Towards a Trace-Preserving Tensor Network Representation of Quantum Channels

Siddarth Srinivasan
University of Washington

Sandesh Adhikary
University of Washington

Jacob Miller
Université de Montréal

Bibek Pokharel
University of Southern California

Guillaume Rabusseau
Université de Montréal

Byron Boots
University of Washington

Abstract

The problem of characterizing quantum channels arises in a number of contexts such as quantum process tomography and quantum error correction. However, direct approaches to parameterizing and optimizing the Choi matrix representation of quantum channels face a curse of dimensionality: the number of parameters scales exponentially in the number of qubits. Recently, Torlai et al. [2020] proposed using locally purified density operators (LPDOs), a tensor network representation of Choi matrices, to overcome the unfavourable scaling in parameters. While the LPDO structure allows it to satisfy a 'complete positivity' (CP) constraint required of physically valid quantum channels, it makes no guarantees about a similarly required 'trace preservation' (TP) constraint. In practice, the TP constraint is violated, and the learned quantum channel may even be trace-increasing, which is non-physical. In this work, we present the problem of optimizing over TP LPDOs, discuss two approaches to characterizing the TP constraints on LPDOs, and outline the next steps for developing an optimization scheme.

1 Introduction

Quantum channels describe the physical evolution of quantum systems, which are generally represented as unit-trace positive semi-definite (PSD) *density matrices*. A physically valid quantum channel must preserve the constraints on input density matrices by satisfying a 'complete positivity' (CP) condition (which ensures that PSD inputs are mapped to PSD outputs), and a 'trace preservation' (TP) condition (which ensures that the channel preserves the trace of the inputs). The problem of optimization over physical CPTP quantum channels is encountered in a variety of applications. Quantum process tomography (QPT) aims to characterize the quantum channel by learning it from measurement data. In quantum error correction, we may wish to optimize over quantum channels to design error-correcting codes for a given error model. *Choi matrices* are a convenient representation of quantum channels for such optimization problems since the CPTP constraints can be easily specified: the CP condition amounts to a PSD requirement on the matrix, and the TP condition is the requirement that the partial trace over the output subsystems is the identity on the input subsystems.

However, the number of parameters characterizing a Choi matrix scales exponentially with the number of qubits. For an N-qubit quantum channel, the Choi matrix has dimension $2^{2N} \times 2^{2N}$, making it intractable to directly optimize over the full matrix for a large number of qubits. One approach that avoids this curse of dimensionality is to factorize the Choi matrix into a *matrix product operator* (MPO) consisting of a series of N compact tensor cores. However, such an MPO must still represent quantum channels which are CP and TP, both of which are non-trivial constraints that must be explicitly satisfied in the optimization algorithm. Recently, Torlai et al. [2020] proposed parameterizing and optimizing quantum channels in the process tomography setting using locally purified density operators (LPDOs),

Second Workshop on Quantum Tensor Networks in Machine Learning, 35th Conference on Neural Information Processing Systems (NeurIPS 2021)

a tensor network factorization that is guaranteed to satisfy the CP requirement *by design* [Werner et al., 2016]. They show how LPDO representations of quantum channels can be learned using gradient descent with a maximium likelihood objective on several synthetic datasets.

While the LPDO structure forces its corresponding Choi matrix to satisfy the CP constraint, it makes no guarantees about the TP constraint. Torlai et al. [2020] argue that as the amount of data increases, their approach will naturally produce a TP LPDO. In practice, they recommend penalizing the maximum likelihood objective by an amount proportional to the TP violation. However, as demonstrated on some small-scale synthetic experiments, this strategy can still only approximately satisfy the TP constraint; the TP violation is many orders of magnitude above machine precision. Moreover, there is no known scheme to even project an arbitrary LPDO onto the set of *trace-preserving* LPDOs or Choi matrices without first reconstructing the full Choi matrix, which would defeat the purpose of the tensor network approximation. For various applications, it may be desirable to parameterize and learn a physically valid LPDO structure that also satisfies the TP constraint. This brings us to the problem we consider in this paper: how can we parameterize and optimize trace-preserving LPDOs? Answering this question will unlock the ability to employ *physical* LPDO approximations of quantum channels for various tasks including quantum process tomography and quantum error correction, particularly for large quantum systems with many qubits.

We make the following contributions in this paper: 1) we pose and discuss the problem of parameterizing TP LPDOs, 2) we show that a naive constraint imposed independently on each LPDO tensor core is sufficient to guarantee the TP property of the LPDO, but is too strict to capture simple 2-qubit gates, 3) we provide an alternative characterization of the constraints on each LPDO tensor core, including a result that the tensor cores of TP LPDOs must be orthonormal 2-frames under a metric induced by the other cores, and 4) and chart the path forward for developing an optimization scheme for TP LPDOs. We intend for this work to discuss the prospects of developing an optimization algorithm for LPDO decompositions of physically valid quantum channels.

The rest of this paper is structured as follows: in Section 2, we provide some preliminaries on quantum channels, Choi matrices, and LPDOs; in Section 3, we discuss TP LPDOs and two different approaches to constraining LPDOs to satisfy the TP constraint, and we conclude in Section 4 with some thoughts on directions for future work.

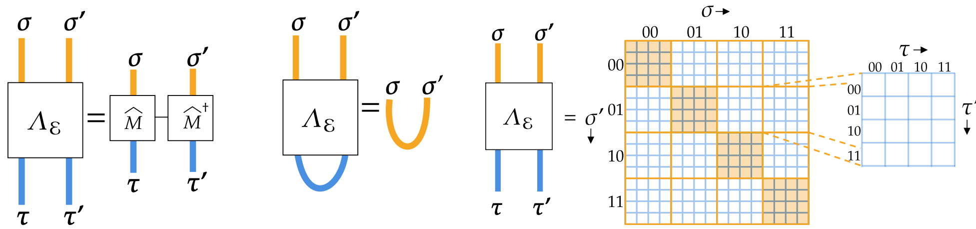
2 Quantum Channels, Choi Matrices, and LPDOs

Quantum Channels We consider a quantum channel acting on a system of N qubits. Let $\mathcal{H}_A$ denote the 2^N-dimensional Hilbert space of pure input states with basis states $\{|\boldsymbol{\sigma}\rangle\}$ and $\mathcal{H}_B$ denote the 2^N-dimensional Hilbert space of pure output states with basis states $\{|\boldsymbol{\tau}\rangle\}$. Mixed input and output states live in $L(\mathcal{H}_A)$ and $L(\mathcal{H}_B)$, which are spaces of unit-trace, positive semi-definite (PSD) density matrices. A *quantum channel* is a linear map $\mathcal{E}: L(\mathcal{H}_A) \to L(\mathcal{H}_B)$ from density matrices in $L(\mathcal{H}_A)$ to density matrices in $L(\mathcal{H}_B)$. Such a map satisfies two[1] constraints: 1) it must be completely positive (CP), i.e., preserve not only the PSD property of input density matrices, but also the PSD property of input density matrices when coupled with an arbitrary dimensional ancilla system, and 2) it must be trace-preserving (TP), i.e., leave the trace of input density matrices unchanged.

Choi Matrices A quantum channel $\mathcal{E} : L(\mathcal{H}_A) \to L(\mathcal{H}_B)$ can be represented by a $2^{2N} \times 2^{2N}$ complex-valued *Choi matrix* as $\mathbf{\Lambda}_{\mathcal{E}} = \sum_{\boldsymbol{\sigma},\boldsymbol{\tau}} |\boldsymbol{\tau}\rangle\langle\boldsymbol{\tau}'| \otimes \mathcal{E}(|\boldsymbol{\sigma}\rangle\langle\boldsymbol{\sigma}'|)$. This permits a simple characterization of the CP-TP constraints: the CP conditions amount to a requirement that $\mathbf{\Lambda}_{\mathcal{E}}$ be a PSD matrix, and the TP condition requires that the partial trace over the output states be identity $\mathrm{tr}_B(\mathbf{\Lambda}_{\mathcal{E}}) = \mathbb{I}_{2^N}$. These constraints are illustrated in the language of tensor diagrams in Figure 1a. We can think of the Choi matrix as having a block structure, where the input basis states $\boldsymbol{\sigma}$, $\boldsymbol{\sigma}'$ pick out a block of the Choi matrix, and output basis states $\boldsymbol{\tau}, \boldsymbol{\tau}'$ pick out a specific entry within the block. This block structure is illustrated in Figure 1b.

The task of estimating the Choi matrix corresponding to a quantum channel (process tomography) is more complicated than learning quantum states (state tomography). While the Choi-Jamiolskowski isomorphism [Choi, 1975, Jamiołkowski, 1972] gives a connection between bipartite states and Choi matrices, this is not a true isomorphism [Jiang et al., 2013] since the TP condition requires the input

[1]The channel must also preserve the Hermiticity of input density matrices, but this is already guaranteed by the CP constraint.



(a) The complete positivity (**left**) and trace preservation (**right**) constraints on Choi matrices.

(b) Block structure of Choi matrix; indices (σ, σ') pick out matrix blocks and indices (τ, τ') pick out elements within the blocks.

Figure 1: Illustrations of Choi matrices and their properties

subsystem of the bipartite state corresponding to a Choi matrix to be the identity. In other words, diagonal blocks of the Choi matrix must be trace unit trace, and off-diagonal blocks must be zero trace. Thus, the problem is not as simple as learning parameters first and renormalizing with a constant at the end, since the diagonal blocks and off-diagonal blocks are (trace) normalized differently. One solution is to simply learn a Choi matrix via unconstrained optimization and project it onto the set of CP-TP matrices at the end [Knee et al., 2018, Surawy-Stepney et al., 2021]. Another solution is to observe that a Choi matrix is CP-TP if and only if the tensor core $\hat{\mathbf{M}}$ in Figure 1a (square root of the Choi matrix $[\boldsymbol{\Lambda}_{\mathcal{E}}]^{\boldsymbol{\sigma}',\boldsymbol{\tau}'}_{\boldsymbol{\sigma},\boldsymbol{\tau}} = (\hat{\mathbf{M}}^{\boldsymbol{\sigma},\boldsymbol{\tau}}_{\mu})^{\dagger}\hat{\mathbf{M}}^{\boldsymbol{\sigma}',\boldsymbol{\tau}'}_{\mu}$) lives on a Stiefel manifold[2], and use Riemannian optimization techniques [Luchnikov et al., 2021] to learn the Choi matrix. This is an important characterization of the TP constraint, so we state it formally below for later use.

Remark 1. *Let* $[\boldsymbol{\Lambda}_{\mathcal{E}}]^{\boldsymbol{\sigma}',\boldsymbol{\tau}'}_{\boldsymbol{\sigma},\boldsymbol{\tau}} = ([\hat{\mathbf{M}}]^{\boldsymbol{\sigma},\boldsymbol{\tau}}_{\mu})^{\dagger}[\hat{\mathbf{M}}]^{\boldsymbol{\sigma}',\boldsymbol{\tau}'}_{\mu}$ *and* $[\mathbf{M}]^{\boldsymbol{\sigma}}_{\boldsymbol{\tau},\mu}$ *be a reshape of* $[\hat{\mathbf{M}}]^{\boldsymbol{\sigma},\boldsymbol{\tau}}_{\mu}$*. Then, the TP condition on the Choi matrix is equivalently stated as* $\mathbf{M}^{\dagger}\mathbf{M} = \mathbb{I}$*.*

However, the other complication in estimating Choi matrices from data is more severe: since the number of parameters scales exponentially with the number of qubits, learning the Choi matrix quickly becomes intractable for large quantum systems. This general problem setting is fertile ground for tensor network solutions, which we consider next.

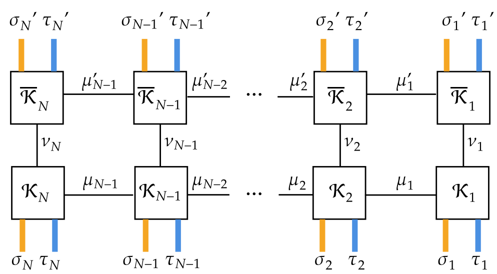


Figure 2: Tensor network diagram of an LPDO

Locally Purified Density Operators The Locally Purified Density Operator (LPDO) is a tensor network proposal for modeling Choi matrices [Werner et al., 2016, Torlai et al., 2020]. It consists of $2N$ cores (N cores and their conjugates) in the structure shown in Figure 2. The j-th tensor core of an LPDO contains the following modes: one input mode σ_j or σ'_j, one output mode τ_j or τ'_j, one purification mode ν_j, and one or two bond modes μ_{j-1}, μ_j or μ'_{j-1}, μ'_j. Denoting the collection of core tensor parameters as θ, the Choi matrix corresponding to such a tensor network is given written as:

$$[\boldsymbol{\Lambda}_{\theta}]^{\boldsymbol{\sigma}',\boldsymbol{\tau}'}_{\boldsymbol{\sigma},\boldsymbol{\tau}} = \sum_{\boldsymbol{\mu},\boldsymbol{\mu}',\boldsymbol{\nu}} \prod_{j=1}^{N} [\mathcal{K}_j]^{\nu_j,\mu_j}_{\sigma_j,\tau_j,\mu_{j-1},} [\overline{\mathcal{K}_j}]^{\sigma'_j,\tau'_j,\mu'_j}_{\nu_j,\mu'_{j-1}} \tag{1}$$

For ease of exposition, in the rest of this paper we fix the dimension of the tensor modes as $|\sigma_j| = |\sigma'_j| = 2$, $|\tau_j| = |\tau'_j| = 2$, $|\nu_j| = P$, and $|\mu_j| = D$ for all j. Each core has $\mathcal{O}(PD^2)$ parameters,

[2] A Stiefel manifold $\mathcal{V}(\mathbb{C}^{n\times k})$ is the set of matrices $\{\kappa \in \mathbb{C}^{n\times k} | \kappa^{\dagger}\kappa = \mathbb{I}\}$

so the LPDO tensor network decomposition approximates a $2^{2N} \times 2^{2N}$ Choi matrix with $\mathcal{O}(NPD^2)$ parameters. Additionally, an LPDO decomposition with sufficiently large bond and purification dimensions can represent any Choi matrix [De las Cuevas et al., 2013]. One advantage of the LPDO tensor network is that it guarantees that the corresponding Choi matrix is PSD (i.e., represents a CP quantum channel) *by design*. From an optimization perspective, this means that a learning algorithm for LPDOs does not need to explicitly worry about satisfying the CP constraint.

Leveraging these properties, Torlai et al. [2020] recently proposed a gradient-descent based maximum-likelihood algorithm to learn LPDO approximations of Choi matrices for quantum process tomography. While this procedure yields LPDOs that are always CP, there is no guarantee that the learned LPDO satisfies the TP condition. Torlai et al. [2020] propose adding a TP violation penalty term to their maximum likelihood objective to push the solution towards TP LPDOs, and note that the TP violation shrinks to zero in the limit of infinite data. However, in practice, with limited data the TP violation can still be non-negligible and the learned LPDO can be a non-physical *trace-increasing* channel. Using this method to learn some simple 2-qubit gates, we find that the TP violation (measured as the Frobenius norm error of the partial trace of the learned Choi from the identity) is on the order of 10^{-2}, and increasing the penalty for TP violation slows the convergence of the algorithm without significantly reducing TP violation error (see Figure 7 in Appendix B). There is also no known scheme to project the learned LPDO to the nearest TP LPDO without reconstructing the full Choi matrix, which would defeat the purpose of using the LPDO decomposition in the first place. Typical projection schemes alternate between projections onto the CP set (violating TP) and projections onto the TP set (violating CP), but the LPDO structure does not allow CP violations.

3 Trace Preserving Locally Purified Density Operators

In order to learn a physically valid Choi matrix through its LPDO decomposition, we need to place additional constraints on the LPDO structure. We first explore a naive constraint we can impose independently on each tensor core that is sufficient to guarantee that the LPDO satisfies a global TP condition. In practice however, this constraint is too strong and is unable to capture simple 2-qubit unitary gates, let alone more complicated channels. We then present a more involved discussion of the TP condition, and provide a more general characterization of TP LPDOs.

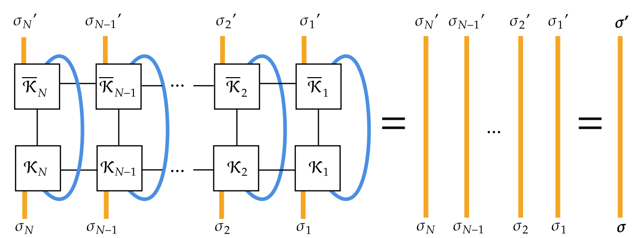


Figure 3: The TP constraint on an LPDO representing a quantum channel. The blue connections correspond to taking the partial trace with respect to the output indices, which results in the identity.

3.1 A Simple Sufficient Condition for TP LPDOs

One simple approach to ensuring TP LPDOs is to force every core $\mathcal{K}_j, \overline{\mathcal{K}}_j$ to function as a local quantum channel acting on the "outputs" from cores $\mathcal{K}_{j-1}, \overline{\mathcal{K}}_{j-1}$. To do this, we will use the Kraus operator formulation of quantum channels: every CP-TP quantum channel can be written using a set of Kraus operators $\mathcal{E}(\rho) = \sum_i \mathbf{K}_i \rho \mathbf{K}_i^\dagger$, where the set of Kraus operators must satisfy the condition $\sum_i \mathbf{K}_i^\dagger \mathbf{K}_i = \mathbb{I}$. As noted by various authors [Bruzda et al., 2009, Pechen and Rabitz, 2014, Srinivasan et al., 2018, Adhikary et al., 2020], vertically stacking a set of Kraus operators produces a matrix that lies on a Stiefel manifold of dimension $2PD \times 2D$, i.e., if $\kappa = [K_1, ..., K_W]^T$, then $\sum_i K_i^\dagger K_i = \kappa^\dagger \kappa = \mathbb{I}$ [Kraus, 1971, Gheondea, 2010]. Now, the idea is to interpret every *core* as a set of Kraus operators on a Stiefel manifold (allowing us to use Riemannian optimization techniques on each tensor core), so the application of a sequence of local quantum channels is also a global quantum channel and recovers the global TP condition. To achieve this, we make a slight modification to the LPDO structure by adding an extra mode to the cores at the N-site, and we call this extended LPDO an xLPDO (Figure 4). Note that the extended edge of the xLPDO can be fused with the core's purification dimension to recover the LPDO structure, so it is still in the same class of tensor networks. We state the sufficient condition formally below. We give a diagrammatic proof in Figure 4 and leave the algebraic proof to the appendix.

Lemma 1 (Sufficient Constraint for TP LPDO)**.** *Let κ_j be a flattening of the tensor core $\mathcal{K}_j$ into a matrix with rows associated with $\nu_j\tau_j\mu_j$ and columns with $\sigma_j\mu_{j-1}$. If each κ_j satisfies $\kappa_j^\dagger\kappa_j=\mathbb{I}$ for $1\leq j\leq N$, , i.e., each κ_j is on the Stiefel manifold, the LPDO satisfies the global TP constraint.*

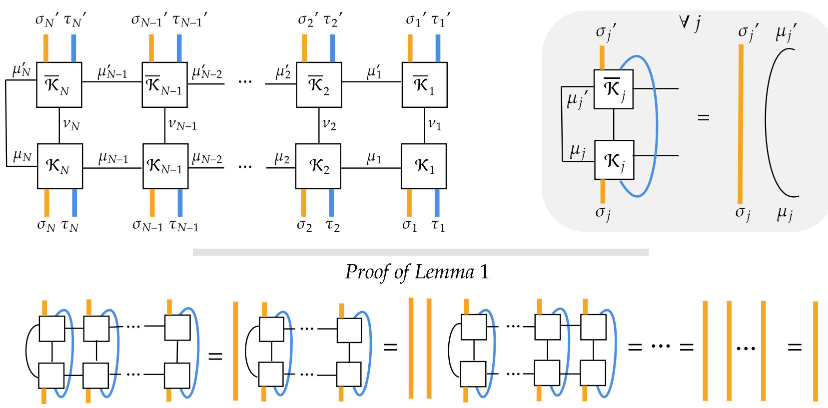


Figure 4: **Top:** Illustration of xLPDO (left) and sufficient local Stiefel manifold constraint on each core for TP LPDO (right), **Bottom:** Graphical proof of Lemma 1

While Lemma 1 gives a simple constraint that can easily parameterize TP LPDOs, it is too strict. Using a maximum likelihood approach similar to Torlai et al. [2020] but *constraining the LPDO cores to lie on a Stiefel manifold*[3], we find that the procedure cannot learn even elementary 2-qubit gates (we compare the true Choi matrix of a 2-qubit CNOT and the learned Choi matrix in Figure 8 in Appendix C). This assumption forces each core to *independently* contribute to the TP requirement, preventing a TP violation in one core to be 'compensated' by another core. Thus, such a restriction to independent constraints on each core does not produce a sufficiently general parameterization of Choi matrices.

3.2 A Necessary and Sufficient Constraint for LPDOs

Here, we propose a *more general* pair of conditions on the tensor cores so the LPDO satisfies a global TP constraint. As we will see, while these constraints are necessary and sufficient for the LPDO to be TP, translating these constraints into a parameterization remains an open problem.

We begin by considering the 'bottom half' of the LPDO, consisting of cores $\mathcal{K}_1,\ldots,\mathcal{K}_N$ without their conjugates. If we fuse indices ν_j,τ_j into a single index ω_j so each boundary (non-boundary) core is a third (fourth) order tensor, we can view the resulting tensor network $\mathcal{M}$ as a matrix product operator (MPO) [Murg et al., 2008] with cores $\mathcal{M}_j$. If we treat $\mathcal{M}$ as the tensor network decomposition of a matrix $[\mathbf{M}]_{\boldsymbol{\omega}}^{\boldsymbol{\sigma}}$ with rows associated with $\boldsymbol{\omega}$ and columns associated with $\boldsymbol{\sigma}$, then as in Remark 1, the TP constraint can be stated as the condition $\mathbf{M}^\dagger\mathbf{M}=\mathbb{I}$. Thus, we can also think of the problem of finding TP LPDOs

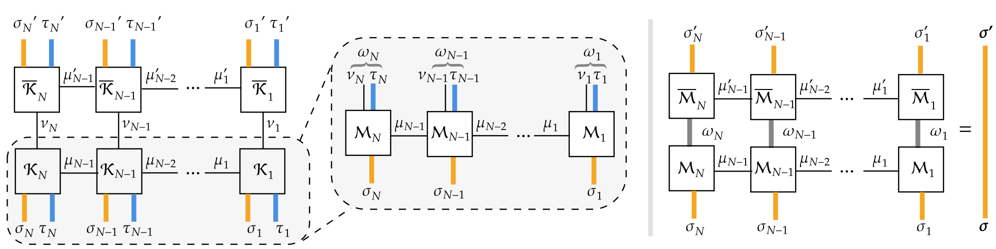
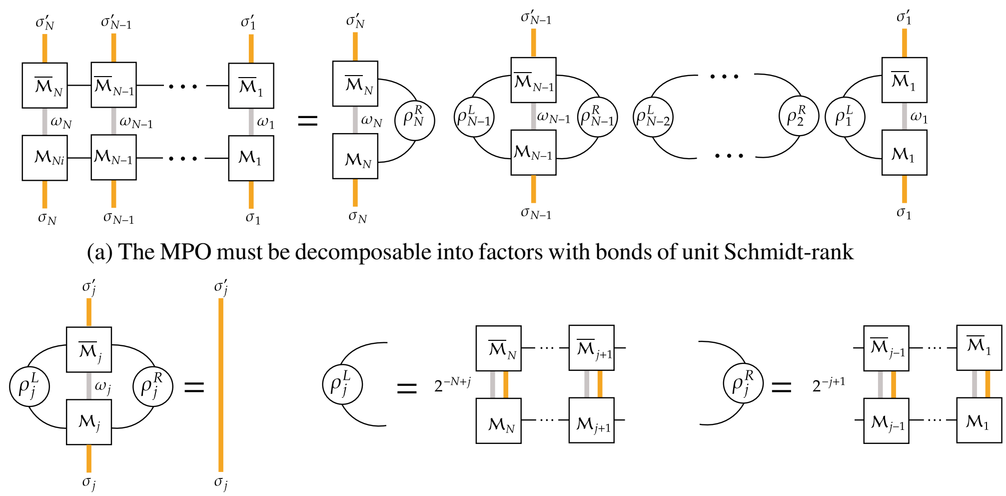


Figure 5: **Left:** Viewing the 'bottom-half' of an LPDO as an MPO by fusing the indices $\{\nu_i,\tau_i\}$ into a single index ω_i, **Right:** TP constraint specified on MPO

[3]We compute the gradient of the log-likelihood with respect to each tensor core and apply the algorithm from Wen and Yin [2013] to retract the gradient at each core onto the local Stiefel manifold.

as one of finding an MPO decomposition $\mathcal{M}$ of an isometry $\mathbf{M}$ living on the Stiefel manifold. Now, we state an alternative characterization of the TP constraint[4], followed by a geometric interpretation.

Theorem 1 (Necessary and Sufficient Condition for TP LPDO)**.** *Let $[\mathcal{M}_j]^{\sigma_j}_{\omega_j,\mu_j,\mu_{j-1}}$ be an MPO core associated with a reshaped LPDO core $[\mathcal{K}_j]^{\sigma_j}_{\tau_j,\nu_j,\mu_j,\mu_{j-1}}$ where τ_j,ν_j are fused into a single index ω_j. Then, the LPDO is TP if and only if the MPO cores $\mathcal{M}_j$ satisfy the conditions in Figure 6a and 6b.*



(a) The MPO must be decomposable into factors with bonds of unit Schmidt-rank

(b) Factors must be identity

(c) Definitions of ρ^L_j and ρ^R_j

Figure 6: (a-b) Pair of conditions which together are necessary and sufficient for MPO $\mathcal{M}$ of an LPDO to be TP. (c) Definition of the matrices ρ^L_j and ρ^R_j appearing in the above conditions, which are given by a weighted trace over all σ_j indices to the left or right of the i-th core.

Remark 2. *Theorem 1 states that if an LPDO $\mathbf{\Lambda}_\theta$ is TP, its corresponding MPO $\mathcal{M}$ must satisfy the condition in Figure 6b. This condition can be written as the requirement that $\mathbf{M}_j^\dagger \mathbf{G}_j \mathbf{M}_j = \mathbb{I}$ where $\mathbf{G}_j = (\boldsymbol{\rho}^L_j \otimes \mathbb{I} \otimes \boldsymbol{\rho}^R_j)$. Thus, the TP condition on LPDOs requires that each MPO core $\mathbf{M}_j$ be an orthonormal 2-frame under the metric $\mathbf{G}_j$ induced by the other tensor cores.*

Theorem 1 illustrates the difficulty in extracting a simple parameterization of TP LPDOs; modifying any core $\mathcal{M}_j$ affects ρ^L_{-j} and ρ^R_{-j} on *all* the other cores, so it is unclear how to modify a core to preserve TP. Thus, the problem of finding a TP parameterization of LPDOs remains open. Nevertheless, our characterization in Theorem 1 moves us closer towards a parameterization of TP LPDOs. If we are able to describe how the parameters at one core of a TP LPDO may be varied without violating the TP constraint, we speculate that a natural approach would be a 'sweeping' optimization scheme that holds other cores fixed while optimizing each core sequentially.

4 Conclusion and Future Work

We discussed the potential of using locally purified density operators (LPDOs) for tractably modeling quantum channels. Although LPDOs satisfy the CP requirement of quantum channels by design, we showed that imposing the TP constraint presents further challenges. We showed that imposing a simple condition that each LPDO tensor core independently be on the Stiefel manifold is sufficient to guarantee TP, but is too strict to capture even simple 2-qubit gates. We also gave a necessary and sufficient description of TP LPDOs with a geometrical interpretation of the constraints on each core. However, converting this characterization of TP constraints into a parameterization in an optimization scheme remains an open question. Resolving this could unlock the potential of using LPDOs to tractably model quantum channels for applications such as quantum process tomography or quantum error correction.

[4]A very similar pair of constraints were given in Cirac et al. [2017] and Şahinoğlu et al. [2018] for MPOs representing unitary matrices, but under the assumption that all cores $\mathcal{M}_j$ are identical and the MPO has periodic boundary conditions. In that setting, these constraints only parameterize a subset of "simple" unitary MPOs, and require slightly stronger assumptions (and different proof techniques) to prove necessity.

References

S. Adhikary, S. Srinivasan, G. Gordon, and B. Boots. Expressiveness and learning of hidden quantum markov models. In *International Conference on Artificial Intelligence and Statistics*, pages 4151–4161. PMLR, 2020.

W. Bruzda, V. Cappellini, H.-J. Sommers, and K. Życzkowski. Random quantum operations. *Physics Letters A*, 373(3):320–324, 2009.

M.-D. Choi. Completely positive linear maps on complex matrices. *Linear algebra and its applications*, 10(3):285–290, 1975.

J. I. Cirac, D. Perez-Garcia, N. Schuch, and F. Verstraete. Matrix product unitaries: structure, symmetries, and topological invariants. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(8):083105, 2017.

G. De las Cuevas, N. Schuch, D. Pérez-García, and J. I. Cirac. Purifications of multipartite states: limitations and constructive methods. *New Journal of Physics*, 15(12):123021, 2013.

A. Gheondea. The three equivalent forms of completely positive maps on matrices. *Annals of the University of Bucharest*, 2010.

A. Jamiołkowski. Linear transformations which preserve trace and positive semidefiniteness of operators. *Reports on Mathematical Physics*, 3(4):275–278, 1972.

M. Jiang, S. Luo, and S. Fu. Channel-state duality. *Physical Review A*, 87(2):022310, 2013.

G. C. Knee, E. Bolduc, J. Leach, and E. M. Gauger. Quantum process tomography via completely positive and trace-preserving projection. *Physical Review A*, 98(6):062336, 2018.

K. Kraus. General state changes in quantum theory. *Annals of Physics*, 64(2):311–335, 1971.

I. A. Luchnikov, A. Ryzhov, S. N. Filippov, and H. Ouerdane. QGOpt: Riemannian optimization for quantum technologies. *SciPost Phys.*, 10:79, 2021. doi: 10.21468/SciPostPhys.10.3.079. URL `https://scipost.org/10.21468/SciPostPhys.10.3.079`.

V. Murg, J. Cirac, B. Pirvu, and F. Verstraete. Matrix product operator representations. *New Journal of Physics*, 12:025012, 2008.

A. N. Pechen and H. A. Rabitz. Incoherent control of open quantum systems. *Journal of Mathematical Sciences*, 199:695–701, 2014.

D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac. Matrix product state representations. *arXiv preprint quant-ph/0608197*, 2006.

M. B. Şahinoğlu, S. K. Shukla, F. Bi, and X. Chen. Matrix product representation of locality preserving unitaries. *Physical Review B*, 98(24):245122, 2018.

S. Srinivasan, G. Gordon, and B. Boots. Learning hidden quantum markov models. In *International Conference on Artificial Intelligence and Statistics*, pages 1979–1987. PMLR, 2018.

T. Surawy-Stepney, J. Kahn, R. Kueng, and M. Guta. Projected least-squares quantum process tomography. *arXiv preprint arXiv:2107.01060*, 2021.

G. Torlai, C. J. Wood, A. Acharya, G. Carleo, J. Carrasquilla, and L. Aolita. Quantum process tomography with unsupervised learning and tensor networks. *arXiv preprint arXiv:2006.02424*, 2020.

Z. Wen and W. Yin. A feasible method for optimization with orthogonality constraints. *Mathematical Programming*, 142(1):397–434, 2013.

A. H. Werner, D. Jaschke, P. Silvi, M. Kliesch, T. Calarco, J. Eisert, and S. Montangero. Positive tensor network approach for simulating open quantum many-body systems. *Physical review letters*, 116 (23):237201, 2016.

A Proofs

Proof of Lemma 1. First, we observe that Equation 1 can be written as follows:

$$[\mathbf{\Lambda}_\theta]^{\boldsymbol{\tau}',\boldsymbol{\tau}}_{\boldsymbol{\sigma}',\boldsymbol{\sigma}} = \sum_{\boldsymbol{\nu}_j} (\mathbf{K}^{\sigma'_1}_{\tau'_1,\nu_1})^\dagger \cdots (\mathbf{K}^{\sigma'_N}_{\tau'_N,\nu_N})^\dagger \mathbf{K}^{\sigma_N}_{\tau_N,\nu_N} \cdots \mathbf{K}^{\sigma_1}_{\tau_1,\nu_1} \tag{2}$$

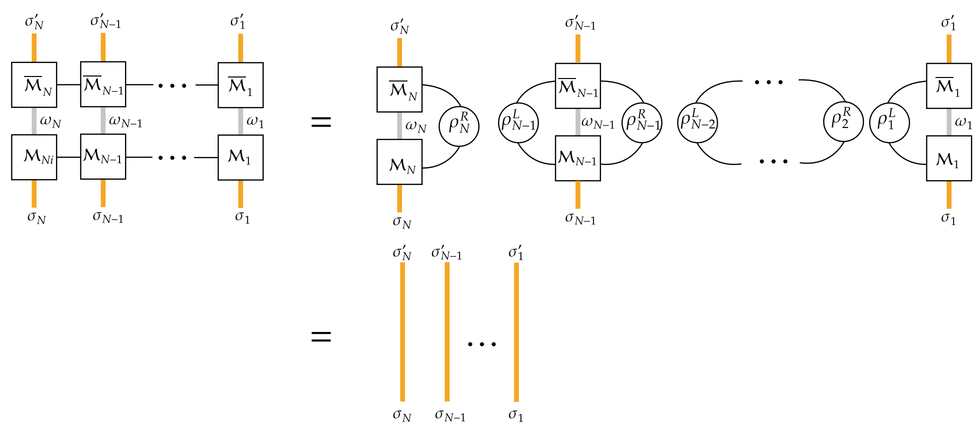
The TP requirement on Choi matrices is that $\mathrm{tr}_B(\mathbf{\Lambda}_\theta) = \mathbb{I}$, or equivalently, the trace of diagonal blocks of the Choi matrix is 1 and trace of off-diagonal blocks is 0: $\sum_{\boldsymbol{\tau}=\boldsymbol{\tau}'} [\mathbf{\Lambda}_\theta]^{\boldsymbol{\tau},\boldsymbol{\tau}'}_{\boldsymbol{\sigma}'=\boldsymbol{\sigma}} = 1$ and $\sum_{\boldsymbol{\tau}=\boldsymbol{\tau}'} [\mathbf{\Lambda}_\theta]^{\boldsymbol{\tau},\boldsymbol{\tau}'}_{\boldsymbol{\sigma}'\neq\boldsymbol{\sigma}} = 0$. Observe that:

$$\begin{aligned}\sum_{\boldsymbol{\tau}=\boldsymbol{\tau}'} [\mathbf{\Lambda}_\theta]^{\boldsymbol{\tau}',\boldsymbol{\tau}}_{\boldsymbol{\sigma}'=\boldsymbol{\sigma}} &= \sum_{\tau_1,\dots,\tau_{N-1}} \sum_{\nu_1,\dots,\nu_{N-1}} (\mathbf{K}^{\sigma_1}_{\tau_1,\nu_1})^\dagger \cdots \left(\sum_{\tau_N,\nu_N} (\mathbf{K}^{\sigma_N}_{\tau_N,\nu_N})^\dagger \mathbf{K}^{\sigma_N}_{\tau_N,\nu_N} \right) \cdots \mathbf{K}^{\sigma_1}_{\tau_1,\nu_1} \\ &= \sum_{\tau_1,\dots,\tau_{N-1}} \sum_{\nu_1,\dots,\nu_{N-1}} (\mathbf{K}^{\sigma_1}_{\tau_1,\nu_1})^\dagger \cdots \left(\kappa^{\sigma_N \dagger} \kappa^{\sigma_N} \right) \cdots \mathbf{K}^{\sigma_1}_{\tau_1,\nu_1}\end{aligned} \tag{3}$$

But $\kappa^{\sigma_N^\dagger} \kappa^{\sigma_N} = \mathbb{I}$, so we can identically repeatedly simplify this sum until we have $\sum_{\tau_1} \sum_{\nu_1} (\mathbf{K}^{\sigma_1}_{\tau_1,\nu_1})^\dagger \mathbf{K}^{\sigma_1}_{\tau_1,\nu_1} = 1$, showing that diagonal blocks trace to 1. Taking the trace of the off-diagonal blocks proceeds similarly; however, since $\sigma_j \neq \sigma'_j$ for some j, we will eventually have the term $\kappa^{\sigma'_j \dagger} \kappa^{\sigma_j} = 0$, which will send the trace of the off-diagonal block to zero. □

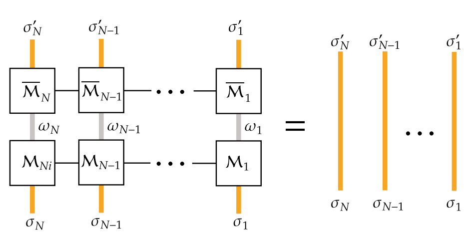
Proof of Theorem 1.

(⇐) Assume the conditions in Figures 6a and 6b hold. Then,

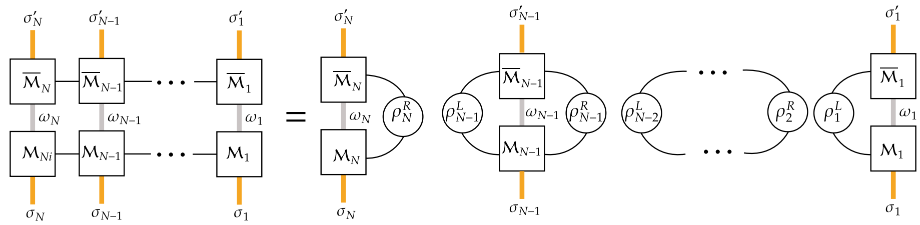


(⇒)

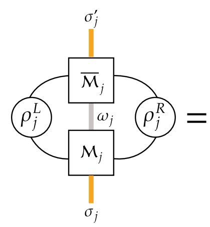
We start by assuming that that the LPDO is trace-preserving, and thus satisfies the following condition.



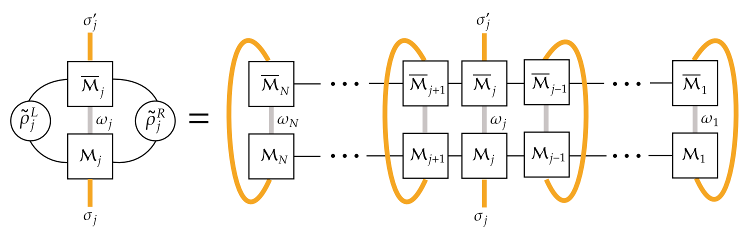
Now, from Theorem 2 in Perez-Garcia et al. [2006], we know that for every i-th core, there exist matrices ρ^L_j and ρ^R_j such that our LPDO can be written as



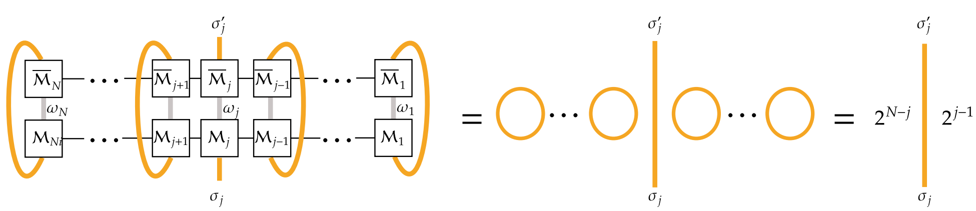
with



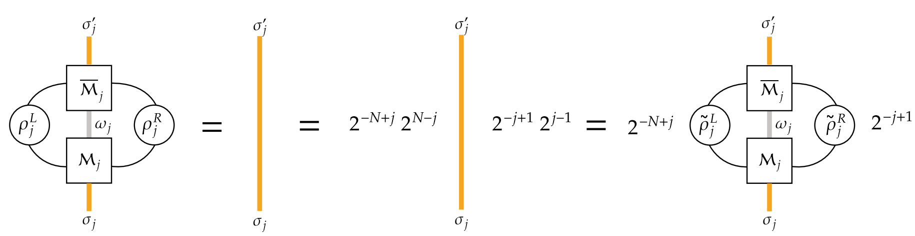
Now, to identify ρ_j^L and ρ_j^R, let us define $\tilde{\rho}_j^L$ and $\tilde{\rho_j}^R$ without the scalar factors 2^{-N+j} and 2^{-j+1} in Figure 6c, and consider the contracted LPDO representation over σ'_{-j}:



But by the TP assumption,



Thus, $\rho_j^L = 2^{-N+j}\ \tilde{\rho}_j^L$ and $\rho_j^R = 2^{-j+1}\ \tilde{\rho}_j^R$. So, we have



□

B TP Violation of Unconstrained LPDO

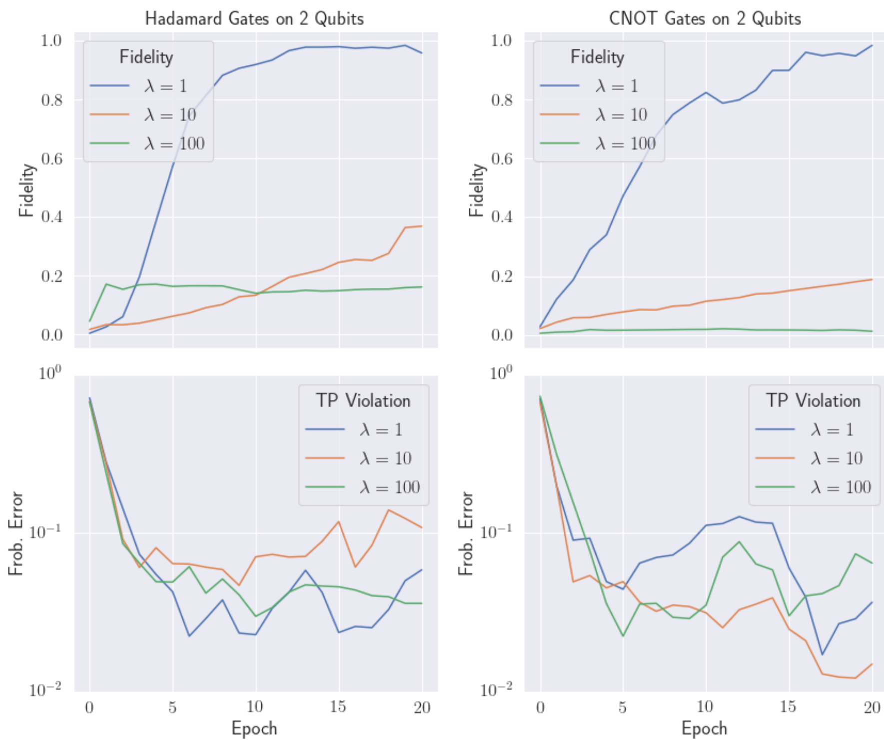


Figure 7: Learning simple quantum channels using the method of Torlai et al. [2020] with various penalties λ on TP violation. **Top**: Fidelity of the learned Choi matrix and the true Choi matrix, **Bottom**: TP Violation as Frobenius norm distance of partial trace of learned Choi matrix from identity. Increasing the penalty term does not affect TP violation error, but does slow convergence.

C Lemma 1 is Too Strict

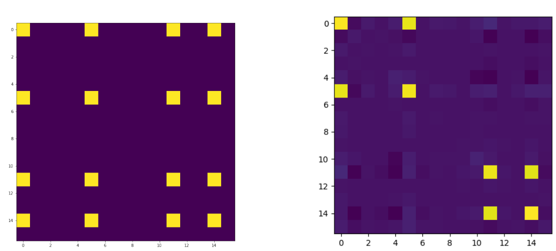


Figure 8: **Left:** Choi matrix of 2-qubit CNOT gate. The yellow and dark-blue elements correspond to non-zero and zero entries of the matrix. **Right:** Choi matrix of TP LPDO learned from tomography data whose cores are constrained according to Lemma 1. The learned matrix fails to capture the non-zero elements in the off-diagonal blocks.