Real-time calibration of coherent-state receivers: learning by trial and error [1]

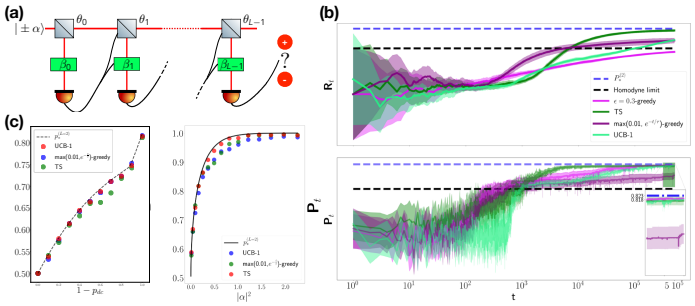
M. Bilkis,[1] M. Rosati,[1] R. Morral Yepes,[1] and J. Calsamiglia[1]

[1] *Física Teòrica: Informació i Fenòmens Quàntics, Departament de Física, Universitat Autònoma de Barcelona, ES-08193 Bellaterra (Barcelona), Spain*

We consider the discrimination of two coherent states of light with passive linear optics, displacements, on/off photodetectors and adaptive control. This is a prototypical problem in quantum information theory, of great technological significance for long-distance communication [4, 6]: the optimal receiver for binary coherent states requires infinite control rounds and it is still demanding at present (Dolinar receiver [2, 3]). Moreover, its extension to multiple states would allow to reach the Holevo communication capacity of real-world channels.

We propose an innovative and experimentally-appealing approach: the search for optimal discrimination strategies is cast as a test-bed for reinforcement learning (RL), studying how well an agent can calibrate a receiver without prior knowledge of: (i) the coherent states' energy; (ii) the actual receiver setup; (iii) the physical laws governing the system. The nature of our approach is particularly appealing for scenarios where an accurate description of the system is not possible, e.g., due to complexity of the receiver structure, experimental constraints or imperfections, untrusted devices or unknown communication channels. By trial and error, the RL agent has to sequentially choose actions, corresponding to different setups, based on previous photodetector outcomes and at a final stage guess for one of the hypotheses (see Fig. (a)). A non-zero reward is given only if the guess is correct. By repeating the procedure over several experiments, the agent earns experience and learns a near-optimal receiver setup and guessing rule with the resources at its disposal.

Let us stress that the RL problem induced by real-time state discrimination is characterized by intrinsically stochastic rewards. Indeed, even when performing a good set of actions and guessing rule, an agent might still not be rewarded. This is due to two crucial factors: (i) quantum states are intrinsically indistinguishable, i.e., even the best receiver has a non-zero probability of discrimination error; (ii) our methods can be applied in real time to the experiment, hence the binary reward received for a given set of actions is stochastic and thereby not sufficient to estimate the success probability of the corresponding receiver. Still, the best among our agents are able to reach good configurations in a number of experiments which is roughly sufficient to try each set of actions only once. Even more strikingly, the agents are also able to exploit the discovered configurations in real time and increase their success rate during the experiments.



In Fig. (b) we compare the average performance of 24 agents with the best Gaussian receiver (homodyne) [5] and the optimal receiver as a function of the number of experiments t, as measured by: (i) the real-time mean reward $\mathbf{R}_t$; (ii) the success probability of the best setup according to the agent $\mathbf{P}_t$. Some methods (upper confidence bound (UCB) and 0.3-greedy) favour exploration of the action space rather than exploitation of previously-found good setups. Hence, they are able to discover configurations very close to optimal (large $\mathbf{P}_t$) but not to exploit them in real time (small $\mathbf{R}_t$). On the contrary, other methods (Thomson sampling (TS) and exponentially-decaying ϵ-greedy) favour exploitation over exploration. For our problem TS has the most profitable exploration/exploitation balance. Indeed, in a search space of $\sim 3\cdot 10^3$ possible receiver setups, whose performance cannot be evaluated within a single experiment, TS discovers a receiver beating the homodyne after $\sim 3\cdot 10^2$ experiments, surpasses its real-time success rate after $\sim 10^3$ experiments and finally achieves near-optimal performance after $\sim 10^5$ experiments. Finally, in Fig. (c) we show robustness of our agents' learning performance as a function of the signal's energy and the dark-count probability.

[1] M. Bilkis, M. Rosati, R. Morral Yepes, and J. Calsamiglia. Real-time calibration of coherent-state receivers: learning by trial and error. preprint at arXiv 2001.10283, accepted for publication in Phys. Rev. Research.

[2] M. T. Dimario and F. E. Becerra. Robust Measurement for the Discrimination of Binary Coherent States. *Phys. Rev. Lett.*, 121(2):023603, jul 2018.
[3] S. J. Dolinar. Communication and sciences engineering. *Q. Prog. Rep. (Research Lab. Electron.*, 111:115, 1973.
[4] Saikat Guha. Structured optical receivers to attain superadditive capacity and the Holevo limit. *Phys. Rev. Lett.*, 106(24):1–4, 2011.
[5] Masahiro Takeoka and Masahide Sasaki. Discrimination of the binary coherent signal: Gaussian-operation limit and simple non-Gaussian near-optimal receivers. *Phys. Rev. A - At. Mol. Opt. Phys.*, 78(2):1–7, 2008.
[6] Atsushi Waseda, Masahide Sasaki, Masahiro Takeoka, Mikio Fujiwara, Morio Toyoshima, and Antonio Assalini. Numerical Evaluation of PPM for Deep-Space Links. *J. Opt. Commun. Netw.*, 3(6):514, 2011.