Dear Editor and Referees,

We would like to thank you for your insightful comments on our manuscript. The comments are very helpful and enable us to improve the quality of the manuscript.

We have extensively revised our manuscript according to the comments of the referees. Especially, to make the paper easy to read for the broad audience, such as the readership of Phys. Rev. Lett., we have clarified the definitions of notations, have resorted to vague expressions, and have provided a pedagogical example demonstrating the utility of the main theorem. Because of space limitation, some results have been moved to the Supplemental Material (SM). We expect that the revised manuscript can meet the requirement.

The changes made according to Referee A are highlighted in red, those according to Referee B in blue, and those according to both referees in purple. The comments of the referees are written in **boldface**, and the replies are written in lightface. We have also made some changes (highlighted in green) as will be described later. Quotations from the original and/or revised manuscripts are framed. The page numbers, etc., refer to the revised manuscript if we do not specify them explicitly.

Response to Referee A's comments

I. Summary

In this paper, the author demonstrates that the finding the maximum success probability for discriminating quantum process with restricted strategies is a convex optimization problem in which the Lagrange dual problem has zero duality gap. The author also derives a necessary and sufficient condition for achieving an optimal restricted strategy. Within this scheme, the authors shows that the adaptive strategies are not necessary for optimal discrimination when a problem has a symmetry. Furthermore, the author demonstrates that the discrimination problem with optimal restricted process can be formulated in terms of a robustness measure.

II. Suggestions/Comments

This paper has a readability issue, and I hope that the author could consider the improvement. My comments are shown below

We agree with the referee's assessment that we should improve the readability of our manuscript.

1. Diagrams (Fig. 1 and Fig. 3): I think that for the broad audiences in Phys. Rev. Lett, people are not familiar with these diagrams. It would be great if the author could explain how to interpret these diagrams explicitly in the equations. Especially, Fig.3 is very hard to get what is going on, while there are explanations in the captions. But, I do not think that it is enough.

Agreed. We have added the explanation of diagrams in the captions of Figs. 1 and 3 [note that Fig. 3(a),(c) of the original manuscript has been moved to the SM]:

> Caption of Fig. 1:
> **(Original text)**
> General protocol of quantum process discrimination.
> **(Revised text)**
> General protocol for the discrimination of quantum combs $\{\hat{\mathcal{E}}_m = \hat{\Lambda}_m^{(T)} \circledast \cdots \circledast \hat{\Lambda}_m^{(1)}\}_{m=1}^M$ (plotted in black) with T time steps, where $W'_1, \dots, W'_{T-1}$ are ancillary systems. Any physically allowed discrimination strategy can be represented by a tester (plotted in blue), which consists of a state $\hat{\sigma}_1$, channels $\hat{\sigma}_2, \dots, \hat{\sigma}_T$, and a measurement $\{\hat{\Pi}_k\}_{k=1}^M$.

> Caption of Fig. 3:
> **(Original text)**

Examples of three types of restricted testers. The left and right figures, respectively, show the diagrammatic representations of testers and their schematic diagrams. (a) A nonadaptive tester, which consists of a state $\hat{\rho}$ and a measurement $\{\hat{\Pi}_k\}_k$, is applied to a given process $\hat{\mathcal{E}}$. (b) A tester is performed by two parties, Alice and Bob, in which only one-way classical communication from Alice to Bob is allowed. Such a tester consists of two sequentially connected single-shot testers. Alice prepares a state $\hat{\rho}_{\mathrm{A}}$, performs a measurement $\{\hat{\Psi}_i\}_i$, and communicates her outcome i to Bob. Then, Bob prepares a state $\hat{\rho}_{\mathrm{B}}^{(i)}$ and performs a measurement $\{\hat{\Pi}_k^{(i)}\}_k$. (c) A tester is performed by Alice and Bob, in which only one-way classical communication from Alice to Bob is allowed. Alice prepares a state $\hat{\rho}_i$ with a probability q_i and sends its one part to the process $\hat{\mathcal{E}}$. She also sends i to Bob, who performs a channel $\hat{\sigma}_i$. Alice finally performs a measurement $\{\hat{\Pi}_k^{(i)}\}_k$.

(Revised text)

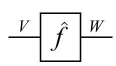
Discrimination scheme for $\{\hat{\mathcal{E}}_m = \hat{\Lambda}_m \circledast \hat{\Lambda}_m\}_{m=1}^3$ with a sequential tester. Consider the task to be performed by two parties, Alice and Bob. Alice prepares a quantum state $\hat{\rho}_{\mathrm{A}} \in \mathsf{Den}_{V_1 \otimes V_1'}$, feeds the system V_1 into $\hat{\Lambda}_m$, and performs a measurement $\{\hat{\Psi}_j\}_{j \in \mathcal{J}}$ on $W_1 \otimes V_1'$, where $\mathcal{J}$ is a set, which may contain any number of elements. According to its outcome, j, Bob prepares a state $\hat{\rho}_{\mathrm{B}}^{(j)} \in \mathsf{Den}_{V_2 \otimes V_2'}$, feeds the system V_2 into $\hat{\Lambda}_m$, and performs a measurement $\{\hat{\Pi}_k^{(j)}\}_{k=1}^3$ on $W_2 \otimes V_2'$. $\{\hat{A}_j := \hat{\Psi}_j \circledast \hat{\rho}_{\mathrm{A}}\}_{j \in \mathcal{J}}$ and $\{\hat{B}_k^{(j)} := \hat{\Pi}_k^{(j)} \circledast \hat{\rho}_{\mathrm{B}}^{(j)}\}_{k=1}^3$ $(\forall j \in \mathcal{J})$ are testers.

Also, we have changed the diagrams to make them easy to understand for readers. In Sec. I.B of the SM, we have provided information on how to interpret the diagrams shown in this manuscript:

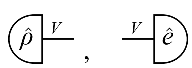
Sec. I.B of the SM:

(Revised text)

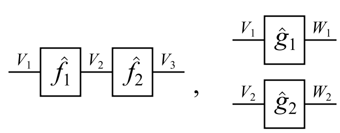
In this manuscript, we often use diagrammatic representations to provide an intuitive understanding. We here briefly review the diagrammatic representation of quantum processes. A single-step process $\hat{f} \in \mathsf{Pos}(V, W)$ is depicted by



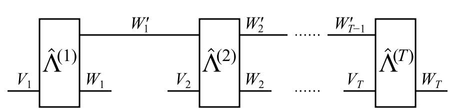
The trivial system $\mathbb{C}$ is represented by 'no wire'. For example, $\hat{\rho} \in \mathsf{Pos}_V$ and $\hat{e} \in \mathsf{Pos}(V, \mathbb{C})$ are diagrammatically represented as



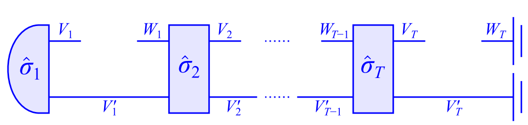
Single-step processes can be linked sequentially or in parallel. The sequential concatenation of $f_1 \in \mathsf{Pos}(V_1, V_2)$ and $f_2 \in \mathsf{Pos}(V_2, V_3)$ is a single-step process in $\mathsf{Pos}(V_1, V_3)$, denoted as $f_2 \circ f_1$. Also, the parallel concatenation of $g_1 \in \mathsf{Pos}(V_1, W_1)$ and $g_2 \in \mathsf{Pos}(V_2, W_2)$ is a single-step process in $\mathsf{Pos}(V_1 \otimes V_2, W_1 \otimes W_2)$, denoted as $g_1 \otimes g_2$. In diagrammatic terms, they are depicted as



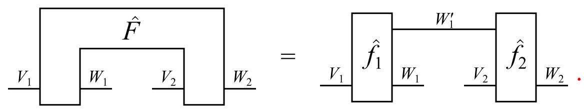
A quantum process is a concatenation of single-step processes; in particular, a quantum comb is a concatenation of channels. Each element of $\mathsf{Comb}_{W_T, V_T, \ldots, W_1, V_1}$ corresponds to a comb expressed in the form



where $\hat{\Lambda}^{(1)}, \ldots, \hat{\Lambda}^{(T)}$ are channels. Each element of $\mathsf{Comb}^*_{W_T, V_T, \ldots, W_1, V_1}$ corresponds to a comb expressed in the form



where $\hat{\sigma}_1, \dots, \hat{\sigma}_T$ are channels (in particular, $\hat{\sigma}_1$ is a state) and "$\dashv\!\vdash$" denotes the trace. Any tester for a process in $\mathsf{Comb}_{W_T,V_T,\dots,W_1,V_1}$ is represented by $\{\Phi_m\}_{m=1}^M \in \mathsf{Pos}_{\tilde{V}}^M =: C_{\mathrm{G}}$ satisfying $\sum_{m=1}^M \Phi_m \in \mathsf{Comb}^*_{W_T,V_T,\dots,W_1,V_1} =: \mathcal{S}_{\mathrm{G}}$. In our manuscript, processes belonging to $\mathsf{Comb}^*_{W_T,V_T,\dots,W_1,V_1}$ and tester elements are diagrammatically depicted in blue. The concatenation of two single-step processes $\hat{f}_1 \in \mathsf{Pos}(V_1, W'_1 \otimes W_1)$ and $\hat{f}_2 \in \mathsf{Pos}(W'_1 \otimes V_2, W_2)$, denoted by the process $\hat{F} := \hat{f}_1 \circledast \hat{f}_2$ (where $\circledast$ denotes the concatenation), is often depicted as



2. Examples: Since the paper is very abstract, in order to get the broad audience's attention, it would be great if the author could explain the utility of the main results with pedagogical examples (1-qubit or 2-qubit case, etc), and implement the detailed calculations in order to make the audience understand better what is going on. These examples can be included in the supplemental materials, so that the readers can follow what happens.

Agreed. We have added a pedagogical example of discriminating unitary qubit channels with a sequential tester:

Page 3, left column:
(Revised text)

Example — We illustrate the use of Theorem 1 in the following example. Let us consider the problem of discriminating three qubit channels $\hat{\Lambda}_1, \hat{\Lambda}_2, \hat{\Lambda}_3$ with $T = 2$ uses, in which case V_1, W_1, V_2, and W_2 are all qubit systems and $\hat{\mathcal{E}}_m = \hat{\Lambda}_m \circledast \hat{\Lambda}_m$ (i.e., $\mathcal{E}_m = \Lambda_m \otimes \Lambda_m$) holds. Assume that the prior probabilities are equal and that each $\hat{\Lambda}_m$ is the unitary channel represented by $\hat{\Lambda}_m(\rho) = U^m \rho U^{-m}$, where $U := \mathrm{diag}(1, \omega)$ and $\omega := \exp(2\pi\sqrt{-1}/3)$. Then, we have

$$\Lambda_m = \begin{bmatrix} 1 & 0 & 0 & \omega^{-m} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \omega^m & 0 & 0 & 1 \end{bmatrix}.$$

We consider the case where a tester is restricted to a sequential one. The set of all sequential testers in $\mathcal{P}_{\mathrm{G}}$, denoted by $\mathcal{P}_{\mathrm{seq}}$, is (see Fig. 3 and Sec. III of the SM)

$$\mathcal{P}_{\mathrm{seq}} := \left\{ \left\{ \sum_j B_m^{(j)} \otimes A_j \right\}_{m=1}^3 : \{A_j\} \in \mathsf{Test},\ \{B_m^{(j)}\}_m \in \mathsf{Test}_3 \right\},$$

where

$$\mathsf{Test}_M := \left\{ \{B_m\}_{m=1}^M \subset \mathsf{Pos}_4 : \sum_{m=1}^M B_m = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes \rho,\ \rho \in \mathsf{Den}_2 \right\},$$

$\mathsf{Test} := \bigcup_{M=1}^\infty \mathsf{Test}_M$, and Pos_n and Den_n are, respectively, the sets of all positive semidefinite and density matrices of order n. Problem (P) with $\mathcal{P} = \mathcal{P}_{\mathrm{seq}}$ can be written as the following non-convex programming problem:

$$\begin{aligned} &\text{maximize} && \frac{1}{3} \sum_{m=1}^3 \sum_j \langle B_m^{(j)}, \Lambda_m \rangle \langle A_j, \Lambda_m \rangle \\ &\text{subject to} && \{A_j\} \in \mathsf{Test},\ \{B_m^{(j)}\}_m \in \mathsf{Test}_3. \end{aligned} \tag{3}$$

This problem is very hard to solve due to two main reasons: i) both $\{A_j\}$ and $\{B_m^{(j)}\}_m$ $(\forall j)$ need to be optimized and ii) how many elements an optimal tester $\{A_j\}$ has is unknown. Here, we pay attention to the fact that Eq. (2) with

$$C := \left\{ \left\{ \sum_j B_m^{(j)} \otimes A_j \right\}_{m=1}^{3} : A_j \in \mathsf{Pos}_4,\ \{B_m^{(j)}\}_m \in \mathsf{Test}_3 \right\},$$
$$\mathcal{S} := \mathcal{S}_{\mathrm{G}} \tag{4}$$

holds (see Sec. III of the SM). In this situation, Problem (D) is expressed as

$$\begin{aligned} \text{minimize} \quad & D_{\mathcal{S}_{\mathrm{G}}}(\chi) \\ \text{subject to} \quad & \sum_{m=1}^{3} \left\langle \sum_j B_m^{(j)} \otimes A_j, \chi - \frac{1}{3} \Lambda_m \otimes \Lambda_m \right\rangle \geq 0 \\ & \left[\forall A_j \in \mathsf{Pos}_4,\ \{B_m^{(j)}\}_m \in \mathsf{Test}_3 \right] \end{aligned} \tag{5}$$

with $\chi \in \mathsf{Her}_{\tilde{V}}$. After some algebra, this problem is reduced to (see Sec. III of the SM)

$$\begin{aligned} \text{minimize} \quad & \lambda \\ \text{subject to} \quad & \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \geq \frac{1}{3} \sum_{m=1}^{3} \langle B_m, \Lambda_m \rangle \begin{bmatrix} 1 & \omega^{-m} \\ \omega^m & 1 \end{bmatrix} \\ & (\forall \{B_m\} \in \mathsf{Test}_3) \end{aligned} \tag{6}$$

with $\lambda \in \mathbf{R}_+$. This problem is much easier to solve than Problem (3). Also, Theorem 1 guarantees that the optimal value of Problem (6), which is numerically found to be around 0.933, is equal to the maximum success probability. Note that in the case where any physically allowed discrimination strategy can be used, we can easily see that the three channels can be perfectly distinguished with two uses (see Sec. III of the SM). We restrict our discussion here to the discrimination problem for symmetric unitary qubit channels, but our method can be applied to more general combs. Other examples of different restricted strategies are shown in Sec. IV of the SM.

Also, to deal with the example discussed in the main paper, we have revised the text in Sec. VI of the SM.

3. Other comments:

(a) It would be great if the author rewrite the manuscript by paying more attentions to how the notations are introduced. Let me show the example. In Eq. ($\mathrm{P_G}$), the author introduces $\mathcal{P}_{\mathrm{G}}$. However, the definition or the concept of $\mathcal{P}_{\mathrm{G}}$ is not explained before introducing this equation. The author explains it in the next section, and says "$\mathcal{P}_{\mathrm{G}}$ can be written as...". This means that $\mathcal{P}_{\mathrm{G}}$ has its general definition, and it has a form from different candidates of possible expressible forms. Therefore, the author has to first introduce the definition of the notation, and then use it in the rest of the manuscript in order to make the manuscript easy to read.

We apologize for not defining some terms. We also carefully revised our manuscript to use not-so-standard notions after defining them. As for $\mathcal{P}_{\mathrm{G}}$, we have changed the text as follows:

Page 2, left column:
(Original text)
The probability that a tester $\Phi := \{\Phi_m\}$ gives the outcome k for the process $\mathcal{E}_m$ is given by

$$\langle \Phi_k, \mathcal{E}_m \rangle := \mathrm{Tr}(\Phi_k \mathcal{E}_m) = \hat{\Pi}_k \circ \hat{\Lambda}_m^{(2)} \circ \hat{\sigma}_2 \circ \hat{\Lambda}_m^{(1)} \circ \hat{\sigma}_1,$$

where $\circ$ denotes the map composition. The task of finding the maximum success probability for discriminating the given quantum processes $\{p_m, \mathcal{E}_m\}_{m=1}^{M}$ (where p_m is the prior probability of $\mathcal{E}_m$) can be formulated by as an optimization problem, namely [20]

$$\begin{aligned} \text{maximize} \quad & P(\Phi) := \sum_{m=1}^{M} p_m \langle \Phi_m, \mathcal{E}_m \rangle \\ \text{subject to} \quad & \Phi \in \mathcal{P}_{\mathrm{G}}. \end{aligned} \tag{P$_{\mathrm{G}}$}$$

(Revised text)

Let $\mathcal{P}_{\mathrm{G}}$ be the set of all such testers $\Phi := \{\Phi_k\}_{k=1}^{M}$, which can be written as (see [41] for details)

$$\mathcal{P}_{\mathrm{G}} = \left\{ \{\Phi_m\}_{m=1}^{M} \subset \mathsf{Pos}_{\tilde{V}} : \sum_{m=1}^{M} \Phi_m \in \mathcal{S}_{\mathrm{G}} \right\}, \tag{1}$$

where

$$\begin{aligned} \mathcal{S}_{\mathrm{G}} := \{ I_{W_2} \otimes \tau_2 : \; & \tau_2 \in \mathsf{Pos}_{V_2 \otimes W_1 \otimes V_1}, \\ & \tau_1 \in \mathsf{Den}_{V_1}, \;\; \mathrm{Tr}_{V_2} \tau_2 = I_{W_1} \otimes \tau_1 \}. \end{aligned}$$

The probability that a tester Φ gives the outcome k for the comb $\mathcal{E}_m$ is given by $\langle \Phi_k, \mathcal{E}_m \rangle := \mathrm{Tr}(\Phi_k \mathcal{E}_m)$. The task of finding the maximum success probability for discriminating the given quantum combs $\{\mathcal{E}_m\}_{m=1}^{M}$ with prior probabilities $\{p_m\}_{m=1}^{M}$ can be formulated as an optimization problem, namely [20]

$$\begin{aligned} \text{maximize} \quad & P(\Phi) := \sum_{m=1}^{M} p_m \langle \Phi_m, \mathcal{E}_m \rangle \\ \text{subject to} \quad & \Phi \in \mathcal{P}_{\mathrm{G}}. \end{aligned} \tag{P$_{\mathrm{G}}$}$$

(b) Regarding $W'_1 = \cdots = W'_{T-1}$, the author wrote that these are the internal systems, but these systems belong to C, which is the set of complex numbers. This will confuse the reader a little bit. So, it would be great if the author could clarify this point, and explain.

We apologize for the lack of clarity. We wanted to mention that a channel discrimination problem is a special case of a process discrimination problem with $W'_1 = \cdots = W'_{T-1} = \mathbb{C}$. We have revised as follows:

Page 1, right column:

(Original text)

For example, the T-shot discrimination problem of channels $\hat{\Lambda}_1, \ldots, \hat{\Lambda}_M$ is a special case of $\hat{\Lambda}_m^{(t)} = \hat{\Lambda}_m$ $(\forall t)$ and $W'_1 = \cdots = W'_{T-1} = \mathbb{C}$.

(Revised text)

In the particular case where, for each m, $\hat{\mathcal{E}}_m$ has no ancillary system and $\hat{\Lambda}_m^{(1)}, \ldots, \hat{\Lambda}_m^{(T)}$ are the same channel, denoted by $\hat{\Lambda}_m$, the problem reduces to the problem of discriminating M channels $\hat{\Lambda}_1, \ldots, \hat{\Lambda}_M$ with T uses.

(c) Regarding $\hat{\sigma}_1$, below Fig. 1, the author says that $\hat{\sigma}_1$ is a initial state. However, $\hat{\sigma}_2$ is a quantum channel, and then the author also says $\hat{\sigma}_1$ and $\hat{\sigma}_2$ are the quantum process. These descriptions actually confuse the readers. So, I hope that author could review the definitions of every notion.

We apologize for the lack of clarity. We use the term "process" as a general term, including quantum states, channels, and measurement elements; however, this would make the readers confused. We have revised as follows:

Page 1, right column, the last line:
(Original text)

To discriminate between given processes, we first prepare a bipartite system $V_1 \otimes V_1'$ in an initial state $\hat{\sigma}_1$. One part V_1 is sent through the channel $\hat{\Lambda}_m^{(1)}$, followed by a channel $\hat{\sigma}_2$. After that, we send the system V_2 through the channel $\hat{\Lambda}_m^{(2)}$ and perform a measurement $\{\hat{\Pi}_k\}_{k=1}^M$ on the system $W_2 \otimes V_2'$. Such a sequence of processes $\hat{\sigma}_1, \hat{\sigma}_2, \{\hat{\Pi}_k\}_{k=1}^M$ allows us to represent any discrimination strategy, i.e., a tester, which includes an entanglement-assisted tester and an adaptive tester.

(Revised text)

As shown in Fig. 1, to discriminate between given combs, we first prepare a bipartite system $V_1 \otimes V_1'$ in an initial state $\hat{\sigma}_1$. One part V_1 is sent through the channel $\hat{\Lambda}_m^{(1)}$, followed by a channel $\hat{\sigma}_2$. After that, we send the system V_2 through the channel $\hat{\Lambda}_m^{(2)}$ and perform a measurement $\{\hat{\Pi}_k\}_{k=1}^M$ on the system $W_2 \otimes V_2'$. Such a collection of $\{\hat{\sigma}_1, \hat{\sigma}_2, \{\hat{\Pi}_k\}_{k=1}^M\}$, which can be expressed as $\{\hat{\Phi}_k := \hat{\Pi}_k \circledast \hat{\sigma}_2 \circledast \hat{\sigma}_1\}_{k=1}^M$, is called a tester. Any discrimination strategy, including an entanglement-assisted strategy and an adaptive strategy, can be represented by a tester.

We also have reviewed the definition of every notion.

III. Recommendations

The results of the paper are important and timely, as this paper provides an essential step toward to solving the quantum process discrimination problem. Verifying the characteristics of the problem as the convex optimization problem where the Lagrange dual problem has zero duality gap, I think, is an important achievement. Therefore, I suggest that the results of the paper are publishable. However, in order to publish in Phys. Rev. Lett, which aims to the broad audience in the physics community, the presentation of this paper has to be polished extensively because the style of the current version is hard to read for the broad audience. Therefore, I suggest that the author should clarify the definitions of the notations, details of the diagrams, and provide a pedagogical example, to make the manuscript easy to read. In conclusion, I suggest the extensive revision of current version before considering the publishability to Phys. Rev. Lett.

We are pleased to hear that the referee appreciates the importance of our results. We have polished the manuscript to make it easy to read for the broad audience. We have clarified the definitions of the not-so-standard terms and details of the diagrams and have provided a pedagogical example.

Response to Referee B's comments

General Comments

The paper addresses and important question of discrimination of quantum processes that have a causal structure, such processes are described by quantum combs. This is done under any restrictions on the allowed discrimination strategies. The optimal success probability is formulated as a convex optimization problem so that the techniques of convex programming can be applied. The results of the paper are as follows:

1. The Lagrangian dual is found and it is shown that it has a zero duality gap. This is applied to three examples: restriction to nonadaptive strategies, not using the causal structure of the process, and two types of sequential strategies.

2. Conditions are obtained for existence of a restricted strategy that is globally optimal.

3. It is shown that under certain covariance properties a nonadaptive strategy is globally optimal.

4. The success probability in each such restricted discrimination problem is related to a robustness measure known in resource theories of quantum processes.

The problem of quantum process discrimination is notoriously hard and complicated, especially when various possible restrictions on the testers are taken into account. This paper is a valuable contribution and certainly deserves publication. However, I have some doubts on its suitability for PRL, due to the following issues:

-The use of SDP or convex programming techniques in quantum discrimination problems is quite standard, in discrimination of processes (e.g. Ref. [20]) or discrimination of states by restricted measurements (e.g. also in some papers of the present author). The methods applied here are very similar to those previously used in this context. The dual formulation of the problem might give some computational advantages (not so much demonstrated here), but no interpretation of the dual is provided and little new insight is gained. Some insight is provided by the relation to the generalized robustness measure, which in fact is a consequence of this duality, but this is not recognizable from the main text. So I doubt that the dual formulation can be appreciated by non-experts.

We apologize for the unclear presentation. The main contributions of our work are:

(1) We provide a new general formulation of the task of finding the maximum success probability for discriminating quantum processes under a restricted discrimination strategy as a convex optimization problem whose Lagrange dual problem has zero duality gap. Our formulation is applicable to any quantum processes and any restricted strategy. In some scenarios, the Lagrange dual problem of our formulation can be much easier to solve than the original problem.

(2) We show that the maximum success probability can be written in terms of a generalized robustness measure.

We have revised our manuscript to stress these points; for example, the abstract has been changed as follows:

> Abstract:
> **(Original text)**
> In this Letter, we show that the task of finding the maximum success probability for discriminating any quantum processes with any restricted strategy can always be formulated as a convex optimization problem whose Lagrange dual problem exhibits zero duality gap. We also derive a necessary and sufficient condition for an optimal restricted strategy to be optimal within the set of all strategies. As an application of this result, it is shown that adaptive strategies are not necessary for optimal discrimination if a problem has a certain symmetry. Moreover, we show that the optimal performance of each restricted process discrimination problem can be written in terms of a certain robustness measure.
> **(Revised text)**
> In this Letter, we present a general formulation of the task of finding the maximum success probability for discriminating quantum processes as a convex optimization problem whose Lagrange dual problem exhibits zero duality gap. The proposed formulation can be applied to any restricted strategy. We provide a simple example in which the dual problem given by our formulation can be much easier to solve than the original problem. We also show that the optimal performance of each restricted process discrimination problem can be written in terms of a certain robustness measure.

Regarding Statement (1), we agree with the referee that the use of SDP techniques in quantum discrimination problems is quite standard (note that the technique for proving the zero duality gap would also be standard), while non-SDP convex programming techniques may not be so standard in quantum discrimination problems. We believe that the proposed formulation is novel and quite general; we should emphasize that, to our knowledge, such a general formulation has not previously been reported even if we restrict ourselves to quantum state discrimination problems. One of our key ideas is to interpret each restricted tester Φ as a vector in some convex cone C such that the sum $\sum_{m=1}^{M} \Phi_m$ is in some convex set $\mathcal{S}$. We also believe that our formulation is practical, at least for some applications. We have revised our manuscript as follows:

> Page 1, left column:
> **(Original text)**
> In this paper, we provide a general method to analyze quantum process discrimination problems in which discrimination testers are restricted to given types of testers. We show that the task of finding the maximum success probability for discriminating any quantum processes with any restricted tester can be formulated as a convex optimization problem and that its Lagrange dual problem has zero duality gap. The dual problem is often easier to solve analytically or numerically than the original problem.

(Revised text)

In this Letter, we provide a general method to analyze quantum process discrimination problems in which discrimination testers are restricted to given types of testers. We show that the task of finding the maximum success probability for discriminating any quantum processes can be formulated as a convex optimization problem even if the allowed testers are restricted to any subset of all testers and that its Lagrange dual problem has zero duality gap. It should be mentioned that, to our knowledge, a convex programming formulation applicable to any restricted strategy has not yet been reported even in quantum state discrimination problems. In some scenarios, the dual problem can be much easier to solve analytically or numerically than the original problem, as we will demonstrate through a simple example.

Page 2, right column:

(Original text)

We became aware that the set of all testers $\mathcal{P}_{\mathrm{G}}$ can be written as

$$\mathcal{P}_{\mathrm{G}} = \left\{ \Phi \in \mathcal{C}_{\mathrm{G}} : \sum_{m=1}^{M} \Phi_m \in \mathcal{S}_{\mathrm{G}} \right\},$$

where $\mathcal{C}_{\mathrm{G}} := \mathsf{Pos}_{\tilde{V}}^{M}$ and $\mathcal{S}_{\mathrm{G}}$ is the set of all combs in $\mathsf{Pos}_{\tilde{V}}^{M}$, which means that $\mathcal{P}_{\mathrm{G}}$ is fully characterized by the closed convex cone $\mathcal{C}_{\mathrm{G}}$ and the closed convex set $\mathcal{S}_{\mathrm{G}}$ [44]. Similarly, for any feasible set $\mathcal{P}$, we can choose a closed convex cone $\mathcal{C}$ and a closed convex set $\mathcal{S}$ such that (see Fig. 2)

$$\overline{\mathsf{co}}\,\mathcal{P} = \left\{ \Phi \in \mathcal{C} : \sum_{m=1}^{M} \Phi_m \in \mathcal{S} \right\}, \quad \mathcal{C} \subseteq \mathcal{C}_{\mathrm{G}}, \quad \mathcal{S} \subseteq \mathcal{S}_{\mathrm{G}}. \tag{1}$$

(Revised text)

We find that, for any feasible set $\mathcal{P}$, each tester $\Phi \in \mathcal{P}$ can be interpreted as an element in some convex cone such that $\sum_{m=1}^{M} \Phi_m$ is in some convex set. Specifically, we can choose a closed convex cone $\mathcal{C}$ and a closed convex set $\mathcal{S}$ such that (see Fig. 2)

$$\overline{\mathsf{co}}\,\mathcal{P} = \left\{ \Phi \in \mathcal{C} : \sum_{m=1}^{M} \Phi_m \in \mathcal{S} \right\}, \quad \mathcal{C} \subseteq \mathcal{C}_{\mathrm{G}}, \quad \mathcal{S} \subseteq \mathcal{S}_{\mathrm{G}}, \tag{2}$$

where $\mathcal{C}_{\mathrm{G}} := \mathsf{Pos}_{\tilde{V}}^{M}$.

Moreover, we provide an example in our main paper to demonstrate the computational advantages of the dual formulation (see the section entitled "Example" in page 3).

Regarding Statement (2), as the referee points out, the relation between the optimal value of Problem (P) and the generalized robustness measure is a consequence of the duality. In order for the readers to be recognizable this fact, we provide the proof of Corollary 2 (of our revised manuscript) in our main paper:

Page 4, left column:

(Original text)

Let

$$\mathcal{K} := \left\{ Y \in \mathsf{Her}_{W_{\mathrm{A}} \otimes \tilde{V}} : \sum_{m=1}^{M} \langle |m\rangle\langle m| \otimes \Phi_m, Y \rangle \geq 0 \ (\forall \Phi \in \mathcal{C}) \right\},$$
$$\mathcal{F} := \{ I_{W_{\mathrm{A}}} \otimes \chi' : \chi' \in \mathsf{Her}_{\tilde{V}},\ D_{\mathcal{S}}(\chi') \leq 1/M \};$$

then, we can easily prove that the optimal value of Problem (P) is equal to $[1 + R_{\mathcal{K}}^{\mathcal{F}}(\mathcal{E}^{\mathrm{ex}})]/M$ (see Sec. VIII of the SM).

(Revised text)

Using Theorem 1, we can see that the optimal value of Problem (P) is characterized by a robustness measure.

Corollary 2 Let

$$\mathcal{K} := \left\{ Y \in \mathsf{Her}_{W_A \otimes \tilde{V}} : \sum_{m=1}^{M} \langle |m\rangle\langle m| \otimes \Phi_m, Y \rangle \ge 0 \ (\forall \Phi \in C) \right\},$$
$$\mathcal{F} := \{ I_{W_A} \otimes \chi' : \chi' \in \mathsf{Her}_{\tilde{V}},\ D_{\mathcal{S}}(\chi') = 1/M \};$$

then, the optimal value of Problem (P) is equal to $[1 + R_{\mathcal{K}}^{\mathcal{F}}(\mathcal{E}^{\mathrm{ex}})]/M$.

Proof From Eq. (7), we have

$$\frac{1 + R_{\mathcal{K}}^{\mathcal{F}}(\mathcal{E}^{\mathrm{ex}})}{M} = \inf \left\{ \frac{1+\lambda}{M} : \frac{\mathcal{E}^{\mathrm{ex}} + \delta}{1+\lambda} = I_{W_A} \otimes \chi',\ \chi' \in \mathsf{Her}_{\tilde{V}}, \right.$$
$$\left. D_{\mathcal{S}}(\chi') = \frac{1}{M},\ \delta \in \mathcal{K} \right\}.$$

Letting $\chi := (1+\lambda)\chi'$ and using some algebra, the right-hand side becomes

$$\begin{aligned}
&\inf \{ D_{\mathcal{S}}(\chi) : \chi \in \mathsf{Her}_{\tilde{V}},\ I_{W_A} \otimes \chi - \mathcal{E}^{\mathrm{ex}} \in \mathcal{K} \} \\
&\quad = \inf \left\{ D_{\mathcal{S}}(\chi) : \chi \in \mathsf{Her}_{\tilde{V}},\ \sum_{m=1}^{M} |m\rangle\langle m| \otimes (\chi - p_m \mathcal{E}_m) \in \mathcal{K} \right\} \\
&\quad = \inf \{ D_{\mathcal{S}}(\chi) : \chi \in \mathcal{D}_C \} = D^\star,
\end{aligned}$$

where $D^\star$ is the optimal value of Problem (D). Thus, Theorem 1 completes the proof. ■

Also, to let the readers gain some new insights, we give a simple derivation of the dual problem:

Page 2, right column:

(Original text)

We obtain the following statement [proved in Sec. II of the Supplemental Material (SM)]:

Theorem 1 Let us arbitrarily choose a closed convex cone C and a closed convex set $\mathcal{S}$ satisfying Eq. (1); then, the optimal value of Problem (P) coincides with that of the following Lagrange dual problem:

$$\begin{aligned}
&\text{minimize} && D_{\mathcal{S}}(\chi) := \max_{\varphi \in \mathcal{S}} \langle \varphi, \chi \rangle \\
&\text{subject to} && \chi \in \mathcal{D}_C,
\end{aligned} \tag{D}$$

where

$$\mathcal{D}_C := \left\{ \chi \in \mathsf{Her}_{\tilde{V}} : \sum_{m=1}^{M} \langle \Phi_m, \chi - p_m \mathcal{E}_m \rangle \ge 0 \ (\forall \Phi \in C) \right\}.$$

(Revised text)

Let

$$\mathcal{D}_C := \left\{ \chi \in \mathsf{Her}_{\tilde{V}} : \sum_{m=1}^{M} \langle \Phi_m, \chi - p_m \mathcal{E}_m \rangle \ge 0 \ (\forall \Phi \in C) \right\};$$

then, we can easily verify that

$$D_{\mathcal{S}}(\chi) := \max_{\varphi \in \mathcal{S}} \langle \varphi, \chi \rangle \ge \sum_{m=1}^{M} \langle \Phi_m^\star, \chi \rangle \ge P(\Phi^\star)$$

holds for any $\chi \in \mathcal{D}_C$. The first and second inequalities follow from $\sum_{m=1}^{M} \Phi_m^\star \in \mathcal{S}$ and $\Phi^\star \in C$, respectively. Thus, the optimal value of the following problem

$$\begin{array}{ll} \text{minimize} & D_\mathcal{S}(\chi) \\ \text{subject to} & \chi \in \mathcal{D}_C \end{array} \tag{D}$$

is not less than that of Problem (P). We can see that Problem (D), which is the so-called Lagrange dual problem of Problem (P), has zero duality gap, as shown in the following theorem [proved in Sec. II of the Supplemental Material (SM)]:

Theorem 1 Let us arbitrarily choose a closed convex cone C and a closed convex set $\mathcal{S}$ satisfying Eq. (2); then, the optimal values of Problems (P) and (D) are the same.

- Some of the results (3.,4.) were already observed before (in less general situations), some citations are missing.

We apologize for the missing citations relating to Result 3.; we have included several citations in Sec. VI of the SM:

Sec. VI of the SM:
(Revised text)
Note that several related results in particular cases have been reported [S3–S5].

Regarding Result 4., please see the reply to Specific Comment 6.

- The paper is not very well written, with definitions omitted, unclear notations or vague expressions. See some of the specific comments below.

I think it would be better to publish the results as a regular article, with more space given to necessary definitions and explanations.

We apologize for the writing issues. According to the referee's specific comments, we have extensively revised our manuscript to address these issues.

Specific Comments

1. It is stated in the abstract that "We also derive a necessary and sufficient condition for an optimal restricted strategy to be optimal within the set of all strategies". This is also repeated in the Introduction. This is slightly confusing: what is proved in Prop. 2 (and also in Sec. V of SM) is a necessary and sufficient condition for -existence- of a restricted strategy that is globally optimal. Only a sufficient condition for a -specific- optimal restricted strategy to be globally optimal is given.

We apologize for the confusion. In our original manuscript, the statement in the abstract is essentially equivalent to that in Proposition 2. Let $P^\star$ be the globally maximum success probability. If there exists a restricted strategy, denoted by $\Phi \in \mathcal{P}$, that is globally optimal [i.e., $P(\Phi) = P^\star$], then any optimal restricted strategy, denoted by $\Phi' \in \mathcal{P}$, must satisfy $P(\Phi') \ge P(\Phi) = P^\star$. Thus, from $P(\Phi') \le P^\star$, we have $P(\Phi') = P^\star$, which yields that Φ' is also globally optimal.
Because of the limited space, we have moved this result to the SM.

2. p.1, col. 2: "superchannels" is used in some papers for processes transforming channels to channels, I am not sure that this is a standard notion, so it would be better to explain

Agreed. We have added the explanation as follows:

Page 1, right column:
(Original text)
States, channels, and superchannels are special cases of quantum processes.
(Revised text)
States, channels, and superchannels, which are processes that transform quantum channels to quantum channels, are special cases of quantum combs.

3. p.2, col. 1: $\mathcal{P}_{\mathrm{G}}$ is not defined

We apologize for not defining $\mathcal{P}_{\mathrm{G}}$. We have corrected it. Please see the reply to Comment 3(a) of Referee A.

4. p.2, col. 2: "...the set of all combs in...": without specification of the input and output spaces, this is quite ambiguous. For example, any state is a comb, or any (Choi matrix of) a channel with any choice of the input/output spaces in the given composite space $\tilde{V}$. These are very different sets. I guess what is meant here is the set $\mathsf{Comb}_{W_T,V_T,\ldots,W_1,V_1}$ defined in the SM. Note here that according to the original definition of a quantum comb in Ref. [41] this set is in fact the set of all combs where the first input and the last output spaces are trivial (so one should add the spaces $W_0 = V_{T+1} = \mathbb{C}$).

We apologize for the vague statement. As the referee points out, "the set of all combs in ..." here means the set $\mathsf{Comb}_{W_T,V_T,\ldots,W_1,V_1}$ defined in the SM of the original manuscript.
To provide a clear definition, we have modified our main text as follows:

Page 2, left column:
(Original text)
We became aware that the set of all testers $\mathcal{P}_{\mathrm{G}}$ can be written as

$$\mathcal{P}_{\mathrm{G}} = \left\{ \Phi \in C_{\mathrm{G}} : \sum_{m=1}^{M} \Phi_m \in \mathcal{S}_{\mathrm{G}} \right\},$$

where $C_{\mathrm{G}} := \mathsf{Pos}_{\tilde{V}}^{M}$ and $\mathcal{S}_{\mathrm{G}}$ is the set of all combs in $\mathsf{Pos}_{\tilde{V}}^{M}$, which means that $\mathcal{P}_{\mathrm{G}}$ is fully characterized by the closed convex cone C_{G} and the closed convex set $\mathcal{S}_{\mathrm{G}}$ [44].
(Revised text)
Let $\mathcal{P}_{\mathrm{G}}$ be the set of all such testers $\Phi := \{\Phi_k\}_{k=1}^{M}$, which can be written as (see [41] for details)

$$\mathcal{P}_{\mathrm{G}} = \left\{ \{\Phi_m\}_{m=1}^{M} \subset \mathsf{Pos}_{\tilde{V}} : \sum_{m=1}^{M} \Phi_m \in \mathcal{S}_{\mathrm{G}} \right\}, \tag{1}$$

where

$$\begin{aligned} \mathcal{S}_{\mathrm{G}} := \{ I_{W_2} \otimes \tau_2 : \tau_2 \in \mathsf{Pos}_{V_2 \otimes W_1 \otimes V_1}, \\ \tau_1 \in \mathsf{Den}_{V_1},\ \operatorname{Tr}_{V_2} \tau_2 = I_{W_1} \otimes \tau_1 \}. \end{aligned}$$

Also, to give a more natural definition of $\mathsf{Comb}_{W_T,V_T,\ldots,W_1,V_1}$, we have revised our manuscript as follows (note that $\mathsf{Comb}_{W_T,V_T,\ldots,W_1,V_1}$ of the original manuscript is equivalent to $\mathsf{Comb}^{*}_{W_T,V_T,\ldots,W_1,V_1}$ of the revised manuscript):

Sec. I.A of the SM:
(Original text)

$\mathsf{Comb}_{W_T,V_T,\dots,W_1,V_1}$ denotes the set of all $\tau \in \mathsf{Pos}_{W_T\otimes V_T\otimes\cdots\otimes W_1\otimes V_1}$ expressed in the form

$$\tau = I_{W_T} \otimes \tau^{(T)},$$
$$\mathrm{Tr}_{V_t}\, \tau^{(t)} = I_{W_{t-1}} \otimes \tau^{(t-1)}, \quad \forall 2 \le t \le T$$

with some $\tau^{(1)} \in \mathsf{Den}_{V_1}$ and $\tau^{(2)}, \dots, \tau^{(T)}$.

(Revised text)

$\mathsf{Comb}_{W_T,V_T,\dots,W_1,V_1}$ denotes the set of all $\tau \in \mathsf{Pos}_{W_T\otimes V_T\otimes\cdots\otimes W_1\otimes V_1}$ such that there exists $\{\tau^{(t)} \in \mathsf{Pos}_{W_t\otimes V_t\otimes\cdots\otimes W_1\otimes V_1}\}_{t=1}^{T-1}$ satisfying

$$\mathrm{Tr}_{W_t}\, \tau^{(t)} = I_{V_t} \otimes \tau^{(t-1)}, \quad \forall 1 \le t \le T,$$

where $\tau^{(0)} := 1$ and $\tau^{(T)} := \tau$. In particular, $\mathsf{Comb}_{\mathbb{C},W_T,V_T,\dots,W_1,V_1,\mathbb{C}}$ is denoted by $\mathsf{Comb}^*_{W_T,V_T,\dots,W_1,V_1}$, which is the set of all $\tau \in \mathsf{Pos}_{W_T\otimes V_T\otimes\cdots\otimes W_1\otimes V_1}$ such that there exist $\tau^{(1)} \in \mathsf{Den}_{V_1}$ and $\{\tau^{(t)} \in \mathsf{Pos}_{V_t\otimes W_{t-1}\otimes V_{t-1}\otimes\cdots\otimes W_1\otimes V_1}\}_{t=2}^{T}$ satisfying

$$\tau = I_{W_T} \otimes \tau^{(T)},$$
$$\mathrm{Tr}_{V_t}\, \tau^{(t)} = I_{W_{t-1}} \otimes \tau^{(t-1)}, \quad \forall 2 \le t \le T. \tag{S1}$$

We have

$$\langle \sigma, \tau \rangle = 1, \quad \forall \tau \in \mathsf{Comb}_{W_T,V_T,\dots,W_1,V_1},\ \sigma \in \mathsf{Comb}^*_{W_T,V_T,\dots,W_1,V_1}. \tag{S2}$$

5. p. 3/4, Prop. 2: "... is proportional to some quantum comb" the same remark as above. In addition, as I understand, here the "quantum comb" is not an element of $\mathsf{Comb}_{W_T,V_T,\dots,W_1,V_1}$ as before, but rather a quantum comb with inputs $V_1, \dots, V_T$ and outputs $W_1, \dots, W_T$, as are the processes $\mathcal{E}_m$. I guess some suitable notations for the "sets of combs" is needed.

Agreed. We have provided the definition of $\mathsf{Comb}_{W_T,V_T,\dots,W_1,V_1}$ as mentioned above and have added the following text:

Sec. I.A of the SM:

(Revised text)

We refer to a feasible solution, χ, to Problem (D) as proportional to some quantum comb if χ is expressed in the form $\chi = \lambda\tilde{\chi}$, with $\lambda \in \mathbf{R}_+$ and $\tilde{\chi} \in \mathsf{Comb}_{W_T,V_T,\dots,W_1,V_1}$.

6. p. 4, Col. 2: it seems that some related results for relations of robustness to discrimination of quantum channels were obtained in arxiv:1901.08127

As we noted in the introduction, the paper arxiv:1901.08127 (i.e., Ref. [39] in our original manuscript) claims that the robustness of a process can be seen as a measure of its advantage over all resource-free processes in some discrimination task, which is somewhat different from the task of discriminating quantum processes that our paper deals with. We have revised as follows:

Page 4, left column:

(Original text)

As already mentioned in the introduction, the robustness of $\mathcal{E}$ can be interpreted as a measure of the advantage of $\mathcal{E}$ over all the processes in $\mathcal{F}$ in some discrimination tasks (see Sec. VII of the SM).

(Revised text)

As already mentioned in the introduction, it has been shown that the robustness of $\mathcal{E}$ is characterized as a measure of the advantage of $\mathcal{E}$ over all the processes in $\mathcal{F}$ in some discrimination problem [33–40] (see also Sec. VII of the SM). Note that this problem is somewhat different from a process discrimination problem that this Letter deals with.

7. SM, p.3: in the diagram (S5) and below, some of the input/output systems are not labelled correctly (e.g. V should be V_1 or V_2 in (S5))

We apologize for the display error. This error may be due to a font issue. To avoid this, we have converted the fonts in the figure files to outlines.

8. SM, Sec. IIIB: perhaps it should be noted that this example is somewhat restricted: all strategies are non-adaptive, since the channel $\hat{\sigma}_2$ in Fig. S1 (a) can always be included in the final POVM $\hat{\Pi}$. A sequential strategy is obtained by a specific choice of the final measurement.

Although the referee correctly points out that the channel $\hat{\sigma}_2$ in Fig. S1(a) can always be included in the final POVM $\hat{\Pi}$, any strategy for discriminating $\{\hat{\mathcal{E}}_m = \hat{A}_{q_m} \otimes \hat{\rho}_m\}$ can be expressed in the form of Fig. S1(a). In this example, any adaptive strategy can be represented by a "collective" strategy (as well as any adaptive POVM can be represented by a collective POVM). We have removed this example since another simple example of a sequential strategy is presented in our revised manuscript.

9. SM, Sec. VI: note that a result related to Sec. B (especially Cor. S6) was already obtained in arxiv:1209.2329.

Although we cannot find a related result in arxiv:1209.2329, we have added some citations in Sec. VI of the SM, as already mentioned above.

Other changes

We have made other changes in green as follows:

(1)
We have revised as:

Page 1, left column:
(Original text)
In this paper
(Revised text)
In this Letter

(2)
To make our manuscript easier to understand, we have added the following text

Just after Eq. (S29):
(Revised text)
Note that η is a convex function.

and have changed as follows:

Just before Lemma S7:
(Original text)
Since the case $Z = \mathbb{0}$ is obvious, we may assume $Z \neq \mathbb{0}$.
(Revised text)
Since the case $\langle \varphi, Z \rangle \leq 0$ is obvious, we may assume $\langle \varphi, Z \rangle > 0$.

Thank you once again for your valuable comments and suggestions.