

# Quantum Channel Learning

Mikhail Gennadievich Belov\*

*Lomonosov Moscow State University, Faculty of Mechanics and Mathematics,*

*GSP-1, Moscow, Vorob'evy Gory, 119991, Russia*

Victor Victorovich Dubov<sup>†</sup> and Alexey Vladimirovich Filimonov<sup>‡</sup>

*Peter the Great St. Petersburg Polytechnic University, 195251, Russia*

Vladislav Gennadievich Malyshkin<sup>§</sup>

*Ioffe Institute, Politekhnicheskaya 26, St Petersburg, 194021, Russia*

(Dated: June, 6, 2024)

\$Id: QuantumChannelLearning.tex,v 1.404 2025/01/03 17:31:18 mal Exp \$

The problem of an optimal mapping between Hilbert spaces  $IN$  and  $OUT$ , based on a series of density matrix mapping measurements  $\rho^{(l)} \rightarrow \varrho^{(l)}$ ,  $l = 1 \dots M$ , is formulated as an optimization problem maximizing the total fidelity  $\mathcal{F} = \sum_{l=1}^M \omega^{(l)} F(\varrho^{(l)}, \sum_s B_s \rho^{(l)} B_s^\dagger)$  subject to probability preservation constraints on Kraus operators  $B_s$ . For  $F(\varrho, \sigma)$  in the form that total fidelity can be represented as a quadratic form with superoperator  $\mathcal{F} = \sum_s \langle B_s | S | B_s \rangle$  (either exactly or as an approximation) an iterative algorithm is developed. The work introduces two important generalizations of unitary learning: 1.  $IN/OUT$  states are represented as density matrices. 2. The mapping itself is formulated as a mixed unitary quantum channel  $A^{OUT} = \sum_s |w_s|^2 \mathcal{U}_s A^{IN} \mathcal{U}_s^\dagger$  (no general quantum channel yet). This marks a crucial advancement from the commonly studied unitary mapping of pure states  $\phi_l = \mathcal{U} \psi_l$  to a quantum channel, what allows us to distinguish probabilistic mixture of states and their superposition. An application of the approach is demonstrated on unitary learning of density matrix mapping  $\varrho^{(l)} = \mathcal{U} \rho^{(l)} \mathcal{U}^\dagger$ , in this case a quadratic on  $\mathcal{U}$  fidelity can be constructed by considering  $\sqrt{\rho^{(l)}} \rightarrow \sqrt{\varrho^{(l)}}$  mapping, and on a quantum channel, where quadratic on  $B_s$  fidelity is an approximation — a quantum channel is then obtained as a hierarchy of unitary mappings, a mixed unitary channel. The approach can be applied to studying quantum inverse problems, variational quantum algorithms, quantum tomography, and more. A software product implementing the algorithm is available from the authors.

DOI: 10.1103/PhysRevE.111.015302

---

\* [mikhail.belov@tafs.pro](mailto:mikhail.belov@tafs.pro)

† [dubov@spbstu.ru](mailto:dubov@spbstu.ru)

‡ [filimonov@rphf.spbstu.ru](mailto:filimonov@rphf.spbstu.ru)

§ [malyshki@ton.ioffe.ru](mailto:malyshki@ton.ioffe.ru)

Dedicated to the memory of Ivan Anatol'evich Komarchev

## I. INTRODUCTION

The form of knowledge representation stands out as a crucial and distinctive feature in any approach to Machine Learning (ML). The utilization of unitary operators for knowledge representation has garnered increasing attention recently[1–3]. In nature most of dynamic equations are equivalent to a sequence of infinitesimal unitary transformations: Newton, Maxwell, Schrödinger equations. This inherent connection with unitary transformations makes it particularly appealing to represent knowledge in this form. Most of the existing works consider algorithms that take wavefunction as input and construct a unitary operator providing a high value of mapping fidelity[2–4]; they are mostly different in parametrization of unitary operator [5] and optimization details. Such a pure state to pure state mapping is a limited form of quantum evolution. There are a few works[6–9] that consider unitary learning with density matrix input. Real systems probability should be described by a mixed state (density matrix) what allows to distinguish probabilistic mixture of states and their superposition. There are two sources of mixed states: 1. The input data itself can be in mixed state. 2.  $IN \rightarrow OUT$  mapping of a general quantum channel can transform a pure state input into a mixed state output. Unitary mapping is a simple example of a quantum channel converting a pure state into a pure state.

In our previous works[10, 11] the quantum mechanics inverse problem of optimal unitary mapping of pure states was converted to a **QCQP** (Quadratically Constrained Quadratic Program) problem, a novel algebraic problem (43) was formulated, and an efficient iterative global optimization algorithm was developed. This algorithm can be applied to quantum inverse problems, variational quantum algorithms [12–14] with a cost function in the form  $C(\theta) = \text{Tr } OU(\theta)\rho_0U^\dagger(\theta)$ , quantum tomography[15–18], various classical problems, and many others. We do not have a formal proof of the developed algorithm's convergence, but among the millions of test runs, only a few did not converge to the global maximum. This could possibly be caused by numerical instability. In the current work, we generalize our algorithm to mixed states and further extend it to quantum channels. This generalization is possible as long as the total fidelity can be represented as a quadratic form on quantum channel mapping operators.

In the case of unitary mapping ( $D = n$ ,  $N_s = 1$ ) of mixed states there are several good choices for fidelity, for example a quantum channel mapping the square root of density matrix  $\sqrt{\rho} \rightarrow \sqrt{\bar{\rho}}$  with a unitary operator (3), it provides the exact fidelity as a quadratic form. Together with the available iterative algorithm, this problem may be considered mostly solved.

In the case of a general quantum channel (4) constructing fidelity as the quadratic form of Eq. (16) on quantum channel mapping operators can be challenging. We have developed several approximations that can be used with the iterative algorithm and have constructed a hierarchy of mapping operators (52) as a mixed unitary channel. A demonstration of several approximations is presented in Section IV A. The problem of finding a good quadratic (on mapping operators) fidelity for Kraus rank  $N_s > 1$  quantum channels requires more research and is related to a physical meaning of quantum channel mapping. Even when fidelity is expressed as the quadratic form of Eq. (16), we currently only have a numerical algorithm to approximate the data using a mixed unitary channel. A general reconstruction of quantum channels is not yet available.

The paper is organized as follows. After formulating the problem, the unitary mapping of mixed states is considered in Section III. Section IV focuses on the construction of a hierarchy of unitary operators. In the conclusion, we discuss the results obtained for the problem of quantum channel reconstruction. Appendix A provides a detailed description of the numerical solution that uses algebraic techniques. Appendix B presents a generalization of the novel algebraic problem (43) to the nonstationary case and introduces a time-dependent Schrödinger-like equation for unitary operator dynamics (B1). Appendix C explores potential generalizations to states with memory, while Appendix D estimates the algorithm's computational complexity.

This paper is accompanied by a software which is available from Ref. [19]; all references to code in the paper correspond to this software.

## II. FORMULATION OF THE PROBLEM

In [11], we considered data of vector-to-vector pure state mapping,  $l = 1 \dots M$ ,

$$\psi_l(\mathbf{x}) \rightarrow \phi_l(\mathbf{f}) \quad \text{weight } \omega^{(l)} \quad (1)$$

The weights  $\omega^{(l)}$  are typically all equal to 1.<sup>1</sup> However, they can be set to different values if the observations are made with different accuracy, this is particularly convenient for classical systems. The usefulness of  $\omega^{(l)}$  weights also arises in models that comprise a number of identical mappings; in this case one may consider only distinct mappings and set  $\omega^{(l)}$  equal to the number of times the observation was encountered in the sample. The problem of maximizing total fidelity

$$\mathcal{F} = \sum_{l=1}^M \omega^{(l)} \left| \langle \phi_l | \mathcal{U} | \psi_l \rangle \right|^2 \xrightarrow{\mathcal{U}} \max \quad (2)$$

was considered, and a numerical algorithm finding the global maximum of  $\mathcal{F}$  was developed. Found solution  $\mathcal{U}$ , a partially unitary matrix of  $\dim(OUT) \times \dim(IN)$  ( $D \times n$ )

$$A^{OUT} = \mathcal{U} A^{IN} \mathcal{U}^\dagger \quad (3)$$

converts any operator between two Hilbert spaces, for example it converts a pure state  $A^{IN} = |\psi\rangle \langle\psi|$  into a pure state  $A^{OUT} = |\mathcal{U}\psi\rangle \langle\psi|\mathcal{U}^\dagger|$ . For  $D = n$  it is a trace preserving map, and for  $D < n$  it is a **trace decreasing map** quantum channel. The input data (1) represent pure state to pure state mapping. This type of mapping is not the most general, it cannot describe systems with effects like **quantum decoherence**. For example a simple **Markov chain** system converts a pure state  $|\psi\rangle \langle\psi|$  into a density matrix state, see Appendix I of Ref. [20].

In this work a more general mapping between Hilbert spaces  $IN$  of  $|\psi\rangle$  (dimension  $n$ ) and  $OUT$  of  $|\phi\rangle$  (dimension  $D$ ) is considered. **Kraus' theorem** determines the most general form of mapping between Hilbert spaces[21–23]:

$$A^{OUT} = \sum_{s=0}^{N_s-1} B_s A^{IN} B_s^\dagger \quad (4)$$

the number of terms in the sum  $N_s$  is called the Kraus rank, see Choi's theorem [24] and Belavkin's Radon-Nikodym theorem for completely positive maps [25]. The Kraus operators  $B_s$  satisfy the constraints that unit  $A^{IN}$  is converted to unit  $A^{OUT}$ . There are other options to construct a quantum channel, in [10, 11] we considered a quantum channel that transforms the Gram matrix in space  $|\psi\rangle$  into the Gram matrix in space  $|\phi\rangle$ ; with proper regularization

---

<sup>1</sup> Some authors use  $\omega^{(l)} = 1/M$  to normalize the fidelity to the range  $[0 : 1]$ . However, this approach is inconvenient because  $\mathcal{F}$  is no longer an **extensive quantity** in the sense that, for two sets of observations, the total fidelity is no longer the sum of the two — losing its additive property. For this reason, normalizing to the number of observations is preferred over normalizing to  $[0 : 1]$ .

it can be reduced to the same form.

$$\sum_{s=0}^{N_s-1} B_s B_s^\dagger = \mathbb{1} \quad (5)$$

We further generalize Kraus operators  $B_s$  by considering them to be rectangular  $D \times n$  matrices instead of being limited to Hilbert space mapping into itself. The term “partially Kraus” is used in the same sense as “partially unitary” — we consider rectangular matrices  $D \leq n$  mapping Hilbert spaces of different dimensions.<sup>2</sup> For  $D < n$  a partially unitary quantum channel (3) does not preserve the density matrix trace. For a large enough  $N_s$ , as in the condition (56), the form of Eq. (4) allows us to construct a trace preserving quantum channel, for example to implement an operation of **partial trace**.

Kraus operators selection is not unique. In addition to the familiar  $\exp(i\varphi)$  phase factor ( $\pm 1$  for real space), there is a **gauge** that regulates redundant degrees of freedom. The most well-known is the canonical form:

$$\text{Tr} B_s B_t^\dagger = 0 \quad \text{for } s \neq t \quad (6)$$

However, other gauges can be used. See Appendix A 1 for an algorithm transforming Kraus operators to canonical form.

Back in [10, 20] we considered, instead of (1), the data of density matrix to density matrix mapping (7) — just replace  $S_{jk;j'k'}$  from (19) by (18) or (30). Whereas input/output data in density matrix form distinguishes probabilistic mixture of states and their superposition, the transform (3) cannot map such data exactly, it has insufficient expressive power. Only a general quantum channel (4) mapping has sufficient expressive power to construct learning models from general data, for example with a decoherence effect, or a much more seldom but very interesting effects of pure state formation from mixed state input — spontaneous coherence, exhibited in entropy decrease, synchronizing, coherent responses of  $(\psi_A + \psi_B)^2 \gg \psi_A^2 + \psi_B^2$  type, etc.

---

<sup>2</sup> Here, instead of the usual constraint  $\sum_{s=0}^{N_s-1} B_s^\dagger B_s = \mathbb{1}$  (55) we use the constraint (5) to fully utilize our numerical algorithm in the case of  $N_s = 1$ . Constraint (5) imposes  $D(D+1)/2$  independent conditions, whereas constraint (55) imposes  $n(n+1)/2$  independent conditions. When  $D = n$  both definitions coincide. However, for  $N_s = 1$  and  $D < n$ , the constraint (55) fails to satisfy the minimal Kraus rank condition (56), but the constraints (5) are applicable in the numerical study of trace-decreasing maps when  $N_s = 1$  and  $D < n$ . Additionally, the Lagrange multiplier calculation method described in Appendix A 4, along with the convergence-helper constraints in Appendix A 5, facilitates the incorporation of quadratic form constraints of any type.

Consider the density matrix mapping of  $l = 1 \dots M$  observations

$$\rho^{(l)} \rightarrow \varrho^{(l)} \quad \text{weight } \omega^{(l)} \quad (7)$$

The  $\rho^{(l)}$  is a measured Hermitian density matrix  $n \times n$  in the  $|\psi\rangle$  space, while the  $\varrho^{(l)}$  is a measured Hermitian density matrix  $D \times D$  in the  $|\phi\rangle$  space. Both  $\rho$  and  $\varrho$  density operators are Hermitian matrices with positive eigenvalues and having unit **trace** (the sum of diagonal elements). Previously considered data (1) corresponded to  $\rho^{(l)} = |\psi_l\rangle \langle \psi_l|$  and  $\varrho^{(l)} = |\phi_l\rangle \langle \phi_l|$ . The optimization problem becomes

$$\mathcal{F} = \sum_{l=1}^M \omega^{(l)} F(\varrho^{(l)}, \sigma^{(l)}) \xrightarrow{B} \max \quad (8)$$

where  $\sigma = A^{OUT}$  and  $\rho = A^{IN}$  — the  $\sigma$  is the  $\rho$  passed through the quantum channel.

It should be noted that while for wavefunction mapping (1), in the seldom case of a classic system with known wavefunctions phases, a regression type technique is sometimes applicable, for density matrix mapping (7), which is always quadratic on  $\mathcal{U}$  (3), no regression can possibly be applied.

We need to define the total fidelity  $\mathcal{F}$  of the (4) transform, which is a sum of contributions from all  $M$  observations. For each observation of the  $IN$  to  $OUT$  mapping, an accuracy factor  $F$  ranging  $[0 : 1]$  should be defined. For the theory presented in this paper to be applicable, this factor must be quadratic with respect to the operator  $\mathcal{U}$  (an instance of  $B_s$ ). For pure state to pure state mapping (1) it is a squared projection, the **fidelity**

$$F = \left| \langle \phi | \mathcal{U} | \psi \rangle \right|^2 \quad (9)$$

For mixed state mapping (7) a similar expression

$$F^{\varrho\sigma}(\varrho, \sigma) = \text{Tr } \varrho\sigma \quad (10)$$

is bounded between 0 and  $\text{Tr } \varrho^2 \leq 1$ , it has the meaning of a measured probability of probability that reaches the maximal value 1 only in pure states. A standard definition of fidelity between mixed states  $\varrho, \sigma$ , [26] p. 409, [27] p. 285,

$$F^{prop}(\varrho, \sigma) = \text{Tr} \sqrt{\varrho^{1/2} \sigma \varrho^{1/2}} = \text{Tr} \sqrt{\varrho} \sqrt{\sigma} \quad (11)$$

creates calculation difficulties since it requires taking matrix square root of an expression with  $B_s$  sum, and it is not explicitly quadratic. There is an option to use a quality criterion

based on the conditional entropy of  $OUT|IN$  states [28]. In the general case, this requires a joint density matrix  $\rho(OUT, IN)$  [29]. The tensor  $S_{jk;j'k'}$  (18) can be viewed as a joint density matrix  $\rho(OUT, IN)$  averaged over  $M$  observations in the sample. However, in the case where  $N_s > 1$ , we were unable to express the conditional entropy in quadratic form with respect to  $B_s$ , and we will leave this aspect of conditional entropy as a quality criterion for future research. In the  $N_s = 1$  case, the Kullback-Leibler divergence (Eq. 32) can be directly applied.

There is a noticeable feature of Hermitian operators: if one vectorizes  $D \times D$  Hermitian operators  $\varrho_{ij}$  and  $\sigma_{ij}$ , then one can verify that

$$\text{Tr} \varrho \sigma = \sum_{i,j=0}^{D-1} \varrho_{ij} \sigma_{ij}^* \quad (12)$$

i.e. the  $L^2$  of a vector obtained from all  $D^2$  matrix elements of  $\varrho$  gives the trace of  $\varrho^2$ . This provides a closeness criterion,  $F^c$ , which acts as a “correlation” between the density matrices  $\varrho_{ij}$  and  $\sigma_{ij}$ , treated as if they were vectors.

$$F^c(\varrho, \sigma) = \frac{\sum_{i,j=0}^{D-1} \varrho_{ij} \sigma_{ij}^*}{\sqrt{\sum_{i,j=0}^{D-1} |\varrho_{ij}|^2} \sqrt{\sum_{i,j=0}^{D-1} |\sigma_{ij}|^2}} = \frac{\text{Tr} \varrho \sigma}{\sqrt{\text{Tr} \varrho^2} \sqrt{\text{Tr} \sigma^2}} \quad (13)$$

It can be viewed as  $F^{\varrho\sigma}$  (10) with an adjusted contribution for mixed states. The practical issue with it is that  $\sigma$ , in the denominator, depends on the quantum channel operators  $B_s$ ; to obtain a workable expression, one can replace the  $\sigma$ -normalized denominator by the input density matrix  $\rho$

$$F^v = \frac{\sum_{i,j=0}^{D-1} \varrho_{ij} \sigma_{ij}^*(\rho)}{\sqrt{\sum_{i,j=0}^{D-1} |\varrho_{ij}|^2} \sqrt{\sum_{k,q=0}^{n-1} |\rho_{kq}|^2}} \quad (14)$$

Here, the entity denoted as  $\sigma(\rho)$  is obtained by passing  $\rho$  through the quantum channel using either the transformation in (4) or (3). Since  $\sigma$  does not appear in the denominator, Eq. (14) is quadratic in the quantum channel mapping operators. This expression uses (12) and treats density matrices as if they were vectors. It is linear in  $\sigma$  and quadratic in  $\mathcal{U}$ ; for unitary mapping ( $D = n$ ,  $N_s = 1$ ), we have  $\text{Tr} \rho^2 = \text{Tr} \sigma^2$ , and thus  $F^v$  matches  $F^c$ . An alternative is

to consider a surrogate closeness by normalizing it only on  $\varrho$  — to the degree of output state purity

$$F^{N\varrho^2}(\varrho, \sigma) = \frac{\text{Tr}\varrho\sigma}{\text{Tr}\varrho^2} \quad (15)$$

but this one performs poorly, especially in the case when  $N_s > 1$ . Note that if  $\varrho$  is a pure state  $|\phi\rangle\langle\phi|$  and  $\rho$  is a mixed state then (10), (11), and (15) are the same. Whereas the closeness criteria  $F^v$  (14) and  $F^{N\varrho^2}$  (15) are quadratic in quantum channel operators, their main drawback is the lack of clear physical meaning. This can lead to artifacts in the obtained solution, particularly for noisy data in the case of a density matrix passing through a general quantum channel. The criterion  $F^{\varrho\sigma}$  (10) is quadratic in quantum channel operators but represents the square of a probability, which overestimates the contribution of pure states[20]. As a result, it is poorly suited for density matrix mappings. The criteria  $F^{prop}$  (11) and  $F^c$  (13) have a clear physical meaning, but they are not quadratic in quantum channel operators. Of particular interest is the criterion  $F^{\sqrt{\varrho}\sqrt{\rho}}$  (31), considered below, which is quadratic in quantum channel operators and matches  $F^{prop}$  exactly in the case of unitary mapping of density matrices.

For the theory presented in this paper to be applicable, the only required feature of the mapping accuracy  $F$  in (8) is that it must be a *quadratic* form on  $B_s$  with a superoperator  $S_{jk;j'k'}$ .

$$\mathcal{F} = \sum_{s=0}^{N_s-1} \langle B_s | S | B_s \rangle \quad (16)$$

The tensor  $S_{jk;j'k'}$  depends on the choice of  $F$  and observation weights  $\omega^{(l)}$ . They are problem specific and do not change our considerations. Expanding the sums in (16) we obtain (for a simple choice of  $F$  (10))

$$\mathcal{F} = \sum_{l=1}^M \omega^{(l)} \sum_{s=0}^{N_s-1} \sum_{i,j=0}^{D-1} \sum_{k,k'=0}^{n-1} \varrho_{ij}^{(l)} b_{s,ik} \rho_{kk'}^{(l)} b_{s,jk'}^* \quad (17)$$

where the  $b_{s,ik}$  are  $s = 0 \dots N_s - 1$  matrices of dimension  $D \times n$  corresponding to partially Kraus operators  $B_s$ . One can introduce a tensor

$$S_{jk;j'k'} = \sum_{l=1}^M \omega^{(l)} \varrho_{jj'}^{(l)} \rho_{kk'}^{(l)} \quad (18)$$



to obtain an expression similar to the one in [11] for  $\mathbf{x} \rightarrow \mathbf{f}$  Hilbert spaces vector mapping. With the density matrix  $|\mathbf{x}\rangle\langle\mathbf{x}| \rightarrow |\mathbf{f}\rangle\langle\mathbf{f}|$  (18) takes a familiar form

$$S_{jk;j'k'} = \sum_{l=1}^M \omega^{(l)} f_j^{(l)} x_k^{(l)} f_{j'}^{(l)*} x_{k'}^{(l)*} \quad (19)$$

Distinguishing features of the current work include:

- Using density matrix states mapping (7) as input.
- Go beyond unitary mapping (3) to consider quantum channel mapping (4).

We use them to take into account two kinds of probabilities: probabilistic mixture of states and states superposition. For the latter, probability is the square of the sum of amplitudes, while for the former, probability is the sum of squared amplitudes. Both effects are present in real life data. There are a number of studies suggesting a limitation of unitary scalarization [30, 31], the other reevaluate recent research [32]. It should be noted how researchers commonly use unitarity — they take a vector, normalize its  $L^2$  norm, and consider the squared Euclidean projections obtained as if they were probabilities. Existing approaches do not distinguish probabilistic mixture of states from their superposition. This work is trying to overcome this deficiency by using input data in the form of density matrices (7) and mapping them with a general quantum channel (4). Consider a few demonstrations.

The partially unitary mapping (3) (when  $N_s = 1$ ) is a **trace decreasing map** quantum channel for  $D < n$ . A simple  $N_s > 1$  example: Let  $D = 1$  and we want to construct a quantum channel with operators  $B_s$  calculating the trace of  $A^{IN}$  with (4). The matrices  $b_{s,jk}$  are of  $1 \times n$  dimension for all  $s$ . Any orthogonal basis  $|x_s\rangle$  in  $|\psi\rangle$  solves the problem:

$$B_s = |f_0\rangle\langle x_s| \quad (20)$$

with  $N_s = n$ ,  $s = 0 \dots n - 1$ . Any orthogonal basis  $|x_s\rangle$  creates a solution of the form (20) that satisfies the canonical form constraints (25), this degeneracy may cause difficulties in numerical methods. Another example. A quantum channel parametrized by two orthogonal bases  $|x_k\rangle$ ,  $|f_j\rangle$ , and a matrix  $\mathcal{M}_{jk}$  creating  $N_s = Dn$  rank one operators

$$B_s = |f_j\rangle\mathcal{M}_{jk}\langle x_k| \quad (21)$$

The index  $s = 0 \dots Dn - 1$  enumerates all  $(j, k)$  pairs. The actual Kraus rank is less than or equal to  $Dn$ . The trace-preservation condition (55) imposes  $k = 0 \dots n - 1$  constraints:

$$1 = \sum_{j=0}^{D-1} |\mathcal{M}_{jk}|^2 \quad (22)$$

The matrix  $\mathcal{M}_{jk}$  defines a quantum channel. Equations (20) and (21) are two examples of trace-preserving quantum channels, constructed as a sum of rank one operators  $|f_j\rangle\langle x_k|$ . Such channels are defined using two bases,  $|f_j\rangle$ ,  $|x_k\rangle$ , along with the mapping matrix  $\mathcal{M}_{jk}$  between them. These channels are much easier to analyze (and work numerically) compared to a sum of general operators  $B_s$ . At the same time, they possess sufficiently high expressive power for trace-type mappings. However, a unitary mapping cannot be represented by (21).

For a general form of quantum channel see [17, 33] for a representation of a quantum channel with Choi matrix of the channel[24]. Let us defer the study of this Choi-style  $J(\Phi) = \sum_{kk'} |k\rangle\langle k'| \otimes \Phi(|k\rangle\langle k'|)$  representation of a quantum channel  $\Phi$  to future research and focus on the original problem.

Mathematically, the problem becomes: optimize (23) subject to (24) and (25) constraints.

$$\mathcal{F} = \sum_{s=0}^{N_s-1} \sum_{j,j'=0}^{D-1} \sum_{k,k'=0}^{n-1} b_{s,jk} S_{jk;j'k'} b_{s,j'k'}^* \xrightarrow{b} \max \quad (23)$$

$$\delta_{jj'} = \sum_{s=0}^{N_s-1} \sum_{k=0}^{n-1} b_{s,jk} b_{s,j'k}^* \quad j, j' = 0 \dots D - 1 \quad (24)$$

A selection of the gauge is required to avoid problem degeneracy, for example take the canonical form.

$$0 = \sum_{j=0}^{D-1} \sum_{k=0}^{n-1} b_{s,jk} b_{s',jk}^* \quad s \neq s', s = 0 \dots N_s - 1 \quad (25)$$

Other gauges such as Cholesky decomposition (on index  $s$ ) can possibly be used instead.

The main feature of the numerical method we developed consists of using eigenvalue problem as the algorithm's building block. If we consider a subset of (24) constraints, then the optimization problem can be readily solved. Consider the squared Frobenius norm of  $b_{s,jk}$  to be a "simplified constraint":

$$\sum_{s=0}^{N_s-1} \sum_{j=0}^{D-1} \sum_{k=0}^{n-1} |b_{s,jk}|^2 = D \quad (26)$$

This is a partial constraint (it is the sum of all diagonal elements in (24)). For this partial constraint the optimization problem (23) is equivalent to an eigenvalue problem — it can be directly solved by considering a vector of dimension  $N_s D n$  obtained from the  $b_{s,jk}$  operator by saving all its components to a **single vector**, row by row. With  $s$ -independent  $S_{jk;j'k'}$  each eigenvalue is degenerate  $N_s$  times; the Lagrange multipliers  $\nu_{ss'}$  in (A2) may potentially remove this degeneracy. This partially constrained problem is the main building block of our numerical algorithm. Whereas most existing learning algorithms use either first order gradient-style methods or second order Newtonian methods for optimization, an eigenvalue problem is the building block of the algorithm in this paper.<sup>3</sup> This represents a transition from mathematical analysis tools (e.g., gradient, derivative, etc.) to using algebraic tools (eigenproblem). This transition enables the move from single solution methods to multiple solutions (eigenvectors). This transition from analysis to algebra makes finding the global maximum much more likely, not to mention providing a better understanding of the problem itself.

### III. A UNITARY MAPPING OF MIXED STATES

In quantum channel learning there are two sources of mixed states. First, the original input density matrices (7) can be in a mixed state. Secondly, the quantum channel itself, through the transformation (4), can create mixed states. In this section we consider the first source. Let the data be of mixed states according to mapping (7), while the quantum channel itself is assumed to be unitary (3) or possibly partially unitary with  $D \leq n$ . To apply our theory and software we need to obtain a tensor  $S_{jk;j'k'}$  with which to solve the optimization problem (23), now with  $N_s = 1$ .

$$\mathcal{F} = \sum_{j,j'=0}^{D-1} \sum_{k,k'=0}^{n-1} u_{jk} S_{jk;j'k'} u_{j'k'}^* \xrightarrow{u} \max \quad (27)$$

$$\delta_{jj'} = \sum_{k=0}^{n-1} u_{jk} u_{j'k}^* \quad j, j' = 0 \dots D-1 \quad (28)$$

---

<sup>3</sup> We think that the reasons why first order gradient methods are preferred in neural networks over second order Newtonian-type methods are: 1. The problem may be of very high dimension, and first order methods do not need to store a large Hessian matrix. 2. The biggest problem in learning optimization is not local maximums, but saddle points[34]. First order methods are less likely to get stuck in a saddle point. 3. For some tasks, first-order gradient-type methods may offer better computational complexity, see Appendix D below.

If we take the tensor (18) as  $S_{jk;j'k'}$  — then we underestimated fidelity, which leads to unusual effects, see Table I row  $F^{\varrho\sigma}$ .

To obtain a proper estimation of the closeness of states for density matrix input, consider the mapping of the square roots of density matrices.

$$\sqrt{\rho^{(l)}} \rightarrow \sqrt{\varrho^{(l)}} \quad \text{weight } \omega^{(l)} \quad (29)$$

and use it instead of (7), as if it were the actual density matrix mapping. For a unitary mapping  $\mathcal{U}$  the same quantum channel converts both the density matrix and its square root. If  $\sqrt{\varrho} = \mathcal{U}\sqrt{\rho}\mathcal{U}^\dagger$  then  $\varrho = \mathcal{U}\sqrt{\rho}\mathcal{U}^\dagger\mathcal{U}\sqrt{\rho}\mathcal{U}^\dagger = \mathcal{U}\rho\mathcal{U}^\dagger$ ; also see Exercise 9.14 from [26], p. 410: for any positive operator  $A$ ,  $\sqrt{UAU^\dagger} = U\sqrt{A}U^\dagger$ . Thus we can use<sup>4</sup>

$$S_{jk;j'k'} = \sum_{l=1}^M \omega^{(l)} \left( \sqrt{\varrho^{(l)}} \right)_{jj'} \left( \sqrt{\rho^{(l)}} \right)_{kk'} \quad (30)$$

in the optimization problem if we apply the closeness (10) to the mapping of density matrix square root (29). Let us denote it as  $F^{\sqrt{\varrho}\sqrt{\rho}}$ .

$$F^{\sqrt{\varrho}\sqrt{\rho}} = \text{Tr} \sqrt{\varrho} \sigma(\sqrt{\rho}) \quad (31)$$

Here the entity denoted as  $\sigma(\sqrt{\rho})$  is obtained by passing  $\sqrt{\rho}$  through the quantum channel using either the transformation in (4) or (3). In the case (3), where  $N_s = 1$ , this  $F^{\sqrt{\varrho}\sqrt{\rho}}$  is equal to the proper fidelity  $F^{prop}$  (11). The approach creates no difficulty in implementation, as the square roots of density matrices are calculated upfront and then used as if they were actual measurements. The  $S_{jk;j'k'}$  from (30) corresponds to the density matrix square root mapping in (29). For a pure state  $\rho = |\psi\rangle\langle\psi|$ , we have  $\rho = \sqrt{\rho}$ , and (31) gives the same result as (9).

### A. A Conditional Entropy Based Similarity

In the case of density matrix unitary mapping, there is a possible alternative to the square root mapping  $\sqrt{\rho} \rightarrow \sqrt{\varrho}$  (29) that has been considered above. Now consider the mutual

<sup>4</sup> Note that from the invariance property of unitaries,  $Uf(A)U^\dagger = f(UAU^\dagger)$ , for any positive operator  $A$  and  $p \geq 0$ , we have  $(UAU^\dagger)^p = UA^pU^\dagger$ . Thus, a form of  $S_{jk;j'k'} = \sum_{l=1}^M \omega^{(l)} (\varrho^{(l)})_{jj'}^p (\rho^{(l)})_{kk'}^q$  with  $p + q = 1$  can be considered. However, an arbitrary  $p$  and  $q$  does not correspond to the fidelity in Eq. (11), which has  $p = q = 1/2$ , resulting in Eq. (30). For an arbitrary  $N_s$ , any form with  $q = 1$  can be reduced to a QCQP optimization problem. Equation (10) corresponds to  $p = q = 1$ . One can consider a form with  $q = 1$  and some  $p$  as a quadratic form proxy for the general quantum channel fidelity.

information between the density matrices  $\varrho$  and  $\sigma$ : the **Kullback-Leibler divergence**[35–37]. Given two density matrices  $\varrho$  and  $\sigma$ , the conditional entropy is

$$S(\varrho||\sigma) = \text{Tr}(\varrho \ln \varrho - \varrho \ln \sigma) \quad (32)$$

An interpretation of the Kullback Leibler divergence of  $\varrho$  from  $\sigma$  is the expected excess surprise from using  $\sigma$  as a model instead of  $\varrho$  when the actual distribution is  $\varrho$ . It satisfies the Gibbs inequality:  $S(\varrho||\sigma) \geq 0$ ,  $S(\varrho||\sigma) = 0$  iff  $\varrho = \sigma$ . For  $\sigma$  obtained as a unitary mapping (3) from  $\rho$ , the  $\ln \sigma = \ln(U\rho U^\dagger)$  can be calculated using the **invariance property** of the unitaries. Taking into account that  $\text{Tr}(\varrho \ln \varrho)$  is a constant one may consider the mapping

$$\ln \rho^{(l)} \rightarrow \varrho^{(l)} \quad \text{weight } \omega^{(l)} \quad (33)$$

and use the tensor

$$S_{jk;j'k'} = \sum_{l=1}^M \omega^{(l)} (\varrho^{(l)})_{jj'} (\ln \rho^{(l)})_{kk'} \quad (34)$$

as if it were the actual density matrix mapping. With this  $S_{jk;j'k'}$ , the functional  $\mathcal{F}$  no longer has the meaning of the number of observations, and special care should be taken when regularizing  $\rho$  states with eigenvalues equal to zero due to logarithm calculations, the  $\varrho$  must also be zero in this state. This information divergence measure is expected to be a good alternative to the square root mapping (30), but it requires more research, especially regarding the possibility of generalizing it to a quadratic form for quantum channels with Kraus rank  $N_s > 1$  and its property of not satisfying the triangle inequality.

## B. A Demonstration Of Density Matrix Square Root Mapping

Let us demonstrate the advantages of using the square root of the density matrix in quantum channel learning. Here and below we need a number of random density matrix input states  $\rho^{(l)}$  (of a given rank  $N_r$ , all real for simplicity) that we map with the quantum channel of form (3) or (4) to  $\varrho^{(l)}$ . The states are created from random vectors  $v_{r,k}^{(l)}$  as follows:

$$\rho_{kk'}^{(l)} = \frac{1}{\text{Norm}^{(l)}} \sum_{r=0}^{N_r-1} v_{r,k}^{(l)} v_{r,k'}^{(l)} \quad (35)$$

For every  $l$  we generate  $N_r$  random real vectors  $v_{r,k}^{(l)}$  of dimension  $n$ . The density matrix is obtained as a  $r = 0 \dots N_r - 1$  sum of the **dyadic product** of vector  $v_{r,k}^{(l)}$  with itself, and then it

is normalized to  $1 = \sum_{k=0}^{n-1} \rho_{kk}^{(l)}$ . These  $\rho_{kk}^{(l)}$  are used as input density matrices in our numerical experiments.

A demonstration is performed as follows. We generate a quantum channel (4) of dimension  $n = D = 20$  with a given Kraus rank  $N_s$ . Then for  $N_r = 1 \dots 20$ , we generate  $M = 1000$  input density matrices  $\rho_{kk'}^{(l)}$  (35) of rank  $N_r$  and for each of them, we obtain the output density matrix  $\varrho_{jj'}^{(l)}$  by passing  $\rho_{kk'}^{(l)}$  through the quantum channel. On this  $\rho_{kk'}^{(l)} \rightarrow \varrho_{jj'}^{(l)}$  mapping we evaluate the total fidelity  $\mathcal{F}$  (8) for different definitions of  $F$  (with the same  $B_s$  defining the quantum channel,  $\sigma = \sum_{s=0}^{N_s-1} B_s \rho B_s^\dagger$ ). The result is presented in Fig. 1; see `com/polytechnik/algorithms/DemoDMSqrtRhoMappingTest.java` for an implementation. For the  $F^{prop}$  (11) the total fidelity (8) is equal to the maximal value  $\mathcal{F}^{prop} = M$  since the same  $B_s$  were used both in construction and in evaluation of the quantum channel. However, a QCQP optimization problem that we can solve, an algebraic problem (43), must be a quadratic function on  $\mathcal{U}$ . The  $F^{prop}$  is not such a function, which makes it problematic to use. A simple  $F^{\varrho\sigma}$  as density matrices trace (10) is a quadratic function, but it is not normalized for  $N_r > 1$  or  $N_s > 1$ : it produces the exact result only for pure state input ( $N_r = 1$ ) and unitary ( $N_s = 1$ ) mappings, see Fig. 1a. The mapping  $\sqrt{\rho_{kk'}^{(l)}} \rightarrow \sqrt{\varrho_{jj'}^{(l)}}$  allows us to introduce  $F^{\sqrt{\varrho}\sqrt{\rho}}$  (31) considering  $\sqrt{\varrho}$  and  $\sqrt{\rho}$  as if they were actually measured density operators (29). In the unitary case  $N_s = 1$  it produces the correct value  $\mathcal{F} = M$  for any rank of the input density matrix  $\rho$ , this is an important result of this work. In this  $N_s = 1$  case for each  $N_r$  we also run the operator  $\mathcal{U}$  reconstruction algorithm of Appendix A — it is correctly (with zero error) recovered (up to a sign) for any  $N_r$  and any measure of closeness used:  $F^{\varrho\sigma}$ ,  $F^v$ ,  $F^{N\varrho^2}$ , or  $F^{\sqrt{\varrho}\sqrt{\rho}}$ , regardless of the specifics of  $F$ . This occurs only for exact mapping; when a noise is present in the data — the low values of  $F^{\varrho\sigma}$  make it problematic to reconstruct the quantum channel from input density matrices. The  $F^{\sqrt{\varrho}\sqrt{\rho}}$  does not have this limitation as for  $N_s = 1$  it is equal to proper fidelity  $F^{prop}$  (11). We see  $F^{\sqrt{\varrho}\sqrt{\rho}}$  as a very promising approach to the problem of reconstructing a unitary operator from a set of mixed states density matrix mapping. Another one is  $F^v$  (14), which treats density matrices as “vectors”, but it does not have a clear physical meaning.

The situation with quantum channels of Kraus rank  $N_s > 1$  is more problematic. The  $F^{\varrho\sigma}$  strongly underestimates the fidelity, and the  $F^{\sqrt{\varrho}\sqrt{\rho}}$  is no longer correct because when there is more than one term in the sum of (4), no quantum channel simultaneously converts  $\rho$  and  $\sqrt{\rho}$  between Hilbert spaces. The main problem we encountered is constructing a fidelity that

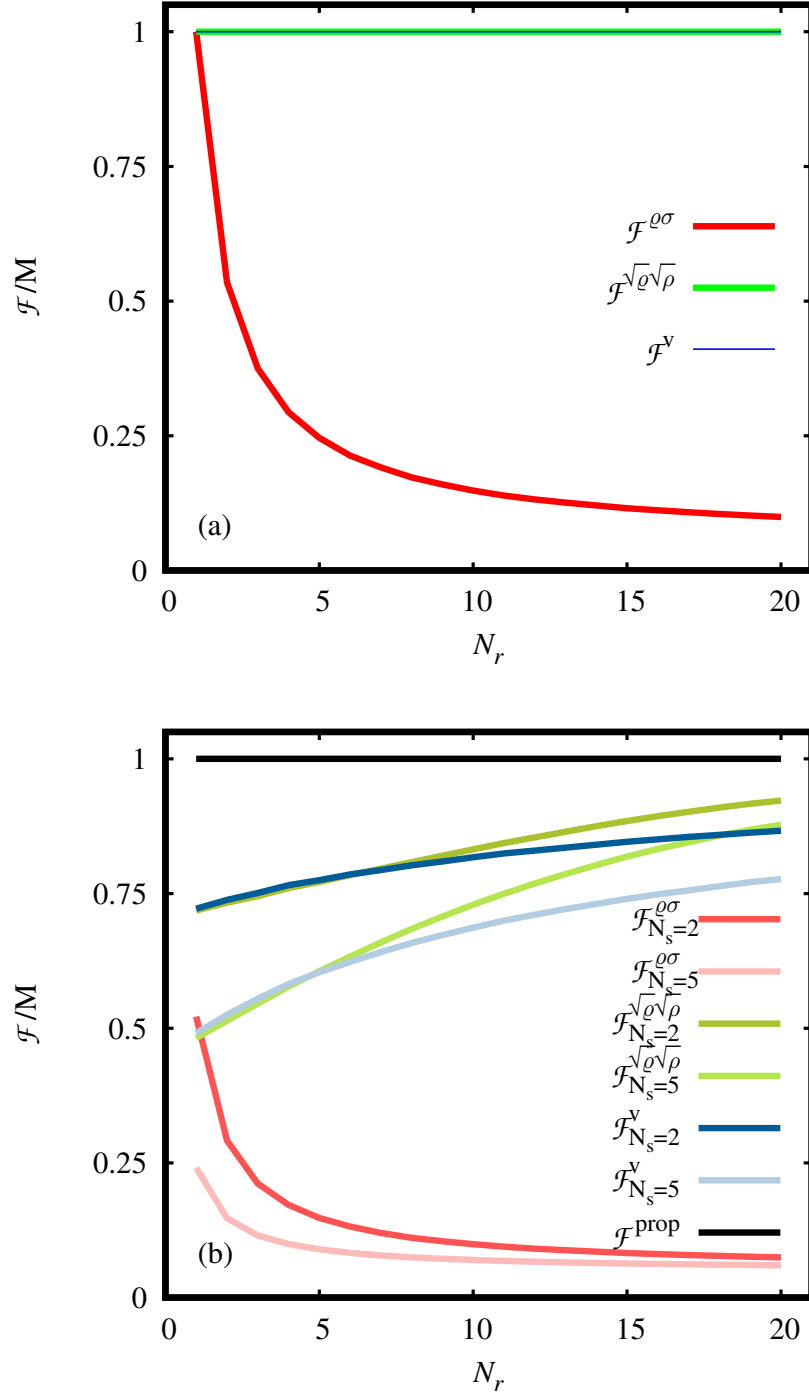


FIG. 1. For known density matrix mappings ( $\mathcal{F}^{prop} = M = 1000$ ) evaluate the total fidelity  $\mathcal{F}$  (8) (with different definitions of closeness  $F$ ) as a function of the rank  $N_r$  of the input density matrix and the quantum channel's Kraus rank  $N_s$ . Problem dimension is  $n = D = 20$ . (a) Unitary mapping  $N_s = 1$ : The  $\mathcal{F}^{q\sigma}$  (10) strongly depends on the rank of the input density matrices. The  $\mathcal{F}^{\sqrt{q}\sqrt{p}}$  (corresponding to (29) mapping) and  $\mathcal{F}^v$  (14) produce the exact result  $\mathcal{F} = M$  in the unitary mapping case. (b) Multiple terms  $N_s > 1$ : The  $\mathcal{F}^{q\sigma}$  decreases even more strongly, while  $\mathcal{F}^{\sqrt{q}\sqrt{p}}$  and  $\mathcal{F}^v$  are no longer exact.

can be expressed as a quadratic function of  $B_s$  in the form of (16). Not every fidelity can be expressed in this form. Using a proxy, such as  $F^{e\sigma}$ ,  $F^{N\varrho^2}$ , etc., creates difficulties – for example it may become possible to obtain a fidelity value greater than the one calculated with the exact mapping of the quantum channel. This is the main requirement for a “fidelity proxy”: it should reach the maximal value at the exact mapping of the quantum channel, see Section IV A below for counterexamples.

#### IV. A HIERARCHY OF UNITARY OPERATORS

As discussed below in Appendix A, a direct numerical attempt to obtain Kraus operators (4) in canonical gauge (25) does not work due to the degeneracy caused by the increased eigenproblem dimension from  $Dn$  to  $N_s Dn$  and the difficulty in formulating convergence helper constraints. We previously attempted to use a completely different approach for finding Kraus operators that involved a special parametrization of operators. However, it turned out to be not very stable numerically and applicable only to problems of small dimensions,  $n \lesssim 3$ , such as  $\text{SO}(3)$ , see [38] for a similar consideration.

Let us approach the problem of quantum channel reconstruction from a “numerical algorithm” perspective. What quantum channel problem we can efficiently solve for high dimension, i.e.  $D \leq n > 50$ ? Actually — only the problem of finding the optimal unitary mapping (3) that maximizes (27) subject to partial unitarity constraints (28).

Our algorithm [11] has good stability and convergence. For an updated version see Appendix A and set  $N_s = 1$  in all formulas. Now, instead of optimizing the Kraus problem (23) subject to constraints (24) and (25), consider the unitary hierarchy. A quantum channel is built as<sup>5</sup>

$$A^{OUT} = \sum_{s=0}^{N_s-1} |w_s|^2 \mathcal{U}^{[s]} A^{IN} \mathcal{U}^{[s]\dagger} \quad (36)$$

$$1 = \sum_{s=0}^{N_s-1} |w_s|^2 \quad (37)$$

where  $\mathcal{U}^{[s]}$  are partially unitary operators satisfying the constraints (28) and  $|w_s|^2$  are positive real weights. Since the weights are positive, the channel is a convex combination of unitary

<sup>5</sup> This quantum channel has an interesting physical interpretation. If (36) represents a time evolution, then one may think of it as a quantum system evolving with several Hamiltonians at once,  $\mathcal{U}^{[s]} = \exp[-i\frac{t}{\hbar}H_s]$ , rather than as a system evolving with the single Hamiltonian  $H = \sum_s H_s$ .



channels — a **mixed unitary channel**. Note that not every quantum channel can be represented as a mixed-unitary channel[39]. The authors [40] proved that for every mixed-unitary channel, the following rank inequality holds between the Kraus rank  $N_s^K$  (4) and the mixed-unitary rank  $N_s^U$  (the number of terms in (36)).

$$N_s^U \leq (N_s^K)^2 - N_s^K + 1 \quad (38)$$

It may look as the sum (36) exactly represents the Kraus quantum channel (4) with  $B_s = w_s \mathcal{U}^{[s]}$  (they are not necessary in the canonical form (6), see Appendix A 1 for a transformation), but actual Kraus rank can be lower than the number of terms in the sum (36). However, if we replace the sum of general Kraus operators  $B_s$  by a sum of partially unitary operators  $\mathcal{U}^{[s]}$ , then we can apply our numerical algorithm to build the quantum channel as a hierarchy of partially unitary operators. Considering this convex combination of unitary channels we can solve the problem incrementally. This is the cost required to apply our numerical method[11] to quantum channel reconstruction.

The idea of finding  $\mathcal{U}^{[s]}$  is similar to density matrix reconstruction. Assume we have a density matrix  $\varrho$  of dimension  $N_s$  with some eigenvectors  $\phi^{[s]}$  and eigenvalues  $P^{[s]}$ , a convex combination of pure states.

$$\varrho = \sum_{s=0}^{N_s-1} P^{[s]} |\phi^{[s]}\rangle \langle \phi^{[s]}| \quad (39)$$

$$|\varrho|\phi^{[s]}\rangle = P^{[s]} |\phi^{[s]}\rangle \quad (40)$$

To recover  $|\phi^{[s]}\rangle$  and  $P^{[s]}$  one may solve a sequence of  $N_s$  constrained optimization problems.

$$P = \frac{\langle \phi | \varrho | \phi \rangle}{\langle \phi | \phi \rangle} \xrightarrow[\phi]{} \max \quad (41)$$

$$0 = \langle \phi | \phi^{[s']}\rangle \quad s' < s \quad (42)$$

On each step a pair  $(P^{[s]}, |\phi^{[s]}\rangle)$  is obtained from the optimization problem (41) subject to the homogeneous linear constraints (42). The probabilities  $P^{[s]}$  decrease with  $s$  and the obtained pairs form a hierarchy of states from most probable to less probable.

In [10, 11, 20] we formulated a novel algebraic problem

$$S\mathcal{U} = \lambda\mathcal{U} \quad (43)$$

where superoperator  $S$  is a tensor  $S_{jk;j'k'}$ , “eigenvector”  $\mathcal{U}$  is a partially unitary operator (3) represented by a  $D \times n$  matrix  $u_{jk}$ , and “eigenvalue”  $\lambda$  is a Hermitian  $D \times D$  matrix of Lagrange multipliers  $\lambda_{ij}$  (A26). See Appendix B below for its time-dependent form (B1).

To construct a hierarchy of (Lagrange multipliers, operator) pairs  $(\lambda^{[s]}, \mathcal{U}^{[s]})$  of decreasing fidelity  $\mathcal{F}^{[s]} = \text{Tr} \lambda^{[s]}$  we need an analogue of the already found states’ orthogonality condition. Whereas in (42) for a regular eigenproblem with a scalar eigenvalue the result is the same regardless of performing an inner product with the “numerator”  $P^{[s]} |\phi^{[s]}\rangle$  or “denominator”  $|\phi^{[s]}\rangle$  terms, in the algebraic problem (43) with  $\lambda_{ij}$  being Hermitian matrix this is no longer the case. We have a number of options for orthogonality conditions for already found states  $s' < s$ , for example (44a) is an analogue of the “numerator” and (44b) of the “denominator”.

$$0 = \sum_{i,j=0}^{D-1} \sum_{k=0}^{n-1} u_{ik} \lambda_{ij}^{[s']} u_{jk}^{[s']*} = \langle \mathcal{U} | \lambda^{[s']} | \mathcal{U}^{[s']} \rangle \quad (44a)$$

$$0 = \sum_{j=0}^{D-1} \sum_{k=0}^{n-1} u_{jk} u_{jk}^{[s']*} = \langle \mathcal{U} | \mathcal{U}^{[s']} \rangle \quad (44b)$$

However the most convenient constraints correspond to a hierarchy of orthogonal quantum channels:

$$0 = \sum_{j,j'=0}^{D-1} \sum_{k=k'=0}^{n-1} u_{jk} S_{jk;j'k'} u_{j'k'}^{[s']*} = \langle \mathcal{U} | S | \mathcal{U}^{[s']} \rangle \quad (45)$$

Our numerical algorithm allows any homogeneous linear constraint to be incorporated into (A4). We believe that the appropriate orthogonality constraint on the previous state is the homogeneous linear constraint (45). We also use the “denominator”-type constraint (44b) in Appendix B below to construct a density tensor. The reason why we choose the (45) form — assume an operator  $\mathcal{V}$  is built as

$$\mathcal{V} = \sum_{s=0}^{N_s-1} w_s \mathcal{U}^{[s]} \quad (46)$$

Then, using the constraint (45), obtain

$$\langle \mathcal{V} | S | \mathcal{V} \rangle = \sum_{s=0}^{N_s-1} |w_s|^2 \langle \mathcal{U}^{[s]} | S | \mathcal{U}^{[s]} \rangle \quad (47)$$

i.e. a single nonunitary operator  $\mathcal{V}$  can completely define a quantum channel (36) with  $N_s > 1$ .

For a given  $\mathcal{V}$  the expansion weights in (46) are obtained as

$$w_s = \frac{\langle \mathcal{V} | S | \mathcal{U}^{[s]} \rangle}{\langle \mathcal{U}^{[s]} | S | \mathcal{U}^{[s]} \rangle} \quad (48)$$

A hierarchy of (Lagrange multipliers, operator) pairs is constructed by performing (27) partially unitary optimization (not Kraus!)  $N_s$  times. For each  $s$  include the constraints (45) from all previous optimizations  $s' = 0 \dots s - 1$  into the full set (A4). The result of this procedure is a hierarchy of  $N_s$  pairs  $\left(\lambda_{ij}^{[s]}, u_{jk}^{[s]}\right)$  with decreasing fidelity.

The tensor  $S_{jk;j'k'}$  can be expanded:

$$S \approx \sum_{s=0}^{N_s-1} \frac{1}{\mathcal{F}^{[s]}} |S| \mathcal{U}^{[s]} \langle \mathcal{U}^{[s]} | S | \quad (49)$$

$$\mathcal{F}^{[s]} = \sum_{i=0}^{D-1} \lambda_{ii}^{[s]} = \langle \mathcal{U}^{[s]} | S | \mathcal{U}^{[s]} \rangle \quad (50)$$

$$D = \langle \mathcal{U}^{[s]} | \mathcal{U}^{[s]} \rangle \quad (51)$$

$$\mathcal{F}^{[s]} \delta_{ss'} = \langle \mathcal{U}^{[s]} | S | \mathcal{U}^{[s']} \rangle \quad (52)$$

This expansion is similar to the eigenvector expansion (39), but ket  $|\cdot\rangle$  and bra  $\langle \cdot|$  are now operators, and  $S$  is called a **superoperator**. Note that usual orthogonality does not hold  $\langle \mathcal{U}^{[s]} | \mathcal{U}^{[s']} \rangle \neq \delta_{ss'}$ , (51) follows from the partial unitarity constraints (28), and (52) follows from the hierarchy constraints (45). By continuously solving ( $s = 0, 1, 2, \dots$ ) the optimization problem (27) subject to the constraints (28) and (45) we can build the required hierarchy of solutions  $\left(\lambda_{ij}^{[s]}, u_{jk}^{[s]}\right)$  to reconstruct the quantum channel from it, see `com/polytechnik/kgo/KGOHierarchy.java` for an implementation. A possible issue with the constraint (45) is that for the solution  $s$  in the hierarchy, Eq. (43) is not satisfied for some projections, the number of which equals the number of previous states in the hierarchy; for the ground state  $s = 0$ , all projections are satisfied. However, the main advantage of (45) is that the matrix (52) for an approximated  $S$  from (49) with the full  $N_s = Dn$  basis is equal to the corresponding matrix with the exact  $S$ . For the reconstruction of a quantum channel as a mixed unitary channel, a few solutions with high  $\mathcal{F}^{[s]}$  may be sufficient in a number of practical ML problems. The obtained hierarchy may have up to  $Dn$  solutions, which is the maximum possible number  $N_s$  of terms in the Kraus sum (4).

An application of the constructed hierarchy typically occurs when a nonunitary operator  $\mathcal{V}$  is available that provides a high value of fidelity. An example of such an operator can be a solution to the optimization problem (27) subject to partial constraints (Eq. (26) with  $N_s = 1$ ), which does not satisfy the full set of constraints (28). The expansion (48) allows constructing a mixed unitary channel that provides a similar value of fidelity. The problem

of constructing a quantum channel is reduced to two steps: 1. Find a nonunitary operator to estimate the maximum possible fidelity, and then 2. Find a mixed unitary channel that gives approximately the same fidelity. This approach allows us to consider a sequence of  $N_s = 1$  problems instead of a single problem with a high Kraus rank  $N_s$ . The only factor that limits this program is the formulation of the actual objective function as a quadratic form on quantum channel operators. In most cases, this is possible only through some kind of approximation, which significantly limits its applicability.

### A. A Demonstration of Constructed Hierarchy of Operators

Let us demonstrate the construction of operators hierarchy. First, we would like to mention that this problem is more difficult for numerical solution than the original (27). With the “external” constraints (45) added, we were able to solve the problem with  $N_s$  not greater than 5 to 10, depending on the values of  $D$  and  $n$ . The first difficulty arises from numerical instability and the need for a new **adjustment algorithm** to satisfy both the (28) and (45) constraints, however this problem is technical and can be resolved with a little effort. The second difficulty is a significant one. The optimization problem we are able to solve is a quadratic functional (16) optimization. However, mixed states fidelity (11) is not of this form, thus it cannot be directly applied to a general quantum channel. In the unitary case where  $N_s = 1$ , there are several good options, the most noticeable being  $F^{\sqrt{\varrho}\sqrt{\rho}}$  (31). In cases where  $N_s > 1$ , there are no good options, as depicted in Fig.1b. For this reason we try a number of “proxies” for total fidelity (8). Specifically  $F^{\varrho\sigma}$  (10),  $F^v$  (14),  $F^{N\varrho^2}$  (15) and  $F^{\sqrt{\varrho}\sqrt{\rho}}$  (31). These four expressions are quadratic in  $\mathcal{U}$  what allows us to build a  $S_{jk;j'k'}$  to which the optimization technique can be applied.

Consider a demonstration. We generate a quantum channel (4) of dimension  $D = n = 10$  with Kraus rank  $N_s = 3$ . Then generate  $M = 1000$  random states  $\rho_{kk'}^{(l)}$  (35) of rank  $N_r = 1$  (pure states). Passing them through the quantum channel obtain the  $\varrho_{jj'}^{(l)}$  of rank 3. If we calculate the total fidelity (8) with the state closeness  $F^{prop}$  (11) of this  $\rho_{kk'}^{(l)} \rightarrow \varrho_{jj'}^{(l)}$  mapping on  $B_s$  of the quantum channel, then the result will be  $\mathcal{F}^{prop}(B_s^{exact}) = M = 1000$  since these  $B_s$  make an exact mapping of the density matrix for all  $l$ . However,  $F^{prop}$  is not a quadratic function on  $B_s$  because of the square root taken from the sum (4) — our optimization algorithm cannot be applied. Instead, for the four aforementioned fidelity proxies

TABLE I. A demonstration of a unitary hierarchy for a quantum channel of dimension  $D = n = 10$  with Kraus rank  $N_s = 3$ . A total of  $M = 1000$  random density matrices  $\rho_{kk'}^{(l)}$  (35), each of rank  $N_r = 1$  (pure states), are generated and mapped to  $\varrho_{jj'}^{(l)}$  using the quantum channel with Kraus rank  $N_s = 3$ . The mapping gives  $\mathcal{F}^{prop}(B_s^{exact}) = M = 1000$  for proper closeness (11), but we need a quadratic expression (16) for total fidelity. Four fidelity proxies are considered:  $F^{\varrho\sigma}$  (10),  $F^v$  (14),  $F^{N\varrho^2}$  (15) and  $F^{\sqrt{\varrho}\sqrt{\rho}}$  (31) to construct four different hierarchies. For  $N_s > 1$ , there always exists a single unitary mapping that produces a proxy-fidelity greater than that of the entire quantum channel, e.g. for  $F^{\sqrt{\varrho}\sqrt{\rho}}$  we have  $\mathcal{F}(B_s^{exact}) = 617.13 < \mathcal{F}^{[0]} = 627.04$ , whereas  $\mathcal{F}^{prop}(B_s^{exact}) = 1000 > \mathcal{F}^{prop}(\mathcal{U}^{[0]}) = 627.04$ . Also, note that among all four proxies,  $\mathcal{F}^{prop}(\mathcal{U}^{[i]})$  is exactly equal to  $\mathcal{F}^{[i]}$  only for the proxy-fidelity  $F^{\sqrt{\varrho}\sqrt{\rho}}$ , this follows from matching (31) to (11) in the case  $N_s = 1$ .

$F$ constr.	$\mathcal{F}(B_s^{exact})$	$\mathcal{F}^{[0]}$	$\mathcal{F}^{prop}(\mathcal{U}^{[0]})$	$\mathcal{F}^{[1]}$	$\mathcal{F}^{prop}(\mathcal{U}^{[1]})$	$\mathcal{F}^{[2]}$	$\mathcal{F}^{prop}(\mathcal{U}^{[2]})$
$F^{\varrho\sigma}$	393.83	410.59	625.00	344.94	558.10	303.17	507.44
$F^v$	626.80	652.47	625.18	552.36	559.80	487.67	510.98
$F^{N\varrho^2}$	1000.00	1039.44	625.33	886.38	561.45	786.07	514.41
$F^{\sqrt{\varrho}\sqrt{\rho}}$	617.13	627.04	627.04	571.86	571.86	545.94	545.94

that produce a quadratic target functional (16), we calculate a superoperator  $S_{jk;j'k'}$  from density matrix mapping, perform optimization, and construct  $N_s = 3$  unitary hierarchy. The result is presented in Table I, see `com/polytechnik/algorithms/DemoDMGeneralMappingTest.java` for an implementation and run it as `java com/polytechnik/algorithms/DemoDMGeneralMappingTest 2>&1 | grep Proxy=`.

First, we would note that all proxies (except  $F^{N\varrho^2}$ ) do not give  $M$  on exact quantum channel mapping  $\mathcal{F}(B_s^{exact})$ ; but this may be fixed by normalizing. The major difficulty is the fact that for all four proxies the maximal total proxy-fidelity is reached not on the exact mapping  $B_s$  of quantum channel. See the column  $\mathcal{F}^{[0]}$  — the total proxy-fidelity calculated on a single unitary operator  $\mathcal{U}^{[0]}$  exceeds the proxy-fidelity  $\mathcal{F}(B_s^{exact})$  of the entire quantum channel  $B_s$ . This is a general property of all quadratic on  $B_s$  fidelity forms (16) we have tested, the four listed in Table I as well as a few others: There exists a single unitary operator, the  $\mathcal{U}^{[0]}$  in (52) hierarchy, that yields proxy-fidelity exceeding that of the entire quantum channel. This is the primary issue encountered when applying our QCQP algorithm to reconstruct

a quantum channel with a Kraus rank  $N_s > 1$ . Whereas for  $N_s = 1$  density matrix unitary mapping we have a number of good options for quadratic on  $\mathcal{U}$  fidelity, in the case of  $N_s > 1$ , for all proxy-fidelities we have tested, there always exists a single unitary mapping that yields a proxy-fidelity greater than that of the entire quantum channel.

A quantum channel is always constructed with a preset mapping. The constraints (5) mean that a unit matrix from the Hilbert space  $IN$  should be mapped to a unit matrix in the Hilbert space  $OUT$ . In [10, 11] a quantum channel converting the Gram matrix in  $IN$  into the Gram matrix in  $OUT$  was considered, with a regularization it is equivalent to the same (5). In Section III a quantum channel mapping  $\sqrt{\rho} \rightarrow \sqrt{\varrho}$  was considered, in the  $N_s = 1$  unitary case the same quantum channel maps both  $\rho$  and  $\sqrt{\rho}$  what allowed us to apply unitary learning to density matrix mappings. The problem of constructing a good quadratic proxy-fidelity in the case of  $N_s > 1$  is a subject for future research.

### B. Trace Preserving Maps

A general quantum channel (4) is a mapping between two Hilbert spaces. In this paper, we primarily use it to convert the density matrix. Such a conversion may not preserve the matrix trace. For example, a partially unitary mapping with  $D < n$  and  $N_s = 1$  was considered in [11]. Mappings that preserve the matrix trace have **special importance** in quantum channel studies. A question arises: when does a general quantum channel preserve the matrix trace? We are interested in formulating the optimization problem (23) with quadratic constraints, e.g. (24). Consider a density matrix  $\rho_{kk'}$ . Applying a quantum channel with the operators  $b_{s,jk}$  to it and taking the trace yields the  $\text{Tr}\varrho$ .

$$\text{Tr}\varrho = \sum_{s=0}^{N_s-1} \sum_{j=0}^{D-1} \sum_{k,k'=0}^{n-1} b_{s,jk} \rho_{kk'} b_{s,jk'}^* \quad (53)$$

For the expression to ensure  $\text{Tr}\varrho = \text{Tr}\rho$  for an arbitrary  $\rho$ , these conditions should be satisfied:

$$\delta_{kk'} = \sum_{s=0}^{N_s-1} \sum_{j=0}^{D-1} b_{s,jk} b_{s,jk'}^* \quad (54)$$

Equation (54) is a familiar trace preservation condition

$$\sum_{s=0}^{N_s-1} B_s^\dagger B_s = \mathbb{1} \quad (55)$$

There are  $n(n+1)/2$  independent constraints on the quantum channel (4) to **preserve** the trace. For  $N_s = 1$  it immediately gives  $n = D$ . For  $n = D$  and an arbitrary  $N_s$ , (5) and (55) are identical. In the convex combination of unitary channels gauge (36) with  $D = n$  it gives (37). For trace calculating map (where  $D = 1$ ) it gives  $N_s = n$ . For these trace-preserving constraints their linear forms (A30), totaling  $(n-1)(n+2)/2$ , can be **added** to the constraints (A4) of the numerical method from Appendix A.

For  $n = D$ , the constraints (54) are equivalent to (24). For  $D < n$  the latter has fewer constraints than the former. When  $D < n$ , the value of  $N_s$  should be chosen to be large enough so that the Gram matrix (A20) is not degenerate, see (20) as an example of a trace-calculating quantum channel having  $D = 1$  and  $N_s = n$ . For  $D < n$  the matrix  $B_s$  has dimensions  $D \times n$ , which gives the matrix  $B_s^\dagger B_s$  a rank of  $D$ . Thus, for  $D < n$ , in the case of trace-preserving maps, the Kraus rank should be at least

$$N_s \geq n - D + 1 \quad (56)$$

which is the minimum Kraus rank. Otherwise, the Gram matrix (A20) becomes degenerate, and the trace preservation condition (55) cannot be satisfied.

## V. CONCLUSION

We construct “quantum mechanics over quantum mechanics” by generalizing eigenstates to quantum channel mapping operators. While in traditional quantum mechanics the stationary Schrödinger equation determines system eigenstates, in our approach the algebraic problem (43) determines quantum channel mapping operators. The total fidelity must be a quadratic function on operators (16) for the problem to be represented in the form (43).

The technique was applied to the unitary mapping of density matrices. Since, for unitary mapping, the same quantum channel converts both the density matrix and its square root, the most promising approach is to convert the density matrix mapping (7) to the density matrix square root mapping (29), to which we can apply our QCQP optimization algorithm. This allows us to employ unitary learning in the application to density matrix mapping, representing an important advancement from the commonly studied unitary mapping of pure states  $\phi_l = \mathcal{U}\psi_l$ ; it allows us to distinguish between a probabilistic mixture of states and their superposition. The technique was tested on a number of randomly generated density matrices

$\rho^{(l)}$  of different ranks unitary mapped to  $\varrho^{(l)}$ ; in all cases the unitary quantum channel (3) was perfectly recovered.

The problem was then generalized to quantum channels (4) with Kraus rank  $N_s > 1$ . Reconstructing a general quantum channel mapping appears to be significantly more challenging. First, we were unable to represent the proper fidelity (11) as a quadratic form of the mapping operators, necessitating the use of approximations discussed in Section II. Developing a better quadratic fidelity representation (for mapping operators) is a subject for future research and is closely tied to the physical interpretation of quantum channels. Second, even when an approximate quadratic form for fidelity is obtained, the algorithm described in Appendix A fails to converge for  $N_s > 1$ . We expect, however, that implementing advanced constraint methods could lead to improvements. If, instead of reconstructing an arbitrary quantum channel, we restrict the problem to constructing a mixed unitary channel (36) using a hierarchy of unitary operators, then the problem becomes solvable, as detailed in Section IV. The results, however, are less satisfactory than desired due to a “double approximation”: first, approximating fidelity with a quadratic form, and second, relying on a mixed unitary quantum channel instead of a general one.

In this work, we have studied quantum channels that convert: a unit matrix from *IN* to *OUT* (standard definition), a Gram matrix from *IN* to *OUT*, and a  $\sqrt{\rho}$  to  $\sqrt{\varrho}$ . We developed a method for finding the global maximum by solving a novel algebraic problem (43), see Appendix B, which presents a generalization of this algebraic problem to the nonstationary case and introduces a time-dependent Schrödinger-like equation for operator  $\mathcal{U}(t)$ . We anticipate that applying memoryless quantum channel mappings to other problems could provide a solid foundation for a new form of machine learning knowledge representation. For initial insights into quantum channels with memory, see Appendix C below.

## ACKNOWLEDGMENTS

This research was supported by Autrotech Group, [www.atpetek.pph](http://www.atpetek.pph), a resident company of the Skolkovo Technopark. We thank our colleagues from the Autrotech R&D department who provided insight and expertise that greatly assisted the research. Our grateful thanks are also extended to Mr. Gennady Belov for his methodological support in doing the data analysis.



In this work, we study quantum channels as a positive map between two Hilbert spaces. In the late 80s to early 90s, Ivan Anatol’evich Komarchev brought to V.M.’s attention the problem of positive maps between Hilbert spaces and their relation to the Radon-Nikodym derivative. This became the origin of this entire theme. His unexpected death in 2022 was a significant loss to the Department of Mathematics at [St.Petersburg Polytechnic University](#), his colleagues who collaborated with him, and everyone who knew Ivan Anatol’evich. This work is dedicated to his memory.

## Appendix A: Numerical Solution

The described numerical algorithm is a further development of the one presented in [\[11\]](#). The advancements over the previous version include:

- The ability to simultaneously include quadratic constraints of different types, such as [\(24\)](#) and [\(25\)](#), or alternatively, [\(54\)](#) and [\(25\)](#), requires the introduction of two kinds of Lagrange multipliers:  $\lambda_{ij}$  and  $\nu_{ss'}$ . These multipliers are obtained from the solution of the linear system [\(A25\)](#).
- Different groups of quadratic constraints create different convergence-helper linear constraints. A general method, which does not involve the use of a special basis as in [\[11\]](#), is formulated in [Appendix A 5](#). This method involves varying the total  $N_c$  quadratic constraints to obtain  $N_c - 1$  homogeneous linear constraints and a single “simplified” (partial) quadratic constraint [\(26\)](#), which facilitates the use of a standard eigenvalue problem solver as an iterative algorithm building block.
- In addition to the convergence-helper constraints, we can now include “external” linear constraints. These external constraints can be used to construct a hierarchy of solutions for the novel algebraic problem [\(43\)](#).
- Now that we have several types of constraints, especially “external” and convergence-helper constraints, the procedure for adjusting the solution to satisfy all constraints becomes more complicated. This is because adjustments for different groups of constraints may conflict with each other. Therefore, a conflict resolution strategy is required, such as the one discussed in [Appendix A 2](#).

- The algorithm described below converges only when  $N_s = 1$ : maximize (27) subject to the constraints (28). Currently, it does not work for a general quantum channel where  $N_s > 1$ , as the problem becomes degenerate and the algorithm fails to converge. However, the calculations below are presented for an arbitrary  $N_s$  because the technique provides a clear approach for adding constraints of various types. The current version of the algorithm converges only for unitary (and partially unitary) learning when  $N_s = 1$ , with the “external” constraints possibly used to build a **hierarchy of unitary operators**.

Consider the optimization problem: optimize (23) subject to (24) and the canonical gauge (25) constraints. For a general  $N_s$  the solution to the optimization problem is a matrix  $b_{s,jk}$  of dimensions  $N_s \times D \times n$ , corresponding to  $N_s$  Kraus operators  $B_s$  (4) that satisfy the orthogonality constraints (5) and the gauge constraints, i.e. (6). This is the most general form of mapping a Hilbert space  $|\psi\rangle$  of dimension  $n$  into a Hilbert space  $|\phi\rangle$  of dimension  $D$ . This is a variant of the **QCQP** problem and the technique we used in [11] can also be applied here. The tensor  $S_{jk;j'k'} = S_{j'k';jk}^*$  is Hermitian, it does not depend on the Kraus index  $s$ ; the calculations below can be generalized to  $s$ -dependent  $S_{s,jk;s',j'k'}$  except for Kraus operators transformation to a canonical form with Appendix A1. For simplicity we consider  $S_{jk;j'k'}$  and  $b_{s,jk}$  to be real and do not write the complex conjugated  $*$  below, a generalization to complex values is straightforward. Consider a Lagrangian

$$\begin{aligned}
\mathcal{L} = & \sum_{s=0}^{N_s-1} \sum_{j,j'=0}^{D-1} \sum_{k,k'=0}^{n-1} b_{s,jk} S_{jk;j'k'} b_{s,j'k'} \\
& + \sum_{s=0}^{N_s-1} \sum_{j,j'=0}^{D-1} \lambda_{jj'} \left[ \delta_{jj'} - \sum_{k'=0}^{n-1} b_{s,jk'} b_{s,j'k'} \right] \\
& - \sum_{s \neq s'=0}^{N_s-1} \sum_{j=0}^{D-1} \sum_{k=0}^{n-1} \nu_{ss'} b_{s,jk} b_{s',jk} \xrightarrow{b} \max
\end{aligned} \tag{A1}$$

There are  $D(D+1)/2$  constraints (24) and  $N_s(N_s - 1)$  constraints (25). Lagrange multipliers  $\lambda_{jj'}$  and  $\nu_{ss'}$  are Hermitian matrices with corresponding number of independent elements; the  $\nu_{ss'}$  matrix has all diagonal elements equal to zero. Introduce a matrix  $\mathcal{S}_{jk;j'k'}$  to consider

the quadratic form maximization problem obtained with the partial constraint (26)

$$\mathcal{S}_{s,jk;s',j'k'} = \delta_{ss'} \mathcal{S}_{jk;j'k'} - \lambda_{jj'} \delta_{ss'} \delta_{kk'} - \nu_{ss'} \delta_{jj'} \delta_{kk'} \quad (\text{A2})$$

$$\frac{\sum_{s,s'=0}^{N_s-1} \sum_{j,j'=0}^{D-1} \sum_{k,k'=0}^{n-1} b_{s,jk} \mathcal{S}_{s,jk;s',j'k'} b_{s',j'k'}}{\frac{1}{D} \sum_{s=0}^{N_s-1} \sum_{j=0}^{D-1} \sum_{k=0}^{n-1} b_{s,jk}^2} \xrightarrow{b} \max \quad (\text{A3})$$

Following [11], we calculate Lagrange multipliers  $\lambda_{jj'}$  and  $\nu_{ss'}$  from the current iteration  $b_{s,jk}$ , see Appendix A 4 below — then  $\mathcal{S}_{s,jk;s',j'k'}$  is fixed and the problem (A3) can be considered as an eigenvalue problem. Additional  $N_d$  linear constraints on  $b_{s,jk}$  can be incorporated to improve convergence or obtain a sequence of solutions, see Appendix A 5 for convergence improving constraints and (45) for obtaining a solution sequence.

$$0 = \sum_{s=0}^{N_s-1} \sum_{j=0}^{D-1} \sum_{k=0}^{n-1} C_{d;s,jk} b_{s,jk} \quad d = 0 \dots N_d - 1 \quad (\text{A4})$$

A common method of solving the eigenproblem (A3) with **homogeneous** linear constraints (A4) is Lagrange multipliers method[41]. Since we have a large number of constraints it is better to use direct elimination instead. From  $N_s D n$  independent components of  $b_{s,jk}$  select some general variables  $V_p$ ,  $p = 0 \dots N_V - 1$

$$N_V = N_s D n - \text{rank}(C_{d;s,jk}) \quad (\text{A5})$$

$$b_{s,jk} = \sum_{p=0}^{N_V-1} M_{s,jk;p} V_p \quad (\text{A6})$$

where the constraints in (A4) are eliminated. The method of selection could be **Gaussian elimination**, **QR decomposition**, or a similar technique. A simple implementation with row and column pivoting is used in `com/polytechnik/Utils/EliminateLinearConstraints_HomegrownLUFactorization.java`. The result is a matrix  $M_{s,jk;p}$  that converts  $N_V$  independent variables  $V_p$  to  $b_{s,jk}$  of  $N_s D n$  components satisfying all the constraints (A4). A new eigenproblem has the matrices  $\mathcal{S}_{p;p'}$  and  $Q_{p;p'}$  in the numerator and denominator respectively.

$$\mathcal{S}_{p;p'} = \sum_{s,s'=0}^{N_s-1} \sum_{j,j'=0}^{D-1} \sum_{k,k'=0}^{n-1} M_{s,jk;p} \mathcal{S}_{s,jk;s',j'k'} M_{s',j'k';p'} \quad (\text{A7a})$$

$$Q_{p;p'} = \sum_{s=0}^{N_s-1} \sum_{j=0}^{D-1} \sum_{k=0}^{n-1} M_{s,jk;p} M_{s,jk;p'} \quad (\text{A7b})$$

Then we can write the eigenproblem (A3) in the form,

$$\frac{\sum_{p,p'=0}^{N_V-1} V_p \mathcal{S}_{p;p'} V_{p'}}{\sum_{p,p'=0}^{N_V-1} V_p Q_{p;p'} V_{p'}} \xrightarrow{V} \max \quad (\text{A8})$$

$$\sum_{p'=0}^{N_V-1} \mathcal{S}_{p;p'} V_{p'}^{[i]} = \mu^{[i]} \sum_{p'=0}^{N_V-1} Q_{p;p'} V_{p'}^{[i]} \quad (\text{A9})$$

which has  $\text{rank}(C_{d;s,jk})$  less dimension and no linear constraints. The matrix  $Q_{p;p'}$  in the denominator is no longer a unit matrix. This is not an issue since any modern linear algebra package internally converts a generalized eigenproblem to a regular one, see e.g. [DSYGST](#), [DSPGST](#), [DPBSTF](#) and similar subroutines. Combining all together the algorithm becomes:

1. Take initial  $\lambda_{ij}$ ,  $\nu_{ss'}$  and linear constraints  $C_{d;s,jk}$  to solve the optimization problem (A8) with respect to  $V_p$ . The solution method involves solving an eigenvalue problem of dimension  $N_V$ , the number of columns in  $M_{s,jk;p}$  matrix. A new  $b_{s,jk}$  is obtained from  $V_p$  using (A6). The result:  $i = 0 \dots N_V - 1$  eigenvalues  $\mu^{[i]}$  and corresponding matrices  $b_{s,jk}^{[i]}$  reconstructed from  $V_p^{[i]}$ . The value of  $N_V$  is typically  $N_s D n - (D - 1)(D + 2)/2 - N_s(N_s - 1)/2$ . Additional constraints (further reducing  $N_V$ ) are added when constructing the operators hierarchy.
2. A heuristic is required to select the  $b_{s,jk}$  among all  $N_V$  eigenstates. There is a discussion about this in [11]. Our numerical experiments show that in most cases it is sufficient to always take the state of the maximal  $\mu^{[i]}$  to reach the global maximum. In the current implementation we always select this state.
3. Obtained  $b_{s,jk}$  satisfies only partial constraint (26). We need to adjust it to satisfy all the required constraints of orthogonality and canonical form. Adjust  $b_{s,jk}$  with Appendix A 2 to satisfy orthogonality (24), then apply the result to Appendix A 1 to convert the Kraus operators to canonical form (25). For  $N_s = 1$ , transformation to the canonical form is not required. With the resulting  $b_{s,jk}$  calculate Lagrange multipliers  $\lambda_{ij}$  and  $\nu_{ss'}$  for the next iteration as described in Appendix A 4.
4. For good convergence, in addition to Lagrange multipliers, we need to select a subspace for the next iteration variation. Using full size basis leads to poor convergence [10]. Use

Appendix A5 to obtain (from  $b_{s,jk}$ ) the linear constraints coefficients  $C_{d;s,jk}$  used in (A4). There is an important feature of this additional linear constraints approach. In addition to the constraints from Appendix A5, we can manually add some external constraints, such as requiring zero projection (45) onto already found solutions. This way a hierarchy of high  $\mathcal{F}$  solutions can be obtained.

5. Put these new  $\lambda_{ij}$  and  $\nu_{ss'}$  into (A2) and, using the basis  $V_p$  obtained (A6) from  $C_{d;s,jk}$ , calculate the generalized eigenproblem numerator and denominator matrices (A7) to be used in the next iteration. Repeat the iteration process until convergence to a maximum (presumably global) of  $\mathcal{F}$  with  $b_{s,jk}$  satisfying the constraints (24) and (25). If convergence is achieved, the Lagrange multipliers stop changing from iteration to iteration, and the  $\mu^{[i]}$  of the selected state in step 2 is close to zero. On the first iteration take zero initial values for the Lagrange multipliers and have no linear constraints.

For a general  $N_s$  implementation of this algorithm see `com/polytechnik/kg0/IterationalSubspaceLinearConstraintsNaiveKraus.java`. There is a special  $N_s = 1$  implementation of it `com/polytechnik/kg0/KG0IterationalSubspaceLinearConstraintsB.java`; they share a codebase and unit tests. Currently the algorithm converges only if  $N_s = 1$ ; in this case, except for implementation improvements and optimizations, it has properties very similar to the `com/polytechnik/kg0/KG0IterationalSubspaceLinearConstraints.java` implementation in [11].

## 1. Transforming Kraus Operators to Canonical Form

The selection of Kraus operators  $B_s$  in which to evaluate (23) is nonunique, there should be a gauge that regulates redundant degrees of freedom. For example consider data of exact unitary mapping  $|\psi_l\rangle\langle\psi_l| \rightarrow |\phi_l\rangle\langle\phi_l|$  with an operator  $\mathcal{U}$  providing perfect coverage in (2). Let we want to describe this data with (4) mapping of Kraus rank  $N_s = 2$ . Then any combination  $B_s = w_s \mathcal{U}$  with  $1 = \sum_s |w_s|^2$  is a solution. The problem becomes degenerate. This creates difficulty in both analyzing the results and achieving convergence in the iterative algorithm – any degeneracy greatly degrades it. It is common to consider Kraus operators  $B_s$  in the canonical form (6). The problem becomes: converting a given  $b_{s,jk}$  matrix to the canonical form of Eq. (25).

Kraus operators can be considered as some superoperator states, similar to density matrix states, but ket  $|\cdot\rangle$  and bra  $\langle\cdot|$  are now operators:  $\sum_s |B_s\rangle \langle B_s|$ , not vectors. This algebra can be used in the expansion of tensors, such as  $S_{jk;j'k'}$  (49) — then evaluating  $\mathcal{F}$  is an inner product of superoperators. In this work we are not going to discuss the construction of superoperators algebra and its applications. The most important application is the construction of operators hierarchy in Section IV, what allows us to reconstruct a quantum channel from observable data. There are a number of other interesting features that arise on this path and we hope to present a detailed discussion elsewhere. In this appendix we limit ourselves to one small but important problem.

Assume we obtained a partially Kraus operator in the form of a matrix  $b_{s,jk}$  that evaluates the fidelity to some  $\mathcal{F}$  (23). This  $b_{s,jk}$  does not satisfy the (25) constraints. The problem is to convert  $b_{s,jk}$  to a new  $\tilde{b}_{s,jk}$  satisfying the (25) while maintaining the same value of the functional  $\mathcal{F}$ . The solution is not unique and requires a gauge, akin to basis orthogonalization. Since the tensor  $S_{jk;j'k'}$  does not depend on the Kraus index  $s$ , a solution can be readily found. We use the same technique previously employed in adjusting for partial unitarity[11], as outlined in Appendix A 2. It's even simpler to apply it here since diagonal elements do not enter into the constraints defined in (25). Consider the Gram matrix in the Kraus space

$$G_{ss'} = \sum_{j=0}^{D-1} \sum_{k=0}^{n-1} b_{s,jk} b_{s',jk} \quad (\text{A10})$$

and solve an eigenproblem with it

$$|G|g^{[i]}\rangle = \lambda_G^{[i]} |g^{[i]}\rangle \quad i = 0 \dots N_s - 1 \quad (\text{A11})$$

Since the  $S_{jk;j'k'}$  does not depend on  $s$  one can verify that

$$\tilde{b}_{s,jk} = \sum_{s'=0}^{N_s-1} g_{s'}^{[s]} b_{s',jk} \quad (\text{A12})$$

satisfies the (25) constraints, whereas the value of  $\mathcal{F}$  (23) stays the same. See `com/polytechnique/kg/TransformToCanonicalFormKraus.java` for an implementation. The implementation is straightforward: calculate the Gram matrix (A10), solve the eigenproblem (A11), then convert the original  $b_{s,jk}$  to the basis of the found eigenvectors (A12);  $\mathcal{F}$  does not change with this transformation.

## 2. An Adjustment of Operators to Orthogonal

A  $b_{s,jk}$ , found at some iteration, may not satisfy all the required constraints. In Appendix A 1 we considered canonical form constraints (25), now consider orthogonality constraints (24). It is sufficient to consider a single  $u_{jk}$  since the same transform is applied to all  $b_{s,jk}$ . The Gram matrix

$$G_{jj'} = \sum_{k=0}^{n-1} u_{jk} u_{j'k} \quad (\text{A13})$$

(or  $G_{jj'} = \sum_{s=0}^{N_s-1} \sum_{k=0}^{n-1} b_{s,jk} b_{s,j'k}$  for  $N_s > 1$ ) is a unit matrix when all the constraints (24) are satisfied. When this is not the case — we want to change the  $u_{jk}$  in a way that the change is as small as possible. Following [11] we apply the same  $G_{jj'}^{-1/2}$  technique. Specifically, solve the eigenproblem

$$|G|g^{[i]}\rangle = \lambda_G^{[i]} |g^{[i]}\rangle \quad i = 0 \dots D-1 \quad (\text{A14})$$

and calculate the inverse square root of the Gram matrix

$$\|G^{-1/2}\| = \sum_{i=0}^{D-1} \frac{\pm 1}{\sqrt{\lambda_G^{[i]}}} |g^{[i]}\rangle \langle g^{[i]}| \quad (\text{A15})$$

By checking the result, one can verify that for any  $u_{jk}$  producing nondegenerated Gram matrix (A13) the matrix

$$\tilde{u}_{jk} = \sum_{i=0}^{D-1} G_{ji}^{-1/2} u_{ik} \quad (\text{A16})$$

satisfies all (28) constraints (or  $\tilde{b}_{s,jk} = \sum_{i=0}^{D-1} G_{ji}^{-1/2} b_{s,ik}$  and constraints (24) for  $N_s > 1$ ; in this transform we apply the same  $G_{ji}^{-1/2}$  to every Kraus operator  $B_s$  — the adjustment does not depend on  $s$ ). Previously considered transformation to the canonical form (A12) has an important difference from (A16). The transformation (A12) does not change any observable, it just reshuffles  $B_s$  among themselves in the (4) sum (gauge transform) in a way that the obtained  $B_s$  satisfy the canonical form constraints (6), specifically (25). But Eq. (A16) actually changes the solution, it is not possible to satisfy (24) without changing the solution itself. We will not repeat the lengthy discussion [11] about this  $G_{ji}^{-1/2}$  adjustment technique, just note that it introduces a “minimal disturbance” to a solution, which is very

advantageous for an iterative algorithm. This adjustment technique is implemented in `com/polytechnik/kg/AdjustToOrthogonalKraus.java`.

There is an issue that arises when constructing operators hierarchy in Section IV — there we do have external linear constraints (45) to be added to the iterative algorithm. Whereas the convergence-helper constraints of Appendix A5 are themselves calculated from the adjusted  $\tilde{u}_{jk}$  (A16), external constraints are different in the sense that they are preset. The application of  $G_{jj'}^{-1/2}$  to the current iteration  $u_{jk}$ , which satisfies external constraints due to (A6) incorporating all the linear constraints, may create a  $\tilde{u}_{jk}$  that does not satisfy external constraints. In most cases, this problem can be ignored since as iterations progress, the solution converges to the required subspace. However, it would be beneficial to modify the adjustment procedure to explicitly incorporate externally defined homogeneous linear constraints.

Consider external constraints

$$0 = \sum_{j=0}^{D-1} \sum_{k=0}^{n-1} C_{d;jk}^e u_{jk} \quad (\text{A17})$$

here  $C_{d;jk}^e$  represents, for example, constraints (45), where index  $d$  enumerates the constraints; they do not include convergence-helper constraints from Appendix A5, which are determined later, after the constraint-adjusted  $u_{jk}$  is obtained. To simplify projections bellow, consider  $C_{d;jk}^e$  and  $u_{jk}$  as vectors, and convert  $C_{d;jk}^e$  to an orthogonal form. One can either use Gram-Schmidt orthogonalization or, alternatively, apply the same  $G^{-1/2}$  technique we previously [used](#) for a different problem. Now we convert (A17)

$$G_{dd'} = \sum_{j=0}^{D-1} \sum_{k=0}^{n-1} C_{d;jk}^e C_{d';jk}^e \quad (\text{A18a})$$

$$\tilde{C}_{d;jk}^e = \sum_{d'} G_{dd'}^{-1/2} C_{d';jk}^e \quad (\text{A18b})$$

to an orthogonalized form  $\delta_{dd'} = \sum_{j=0}^{D-1} \sum_{k=0}^{n-1} \tilde{C}_{d;jk}^e \tilde{C}_{d';jk}^e$ ; the only purpose of this orthogonalization is to have a simple projection formula (A19) below. Consider the following iteration:

- Take the original  $u_{jk}$  and convert it to the form that satisfies orthogonality constraints (A16).



- This  $\tilde{u}_{jk}$  satisfies (28) but may not satisfy external linear constraints (A17). Remove the projections

$$\tilde{\tilde{u}}_{jk} = \tilde{u}_{jk} - \sum_d \tilde{C}_{d;jk}^e \sum_{j'=0}^{D-1} \sum_{k'=0}^{n-1} \tilde{C}_{d;j'k'}^e \tilde{u}_{j'k'} \quad (\text{A19})$$

Obtained  $\tilde{\tilde{u}}_{jk}$  satisfies external constraints (A17) but may not satisfy the orthogonality constraint (28).

- Repeat the process by performing a number of iterations. This iteration-adjustment algorithm has poor convergence per se and requires about a hundred iterations to obtain a  $u_{jk}$  that satisfies both the constraints (28) and (A17) exactly. However, since this iteration-adjustment is only step 3 of the main iterative algorithm, we do not necessarily need an exact solution. Numerical experiments show that 3 to 6 iterations are sufficient for the main iterative algorithm to converge in the presence of external linear constraints.

See `com/polytechnik/kg/AdjustToUnitaryWithEV_SubjectToLinearConstraints.java` for an implementation. As emphasized above, several iterations of these sequential adjustments are sufficient to achieve good convergence of the main algorithm. However, significant improvements are possible, and further study of this problem is necessary.

### 3. An Adjustment of a Quantum Channel to Preserve the Trace

A  $b_{s,jk}$ , found at some iteration, may not satisfy the constraints (54) of trace preservation. We can apply the same  $G^{-1/2}$  technique from Appendix A2 to adjust the quantum channel operators  $b_{s,jk}$  to preserve the trace. Consider the Gram matrix

$$G_{kk'} = \sum_{s=0}^{N_s-1} \sum_{j=0}^{D-1} b_{s,jk} b_{s,jk'} \quad (\text{A20})$$

and solve an eigenproblem with it

$$|G|g^{[i]}\rangle = \lambda_G^{[i]} |g^{[i]}\rangle \quad i = 0 \dots n-1 \quad (\text{A21})$$

Since we typically have  $D \leq n$  the Gram matrix (A20) may become degenerated. To avoid a degeneracy the number of terms  $N_s$  in the quantum channel should be large enough, there is

a minimum Kraus rank (56). For a nondegenerated  $G$  calculate its inverse square root:

$$\|G^{-1/2}\| = \sum_{i=0}^{n-1} \frac{\pm 1}{\sqrt{\lambda_G^{[i]}}} |g^{[i]}\rangle \langle g^{[i]}| \quad (\text{A22})$$

And apply it to the current solution  $b_{s,jk}$

$$\tilde{b}_{s,jk} = \sum_{q=0}^{n-1} G_{kq}^{-1/2} b_{s,jq} \quad (\text{A23})$$

By checking the result one can verify that  $\tilde{b}_{s,jk}$  satisfies all the trace preservation constraints (54), see `com/polytechnik/kg/AdjustToTracePreservingKraus.java` for an implementation. Note that the problem may become degenerate, which requires special attention. Also note that (A23) is similar to (A16) and, contrary to (A12), actually changes the  $b_{s,jk}$ . A possible conflict between different adjustments can potentially be resolved similarly to the resolution of a conflict between linear constraints and the adjustment (A16).

#### 4. Lagrange Multipliers Calculation

The variation of the Lagrangian  $\mathcal{L}$  (A1) must be zero in the iteration state  $b_{s,jk}$

$$\begin{aligned} 0 = \frac{1}{2} \frac{\delta \mathcal{L}}{\delta b_{s,iq}} &= \sum_{j'=0}^{D-1} \sum_{k'=0}^{n-1} S_{iq;j'k'} b_{s,j'k'} \\ &\quad - \sum_{j'=0}^{D-1} \lambda_{ij'} b_{s,j'q} - \sum_{s'=0}^{N_s-1} \nu_{ss'} b_{s',iq} \end{aligned} \quad (\text{A24})$$

There are a total of  $N_s D n$  equations. They are all satisfied if  $b_{s,jk}$  is extremal in (A3). The  $b_{s,jk}$  used, however, has the orthogonality adjustment procedure of Appendix A2 applied to it and the Lagrangian variation is no longer zero. Lagrange multipliers are Hermitian matrices  $\lambda_{ij}$  and  $\nu_{ss'}$ , they have  $D(D+1)/2$  and  $N_s(N_s-1)/2$  independent values thus all the  $N_s D n$  equations cannot be simultaneously satisfied. We need to select the  $\lambda_{ij}$  and  $\nu_{ss'}$  to ensure they best satisfy the zero variation condition for a given  $b_{s,jk}$ . Consider the  $L^2$  norm of (A24) and find the  $\lambda_{ij}$  and  $\nu_{ss'}$  minimizing the sum of squares.

$$\sum_{s=0}^{N_s-1} \sum_{i=0}^{D-1} \sum_{q=0}^{n-1} \left| \frac{\delta \mathcal{L}}{\delta b_{s,iq}} \right|^2 \xrightarrow{\lambda_{ij}, \nu_{ss'}} \min \quad (\text{A25})$$

To simplify the minimization of (A25) it is convenient to map  $\lambda_{ij}$  into a vector of independent components  $\lambda_r$ . One can note that for  $j \leq i$  the vector index can be taken as  $r = i(i+1)/2 + j$  (if  $j > i$  swap them). Similarly for  $\nu_r$  and  $\nu_{ss'}$  take the vector index as  $r = s(s-1)/2 + s'$  for  $s' < s$  otherwise swap the indexes. Then differentiate (A25) over  $\lambda_r$  and  $\nu_r$  to obtain a linear system with respect to them. Back in [11] a similar linear system was analytically solved and an explicit expression for  $\lambda_{ij}$  was obtained.

$$\lambda_{ij} = \text{Herm} \sum_{j'=0}^{D-1} \sum_{k,k'=0}^{n-1} u_{ik} S_{jk;j'k'} u_{j'k'} \quad (\text{A26})$$

This is an analytic solution in the case  $N_s = 1$  with  $u_{jk} = b_{0,jk}$  satisfying the (28) constraints, see `com/polytechnik/kg/LagrangeMultipliersPartialSubspace.java:calculateRegularLambda` for an implementation of this special case. In the general case we have two types of Lagrange multipliers:  $\lambda_{ij}$  and  $\nu_{ss'}$ , which prevents us from obtaining an analytic solution. But the problem is straightforward: take each term from (A24), square it, and sum over all  $s, i, q$ . Differentiate this sum of squares with respect to the vectors  $\lambda_r$  and  $\nu_r$  to obtain a linear system, then solve it. The calculations are straightforward but lengthy, see `com/polytechnik/kg/LagrangeMultipliersNaiveKraus.java:getLagrangeMultipliers` for an implementation that, given  $b_{s,jk}$  and  $S_{jk;j'k'}$ , calculates the Lagrange multiplier matrices  $\lambda_{ij}$  and  $\nu_{ss'}$ .

This approach differs from the commonly used one in that we only utilize the Lagrangian variation (A24) to calculate Lagrange multipliers. The  $b_{s,jk}$  is then determined from the eigenproblem solution with these Lagrange multipliers used in (A2). This allows us to apply constraint satisfying adjustments to the  $b_{s,jk}$  before using it in the calculation of Lagrange multipliers. We cannot simultaneously solve the optimization problem for Lagrange multipliers  $\lambda_{ij}$ ,  $\nu_{ss'}$  and the solution  $b_{s,jk}$ . Instead, on every iteration, we adjust the  $b_{s,jk}$  to satisfy all required constraints, solve the linear system to find the Lagrange multipliers, and then use them in an eigenproblem to find the next  $b_{s,jk}$ .

## 5. Convergence-Helper Linear Constraints

The problem we study involves quadratic constraints (24) and (25). These constraints cannot be directly incorporated into the eigenproblem (A3). However, homogeneous linear constraints can be easily incorporated. Let's construct the linear constraints corresponding to

the constrained local variation of the current iteration  $b_{s,jk}$ . The concept can be illustrated by varying a quadratic constraint  $\langle \mathbf{x} | \mathbf{y} \rangle = 0$  to obtain a linear constraint on  $\delta \mathbf{x}$  and  $\delta \mathbf{y}$  at fixed  $\mathbf{x} = u_{j*}$  and  $\mathbf{y} = u_{j'*}$ :  $\delta \langle \mathbf{x} | \mathbf{y} \rangle = \langle \delta \mathbf{x} | \mathbf{y} \rangle + \langle \mathbf{x} | \delta \mathbf{y} \rangle = 0$ ; similarly, two expressions  $1 = \langle \mathbf{x} | \mathbf{x} \rangle = \langle \mathbf{y} | \mathbf{y} \rangle$  are replaced by  $\langle \mathbf{x} | \mathbf{x} \rangle - \langle \mathbf{y} | \mathbf{y} \rangle = 0$ . By varying it, obtain a linear constraint on  $\delta \mathbf{x}$  and  $\delta \mathbf{y}$ :  $\langle \delta \mathbf{x} | \mathbf{x} \rangle + \langle \mathbf{x} | \delta \mathbf{x} \rangle - \langle \delta \mathbf{y} | \mathbf{y} \rangle - \langle \mathbf{y} | \delta \mathbf{y} \rangle = 0$ . In the case of unitary mapping ( $D = n$ ,  $N_s = 1$ ), this corresponds to replacing the  $n(n+1)/2$  quadratic constraints of the unitary property (28) with  $n(n+1)/2 - 1$  homogeneous linear constraints (A4), that simply reduce the search space dimension (A6), and a single quadratic constraint (26), referred to as a simplified (partial) constraint. This problem can then be iteratively solved with excellent convergence using an eigenvalue problem solver as a building block.

Consider (24)  $j \neq j'$  off-diagonal elements. There are  $D(D-1)/2$  total distinct ones. Varying (24), we obtain

$$C_{d;s,jk} = b_{s,j'k} \quad (\text{A27a})$$

$$C_{d;s,j'k} = b_{s,jk} \quad (\text{A27b})$$

Two equations set different elements in  $C_{d;s,jk}$  for the same  $d$ , they may be viewed as two initialization commands for the matrix  $C_{d;s,jk}$ . In (A27), the constraint index  $d$  takes  $D(D-1)/2$  distinct values corresponding to all  $j' < j$  pairs.

Consider the inhomogeneous constraints (24) corresponding to the diagonal elements with  $j = j'$ . There are  $D$  of them. Since the partial constraint (26) preserves the total norm, it is sufficient for all diagonal elements to be equal. Equality of diagonal elements constitutes a homogeneous constraint, yielding  $D-1$  constraints for a given  $b_{s,jk}$ .

$$C_{d;s,jk} = b_{s,jk} \quad (\text{A28a})$$

$$C_{d;s,j-1k} = -b_{s,j-1k} \quad (\text{A28b})$$

Similarly to the previous case, two equations set different elements in  $C_{d;s,jk}$  for the same  $d$ . In (A28) the constraint index  $d$  takes  $D-1$  distinct values corresponding to  $j = 1 \dots D-1$ . For an implementation of (A27) and (A28) see `com/polytechnik/kg0/LinearConstraintsKraus.java:getOrthogonalOffdiag0DiagEq`.

Consider the canonical form constraints on Kraus operators (25). There are no diagonal

elements, resulting in a total of  $N_s(N_s - 1)/2$  distinct constraints. Similarly obtain

$$C_{d;s,jk} = b_{s',jk} \quad (\text{A29a})$$

$$C_{d;s',jk} = b_{s,jk} \quad (\text{A29b})$$

Two equations set different elements in  $C_{d;s,jk}$  for the same  $d$ , they may be viewed as two initialization commands for the matrix  $C_{d;s,jk}$ . In (A29) constraint index  $d$  takes  $N_s(N_s - 1)/2$  distinct values corresponding to all  $s' < s$  pairs; for  $N_s = 1$  the constraints vanish. For an implementation of (A29) see `com/polytechnik/kgol/LinearConstraintsKraus.java:getOrthogonalKrausOffdiag0`.

Quadratic constraints for trace preservation (54) can generate  $(n-1)(n+2)/2$  homogeneous linear constraints, similar to how (A27) and (A28) were derived from the orthogonality constraint (24).

$$C_{d;s,jk} = b_{s,jk'} \quad (\text{A30a})$$

$$C_{d;s,jk'} = b_{s,jk} \quad (\text{A30b})$$

and

$$C_{d;s,jk} = b_{s,jk} \quad (\text{A30c})$$

$$C_{d;s,jk-1} = -b_{s,jk-1} \quad (\text{A30d})$$

see `com/polytechnik/kgol/LinearConstraintsKraus.java:getTracePreservingConstraints` for an implementation.

These linear homogeneous convergence-helper constraints can be summarized as follows:

- (A27) —  $D(D-1)/2$  in total, corresponding to (24), ensures that off-diagonal elements are zero.
- (A28) —  $D-1$  in total, corresponding to (24), ensures that the diagonal elements are equal.
- (A29) —  $N_s(N_s - 1)/2$  in total, corresponding to (25).
- (A30) —  $(n-1)(n+2)/2$  in total, corresponding to (54), optionally used instead of (A27) and (A28).

The combined set of linear constraints can be put to (A4) with  $N_d = (D - 1)(D + 2)/2 + N_s(N_s - 1)/2$  to restrict<sup>6</sup> the variation subspace of  $b_{s,jk}$ . All  $C_{d;s,jk}$  are calculated from  $b_{s,jk}$ , similar to the calculation of Lagrange multipliers in Appendix A4. Our algorithm uses, as an iteration state, not a pair: approximation, Lagrange multipliers  $(b_{s,jk}, \{\lambda_{ij}, \nu_{ss'}\})$ , but a triple: approximation, Lagrange multipliers, homogeneous linear constraints:  $(b_{s,jk}, \{\lambda_{ij}, \nu_{ss'}\}, C_{d;s,jk})$ ; it is the linear constraints themselves that ensure algorithm convergence — current work and previous [11] results show their critical importance.

## Appendix B: A Time-Dependent Schrödinger-like Equation

A formulated novel algebraic problem (43) is a generalization of the eigenvalue problem (time-independent Schrödinger equation). A question arises regarding the generalization of (43) to a time-dependent form. Consider the equation

$$i\hbar \frac{\partial \mathcal{U}}{\partial t} = S\mathcal{U} \quad (\text{B1})$$

where superoperator  $S$  is a Hermitian tensor  $S_{jk;j'k'}$  and  $\mathcal{U}$  is a unitary operator  $u_{jk}$  where  $D = n$ . This equation is different from the dynamic equation for the density matrix

$$i\hbar \frac{\partial \rho}{\partial t} = \mathcal{L}\rho \quad (\text{B2})$$

where the Liouville operator  $\mathcal{L}\rho = H\rho - \rho H$ . This equation has the Hermitian density matrix  $\rho$  as its solution. The distinction arises from the fact that the operators  $\mathcal{L}$  and  $S$  have completely different structures, and the solution  $\mathcal{U}$  is now a unitary operator, not a Hermitian matrix  $\rho$ . We can, however, employ a transition from the wavefunction Schrödinger equation to density matrix **Liouville – von Neumann** equation (B2) to construct an equation for the density supermatrix  $\Upsilon$ , which represents the density tensor. The density matrix is a convex combination of pure states (Eq. 39); this results in a transition from a vector  $|\psi\rangle$  of dimension  $n$  to a Hermitian matrix  $\rho$  of dimension  $n \times n$ . Similarly, consider a convex combination of unitary channels

$$\Upsilon = \sum_{s=0}^{Dn-1} P^{[s]} |\mathcal{U}^{[s]}\rangle \langle \mathcal{U}^{[s]}| \quad (\text{B3})$$

---

<sup>6</sup> From the  $n(n + 1)/2$  trace preservation constraints (54), we obtained  $(n - 1)(n + 2)/2$  linear constraints (A30). These can be used instead of (A27) and (A28). For  $n = D$  both sets are equivalent. When, for some reason, they are used together, it is possible to have redundant constraints (e.g., in the  $D = n$  unitary case where  $N_s = 1$ ). In such cases Gaussian elimination retains only the rank  $C_{d;s,jk}$  of them.

This results in a transition from a  $D \times n$  unitary matrix  $\mathcal{U}$  (where  $D = n$ ) to a  $Dn \times Dn$  density tensor  $\Upsilon$ , which has the same structure as the Hermitian tensor  $S_{jk;j'k'}$ . A general quantum channel is described by a density tensor  $\Upsilon$ . Then, similarly to Eq. (B2), we can assume that  $\Upsilon$  satisfies its own Liouville equation

$$i\hbar \frac{\partial \Upsilon}{\partial t} = \mathcal{L}\Upsilon \quad (\text{B4})$$

Equation (B4) is analogous to Eq. (B2), but describes dynamics where a unitary operator  $\mathcal{U}$  replaces the vector wavefunction  $|\psi\rangle$ , and the density tensor  $\Upsilon$  replaces the density matrix  $\rho$ . The specific form of the operator  $\mathcal{L}$  requires further research; the first form to consider is, evidently,

$$\mathcal{L}\Upsilon = S\Upsilon - \Upsilon S \quad (\text{B5})$$

The commutator is possible because  $S$  and  $\Upsilon$  have the same tensor structure. This form, however, differs from regular density matrix dynamics. For a pure superstate  $\Upsilon = |\mathcal{U}\rangle \langle \mathcal{U}|$ , where  $\mathcal{U}$  is the solution of (43),  $\mathcal{L}\Upsilon$  from (B5) is not zero:  $(\mathcal{L}\Upsilon)_{ik;jq} = \sum_{j'=0}^{D-1} \lambda_{ij'} u_{j'k} u_{jq}^* - \sum_{j'=0}^{D-1} u_{ik} u_{j'q}^* \lambda_{jj'}^*$ . This issue does not arise in regular density matrix dynamics, where for a Hamiltonian eigenvector  $|\psi\rangle$ , the density matrix  $\rho = |\psi\rangle \langle \psi|$  satisfies  $0 = H\rho - \rho H$ .

If the tensor  $\Upsilon_{jk;j'k'}$  is known only in matrix form, the expansion (B3) can be obtained by applying the hierarchy construction from Section IV to  $\Upsilon_{jk;j'k'}$ . Note that various orthogonality constraints can be applied to the operators  $\mathcal{U}^{[s]}$  in (B3). For example, the constraint (44b) can be used instead of (45), which was previously employed in Section IV. The operators  $\mathcal{U}^{[s]}$  in (B3) then satisfy the “denominator”-type orthogonality condition (44b)

$$D\delta_{ss'} = \left\langle \mathcal{U}^{[s]} \left| \mathcal{U}^{[s']} \right. \right\rangle \quad (\text{B6})$$

We will defer the study of the equation for the density tensor  $\Upsilon$  to future research and focus here on a simple example of the ground state time evolution of Eq. (B1). This equation is a generalization of the Schrödinger equation from wavefunction vector space to the space of unitary operators. The Schrödinger equation itself can be derived from Brownian motion, Fisher information, the Hamilton-Jacobi equation, and various other approaches to direct problems (see references [21-34] in [42]). Our equation (B1) could likely be derived from various inverse problems.

Consider a ground state stationary solution of (B1) corresponding to  $\mathcal{U}^{[0]}$ , i.e. obtained by maximizing the fidelity  $\mathcal{F}$  without using the constraints from (45). This solution, represented

by the matrix  $u_{jk}^{[0]}$ , satisfies (43) with some  $\lambda_{ij}^{[0]}$ . The matrix  $\lambda_{ij}^{[0]}$  is Hermitian but not necessarily diagonal. Let us convert the problem to the basis in which  $\lambda_{ij}^{[0]}$  is diagonal. Consider the basis of  $\lambda_{ij}$  eigenvectors

$$\sum_{j=0}^{D-1} \lambda_{ij} \beta_j^{[p]} = \lambda_i^{[p]} \beta_i^{[p]} \quad (\text{B7})$$

Then, if we formulate the original problem (43) in basis

$$v_{pk} = \sum_{j=0}^{D-1} \beta_j^{[p]} u_{jk} \quad (\text{B8})$$

the obtained solution is  $v_{pk}^{[0]}$ , and the corresponding  $\lambda_{pp'}$  is diagonal with the diagonal elements equal to the eigenvalues  $\lambda^{[p]}$  in (B7). We considered a similar basis transformation in [10] (see Appendix A.2 therein, “On Iteration Step Without Using the SVD”), where  $S_{jk;j'k'}$  and  $u_{jk}$  are converted between bases for the purpose of improving the convergence of an algorithm. However, later [11] we found an algorithm that ensures convergence in any basis. Here, this transformation is performed solely to obtain the Lagrange multipliers  $\lambda_{ij}$  (A26) in diagonal form, which is necessary to find a stationary state solution of the time-dependent equation (B1). Thus, without loss of generality, the  $\lambda_{ij}$  can be considered diagonal; otherwise, the original problem basis should be changed to (B8).

Let us also rewrite (B1) by replacing  $S$  with  $\lambda_{jj'} \delta_{kk'}$ , and referring to this  $S$  as the single Hamiltonian approximation.

$$S_{jk;j'k'} \approx \lambda_{jj'} \delta_{kk'} \quad (\text{B9})$$

In this approximation, a general  $S$ , which encompasses many different Hamiltonians, is replaced by a form containing only a single Hamiltonian  $\lambda$ , typically corresponding to the ground state. The maximal fidelity  $\mathcal{F} = \langle \mathcal{U} | \lambda | \mathcal{U} \rangle = \text{Tr} \lambda$  then holds for an arbitrary unitary operator  $\mathcal{U}$ , not just for the ground state solution of the equation (43). By focusing solely on the ground state quantum channel the dynamic equation (B1) becomes

$$i\hbar \frac{\partial \mathcal{U}}{\partial t} = \lambda \mathcal{U} \quad (\text{B10})$$

This equation is akin to a Schrödinger equation with the Hamiltonian  $\lambda$ , where all possible solutions are encompassed in  $\mathcal{U}$ . The equation describes the simultaneous time evolution of all possible solutions. The full superoperator  $S_{jk;j'k'}$  contains many Hamiltonians. For each



solution of (43), the Lagrange multiplier matrix  $\lambda$  can be considered as a Hamiltonian of some quantum system. For  $u_{pk}$  being the ground state solution  $u_{pk}^{[0]}$  of (43) and  $\lambda_{pj}$  being diagonal  $\lambda^{[p]}\delta_{pj}$ , the time-dependent solution of it is

$$u_{pk}(t) = \exp\left(-\frac{i}{\hbar}\lambda^{[p]}t\right)u_{pk}(t=0) \quad (\text{B11})$$

This can be viewed as  $u_{p*}$  vectors evolving each with its own phase  $\exp(-it\lambda^{[p]}/\hbar)$ . If we consider the operator  $u_{pk}$  as  $D$  “wavefunctions”  $u_{p*}$  of dimension  $n$ , then each of them evolves with its own “energy”  $\lambda^{[p]}$ . This represents the time evolution of the ground state solution of (43). The difference from traditional quantum mechanics is that the ground state  $u_{pk}$  now contains multiple vectors, each evolving with its own exponent  $\lambda^{[p]}$ . The density tensor corresponding to the ground state is

$$\begin{aligned} (|\mathcal{U}\rangle\langle\mathcal{U}|)_{pk;p'k'}(t) = \\ \exp\left(-\frac{i}{\hbar}\left(\lambda^{[p]} - \lambda^{[p']}\right)t\right)u_{pk}(t=0)u_{p'k'}^*(t=0) \end{aligned} \quad (\text{B12})$$

Unlike the density matrix of a traditional quantum system’s ground state, which does not depend on time, this density tensor has only the diagonal elements,  $p = p'$ , that are independent of time. This explains why (B5) is nonzero for a pure superstate  $\Upsilon = |\mathcal{U}\rangle\langle\mathcal{U}|$ , where  $\mathcal{U}$  is the solution to (43).

A **heat transfer**-like equation can be obtained by removing the imaginary unit  $i$  from (B1), resulting in  $\varkappa\partial\mathcal{U}/\partial t = S\mathcal{U}$ . In this case, the solution (B11) results in an exponentially growing or decaying  $u_{pk}(t)$ .

The time evolution of a superposition of two solutions is more complicated compared to the Schrödinger equation. For the Schrödinger equation, a superposition of two eigenstates  $a\psi^{[0]} + b\psi^{[1]}$  of a Hamiltonian time-evolves as  $a\psi^{[0]}\exp(-itE_0/\hbar) + b\psi^{[1]}\exp(-itE_1/\hbar)$ . At the same time, if we consider a superposition of two solutions  $au_{jk}^{[0]} + bu_{jk}^{[1]}$  of (43), the result may not be unitary. The violation of unitarity resulting from such a superposition has deep physical significance. Such a superposition does not describe a physical state and, therefore, should not be considered. A proper generalization is the introduction of the density supermatrix  $\Upsilon$  (B3), which describes a mixed type of state.

Another feature that may differ between wavefunctions and quantum channels is the state post-measurement destruction rules. In traditional quantum mechanics, the state  $|\psi\rangle$  is destroyed (at least partially) after the measurement act  $\langle\psi|R|\psi\rangle$ . In contrast, the

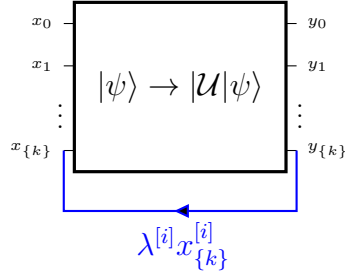
measurement  $\langle \mathcal{U} | R | \mathcal{U} \rangle$  involves a quantum channel  $\mathcal{U}$  as the state. The post-measurement destruction rules for a quantum channel  $\mathcal{U}$  remain an open question. It is possible that these rules differ between wavefunctions and quantum channels. A very important question is what the measurable unit is in quantum channel dynamics: whether it is a scalar or a Hermitian matrix? Since the corresponding algebraic problem (43) has eigenmatrices as its spectrum, rather than the usual eigenvalues (scalars), we are inclined to believe that the measurement unit is a Hermitian matrix. As the ground state energy of a quantum system can be obtained as a result of a single measurement act, a quantum channel single measurement could potentially provide the Hamiltonian  $\lambda^{[0]}$  corresponding to the solution of (43) with the maximal fidelity.

For these reasons, we will leave a detailed consideration of the time-dependent equations (B1) and (B4) for future research and limit ourselves to the time evolution (B11) of the ground state, which corresponds to each  $u_{p*}$  vector evolving with its own “energy”  $\lambda^{[p]}$ . The total fidelity is equal to the sum of all  $\lambda^{[p]}$ , and the ground state is the state with the maximum total fidelity. The ground state solution corresponds to a unitary operator  $\mathcal{U}(t)$  that itself depends on time (B11).

### Appendix C: Parametrizing states with memory using a feedback loop

In this and previous works, we considered a **memoryless quantum channel** described by a set of Kraus operators  $B_s$  or, in the simplest case where  $N_s = 1$ , by a single unitary operator  $\mathcal{U}$ . A question arises about how memory can be introduced into the model. Modeling a system with even a single bit of internal state is a daunting task, see Appendix H of [20], where we discuss the problem of modeling a **synchronous positive-edge-triggered D flip-flop** (D trigger) with ML. Any memory can be represented by a feedback loop, as seen in, for example, a D-trigger or a **recurrent neural network**. The question is how to introduce memory into the quantum channel (3) described by an operator  $\mathcal{U}$ ? For pure states, the channel transforms an input state  $|\psi\rangle$  into the output state  $|\mathcal{U}|\psi\rangle$ . Let’s assume there is a feedback loop that requires some (but not all) components of the output state to be equal (within a phase) to

the same components of the input state.



$$\psi_{\{k\}}^{[i]} = \lambda^{[i]} \sum_{k=0}^{n-1} u_{\{k\}k} \psi_k^{[i]} \quad (\text{C1})$$

Here the  $\{k\}$  denotes these selected “feedbacked” components. If  $\{k\}$  is an empty set, we obtain the problem we considered in [11]. If  $\{k\}$  includes *all*  $n$  components — then (C1) becomes a regular eigenproblem with  $\lambda$  being  $\exp(i\zeta)$  (or  $\pm 1$  for an orthogonal  $u_{jk}$ ). In this case the problem does not have any input and the internal state can be represented as a superposition of  $n$  eigenstates of eigenproblem (C1). Now consider  $\{k\}$  to be a subset of  $0 \dots n-1$  indices with a total of  $n_m$  elements. Then the problem (C1) represents an eigenvalue-like problem of dimension  $n_m$ ,  $i = 0 \dots n_m - 1$ , where an “eigenvector”  $|\psi^{[i]}\rangle$  (there are  $n_m$  total) has  $n - n_m$  components set externally and  $n_m$  components being an internal state. This “partial eigenvalue” (C1) problem is a simple way to represent a quantum channel with an internal state (memory). A given state  $|\psi\rangle$  has  $n - n_m$  free components (input) and  $n_m$  components of the internal state. A change in the input components can potentially cause changes in the internal state. The study of unitary quantum channels with internal states is a subject for our future research.

A different application of these “feedbacked” states of a unitary operator relates to the problem of integer factorization. An integer can be encoded into a state  $|\psi\rangle$ , and the multiplication of the integer modulo  $N$  by a number  $a$  can be represented as a unitary transformation  $\mathcal{U}$ . In Shor’s algorithm [43], a major step in the prime factorization of a number  $N$  involves finding the period  $r$  of

$$a^r \equiv 1 \bmod N \quad (\text{C2})$$

If multiplication by  $a$  is encoded as a unitary transformation  $\mathcal{U}$  acting on the state  $|\psi\rangle$ , then the problem (C2) can be framed as finding the  $r$  such that  $|\mathcal{U}^r|\psi_0\rangle$  equals the initial state  $|\psi_0\rangle$  up to a phase factor. This is similar to the concept of the feedback loop shown in (C1).

For example, if we were able to build an actual quantum system with the Hamiltonian

$$H = i\frac{\hbar}{\tau} \ln \mathcal{U} \quad (\text{C3})$$

and initial state  $|\psi^{(t=0)}\rangle = |\psi_0\rangle$  then, since the system's time evolution

$$U = \exp \left[ -i\frac{t}{\hbar} H \right] \quad (\text{C4})$$

$$|\psi^{(t)}\rangle = |U|\psi^{(t=0)}\rangle \quad (\text{C5})$$

the calculation of  $r$  can be reduced to simply observing the quantum system (C3) with  $\mathcal{U} = U(t = \tau)$  at time moments  $t = \tau l$  and waiting until  $|\psi^{(t=\tau l)}\rangle$  matches  $|\psi_0\rangle$  up to a phase. The index  $l$  of this time moment gives the period  $r$ . For this quantum system, the “Boltzmann time” of spontaneous return to the exact initial configuration provides the sought period  $r$ .

#### Appendix D: An estimate of computational complexity

The described class of algorithms solves QCQP problems that arise in the maximization of quadratic fidelity, subject to quadratic constraints on mapping operators (e.g., unitarity). This problem is equivalent to an algebraic problem; for example, in the case of unitary learning, it corresponds to the algebraic problem (43) originally introduced in [20]. This represents the simplest problem of this type. In the general case, this is a new algebraic problem of dimension  $N_s D n$ . It can be solved numerically using an iterative algorithm that, on each iteration, replaces  $N_c$  quadratic constraints with  $N_c - 1$  homogeneous linear constraints and a single quadratic constraint (26), referred to as a simplified (partial) constraint, see Appendix A 5 above. Since a QCQP problem with a single quadratic constraint is equivalent to an eigenvalue problem, a regular eigenvalue problem solver is applied to a problem of dimension  $N_V = N_s D n - N_c + 1$ , this is step 1 of the algorithm. As discussed in Appendix B of [11], this is the most computationally intensive step. Its computational complexity in the unitary learning case ( $D = n$ ,  $N_s = 1$ ) involves solving an eigenproblem with a matrix of dimension  $N_V = n^2 - n(n+1)/2 + 1$ . This complexity can be estimated as  $O(n^4)$  when using a specialized solver optimized to find only the maximal eigenvalue and  $O(n^6)$  when using a general-purpose solver. Our current implementation [19] uses a general-purpose eigenvalue solver.

Let us compare the computational complexity of this “algebraic” algorithm with that of the “mathematical analysis” ones: second-order Newton’s method and first-order gradient method. A unitary matrix of dimension  $n$  can be completely parametrized with  $N = n(n + 1)/2$  parameters. There are  $n$  fewer parameters if we do not need unmeasurable phases. The number of parameters is always greater than or equal to  $n(n - 1)/2$  (e.g., Euler angles), see (B11), which gives a time-dependent ground state unitary operator  $\mathcal{U}(t)$  such that, at every time moment  $t$ , the matrix  $\mathcal{U}$  gives the maximal value of fidelity  $\mathcal{F}$  (27). The complexity of the classical Newton’s method for nonlinear systems of  $N$  equations is  $O(N_{it}N^3)$ , as a linear system needs to be solved at each iteration. Some tricks, such as updating the Jacobian only once every  $m$  iterations, can reduce the complexity to  $O(N_{it}N^2 + N_{it}N^3/m)$ . Since the number of iterations,  $N_{it}$ , for Newton’s method is often independent of  $N$ , the practical complexity of the Newton’s method algorithm can be estimated to lie between  $O(n^4)$  and  $O(n^6)$ ,<sup>7</sup> depending on the Jacobian updating strategy. While the computational complexity of the proposed algebraic algorithm is comparable to that of the second-order Newton’s method, using the generalized eigenvalue problem as the algorithm’s building block offers the advantage of obtaining multiple solution candidates (eigenvectors). Numerical experiments[11] demonstrate that this significantly increases the chances of finding the global maximum.

The computational complexity of the gradient method in the general case is  $O(N_{it}MN)$ . If the  $M$ -sum can be factored out from the objective functional (e.g., into a form like (27)), the complexity reduces to  $O(N_{it}N^2) = O(N_{it}n^4)$ . However, in the gradient method, the number of iterations  $N_{it}$  is not guaranteed to be small, nor is the algorithm guaranteed to converge, especially for a global optimization problem that is nonconvex, with local extrema and multiple saddle points.

The developed algorithm exhibits computational complexity on par with second-order Newton’s method and higher than that of the first-order gradient method. Its main feature is the algebraic approach, which enables the construction of a globally converging algorithm. While we lack a formal proof of its convergence, among millions of test runs, only a few failed to converge to the global maximum. In this work, we aimed to address a practical problem that can be validated using a classical computer. Consider a scenario where we have input data (e.g., in the form of files on disk), and we seek to recover the corresponding quantum mapping.

<sup>7</sup> If all  $M$  observations are not combined into a single entity, such as  $S_{jk;j'k'}$ , then the complexity may range between  $O(Mn^4)$  and  $O(Mn^6)$ , which significantly worsens the result.

This problem frequently arises in Machine Learning for large dimensions, as discussed in [2] and over 1,000 subsequent papers citing it. In the notations of the submitted paper, typical dimensions of the ML problem are  $N_s = 1$  and  $D = n \sim 10^3$  or greater. The solution involves a form of unitary parametrization[5], such as representing the unitary as a series of elementary rotations (e.g., Euler angles) and employing a gradient-based optimization method. The resulting solution is obtained through this process, but there is no guarantee that it corresponds to the global maximum. Our primary focus was on developing an algorithm capable of providing the *global* maximum solution. For this reason we considered an exact unitary mapping and aimed to recover it from a sample, this is why we did not include noise or other common features in our study. This is a highly challenging nonconvex problem, as the number of parameters grows quadratically with  $n$  [44]. To the best of our knowledge, practical solutions are feasible only for small  $n$  (fewer than a dozen) or under specific setups [14, 18, 38]. In our numerical experiments for unitary learning, the problem’s dimensionality appears to be constrained solely by computational complexity. At each iteration step, an eigenvalue problem of dimension  $N_V = n^2 - n(n + 1)/2 + 1$  must be solved, with only the eigenstate corresponding to the largest eigenvalue being required. Given the extensive global effort devoted to developing efficient numerical solvers for eigenproblems, we anticipate that the dimensionality of the problem can be significantly increased. Currently, however, we lack access to hardware capable of solving eigenproblems for matrices larger than dimension 1000. Our original goal was to push the problem’s dimensionality to a point where exact recovery of orthogonal (real unitary) matrices would become infeasible. However, computational limits were encountered first. In our experiments, we tested the exact recovery of thousands of randomly generated orthogonal matrices with dimensions under 50 (corresponding to an eigenproblem size of 1226). The ability to find the global maximum remains the algorithm’s most significant feature.

The computational complexity of constructing the tensor  $S_{jk;j'k'}$  (19) was insignificant for the pure state unitary mapping (1) discussed in [11]. However, for the unitary mapping (3) of mixed states (7) considered in the present work, obtaining the fidelity in quadratic form requires transforming the mapping to a density matrix square root mapping (29). This transformation may significantly increase the computational complexity of calculating  $S_{jk;j'k'}$  (30). Using an eigenvalue problem to calculate the square root of a matrix would require solving  $M$  eigenproblems with the matrices  $\rho$  and  $\varrho$ , each of dimension  $n$ , resulting

in a complexity of the order  $O(Mn^3)$ . This cannot be reduced as in the QCQP case above, where only a single eigenvector was needed. For the density matrix square root, we need all eigenvectors, and the calculation of (30) probably cannot be done any better than  $O(Mn^3)$ . The factor before that can, however, be reduced. When the number of observations  $M$  is large, instead of using the regular method of calculating the matrix square root by solving an eigenproblem (as in (A15) and (A22) above), one can apply alternative methods to compute the square root of a positively definite Hermitian matrix without solving an eigenproblem [45, 46]. These approaches, like the eigenvalue problem, have a dominant complexity of  $O(n^3)$  for the square root computation. However, in practice, it can significantly improve computational efficiency when creating the mapping (29) from the original data (7). Despite the complexity of calculating  $S_{jk;j'k'}$  for a large  $M$ , this creates no difficulty in applications since the calculations are similar to those used, for example, in covariance matrix computation. Each component of the tensor is a sum over all  $M$  observations, and this task can be trivially parallelized.

The general Kraus case,  $N_s > 1$ , is significantly more difficult, since the fidelity (16), as a quadratic form, can only be obtained as an approximation. However, once the approximation is obtained, in the general case, it requires solving an eigenproblem of dimension  $N_V = N_s Dn - N_c + 1$  at each iteration (step 1 of the algorithm), with only a single eigenvector corresponding to the maximal eigenvalue being needed. For a general quantum channel (4), we were not able to construct a converging algorithm, not to mention the increased computational complexity (eigenvalue problem dimension is increased in Kraus rank times). However, if we limit ourselves to mixed-unitary channels (36), then this type of quantum channel can be reconstructed as a hierarchy of unitary operators. Computationally, the problem in Section IV is much simpler than a general quantum channel reconstruction. Instead of dealing with a single complicated general quantum channel problem, the problem of unitary mapping needs to be solved several times. Numerically, we were able to obtain a hierarchy of no more than 7 operators, which is probably due to the suboptimal algorithm for constraint conflict resolution used in Appendix A 2. The result is a mixed unitary channel (36). However, contrary to the  $N_s = 1$  case, for  $N_s > 1$ , this is only an approximation, since we were not able to represent

the proper fidelity (11) as the quadratic form of Eq. (16) in Kraus operators  $B_s$  when  $N_s > 1$ .

- 
- [1] A. Bisio, G. Chiribella, G. M. D’Ariano, S. Facchini, and P. Perinotti, Optimal quantum learning of a unitary transformation, *Physical Review A* **81**, 032324 (2010).
  - [2] M. Arjovsky, A. Shah, and Y. Bengio, Unitary evolution recurrent neural networks, in *International conference on machine learning, NY, USA, 2016* (Proceedings of Machine Learning Research (proceedings.mlr.press), 2016) pp. 1120–1128.
  - [3] S. Hyland and G. Rätsch, Learning unitary operators with help from  $u(n)$ , in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 31 (Association for the Advancement of Artificial Intelligence (aaai.org), 2017).
  - [4] S. Lloyd and R. Maity, Efficient implementation of unitary transformations, arXiv preprint arXiv:1901.03431 [10.48550/arXiv.1901.03431](https://arxiv.org/abs/10.48550/arXiv.1901.03431) (2019).
  - [5] A. Raza, *Learning unitary matrices* (2020), <https://araza6.github.io/posts/unitary-learning/>.
  - [6] S. Pai, B. Bartlett, O. Solgaard, and D. A. B. Miller, Matrix optimization on universal unitary photonic devices, *Physical review applied* **11**, 064044 (2019).
  - [7] Y.-M. Huang, X.-Y. Li, Y.-X. Zhu, H. Lei, Q.-S. Zhu, and S. Yang, Learning Unitary Transformation by Quantum Machine Learning Model., *Computers, Materials & Continua* **68**, [10.32604/cmc.2021.016663](https://doi.org/10.32604/cmc.2021.016663) (2021).
  - [8] F. A. González, A. Gallego, S. Toledo-Cortés, and V. Vargas-Calderón, Learning with density matrices and random features, *Quantum Machine Intelligence* **4**, 23 (2022).
  - [9] Z. Yu, X. Zhao, B. Zhao, and X. Wang, Optimal quantum dataset for learning a unitary transformation, *Physical Review Applied* **19**, 034017 (2023).
  - [10] V. G. Malyskin, On Machine Learning Knowledge Representation In The Form Of Partially Unitary Operator. Knowledge Generalizing Operator, arXiv preprint arXiv:2212.14810 [10.48550/arXiv.2212.14810](https://arxiv.org/abs/10.48550/arXiv.2212.14810) (2022).
  - [11] M. G. Belov and V. G. Malyskin, Partially unitary learning, *Phys. Rev. E* **110**, 055306 (2024).
  - [12] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, *et al.*, Variational quantum algorithms, *Nature Reviews Physics* **3**, 625 (2021).



- [13] C.-Y. Park and N. Killoran, Hamiltonian variational ansatz without barren plateaus, *Quantum* **8**, 1239 (2024).
- [14] X. Wang, C. Yang, and M. Gu, Variational Quantum Circuit Decoupling, *Phys. Rev. Lett.* **133**, 130602 (2024).
- [15] G. M. D'Ariano, M. G. A. Paris, and M. F. Sacchi, Quantum Tomography (Elsevier, 2003) pp. 205–308, *Advances in Imaging and Electron Physics*, Vol 128.
- [16] A. I. Lvovsky and M. G. Raymer, Continuous-variable optical quantum-state tomography, *Reviews of modern physics* **81**, 299 (2009).
- [17] G. Torlai, C. J. Wood, A. Acharya, G. Carleo, J. Carrasquilla, and L. Aolita, Quantum process tomography with unsupervised learning and tensor networks, *Nature Communications* **14**, 2858 (2023).
- [18] S. Ahmed, F. Quijandría, and A. F. Kockum, Gradient-descent quantum process tomography by learning Kraus operators, *Physical Review Letters* **130**, 150402 (2023).
- [19] V. G. Malyshkin, *The code for polynomials calculation* (2014), <http://www.ioffe.ru/LNEPS/malyshkin/code.html> and an *alternative location*.
- [20] V. G. Malyshkin, On The Radon-Nikodym Spectral Approach With Optimal Clustering, arXiv preprint arXiv:1906.00460 [10.48550/arXiv.1906.00460](https://arxiv.org/abs/1906.00460) (2019).
- [21] A. Jamiołkowski, Linear transformations which preserve trace and positive semidefiniteness of operators, *Reports on mathematical physics* **3**, 275 (1972).
- [22] K. Kraus, *States, Effects, and Operations: Fundamental Notions of Quantum Theory*, Lecture Notes in Physics, Vol. 190 (Springer-Verlag, 1983) Lectures in Mathematical Physics at the University of Texas at Austin.
- [23] E. Joos, H. D. Zeh, C. Kiefer, D. J. Giulini, J. Kupsch, and I.-O. Stamatescu, *Decoherence and the appearance of a classical world in quantum theory* (Springer Science & Business Media, 2013).
- [24] M.-D. Choi, Completely positive linear maps on complex matrices, *Linear algebra and its applications* **10**, 285 (1975).
- [25] V. P. Belavkin and P. Staszewski, A Radon-Nikodym theorem for completely positive maps, *Reports on mathematical physics* **24**, 49 (1986).
- [26] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information* (Cambridge university press, 2010).

- [27] M. M. Wilde, From classical to quantum Shannon theory, arXiv preprint arXiv:1106.1445 [10.48550/arXiv.1106.1445](#) (2011).
- [28] A. Jenčová, Recoverability of quantum channels via hypothesis testing, [Letters in Mathematical Physics](#) **114**, 31 (2024).
- [29] L. B. Levitin, Conditional entropy and information in quantum systems, [Chaos, Solitons & Fractals](#) **10**, 1651 (1999).
- [30] A. Kendall, Y. Gal, and R. Cipolla, Multi-task learning using uncertainty to weigh losses for scene geometry and semantics, in [Proceedings of the IEEE conference on computer vision and pattern recognition](#) (2018) pp. 7482–7491.
- [31] O. Sener and V. Koltun, Multi-task learning as multi-objective optimization, Advances in neural information processing systems **31**, [10.48550/arXiv.1810.04650](#) (2018).
- [32] V. Kurin, A. De Palma, I. Kostrikov, S. Whiteson, and P. K. Mudigonda, In defense of the unitary scalarization for deep multi-task learning, [Advances in Neural Information Processing Systems](#) **35**, 12169 (2022).
- [33] R. Duvenhage, Optimal quantum channels, [Physical Review A](#) **104**, 032604 (2021).
- [34] M. Arjovsky, Saddle-free Hessian-free optimization, arXiv preprint arXiv:1506.00059 [10.48550/arXiv.1506.00059](#) (2015).
- [35] T. Van Erven and P. Harremos, Rényi divergence and Kullback-Leibler divergence, [IEEE Transactions on Information Theory](#) **60**, 3797 (2014).
- [36] N. Galke, L. Van Luijk, and H. Wilming, Sufficiency of Rényi divergences, [IEEE Transactions on Information Theory](#) [10.1109/TIT.2024.3376395](#) (2024).
- [37] P. Lipka-Bartosik, H. Wilming, and N. H. Ng, Catalysis in quantum information theory, [Reviews of Modern Physics](#) **96**, 025005 (2024).
- [38] K. Schäfers, M. Peardon, and M. Günther, A modified Cayley transform for SU(3), arXiv preprint arXiv:2406.11337 [10.48550/arXiv.2406.11337](#) (2024).
- [39] C. D.-Y. Lee and J. Watrous, Detecting mixed-unitary quantum channels is NP-hard, [Quantum](#) **4**, 253 (2020).
- [40] M. Girard, D. Leung, J. Levick, C.-K. Li, V. Paulsen, Y. T. Poon, and J. Watrous, On the mixed-unitary rank of quantum channels, [Communications in Mathematical Physics](#) **394**, 919 (2022).
- [41] G. H. Golub, Some modified matrix eigenvalue problems, [Siam Review](#) **15**, 318 (1973).

- [42] S. Efthimiades, Derivation of the Schrödinger equation from QED, *European Journal of Physics* **45**, 035401 (2024).
- [43] P. W. Shor, Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer, *SIAM review* **41**, 303 (1999).
- [44] E. Andersson, J. D. Cresser, and M. J. Hall, Finding the Kraus decomposition from a master equation and vice versa, *Journal of Modern Optics* **54**, 1695 (2007).
- [45] Å. Björck, *Numerical methods for least squares problems* (SIAM, 2024).
- [46] N. J. Higham, Newton's method for the matrix square root, *Mathematics of computation* **46**, 537 (1986).