6. Parallel Repetition of MIP(2,1) Systems

Clemens Gröpl, Martin Skutella

6.1 Prologue

Two men standing trial for a joint crime try to convince a judge of their innocence. Since the judge does not want to spend too much time on verifying their joint alibi, he selects a pair of questions at random. Then each of the suspects gets only one of the questions and gives an answer. Based on the coincidence of the two answers, the judge decides on the guilt or innocence of the men.

The judge is convinced of the fairness of this procedure, because once the questioning starts the suspects can neither talk to each other as they are kept in separate rooms nor anticipate the randomized questions he may ask them. If the two guys are innocent an optimal strategy is to convince the judge by telling the truth. However, if they have actually jointly committed the crime, their answers will agree with probability at most ε, regardless of the strategy they use.

To reduce the error ε the judge decides to repeat the random questioning k times and to declare the two men innocent if all k pairs of answers coincide. This obviously reduces the error to ε^k. So he randomly chooses k pairs of questions and writes them on two lists, one for each of the accused.

The judge does not want to ask the questions one after another but, to make things easier, hands out the two lists to the accused and asks them to write down their answers. Of course he does not allow them to communicate with each other while answering the questions. Waiting for the answers, he once more thinks about the error probability. Can the two men benefit from knowing all the questions in advance?

6.2 Introduction

In mathematical terms, the situation described in the prologue fits into the context of *two-prover one-round proof systems*. In this chapter, we give an introduction into basic definitions and characteristics of two-prover one-round proof systems and the complexity class MIP$(2,1)$. (The letters MIP stand for *m*ulti-prover *i*nteractive *p*roof system). Furthermore, we illustrate the central ideas of the proof of the Parallel Repetition Theorem. Our approach is mainly based on the papers of Raz [Raz95, Raz97] and Feige [Fei95].

Multi-prover interactive proofs were introduced by Ben-Or, Goldwasser, Kilian, and Wigderson in [BGKW88], motivated by the necessity to find new foundations to the design of secure cryptography. The story told in our prologue is an adaption of an example given in [BGKW88].

The chapter is organized as follows: We formally introduce MIP$(2,1)$ proof systems in Section 6.3. Then, in Section 6.4, we explain the problem of error reduction by parallel repetition of MIP$(2,1)$ proof systems and state Raz' Parallel Repetition Theorem. Its proof is based on the investigation of coordinate games, which are introduced in Section 6.5. Sections 6.6 and 6.7 give an overview of the proof of the Parallel Repetition Theorem.

6.3 Two-Prover One-Round Proof Systems

In a MIP$(2,1)$ proof system G, two computationally unbounded *provers* P_1, P_2 try to convince a probabilistic polynomial time *verifier* that a certain input I belongs to a pre-specified language L. The proof system has one round: Based on I and a random string τ, the verifier randomly generates a pair of questions (x,y) from a pool $X \times Y$ which depends on the input I and sends x to prover P_1 and y to prover P_2. The first prover responds by sending $u = u(x) \in U$ to the verifier, the second prover responds with $v = v(x) \in V$. Here U and V are pre-specified sets of possible answers to the questions in X resp. Y. Since the verifier is polynomially bounded, only a polynomial number of bits can be exchanged. In particular, the size of all possible questions and answers must be polynomially bounded in the size of the input I.

The two provers know the input I and can agree upon a strategy (i. e., mappings $u : X \to U$ and $v : Y \to V$) in advance, but they are not allowed to communicate about the particular random choice of questions the verifier actually asks. Based on x, y, u, and v, the verifier decides whether to accept or reject the input I. This is done by evaluating an *acceptance predicate* $Q : X \times Y \times U \times V \to \{0,1\}$ where output 1 means acceptance and 0 rejection. We think of Q also as a subset of $Z := X \times Y \times U \times V$, i.e., $Q = \{(x,y,u,v) \in Z \mid Q(x,y,u,v) = 1\}$. The probability distribution of the question pairs (x,y) chosen from $X \times Y$ by the verifier is denoted by μ. The strategy of the verifier (i. e., his choice of μ and Q based on I) is called the *proof system*, while the strategy of the provers (i. e., their choice of u and v, based on the knowledge of I, μ and Q) is called the *protocol*.

For fixed input I, a proof system G can be interpreted as a *game* for two players (provers). Thus, if we talk of $G = G(I)$ as a game we implicitly consider a fixed input I. A game G is described by sets X, Y, U, V, a predicate Q, a probability measure μ (which makes use of a random string τ), and mappings u and v. We will write $\mu(x,y)$ for $\mu\big((x,y)\big)$. Through their choices of u and v the two players aim to maximize the μ-probability that Q is satisfied, which is given by

$$\sum_{X\times Y} \mu(x,y)\, Q\big(x,y,u(x),v(y)\big)\,.$$

This probability is called the *value of the protocol* (u,v). The *value* $w(G)$ *of the game* G is defined as the maximal value of a protocol, i. e.,

$$w(G) := \max_{\substack{u:X\to U\\ v:Y\to V}} \sum_{X\times Y} \mu(x,y)\, Q\big(x,y,u(x),v(y)\big)\,.$$

A strategy (u,v) of the two players for which the maximum is attained is called an *optimal protocol*. A protocol with value 1 is also called a *winning strategy*.

We can now formally define MIP(2, 1) proof systems.

Definition 6.1. *A verifier and two provers build a* MIP(2, 1) *proof system for a language* L *with* error probability ε *if:*

1. Perfect completeness*: For all* $I \in L$*, there exists a protocol with value* 1 *(i. e., the provers have a winning strategy);*
2. Soundness: *For all* $I \notin L$*, the value of the proof system is* $w(G) \leqslant \varepsilon$ *(i. e., the provers can only succeed with probability at most* ε*).*

(The largest value c such that for all $I \in L$, the value of the corresponding game is at least c, is called *completeness*.) The following example is one of the most important applications of MIP(2, 1) proof systems in the context of non-approximability results. It will be applied in Chapters 7, 9, and 10.

Example 6.2. Consider an RobE3Sat formula $\varphi = C_1 \wedge \ldots \wedge C_m$. Let z_{a_j}, z_{b_j} and z_{c_j} denote the variables in clause C_j, $j = 1,\ldots,m$. Then we can define the following two-prover interactive proof system:

1. The verifier chooses a clause with index $x \in \{1,\ldots,m\}$ and a variable with index $y \in \{a_x, b_x, c_x\}$ at random. Question x is sent to prover P_1 and question y is sent to prover P_2.
2. Prover P_1 answers with an assignment for the variables z_{a_x}, z_{b_x}, and z_{c_x}, and prover P_2 answers with an assignment for z_y.
3. The verifier accepts if and only if C_x is satisfied by P_1's assignment and the two provers return the same value for z_y.

For each protocol, the strategy v of prover P_2 induces an assignment to the variables in the obvious way. Moreover, to get an optimal protocol prover P_1 must answer in accordance with this assignment unless the given clause C_x is not satisfied. If C_x is not satisfied, P_1 must change the value of at least one variable in the clause. Thus, if φ is satisfiable the value of the proof system is 1. Otherwise, with probability at least ε the verifier chooses a clause C_x not satisfied by P_2's assignment and P_1 has to change the value of a variable $z_{y'}$ where $y' \in \{a_x, b_x, c_x\}$. Since $z_y = z_{y'}$ with conditional probability at least $1/3$ the value of the protocol is at most $1 - \varepsilon/3$.

As a result of the PCP-Theorem 4.1 we know that we can translate an arbitrary instance of a problem in $\mathcal{NP}$ to an instance of RobE3Sat. The previous example is a MIP$(2,1)$ proof system for RobE3Sat with error probability $1 - \varepsilon/3$.

We briefly mention some other results on MIPs. The classes MIP(k,r) are generalizations of MIP$(2,1)$ with k provers and r rounds of questioning. Both the verifier and the provers are allowed to use their knowledge of earlier questions and answers. Obviously, MIP$(k,r) \subseteq$ MIP(k',r') holds, if $k \leqslant k'$ and $r \leqslant r'$. Feige and Lovász [FL92] proved that MIP$(2,1)$ = MIP(k,r) = $\mathcal{NEXP}$ for $k \geqslant 2$ and $r \geqslant 1$. The result that MIP$(1, \text{poly}(n))$ = $\mathcal{PSPACE}$ is known as Shamir's Theorem, see [Pap94]. It follows from results of [BGKW88] that multi-prover interactive proof systems with two-sided error (i. e., completeness < 1) behave similar to MIPs with one-sided error.

6.4 Reducing the Error Probability

The main application of MIPs in the context of approximability results is the construction of probabilistically checkable proofs with small error probability for $\mathcal{NP}$-hard optimization problems, see Chapters 7, 9, and 10. Without giving details, we shortly describe the basic idea behind this construction. One forces the two provers to write down their answers to all possible questions in a special encoding, the so-called long code. The resulting string serves as a proof in the context of PCPs. Thereby the soundness of the PCP directly depends on the error probability of the MIP. Thus, reducing the error in MIP$(2,1)$ proof systems is an important, however subtle issue.

A straightforward approach is to repeat an MIP$(2,1)$ protocol k times and to accept only if all executions are accepting. Ideally one would hope that this method reduces the error to ε^k. This is indeed true if the executions are performed sequentially and in an oblivious way, i. e., each prover must answer each question online before seeing the question for the next execution, and is not allowed to use his knowledge about the questions he was asked before.

Instead, we will consider parallel repetition, where each prover sends out its answers only after receiving all its questions. Such a k-fold parallel repetition of

G can be seen as a new two-prover one-round proof system with k *coordinates*, denoted by $G^{\otimes k}$. The verifier treats each coordinate of $G^{\otimes k}$ as an independent copy of the original game G and accepts in $G^{\otimes k}$ only if it would have accepted all the k copies of G. We will elaborate on the decomposition into coordinates in Section 6.5.

The *k-fold parallel repetition* $G^{\otimes k}$ of the game G consists of the four sets X^k, Y^k, U^k, and V^k, the probability measure $\mu^{\otimes k}$ defined on $\Omega^k := X^k \times Y^k$, and the acceptance predicate $Q^{\otimes k}$. For $\bar{x} \in X^k$ and $\bar{y} \in Y^k$ we write $\bar{x} = (x_1, \ldots, x_k)$ resp. $\bar{y} = (y_1, \ldots, y_k)$. The probability measure $\mu^{\otimes k}$ is defined by

$$\mu^{\otimes k}(\bar{x}, \bar{y}) := \prod_{i=1}^{k} \mu(x_i, y_i) \, .$$

The acceptance predicate is

$$Q^{\otimes k}(\bar{x}, \bar{y}, \bar{u}, \bar{v}) := \prod_{i=1}^{k} Q(x_i, y_i, u_i, v_i) \, .$$

A protocol for $\bar{G}$ consists of two mappings $\bar{u} : X^k \to U^k$ and $\bar{v} : Y^k \to V^k$. The value of $G^{\otimes k}$ is denoted by $w\big(G^{\otimes k}\big)$. For simplicity, we will also use the notations $\bar{G} := G^{\otimes k}$, $\bar{X} := X^k$, $\bar{Y} := Y^k$, $\bar{U} := U^k$, $\bar{V} := V^k$, $\bar{\mu} = \mu^{\otimes k}$, $\bar{\Omega} := \bar{X} \times \bar{Y}$, and $\bar{Q} = Q^{\otimes k}$. As before, let $\bar{Z} := \bar{X} \times \bar{Y} \times \bar{U} \times \bar{V}$.

Note that, in contrast to the case of sequential repetition, prover P_1 is allowed to take all the questions $x_1, \ldots, x_k$ into account (and not only x_i), when it responds to question x_i. The same holds for prover P_2. If the input I belongs to the language L, the provers already have a winning strategy in G, so knowing also the questions of other coordinates they can do no better. The problem with parallel repetition is therefore, to which amount this side information can help the provers to cheat, if the input I does *not* belong to the language L.

At first, it was believed that, as in the sequential case, repeating a proof system k times in parallel reduces the error to ε^k, see [FRS88]. Later, Fortnow constructed an example where $w(G^{\otimes 2}) > w(G)^2$, showing that the two provers can benefit from their knowledge of all the questions in advance, see [FRS90]. Feige [Fei91] presented the following game G which has the same value as its 2-fold parallel repetition $G^{\otimes 2}$, see also Exercise 6.2.

Example 6.3. The verifier selects two integers x and y independently at random from $\{0,1\}$ and sends x to prover P_1 and y to P_2. The provers have to reply with numbers u resp. v from $\{0,1\}$. Moreover each of them must either point at itself or at the other prover. The verifier accepts if

1. both point at the same prover and
2. the prover which both point at replies with the number it was asked and
3. $u + v \equiv 0 \mod 2$.

For some years, it was an open conjecture whether parallel repetition is suficent for every games G to reduce the error probability below arbitrarily small constants. This was proved by Verbitsky in [Ver94, Ver96] for the case where μ is a *uniform* measure supported by a subset of Ω. He pointed out that parallel repetition is connected to a density version of Hales-Jewett's theorem in Ramsey theory, proved by Furstenberg and Katznelson [FK91]. Unfortunately, the proof technique used gives no constructive bound on the number of required repetitions.

Recently, Raz [Raz95] showed that parallel repetition reduces the error probability of any MIP$(2,1)$ system even at an exponential rate, thus proving the so-called *Parallel Repetition Conjecture* [FL92]. The constant in the exponent (in his analysis) depends only on $w(G)$ and the *answer size* $s = s(G) := |U| \cdot |V|$ (of the original problem). His result is now known as the *Parallel Repetition Theorem*:

Theorem 6.4 (Parallel Repetition Theorem [Raz95]**).** *There exists a function $W : [0,1] \to [0,1]$ with $W(z) < 1$ for all $z < 1$, such that for all games G with value $w(G)$ and answer size $s = s(G) \geqslant 2$ and all $k \in \mathbb{N}$ holds*

$$w\left(G^{\otimes k}\right) \leqslant W\left(w(G)\right)^{k/\log s}.$$

For some applications, only the fact that parallel repetition reduces the error probability below arbitrarily small *constants* is used. Let us refer to this fact as the *weak parallel repetition theorem.*

In the meantime, Feige [Fei95] showed that Raz' theorem is nearly best possible. He proved that there is no universal constant $\alpha > 0$ such that

$$w\left(G^{\otimes k}\right) \leqslant w(G)^{\alpha k}$$

for all MIP$(2,1)$ proof systems. Hence, the exponent in Theorem 6.4 must depend on G. Its inverse logarithmic dependency on $s(G)$ was shown to be almost best possible by [FV96].

In the remaining sections of this chapter, we give an overview of the proof of the Parallel Repetition Theorem, based on [Raz95, Raz97, Fei95].

6.5 Coordinate Games

In the last section, we introduced $\bar{G}$ as the parallel repetition of G. In fact, $\bar{G}$ is best viewed as not being simply a repetition of the game G, but as a "simultaneous execution" of its *coordinates*, which will be defined next.

The *coordinate game* $\bar{G}^i$, where $i \in \{1, \ldots, k\}$, consists of:

- the sets $\bar{X}$, $\bar{Y}$, U, V;
- the probability measure $\bar{\mu}$;
- and the acceptance predicate $\bar{Q}^i$ defined by $\bar{Q}^i(\bar{x},\bar{y},u,v) := Q(x_i,y_i,u,v)$.

The value of $\bar{G}^i$ will be denoted by $w_i(\bar{G})$.

Thus, in the parallel repetition game $\bar{G}$ the verifier accepts if and only if all the predicates $\bar{Q}^i$, $i = 1,\ldots,k$, of the coordinate games $\bar{G}^i$ are satisfied. We will also say that $\bar{G}^i$ is the restriction of $\bar{G}$ to coordinate i. A protocol $\bar{u}$, $\bar{v}$ for the game $\bar{G}$ induces a protocol u_i, v_i for the game $\bar{G}^i$, where $u_i(\bar{x})$ and $v_i(\bar{y})$ are the i-th coordinates of the vectors $\bar{u}(\bar{x})$ resp. $\bar{v}(\bar{y})$.

Note that it is not required that the verifier will ask every question pair in $X \times Y$ with positive probability. In the proof of Theorem 6.4, we will consider restrictions of the game $\bar{G}$ to a question set $A \subseteq \bar{\Omega}$, which has the form of a cartesian product $A = A_X \times A_Y$, where $A_X \subseteq \bar{X}$ and $A_Y \subseteq \bar{Y}$. For $\mu(A) > 0$, we define

$$\bar{\mu}_A(\bar{x},\bar{y}) := \begin{cases} \dfrac{\bar{\mu}(\bar{x},\bar{y})}{\bar{\mu}(A)}, & (\bar{x},\bar{y}) \in A; \\ 0, & (\bar{x},\bar{y}) \notin A. \end{cases}$$

The game $\bar{G}_A$ is defined similarly to the game $\bar{G}$, but with the probability measure $\bar{\mu}_A$ instead of $\bar{\mu}$. The game $\bar{G}_A$ will be called the *restriction* of $\bar{G}$ to the question set A and is sometimes also denoted by $\bar{G}_{\bar{\mu}_A}$. If $\bar{\mu}(A) = 0$, we set $\bar{\mu}_A := 0$.

The following main technical theorem says that, provided $\bar{\mu}(A)$ is not "too" small, there always exists a coordinate i such that the value $w(\bar{G}^i_A)$ of the coordinate game $\bar{G}^i_A$ is not "too" large, compared with $w(G)$.

Theorem 6.5. *There exists a function $W_2 : [0,1] \to [0,1]$ with $W_2(z) < 1$ for all $z < 1$ and a constant c_0 such that the following holds. For all games G, dimensions k, and $A = A_X \times A_Y$, where $A_X \subseteq \bar{X}$ and $A_Y \subseteq \bar{Y}$, such that $-\log\bar{\mu}(A)/k \leqslant 1$ (i. e., $\bar{\mu}(A) \geqslant 2^{-k}$), there exists a coordinate i of $\bar{G}_A$ such that*

$$w(\bar{G}^i_A) \leqslant W_2\big(w(G)\big) + c_0 \left(\frac{-\log\bar{\mu}(A)}{k}\right)^{\frac{1}{16}}.$$

In Section 6.7 we will give some remarks on the proof of Theorem 6.5. But first let us see how Theorem 6.5 is applied in the proof of the Parallel Repetition Theorem 6.4

6.6 How to Prove the Parallel Repetition Theorem (I)

In order to prove Theorem 6.4, we will apply an inductive argument that yields a slightly stronger result. Let us define

$$C(k,r) := \max_{\substack{A=A_X\times A_Y\\ -\log\bar{\mu}(A)\leqslant r}} w(G_A^{\otimes k})$$

for $k \in \mathbb{N}$ and $r \in \mathbb{R}_+$. It will be convenient to define $C(0,r) := 1$. $C(k,r)$ is an upper bound for the value (i. e., error probability) of any restriction of k parallel repetitions of G to a set A, that has the form of a cartesian product $A = A_X \times A_Y$ and whose size $\bar{\mu}(A)$ is not very small, where "small" is quantified by r. Observe that $\bar{G} = \bar{G}_{\bar{\Omega}}$ and $\bar{\mu}(\bar{\Omega}) = 1$, so $C(k,r)$ is an upper bound for $w(\bar{G})$, and in particular, $C(k,0) = w(\bar{G})$.

For technical reasons, we will not apply Theorem 6.5, but the following restatement of it.

Theorem 6.6. *There exists a function* $W_2 : [0,1] \to [0,1]$ *with* $W_2(z) < 1$ *for all* $z < 1$ *and a constant* c_0 *such that the following holds. For all games* G, *dimensions* k, *and* $A = A_X \times A_Y$, *where* $A_X \subseteq \bar{X}$ *and* $A_Y \subseteq \bar{Y}$, *such that for all* Δ *with* $-\log\bar{\mu}(A)/k \leqslant \Delta \leqslant 1$ *(i. e.,* $\bar{\mu}(A) \geqslant 2^{-\Delta k}$*), there exists a coordinate* i *of* $\bar{G}_A$ *such that*

$$w(\bar{G}_A^i) \leqslant W_2\big(w(G)\big) + c_0\Delta^{\frac{1}{16}}.$$

It will be necessary to choose Δ carefully such that certain assumptions made in the proof are satisfied.

Now we show how Raz derives Theorem 6.4 from Theorem 6.6. Let us denote the upper bound from Theorem 6.6 by

$$\hat{w} := W_2\big(w(G)\big) + c_0\Delta^{\frac{1}{16}}.$$

and assume that $r \leqslant \Delta k$. We choose an $A = A_X \times A_Y \subseteq \bar{\Omega}$ with $-\log\bar{\mu}(A) \leqslant r$ that maximizes $w(\bar{G}_A)$, i.e., $C(k,r) = w(\bar{G}_A)$. (So $\bar{\mu}(A) > 0$.) By Theorem 6.6, we know that there must be a coordinate i (without loss of generality, we will assume that $i = 1$) whose value is not too big, namely $w(\bar{G}_A^i) \leqslant \hat{w}$. We will use this "good" coordinate in order to make an induction step in the following way. Recall that the value $w(\bar{G}_A)$ is defined as the maximal expected error probability with respect to the probability measure $\bar{\mu}_A$, taken over all protocols $\bar{u} : \bar{X} \to \bar{U}$, $\bar{v} : \bar{Y} \to \bar{V}$ the provers can agree upon. This means that an optimal protocol $\bar{u}$, $\bar{v}$ (for the provers) satisfies

$$w(\bar{G}_A) = \sum_{(\bar{x},\bar{y})\in\bar{\Omega}} \bar{\mu}_A(\bar{x},\bar{y})\,\bar{Q}\big(\bar{x},\bar{y},\bar{u}(\bar{x}),\bar{v}(\bar{y})\big)\,.$$

Let us fix such an optimal protocol $\bar{u}$, $\bar{v}$. Next we partition A according to the behavior of $\bar{u}$ and $\bar{v}$ on the first coordinate. For every point $z = (x, y, u, v) \in Z$, we define a partition class $A(z) \subseteq A$ by

$$A(x,y,u,v) := \{(\bar{x},\bar{y}) \in A \mid x_1 = x \wedge y_1 = y \wedge u_1(\bar{x}) = u \wedge v_1(\bar{y}) = v\}.$$

Note that the partition classes $A(z)$ again have the form of cartesian products as required for the application of Theorem 6.6. We denote the subset of questions of $A(z)$ such that the acceptance predicate $\bar{Q}$ is satisfied by

$$B(x,y,u,v) := \{(\bar{x},\bar{y}) \in A(x,y,u,v) \mid \bar{Q}(\bar{x},\bar{y},\bar{u}(\bar{x}),\bar{v}(\bar{y})) = 1\}.$$

The sets $B(z)$ will be useful later. The size of the partition $\{A(z) \mid z \in Z\}$ is not too big, and therefore the average size of a partition class is not too small. Also, in many of these subsets $A(z)$ of A the protocol $\bar{u}$, $\bar{v}$ fails to satisfy the predicate $\bar{Q}^1$, because $i = 1$ is a good coordinate. If $\bar{Q}^1$ is not satisfied, then $\bar{Q}$ is also not satisfied, so we can forget about these subsets. This is a consequence of the fact that $\bar{Q}$ can be viewed as the product of $\bar{Q}^1$ and a $(k-1)$-dimensional predicate and that $\bar{Q}^1$ is constant on each $A(z)$. Let us deduce this fact formally. If $Q(z) = 0$, then $B(z) = \emptyset$, because $(\bar{x},\bar{y}) \in A(z)$ implies

$$\bar{Q}(\bar{x},\bar{y},\bar{u}(\bar{x}),\bar{v}(\bar{y})) \leqslant Q(x_1,y_1,u_1(\bar{x}),v_1(\bar{y})) = Q(z) = 0.$$

If $Q(z) = 1$, then $(\bar{x},\bar{y}) \in A(z)$ implies

$$\bar{Q}(\bar{x},\bar{y},\bar{u}(\bar{x}),\bar{v}(\bar{y})) = \underbrace{Q(z)}_{=1} [t] \prod_{j=2}^{k} Q(x_j,y_j,u_j(\bar{x}),v_j(\bar{y})).$$

These facts enable us to carry out an inductive argument.

We denote the conditional probability with which a random $(\bar{x},\bar{y}) \in A$ that was chosen according to the distribution $\bar{\mu}_A$ lies in $A(z)$, depending on $z \in Z$, by

$$\alpha(z) := \bar{\mu}_A(A(z)) = \frac{\bar{\mu}(A(z))}{\bar{\mu}(A)}.$$

Of course, $\alpha(Z) = 1$, because $\{A(z) \mid z \in Z\}$ is a partition of A. Note that the event $(\bar{x},\bar{y}) \in A(z)$ depends only on the first coordinate of $\bar{G}_A$, which is $(x_1,y_1,u_1(\bar{x}),v_1(\bar{y}))$. For some of the partition classes $A(z)$, the predicate $\bar{Q}^1(x_1,y_1,u_1(\bar{x}),v_1(\bar{y}))$ is not satisfied. Summing the α-probability of the other partition classes gives us the value of the first coordinate game. Thus the following claim holds.

Claim 6.1:

$$\alpha(Q) = w(\bar{G}_A^1) \leqslant \hat{w}.$$

This is the point where Theorem 6.6 is applied. But to get an assertion about the parallel repetition game $\bar{G}_A$, we also have to consider the probability that a random $(\bar{x}, \bar{y}) \in A(z)$ leads to acceptance in the game $\bar{G}_{A(z)}$. Let us denote this probability by

$$\beta(z) := \bar{\mu}_{A(z)}(\bar{Q}) = \frac{\bar{\mu}(B(z))}{\bar{\mu}(A(z))}.$$

If $\bar{\mu}(A(z)) = 0$, we set $\beta(z) := 0$. Now a moment's thought shows that the following claim is true.

Claim 6.2:

$$\sum_{z \in Q} \alpha(z)\, \beta(z) = w(\bar{G}_A) = C(k, r).$$

Next, we deduce an upper bound for $\beta(z)$. Observe that in fact, $A(z)$ and $B(z)$ are only $(k-1)$-dimensional sets, so it makes good sense to define

$$A'(z) := \left\{ \left((x_2, \ldots, x_k), (y_2, \ldots, y_k) \right) \,\middle|\, (\bar{x}, \bar{y}) \in A(z) \right\}.$$

and

$$B'(z) := \left\{ \left((x_2, \ldots, x_k), (y_2, \ldots, y_k) \right) \,\middle|\, (\bar{x}, \bar{y}) \in B(z) \right\}.$$

In a similar way, we define a protocol for the game $\bar{G}_{A'(z)}^{\otimes(k-1)}$ by

$$u'(x_2, \ldots, x_k) := \left(u_j(x, x_2, \ldots, x_k) \,\middle|\, j = 2, \ldots, k \right)$$

and

$$v'(y_2, \ldots, y_k) := \left(v_j(y, y_2, \ldots, y_k) \,\middle|\, j = 2, \ldots, k \right).$$

Since the predicate Q^{k-1} of the game $G_{A'(z)}^{\otimes(k-1)}$ is satisfied precisely at the elements of $B'(z)$ and the protocol $\bar{u}$, $\bar{v}$ was chosen optimal for $\bar{G}_A$, it follows that

$$w\left(G_{A'(z)}^{\otimes(k-1)} \right) \geqslant \frac{\mu^{\otimes(k-1)}(B'(z))}{\mu^{\otimes(k-1)}(A'(z))} = \frac{\bar{\mu}(B(z))}{\bar{\mu}(A(z))} = \beta(z).$$

A trivial upper bound for $w\left(G_{A'(z)}^{\otimes(k-1)} \right)$ that holds by definition is

$$w\left(G_{A'(z)}^{\otimes(k-1)} \right) \leqslant C\Big(k-1,\ -\log \mu^{\otimes(k-1)}(A'(z)) \Big).$$

Also, we have

$$\mu^{\otimes(k-1)}(A'(z)) = \frac{\bar{\mu}(A(z))}{\mu(x,y)} = \frac{\alpha(z)\, \bar{\mu}(A)}{\mu(x,y)} \geqslant 2^{-r} \frac{\alpha(z)}{\mu(x,y)}.$$

If we put these inequalities together, using the fact that the function $C(\cdot,\cdot)$ is monotone increasing in the second argument, we see that the following claim is true.

Claim 6.3:
For all $z = (x, y, u, v) \in Q$ with $\alpha(z) > 0$,

$$\beta(z) \leqslant C\left(k-1,\ r - \log\frac{\alpha(z)}{\mu(x,y)}\right).$$

The claims 6.1, 6.2, and 6.3 lead to a recursive inequality for C. Let $T := \{z \in Q \mid \alpha(z) > 0\}$ be the set of all realizations of the coordinate game $\bar{G}_A^1$ that occur with positive probability. Then it holds

$$C(k,r) = \sum_{z \in T} \alpha(z)\,\beta(z) \leqslant \sum_{z \in T} \alpha(z)\, C\left(k-1,\ r - \log\frac{\alpha(z)}{\mu(x,y)}\right).$$

It remains to carry out a recursive estimation, using the bound for $C(k-1,\cdot)$ to prove the bound for $C(k,\cdot)$. Define

$$c := \hat{w}^{\frac{1}{2(\log s + \Delta)}}$$

for abbreviation. (So c is a function of $w(G)$, $s(G)$, and Δ.) The heart of the proof of the following claim is a clever application of Jensen's inequality.

Claim 6.4:

$$C(k,r) \leqslant c^{\Delta k - r}.$$

provided $0 \leqslant \hat{w} < 1$ and $\frac{1}{\sqrt{2}} < c < 1$.

Since $0 < W_2\big(w(G)\big) < 1$ and $\hat{w}$ is monotone in Δ, we can find an appropriate Δ that satisfies the assumptions we made during the proof by starting from $\Delta = 0$ and increasing Δ until the conditions hold.

Finally, we show how Theorem 6.4 follows from claim 6.4. Note that $\Delta < 2\log s$ (because $\Delta < 1$ and $s > 2$). Therefore, under the assumptions made above,

$$w(\bar{G}) = C(k,0) \leqslant c^{\Delta k} = \hat{w}^{\frac{1}{2(\log s+\Delta)}\Delta k} \leqslant \hat{w}^{\frac{1}{4\log s}\Delta k} = \left(\hat{w}^{\Delta/4}\right)^{k/\log s}.$$

So Theorem 6.4 holds with

$$W\big(w(G)\big) := \inf\left(\hat{w}^{\Delta/4}\right),$$

where the infimum is taken over all Δ satisfying the assumptions.

6.7 How to Prove the Parallel Repetition Theorem (II)

In this section we give an overview of basic ideas and techniques that are used in the proof of Theorem 6.6. We do not try to explain how to get the exact quantitative result of the theorem but rather aim to motivate the qualitative statement. The entire proof covers more than 30 pages and can be found in an extended version [Raz97] of [Raz95]. Our description is also based on the short overview of the proof given by Feige in [Fei95]. In particular we use Feige's notion of "good" coordinates and their characterization by certain properties, see Properties 1–3 below.

For the following considerations we keep X, Y, U, V, Q, μ, k and the corresponding game G fixed. Moreover, we will consider games consisting of X, Y, U, V, Q resp. $\bar{X}, \bar{Y}, \bar{U}, \bar{V}, \bar{Q}$ together with probability distributions other than μ resp. $\bar{\mu}$. For arbitrary probability measures $\alpha : \Omega \rightarrow \mathbb{R}$ and $\bar{\alpha} : \bar{\Omega} \rightarrow \mathbb{R}$ we denote the corresponding games by G_α resp. $\bar{G}_{\bar{\alpha}}$. Thus, we can denote the restricted game $\bar{G}_A$ alternatively by $\bar{G}_{\bar{\mu}_A}$ where $A = A_X \times A_Y$ is a fixed subset of $\bar{\Omega}$. To simplify notation we denote the probability measure $\bar{\mu}_A$ by $\bar{\pi}$.

We think of w as a function from the set of all probability measures on Ω resp. $\bar{\Omega}$ to $\mathbb{R}$ and denote the values of the games G_α and $\bar{G}_{\bar{\alpha}}$ by $w(\alpha)$ resp. $w(\bar{\alpha})$. In the following lemma we state two basic properties of the function w. The proof is left to the reader, see exercise 6.5.

Lemma 6.7. *The function w is continuous and concave.*

For $\bar{\alpha} : \bar{\Omega} \rightarrow \mathbb{R}$ define $\bar{\alpha}^i$ to be the projection of $\bar{\alpha}$ on the i-th coordinate, i. e., for $(x, y) \in \Omega$ let

$$\bar{\alpha}^i(x,y) := \sum_{\substack{(\bar{x},\bar{y})\in\bar{\Omega}: \\ (x_i,y_i)=(x,y)}} \bar{\alpha}(\bar{x},\bar{y}) .$$

In particular $\bar{\alpha}^i$ is a probability measure on Ω. We will consider the following games:

- The game $\bar{G}_{\bar{\alpha}}$ as introduced above consisting of $\bar{X}, \bar{Y}, \bar{U}, \bar{V}, \bar{Q}$, with the measure $\bar{\alpha}$ and value $w(\bar{\alpha})$.
- The coordinate game $\bar{G}^i_{\bar{\alpha}}$ of $\bar{G}_{\bar{\alpha}}$ as defined in Section 6.5 consisting of the four sets $\bar{X}, \bar{Y}, U, V$, with the acceptance predicate $\bar{Q}^i$ and with the measure $\bar{\alpha}$. We denote the value of this game by $w_i(\bar{\alpha})$.
- The one-dimensional game $G_{\bar{\alpha}^i}$ induced by the measure $\bar{\alpha}^i$ on X, Y, U, V, Q with value $w(\bar{\alpha}^i)$.

Theorem 6.6 claims that for "large" $A = A_X \times A_Y$ (with respect to the measure $\bar{\mu}$) there exists a coordinate i such that the provers succeed in the corresponding coordinate game $\bar{G}^i_{\bar{\pi}}$ with probability "not much higher" than $w(G)$. We keep such a large subset A fixed for the rest of this section. To prove Theorem 6.6 Raz considers coordinates that satisfy certain properties which lead to the property stated in the theorem. As proposed by Feige we call coordinates with these properties "good". In the rest of this section we characterize good coordinates in an informal way.

A natural requirement on a good coordinate i seems to be:

Property 1: The projection $\bar{\pi}^i$ of $\bar{\pi}$ on coordinate i is very close to the original probability distribution μ.

To be more precise we say that $\bar{\pi}^i$ is close to μ if the *informational divergence* $\mathbf{D}(\bar{\pi}^i \,||\, \mu)$ of $\bar{\pi}^i$ with respect to μ is small. The informational divergence (or *relative entropy*) is a basic tool of information theory and is defined by

$$\mathbf{D}(\bar{\pi}^i \,||\, \mu) := \sum_{z \in \Omega} \bar{\pi}^i(z) \log \frac{\bar{\pi}^i(z)}{\mu(z)}\,.$$

The informational divergence is always non-negative and if it is small then the L_1 distance between the two measures is also small. A short discussion of basic properties can be found in [Raz95], for further information see [Gra90, CK81].

It is easy to show that for large A Property 1 holds for many coordinates i. As a consequence of Lemma 6.7 we know that the value $w(\bar{\pi}^i)$ of the game $G_{\bar{\pi}^i}$ is not much larger than $w(\mu)$ if the coordinate i satisfies Property 1. Thus it suffices to show that for one such coordinate i, the value $w_i(\bar{\pi})$ of the coordinate game $\bar{G}^i_{\bar{\pi}}$ is bounded by some "well behaved" function of $w(\bar{\pi}^i)$. On the other hand it is always true that

$$w(\bar{\pi}^i) \leqslant w_i(\bar{\pi})\,, \tag{6.1}$$

because any protocol for the one-dimensional game $G_{\bar{\pi}^i}$ induces in a canonical way a protocol with the same value for the coordinate game $\bar{G}^i_{\bar{\pi}}$. Remember that in the one-dimensional game $G_{\bar{\pi}^i}$ on X, Y, U, V each prover is asked only one question. However, in the game $\bar{G}^i_{\bar{\pi}}$, each prover gets k questions $\bar{x}$ resp. $\bar{y}$ but only has to answer the i-th question x_i resp. y_i. Thus, ignoring all questions but the i-th, the provers of the game $\bar{G}^i_{\bar{\pi}}$ can apply the optimal strategy of the one-dimensional game $G_{\bar{\pi}^i}$. Moreover, since by definition of the projection $\bar{\pi}^i$ the probability distribution of questions (x, y) in the one-dimensional game exactly equals the distribution of the i-th question (x_i, y_i) in the game $\bar{G}^i_{\bar{\pi}}$, the value of this protocol is the same for both games.

The following lemma describes a special case where an optimal protocol for the one-dimensional game $G_{\bar{\pi}^i}$ is also optimal for the game $\bar{G}^i_{\bar{\pi}}$ (see [Raz95, Lemma 4.1]).

Lemma 6.8. *If there exist functions* $\alpha_1 : X \times Y \to \mathbb{R}$, $\alpha_2 : \bar{X} \to \mathbb{R}$, $\alpha_3 : \bar{Y} \to \mathbb{R}$ *such that for all* $(\bar{x}, \bar{y}) \in \bar{X} \times \bar{Y}$

$$\bar{\pi}(\bar{x}, \bar{y}) = \alpha_1(x_i, y_i)\, \alpha_2(\bar{x})\, \alpha_3(\bar{y})$$

then

$$w(\bar{\pi}^i) = w_i(\bar{\pi})\,.$$

Sketch of Proof. Because of (6.1), it remains to show that $w(\bar{\pi}^i) \geqslant w_i(\bar{\pi})$. Inverting the argument given above we have to show that a protocol for $\bar{G}^i_{\bar{\pi}}$ can be used to define a corresponding protocol with the same value for $G_{\bar{\pi}^i}$. The basic idea is that in this special case the provers of the game $G_{\bar{\pi}^i}$ can simulate a k-dimensional input $(\bar{x}, \bar{y})$ with distribution $\bar{\pi}$ from the given input (x_i, y_i) with distribution $\bar{\pi}^i$. Thus, Exercise 6.4 completes the proof. ■

Corollary 6.9. *If μ is a product measure, i. e.,*

$$\mu(x, y) = \beta(x)\,\gamma(y)$$

where $\beta : X \to \mathbb{R}$ and $\gamma : Y \to \mathbb{R}$ are arbitrary probability measures, then Property 1 suffices to define good coordinates and to prove Theorem 6.6 for this special case.

Proof. If μ is a product measure then the same holds for $\bar{\pi}$ by definition. Moreover, it is easy to see that the condition of Lemma 6.8 is satisfied in this case. This, together with the remarks after Property 1, completes the proof of Theorem 6.6. ■

Unfortunately, in general we do not get an optimal protocol for $\bar{G}^i_{\bar{\pi}}$ by taking one for $G_{\bar{\pi}^i}$. The reason is that in the game $\bar{G}^i_{\bar{\pi}}$ a prover can loose important information by ignoring all but the i-th question. Recall that in the game $\bar{G}$ with measure $\bar{\mu}$ the question pairs selected at different coordinates are independent. This is no longer true in the restricted game $\bar{G}_{\bar{\pi}}$ with measure $\bar{\pi}$. Thus the question which a prover receives on a coordinate j different from i may already provide information on its i-th question. Feige defines the side information for coordinate i as the questions a prover receives on coordinates other than i.

The situation gets even worse if the question a prover receives in coordinate j is correlated with the question that the other prover receives on this coordinate. Then the side information for coordinate i may help him to guess the i-th question of the other prover. As a consequence the side information can obviously help the provers to succeed on coordinate i. Thus a good coordinate should not only meet Property 1, in addition we should require that the side information for good coordinates are in a way useless:

Property 2: The side information for coordinate i available to each prover conveys almost no information on the question that the other prover receives on this coordinate (beyond the direct information available to the prover through its own question on coordinate i and the description of the underlying probability measure used by the verifier).

Unfortunately, even this condition does not suffice to define good coordinates and to get the desired result. Consider the following example which is given in [Fei95]:

Example 6.10. Define a game G by $X = Y = U = V = \{0,1\}$. For $x \in X$, $y \in Y$, $u \in U$, and $v \in V$ the acceptance predicate Q is given by $x \wedge y = u \oplus v$ (where $\oplus$ denotes *exclusive or*). The probability distribution μ is defined by $\mu(x,y) = 1/4$ for all $(x,y) \in X \times Y$, i. e., the two questions are drawn independently from each other and all choices are equally likely.

This game is not trivial, i. e., its value is not 1, see Exercise 6.1. However, for the game $\bar{G} = G^{\otimes 2}$ consider the subset A of $X^2 \times Y^2$ that includes eight question pairs, written as (x_1x_2, y_1y_2):

$$A = \{(00,00),(01,01),(00,10),(01,11),(10,00),(11,01),(10,11),(11,10)\}$$

It is an easy observation that the projection of A on its first coordinate gives the uniform distribution over question pairs (x,y). Moreover, the question y_1 is independent of the two questions x_1x_2 and x_1 is independent of y_1y_2, see [Fei95, Proposition 4]. Thus Properties 1 and 2 hold for coordinate 1. Nevertheless, there exists a perfect strategy for the provers on the first coordinate. Just observe that $x_1 \wedge y_1 = x_2 \oplus y_2$ for all question pairs in A. Thus we get a perfect strategy for the first coordinate if the first prover gives the answer x_2 while the second prover gives the answer y_2.

Example 6.10 seems to contradict the statement of Corollary 6.9. The reason is that the set A given above cannot be written as $A_X \times A_Y$ with $A_X \subseteq X^2$ and $A_Y \subseteq Y^2$. Nevertheless, the game G can be extended to a new game G' which, together with a subset $A' = A'_X \times A'_Y$, shows that Property 2 together with Property 1 does not suffice for a coordinate i to meet the property claimed in Theorem 6.6.

To overcome this drawback one has to substitute Property 2 by a somewhat stronger condition on good coordinates. The idea is to reduce the problem in a way to probability measures as considered in Lemma 6.8. One represents $\bar{\pi}$ as a convex combination of measures that satisfy the condition of Lemma 6.8. Therefore we need the following definition which Raz considers as "probably the most important notion" for his proof.

Definition 6.11. *For a fixed coordinate $i \in \{1,\ldots,k\}$ a set M of type $\mathcal{M}^i$ is given by*

1. *a partition of the set of coordinates $\{1,\ldots,k\} - \{i\}$ into $J \cup L$ and*
2. *values $a_j \in X$, for all $j \in J$, and $b_\ell \in Y$, for all $\ell \in L$.*

Then M is given by

$$M = \{(\bar{x},\bar{y}) \in \bar{X} \times \bar{Y} \mid x_j = a_j \textit{ for all } j \in J,\ y_\ell = b_\ell \textit{ for all } \ell \in L\}.$$

$\mathcal{M}^i$ denotes the family of all sets M of type $\mathcal{M}^i$.

As a simple application of Lemma 6.8 one can prove that

$$w_i(\bar{\pi}_M) = w(\bar{\pi}^i_M) \tag{6.2}$$

for a set M of type $\mathcal{M}^i$. Moreover, the probability measure $\bar{\pi}$ induces in a natural way a measure $\rho_i : \mathcal{M}^i \to \mathbb{R}$ by

$$\rho_i(M) = \frac{\bar{\pi}(M)}{2^{k-1}}.$$

Since $\mathcal{M}^i$ is a cover of $\bar{X} \times \bar{Y}$ and each element $(\bar{x}, \bar{y})$ is covered exactly 2^{k-1} times, ρ_i is in fact a probability measure for $\mathcal{M}^i$ and we can write $\bar{\pi}$ as the convex combination

$$\bar{\pi} = \sum_{M \in \mathcal{M}^i} \rho_i(M)\bar{\pi}_M \tag{6.3}$$

At this point we can imagine the importance of the set $\mathcal{M}^i$. It finally enables us to write the probability measure $\bar{\pi}$ as the convex combination (6.3) of measures $\bar{\pi}_M$ with the nice property (6.2).

Moreover, we can now bound the value $w_i(\bar{\pi})$ of the coordinate game $\bar{G}^i_{\bar{\pi}}$. First note that a protocol for $\bar{G}^i_{\bar{\pi}}$ is also a protocol for each game $\bar{G}^i_{\bar{\pi}_M}$. Thus the concavity of the function w_i (see Exercise 6.5) yields

$$w_i(\bar{\pi}) \leqslant \sum_{M \in \mathcal{M}^i} \rho_i(M) w_i(\bar{\pi}_M).$$

The right hand side of this inequality can be interpreted as the expectation $\mathbf{E}_{\rho_i}\big(w_i(\bar{\pi}_M)\big)$. This together with equation (6.2) yields the upper bound

$$w_i(\bar{\pi}) \leqslant \mathbf{E}_{\rho_i}\big(w(\bar{\pi}^i_M)\big).$$

We need the following notation. For a probability measure $\alpha : X \times Y \to \mathbb{R}$, define $\alpha(x, \cdot) : Y \to \mathbb{R}$ to be the induced measure on Y for fixed $x \in X$, i.e., for $y \in Y$

$$\alpha(x, \cdot)(y) = \frac{\alpha(x, y)}{\alpha(x)}$$

where $\alpha(x) = \sum_{y \in Y} \alpha(x, y)$. If $\alpha(x) = 0$ set $\alpha(x, \cdot)$ to be identically 0. Define $\alpha(\cdot, y)$ and $\alpha(y)$ in the same way.

The following lemma is a qualitative version of [Raz95, Lemma 4.3]. It defines a condition on coordinate i that suffices to upper bound $\mathbf{E}_{\rho_i}\big(w(\bar{\pi}^i_M)\big)$.

Lemma 6.12. *Let coordinate i satisfy Property 1. If*

$$\begin{array}{rl} & \mathbf{E}_{\rho_i}\Big(\sum_{x \in X} \bar{\pi}^i_M(x)\, \mathbf{D}\left(\bar{\pi}^i_M(x, \cdot) \,\|\, \mu(x, \cdot)\right)\Big) \\ \textit{and} & \mathbf{E}_{\rho_i}\Big(\sum_{y \in Y} \bar{\pi}^i_M(y)\, \mathbf{D}\left(\bar{\pi}^i_M(\cdot, y) \,\|\, \mu(\cdot, y)\right)\Big) \end{array} \tag{6.4}$$

are "small" then $\mathbf{E}_{\rho_i}\big(w(\bar{\pi}^i_M)\big)$ is upper bounded by some well behaved function of $w(\mu)$.

We will neither give the proof for Lemma 6.12 nor the proof for the existence of such a good coordinate (see [Raz95, Section 5]). Instead we give a more intuitive formulation for the condition in Lemma 6.12 similar to the one given by Feige.

Consider a fixed $M \in \mathcal{M}^i$ with corresponding partition $\{1,\ldots,k\} - \{i\} = J \cup L$ and questions $a_j \in X$ for $j \in J$ and $b_\ell \in Y$ for $\ell \in L$. The term $\mathbf{D}(\bar{\pi}^i_M(x,\cdot) \,||\, \mu(x,\cdot))$ can be interpreted as a measure for the information that the first prover can get on the i-th question of the other prover from the knowledge of the questions $x_j = a_j$ for $j \in J$ and $y_\ell = b_\ell$ for $\ell \in L$. Using a notion of Feige we call this information the extended side information for coordinate i with regard to M. Thus the condition given in Lemma 6.12 can be reformulated as:

Property 3: On an average (with regard to ρ_i), the extended side information for coordinate i conveys almost no additional information on the question that the other prover receives on coordinate i.

As mentioned above with Properties 1 and 3 defining good coordinates, Raz showed that for large A at least one good coordinate i exists and the value of the coordinate game $\bar{G}^i_{\bar{\pi}}$ is bounded as claimed in Theorem 6.6.

Exercises

Exercise 6.1. What is the value of the game given in Example 6.10?

Exercise 6.2. What is the value $w(G)$ of the game G given in Example 6.3? Show that 2-fold parallel repetition does not reduce the error probability, i. e., $w(G^{\otimes 2}) = w(G)$.

Exercise 6.3. Multiple-prover games are a natural extension of two prover games. Can you generalize the game given in Example 6.3 to the case of k provers, for $k > 2$, such that $w(G^{\otimes k}) = w(G)$?

Exercise 6.4. Can the value of a game be increased by allowing the provers to use random strategies?

Exercise 6.5. For fixed X, Y, U, V, Q, consider the value $w(G)$ of the game G as a function $w(\mu)$ of the measure $\mu : X \times Y \to [0,1]$. Show that this function is concave and continuous with Lipschitz constant 1.