

# Transduction and Active Learning in the Quantum Learning of Unitary Transformations

Peter Wittek<sup>1</sup>

<sup>1</sup> *University of Borås Allegatan 1, Borås Sweden 50190*

**Abstract.** Quantum learning of a unitary transformation estimates a quantum channel in a process similar to process tomography. The classical counterpart of this goal, finding an unknown function, is regression, although the methodology hardly resembles the outline of classical algorithms. To gain a better understanding what such a methodology means to learning theory, we anchor it to the familiar concepts of active learning and transduction. Learning the unitary translates to storing it in quantum memory, but the procedure also requires an optimal maximally entangled input state; this resembles active learning. The retrieval strategy is a blend of inductive and transductive learning.

**Keywords:** Unknown Unitary, Quantum Process Tomography, Active Learning, Transduction

## 1 Introduction

Quantum learning of a unitary transformation estimates a quantum channel in a process similar to quantum process tomography: in tomography, one tries to infer a classical description of the unknown gate, whereas in learning, the goal is to simulate the application of the gate on a new input state, without necessarily having a classical description [1]. The classical counterpart of this goal, finding an unknown function, is known as regression, and then applying it to a new data point. In a classical setting, we define an objective function, and we seek an optimum subject to constraints and assumptions, typically from a parametric family of functions. The assumption in learning by quantum process tomography is that the channel is unitary and that the unitary transformation is drawn from a group – that is, it meets basic symmetry conditions [3]. The objective function is replaced by the fidelity of quantum states.

Apart from these similarities, the rest of the learning process does not resemble the classical variant. Unlike in the classical setting, learning a unitary requires a double maximization: we need an optimal measuring strategy that optimally approximates the unitary, and we also need an optimal input state that best captures the information of the unitary [4]. Using the learned unitary has different strategies, which may differ from the classical application of the estimated function.

In indirect process tomography, we use the Choi-Jamiołkowski isomorphism to imprint the unitary on a state. The key steps are as follows [2] (see also Figure 1):

- Learning the unitary translates to the optimal storage and parallel application of the unitary on a suitable input state.
- It requires an optimal input state, a superposition of maximally entangled states. This resembles active learning.
- Applying the learned unitary either with a coherent strategy, that is, retrieving from quantum memory, or with an incoherent strategy, that is, after measurement and retrieving it from classical memory.

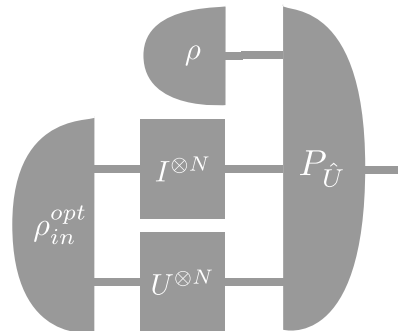


Figure 1: Outline of learning a unitary.

The optimal incoherent strategy is a mixture of inductive and transductive learning, whereas the sub-optimal coherent strategy is purely transduction.

To gain a better understanding what such a methodology means to learning theory, we anchor it to two concepts familiar from this field: active learning [5] and transduction [6].

## 2 Active Learning

Active learning is a variant of semi-supervised learning in which the learning algorithm is able to solicit labels for problematic unlabelled instances from an appropriate information source. Some labels are available, but most of the examples are unlabelled. The task in a learning iteration is to choose the optimal set of unlabelled examples for which the algorithm solicits labels from an appropriate information source, for instance, from a human annotator [5]. Some typical classical strategies are as follows:

- Uncertainty sampling: the selected set corresponds to those data instances where the confidence is low.
- Query by committee: train a simple ensemble that casts votes on data instances, and select those which are most ambiguous.

- Expected model change: select those data instances that would change the current model the most if the learner knew its label. This approach is particularly fruitful in gradient-descent-based models, where the expected change is easy to quantify by the length of the gradient.
- Expected error reduction: select those data instances where the model performs poorly, that is, where the generalization error is most likely to be reduced.
- Variance reduction: generalization performance is hard to measure, whereas minimizing output variance is far more feasible; select those data instances which minimize output variance.
- Density-weighted methods: the selected instances should not only be uncertain, but also representative of the underlying distribution.

Optimal quantum learning of unitaries is similar to active learning in a sense: it requires an *optimal* input state. Since the learner has access to  $U$ , by calling the transformation on an optimal input state, the learner ensures that the most important characteristics are imprinted on the approximation state of the unitary. The optimal input state for storage can be taken of the form

$$|\phi\rangle = \bigoplus_{j \in \text{Irr}(U^{\otimes N})} \sqrt{\frac{p_j}{d_j}} |\mathbb{I}_j\rangle \in \tilde{\mathcal{H}},$$

where  $p_j$  are probabilities, the index  $j$  runs over the set  $\text{Irr}(U^{\otimes N})$  of all irreducible representations  $\{U_j\}$  contained in the decomposition of  $\{U^{\otimes N}\}$ ,  $d_j$  is the dimension of the corresponding subspace  $H_j$ , and  $\tilde{\mathcal{H}} = \bigoplus_{j \in \text{Irr}(U^{\otimes N})} (H_j \otimes H_j)$  is a subspace of  $H_o \otimes H_i$  carrying the representation  $\tilde{U} = \bigoplus_{j \in \text{Irr}(U^{\otimes N})} (U_j \otimes \mathbb{I}_j)$ ,  $\mathbb{I}_j$  being the identity in  $H_j$ .

Binary classification scheme based on the quantum state tomography of state classes through Helstrom measurements also requires an optimal input state [7].

### 3 Transduction

Most classical learning models are inductive: based on a set of data points – labelled or unlabelled – we infer a function that will be applied to unseen data points. Transduction avoids this inference to the more general case, and it infers from particular instances to particular instances [8]. This way, transduction asks for less: an inductive function implies a transductive one. Transduction is similar to instance-based learning, a family of algorithms that compares new problem instances with training instances. The goal in transductive learning is actually to minimize test error, instead of the more abstract goal of maximizing generalization performance [6].

Two different retrieval strategies apply when we would like to use the learned unitary transformation: a coherent strategy, which stores the unitary in quantum memory, and an incoherent one, which measures the unitary and stores it in classical memory; the latter strategy is

considered optimal. The incoherent strategy is a blend of inductive and transductive learning, as the optimal input state depends on the number of target states on which the transformation should be applied, yet once it is learned, the transformation can be used an arbitrary number of times.

The optimal retrieving of  $U$  from the memory state  $|\phi_U\rangle$  is achieved by measuring the ancilla with the optimal covariant POVM in the form  $\Xi = |\eta\rangle\langle\eta|$  [2], namely  $P_{\hat{U}} = |\eta_{\hat{U}}\rangle\langle\eta_{\hat{U}}|$ , where  $|\eta_{\hat{U}}\rangle = \bigoplus_j \sqrt{d_j} |U_j\rangle$ , and, conditionally on outcome  $\hat{U}$ , by performing the unitary  $\hat{U}$  on the new input system. The optimal probability coefficients can be written as

$$p_j = \frac{d_j m_j}{d^N},$$

where  $m_j$  is the multiplicity of the corresponding space [3]. This true for the  $K = 1$  case. The more generic case for arbitrary  $K$  takes the global fidelity between the estimate channel  $C_U$  and  $U^{\otimes K}$ , that is, the objective function averages over  $(|\text{tr}(U^\dagger C_U)|/d)^2$ . Hence the exact values of  $p_j$  always depend on  $K$ , tying the optimal input state to the number of applications, making it a case of transduction.

### References

- [1] Bisio, A., Chiribella, G., D’Ariano, G. M., Facchini, S., and Perinotti, P., *Phys. Rev. A* **81** (2010) 032324.
- [2] Bisio, A., D’Ariano, G. M., Perinotti, P., and Sedláč, M., *Phys. Let. A*, **375** (2011) 3425–3434.
- [3] Chiribella, G., Dariano, G. M., and Sacchi, M. F., *Phys. Rev. A* **72** (2005) 042338.
- [4] Acín, A., Jané, E., and Vidal, G., *Phys. Rev. A* **64** (2001) 050302.
- [5] Settles, B., Active learning literature survey, Technical Report 1648, University of Wisconsin–Madison, 2009.
- [6] El-Yaniv, R. and Pechyony, D., Transductive Rademacher complexity and its applications, in *Proceedings of COLT-07, 20th Annual Conference on Learning Theory*, edited by Bshouty, N. H. and Gentile, C., pages 157–171, 2007.
- [7] Guță, M. and Kotłowski, W., *New J. Phys.* **12** (2010) 123032.
- [8] Gammernan, A., Vovk, V., and Vapnik, V., Learning by transduction, in *Proceedings of UAI-98, 14th Conference on Uncertainty in Artificial Intelligence*, pages 148–155, 1998.