that, if an eigenvalue $\lambda(A)$ of a matrix A is simple, the gradient of λ with respect to A is $\frac{\partial \lambda}{\partial A_{i,j}} = \frac{v_i u_j}{u^T v}$, where u and v are the left and right eigenvectors associatedd with the eigenvalue λ. Furthermore, when A is strictly positive, its largest eigenvalue is guaranteed to be simple by the Perron-Frobenius theorem.[8] Adding a small constant ϵ to the nonnegative matrix $W \odot W$ can make the function $\rho(W \odot W + \epsilon)$ strictly positive and therefore differentiable with a well-defined gradient, which can be computed as $\frac{\partial \rho(W \odot W + \epsilon)}{\partial W_{i,j}} = 2\frac{W_{ij} v_i u_j}{u^T v}$, where u and v are left and right eigenvectors of $W \odot W + \epsilon$.

References

1. A. Califano and M. J. Alvarez, The recurrent architecture of tumour initiation, progression and drug sensitivity, *Nature reviews Cancer* **17**, p. 116 (2017).
2. N. Friedman, M. Linial, I. Nachman and D. Pe'er, Using bayesian networks to analyze expression data, *Journal of computational biology* **7**, 601 (2000).
3. B. Zhang, C. Gaiteri, L.-G. Bodea, Z. Wang, J. McElwee, A. A. Podtelezhnikov, C. Zhang, T. Xie, L. Tran, R. Dobrin *et al.*, Integrated systems approach identifies genetic nodes and networks in late-onset alzheimers disease, *Cell* **153**, 707 (2013).
4. I.-M. Wang, B. Zhang, X. Yang, J. Zhu, S. Stepaniants, C. Zhang, Q. Meng, M. Peters, Y. He, C. Ni *et al.*, Systems analysis of eleven rodent disease models reveals an inflammatome signature and key drivers, *Molecular systems biology* **8** (2012).
5. D. M. Chickering, D. Heckerman and C. Meek, Large-sample learning of bayesian networks is np-hard, *Journal of Machine Learning Research* **5**, 1287 (2004).
6. X. Zheng, B. Aragam, P. K. Ravikumar and E. P. Xing, Dags with no tears: Continuous optimization for structure learning, in *Advances in Neural Information Processing Systems*, 2018.
7. J. C. Liao, R. Boscolo, Y.-L. Yang, L. M. Tran, C. Sabatti and V. P. Roychowdhury, Network component analysis: reconstruction of regulatory signals in biological systems, *Proceedings of the National Academy of Sciences* **100**, 15522 (2003).
8. C. D. Meyer, *Matrix analysis and applied linear algebra* (Siam, 2000).
9. T. Miyato, T. Kataoka, M. Koyama and Y. Yoshida, Spectral normalization for generative adversarial networks, *arXiv preprint arXiv:1802.05957* (2018).
10. M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin *et al.*, Tensorflow: Large-scale machine learning on heterogeneous distributed systems, *arXiv preprint arXiv:1603.04467* (2016).
11. D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980* (2014).
12. A. Tarvainen and H. Valpola, Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results, in *Advances in neural information processing systems*, 2017.
13. A. Irrthum, L. Wehenkel, P. Geurts *et al.*, Inferring regulatory networks from expression data using tree-based methods, *PloS one* **5**, p. e12776 (2010).
14. T. Van den Bulcke, K. Van Leemput, B. Naudts, P. van Remortel, H. Ma, A. Verschoren, B. De Moor and K. Marchal, Syntren: a generator of synthetic gene expression data for design and analysis of structure learning algorithms, *BMC bioinformatics* **7**, p. 43 (2006).
15. T. Schaffter, D. Marbach and D. Floreano, Genenetweaver: in silico benchmark generation and performance profiling of network inference methods, *Bioinformatics* **27**, 2263 (2011).
16. A. Statnikov, S. Ma, M. Henaff, N. Lytkin, E. Efstathiadis, E. R. Peskin and C. F. Aliferis, Ultra-scalable and efficient methods for hybrid observational and experimental local causal pathway discovery, *The Journal of Machine Learning Research* **16**, 3219 (2015).