Quantum Change Point

Gael Sentís,[1] Emilio Bagan,[2] John Calsamiglia,[2] Giulio Chiribella,[3,4] and Ramon Muñoz-Tapia[2]

[1]*Departamento de Física Teórica e Historia de la Ciencia, Universidad del País Vasco UPV/EHU, E-48080 Bilbao, Spain*
[2]*Física Teòrica: Informació i Fenòmens Quàntics, Departament de Física,*
Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain
[3]*Department of Computer Science, The University of Hong Kong, Pokfulam Road, Hong Kong*
[4]*Canadian Institute for Advanced Research, CIFAR Program in Quantum Information Science, Toronto, Ontario M5G 1Z8*

Sudden changes are ubiquitous in nature. Identifying them is crucial for a number of applications in biology, medicine, and social sciences. Here we take the problem of detecting sudden changes to the quantum domain. We consider a source that emits quantum particles in a default state, until a point where a mutation occurs that causes the source to switch to another state. The problem is then to find out where the change occurred. We determine the maximum probability of correctly identifying the change point, allowing for collective measurements on the whole sequence of particles emitted by the source. Then, we devise online strategies where the particles are measured individually and an answer is provided as soon as a new particle is received. We show that these online strategies substantially underperform the optimal quantum measurement, indicating that quantum sudden changes, although happening locally, are better detected globally.

The detection of sudden changes in a sequence of random variables is a pivotal topic in statistics, known as the *change point problem* [1–3]. The problem has widespread applications, including the study of stock market variations [4], protein folding [5], and landscape changes [6]. In general, identifying change points plays a crucial role in all problems involving the analysis of samples collected over time [2,7], because such analysis requires the stability of the system parameters [8]. If changes are correctly detected, the sample can be conveniently divided in subsamples, which can then be analyzed by the standard statistical techniques. The detection of change points can also be viewed as a border problem [9], namely, a problem where one wants to draw a separation between two (or more) different configurations—a task that plays a central role in machine learning [10].

The simplest example of a change point problem is that of a coin with variable bias. Imagine that a game of Heads or Tails is played with a fair coin, but after a few rounds one player suspects that the other has replaced the coin with a biased one. After inspection of the coin, the suspicion is confirmed: the coin has now a bias. Can we identify when the coin was changed based only on the sequence of outcomes? This classical problem has a natural extension to the quantum realm, illustrated in Fig. 1: A source is promised to prepare quantum particles in some default state. At some point, however, the source undergoes a mutation and starts to produce copies of a different state. Given the sequence of particles emitted by the source, the problem is to find out when the change took place. In the basic version of the problem, the initial and final states are known, as in the classical example of the coin. No prior

information is given about the location of the change: *a priori*, every point of the sequence is equally likely to be the change point. For simplicity, we assume the quantum states to be pure. More elaborate variations can be considered, e.g., with unknown states, nonuniform priors, mixed states, and multiple change points. However, as we will see in this Letter, the basic scenario already captures the intriguing features that distinguish the quantum change point problem from its classical version. An example where the change can happen only at two possible positions and the states are completely unknown was studied in Ref. [11], where the problem was shown to be equivalent to programmable discrimination [12].

The quantum change point problem can be formulated as a problem of state discrimination. For a sequence of n particles emitted by the source, the problem is to distinguish among n quantum states, the kth state having the change point in the kth position. Notably, state discrimination problems with multiple quantum states have no
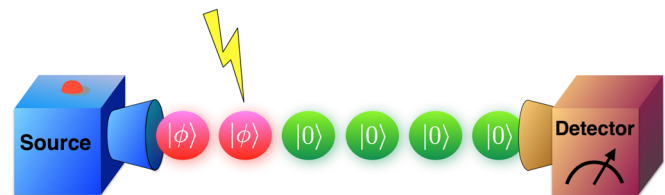

FIG. 1.   The quantum change point problem. A quantum source emits particles in a default state $|0\rangle$, until the point where a mutation occurs, causing the source to emit particles in a different state $|\phi\rangle$. A detector receives the stream of particles emitted from the source and measures them, producing an estimate of the point where the mutation occurred.

closed-form solution in general (see Ref. [13] for some recent progress). A complete solution is known only for the two-state case, a result that dates back to Helstrom's seminal work four decades ago [14]—and even in this case computing the success probability may not be straightforward, see, e.g., the derivation of the quantum Chernoff bound [15]. For $n \geq 3$, the only cases where a solution is known are those of pure states with a high degree of symmetry. This includes the symmetric states [16], generated by the action of a unitary operation U satisfying $U^n = \mathbb{1}$, and, more generally, states that are generated by a group of unitaries [17,18].

Interestingly, the change point problem does not fall into any of the above categories. In spite of this, we show that the problem can be completely solved in the asymptotic regime: in the limit of long sequences, the maximum probability of success takes the elegant form

$$P_{\max} = \frac{4(1 - c^2)}{\pi^2} K^2(c^2) + O\left(\frac{1}{n^{1-\epsilon}}\right), \qquad (1)$$

where c is the overlap between the default state and the alternate state, $K(x)$ is the complete elliptic function of the first kind [19], and $\epsilon > 0$ is an arbitrary constant. Quite remarkably, the limit probability has a nonzero value despite the fact that the number of states to be distinguished tends to infinity.

Equation (1) characterizes the ultimate quantum limit to the detection of a change point. Achieving the limit requires a nonlocal measurement, performed jointly on all the particles in the sequence. To perform such measurement one needs a quantum memory, wherein the states received from the source can be stored as they arrive to the detector. Since quantum memories are challenging to implement, it is also interesting to consider local strategies, where the particles are measured as soon as they arrive to the detector, possibly adapting the measurement settings at one step based on the outcomes of the previous measurements steps. Local strategies are interesting also because they can provide an online answer: they have a chance to identify the change point as soon as it occurs, without having to wait until all the n particles are scanned. When the online strategies are compared with the optimal quantum strategy, our results indicate the presence of a gap, showing that the availability of a quantum memory and the ability to perform nonlocal measurements offer an advantage in the identification of the change point.

Let us derive the optimal quantum strategy. We denote by $|0\rangle$ the default state and by $|\phi\rangle = c|0\rangle + s|1\rangle$ the state after the change. Without loss of generality, we choose c to be real and positive. If the change occurs in the position k, the state of the n particles is

$$|\Psi_k\rangle = |0\rangle^{\otimes k-1} |\phi\rangle^{\otimes n-k+1}. \qquad (2)$$

We call the above states the *source states*. Note that the source states are linearly independent, except in the trivial case when the states $|\phi\rangle$ and $|0\rangle$ are equal. In principle, the change can occur in any position, meaning that every source state has the same *a priori* probability $1/n$ [20]. The detector is described by a positive operator-valued measure (POVM), namely, a set of operators $\{M_k\}_{k=1}^n$ satisfying the positivity condition $M_k \geq 0$ and the completeness relation $\sum_{k=1}^n M_k = \mathbb{1}$, where $\mathbb{1}$ denotes the identity in the space $\mathcal{S}$ spanned by the source states. The average probability of successfully identifying the change point is $P = (1/n) \sum_{k=1}^n \langle \Psi_k | M_k | \Psi_k \rangle$ and our goal is to maximize it over all possible POVMs.

As mentioned above, the source states do not fall into any of the categories of states that admit a closed-form solution to the detection problem. Still, we now show how an optimal solution can be constructed in the large n limit. The key to our argument is a general result about the discrimination of linearly independent pure states, which is of independent interest:

Theorem.—Let $\{|\Psi_k\rangle\}_{k=1}^n$ be a set of linearly independent states and let

$$G_{ij} = \langle \Psi_i | \Psi_j \rangle \qquad (3)$$

be the components of the corresponding Gram matrix. The maximum probability of correctly identifying a state drawn uniformly at random from the set $\{|\Psi_k\rangle\}_{k=1}^n$ satisfies the bounds

$$P_{\max} \geq \left(\frac{\mathrm{tr}\sqrt{G}}{n}\right)^2 \qquad (4)$$

and

$$P_{\max} \leq \left(\frac{\mathrm{tr}\sqrt{G}}{n}\right)^2 + \sqrt{\lambda_{\max}} \|\mathbf{q} - \mathbf{u}\|_1, \qquad (5)$$

where $\lambda_{\max}$ is the maximum eigenvalue of G, $\mathbf{q} = \{q_k\}$ is the probability distribution defined by $q_k := (\sqrt{G})_{kk}/\mathrm{tr}[\sqrt{G}]$, $\mathbf{u} = \{u_k\}$ is the uniform distribution ($u_k = 1/n$ for all k), and $\|\mathbf{q} - \mathbf{u}\|_1 := \sum_k |q_k - u_k|$ is the trace norm.

The proof of the Theorem is provided in the Supplemental Material [21], where we further extend the result to nonuniform prior distributions. Note that the two bounds (4) and (5) are exactly equal when the diagonal matrix elements of $\sqrt{G}$ are all equal to each other. In this case, the Theorem yields the exact value of the success probability, reproducing a recent result of Ref. [23].

We now evaluate the bounds (4) and (5) for the change point problem, showing that the two bounds match at the leading order. We start by evaluating the trace of $\sqrt{G}$. First, we observe that the Gram matrix has matrix elements $G_{ij} = c^{|i-j|}$ and that its inverse has the simple form

$$G^{-1} = \frac{1+c^2}{1-c^2}\mathbb{1} - \frac{c}{1-c^2}H, \qquad (6)$$

where $H_{ij} = \delta_{ij+1} + \delta_{ji+1} + c(\delta_{i1}\delta_{j1} + \delta_{in}\delta_{jn})$. Luckily, the eigenvalues and eigenvectors of H can be constructed explicitly: in the Supplemental Material we show that the eigenvalues have the form $2\cos\theta_l$, where θ_l is a suitable angle in the interval of size π/n centered around $\pi l/n$ [21]. Equation (6) then implies that the eigenvalues of the Gram matrix G are

$$\lambda_l = \frac{1-c^2}{1-2c\cos\theta_l+c^2}, \qquad (7)$$

so that we have

$$\frac{\mathrm{tr}\sqrt{G}}{n} = \frac{1}{n}\sum_{l=1}^{n}\sqrt{\frac{1-c^2}{1-2c\cos\theta_l+c^2}}. \qquad (8)$$

Since the angles θ_l are distributed in intervals of equal size, forming a partition of the interval $[0,\pi)$, the sum can be replaced by an integral in the large n limit, yielding the asymptotic equality

$$\begin{aligned}
\frac{\mathrm{tr}\sqrt{G}}{n} &= \frac{\sqrt{1-c^2}}{\pi}\int_0^\pi \frac{d\theta}{\sqrt{1-2c\cos\theta+c^2}} \\
&= \frac{2\sqrt{1-c^2}}{\pi}K(c^2),
\end{aligned} \qquad (9)$$

valid up to an error of size $1/n^{1-\epsilon}$ [21]. According to Eq. (4), the square of the right-hand side is a lower bound for the maximum success probability.

Let us now evaluate the upper bound (5). First, note that we have $\lambda_{\max} \leq (1+c)/(1-c)$, as one can easily read out from Eq. (7). Moreover, it is possible to show that the probability distribution **q** is approximately uniform, with the bound $\|\mathbf{q}-\mathbf{u}\|_1 \leq 4(1+c)/(1-c)1/n^{1-\epsilon}$ holding at the leading order in n [21]. Hence, the upper bound (5) yields the inequality

$$P_{\max} \leq (\mathrm{tr}\sqrt{G}/n)^2 + 4\left(\frac{1+c}{1-c}\right)^{3/2}\frac{1}{n^{1-\epsilon}}. \qquad (10)$$

In summary, the bounds (4) and (5) match in the asymptotic limit, up to an error of size $1/n^{1-\epsilon}$. This establishes the validity of Eq. (1).

Asymptotically, the maximum success probability is attained by the square root measurement [24,25]. Indeed, it is possible to show [21] that the success probability of the square root measurement, denoted by P_{SQ}, satisfies the bound

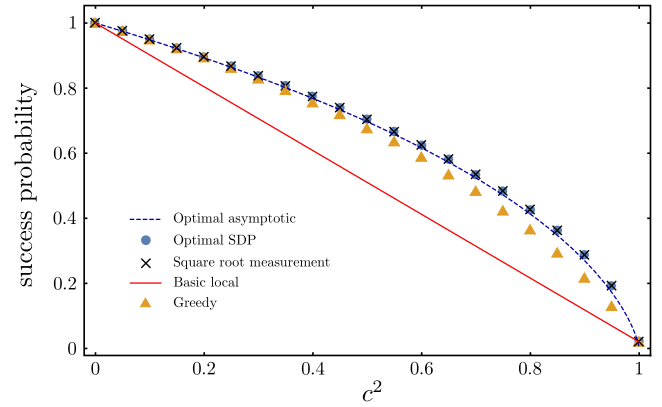$$P_{\mathrm{SQ}} \geq (\mathrm{tr}\sqrt{G}/n)^2, \qquad (11)$$

FIG. 2. Probability of correct identification of the change point as a function of $c^2 = |\langle 0|\phi\rangle|^2$. The (blue) dashed line is the asymptotic result in Eq. (1). For a sequence of $n = 50$ states, we also plot the maximum probability, obtained by SDP optimization, (blue) solid dots, and the results corresponding to the square root measurement, (black) crosses (which lie virtually on top of the dots); the basic local strategy, (red) straight line; and the greedy strategy, (orange) triangles.

and therefore is equal (at the leading order) to the maximum success probability. We also performed a numerical analysis, revealing that the square root measurement is an extremely good approximation already for short sequences ($n \gtrsim 10$), with a difference with respect to the optimal success probability of less than 10^{-3}. In Fig. 2 we compare the asymptotic result in Eq. (1) with the results for $n = 50$ corresponding to the square root measurement and to the optimal measurement obtained via semidefinite programming [26]. As is apparent from Fig. 2, the agreement is strikingly good. The figure also includes the success probability of various local measurement strategies that will be discussed below. Notice that for $c = 0$ the source states are orthogonal and perfect identification is possible, while in the limit $c \to 1$ the source states become indistinguishable and the success probability is given by random guessing, $P_{\max} = 1/n \to 0$ for $n \to \infty$. It is also patent from the figure that Eq. (1) is a lower bound that becomes tight as n goes to infinity, in agreement with the bounds (4) and (5). A numerical fit reveals that the correction to the leading order in Eq. (1) is of order $1/n$, again, consistently with our estimate.

Both the optimal measurement and the square root measurement involve global operations on all the n particles. This means that one has to scan the whole sequence of particles before getting an estimate of the change point. We now analyze the performance of online strategies, where each particle is individually measured as soon as it reaches the detector. The simplest such strategy consists of measuring each particle in the computational basis $\{|0\rangle, |1\rangle\}$. The measurements are performed sequentially until the outcome 1 is obtained for the first time, say at the rth step. At this point, we will know for sure that the measured particle was in the state $|\phi\rangle$, meaning that the

change must have occurred at some position $k \leq r$. Our best guess for the change point is then $\hat{k} = r$, since this is the most likely hypothesis given the observed data. For the success probability one has the exact expression $P_{BL} = 1 - c^2 + c^2/n$, where BL stands for "basic local." This strategy is suboptimal for $0 < c < 1$, and remains suboptimal for $n \to \infty$. The relative difference between the two success probabilities can be up to 50% for suitable values of the overlap c.

It is intriguing to explore whether more general online strategies can increase the success probability. Let us consider a scenario where a classical learning agent is asked to guess when the change point occurs. The agent starts with a uniform prior $p(k) = 1/n$ about the location of the change point and updates her expectation as new data become available. In order to update her information at the sth step, the agent must perform a measurement, which generally depends on the results, $r_1, r_2, \ldots, r_{s-1}$ obtained in the previous steps. Here we focus on *greedy* strategies [27], i.e., strategies that maximize the success probability at every step [28]. For these strategies, we determine the optimal measurement and the optimal guessing rule. The optimal strategy works as follows: At step s, the agent has to perform the Helstrom measurement [14] that distinguishes between the states $|0\rangle$ and $|\phi\rangle$, given with (unnormalized) prior probabilities [21]

$$p_0^{(s)} := \max_k \{p(k|r_1, \ldots, r_{s-1})\}_{k=s+1}^n$$
$$p_\phi^{(s)} := \max_k \{p(k|r_1, \ldots, r_{s-1})\}_{k=1}^s. \qquad (12)$$

Here, $p_0^{(s)}$ ($p_\phi^{(s)}$) is the probability of the most likely sequence that has the particle at position s in the state $|0\rangle$ ($|\phi\rangle$). The agent can deduce these probabilities from the posterior probabilities, updated at step s. After the sth measurement has been performed, the prior is updated in accordance with the measurement result, using Bayes' update rule: adopting the shorthand notation $\eta_k^{(s)} := p(k|r_1, \ldots, r_{s-1})$, we have

$$\eta_k^{(s+1)} = \frac{p(r_s|k)\eta_k^{(s)}}{\sum_{l=1}^n p(r_s|l)\eta_l^{(s)}}. \qquad (13)$$

After the last measurement, the agent updates the prior to $\eta_k^{(n+1)}$ and produces the guess $\hat{k}$ that maximizes $\eta_k^{(n+1)}$ for the change point.

For the greedy strategy, the full optimization over the measurements has been carried out analytically. However, a direct quantification of the average performance is intractable, because the number of possible sequences of outcomes grows exponentially with n. In order to compute the average success probability we used a Monte Carlo simulation, leading to the values plotted in Fig. 2 (orange triangles). The figure shows a significant enhancement of performance

over the basic local strategy. Still, the optimal greedy strategy does *not* attain the optimal quantum performance, as the gap with the optimal collective strategy remains even for large values of n. In short, this means that a learner with quantum memory outperforms a (greedy) learner with classical memory in the task of detecting change points.

Having observed a gap between the optimal greedy strategy and the optimal quantum strategy, it is interesting to check whether the gap can be closed by using arbitrary local strategies, where each particle can be measured multiple times and the measurement settings can depend on the outcomes of all previous measurements. Note that here the learner is allowed to use a quantum memory, but is limited to perform individual measurements on the particles. Unfortunately, optimizing over arbitrary local strategies is a daring task. Nevertheless, we can provide an upper bound to the success probability by considering POVM operators that are positive under partial transposition (PPT). In this case, a numerical optimization proves the existence of a gap between the local strategies and the optimal collective strategy for every value of n up to $n = 7$.

In this Letter we introduced the quantum change point problem—a quantum version of the problem of identifying changes in a sequence of random variables. In the quantum change point problem, a source emits particles in a default state until a point where a mutation occurs, causing the source to switch to a different state. For pure states, we determined the maximum probability of correctly identifying the change point, showing that, for large sequences of particles, the optimal performance is attained by the square root measurement. We also investigated online strategies, where each particle is measured individually as soon as it is received from the source. Among the online strategies, we identified the optimal greedy strategy, which provides the best online guess at each step. Our calculations show a gap between the greedy strategy and the optimal quantum strategy based on a global measurement. Further numerical optimization shows that the gap remains open even for arbitrary local strategies, indicating that local operations cannot match the performance of the optimal quantum protocol. In particular, this shows that a machine with quantum memory can outperform all machines with classical memory.

[1] E. S. Page, Biometrika **41**, 100 (1954); **42**, 523 (1955).

[2] B. E. Brodsky and B. S. Darkhovsky, *Non-Parametric Statistical Diagnosis* (Springer, Science+Business Media, B.V, Dordrecht, 2000).

[3] M. Baseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application* (Prentice Hall, Information and System Sciences Series, New Jersey, 1993).

[4] J. Chen and A. K. Gupta, J. Am. Stat. Assoc. **92**, 438 (1997).

[5] M. Pirchi, G. Ziv, I. Riven, S. Sedghani Cohen, N. Zohar, Y. Barak, and G. Haran, Nat. Commun. **2**, 493 (2011).

[6] G. Francesco Ficetola and M. Denoel, Ecography **32**, 1075 (2009).

[7] J. Chen and A. K. Gupta, *Parametric Statistical Change Point Analysis*, 2nd ed. (Birkhäuser, Boston, 2012).

[8] D. M. Hawkins, J. R. Stat. Soc. C **25**, 51 (1976).

[9] E. Rosten, R. Porter, and T. Drummond, IEEE Trans. Pattern Anal. Mach. Intell. **32**, 105 (2010).

[10] J. Takeuchi and K. Yamanishi, IEEE Trans. Knowledge Data Eng. **18**, 482 (2006).

[11] D. Akimoto and M. Hayashi, Phys. Rev. A **83**, 052328 (2011).

[12] G. Sentis, E. Bagan, J. Calsamiglia, and R. Munoz-Tapia, Phys. Rev. A **82**, 042312 (2010); **83**, 039909(E) (2011).

[13] K. Nakahira, K. Kato, and T. S. Usuda, Phys. Rev. A **91**, 052304 (2015).

[14] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).

[15] K. M. R. Audenaert, J. Calsamiglia, R. Muñoz-Tapia, E. Bagan, Ll. Masanes, A. Acin, and F. Verstraete Phys. Rev. Lett. **98**, 160501 (2007); J. Calsamigilia, R. Muñoz-Tapia, Ll. Masanes, A. Acin, and E. Bagan, Phys. Rev. A **77**, 032311 (2008).

[16] S. M. Barnett, Phys. Rev. A **64**, 030303R (2001).

[17] G. Chiribella, G. M. D'Ariano, P. Perinotti, and M. F. Sacchi, Phys. Rev. A **70**, 062105 (2004).

[18] G. Chiribella, G. M. D'Ariano, P. Perinotti, and M. F. Sacchi, Int. J. Quantum. Inform. **04**, 453 (2006).

[19] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1970).

[20] It is easy to include the case where the change does not occur at all, by adding the vector $|\Psi_0\rangle = |0\rangle^{\otimes n}$ and replacing the probability with $1/(n+1)$.

[21] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.117.150502 for technical details about the Gram matrix eigenvectors and eigenvalues, demonstration of the Theorem, and the calculation of asymptotic limits, which includes Ref. [22].

[22] V. P. Belavkin, Stochastics **1**, 315 (1975).

[23] N. Dalla Pozza and G. Pierobon, Phys. Rev. A **91**, 042334 (2015).

[24] P. Hausladen and W. K. Wootters, J. Mod. Opt. **41**, 2385 (1994).

[25] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, and W. K. Wootters, Phys. Rev. A **54**, 1869 (1996).

[26] S. Boyd and L. Vanderberghe, *Convex Optimization* (Cambridge University Press, Cambridge, England, 2004).

[27] P. E. Black, in *Dictionary of Algorithms and Data Structures*, edited by V. Pieterse and P. E. Black (2005), https://xlinux.nist.gov/dads/HTML/greedyalgo.html.

[28] In the considered scenario, the most general strategy can be optimized by dynamical programming, an optimization technique based on backward induction [29]. Unfortunately, for the change point, the numerical overhead is prohibitive, as one needs to accurately approximate n-multidimensional functions. In practice, such technique is only suitable for problems with a limited number of hypothesis [30].

[29] R. E. Bellman, *Dynamic Programming* (Princeton University Press, Princeton, 2010).

[30] J. Calsamiglia, J. I. de Vicente, R. Munoz-Tapia, and E. Bagan, Phys. Rev. Lett. **105**, 080504 (2010).