A neural network oracle for quantum nonlocality problems in networks

Tamás Kriváchy
Department of Applied Physics
University of Geneva
CH-1211 Geneva, Switzerland
`tamas.krivachy@unige.ch`

Yu Cai
Department of Applied Physics
University of Geneva
CH-1211 Geneva, Switzerland

Daniel Cavalcanti
ICFO, The Institute of Photonic Sciences
University of Geneva
08860 Castelldefels (Barcelona), Spain

Arash Tavakoli
Dyson School of Design Engineering
Imperial College London
London SW7 2AZ, UK

Nicolas Gisin
Department of Applied Physics
University of Geneva
CH-1211 Geneva, Switzerland

Nicolas Brunner
Department of Applied Physics
University of Geneva
CH-1211 Geneva, Switzerland

1 Introduction

The possibility of creating stronger than classical correlations between distant parties has deep implications for both the foundations and applications of quantum theory. These ideas have been initiated by Bell [1], with subsequent research leading to the theory of Bell nonlocality [2]. In the Bell scenario multiple parties jointly share a single classical or quantum source, often referred to as local and nonlocal sources, respectively[1]. Recently, interest in more elaborate causal structures, in which several independent sources are shared among the parties over a network, has been on the rise [3–5]. Contrary to the Bell scenario, in even slightly more complex networks the boundary between local and nonlocal correlations becomes nonlinear and the local set non-convex, greatly perplexing rigorous analysis. Though some progress has been made [6–20], we still lack a robust set of tools to investigate generic networks from an analytic and numerical perspective.

---

[1]Note that in this paper by the notion of (non)locality we refer to the (non)locality in the context of Bell nonlocality, which is conceptually different from that of (non)locality in condensed matter literature.
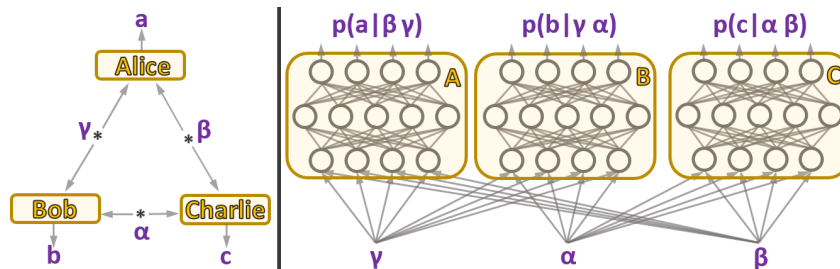
Figure 1: (Left) Triangle network configuration. (Right) Neural network which reproduces local distributions compatible with the triangle configuration.

Here we explore the use of machine learning in these problems. In particular we tackle the membership problem for causal structures, i.e. given a network and a distribution over the observed outputs, we must decide whether it could have been produced by using exclusively local resources. We encode the causal structure into a neural network and ask the network to reproduce the target distribution. By doing so, we approximate the question "does a local causal model *exist*?" with "is a local causal model *learnable*?". Neural networks have proven to be useful ansätze for generic nonlinear functions in terms of expressivity, ease of learning and robustness, both in- and outside the domain of physical sciences [21–25]. They have also been used in the study of nonlocality, however, we note that our method is significantly different from previous ones [26, 27], in particular because we use generative networks which allow us to obtain explicit local models in previously unreachable domains.

In our approach we exploit that information flow in both causal structures and feedforward neural networks is determined by a directed acyclic graph. For any given distribution over observed variables and an ansatz causal structure, we train a neural network which respects that causal structure to reproduce the target distribution. This is equivalent to having a neural network learn the local responses of the parties to their inputs. If the target distribution is inside the local set, then a sufficiently expressive neural network should be able to learn the appropriate response functions and reproduce it. For distributions outside the local set, we should see that the machine can not approximate the given target. This gives us a criterion for deciding whether a target distribution is inside the local set or not. In particular, if a given distribution is truly outside the local set, then by adding noise in a physically relevant way we should see a clear transition in the machine's behavior when entering the set of local correlations. The application for a quantum information theorist is then straightforward: dream up a set of quantum states and measurements in a network and then check whether it is genuinely quantum or not, i.e. whether it can not be reproduced using local resources or whether it can.

We explore the strength of this method by examining a notorious causal structure, the so-called 'triangle' network, depicted in Fig. 1. The triangle configuration is among the simplest tripartite networks, yet it poses immense challenges theoretically and numerically. After checking for the consistency of our method with known results, we examine the distribution proposed in [28], which we refer to as the Elegant distribution from here on. Our method gives strong evidence that the Elegant distribution is outside the local set, as originally conjectured. We also use our method to get an estimate of the noise robustness of this nonlocal distribution. Astonishingly, the method gives us a new conjecture of quantum nonlocality for a set of distributions: when examining the distribution in [17], we find that our causal inference strongly suggests the proposed distribution to be nonlocal in a regime where intuition would suggest that it is local. This exhibits that the method is not only useful for verifying previous conjectures, but also for creating new ones, and thus allows for very intricate collaboration between physicists and artificial intelligence. Our implementation of the algorithm is available at https://github.com/tkrivachy/neural-network-for-nonlocality-in-networks.

2   Encoding causal structures into neural networks

The methods developed in this work are in principle applicable to any causal structure. We will demonstrate how to encode a network nonlocality configuration into a neural network on the simple, yet non-trivial example of the triangle network with quaternary outputs and no inputs, since it is one of the key examples which continues to resist other methods. In this scenario three sources, α, β, γ, otherwise known as latent variables, send information through either a classical or a quantum channel to three parties, Alice, Bob and Charlie. Flow of information is constrained such that the sources are independent from each other, and each one only sends information to two parties of the three, as depicted in Fig. 1. In each round, Alice, Bob and Charlie process their inputs with arbitrary local response functions, and they each output a number $a, b, c \in \{0, 1, 2, 3\}$, respectively. Under the assumption that each source is independent and identically distributed from round to round, and that the local response functions are fixed, such a scenario is well characterized by the probability distribution $p(abc)$ over the random variables of the outputs.

For the classical (local) setup we can assume without loss of generality that the sources each send a random variable drawn from a uniform distribution on the continuous interval between 0 and 1. If we also incorporate the network constraint then the probability distribution over the parties' outputs can

be written as

$$p(abc) = \int_0^1 d\alpha d\beta d\gamma \, p_A(a|\beta\gamma)p_B(b|\gamma\alpha)p_C(c|\alpha\beta). \tag{1}$$

We now construct a neural network which is able to approximate a distribution of the form (1). We use a feedforward neural network, since it is described by a directed acyclic graph, similarly to a causal structure [29–31]. This allows for a seamless transfer from the causal structure to the neural network model. We simply take the hidden variables to be inputs and the conditional probabilities $p_A(a|\beta\gamma), p_B(b|\gamma\alpha)$ and $p_C(c|\alpha\beta)$ to be the outputs, for each possible value of a, b, c. So as to respect the communication constraints of the triangle, the neural network is not fully connected, as shown in Fig. 1. We evaluate the neural network for N_{batch} values of α, β, γ in order to approximate the joint probability distribution (1) with a Monte Carlo approximation,

$$p_M(abc) = \sum_{i=1}^{N_{\text{batch}}} p_A(a|\beta_i\gamma_i)p_B(b|\gamma_i\alpha_i)p_C(c|\alpha_i\beta_i). \tag{2}$$

The cost function can be any measure of discrepancy between the target distribution p_t and the neural network's output p_M, such as the Kullback–Leibler divergence. In order to train the neural network we synthetically generate uniform random numbers for the hidden variables, the inputs. We then adjust the weights of the network after evaluating a minibatch of size N_{batch} according to (2) using conventional neural network optimization methods [29].

By encoding the causal structure in a neural network like this, we can train the neural network to try to reproduce a given target distribution. The procedure generalizes in a straight-forward manner to any causal structure, and is thus in principle applicable to any quantum nonlocality network problem. After having developed the method we found that in the machine learning community a similar technique has been explored for continuous random variables on causal networks, which further strengthens the validity of the technique [32].

3 Results

The most informative way to use the method is to search for transitions in the machine's behavior as we ask it to learn different target distributions. We start from a target distribution which we are truly
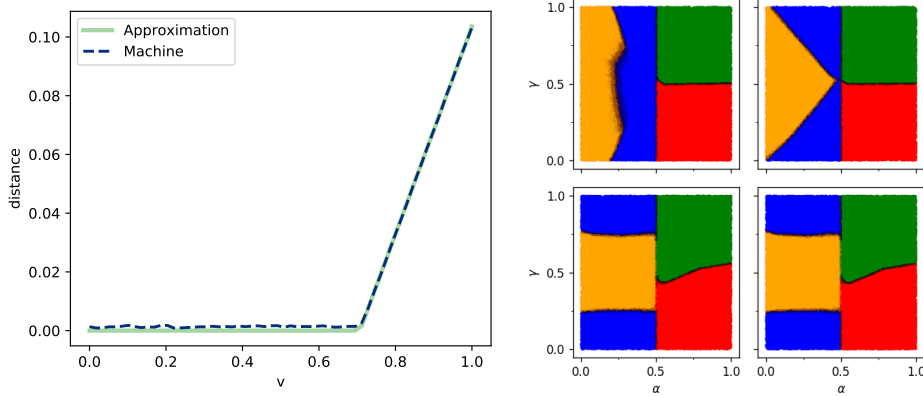

Figure 2: Fritz distribution results. (Left) Plot of the distance perceived by the machine (dashed blue) and the approximation of the distance (solid green) for $\hat{v}^* = 1/\sqrt{2}$ and $\hat{\theta} = 90°$. (Right) Response functions of Bob for $v = 0, 0.44, 0.71, 1$, from top left to bottom right, respectively. We sample several thousand pairs of $\{\alpha, \gamma\}$ and for each pair we sample 30 times from Bob's response $p_B(b|\alpha, \gamma)$ and color-code the 4 possible outcomes, $b = 0, 1, 2, 3$ in red, green, blue and yellow, respectively, with some finite opacity. As a result, solid colors exhibit deterministic responses of Bob to his latent variables and more messy, brown colors show stochastic responses. Note how the complete response maps for $v > \hat{v}^*$ are the same.
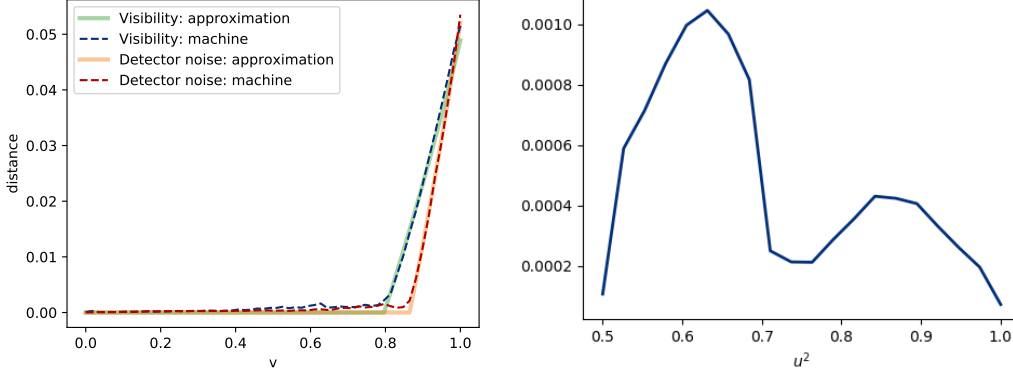
Figure 3: (Left) Elegant distribution: comparison of the distance perceived by the machine, $d_M(v)$ and the approximation of the distance for visibility and detector noise models. (Right) We examine the distance from the local set as perceived by the machine for the family of distributions proposed by Renou *et al.*, without adding any noise this time. The large distance for $0.5 < u^2 < 0.785$ is a surprising feature leading to conjectured nonlocality in that regime.

interested in $p_t(v = 1)$, and by adding noise characterized by the parameter $v \in [0, 1]$, we force the target distribution to become more local until it becomes the completely noisy distribution $p_t(v = 0)$. If the distribution $p_t(v = 1)$ was truly nonlocal then at some transition point v^* the curve $p_t(v)$ enters the local set. This should be observable in the machine's behavior since, by construction, it is able to reproduce only local distributions. We examine the Euclidean distance between the target distribution $p_t(v)$ and the learned distribution $p_M(v)$ and compare it to an approximation of what we should see if the curve $p_t(v)$ enters the local set at v^* and at an angle θ.

First let us consider the quantum distribution proposed by Fritz [5], which is well understood and can be viewed as the CHSH Bell scenario wrapped into the triangle topology. In Fig. 2 we plot the distance discussed previously and compare it to the approximation with the two fitting parameters $\hat{\theta} = 90°$ and $\hat{v}^* = \frac{1}{\sqrt{2}}$, which is indeed what we expect for this scenario. The coincidence of the two curves is already good evidence that the machine finds the closest local distributions to the target distributions. Upon examining the response functions of Alice, Bob and Charlie, (Bob's responses displayed in Fig. 2), we see that they do not change above $\hat{v}^*$, which means that the machine finds the same distributions for target distributions outside the local set. This is in line with our expectations. Due to the connection with the CHSH Bell scenario, we believe the curve $p_t(v)$ exits the local set perpendicularly. It is peculiar how the machine often prefers horizontal and vertical separations of the latent variable space, with very clean, deterministic responses, similarly to how we would do it intuitively. These results reaffirm that our algorithm functions well.

Next we turn our attention to a distribution which is more native to the triangle structure, as it combines entangled states and entangled measurements, namely the Elegant distribution, which is conjectured in [28] to be outside the local set. It is also more interesting since it is an open question whether this distribution is nonlocal or not. We examine two noise models - one at the sources (α, β, γ), which we call the visibility noise model, and one at the detectors (Alice, Bob and Charlie), the detector noise model. For both noise models we see a transition in the distance $d_M(v)$, depicted in Fig. 3, giving us strong evidence that the conjectured distribution is indeed nonlocal. Through this examination we gain insight into the noise robustness of the Elegant distribution as well. It seems that for visibilities above $\hat{v}^* \approx 0.80$, or for detectors functioning above $\hat{v}^* \approx 0.86$, the distribution is still nonlocal.

Finally we consider the one-parameter family of distributions proposed in [17] characterized by the parameter $u \in [\frac{1}{\sqrt{2}}, 1]$. It also utilizes entangled states and measurements, as the Elegant distribution. Note that here u is not a noise parameter now, but each u defines a distinct, interesting target distribution. These distributions are proven to be nonlocal for $u^2_{\text{crit.}} < u^2 < 1$ (with $u^2_{\text{crit.}} \approx 0.785$) and local for $u^2 \in \{0.5, u^2_{\text{crit.}}, 1\}$. Since the distribution is nonlocal above $u_{\text{crit.}}$ and local at precisely that value, one might be inclined to think that below $u_{\text{crit.}}$ the distribution is local. However as is

4

illustrated in Fig 3, the machine finds those points to be *even more nonlocal* than those for which we have a proof of nonlocality, in the sense that they are farther from the local set according to the Euclidean distance. This leads us to conjecture that this distribution is nonlocal also for values of $0.5 < u^2 < u^2_{\text{crit.}}$.

4 Summary

In conclusion, we demonstrate a method for causal inference over discrete observed random variables, which is particularly useful for quantum network nonlocality problems, in which we wish to know whether a given target distribution is nonlocal or not. We deduce nonlocality of a distribution based on whether the machine manages to learn the distribution or not. The task is an otherwise infeasible optimization problem, however using the technique introduced here we could apply this method to an open problem, the nonlocality of the Elegant distribution. Moreover, we obtained a conjecture of quantum nonlocality from the machine by examining distributions proposed in [17]. These two examples show that this generative method promises to provide more fruitful results in the future, especially when used in collaboration with domain-specific analytical insight, while still giving interpretable results.

References

[1] J. S. Bell. On the Einstein Podolsky Rosen paradox. *Physics Physique Fizika*, 1(3):195–200, November 1964.

[2] Nicolas Brunner, Daniel Cavalcanti, Stefano Pironio, Valerio Scarani, and Stephanie Wehner. Bell nonlocality. *Reviews of Modern Physics*, 86(2):419–478, April 2014.

[3] C. Branciard, N. Gisin, and S. Pironio. Characterizing the Nonlocal Correlations Created via Entanglement Swapping. *Physical Review Letters*, 104(17):170401, April 2010.

[4] Cyril Branciard, Denis Rosset, Nicolas Gisin, and Stefano Pironio. Bilocal versus nonbilocal correlations in entanglement-swapping experiments. *Physical Review A*, 85(3):032119, March 2012.

[5] Tobias Fritz. Beyond Bells theorem: correlation scenarios. *New Journal of Physics*, 14(10):103001, October 2012.

[6] Joe Henson, Raymond Lal, and Matthew F. Pusey. Theory-independent limits on correlations from generalized Bayesian networks. *New Journal of Physics*, 16(11):113043, November 2014.

[7] Armin Tavakoli, Paul Skrzypczyk, Daniel Cavalcanti, and Antonio Acín. Nonlocal correlations in the star-network configuration. *Physical Review A*, 90(6):062109, December 2014.

[8] Rafael Chaves, Christian Majenz, and David Gross. Information–theoretic implications of quantum causal structures. *Nature Communications*, 6:5766, January 2015.

[9] Elie Wolfe, Robert W. Spekkens, and Tobias Fritz. The Inflation Technique for Causal Inference with Latent Variables. *arXiv:1609.00672 [quant-ph, stat]*, September 2016. arXiv: 1609.00672.

[10] Denis Rosset, Cyril Branciard, Tomer Jack Barnea, Gilles Pütz, Nicolas Brunner, and Nicolas Gisin. Nonlinear Bell Inequalities Tailored for Quantum Networks. *Physical Review Letters*, 116(1):010403, January 2016.

[11] Miguel Navascues and Elie Wolfe. The Inflation Technique Completely Solves the Causal Compatibility Problem. *arXiv:1707.06476 [quant-ph, stat]*, July 2017. arXiv: 1707.06476.

[12] Denis Rosset, Nicolas Gisin, and Elie Wolfe. Universal bound on the cardinality of local hidden variables in networks. *Quantum Information and Computation*, 18(11&12), September 2017. arXiv: 1709.00707.

[13] Rafael Chaves. Polynomial Bell Inequalities. *Physical Review Letters*, 116(1):010402, January 2016.

[14] Thomas C. Fraser and Elie Wolfe. Causal compatibility inequalities admitting quantum violations in the triangle structure. *Physical Review A*, 98(2):022113, August 2018.

[15] Mirjam Weilenmann and Roger Colbeck. Non-Shannon inequalities in the entropy vector approach to causal structures. *Quantum*, 2:57, March 2018.

[16] Ming-Xing Luo. Computationally Efficient Nonlinear Bell Inequalities for Quantum Networks. *Physical Review Letters*, 120(14):140402, April 2018.

[17] Marc-Olivier Renou, Elisa Bäumer, Sadra Boreiri, Nicolas Brunner, Nicolas Gisin, and Salman Beigi. Genuine quantum nonlocality in the triangle network. *arXiv:1905.04902 [quant-ph]*, May 2019. arXiv: 1905.04902.

[18] Nicolas Gisin, Jean-Daniel Bancal, Yu Cai, Armin Tavakoli, Emmanuel Zambrini Cruzeiro, Sandu Popescu, and Nicolas Brunner. Constraints on nonlocality in networks from no-signaling and independence. *arXiv:1906.06495 [quant-ph]*, June 2019. arXiv: 1906.06495.

[19] Marc-Olivier Renou, Yuyi Wang, Sadra Boreiri, Salman Beigi, Nicolas Gisin, and Nicolas Brunner. Limits on correlations in networks for quantum and no-signaling resources. *arXiv:1901.08287 [quant-ph]*, January 2019. arXiv: 1901.08287.

[20] Alejandro Pozas-Kerstjens, Rafael Rabelo, Łukasz Rudnicki, Rafael Chaves, Daniel Cavalcanti, Miguel Navascues, and Antonio Acín. Bounding the sets of classical and quantum correlations in networks. *arXiv:1904.08943 [quant-ph]*, April 2019. arXiv: 1904.08943.

[21] Roger G. Melko, Giuseppe Carleo, Juan Carrasquilla, and J. Ignacio Cirac. Restricted Boltzmann machines in quantum physics. *Nature Physics*, page 1, June 2019.

[22] Raban Iten, Tony Metger, Henrik Wilming, Lidia del Rio, and Renato Renner. Discovering physical concepts with neural networks. *arXiv:1807.10300 [physics, physics:quant-ph]*, July 2018. arXiv: 1807.10300.

[23] Alexey A. Melnikov, Hendrik Poulsen Nautrup, Mario Krenn, Vedran Dunjko, Markus Tiersch, Anton Zeilinger, and Hans J. Briegel. Active learning machine learns to create new quantum experiments. *Proceedings of the National Academy of Sciences*, 115(6):1221–1226, February 2018.

[24] Evert P. L. van Nieuwenburg, Ye-Hua Liu, and Sebastian D. Huber. Learning phase transitions by confusion. *Nature Physics*, 13(5):435–439, May 2017.

[25] Juan Carrasquilla and Roger G. Melko. Machine learning phases of matter. *Nature Physics*, 13(5):431–434, May 2017.

[26] Dong-Ling Deng. Machine Learning Detection of Bell Nonlocality in Quantum Many-Body Systems. *Physical Review Letters*, 120(24):240402, June 2018.

[27] Askery Canabarro, Samuraí Brito, and Rafael Chaves. Machine Learning Nonlocal Correlations. *Physical Review Letters*, 122(20):200401, May 2019.

[28] Nicolas Gisin. Entanglement 25 Years after Quantum Teleportation: Testing Joint Measurements in Quantum Networks. *Entropy*, 21(3):325, March 2019.

[29] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.

[30] Judea Pearl. *Causality : models reasoning and inference*. Cambridge University Press, 2000.

[31] Daphne Koller and Nir Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

[32] Olivier Goudet, Diviyan Kalainathan, Philippe Caillou, Isabelle Guyon, David Lopez-Paz, and Michèle Sebag. Learning Functional Causal Models with Generative Neural Networks. In Hugo Jair Escalante, Sergio Escalera, Isabelle Guyon, Xavier Baró, Yağmur Güçlütürk, Umut Güçlü, and Marcel van Gerven, editors, *Explainable and Interpretable Models in Computer Vision and Machine Learning*, The Springer Series on Challenges in Machine Learning, pages 39–80. Springer International Publishing, Cham, 2018.