Linearity testing with entangled provers

Thomas Vidick
(Based on joint work with T. Ito)

We first recall the definition of the linearity test, and give a brief proof of its soundness for the case of classical players, in Section 1. In Section 2 we give a "beginner's introduction" to the quantum formalism used to describe entangled players. In Section 3 we state the "entangled-prover linearity test", taking the opportunity to explain some of the challenges that arise in the analysis of entangled games. Finally, in Section 3.2 we give a proof of the soundness of that test, emphasizing the important tools and techniques that it makes use of.

1 The linearity test

Blum, Luby and Rubinfeld's linearity test is a game played with three provers. The verifier's questions are elements of $\mathbb{F}_2^n$, for some integer n, and he expects answers in $\mathbb{F}_2$. The test is designed to verify that each prover answers the verifier's question according to a *linear* function $f : \mathbb{F}_2^n \to \mathbb{F}_2$, i.e. one that can be written as $f(x) = u \cdot x$ for some $u \in \mathbb{F}_2^n$. The test is as follows:

Linearity test. Perform either of the following with probability $1/2$ each:

1. *(Consistency.)* Select a random $x \in \mathbb{F}_2^n$, and send x to each of the provers. Accept if and only if all three provide the same answer.
2. *(Linearity.)* Select two random points $x, y \in \mathbb{F}_2^n$, and set $z := x + y$. Send x to the first prover, y to the second, and z to the third. Expect three answers $a, b, c \in \mathbb{F}_2$, and accept if and only if $a + b = c$.

This test has *perfect completeness*: if the provers answer according to the same linear function, then they succeed in the test with certainty. Note also that the marginal distribution on each prover's questions being the same in both parts of the test, there is no way for the provers to determine locally which part they are being tested on by the verifier: they must use the same strategy in both cases. BLR show the following.

Theorem 1 (BLR)**.** *Suppose that three* deterministic *provers succeed in the linearity test with probability $1 - \varepsilon$, and let $f_1, f_2, f_3 : \mathbb{F}_2^n \to \mathbb{F}_2$ be the functions describing their respective strategies. Then there is an $u \in \mathbb{F}_2^n$ such that, for each $i \in \{1, 2, 3\}$, $f_i(x) = u \cdot x$ for all but a fraction at most $8\,\varepsilon$ of $x \in \mathbb{F}_2^n$.*

Theorem 1 starts from the assumption that the three provers are deterministic. In the case of classical, randomized provers this always holds without loss of generality by "fixing the randomness": among all shared random strings that the provers may use, there is always one that gives them at least as good a success probability as on the average, and we may as well assume that they are using the corresponding deterministic strategies. In the case of entangled players this step will not be possible.

1.0.1 Analysis of the linearity test.

We refresh our reader's memory by giving a classic Fourier-analytic proof of Theorem 1. The proof for the case of entangled players will follow the same outline, and readers new to linearity testing may find it useful to familiarize themselves with the classical proof first.

By assumption the provers succeed with probability $1-\varepsilon$ in the linearity test, so they must succeed in the "consistency" and "linearity" parts of the test with probability at least $1-2\varepsilon$ each. From the "consistency" part we can infer that there is at most a 2ε fraction of $x \in \mathbb{F}_2^n$ such that $f_2(x) \neq f_1(x)$ or $f_3(x) \neq f_1(x)$; call them "bad" x. In the "linearity" part of the test, the probability that either of the two questions y or z is bad is at most 4ε. Hence we may as well assume that all three provers answer according to the same function $f := f_1$, in which case their success in the linearity part of the test should be at least $1-2\varepsilon-4\varepsilon = 1-6\,\varepsilon$.

Instead of working directly with f, it will be convenient to introduce the function $g : \mathbb{F}_2^n \to \{-1,1\}$ defined as $g(x) = (-1)^{f(x)}$ for every $x \in \mathbb{F}_2^n$. For any $u \in \mathbb{F}_2^n$, define the Fourier coefficient of g at u as $\widehat{g}(u) = \mathrm{E}_x (-1)^{u\cdot x} g(x)$. Parseval's identity states that

$$\sum_u \big(\widehat{g}(u)\big)^2 = \sum_u \mathrm{E}_{x,y}\Big[(-1)^{u\cdot(x+y)} g(x)g(y)\Big] = \mathrm{E}_x\left[g(x)^2\right] = 1.$$

It is not hard to see that for any $\varepsilon' > 0$ the provers having success probability at least $1-\varepsilon'$ in the "linearity" part of the test is equivalent to the following equation on g:

$$\frac{1}{2} + \frac{1}{2}\,\mathrm{E}_{x,y}\left[\,g(x)g(y)g(x+y)\,\right] \;\geq\; 1 - 2\,\varepsilon'. \tag{1}$$

The key claim in the proof of Theorem 1 is the following.

Claim 2. *Suppose deterministic provers applying the same function f succeed in the linearity test with probability at least $1-\varepsilon'$, and let $g = (-1)^f$. Then*

$$\sum_u \big(\widehat{g}(u)\big)^3 \;\geq\; 1 - 2\,\varepsilon'. \tag{2}$$

Proof. Expand

$$\begin{aligned}\sum_u \big(\widehat{g}(u)\big)^3 &= \sum_u \mathrm{E}_{x,y,z}\Big[\,(-1)^{u\cdot(x+y+z)} g(x)g(y)g(z)\,\Big] \\ &= \mathrm{E}_{x,y}\left[\,g(x)g(y)g(x+y)\,\right],\end{aligned}$$

since $\sum_u (-1)^{u\cdot(x+y+z)} = 0$ whenever $z \neq x+y$ in $\mathbb{F}_2^n$. The claim then follows directly from (1). □

As a consequence of the bound proven in Claim 2, one can see that g must have a large Fourier coefficient:

$$1 - 12\,\varepsilon \;\leq\; \sum_u \big(\widehat{g}(u)\big)^3 \;\leq\; \Big(\max_u \big|\widehat{g}(u)\big|\Big)\Big(\sum_u \big(\widehat{g}(u)\big)^2\Big) \;=\; \max_u \big|\widehat{g}(u)\big|$$

by Parseval's identity. Let u_0 be such that $\big|\widehat{g}(u_0)\big| \geq 1 - 12\varepsilon$. Then by definition

$$\big|\,\mathrm{E}_x\left[\,(-1)^{u_0\cdot x} g(x)\,\right]\big| \;=\; \big|\widehat{g}(u_0)\big| \;\geq\; 1 - 12\,\varepsilon.$$

Recalling the definition of g, this bound immediately implies that the functions f and $x \mapsto (u_0 \cdot x)$ can differ on at most a fraction 6ε of coordinates, proving Theorem 1.

Recapitulating, the proof of Theorem 1 has three main steps. First we argued that, since the provers had a high success probability in the "consistency" part of the test, we could assume that they were using the same function f to compute their answers. Then we proved Claim 2, which puts a lower bound on the sum of the third powers of the Fourier coefficients of any function that passes the "linearity" part of the test with high probability. Finally, from that bound we deduced that there must exist a *linear* function ℓ (the function $x \mapsto u_0 \cdot x$) that differs from the prover's function f on at most a small fraction of questions, implying that if we *replaced* the provers by ones answering according to ℓ rather than f then the verifier would see little difference — even if this replacement is done as part of a larger protocol in which the linearity test is only a subroutine (provided that, in the larger protocol, the marginal distribution of the questions to each of the "linear" provers is uniform in $\mathbb{F}_2^n$).

2 Entangled strategies

In this section we introduce some notation and concepts from quantum information theory that are needed to describe *what an entangled strategy is*. Even though the linearity test uses three provers, for clarity we focus on the setting of two provers; everything that we say here has a natural extension to the case of more provers.

The reader may already be familiar with *pure* quantum states, which are described by unit vectors $|\Psi\rangle \in \mathbb{C}^d$,[1] for some dimension d which is usually a power of 2 (if the system is represented by qubits, then the number of qubits is $\log_2 d$). The reader may also have prior experience with orthogonal measurements: a measurement is described by the choice of an orthonormal basis $\{|e_i\rangle,\ i = 1, \ldots, d\}$ for the space $\mathbb{C}^d$. Upon measuring in that basis the i-th outcome is observed with probability $|\langle e_i|\Psi\rangle|^2$, while the state of the system is projected to its *post-measurement state* $|e_i\rangle$.

In the remainder of this section we introduce generalizations of these two fundamental concepts that will be necessary to describe entangled-prover strategies.

Density matrices. While pure states provide a convenient way to describe isolated systems, such as the joint state of *all* provers in a multiplayer game, we will sometimes need to work with more general, *non-isolated* systems, such as the first prover's subsystem alone. In full generality, a quantum system is described by a *probabilistic mixture* of pure states $(p_i, |\Psi_i\rangle)$. This mixture is represented by a corresponding *density matrix* $\rho = \sum_i p_i |\Psi_i\rangle\langle\Psi_i|$.[2] Since the p_i are a distribution, and the $|\Psi_i\rangle$ are normalized, ρ is a positive matrix with trace 1.

How does one compute the density matrix representing the first prover's subsystem, given that both provers are jointly in the pure state $|\Psi\rangle$? Suppose the second prover was to measure

[1] Recall that Dirac's very convenient "ket" notation indicates how we think of the vector Ψ: a ket $|\Psi\rangle$ is a column vector, while a bra $\langle\Psi|$ is a line vector: $\langle\Psi| = |\Psi\rangle^\dagger$.

[2] The careful reader may have noticed that different distributions can give rise to the same density matrix. But quantum mechanics states that the formalism of density matrices *does* give a full description of a given subsystem's state: different distributions giving rise to the same density matrix are *indistinguishable*.

his system using an arbitrary fixed orthogonal measurement, described by an orthonormal basis $\{|e_i\rangle, i = 1, \ldots, d\}$. Given such a basis, one may always express $|\Psi\rangle$ as

$$|\Psi\rangle \;=\; \sum_i \sqrt{p_i}|\Psi_i\rangle|e_i\rangle,$$

where the p_i are a distribution, and the $|\Psi_i\rangle$ normalized (but not necessarily orthogonal). Hence once the second prover measures, he obtains outcome i with probability p_i, and the *whole* system gets projected in the state $|\Psi_i\rangle|e_i\rangle$, implying that the first prover is then in state $|\Psi_i\rangle$. But of course, by the no-signaling principle, the state of the first prover should be *independent* of whether the second prover makes the measurement or not — hence the first prover can be accurately described as being in state $|\Psi_i\rangle$ with probability p_i, i.e. the state of his subsystem is represented by the density matrix

$$\rho_1 \;=\; \sum_i p_i|\Psi_i\rangle\langle\Psi_i|.$$

The skeptical reader should check that this matrix is *independent* of our choice of basis for the measurement on the second prover's subsystem, as it should be.

The operation of finding a description of a subsystem given a description of the whole is called "tracing out", and it will be important for us. It can be performed in the same way as described above even starting from a density matrix representation of the whole system: if both provers are in a joint state σ, then the first prover's reduced state, obtained by tracing out the second prover, is given by

$$\rho_1 \;=\; \mathrm{Tr}_2(\sigma) \;=\; \sum_i (\mathrm{Id} \otimes \langle e_i|)\, \sigma\, (\mathrm{Id} \otimes |e_i\rangle). \tag{3}$$

Generalized measurements. Just as we generalized our notion of quantum state from pure states to density matrices, we'll need to generalize measurements to allow "high-rank" measurements, which have fewer possible outcomes than the system's dimension. While we used to think of a measurement as projecting a state on a *basis*, a *projective measurement* corresponds to projecting a state on a *subspace*. Hence a projective measurement is given by a set of *orthogonal projectors* $A_1, \ldots, A_k$ such that $A_1 + \ldots + A_k = \mathrm{Id}$. The "measurement rule" is that when $\{A_i\}$ is performed on the quantum state ρ, outcome i will be observed with probability $\mathrm{Tr}(A_i\rho)$, the "overlap" of the density ρ on A_i. Using a decomposition $\rho = \sum_j p_j|\Psi_k\rangle\langle\Psi_j|$, this probability also reads $\mathrm{Tr}(A_i\rho) = \sum_j p_j\langle\Psi_j|A_i|\Psi_j\rangle$, which is the average, according to the distribution $\{p_j\}$, of the overlap of the state $|\Psi_j\rangle$ on the subspace on which A_i projects.

Entangled strategies. We are ready to put our newly-learned notions from quantum information theory to practice in describing an entangled-prover strategy in a multiplayer game. Consider a two-prover game, in which the first (resp. second) prover is sent a question x (resp. y), and has to provide an answer a (resp. b). For simplicity, assume that the provers' answers are bits, as will be the case in the linearity test. Let σ be the density matrix describing the joint state of the two provers. Upon receiving his question x, the first prover measures his subsystem using a two-outcome measurement, described by a pair of orthogonal projectors A_x^0 and A_x^1 such that $A_x^0 + A_x^1 = \mathrm{Id}$; this measurement can also be succinctly described through the corresponding *observable* $A_x = A_x^0 - A_x^1$.[3] As a result, the prover obtains an outcome $a \in \{0, 1\}$, and he sends

[3] Going from the pair (A_x^0, A_x^1) to A_x is the quantum analogue of going from a $\{0,1\}$-valued function f to the $\{-1,1\}$-valued function $g = (-1)^f$.

it back to the verifier as his answer.[4] Similarly, upon receiving y the second prover makes the measurement $\{B_y^0, B_y^1\}$ on his share of σ, and sends his outcome back to the verifier.

In order to complete our picture we need to give the rule describing the *joint* probability $p(a,b|x,y)$ that the two provers answer the questions (x,y) with (a,b). The way to compute this is to imagine the two provers as making a single, joint measurement, described by four projectors obtained by taking the tensor product of both prover's measurement operators: $A_x^0 \otimes B_y^0$, $A_x^0 \otimes B_x^1$, etc. This leads to the following definition:

$$p(a,b|x,y) \;:=\; \mathrm{Tr}\big(\big(A_x^a \otimes B_y^b\big)\,\sigma\big) \;=\; \sum_{(r,s),(r',s')} \big(A_x^a\big)_{r,r'}\,\big(B_y^b\big)_{s,s'}\,\sigma_{(r,s),(r',s')}.$$

To make sure one understands this equation, one can think of how one would describe classical provers using shared randomness in this formalism. In that case, σ would be a diagonal matrix representing the prior probability $q(r)$ of each shared random string $r \in [d]$: σ's rows can be indexed by pairs $(r,s) \in [d]^2$, and it would have a coefficient $q(r)$ in the diagonal entry corresponding to the (r,r) row; all other coefficients (diagonal and otherwise) would be 0.[5] Let f_r (resp. g_r) be the function that the first (resp. second) prover would use if the shared random string was r. Then for any question x and answer a the prover's measurement matrix A_x^a (resp. B_y^b) would also be diagonal, and contain a 1 in each diagonal entry (r,r) such that $f_r(x) = a$ (resp. $g_r(y) = b$). One can now check that with this setup

$$\mathrm{Tr}\big(\big(A_x^a \otimes B_y^b\big)\,\sigma\big) \;=\; \sum_{r\in[d]} q(r)\,1_{f_r(x)=a}\,1_{g_r(y)=b}\,,$$

as should be the case. Of course, general entangled strategies will differ from the one constructed above in a key aspect: the prover's measurements corresponding to different questions will not in general commute, hence they will not be diagonal in the same basis.

3 Linearity testing of entangled provers

The first difficulty in extending the linearity test to the case of entangled provers is to determine what its precise statement should be. Indeed, in the presence of entanglement (as in the presence of shared randomness between the provers), there is no hope of extracting a *single* linear function from the prover's strategy, as was done in Theorem 1. What does it even mean for entangled provers to be *linear*, if their strategy cannot be tied to a single function? The following informal theorem suggests an answer:

Theorem 3 (Entangled-prover linearity test, informal)**.** *Suppose that three entangled provers, using a strategy described by measurements $\{A_x^a\}$, $\{B_y^b\}$, $\{C_z^c\}$ and a shared entangled state σ, succeed in the linearity test with probability $1-\varepsilon$. For $u \in \mathbb{F}_2^n$, let*

$$\widehat{A}_u = \mathrm{E}_x (-1)^{x\cdot u}\big(A_x^0 - A_x^1\big)$$

[4]The measurement formalism that we have described also encompasses seemingly more complex strategies, in which the prover would make a measurement, then do some classical processing on the outcomes, maybe another measurement, etc. — all these operations are taken into account by the projectors A_x^0 and A_x^1, which describe his final answer.

[5]As a side remark, note that choosing $\sigma = d^{-2}\mathrm{Id}$ would not correspond to shared randomness — indeed, that state has a tensor product form $(d^{-1}\mathrm{Id}_d) \otimes (d^{-1}\mathrm{Id}_d)$, so that it corresponds to a setting where there would be no correlations between the provers at all.

be the matrix Fourier coefficient associated to the first prover's strategy. For every u, define

$$M^u := \left(\widehat{A}_u\right)^2. \tag{4}$$

Then $\{M^u\}$ is a proper quantum measurement. Moreover, the original prover's strategy is almost indistinguishable *from that of three "oblivious" provers whom would behave as follows:*

1. *Measure their share of the entangled state using $\{M^u\}$, each obtaining an outcome u_i, for $i \in \{1,2,3\}$,*
2. *Upon receiving the prover's question x, answer with $u_i \cdot x$.*

The key point in Theorem 3 is Eq. (4), which, rather than defining a *single* linear function to which the provers would be close, constructs a *global* measurement, *independent* of the prover's questions, and then claims that this measurement faithfully reproduces the original provers' strategy. It does this by considering new, "oblivious" provers, who *first* measure according to the constructed measurement, obtaining the label u of a linear function, and *then* apply that function to their question. Hence the measurement might as well have been made before the start the protocol: the oblivious provers are effectively reduced to being classical, using shared randomness to determine the linear function they will use.

The definition of M^u has a simple interpretation in the special case where the provers are classical, but may use shared randomness. It corresponds to the following definition of an "oblivious" strategy: the prover simply looks at his random string r, pointing to a function f_r according to which the "original" prover would answer his questions. Instead, the oblivious prover samples a *function* $\ell : x \mapsto u \cdot x$, where u is chosen according to the distribution suggested by f_r's Fourier spectrum.[6] When his question x arrives, he answers with $\ell(x) = u \cdot x$.

Why is this a good strategy? Recall that in the soundness proof for the classical, deterministic case we had proved Eq. (12), which stated that the Fourier coefficients of $g = (-1)^f$ were sharply concentrated. This justified our "rounding" of the provers' strategy to the linear function corresponding to the largest Fourier coefficient. In the case of a randomized strategy it will still be the case that most functions f_r have a large Fourier coefficient. Even though *which* coefficient might depend on the random string r, the "linear" strategy defined above intrinsically accounts for that possibility, and it is not hard to see that it will indeed faithfully reproduce the original prover's actions.

Returning to Theorem 3, what is maybe more surprising is that essentially the *same* definition of a new strategy will also work in the case where the original provers use an arbitrary entangled strategy! Before showing that this is indeed the case, we should make precise what we mean by two entangled-prover strategies being "almost indistinguishable". In the classical case this was taken to mean "the new provers' strategy differs from the original one in a fraction at most $O(\varepsilon)$ of questions", and we give a definition formalizing a similar intuition in the case of entangled provers in the next section. We then give a more precise statement of Theorem 3, as well as its proof, in Section 3.2.

[6] Letting $g_r = (-1)^{f_r}$, this is the distribution induced by the $|\widehat{g_r}(u)|^2$ (Parseval's identity shows that this is indeed a distribution).

3.1 Measuring the distance between provers

In the classical analysis of the linearity test, from three functions (f, g, h) having high success in the test one constructs a linear function ℓ such that f, g, h differ from ℓ in a small fraction of points $x \in \mathbb{F}_2^n$. This ensures that the linearity test can be performed as part of a bigger protocol, possibly involving a larger number of provers: its goal is to constrain a subset of the provers to answer according to a linear function. Provided in the larger protocol the verifier's question to each prover is distributed as in the linearity test (uniformly in $\mathbb{F}_2^n$), replacing them with their linear approximation will not affect the provers' success probability in the overall protocol much, while potentially making the analysis much simpler.

We would like to achieve the same result in the case of entangled provers. In order to make a meaningful statement, we need an appropriate measure of what it means for two distinct strategies of a single prover to be "indistinguishable". Any measure that we use should be *strong enough* that a small distance implies that the type of "prover replacement" described above does not affect the provers' success probability in the overall protocol too much, while still being *weak enough* that one is able to prove bounds on that distance simply from the fact that the provers have a high success in the linearity test.[7]

Consider two distinct measurements, $\{A_x^a\}$ and $\{\tilde{A}_x^a\}$, that the first prover could apply on his share of the entangled state, which we'll take to be a pure state $|\Psi\rangle$ for simplicity. Fix a question $x \in \mathbb{F}_2^n$. Quantum mechanics dictates that, once the first prover has applied the measurement $\{A_x^0, A_x^1\}$ on his share of $|\Psi\rangle$ and obtained an outcome $a \in \{0,1\}$, the entangled state gets projected to $|\Psi'\rangle = \big(A_x^a \otimes \mathrm{Id} \otimes \mathrm{Id}\big)|\Psi\rangle$, where the identity terms are meant to indicate that the other two provers have not performed any action yet.[8] This means that, for this fixed x, the global states resulting from the first prover measuring using either $\{A_x^0, A_x^1\}$ or $\{\tilde{A}_x^0, \tilde{A}_x^1\}$, and conditioning on either answer being obtained, will be close if the following quantity is small:

$$\begin{aligned}\sum_a \big\|(A_x^a \otimes \mathrm{Id} \otimes \mathrm{Id})|\Psi\rangle - (\tilde{A}_x^a \otimes \mathrm{Id} \otimes \mathrm{Id})|\Psi\rangle\big\|^2 &= \sum_a \langle\Psi|\big(\big(A_x^a - \tilde{A}_x^a\big)^2 \otimes \mathrm{Id} \otimes \mathrm{Id}\big)|\Psi\rangle \\ &= \sum_a \mathrm{Tr}\big(\big(A_x^a - \tilde{A}_x^a\big)^2 \rho\big), \end{aligned} \tag{5}$$

where ρ is the reduced density of the state $|\Psi\rangle$ on the first prover's register (cf. Eq. (3) for a definition). The magic of this last equation is that the systems corresponding to the second and third provers have disappeared; this was made possible by the fact that they act on subsystems separated from the first prover's. Nevertheless, we have shown that if the quantity in (5) small, then the provers' shared state is almost the same *after* the first prover has measured his subsystem using *either* A or $\tilde{A}$, and obtained an answer a. Hence whatever happens in the remainder of the protocol, the probabilities that arise from other provers' measurements will be essentially the same irrespective of which of A or $\tilde{A}$ the first prover applied.

Incorporating the choice of the question x, we are ready to define our distance measure on strategies: we'll say that the strategies described by $\{A_x^a\}$ and $\{\tilde{A}_x^a\}$, together with the entangled

[7] This implies for instance that the operator norm on the provers' measurements would *not* be appropriate, as success in the test does not put constraints directly on the provers' measurements themselves, but only on their probability of obtaining certain outcomes *when applied on the entangled state* σ.

[8] Observe that the squared norm of the state after the prover's measurement is $\|(A_x^a \otimes \mathrm{Id} \otimes \mathrm{Id})|\Psi\rangle\|^2 = \mathrm{Tr}\big(((A_x^a)^2 \otimes \mathrm{Id} \otimes \mathrm{Id})(|\Psi\rangle\langle\Psi|)\big) = \langle\Psi|(A_x^a \otimes \mathrm{Id} \otimes \mathrm{Id}|\Psi\rangle$, since A_x^a is a projector: it is the probability of obtaining outcome a when measuring Ψ with $\{A_x^a\}$. As such, the post-measurement state is not normalized.

state σ with reduced density ρ on the first prover's subsystem, are δ-close if

$$d_\rho(A,B) \;:=\; \Big(\mathrm{E}_x \,\mathrm{Tr}\big(\big(A_x^a - \tilde{A}_x^a\big)^2 \rho\big)\Big)^{1/2} \;\leq\; \delta.$$

It is not too hard to see that d_ρ is indeed a distance measure (it is non-negative and satisfies the triangle inequality), and that if $\{A_x^a\}$ and $\{\tilde{A}_x^a\}$ are measurements then it is bounded between 0 and $\sqrt{2}$.

We argued that the distance measure d_ρ was strong enough to ensure that switching from one of two strategies close in that distance to the other would have a small effect on the overall performance of the provers in any entangled game. The analysis of the linearity test in the next section will show that it is also weak enough, in the sense described above: one can place bounds on d_ρ from the seemingly weak assumption that provers have a high success probability in a certain well-chosen test that the verifier plays with them.

3.2 The quantum analysis

Before stating our theorem, we observe that, given that the linearity test is symmetric under permutation of the three provers, one may assume without loss of generality that the following hold of the provers' strategies:

1. All three provers are applying the same set of measurements $\{A_x^a\}$,
2. The provers' shared state σ is invariant with respect to any permutation of its three subsystems.

This observation follows from a symmetrization argument (omitted). In practice, it means that the probabilities $p(a,b,c|x,y,z)$ do not depend on which prover applied the measurement corresponding to each question:

$$\mathrm{Tr}\big(\big(A_x^a \otimes A_y^b \otimes A_z^c\big)\,\sigma\big) \;=\; \mathrm{Tr}\big(\big(A_y^b \otimes A_z^c \otimes A_x^a\big)\,\sigma\big) \;=\; \ldots \;=\; \mathrm{Tr}\big(\big(A_z^c \otimes A_y^b \otimes A_x^a\big),$$

a convenient property we will make frequent use of.

The following theorem, a precise reformulation of Theorem 3, is the main result of this chapter.

Theorem 4 (Linearity test with entangled provers)**.** *Suppose three entangled provers succeed in the* linearity test *with probability at least* $1-\varepsilon$ *using a symmetric strategy*[9] *with measurements* $\{A_x^a\}$ *and entangled state* σ*. Then there exists a measurement* $\{B^u\}$*, independent of* x *and indexed by outcomes* $u \in \mathbb{F}_2^n$*, such that if we let* $B_x^a := \sum_{u:\,u\cdot x=a} B^u$ *then*

$$\big(d_\rho(A,B)\big)^2 \;=\; \mathrm{E}_x \sum_a \mathrm{Tr}\Big(\Big(A_x^a - \sum_{u:\,u\cdot x=a} B^u\Big)^2 \rho\Big) \;\leq\; 8\sqrt{\varepsilon}. \tag{6}$$

[9] Even though this symmetry implies that all three provers are using the same set of measurements $\{A_x^a\}$, and their entangled state is invariant with respect to any permutation of the provers' subsystems, this assumption alone is not sufficient to guarantee that they will succeed with certainty in the "consistency" part of the linearity test. Indeed, the assumption of symmetry does not for instance preclude randomized strategies in which the random strings would be triples of bits (a,b,c) and the i-th prover would answer with the bit contained in the i-th position.

As discussed in Section 3.1, through Eq. (6) the theorem asserts that "oblivious" provers, who would first measure according to $\{B^u\}$, obtain an outcome u, and then answer their question x with $u \cdot x$, are almost *indistinguishable* from the original provers, in the strong sense that one may *replace* the original provers by the new ones while only affecting their success probability in *any* overall protocol,[10] of which the linearity test might only be a subroutine, by $O\big(\sqrt{\varepsilon}\big)$.

We now turn to the proof of Theorem 4. Following the classical proof given in Section 1, we will use Fourier analysis directly on the prover's observables $A_x := A_x^0 - A_x^1$: for every $u \in \mathbb{F}_2^n$ one may define

$$\widehat{A}_u \;:=\; \mathrm{E}_x \left[\, (-1)^{u\cdot x} A_x \,\right].$$

In general, $\widehat{A}_u$ is Hermitian, with eigenvalues in $[-1,1]$. Indeed, a variant of Parseval's identity also holds in this setting:

$$\sum_u \big(\widehat{A}_u\big)^2 \;=\; \sum_u \mathrm{E}_{x,y} \left[\, (-1)^{u\cdot(x+y)} A_x A_y \,\right] \;=\; \mathrm{E}_x \left[\, A_x^2 \,\right] \;=\; \mathrm{Id}, \tag{7}$$

where the last equality uses that the A_x are observables.

As in the classical case (cf. (1)), the following two equations re-formulate the fact that the provers must succeed in each of the "consistency" and the "linearity" parts of the tests with probability at least $1 - 2\varepsilon$.[11] (Recall that we use $p(a,b,c|x,y,z)$ to denote the probability that the provers answer (a,b,c) to questions (x,y,z).)

$$\mathrm{E}_x \sum_{a,b} p(a,a,b|x,x,x) \;=\; \mathrm{E}_x \,\mathrm{Tr}\big(\big(A_x \otimes A_x \otimes \mathrm{Id}\big)\,\sigma\big) \;\geq\; 1 - 4\varepsilon, \tag{8}$$

$$\mathrm{E}_{x,y} \sum_{a,b} p(a,b,a+b|x,y,z) \;=\; \mathrm{E}_{x,y} \,\mathrm{Tr}\big(\big(A_x \otimes A_y \otimes A_{x+y}\big)\,\sigma\big) \;\geq\; 1 - 4\varepsilon. \tag{9}$$

The proof of both equations is exactly similar to the classical case, and we postpone it until Section 3.3. Still in complete analogy with the classical setting, one can translate Eqs. (8) and (9) into conditions on the Fourier coefficients $\widehat{A}_u$ that we associated with the observables A_x:

$$\sum_u \mathrm{Tr}\big(\big(\widehat{A}_u \otimes \widehat{A}_u \otimes \mathrm{Id}\big)\,\sigma\big) \geq 1 - 4\varepsilon, \tag{10}$$

$$\sum_u \mathrm{Tr}\big(\big(\widehat{A}_u \otimes \widehat{A}_u \otimes \widehat{A}_u\big)\,\sigma\big) \geq 1 - 4\varepsilon. \tag{11}$$

The proof of both equations follows from the definition of $\widehat{A}_u$ and Eqs. (8) and (9). In the classical setting, (11) would already be a proof of Claim 2: since the "entangled state" in that case is one-dimensional, the tensor product becomes a product, and the sum of the third powers of the Fourier coefficients appears.

In the presence of entanglement, however, it seems that we are stuck: each prover acts on his own subsystem only; how could one bring different subsystems together? The following claim shows that this is, in fact, possible as a consequence of the "consistency" part of the linearity test:

[10] As we already mentioned, this is only possible provided the marginal distribution on the "linear" provers' questions in the overall protocol is as it is in the linearity test, uniform over $\mathbb{F}_2^n$.

[11] Note that, for the first equation, we write the probability of the provers succeeding as if the verifier only checked that two out of the three answers were consistent; a weaker but sufficient requirement for our purposes.

a measurement $\{A_x^a\}$ performed on the first subsystem can be "replaced" by the same measurement performed on the second subsystem, without affecting the provers' shared state, after that measurement has been performed, by much.[12] Note that this property is distinct from that of the strategy's symmetry: while symmetry dictates that the *distribution* of outcomes should be the same irrespective of which measurement is performed on which subsystem, here we are showing that the whole post-measurement state is almost the same in both cases. Claim 5 provides an important tool to manipulate entangled strategies, and we will subsequently see how it lets us deduce an analogue of Claim 2 from (11).

Claim 5. *Suppose the provers succeed in the consistency test with probability $1-\varepsilon$. Then*

$$\sum_u \mathrm{Tr}\big((\widehat{A}_u \otimes \mathrm{Id} \otimes \mathrm{Id} - \mathrm{Id} \otimes \widehat{A}_u \otimes \mathrm{Id})^2 \sigma\big) \;\leq\; 8\varepsilon,$$

and the same holds under arbitrary permutation of the registers.

Proof. It suffices to expand the expression on the left-hand side as

$$\begin{aligned}
&\sum_u \mathrm{Tr}\big((\widehat{A}_u \otimes \mathrm{Id} \otimes \mathrm{Id} - \mathrm{Id} \otimes \widehat{A}_u \otimes \mathrm{Id})^2 \sigma\big) \\
&\qquad = \sum_u \Big(\mathrm{Tr}\big((\widehat{A}_u^2 \otimes \mathrm{Id} \otimes \mathrm{Id})\,\sigma\big) + \mathrm{Tr}\big((\mathrm{Id} \otimes \widehat{A}_u^2 \otimes \mathrm{Id})\,\sigma\big) - 2\mathrm{Tr}\big((\widehat{A}_u \otimes \widehat{A}_u \otimes \mathrm{Id})\,\sigma\big)\Big) \\
&\qquad \leq 2 - 2(1-4\varepsilon),
\end{aligned}$$

where we used Parseval's identity (7) to compute the first two terms, and (10) to lower-bound the last term. □

Writing

$$\begin{aligned}
\widehat{A}_u^3 \otimes \mathrm{Id} \otimes \mathrm{Id} - \widehat{A}_u \otimes \widehat{A}_u \otimes \widehat{A}_u = {}& (\widehat{A}_u^2 \otimes \mathrm{Id} \otimes \mathrm{Id})(\widehat{A}_u \otimes \mathrm{Id} \otimes \mathrm{Id} - \mathrm{Id} \otimes \widehat{A}_u \otimes \mathrm{Id}) \\
&+ (\widehat{A}_u \otimes \widehat{A}_u \otimes \mathrm{Id}) \cdot (\widehat{A}_u \otimes \mathrm{Id} \otimes \mathrm{Id} - \mathrm{Id} \otimes \mathrm{Id} \otimes \widehat{A}_u),
\end{aligned}$$

Claim 5 together with Eq. (11) and the Cauchy-Schwarz inequality let us obtain the following quantum analogue of Claim 2.

Claim 6. *The following holds*

$$\sum_u \mathrm{Tr}\big(\widehat{A}_u^3 \rho\big) \geq 1 - 8\sqrt{\varepsilon}. \tag{12}$$

While in the classical setting Eq. (2), together with Parseval's identity, immediately implied the existence of a single large Fourier coefficient for g, in the presence of entanglement Eq. (12) does not imply such a strong statement. Indeed, even in the case of provers using shared randomness, (12) only states that *for most random strings* there should be a corresponding large Fourier coefficient — but which coefficient it is may well depend on the random string itself.

[12] The claim proves this statement for the Fourier operator $\widehat{A}_u$, but a similar bound can be proven directly for the observable A_x itself.

Instead, as described at the beginning of Section 3 we define a measurement $\{M^u\}$ as

$$M^u := \big(\widehat{A}_u\big)^2.$$

Each M^u is non-negative, and Parseval's identity shows that $\sum_u M^u = \mathrm{Id}$: the M^u form a proper quantum measurement. To show that they satisfy the requirement of the theorem, let $C_x = A_x - \sum_u (-1)^{u\cdot x} M^u$, and observe that the Fourier coefficient of C_x at u is

$$\widehat{C}_u \;=\; \widehat{A}_u - \mathrm{E}_x \sum_u (-1)^{u\cdot x} M^u \;=\; \widehat{A}_u - \big(\widehat{A}_u\big)^2,$$

so that by Parseval's identity

$$\mathrm{E}_x\, C_x^2 \;=\; \sum_u \big(\widehat{C}_u\big)^2 = \sum_u \widehat{A}_u^2 \big(\mathrm{Id} - \widehat{A}_u\big)^2 \;\le\; 2\sum_u \widehat{A}_u^2 \big(\mathrm{Id} - \widehat{A}_u\big) \;=\; 2\sum_u \big(\mathrm{Id} - \widehat{A}_u^3\big),$$

again as a consequence of Parseval's identity. Hence

$$\begin{aligned}\mathrm{E}_x \,\mathrm{Tr}\Big(\Big(A_x - \sum_u (-1)^{u\cdot x} M^u\Big)^2 \rho\Big) &\le 2 - 2\sum_u \mathrm{Tr}\big(\widehat{A}_u^3\big) \\ &\le 16\sqrt{\varepsilon}\end{aligned}$$

by (12). This proves the theorem since

$$A_x - \sum_u (-1)^{u\cdot x} M^u = \frac{1}{2}\big(\mathrm{Id} + A_x^0\big) - \frac{1}{2}\Big(\mathrm{Id} + \sum_{u\cdot x = 0} M^u\Big) = \frac{1}{2}\Big(A_x^0 - \sum_{u\cdot x = 0} M^u\Big),$$

and a similar equation holds after replacing $'0'$ by $'1'$.

3.3 Omitted proofs

We give details of the proofs that were omitted from our analysis of the linearity test in the presence of entangled strategies.

Proof of Eqs. (8) and (9). The player's success probability in the consistency test is

$$\mathrm{E}_{x,y} \sum_a \frac{1}{3}\Big(\mathrm{Tr}\big(\big(A_x^a \otimes A_x^a \otimes \mathrm{Id}\big)\,\sigma\big) + \mathrm{Tr}\big(\big(A_x^a \otimes \mathrm{Id} \otimes A_x^a\big)\,\sigma\big) + \mathrm{Tr}\big(\big(\mathrm{Id} \otimes A_x^a \otimes A_x^a\big)\,\sigma\big)\Big) \ge 1 - 2\varepsilon \quad (13)$$

By symmetry, all three terms inside the summation are the same. By definition of $A_x = A_x^0 - A_x^1$, we have

$$\mathrm{Tr}\big(\big(A_x \otimes A_x \otimes \mathrm{Id}\big)\,\sigma\big) = \sum_{a=a'} \mathrm{Tr}\big(\big(A_x^a \otimes A_x^{a'} \otimes \mathrm{Id}\big)\,\sigma\big) - \sum_{a\neq a'} \mathrm{Tr}\big(\big(A_x^a \otimes A_x^{a'} \otimes \mathrm{Id}\big)\,\sigma\big).$$

Using that $A_x^0 + A_x^1 = \mathrm{Id}$, the sum (instead of the difference) of the two terms on the right-hand side is 1. Combining this observation with (13) proves (8), and (9) is proved in a similar way. □

Proof of Claim 6.

$$
\begin{aligned}
&\sum_u \mathrm{Tr}\big((\widehat{A}_u^3 \otimes \mathrm{Id} \otimes \mathrm{Id} - \widehat{A}_u \otimes \widehat{A}_u \otimes \widehat{A}_u)\,\sigma\big) \\
&= \sum_u \mathrm{Tr}\big((\widehat{A}_u^3 \otimes \mathrm{Id} \otimes \mathrm{Id} - \widehat{A}_u^2 \otimes \widehat{A}_u \otimes \mathrm{Id})\,\sigma\big) + \sum_u \mathrm{Tr}\big((\widehat{A}_u^2 \otimes \widehat{A}_u \otimes \mathrm{Id} - \widehat{A}_u \otimes \widehat{A}_u \otimes \widehat{A}_u)\,\sigma\big) \\
&= \sum_u \mathrm{Tr}\big((\widehat{A}_u^2 \otimes \mathrm{Id} \otimes \mathrm{Id}) \cdot (\widehat{A}_u \otimes \mathrm{Id} \otimes \mathrm{Id} - \mathrm{Id} \otimes \widehat{A}_u \otimes \mathrm{Id})\,\sigma\big) \\
&\qquad + \sum_u \mathrm{Tr}\big((\widehat{A}_u \otimes \widehat{A}_u \otimes \mathrm{Id}) \cdot (\widehat{A}_u \otimes \mathrm{Id} \otimes \mathrm{Id} - \mathrm{Id} \otimes \mathrm{Id} \otimes \widehat{A}_u)\,\sigma\big) \\
&\leq \Big(\sum_u \mathrm{Tr}\big((\widehat{A}_u \otimes \mathrm{Id} \otimes \mathrm{Id} - \mathrm{Id} \otimes \widehat{A}_u \otimes \mathrm{Id})^2 \sigma\big)\Big)^{1/2} \Big(\sum_u \mathrm{Tr}\big(\widehat{A}_u^4\,\rho\big) + \mathrm{Tr}\big((\widehat{A}_u^2 \otimes \widehat{A}_u^2 \otimes \mathrm{Id})\,\sigma\big)\Big)^{1/2} \\
&\leq \sqrt{8\varepsilon} \cdot \sqrt{2},
\end{aligned}
$$

where the first inequality is the Cauchy-Schwarz inequality, and the last uses Parseval's identity $\sum_u \hat{A}_u^2 = \mathrm{Id}$. Eq. (11) lets us conclude the proof. □