured variational distributions in VIBES. In *Proceedings Artificial Intelligence and Statistics*, Key West, Florida, 2003. Society for Artificial Intelligence and Statistics.
- C. M. Bishop, J. M. Winn, and D. Spiegelhalter. VIBES: A variational inference engine for Bayesian networks. In *Advances in Neural Information Processing Systems*, volume 15, 2002.
- M. J. Black and Y. Yacoob. Recognizing facial expressions under rigid and non-rigid facial motions. In *International Workshop on Automatic Face and Gesture Recognition, Zurich*, pages 12–17, 1995.
- C. Bregler and S.M. Omohundro. Nonlinear manifold learning for visual speech recognition. In *Fifth International Conference on Computer Vision*, pages 494–499, Boston, Jun 1995.
- J. Buhler, T. Ideker, and D. Haynor. Dapple: Improved techniques for finding spots on DNA microarrays. Technical report, University of Washington, 2000.
- R. Choudrey, W. Penny, and S. Roberts. An ensemble learning approach to independent component analysis. In *IEEE International Workshop on Neural Networks for Signal Processing*, 2000.
- G. F. Cooper. The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence*, 42:393–405, 1990.
- T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models — their training and application. In *Computer vision, graphics and image understanding*, volume 61, pages 38–59, 1995.
- R. G. Cowell, A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter. *Probabilistic Networks and Expert Systems*. Statistics for Engineering and Information Science. Springer-Verlag, 1999.
- R. T. Cox. Probability, frequency and reasonable expectation. *American Journal of Physics*, 14(1):1–13, 1946.
- P. Dagum and M. Luby. Approximating probabilistic inference in Bayesian belief networks is NP-hard. *Artificial Intelligence*, 60:141–153, 1993.
- A. Darwiche. Conditioning methods for exact and approximate inference in causal networks. In *Eleventh Annual Conference on Uncertainty in Artificial Intelligence*. Morgan Kaufmann, August 1995.

- S. Dudoit, Y. H. Yang, Matthew J. Callow, and T. P. Speed. Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. Technical report, Department of Biochemistry, Stanford University School of Medicine, 2000.
- M. Eisen, P. Spellman, D. Botstein, and P. Brown. Cluster analysis and display of genome-wide expression patterns. In *Proceedings of National Academy of Science*, volume 95, pages 14863–14867, 1998.
- M.B. Eisen and P.O. Brown. DNA arrays for analysis of gene expression. *Methods in Enzymology*, 303:179–205, 1999.
- B.S. Everitt and D.J. Hand. *Finite Mixture Distributions*. Chapman and Hall, London, 1981.
- R.P. Feynman. *Statistical Mechanics*. W. A. Benjamin, Inc., MA, 1972.
- B. Frey. *Graphical models for machine learning and digital communications*. MIT Press, Cambridge, MA, 1998.
- B. Frey and N. Jojic. Transformed component analysis: joint estimation of spatial transformations and image components. In *Seventh International Conference on Computer Vision*, pages 1190–1196, 1999.
- B. Frey, F. Kschischang, H. Loeliger, and N. Wiberg. Factor graphs and algorithms. In *Proceedings of the 35th Allerton Conference on Communication, Control and Computing 1997*, 1998.
- N. Friedman, M. Linial, I. Nachman, and D. Pe’er. Using Bayesian networks to analyze expression data. In *RECOMB*, pages 127–135, 2000.
- R.G. Gallager. Low density parity check codes. *IRE Trans. Info. Theory*, IT-8:21–28, Jan 1962.
- R.G. Gallager. *Low density parity check codes*. Number 21 in Research monograph series. MIT Press, Cambridge, MA, 1963.
- S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(1): 721–741, 1984.
- Z. Ghahramani and M. J. Beal. Variational inference for Bayesian mixture of factor analysers. In *Advances in Neural Information Processing Systems*, volume 12, 1999.
- Z. Ghahramani and M. J. Beal. Propagation algorithms for variational Bayesian learning. In T. K. Leen, T. Dietterich, and V. Tresp, editors, *Advances in Neural Information Processing Systems*, volume 13, Cambridge MA, 2001. MIT Press.

- W. R. Gilks and P. Wild. Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, 41(2):337–348, 1992.
- A. Hartemink, D. Gifford, T. Jaakkola, and R. Young. Using graphical models and genomic expression data to statistically validate models of genetic regulatory networks. In *Pacific Symposium on Biocomputing*, volume 6, pages 422–433, 2001.
- T. Heap and D. Hogg. Wormholes in shape space: Tracking through discontinuous changes in shape. In *Sixth International Conference on Computer Vision*, pages 344–349, 1998.
- P. Hegde, R. Qi, R. Abernathy, C. Gay, S. Dharap, R. Gaspard R, J. Earle-Hughes, E. Snesrud, N. H. Lee, and J. Quackenbush. A concise guide to cDNA microarray analysis. *Biotechniques*, 29(3):548–562, 2000.
- T. Heskes. Stable fixed points of loopy belief propagation are minima of the Bethe free energy, 2002.
- G. E. Hinton and D. van Camp. Keeping neural networks simple by minimizing the description length of the weights. In *Proceedings of the Sixth Annual Conference on Computational Learning Theory*, pages 5–13, 1993.
- G. E. Hinton and R. S. Zemel. Autoencoders, minimum description length, and Helmholtz free energy. In *Advances in Neural Information Processing Systems*, volume 6, 1994.
- G. Hori, M. Inoue, S. Nishimura, and H. Nakahara. Blind gene classification on ICA of microarray data. In *ICA 2001*, pages 332–336, 2001.
- T. Jaakkola. *Variational Methods for Inference and Estimation in Graphical Models*. PhD thesis, MIT, 1997.
- F. Jensen. *An introduction to Bayesian networks*. UCL Press, 1996.
- M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul. An introduction to variational methods for graphical models. In M. I. Jordan, editor, *Learning in Graphical Models*, pages 105–162. Kluwer, 1998.
- N. Kambhatla and T.K. Leen. Dimension reduction by local principal component analysis. *Neural Computation*, 9(7):1493–1516, 1997.
- R. Kinderman and J. L. Snell. Markov random fields and their applications. *American Mathematical Society*, 1:1–142, 1980.
- F. R. Kschischang, B. Frey, and H.-A. Loeliger. Factor graphs and the sum-product algorithm. *IEEE Trans. Inform. Theory*, 47(2):498–519, 2001.
- S. Kullback. *Information Theory and Statistics*. Dover Publications, New York, 1959.

- S. Kullback and R. A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22:79–86, 1951.
- S. L. Lauritzen and D.J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society*, 50:157–224, 1988.
- S. L. Lauritzen and N. Wermuth. Graphical models for associations between variables, some of which are qualitative and some quantitative. *Annals of Statistics*, 17:31–57, 1989.
- N. Lawrence, M. Milo, M. Niranjana, P. Rashbass, and S. Soullier. Reducing the variability in microarray image processing by Bayesian inference. Technical report, Department of Computer Science, University of Sheffield, 2002.
- D. B. Lenat. CYC: A large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38(11):33–38, 1995.
- D. J. Lunn, A. Thomas, N. G. Best, and D. J. Spiegelhalter. WinBUGS – a Bayesian modelling framework: concepts, structure and extensibility. *Statistics and Computing*, 10:321–333, 2000. <http://www.mrc-bsu.cam.ac.uk/bugs/>.
- D. J. C. MacKay. Probable networks and plausible predictions – a review of practical Bayesian methods for supervised neural networks. *Network: Computation in Neural Systems*, 6(3): 469–505, 1995.
- D. J. C. MacKay. Ensemble learning for hidden Markov models, 1997. Unpublished manuscript, Department of Physics, University of Cambridge.
- D. J. C. MacKay. Introduction to Monte Carlo methods. In M. I. Jordan, editor, *Learning in Graphical Models*. Kluwer, 1998.
- D. J. C. MacKay. *Information theory, inference, and learning algorithms*. Cambridge University Press, Cambridge, UK, 2003.
- D. J. C. MacKay and R. M. Neal. Good codes based on very sparse matrices. In *IMA: IMA Conference on Cryptography and Coding, LNCS lately (earlier: Cryptography and Coding II, Edited by Chris Mitchell, Clarendon Press, 1992)*, 1995.
- A. Martoglio, J. W. Miskin, S. K. Smith, and D. J. C. MacKay. A decomposition model to track gene expression signatures: preview on observer-independent classification of ovarian cancer. *Bioinformatics*, 18:1617–1624, 2002.
- A. Martoglio, B. D. Tom, M. Starkey, A. N. Corps, S. Charnock-Jones, and S. K. Smith. Changes in tumorigenesis- and angiogenesis-related gene transcript abundance profiles in ovarian cancer detected by tailored high density cDNA arrays. *Molecular Medicine*, 6(9): 750–765, 2000.

- R. J. McEliece, D. J. C. MacKay, and J.-F. Cheng. Turbo decoding as an instance of Pearl's Belief Propagation algorithm. *IEEE Journal on selected areas in communication*, 1997.
- G. S. Michaels, D. B. Carr, M. Askenazi, S. Fuhrman, X. Wen, and R. Somogyi. Cluster analysis and data visualization of large-scale gene expression data. In *Pacific Symposium on Biocomputing*, volume 3, pages 42–53, 1998.
- T. P. Minka. Expectation propagation for approximate Bayesian inference. In *Proceedings of the 17th Annual Conference on Uncertainty in Artificial Intelligence*, pages 362–369. Morgan Kaufmann, 2001a.
- T. P. Minka. *A family of algorithms for approximate Bayesian inference*. PhD thesis, MIT, 2001b.
- J. W. Miskin. *Ensemble Learning for Independent Component Analysis*. PhD thesis, University of Cambridge, 2000.
- J. W. Miskin and D. J. C. MacKay. Ensemble learning for blind source separation. In S. J. Roberts and R. M. Everson, editors, *ICA: Principles and Practice*. Cambridge University Press, 2000.
- B. Moghaddam. Principal manifolds and Bayesian subspaces for visual recognition. In *Seventh International Conference on Computer Vision*, pages 1131–1136, 1999.
- B. Moghaddam and A. Pentland. Probabilistic visual learning for object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):696–710, 1997.
- V.S. Nalwa. *A Guided Tour of Computer Vision*. Addison-Wesley, 1993.
- R. M. Neal. Connectionist learning of belief networks. *Artificial Intelligence*, 56:71–113, 1992.
- R. M. Neal. Probabilistic inference using Markov chain Monte Carlo methods. Technical Report CRG-TR-93-1, Department of Computer Science, University of Toronto, Canada, 1993.
- R. M. Neal. *Bayesian Learning for Neural Networks*. PhD thesis, University of Toronto, Canada, 1994.
- R. M. Neal and G. E. Hinton. A new view of the EM algorithm that justifies incremental and other variants. In M. I. Jordan, editor, *Learning in Graphical Models*, pages 355–368. Kluwer, 1998.
- J. Pearl. Fusion, propagation and structuring in belief networks. *Artificial Intelligence*, 29: 241–288, 1986.
- J. Pearl. Evidential reasoning using stochastic simulation of causal models. *Artificial Intelligence*, 32:245–257, 1987.

- J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Francisco, 1988.
- S. Raychaudhuri, J. Stuart, and R. Altman. Principal Components Analysis to summarize microarray experiments: application to sporulation time series. In *Pacific Symposium on Biocomputing*, volume 5, 2000.
- S. Roweis. EM algorithms for PCA and SPCA. In Michael I. Jordan, Michael J. Kearns, and Sara A. Solla, editors, *Advances in Neural Information Processing Systems*, volume 10. MIT Press, 1998.
- J. Rustagi. *Variational Methods in Statistics*. Academic Press, New York, 1976.
- J. Sakurai. *Modern Quantum Mechanics*. Addison-Wesley, Redwood City, CA, 1985.
- L. K. Saul and M. I. Jordan. Exploiting tractable substructures in intractable networks. In D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, editors, *Advances in Neural Information Processing Systems*, volume 8, pages 486–492. MIT Press, 1996.
- E. H. Shortcliffe. *Computer-Based Medical Consultations: MYCIN*. Elsevier Science, New York, 1976.
- P. T. Spellman, G. Sherlock, M. Q. Zhang, V. R. Iyer, K. Anders, M. B. Eisen, P. O. Brown, D. Botstein, and B. Futcher. Comprehensive identification of cell cycle-regulated genes of the yeast *saccharomyces cerevisiae* by microarray hybridization. In *Molecular Biology of the Cell*, volume 9, pages 3273–3297, 1998.
- D. J. Spiegelhalter. Probabilistic reasoning in predictive expert systems. In L. N. Kanal and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence*, pages 47–68, Amsterdam, 1986. North Holland.
- P. Tamayo, D. Slonim, J. Mesirov, Q. Zhu, S. Kitareewan, E. Dmitrovsky, E. Lander, and T. Golub. Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. In *Proceedings of the National Academy of Science*, volume 96, pages 2907–2912, 1999.
- A. Thomas, D. J. Spiegelhalter, and W. R. Gilks. BUGS: A program to perform Bayesian inference using Gibbs sampling. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics*, Oxford: Clarendon Press, 1992.
- M. E. Tipping and C. M. Bishop. Mixtures of probabilistic principal component analyzers. *Neural Computation*, 11(2):443–482, 1999a.
- M. E. Tipping and C. M. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society, Series B*, 21(3):611–622, 1999b.

- M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.
- N. Ueda, R. Nakano, Z. Ghahramani, and G. E. Hinton. SMEM algorithm for mixture models. In *Advances in Neural Information Processing Systems*, volume 11, 1999.
- P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, 2001.
- Niclas Wiberg. *Codes and Decoding on General Graphs*. PhD thesis, Linköping University, 1996.
- W. Wiegnerinck. Variational approximations between mean field theory and the junction tree algorithm. In *Uncertainty in Artificial Intelligence*. Morgan Kauffmann, 2000.
- C. K. I. Williams and C. E. Rasmussen. Gaussian processes for regression. In D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, editors, *Advances in Neural Information Processing Systems*, volume 8, pages 514–520. MIT Press, 1996.
- P. H. Winston. *Artificial Intelligence*. Addison-Wesley, third edition, 1992.
- E. P. Xing, M. I. Jordan, and S. Russell. A generalized mean field algorithm for variational inference in exponential families. In *Uncertainty in Artificial Intelligence*. Morgan Kauffmann, 2003.
- J. Yedidia, W. Freeman, and Y. Weiss. Understanding belief propagation and its generalizations. In G. Lakemeyer and B. Nebel, editors, *Exploring Artificial Intelligence in the New Millennium*. Morgan Kaufmann, 2002.
- K. Yeung, C. Fraley, A. Murua, A. Raftery, and W. Ruzzo. Model-based clustering and data transformations for gene expression data. *Bioinformatics*, 17(10):977–987, 2001.