CYBERNETICS AND PHYSICS, VOL. 12, NO. 2, 2023, 152–156

TRAINABLE UNRAVELLING FOR QUANTUM STATE DISCRIMINATION

V.A. Tomilin
Institute of Automation and Electrometry SB RAS
630090, Novosibirsk, Russia
8342tomilin@mail.ru

L.V. Il'ichov
Institute of Automation and Electrometry SB RAS
630090, Novosibirsk, Russia
leonid@iae.nsk.su

Article history:
Received 02.07.2023, Accepted 15.09.2023

Abstract

We propose a way for partial compensation of maladaption of decoder to non-ideal quantum communication channel by means of optimal choice of unravelling of the decoding operation. No physical modification of the decoder itself is required. We show that it is sufficient to add an interface that modifies the decoder-environment interaction. Tuning of this interface can be done by methods of (quantum-inspired) machine learning. We suggest a search algorithm for an optimal unravelling, as an alternative for the classical gradient descent method.

Key words

Quantum state discrimination, quantum-inspired machine learning

1 Introduction

One of the greatest challenges in communication theory is the decoding of the distorted transmitted message. It is reduced to the most accurate recognition of the signal corresponding to each symbol of the alphabet used. For quantum communications, the task is to discriminate between non-orthogonal and generally mixed states from the set $\{\hat{\varrho}(i)\}_{i\in I}$, indexed by the items of the alphabet I. In the simplest case of binary alphabet the message is a sequence of symbols $'0'$ and $'1'$. We assume that each symbol is encoded during transmission by one of pure orthogonal states, $|0\rangle$ and $|1\rangle$. Because of the non-ideal transmission through the quantum channel (due to its coupling to the environment), the states are transformed

$$|i\rangle\langle i| \mapsto \hat{\varrho}(i) = \Lambda[|i\rangle\langle i|] \quad i \in \{0,1\} \tag{1}$$

where Λ is a completely positive trace-preserving (CPTP) map [Nielsen and Chuang, 2000] reflecting the transmission properties of the channel.

The decoding device should be able to distinguish the states $\hat{\varrho}(0)$ and $\hat{\varrho}(1)$ with minimal error, i.e. to most faithfully reproduce the optimal measurement that discriminates the states most accurately (but in general this discrimination is still non-deterministic) [Helstrom, 1979; Fuchs, 1996]. The decoder may be represented by another CPTP map $\mathcal{E}_{dec}$:

$$\begin{aligned} &\mathcal{E}_{dec}[\hat{\varrho}] = \hat{E}_0\hat{\varrho}\hat{E}_0^\dagger + \hat{E}_1\hat{\varrho}\hat{E}_1^\dagger, \\ &\hat{E}_0^\dagger\hat{E}_0 + \hat{E}_1^\dagger\hat{E}_1 = \hat{1}. \end{aligned} \tag{2}$$

Here indices '0' and '1' correspond to different measurement outcomes. Operators $\hat{E}_0$ and $\hat{E}_1$ are known as Kraus operators [Kraus, 1983], and the probabilities of different decoding results are given by $Tr(\hat{E}_i^\dagger\hat{E}_i\hat{\varrho})$.

To minimize the decoding error, the Kraus operators $\hat{E}_0$ and $\hat{E}_1$ should be close to the projectors of the aforementioned optimal measurement aimed for discrimination of $\hat{\varrho}(0)$ and $\hat{\varrho}(1)$. The latter appear as a result of Λ operation; therefore, $\hat{E}_0$ and $\hat{E}_1$ are bound to it. Consider a situation of the properties of the channel (and hence the operation Λ) are altered for any reason: $\Lambda \mapsto \Lambda'$. Then the decoding based on $\hat{E}_0$ and $\hat{E}_1$ might become quite different from the one aimed for discrimination of $\hat{\varrho}'(0) = \Lambda'[|0\rangle\langle 0|]$ and $\hat{\varrho}'(1) = \Lambda'[|1\rangle\langle 1|]$. Implementation of the necessary change, $\mathcal{E}_{dec} \mapsto \mathcal{E}'_{dec}$, might be a challenging task. It is also worth noting that finding an optimal decoding for arbitrary mixed states is already a significant challenge and remains an active area of research [Sych, 2016; Rosati, 2017; Weir, 2017; DiMario, 2018]. Therefore a strategy that does not involve rebuilding of the decoder is desired.

Another situation that also arises in quantum communication scenarios is fundamentally incomplete knowledge of the transmission channel from the very beginning (e.g. due to noise). In this case, discrimination should be performed between the elements of an unknown set. For some classes of channels it is possible to use the same measurement that discriminates between initial input states [Kechrimparis, 2019]; it is also possible to specifically choose the input states encoding to a set that is invariant under action of a specific channel or a class of channels (so-called decoherence-free subspace [Lidar, 1997; Ticozzi, 2007]). However, generally a quantum process tomography is required [Chuang, 1997; Poyatos, 1997].

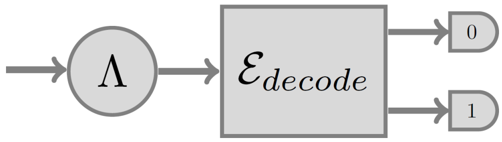

Figure 1. The scheme of message transmission and decoding.

2 Formulation of the problem

In the present work we propose a strategy based on fixed structure of decoding, hence no change of $\mathcal{E}_{dec}$ (i.e. $\mathcal{E}_{dec} = \mathcal{E}'_{dec}$). Instead, the interpretation of information coming from the decoder to the environment may be changed. It requires a sort of 'interface' between the decoder and the the set of detectors in which the results of decoding are registered as physical events. Tuning the interface is much simpler than rebuilding the decoder. This tuning allows for optimal *unravelling* [Breslin, 1997; Il'ichov, 2013] of the decoding operation, i.e. the choice of a pair $\{\hat{E}'_0, \hat{E}'_1\}$ that yields the same $\mathcal{E}_{dec}$,

$$\mathcal{E}_{dec}[\hat{\varrho}] = \hat{E}'_0 \hat{\varrho} \hat{E}'^{\dagger}_0 + \hat{E}'_1 \hat{\varrho} \hat{E}'^{\dagger}_1, \tag{3}$$

and in the same time gives the minimal decoding error, among all other unravellings[1]. Clearly, all possible unravellings of $\mathcal{E}_{dec}$ are given by linear transforms of the initial pair:

$$\hat{E}'_i = \hat{E}_i(U) \doteq \sum_{j=0,1} U_{ij} \hat{E}_j, \tag{4}$$

where U_{ij} are the elements of 2×2 unitary matrix. It is sufficient to consider only unit-determinant matrices.

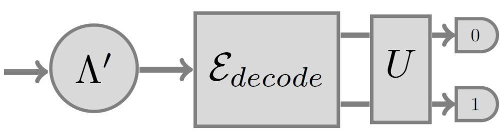

Figure 2. The scheme of the decoding with tunable unravelling, after the transmission properties of the channel are changed.

[1]Non-trivial applications of unravelling have been suggested in quantum states engineering [Il'ichov, 2003] and quantum feedback systems [Tomilin, 2022].

How to implement the arbitrary choice of unravelling is a separate problem. As a possible variant, the decoder signals the result by emitting a photon into one of two detectors (Fig. 1). The tunable beam-splitter placed between the decoder and the detectors implements an unravelling (Fig. 2). It is of utmost importance that the emitted photon is the only channel of transmitting the decoding results to the environment. The beam-splitter organizes controllable interference of two photon paths.

In essence, the decoder+interface system is a binary classifier that sorts the input signals into two categories. The search for optimal unravelling, i.e. the matrix $U^{(opt)}_{ij}$, can be done in the form of machine learning[2], similar to training a classical perceptron [Haykin, 1999]. The training data set is

$$\mathcal{D} = \{i_n, \hat{\varrho}_n\}_{n=1}^{N}, \tag{5}$$

where $i_n \in \{0, 1\}$ is the content of the signal, and $\hat{\varrho}_n = \hat{\varrho}(i_n)$ is a quantum state representing it[3].

The cost functional quantifying the training quality should include $\mathcal{D}$ and the unravelling matrix U. We choose Kullback-Leibler divergence (KL-divergence) as a measure of difference between the desired operation of the decoder and the real one. Other choices of divergence are possible, but KL-divergence is possibly the simplest one and, as will be shown later, greatly reduces the calculations needed. The following expression

$$S_{KL}(p||q) = \frac{1}{N} \sum_{n=1}^{N} \sum_{j=0,1} p(j|\hat{\varrho}_n) \ln \frac{p(j|\hat{\varrho}_n)}{q(j|\hat{\varrho}_n, U)} \tag{6}$$

is an equal-weight sum of KL-divergences calculated for different elements of $\mathcal{D}$. It contains two types of conditional distributions – $p(j|\hat{\varrho})$ and $q(j|\hat{\varrho}, U)$. The first one, $p(j|\hat{\varrho})$, is taken to be a reference and will be discussed later in the text, and the latter one, $q(j|\hat{\varrho}, U)$, is the probability of registering j outcome given the $\hat{\varrho}$ input state:

$$q(j|\hat{\varrho}, U) = Tr(\hat{E}^{\dagger}_j(U) \hat{E}_j(U) \hat{\varrho}). \tag{7}$$

Kullback-Leibler divergence is non-negative and becomes zero for equal p and q [Gardiner, 1985].

Minimizing $S_{KL}(p||q)$ by a proper choice of unravelling U, it is possible to make $q(j|\hat{\varrho}, U)$ close to a reference distribution $p(j|\hat{\varrho})$. A natural choice for the latter is

[2]In the course of learning, parameters of the unravelling as a quantum information processing are established. But, in fact, this processing is implemented by means of a classical linear optical system. Hence, it belongs to the so-called 'quantum-inspired machine learning' [Huynh, 2023].

[3]The set $\mathcal{D}$ contains the states $\hat{\varrho}'_n = \Lambda'[|i_n\rangle\langle i_n|]$ formed by the transmission channel after alteration. In (5) and further the asterisk is dropped.

such that $p(j|\hat{\varrho}(i))$ reaches maximum for $i = j$. The decoding probabilities for optimal measurement (inaccessible by assumption) distinguishing $\hat{\varrho}(0)$ and $\hat{\varrho}(1)$ can serve as an example. It can certainly be chosen as a reference in (6), but for simplicity reasons an ideal decoding will be used:

$$p(j|\hat{\varrho}(i)) = \delta_{ij}, \tag{8}$$

similar to the one widely used in classical machine learning [Bishop, 2006]. For simplicity, we will only leave the part of (6) that depends on U, and introduce the functional

$$F(\mathcal{D}, U) \doteq \frac{1}{N}\sum_{n=1}^{N}\sum_{j=0,1} p(j|\hat{\varrho}(i_n))\ln q(j|\hat{\varrho}(i_n), U). \tag{9}$$

The choice of (9) corresponds to the maximum of F for close p and q. With (8), one gets

$$F(\mathcal{D}, U) = \pi_0 \ln q(0|\hat{\varrho}(0), U) + \pi_1 \ln q(1|\hat{\varrho}(1), U), \tag{10}$$

where π_0 and π_1 are fractions of zeroes and ones in the training set $\mathcal{D}$. It is reasonable to choose these fractions equal to the relative frequencies of symbols $'0'$ and $'1'$ occurring in a typical text transmitted through the channel.

As mentioned earlier, the training of interface aims for finding the matrix

$$U^{(opt)} = \arg\max_{U} F(\mathcal{D}, U), \tag{11}$$

that defines the optimal decoding unravelling. These matrices belonging to $SU(2)$ group have one-to-one correspondence to points of 3D sphere of a unit radius in a 4D Euclidian space. Coordinates on a sphere are the needed parameters. The functional $-F(\mathcal{D}, U)$ can be used in a traditional gradient descent scheme, moving on a sphere in search of $U^{(opt)}$. In what follows we consider an alternative approach, with U being treated as an integral matrix object. It allows to reformulate the search for $U^{(opt)}$ in terms of solution of an implicit equation.

3 The variational problem

Conditional probabilities (10) can be represented as follows:

$$q(i|\hat{\varrho}(i), U) = \sum_{j=0,1}\sum_{k=0,1} U^*_{ij} U_{ik} M_{kj}(i), \tag{12}$$

with

$$M_{kj}(i) \doteq Tr\left(\hat{E}^\dagger_j \hat{E}_k \hat{\varrho}(i)\right). \tag{13}$$

Since U is unitary, matrices $M(0)$ and $M(1)$ are Hermitian; in case of $\hat{E}_0$ and $\hat{E}_1$ being projectors on eigenstates of the optimal measurement (for a previously existing parameters of the channel), matrices $M(0)$ and $M(1)$ are diagonal.

The search for extremum of $F(\mathcal{D}, U)$ constrained by unitarity of U requires considering the following functional

$$\sum_{i=0,1}\sum_{j=0,1}\sum_{k=0,1}\left[\pi_i \ln\left(U^*_{ij} U_{ik} M_{kj}(i)\right) - U^*_{ji} U_{ki} L_{jk}\right], \tag{14}$$

where (10) is augmented by terms with a matrix of Lagrange multipliers L_{jk}. Their proper choice will ensure that the unitarity condition will be always fulfilled during solution of variational problem. Note that the matrix L should be Hermitian for the added terms to be positive. Assume a small variation of U: $U_{ij} \mapsto U_{ij} + \varepsilon_{ij}$. The variation (14), up to linear terms in ε and ε^*, takes the form

$$\sum_{i=0,1}\sum_{j=0,1}\sum_{k=0,1}\left[\frac{\pi_i}{q(i|\hat{\varrho}(i), U)}\left(U_{ik} M_{kj}(i)\varepsilon^*_{ij} + U^*_{ij} M_{kj}(i)\varepsilon_{ik}\right) - \left(U_{ki}\varepsilon^*_{ji} + U^*_{ji}\varepsilon_{ki}\right)L_{jk}\right]. \tag{15}$$

Equaling the terms with ε^*_{ij} to zero, one gets

$$\nu_i (UM(i))_{ij} = (LU)_{ij}. \tag{16}$$

Here

$$\nu_i \doteq \frac{\pi_i}{q(i|\hat{\varrho}(i), U)}. \tag{17}$$

In a compact matrix form, the relation (16) is equivalent to

$$\nu_i (UM(i)U^\dagger)_{ij} = L_{ij}. \tag{18}$$

Due to L being Hermitian,

$$\nu_0 (UM(0)U^\dagger)_{01} = \nu_1 (UM(1)U^\dagger)_{01}, \tag{19}$$

and from $M(0)$ and $M(1)$ being Hermitian it follows that

$$\nu_0 (UM(0)U^\dagger)_{10} = \nu_1 (UM(1)U^\dagger)_{10}. \tag{20}$$

From (19) and (20) the equality of corresponding matrix elements of the matrices $\nu_0 UM(0)U^\dagger$ and $\nu_1 UM(1)U^\dagger$ follows, i.e. the matrix $U(\nu_0 M(0) - \nu_1 M(1))U^\dagger$ appears to be diagonal. The search for $U^{(opt)}$ is then reduced to finding the eigenvectors of $\mathfrak{M} \doteq \nu_0 M(0) - \nu_1 M(1)$:

$$(U\mathfrak{M})_{ij} = \lambda_i U_{ij}. \tag{21}$$

Since the matrix $\mathfrak{M}$ depends on U through ν_0 and ν_1, expression (21) becomes an implicit equation on $U^{(opt)}$.

4 Discussion

Equation (21) allows for iterative computation of $U^{(n)}$ $n = 0, 1, 2, ...$ for a given initial unravelling $U^{(0)}$:

$$(U^{(n+1)}\mathfrak{M}(U^{(n)}))_{ij} = \lambda_i^{(n+1)} U_{ij}^{(n+1)}. \quad (22)$$

If this series converges, then it is possible to state the advantages of the proposed method over conventional gradient descent method, for which the step size in the unravelling parameter space should be declared explicitly.

The Kullbakc-Leibler divergence is not symmetric under permutation of distributions: $S_{KL}(p||q) \neq S_{KL}(q||p)$. The choice of the first variant in (6) was dictated by a further usage of model reference distribution (8). The second variant

$$S_{KL}(q||p) = \frac{1}{N}\sum_{n=1}^{N}\sum_{j=0,1} q(j|\hat{\varrho}_n, U)\times \ln\frac{q(j|\hat{\varrho}_n, U)}{p(j|\hat{\varrho}_n)} \quad (23)$$

in this case is not an option due to the presence of zero probabilities in (8) which leads to the divergence $S_{KL}(q||p)$ [4]. If instead of (8) one takes a distribution originating from optimal discrimination measurement (for $\hat{\varrho}(0)$ and $\hat{\varrho}(1)$), then probabilities $p(1|\hat{\varrho}(0))$ and $p(0|\hat{\varrho}(1))$ are generally non-zero. This enables usage of the functional (23) for search of optimal unravelling. In further investigations it would be interesting to compare unravellings that minimize $S_{KL}(q||p)$ and $S_{KL}(p||q)$.

A certain advantage of choosing $S_{KL}(q||p)$ over $S_{KL}(p||q)$ appears for close values of $q(j|\hat{\varrho}, U)$ and $p(j|\hat{\varrho})$. In this case,

$$S_{KL}(q||p) + 1 \simeq \frac{1}{N}\sum_{n=1}^{N}\sum_{j=0,1}\frac{q^2(j|\hat{\varrho}_n, U)}{p(j|\hat{\varrho}_n)}, \quad (24)$$

and the solution of variational problem for minimization of the right-hand side of (24) becomes even easier than described above.

The training set (5), with $\hat{\varrho}_n = \Lambda[|i_n\rangle\langle i_n|]$, is the simplest one and allows for analytical study. Note that in general the properties of the channel may be changed during the engineering of $\mathcal{D}$:

$$\hat{\varrho}_n = \Lambda_n[|i_n\rangle\langle i_n|]. \quad (25)$$

Formulation of the optimal unravelling as a solution of machine learning problem with functionals (6) and (23) remains valid even in this general case.

The scheme investigated in the present work is not internally constrained by the form of the maps Λ and Λ'; the only concern is the convergence of the series (22). It has a certain advantage over other proposals that use machine learning methods for discrimination of quantum states from the specific parametric sets [Patterson, 2021; Chen, 2021]. Also, current scheme is more transparent than the schemes based on neural networks [Magesan, 2015; Quiroga, 2021]. In case of quantum communication with polarization-encoded pulses, different unravellings of the decoder can be implemented by means of linear-optical elements, allowing for relatively simple experimental realization. The only necessary condition is the possibility to insert the interface that processes the quantum information transmitted through the channel *before* it is registered as classical measurement outcome. It is precisely this stage – after transmission, but before detection – where the apparatus implementing the unravelling operates.

The subject of the present work was limited to a classical decoding strategy, with a binary outcome. The goal is to minimize the error probability in every single measurement act. However, there exist strategies that allow for perfect state discrimination, at the cost of being able to do so only in a fraction of measurement series (so-called unambiguous state-discrimination [Peres, 1988; Ivanovic, 1987; Dieks, 1988; Zhang, 2009]). This can be done by allowing for additional decoding result – the inconclusive one. Optimization strategy in this case would aim at minimizing the number of such results. The scheme proposed in the present work can indeed be adapted to this paradigm, and investigation of its capabilities in such a setting is certainly of interest.

[4] Note that the nature of $q(j|\hat{\varrho}, U)$ distribution generally prohibits zero probabilities and divergence of $S_{KL}(p||q)$.

5 Conclusion

In conclusion, we proposed a strategy for minimizing the decoding error if the decoder is maladapted to the transmission channel. If the interaction of the decoder with the environment is interfaced by a device that implements a specific unravelling, than the decoding performance can be improved. This unravelling can be found during machine learning procedure, and the implicit equation on it was derived.

This work was supported by the State order (project AAAA-A21-121021800168-4) at the Institute of Automation and Electrometry SB RAS.

References

Nielsen, M.A., Chuang, I.L. (2000). *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge.

Helstrom, C.W. (1979). Quantum Detection and Estimation Theory. In *Mathematics in Science and Engineering*, **123**. Academic, New York.

Fuchs, C.A. (1996). Distinguishability and Accessible Information in Quantum Theory. quant-ph/9601020.

Kraus, K. (1983). *States, Effects, and Operations: Fundamental Notions of Quantum Theory*. Springer-Verlag, Berlin.

Sych, D. and Leuchs, G. (2016). Practical Receiver for Optimal Discrimination of Binary Coherent Signals. *Phys. Rev. Lett.*, **117**, 200501.

Rosati, M. et al. (2017). Optimal quantum state discrimination via nested binary measurements. *Phys. Rev. A*, **95**, 042307.

Weir, G., Barnett, S.M., and Croke, S. (2017). Optimal discrimination of single-qubit mixed states. *Phys. Rev. A*, **96**, 022312.

DiMario, M.T. and Becerra, F.E. (2018). Robust Measurement for the Discrimination of Binary Coherent States. *Phys. Rev. Lett.*, **121**, 023603.

Kechrimparis, S. (2019). Preserving measurements for optimal state discrimination over quantum channels. *Phys. Rev. A*, **99**, 062302.

Lidar, D.A., Chuang, I.L., and Whaley, K.B. (1997). Decoherencefree subspaces for quantum computation. *Phys. Rev. Lett.*, **81**, pp. 2594–2597.

Ticozzi, F., Ferrante, A. (2007). Finding quantum noiseless subsystems: A linear-algebraic approach. *3rd International Conference "Physics and Control" (PhysCon 2007).* Potsdam, Germany, Sep. 3-7.

Chuang, I.L. and Nielsen, M.A. (1997). Prescription for experimental determination of the dynamics of a quantum black box. *J. Mod. Opt.*, **44**, pp. 2455–2467.

Poyatos, J.F., Cirac, J.I., and Zoller, P. (1997). Complete Characterization of a Quantum Process: The Two-Bit Quantum Gate. *Phys. Rev. Lett.*, **78**, 390.

Breslin, J.K. and Milburn, G.J. (1997). Optimal quantum trajectories for discrete measurements. *J. Mod. Opt.*, **44**, pp. 2469–2484.

Il'ichov, L.V. (2013). Quantum Unravelling of Perception by Active Consciousness. *NeuroQuantology*, **11**, pp. 110–115.

Il'ichov, L.V. (2003). Practical implementation of a nontrivial evolution "unravelling" for a resonantly fluorescing atom. *JETP Lett.*, **77**, pp. 196–198.

Tomilin, V.A., Il'ichov, L.V. (2022) Quantum Feedback Control Outside of the Controlled System. *JETP Lett.*, **116**, pp. 649–654.

Huynh, L. et al. (2023). Quantum-Inspired Machine Learning: a Survey. https://arxiv.org/abs/2308.11269v2.

Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*. Prentice Hall, Upper Saddle River.

Gardiner, C.W. (1985). *Handbook of Stochastic Methods*. Springer-Verlag, Berlin.

Bishop, C.M. (2006). *Pattern Recoginition and Machine Learning*. Springer, New York.

Patterson, A. et al. (2021) Quantum state discrimination using noisy quantum neural networks. *Phys. Rev. Research*, **3**, 013603.

Chen, H. et al. (2021). Universal discriminative quantum neural networks. *Quantum Mach. Intell.*, **3**, 1.

Magesan, E. et al. (2015). Machine Learning for Discriminating Quantum Measurement Trajectories and Improving Readout. *Phys. Rev. Lett.*, **114**, 200501.

Quiroga, D., Date, P., and Pooser, R. (2021). Discriminating Quantum States with Quantum Machine Learning. arXiv:2112.00313v1 [quant-ph].

Peres, A. (1988). How to differentiate between nonorthogonal states. *Phys. Lett. A*, **128**, p. 19.

Ivanovic, I.D. (1987). How to differentiate between nonorthogonal states. *Phys. Lett. A*, **123**, pp. 257–259.

Dieks, D. (1988). Overlap and distinguishability of quantum states. *Phys. Lett. A*, **126**, pp. 303–306.

Zhang, M. et al. (2009). Optimal unambiguous discrimination problems under different a priori knowledge. *4th International Conference on Physics and Control (PhysCon 2009).* Catania, Italy, Sep. 1-9.