Graphical Models, Exponential Families, and Variational Inference

Graphical Models, Exponential Families, and Variational Inference

Martin J. Wainwright

University of California
Berkeley 94720
USA
wainwrig@stat.berkeley.edu

Michael I. Jordan

University of California
Berkeley 94720
USA
jordan@stat.berkeley.edu

the essence of knowledge

Boston – Delft

Foundations and Trends® in Machine Learning

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is M. J. Wainwright and M. I. Jordan, Graphical Models, Exponential Families, and Variational Inference, Foundation and Trends® in Machine Learning, vol 1, nos 1–2, pp 1–305, 2008

ISBN: 978-1-60198-184-4

Foundations and Trends® in Machine Learning

Volume 1 Issue 1–2, 2008

Editorial Board

Editorial Scope

Foundations and Trends® in Machine Learning publishes survey and tutorial articles in the following topics:

- Adaptive control and signal processing
- Applications and case studies
- Behavioral, cognitive and neural learning
- Bayesian learning
- Classification and prediction
- Clustering
- Data mining
- Dimensionality reduction
- Evaluation
- Game theoretic learning
- Graphical models
- Independent component analysis
- Inductive logic programming
- Kernel methods
- Markov chain Monte Carlo
- Model choice
- Nonparametric methods
- Online learning
- Optimization
- Reinforcement learning
- Relational learning
- Robustness
- Spectral methods
- Statistical learning theory
- Variational inference
- Visualization

Information for Librarians

Foundations and Trends® in Machine Learning, 2008, Volume 1, 4 issues. ISSN paper version 1935-8237. ISSN online version 1935-8245. Also available as a combined paper and online subscription.

Foundations and Trends® in
Machine Learning
Vol. 1, Nos. 1–2 (2008) 1–305

DOI: 10.1561/2200000001

Graphical Models, Exponential Families, and Variational Inference

Martin J. Wainwright[1] and Michael I. Jordan[2]

[1] *Department of Statistics, and Department of Electrical Engineering and Computer Science, University of California, Berkeley 94720, USA, wainwrig@stat.berkeley.edu*

[2] *Department of Statistics, and Department of Electrical Engineering and Computer Science, University of California, Berkeley 94720, USA, jordan@stat.berkeley.edu*

Abstract

The formalism of probabilistic graphical models provides a unifying framework for capturing complex dependencies among random variables, and building large-scale multivariate statistical models. Graphical models have become a focus of research in many statistical, computational and mathematical fields, including bioinformatics, communication theory, statistical physics, combinatorial optimization, signal and image processing, information retrieval and statistical machine learning. Many problems that arise in specific instances — including the key problems of computing marginals and modes of probability distributions — are best studied in the general setting. Working with exponential family representations, and exploiting the conjugate duality between the cumulant function and the entropy for exponential families, we develop general variational representations of the problems of computing likelihoods, marginal probabilities and most probable configurations. We describe how a wide variety

of algorithms — among them sum-product, cluster variational methods, expectation-propagation, mean field methods, max-product and linear programming relaxation, as well as conic programming relaxations — can all be understood in terms of exact or approximate forms of these variational representations. The variational approach provides a complementary alternative to Markov chain Monte Carlo as a general source of approximation methods for inference in large-scale statistical models.

Contents

1

Introduction

Graphical models bring together graph theory and probability theory in a powerful formalism for multivariate statistical modeling. In various applied fields including bioinformatics, speech processing, image processing and control theory, statistical models have long been formulated in terms of graphs, and algorithms for computing basic statistical quantities such as likelihoods and score functions have often been expressed in terms of recursions operating on these graphs; examples include phylogenies, pedigrees, hidden Markov models, Markov random fields, and Kalman filters. These ideas can be understood, unified, and generalized within the formalism of graphical models. Indeed, graphical models provide a natural tool for formulating variations on these classical architectures, as well as for exploring entirely new families of statistical models. Accordingly, in fields that involve the study of large numbers of interacting variables, graphical models are increasingly in evidence.

Graph theory plays an important role in many computationally oriented fields, including combinatorial optimization, statistical physics, and economics. Beyond its use as a language for formulating models, graph theory also plays a fundamental role in assessing computational complexity and feasibility. In particular, the running time of an algorithm or the magnitude of an error bound can often be characterized in terms of structural properties of a graph. This statement is also true in the context of graphical models. Indeed, as we discuss, the computational complexity of a fundamental method known as the *junction tree algorithm* — which generalizes many of the recursive algorithms on graphs cited above — can be characterized in terms of a natural graph-theoretic measure of interaction among variables. For suitably sparse graphs, the junction tree algorithm provides a systematic solution to the problem of computing likelihoods and other statistical quantities associated with a graphical model.

Unfortunately, many graphical models of practical interest are not "suitably sparse," so that the junction tree algorithm no longer provides a viable computational framework. One popular source of methods for attempting to cope with such cases is the *Markov chain Monte Carlo* (MCMC) framework, and indeed there is a significant literature on the application of MCMC methods to graphical models [e.g., 28, 93, 202]. Our focus in this survey is rather different: we present an alternative computational methodology for statistical inference that is based on *variational methods*. These techniques provide a general class of alternatives to MCMC, and have applications outside of the graphical model framework. As we will see, however, they are particularly natural in their application to graphical models, due to their relationships with the structural properties of graphs.

The phrase "variational" itself is an umbrella term that refers to various mathematical tools for optimization-based formulations of problems, as well as associated techniques for their solution. The general idea is to express a quantity of interest as the solution of an optimization problem. The optimization problem can then be "relaxed" in various ways, either by approximating the function to be optimized or by approximating the set over which the optimization takes place. Such relaxations, in turn, provide a means of approximating the original quantity of interest.

The roots of both MCMC methods and variational methods lie in statistical physics. Indeed, the successful deployment of MCMC methods in statistical physics motivated and predated their entry into statistics. However, the development of MCMC methodology specifically designed for statistical problems has played an important role in sparking widespread application of such methods in statistics [88]. A similar development in the case of variational methodology would be of significant interest. In our view, the most promising avenue toward a variational methodology tuned to statistics is to build on existing links between variational analysis and the exponential family of distributions [4, 11, 43, 74]. Indeed, the notions of convexity that lie at the heart of the statistical theory of the exponential family have immediate implications for the design of variational relaxations. Moreover, these variational relaxations have particularly interesting algorithmic consequences in the setting of graphical models, where they again lead to recursions on graphs.

Thus, we present a story with three interrelated themes. We begin in Section 2 with a discussion of graphical models, providing both an overview of the general mathematical framework, and also presenting several specific examples. All of these examples, as well as the majority of current applications of graphical models, involve distributions in the exponential family. Accordingly, Section 3 is devoted to a discussion of exponential families, focusing on the mathematical links to convex analysis, and thus anticipating our development of variational methods. In particular, the principal object of interest in our exposition is a certain *conjugate dual* relation associated with exponential families. From this foundation of conjugate duality, we develop a general variational representation for computing likelihoods and marginal probabilities in exponential families. Subsequent sections are devoted to the exploration of various instantiations of this variational principle, both in exact and approximate forms, which in turn yield various algorithms for computing exact and approximate marginal probabilities, respectively. In Section 4, we discuss the connection between the Bethe approximation and the sum-product algorithm, including both its exact form for trees and approximate form for graphs with cycles. We also develop the connections between Bethe-like approximations and other algorithms, including generalized sum-product, expectation-propagation and related moment-matching methods. In Section 5, we discuss the class of mean field methods, which arise from a qualitatively different approximation to the exact variational principle, with the added benefit of generating lower bounds on the likelihood. In Section 6, we discuss the role of variational methods in parameter estimation, including both the fully observed and partially observed cases, as well as both frequentist and Bayesian settings. Both Bethe-type and mean field methods are based on nonconvex optimization problems, which typically have multiple solutions. In contrast, Section 7 discusses variational methods based on convex relaxations of the exact variational principle, many of which are also guaranteed to yield upper bounds on the log likelihood. Section 8 is devoted to the problem of mode computation, with particular emphasis on the case of discrete random variables, in which context computing the mode requires solving an integer programming problem. We develop connections between (reweighted) max-product algorithms and hierarchies of linear programming relaxations. In Section 9, we discuss the broader class of conic programming relaxations, and show how they can be understood in terms of semidefinite constraints imposed via moment matrices. We conclude with a discussion in Section 10.

The scope of this survey is limited in the following sense: given a distribution represented as a graphical model, we are concerned with the problem of computing marginal probabilities (including likelihoods), as well as the problem of computing modes. We refer to such computational tasks as problems of "probabilistic inference," or "inference" for short. As with presentations of MCMC methods, such a limited focus may appear to aim most directly at applications in Bayesian statistics. While Bayesian statistics is indeed a natural terrain for deploying many of the methods that we present here, we see these methods as having applications throughout statistics, within both the frequentist and Bayesian paradigms, and we indicate some of these applications at various junctures in the survey.

References

[1] A. Agresti, *Categorical Data Analysis.* New York: Wiley, 2002.

[2] S. M. Aji and R. J. McEliece, "The generalized distributive law," *IEEE Transactions on Information Theory*, vol. 46, pp. 325–343, 2000.

[3] N. I. Akhiezer, *The Classical Moment Problem and Some Related Questions in Analysis.* New York: Hafner Publishing Company, 1966.

[4] S. Amari, "Differential geometry of curved exponential families — curvatures and information loss," *Annals of Statistics*, vol. 10, no. 2, pp. 357–385, 1982.

[5] S. Amari and H. Nagaoka, *Methods of Information Geometry.* Providence, RI: AMS, 2000.

[6] G. An, "A note on the cluster variation method," *Journal of Statistical Physics*, vol. 52, no. 3, pp. 727–734, 1988.

[7] A. Bandyopadhyay and D. Gamarnik, "Counting without sampling: New algorithms for enumeration problems wiusing statistical physics," in *Proceedings of the 17th ACM-SIAM Symposium on Discrete Algorithms*, 2006.

[8] O. Banerjee, L. El Ghaoui, and A. d'Aspremont, "Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data," *Journal of Machine Learning Research*, vol. 9, pp. 485–516, 2008.

[9] F. Barahona and M. Groetschel, "On the cycle polytope of a binary matroid," *Journal on Combination Theory, Series B*, vol. 40, pp. 40–62, 1986.

[10] D. Barber and W. Wiegerinck, "Tractable variational structures for approximating graphical models," in *Advances in Neural Information Processing Systems*, pp. 183–189, Cambridge, MA: MIT Press, 1999.

[11] O. E. Barndorff-Nielsen, *Information and Exponential Families.* Chichester, UK: Wiley, 1978.

[12] R. J. Baxter, *Exactly Solved Models in Statistical Mechanics.* New York: Academic Press, 1982.

[13] M. Bayati, C. Borgs, J. Chayes, and R. Zecchina, "Belief-propagation for weighted b-matchings on arbitrary graphs and its relation to linear programs with integer solutions," Technical Report arXiv: 0709 1190, Microsoft Research, 2007.

[14] M. Bayati and C. Nair, "A rigorous proof of the cavity method for counting matchings," in *Proceedings of the Allerton Conference on Control, Communication and Computing*, Monticello, IL, 2007.

[15] M. Bayati, D. Shah, and M. Sharma, "Maximum weight matching for max-product belief propagation," in *International Symposium on Information Theory*, Adelaide, Australia, 2005.

[16] M. J. Beal, "Variational algorithms for approximate Bayesian inference," PhD thesis, Gatsby Computational Neuroscience Unit, University College, London, 2003.

[17] C. Berge, *The Theory of Graphs and its Applications.* New York: Wiley, 1964.

[18] E. Berlekamp, R. McEliece, and H. van Tilborg, "On the inherent intractability of certain coding problems," *IEEE Transactions on Information Theory*, vol. 24, pp. 384–386, 1978.

[19] U. Bertele and F. Brioschi, *Nonserial Dynamic Programming.* New York: Academic Press, 1972.

[20] D. P. Bertsekas, *Dynamic Programming and Stochastic Control.* Vol. 1. Belmont, MA: Athena Scientific, 1995.

[21] D. P. Bertsekas, *Nonlinear Programming.* Belmont, MA: Athena Scientific, 1995.

[22] D. P. Bertsekas, *Network Optimization: Continuous and Discrete Methods.* Belmont, MA: Athena Scientific, 1998.

[23] D. P. Bertsekas, *Convex Analysis and Optimization.* Belmont, MA: Athena Scientific, 2003.

[24] D. Bertsimas and J. N. Tsitsiklis, *Introduction to Linear Optimization.* Belmont, MA: Athena Scientific, 1997.

[25] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *Journal of the Royal Statistical Society, Series B*, vol. 36, pp. 192–236, 1974.

[26] J. Besag, "Statistical analysis of non-lattice data," *The Statistician*, vol. 24, no. 3, pp. 179–195, 1975.

[27] J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society, Series B*, vol. 48, no. 3, pp. 259–279, 1986.

[28] J. Besag and P. J. Green, "Spatial statistics and Bayesian computation," *Journal of the Royal Statistical Society, Series B*, vol. 55, no. 1, pp. 25–37, 1993.

[29] H. A. Bethe, "Statistics theory of superlattices," *Proceedings of Royal Society London, Series A*, vol. 150, no. 871, pp. 552–575, 1935.

[30] P. J. Bickel and K. A. Doksum, *Mathematical statistics: basic ideas and selected topics.* Upper Saddle River, N.J.: Prentice Hall, 2001.

[31] D. M. Blei and M. I. Jordan, "Variational inference for Dirichlet process mixtures," *Bayesian Analysis*, vol. 1, pp. 121–144, 2005.

[32] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[33] H. Bodlaender, "A tourist guide through treewidth," *Acta Cybernetica*, vol. 11, pp. 1–21, 1993.

[34] B. Bollobás, *Graph Theory: An Introductory Course.* New York: Springer-Verlag, 1979.

[35] B. Bollobás, *Modern Graph Theory.* New York: Springer-Verlag, 1998.

[36] E. Boros, Y. Crama, and P. L. Hammer, "Upper bounds for quadratic 0-1 maximization," *Operations Research Letters*, vol. 9, pp. 73–79, 1990.

[37] E. Boros and P. L. Hammer, "Pseudo-boolean optimization," *Discrete Applied Mathematics*, vol. 123, pp. 155–225, 2002.

[38] J. Borwein and A. Lewis, *Convex Analysis.* New York: Springer-Verlag, 1999.

[39] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge, UK: Cambridge University Press, 2004.

[40] X. Boyen and D. Koller, "Tractable inference for complex stochastic processes," in *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, pp. 33–42, San Francisco, CA: Morgan Kaufmann, 1998.

[41] A. Braunstein, M. Mézard, and R. Zecchina, "Survey propagation: An algorithm for satisfiability," Technical Report, arXiv:cs.CC/02122002 v2, 2003.

[42] L. M. Bregman, "The relaxation method for finding the common point of convex sets and its application to the solution of problems in convex programming," *USSR Computational Mathematics and Mathematical Physics*, vol. 7, pp. 191–204, 1967.

[43] L. D. Brown, *Fundamentals of Statistical Exponential Families.* Hayward, CA: Institute of Mathematical Statistics, 1986.

[44] C. B. Burge and S. Karlin, "Finding the genes in genomic DNA," *Current Opinion in Structural Biology*, vol. 8, pp. 346–354, 1998.

[45] Y. Censor and S. A. Zenios, *Parallel Optimization: Theory, Algorithms, and Applications.* Oxford: Oxford University Press, 1988.

[46] M. Cetin, L. Chen, J. W. Fisher, A. T. Ihler, R. L. Moses, M. J. Wainwright, and A. S. Willsky, "Distributed fusion in sensor networks," *IEEE Signal Processing Magazine*, vol. 23, pp. 42–55, 2006.

[47] D. Chandler, *Introduction to Modern Statistical Mechanics.* Oxford: Oxford University Press, 1987.

[48] C. Chekuri, S. Khanna, J. Naor, and L. Zosin, "A linear programming formulation and approximation algorithms for the metric labeling problem," *SIAM Journal on Discrete Mathematics*, vol. 18, no. 3, pp. 608–625, 2005.

[49] L. Chen, M. J. Wainwright, M. Cetin, and A. Willsky, "Multitarget-multisensor data association using the tree-reweighted max-product algorithm," in *SPIE Aerosense Conference*, Orlando, FL, 2003.

[50] M. Chertkov and V. Y. Chernyak, "Loop calculus helps to improve belief propagation and linear programming decoding of LDPC codes," in *Proceedings of the Allerton Conference on Control, Communication and Computing*, Monticello, IL, 2006.

[51] M. Chertkov and V. Y. Chernyak, "Loop series for discrete statistical models on graphs," *Journal of Statistical Mechanics*, p. P06009, 2006.

[52] M. Chertkov and V. Y. Chernyak, "Loop calculus helps to improve belief propagation and linear programming decodings of low density parity check codes," in *Proceedings of the Allerton Conference on Control, Communication and Computing*, Monticello, IL, 2007.

[53] M. Chertkov and M. G. Stepanov, "An efficient pseudo-codeword search algorithm for linear programming decoding of LDPC codes," Technical Report arXiv:cs.IT/0601113, Los Alamos National Laboratories, 2006.

[54] S. Chopra, "On the spanning tree polyhedron," *Operations Research Letters*, vol. 8, pp. 25–29, 1989.

[55] S. Cook, "The complexity of theorem-proving procedures," in *Proceedings of the Third Annual ACM Symposium on Theory of Computing*, pp. 151–158, 1971.

[56] T. H. Cormen, C. E. Leiserson, and R. L. Rivest, *Introduction to Algorithms*. Cambridge, MA: MIT Press, 1990.

[57] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley and Sons, 1991.

[58] G. Cross and A. Jain, "Markov random field texture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 5, pp. 25–39, 1983.

[59] I. Csiszár, "A geometric interpretation of Darroch and Ratcliff's generalized iterative scaling," *Annals of Statistics*, vol. 17, no. 3, pp. 1409–1413, 1989.

[60] I. Csiszár and G. Tusnády, "Information geometry and alternating minimization procedures," *Statistics and Decisions, Supplemental Issue 1*, pp. 205–237, 1984.

[61] J. N. Darroch and D. Ratcliff, "Generalized iterative scaling for log-linear models," *Annals of Mathematical Statistics*, vol. 43, pp. 1470–1480, 1972.

[62] C. Daskalakis, A. G. Dimakis, R. M. Karp, and M. J. Wainwright, "Probabilistic analysis of linear programming decoding," *IEEE Transactions on Information Theory*, vol. 54, no. 8, pp. 3565–3578, 2008.

[63] A. d'Aspremont, O. Banerjee, and L. El Ghaoui, "First order methods for sparse covariance selection," *SIAM Journal on Matrix Analysis and its Applications*, vol. 30, no. 1, pp. 55–66, 2008.

[64] J. Dauwels, H. A. Loeliger, P. Merkli, and M. Ostojic, "On structured-summary propagation, LFSR synchronization, and low-complexity trellis decoding," in *Proceedings of the Allerton Conference on Control, Communication and Computing*, Monticello, IL, 2003.

[65] A. P. Dawid, "Applications of a general propagation algorithm for probabilistic expert systems," *Statistics and Computing*, vol. 2, pp. 25–36, 1992.

[66] R. Dechter, *Constraint Processing*. San Francisco, CA: Morgan Kaufmann, 2003.

[67] J. W. Demmel, *Applied Numerical Linear Algebra*. Philadelphia, PA: SIAM, 1997.

[68] 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*, vol. 39, pp. 1–38, 1977.

[69] M. Deza and M. Laurent, *Geometry of Cuts and Metric Embeddings*. New York: Springer-Verlag, 1997.

[70] A. G. Dimakis and M. J. Wainwright, "Guessing facets: Improved LP decoding and polytope structure," in *International Symposium on Information Theory*, Seattle, WA, 2006.

[71] M. Dudik, S. J. Phillips, and R. E. Schapire, "Maximum entropy density estimation with generalized regularization and an application to species distribution modeling," *Journal of Machine Learning Research*, vol. 8, pp. 1217–1260, 2007.

[72] R. Durbin, S. Eddy, A. Krogh, and G. Mitchison, eds., *Biological Sequence Analysis*. Cambridge, UK: Cambridge University Press, 1998.

[73] J. Edmonds, "Matroids and the greedy algorithm," *Mathematical Programming*, vol. 1, pp. 127–136, 1971.

[74] B. Efron, "The geometry of exponential families," *Annals of Statistics*, vol. 6, pp. 362–376, 1978.

[75] G. Elidan, I. McGraw, and D. Koller, "Residual belief propagation: Informed scheduling for asynchronous message-passing," in *Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence*, Arlington, VA: AUAI Press, 2006.

[76] J. Feldman, D. R. Karger, and M. J. Wainwright, "Linear programming-based decoding of turbo-like codes and its relation to iterative approaches," in *Proceedings of the Allerton Conference on Control, Communication and Computing*, Monticello, IL, 2002.

[77] J. Feldman, T. Malkin, R. A. Servedio, C. Stein, and M. J. Wainwright, "LP decoding corrects a constant fraction of errors," *IEEE Transactions on Information Theory*, vol. 53, no. 1, pp. 82–89, 2007.

[78] J. Feldman, M. J. Wainwright, and D. R. Karger, "Using linear programming to decode binary linear codes," *IEEE Transactions on Information Theory*, vol. 51, pp. 954–972, 2005.

[79] J. Felsenstein, "Evolutionary trees from DNA sequences: A maximum likelihood approach," *Journal of Molecular Evolution*, vol. 17, pp. 368–376, 1981.

[80] S. E. Fienberg, "Contingency tables and log-linear models: Basic results and new developments," *Journal of the American Statistical Association*, vol. 95, no. 450, pp. 643–647, 2000.

[81] M. E. Fisher, "On the dimer solution of planar Ising models," *Journal of Mathematical Physics*, vol. 7, pp. 1776–1781, 1966.

[82] G. D. Forney, Jr., "The Viterbi algorithm," *Proceedings of the IEEE*, vol. 61, pp. 268–277, 1973.

[83] G. D. Forney, Jr., R. Koetter, F. R. Kschischang, and A. Reznick, "On the effective weights of pseudocodewords for codes defined on graphs with cycles," in *Codes, Systems and Graphical Models*, pp. 101–112, New York: Springer, 2001.

[84] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael, "Learning low-level vision," *International Journal on Computer Vision*, vol. 40, no. 1, pp. 25–47, 2000.

[85] B. J. Frey, R. Koetter, and N. Petrovic, "Very loopy belief propagation for unwrapping phase images," in *Advances in Neural Information Processing Systems*, pp. 737–743, Cambridge, MA: MIT Press, 2001.

[86] B. J. Frey, R. Koetter, and A. Vardy, "Signal-space characterization of iterative decoding," *IEEE Transactions on Information Theory*, vol. 47, pp. 766–781, 2001.

[87] R. G. Gallager, *Low-Density Parity Check Codes.* Cambridge, MA: MIT Press, 1963.

[88] A. E. Gelfand and A. F. M. Smith, "Sampling-based approaches to calculating marginal densities," *Journal of the American Statistical Association*, vol. 85, pp. 398–409, 1990.

[89] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721–741, 1984.

[90] H. O. Georgii, *Gibbs Measures and Phase Transitions.* New York: De Gruyter, 1988.

[91] Z. Ghahramani and M. J. Beal, "Propagation algorithms for variational Bayesian learning," in *Advances in Neural Information Processing Systems*, pp. 507–513, Cambridge, MA: MIT Press, 2001.

[92] Z. Ghahramani and M. I. Jordan, "Factorial hidden Markov models," *Machine Learning*, vol. 29, pp. 245–273, 1997.

[93] W. Gilks, S. Richardson, and D. Spiegelhalter, eds., *Markov Chain Monte Carlo in Practice.* New York: Chapman and Hall, 1996.

[94] A. Globerson and T. Jaakkola, "Approximate inference using planar graph decomposition," in *Advances in Neural Information Processing Systems*, pp. 473–480, Cambridge, MA: MIT Press, 2006.

[95] A. Globerson and T. Jaakkola, "Approximate inference using conditional entropy decompositions," in *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics*, San Juan, Puerto Rico, 2007.

[96] A. Globerson and T. Jaakkola, "Convergent propagation algorithms via oriented trees," in *Proceedings of the 23rd Conference on Uncertainty in Artificial Intelligence*, Arlington, VA: AUAI Press, 2007.

[97] A. Globerson and T. Jaakkola, "Fixing max-product: Convergent message passing algorithms for MAP, LP-relaxations," in *Advances in Neural Information Processing Systems*, pp. 553–560, Cambridge, MA: MIT Press, 2007.

[98] M. X. Goemans and D. P. Williamson, "Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming," *Journal of the ACM*, vol. 42, pp. 1115–1145, 1995.

[99] G. Golub and C. Van Loan, *Matrix Computations.* Baltimore: Johns Hopkins University Press, 1996.

[100] V. Gómez, J. M. Mooij, and H. J. Kappen, "Truncating the loop series expansion for BP," *Journal of Machine Learning Research*, vol. 8, pp. 1987–2016, 2007.

[101] D. M. Greig, B. T. Porteous, and A. H. Seheuly, "Exact maximum a posteriori estimation for binary images," *Journal of the Royal Statistical Society B*, vol. 51, pp. 271–279, 1989.

[102] G. R. Grimmett, "A theorem about random fields," *Bulletin of the London Mathematical Society*, vol. 5, pp. 81–84, 1973.

[103] M. Grötschel, L. Lovász, and A. Schrijver, *Geometric Algorithms and Combinatorial Optimization.* Berlin: Springer-Verlag, 1993.

[104] M. Grötschel and K. Truemper, "Decomposition and optimization over cycles in binary matroids," *Journal on Combination Theory, Series B*, vol. 46, pp. 306–337, 1989.

[105] P. L. Hammer, P. Hansen, and B. Simeone, "Roof duality, complementation, and persistency in quadratic 0-1 optimization," *Mathematical Programming*, vol. 28, pp. 121–155, 1984.

[106] J. M. Hammersley and P. Clifford, "Markov fields on finite graphs and lattices," Unpublished, 1971.

[107] M. Hassner and J. Sklansky, "The use of Markov random fields as models of texture," *Computer Graphics and Image Processing*, vol. 12, pp. 357–370, 1980.

[108] T. Hazan and A. Shashua, "Convergent message-passing algorithms for inference over general graphs with convex free energy," in *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence*, Arlington, VA: AUAI Press, 2008.

[109] T. Heskes, "On the uniqueness of loopy belief propagation fixed points," *Neural Computation*, vol. 16, pp. 2379–2413, 2004.

[110] T. Heskes, "Convexity arguments for efficient minimization of the Bethe and Kikuchi free energies," *Journal of Artificial Intelligence Research*, vol. 26, pp. 153–190, 2006.

[111] T. Heskes, K. Albers, and B. Kappen, "Approximate inference and constrained optimization," in *Proceedings of the 19th Conference on Uncertainty in Artificial Intelligence*, pp. 313–320, San Francisco, CA: Morgan Kaufmann, 2003.

[112] J. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms*. Vol. 1, New York: Springer-Verlag, 1993.

[113] J. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of Convex Analysis*. Springer-Verlag: New York, 2001.

[114] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, UK: Cambridge University Press, 1985.

[115] J. Hu, H.-A. Loeliger, J. Dauwels, and F. Kschischang, "A general computation rule for lossy summaries/messages with examples from equalization," in *Proceedings of the Allerton Conference on Control, Communication and Computing*, Monticello, IL, 2005.

[116] B. Huang and T. Jebara, "Loopy belief propagation for bipartite maximum weight b-matching," in *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics*, San Juan, Puerto Rico, 2007.

[117] X. Huang, A. Acero, H.-W. Hon, and R. Reddy, *Spoken Language Processing*. New York: Prentice Hall, 2001.

[118] A. Ihler, J. Fisher, and A. S. Willsky, "Loopy belief propagation: Convergence and effects of message errors," *Journal of Machine Learning Research*, vol. 6, pp. 905–936, 2005.

[119] E. Ising, "Beitrag zur theorie der ferromagnetismus," *Zeitschrift für Physik*, vol. 31, pp. 253–258, 1925.

[120] T. S. Jaakkola, "Tutorial on variational approximation methods," in *Advanced Mean Field Methods: Theory and Practice*, (M. Opper and D. Saad, eds.), pp. 129–160, Cambridge, MA: MIT Press, 2001.

[121] T. S. Jaakkola and M. I. Jordan, "Improving the mean field approximation via the use of mixture distributions," in *Learning in Graphical Models*, (M. I. Jordan, ed.), pp. 105–161, MIT Press, 1999.

[122] T. S. Jaakkola and M. I. Jordan, "Variational probabilistic inference and the QMR-DT network," *Journal of Artificial Intelligence Research*, vol. 10, pp. 291–322, 1999.

[123] E. T. Jaynes, "Information theory and statistical mechanics," *Physical Review*, vol. 106, pp. 620–630, 1957.

[124] J. K. Johnson, D. M. Malioutov, and A. S. Willsky, "Lagrangian relaxation for MAP estimation in graphical models," in *Proceedings of the Allerton Conference on Control, Communication and Computing*, Monticello, IL, 2007.

[125] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. Saul, "An introduction to variational methods for graphical models," *Machine Learning*, vol. 37, pp. 183–233, 1999.

[126] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Englewood Cliffs, NJ: Prentice Hall, 2000.

[127] H. Kappen and P. Rodriguez, "Efficient learning in Boltzmann machines using linear response theory," *Neural Computation*, vol. 10, pp. 1137–1156, 1998.

[128] H. J. Kappen and W. Wiegerinck, "Novel iteration schemes for the cluster variation method," in *Advances in Neural Information Processing Systems*, pp. 415–422, Cambridge, MA: MIT Press, 2002.

[129] D. Karger and N. Srebro, "Learning Markov networks: Maximum bounded tree-width graphs," in *Symposium on Discrete Algorithms*, pp. 392–401, 2001.

[130] S. Karlin and W. Studden, *Tchebycheff Systems, with Applications in Analysis and Statistics*. New York: Interscience Publishers, 1966.

[131] R. Karp, "Reducibility among combinatorial problems," in *Complexity of Computer Computations*, pp. 85–103, New York: Plenum Press, 1972.

[132] P. W. Kastelyn, "Dimer statistics and phase transitions," *Journal of Mathematical Physics*, vol. 4, pp. 287–293, 1963.

[133] R. Kikuchi, "The theory of cooperative phenomena," *Physical Review*, vol. 81, pp. 988–1003, 1951.

[134] S. Kim and M. Kojima, "Second order cone programming relaxation of nonconvex quadratic optimization problems," Technical report, Tokyo Institute of Technology, July 2000.

[135] J. Kleinberg and E. Tardos, "Approximation algorithms for classification problems with pairwise relationships: Metric labeling and Markov random fields," *Journal of the ACM*, vol. 49, pp. 616–639, 2002.

[136] R. Koetter and P. O. Vontobel, "Graph-covers and iterative decoding of finite length codes," in *Proceedings of the 3rd International Symposium on Turbo Codes*, pp. 75–82, Brest, France, 2003.

[137] V. Kolmogorov, "Convergent tree-reweighted message-passing for energy minimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1568–1583, 2006.

[138] V. Kolmogorov and M. J. Wainwright, "On optimality properties of tree-reweighted message-passing," in *Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence*, pp. 316–322, Arlington, VA: AUAI Press, 2005.

[139] N. Komodakis, N. Paragios, and G. Tziritas, "MRF optimization via dual decomposition: Message-passing revisited," in *International Conference on Computer Vision*, Rio de Janeiro, Brazil, 2007.

[140] V. K. Koval and M. I. Schlesinger, "Two-dimensional programming in image analysis problems," *USSR Academy of Science, Automatics and Telemechanics*, vol. 8, pp. 149–168, 1976.

[141] A. Krogh, B. Larsson, G. von Heijne, and E. L. L. Sonnhammer, "Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes," *Journal of Molecular Biology*, vol. 305, no. 3, pp. 567–580, 2001.

[142] F. R. Kschischang and B. J. Frey, "Iterative decoding of compound codes by probability propagation in graphical models," *IEEE Selected Areas in Communications*, vol. 16, no. 2, pp. 219–230, 1998.

[143] 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, pp. 498–519, 2001.

[144] A. Kulesza and F. Pereira, "Structured learning with approximate inference," in *Advances in Neural Information Processing Systems*, pp. 785–792, Cambridge, MA: MIT Press, 2008.

[145] P. Kumar, V. Kolmogorov, and P. H. S. Torr, "An analysis of convex relaxations for MAP estimation," in *Advances in Neural Information Processing Systems*, pp. 1041–1048, Cambridge, MA: MIT Press, 2008.

[146] P. Kumar, P. H. S. Torr, and A. Zisserman, "Solving Markov random fields using second order cone programming," *IEEE Conference of Computer Vision and Pattern Recognition*, pp. 1045–1052, 2006.

[147] J. B. Lasserre, "An explicit equivalent positive semidefinite program for nonlinear 0–1 programs," *SIAM Journal on Optimization*, vol. 12, pp. 756–769, 2001.

[148] J. B. Lasserre, "Global optimization with polynomials and the problem of moments," *SIAM Journal on Optimization*, vol. 11, no. 3, pp. 796–817, 2001.

[149] M. Laurent, "Semidefinite relaxations for Max-Cut," in *The Sharpest Cut: Festschrift in Honor of M. Padberg's 60th Birthday*, New York: MPS-SIAM Series in Optimization, 2002.

[150] M. Laurent, "A comparison of the Sherali-Adams, Lovász-Schrijver and Lasserre relaxations for 0-1 programming," *Mathematics of Operations Research*, vol. 28, pp. 470–496, 2003.

[151] S. L. Lauritzen, *Lectures on Contingency Tables.* Department of Mathematics, Aalborg University, 1989.

[152] S. L. Lauritzen, "Propagation of probabilities, means and variances in mixed graphical association models," *Journal of the American Statistical Association*, vol. 87, pp. 1098–1108, 1992.

[153] S. L. Lauritzen, *Graphical Models.* Oxford: Oxford University Press, 1996.

[154] 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, Series B*, vol. 50, pp. 155–224, 1988.

[155] M. A. R. Leisink and H. J. Kappen, "Learning in higher order Boltzmann machines using linear response," *Neural Networks*, vol. 13, pp. 329–335, 2000.

[156] M. A. R. Leisink and H. J. Kappen, "A tighter bound for graphical models," in *Advances in Neural Information Processing Systems*, pp. 266–272, Cambridge, MA: MIT Press, 2001.

[157] H. A. Loeliger, "An introduction to factor graphs," *IEEE Signal Processing Magazine*, vol. 21, pp. 28–41, 2004.

[158] L. Lovász, "Submodular functions and convexity," in *Mathematical Programming: The State of the Art*, (A. Bachem, M. Grötschel, and B. Korte, eds.), pp. 235–257, New York: Springer-Verlag, 1983.

[159] L. Lovász and A. Schrijver, "Cones of matrices, set-functions and 0-1 optimization," *SIAM Journal of Optimization*, vol. 1, pp. 166–190, 1991.

[160] M. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. Spielman, "Improved low-density parity check codes using irregular graphs," *IEEE Transactions on Information Theory*, vol. 47, pp. 585–598, 2001.

[161] D. M. Malioutov, J. M. Johnson, and A. S. Willsky, "Walk-sums and belief propagation in Gaussian graphical models," *Journal of Machine Learning Research*, vol. 7, pp. 2013–2064, 2006.

[162] E. Maneva, E. Mossel, and M. J. Wainwright, "A new look at survey propagation and its generalizations," *Journal of the ACM*, vol. 54, no. 4, pp. 2–41, 2007.

[163] C. D. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press, 1999.

[164] S. Maybeck, *Stochastic Models, Estimation, and Control*. New York: Academic Press, 1982.

[165] J. McAuliffe, L. Pachter, and M. I. Jordan, "Multiple-sequence functional annotation and the generalized hidden Markov phylogeny," *Bioinformatics*, vol. 20, pp. 1850–1860, 2004.

[166] R. J. McEliece, D. J. C. McKay, and J. F. Cheng, "Turbo decoding as an instance of Pearl's belief propagation algorithm," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 2, no. 2, pp. 140–152, 1998.

[167] R. J. McEliece and M. Yildirim, "Belief propagation on partially ordered sets," in *Mathematical Theory of Systems and Networks*, (D. Gilliam and J. Rosenthal, eds.), Minneapolis, MN: Institute for Mathematics and its Applications, 2002.

[168] T. Meltzer, C. Yanover, and Y. Weiss, "Globally optimal solutions for energy minimization in stereo vision using reweighted belief propagation," in *International Conference on Computer Vision*, pp. 428–435, Silver Springs, MD: IEEE Computer Society, 2005.

[169] M. Mézard and A. Montanari, *Information, Physics and Computation*. Oxford: Oxford University Press, 2008.

[170] M. Mézard, G. Parisi, and R. Zecchina, "Analytic and algorithmic solution of random satisfiability problems," *Science*, vol. 297, p. 812, 2002.

[171] M. Mézard and R. Zecchina, "Random K-satisfiability: From an analytic solution to an efficient algorithm," *Physical Review E*, vol. 66, p. 056126, 2002.

[172] T. Minka, "Expectation propagation and approximate Bayesian inference," in *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence*, pp. 362–369, San Francisco, CA: Morgan Kaufmann, 2001.

[173] T. Minka, "Power EP," Technical Report MSR-TR-2004-149, Microsoft Research, October 2004.
[174] T. Minka and Y. Qi, "Tree-structured approximations by expectation propagation," in *Advances in Neural Information Processing Systems*, pp. 193–200, Cambridge, MA: MIT Press, 2004.
[175] T. P. Minka, "A family of algorithms for approximate Bayesian inference," PhD thesis, MIT, January 2001.
[176] C. Moallemi and B. van Roy, "Convergence of the min-sum message-passing algorithm for quadratic optimization," Technical Report, Stanford University, March 2006.
[177] C. Moallemi and B. van Roy, "Convergence of the min-sum algorithm for convex optimization," Technical Report, Stanford University, May 2007.
[178] J. M. Mooij and H. J. Kappen, "Sufficient conditions for convergence of loopy belief propagation," in *Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence*, pp. 396–403, Arlington, VA: AUAI Press, 2005.
[179] R. Neal and G. E. Hinton, "A view of the EM algorithm that justifies incremental, sparse, and other variants," in *Learning in Graphical Models*, (M. I. Jordan, ed.), Cambridge, MA: MIT Press, 1999.
[180] G. L. Nemhauser and L. A. Wolsey, *Integer and Combinatorial Optimization*. New York: Wiley-Interscience, 1999.
[181] Y. Nesterov, "Semidefinite relaxation and non-convex quadratic optimization," *Optimization methods and software*, vol. 12, pp. 1–20, 1997.
[182] M. Opper and D. Saad, "Adaptive TAP equations," in *Advanced Mean Field Methods: Theory and Practice*, (M. Opper and D. Saad, eds.), pp. 85–98, Cambridge, MA: MIT Press, 2001.
[183] M. Opper and O. Winther, "Gaussian processes for classification: Mean field algorithms," *Neural Computation*, vol. 12, no. 11, pp. 2177–2204, 2000.
[184] M. Opper and O. Winther, "Tractable approximations for probabilistic models: The adaptive Thouless-Anderson-Palmer approach," *Physical Review Letters*, vol. 64, p. 3695, 2001.
[185] M. Opper and O. Winther, "Expectation-consistent approximate inference," *Journal of Machine Learning Research*, vol. 6, pp. 2177–2204, 2005.
[186] J. G. Oxley, *Matroid Theory*. Oxford: Oxford University Press, 1992.
[187] M. Padberg, "The Boolean quadric polytope: Some characteristics, facets and relatives," *Mathematical Programming*, vol. 45, pp. 139–172, 1989.
[188] P. Pakzad and V. Anantharam, "Iterative algorithms and free energy minimization," in *Annual Conference on Information Sciences and Systems*, Princeton, NJ, 2002.
[189] P. Pakzad and V. Anantharam, "Estimation and marginalization using Kikuchi approximation methods," *Neural Computation*, vol. 17, no. 8, pp. 1836–1873, 2005.
[190] G. Parisi, *Statistical Field Theory*. New York: Addison-Wesley, 1988.
[191] P. Parrilo, "Semidefinite programming relaxations for semialgebraic problems," *Mathematical Programming, Series B*, vol. 96, pp. 293–320, 2003.
[192] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*. San Francisco, CA: Morgan Kaufmann, 1988.

[193] J. S. Pedersen and J. Hein, "Gene finding with a hidden Markov model of genome structure and evolution," *Bioinformatics*, vol. 19, pp. 219–227, 2003.
[194] P. Pevzner, *Computational Molecular Biology: An Algorithmic Approach*. Cambridge, MA: MIT Press, 2000.
[195] T. Plefka, "Convergence condition of the TAP equation for the infinite-ranged Ising model," *Journal of Physics A*, vol. 15, no. 6, pp. 1971–1978, 1982.
[196] L. R. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ: Prentice Hall, 1993.
[197] P. Ravikumar, A. Agarwal, and M. J. Wainwright, "Message-passing for graph-structured linear programs: Proximal projections, convergence and rounding schemes," in *International Conference on Machine Learning*, pp. 800–807, New York: ACM Press, 2008.
[198] P. Ravikumar and J. Lafferty, "Quadratic programming relaxations for metric labeling and Markov random field map estimation," *International Conference on Machine Learning*, pp. 737–744, 2006.
[199] T. Richardson and R. Urbanke, "The capacity of low-density parity check codes under message-passing decoding," *IEEE Transactions on Information Theory*, vol. 47, pp. 599–618, 2001.
[200] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge, UK: Cambridge University Press, 2008.
[201] B. D. Ripley, *Spatial Statistics*. New York: Wiley, 1981.
[202] C. P. Robert and G. Casella, *Monte Carlo statistical methods*. *Springer texts in statistics*, New York, NY: Springer-Verlag, 1999.
[203] G. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton University Press, 1970.
[204] T. G. Roosta, M. J. Wainwright, and S. S. Sastry, "Convergence analysis of reweighted sum-product algorithms," *IEEE Transactions on Signal Processing*, vol. 56, no. 9, pp. 4293–4305, 2008.
[205] P. Rusmevichientong and B. Van Roy, "An analysis of turbo decoding with Gaussian densities," in *Advances in Neural Information Processing Systems*, pp. 575–581, Cambridge, MA: MIT Press, 2000.
[206] S. Sanghavi, D. Malioutov, and A. Willsky, "Linear programming analysis of loopy belief propagation for weighted matching," in *Advances in Neural Information Processing Systems*, pp. 1273–1280, Cambridge, MA: MIT Press, 2007.
[207] S. Sanghavi, D. Shah, and A. Willsky, "Message-passing for max-weight independent set," in *Advances in Neural Information Processing Systems*, pp. 1281–1288, Cambridge, MA: MIT Press, 2007.
[208] L. K. Saul and M. I. Jordan, "Boltzmann chains and hidden Markov models," in *Advances in Neural Information Processing Systems*, pp. 435–442, Cambridge, MA: MIT Press, 1995.
[209] L. K. Saul and M. I. Jordan, "Exploiting tractable substructures in intractable networks," in *Advances in Neural Information Processing Systems*, pp. 486–492, Cambridge, MA: MIT Press, 1996.
[210] M. I. Schlesinger, "Syntactic analysis of two-dimensional visual signals in noisy conditions," *Kibernetika*, vol. 4, pp. 113–130, 1976.

[211] A. Schrijver, *Theory of Linear and Integer Programming*. New York: Wiley-Interscience Series in Discrete Mathematics, 1989.

[212] A. Schrijver, *Combinatorial Optimization: Polyhedra and Efficiency*. New York: Springer-Verlag, 2003.

[213] M. Seeger, "Expectation propagation for exponential families," Technical Report, Max Planck Institute, Tuebingen, November 2005.

[214] M. Seeger, F. Steinke, and K. Tsuda, "Bayesian inference and optimal design in the sparse linear model," in *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics*, San Juan, Puerto Rico, 2007.

[215] P. D. Seymour, "Matroids and multi-commodity flows," *European Journal on Combinatorics*, vol. 2, pp. 257–290, 1981.

[216] G. R. Shafer and P. P. Shenoy, "Probability propagation," *Annals of Mathematics and Artificial Intelligence*, vol. 2, pp. 327–352, 1990.

[217] H. D. Sherali and W. P. Adams, "A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems," *SIAM Journal on Discrete Mathematics*, vol. 3, pp. 411–430, 1990.

[218] A. Siepel and D. Haussler, "Combining phylogenetic and hidden Markov models in biosequence analysis," in *Proceedings of the Seventh Annual International Conference on Computational Biology*, pp. 277–286, 2003.

[219] D. Sontag and T. Jaakkola, "New outer bounds on the marginal polytope," in *Advances in Neural Information Processing Systems*, pp. 1393–1400, Cambridge, MA: MIT Press, 2007.

[220] T. P. Speed and H. T. Kiiveri, "Gaussian Markov distributions over finite graphs," *Annals of Statistics*, vol. 14, no. 1, pp. 138–150, 1986.

[221] R. P. Stanley, *Enumerative Combinatorics*. Vol. 1, Cambridge, UK: Cambridge University Press, 1997.

[222] R. P. Stanley, *Enumerative Combinatorics*. Vol. 2, Cambridge, UK: Cambridge University Press, 1997.

[223] E. B. Sudderth, M. J. Wainwright, and A. S. Willsky, "Embedded trees: Estimation of Gaussian processes on graphs with cycles," *IEEE Transactions on Signal Processing*, vol. 52, no. 11, pp. 3136–3150, 2004.

[224] E. B. Sudderth, M. J. Wainwright, and A. S. Willsky, "Loop series and Bethe variational bounds for attractive graphical models," in *Advances in Neural Information Processing Systems*, pp. 1425–1432, Cambridge, MA: MIT Press, 2008.

[225] C. Sutton and A. McCallum, "Piecewise training of undirected models," in *Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence*, pp. 568–575, San Francisco, CA: Morgan Kaufmann, 2005.

[226] C. Sutton and A. McCallum, "Improved dyamic schedules for belief propagation," in *Proceedings of the 23rd Conference on Uncertainty in Artificial Intelligence*, San Francisco, CA: Morgan Kaufmann, 2007.

[227] R. Szeliski, R. Zabih, D. Scharstein, O. Veskler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother, "Comparative study of energy minimization methods for Markov random fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 1068–1080, 2008.

[228] M. H. Taghavi and P. H. Siegel, "Adaptive linear programming decoding," in *International Symposium on Information Theory*, Seattle, WA, 2006.

[229] K. Tanaka and T. Morita, "Cluster variation method and image restoration problem," *Physics Letters A*, vol. 203, pp. 122–128, 1995.

[230] S. Tatikonda and M. I. Jordan, "Loopy belief propagation and Gibbs measures," in *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence*, pp. 493–500, San Francisco, CA: Morgan Kaufmann, 2002.

[231] Y. W. Teh and M. Welling, "On improving the efficiency of the iterative proportional fitting procedure," in *Proceedings of the Ninth International Conference on Artificial Intelligence and Statistics*, Key West, FL, 2003.

[232] A. Thomas, A. Gutin, V. Abkevich, and A. Bansal, "Multilocus linkage analysis by blocked Gibbs sampling," *Statistics and Computing*, vol. 10, pp. 259–269, 2000.

[233] A. W. van der Vaart, *Asymptotic Statistics*. Cambridge, UK: Cambridge University Press, 1998.

[234] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Review*, vol. 38, pp. 49–95, 1996.

[235] L. Vandenberghe, S. Boyd, and S. Wu, "Determinant maximization with linear matrix inequality constraints," *SIAM Journal on Matrix Analysis and Applications*, vol. 19, pp. 499–533, 1998.

[236] V. V. Vazirani, *Approximation Algorithms*. New York: Springer-Verlag, 2003.

[237] S. Verdú and H. V. Poor, "Abstract dynamic programming models under commutativity conditions," *SIAM Journal of Control and Optimization*, vol. 25, no. 4, pp. 990–1006, 1987.

[238] P. O. Vontobel and R. Koetter, "Lower bounds on the minimum pseudo-weight of linear codes," in *International Symposium on Information Theory*, Chicago, IL, 2004.

[239] P. O. Vontobel and R. Koetter, "Towards low-complexity linear-programming decoding," in *Proceedings of the International Conference on Turbo Codes and Related Topics*, Munich, Germany, 2006.

[240] M. J. Wainwright, "Stochastic processes on graphs with cycles: Geometric and variational approaches," PhD thesis, MIT, May 2002.

[241] M. J. Wainwright, "Estimating the "wrong" graphical model: Benefits in the computation-limited regime," *Journal of Machine Learning Research*, vol. 7, pp. 1829–1859, 2006.

[242] M. J. Wainwright, T. S. Jaakkola, and A. S. Willsky, "Tree-based reparameterization framework for analysis of sum-product and related algorithms," *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1120–1146, 2003.

[243] M. J. Wainwright, T. S. Jaakkola, and A. S. Willsky, "Tree-reweighted belief propagation algorithms and approximate ML, estimation by pseudomoment matching," in *Proceedings of the Ninth International Conference on Artificial Intelligence and Statistics*, 2003.

[244] M. J. Wainwright, T. S. Jaakkola, and A. S. Willsky, "Tree consistency and bounds on the max-product algorithm and its generalizations," *Statistics and Computing*, vol. 14, pp. 143–166, 2004.

[245] M. J. Wainwright, T. S. Jaakkola, and A. S. Willsky, "Exact MAP estimates via agreement on (hyper)trees: Linear programming and message-passing," *IEEE Transactions on Information Theory*, vol. 51, no. 11, pp. 3697–3717, 2005.

[246] M. J. Wainwright, T. S. Jaakkola, and A. S. Willsky, "A new class of upper bounds on the log partition function," *IEEE Transactions on Information Theory*, vol. 51, no. 7, pp. 2313–2335, 2005.

[247] M. J. Wainwright and M. I. Jordan, "Treewidth-based conditions for exactness of the Sherali-Adams and Lasserre relaxations," Technical Report 671, University of California, Berkeley, Department of Statistics, September 2004.

[248] M. J. Wainwright and M. I. Jordan, "Log-determinant relaxation for approximate inference in discrete Markov random fields," *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 2099–2109, 2006.

[249] Y. Weiss, "Correctness of local probability propagation in graphical models with loops," *Neural Computation*, vol. 12, pp. 1–41, 2000.

[250] Y. Weiss and W. T. Freeman, "Correctness of belief propagation in Gaussian graphical models of arbitrary topology," in *Advances in Neural Information Processing Systems*, pp. 673–679, Cambridge, MA: MIT Press, 2000.

[251] Y. Weiss, C. Yanover, and T. Meltzer, "MAP estimation, linear programming, and belief propagation with convex free energies," in *Proceedings of the 23rd Conference on Uncertainty in Artificial Intelligence*, Arlington, VA: AUAI Press, 2007.

[252] M. Welling, T. Minka, and Y. W. Teh, "Structured region graphs: Morphing EP into GBP," in *Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence*, pp. 609–614, Arlington, VA: AUAI Press, 2005.

[253] M. Welling and S. Parise, "Bayesian random fields: The Bethe-Laplace approximation," in *Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence*, Arlington, VA: AUAI Press, 2006.

[254] M. Welling and Y. W. Teh, "Belief optimization: A stable alternative to loopy belief propagation," in *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence*, pp. 554–561, San Francisco, CA: Morgan Kaufmann, 2001.

[255] M. Welling and Y. W. Teh, "Linear response for approximate inference," in *Advances in Neural Information Processing Systems*, pp. 361–368, Cambridge, MA: MIT Press, 2004.

[256] T. Werner, "A linear programming approach to max-sum problem: A review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 7, pp. 1165–1179, 2007.

[257] N. Wiberg, "Codes and decoding on general graphs," PhD thesis, University of Linkoping, Sweden, 1996.

[258] N. Wiberg, H. A. Loeliger, and R. Koetter, "Codes and iterative decoding on general graphs," *European Transactions on Telecommunications*, vol. 6, pp. 513–526, 1995.

[259] W. Wiegerinck, "Variational approximations between mean field theory and the junction tree algorithm," in *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence*, pp. 626–633, San Francisco, CA: Morgan Kaufmann Publishers, 2000.

[260] W. Wiegerinck, "Approximations with reweighted generalized belief propagation," in *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics*, pp. 421–428, Barbardos, 2005.

[261] W. Wiegerinck and T. Heskes, "Fractional belief propagation," in *Advances in Neural Information Processing Systems*, pp. 438–445, Cambridge, MA: MIT Press, 2002.

[262] A. S. Willsky, "Multiresolution Markov models for signal and image processing," *Proceedings of the IEEE*, vol. 90, no. 8, pp. 1396–1458, 2002.

[263] J. W. Woods, "Markov image modeling," *IEEE Transactions on Automatic Control*, vol. 23, pp. 846–850, 1978.

[264] N. Wu, *The Maximum Entropy Method.* New York: Springer, 1997.

[265] C. Yanover, T. Meltzer, and Y. Weiss, "Linear programming relaxations and belief propagation: An empirical study," *Journal of Machine Learning Research*, vol. 7, pp. 1887–1907, 2006.

[266] C. Yanover, O. Schueler-Furman, and Y. Weiss, "Minimizing and learning energy functions for side-chain prediction," in *Eleventh Annual Conference on Research in Computational Molecular Biology*, pp. 381–395, San Francisco, CA, 2007.

[267] J. S. Yedidia, "An idiosyncratic journey beyond mean field theory," in *Advanced Mean Field Methods: Theory and Practice*, (M. Opper and D. Saad, eds.), pp. 21–36, Cambridge, MA: MIT Press, 2001.

[268] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Generalized belief propagation," in *Advances in Neural Information Processing Systems*, pp. 689–695, Cambridge, MA: MIT Press, 2001.

[269] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Constructing free energy approximations and generalized belief propagation algorithms," *IEEE Transactions on Information Theory*, vol. 51, no. 7, pp. 2282–2312, 2005.

[270] A. Yuille, "CCCP algorithms to minimize the Bethe and Kikuchi free energies: Convergent alternatives to belief propagation," *Neural Computation*, vol. 14, pp. 1691–1722, 2002.

[271] G. M. Ziegler, *Lectures on Polytopes.* New York: Springer-Verlag, 1995.