Méthodes Combinatoires et Algébriques en Complexité de la Communication

Marc Kaplan

▶ **To cite this version:**

Marc Kaplan. Méthodes Combinatoires et Algébriques en Complexité de la Communication. Informatique [cs]. Université Paris Sud - Paris XI, 2009. Français. NNT : . tel-00439929

HAL Id: tel-00439929
https://theses.hal.science/tel-00439929

Submitted on 9 Dec 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Paris-Sud XI
Laboratoire de recherche en informatique

Méthodes Combinatoires et Algébriques en Complexité de la Communication

Marc Kaplan

Thèse présentée pour obtenir le grade de
Docteur en Sciences

soutenue le 28 Septembre 2009 à Orsay devant le jury composé de :

Mr Christoph Dürr	Chargé de Recherche, CNRS	Rapporteur
Mr Serge Massar	Maître de Recherche, FNRS	Rapporteur
Mme Sophie Laplante	Professeur, Université Paris XI	Directeur de thèse
Mme Nicole Bidoit	Professeur, Université Paris XI	Examinateur
Mr Alain Tapp	Professeur Agrégé, Université de Montréal	Examinateur

The medium is the message. This is merely to say that the personal and social consequences of any medium - that is, of any extension of ourselves - result from the new scale that is introduced into our affairs by each extension of ourselves, or by any new technology.

M. McLuhan, Understanding Media : The Extensions of Man

Table des matières

Remerciements

La personne dont le travail présenté ici est le plus redevable est sans aucun doute Sophie Laplante. J'ai eu la chance d'être son étudiant pendant quatre années au cours desquelles j'ai été formé au métier de chercheur. Sa gentillesse, sa générosité et sa disponibilité m'ont permis d'acquérir de nombreuses aptitudes : des compétences scientifiques d'abord, en me permettant d'apprendre l'informatique après des études de mathématiques ; des compétences techniques ensuite, en me montrant les nombreuses tâches de l'enseignant chercheur ; et au-delà de cela, Sophie m'a transmis l'éthique de sont métier. Rétrospectivement, je peux dire que j'ai eu autant de plaisir pendant ma thèse à pratiquer la recherche scientifique à proprement parler qu'à cultiver des valeurs telles que que l'honnêteté intellectuelle, le désintéressement, la curiosité scientifique, l'humilité devant le savoir, etc... En me transmettant ces valeurs, Sophie m'a fait un cadeau inestimable. Par ailleurs, Sophie a coutume de dire que les chercheurs en complexité sont des gens sympathiques et agréables à fréquenter. Je pense qu'elle en est l'un des plus parfaits exemples. Pour toutes ces raisons, l'évocation de Sophie Laplante m'inspirera toujours respect et admiration, tant d'un point de vue professionnel que personnel.

Ensuite, ce travail n'aurait pu être fait sans la collaboration de chacun de mes coauteurs. J'ai en effet eu la chance de travailler avec des personnes d'horizons scientifiques très différents et j'ai appris de chacun d'eux des choses différentes. J'ai rencontré Julien Degorre au début de mes études. Il avait déjà ce grain de folie particulier qui rend le travail avec lui à la fois riche et sympathique. Sa présence à Orsay est l'un des éléments qui m'a amené à rejoindre l'équipe algorithmique et complexité. Quant à Jérémie Roland, son dynamisme et son talent sont exemplaires, et j'ai beaucoup apprécier travailler avec lui. Je le remercie tout particulièrement pour son accueil lors de la visite que je lui ai rendu chez NEC. De plus, je dois à Jérémie et à Julien à peu près tout ce que je sais de physique. Le talent de chercheur de Iordanis Kerenidis n'est pas à démontrer. Il est de plus la preuve qu'un tel talent peut se conjuguer avec une sympathie illimitée. Il a fait preuve d'une grande patience pour m'expliquer son travail, et je me considère comme très chanceux d'avoir eu l'occasion d'interagir avec lui.

La dynamisme de l'équipe algorithmique et complexité du LRI doit probablement beaucoup à son responsable Miklos Santha. Ses connaissances et sa rigueur sont reconnues dans le milieux scientifique. J'ai trouvé les conversations que nous avons eues, scientifiques ou non, d'un grand intérêt, mais surtout trop rares. J'espère en avoir bien d'autres dans le futur, et sur de nombreux continents. L'équipe a la chance d'avoir un autre personnage de grande qualité avec Frédéric Magniez. Ses conseils m'auront été particulièrement utiles et auront largement contribué à rendre mon séjour dans l'équipe des plus agréables. L'étendu de ses connaissances m'aura également aidé à réaliser mon épanouissement scientifique.

De nombreux professeurs ont jalonné la route qui m'a mené des mathématiques à

la logique, puis à l'informatique théorique. Il m'est impossible d'en dresser une liste complète, et je me restreindrait à ceux dont l'apport est le plus facile à reconnaître. Richard Lassaigne d'abord, car il est le premier à m'avoir donné un cour sur la théorie de la complexité. Michel de Rougemont, qui a été mon enseignant à Paris 7, m'a également aidé à passer de Paris à Orsay.

J'ai eu pendant ma thèse l'occasion de discuter avec de nombreuses personnes. Même si cela n'a pas débouché sur des collaborations concrètes, toutes ces conversation ont été très stimulantes, et sont une partie importante de mon travail. Je me dois de commencer par Troy Lee, tant j'ai appris de lui. Je lui dois par exemple quasiment tout ce que je sais sur les normes de factorisation. De plus, son accueil lors de mon séjour à Rutgers a démontré une borne inférieure très large sur son hospitalité. Je ne peux pas garantir de n'oublier aucune des personnes qui ont ainsi indirectement participé à ma thèse, mais de chacune des personnes de la liste suivante, je garde le souvenir d'une conversation au cours de laquelle j'ai appris quelque chose : Julia Kempe, Oded Regev, Ben Toner, Ashwin Nayak, Stefano Pironio, Peter Høyer, Dan Browne, Simon Perdrix, Stephanie Wehner, Vincent Nesme, Thomas Vidick, André Chailloux, Loïck Magnin, David Steurer, Martin Roettler, Adi Rosen, Christoph Dürr, Sylvie Corteel, Pascal Ochem, Matthieu Josuat-Vergès, Matthieu Tracol, Philippe Nadeau, Adrien Vieilleribière, Abdelfattah Abouelaoualim. J'ajoute une mention spéciale pour Nisheet Vishnoi, parce nos conversations ont été nombreuses. Il a su faire preuve avec moi d'un sens pédagogique remarquable, et je considère les conversations que nous avons eues, quoique là encore trop rares, comme particulièrement enrichissantes.

Je tiens ensuite à remercier ma famille. J'aime en particulier penser que j'ai hérité ma passion pour le savoir de mon grand-père, qui est pour bien des raisons une personne qui compte beaucoup pour moi. Je suis également très redevable à mes parents de m'avoir soutenu et encouragé pendant toutes mes études. Leur dernière tâche sera de m'aider à choisir entre la confection homme et la confection femme, probablement le plus crucial des choix que j'aurai à faire. J'ai évidemment une pensée émue pour ma soeur Lydia, avec qui j'ai toujours aimé partager mes expériences. Grâce à elle, j'aurais vu ma famille accueillir deux nouveaux membres pendant ma thèse : Luc-Antoine et Maxime, respectivement son époux et son fils. Ma famille s'est agrandit d'un autre côté à l'occasion de mon propre mariage. Les Bongrand forment une sympathique communauté dans laquelle j'ai eu le plaisir et l'honneur de m'introduire. D'ailleurs, quiconque a eu l'occasion de les fréquenter a eu la même sensation d'être irrésistiblement charmé par leur sympathie et leur hospitalité.

Je ne peux pas faire la liste de tous les amis qui ont compté pendant cette période. Je me bornerais à citer Laurent et Julien, avec qui j'ai commencé mes études et qui sont aujourd'hui docteurs comme moi. J'ai profité de leurs conseils avisés dont je les remercie aujourd'hui. Dans un autre registre, voir chez mon ami Nicolas se mélanger la passion avec la rigueur et l'honnêteté a toujours été une grande source d'admiration et d'exigence.

Je terminerai comme j'ai commencé avec l'autre personne qui a été la plus importante dans l'élaboration de cette thèse. Bérengère est bien entendu celle qui a été la plus proche de moi pendant tout ce temps, celle que j'ai associé à mes joies et qui m'a aidé à surmonter mes angoisses. La forme finale de ce manuscrit doit beaucoup à ses nombreuses relectures. Le travail que je présente ici est peut-être le mien, mais l'histoire à laquelle il appartient est la notre.

Introduction

L'informatique comme science exacte

Il existe un moyen élémentaire pour aider un voyageur égaré à retrouver son chemin. Il suffit de lui indiquer sa position sur une carte. Une autre manière de le sortir de l'embarras est de lui indiquer la route menant à un lieu qu'il connaît. Ces deux procédés nous semblent tout à fait naturels, mais supposent de nombreuses étapes préalables. Il aura fallu que soit développée la science de la cartographie, que des explorateurs parcourent le monde pour tracer des cartes, que des pouvoirs publics investissent dans la construction de routes, etc...

Existe-t-il un procédé analogue pour situer un domaine scientifique dans l'ensemble des connaissances humaines ? Il y a en fait plusieurs manières de procéder. Celle qui est la plus courante est la manière dont un bibliothécaire classe un nouveau livre sur les étagères d'une bibliothèque. Une partie de la cote d'un livre est en général obtenue en classant celui-ci de manière thématique et par niveau de spécialisation croissant. Pour reprendre l'analogie avec la géographie, ceci correspond à repérer un lieu en donnant par exemple, le continent, puis le pays, la région, le département, etc...

Cette méthode de classification est peut être adaptée à une bibliothèque, mais pour le but que nous nous sommes fixé, nous avons besoin de donner plus de précisions. En effet, notre travail porte sur des disciplines scientifiques récentes, et l'existence même de celles-ci n'est pas évidente a priori. Il est donc nécessaire de justifier l'intérêt qu'on leur porte. Une manière de le faire est de décrire les liens qu'elles entretiennent avec des disciplines connexes, en positionnant chaque discipline relativement à celles avec lesquelles elle entretient des relations épistémologiques. Là encore, on retrouve la métaphore géographique puisqu'il s'agit de tracer des routes et des ponts entre ce qu'on souhaite repérer et d'autres lieux mieux connus.

Bien entendu, cette méthode n'est pas toujours facile à mettre en oeuvre. Il peut ne pas y avoir de consensus sur une une définition d'une discipline comme sous partie d'une autre. L'objectif du travail présent étant de faire progresser la connaissance d'une discipline précise, nous devons d'abord nous efforcer de rendre compte de son état de la manière la plus objective possible. En général, la science procède par spécialisation successive. Cette manière de localiser les connaissances fait apparaître une perspective historique. L'autre méthode, la localisation relativement à d'autres disciplines permet en quelque sorte de mesurer l'importance d'un domaine, comme si on mesurait les longueurs des frontières entre un état et ses voisins.

La discipline qu'on cherche à situer ici est la complexité de la communication. D'abord, le domaine dans lequel nous nous trouvons est l'informatique. Il n'est pas inutile de rappeler comment celui-ci s'est formé. Concrètement, il s'est détaché des *Mathématiques*

au cours du vingtième siècle, comme l'avait fait la *Physique* au cours du dix-huitième. Le problème qui, ayant pris une telle importance, avait conduit à considérer la physique comme une science à part entière était l'étude des mouvements des corps. De manière analogue, la question qui a conduit l'informatique à devenir autonome est de comprendre ce qu'est un calcul. Le travail que nous présentons ici se situe dans la filiation de cette question et étudie certains de ses aspects.

L'informatique partage par ailleurs certains objectifs avec la physique. En effet, si la science du calcul a aujourd'hui une telle importance, c'est qu'elle a apporté beaucoup à la connaissance du monde. La notion de calcul est considéré aujourd'hui dans le contexte plus large des processus de traitement de l'information. On appelle aussi sciences de l'information les sciences qui traitent de tels problèmes. Même si les méthodologies de l'informatique et de la physique sont différentes, elles partagent des objectifs similaires et des techniques héritées des mathématiques.

Théorie de la complexité et modèles de calculs

Comme sa cousine la physique, l'informatique se divise en une partie théorique et une partie appliquée. Nous nous trouvons ici du côté de la première. La discipline qu'on cherche maintenant à situer est la théorie de la complexité. Une manière d'y parvenir est de faire un détour par une discipline plus ancienne mais tout aussi importante, la théorie de la calculabilité. Celle-ci a pour objet de distinguer ce qui est calculable de ce qui ne l'est pas. Il s'agit donc d'une approche qualitative. Ce qu'apporte la théorie de la complexité, c'est une approche quantitative. Si la calculabilité a pour but de prouver qu'un calcul est réalisable par un ordinateur, la complexité cherche à montrer que celui-ci va s'achever en un temps raisonnable, par exemple avant la mort du système solaire.

Donnons un exemple pour illustrer l'intérêt de cette question. Nous allons pour cela tracer une nouvelle route, afin d'atteindre une discipline qu'on appelle la cryptographie. Celle-ci étudie les mécanismes permettant de chiffrer et de déchiffrer l'information. Dans les années 1970, la cryptographie a connu une véritable révolution avec la découverte de la cryptographie à clé publique. La cryptographie à clé secrète est pratiquée depuis l'antiquité, mais s'est révélée peu adaptée aux structures de communication modernes. En effet, la sécurité de la cryptographie à clé secrète implique toujours le transport d'une information qui doit rester inconnue d'un éventuel curieux. Cela peut être une clé de déchiffrement, ou la méthode de chiffrement elle-même ; garantir cette sécurité peut s'avérer difficile et coûteux. C'est véritablement la découverte de la cryptographie à clé publique [DH76] qui a permis d'introduire la cryptographie dans la vie courante, et en particulier sur internet. Sans rentrer dans les détails techniques du fonctionnement du chiffrement, ce changement de paradigme s'accompagne également d'un changement important de la définition même de sécurité. La sécurité des systèmes modernes repose en effet sur l'idée que le déchiffrement d'un message sans sa clé demande un temps trop long pour être réalisé en pratique, même s'il est théoriquement possible. La sécurité des systèmes cryptographiques à clé publique repose sur l'hypothèse que certains problèmes sont difficiles à résoudre. Ces résultats de complexité n'ont par ailleurs pas été prouvés à ce jour.

La théorie de la complexité aborde donc la notion de calcul d'une manière quantitative. Si on veut quantifier le coût d'un calcul, il faut préciser les opérations élémentaires auto-

risées, celles dont le coût est unitaire. Claude Shannon a été l'un des premiers à proposer un tel formalisme [Sha49]. Son idée a été de caractériser la complexité d'un problème par la taille du plus petit circuit booléen permettant de le résoudre. Précisément, il s'agit du nombre de portes logiques requises pour résoudre le problème. En fixant ainsi les opérations élémentaires et les règles de calcul, on définit ce qu'on appelle un modèle de calcul.

Avant de définir plus précisément le modèle qui va nous occuper ici, donnons un second exemple. L'un des modèles de calcul les plus célèbres et les plus utiles est la machine de Turing [Tur36]. Il s'agit d'une machine abstraite qui exécute des programmes en lisant et en écrivant des données sur un ou plusieurs rubans constitués de cases. En fixant des règles définissant les opérations que la machine doit réaliser en fonction de son état à un instant donné, on peut programmer celle-ci pour réaliser des calculs et résoudre des problèmes. L'importance de ce modèle en informatique théorique est considérable. En théorie de la calculabilité, la thèse de Church-Turing postule que tout ce qui est calculable peut l'être par une machine de Turing. Il ne s'agit pas d'un énoncé mathématique, mais de la croyance que ce modèle de calcul définit une notion de calcule conforme à l'intuition qu'on en a. En théorie de la complexité, les machines de Turing permettent de définir formellement deux notions extrêmement importantes :

1. le temps de calcul, défini comme le nombre total d'opérations réalisées par la machine,
2. l'espace utilisé, défini comme le nombre maximum de cases utilisées par la machine lors de l'exécution d'un programme.

Ces deux ressources sont évidemment très importantes, tant les notions qu'elles définissent semblent intuitives et utiles. On peut même aller plus loin et s'intéresser au compromis entre temps et espace. On cherche alors à étudier le principe bien connu des possesseurs d'ordinateurs qui consiste à ajouter de la mémoire pour le rendre plus rapide.

En général, on appelle algorithme un programme exécutable par une machine de Turing. Concrètement, pour une ressource donnée, la complexité d'un algorithme est son coût maximal sur toutes les instances du problème qu'il est supposé résoudre. On fixe en général l'espace nécessaire pour encoder la donnée dans un modèle informatique raisonnable, et on exprime la complexité en fonction de ce paramètre. La complexité d'un problème est alors précisément la complexité du meilleur algorithme qui le résout. L'un des objectifs majeurs de la théorie de la complexité est de classer les problèmes en fonction de leur difficulté. Un tel classement est possible en considérant le comportement asymptotique de la complexité.

Complexité de la communication

Intéressons nous maintenant plus précisément à la complexité de la communication. Dans cette discipline, on étudie des problèmes dont les entrées sont réparties entre plusieurs agents. Généralement, les agents sont appelés des joueurs, et lorsqu'il n'y en a que deux, on les nomme Alice et Bob. Le cas le plus simple est celui de l'évaluation d'une fonction booléenne. On fixe une fonction $f : X \times Y \to Z$, Alice reçoit une entrée x dans X et Bob une entrée y dans Y. Leur but est alors de calculer la valeur de la fonction f sur l'entrée x, y. Pour y arriver, il vont devoir s'échanger des informations sur leurs entrées

respectives, bit après bit, jusqu'à ce qu'un joueur soit capable de donner la valeur de la fonction. Chaque message qu'un joueur envoie à l'autre est une fonction de son entrée et des messages qu'il a précédemment reçu. L'ensemble des règles qui définissent les messages envoyés par un joueur s'appelle un protocole de communication. Il n'y a aucune restriction sur la complexité de ces règles. Les joueurs peuvent réaliser des calculs impossibles à faire en pratique, et même calculer des choses qui ne sont pas calculables au sens de Church-Turing. Seule compte la quantité de communication échangée par les joueurs. La complexité d'un protocole de communication est alors le nombre de bits échangés par les joueurs lorsqu'ils exécutent celui-ci. La complexité de la communication d'une fonction est alors la complexité du meilleur protocole de communication permettant à Alice et Bob de calculer celle-ci.

Ce modèle a été proposé par Andrew C.C. Yao en 1979 [Yao79]. L'étude de la complexité de la communication a de nombreuses applications à d'autres problèmes d'informatique parmi lesquels la profondeur des circuits booléens [KW90], la conception de circuit VLSI [Len90], les algorithmes streaming [AMS99] et l'étude des compromis temps-espace [BNS92]. Toutes ces applications ont conduit la complexité de la communication à devenir un champs de recherche autonome et important. On a ainsi un exemple manifeste de spécialisation de la science, où un problème est d'abord étudié pour ses applications, qui sont si importantes qu'il devient finalement une discipline à part entière.

Le modèle de Yao est suffisamment précis pour donner une définition de la communication conforme à l'intuition. Mais il est également assez général pour qu'on puisse considérer de nombreuses variations. On peut en effet envisager de nombreuses variantes dans les règles d'envoi des messages. On peut restreindre par exemple le nombre de tours, ou ajouter un troisième joueur qui ne reçoit pas de variable, mais qui reçoit les messages des joueurs et donne la valeur de la fonction. Nous allons maintenant présenter des variantes du modèle qui sont beaucoup plus importantes. En donnant le modèle de calcul, nous n'avons jamais eu à préciser ce qu'on attendait en sortie. Bien entendu, on attend d'un algorithme qu'il résolve le problème pour lequel il a été conçu, mais nous allons voir maintenant qu'il y a différentes manières de définir ce qu'on entend par "résoudre un problème".

Calcul probabiliste et bornes inférieures

Dans le modèle que nous avons défini plus haut, nous avons supposé que chaque bit envoyé d'un joueur à l'autre dépendait de son entrée et des messages déjà reçu. Ceci sous-entend que l'application d'un protocole est déterministe. C'est précisément ce point que nous allons changer en introduisant le calcul probabiliste. Même si formellement, il s'agit d'un nouveau modèle de calcul, le changement est si important qu'on parlera plutôt d'un nouveau de paradigme de calcul.

Une machine de Turing probabiliste est une machine dont certaines opérations peuvent être choisies aléatoirement. Pour un protocole de communication, le caractère probabiliste signifie que pour déterminer le message à envoyer, chaque joueur peut choisir son message de manière aléatoire. Bien entendu, cela ne signifie pas que les messages ne contiennent plus d'information sur les entrées, ou sur les messages précédents, mais l'aléa vient s'ajouter aux autres informations déterminant chaque message envoyé. De plus, on ne demande plus aux joueurs de calculer exactement la fonction. Ils sont autorisés à se tromper avec

une petite probabilité.

La notion de calcul probabiliste est fondamentale en informatique théorique. Cette notion est apparue à la fin des années soixante-dix. Si elle a suscité de l'intérêt, c'est qu'elle a permis de résoudre des problèmes qu'on ne savait pas résoudre efficacement de manière déterministe. L'un des algorithmes probabilistes les plus célèbres est le test de primalité de Solovay-Strassen [SS77]. Cet algorithme permet de déterminer, avec une grande probabilité de succès, si un nombre est premier. Cet algorithme est très efficace, et peut facilement être mis en oeuvre en pratique. De plus, à l'époque de sa découverte, on ne savait pas que ce problème pouvait être résolu efficacement par un algorithme déterministe. Même si on sait aujourd'hui que cela est possible [AKS04], on continue à utiliser l'algorithme de Solovay-Strassen pour sa simplicité. On pourrait penser qu'un algorithme déterministe est plus sûr qu'un algorithme probabiliste car celui-ci ne donne pas toujours la bonne réponse. Dans certains cas critiques, comme le contrôle aérien, ou le contrôle d'une centrale nucléaire, on préférera s'assurer de la certitude des réponses. En réalité, tout l'art des techniques probabilistes est de permettre de contrôler les probabilités d'erreur, pour les rendre arbitrairement petites. On peut s'assurer, par exemple, que le calcul a plus de chance d'être faux à cause d'un défaillance de matériel que d'une erreur de l'algorithme. En pratique, ceci conduit à une notion robuste et raisonnable de calcul, qui permet souvent la mise en oeuvre d'algorithmes simples et efficaces.

L'informatique théorique étudie de nombreux modèles de complexité différents. Quelles sont dans cette situation les questions que se pose un théoricien de la complexité ? Une question très importante est de comparer, pour chacun de ces modèles, leurs puissances calculatoires respectives. Par exemple, quel temps de calcul peut-on espérer gagner avec un algorithme probabiliste par rapport à un algorithme déterministe ? Une telle différence entre deux modèles s'appelle une séparation. Pour prouver une séparation entre deux modèles, on pourrait procéder suivant ce raisonnement :

1. choisir un problème et trouver un algorithme probabiliste efficace pour celui-ci,
2. montrer que tous les algorithmes déterministes sont moins efficaces que l'algorithme probabiliste fixé.

La différence entre ces deux points est très importante. Dans le premier cas, on doit trouver un algorithme, alors que dans le second, il s'agit de prouver une propriété vérifiée par tous les algorithmes résolvant le problème. En général, prouver qu'un problème est difficile, c'est-à-dire qu'il n'existe pas d'algorithme efficace, est en soi une question difficile. Montrer qu'un problème est difficile, c'est précisément prouver une borne inférieure sur sa complexité.

En fonction des modèles considérés, on ne sait pas toujours si de telles séparations existent. Pour le temps de calcul par exemple, on conjecture que la plus grande séparation possible entre le calcul déterministe et le calcul probabiliste est au plus polynomial en la taille de l'entrée. Pour la complexité en circuit, la situation est encore plus complexe. Claude Shannon a souligné le résultat suivant [Sha49]. Si on tire une fonction au hasard parmi toutes les fonctions booléennes, il est facile de voir qu'elle nécessite un nombre exponentiel de portes logiques. Cela se prouve en comparant le nombre total de fonctions avec le nombre total de circuits de taille sous-exponentielle. A côté de cela, la meilleure borne inférieure pour une fonction explicite est linéaire [IM02].

Une chose remarquable en complexité de la communication est qu'on sait prouver de

larges séparations entre les différents modèles. On a en effet des techniques efficaces pour prouver des bornes inférieures. Mais souvent, des techniques ad hoc sont développées pour des problèmes spécifiques. Prouver une séparation entre deux modèles permet de mieux comprendre la puissance d'expression de chacun d'eux et l'un des objectifs du travail présenté ici est de formaliser les différentes techniques utilisées pour prouver des bornes inférieures en complexité de la communication.

Calcul quantique

L'autre paradigme de calcul important que nous allons considérer est la complexité de la communication quantique. Ici, ce n'est pas la définition du résultat du calcul qui change, mais les opérations élémentaires sont radicalement différentes de celles permises dans les autres modèles. Commençons par rappeler l'origine du problème. En informatique classique, la loi de Moore énonce que le nombre de transistors dans les microprocesseurs double tous les deux ans (à encombrement constant) [Moo65]. Bien qu'il s'agisse d'une loi empirique, ses prédictions se sont assez bien vérifiées. La densité de transistors a des conséquences sur la mémoire, la puissance, ou la vitesse de calcul. De ce point de vue, les conséquences des prédictions de la loi de Moore sont positives. Néanmoins, cette loi signifie également qu'il y a une miniaturisation des composants. Or les lois physiques qui sont utilisées pour décrire le monde à l'échelle macroscopique ne sont plus valables lorsqu'on réduit l'échelle jusqu'au niveau atomique. On ne peut plus alors ignorer les phénomènes d'interférences quantiques, qui restent négligeables à grande échelle. Ceci implique que la loi de Moore ne peut se poursuivre, car la miniaturisation rencontre là une limite. La loi de Moore prévoit que ceci se produira vers 2020.

Au début des années quatre-vingt, le physicien américain Richard Feynman a eu l'idée d'utiliser les phénomènes quantiques à des fins calculatoires [Fey82]. Il s'agit précisément de tirer parti des phénomènes qui font obstacle à la poursuite de l'augmentation de la puissance de calcul des ordinateurs. Non seulement la physique quantique permet de coder l'information, mais elle fournit également de nouveaux outils pour la manipuler. Il y a plusieurs candidats pour être le support de l'information quantique. Cela peut être par exemple le spin d'un électron ou le niveau d'énergie d'un atome. La plus petite unité d'information quantique s'appelle un qubit. L'un des postulats de la physique quantique affirme que l'action de mesurer l'état d'un système le perturbe. Cette théorie a donc dû renoncer à donner une description déterministe de la dynamique au niveau atomique. Elle en donne au contraire une description statistique. Par exemple, un atome d'hydrogène a deux niveaux d'énergie possibles. On décrit l'évolution de l'état de l'atome comme une superposition de ces deux niveaux d'énergie. L'action de mesurer l'état de l'atome pour connaître son niveau d'énergie est alors un processus probabiliste. Les deux résultats sont a priori possibles, et l'un des objectifs de la physique quantique est de quantifier les probabilités de chaque résultat, en fonction de l'état du système. De plus, après une mesure, le système est dans l'état qu'on a observé. Ce phénomène de superposition est le premier point crucial qui fait du calcul quantique un paradigme de calcul radicalement différent du modèle classique.

La seconde originalité du calcul quantique est la manière dont les qubits se combinent. Il est en effet possible d'intriquer deux particules, de manière à ce que le résultat de la mesure de leurs états soient les mêmes. Une telle corrélation est d'abord appa-

rue comme paradoxale, et a été à l'origine de son rejet par de nombreux physiciens. En effet, l'interprétation classique de la théorie quantique, appelée interprétation de Copenhague, affirme qu'il est impossible de connaître l'état du système avant de faire la mesure. Einstein, Podolsky et Rosen [EPR35] ont souligné en 1935 le paradoxe qui porte leurs initiales, EPR. En effet, s'il est impossible de connaître l'état d'un qubit avant sa mesure, et si deux qubits peuvent être corrélés de manière à ce qu'une mesure sur le premier détermine le résultat d'une mesure sur le second, alors il semble y avoir un transfert d'information instantané entre les qubits. Ceci est alors en contradiction avec la théorie de la relativité, qui affirme qu'aucune information ne peut se propager plus rapidement que la vitesse de la lumière, ou avec le postulat de causalité qui affirme qu'aucun effet ne peut précéder les causes qui l'ont engendré.

La conséquence que voulurent en tirer Einstein, Podolsky et Rosen était que la physique quantique était une théorie incomplète. Leur idée était qu'il devait être possible d'ajouter des variables supplémentaires à la théorie pour décrire le comportement des particules. L'état de la particule serait fixé avant la mesure, mais la théorie quantique serait trop faible pour le décrire. C'est en 1964 que John Bell a montré que cette solution n'était pas possible, résolvant en théorie ce paradoxe [Bel64]. Il a montré que l'ajout de variables supplémentaires ne permettait pas de retrouver les prédictions de la physique quantique. La preuve de Bell est basée sur l'étude d'inégalités, dites de Bell, qui permettent de séparer la physique classique de la physique quantique. Ce résultat de Bell a plus tard été confirmé par une série d'expériences, dont les plus célèbres sont celle de Freedman et Clauser [FC72], puis celles menées par Aspect [AGR81, AGR82, ADR82].

La première conséquence du théorème de Bell est de prouver l'existence de théories dites non-locales, c'est-à-dire qu'on ne peut considérer deux états intriqués indépendamment l'un de l'autre, ni corrélés classiquement. Plus important encore, on peut expliquer la raison pour laquelle la théorie quantique ne permet pas de transfert instantané d'information. En effet, le paradoxe disparaît lorsque les mesures faites sur chaque particule sont inconnues à l'avance. Imaginons qu'on choisisse indépendamment les deux mesures, et qu'on mesure deux particules intriquées en même temps. L'important n'est pas alors que le résultat de la première mesure puisse permettre d'apprendre quelque chose sur le résultat de la seconde. L'important est que le résultat d'une mesure ne donne aucune information sur le choix de la mesure qui a été faite sur l'autre particule. Autrement dit, l'utilisation de particules intriquées ne permet pas de communiquer ce choix.

D'un point de vue algorithmique, l'intérêt principal qu'on porte à l'informatique quantique vient du fait qu'elle permet de réaliser des tâches qu'on ne sait pas réaliser de manière aussi efficace avec des ressources classiques. C'est en cryptographie que sont venues les premières avancées importantes. La technique du codage quantique [Wie83], puis le protocole de Bennett et Brassard [BB84] ont donné de nouveaux horizons à la cryptographie. Ce dernier permet en effet de distribuer de manière sécurisée et efficace des clés pour le chiffrement à clé secrète. Ceci permet de concevoir des protocoles de chiffrement quantique inconditionnellement sûrs. C'est d'ailleurs la partie de l'informatique quantique la plus opérationnelle, et il existe aujourd'hui des applications pratiques et même commerciales de ce protocole.

Si d'un côté, l'informatique quantique offre de nouvelles possibilités en cryptographie, elle fait peser une menace sur la cryptographie classique. L'agorithme de Shor permet de résoudre de manière efficace le problème de la factorisation [Sho97], et par là de casser le

système de chiffrement RSA, un des protocoles de cryptographie à clé publique les plus utilisés. C'est véritablement tout le paysage de la cryptographie qui se trouverait modifié par la réalisation pratique d'un ordinateur quantique.

Il existe un second résultat de complexité en informatique quantique très intéressant. En 1996, L.K. Grover a montré qu'il existait une séparation quadratique entre le calcul classique et le calcul quantique, pour le problème de la recherche dans une base de données non triée [Gro96]. Ce problème est d'une part très général, et d'autre part, la borne inférieure classique est très facile à prouver, ce qui rend le résultat de Grover d'autant plus surprenant. La théorie de la complexité a fortement contribué à rendre l'informatique quantique attrayante. Il y a donc un lien très fort entre ces deux disciplines.

Au long de cet exposé sommaire du calcul quantique, nous avons rencontré les mots "communication" et "information". Nous allons voir que le lien entre le calcul quantique et la complexité de la communication est fondamental. Mais avant, nous allons présenter comment on formalise, en informatique, la notion d'information et comment elle est utilisée en complexité de la communication.

Mesurer l'information

Une manière de prouver des bornes inférieures en complexité de la communication est de considérer le modèle sous l'angle de la théorie de de l'information. L'idée est simple. Il s'agit pour un problème donné de quantifier l'information que les joueurs doivent échanger sur leurs entrées pour résoudre le problème. Voici une présentation informelle de la manière d'articuler cette idée. Le problème de l'égalité est défini de la manière suivante : Alice et Bob reçoivent respectivement des chaînes binaires x et y, et leur but est de déterminer si celles-ci sont égales ou différentes. Supposons que les joueurs reçoivent soit des entrées partout égales sauf sur un seul indice inconnu, soit des entrées partout égales. L'information échangée doit permettre de distinguer des entrées qui conduisent à des réponses différentes. Or, un bit de communication envoyé par Alice ne permet à Bob d'identifier qu'au plus un bit de la chaîne d'Alice. Donc, tant qu'il n'a pas reçu suffisamment d'information pour identifier toute la chaîne d'Alice, Bob ne peut conclure. Autrement dit, dans le cas déterministe, le meilleur protocole consiste, pour Alice, à envoyer toute sa chaîne à Bob pour qu'il puisse décider si elle est égale à la sienne ou non. Ceci ne constitue pas une preuve au sens strict mais on peut formaliser cet argument et prouver le résultat en utilisant cette intuition.

Pour formaliser la notion d'information, on peut utiliser la théorie de l'information de Shannon [Sha48]. L'objectif de Shannon était d'étudier des problèmes de communication, il n'est donc pas étonnant que sa théorie s'applique à la complexité de la communication. Cette théorie a en effet été utilisée ad hoc pour prouver des bornes inférieures pour des problèmes donnés [BYJKS02]. Il existe aussi une méthode générale basée sur cette idée, appelée *coût en information* [BYJKS04]. Cette méthode a plusieurs applications [JKS03]. Elle permet de formaliser l'idée présentée au paragraphe précédent. Plus récemment, la *borne de sous-distribution* [JKN08] a été utilisée pour formaliser et généraliser la méthode des rectangles, une méthode élémentaire de borne inférieure basée sur la structure combinatoire des problèmes.

Au chapitre 2, nous allons utiliser un autre concept pour formaliser la notion d'information. Il s'agit de la complexité de Kolmogorov [Sol64, Kol65, Cha69]. Cette idée définit

la quantité d'information contenue dans une chaîne booléenne comme la longueur du plus court programme qui calcule cette chaîne. L'ambition de Kolmogorov était de donner une définition robuste à la notion d'aléatoire. Supposons qu'on forme une chaîne binaire en tirant 20 fois une pièce de monnaie et en notant 0 pour face et 1 pour pile. La théorie classique des probabilités ne fait pas de distinction entre les chaînes 00000000000000000000 et 01001000101100110110. L'expérience a la même probabilité de donner ces deux résultats, alors que personne ne peut croire raisonnablement que le premier résultat a été obtenu en utilisant une pièce non-biaisée.

La complexité de Kolmogorov permet d'opérer une distinction entre ces deux chaînes. Intuitivement, la première chaîne ne semble pas issue d'une série de tirages aléatoires indépendants. La raison est qu'on peut en donner description particulièrement simple. En utilisant une machine de Turing, on peut la produire avec l'algorithme "écrire 20 fois 0". En revanche, il ne semble pas y avoir de meilleur programme pour décrire la seconde chaîne que de donner successivement tous les bits de celle-ci. On dit alors que cette chaîne est incompressible. L'idée de la complexité de Kolmogorov est qu'un tirage aléatoire est imprévisible, et par conséquent qu'une série de tels tirages doit être difficile à décrire. Autrement dit, une chaîne aléatoire doit contenir beaucoup d'information.

La complexité de Kolmogorov présente de grandes similitudes avec la théorie de l'information. Néanmoins, la mise en oeuvre de méthodes basées sur la théorie de l'information pour prouver des bornes inférieures en complexité de la communication nécessite souvent l'utilisation d'outils statistiques profonds. Notre objectif en utilisant la complexité de Kolmogorov est de prouver des bornes inférieures en dégageant la nature combinatoire des preuves.

La complexité de Kolmogorov a déjà été utilisée pour prouver des bornes inférieures de complexité dans différents modèles : nombre de requêtes classiques et quantiques [LM08], temps de calcul [PSS84], complexité de la communication [BJLV00] et bien d'autres. La méthode qui permet de prouver des bornes inférieures en utilisant la complexité de Kolmogorov s'appelle la méthode d'incompressibilité (voir à ce sujet [LV08]). L'un des buts du chapitre 2 est de formaliser la méthode d'incompressibilité spécifiquement pour la complexité de la communication. Ainsi, nous utilisons la complexité de Kolmogorov pour donner une borne inférieure générale sur la complexité de la communication et pour identifier les entrées difficiles à calculer.

Nous faisons également une troisième utilisation de la complexité de Kolmogorov pour donner une version alternative du principe du min-max de Yao. Ce principe, énoncé par Yao en 1977 [Yao77], permet de transformer un protocole probabiliste en un protocole déterministe faisant des erreurs. Ce principe est l'un des résultats les plus importants en informatique théorique, et est à la base de nombreuses preuves de bornes inférieures. L'idée en utilisant la complexité de Kolmogorov est de choisir comme aléa une chaîne booléenne incompressible. En choisissant l'aléa et les entrées de manière indépendante, nous obtenons des propriétés qui permettent de simplifier les preuves.

La communication comme mesure de la non-localité

Le paradoxe EPR que nous avons présenté plus haut peut être présenté dans un formalisme proche de la complexité de la communication. Le scénario implique deux joueurs et la règle est la suivante :

– Alice et Bob partagent deux états quantiques intriqués.
– Chaque joueur reçoit une entrée spécifiant une mesure à faire sur son état.
– Chaque joueur donne le résultat de sa mesure.

La mesure étant un processus probabiliste, on obtient pour chaque paire d'entrées une distribution de probabilité sur les réponses possibles des joueurs. Cette distribution vérifie la causalité, c'est-à-dire que la réponse d'un joueur ne contient pas d'information sur l'entrée de l'autre.

Le résultat de Bell peut également être traduit dans ce modèle. L'ajout de variables supplémentaires revient ici à remplacer l'intrication par de l'aléa partagé entre les joueurs. Le théorème affirme alors qu'il existe des distributions obtenues en mesurant des états intriqués qui ne peuvent pas être reproduites avec seulement de l'aléa partagé. Dans quelle mesure la théorie quantique est-elle éloignée de la théorie classique ? Peut-on quantifier la non-localité de la physique quantique ?

En 1992, le philosophe Tim Maudlin a proposé un moyen pour quantifier la non-localité [Mau92]. Puisqu'il existe des distributions de probabilité issues d'une expérience quantique qu'on ne peut simuler de manière classique, Maudlin a proposé d'ajouter des ressources supplémentaires, par exemple de la communication. La quantité de communication à ajouter devient alors une mesure de la non-localité. Plusieurs auteurs ont par la suite, parfois indépendamment les uns des autres, poursuivi ou redécouvert ce travail. Le tableau 1 résume les différents résultats de simulation des distributions quantiques.

Auteur(s)	Année	Com.	Configuration	dim.	Type
Maudlin [Mau92]	1992	1.17	moyenne	2	plan
Brassard, Cleve, Tapp [BCT99]	1999	8	pire cas	2	sphère
Steiner [Ste00]	2000	1.48	moyenne	2	plan
Cerf, Gisin, Massar [CGM00]	2000	1.19	moyenne	2	sphère
Toner, Bacon [TB03]	2003	1	pire cas	2	sphère
Regev, Toner [RT07]	2007	2	pire cas	≥ 2	sphère

Tab. 1 – Résultats de simulation des distributions quantiques - La colonne configuration spécifie si la quantité de communication est comptée en moyenne sur les entrées, ou en pire cas ; La colonne type d'état précise sur quel état les mesures sont effectuées ; La dernière colonne indique le type de mesure qu'on considère.

Toutes ces mesures vérifient deux propriétés supplémentaires qui ne sont pas spécifiées dans le tableau. Tout d'abord, les mesures sont dites binaires, c'est-à-dire qu'elles n'ont que deux résultats possibles. La seconde demande quelques explications préliminaires sur la structure des distributions de probabilités quantiques. Supposons que les deux résultats possibles des mesures sont -1 et 1, et notons A et B les variables aléatoires qui désignent les résultats des mesures de chaque joueur. L'espérance de la variable aléatoire A est appelée la marginale d'Alice. Intuitivement, il s'agit de la valeur moyenne qu'Alice obtient en répétant l'expérience. On définit la marginale de Bob de manière identique. L'espérance du produit AB est appelée la corrélation des deux variables. Si la corrélation vaut 1, alors les deux variables sont toujours égales. Si elle vaut -1, elles sont toujours différentes. Nous allons montrer plus loin que toute distribution causale binaire est parfaitement définie par sa corrélation et ses marginales. La deuxième propriété commune aux protocoles de

la table 1 est de reproduire uniquement les corrélations, sans les marginales.

Pour le protocole de Regev et Toner par exemple, il y a deux cas possibles. Soit la distribution quantique a des marginales uniformes, et dans ce cas, le protocole reproduit exactement la distribution ; pour toute autre distribution, le protocole ne reproduit que partiellement celle-ci. Nous verrons plus loin que les distributions à marginales uniformes jouent un rôle fondamental en physique quantique. Néanmoins, dans le cas général, on ne sait pas s'il est possible de simuler les distributions quantiques avec une quantité finie de communication. A première vue, il semble qu'une fois la corrélation simulée, les marginales sont issues d'une transformation locale, et devraient être faciles à modifier. Pourtant, ce problème est toujours ouvert. Dans le cas particulier où Alice et Bob partagent des qubits, Toner et Bacon ont également proposé un protocole pour simuler la distribution complète [TB03] avec deux bits de communication.

Simuler les distributions causales : un cadre de calcul général

On a vu à la section précédente que le problème de la simulation des distributions quantiques était loin d'être résolu dans le cas général. Ce qu'on cherche à comprendre en étudiant cette question, c'est de manière générale la façon dont la physique quantique permet de manipuler l'information. Par exemple, on a dit plus haut que l'utilisation d'états intriqués ne permettait pas de communiquer instantanément. Cette conséquence de la causalité est exactement le type de phénomène qu'on souhaite étudier.

Les mécanismes de traitement de l'information relèvent par définition de l'informatique. Si l'informatique générale permet de comprendre certains aspects calculatoires de la physique quantique, la complexité de la communication offre un cadre très naturel pour étudier les expériences de type EPR. Une démarche scientifique naturelle est, pour saisir les spécificités du problème étudié, de le prendre sous la forme la plus générale et la plus conceptuelle possible. C'est ce que nous tentons de faire en étudiant la complexité de la communication des distributions causales.

Le premier axiome que doit vérifier une théorie physique est probablement la causalité, c'est-à-dire le fait que les causes précèdent leurs effets. Il est en effet acquis que toute théorie physique raisonnable doit vérifier cette propriété. Nous avons dit plus haut que les distributions issues de mesures sur des états intriqués vérifiaient cette propriété. De même, les distributions locales, issues d'expériences classiques la vérifient également. De plus, Boris Tsirelson a prouvé en 1985 que les distributions quantiques vérifiaient des inégalités similaires aux inégalités de Bell [Tsi85]. Ce résultat de Tsirelson implique directement l'existence de distributions causales qui ne sont pas quantiques.

Des résultats comme celui de Tsirelson donnent un aspect de la structure du problème considéré. Pour notre problème, qui est de simuler les distributions causales, l'étude de cette structure est très importante ; elle nous est en effet utile pour prouver des bornes inférieures et supérieures. Il existe une littérature riche dans ce domaine. On sait par exemple que l'ensemble des distributions causales peut être représenté par un polytope dans un espace vectoriel. Parmi les résultats importants, il y a d'abord la caractérisation des points extrémaux de ce polytope dans certains cas particuliers [BLM$^+$05, JM05]. Un autre résultat très original est la caractérisation des distributions causales comme enveloppe affine des distributions locales. Comme le problème de Maudlin cité plus haut, ce résultat a été prouvé plusieurs fois de manière indépendante [RF81, FR81, Gro85, Wil92,

DKLR09]. Il s'agit aujourd'hui d'un résultat important, qui a été également été utilisé en pratique [Bar07, BBLW07]. Bien que nous n'utilisions pas ce résultat directement ici, les propriétés que nous étudions semblent lui être liées.

Le modèle de complexité que nous définissons permet aussi de d'étudier la complexité de la communication des fonctions booléennes. Appliquée à celles-ci, notre méthode se trouve être équivalente à la borne inférieure de Linial et Shraibman basée sur les normes de factorisation [LS08b]. Cette méthode de borne inférieure a été proposée en 2006, et permet de généraliser de nombreuses méthodes connues : Discrepancy [CG88], coefficients de Fourrier [Raz95, Kla07], norme Ky-Fan [Kla07], et récemment méthode du rang approché [LS09]. Le point commun entre toutes ces méthodes est que pour définir le problème, on a besoin d'une structure algébrique. Par exemple, une fonction booléenne à deux variables peut être vue comme une matrice dont les coefficients sont 0 ou 1.

L'approche de Linial et Shraibman est basée sur la géométrie des espaces de Banach. La notre est, au sens de la programmation semi-définie, duale de celle de Linial et Shraibman. Intuitivement, la preuve de notre borne inférieure est basée sur l'idée que toute distribution de probabilité causale peut être diluée en ajoutant du bruit, jusqu'à devenir locale. Une distribution ainsi diluée peut être simulée sans communication. Notre borne inférieure est la quantité de bruit à ajouter pour rendre une distribution locale. Cette grandeur est une mesure de la distance entre une distribution et l'ensemble des distributions locales.

Il existe peu de bornes inférieures sur la simulation des distributions quantiques. Le protocole de Regev et Toner permet de simuler les distributions obtenues par des mesures binaires sur des états maximalement intriqués en utilisant deux bits de communication. Il a récemment été prouvé que ce protocole était optimal [VB09]. Toutefois, ceci est prouvé pour un espace de mesures continu, un cas que nous ne traitons pas ici. Brassard, Cleve et Tapp [BCT99] ont donné un exemple de distribution nécessitant, en fonction de la taille de l'entrée, une quantité linéaire de communication pour être simulée exactement. La preuve de ce résultat est basée sur une séparation entre communication classique et quantique pour un problème partiel [BCW98]. Récemment Gavinsky a également prouvé un borne inférieure sur la communication nécessaire pour simuler une distribution, là encore en prouvant d'abord une séparation entre communication classique et quantique pour un problème relationnel [Gav09].

Nous montrons également des bornes supérieures sur la quantité de communication requise pour approximer les distributions causale. Dans le cas quantique, notre méthode est équivalente à celle de Shi et Zhu [SZ08]. Toutefois, nos résultats utilisent encore la géométrie de la structure des distributions, là où Shi et Zhu utilisaient les normes d'opérateurs, qui sont des objets bien plus complexes.

Pour prouver ces bornes, nous montrons un nouveau résultat sur la structure des distributions. Dans le cas des distributions binaires à marginales uniformes, on peut borner la distance entre l'ensemble des distributions quantiques et classiques en utilisant l'inégalité de Grothendieck. Mais les éléments structurels spécifiques à ce cas ne s'étendent pas au cas général, ce qui interdit d'appliquer directement l'inégalité de Grothendieck. Notre résultat principal est de donner une borne supérieure sur cette distance dans le cas général, étendant ainsi cette application de l'inégalité de Grothendieck au cas général.

Au delà de théorie quantique, la non-localité comme ressource de calcul

Nous allons voir plus loin que l'utilisation de ressources quantiques permet, pour certains problèmes, de baisser la complexité de la communication. Néanmoins, il existe également des problèmes pour lesquels il n'y a pas de grande séparation entre complexité classique et quantique. La première limite de la physique quantique dans sa capacité à manipuler l'information est la causalité. Des mesures sur des états intriqués ne permettent pas de transmettre de l'information de manière instantanée. Par ailleurs, nous avons dit que le théorème de Tsirelson, en impliquant l'existence de distributions causales non-quantiques, donnait une idée de la structure des distributions causales. Quelles sont les implications de ce théorème sur le traitement de l'information en physique quantique ? On peut étudier cette question en introduisant une nouvelle ressource théorique, les boîtes non-locales. Avant cela, il nous faut légèrement changer de contexte.

On peut présenter les résultats de Bell et de Tsirelson de manière légèrement différente en considérant les jeux multiprouveurs sans interaction. En particulier, un célèbre résultat fait apparaître une séparation entre les ressources classiques et les ressources quantiques. Il s'agit du jeu CHSH [CHSH69]. Dans ce jeu, un vérifieur envoie deux questions à deux joueurs. Ces questions sont deux bits x et y. Les questions sont choisies uniformément au hasard parmi les questions possibles. Les joueurs répondent par un bit chacun, respectivement a et b, et ils gagnent si la parité des sorties $a \oplus b$ est égale à $x \wedge y$. La probabilité de gagner en utilisant des ressources classiques est de 0,75 alors que celle de gagner en utilisant des ressources quantiques est environ de 0,85. De plus, Tsirelson a montré que cette probabilité était optimale [Tsi85] et qu'aucun protocole quantique ne pouvait faire mieux. Un question fondamentale en physique quantique est de comprendre la signification de cette limite. Quelle propriété empêche d'aller au delà ? Les distributions qui la dépassent respectent-elles toujours la causalité par exemple ?

Popescu et Rohrlich ont défini une ressource permettant de gagner avec certitude au jeu CHSH, tout en respectant la causalité [PR94]. Cette ressource s'appelle une boîte non-locale. Par le théorème de Tsirelson, on sait que cette ressource n'est pas quantique. La théorie des boîtes non-locales est parfaitement correcte d'un point de vue mathématique. On pense pourtant qu'elle est fausse d'un point de vue physique, c'est-à-dire que les boîtes non-locales ne peuvent exister dans la nature.

Pour montrer que cette théorie n'est pas réaliste, il suffit de montrer qu'elle entraîne des contradictions, ou du moins des conséquences invraisemblables. Et c'est précisément ce qui a été fait avec les boîtes non-locales. En effet, van Dam a montré que si une telle ressource existait dans la nature, tous les problèmes de communication pourraient être résolus avec un seul bit de communication [vD05]. Ceci semble contredire la perception que l'on a du monde. Ce raisonnement est tout à fait remarquable. Alors qu'à l'origine, on voulait comprendre la puissance calculatoire de l'informatique quantique, nous voilà avec le raisonnement inverse, cherchant à montrer que la physique quantique est la seule qui permette un traitement de l'information conforme à l'intuition qu'on en a. Les boîtes non-locales n'étant qu'une théorie parmi les théories plus fortes que la physique quantique, l'objectif suivant est de montrer que toutes ces théories impliquent des conséquences aussi invraisemblables. Plusieurs résultats ont depuis été prouvés dans ce sens [BBL$^+$06, FWW09, BS09].

Tout aussi remarquable qu'il soit, ce raisonnement nécessite de prendre quelques précautions. La complexité de la communication n'est pas a priori un modèle de calcul réaliste. En effet, ce modèle se concentre exclusivement sur la communication et donne aux joueurs une puissance de calcul illimitée. Ils peuvent résoudre des problèmes qui ne sont pas calculables au sens de Church-Turing. Si on veut comparer les résultats de complexité de la communication avec l'expérience qu'on a du monde réel, il parait naturel de se limiter à des protocoles de communication réalistes. Ce qui paraît réaliste du point de vue de la théorie de la complexité, ce sont des algorithmes dont le coût est polynomial en la taille des entrées. La prise en compte de telles limitations conduit à étudier l'utilisation des boîtes non-locales du point de vue de la théorie de la complexité, et c'est ce que nous faisons au chapitre 4. L'argument de van Dam semble impliquer que le modèle de complexité en boîte non-locale n'est pas réaliste. Ce que nous voulons vérifier, c'est si c'est bien le modèle qui est irréaliste et pas les algorithmes proposés.

Il y a une autre motivation à ce travail. Au chapitre 3, nous étudions la complexité de la communication des distributions causales. Or, un protocole de communication n'est évidemment pas un processus causal, car il y a manifestement un transfert d'information entre les joueurs. Il semble naturel alors, pour quantifier la non-localité d'un processus causal, d'utiliser une ressource qui vérifie la causalité. Les boîtes non-locales semblent être une ressource naturelle pour cela. L'idée est donc de mesurer la non-localité en comptant le nombre de boîtes non-locales requises pour simuler une distribution causale [BP05, CGMP05].

Enfin, l'utilisation de boîtes non-locales comme ressource pour le calcul permet de réaliser une tâche cryptographique. En effet, puisque les boîtes non-locales ne permettent pas de transférer de l'information entre les joueurs, l'évaluation d'une fonction booléenne avec de telles boîtes interdit à un joueur d'apprendre la moindre information sur l'entrée de l'autre. En cryptographie, on dit que l'évaluation de la fonction est sécurisée.

Ce lien entre boîtes non-locales et cryptographie a déjà été plusieurs fois étudié [WW05, BCU$^+$06]. L'objectif ici est d'appliquer les conséquences de l'étude de la complexité en boîte non-locale à la complexité de l'évaluation sécurisée. Traditionnellement, les ressources qui permettent de réaliser cette tâche sont soit le ET sécurisé, soit un boîte qui réalise un tâche appelée "oblivious transfer". Notre approche permet de donner les bornes sur la quantité de ressources nécessaires pour réaliser l'évaluation sécurisée, repondant à une question posée par Beimel et Malkin en 2004 [BM04].

Bornes inférieures algébriques et combinatoires

Pour prouver des bornes inférieures en complexité de la communication, nous avons étudié deux méthodes. La première, basée sur la complexité de Kolmogorov, utilise des concepts proches de la théorie de l'information. Elle généralise plusieurs méthodes connues et permet de dégager la structure combinatoire des preuves. La seconde utilise la structure géométrique des distributions causales. Elle est définie pour toutes les distributions causales, et permet en particulier de retrouver la borne inférieure de Linial et Shraibman pour les fonctions booléennes, qui à son tour généralise de nombreuses techniques connues [LS08b]. La preuve en elle-même présente un intérêt, car elle considère le dual de la quantité utilisée par Linial et Shraibman.

Ces deux méthodes ont en commun le fait d'être très générales, dans le sens où elles

permettent de généraliser plusieurs méthodes connues. Une question naturelle est de demander comment les deux se comparent l'une à l'autre. Nous n'avons malheureusement pas de réponse précise à cette question mais seulement quelques indications. Certains résultats permettent de comparer des méthodes déduites de nos méthodes principales.

- La distribution quantique de Deutsch-Josza [BCT99] nécessite une quantité linéaire de communication. Nous prouvons que les méthodes algébriques ne permettent pas de retrouver ce résultat. La preuve originale de la borne inférieure est basée sur une séparation exponentielle entre communication classique et quantique, pour le problème EQ', un problème partiel [BCW98]. Or cette séparation est prouvée justement en utilisant la méthode des rectangles, une méthode déduite de la méthode de Kolmogorov.
- La VC-dimension d'une matrice permet de donner une borne sur la complexité de la communication à sens unique du problème de communication associé. De plus, cette méthode est déduite de la méthode de Kolmogorov. Linial et Shraibman ont prouvé qu'elle est incomparable avec leur méthode.
- La borne de corruption, formalisée après avoir servi plusieurs fois de manière implicite [Raz92, KS92], est également déduite de la méthode de Kolmogorov. Or on sait que pour tout problème cette méthode est au moins aussi bonne que la méthode "discrepancy" [BPSW06], qui est une technique algébrique qu'on peut retrouver en utilisant le résultat de Linial et Shraibman [LS08b].

Le premier résultat est sur les distributions de probabilité, tandis que les deux suivants portent sur les fonctions booléennes. Ces résultats semblent indiquer que les méthodes combinatoires sont plus fortes dans le modèle probabiliste. En effet, elles peuvent être utilisées pour prouver des séparations plus grandes. Pourtant, il est difficile d'imaginer comment comparer les méthodes générales.

Une des ces deux méthodes peut-elle caractériser la complexité de la communication ? Si on savait montrer qu'une technique était meilleure que l'autre, cela nous donnerait un meilleur candidat pour cette caractérisation. L'une des conjectures les plus importantes en complexité de la communication est celle du *log rank* [MS82]. Elle affirme que la complexité de la communication est équivalente, à un polynôme près, au logarithme du rang de la matrice qui définit le problème. La force de cette conjecture est qu'elle peut être formulée, sous des formes légèrement différentes, à toutes les variantes du modèle de base. Par exemple, la même conjecture existe pour la complexité de la communication probabiliste, mais avec une version approchée du rang. Or, il a été prouvé que la méthode de Linial et Shraibman était équivalente (à un polynôme près) au rang approché [LS09]. Ainsi, la conjecture est équivalente à montrer que la méthode de Linial et Shraibman caractérise la complexité de la communication des fonctions booléennes. Ceci impliquerait alors qu'il n'existe pas de grande séparation entre communication classique et quantique. Dans le cas général, on sait que ceci est faux car on connaît des fonctions (partielles) pour lesquelles de larges séparations existent. Néanmoins on conjecture que c'est vrai pour les fonctions totales. L'une des raisons qui rend cette conjecture réaliste, c'est qu'elle est vraie dans un autre modèle, la complexité en requête [BBC+01]. On ne sait pas vraiment comment interpréter le fait que les fonctions sont totales en terme d'information. Les méthodes algébriques pourraient fournir une interprétation structurelle de cette propriété.

On voit ainsi que l'étude de ces mesures de complexité a une portée très importante. La partie visible, immédiate et appliquée à l'informatique est sans doute la question

des séparations entre la communication classique et la communication quantique. Mais ce n'est qu'un cas particulier d'une question plus large. Le principal enjeu derrière ces questions est de mieux comprendre ce modèle de calcul, et à travers cela la physique quantique comme processus de traitement de l'information.

Résumé des résultats

Nous résumons ici les résultats présentés dans la suite. Les théorèmes, lemmes, corollaires et propositions écrits en gras sont, à notre connaissance, nouveaux.

Chapitre 1

Dans le premier chapitre, nous allons exposer les concepts qui nous seront utiles dans la suite. Ceci recouvre la complexité de la communication, la complexité de Kolmogorov et la théorie des distributions causales.

La section 1.1 présente la complexité de la communication. Nous définissons d'abord le modèle déterministe, et détaillons la structure combinatoire de celui-ci. La proposition 1.1 et le corollaire 1.1 rappellent le lien entre complexité de la communication déterministe et partition en rectangles monochromatiques. Nous donnons également la définition de rectangle ε-monochromatiques.

Nous définissons ensuite les modèles probabiliste et distributionnel. Dans le cas probabilitste, nous rappelons le théorème de Newman (théorème 1.1) sur l'aléa d'un protocole de communication probabiliste. Nous rappelons ensuite le théorème du minmax de Yao (théorème 1.2), qui montre que les deux modèles sont équivalents.

Le troisième modèle que nous définissons est le modèle de communication quantique. Pour cela, nous donnons d'abord une courte introduction à l'informatique quantique. Ceci nous permet de présenter le concept d'intrication, et de donner un premier aperçu du paradoxe EPR et du théorème de Bell. Nous pouvons ensuite définir les deux modèles de communication quantique : le modèle de Yao, et le modèle de Buhrman et Cleve. Nous terminons cette introduction en rappelant un résultat important de communication quantique : le théorème de factorisation de Yao, Kremer et Klauck. Nous terminons la section sur la complexité de la communication en rappelant quelques méthodes de bornes inférieures connues.

La section 1.2 donne les bases de la complexité de Kolmogorov. Nous définissons directement la complexité préfixe, plus pratique pour nos applications. La proposition 1.3 donne une borne supérieure sur la complexité. Cette borne est optimale à une constante additive près, comme le montre l'existence d'éléments incompressibles (proposition 1.4). Nous donnons ensuite quelques manipulations élémentaires utiles sur la complexité de Kolmogorov (proposition 1.5 et corollaire 1.2). Enfin, nous montrons comment intégrer des éléments de théorie du codage à la complexité de Kolmogorov (proposition 1.6 et 1.7).

Dans la section 1.3, nous donnons quelques éléments de la théorie des distributions causales. Après avoir défini le problème de la simulation, nous donnons un aperçu de la structure géométrique du problème, en définissant les probabilités locales et quantiques.

Nous détaillons plus particulièrement la structure des distributions binaires, pour lesquelles nous donnons une représentation vectorielle spécifique (proposition 1.9). Nous montrons les liens entre ces distributions et les fonctions booléennes, et rappelons le théorème de Tsirelson (théorème 1.10), qui caractérise les distributions quantiques. Nous donnons également la caractérisation des points extrémaux du polytope causal dans le cas des distributions binaires à marginales uniformes (théorème 1.11).

Enfin, dans une dernière section, nous donnons l'inégalité de Hoeffding (théorème 1.12) qui permet de borner la déviation d'une somme de variables aléatoires. Nous en donnons ensuite un corollaire appliqué à l'informatique (corollaire 1.3).

Chapitre 2

Au chapitre 2, nous étudions les propriétés combinatoires de la complexité de la communication. Les résultats de cette section sont présentés dans l'article *Kolmogorov complexity and combinatorial methods in communication complexity* [KL09], écrit en collaboration avec S. Laplante.

Nous donnons d'abord une borne inférieure générale sur la complexité de la communication (**théorème 2.1**). Ce théorème utilise la notion d'information mutuelle au sens de Kolmogorov. Nous appliquons ensuite notre théorème pour prouver la borne inférieure des rectangles (théorème 1.7) et la borne de corruption (théorème 1.8).

Dans la section 2.3, nous nous intéressons à la complexité de la communication probabiliste. Nous donnons une borne inférieure générale pour ce modèle (**théorème 2.3**), ainsi qu'une alternative au principe de Yao, basée sur la complexité de Kolmogorov (**lemme 2.1**).

Nous donnons ensuite deux applications de notre méthode dans le cas probabiliste. Dans la section 2.4.1, nous montrons comment prouver les bornes inférieures VC-dimension et coefficients d'éclatement (théorème 2.4). Dans la section 2.4.2, nous appliquons notre méthode à un problème concret : le problème du couplage caché. Nous montrons qu'on peut retrouver la borne inférieure en $\sqrt{n}$ (théorème 2.5). Notre preuve se base sur une propriété des graphes aléatoires, que nous avons développée pour ce problème (**theorème 2.7**). Ce résultat est présenté à la section 2.4.3

Enfin, à la section 2.5, nous montrons un résultat comparant la communication à sens unique avec la communication simultanée. Nous montrons que les deux modèles sont équivalents dans le cas de la complexité distributionnelle à marginales uniformes. Comparé aux résultats précédemment connus, notre analyse améliore sensiblement le traitement des erreurs des protocoles (**théorème 2.8**).

Chapitre 3

Les résultats du chapitre 3 sont issus de l'article *The communication complexity of non-signaling distributions* [DKLR09], écrit en collaboration avec J. Degorre, S. Laplante et J. Roland.

La première section définit le concept de dilution d'une distribution. Il s'agit de l'opération consistant à mélanger une distribution de probabilité avec du bruit. Nous montrons qu'on peut ajouter du bruit à une distribution de manière à la rendre locale

(**théorème 3.1**). Nous utilisons cette propriété pour prouver les bornes inférieures de la section 3.2 (**théorème 3.2**).

Dans la section 3.3, nous étudions en particulier la simulation des distributions binaires à marginales uniformes. Nous donnons quelques résultats de structure spécifiques à ce cas là (**proposition 3.2 et lemme 3.1**). Ceci nous permet de donner une nouvelle expression de nos mesures de complexité comme jauge des ensembles des distributions locales et quantiques (**propositions 3.3 et 3.4**). Enfin, les symétries spécifiques au cas des marginales uniformes permettent d'améliorer les bornes inférieures obtenues dans le cas général (**théorèmes 3.4, 3.5 et 3.6**).

Dans la section 3.4, nous donnons une expression duale de nos bornes (**théorème 3.7**). Pour prouver cette expression, nous utilisons la dualité au sens de Lagrange. Ces expressions nous permettent d'interpréter nos mesures de complexité de manière très naturelle en termes de violation d'inégalités de Bell et de Tsirelson. Nous appliquons ensuite nos travaux au jeux XOR. Nous montrons qu'il existe une bijection entre les jeux XOR et les inégalités de Bell sur les corrélation (**lemme 3.3**). Ceci nous permet de relier notre borne inférieure avec la valeur des jeux XOR (**théorème 3.8**).

Dans la section 3.5, nous comparons dans le cas général les violations des inégalités de Bell et de Tsirelson. Dans le cas des distributions binaires à marginales uniformes, on peut faire la comparaison des deux en utilisant l'inégalité de Grothendieck (théorème 3.9). Nous étendons cette analyse au cas général (**théorème 3.10**). Ceci est un nouveau théorème de structure sur l'ensemble des distributions causales.

Enfin, dans une dernière partie, nous donnons des bornes supérieures sur le problème de la simulation des distributions causales avec erreur. Nous donnons d'abord deux nouveaux protocoles pour calculer des fonctions booléennes (théorèmes 3.12 et 3.13). Nous donnons ensuite deux protocoles pour simuler les distributions arbitraires : le premier dans le modèle classique simultané avec aléa partagé, le second dans le modèle quantique simultané sans ressource partagée(**théorème 3.14**). En combinant ces résultats avec les bornes inférieures et le théorème 3.10, cela nous donne des protocoles pour simuler sans interaction la communication quantique avec intrication (**théorème 3.3**).

Chapitre 4

Dans le chapitre 4, nous traitons de la complexité en boîte non-locale. Les résultats présentés font l'objet d'un rapport technique dont le titre est *Non-local box complexity and secure function evaluation* [KKLR09], écrit en collaboration avec I. Kerenidis, S. Laplante et J. Roland.

Dans la section 4.1, nous définissons les boîtes non-locales, et les modèles de complexité en boîte non-locale déterministe et probabiliste.

Dans la section 4.2, nous analysons la complexité déterministe. Nous introduisons la notion de rang d'une matrice sur corps fini, et rappelons le lien entre rang et interpolation polynomiale (lemme 4.2). Nous analysons ensuite le protocole de van Dam permettant de calculer toute fonction en utilisant des boîtes non-locales et un seul bit de communication. Nous montrons que ce protocole est optimal pour la complexité simultanée. Ceci nous permet de caractériser exactement le modèle de complexité en boîte non-locale déterministe simultanée (**théorème 4.2**). Nous donnons ensuite des exemples de fonction

pour lesquelles il existe un écart exponentiel entre la complexité simultanée et le modèle général (proposition 4.3). Pour terminer, nous comparons la complexité en boîte non-locale et la complexité de la communication dans le modèle déterministe (**corollaire 4.1 et proposition 4.4**).

Dans la section 4.3, nous étudions la complexité en boîte non-locale probabiliste. Nous donnons des bornes en fonction :

- du rang approché sur $\mathbb{Z}/2\mathbb{Z}$ pour la complexité simultanée (**théorème 4.3**),
- de la complexité de la communication simultanée pour le modèle général (**théorème 4.5**),
- des normes de factorisation pour le modèle général également (**corollaire 4.2**).

Dans la section 4.4, nous montrons comment utiliser les boîtes non-locales pour simuler les corrélations quantiques et au-delà. Cela permet de simuler un processus causal en utilisant une ressource qui respecte également la causalité (**théorème 4.6**).

Enfin, dans la section 4.5, nous étudions les conséquences de notre travail sur l'évaluation sécurisée des fonctions booléennes. Dans le modèle honnête mais curieux (déterministe), nous étudions le ET sécurisé. Nous caractérisons le nombre de ET nécessaire et suffisant pour évaluer une fonction de manière sécurisée (**théorème 4.7**). Dans le modèle malicieux (probabiliste), nous donnons une borne supérieure sur le nombre suffisant de boîtes OT en fonction de la complexité en boîte non-locale (**théorème 4.8**). Nous donnons une borne inférieure sur le nombre de boîtes OT requises en fonction de la complexité de la communication (**théorème 4.9**). La preuve de ce point utilise également les boîtes non-locales pour réaliser l'inversion des boîtes OT.

Chapitre 1

Concepts

Dans ce chapitre, nous allons présenter les concepts informatiques sur lesquels nous allons travailler. Le premier de ceux-ci est le modèle de complexité proposé par A.C.C. Yao pour modéliser la communication. Nous définirons le modèle dans trois paradigmes de calcul : déterministe, probabiliste et quantique. Pour définir ce troisième cas formellement, nous allons également donner quelques bases de calcul quantique général. Ensuite, nous allons donner les bases de la complexité de Kolmogorov, une branche de la logique qui a pour but de donner une caractérisation algorithmique de la notion d'aléatoire. La troisième partie sera consacrée à la définition des distributions de probabilité causales. Nous en donnerons la définition et quelques propriétés élémentaires. Nous donnerons également des exemples d'expériences donnant lieu à de telles distributions. Ceci nous permettra de donner une idée de la structure géométrique sous-jacente.

1.1 Complexité de la communication

La complexité de la communication est un modèle de calcul qui considère des fonctions dont des entrées sont réparties entre plusieurs agents. Ce modèle a été introduit en 1979 par A.C.C. Yao [Yao79]. L'importance de ce modèle vient de ses nombreuses applications, parmi lesquelles les circuits VLSI [Yao79, Len90], la profondeur des circuits booléens [KW90], les compromis temps-espace [BNS92], les bornes inférieures sur la complexité des algorithmes streaming [AMS99]...

Dans la section 1.1.1, nous commençons par définir la complexité de la communication déterministe. Nous expliciterons les liens entre la complexité de la communication et les partitions d'ensemble. Pour cela, nous définirons formellement la notion de rectangle monochromatique et ε-monochromatique.

Dans la section 1.1.2, nous définirons les modèles probabilistes et distributionnels. Nous rappellerons le théorème du minmax de Yao [Yao83], qui montre que ces deux modèles sont équivalents. Nous donnerons également le théorème de Newmann [New91] qui permet de comparer la communication dans les cas de l'aléa partagé et de l'aléa privé.

Enfin, dans la section 1.1.3, nous donnerons les bases du calcul quantique. Cette introduction sera limitée à ce qui est nécessaire pour définir la complexité de la communication quantique, et le théorème de Bell [Bel64]. Nous rappellerons le principe de la téléportation quantique [BBC$^+$93] et du codage super-dense [BW92] et donnerons leurs conséquences sur la complexité de la communication. Enfin, nous rappellerons ce qui est sans doute le

résultat le plus important de complexité de la communication quantique : le lemme de factorisation de Kremer-Klauck [Kre95, Kla07], également prouvé par Yao [Yao93].

1.1.1 Modèle deterministe et partitions

Nous allons définir le modèle de communication à deux joueurs. Le modèle avec un nombre supérieur de joueur est défini de manière similaire. On fixe deux ensembles finis X et Y, tous les deux de taille N, et $n = \log N$. Ici et dans la suite, la fonction log sera toujours le logarithme à base 2. En général, k bits permettent de compter jusqu'à 2^k. Inversement, si on veut donner un code binaire pour tous les éléments d'un ensemble de taille K, on a besoin de $\lceil \log K \rceil$ bits, où $\lceil x \rceil$ est le plus petit entier supérieur ou égal à x.

Soit f une fonction de $X \times Y$ dans $\{0,1\}$. Notons que cette fonction peut être représentée par une matrice M_f de taille $|X \times Y|$, et telle que $M_f[x,y] = f(x,y)$. On appelle cette matrice la *matrice de communication* du problème.

Le scénario qu'on étudie est le suivant. Deux joueurs, Alice et Bob, reçoivent respectivement un élément $x \in X$ et $y \in Y$. Leur but est de calculer $f(x,y)$ et pour cela, ils vont s'échanger des messages suivant une procédure appelée protocole de communication. Les messages échangés par les joueurs dépendent de leur propre entrée et des messages précédemment reçus. Supposons que chaque message envoyé se limite à un seul bit. A chaque étape d'un protocole, on peut représenter la communication par une chaîne booléenne obtenue en concaténant tous les messages envoyés. Ainsi, chaque message d'Alice est fonction de son entrée $x \in X$ et d'une chaîne booléenne v.

On peut ainsi représenter un protocole de communication par un arbre binaire, pas nécessairement complet, dont chaque noeud interne est étiqueté par une chaîne booléenne v qui représente la communication passée et par une fonction qui représente le message envoyé d'un joueur à l'autre. Cette fonction peut être soit $a_v : X \to \{0,1\}$, soit $b_v : Y \to \{0,1\}$. Enfin, les feuilles de l'arbre sont étiquetées par une valeur dans $\{0,1\}$.

L'exécution d'un protocole $\mathcal{P}$ sur une entrée (x,y) se déroule de la manière suivante. On commence à la racine et on parcourt l'arbre d'après la règle suivante : à chaque noeud v étiqueté par une fonction a_v, on continue le parcours par l'arête de gauche si $a_v(x) = 0$ et l'arête de droite sinon. On fait de même pour les arêtes étiquetées par une fonction b_v, en fonction de la valeur de $b_v(y)$. La valeur calculée est la valeur de la feuille à laquelle on arrive en fin de parcours. On la note $\mathcal{P}(x,y)$. Si pour tout $(x,y) \in X \times Y$, on a $\mathcal{P}(x,y) = f(x,y)$, on dit que $\mathcal{P}$ calcule f.

On appelle transcription d'un protocole sur une entrée la concaténation de tous les messages envoyés. Cela correspond à un parcours complet de l'arbre, de la racine à une feuille. La complexité en pire cas d'un protocole de communication est la longueur maximale d'une transcription parmi toutes les entrées, c'est-à-dire la profondeur de l'arbre correspondant. La complexité de la communication d'une fonction f est le coût minimum d'un protocole calculant f. On la note $D(f)$.

Ce modèle a une structure combinatoire très riche et très étudiée. La proposition suivante montre le lien entre la complexité de la communication et l'étude des partitions en rectangles monochromatiques. Un ensemble $S \subseteq X \times Y$ est appelé un rectangle s'il existe $R \subseteq X$ et $T \subseteq Y$ tels que $S = R \times T$. On dit de plus que S est monochromatique pour f (ou f-monochromatique, ou juste monochromatique s'il n'y a pas d'ambiguïté) si f est constante sur S.

Proposition 1.1 ([Yao79])**.** *Soit $\mathcal{P}$ un protocole de communication déterministe calculant une fonction f. Fixons une transcription du protocole T. Soit R l'ensemble des entrées donnant lieu à la transcription T, alors R est un rectangle f-monochromatique.*

En fixant une transcription particulière, on fixe a fortiori la réponse du protocole. Comme le protocole est déterministe, la fonction est calculée sans erreur. C'est ainsi qu'on déduit que l'ensemble des entrées conduisant à une transcription donnée est monochromatique. Le fait que cet ensemble est un rectangle vient du fait que chaque message est indépendant de l'entrée de l'autre joueur.

La propriété précédente est au coeur de nombreuses méthodes de bornes inférieures. Elle permet en effet de déduire le corollaire suivant, qui relie la complexité de la communication avec le nombre de rectangles monochromatiques nécessaires pour partitionner l'ensemble des entrées..

Corollaire 1.1. *Un protocole de communication de complexité t pour f induit une partition de $X \times Y$ en au plus 2^t rectangles monochromatiques pour f.*

Si on considère l'arbre de communication qui représente un protocole de communication, t est la hauteur de l'arbre. L'arbre étant binaire, 2^t est donc le nombre maximal de feuilles.

Le corollaire précédent permet de donner une méthode élémentaire pour prouver des bornes inférieures. La contraposée nous dit en effet que si $X \times Y$ ne peut être partitionné en moins de t rectangles monochromatiques pour f, alors $D(f) \geq \log t$. On ne sait pas en revanche si la réciproque est vraie. Le problème vient du fait qu'il existe des partitions qui ne peuvent pas être obtenues avec un protocole de communication (voir à ce sujet [KN97]). Il est néanmoins possible qu'une partition optimale puisse toujours être atteinte par un protocole de communication, ce qui montrerait que la réciproque est vraie.

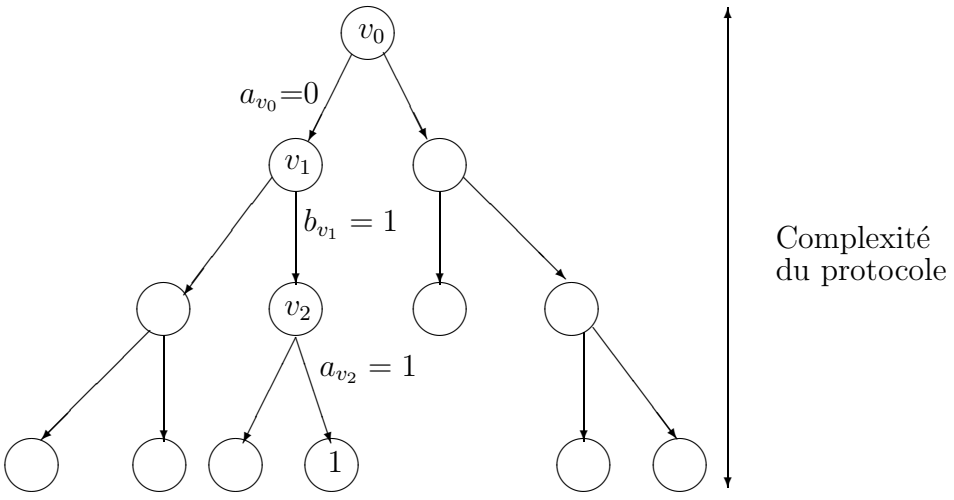


FIG. 1.1 – Un arbre représentant un protocole de communication - La transcription 011 conduit à la sortie 1

Plus loin, nous allons introduire les modèles de complexité de la communication probabiliste et distributionnel. Dans ces cas là, au lieu de considérer des partitions en rectangles monochromatiques, on considérera des partitions imparfaites, c'est à dire avec des rectangles "presque" monochromatiques. Nous en donnons la définition formelle.

Définition 1.1. *Fixons une fonction $f : X \times Y \to \{0,1\}$, une distribution μ sur $X \times Y$, et une constante $\varepsilon > 0$. Un rectangle $R = S \times T$ est dit (μ, ε)-monochromatique s'il existe $b \in \{0,1\}$ tel que $\mu(\{(x,y) \in R : f(x,y) = b\}) > (1-\varepsilon)\mu(R)$.*

Enfin, avant de conclure, nous donnons la liste des principales variantes du modèle que nous allons rencontrer. Les deux premières variantes imposent des restrictions sur les règles d'envoi des messages. La dernière concerne le type de problème qu'on cherche à résoudre.

- *Communication à sens unique :* un joueur envoie un unique message à l'autre, qui doit donner la valeur de la fonction. On note la complexité de la communication à sens unique $D^{A \to B}$ ou $D^{B \to A}$, suivant la direction de la communication.
- *Communication simultanée :* les deux joueurs envoient simultanément un message à un arbitre qui doit donner la valeur de la fonction. Dans ce cas, on note la complexité de la communication $D^{||}$.
- *Complexité d'une relation :* on fixe une relation $R \subseteq X \times Y \times Z$. Les joueurs reçoivent respectivement $x \in X$ et $y \in Y$ et leur but est de donner une valeur de z telle que $(x,y,z) \in R$. On suppose qu'il existe toujours un tel z.

1.1.2 Modèles de communication probabiliste et distributionnel

Il y a deux manières de rendre probabiliste le modèle initial. La première consiste à donner aux joueurs la possibilité de tirer à pile ou face. On dit alors que le protocole est probabiliste. Dans ce cas, on peut supposer sans perte de généralité que tous les tirages sont faits avant le début de l'exécution de celui-ci. Les tirages peuvent être privés, ou bien être visibles par les deux joueurs. On dit alors que l'aléa est partagé ou public. Si on fixe l'aléa d'un protocole, celui-ci devient par définition déterministe. On peut donc voir un protocole probabiliste comme une distribution sur des protocoles probabilistes. La définition de la correction est que pour chaque entrée, le protocole fait peu d'erreur, en moyenne sur l'aléa des joueurs.

La seconde possibilité est de mettre une distribution de probabilité sur les entrées. Les joueurs reçoivent alors leurs entrées suivant cette distribution. Ils exécutent ensuite un protocole déterministe qui, pour chaque entrée, donne une bonne ou une mauvaise réponse. La définition de la correction d'un protocole est alors que celui-ci fait peu d'erreurs, en moyenne sur les entrées, suivant la distribution fixée au départ.

On va supposer qu'avant l'exécution d'un protocole probabiliste, chaque joueur reçoit avant de démarrer une chaîne aléatoire, $r_A \in R_A$ pour Alice et $r_B \in R_B$ pour Bob, où les ensembles R_A et R_B sont finis. Sans perte de généralité, on suppose que les chaînes sont uniformément distribuées. Ainsi, chaque message qu'un joueur envoie à l'autre est fonction de son entrée, de l'historique, et de cette chaîne. Si $\mathcal{P}$ est un protocole probabiliste, on notera $\mathcal{P}^{r_A, r_B}$ le protocole déterministe obtenu par l'exécution de $\mathcal{P}$ en utilisant les chaînes aléatoires r_A et r_B.

Définition 1.2. *Soit $f : X \times Y \to \{0,1\}$ une fonction booléenne et $\mathcal{P}$ un protocole probabiliste. $\mathcal{P}$ est dit ε-correct pour f si pour tout $(x,y) \in \{0,1\}^n \times \{0,1\}^n$, $\mathrm{Prob}_{r_A,r_B}(f(x,y) = \mathcal{P}^{r_A,r_B}(x,y)) \geq 1-\varepsilon$.*

La complexité de la communication probabiliste est alors le coût du meilleur protocole ε-correct pour f. On la note $R_\varepsilon(f)$. Dans le cas de l'aléa partagé, on considère que les joueurs reçoivent une chaîne aléatoire $r \in R$. Dans ce cas, on note la complexité de la communication $R^{\mathrm{pub}}_\varepsilon(f)$. Bien entendu, l'aléa partagé ajoute de la puissance de calcul au modèle. Les joueurs partagent de l'information qui peut éventuellement les aider, sans avoir à payer pour se l'échanger. Concrètement, les joueurs peuvent simuler l'aléa privé avec de l'aléa partagé, en ignorant simplement que celui-ci est public. Pour toute fonction f, on a donc $R^{\mathrm{pub}}_\varepsilon(f) \leq R_\varepsilon(f)$. Dans l'autre sens, le théorème suivant, dû à I. Newman, donne une majoration de la différence maximale entre les deux modèles.

Théorème 1.1 ([New91])**.** *Soit $f : X \times Y \to \{0,1\}$ une fonction booléenne. Pour tout $\delta > 0$ et tout $\varepsilon > 0$, on a $R_{\varepsilon+\delta}(f) \leq R^{\mathrm{pub}}_\varepsilon(f) + O(\log n + \log \delta^{-1})$, où $n = \log N$ et $N = \max\{|X|, |Y|\}$.*

La preuve du théorème précédent est constructive et montre comment concevoir un protocole avec aléa privé à partir d'un protocole dont l'aléa est public. La preuve est basée sur l'argument suivant, qu'on utilisera plus tard. il existe un petit ensemble de chaînes aléatoires - de l'ordre de n éléments - tel que si le choix de l'aléa est restreint à cet ensemble, l'erreur augmente peu. Il suffit alors à Alice de tirer elle-même l'aléa, et d'envoyer l'indice de sa chaîne à Bob. Ceci définit un protocole avec aléa privé, avec une légère augmentation de l'erreur. Il faut bien noter que cet argument fonctionne bien dans le cas de la communication à sens unique, mais n'est plus vrai dans le cas de la communication simultanée. C'est cette interprétation du théorème que nous allons utiliser dans les démonstrations à suivre.

Dans le cas de la complexité de la communication distributionnelle, on ne mesure plus la correction d'un protocole suivant son aléa, mais suivant la distribution sur les entrées.

Définition 1.3. *Soit $f : X \times Y \to \{0,1\}$ une fonction booléenne, μ une distribution de probabilité sur $X \times Y$ et $\mathcal{P}$ un protocole déterministe. $\mathcal{P}$ est dit (μ,ε)-correct pour f si $Prob_\mu(f(x,y) = \mathcal{P}(x,y)) \geq 1-\varepsilon$.*

Pour une distribution μ fixée, on note $D^\mu_\varepsilon(f)$ le coût du meilleur protocole (μ,ε)-correct pour f. La complexité de la communication distributionnelle est alors définie par $D_\varepsilon(f) = \max_\mu D^\mu_\varepsilon(f)$.

Le théorème du minmax de Yao montre que ces deux modèles sont strictement identiques. L'intérêt du théorème est qu'il permet de ramener l'étude d'une structure probabiliste, objet difficile à appréhender en général, à une structure déterministe fixe. Par exemple, pour prouver une borne inférieure sur la complexité probabiliste, il suffit de fixer une distribution et d'étudier la complexité déterministe. Intuitivement, on veut fixer une distribution qui met plus de poids sur les entrées difficiles à calculer.

Théorème 1.2. *[Yao83] Pour toute fonction $f : X \times Y \to \{0,1\}$, $R^{\mathrm{pub}}_\varepsilon(f) = D_\varepsilon(f)$.*

Les variantes que nous avons donné dans le cas déterministe peuvent être envisagées dans le cas probabiliste. Il s'agissait de la complexité à sens unique, de la complexité simultanée, et du calcul des relations.

Pour la complexité de la communication distributionnelle, il existe une autre variante qui consiste à limiter le type de distribution qu'on considère sur les entrées. Le cas intéressant est de se limiter aux distributions produits, c'est-à-dire les distributions μ sur $X \times Y$, telles qu'il existe μ_1 sur X et μ_2 sur Y avec $\mu = \mu_1 \otimes \mu_2$. On note la complexité de la communication distributionnelle d'une fonction f restreinte aux distributions produits $D^{\square}_{\varepsilon}(f)$

1.1.3 Modèle de communication quantique

Eléments de calcul quantique

Nous allons commencer par donner quelques notions de calcul quantique. Cette introduction sera limitée aux axiomes fondamentaux, et aux opérations les plus simples. L'objectif est de pouvoir définir formellement la complexité de la communication quantique, et de pouvoir présenter le théorème de Bell mentionné dans l'introduction.

Représentation des états : un état quantique est un vecteur unitaire dans un espace de Hilbert complexe $\mathcal{H}$. En physique quantique, on utilise la notation bra-ket pour les représenter. Par exemple, si $\mathcal{H} = \mathbb{C}^2$, on appelle les états des qubits. Fixons une base orthonormée de $\mathbb{C}^2$ notée $\{|0\rangle, |1\rangle\}$. Par exemple, si la base considérée est la base canonique, on a $|0\rangle = (0,1)$ et $|1\rangle = (1,0)$. Un qubit peut alors toujours être décomposé sous la forme $|\varphi\rangle = \alpha|0\rangle + \beta|\varphi\rangle$, où α et β sont des nombres complexes vérifiant $|\alpha|^2 + |\beta|^2 = 1$. De manière analogue au calcul classique, $|0\rangle$ et $|1\rangle$ peuvent être utilisés pour représenter les booléens 0 et 1. Mais de manière générale, on dit qu'un qubit est une superposition des états $|0\rangle$ et $|1\rangle$. La notation $\langle\varphi|$ est utilisée pour noter le transconjugué de $|\varphi\rangle$ en tant que vecteur de $\mathcal{H}$. Par suite, pour tout état quantique $|\varphi\rangle$, on a $\langle\varphi|\varphi\rangle = 1$.

Il existe d'autres bases que $\{|0\rangle, |1\rangle\}$. Par exemple, $\{\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)\}$ est également une base de $\mathbb{C}^2$. On la note $\{|+\rangle, |-\rangle\}$.

Lorsqu'on a deux qubits, ils appartiennent à l'espace de Hilbert $\mathbb{C}^2 \otimes \mathbb{C}^2$, l'espace vectoriel engendré par les éléments $|u\rangle \otimes |v\rangle$ avec $(|u\rangle, |v\rangle) \in \mathbb{C}^2 \times \mathbb{C}^2$. Les notations $|u\rangle \otimes |v\rangle$, $|u\rangle|v\rangle$ et $|uv\rangle$ sont équivalentes.

Evolution : L'évolution d'un état quantique peut être décrite par une transformation unitaire. L'effet d'une transformation peut être décrit par son action sur une base. Par exemple, les trois transformations X, Y et Z suivantes sont appelées les opérateurs de Pauli :

- la transformation X est défini par $X|0\rangle = |1\rangle$ et $X|1\rangle = |0\rangle$,
- la transformation Y est défini par $Y|0\rangle = i|1\rangle$ et $Y|1\rangle = i|0\rangle$,
- la transformation Z est défini par $Z|0\rangle = |0\rangle$ et $Z|1\rangle = -|1\rangle$,
- enfin, la transformation de Hadamard H est défini par $H|0\rangle = |+\rangle$, et $H|1\rangle = |-\rangle$.

Il existe également des opérations portant sur 2 qubits. Par exemple, l'opérateur CNOT (ou porte CNOT) est définie par $CNOT|0u\rangle = |0u\rangle$ et $CNOT|1u\rangle = |1\rangle \otimes (X|u\rangle)$. Intuitivement, cette porte inverse le second qubit si le premier est 1 et le laisse inchangé sinon.

En ajoutant un petit nombre de transformations à celles que nous venons de décrire, on obtient un véritable système de calcul. En effet, il a été prouvé que cet ensemble d'opérateurs permet d'approximer toute transformation unitaire [Kit97, Sol95]. On peut

ainsi simuler toutes évolutions avec un petit nombre d'opérations élémentaires.

Mesure : Contrairement à l'information classique, la mesure d'un système quantique perturbe celui-ci. Intuitivement, une mesure d'un état $|\varphi\rangle$ est la donnée d'une décomposition de l'espace de Hilbert sous-jacent au problème en sous-espace. La mesure de $|\varphi\rangle$ donne comme résultat un sous-espace de cette décomposition avec une probabilité proportionnelle au module du coefficient de $|\varphi\rangle$ dans la décomposition. De plus, après la mesure, $|\varphi\rangle$ est projeté sur le sous-espace mesuré. La définition formelle est la suivante.

Définition 1.4. *Une mesure E est définie par un ensemble d'opérateurs $\{E_a : a \in A\}$ sur un espace de Hilbert $\mathcal{H}$ vérifiant les axiomes suivants :*

1. $E_a^\dagger = E_a$
2. $E_a E_{a'} = \delta_{a=a'} E_a$
3. $\sum_a E_a = Id_{\mathcal{H}}$

Ainsi, quand on effectue la mesure E sur un état $|\varphi\rangle$, la probabilité que le résultat soit a est alors $p(a) = \langle\varphi|E_a^\dagger E_a|\varphi\rangle$. Après la mesure, l'état $|\varphi\rangle$ est transformé en l'état $\frac{E_a|\varphi\rangle}{\sqrt{p(a)}}$.

L'exemple le plus simple est celui de la mesure d'un qubit dans une base. Supposons qu'on mesure $|\varphi\rangle$ dans la base $\{|0\rangle, |1\rangle\}$. φ se décompose sous la forme $\alpha|0\rangle + \beta|1\rangle$. La probabilité de mesurer 0 est alors $|\alpha|^2$ et celle de mesurer 1 est $|\beta|^2$. Après la mesure, $|\varphi\rangle$ se retrouve dans un état consistant avec le résultat de celle-ci, c'est à dire $|0\rangle$ si on a mesuré 0, et $|1\rangle$ sinon.

Nous donnerons plus loin une définition plus générale des mesures. Cette définition nous servira à définir les distributions de probabilité quantiques.

Etats intriqués, paradoxe EPR et théorème de Bell

L'existence de l'intrication vient de la structure de produit tensoriel dans laquelle vivent les qubits multiples. Le produit tensoriel d'ensembles est engendré par les produits d'éléments. Tout élément d'un produit tensoriel d'ensemble peut donc s'écrire comme combinaison linéaire de produits d'éléments. Cependant, un élément d'un produit tensoriel d'ensembles ne peut pas nécessairement être décomposé lui-même comme produit.

Considérons l'état quantique $\frac{|00\rangle+|11\rangle}{\sqrt{2}}$. On appelle cet état *l'état de Bell* ou *paire EPR*. Rappelons que par définition $|00\rangle = |0\rangle \otimes |0\rangle$, et $|11\rangle$ est défini de manière similaire. Ainsi, la paire EPR est bien une combinaison linéaire d'états produits, mais on peut montrer qu'elle n'est pas elle même un produit tensoriel d'état.

Lorsqu'on mesure le premier qubit d'une paire EPR dans la base $\{|0\rangle, |1\rangle\}$, on obtient 0 ou 1 tous les deux avec probabilité 1/2. De plus, si on a obtenu 0, l'état est projeté sur $|00\rangle$, et sinon sur $|11\rangle$. On approche ici du paradoxe EPR, car si on mesure maintenant le second qubit dans la même base, on obtient toujours un résultat identique à la première mesure. Imaginons alors l'expérience suivante : Alice et Bob préparent une paire EPR dans un laboratoire, et chacun rentre chez lui en emmenant un qubit de cette paire. Lorsqu'Alice rentre chez elle, si elle mesure son qubit, le second se retrouvera dans un état consistant avec sa mesure. On pourrait avoir l'impression qu'Alice agit à distance sur le qubit de Bob, et de manière instantanée.

Ce qu'il faut remarquer dans l'expérience que nous venons de décrire, c'est qu'il n'y a pas de transfert d'information entre Alice et Bob. En effet, si Bob connaît à l'avance la mesure qu'Alice va faire sur son qubit, toute l'information est connue à l'avance. La corrélation entre les valeurs observées par les deux joueurs peut alors être reproduite avec des ressources classiques, comme l'aléa partagé.

Supposons maintenant que les joueurs choisissent la mesure qu'ils vont faire sur leur qubit en secret. Par exemple, avec probabilité 1/2, Alice va mesurer dans la base $\{|0\rangle, |1\rangle\}$ et avec probabilité 1/2 dans la base $\{|+\rangle, |-\rangle\}$ décrite plus haut. Peut-on alors reproduire les corrélations entre les résultats des mesures avec de l'aléa partagé ? La réponse à cette question est négative, et c'est précisément ce qu'affirme le théorème de Bell [Bel64].

Ce théorème répond aux travaux d'Einstein, Podolsky et Rosen qui ont les premiers imaginé l'expérience que nous venons de décrire [EPR35]. D'après ces auteurs, cette expérience semblait contredire d'autres principes de physique. L'action à distance entre les états quantiques semblait en effet contredire la relativité générale, qui affirme qu'aucune information ne peut voyager plus rapidement que la vitesse de la lumière. La proposition de ces auteurs était, pour remédier à ce paradoxe, d'ajouter des variables cachées locales à la théorie quantique qu'ils considéraient donc comme incomplète.

Le théorème de Bell affirme que l'utilisation de variables cachées locales - c'est-à-dire d'aléa partagé dans notre contexte - ne permet pas de retrouver les prédictions de la physique quantique. De plus, il n'y pas non plus de contradiction avec la théorie de la relativité. En effet, il n'y a pas de transfert d'information d'un joueurs à l'autre. Bob n'apprend rien sur la mesure qu'Alice a choisi de faire, et vice versa. La causalité est bien respectée. Nous reviendrons en détail sur cette notion à la section 1.3.

Du théorème de Bell, on peut déduire que la physique quantique est non-locale. Les deux état intriqués ne peuvent être considérés comme deux entités indépendantes, et l'action de mesurer l'un a un effet immédiat sur l'autre, bien qu'il n'y ait pas de transfert d'information.

D'un point de vue informatique, l'intrication peut être utilisée pour simuler de l'aléa partagé. Il suffit pour cela de mesurer les deux qubits d'une paire EPR dans la base $\{|0\rangle, |1\rangle\}$. Mais d'après le théorème de Bell, la réciproque est fausse. On ne peut pas utiliser de l'aléa pour simuler le résultat des mesures sur une paire EPR. C'est pourquoi la puissance calculatoire de l'intrication semble supérieure à celle de l'aléa partagé.

Communication quantique

Il y a deux façons d'envisager la communication quantique :

- Au lieu d'envoyer des bits classiques, les joueurs échangent des qubits. Ce modèle a été proposé par Yao [Yao93].
- Alice et Bob partagent un état quantique intriqué en dimension arbitrairement grande, mais la communication reste classique. Cette variante a été introduite par Buhrmann et Cleve [BC97].

On peut également combiner ces deux variantes, mais nous allons voir tout de suite que l'intérêt du modèle ainsi obtenu est limité. Il est clair que les qubits peuvent simuler des bits classiques, et que l'intrication permet de simuler de l'aléa partagé.

On note $Q_0(f)$ (resp. $Q_\varepsilon(f)$) le coût du meilleur protocole de communication avec échange de qubits, qui calcule f exactement (resp. avec probabilité ε). Lorsque les joueurs

partagent de l'intrication, on ajoute l'exposant ent, définissant ainsi Q_0^{ent} et $Q_\varepsilon^{\text{ent}}$. Lorsqu'on considère que la communication est classique, mais que les joueurs partagent de l'intrication, on note les mesure de complexité associée R_0^{ent} et $R_\varepsilon^{\text{ent}}$.

Deux éléments de calculs quantiques ont des applications directes à la complexité de la communication. La téléportation quantique [BBC+93] permet de transmettre un état quantique d'un joueur à l'autre en utilisant deux bits de communication classique et une paire EPR. Inversement, le codage super-dense [BW92] permet de coder deux bits classiques à l'intérieur d'un seul qubit. Le décodage nécessite lui aussi une paire EPR. Par conséquent, si les joueurs sont autorisés à partager une quantité arbitraire d'intrication, la nature des messages importe peu. Suivant que les messages sont classiques ou quantiques, la communication ne varie que d'un facteur 2. On obtient donc la proposition suivante.

Proposition 1.2. $R_\varepsilon^{\text{ent}}(f) = 2Q_\varepsilon^{\text{ent}}(f)$.

Un autre résultat montre une limite à la communication quantique. Il s'agit du théorème de Holevo [Hol73]. Celui-ci affirme que pour transmettre l'équivalent de n bits d'information classique, n qubits sont nécessaires. Cela semble impliquer qu'il ne peut y avoir de différence entre communication classique et quantique, mais cette interprétation est fausse. Nous verrons plus loin des exemples de séparation entre ces deux modèles. L'idée est qu'en utilisant des ressources quantiques, Alice peut envoyer plusieurs bits d'information en superposition. D'après le théorème de Holevo, Bob ne pourra en tirer que l'équivalent d'un bit classique, c'est-à-dire que toute autre information sera détruite. Cependant, grâce à la superposition, Bob peut d'une certaine manière choisir l'information qu'il veut obtenir, sans avoir à la demander à Alice.

La comparaison de la communication classique et quantique est un sujet important en complexité de la communication. Au chapitre 2, nous donnons un historique des résultats de séparation entre ces deux modèles. L'une des questions ouvertes les plus importantes de ce domaine est de savoir s'il existe ou non une séparation exponentielle entre communication classique et quantique pour une fonction booléenne totale. En effet, une telle séparation n'a pas été trouvée. L'un des éléments qui rend cette question pertinente est que sous les mêmes hypothèses, la différence entre les complexités en requête classique et quantique est au plus polynômiale [BBC+01].

En dépit des efforts deployés [dGdW02], on ne connaît pas d'équivalent quantique du principe du minmax de Yao. Il existe néanmoins un théorème de structure sur les distributions issues d'un protocole quantique. Nous étendrons plus tard celui ci au cas de la simulation des distributions causales.

Théorème 1.3 ([Kre95, Kla07, Yao93])**.** *Soit f une fonction booléenne et $\mathcal{P}$ un protocole quantique ε-correct dans lequel Alice et Bob échangent q qubits, et partagent une quantité arbitraire d'intrication. Alors il existe une constante d et deux familles de vecteurs a_x et b_y dans $\mathbb{R}^d$ tels que $\| a_x \| . \| b_y \| \leq 2^q$ et $|f(x,y) - a_x.b_y| < \varepsilon$ pour tout x, y.*

1.1.4 Méthodes de bornes inférieures en complexité de la communication

L'un des nos objectifs est de classer les méthodes de bornes inférieures. Nous présentons donc ici quelques méthodes connues pour prouver des bornes inférieures sur la complexité de la communication, en précisant le modèle dans lequel elles fonctionnent.

Méthode du rang

L'une des méthodes les plus anciennes, mais aussi les plus fortes est le rang. Le rang qu'on considère est celui de la matrice de communication associée au problème. On considère de plus que le corps de base est $\mathbb{R}$.

Théorème 1.4 ([MS82]). *Soit $f : X \times Y \to \{0,1\}$. On a $D(f) \geq \log \operatorname{rang}(M_f)$.*

L'une des conjectures les plus importantes depuis que ce résultat est connu est la conjecture du *log rank*. Il s'agit de savoir s'il existe une constante k telle que pour toute fonction f, on a $D(f) \leq (\log \operatorname{rang}(M_f))^k$ [LS93].

La méthode du rang peut être étendue à la complexité probabiliste. Le rang qu'on considère alors n'est plus le rang exact, mais le rang approché.

Définition 1.5. *Pour tout $\varepsilon \geq 0$, le rang approché d'une matrice A est défini par*

$$\varepsilon\text{−rang}(A) = \min_{A' : 1 \leq A[i,j]A'[i,j] \leq \varepsilon} \operatorname{rang}(A').$$

La rang approché permet de prouver une borne inférieure sur la complexité de la communication quantique. Cette borne a d'abord été prouvée dans le modèle sans intrication [BdW01], et a été récemment étendue au modèle avec intrication partagée.

Théorème 1.5 ([LS09]). *Pour toute fonction $f : X \times Y \to \{-1,1\}$ et tout $\varepsilon > 0$, on a $Q_\varepsilon^{\text{ent}}(f) \geq \frac{1}{6}\log(\alpha_\varepsilon\text{−rang}(M_f)) - \frac{1}{2}\log\log|X||Y| - 2\log\alpha_\varepsilon + \log(\alpha_\varepsilon - 1) - O(1)$, où $\alpha_\varepsilon = \frac{1}{1-2\varepsilon}$.*

Norme de factorisation

Linial et Shraibman ont proposé une méthode pour prouver des bornes inférieures basées sur les propriétés algébriques de la matrice de communication. Ils considèrent la norme de matrice suivante :

Définition 1.6. *Pour toute matrice A, on note $\operatorname{col}(A)$ la plus grande norme euclidienne des vecteurs obtenus en considérant les colonnes de la matrice A. La norme $\gamma_2(A)$ est alors définie par :*

$$\gamma_2(A) = \min_{X^T \cdot Y = A} \operatorname{col}(X).\operatorname{col}(Y).$$

La version approchée est définie par :

$$\gamma_2^\alpha(A) = \min_{A' : 1 \leq A[i,j]A'[i,j] \leq \alpha} \gamma_2(A).$$

Cette mesure permet de prouver des bornes inférieures dans plusieurs modèles de complexité. Dans le cas quantique, la preuve est basée sur le théorème de Kremer, Klauck et Yao cité plus haut.

Théorème 1.6 ([LS08b]). *Pour toute fonction $f : X \times Y \to \{-1,1\}$, et tout $\varepsilon > 0$, on a :*

$$R_\varepsilon(f) \geq 2\gamma_2^{\alpha_\varepsilon}(A) - 2\log\alpha_\varepsilon,$$

et

$$Q_\varepsilon^{\text{ent}}(f) \geq \gamma_2^{\alpha_\varepsilon}(A) - \log\alpha_\varepsilon - 2,$$

où $\alpha_\varepsilon = \frac{1}{1-2\varepsilon}$.

Nous reviendrons en détail sur cette méthode dans le chapitre 3. Un des intérêts de celle-ci est qu'elle permet de retrouver de nombreuses autres méthodes connues, parmi lesquelles la méthode "discrepancy" [CG88], la norme trace [Raz03], la transformée de Fourier [Raz03, Kla07], etc... Il a également été prouvé que cette méthode était équivalente à celle du rang approché [LS09]. C'est ce qui a permis d'étendre la borne inférieure du rang approché au cas de la communication quantique avec intrication partagée.

Toutes ces méthodes peuvent être qualifiées d'algébriques dans le sens où elles prennent en compte des propriétés d'une représentation du problème : rang de la matrice de communication, coefficient de Fourier de la fonction, etc.. Ceci suppose d'avoir une structure algébrique sous-jacente. Nous allons maintenant présenter d'autre méthodes, basées sur les propriétés combinatoires de la fonction. Dans ce cas, les propriétés qu'on étudie sont indépendantes de la représentation choisie du problème.

La méthode des rectangles et ses dérivées

Dans le cas déterministe, on a donné plus haut une méthode pour prouver des bornes inférieures. On a dit qu'un protocole de communication induisait une partition de l'ensemble des entrées. En prenant la constraposée, on obtient le théorème suivant.

Théorème 1.7 ([Yao79])**.** *Soit $f : X \times Y \to \{0,1\}$ une fonction booléenne. Soit $\lambda \geq 0$ tel que pour tout rectangle $S \times T$ monochromatique pour f, $|S \times T| \leq \lambda$, alors $D(f) \geq \frac{|X \times Y|}{\lambda}$.*

On peut également étendre cette borne à la complexité probabiliste. Dans ce cas, on ne considère pas les rectangles monochromatiques, mais les rectangles "presque" monochromatiques.

Définition 1.7. *Pour une fonction $f : X \times Y \to \{0,1\}$, une distribution μ sur $X \times Y$ et une constante $\varepsilon > 0$, on définit :*

$$mono_\mu(f, \varepsilon) = \max\{\mu(R) : R \text{ est } (\mu, \varepsilon)\text{-monochromatique pour } f\}.$$

Théorème 1.8 (Borne de corruption [BPSW06])**.** *Soit $f : X \times Y \to \{0,1\}$ une fonction booléenne, μ une distribution sur $X \times Y$ et $1/2 > \varepsilon > 0$. On a alors*

$$D^\mu_\varepsilon(f) \geq \log \frac{1}{mono(f, 2\varepsilon)}.$$

Nous prouverons ces deux résultats par au chapitre 2. L'avantage de la borne de corruption est qu'elle semble ne pas s'étendre au cas quantique. Cette méthode pourrait donc servir à prouver des séparations entre communication classique et quantique.

Méthodes basées sur la théorie de l'information

L'objectif du chapitre 2 est de proposer une alternative aux méthodes basées sur la théorie de l'information. La plupart des arguments utilisant la théorie de l'information sont développés ad hoc pour un problème spécifique. Nous donnons ici un exemple de méthode utilisant la théorie de l'information.

Pour deux variables aléatoires $\mathbf{U}$ et $\mathbf{V}$, on note $H(\mathbf{U})$ l'entropie de Shannon, $H(\mathbf{U}|\mathbf{V})$ l'entropie conditionnelle de $\mathbf{U}$ sachant $\mathbf{V}$. L'information mutuelle entre $\mathbf{U}$ et $\mathbf{V}$ est par définition $I(\mathbf{U};\mathbf{V}) = H(\mathbf{U}) - H(\mathbf{U}|\mathbf{V})$.

Définition 1.8. *Soit $f : X \times Y\{0,1\}$. Soit μ une distribution sur $X \times Y$ et $(\mathbf{X}, \mathbf{Y})$ un couple de variables aléatoires distribué suivant μ. Pour un protocole $\mathcal{P}$, on note $\mathcal{P}(\mathbf{X}, \mathbf{Y})$ la variable aléatoire sur l'ensemble des transcriptions dont la distribution est induite par le protocole lorsque les entrées sont choisies suivant μ. Pour toute constante $\varepsilon > 0$, le coût en information de f suivant μ est le minimum, sur tous les protocoles probabilistes ε-corrects pour f de la quantité $I(\mathbf{X}, \mathbf{Y} : \mathcal{P}(\mathbf{X}, \mathbf{Y}))$. On la note $IC_{\varepsilon,\mu}(f)$.*

Théorème 1.9 ([BYJKS04])**.** *Soit $f : X \times Y \to \{0,1\}$. Pour toute distribution μ sur $X \times Y$ et toute constante $\varepsilon > 0$, on a $R_\varepsilon(f) \geq IC_{\varepsilon,\mu}(f)$.*

Jain, Klauck et Nayak ont proposé une autre méthode générale basée sur la théorie de l'information [JKN08]. Le point important de celle-ci est qu'elle est équivalente à la borne corruption présentée plus haut.

1.2 Complexité de Kolmogorov

Comme nous l'avons dit dans l'introduction, la complexité de Kolmogorov a été introduite indépendamment par Solomonoff, Kolmogorov, et Chaitin pour modéliser la notion d'aléatoire [Sol64, Kol65, Cha69] dans des objets donnés. Dans nos applications, nous allons utiliser directement la complexité préfixe. Cette extension de la complexité de Kolmogorov a été introduite simultanément par Levin, Gács et Chaitin [Lev74, Gác74, Cha75]. Ceci permet de gagner un terme logarithmique dans les bornes inférieures. De plus, chaque fois qu'on parlera d'ensemble, on sous entendra que celui ci est calculable.

Définition 1.9. *Un ensemble de chaînes booléennes est dit sans préfixe si aucune chaîne n'est le préfixe d'une autre.*

Définition 1.10. *Soit φ une machine de Turing universelle et $\mathbb{P}$ un ensemble sans préfixe. La complexité de Kolmogorov préfixe d'une chaîne x sachant y est $K_\varphi(x|y) = min\{|p| : p \in \mathbb{P} et\, \varphi(p, y) = x\}$. En particulier la complexité de Kolmogorov de x est $K_\varphi(x) = K_\varphi(x|\theta)$ où θ est la chaîne vide.*

Ainsi définie, la complexité d'une chaîne dépend de la machine de Turing φ et de l'ensemble $\mathbb{P}$. Dans le cas de la complexité plein, c'est à dire sans condition sur l'ensemble des programmes, Kolmogorov a montré qu'on pouvait lever cette ambiguïté. Ce théorème, appelé théorème d'universalité, montre qu'on peut fixer une machine de Turing φ de manière universelle. Précisément, tout autre choix peut être simulé avec la machine universelle φ, en augmentant la complexité de Kolmogorov d'au plus une constante additive. Ainsi, le choix de φ définit la complexité de Kolmogorov à une constante additive près. Ce théorème a plus tard été étendu à la complexité préfixe par ceux qui l'ont introduite.

Dans toute la suite, on fixe donc une machine de Turing universelle et un ensemble préfixe. On écrira ainsi $K(x)$ et $K(x|y)$ pour $K_\varphi(x)$ et $K_\varphi(x|y)$. Pour un couple, Ainsi, $K(x, y|z) = K(\langle x, y\rangle|z)$, ou $\langle x, y\rangle$ est un code non-ambigu du couple (x, y).

Nous présentons maintenant quelques propriétés élémentaires de la complexité de Kolmogorov. Les preuves des propositions 1.3, 1.4, 1.6 et 1.7 peuvent être trouvées dans la littérature consacrée [LV08]. Dans la proposition suivante, on utilise pour coder un élément x dans un ensemble X un programme très simple. On fixe une énumération de X et on donne un indice de l'élément cherché.

Proposition 1.3. *Soit X un ensemble fini et $\sigma \in \{0,1\}^*$. Il existe une constante c telle que pour tout $x \in X$, $K(x|\sigma) \leq \log|X| + c$.*

La proposition suivante précise que le codage utilisé précédemment est en quelque sorte optimal. Cette proposition se prouve aisément par un argument de comptage et en utilisant le principe des tiroirs de Dirichlet.

Proposition 1.4. *Soit X un ensemble fini et $\sigma \in \{0,1\}^*$. Il existe un élément $x \in X$ tel que $K(x|\sigma) \geq \log|X|$. Un tel élément est dit incompressible.*

La proposition suivante est un exemple de propriété qu'on peut établir avec des manipulations élémentaires.

Proposition 1.5. *Il existe une constante c telle que pour tout x, y et $\sigma \in \{0,1\}^*$, $K(x|\sigma) \leq K(x|y,\sigma) + K(y) + c$.*

Démonstration. Soit p_1 le plus court programme calculant x connaissant y et σ. Soit p_2 le plus court programme calculant y. Connaissant σ, on peut calculer x en exécutant p_2 pour calculer y, puis en exécutant p_1. Ainsi, la taille du plus court programme calculant x est au plus $|p_1| + |p_2| + c$, où c est le coût de la simulation de programme que nous venons de décrire, sur la machine de Turing universelle. □

Il existe une façon simple de rendre la constante apparaissant dans la proposition précédente nulle. En effet, celle-ci est la longueur du programme décrivant la série d'opérations nécessaires pour calculer x. Il suffit, pour annuler cette constante, de coder cette information dans la chaîne auxiliaire σ.

En utilisant la proposition 1.5, on peut généraliser la proposition 1.4 au choix de deux chaînes incompressibles.

Corollaire 1.2. *Soit X et Y deux ensembles. Pour tout $x \in X$, $y \in Y$ et $\sigma \in \{0,1\}^*$, Si $K(x,y|\sigma) \geq \log|X| + \log|Y|$, alors $K(x|\sigma) \geq \log|X|$ et $K(y|\sigma) \geq \log|Y|$. Deux tels x et y sont dits mutuellement incompressibles.*

Dans les propositions 1.3 et 1.4, nous avons utilisé une machine de Turing pour calculer des entrées dans un ensemble fini arbitraire. Une extension intéressante de cette idée est le cas où l'ensemble est muni d'une distribution de probabilité. Dans ce cas, il est bien connu qu'on peut utiliser celle-ci pour coder les éléments de l'ensemble. On utilise ce codage, dit de Shannon-Fano dans la proposition suivante :

Proposition 1.6. *Soit X un ensemble fini, et μ une distribution de probabilité sur X. Il existe une constante c telle que pour tout $\sigma \in \{0,1\}^*$, et $x \in X$ tel que $\mu(x) > 0$, $K(x|\sigma) \leq \log \frac{1}{\mu(x)} + c$.*

Là encore, ce codage est en un sens optimal, comme le précise la proposition suivante :

Proposition 1.7. *Soit X un ensemble fini et μ une distribution de probabilité sur X. Pour tout $\sigma \in \{0,1\}^*$, il existe un élément $x \in X$ tel que $\mu(x) > 0$ et $K(x|\sigma) \geq \log \frac{1}{\mu(x)}$.*

Les propositions 1.6 et 1.7 donnent un aperçu des liens qui peuvent existent entre théorie de l'information et complexité de Kolmogorov. Nous aurons besoin, dans la suite de la proposition suivante, qu'on peut démontrer en utilisant la formule de Stirling. La preuve complète se trouve dans [KN97].

Proposition 1.8. *Soient $n \in \mathbb{N}$ et $\varepsilon \in]0,1[$.*

$$\log \binom{n}{\lceil \varepsilon n \rceil} \sim nH_2(\varepsilon)$$

où $H_2(\varepsilon)$ est l'entropie d'une variable aléatoire suivant une loi de Bernoulli de paramêtre ε.

1.3 Distributions de probabilité causales

Nous allons maintenant présenter la théorie des distributions de probabilité causales. Après avoir présenté les définitions élémentaires, nous donnerons des exemples importants de distributions, révélant quelques aspects de la riche structure de ces distributions.

Soient A et B deux ensembles finis. On considère une famille de distributions $p_{x,y}$ indexées par les éléments de deux ensembles finis X et Y. Chaque distribution porte sur les éléments de $A \times B$. Par convention, on note $p_{x,y}(a,b) = p(a,b|x,y)$. En général, il n'y a pas de distribution sur $X \times Y$, donc la notation ne fait pas référence à une probabilité conditionnelle. Cependant, dans les cas où il y aura également une distribution sur $X \times Y$, notamment lorsque nous aborderons les jeux à deux prouveurs sans interaction, nous conserverons la même notation. Pour résumer, on utilisera un vocabulaire spécifique aux distributions causales en écrivant qu'une telle distribution $\mathbf{p}$ est définie sur $X \times Y$ et à valeurs dans $A \times B$.

Nous allons maintenant définir la causalité. On peut dire de manière imagée que les distributions que nous étudions sont scindées entre deux joueurs. Alice possède l'entrée x et produit la sortie a, et Bob possède y et produit b. Dans ces conditions, la causalité exprime le fait qu'en regardant a, on ne puisse pas obtenir d'information sur y, et vice versa. Formellement, cela se traduit en disant que les distributions marginales d'un joueur sont indépendantes de l'entrée de l'autre joueur.

Définition 1.11. *Soit $\mathbf{p}$ une distribution définie sur $X \times Y$ et à valeurs dans $A \times B$, où X, Y, A et B sont des ensembles finis. On dit que $\mathbf{p}$ est causale si et seulement si pour tout x, x', y, y'*

- $\sum_{b \in B} p(a,b|x,y) = \sum_{b \in B} p(a,b|x,y')$,
- $\sum_{a \in A} p(a,b|x,y) = \sum_{a \in A} p(a,b|x',y)$.

Dans le cas d'une distribution causale, la marginale sur A ne dépend pas de y et inversement, la marginale sur B ne dépend pas de x. On peut donc écrire sans ambiguïté $p(a|x)$ et $p(b|y)$. Dans la suite, on notera $\mathcal{C}$ l'ensemble des distributions causales.

On généralise la complexité de la communication de la manière suivante. Alice reçoit une entrée $x \in X$, Bob une entrée $y \in Y$. Ils communiquent afin de produire une paire de sorties $(a,b) \in A \times B$ suivant une distribution $\mathbf{p}$ définie sur $X \times Y$ et à valeur dans $A \times B$. On a les mesures de complexité associées. $R_0(\mathbf{p})$, $R_\varepsilon(\mathbf{p})$ et $Q_\varepsilon(\mathbf{p})$ désignent respectivement la complexité de communication probabiliste sans erreur, probabiliste avec erreur et quantique d'une distribution de probabilité $\mathbf{p}$. Dans les cas avec erreurs, on utilise pour mesurer la distance entre deux distributions la variation totale. Celle-ci est définie, pour deux distributions de probabilité standards p et p' sur un espace topologique U par $\delta(p,p') = \sup_{V \subseteq U}\{|p(V) - p'(V)|\}$ où V parcourt les boreliens de U. Pour les distributions causales, on a la définition suivante.

Définition 1.12. *Soit* $\mathbf{p}$ *et* $\mathbf{p}'$ *deux distributions causales définies sur* $X \times Y$ *et à valeur dans* $A \times B$*. Pour chaque* x, y*, on note* $p_{x,y}$ *et* $p'_{x,y}$ *les distributions induites par* $\mathbf{p}$ *et* $\mathbf{p}'$ *sur* $A \times B$*. La variation totale entre* $\mathbf{p}$ *et* $\mathbf{p}'$ *est définie par* $\delta(\mathbf{p}, \mathbf{p}') = \max_{x,y} \delta(p_{x,y}, p'_{x,y})$.

Le problème de la simulation d'une distribution avec erreur consiste alors à simuler une distribution $\mathbf{p}'$ telle que $\delta(\mathbf{p}, \mathbf{p}') \leq \varepsilon$. L'intérêt de la distance en variation est que dans le cas particulier où on représente les fonctions booléennes sous la forme de distributions causales, on obtient une définition consistante avec la définition traditionnelle de complexité de la communication avec erreur.

Dans le contexte où on cherche à simuler une distribution, le modèle déterministe est trop faible pour être réellement intéressant. En effet, le problème est de simuler une distribution de probabilité, et on autorise donc toujours les joueurs à utiliser de l'aléa. En général celui-ci est public, mais si on veut le spécifier explicitement, on pourra ajouter l'exposant "pub", ou inversement "priv" pour spécifier que celui-ci est privé. On ajoutera l'exposant "ent" pour spécifier que les joueurs partagent un état intriqué en dimension arbitrairement grand. Rappelons également que l'intrication permet de simuler de l'aléa partagé.

1.3.1 Structure des distributions causales

Afin de clarifier cette définition, nous allons donner des exemples de distributions causales. Pour cela, on décrit des procédures effectives donnant lieu à des distributions causales. L'étude de ces exemples va nous donner une idée de la riche structure de l'ensemble des distributions causales. Cette structure a largement été étudiée comme modèle de la non-localité. Rappelons qu'on note $\mathcal{C}$ l'ensemble des distributions causales. Le premier exemple qu'on va voir est celui des distributions classiques, c'est-à-dire celles qui sont produites lorsque les joueurs ne partagent que des ressources classiques.

Le cas le plus simple est celui où Alice et Bob produisent leurs sorties de manière déterministe et sans aucune autre ressource. Dans ce cas, la stratégie de chaque joueur peut être représentée par une fonction booléenne appliquée à leurs entrées.

Définition 1.13. *Une distribution causale* $\mathbf{p}$ *définie sur* $X \times Y$ *à valeur dans* $A \times B$ *est dite locale déterministe s'il existe deux fonctions* $u : X \to A$ *et* $v : Y \to B$ *telles que :*

$$p(a, b|x, y) = \begin{cases} 1 & \textit{si } u(x){=}a \textit{ et } v(y){=}b \\ 0 & \textit{sinon} \end{cases}$$

On peut également définir les distributions locales déterministes en utilisant le symbole de Kronecker. Soit Q un prédicat, et δ_Q la fonction qui vaut 1 si Q est vrai et 0 sinon. L'expression précedente peut alors s'écrire : $p(a, b|x, y) = \delta_{u(x)=a}\delta_{v(y)=b}$.

Le second cas qu'on considère est celui où Alice et Bob partagent une chaîne aléatoire. Dans ce cas, la distribution calculée par les joueurs est une combinaison convexe de stratégies locales déterministes. Considérer l'aléa partagé revient, dans le contexte du théorème de Bell, à ajouter des variables cachées à la théorie. Ce que prouve le théorème de Bell, c'est que les distributions locales ainsi obtenues sont strictement incluses dans l'ensemble des distributions quantiques, qu'on présente plus loin.

Définition 1.14. *Une distribution causale* $\mathbf{p}$ *définie sur* $X \times Y$ *à valeur dans* $A \times B$ *est dite locale s'il existe une distribution de probabilité* μ *ayant pour support un ensemble fini* Λ *et une famille de distribution locale déterministe* $\{\mathbf{p}_\lambda\}_{\lambda\in\Lambda}$ *tel que*

$$p(a,b|x,y) = \sum_\lambda \mu(\lambda) p_\lambda(a,b|x,y).$$

En utilisant le symbole de Kronecker, on peut exprimer une distribution locale comme combinaison convexe de distributions locales déterministes :

$$p(a,b|x,y) = \sum_\lambda \mu(\lambda) \delta_{u_\lambda(x)=a} \delta_{v_\lambda(y)=b}.$$

Dans la suite, on notera $\mathcal{L}$ l'ensemble des distributions locales. Notons que ces distributions peuvent être, par définition, simulées sans communication.

Le dernier scénario est celui où Alice et Bob partagent un état intriquée. Les deux joueurs commencent alors par appliquer une transformation unitaire, puis font une mesure sur leur partie de l'état. On n'impose pas de limite sur la dimension du système.

En utilisant la définition 1.4, on a la définition suivante.

Définition 1.15. *Une distribution est dite quantique s'il existe un état quantique* $|\psi\rangle$ *dans un espace de Hilbert* $\mathcal{H}$ *et pour chaque* $(x,y) \in X \times Y$*, deux mesures* $\{E_a(x) : a \in A\}$ *et* $\{E_b(y) : b \in B\}$ *telles que :*

1. $E_a(x)E_b(y) = E_b(y)E_a(x)$,
2. $p(a,b|x,y) = \langle\psi|E_a(x)E_b(y)|\psi\rangle$.

Dans la suite, on notera $\mathcal{Q}$ l'ensemble des distributions quantiques. Il est aisé de vérifier en utilisant les axiomes de la définition 1.4 qu'une telle distribution respecte la causalité.

Cette définition a pour but de traduire l'expérience EPR. Pourtant, intuitivement, la première condition peut sembler trop forte. En effet, la situation qu'on veut représenter est celle où les mesures d'Alice et de Bob agissent sur des parties différentes de $|\psi\rangle$. Or, la définition dit que les opérateurs $E_a(x)$ et $E_b(y)$ commutent. Il suffirait donc pour représenter ceci d'avoir une factorisation de l'espace de Hilbert $\mathcal{H} = \mathcal{H}_A \otimes H_B$ et des opérateurs $E_a(x) = E'_a(x) \otimes Id_{\mathcal{H}_B}$ et $E_b(y) = Id_{\mathcal{H}_A} \otimes E'_b(y)$. De plus, il est évident que cette propriété implique la commutativité.

Dans le cas où $\mathcal{H}$ est de dimension finie, on sait que ces deux conditions sont équivalentes, mais le problème est toujours ouvert en dimension infinie. On conjecture toutefois que les deux conditions sont également équivalentes dans ce cas là [DLTW08, NPA08]. Cette propriété est largement utilisée pour définir la mesure. Elle est par ailleurs plus facile à manipuler. Bien que ce changement n'ait pas d'incidence sur notre travail, on retient cette propriété pour notre définition.

L'étude de la structure des distributions causales a donné lieu à de nombreuses recherches. Par exemple, il a été montré que $\mathcal{C}$ était exactement l'enveloppe affine de $\mathcal{L}$. Ce résultat a été prouvé plusieurs fois, parfois de manière indépendante, et dans différents contextes [RF81, FR81, KRF87, Wil92]. Aujourd'hui, ce résultat est mieux connu dans la communauté de l'informatique quantique, et est utilisé pour montrer que

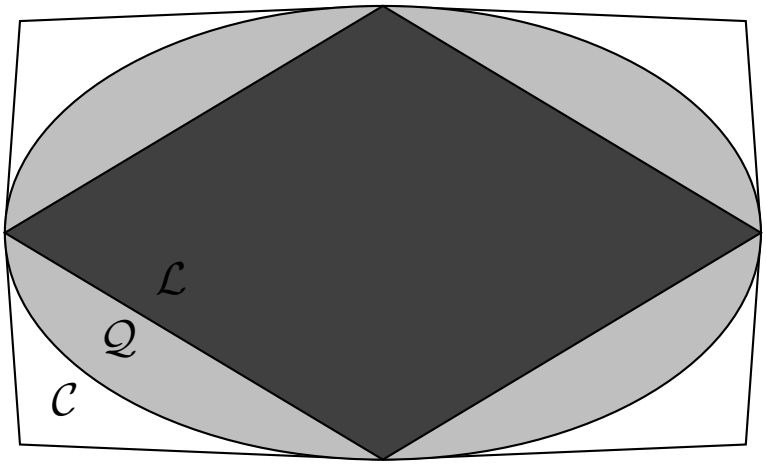


FIG. 1.2 – Représentation schématique des ensembles $\mathcal{L}$, $\mathcal{Q}$, et $\mathcal{C}$. $\mathcal{L}$ et $\mathcal{C}$ sont des polytopes, $\mathcal{Q}$ est un convexe. On a $\mathcal{L} \subset \mathcal{Q} \subset \mathcal{C}$.

certains processus quantiques sont vrais dans le cadre d'une théorie non-locale généralisée [Bar07, BBLW07].

L'autre point qui va nous intéresser est la caractérisation des points extrémaux. En effet, il est possible de voir que l'ensemble des distributions causales est convexe. De plus, l'aléa partagé est considéré comme gratuit dans notre modèle. Or, ce que permet l'aléa partagé d'un point de vue opérationnel, c'est précisément de simuler des combinaisons convexes de distributions. Ainsi, si deux distributions $\mathbf{p}_1$ et $\mathbf{p}_2$ ont une complexité au plus t, toutes les distributions $\alpha\mathbf{p}_1 + (1-\alpha)\mathbf{p}_2$, pour $\alpha \in [0,1]$, ont également une complexité au plus t. Il suffit aux joueurs de choisir quelle distribution échantillonner en utilisant leur aléa partagé. L'étude des point extrémaux nous permet donc de ramener l'étude de l'ensemble des distributions à un petit nombre de celle-ci. Le tableau suivant donne les cas où les points extrémaux sont connus.

Nombre de joueurs	Entrées	Sorties	Référence
2	2	Arbitraire	[BLM$^+$05]
2	Arbitraire	2	[BP05, JM05]

FIG. 1.3 – Points extrémaux du polytope causale

Nous donnerons dans la section suivante les points extrémaux dans le cas des sorties binaires. Il est par ailleurs bien connu que l'ensemble $\mathcal{Q}$ est également convexe.

1.3.2 Le cas des sorties binaires

Nous allons maintenant préciser les choses dans le cas où $|A| = |B| = 2$. Celui-ci qui nous permet de traiter également des fonctions booléennes. Par convention, et parce que cela permet de simplifier les notations, on va supposer que $A = B = \{-1, 1\}$.

Dans le cas binaire, une distribution causale peut être décrite par trois fonctions faciles à interpréter. (x, y) étant fixé, la distribution sur (a, b) peut en effet être décrite

par l'espérance de a, celle b, et celle du produit $a.b$. Ces trois quantités sont appelées respectivement la marginale d'Alice, celle de Bob et la corrélation des deux variables. Elles sont notées $M_A(x)$, $M_B(y)$ et $C(x,y)$.

Proposition 1.9. *Soit $\mathcal{F}$ l'ensemble des triplets de fonctions $C : X \times Y \to [-1,1]$, $M_A : X \to [-1,1]$ et $M_B : Y \to [-1,1]$, telles que $1 + abC(x,y) + aM_A(x) + bM_B(y) \geq 0$ pour tout $(x,y) \in X \times Y$ et $(a,b) \in \{-1,1\} \times \{-1,1\}$. Alors il existe une bijection entre $\mathcal{F}$ et $\mathcal{C}$.*

Démonstration. Soit $\mathbf{p}$ une distribution causale binaire définie sur $X \times Y$. On définit les fonctions C, M_A et M_B par le système suivant :

$$\begin{pmatrix} 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} p(+1,+1|x,y) \\ p(+1,-1|x,y) \\ p(-1,+1|x,y) \\ p(-1,-1|x,y) \end{pmatrix} = \begin{pmatrix} C(x,y) \\ M_A(x) \\ M_B(y) \\ 1 \end{pmatrix}.$$

En inversant le système précédent, on obtient $p(a,b|x,y) = \frac{1}{4}(1+abC(x,y)+aM_A(x)+bM_B(y))$. Ainsi, si $\mathbf{p}$ est une distribution de probabilité, on a bien $(C, M_A, M_B) \in \mathcal{F}$. Inversement, on vérifie immédiatement que si $(C, M_A, M_B) \in \mathcal{F}$, alors l'inverse du système définit bien une distribution causale. □

Puisqu'il y a une bijection entre les distributions causales et les fonctions définies plus haut, nous utiliserons l'une ou l'autre des deux représentations. Hormis le fait que les distributions binaires sont sans aucun doute le type le plus simple de distribution, l'intérêt particulier que nous leur portons est qu'elles généralisent d'un certaine façon les fonctions booléennes.

Définition 1.16. *Soit $f : X \times Y \to \{-1,1\}$ et M_f la matrice de communication associée. On définit la distribution binaire associée $\mathbf{p}_f$ par $p_f(a,b|x,y) = \frac{1}{4}(1 + abM_f[x,y])$. De façon équivalente $\mathbf{p}_f = (M_f, 0, 0)$, où M_f est la matrice de communication de la fonction f.*

La proposition suivante montre que du point de vue de la complexité de la communication, cette généralisation ne pose pas de problème. La preuve est une réduction de protocole à protocole. On énonce la proposition dans le cas probabiliste sans erreur, mais elle est également valable dans les cas avec erreurs et quantique.

Proposition 1.10. *Soit f une fonction booléenne définie sur $X \times Y$, et $\varepsilon \geq 0$. On a*

$$R_\varepsilon(\mathbf{p}_f) \leq R_{2\varepsilon}(f) \leq R_\varepsilon(\mathbf{p}_f) + 1.$$

Démonstration. Soit $\mathcal{P}$ un protocole pour f. On suppose qu'après l'execution de $\mathcal{P}$, Bob répond par un bit b tel que $\text{Prob}[b = f(x,y)] \geq 1 - 2\varepsilon$. En utilisant l'aléa partagé, Alice répond par une valeur r uniformément choisie dans $\{-1,1\}$. Bob donne la valeur rb. Le produit des sorties est alors toujours b. Soit $\mathbf{p}'$ la distributions sur les sorties ; elle vérifie $\delta(\mathbf{p}_f, \mathbf{p}) \leq \varepsilon$.

Soit $\mathcal{P}$ un protocole pour p_f. Supposons qu'à la fin, Alice répond par un bit $a \in \{-1,1\}$ et Bob $b \in \{-1,1\}$ suivant une distribution $\mathbf{p}'$ telle que $\delta(\mathbf{p}_f, \mathbf{p}') \leq \varepsilon$. Alors Alice n'a qu'à envoyer son bit à Bob qui répond par le produit ab. Comme la distribution $\mathbf{p}'$ vérifie $\text{Prob}[ab = f(x,y)] \geq 1 - 2\varepsilon$, on a bien un protocole 2ε-correct pour f. □

Enfin, dans le cas des distributions binaires, on peut simplifier la définition de l'ensemble $\mathcal{Q}$. En effet, dans ce cas, les corrélations peuvent s'exprimer comme un produit de vecteurs : il suffit pour cela d'effectuer le produit de l'état et de l'opérateur de mesure. Réciproquement, Tsirelson a prouvé que pour un produit de vecteurs, il existait un état quantique et une mesure qui permettaient de retrouver la corrélation correspondante.

Théorème 1.10 ([Tsi85])**.** *Si une distribution* $\mathbf{p} = (C, M_A, M_B)$ *est quantique alors il existe deux familles de vecteurs unitaires* a_x *et* b_y *telle que* $C[x, y] = a_x.b_y$*.*

Réciproquement, si a_x *et* b_y *sont deux familles de vecteurs unitaires, alors la distribution* $(C, 0, 0)$ *avec* $C[x, y] = a_x.b_y$ *est quantique.*

Les distributions causales permettent de représenter les fonctions booléennes par l'intermédiaire de leur matrice de communication. Mais les distributions représentant les fonctions booléennes ont un autre intérêt : elles permettent de décrire les points extrémaux du polytope causal. On donne le théorème pour les distributions à marginales uniformes, bien qu'il puisse être étendu au cas général.

Théorème 1.11 ([BP05])**.** *Les points extrémaux du polytope causale* $\mathcal{C}$ *dans le cas des distributions binaires à marginales uniformes sont les distributions* $(C, 0, 0)$ *où* C *est une matrice signe.*

Autrement dit, les points extrémaux sont les distributions associées à des fonctions booléennes. Notons que ce théorème décrit tous les points extrémaux de $\mathcal{C}$, et en particulier les distributions locales.

1.4 Inégalités probabilistes

On présente dans cette section un résultat essentiel de probabilité qui permet de borner la déviation d'une somme de variables aléatoires par rapport à sa moyenne. Il s'agit de l'inégalité de Hoeffding. On utilise ensuite cette inégalité pour montrer comment améliorer la probabilité de succès d'un algorithme probabiliste.

Théorème 1.12 (inégalité de Hoeffding [McD91])**.** *Soient* $X_1, \dots X_k$ *des variables aléatoires indépendantes et telles que pour tout* i, $\text{Prob}[a_i \leq X_i \leq b_i] = 1$*. Soit* $S = \sum_i X_i$*, on a pour tout* $\delta \geq 0$ *:*

- $\text{Prob}(S - \mathbf{E}S \geq \delta) \leq \exp(-\frac{2\delta^2}{\sum_i (b_i - a_i)})$
- $\text{Prob}(|S - \mathbf{E}S| \geq \delta) \leq 2\exp(-\frac{2\delta^2}{\sum_i (b_i - a_i)})$

Corollaire 1.3. *Soit* $\mathcal{P}$ *un protocole de communication probabiliste de complexité* t *pour une fonction booléenne* f*. Supposons que* $\text{Prob}[\mathcal{P}(x, y) = f(x, y)] = 1/2 + \delta$*, alors pour tout* $\varepsilon > 0$ *tel que* $\varepsilon < \delta$*,* $R_\varepsilon(f) = O(\frac{t}{\delta^2} \log \frac{1}{\varepsilon})$*.*

Démonstration. L'idée est de répéter le protocole $\mathcal{P}$ et de prendre la majorité. On va exprimer la majorité comme une somme pour pouvoir appliquer l'inégalité de Hoeffding. Notons $s_1, \dots, s_k \in \{0, 1\}$ les résultats de k itérations du protocole $\mathcal{P}$ et $S_k = \sum_i (s_i - \frac{1}{2})$. Notons qu'on a $\text{Prob}[-\frac{1}{2} \leq s_i - \frac{1}{2} \leq \frac{1}{2}]$. Le nouveau protocole est simplement de répéter k fois $\mathcal{P}$, de calculer S_k et de répondre 0 si S_k est négatif, et 1 sinon.

Supposons que $f(x,y) = 1$, le cas $f(x,y) = 0$ étant similaire. Dans ce cas, on a $\mathbf{E}[S_k] \geq k\delta$. Le nouveau protocole se trompe si $S_k \leq 0$. Calculons donc la probabilité de se tromper.

$$\begin{aligned} \text{Prob}[S_k \leq 0] &\leq \text{Prob}[|S_k - k\delta| \geq k\delta] \\ &\leq 2e^{-\frac{2k^2\delta^2}{k}} \text{ d'après le théorème 1.12} \\ &\leq 2e^{-2k\delta^2}. \end{aligned}$$

On peut maintenant choisir k pour que cette probabilité soit plus petite que ε. Il suffit pour cela de prendre $k \geq \frac{1}{2\delta^2} \ln \frac{1}{\varepsilon}$. On a donc bien définit un protocole de ε-correct pour f, de complexité $O(\frac{t}{\delta^2} \log \frac{1}{\varepsilon})$. □

Chapitre 2

Méthodes combinatoires, bornes inférieures et complexité de Kolmogorov.

Dans cette section, nous présentons une méthode de borne inférieure sur la complexité de la communication. Cette méthode est basée sur la complexité de Kolmogorov, dont l'objectif est de formaliser la notion d'aléatoire dans les chaînes fixes. Dans un premier temps, nous présentons la méthode générale appliquée au cas déterministe. Celle-ci utilise en particulier la notion d'information mutuelle au sens de Kolmogorov. Nous utilisons la complexité de Kolmogorov dans un second temps pour choisir les entrées de manière incompressible. Intuitivement, les entrées incompressibles correspondent au cas où un joueur ne peut pas compresser l'information contenue dans son entrée. Nous présentons des applications simples de notre méthode. En particulier, nous montrons qu'elle permet de retrouver la borne de corruption définie au chapitre 1.

Pour appliquer notre méthode dans le cas probabiliste, nous prouvons une alternative au principe du minmax de Yao, dont la formulation traditionnelle a été donnée au chapitre 1. Ceci nous permet de choisir l'aléa d'un protocole probabiliste et les entrées de manière indépendante et incompressible. Nous appliquons ensuite notre méthode à deux exemples concrets. Tout d'abord, nous montrons qu'elle permet de retrouver la borne inférieure VC-dimension [KNR99]. Ensuite, nous montrons une borne inférieure sur le problème du couplage caché [BYJK08], une problème relationnel qui a permis d'exhiber un écart exponentiel entre la communication classique et la communication quantique.

Pour la VC-dimension, on donne également une deuxième preuve, entièrement combinatoire. Dans un premier temps, nous avons simplifié la preuve du théorème en remplaçant la théorie de l'information par la complexité de Kolmogorov. Alors que la théorie de l'information procède en considérant des distributions sur les entrées, nous n'avons qu'à considérer des éléments fixes, en les choisissant incompressibles. Le reste de la preuve consiste à exploiter les propriétés combinatoires de la fonction pour prouver des résultats sur la complexité de Kolmogorov de ces instances. L'existence d'éléments incompressibles repose essentiellement sur le principe des tiroirs de Dirichlet. Dans le cas de la VC-dimension, on peut franchir un palier supplémentaire en supprimant la complexité de Kolmogorov pour ne garder que le principe des tiroirs. On réduit ainsi la preuve initiale à un simple argument combinatoire.

Nous terminons ce chapitre en montrant une autre application des méthodes combinatoires. On applique cette fois les techniques combinatoires à la complexité multipartite, c'est-à-dire à plus de deux joueurs. On cherche à comparer la complexité simultanée et la complexité à sens unique. On montre alors que pour la complexité distributionnelle, restreinte aux distributions rectangulaires, les deux modèles de communication sont équivalents. Auparavant, ce résultat était prouvé en utilisant la théorie de l'information [BYJKS04]. La preuve qu'on en donne ici améliore significativement le résultat connu.

Rappelons que nous avons défini la complexité de Kolmogorov formellement à la section 1.2. Dans ce même chapitre, nous avons présenté les propriétés de la complexité de Kolmogorov qui nous seront utiles, ainsi que la notion d'élément incompressible.

2.1 Information mutuelle entre deux chaînes

La complexité de Kolmogorov, comme la théorie de l'information, permet de donner une définition formelle de la notion d'information. Plus encore, ce qui nous intéresse est un moyen de la quantifier. La question qu'on cherche à formaliser est la suivante : quelle quantité d'information les joueurs ont-ils besoin d'échanger ? La réponse évidente est : au moins assez pour résoudre le problème. Si on cherche à donner une borne optimale, on pourrait même dire : juste assez pour résoudre le problème. Dans le modèle de complexité de la communication, l'information donnée à un joueur est son entrée respectives. L'information échangée par les joueurs est contenue dans la transcription. D'une part, s'il existe une transcription longue, c'est qu'il y a beaucoup de transcriptions différentes, sinon l'information contenue dans les transcriptions pourraient être compressée. Mais d'autre part, s'il y a beaucoup de transcriptions différentes, alors le nombre d'entrées donnant lieu à une transcription est en moyenne petit. Dans ce cas, on peut dire intuitivement que les transcriptions contiennent beaucoup d'information sur les entrées qui l'ont engendré.

La notion qui traduit ce raisonnement dans le langage de la complexité de Kolmogorov est l'information mutuelle[1]. Intuitivement, celle-ci est définie de la manière suivante. Pour deux chaînes x et y, l'information mutuelle entre x et y mesure le nombre de bits que la connaissance de y nous permet d'économiser dans le calcul de x.

Définition 2.1. *Soit $x, y, \sigma \in \{0,1\}^*$. On note $I_K(x : y|\sigma) = K(x|y,\sigma) - K(x|\sigma)$. On appelle $I_K(x, y|\sigma)$ l'information mutuelle entre x et y sachant σ.*

Pour rendre cette définition plus facile à comprendre, regardons l'expression $K(x) - K(x|y)$. On oublie momentanément pour simplifier l'information supplémentaire σ. Intuitivement, si y comporte beaucoup d'information sur x, alors le calcul de x sachant y est facile. $K(x|y)$ sera donc proche de 0, et l'information mutuelle proche de $K(x)$. Inversement, si y contient peu d'information sur x, la connaissance de y n'aide pas beaucoup pour le calcul de x. $K(x|y)$ est donc proche de $K(x)$. Finalement l'information mutuelle est proche de 0.

[1] On devrait préciser ici qu'il s'agit d'information au sens de Kolmogorov par opposition à la théorie de l'information. Toutefois, comme nous n'utiliserons plus la théorie de l'information, il n'y aura pas d'ambiguïté.

L'une des propriétés les plus remarquables de l'information mutuelle est qu'elle est symétrique. On a ainsi, à une constante additive près, $I_K(x : y) = I_K(y : x)$. Nous n'utilisons pas cette propriété, mais il semble important de la mentionner pour justifier que la définition précédente formalise bien l'intuition qu'on a de l'information.

2.2 Le cas déterministe

Dans cette section, nous allons montrer une borne inférieure générale sur la complexité de la communication déterministe. Fixons un protocole, la borne inférieure est alors exactement le maximum sur les entrées de l'information mutuelle entre les entrées et la transcription à laquelle elles donnent lieu. La borne ainsi exprimée dépend du protocole fixé au départ.

Dans les applications qu'on donne de notre méthode, on utilise la structure combinatoire du problème pour borner l'information mutuelle d'un protocole optimal. Par exemple, on peut dire que tout protocole qui résout le problème doit distinguer des entrées conduisant à des réponses différentes. Cette réponse d'apparence simple est à l'origine de méthodes sophistiquées dans d'autres modèles, comme les méthodes d'adversaire en complexité en requête classique et quantique [Amb02]. Dans ce modèle, cette idée a également servi à donner des bornes inférieures basées sur la complexité de Kolmogorov [LM08].

On pourra comparer la forme du théorème suivant avec celle du théorème 1.9, basé sur la théorie de l'information. L'intuition derrière la théorie de l'information est similaire à celle de la complexité de Kolmogorov. On voit ici que les mécanismes derrières ces méthodes présentent d'autres similitudes, ce qui donne deux bornes inférieures très similaires dans leurs expressions.

Théorème 2.1. *Soit $\mathcal{P}$ un protocole de communication déterministe optimal pour une fonction $f : X \times Y \to \{0,1\}$. Notons $T(x,y)$ la transcription du protocole $\mathcal{P}$ sur l'entrée (x,y). Alors, pour toute chaîne $\sigma \in \{0,1\}^*$,*

$$D(f) \geq max_{(x,y)\in X\times Y} I_K(x,y : T(x,y)|\sigma).$$

Démonstration. Fixons $(x,y) \in X \times Y$. En utilisant la proposition 1.5, on obtient

$$K(x,y|\sigma) \leq K(x,y|T(x,y),\sigma) + K(T(x,y)|\sigma).$$

De plus, d'après la proposition 1.3 et la définition de la complexité de la communication,

$$K(T(x,y)|\sigma) \leq |T(x,y)| \leq D(\mathcal{P}).$$

En combinant ces deux inégalités, on obtient $D(\mathcal{P}) \geq K(x,y|\sigma) - K(x,y|\ T(x,y)\sigma)$, ce qui prouve le théorème. □

Remarquons que dans le théorème précédent, on fait seulement l'hypothèse que le protocole est déterministe. On ne suppose pas en revanche que le protocole permet de calculer la fonction parfaitement. Ceci permet d'appliquer le théorème aussi bien à la complexité déterministe qu'à la complexité distributionnelle.

Ce théorème nous donne un mécanisme pour prouver des bornes inférieures. La communication est minorée par $K(x,y|\sigma) - K(x,y|T(x,y),\sigma)$. D'un côté, on va choisir des

entrées dont la complexité est maximale, c'est à dire des entrées incompressibles. De l'autre, on cherche à prouver une borne supérieure sur $K(x, y|T(x, y), \sigma)$. En tenant compte de la définition de la complexité de Kolmogorov, cela signifie qu'on doit trouver un algorithme qui, connaissant $T(x, y)$ et σ, calcule l'entrée (x, y).

Comme première application, nous montrons comment utiliser le théorème 2.1 pour retrouver la méthode des rectangles présentée à la section 1.1. D'après la proposition 1.1, un protocole de communication induit à une partition des entrées en rectangles monochromatiques. Cette propriété nous donne un exemple de l'information qu'une transcription contient sur les entrées. En effet, chaque transcription correspond précisément à un rectangle monochromatique de la partition. Connaissant une transcription, on peut produire l'entrée en donnant un indice dans le rectangle. Nous rappelons l'énoncé du théorème puis donnons la preuve.

Théorème 1.7 ([Yao79])**.** *Soit $f : X \times Y \to \{0, 1\}$ une fonction booléenne. Soit $\lambda \geq 0$ tel que pour tout rectangle $S \times T$ monochromatique pour f, $|S \times T| \leq \lambda$, alors $D(f) \geq \frac{|X \times Y|}{\lambda}$.*

Démonstration. Soit $\mathcal{P}$ un protocole optimal pour calculer f sans erreur. On a vu plus tôt que $\mathcal{P}$ induisait une partition de $X \times Y$ en rectangles monochromatiques pour f. En utilisant la proposition 1.4, fixons (x, y) incompressible. On a donc $(x, y) \in X \times Y$ tel que $K(x, y) \geq \log |X \times Y|$. Soit $T(x, y)$ la transcription de $\mathcal{P}$ sur (x, y). D'après la proposition 1.1, l'ensemble des entrées donnant lieu à la transcription $T(x, y)$ est un rectangle monochromatique pour f. Soit $S \times T$ ce rectangle. Par hypothèse, $|S \times T| \leq \lambda$.

Pour calculer (x, y) connaissant $T(x, y)$, il suffit d'en donner un indice dans $S \times T$. D'après la borne sur la taille, on a donc $K(x, y|T(x, y)) \leq \log \lambda$.

Finalement, en utilisant le théorème 2.1, on obtient

$$\begin{aligned} D(f) &\geq K(x, y) - K(x, y|T(x, y)) \\ &\geq \log |X \times Y| - \log \lambda \\ &\geq \log \frac{|X \times Y|}{\lambda}. \end{aligned}$$

□

On peut raffiner le théorème précédent en ajoutant une distribution de probabilité sur les entrées. En effet, une partition peut contenir un grand rectangle et une multitude de petits. Dans ce cas, la complexité de la communication est grande car il y a beaucoup de rectangles, mais la borne est rendue petite par la présence d'un unique grand rectangle. En ajoutant une distribution sur les entrées, on rééquilibre les choses en donnant plus de poids aux entrées dans les petits rectangles.

L'ajout de cette distribution nous permet d'utiliser les outils de codage présentés au chapitre 1. Dans ce cas, au lieu de coder les éléments en donnant leurs indices dans un rectangle, on utilise un code de Shannon-Fano pour coder les éléments. Le théorème précédent correspond alors au cas d'une distribution uniforme sur toutes les entrées.

Théorème 2.2. *Soient $f : X \times Y \to \{0, 1\}$ une fonction booléenne, et μ une distribution de probabilité sur $X \times Y$. Soit $\delta \geq 0$ tel que pour tout rectangle $S \times T$ monochromatique pour f, $\mu(S \times T) \leq \lambda$, alors $D(f) \geq \log \frac{1}{\lambda}$.*

Démonstration. Soit $\mathcal{P}$ un protocole optimal pour calculer f exactement. Soit (x, y) un élément incompressible, choisi suivant la proposition 1.7, c'est-à-dire qu'on a $K(x, y|\mu) \geq \log \mu(x, y)$. Soit $T(x, y)$ la transcription de $\mathcal{P}$ sur (x, y) et $R = S \times T$ l'ensemble des entrées donnant lieu à cette transcription.

Pour coder (x, y) connaissant $T(x, y)$, on utilise un codage de Shannon-Fano avec la distribution induite par μ sur R. Appelons μ_R celle-ci. Ainsi, on a $\mu_R(x, y) = \frac{\mu(x,y)}{\mu(R)}$. D'après la propostion 1.6, on a donc $K(x, y|T(x, y), \mu) \geq \log \frac{\mu(R)}{\mu x,y}$.

Finalement, en utilisant le théorème 2.1, on a :

$$\begin{aligned} D(f) &\geq K(x, y) - K(x, y|T(x, y)) \\ &\geq \log |\mu(x, y)| - \log \frac{\mu(R)}{\mu(x, y)} \\ &\geq \log \frac{1}{\mu(R)} \\ &\geq \log \frac{1}{\lambda}. \end{aligned}$$

□

Jusqu'à présent, on a appliqué ici le théorème 2.1 à des protocoles calculant f sans erreur. Nous allons maintenant nous intéresser au cas où des protocoles déterministes font des erreurs. Il s'agit du modèle de complexité distributionnelle. Il existe alors une mesure de complexité qui généralise la taille des rectangles. Cette mesure est la borne de corruption. Dans le théorème précédent, on a utilisé le fait qu'un protocole déterministe induit une partition des entrées en rectangles monochromatiques. Dans le cas avec erreur, la définition 1.1 formalise le concept de rectangle (μ, ε)-monochromatique. Peut-on dire que la complexité distributionnelle induit des partitions en rectangles (μ, ε)-monochromatiques ? Non, car il est possible que toutes les erreurs d'un protocole soient dans le même rectangle, tandis que tous les autres rectangles sont exactement monochromatiques. Ce qu'on peut dire néanmoins, c'est que dans la majorité des rectangles, on doit trouver peu d'erreurs. Si ce n'est pas le cas, la somme des erreurs est trop grande, ce qui contredit la correction du protocole.

La borne de corruption a été utilisée plusieurs fois de manière implicite. Par exemple, elle a été utilisée par Razborov pour prouver une borne linéaire pour la fonction *disjointness* [Raz92]. Pour Razborov, l'intérêt de cette méthode est de donner une altérnative à la méthode *discrepancy*. Cette dernière est en effet, sous sa forme traditionnelle, insuffisante pour étudier la fonction *disjointness* [Raz03]. Une autre application intéressante de la borne de corruption se trouve dans le travail de Jain, Klauck et Nayak [JKN08] En étudiant celle-ci du point de vue de la théorie de l'information, ils parviennent à prouver des théorème de produit direct sur la complexité de la communication. Un théorème de produit direct est un théorème qui relie la complexité de plusieurs instances d'un problème à la complexité d'une seule instance.

Définition 2.2. *Pour une fonction $f : X \times Y \to \{0, 1\}$, une distribution μ sur $X \times Y$ et une constante $\varepsilon > 0$, on définie*

$$mono_\mu(f, \varepsilon) = \max\{\mu(R) : R \text{ est } (\mu, \varepsilon)\text{-monochromatique pour } f\}.$$

Théorème 1.8 ([BPSW06]). *Fixons $f : X \times Y \to \{0,1\}$, μ une distribution sur $X \times Y$ et $1/2 > \varepsilon > 0$. On a alors*

$$D^{\mu}_{\varepsilon}(f) \geq \log \frac{1}{mono(f, 2\varepsilon)}.$$

Démonstration. Soit $\mathcal{P}$ un protocole distributionnel (μ, ε)-correct pour f optimal. D'après le corollaire 1.1, ce protocole induit une partition des entrées. Notons $\mathcal{R}$ celle-ci. Pour tout ensemble $S \subseteq X \times Y$, on note $Err(S) = \{(x,y) \in S : f(x,y) \neq \mathcal{P}(x,y)\}$. On note maintenant $\tilde{\mathcal{R}} = \{R \in \mathcal{R} : \mu(Err(R)) \geq 2\varepsilon\mu(R)\}$. Ce sont les rectangles sur lesquels le protocole fait beaucoup d'erreurs. Enfin, on note $E = \cup_{R \in \tilde{\mathcal{R}}} R$, l'ensemble des entrées dans $\tilde{\mathcal{R}}$.

Maintenant, on va fixer comme dans la proposition 1.7, une entrée (x,y) telle que $K(x,y) \geq \frac{1}{\mu(x,y)}$. On va prouver qu'une telle entrée ne peut être dans E. Notons d'abord que $\mu(E) < 1/2$. Sinon, en sommant, on trouve $\mu(Err(X \times Y)) \geq 2\varepsilon\mu(E) > \varepsilon$, ce qui contredit l'hypothèse de correction du protocole. Supposons maintenant que $(x,y) \in E$. On peut alors coder (x,y) en utilisant un code de Shannon-Fano comme dans la proposition 1.6. , avec la distribution induite par μ sur E. On a donc $K(x,y) \leq \log \frac{\mu(E)}{\mu(x,y)} < \log \frac{1}{2\mu(x,y)}$, ce qui contredit l'incompressibilité de (x,y).

Finalement, le rectangle $R \in \tilde{\mathcal{R}}$ dans lequel se trouve (x,y) vérifie donc $\mu(R) \leq mono_{\mu}(f, 2\varepsilon)$. On peut donc utiliser pour coder (x,y) sachant $T(x,y)$ un code de Shannon-Fano avec la distribution induite par μ sur R. On a ainsi $K(x,y|T(x,y),\mathcal{P}) \geq \log \frac{\mu(R)}{\mu(x,y)} \leq \log \frac{mono_{\mu}(f,2\varepsilon)}{\mu(x,y)}$. Enfin, en utilisant le théorème 2.1, on a directement $D^{\mu}_{\varepsilon}(f) \geq \log \frac{1}{mono_{\mu}(f,2\varepsilon)}$. □

2.3 Le cas probabiliste

On commence par introduire la version probabiliste du théorème 2.1. Pour prouver le théorème pour une protocole probabiliste, il suffit de remarquer qu'un protocole probabiliste dont on a fixé l'aléa est déterministe. Rappelons que si $\mathcal{P}$ est un protocole déterministe, on note $\mathcal{P}^{r_A, r_B}$ le protocole déterministe obtenu en fixant l'aléa d'Alice à r_A et celui de Bob à r_B.

Théorème 2.3. *Soit $\mathcal{P}$ un protocole probabiliste ε-correct optimal pour une fonction $f : X \times Y \to \{0,1\}$. Notons $T(x,y,r_A,r_B)$ la transcription de $\mathcal{P}^{r_A,r_B}$ sur (x,y). Pour tout $(x,y) \in X \times Y$, $r_A \in R_A$, $r_B \in R_B$ et $\sigma \in \{0,1\}^*$, on a*

$$R_{\varepsilon}(f) \geq I_K(x,y : T(x,y,r_A,r_B)|\sigma).$$

Démonstration. L'idée est d'appliquer la même preuve que celle du théorème 2.1 au protocole $\mathcal{P}^{r_A,r_B}$. En effet, d'après la proposition 1.5, on a $K(x,y|\sigma) \leq K(x,y|T(x,y,r_A,r_B),\sigma) + K(T(x,y,r_A,r_B)|\sigma)$. De plus $R_{\varepsilon}(f) \geq K(T(x,y,r_A,r_B)|\sigma)$, ce qui prouve le théorème. □

Comme dans le cas déterministe, le théorème est une borne inférieure sur la complexité des protocoles. Pour l'appliquer, il faut d'abord décider comment fixer l'aléa. En effet, pour certaines valeurs de l'aléa, la réponse du protocole peut être très éloignée de la fonction à calculer. On cherche donc un moyen de choisir l'aléa de manière à ce que la réponse du protocole contienne suffisamment d'information sur le problème initial.

L'idée est simplement de choisir l'aléa de façon incompressible. En le choisissant ainsi, on montre qu'on peut limiter le nombre d'erreurs. La preuve est similaire à l'une des directions du principe du minmax de Yao. L'intérêt est toutefois de permettre de tirer l'aléa et les entrées mutuellement incompressibles, ce qui est essentiel pour les applications de notre méthode.

Lemme 2.1. *Soit $\mathcal{P}$ un protocole probabiliste ε-correct pour $f : X \times Y \to \{0,1\}$. Soit μ une distribution de probabilité sur $X \times Y$. Pour tout $S \in X \times Y$, on définit :*

$$Err_{r_A,r_B}(S) = |\{(x,y) \in S : \mathcal{P}^{r_A,r_B}(x,y) \neq f(x,y)|.$$

Soit (r_A^*, r_B^*) tel que $K(r_A^*, r_B^*|\mu, \mathcal{P}, S) \geq \log |R_A||R_B|$, on a alors :

$$\mu(Err_{r_A^*,r_B^*}(S) \leq 2\varepsilon\mu(S).$$

Démonstration. Soit $\tilde{R} = \{(r_A, r_B) \in R_A \times R_B : \mu(Err_{r_A,r_B}(S) > 2\varepsilon\}$. Nous allons prouver que $|\tilde{R}| < \frac{|R_A||R_B|}{2}$. Ceci est suffisant pour conclure que $(r_A^*, r_B^*) \notin \tilde{R}$, sinon il suffirait pour le calculer d'en donner un indice dans $\tilde{R}$. Comme $\tilde{R}$ est calculable, cela contredirait l'hypothèse $K(r_A^*, r_B^*|\mu, \mathcal{P}, S) \geq \log |R_A||R_B|$.

$\mathcal{P}$ étant ε-correct, on a en sommant :

$$\sum_{(r_A,r_B)\in R_A\times R_B} \mu(Err_{r_A,r_B}(S)) \leq |R_A||R_B|\varepsilon\mu(S).$$

D'un autre côté, on a

$$\sum_{(r_A,r_B)\in R_A\times R_B} \mu(Err_{r_A,r_B}(S)) \geq \sum_{(r_A,r_B)\in \tilde{R}} \mu(Err_{r_A,r_B}(S)) > 2\varepsilon\mu(S)|\tilde{R}|.$$

En combinant ces 2 inégalités, on en déduit $\mu(Err_{r_A^*,r_B^*}(S)) \leq 2\varepsilon\mu(S)$, ce qui conclut la preuve. □

2.4 Applications

2.4.1 Dimension de Vapnik-Chervonenkis et coefficients de pulvérisation

Dans cette section, on utilise la complexité de Kolmogorov pour prouver une borne inférieure générale sur la complexité de la communication. Cette application concerne la complexité à sens unique. Dans ce cas, la transcription ne dépend que d'une entrée. Par exemple, si la communication va d'Alice vers Bob, la transcription ne dépend que de x et on la note $T(x)$. La borne sur la complexité de la communication devient alors $I_K(x : T(x)|\sigma) = K(x|\sigma) - K(x|T(x), \sigma)$ et pour appliquer notre méthode, il suffit de prouver une borne supérieure sur $K(x|T(x), \sigma)$.

Rappelons la définition de la matrice de communication associée à une fonction f. Pour une fonction $f : X \times Y \to \{0,1\}$, il s'agit de la matrice M_f indexée par $X \times Y$ et telle que $M_f[x,y] = f(x,y)$. Dans le cas déterministe, la complexité à sens unique est

caractérisée par le logarithme du nombre de lignes différentes de la matrice de communication. Une ligne d'une matrice est simplement une chaîne booléenne formée des valeurs de la matrice pour un indice de ligne fixé. Supposons que la matrice de communication contient exactement d lignes différentes. Quel que soit le problème, ll suffit qu'Alice dise à Bob à quel groupe de lignes identiques son entrée appartient. $\log d$ bits de communication suffisent donc à résoudre le problème.

Montrons maintenant que $\log d$ bits sont nécessaires. On va montrer que deux entrées dont les lignes sont différentes conduisent à deux messages différents. Ainsi, il y a au moins d messages différents, ce qui nécessite au moins $\log d$ bits de communication. Soient x_1 et x_2 deux entrées d'Alice dont les lignes correspondantes sont différentes. Supposons qu'Alice envoie le même message sur ces deux entrées. Alors la réponse de Bob est identique quelle que soit son entrée y. Or, par hypothèse, il existe une entrée y telle que $f(x_1, y) \neq f(x_2, y)$. Donc le protocole fait une erreur sur l'une de ces deux entrées, ce qui contredit le fait qu'il est sans erreur.

On voudrait maintenant étendre cette idée au cas probabiliste. On veut utiliser la transcription pour différencier des lignes différentes de la matrice. La différence avec le cas déterministe est qu'ici, le protocole faisant des erreurs, il est possible que deux lignes différentes conduisent à des messages identiques. En revanche, si on est capable de trouver un ensemble suffisamment grand de lignes distinctes, on peut utiliser le lemme 2.1 pour borner les erreurs dans cet ensemble.

Le plus simplement, on peut chercher une sous-matrice qui contient toutes les lignes possibles d'une longueur donnée. Cette définition est en fait celle de la VC-dimension (pour dimension de Vapnik et Chervonenkis), une mesure qui vient de la théorie de l'apprentissage [BEHW89]. Plus généralement, on peut chercher le plus grand nombre possible de lignes différentes d'une longueur donnée. La définition qui formalise cette idée est celle de coefficient de pulvérisation.

Cette seconde notion généralise celle de VC-dimension. Néanmoins, la VC-dimension est importante car il s'agit d'une caractérisation de la complexité distributionnelle, restreinte aux distributions produits [KNR99]. Or, nous savons qu'il peut exister un écart très grand entre la complexité distributionnelle restreinte aux distributions produits et la complexité probabiliste [She08]. Les coefficients de pulvérisation étant également une borne inférieure sur le modèle distributionnel avec distribution produit, on ne peut les utiliser pour prouver de meilleures bornes. Le lien entre les deux mesures peut se déduire du lemme de Sauer [Sau72].

La première preuve montrant qu'on pouvait utiliser la VC-dimension pour minorer la complexité de la communication utilisait un argument de comptage [KNR99], toutefois assez différent de celui que nous présentons. L'extension aux coefficients de pulvérisation utilisait la théorie de l'information [BYJKS02]. On donne ici une nouvelle preuve, utilisant la complexité de Kolmogorov. Nous donnons ensuite une seconde preuve, basée sur du comptage, qui est une traduction combinatoire de la première.

Donnons les définitions formelles de VC-dimension et coefficients de pulvérisation. On identifie chaque ligne de M_f (ou d'une sous matrice de M_f) à une chaîne booléenne. Ainsi, on peut voir M_f comme un ensemble de $|X|$ chaînes de longueur $|Y|$. M_f étant indexée par $X \times Y$, si $U \subseteq X$ et $V \subseteq Y$, on note $M_f|_{U,V}$ la sous matrice de M_f obtenue en restreignant les indices de M_f à $U \times V$.

Définition 2.3. – *La VC-dimension de M_f est la taille du plus grand ensemble $Y_0 \subseteq$*

Y tel qu'il existe un ensemble $X_0 \subseteq X$ de taille $2^{|Y_0|}$ et $M_f|_{X_0,Y_0}$ est exactement l'ensemble des chaînes booléennes de taille $|Y_0|$. On la note $VC(M_f)$.

– *Soit $l > VC(M_f)$. Le l-ème coefficient de pulverisation est la taille du plus grand $X_0 \subseteq X$ tel qu'il existe un $Y_0 \subseteq Y$ de taille l et toutes les lignes de $M_f|_{X_0,Y_0}$ sont différentes. On le note $SC(l, M_f)$.*

Ces deux mesures sont représentées sur la figure 2.1. Dans le premier cas, on a $VC(M_f) = 3$ et dans le second $SC(l, M_f) = 10$. Un sous ensemble de $S \subseteq X \times Y$ tel que $S = U \times V$ et $|U| = SC(l, M_f)$ et $|V| = l$ est appelé un témoin pour $SC(l, M_f)$.

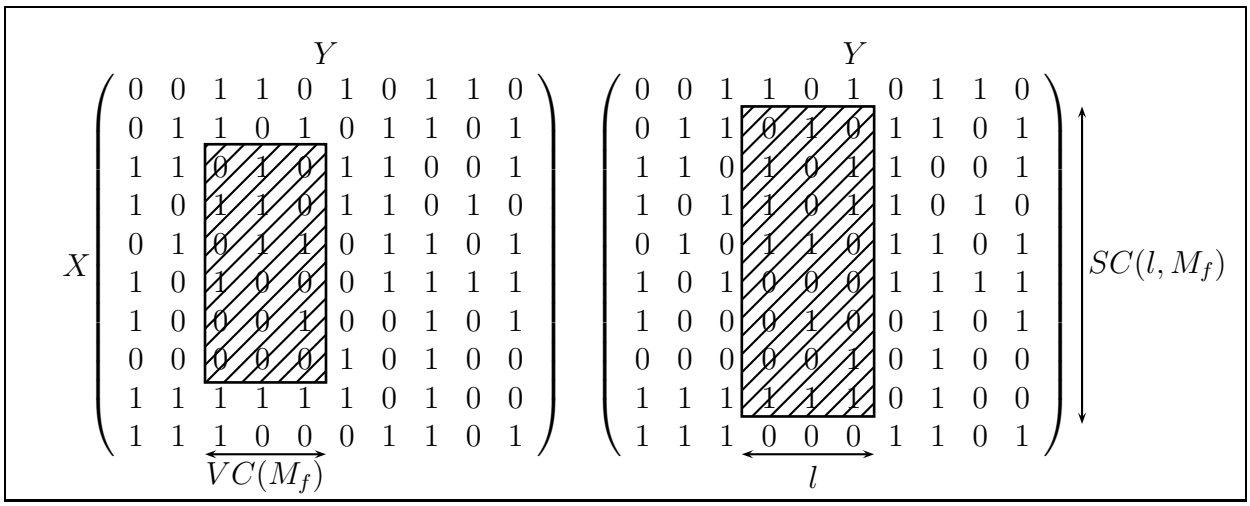


FIG. 2.1 – La VC-dimension et le 4ème coefficient de pulvérisation de la matrice M_f

Théorème 2.4 ([KNR99, BYJKS02])**.** *Pour toute constante $\delta > 0$, il existe un entier L tel que pour toute fonction $f : X \times Y \to \{0, 1\}$ avec $VC(M_f) \geq L$,*

$$R_\varepsilon^{A\to B}(f) \geq VC(M_f)(1 - (1 + \delta)H_2(2\varepsilon))$$

et pour tout $l > VC(M_f)$:

$$R_\varepsilon^{A\to B}(f) \geq \log(SC(l, M_f)) - l(1 + \delta)H_2(2\varepsilon).$$

Démonstration. Remarquons que si on étend la définition de $SC(l, M_f)$ à $l = VC(M_f)$, on a $VC(M_f) = \log(SC(l, M_f))$. Il suffit donc de prouver le second avec $l \geq VC(M_f)$. Soit $\mathcal{P}$ un protocole ε-correct pour f. Soit $S = U \times V$ un témoin pour $SC(l, M_f)$. En utilisant la proposition 1.4, fixons $x^* \in U$ et $(r_A^*, r_B^*) \in R_A \times R_B$ tels que $K(x^*, r_A^*, r_B^*|f, \mathcal{P}, S) \geq \log|U| + \log|R_A| + \log|R_B|$. On a donc choisi l'aléa et l'entrée d'Alice mutuellement incompressibles. On va appliquer le théorème 2.3 sur l'entrée incompressible. En utilisant le corollaire 1.2, on déduit $K(x^*|r_A^*, r_B^*, f, \mathcal{P}, S) \geq \log|U|$ et $K(r_A^*, r_B^*|x^*, f, \mathcal{P}, S) \geq \log|R_A| + \log|R_B|$ Soit $S' = \{x^*\} \times V$. Cet ensemble étant calculable, connaissant x^* et S, on peut appliquer le lemme 2.1 pour déduire $|Err_{r_A,r_B}(S')| \leq 2\varepsilon|S'|$. En choisissant l'aléa incompressible sachant x^*, on s'assure que la ligne correspondant à x^* contient peu d'erreurs. On veut maintenant décrire un algorithme qui permet de retrouver x^* à partir de sa transcription. Comme il y a peu d'erreurs dans la ligne correspondant à x^*, on peut se contenter de les donner comme entrée auxiliaire de l'algorithme.

Soit $T(x^*, r_A^*)$ la transcription du protocole $\mathcal{P}^{r_A^*, r_B^*}$ sur l'entrée x^*. On définit un algorithme pour calculer x^* connaissant $T(x^*, r_A^*)$ et r_B^*.

1. Simuler $\mathcal{P}^{r_A^*, r_B^*}$ en utilisant $T(x^*, r_A^*)$ pour tout $y \in V$.
2. Corriger les erreurs dans S'. L'emplacement des erreurs est donné en entrée du programme.
3. Comparer le résultat avec chaque ligne de S. Comme elles sont toutes différentes, une seule correspond à x^*.

Ce programme utilise $\log \binom{l}{2\varepsilon l}$ bits pour décrire les erreurs. D'après la proposition 1.8, on a pour tout δ et $l \geq L$, $\log \binom{l}{2\varepsilon l} \leq (1+\delta) l H_2(2\varepsilon)$. On a ainsi :

$$K(x^* | T(x^*, r_A^*), r_A^*, r_B^*, f, \mathcal{P}, S) \leq l(1+\delta) H_2(2\varepsilon).$$

Finalement, en appliquant le théorème 2.3, on obtient $R^{A \to B}(f) > \log |U| - l(1+\delta) H_2(2\varepsilon)$. Sachant que $|U| = SC(l, M_f)$, le théorème est prouvé. □

On peut adapter la preuve précédente pour montrer que la VC-dimension est une borne inférieure sur la complexité distributionnelle restreinte aux distributions produits. Pour cela, il faut utiliser le théorème du minmax de Yao (théorème 1.2). Toutefois, la preuve que nous avons donnée permet d'illustrer la mécanique consistant à choisir l'aléa et l'entrée mutuellement incompressible, ce qui sera un argument crucial dans les preuves à venir.

On donne maintenant une seconde preuve de ce théorème qui n'utilise pas la complexité de Kolmogorov. Cette preuve utilise uniquement des arguments combinatoires. Dans le cas étudié dans cette section, il apparait en effet qu'on peut supprimer la complexité de Kolmogorov, en gardant la même structure de preuve. La principale propriété qu'on utilise ici est qu'il existe une injection d'un ensemble E dans un ensemble F si et seulement si $|E| \leq |F|$. L'injection nous permet ici de retrouver l'entrée x étant donnée la transcription et les informations auxiliaires.

Démonstration. Soit $S = U \times V$ un témoin pour $SC(l, M_f)$. Soit μ la distribution uniforme sur S. Le principe du min-max de Yao nous dit que $R_\varepsilon(f) \geq D_\varepsilon^\mu(f)$. Soit $\mathcal{P}$ un protocole (μ, ε)-correct optimal. Soit $U' = \{x \in U : \#\{y \in V : \mathcal{P}(x,y) \neq f(x,y)\} \geq 2\varepsilon l\}$ et $S' = U' \times V$. La distribution étant uniforme, Le nombre d'erreurs de $\mathcal{P}$ sur les entrées de S est au plus $\varepsilon |S|$. En sommant, on obtient donc $|U'| \leq |U|/2$. Soit t la complexité de $\mathcal{P}$ et $\binom{V}{\leq k}$ l'ensemble des sous-ensembles de V de taille au plus k. On définit la fonction suivante :

$$\begin{array}{rcl} \gamma : U' & \longrightarrow & \{0,1\}^t \times \binom{V}{\leq k} \\ x & \longmapsto & (\sigma, E) \end{array}$$

où σ est la transcription de $\mathcal{P}$ sur x et E l'ensemble des erreurs de $\mathcal{P}$ sur $\{x\} \times V$. Nous allons prouver que γ est injective, et donc $|U'| \leq 2^t 2\varepsilon l \binom{l}{2\varepsilon l}$. En appliquant la fonction logarithme et en utilisant la proposition 1.8, on obtient $t \geq \log |U'| - l(1+c) H_2(2\varepsilon) \leq \log |U| - l(1+c) H_2(2\varepsilon) - 1$, ce qui termine la preuve.

Il reste donc à prouver que γ est injective. Soit $x_1 \in U'$ et $x_2 \in U'$ tels que $\gamma(x_1) = \gamma(x_2) = (\sigma, E)$. En utilisant σ, on peut simuler le protocole sur tous les $y \in V$ et corriger les erreurs en fonction de E. On trouve ainsi une ligne complète de S et celles-ci étant toutes différentes, on déduit $x_1 = x_2$. □

2.4.2 Le problème du couplage caché

Nous allons maintenant appliquer notre méthode pour calculer une borne inférieure sur un problème concret. Le problème que nous étudions est le problème du couplage caché [BYJK08]. Ce problème a permis d'exhiber une séparation entre la complexité de la communication classique et quantique.

Rappelons d'abord quelques résultats de séparation importants. On laisse de côté les résultats concernant la communication simultanée. Ceux-ci seront abordés ultérieurement.

- La fonction EQ', une fonction partielle qui est un sous problème de la fonction *égalité*, permet de montrer une séparation exponentielle entre la complexité classique et quantique sans erreur [BCW98]. La complexité quantique est de l'ordre de $\log n$, et la complexité classique de l'ordre de n. La preuve repose sur un argument purement combinatoire utilisée avec la méthode de la taille des rectangles.
- Raz a proposé le problème suivant : Alice reçoit un vecteur unitaire, Bob deux sous-espace vectoriels. Leur but est de déterminer à quel espace appartient le vecteur d'Alice [Raz99]. Raz prouve une séparation pour une version discrétisée de ce problème : Pour la complexité de la communication à sens unique, il propose un protocole quantique avec erreur de complexité $O(\log n)$ et prouve une borne inférieure classique en $\Omega(n^{1/4}/\log n)$.
- Enfin, il existe une façon de modifier le problème du couplage caché en une fonction partielle avec la même complexité [GKK$^+$07]. Pour ce problème, la borne classique est de l'ordre de $\sqrt{n}$ et la borne quantique de l'ordre de $\log n$. La preuve utilise la méthode des rectangles appliquée à la complexité distributionnelle. L'analyse combinatoire du problème utilise l'analyse de Fourier.

Définissons maintenant le problème que nous allons étudier ici.

Définition 2.4 ([BYJK08])**.** *Le problème du couplage caché $HM_n(x, M)$ est défini comme suit :*

- *Alice reçoit une chaîne $x \in \{0,1\}^n$,*
- *Bob reçoit un couplage parfait M sur n sommets,*
- *Bob doit donner comme réponse un triplet (i, j, b) tel que $x_i \oplus x_j = b$ et $(i, j) \in M$*

Rappelons le fonctionnement du protocole quantique à sens unique [BYJK08].

1. Alice envoie à Bob l'état $|\varphi\rangle = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (-1)^{x_i} |i\rangle$
2. Bob mesure dans la base orthonormale $\{\frac{1}{\sqrt{2}}(|k\rangle \pm |l\rangle) : (k, l) \in M\}$

On prouve ensuite que pour Bob, la probabilité que le résultat de la mesure soit $\frac{1}{\sqrt{2}}(|k\rangle - |l\rangle)$ est nulle si $x_k \oplus x_l = 0$. Inversement, la probabilité d'obtenir $\frac{1}{\sqrt{2}}(|k\rangle + |l\rangle)$ est nulle si $x_k \oplus x_l = 1$. La réponse de Bob est donc $(k, l, 1)$ s'il a mesuré $\frac{1}{\sqrt{2}}(|k\rangle - |l\rangle)$ et $(k, l, 0)$ s'il a mesuré $\frac{1}{\sqrt{2}}(|k\rangle + |l\rangle)$. On a donc un protocole sans erreur pour calculer $HM_n(x, M)$ avec une complexité $O(\log n)$.

On montre maintenant la borne inférieure classique pour ce problème, qui est en $\Omega(\sqrt{n})$. Cette borne est de plus optimale. Si Alice envoie $\sqrt{n}$ bits de sa chaîne choisis au hasard à Bob, alors d'après le paradoxe des anniversaires, avec grande probabilité, Bob va être capable de résoudre le problème.

La preuve initiale de ce théorème utilise la théorie de l'information. Alors que celle-ci procédait par éliminations successives des cas défavorables, notre preuve évite cette

itération et est de ce point de vue, plus simple. D'un autre côté, on utilise un résultat sur la structure des graphes que nous avons développé pour ce problème. Ce résultat est présenté à part dans la section 2.4.3.

Théorème 2.5 ([BYJK08])**.**

$$R_{\varepsilon}^{A\to B}(HM_n) \geq \Omega(\sqrt{n}).$$

Démonstration. Soit $\mathcal{P}$ un protocole ε-correct pour HM_n. Soit $\mathcal{M}$ l'ensemble des couplages de la forme $M_i = \{(k, k+i \mod n), k = 0, \dots n-1\}$ pour $i = 0, \dots, n-1$. Une représentation de ces graphes dans le cas $n = 4$ est donnée à la figure 2.2

Nous allons maintenant choisir une chaîne aléatoire, ainsi qu'un ensemble de $\sqrt{n}$ couplages dans $\mathcal{M}$. En utilisant la proposition 1.4, fixons $x^* \in X$, $(r_A^*, r_B^*) \in R_A \times R_B$, et $\mathcal{M}^* \subseteq \mathcal{M}$ tel que $|\mathcal{M}^*| = \sqrt{n}$ et $K(x^*, r_A^*, r_B^*, \mathcal{M}^* | f, \mathcal{P}) \geq \log|X| + |R_A| + |R_B| + \log\binom{n}{\sqrt{n}}$. D'après le corollaire 1.2, on a ainsi $K(r_A, r_B | \mathcal{P}, \{x^*\} \times \mathcal{M}') \geq \log|R_A||R_B|$. En utilisant le lemme 2.1 avec la distribution uniforme sur $\{x^*\} \times \mathcal{M}^*$, on en déduit $Err_{r_A^*, r_B^*}(\{x^*\} \times \mathcal{M}^*) \leq 2\varepsilon\sqrt{n}$.

On cherche donc à reconstruire l'entrée x^* en utilisant sa transcription. L'idée est que celle-ci, pour chaque $M \in \mathcal{M}^*$, permet d'obtenir une équation $x_i \oplus x_j = b$. Nous allons prouver qu'avec $\mathcal{M}^*$ incompressible ce système d'équations est de degré $\Omega(\sqrt{n})$. Soit $E(\mathcal{M}, x^*)$ l'ensemble des arêtes obtenues en exécutant le protocole $\mathcal{P}^{r_A^*, r_B^*}$ sur x^* et chaque $M \in \mathcal{M}$. Notons G le graphe engendré par $E(\mathcal{M}, x^*)$. Les couplages de $\mathcal{M}$ étant indépendants, G a n arêtes. $\sqrt{n}$ sommets de G sont donc suffisants pour engendrer une forêt couvrante pour tous les sommets de G. Une forêt étant par définition sans cycle, les $\sqrt{n}$ arêtes de celle ci correspondent à $\sqrt{n}$ équations indépendantes.

Pour conclure, il reste à prouver qu'en utilisant $\mathcal{M}^*$, on obtient également $\sqrt{n}$ équations indépendantes. En fait, nous allons prouver dans la section suivante un résultat plus fort. Fixons un graphe à n arêtes, et tirons $\sqrt{n}$ arêtes au hasard, uniformément parmi les sous-ensembles de $\sqrt{n}$ arêtes. Soit H le graphe aléatoire généré par ces arêtes. Alors il existe une constante c telle que pour n suffisamment grand, $\text{Prob}\{|V(H)| < c\sqrt{n}\} < 1/2$. La preuve de ce résultat peut être trouvée en section 2.4.3.

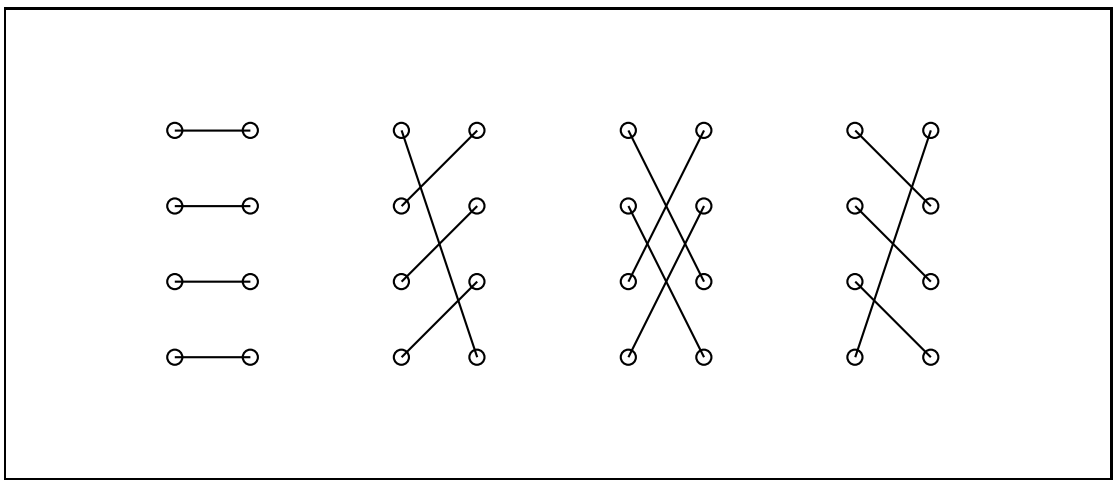


FIG. 2.2 – Couplages parfaits indépendants sur 4 sommets

Rappelons que nous avons défini G, le graphe engendré par $E(\mathcal{M}, x^*)$. Notons $\mathcal{B} = \{H \subset G : |E(H)| = \sqrt{n} \text{ et } |V(H)| < c\sqrt{n}\}$. Notons H le graphe engendré par $E(\mathcal{M}^*, x^*)$. Le précédent argument assure que pour n assez grand, $|\mathcal{B}| < \binom{n}{\sqrt{n}}$. Ainsi, $H \notin \mathcal{B}$, autrement cela contredirait l'hypothèse d'incompressibilité.

On peut maintenant donner un algorithme qui calcule x^* connaissant sa transcription. $\mathcal{P}$ étant un protocole à sens unique, la transcription ne dépend que de l'aléa d'Alice. Soit $T(x^*, r_A^*)$ la transcription de $\mathcal{P}^{r_A^*, r_B^*}$ sur x^*. Pour calculer x^*, on effectue les opérations suivantes :

1. Simuler $\mathcal{P}^{r_A^*, r_B^*}$ avec la transcription $T(x^*, r_A^*)$ pour tout $M \in \mathcal{M}^*$.
2. Corriger les erreurs dans l'ensemble $\{x^*\} \times \mathcal{M}^*$. Celles-ci sont données en entrée de l'algorithme.
3. Résoudre le système d'équations. Celui ci permet de trouver au moins $c\sqrt{n}$ indices de x^*.
4. Les $n - c\sqrt{n}$ indices restants sont donnés en entrée du protocole.

Ce programme utilise $\log \binom{\sqrt{n}}{2\varepsilon\sqrt{n}}$ bits pour décrire les erreurs. D'après la proposition 1.8, pour tout $\delta > 0$ et n suffisamment grand, $\log \binom{\sqrt{n}}{2\varepsilon\sqrt{n}} < (1+\delta)\sqrt{n}H_2(2\varepsilon)$. On ajoute également $n - c\sqrt{n}$ bits pour les indices de x^* qu'on n'a pas calculé. On obtient donc $K(x^*|T(x^*, r_A^*), r_A^*, r_B^*, \mathcal{M}^*, \mathcal{P}, f) < n - (c + (1+\delta)H_2(2\varepsilon))\sqrt{n}$. Enfin, en utilisant le théorème 2.3, on déduit $R^{A\to B}(HM_n) \geq (c + (1+\delta)H_2(2\varepsilon)$, ce qui conclut la preuve.

□

2.4.3 Interlude : un théorème sur les graphes aléatoires

Dans cette section, on prouve un théorème sur les graphes aléatoires. Ce théorème sert à prouver la borne inférieure sur le problème du couplage caché, à la section précédente.

Commençons par rappeler les différentes constructions de graphes aléatoires. Soit $G = (V(G), E(G))$ un graphe tel que $E(G) = n$. Il a deux manières canoniques de construire un sous-graphe aléatoire de G.

- On construit l'ensemble des arêtes en choisissant chaque arête $e \in E(G)$ avec probabilité p. Notons H_p le graphe construit suivant cette procédure.
- On choisit un sous-ensemble de taille m aléatoirement parmi tous les sous-ensembles de taille m de $E(G)$. Notons H_m le graphe construit suivant cette procédure.

Dans la section précédente, on construisait un graphe suivant la deuxième procédure. Mais le premier modèle est en général plus facile à analyser. Quoi qu'il en soit, le théorème suivant montre que dans un certain sens, ces deux modèles sont équivalents.

Théorème 2.6 ([JLR00]). *Soit $\mathbb{Q}$ un ensemble de sous-graphes de G. Alors pour $p = \frac{1}{\sqrt{n}}$, $m = \sqrt{n}$ et n assez grand, on a :*

$$\text{Prob}(H_m \in \mathbb{Q}) \leq \sqrt{2\pi} n^{1/4} \text{Prob}(H_p \in \mathbb{Q})$$

Démonstration.

$$\begin{aligned}
\text{Prob}(H_p \in \mathbb{Q}) &= \sum_{k=0}^{n} \text{Prob}(H_p \in \mathbb{Q} || E(H_p)| = k)\text{Prob}(|E(H_p)| = k) \\
&= \sum_{k=0}^{n} \text{Prob}(H_k \in \mathbb{Q})\text{Prob}(|E(H_p)| = k) \\
&\geq \text{Prob}(H_m \in \mathbb{Q})\text{Prob}(|E(H_p)| = m)
\end{aligned}$$

On peut calculer directement $\text{Prob}(|E(H_p)| = m) = \binom{n}{m} p^m (1-p)^{n-m}$. Pour conclure, il suffit de remarquer que cette dernière expression est équivalente à $\frac{1}{\sqrt{2\pi} n^{1/4}}$ pour $p = \frac{1}{\sqrt{n}}$, ce qui termine la démonstration. □

Théorème 2.7. *Soit G un graphe biparti et H un sous-graphe aléatoire $H = H_{\sqrt{n}}$. Soit $k_H = \{v \in V(H) : deg_H(v) > 0\}$. Il existe une constante c telle que* $\lim_{n\to+\infty} \text{Prob}(k_H \geq c\sqrt{n}) = 1$.

Démonstration. Il suffit de prouver ce théorème pour un graphe $H = H_{1/\sqrt{n}}$ et d'appliquer le théorème 2.6 pour conclure que le résultat est aussi valable pour $H = H_{\sqrt{n}}$. Soient X_1 et X_2 deux ensembles tels que $V(G) = X_1 \cup X_2$ et $E(G) \subseteq X_1 \times X_2$. Notons $X_i^+ = \{v \in X_i : deg(v) > 2\sqrt{n}\}$ et $X_i^- = \overline{X_i^+}$.

Nous allons préciser la structure du graphe en montrant que pour au moins un $i \in \{1, 2\}$, on a les deux propriétés suivantes :

1. $|X_i| \geq \sqrt{n}$.
2. $\sum_{v \in X_i^-} deg(v) > 3n/8$.

Remarquons pour commencer qu'un des côtés du graphe a au moins $\sqrt{n}$ sommets. Supposons maintenant que l'un des côté du graphe a strictement moins de $2\sqrt{n}$ sommets, alors chaque sommet du côté opposé a un degré au plus de $2\sqrt{n}$. En sommant sur X_i^-, on obtient le résultat voulu. Supposons donc que les deux côtés ont plus de $2\sqrt{n}$ sommets.

Soit $E^{\delta,\varphi} = E(G) \cap (X_1^\delta \times X_2^\varphi)$ pour $\delta, \varphi \in \{+, -\}$. Le nombre total d'arêtes est ainsi $n = |E^{++}| + |E^{+-}| + |E^{-+}| + |E^{--}|$. Ensuite, on a en sommant

$$\sum_{v \in X_1^+} deg(v) = |E^{++}| + |E^{+-}|, \text{ et}$$

$$\sum_{v \in X_2^+} deg(v) = |E^{++}| + |E^{+-}|.$$

Enfin, sachant que $\sum_{v \in X_i} deg(v) = n$ pour $i = 1, 2$, on déduit qu'il y a au plus $\sqrt{n}/2$ sommets de degré supérieur à $2\sqrt{n}$, soit $|E^{++}| < \frac{n}{4}$. Finalement,

$$\begin{aligned}
n &= |E^{++}| + |E^{+-}| + |E^{-+}| + |E^{--}| \\
&= \sum_{X_1^+} deg(v) + \sum_{X_2^+} deg(v) - |E^{++}| + |E^{--}| \\
&> \sum_{X_1^+} deg(v) + \sum_{X_2^+} deg(v) - \frac{n}{4} \\
\frac{5n}{4} &> \sum_{X_1^+} deg(v) + \sum_{X_2^+} deg(v).
\end{aligned}$$

Pour un indice i, on a donc $\sum_{v\in X_i^+} deg(v) < 5n/8$, soit pour l'autre côté $\sum_{v\in X_j^-} deg(v) > 3n/8$.

Soit $v_0 = \#\{v : deg_H(v) = 0\}$. Nous allons prouver une borne supérieure sur $\mathbf{E}v_0$. Soit X_i vérifiant les hypothèses précédentes. Pour chaque $v \in X_i$, on a $\text{Prob}(deg(v) = 0) = (1 - \frac{1}{\sqrt{n}})^{deg(v)} \sim e^{-deg(v)/\sqrt{n}}$. Pour des sommets du même côté de G, les degrés dans H sont des variables aléatoires indépendantes, donc $\mathbf{E}v_0 = \sum_v e^{-deg(v)/\sqrt{n}}$.

Soit $\alpha_1 = \frac{1-e^{-2}}{2}$. Pour tout x tel que $0 < x < 2$, $e^{-x} < 1 - \alpha_1 x$. Pour les sommets de X_i^-, c'est-à-dire ceux de petits degrés, on a

$$\sum_{v\in X_i^-} deg(v) < \sum_{v\in X_i^-} (1 - \alpha_1 \frac{deg(v)}{\sqrt{n}}) < |X_i| - \frac{3\alpha_1}{8}\sqrt{n}$$

et pour les autres sommets, ceux de grands degrés

$$\sum_{v\in X_i^+} deg(v) < e^{-2}|X_i^+| \leq \frac{e^{-2}}{2}\sqrt{n}.$$

Posons $c = (3\alpha_1/8 - e^{-2}/2)$. On a donc $\mathbf{E}v_0 < |X_i| - c\sqrt{n}$, soit $\mathbf{E}k_H > c\sqrt{n}$. Finalement, nous avons dit plus haut que v_0, et donc k_H, était une somme de variables aléatoires indépendantes. Il suffit d'appliquer le théorème de Hoeffding (théorème 1.12) pour déduire

$$\text{Prob}(k_H < \frac{c}{2}\sqrt{n}) < e^{c\sqrt{n}/8}.$$

□

2.5 Comparaison de la communication à sens unique et la communication simultanée

Dans cette dernière section, nous allons comparer deux modèles de complexité. Le premier est la complexité mutipartite, modèle *NOH* ("number on hand"). Dans ce modèle, il n'y a plus 2 mais n joueurs. Fixons une fonction $f : X_1, \ldots X_n \to \{0, 1\}$. Un protocole fonctionne de la manière suivante : chaque joueur reçoit une entrée du problème $x_i \in X_i$ et envoie un message à un arbitre qui doit donner la valeur de la fonction. La complexité de la communication simultanée de f est la somme des longueurs de tous les messages et on la note $D^{X_1||\ldots||X_n}(f)$.

Nous allons comparer ce modèle avec la complexité à sens unique. Pour la même fonction f, on considère le problème à 2 joueurs suivants. i étant fixé :

- Alice reçoit x_i
- Bob reçoit toutes les autres entrées $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$. On note cette entrée x_{-i}.

La complexité de la communication déterministe pour de tels protocoles est notée $D^{X_i\to X_{-i}}(f)$. Comme dans le cas général, ces protocoles peuvent être déterministes, probabilistes ou distributionnels.

Dans le modèle déterministe, on peut définir un protocole à sens unique à partir d'un protocole simultané. Il suffit à Alice, par exemple, d'envoyer son message à Bob au lieu de

l'envoyer à l'arbitre. En faisant de même pour Bob, on en déduit $D^{A\to B}(f)+D^{A\leftarrow B}(f) \leq D^{\|}(f)$. Ce résultat s'étend trivialement au modèle multipartite, mais aussi aux protocoles probabilistes ou distributionnels.

Dans cette section, nous allons étudier l'inégalité inverse. L'équation qu'on va chercher à confirmer ou infirmer, dans différents modèles de complexité, est la suivante :

$$D^{A\to B}(f) + D^{A\leftarrow B}(f) \geq D^{\|}(f). \tag{2.1}$$

Dans le cas déterministe, les deux membres de l'équation sont strictement égaux. On a vu dans la section 2.4.1 que la complexité à sens unique déterministe était caractérisée par le nombre de lignes ou de colonnes différentes (suivant la direction de la communication). D'autre part, la complexité simultanée est exactement la somme du nombre de lignes et de colonnes différentes. La raison est la même que celle qu'on a utilisée dans le modèle à sens unique. Soient x_1 et x_2 deux entrées d'Alice dont les lignes de la matrice de communication sont différentes. Il existe donc une entrée y de Bob telle que $f(x_1, y) \neq f(x_2, y)$. Or si Alice envoyait le même message à l'arbitre pour les deux entrées, celui-ci donnerait la même réponse, ce qui contredit la correction du protocole. En faisant le même raisonnement pour Bob, on en déduit que l'inégalité 2.1 est vraie, et donc $D^{A\to B}(f) + D^{A\leftarrow B}(f) = D^{\|}(f)$.

Dans le cas probabiliste, considérons la fonction égalité définie au chapitre 1. Cette fonction est définie par $EQ(x, y) = 1$ si et seulement si $x = y$. On peut montrer pour cette fonction $R_\varepsilon^{A\to B}(Eq) = R_\varepsilon^{A\leftarrow B}(Eq) = O(\log n)$. L'un des moyens pour y parvenir est d'utiliser un code correcteur d'erreurs de Justesen [Jus72]. Un tel code est défini par une fonction $\Psi : \{0,1\}^n \to \{0,1\}^{c.n}$ qui a la propriété suivante : si $x \neq y$, alors $\mathrm{Prob}_i[(\Psi(x))_i = (\Psi(y))_i] \leq \delta$. En utilisant un tel code, il suffit à Alice de tirer un indice i au hasard et de l'envoyer à Bob avec la valeur $(\Psi(x))_i$. En comparant avec $(\Psi(y))_i$, Bob sait avec grande probabilité si les deux entrées sont égales ou différentes.

D'autre part, il a été prouvé que $R_\varepsilon^{\|}(Eq) = \Omega(\sqrt{n})$ [BK97], ce qui montre que l'équation 2.1 est fausse dans le modèle probabiliste. Toutefois, la différence entre les deux tient principalement au fait que l'aléa est privé. En effet, le théorème de Newman, référencé théorème 1.1 au chapitre 1 affirme que la différence entre aléa privé et public est au plus $O(\log n)$. Ce résultat s'applique dans le modèle à sens unique mais pas dans le modèle simultané. Si on autorise l'aléa partagé, le protocole décrit plus haut peut être exécuté avec un seul bit de communication, puisqu'Alice n'a plus à envoyer l'indice qu'elle a tiré au hasard à Bob. Dans le modèle simultané également, il suffit que les joueurs partagent leur aléa et envoient un bit tiré au hasard de leur entrée corrigée. L'arbitre peut, en comparant ces deux bits, donner la bonne réponse avec grande probabilité.

Dans le cas où l'aléa est public, le contre-exemple à l'équation 2.1 est donné par Bar-Yossef *et al.* [BYJKS02]. Il s'agit de la fonction $G : \{0,1\}^{n+\log n} \times \{0,1\}^{n+\log n} \to \{0,1\}$ définie par $G(f, x, g, y) = f(x \oplus y)$ si $f = g$ et 0 sinon. Ils montrent pour celle-ci que $R_\delta^{A\to B}(G) = O(\log n)$ alors que $D_\varepsilon^{\|}(G) = \Omega(\sqrt{n})$. Ceci montre que l'équation 2.1 est fausse dans le cas probabiliste, même avec de l'aléa partagé.

Enfin, le cas quantique est également remarquable, car il met en oeuvre une technique algorithmique bien spécifique au monde quantique. Pour la fonction égalité, il est prouvé que $Q_\varepsilon^{\|}(EQ) = O(\log n)$, bien qu'il n'y ait aucune ressource partagée entre les joueurs et l'arbitre. Ceci est possible grâce à la technique de l'empreinte quantique (*quantum fingerprinting*) [BCWdW01]. Ainsi, l'équation 2.1 pourrait être vraie pour des messages quantiques.

Le modèle que nous regardons ici est la complexité distributionnelle avec distributions produits. Et dans ce cas, on peut prouver l'équivalence entre communication simultanée et à sens unique. Ceci fait donc apparaître une différence avec le modèle probabiliste, qui est équivalent au modèle distributionnel, mais avec des distributions quelconques.

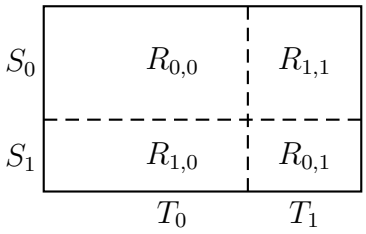
Théorème 2.8. *Soit une fonction booléenne* $f : X_1 \dots X_k \to \{0,1\}$. *Pour* $\varepsilon > (1 + 1/n)\sum_{i=1}^{k} \varepsilon_i$, *on a*

$$D_{\varepsilon}^{[],X_1||\dots X_k}(f) \leq \sum_{i=1}^{k} D_{\varepsilon_i}^{||,X_i \to X_{-i}}(f).$$

Démonstration. Nous allons prouver ce résultat par induction sur le nombre de joueurs. On commence donc par étudier le cas de deux joueurs. Soient $X = X_1$ et $Y = Y_2$ et $\mu = \mu_1 \otimes \mu_2$ une distribution sur $X \times Y$. On fixe deux protocoles à sens unique :

- dans $\mathcal{P}_1$, la communication va du joueur 1 vers le joueur 2,
- dans $\mathcal{P}_2$, la communication va du joueur 1 vers le joueur 1.

Fixons les entrées (x, y). On note $m_1(x)$ la transcription suivant $\mathcal{P}_1$, et $m_2(y)$ la transcription suivant $m_2(y)$. Soient S l'ensemble des entrées de X donnant lieu au message $m_1(x)$ et T l'ensemble des entrées de Y donnant lieu à $m_2(y)$. Soit $R = S \times T$. On partitionne R suivant le diagramme :



On définit les sous-ensembles suivants :

- $S_i = \{x \in S :$ la sortie de $\mathcal{P}_2$ sur x et $m_2(y)$ est $i\}$.
- $T_i = \{y \in T :$ la sortie de $\mathcal{P}_1$ sur y et $m_1(x)$ est $i\}$.
- $R_{i,j} = S_i \times T_j$.

On définit maintenant un protocole simultané à partir de $\mathcal{P}_1$ et $\mathcal{P}_2$. Les deux joueurs envoient leurs messages $m_1(x)$ et $m_2(y)$ à l'arbitre. L'arbitre simule les deux protocoles $\mathcal{P}_1$ et $\mathcal{P}_2$ sur toutes les entrées de R et sort la valeur 0 si $\mu(R_{0,0}) > \mu(R_{1,1})$ et 1 sinon. La complexité de ce protocole est bien la somme des complexités de $\mathcal{P}_1$ et $\mathcal{P}_2$. Il nous reste maintenant à analyser les erreurs. Supposons sans perte de généralité que la sortie de l'arbitre est 0. Regardons les erreurs de chaque sous-ensemble.

- Sur $R_{0,0}$, les sorties de $\mathcal{P}_1$ et de $\mathcal{P}_2$ sont toutes les deux 0. Si le protocole simultané est faux, alors les deux protocoles à sens unique le sont également.
- Sur $R_{i,j}$ avec $i \neq j$, les deux protocoles donnent toujours des valeurs différentes donc pour chaque entrée, l'un des deux se trompe.
- Sur $R_{1,1}$, la sortie du protocole simultané est différente. Ce protocole peut donc se tromper alors que les deux protocoles à sens unique sont corrects.

On voit donc que sur les trois premiers ensembles, le protocole simultané ne fait pas de nouvelles erreurs. En revanche, sur $R_{1,1}$ le protocole simultané crée des erreurs là où n'y en avait pas avant. En sommant sur tous les rectangles, il est facile de voir qu'on obtient $\varepsilon < \varepsilon_1 + \varepsilon_2 + \sum \mu(R_{i,i}) < 2(\varepsilon_1 + \varepsilon_2)$. En effet, le protocole est conçu de manière que pour chaque rectangle, le poids du sous-rectangle $R_{i,i}$ où les nouvelles erreurs sont créés soit borné par $\mu(R_{0,1} \cup R_{1,0})$. On peut affiner cette analyse pour prouver le théorème.

Fixons quelques notations. Soit μ' la distribution induite par μ sur R. Comme μ est par hypothèse une distribution produit, μ' est également une distribution produit. Posons donc $\mu' = \mu'_1 \otimes \mu'_2$. Soit $\alpha_1 = \mu'_1(S_1)$ et $\alpha_2 = \mu'_2(T_1)$. On va maintenant montrer que $\mu(R_{i,i}) < \frac{\mu(R_{0,1} \cup R_{1,0})}{2}$, ou de façon équivalente, $\alpha_1\alpha_2 < \frac{1}{2}(\alpha_1(1-\alpha_2) + (1-\alpha_1)\alpha_2)$. En sommant sur tous les rectangles, cela montre que $\varepsilon < \frac{3}{2}(\varepsilon_1 + \varepsilon_2)$.

La conception du protocole assure que $\alpha_1\alpha_2 < (1-\alpha_1)(1-\alpha_2)$. On peut donc supposer sans perte de généralité que $\alpha_1 < 1-\alpha_1$ (autrement, $\alpha_2 < 1-\alpha_2$). On a maintenant deux cas possibles.

- Soit $\alpha_2 \leq 1-\alpha_2$. Ceci implique $\alpha_1\alpha_2 \leq \alpha_1(1-\alpha_2)$. De même, $\alpha_1\alpha_2 \leq (1-\alpha_1)\alpha_2$. En sommant, on obtient bien $\alpha_1\alpha_2 \leq \frac{1}{2}((1-\alpha_1)\alpha_2 + \alpha_1(1-\alpha_2))$.
- Soit $\alpha_1 > 1-\alpha_1$. Notons alors que $(1-\alpha_1)(1-\alpha_2) - \alpha_1(1-\alpha_2) = (1-2\alpha_1)(1-\alpha_2) < (1-2\alpha_1)\alpha 2$. On a ainsi,

$$\begin{aligned} \alpha_1\alpha_2 &\leq (1-\alpha_1)(1-\alpha_2) \\ &\leq (1-2\alpha_1)\alpha_2 + \alpha_1(1-\alpha_2) \\ 2\alpha_1\alpha_2 &\leq \alpha_1(1-\alpha_2) + \alpha_2(1-\alpha_1) \end{aligned}$$

Ceci clot la démonstration dans le cas de deux joueurs. Soit maintenant $n > 2$ le nombre de joueurs. On fixe donc une distribution produit $\mu = \mu_1 \otimes \ldots \otimes \mu_n$ et des protocoles à sens unique $\mathcal{P}_1, \ldots, \mathcal{P}_n$. Le protocole $\mathcal{P}_i$ s'applique au problème à deux joueurs dans lequel l'un reçoit l'entrée x_i et le second x_{-i}. La communication va du premier au second. Commençons par étendre les notations dans le cas de plusieurs joueurs. On note :

- $m_i(x_i)$ la transcription du protocole $\mathcal{P}_i$,
- $S^i \subseteq X_i$, l'ensemble des entrées donnant lieu à cette transcription,
- $S_{-i} = S^1 \times \ldots \times S^{i-1} \times S^{i+1} \times \ldots S^n$,
- $S^i_a = \{x_{-i} \in S_{-i} :$ la sortie de $\mathcal{P}_i$ sur l'entréé x_{-i} et le message $m_i(x_i)$ est $a\}$.

Le protocole simultané fonctionne comme dans le cas à 2 joueurs. Le i-ème joueur envoie à l'arbitre le message $m_i(x_i)$. L'arbitre sort 0 si $\mu(\prod_i S^i_0) > \mu(\prod_i S^i_1)$ et 1 sinon. De même, la complexité de ce protocole est la somme des complexités protocoles $\mathcal{P}_i$. Nous allons maintenant analyser les erreurs de ce protocole.

Supposons sans perte de généralité que la sortie de l'arbitre est 0. Les erreurs du protocole simultané pour lesquels tous les protocoles à sens unique étaient corrects sont dans l'ensemble $\prod_i S^i_1$. Appelons μ' la distribution induite par μ sur R. μ' étant une distribution produit, on note $\mu' = \mu'_1 \otimes \ldots \otimes \mu'_n$. Enfin, on note $\alpha_i = \mu'_i(S^i)$. Avec ces notations, la propriété que l'on cherche à prouver est

$$n \prod_{i=1}^{n} \alpha_i < \sum_{\substack{S \subseteq [n] \\ S \neq \emptyset, S \neq [n]}} \prod_{i=1}^{n} \alpha_i \prod_{i=1}^{n} \mu(1-\alpha_i).$$

Le membre de gauche de cette équation est précisément $\prod_i \mu(S^i_1)$. A droite, on a la somme des mesures des ensembles tels que les protocoles à sens unique donnent des réponses différentes. Sur ces ensembles en effet, pour chaque entrée l'un au moins des protocoles donne un réponse fausse. On a donc par définition $\sum_{\substack{S \subseteq [n] \\ S \neq \emptyset, S \neq [n]}} \prod_{i=1}^{n} \alpha_i \prod_{i=1}^{n} \mu(1-\alpha_i) =$

$1 - \prod_{i=1}^{n} \alpha_i - \prod_{i=1}^{n}(1-\alpha_i)$, L'hypothèse d'induction s'écrit donc

$$(n-1)\prod_{i=1}^{n-1}\alpha_i < 1 - \prod_{i=1}^{n-1}\alpha_i - \prod_{i=1}^{n-1}(1-\alpha_i) \tag{2.2}$$

Il y a alors encore deux possibilités :

– $\prod_{i=1}^{n-1}\alpha_i \leq \prod_{i=1}^{n-1}(1-\alpha_i)$, alors en utilisant l'équation 2.2,

$$\begin{aligned} n\prod_{i=1}^{n}\alpha_i &= (n-1)\alpha_n\prod_{i=1}^{n-1}\alpha_i + \alpha_n\prod_{i=1}^{n-1}\alpha_i \\ &\leq \alpha_n \sum_{\substack{S\subset[n-1]\\ S\neq 0, S\neq[n-1]}} \prod_{i\in S}\alpha_i \prod_{i\notin S}(1-\alpha_i) + \alpha_n\prod_{i=1}^{n-1}(1-\alpha_i) \\ &\leq \sum_{\substack{S\subset[n]\\ S\neq 0, S\neq[n]}} \prod_{i\in S}\alpha_i \prod_{i\notin S}(1-\alpha_i) \end{aligned}$$

– $\alpha_n \leq 1-\alpha_n$, alors $\prod \alpha_i < (1-\alpha_n)\prod_{i=1}^{n-1}\alpha_i$. En utilisant l'équation 2.2 :

$$\begin{aligned} n\prod_{i=1}^{n}\alpha_i &= \prod_{i=1}^{n}\alpha_i + (n-1)\alpha_n\prod_{i=1}^{n-1}\alpha_i \\ &\leq (1-\alpha_n)\prod_{i=1}^{n-1}\alpha_i + \alpha_n\Big(\sum_{\substack{S\subset[n-1]\\ S\neq 0, S\neq[n-1]}} \prod_{S}\alpha_i \prod_{\overline{S}}(1-\alpha_i)\Big) \\ &\leq \sum_{\substack{S\subset[n]\\ S\neq 0, S\neq[n]}} \prod_{S}\alpha_i \prod_{\overline{S}}(1-\alpha_i) \end{aligned}$$

ce qui termine la preuve.

□

2.6 Conclusion

Dans ce chapitre, nous avons abordé la complexité de la communication sous l'aspect de la combinatoire. Pour cela, nous avons utilisé la complexité de Kolmogorov, qui nous a permis d'identifier les entrées incompressibles, de donner une borne inférieure générale, et de choisir l'aléa également incompressible. Ce dernier point est une alternative à l'utilisation du principe du minmax de Yao, permettant de transformer un protocole probabiliste en protocole déterministe, tout en limitant le nombre d'erreurs.

Nous avons donné plusieurs applications de notre méthode. Dans le cas déterministe, nous avons montré que notre borne inférieure généralisait la taille des rectangles. Nous avons étendu cela au cas distributionnel, en montrant comment prouver la borne de corruption en utilisant la complexité de Kolmogorov. Dans le cas probabiliste, nous avons donné comme application une borne inférieure générale, la VC-dimension, et une borne inférieure sur un problème particulier : le couplage caché.

Le schéma formel de chacune des preuves de borne inférieure utilisant la complexité de Kolmogorov est le suivant.

1. Fixer les entrées (et eventuellement l'aléa) de manière incompressible.
2. Appliquer le théorème principal (théorème 2.1 ou 2.3).
3. Décrire un algorithme qui calcule l'entrée sachant la transcription qu'elle a engendré.

C'est ce dernier point qui nécessite d'exploiter les propriétés combinatoires de la fonction.

Dans le cas de la VC-dimension, nous avons donné une seconde preuve qui n'utilise plus la complexité de Kolmogorov, mais uniquement des propriétés combinatoires. En analysant la structure combinatoire de la communication, nous avons montré comment comparer la communication à sens unique et la communication simultanée, dans le modèle de complexité distributionnel restreint aux distributions produits.

Le point commun de toutes ces méthodes est qu'elles ne considèrent que les propriétés combinatoires, et sont indépendantes de la structure choisie pour représenter les problèmes. On a ainsi l'avantage de placer l'analyse à un niveau d'abstraction assez élevé, ce qui permet de donner des preuves facilement compréhensibles et révèle l'intuition derrière chacun d'elles. En revanche, on ne fait pas l'économie de l'analyse combinatoire, ce qui peut se révéler être d'une relative complexité.

Chapitre 3

Méthodes algébriques et simulation des distributions causales

Dans ce chapitre, nous allons étudier de coût de la simulation des distributions causales. Rappelons brièvement le contexte. Les détails peuvent être trouvés au chapitre 1. On s'intéresse à des distributions bipartites $p(a, b|x, y)$. Le problème est analogue à la complexité de la communication : Alice et Bob reçoivent respectivement des entrées x et y et doivent produire des réponses a et b telles que celles-ci sont distribuées suivant $\mathbf{p}$. Pour y arriver, ils peuvent communiquer. Les joueurs pourront également disposer de ressources additionnelles comme l'aléa partagé ou l'intrication.

On suppose de plus que ces distributions vérifient une propriété physique qu'on appelle la causalité, c'est-à-dire que les causes précèdent leurs effets. Formellement, cela signifie que les distributions marginales $p(a|x, y)$ et $p(b|x, y)$ sont respectivement indépendantes de y et de x. Cette propriété est vraie en particulier pour une distribution obtenue en faisant des mesures bipartites sur des états intriqués. Le modèle étudié ici généralise également la complexité de la communication des fonctions booléennes.

Pour une distribution $\mathbf{p}$, on note $R_0(\mathbf{p})$ le coût de la simulation exacte de $\mathbf{p}$ par un protocole probabiliste et des messages classiques. Lorsque le problème est d'approximer $\mathbf{p}$, les joueurs doivent produire une distribution $\mathbf{p}'$ telle que $\delta(\mathbf{p}', \mathbf{p}) \leq \varepsilon$, ou $\delta(\mathbf{p}, \mathbf{p}')$ est la variation totale entre $\mathbf{p}$ et $\mathbf{p}'$ définie à la section 1.3 (définition 1.12). On note alors la complexité de la communication $R_\varepsilon(\mathbf{p})$. Sauf précision particulière, l'aléa est toujours partagé et illimité. Lorsqu'on veut néanmoins distinguer l'aléa partagé de l'aléa privé, on ajoute les exposant "pub" et "priv". Lorsque les joueurs partagent de l'intrication, on ajoute l'exposant "ent". Enfin, si les messages envoyés sont quantiques, on note la complexité de la communication $Q_\varepsilon(\mathbf{p})$.

Nous donnons à la section 3.2 une borne inférieure générale sur le coût de la simulation des distributions causales. Celle-ci est basée sur la structure géométrique sous-jacente. La section 3.3 est dédiée à l'étude de distributions binaires à marginales uniformes. Cela comprend le cas particulier des fonctions booléennes. Dans ce cas, notre borne est équivalente à celle de Linial et Shraibman [LS08b]. L'étude de la dualité, de son interprétation physique ainsi que les applications aux jeux XOR sont l'objet de la section 3.4. Ceci permet d'interpréter notre méthode en termes de violation d'inégalités de Bell et de Tsirelson. La comparaison entre violation des inégalités de Bell et de Tsirelson dans le cas général est présentée à la section 3.5. Enfin, nous terminons en montrant des bornes supérieures

sur le problème de la simulation avec erreur à la section 3.6.

3.1 Dilution d'une distribution causale.

La définition des distributions causales permet de définir certaines opérations algébriques utiles. Considérons la combinaison convexe de deux distributions causales $\mathbf{p}_1$ et $\mathbf{p}_2$, définies sur $X \times Y$ et à valeurs dans $A \times B$. De manière opérationnelle, si on sait simuler ces deux distributions, une combinaison convexe des deux est obtenue en échantillonnant $\mathbf{p}_1$ avec probabilité θ et $\mathbf{p}_2$ avec probabilité $1-\theta$, où θ est un réel entre 0 et 1. Les joueurs utilisent donc l'aléa partagé pour déterminer, suivant θ, la distribution à échantillonner. L'important ici est que θ est indépendant des entrées des joueurs x et y. On peut vérifier de plus que la combinaison convexe de deux distributions causales est bien causale. Cela se vérifie aisément par linéarité :

$$\begin{aligned}\sum_b (\theta p_1(a,b|x,y) + (1-\theta)p_2(a,b|x,y)) &= \theta p_1(a|x,y) + (1-\theta)p_2(a|x,y), \\ &= \theta p_1(a|x) + (1-\theta)p_2(a|x),\end{aligned}$$

et de même pour b.

Pour prouver une borne inférieure sur la complexité de la communication, on va mélanger la distribution avec du bruit. Ce qu'on prouve maintenant, c'est qu'en ajoutant une quantité suffisante de bruit, la distribution devient locale et peut être simulée sans communication.

Théorème 3.1. *Soit $\mathbf{p}$ une distribution causale définie sur $X \times Y$, à valeurs dans $A \times B$. Supposons que $R_0(\mathbf{p}) \leq t$, alors il existe deux distributions, $\mathbf{p}_A$ définies sur X, à valeur dans A et $\mathbf{p}_B$ définies sur Y et à valeurs dans B, telles que la distribution $\mathbf{p}_l$ définie par*

$$p_l(a,b|x,y) = \frac{1}{2^t} p(a,b|x,y) + \frac{2^t - 1}{2^t} p_A(a|x) p_B(b|y)$$

est locale.

Démonstration. Soit $\mathcal{P}$ un protocole pour simuler $\mathbf{p}$ avec au plus t bits de communication et $\mathcal{T}$ l'ensemble des transcriptions possibles de $\mathcal{P}$. On a $|\mathcal{T}| \leq 2^t$. On suppose que la longueur de toutes les transcriptions est t, quitte à ajouter des bits inutiles à la fin de celles-ci. Fixons d'abord quelques notations. Dans le protocole initial, on suppose que les joueurs tirent une chaînes aléatoire λ. La transcription du protocole est notée $T(x,y,\lambda)$. On suppose qu'à la fin du protocole, Alice choisit sa sortie suivant une distribution $p_P(a|x,\lambda,T)$. De même, Bob choisit la sienne suivant $p_P(b|y,\lambda,T)$.

Nous définirons précisément les distributions $\mathbf{p}_A$ et $\mathbf{p}_B$ plus loin. Notons cependant que la distribution $\mathbf{p}_m$ définie par $p_m(a,b|x,y) = p_A(a|x)p_B(b|y)$ est causale car locale. $\mathbf{p}_l$ étant une combinaison convexe de $\mathbf{p}$ et $\mathbf{p}_m$, elle est également causale. Pour une chaîne booléenne $\sigma \in \{0,1\}^t$, on dit que σ est compatible avec une entrée $x \in X$ s'il existe une entrée $y \in Y$ et une valeur de l'aléa λ telles que $T(x,y,\lambda) = \sigma$. On note $U_{x,\lambda}$ l'ensemble des transcriptions compatibles avec x pour l'aléa λ. De même, on note $V_{y,\lambda}$ l'ensemble des transcriptions compatibles avec y pour l'aléa λ.

On définit maintenant le protocole $\mathcal{P}'$ sans communication pour simuler $\mathbf{p}_l$:

– Alice et Bob commencent par fixer l'aléa du protocole $\mathcal{P}$ en tirant une chaîne aléatoire au hasard.
– Alice et Bob tirent ensuite une chaîne de longueur t au hasard. Soit σ cette chaîne.
– Si σ est compatible avec x, alors Alice donne la valeur qu'elle aurait donnée dans le protocole $\mathcal{P}$, autrement, elle répond au hasard suivant la distribution p_A.
– Bob fait la même chose qu'Alice, en répondant suivant $\mathbf{p}_B$ si la σ n'est pas compatible avec y.

Ce protocole est sans communication. On note μ la distribution sur le choix de la chaîne aléatoire λ et de la transcription σ. Par définition, la distribution obtenue est

$$\begin{aligned} p_l(a,b|x,y) &= \sum_\lambda \mu(\lambda) \left[\sum_{T\in U_{x,\lambda}\cap V_{y,\lambda}} \mu(T) p_P(a|x,\lambda,T) p_P(b|y,\lambda,T) \right. \\ &+ p_B(b|y) \sum_{T\in U_{x,\lambda}\cap \bar{V}_{y,\lambda}} \mu(T) p_P(a|x,\lambda,T) + p_A(a|x) \sum_{T\in \bar{U}_{x,\lambda}\cap V_{y,\lambda}} \mu(T) p_P(b|y,\lambda,T) \\ &+ \left. p_B(b|y) p_A(a|x) \sum_{T\in \bar{U}_{x,\lambda}\cap \bar{V}_{y,\lambda}} \mu(T) \right] \end{aligned}$$

Analysons chacun des termes séparément. Lorsque λ est fixé, il n'y a qu'une seule transcription compatible à la fois avec x et y. On a ainsi

$$\sum_\lambda \mu(\lambda) \sum_{T\in U_{x,\lambda}\cap V_{y,\lambda}} \mu(T) p_P(a|x,\lambda,T) p_P(b|y,\lambda,T) = \frac{1}{2^t} p(a,b|x,y).$$

Soit A_x l'événement qui indique que la transcription d'Alice est compatible avec son entrée x, lorsqu'on choisit λ et T suivant μ. On définit B_y de manière analogue. On note ainsi

$$p_P(a|x, A_x \cap \bar{B}_y) = \frac{\sum_\lambda \mu(\lambda) \sum_{T\in U_{x,\lambda}\cap \bar{V}_{y,\lambda}} \mu(T) p_P(a|x,\lambda,T)}{\mu(A_x \cap \bar{B}_y)},$$

où on a par définition $\mu(A_x \cap \bar{B}_y) = \sum_\lambda \mu(\lambda) \sum_{T\in U_{x,\lambda}\cap \bar{V}_{y,\lambda}} \mu(T)$. Nous allons montrer que cette distribution est indépendante de y, et que la distribution analogue $p_P(b|y, \bar{A}_x \cap B_y)$ pour Bob est indépendante de x. En utilisant ces distributions, on a l'expression suivante pour $\mathbf{p}_l$

$$\begin{aligned} p_l(a,b|x,y) &= \frac{1}{2^t} p(a,b|x,y) + \mu(A_x \cap \bar{B}_y) p_B(b|y) p_P(a|x, A_x \cap \bar{B}_y) \\ &+ \mu(\bar{A}_x \cap B_y) p_A(a|x) p_P(b|x, \bar{A}_x \cap B_y) + \mu(\bar{A}_x \cap \bar{B}_y) p_B(b|y) p_A(a|x) \end{aligned}$$

En sommant sur b et en utilisant le fait que $\mathbf{p}$ et $\mathbf{p}_l$ sont causales, on a

$$\begin{aligned} p_l(a|x) &= \frac{1}{2^t} p(a|x) + \mu(A_x \cap \bar{B}_y) p_P(a|x, A_x \cap \bar{B}_y) \\ &+ \mu(\bar{A}_x \cap B_y) p_A(a|x) + \mu(\bar{A}_x \cap \bar{B}_y) p_A(a|x) \\ &= \frac{1}{2^t} p(a|x) + \mu(A_x \cap \bar{B}_y) p_P(a|x, A_x \cap \bar{B}_y) + \mu(\bar{A}_x) p_A(a|x), \end{aligned}$$

Par définition, $\mu(A_x)$ est indépendant de y, et il en va donc de même pour $\mu(A_x \cap \bar{B}_y) = \mu(A_x) - \mu(A_x \cap B_y) = \mu(A_x) - \frac{1}{2^t}$ De l'expression précédente de $\mathbf{p}_l$, on déduit donc que $p_P(a|x, A_x \cap \bar{B}_y)$ est indépendant de y. On peut donc poser $p_A(a|x) = p_P(a|x, A_x \cap \bar{B}_y)$. En faisant le même raisonnement pour Bob, on peut poser $p_B(b|y) = p_P(b|y, \bar{A}_x \cap B_y)$. On a enfin l'expression suivante pour $\mathbf{p}_l$=

$$\begin{aligned} p_l(a,b|x,y) &= \frac{1}{2^t}p(a,b|x,y) + \mu(A_x \cap \bar{B}_y)p_A(a|x)p_B(b|y) \\ &+ \mu(\bar{A}_x \cap B_y)p_A(a|x)p_B(b|y) + \mu(\bar{A}_x \cap \bar{B}_y)p_A(a|x)p_B(b|y) \\ &= \frac{1}{2^t}p(ab|xy) + (1 - \frac{1}{2^t})p_A(a|x)p_B(b|y). \end{aligned}$$

□

Pour toute distribution, on a toujours une borne triviale $R_0(\mathbf{p}) \leq \log(\min\{|X|, |Y|\})$. En effet, il suffit qu'un joueur envoie son entrée à l'autre. Appliquons alors le théorème à une distribution $\mathbf{p}$ binaire à marginales uniformes. Le produit des marginales est alors la distribution définie par $p_m(a, b|x, y) = \frac{1}{4}$. Le théorème affirme donc que la distribution $\frac{1}{2^n}p(a, b|x, y) + \frac{2^n-1}{2^n}\frac{1}{4}$ est locale.

On a vu au chapitre 1 qu'une distribution binaire à marginales uniformes $\mathbf{p}$ était entièrement déterminée par sa matrice de corrélation. Soit C une matrice de corrélation correspondant à $\mathbf{p}$. La distribution $\mathbf{p}_m$, produit des marginales de $\mathbf{p}$, est uniforme pour tout x, y. On appelle cette distribution le bruit uniforme. Il est facile de vérifier que la matrice de corrélation correspondante est identiquement nulle. Par linéarité, la matrice de corrélation correspondant à $\mathbf{p}_l$ est donc $C/2^n$.

Supposons maintenant que C_f est la matrice de corrélation associée à une fonction booléenne f. Si Alice et Bob simulent la distribution $(\frac{C}{2^t}, 0, 0)$, ils calculent correctement la valeur de la fonction avec probabilité $1/2 + 1/2^n$. On peut interpréter ce résultat autrement en terme de complexité de la communication. Ce qu'on voit, c'est que si on autorise une erreur arbitrairement proche de $1/2$, et pas une erreur constante comme on le fait habituellement, et si l'aléa est public, alors le calcul d'une fonction est toujours trivial. Ce résultat a déjà été utilisé pour montrer que les problèmes de communication deviennent triviaux avec lorsque les joueurs partagent des boites non-locales bruitées [BBL$^+$06].

3.2 Borne inférieure sur la communication

Nous allons maintenant utiliser le théorème 3.1 pour prouver une borne inférieure sur la communication nécessaire pour simuler une distribution causale. La plupart des travaux précédents sur ce problème sont des bornes supérieures sur le coût de la simulation des mesures bipartites sur les états quantiques intriqués. La plupart portent même spécifiquement sur les mesures sur des états maximalement intriqués, ces mesures donnant lieu à des distributions à marginales uniformes. Ces résultats sont récapitulés dans l'introduction, dans le tableau 1, page xvi.

Il existe quelques résultats de bornes inférieures pour des problèmes spécifiques. Notamment, Brassard, Cleve et Tapp ont étudié la distribution suivante, appelée distribution de Deutsch-Josza [BCT99]. Elle est définie comme résultat du protocole suivant : 2 joueurs partagent l'état $|\varphi^+\rangle^{\otimes \log n} = \frac{1}{\sqrt{n}} \sum_{i \in [n]} |i\rangle|i\rangle$, un produit de $\log n$ paires EPR.

Les mesures sont indexées par des chaînes de longueur n. Sur l'entrée $z \in \{0,1\}^{\log n}$, les joueurs vont appliquer successivement sur chacune de leurs parties de paire EPR les deux transformations suivantes :

- un décalage de phase $|i\rangle \mapsto (-1)^{z_i}|i\rangle$,
- une transformée de Hadamard $|i\rangle \mapsto \frac{1}{\sqrt{n}} \sum_{j\in[n]} (-1)^{i.j}|j\rangle$.

Ils mesurent ensuite dans la base $\{|i\rangle, i \in [n]\}$. Les auteurs ont montré que la simulation de cette distribution avec aléa partagé et communication classique nécessitait n bits de communication.

Il est intéressant de noter que ce résultat repose essentiellement sur un résultat précédent. En effet, cette borne inférieure est basée sur une séparation exponentielle entre communications classique et quantique pour le calcul sans erreur de la fonction EQ', un version partielle de la fonction égalité BCW98.

Récemment, Gavinsky a également prouvé une borne inférieure pour une distribution donnée [Gav09]. Là encore, celle-ci est basée sur une séparation exponentielle entre communication classique et quantique pour un problème relationnel. Si l'étude des théories non-locales généralisées a donné lieu à une importante littérature, présentée en introduction, il n'existe pas à notre connaissance de résultats de complexité similaires à ceux introduits ici.

Définition 3.1. *Soit* $\mathbf{p}$ *une distribution causale. On note*

- $\nu(\mathbf{p}) = min\{\sum |q_i| : \exists p_i \in \mathcal{L}, \sum q_i \mathbf{p}_i = \mathbf{p}\}$
- $\gamma_2(\mathbf{p}) = min\{\sum |q_i| : \exists p_i \in \mathcal{Q}, \sum q_i \mathbf{p}_i = \mathbf{p}\}$.

Nous avons dit plus haut que nous avions toujours $R_0(\mathbf{p}) \leq \log(\min\{|X|, |Y|\})$ pour toute distribution causale. On sait donc que la complexité de la communication est finie dès que l'ensemble des entrées est fini. En combinant ceci avec le théorème 3.1, on peut déduire que pour toute distribution causale $\mathbf{p}$, $\nu(\mathbf{p})$ est fini. En effet, on va voir plus loin qu'on peut utiliser ce théorème pour déduire une expression de $\mathbf{p}$ comme combinaison linéaire de distributions locales. Comme $\mathcal{L} \subseteq \mathcal{Q}$, on a $\nu(\mathbf{p}) \geq \gamma_2(\mathbf{p})$ et donc $\gamma_2(\mathbf{p})$ est également fini pour toute distribution causale $\mathbf{p}$.

Fixons une combinaison $(q_i, \mathbf{p}_i)_i$ telle que $\sum q_i \mathbf{p}_i = \mathbf{p}$. En fixant les entrées et en sommant sur les sorties on trouve $\sum q_i = 1$. On en déduit que l'expression de $\mathbf{p}$ est une combinaison affine de distributions locales. La quantité ν représente intuitivement le coût de la meilleure telle décomposition de $\mathbf{p}$. On verra que dans le cas binaire avec marginales uniformes, cette quantité est au sens mathématique une mesure de la distance de $\mathbf{p}$ à $\mathcal{L}$.

On peut faire une autre remarque sur la définition de ν et γ_2. Dans une décomposition $\mathbf{p} = \sum q_i \mathbf{p}_i$, on ne borne pas a priori le nombre de distributions $\mathbf{p}_i$ utilisées. Pour ν, le plus naturel semble de décomposer $\mathbf{p}$ en combinaison affine des points extrémaux du polytope local. Mais dans le cas quantique, le nombre de points extrémaux de $\mathcal{Q}$ est infini. Pour γ_2, mais également pour ν, deux distributions sont en fait suffisantes pour exprimer $\mathbf{p}$. Etant donnée une décomposition $\mathbf{p} = \sum q_i \mathbf{p}_i$, regroupons les p_i en fonction des signes des q_i. On obtient $\mathbf{p} = \sum_{i:q_i \geq 0} q_i \mathbf{p}_i + \sum_{i:q_i<0} q_i \mathbf{p}_i$. Posons maintenant $q^+ = \sum_{i:q_i\geq 0} q_i$, $q^- = -\sum_{i:q_i<0} q_i$, $\mathbf{p}^+ = \sum_{i:q_i\geq 0} \frac{q_i}{q^+}\mathbf{p}_i$ et $\mathbf{p}^- = \sum_{i:q_i<0} \frac{q_i}{q^-}\mathbf{p}_i$, et on a $\mathbf{p} = q^+\mathbf{p}^+ - q^-\mathbf{p}^-$. De plus, $\mathbf{p}^+$ et $\mathbf{p}^-$ sont des combinaisons convexes de $\mathbf{p}_i$, et $\mathcal{L}$ et $\mathcal{Q}$ étant convexes, on en déduit les définitions équivalentes suivantes de ν et γ_2.

Proposition 3.1. *Soit* $\mathbf{p}$ *une distribution causale, on a*

– $\nu(\mathbf{p}) = \{q^+ + q^- : \exists p^+, p^- \in \mathcal{L},\ \mathbf{p} = q^+\mathbf{p}^+ - q^-\mathbf{p}^-,\ q^+, q^- \geq 0\}$.
– $\gamma_2(\mathbf{p}) = \{q^+ + q^- : \exists p^+, p^- \in \mathcal{Q},\ \mathbf{p} = q^+\mathbf{p}^+ - q^-\mathbf{p}^-,\ q^+, q^- \geq 0\}$.

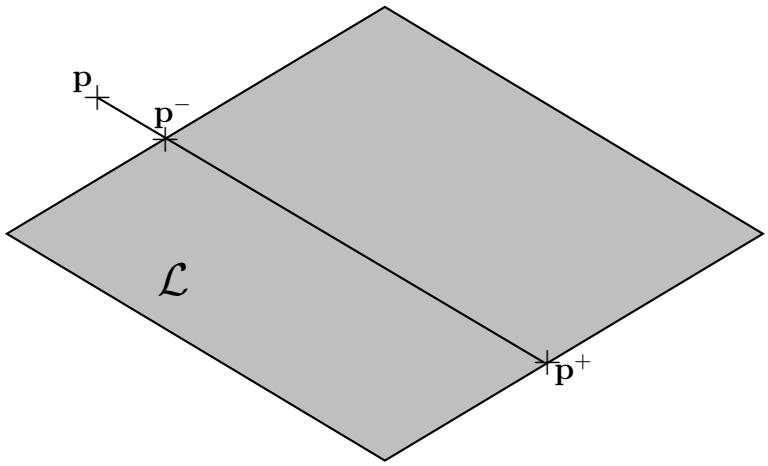


FIG. 3.1 – Décomposition affine de la distribution $\mathbf{p}$ en distributions locales

Etant donnée une distribution causale $\mathbf{p}$, le théorème 3.1 prouvé dans la section précédente induit justement une décomposition affine de $\mathbf{p}$. En écrivant explicitement cette décomposition, on en déduit une borne inférieure sur la communication.

Théorème 3.2. *Pour toute distribution causale* $\mathbf{p}$*,*
– $R_0(\mathbf{p}) \geq \log(\nu(\mathbf{p}) + 1) - 1$.
– $Q_0(\mathbf{p}) \geq \log(\gamma_2(\mathbf{p}) + 1) - 1$.

Démonstration. Soit $t = R_0(\mathbf{p})$. D'après le théorème 3.1, la distribution $\mathbf{p}_l$, définie par

$$p_l(a, b|x, y) = \frac{1}{2^t} p(a, b|x, y) + \frac{2^t - 1}{2^t} p(a|x)p(b|y) \tag{3.1}$$

est locale. Remarquons que la distribution $p(a|x)p(b|y)$ est également locale. En inversant la relation 3.1, on obtient $p(a, b|x, y) = 2^t p_l(a, b|x, y) + (1 - 2^t)p(a|x)p(b|y)$. La somme des coefficients valant 1, on obtient $\mathbf{p}$ comme combinaison affine de distributions locales. Par conséquent $\nu(\mathbf{p}) \leq 2^t + (2^t - 1) = 2^{t+1} - 1$. □

Avant de conclure, on introduit les versions approchées de ν et γ_2 Celles-ci permettent de borner la communication avec erreur. La mesure $\nu^\varepsilon(\mathbf{p})$ est simplement le minimum de ν sur une boule autour de $\mathbf{p}$. La distance qui sert à définir la boule est la distance en variation, définie au chapitre 1.

Définition 3.2. *Soit* $\mathbf{p}$ *une distribution causale. On note*
– $\nu^\varepsilon(\mathbf{p}) = \min\{\nu(\mathbf{p}') : \delta(\mathbf{p}, \mathbf{p}') \ \leq \varepsilon\}$,
– $\gamma_2^\varepsilon(\mathbf{p}) = \min\{\gamma_2(\mathbf{p}') : \delta(\mathbf{p}, \mathbf{p}') \ \leq \varepsilon\}$.

Ces notions approchées donnent des bornes inférieures sur la complexité avec erreur. En effet, si on applique le théorème 3.2 à une une distribution qui minimise ν sur une boule autour d'une distribution $\mathbf{p}$, on obtient immédiatement le théorème suivant.

Théorème 3.3. *Soit* $\mathbf{p}$ *une distribution de probabilité causale. On a*
- $R_\varepsilon(\mathbf{p}) \geq \log(\nu^\varepsilon(\mathbf{p}) + 1) - 1$,
- $Q_\varepsilon(\mathbf{p}) \geq \log(\gamma_2^\varepsilon(\mathbf{p}) + 1) - 1$,

3.3 Distributions binaires à marginales uniformes

Nous traitons à part maintenant le cas des distributions binaires à marginales uniformes. Ce cas est important pour plusieurs raisons. Il inclut la complexité de la communication des fonctions booléennes et les mesures binaires sur états maximalement intriqués. Nous allons voir que pour ces distributions, on peut améliorer la borne donnée plus haut.

Il apparaît dans ce cas des différences structurelles essentielles. Dans le cas général, les distributions causales sont des familles de distributions indexées par des couples d'entrées (x, y). On peut plonger ces distributions dans un espace vectoriel réel de dimension $|X||Y||A||B|$, en identifiant les distributions à des vecteurs de l'espace. La valeur de la coordonnée d'indice (x, y, a, b) est la probabilité $p(a, b|x, y)$.

Dans le cas binaire avec marginales uniformes, nous avons montré au chapitre 1 qu'on pouvait décrire une distribution en donnant sa matrice de corrélation C, définie par $C[x, y] = \sum_{a,b} abp(a, b|x, y)$. Ceci permet en particulier de représenter très facilement les fonctions booléennes. La structure vectorielle présente d'autres avantages. La proposition suivante donne quelques propriétés spécifiques des matrices de corrélation.

Proposition 3.2. *1. C est une matrice de corrélation si et seulement si $\| C \|_\infty \leq 1$.*

2. Pour toute matrice de corrélation C et tout $\lambda > 0$, C/λ est également une matrice de corrélation.

Démonstration. 1. Soit C une matrice de corrélation. Par définition, on a $|C[x, y]| = |\mathbf{E}[ab]| \leq 1$. On a donc $\| C \|_\infty \leq 1$.

Soit maintenant C une matrice telle que $\| C \|_\infty \leq 1$. Soit $\mathbf{p}$ défini par $p(a, b|x, y) = \frac{1}{4}(1 + abC[x, y])$. Il est facile de voir que $\mathbf{p}$ ainsi défini est bien une distribution de probabilité dont la matrice de corrélation est C.

2. Immédiat par application du premier point.

□

La distribution de probabilité du deuxième point de la proposition précédente peut être décrite de manière constructive. Nous avons vu à la section 3.1 que la distribution correspondante pouvait être obtenue en mélangeant C avec du bruit uniforme, c'est-à-dire une matrice de corrélation nulle. Ici, la distribution dont la matrice de corrélation est C/λ est obtenue en échantillonnant la distribution dont la corrélation est C avec probabilité $1/\lambda$ et le bruit uniforme avec probabilité $1 - 1/\lambda$.

L'espace vectoriel dans lequel nous travaillons est donc ici l'ensemble des matrices réelles, et les distributions causales binaires à marginales uniformes en sont un sous-ensemble. On sait que par construction, $\mathcal{L}$ est un convexe de cet espace. Cet ensemble a de plus les bonnes propriétés pour définir une norme, via sa jauge ou fonctionnelle de Minkowski.

Définition 3.3. *La jauge de l'ensemble $\mathcal{L}$ est la fonction*

$$\begin{aligned} j_{\mathcal{L}} : \mathcal{M}_n(\mathbb{R}) &\rightarrow \mathbb{R} \\ C &\rightarrow \min\{\rho > 0 : C/\rho \in \mathcal{L}\}. \end{aligned}$$

Les propriétés de $\mathcal{L}$ qui font de $j_{\mathcal{L}}$ une norme sont les suivantes :
– $\mathcal{L}$ est un convexe compact,
– 0 appartient à l'intérieur de ce compact.
La boule unité de cette norme est alors exactement l'ensemble $\mathcal{L}$. On peut montrer que les mesures de complexité ν et γ_2 définies par Linial et Shraibman [LS08b] sont équivalentes aux jauges des ensembles $\mathcal{L}$ et $\mathcal{Q}$.

En revanche, les mesures ν et γ_2 telles que nous les définissons ne sont pas des normes de matrice. En effet, si on considère la matrice identiquement nulle, qui correspond au bruit, alors on a $\nu(\mathbf{p}) = 1$, et pas $\nu(\mathbf{p}) = 0$. D'après notre définition, on a de même $\nu(\mathbf{p}) = 1$ pour tout $\mathbf{p} \in \mathcal{L}$. Néanmoins, pour des distributions non locales, nous allons montrer que notre définition de ν devient égale à $j_{\mathcal{L}}$ et celle de γ_2 égale à $j_{\mathcal{Q}}$. Pour simplifier les notations, nous utiliserons ν et γ_2 pour noter les grandeurs définies à la définition 3.1, et $j_{\mathcal{L}}$ et $j_{\mathcal{Q}}$ pour les grandeurs définies par Linial et Shraibman, lorsque celles-ci sont différentes.

Plus haut, on a défini ν sur les distributions et non pas sur les corrélations. On étend cette définition de manière évidente en utilisant la proposition 1.9. Si C est une matrice de corrélation correspondant à une distribution binaire $\mathbf{p} = (C, 0, 0)$, on pose simplement $\nu(C) = \nu(\mathbf{p})$.

Proposition 3.3. *Soit $C \notin \mathcal{L}$. On a alors $j_{\mathcal{L}}(C) = \nu(C)$.*

La preuve de cette proposition s'appuie sur le lemme crucial suivant. Toutes les simplifications qu'on a dans le cas des distributions binaires à marginales uniformes viennent du résultat suivant qui assure que l'ensemble $\mathcal{L}$ est symétrique.

Lemme 3.1. *$C \in \mathcal{L} \Leftrightarrow -C \in \mathcal{L}$.*

Démonstration. Soit $C \in \mathcal{L}$. Supposons dans un premier temps que C est un point extrémal de $\mathcal{L}$. Par définition, $\mathcal{L}$ est l'enveloppe convexe des distributions locales déterministes. Il est facile de voir que la matrice de corrélation d'une distribution locale déterministe est une matrice signe de rang un. En effet, par définition, une distribution locale déterministe $\mathbf{p}$ vérifie $p(a, b|x, y) = \delta_{\alpha(x)=a}\delta_{\beta(y)=b}$, pour un couple de fonctions $\alpha : X \rightarrow \{0, 1\}$ et $\beta : Y \rightarrow \{0, 1\}$.

Notons $u(x) = (\delta_{\alpha(x)=1} - \delta_{\alpha(x)=0})$, et $v(x) = (\delta_{\beta(y)=1} - \delta_{\beta(y)=0})$. Par définition, $u \in \{-1, 1\}$ et $v \in \{-1, 1\}$. On peut voir que la matrice de corrélation de $\mathbf{p}$ est définie par par $C[x, y] = u(x)v(y)$. Donc C est bien une matrice signe de rang 1. Ainsi, $-C$ vérifie $-C[x, y] = -u(x)v(y)$ et par conséquent, $-C$ est la matrice de corrélation d'une distribution locale déterministe.

Dans le cas général, il suffit de décomposer C en combinaison convexe des points extrémaux de $\mathcal{L}$. Posons $C = \sum_i q_i C_i$, où $q_i \geq 0$, $\sum_i q_i = 1$ et C_i est une matrice signe de rang 1. On a $-C = \sum_i q_i(-C_i)$. Or d'après la première partie de la preuve, $-C_i$ est locale pour tout i. Par convexité, on en déduit que $-C$ est locale. □

De façon opérationnelle, si on a un protocole pour C, on peut facilement le transformer en un protocole pour $-C$. Il suffit qu'à la fin de l'exécution de celui-ci, Alice sorte la valeur opposée.

Démonstration de la proposition. On a par définition $\nu(C) = \nu(\mathbf{p})$ où $\mathbf{p} = (C, 0, 0)$. La transformation de $\mathbf{p}$ en C étant linéaire, on a $\nu(C) = \min\{\sum |q_i| : \sum q_i = 1, \exists C_i \in \mathcal{L}, \sum q_i C_i = C\}$. D'un autre côté, il est facile de voir que $j_{\mathcal{L}}(C) = \{\sum q_i : q_i > 0, \exists C_i \in \mathcal{L}, \sum q_i C_i = C\}$. Fixons une décomposition $C = \sum q_i C_i$ avec $\sum q_i = 1$. Posons $q'_i = |q_i|$ et $C'_i = sgn(q_i)C_i$. D'après le lemme 3.1, on a $C'_i \in \mathcal{L}$, et la décomposition $C = \sum_i q'_i C'_i$ vérifie donc $\sum q'_i \geq j_{\mathcal{L}}(C)$. On a donc par conséquent $j_{\mathcal{L}}(C) \leq \nu(C)$.

Posons maintenant $j_{\mathcal{L}}(C) = \lambda$. Comme $C \notin \mathcal{L}$, on a $\lambda > 1$ par définition de $j_{\mathcal{L}}$. De plus, on peut écrire C comme la combinaison affine suivante :

$$C = \frac{1+\lambda}{2}\frac{C}{\lambda} + \frac{1-\lambda}{2}\frac{(-C)}{\lambda}.$$

Par définition, $\frac{C}{\lambda}$ et $\frac{-C}{\lambda}$ sont des distributions locales, et on a a $\nu(C) \leq \frac{1+\lambda}{2} + \frac{\lambda-1}{2} = \lambda$. On a bien finalement $\nu(C) = j_{\mathcal{L}}(C)$ pour $C \notin L$. □

Par le même raisonnement, on obtient un résultat analogue pour $\mathcal{Q}$. Bien entendu, pour prouver que $\mathcal{Q}$ est symétrique par rapport à l'origine, on ne raisonne pas avec les points extrémaux. Ce résultat est en revanche trivial en utilisant le théorème de Tsirelson (théorème 1.10).

Proposition 3.4. *Soit $C \notin \mathcal{Q}$. On a alors $j_{\mathcal{Q}}(C) = \gamma_2(C)$.*

Passons enfin à la preuve de la borne inférieure. Dans le cas des distributions binaires à marginales uniformes, on peut améliorer le résultat obtenu précédemment. En utilisant les propositions 3.3 et 3.4, on peut travailler indifféremment avec ν ou $j_{\mathcal{L}}$, et avec γ_2 ou $j_{\mathcal{Q}}$. En effet, le seul cas où ces mesures diffèrent est celui où la complexité de la communication est nulle.

Théorème 3.4. *Soit C la matrice de corrélation d'une distribution causale binaire à marginales uniformes, définie sur $X \times Y$. On a $R_0(C) \geq \log \nu(C)$ et $R_0^{\text{ent}}(C) \geq \log \gamma_2(C)$.*

Démonstration. Commençons par prouver le théorème dans le cas où les joueurs partagent de l'intrication. Soit $\mathcal{P}$ un protocole utilisant de l'intrication partagée et t bits de communication classique. Comme dans le cas général, on dilue la distribution en tirant une transcription T au hasard. Les joueurs vérifient si T est compatible avec leurs entrées. Si elle l'est, ils répondent ce qu'ils auraient répondu suivant $\mathcal{P}$ et sinon, ils choisissent une réponse uniformément au hasard dans $\{-1, 1\}$. La distribution ainsi calculée est toujours binaire, ses marginales sont uniformes et elle a pour matrice de corrélation $C/2^t$. On a donc $C/2^t \in \mathcal{Q}$ et par définition de $j_{\mathcal{Q}}$, on a $t \geq \log j_{\mathcal{Q}}$. La preuve dans le cas classique est identique. □

Enfin, on a vu au chapitre 1 que le codage super dense et la téléportation impliquent $R_0^{\text{ent}}(C) = 2Q_0^{\text{ent}}(C)$ (proposition 1.2). D'après le théorème 3.4, on en déduit que $Q_0^{\text{ent}}(C) \geq \frac{1}{2}\log j_{\mathcal{Q}}(C)$, et par conséquent que $R_0(C) \geq \log j_{\mathcal{Q}}(C)$.

On peut améliorer ce résultat en utilisant les théorèmes de Kremer, Klauck et Yao et de Tsirelson (théorèmes 1.3 et 1.10). Ce théorème est également à la base du théorème de Linial et Shraibman pour les fonctions. On le prouve ici pour toutes les corrélations.

Théorème 3.5. *Soit C la matrice de corrélation d'une distribution causale binaire à marginales uniformes, définie sur $X \times Y$. On a $Q_0^{\text{ent}}(C) \geq \log \gamma_2(C)$.*

Démonstration. Soit $\mathcal{P}$ un protocole quantique sans erreur, utilisant q qubits et une quantité arbitraire d'intrication. D'après le théorème 1.3, il existe deux familles de vecteurs a_x et b_y tels que $a_x.b_y = C[x,y]$ et $\| a_x \| . \| b_y \| \leq 2^q$. Soit A la matrice composée des vecteur a_x en ligne, et B celle composée des b_y en ligne. On a donc $C = A \cdot B^T$. De plus, en utilisant le théorème 1.10 on en déduit qu'on a $C/2^q \in \mathcal{Q}$, et par conséquent $\log j_{\mathcal{Q}} \leq q$. □

En utilisant le codage super-dense, on en déduit le théorème suivant. Il s'agit d'une extension du théorème de Linial et Shraibman à toutes les distributions binaires à marginales uniformes.

Corollaire 3.1. *Soit C une distribution causale binaire à marginales uniformes, définie sur $X \times Y$. On a $R_0(C) \geq 2 \log \gamma_2(C)$.*

Enfin, nous avons également des versions approchées pour ces grandeurs. Par définition du problème de la simulation avec erreur, on en déduit les bornes inférieures correspondantes.

Définition 3.4. *Soit C une matrice de corrélation. On note*
- $\nu^\varepsilon(C) = \min\{\nu(C') : \| C - C' \|_\infty \leq \varepsilon\}$,
- $\gamma_2^\varepsilon(C) = \min\{\gamma_2(C') : \| C - C' \|_\infty \leq \varepsilon\}$.

Cette définition des grandeurs approchées est cohérente avec celle donnée dans le cas général. En revanche, il y a une différence avec la manière dont elles sont définies par Linial et Shraibman. La relaxation qu'ils proposent consiste à minimiser ν sur les matrices C telles que $1 \leq C[x,y]C'[x,y] \leq \alpha$. Cette condition est tout à fait opérationnelle pour des fonctions, c'est-à-dire lorsque C est une matrice de signe, mais pas pour des corrélations en général. Pour des fonctions, la proposition suivante précise le lien entre les deux relaxations.

Proposition 3.5. *Soit $f : X \times Y \to \{-1,1\}$ une fonction booléenne, $\varepsilon > 0$ et $\alpha = \frac{1}{1-2\varepsilon}$. Soit C_f la matrice de corrélation de la distribution $\mathbf{p}_f$. Notons*

$$\widehat{\nu}^\alpha(C_f) = \min\{\nu(C) : 1 \leq C_f[x,y]C[x,y] \leq \alpha\},$$

Si $\nu(C_f) > 1$ alors $\widehat{\nu}^\alpha(C_f) = \frac{\nu^\varepsilon(C_f)}{1-2\varepsilon}$.

Démonstration. Montrons d'abord que $\widehat{\nu}^\alpha(C_f) \leq \frac{\nu^\varepsilon(C_f)}{1-2\varepsilon}$. Soit C une matrice de corrélation telle que $\nu(C) = \nu^\varepsilon(C_f)$ et $\max_{x,y} |C[x,y] - C_f[x,y]| \leq 2\varepsilon$. Comme $C_f[x,y] \in \{-1,1\}$ et C est une corrélation, on a donc $1 - 2\varepsilon \leq |C[x,y]| \leq 1$ et donc $1 \leq C[x,y]C_f[x,y] \leq \frac{1}{1-2\varepsilon}$. On a donc $\nu(C) \geq \widehat{\nu}^\alpha(C_f)$.

Montrons maintenant $\widehat{\nu}^\alpha(C_f) \geq \frac{\nu^\varepsilon(C_f)}{1-2\varepsilon}$. Soit C une matrice de corrélation telle que $\nu(C) = \widehat{\nu}^\alpha(C_f)$ et pour tout x, y, $1 \leq C[x,y]C_f[x,y] \leq \alpha$. On en déduit aisément que $|C_f[x,y] - (1-2\varepsilon)C[x,y]| \leq 2\varepsilon$ et donc $\nu(C) \geq \frac{\nu^\varepsilon(C_f)}{1-2\varepsilon}$. □

Cette proposition nous permet d'établir la borne inférieure générale sur les matrices de corrélation.

Théorème 3.6. *Pour toute matrice de corrélation C, on a*
- $R_\varepsilon(C) \geq 2\log\gamma_2^\varepsilon(C)$
- $Q_\varepsilon^{\text{ent}}(C) \geq \log\gamma_2^\varepsilon(C)$.

Dans la section 3.6, nous aurons un cas particulier dans lequel la relaxation classique est plus utile. Il s'agit dans l'inégalité $1 \leq C[x,y]C'[x,y] \leq \alpha$ de faire tendre α vers l'infini. Pour une fonction booléenne f, cela revient à minimiser pour toutes les matrices C telles que $1 \leq C_f[x,y]C[x,y]$, c'est-à-dire les matrices dont les coefficients ont les même signes que C_f, la matrice de communication associée à f.

Définition 3.5. *Soit $f : X \times Y \to \{-1,1\}$ une fonction booléenne. On note*

$$\begin{aligned}\nu^\infty(f) &= \min\{\nu(C) : 1 \leq C_f[x,y]C[x,y]\}\\ \gamma_2^\infty(f) &= \min\{\gamma_2(C) : 1 \leq C_f[x,y]C[x,y]\}\end{aligned}$$

3.4 Dualité, Inégalités de Bell et de Tsirleson, application aux jeux XOR.

3.4.1 Inégalités de Bell et Tsirelson

Les bornes inférieures sur la communication que nous avons données étaient exprimées comme des problèmes de minimisation. Il est en général plus pertinent d'avoir des bornes inférieures exprimées comme problème de maximisation. En effet, si une borne inférieure est exprimée comme un problème de maximisation d'une fonction f sur un domaine $\mathcal{D}$, alors pour tout élément de $x \in \mathcal{D}$, $f(x)$ est une borne inférieure.

Dans cette section, nous allons donner une formulation duale de nos bornes. Cette formulation est en termes de violation d'inégalité de Bell et de Tsirelson. Les inégalités de Bell sont définies par des fonctions affines, permettant de séparer les distributions locales des distributions non-locales. Bell les a introduites pour montrer l'existence de distributions quantiques non-locales [Bel64], c'est-à-dire violant les inégalités de Bell. La même idée permet de séparer les distributions quantiques des distributions causales générales. Les opérateurs permettant ces séparations sont les inégalités de Tsirelson [Tsi80].

Par la suite, si U est une partie d'un espace vectoriel euclidien, on note aff(U) l'espace affine engendré par U, soit $\text{aff(U)} = \{\lambda\text{u} + \mu\text{v} : \text{u}, \text{v} \in \text{U}, \lambda + \mu = 1\}$. Rappelons également qu'on peut plonger $\mathcal{C}$ dans un espace vectoriel réel de dimension $|X||Y||A||B|$.

Définition 3.6 (Inégalité de Bell normalisée)**.** *Soit B une forme linéaire définie sur* aff($\mathcal{C}$)*. On dit que B est une inégalité de Bell normalisée si B vérifie $|B(\mathbf{p})| \leq 1$ pour tout $\mathbf{p}$ dans $\mathcal{L}$.*

Définition 3.7 (Inégalité de Tsirelson normalisée)**.** *Soit B une forme linéaire définie sur* aff($\mathcal{C}$)*. On dit que B est une inégalité de Tsirelson normalisée si B vérifie $|B(\mathbf{p})| \leq 1$ pour tout $\mathbf{p}$ dans $\mathcal{Q}$.*

Ainsi, on dit qu'une distribution $\mathbf{p}$ viole une inégalité de Bell normalisée B si $B(\mathbf{p}) \geq 1$. Cela implique que $\mathbf{p}$ est non-locale. On peut bien entendu faire le même raisonnement avec les inégalités de Tsirelson pour montrer qu'une distribution n'est pas quantique.

Nous allons maintenant prouver que les bornes inférieures présentées plus haut sont égales aux violations des inégalités de Bell et de Tsirelson normalisée. Cela confirme l'intuition physique qui est que plus la violation est grande, plus la distribution nécessite de communication pour être simulée. Pironio a déjà utilisé cette intuition pour donner des bornes inférieures sur la complexité en moyenne [Pir03]. Nous utilisons ici la même intuition pour la complexité en pire cas.

Théorème 3.7. *Pour toute distribution* $\mathbf{p} \in \mathcal{C}$*,*

1. $\nu(\mathbf{p}) = \max\{B(\mathbf{p}) : B \text{ est une inégalité de Bell normalisée}\}$
2. $\gamma_2(\mathbf{p}) = \max\{B(\mathbf{p}) : B \text{ est une inégalité de Tsirelson normalisée}\}$

Le premier point du théorème se prouve facilement en utilisant la dualité de la programmation linéaire. Pour le second point, on introduit quelques résultats nécessaires sur la structure de $\mathcal{Q}$.

Lemme 3.2. *Soit* $C(\mathcal{Q}) = \{\theta_1\mathbf{p}_1 + \theta_2\mathbf{p}_2 : \theta_i \geq 0, p_i \in \mathcal{Q}\}$*. C'est le plus petit cône contenant* $\mathcal{Q}$*. Soit* $\mathbb{R}^+\mathcal{Q} = \{\lambda\mathbf{p} : (\lambda, \mathbf{p}) \in \mathbb{R}^+ \times \mathcal{Q}\}$*. On a* $C(\mathcal{Q}) = \mathbb{R}^+\mathcal{Q}$*.*

Démonstration. On a immédiatement $\mathbb{R}^+\mathcal{Q} \subseteq C(\mathcal{Q})$. Montrons que l'inverse est également vrai. Soit $\mathbf{p} \in C(\mathcal{Q})$. Soit $(\theta_1, \theta_2, \mathbf{p}_1, \mathbf{p}_2) \in (\mathbb{R}^+)^2 \times (\mathcal{Q})^2$ tel que $\mathbf{p} = \theta_1\mathbf{p}_1 + \theta_2\mathbf{p}_2$. On a donc

$$\begin{aligned} \mathbf{p} &= (\theta_1 + \theta_2)\left(\frac{\theta_1}{\theta_1+\theta_2}\mathbf{p}_1 + \frac{\theta_2}{\theta_1+\theta_2}\mathbf{p}_2\right), \\ &= (\theta_1 + \theta_2)\mathbf{p}', \end{aligned}$$

où $\mathbf{p}' = \frac{\theta_1}{\theta_1+\theta_2}\mathbf{p}_1 + \frac{\theta_2}{\theta_1+\theta_2}\mathbf{p}_2$. $\mathcal{Q}$ étant convexe, on a $\mathbf{p}' \in \mathcal{Q}$, et donc $\mathbf{p} \in \mathbb{R}^+\mathcal{Q}$. □

Corollaire 3.2. *Soit* P *l'ensemble des vecteurs* $\mathbf{p}$ *sur* $A \times B \times X \times Y$ *tel que pour tout* $(x, y) \in X \times Y$*,* $p[a, b, x, y]$ *définit une distribution de probabilité sur* $A \times B$*. On a alors* $\mathcal{Q} = \mathbb{R}^+\mathcal{Q} \cap P$*.*

Notons $\varphi = \frac{1}{|X||Y|}.\mathbf{1}$, où $\mathbf{1}$ est le vecteur dont toutes les coordonnées valent 1. On a pour tout $\mathbf{p} \in \mathbb{R}^+\mathcal{Q}$, $\varphi^T\mathbf{p} = \frac{1}{|X||Y|}\sum_{a,b,x,y} p(a, b|x, y)$. Notons que tout vecteur $\mathbf{p} \in \mathbb{R}^+\mathcal{Q}$ respecte la causalité, dans le sens où $\sum_b p(a, b|x, y) = \sum_b p(a, b|x', y)$ pour tout x, x', y, a. On en déduit que pour tout x, y, x', y', $\sum_{a,b} p(a, b|x, y) = \sum_{a,b} p(a, b|x', y')$, et donc que $\varphi^T\mathbf{p} = \sum_{a,b} p(a, b|x, y)$ pour une paire (x, y) donnée. Les deux propriétés suivantes sont alors évidentes.

Proposition 3.6. *Pour tout* $\mathbf{p} \in \mathbb{R}^+\mathcal{Q}$*, on a*

1. $\varphi^T\mathbf{p} = 1$ *si et seulement si* $\mathbf{p} \in \mathcal{Q}$*,*
2. $\frac{\mathbf{p}}{\varphi^T\mathbf{p}} \in \mathcal{Q}$*.*

Démonstration du théorème. Le premier point se prouve en utilisant la dualité de la programmation linéaire. Pour deux vecteurs x et y, on note $x \succeq 0$ lorsque toutes les coordonnées de x sont positives et $x \succeq y$ lorsque $x - y \succeq 0$. Enfin, $x \preceq y$ est équivalent à $y \succeq x$.

Soient $\mathbf{p}_1, \ldots, \mathbf{p}_l$ les points extrémaux du polytope local et $M_{\mathcal{L}}$ la matrice contenant en ligne les vecteurs $\mathbf{p}_1, \ldots, \mathbf{p}_l$. La définition du ν s'écrit alors :

$$\begin{aligned} \nu(\mathbf{p}) &= \min \sum q_i^+ + q_i^-, \\ &\text{s.à } M_{\mathcal{L}} q^+ - M_{\mathcal{L}} q^- = \mathbf{p}, \\ &\quad q^+, q^- \succeq 0. \end{aligned}$$

On a ainsi exprimé ν comme solution d'un programme linéaire. Le dual de celui-ci est

$$\begin{aligned} \nu^\star(\mathbf{p}) &= \max B^T \mathbf{p}. \\ &\text{s.à } -\mathbf{1} \preceq M_{\mathcal{L}}^T B \preceq \mathbf{1}, \end{aligned}$$

La condition $-\mathbf{1} \preceq M_{\mathcal{L}}^T B \preceq \mathbf{1}$ exprime bien le fait que B, appliqué à une ligne de $M_{\mathcal{L}}$, c'est-à-dire à une distribution locale, est compris entre -1 et 1, autrement dit, que B est une inégalité de Bell normalisée. La faisabilité de la solution d'un programme linéaire implique directement $\nu^\star(\mathbf{p}) = \nu(\mathbf{p})$.

Montrons maintenant le point 2 du théorème. D'après la proposition 3.1 on a $\gamma_2(\mathbf{p}) = \{q^+ + q^- : \exists p^+, p^- \in \mathcal{Q},\ \mathbf{p} = q^+\mathbf{p}^+ - q^-\mathbf{p}^-,\ q^+, q^- \geq 0\}$. De plus, en utilisant le vecteur φ, on a pour tout $\mathbf{p} \in \mathcal{Q}$ et $q \in \mathbb{R}$, $\varphi^T(q\mathbf{p}) = q$. On en déduit donc l'expression suivante pour γ_2 :

$$\begin{aligned} \gamma_2(\mathbf{p}) &= \min(\varphi^T \mathbf{p}^+ + \varphi^T \mathbf{p}^-), \\ &\text{s.à } \mathbf{p}^+ - \mathbf{p}^- = \mathbf{p}, \\ &\quad \mathbf{p}^+, \mathbf{p}^- \in \mathbb{R}^+\mathcal{Q}. \end{aligned}$$

On va maintenant écrire le dual de Lagrange de ce problème. Ceci va nous donner une expression duale, et immédiatement la dualité faible. De plus, lorsqu'on a un programme convexe, comme celui qui définit γ_2, et son dual de Lagrange, il existe des conditions simples pour prouver la dualité forte, c'est-à-dire la propriété recherchée (voir par exemple [BV04]). Soit L la fonction de Lagrange associée au problème précédent. $\mathbf{p}$ étant fixé, elle est définie par :

$$\begin{aligned} L(p^+, p^-, \lambda^+, \lambda^-, \eta) &= \varphi^T p^+ + \varphi^T p^- - \lambda_-^T p^- - \lambda_+^T p^+ + \eta^T(p^+ - p^- - p), \\ &= (\varphi - \lambda_+ + \eta)^T \mathbf{p}^+ + (\varphi - \lambda_- - \eta)^T \mathbf{p}^- - \eta^T \mathbf{p}, \end{aligned}$$

où λ^+, λ^- et η sont des vecteurs réels. On définit de plus la fonction g :

$$\begin{aligned} g(\lambda_+, \lambda_-, \eta) &= \inf_{\mathbf{p}^+, \mathbf{p}^-} L(\mathbf{p}^+, \mathbf{p}^-, \lambda^+, \lambda^-, \eta\}, \\ &= \begin{cases} -\eta^T \mathbf{p} & \text{si } \lambda_+ = \varphi + \eta \text{ et } \lambda_- = \varphi - \eta, \\ -\infty & \text{sinon.} \end{cases} \end{aligned}$$

Le problème dual de Lagrange s'exprime alors :

$$\begin{aligned} \gamma_2^\star(\mathbf{p}) &= \max g(\lambda_+, \lambda_-, \eta) \\ &\text{s.à } \forall \hat{\mathbf{p}} \in \mathbb{R}^+\mathcal{Q}\ , \lambda_+^T \hat{\mathbf{p}} \geq 0 \text{ et } \lambda_-^T \hat{\mathbf{p}} \geq 0. \end{aligned}$$

Grace à la condition sur λ_+ et λ_-, on a l'assurance que la solution du problème dual est une borne inférieure sur la solution du problème primal. De plus, en analysant l'expression algébrique de g, nous avons vu que sans aucune condition supplémentaire sur η, cette borne inférieure était $-\infty$. Ajoutons au problème dual les conditions permettant d'obtenir une borne inférieure finie.

$$\begin{aligned} \gamma_2^\star(\mathbf{p}) &= \max -\eta^T \mathbf{p} \\ &\text{s.à} \quad -\varphi^T \hat{\mathbf{p}} \leq \eta^T \hat{\mathbf{p}} \leq \varphi^T \hat{\mathbf{p}} \text{ pour tout } \hat{\mathbf{p}} \in \mathbb{R}^+\mathcal{Q}. \end{aligned}$$

Enfin, d'après la proposition 3.6, cette condition est équivalente à $-1 \leq \eta^T \hat{\mathbf{p}} \leq 1$ pour tout $\hat{\mathbf{p}} \in Q$. En faisant le changement de variable $\eta \mapsto -\eta$, on obtient bien le programme voulu, c'est-à-dire :

$$\begin{aligned} \gamma_2^\star(\mathbf{p}) &= \max \eta^T \mathbf{p} \\ &\text{s.à} \quad |\eta^T \hat{\mathbf{p}}| \leq 1 \text{ pour tout } \hat{\mathbf{p}} \in \mathcal{Q}. \end{aligned}$$

Par construction, on a la dualité faible $\gamma_2^\star(\mathbf{p}) \leq \gamma_2(\mathbf{p})$.

Montrons qu'on a également la dualité forte $\gamma_2^\star(\mathbf{p}) = \gamma_2(\mathbf{p})$. Le plus simple est de montrer que le programme vérifie les conditions de Slater [Roc70]. Il s'agit ici de trouver un point strictement faisable, c'est-à-dire deux éléments $\mathbf{p}^+, \mathbf{p}^- \in \mathbb{R}^+\mathcal{Q}$ tel que $\mathbf{p} = \mathbf{p}^+ - \mathbf{p}^-$ et $\mathbf{p}^+, \mathbf{p}^- \neq 0$. Nous avons vu dans la preuve du théorème 3.2 qu'on pouvait toujours exprimer une distribution $\mathbf{p}$ telle que $Q_0(\mathbf{p}) > 0$ comme combinaison affine de deux distributions. Pour une distribution $\mathbf{p} \in \mathcal{Q}$, prenons une autre distribution $\mathbf{p}' \in \mathcal{Q}$. Par convexité, $\mathbf{p}'' = \frac{1}{2}\mathbf{p} + \frac{1}{2}\mathbf{p}' \in \mathcal{Q}$. En inversant cette relation, on obtient $\mathbf{p} = 2\mathbf{p}'' - \mathbf{p}'$, ce qui prouve le théorème. □

Le théorème précédent donne une expression duale pour ν et γ_2. Cette formulation est intéressante en théorie, et nous allons en donner des applications dans la section suivante. Qu'en est-il de leur calcul explicite pour une distribution donnée ? Nous avons exprimé ν comme solution d'un programme linéaire. On a ainsi tous les algorithmes de calcul de programme linéaire à notre disposition pour calculer ν. Néanmoins, les points extrémaux du polytope local étant les matrices signes de rang 1, le programme linéaire est de taille exponentiel.

Concernant γ_2, nous ne savons pas si sous cette forme, il existe un algorithme performant pour le calculer. Il existe néanmoins une manière de l'approximer, en utilisant une approximation de l'ensemble $\mathcal{Q}$ [NPA08, DLTW08]. Il existe en effet une suite d'ensembles $\mathcal{Q}^n$ qui possède les propriétés suivantes :

- $\mathcal{Q}^{n+1} \subseteq \mathcal{Q}^n$,
- $\mathcal{Q}^n \to \mathcal{Q}$ quand $n \to \infty$,
- pour tout $n \in \mathbb{N}$, l'appartenance à $\mathcal{Q}$ est décidable par un programme semi-défini.

En utilisant cette hiérarchie, on peut approximer γ_2 par une suite monotone de fonctions $\gamma_2^n = \min\{q^+q^- : \exists \mathbf{p}^+, \mathbf{p}^- \in \mathcal{Q}^n,\ \mathbf{p} = q^+\mathbf{p}^+ - q^-\mathbf{p}^-,\ q^+, q^- \geq 0\}$. On peut montrer par ailleurs que cette suite de fonctions vérifie $\gamma_2^n \to \gamma_2$. De plus, γ_2^n est solution d'un programme semi-défini. On peut donc écrire le dual de chaque fonction γ_2^n en appliquant la dualité de la programmation semi-définie. On obtient ainsi une suite de programmes duaux dont l'optimum tend vers la violation maximale d'une inégalité de Tsirelson. Toutefois, là encore il a été prouvé que cette suite de programmes était de taille exponentielle [KKM+08].

La hiérarchie de programmes semi-définis nous donne ainsi une autre manière de prouver l'expression duale de γ_2, et également un algorithme pour la calculer. En revanche, pour introduire cette hiérarchie et prouver les propriétés nécessaires pour l'exploiter, cela demande d'introduire de nouveaux concepts, dont nous n'aurions pu tirer d'autre application ici. Les détails de cette construction peuvent être trouvés dans la littérature consacrés [NPA08, DLTW08].

3.4.2 Application aux jeux XOR

Nous appliquons les résultats précédents aux jeux XOR. On se limite pour cela au cas des distributions booléennes à marginales uniformes. Les jeux XOR sont importants d'un point de vue théorique car ils font déjà apparaître des différences entre les ressources classiques et quantiques.

Un jeu XOR est défini par la donnée d'une fonction $G : X \times Y \to \{0,1\}$ et une distribution μ sur $X \times Y$. La règle est que 2 joueurs, Alice et Bob, reçoivent respectivement $x \in X$ et $y \in Y$ suivant μ et chacun répond par une valeur, respectivement a et b, dans $\{0,1\}$. Il gagnent le jeu si le XOR de leurs réponses $a \oplus b$ est égal à $G(x,y)$. Pour appliquer nos résultats, on change légèrement la règle traditionnelle. On suppose d'abord que G est à valeurs dans $\{-1,1\}$. Au lieu de répondre par un bit dans $\{0,1\}$, les joueurs répondent dans $\{-1,1\}$, et au lieu de comparer à $G(x,y)$ le XOR de leurs réponses, on compare leur produit. Un jeu XOR est un jeu sans interaction ; c'est-à-dire que les joueurs ne sont pas autorisés à communiquer. En général, la valeur du jeu est la probabilité maximale de gagner, en fonction de la distribution sur les questions. On considérera la valeur classique, dans laquelle les joueurs peuvent utiliser de l'aléa partagé, et la valeur quantique dans laquelle ils partagent de l'intrication.

Considérons le jeu CHSH [CHSH69]. Celui-ci est défini par la distribution uniforme sur $X \times Y$ où $X \times Y = \{0,1\} \times \{0,1\}$, et par la fonction $G(x,y) = (-1)^{x.y}$. Autrement dit, les joueurs doivent donner la même valeur, sauf dans le cas $x = y = 1$, où ils doivent donner des valeurs différentes. Le jeu est sans communication mais permet des ressources additionnelles. Dans le cas de l'aléa partagé, la situation est claire. En effet, par linéarité, la probabilité de gain d'une stratégie probabiliste est une combinaison convexe des gains de stratégies déterministes. La meilleure stratégie probabiliste est donc de tout miser sur la meilleure stratégie déterministe. Par un examen exhaustif de celles-ci, on voit facilement que la meilleure stratégie déterministe est pour les 2 joueurs de toujours répondre 1. Cette stratégie permet de gagner avec probabilité 3/4, et on ne peut donc pas faire mieux. Or, on sait que si on autorise les joueurs à partager un état intriqué, alors il existe une stratégie qui permet de gagner avec probabilité $\cos^2 \frac{\pi}{8} \simeq 0.853$ [CHSH69]. Tsirelson a prouvé par ailleurs que cette probabilité était optimale. Ce jeu fait donc apparaître une différence entre ressource classique et ressource quantique, ainsi qu'une limite au pouvoir des ressources quantiques. L'étude des jeux XOR, et plus généralement des jeux à deux prouveurs sans interaction, est susceptible de donner d'autres résultats pour caractériser la puissance calculatoire de l'intrication.

Une autre question qui se pose sur la théorie des jeux sans interaction est la complexité du calcul de la valeur du jeu. La valeur d'un jeu est l'avantage maximale qu'une stratégie peut donner par rapport à une réponse complètement aléatoire. Pour un jeu XOR avec intrication, Cleve *et al.* ont montré qu'on pouvait calculer la valeur du jeu

avec une précision exponentielle en utilisant la programmation semi-définie [CHTW04]. En comparaison, Håstad a prouvé que le même problème, sans intrication, était NP-difficile [Hås01].

Le cadre général dans lequel se situe l'étude des jeux XOR est celui des jeux sans interaction à deux prouveurs. Plusieurs problèmes classiques de complexité peuvent se traduire dans ce langage. Ainsi, le théorème PCP [ALM+98] et la conjecture des jeux uniques [Kho02] peuvent être interprétés dans un cadre similaire. Le cas des jeux à deux prouveurs avec intrication présente également un intérêt pratique ici. En effet, il a été prouvé que dans le cas général d'un jeu à 2 prouveurs avec intrication, distinguer entre le cas où la valeur du jeu est 1 et le cas où elle est plus petite que $1 - \varepsilon$ est NP-difficile [KKM+08]. Ce résultat a une conséquence directe sur la suite de programmes semi-définis qui approxime l'ensemble $\mathcal{Q}$, mentionné dans la section précédente. En effet, ce résultat sur les jeux sans interaction implique qu'on ne peut construire une telle suite de programmes de taille polynomiale. En effet, si cette suite de programmes était de taille polynomiale, cela donnerait directement un algorithme pour décider si une distribution est quantique, ou si on ne peut la simuler avec une bonne précision.

Revenons aux jeux XOR et leurs liens avec les concepts de ce chapitre. Puisque le jeu est sans communication, la stratégie des joueurs produit une distribution locale binaire à marginales uniformes, qui peut être représentée par une matrice de corrélation locale $S \in \mathcal{L}$. On appelle le biais du gain de la stratégie S la quantité $\epsilon_\mu(G||S) = \sum_{x,y} \mu(x,y)S[x,y]G[x,y]$. La probabilité de gagner au jeu défini par G et μ en jouant la stratégie S est alors $\frac{1+\epsilon_\mu}{2}$. Dans notre contexte, le biais est plus facile à manipuler. Mais on peut passer de la probabilité de gagner au biais avec une bijection.

On a vu plus haut que ν et γ_2 pouvaient être interprétés comme la violation maximale d'une inégalité de Bell et de Tsirelson. Or, la définition de la valeur d'un jeu fait intervenir une forme linéaire, et on peut donc également l'interpréter comme une inégalité de Bell et de Tsirelson. On va montrer un lien concret entre les deux. Nous allons en effet montrer que la valeur classique du jeu nous permet d'avoir une borne inférieure sur ν.

Lemme 3.3. *L'ensemble des jeux XOR est isomorphe à l'ensemble des inégalités de Bell normalisées sur les corrélations, c'est à dire l'ensemble des matrices B tel que pour toute matrice de corrélation locale C, $Tr(B^T \cdot C) \leq 1$.*

Démonstration. Soit (G, μ) un jeu XOR et $\varphi_{G,\mu} : C \mapsto \frac{\epsilon_\mu(G||C)}{\epsilon_\mu(G)}$. Si $C \in \mathcal{L}$, alors $\varphi_{G,\mu}(C) \leq 1$ par définition de ϵ_μ. Comme on vérifie facilement que $\varphi_{G,\mu}$ est une application affine, on a bien défini une inégalité de Bell normalisée.

Inversement, soit B une application affine telle que pour tout $C \in \mathcal{L}$, $B(C) \leq 1$. Pour tout $C \in \mathcal{C}$, on a $B(C) = \sum_{x,y} B[x,y]C[x,y]$. Soit $\text{sign}(x)$ la fonction qui à x associe $+1$ si $x \geq 0$ et -1 sinon. Le jeu est donné par la matrice G définie par $G[x,y] = \text{sign}(B[x,y])$ et la distribution μ sur $X \times Y$ telle que $\mu(x,y) = \frac{B[x,y]}{\sum_{x,y} B[x,y]}$. On a bien défini un jeu XOR (G, μ). □

Cet isomorphisme permet de faire le lien entre les jeux XOR et la mesure ν. En effet, l'utilisation de ce lemme et le théorème 3.7, on établit le théorème suivant. Une preuve de ce théorème apparaît également dans [LSŠ08].

Théorème 3.8. *Soit C la matrice d'un jeu XOR. On a*

– $\nu(C) = \max_{(G,\mu)} \frac{\epsilon_\mu(G||C)}{\epsilon_\mu(G)}$,
– $\nu(C) \geq \frac{1}{\epsilon(C)}$.

Démonstration. Rappelons que d'après le théorème 3.7, on a

$$\nu(C) = \max_B \{\sum_{x,y} B[x,y]C[x,y] : \forall D \in \mathcal{L},\ |\sum_{x,y} B[x,y]D[x,y]| \leq 1\}.$$

En appliquant la bijection $(G,\mu) \mapsto \varphi_{G,\mu}$ définie au lemme 3.3, on a immédiatement la preuve du premier point. Par cette bijection, la violation de l'inégalité de Bell B est en effet égale à la valeur du jeu associé. Ainsi, $\epsilon_\mu(G||C)$ est la violation maximale d'une inégalité de Bell, la normalisation étant assurée par le dénominateur $\epsilon_\mu(G)$.

Pour le second point, il suffit de poser $G = C$ dans le premier point. □

3.5 Comparaison de γ_2 et ν.

L'objectif de cette section est de comparer les bornes ν et γ_2. Dans le cas des corrélations, il est connu qu'on a $K_G\gamma_2(C) \geq \nu(C) \geq \gamma_2(C)$. On prouve ce résultat en utilisant d'une part le théorème de Tsirelson (théorème 1.10), et d'autre part l'inégalité de Grothendieck [Gro56], que nous allons introduire maitenant. L'importance de cette inégalité est apparu à plusieurs endroits dans la théorie de l'information quantique, mais également dans d'autres domaines de l'informatique. A l'origine, cette inégalité vient de l'analyse fonctionnelle. Elle prouve l'existence d'une constance, appelée constante de Grothendieck, liant deux quantités que nous présentons dans la suite. Un article important de Krivine [Kri79] montre que la valeur de cette constante vaut au plus environ 1,85.

L'un des très importantes utilisations de l'inégalité de Grothendieck en informatique sert à analyser un algorithme d'approximation pour la norme appelée *Cut-Norm* [AN06]. Rappelons le schéma de cet algorithme, nous en retrouverons certains éléments par la suite.

1. Relâcher le problème initial pour le transformer en problème géométrique.
2. Résoudre le problème géométrique en utilisant la programmation semi-définie.
3. Arrondir la solution pour retrouver trouver une solution au problème inital.

L'utilisation de l'inégalité de Grothendieck permet de garantir la qualité de l'approximation obtenue par ce processus.

En plus de son utilisation en complexité de la communication par Linial et Shraibman, il y a au moins deux autres occurrences importantes de cette inégalité en théorie du traitement de l'information quantique. La première sur la théorie des jeux à deux prouveurs sans interaction. Cleve *et al.* [CHTW04] ont montré que la valeur quantique d'un jeu XOR était une relaxation semi-définie de sa valeur classique. L'inégalité de Grothendieck permet de montrer une borne supérieure sur la différence entre ces deux valeurs.

La seconde occurrence est dans l'algorithme de Regev et Toner pour simuler les mesures sur les états maximalement intriqués [RT07]. L'une des particularités de cet algorithme est d'utiliser la technique des projections aléatoires. L'inégalité de Grothendieck n'y est pas appliquée directement, mais les techniques de preuves sont inspirées de la borne supérieure prouvée par Krivine.

Théorème 3.9 (Inégalité de Grothendieck)**.** *Il existe une constante K_G telle que pour toute matrice réelle M et tout $k \geq 1$,*

$$\max_{a_i,b_j} \sum M[i,j]a_i.b_j \leq K_G \max_{\varepsilon_i,\delta_i} \sum M[i,j]\varepsilon_i\delta_j,$$

où le maximum du membre de gauche est pris sur les vecteurs unitaires $a_i, b_j \in \mathbb{R}^k$, et le maximum du membre de droite est pris sur les $\varepsilon_i, \delta_j \in \{-1, 1\}$.

Comment interpréter ce théorème dans notre contexte ? Dans le membre de droite d'abord, considérons la matrice N définie par $N[i,j] = \varepsilon_i\delta_j$. N est une matrice signe de rang 1, c'est donc un point extrémal du polytope $\mathcal{L}$. On a donc $\max \sum M[i,j]\varepsilon_i\delta_j = \max_{N\in\mathcal{L}} Tr(M^T \cdot N)$.

De même, dans le membre de droite, on maximise sur les vecteurs unitaires a_i et b_j. Or par le théorème de Tsirelson (théorème 1.10), les produits de vecteurs unitaires sont exactement des distributions de probabilité quantique. On peut donc écrire ce terme $\max \sum M[i,j]a_i.b_j = \max_{N\in\mathcal{Q}} Tr(M^T \cdot N)$.

Rappelons que pour une norme $\| \, . \, \|$ sur un espace vectoriel, la norme duale sur l'espace duale est définie par $\| \, L \, \|^* = \max_{x:\|x\|=1} \| \, L(x) \, \|$ On remarque ainsi qui l'inégalité de Grothendieck donne une relation entre les normes duales de $j_{\mathcal{L}}$ et $j_{\mathcal{Q}}$. En utilisant cela et la proposition 3.3, on obtient la proposition suivante.

Proposition 3.7 ([LS08b])**.** *Pour toute matrice de corrélation C, $K_G\gamma_2(C) \geq \nu(C) \geq \gamma_2(C)$.*

En conséquence, pour une distribution quantique, la méthode γ_2 donne une borne inférieure de $\log K_G$. Comme K_G est inférieure à 2, on a une borne triviale inférieure à 1. On ne peut utiliser cette méthode pour prouver l'optimalité du protocole de Regev et Toner pour simuler les mesures sur les états quantiques maximalement intriqués en dimension arbitraire [RT07].

De plus, on ne peut pas non plus utiliser ces méthodes pour prouver une séparation entre communication classique et quantique. En effet, la borne inférieure donnée par ν est, à une constante additive près, la même que la borne inférieure donnée par γ_2.

Pour les distributions générales, on ne peut appliquer directement l'inégalité de Grothendieck. Nous allons maintenant en donner une borne à la violation maximale d'une inégalité de Bell en fonction de la violation maximale d'une inégalité de Tsirelson. Ceci peut être vu comme une extension de l'inégalité de Grothendieck dans le cas général. L'inégalité de Grothendieck s'applique aux distributions binaires à marginales uniformes. On donne donc d'abord le résultat pour des distributions binaires à marginales arbitraires, puis pour des distributions générales.

Théorème 3.10. *Soit $\mathbf{p}$ une distribution causale sur $X \times Y$ et à valeurs dans $A \times B$.*
- *Si $|A| = |B| = 2$, alors $\nu(\mathbf{p}) \leq (2K_G + 1)\gamma_2(\mathbf{p})$,*
- *Sinon, $\nu(\mathbf{p}) \leq 2[|A||B|(K_G + 1) - 1]\gamma_2(\mathbf{p})$.*

La première implication de ce théorème est négative. Cela montre que comme dans le cas des corrélations, on ne peut utiliser nos méthodes pour prouver des séparations entre communication classique et quantique. Nous avons introduit à la section 3.2 la distribution de Deutsch-Josza [BCT99]. Pour celle-ci, le théorème affirme que $\nu(\mathbf{p}) \leq$

$2|A||B|(K_G+1)-1$. Il a été montré que cette distribution nécessite n bits pour être simulée. Or, d'après sa définition, $|A|=|B|=\log n$. Ainsi, la borne que notre méthode permet de montrer est exponentiellement plus petite que la complexité de la communication. Nous avons dit plus haut que notre méthode confirmait l'intuition selon laquelle la complexité de la communication augmentait avec la violation maximale d'une inégalité de Bell. Cet exemple montre que cette quantité ne suffit pas à caractériser la complexité de la communication.

En revanche, ce résultat donne également une information sur la structure des distributions causales. En effet, si on considère que ν est une bonne distance pour les distributions causales, alors cela donne une idée de la distance maximale entre $\mathcal{L}$ et $\mathcal{Q}$.

Nous prouvons d'abord quelques résultats intermédiaires sur la représentation des distributions.

Lemme 3.4. *Soit* $\mathbf{p}=\sum_i q_i\mathbf{p}_i$, *où pour tout* i, $\mathbf{p}_i\in\mathcal{C}$ *et* $q_i\in\mathbb{R}$. *Alors* $\nu(\mathbf{p})\leq\sum_i|q_i|\nu(\mathbf{p}_i)$.

Démonstration. Par définition de $\mathbf{p}_i$, il existe $\mathbf{p}_i^+$, et $\mathbf{p}_i^-$ et q_i^+,q_i^- tels que $\mathbf{p}_i=q_i^+\mathbf{p}_i^+-q_i^-\mathbf{p}_i^-$ et $|q_i^+|+|q_i^-|=\nu(\mathbf{p}_i)$. Ainsi, $\mathbf{p}=\sum_i q_i(q_i^+\mathbf{p}_i^+-q_i^-\mathbf{p}_i^-)$, et

$$\begin{aligned}\nu(\mathbf{p}) &\leq \sum_i |q_iq_i^+|+|q_iq_i^-|,\\ &\leq \sum_i |q_i|(q_i^++q_i^-),\\ &\leq \sum_i |q_i|\nu(\mathbf{p}_i).\end{aligned}$$

□

Dans le lemme suivant, on montre qu'agrandir le domaine ne permet pas de donner une meilleure représentation affine des distributions.

Lemme 3.5. *Soit* $\mathbf{p}$ *une distribution causale sur* $X\times Y$ *à valeurs dans* $A\times B$. *Notons* A' *et* B' *respectivement les supports des distributions marginales* $p(a|x)$ *et* $p(b|y)$, *et* $\mathbf{p}'$ *la distribution* $\mathbf{p}$ *restreinte à* $A'\times B'$. *Alors* $\nu(\mathbf{p})=\nu(\mathbf{p}')$.

Démonstration. Montrons d'abord que $\nu(\mathbf{p})\leq\nu(\mathbf{p}')$. Soit $\mathbf{p}'=q^+\mathbf{p}^+-q^-\mathbf{p}^-$, avec $\mathbf{p}^+,\mathbf{p}^-\in\mathcal{L}$. On peut étendre le domaine de $\mathbf{p}^+$ et $\mathbf{p}^-$ à tout $A\times B$ en posant $p^+(a,b|x,y)=p^-(a,b|x,y)=0$ pour tout $(a,b)\in(A\times B)\backslash(A'\times B')$. On en déduit une décomposition affine de $\mathbf{p}$ et $\nu(\mathbf{p})\leq q^++q^-=\nu(\mathbf{p}')$.

Montrons maintenant que $\nu(\mathbf{p}')\leq\nu(\mathbf{p})$. Soit $\mathbf{p}=q^+\mathbf{p}^+-q^-\mathbf{p}^-$, tel que $\mathbf{p}^+,\mathbf{p}^-\in\mathcal{L}$. Les distributions $\mathbf{p}^+$ et $\mathbf{p}^-$ ne sont pas nécessairement nulles sur $(A\times B)\backslash(A'\times B')$. On définit la distribution $\mathbf{p}'^+$ suivante :

$$\begin{aligned}p'^+(a,b|x,y) &= p^+(a,b|x,y)+\frac{1}{|A|}\sum_{a'\notin A'}p^+(a',b|x,y)+\frac{1}{|B|}\sum_{b'\notin B'}p^+(a,b'|x,y)\\ &+ \frac{1}{|A||B|}\sum_{a'\notin A',b'\notin B'}p^+(a',b'|x,y),\end{aligned}$$

et $\mathbf{p}'^{-}$ de manière similaire. Ces distributions sont bien locales. Pour simuler $\mathbf{p}'^{+}$ par exemple, il suffit d'appliquer le protocole pour $\mathbf{p}^{+}$ en remplaçant les réponses de $(A \times B)\backslash(A' \times B')$ par une réponse choisie uniformément dans $A' \times B'$. On vérifie enfin facilement que $\mathbf{p}' = q^{+}\mathbf{p}'^{+} - q^{-}\mathbf{p}'^{-}$. On a donc bien $\nu(\mathbf{p}') \leq \nu(\mathbf{p})$. □

Pour une distribution quantique, on a $\gamma_2(\mathbf{p}) = 1$. Si de plus $\mathbf{p}$ est binaire avec marginales uniformes, la proposition 3.7 implique $\nu(\mathbf{p}) \leq K_G$. Dans le prochain théorème, on étend ce résultat à des distributions quantiques avec marginales arbitraires.

Théorème 3.11. *Soit $\mathbf{p} \in \mathcal{Q}$ définie sur $X \times Y$, à valeurs dans $A \times B$.*

1. *$\nu(\mathbf{p}) \leq 2K_G + 1$ si $|A| = |B| = 2$.*
2. *$\nu(\mathbf{p}) \leq 2|A||B|(K_G + 1) - 1$ sinon.*

Démonstration. 1. Soit $\mathbf{p} = (C, M_A, M_B) \in \mathcal{Q}$. On a vu qu'alors $(C, 0, 0) \in \mathcal{Q}$ et en appliquant la proposition 3.7 et la définition de $j_{\mathcal{L}}$, $(C/K_G, 0, 0) \in \mathcal{L}$. On peut ainsi écrire $\mathbf{p}$ comme combinaison affine de distributions locales de la manière suivante :

$$(C, M_A, M_B) = K_G(C/K_G, 0, 0) + (M_A M_B, M_A, M_B) - (M_A M_B, 0, 0) - (K_G - 1)(0, 0, 0).$$

En sommant la valeur absolue des coefficients, on en déduit $\nu(\mathbf{p}) \leq 2K_G + 1$.

2. Dans le cas général, nous allons décomposer les distributions en une somme de distributions binaires et appliquer le point précédent. Pour réduire une distribution quantique à un ensemble de distributions binaires de complexité inférieure, on va augmenter le domaine d'un élément supplémentaire. Soit ω un élément tel que $\omega \notin A \cup B$. Soit $A' = A \cup \{\omega\}$ et $B' = B \cup \{\omega\}$. Pour tout $(\alpha, \beta) \in A \times B$, on définit la distribution $\mathbf{p}_{\alpha,\beta}$ sur $A' \times B'$. Celle-ci consiste pour Alice et Bob à simuler la distribution $\mathbf{p}$ en modifiant leurs sorties. Si la réponse suivant $\mathbf{p}$ est α pour Alice ou β pour Bob, alors ils répondent cette valeur. Sinon, ils répondent par la valeur ω. La distribution obtenue en appliquant ce protocole est alors :

$$p_{\alpha,\beta}(a, b|x, y) = \begin{cases} p(\alpha, \beta|x, y) & \text{si } (a, b) = (\alpha, \beta), \\ p(\alpha|x) - p(\alpha, \beta|x, y) & \text{si } (a, b) = (\alpha, \omega), \\ p(\beta|y) - p(\alpha, \beta|x, y) & \text{si } (a, b) = (\omega, \beta), \\ 1 - p(\alpha|x) - p(\beta|y) - p(\alpha, \beta|x, y) & \text{si } (a, b) = (\omega, \omega), \\ 0 & \text{sinon.} \end{cases}$$

On en déduit donc que $\mathbf{p}_{\alpha,\beta} \in \mathcal{Q}$, et par le point précédent et en restreignant $\mathbf{p}$ à son support, on a $\nu(\mathbf{p}_{\alpha,\beta}) \leq 2K_G + 1$.

Par ailleurs, on définit les 3 distributions suivantes sur $X \times Y$ à valeur dans $A' \times B'$.

– $p_A(a, b|x, y) = p(a|x)\delta_{b=\omega}$,

– $p_B(a, b|x, y) = \delta_{a=\omega}p(b|y)$,

– $p_\omega(a, b|x, y) = \delta_{a=\omega}\delta_{b=\omega}$.

Ces distributions sont définies de manière à avoir l'égalité suivante :

$$\mathbf{p}' = \sum_{(\alpha,\beta)\in A\times B} p_{\alpha,\beta} - (|B| - 1)p_A - (|B| - 1)p_B - (|A||B| - |A| - |B| + 1)p_\omega.$$

On vérifie facilement que $\mathbf{p}'$ est égale à $\mathbf{p}$ sur son support et finalement $\nu(\mathbf{p}) \leq |A||B|(2K_G + 2) - 1$.

□

La preuve du théorème 3.10 se déduit maintenant très simplement.

Démonstration du théorème 3.10. Soit $\mathbf{p} = q^+\mathbf{p}^+ - q^-\mathbf{p}^-$ et $q^+ + q^- = \gamma_2(\mathbf{p})$ et $\mathbf{p}^+, \mathbf{p}^- \in \mathcal{Q}$. D'après le lemme 3.4, on a $\nu(\mathbf{p}) \leq q^+\nu(\mathbf{p}^+) + q^-\nu(\mathbf{p}^-)$. Il suffit enfin d'appliquer le théorème 3.11 pour conclure. □

3.6 Bornes supérieures sur la complexité de la communication

Nous allons maintenant montrer comment utiliser ν et γ_2 pour prouver des bornes supérieures sur la communication. Comme nous l'avons dit, la plupart des résultats antérieurs sur la communication des distributions quantiques sont des résultats de bornes supérieures. Néanmoins, la plupart des protocoles simulent des mesures, soit sur des paires de qubits, soit sur des états maximalement intriqués, donnant des distributions à dont les marginales sont uniformes. Le problème général de la simulation exacte des distributions quantiques est toujours un problème ouvert.

Les résultats que nous présentons dans cette section portent sur les distributions causales générales. En revanche, nous présentons des protocoles pour approximer les distributions et non pour les calculer exactement. Dans le cas des fonctions booléennes, la borne supérieure est identique à celle de Linial et Shraibman [LS08b], mais nous en donnons une nouvelle preuve. Shi et Zhu [SZ08] ont donné un protocole pour approximer les distributions quantiques en utilisant les normes tensorielles. Nous en donnons ici une preuve élémentaire, utilisant les idées que nous avons développées dans cette section.

Les résultats de cette section permettent également de prouver qu'on peut simuler tout protocole de communication dans le modèle simultané, avec une communication exponentielle. Nous étendons ce résultat connu pour les fonctions booléennes au cas des distributions causales.

Nous allons commencer par traiter séparément le cas des fonctions booléennes. Ce cas implique en effet des simplifications qui conduisent à une preuve élémentaire. La borne supérieure utilise la mesure ν^∞, définie page 51.

Théorème 3.12. *Pour toute fonction booléenne $f : X \times Y \to \{-1, 1\}$, et tout $0 < \varepsilon < 1/2$, on a $R^{||}_\varepsilon(f) = O((\nu^\infty(f))^2 \ln \frac{1}{\varepsilon})$.*

Démonstration. Soit C_f la matrice de communication de f, $\lambda = \nu^\infty(f)$ et $\tilde{C}$ une matrice telle que $\nu(C) = \lambda$ et $1 \leq C_f[x, y]\tilde{C}[x, y]$. Par définition, $\lambda \geq 0$, donc les coefficients de C et $\frac{\tilde{C}}{\lambda}$ ont les mêmes signes. Supposons qu'Alice et Bob échantillonnent la distribution correspondant à $\frac{\tilde{C}}{\lambda}$ une seule fois, et envoient leurs réponses à l'arbitre. La réponse de l'arbitre est alors le produit des valeurs envoyées par les joueurs. Soient X et Y les résultats obtenus par Alice et Bob en échantillonnant $\frac{\tilde{C}}{\lambda}$. Supposons $C[x, y] = 1$, l'autre cas étant identique. On a alors par définition $\mathbf{E}[XY] = \frac{\tilde{C}}{\lambda} \geq 1/\lambda$. La probabilité de succès du protocole simultané est alors $\frac{1}{2}(1 + \frac{1}{\lambda})$.

De plus, d'après la proposition 3.3, $\frac{\tilde{C}}{\lambda}$ est locale, et la complexité de ce protocole est donc 2 bits de communication. En utilisant le corollaire du théorème de Hoeffding (corollaire 1.3), on en déduit donc qu'en répétant le protocole $O(\lambda^2 \ln \frac{1}{\varepsilon})$, on a un protocole simultané ε-correct pour f. □

Bien que l'inégalité de Grothendieck permette de déduire du théorème précédent une borne supérieure en $O((\gamma_2^\infty(f))^2)$, nous allons donner une preuve directe de ce résultat. L'avantage de cette preuve est qu'elle ne nécessite aucune connaissance préalable des distributions causales et s'applique directement aux fonctions. Cette preuve est également différente de celle de Linial et Shraibman et utilise des projections aléatoires. Ce protocole s'inspire de celui proposé par Kremer, Nisan et Ron [KNR99] pour calculer le produit scalaire réel de deux vecteurs de norme au plus 1.

Théorème 3.13. *Pour toute fonction $f : X \times Y \to \{-1, 1\}$ et $0 < \varepsilon < 1/2$, on a $R_\varepsilon^{\|}(f) \leq O((\gamma_2^\infty(f))^2)$.*

On utilise pour prouver ce résultat le lemme de Goemans et Williamson [GW95].

Lemme 3.6. *Soient u et v deux vecteurs unitaires dans $\mathbb{R}^d$. Soit r un vecteur aléatoire unitaire dans $\mathbb{R}^d$. On a alors*

$$\text{Prob}(sgn(\langle u, r\rangle) \neq sgn(\langle v, r\rangle)) = \frac{1}{\pi} \arccos\langle u, v\rangle,$$

où $sgn(x) = -1$ si $x < 0$ et 1 sinon.

Démonstration du théorème : Soit C_f la matrice de corrélation de f. On a montré à la section 3.3 que la définition de γ_2 était égale à $j_{\mathcal{Q}}$ dès que $\gamma_2 \geq 1$. Supposons donc que $\gamma_2 \geq 1$. Si ce n'est pas le cas, on a alors $C_f \in \mathcal{Q}$, et comme les coefficients de C_f sont toutes dans $\{-1, 1\}$, on a d'après [CHTW04], $C_f \in \mathcal{L}$. f peut alors être calculée sans communication.

Rappelons la définition de j_Q. On a $j_Q(C_f) = \min\{\lambda \in \mathbb{R} : C_f/\lambda \in \mathcal{Q}\}$. Par conséquent, si on note $\gamma_2^\infty(C_f) = \gamma$, alors, en utilisant la caractérisation de $\mathcal{Q}$ déduite du théorème de Tsirelson (théorème 1.10), on en déduit qu'il existe un entier d deux matrices réelles X et Y de dimension $n \times d$ telles que $X^T \cdot Y = M_f$ et row(X)row(Y) $\leq \gamma$, où row(X) est le maximum parmi les normes euclidiennes des vecteurs formés par les lignes de X.

L'idée du protocole est la suivante. Supposons que la sortie du protocole est 1. Alice reçoit son entrée x, et fixe a comme étant la ligne de X dont l'indice est x. Bob fixe b de la même manière. Par hypothèse, on a donc $sgn(\langle a, b\rangle) = 1$. Soit $a' = \frac{a}{\|a\|_2}$ et $b' = \frac{b}{\|b\|_2}$. Pour connaître le signe de $\langle a, b\rangle$, les joueurs vont faire une projection de leurs vecteurs respectifs sur un vecteur r unitaire aléatoire et envoyer le signe de leur projection à l'arbitre. Celui-ci donnera comme valeur le produit des signes.

Remarquons qu'on a $\langle a', b'\rangle = \frac{1}{\|a\|_2\|b\|_2} \geq \frac{1}{\gamma}$. Par le lemme 3.6, la probabilité d'erreur de ce protocole est alors :

$$\begin{aligned}
\text{Prob}(sgn(\langle a', r\rangle)sgn(\langle b', r\rangle) = -1) &= \text{Prob}(sgn(\langle a', r\rangle) \neq sgn(\langle b', r\rangle)) \\
&= \frac{1}{\pi} \arccos\langle a', b'\rangle \\
&= \frac{1}{2} - \frac{1}{\pi} \arcsin\langle a', b'\rangle \\
&\leq \frac{1}{2} - \frac{1}{\pi}\langle a', b'\rangle \\
&\leq \frac{1}{2} - \frac{1}{\pi\gamma}
\end{aligned}$$

Ainsi, avec un bit envoyé par les joueurs à l'arbitre, ils obtiennent par rapport à une sortie aléatoire, un avantage de $\frac{1}{\pi\gamma}$. On repète ensuite le protocole k fois afin d'atteindre une probabilité de succès plus grande. Comme pour le théorème 3.12, on utilise le corollaire du théorème de Hoeffding (corollaire 1.3) et on en déduit qu'il faut répéter ce protocole $O(\gamma^2 \ln \frac{1}{\varepsilon})$ fois pour atteindre une probabilité de succès $1-\varepsilon$. □

Pour le cas général, on utilise la décomposition affine d'une distribution en distributions locales. Alice, Bob et l'arbitre peuvent se mettre d'accord sur une décomposition, et envoyer des échantillons des distributions locales à l'arbitre. Celui-ci, connaissant la décomposition construit une approximation de la distribution complète.

Théorème 3.14. *Soit* $\mathbf{p}$ *une distribution causale définie sur* $X \times Y$*, à valeur dans* $A \times B$*. Fixons* $\varepsilon, \delta > 0$*. On a alors*

- $R^{||}_{\varepsilon+\delta}(\mathbf{p}) \leq 16[\frac{|A||B|\nu^{\varepsilon}(\mathbf{p})}{\delta}]$,
- $Q^{||}_{\varepsilon+\delta}(\mathbf{p}) = O\left((|A||B|)^5[\frac{\nu^{\varepsilon}(\mathbf{p})}{\delta}]^4 \ln\left(\frac{|A||B|}{\delta}\right) \log n\right)$.

L'idée pour prouver ce théorème est de simuler les distributions qui apparaissent dans les décompositions affines locales et quantiques de la distribution initiale. L'essentiel de la preuve du théorème est de décrire les processus aléatoires utilisés. Avant cela, le lemme suivant montre comment utiliser des échantillons pour construire une approximation d'une distribution de probabilité.

L'idée du lemme 3.7 est la suivante : on cherche à simuler une distribution de probabilité. Pour ce faire, on a un processus probabiliste qui permet d'approximer la distribution avec une bonne précision. Néanmoins, on n'a pas la garantie que les estimateurs forment eux-mêmes une distribution. On doit donc rééchelonner ceux-ci pour qu'ils en forment une. La distribution $\mathbf{p}'$ est alors construite de la manière suivante :

1. on a obtenu une famille d'estimateurs $\{Q_v\}_{v \in V}$ dont on veut se servir pour approximer la distribution $\mathbf{p}$,
2. ceux-ci ne sont pas tous nécessairement positifs, alors on les remplace par $\max\{0, Q_v\}$,
3. la somme n'est pas nécessairement égale à 1, alors on les renormalise de la manière suivante :
   (a) si la somme des estimateurs est plus grande que 1, on divise par la somme,
   (b) si la somme des estimateurs est plus petite que 1, on répartit uniformément le poids manquant sur les probabilités construites.

La distinction entre les cas 3a et 3b vient du fait que l'inégalité de Markov permet de borner la déviation supérieure d'une distribution et pas directement sa déviation inférieure.

Lemme 3.7. *Soient* $\mathbf{p}$ *une distribution de probabilité sur un ensemble* V *et* $e : \mathbb{R}^+ \to \mathbb{R}^+$ *une fonction. Supposons qu'on a une famille de variables aléatoires réelles* $\{Q_v\}_{v \in V}$ *telle que pour tout* $v \in V$ *et tout* $\beta > 0$*,* $\text{Prob}[|Q_v - p(v)| \geq \beta] \leq e(\beta)$*. Notons* $R_v = \max\{0, Q_v\}$ *et pour tout* $U \subseteq V$*,* $R_U = \sum_{u \in U} R_u$*. Alors pour tout* $\beta > 0$*, la distribution* $\mathbf{p}'$ *définie sur* V *par* $p'(v) = \mathbf{E}[S_v]$ *et*

$$S_v = \begin{cases} R_v / R_V & \text{si } R_V \geq 1, \\ R_v + \frac{1}{|V|}(1 - R_V) & \text{sinon,} \end{cases}$$

est bien définie et vérifie $\delta(\mathbf{p}', \mathbf{p}) \leq 2|V|(\beta + e(\beta))$.

Démonstration. Il est facile de voir que par construction, $\mathbf{p}'$ ainsi définie est toujours une distribution de probabilité. On a en effet bien $S_v \geq 0$ et $\sum_v S_v = 1$ et donc a fortiori $p'(v) \geq 0$ et $\sum_v p'(v) = 1$.

Soit pour tout $U \subseteq V$, $R_U = \sum_{u \in U} R_u$. Comme $p(v) \geq 0$, on a $|R_v - p(v)| \leq |Q_v - p(v)|$ et donc

$$\text{Prob}[|R_v - p(v)| \geq \beta] \leq e(\beta).$$

De plus, si $U \subseteq V$, on a :

$$\begin{aligned} \text{Prob}[R_U \geq p(U) + |U|\beta] &= \text{Prob}[\sum_{u \in U}(R_U - p(U)) \leq |U|\beta] \\ &\leq \text{Prob}[\exists u : (R_u - p(U)) \geq \beta] \\ &\leq |U|e(\beta) \end{aligned}$$

En bornant de la même manière la déviation inférieure, on en déduit $\text{Prob}[|R_U - p(U)| \geq |U|\beta] \leq |U|e(\beta)$. On veut maintenant montrer que tout $U \subseteq V$, $|\mathbf{E}[S_v] - p(v)| \leq 2|V|(\beta + e(\beta))$.

Montrons d'abord une borne supérieure sur $\mathbf{E}[S_U]$. Supposons que $R_V \geq 1$. On a dans ce cas $S_U \leq R_U$. Par conséquent, $\text{Prob}[S_U \geq p(U) + |U|\beta] \leq \text{Prob}[R_U \geq p(U) + |U|\beta] \leq |U|e(\beta)$. Comme par définition $S_U \leq 1$, on a :

$$\begin{aligned} \mathbf{E}[S_U] &\leq (p(U) + |U|\beta)\text{Prob}[S_U \leq p(U) + |U|\beta] + 1.\text{Prob}[S_U \geq p(U) + |U|\beta] \\ &\leq p(U) + |U|(\beta + e(\beta)). \end{aligned}$$

Suppsons maintenant que $R_V \leq 1$. On a alors $S_U = R_U + \frac{|U|}{|V|}(1 - R_V)$. On va borner supérieurement chaque partie de la somme. Puisque $R_V \leq 1$, on peut appliquer le même raisonnement que précédemment pour S_U, et on en déduit $\mathbf{E}[R_U] \leq p(U) + |U|(\beta + e(\beta))$.

Par l'inégalité de Markov, on a pour tout $\alpha \geq 0$, $\mathbf{E}[R_V] \geq \alpha.\text{Prob}[R_v \geq \alpha]$. Or on a vu que $\text{Prob}[R_V \geq 1 - |V|\beta] \geq 1 - |V|e(\beta)$. On a donc $\mathbf{E}[R_V] \geq (1 - |V|\beta)(1 - |V|e(\beta)) \geq 1 - |V|[\beta + e(\beta)]$. En combinant les deux parties, on en déduit bien $\mathbf{E}[S_U] \leq p(U) + 2|U|(\beta + e(\beta))$.

Pour la borne inférieure, on va utiliser à nouveau l'inégalité de Markov. Supposons cette fois ci en premier que $R_V \leq 1$. Dans ce cas, on a $S_U \geq R_U$ et par conséquent $\mathbf{E}[S_U] \geq \mathbf{E}[R_U]$, et en utilisant l'inégalité de Markov comme précédemment, $\mathbf{E}[R_U] \geq p(U) - |U|[\beta + e(\beta)]$. On a donc bien $\mathbf{E}[S_U] \geq p(U) - |V|[\beta + e(\beta)]$.

Supposons maintenant que $R_V \geq 1$. On a alors d'un côté $\text{Prob}[R_V \leq 1 + |V|\beta] \geq 1 - |V|e(\beta)$, et de l'autre $\text{Prob}[R_U \geq p(U) - |U|\beta] \leq 1 - |U|e(\beta)$. Ainsi, ces deux événements arrivent simultanément avec une probabilité supérieure à $1 - (|U| + |V|)e(\beta)$. Enfin, comme $\frac{p(U) - |U|\beta}{1 - |V|\beta} \geq (p(U) - |U|\beta)(1 - |V|\beta) \geq p(u) - (|U| + |V|)\beta$, on en déduit

$$\begin{aligned} \text{Prob}[S_U \geq p(u) - (|U| + |V|)\beta] &\geq (1 - |V|e(\beta))(1 - |U|e(\beta)) \\ &\geq 1 - (|U| + |V|)e(\beta). \end{aligned}$$

Finalement, une dernière application de l'inégalité de Markov donne :

$$\begin{aligned} \mathbf{E}[S_v] &\geq (1 - (|U| + |V|)e(\beta))(p(v) + (|U| + \ |V|)\beta) \\ &\geq p(v) - (|U| + |V|)(\beta + e(\beta)). \end{aligned}$$

En combinant les deux inégalités, on a bien $|\mathbf{E}[S_v] - p(v)| \leq 2|V|(\beta + e(\beta))$. □

Démonstration du théorème. Soit $\tilde{\mathbf{p}}$ une distribution telle que $\delta(\mathbf{p}, \tilde{\mathbf{p}}) \leq \varepsilon$ et $\nu(\tilde{\mathbf{p}}) = \nu^{\varepsilon}(\mathbf{p})$. On note $\nu = \nu(\tilde{\mathbf{p}})$. Soit $\tilde{\mathbf{p}} = q_+\mathbf{p}^+ - q_-\mathbf{p}^-$, avec $\mathbf{p}^+, \mathbf{p}^- \in \mathcal{L}$, $q_+ + q_- = \nu$ et $q_+, q_- \geq 0$. Fixons deux protocoles locaux $\mathcal{P}^+$ et $\mathcal{P}^-$ pour $\mathbf{p}^+$ et $\mathbf{p}^-$. On fixe également les entrées (x, y). Les joueurs vont répéter plusieurs fois les protocoles locaux pour approcher $\mathbf{p}^+$ et $\mathbf{p}^-$. On introduit donc des variables aléatoires pour noter les sorties de ces protocoles.

Soit k le nombre d'itérations des protocoles $\mathcal{P}^+$ et $\mathcal{P}^-$. Pour $i = 1, \ldots k$, et $(a, b) \in A \times B$, soit $P^+_{i,a,b}$ la variable aléatoire qui vaut 1 si la sortie de la i-ème itération de $\mathcal{P}^+$ est (a, b) et 0 sinon. On définit $P^-_{i,a,b}$ de façon analogue, et $P_{i,a,b} = q^+P^+_{i,a,b} - q^-P^-_{i,a,b}$. L'arbitre va approximer $p(a, b|x, y)$ par la variable aléatoire $P_{a,b} = \frac{1}{k}\sum_i P_{i,a,b}$.

Vérifions que cette famille de variables aléatoire satisfait bien les hypothèses du lemme 3.7. On a $\mathbf{E}[P_{i,a,b}] = p(a, b|x, y)$ et $P_{i,a,b} \in [-q_-, q_+]$. En appliquant le théorème de Hoeffding (théorème 1.12), on obtient directement $\text{Prob}[\ |P_{a,b} - p(a, b|x, y)| \ \geq \beta] \leq 2e^{-\frac{2k\beta^2}{\nu^2}}$. On applique donc le lemme avec $V = A \times B$, $Q_v = P_{a,b}$ et $e(\beta) = 2e^{-\frac{2k\beta^2}{\nu^2}}$. De plus, on a par construction $\sum_{a,b} P^+_{a,b} = 1$ et $\sum_{a,b} P^-_{a,b} = 1$, et donc $\sum_{a,b} P_{a,b} = 1$. Par conséquent, $\text{Prob}[\sum_{a,b} P_{a,b} = 1] = 1$.

L'arbitre peut donc utiliser ces estimateurs pour échantillonner selon une distribution de probabilité $\mathbf{p}'$ telle que $\delta(\tilde{\mathbf{p}}, \mathbf{p}') \leq 2|A||B|(\beta + 2e^{-\frac{2k\beta^2}{\nu^2}})$. En choisissant $\beta = \frac{\delta}{4|A||B|}$ et $k = 8\left(\frac{|A||B|\nu}{\delta}\right)^2 \ln\left(\frac{8|A||B|}{\delta}\right)$, on obtient $\delta(\mathbf{p}, \mathbf{p}') \leq \varepsilon + \delta$ comme annoncé.

Dans le second point, l'aléa partagé n'est plus disponible, mais les messages sont quantiques. L'idée est alors que les joueurs se mettent d'accord sur les chaînes aléatoires par avance et utilisent la technique l'empreinte quantique [BCWdW01] pour encoder les échantillons correspondants. En effet, d'après le théorème de Newman (théorème 1.1, présenté au chapitre 1), on peut en augmentant l'erreur d'une constante δ, restreindre le nombre de chaînes aléatoires différentes du protocole à n. Bien que ce théorème soit énoncé pour des fonctions booléennes, on peut l'appliquer également pour des fonctions. Cette technique est inspirée de la technique de Yao pour simuler la communication classique avec l'aléa partagé dans le modèle quantique simultané sans ressource partagée [Yao03].

Soit $\mathcal{P}^{+,r}$ le protocole consistant à exécuter $\mathcal{P}^+$ en utilisant la chaîne aléatoire r et $P^{+,r}_{a,b}$ la variable indicatrice de la sortie de celui-ci. Soit R l'ensemble des chaînes aléatoires utilisées par $\mathcal{P}^+$. On suppose, quitte à agrandir l'ensemble que $\mathcal{P}^-$ utilise le même ensemble. En appliquant le théorème de Newman, on en déduit qu'il existe un ensemble $\tilde{R} \subseteq R$ tel que $|\tilde{R}| \leq \frac{4n}{\alpha^2}$ et pour tout $U \subseteq A \times B$, $|\mathbf{E}_{\tilde{R}}[P^{+,r}_U] - \mathbf{E}[P^{+,r}_U]| \leq \alpha$. Si on note $\tilde{\mathbf{p}}^+$ la distribution produite en choisissant une chaîne dans $r \in \tilde{R}$ et en exécutant $\mathcal{P}^{+,r}$. On a prouvé que cette distribution vérifie $\delta(\tilde{\mathbf{p}}^+, \mathbf{p}^+) \leq \alpha$.

Soit $L = |U|$. On note par ailleurs $A^{+,r}_a$ et $B^{+,r}_b$ les fonctions indicatrices des sorties du protocole $\mathcal{P}^+$. Le protocole simultané fonctionne de la manière suivante. Alice prépare un état $|\varphi^+_a\rangle = \frac{1}{\sqrt{L}}\sum_{r_i \in \tilde{R}} |A^{+,r_i}_a\rangle|1\rangle|i\rangle$, et Bob, de même prépare $|\varphi^+_b\rangle = \frac{1}{\sqrt{L}}\sum_{r_i \in \tilde{R}} |B^{+,r_i}_b\rangle|1\rangle|i\rangle$. Ils envoient ensuite k copies de cet état à l'arbitre.

Calculons le produit de ces deux états :

$$\begin{aligned}
\langle\varphi^+_a|\varphi^+_b\rangle &= \frac{1}{L}\sum_{r\in\tilde{R}}\langle A^{+,r}_a|1\rangle\langle 1|B^{+,r}_b\rangle \\
&= \mathbf{E}_{\tilde{R}} A^{+,r}_a B^{+,r}_b \\
&= \tilde{p}^+(a, b|x, y).
\end{aligned}$$

L'arbitre utilise la technique suivante pour estimer le produit scalaire $\langle\varphi_a^+|\varphi_b^+\rangle$. Après avoir reçu les états $|\varphi_a^+\rangle$ et $|\varphi_b^+\rangle$, il exécute la transformation suivante :

$$(H\otimes I)(\text{c}-\text{SWAP})(H\otimes I)|0\rangle|\varphi_a^+\rangle\ |\varphi_b^+\rangle.$$

Soit $Z_{i,a,b}^+$ la variable aléatoire qui représente le résultat de la mesure dans la base $\{|0\rangle,|1\rangle\}$. D'après [BCWdW01], cette variable vérifie $\text{Prob}(Z_{i,a,b}^+=1)=\frac{1-\langle\varphi_a^+|\varphi_b^+\rangle^2}{2}$.

Posons $Z_{a,b}^+=\frac{1}{k}\sum_{i=1}^t Z_{i,a,b}^+$. L'estimateur de l'arbitre est alors

$$Q_{a,b}^+=\begin{cases}\sqrt{1-2Z_{a,b}^+} & \text{si } Z_{a,b}^+\leq 1/2,\\ 0 & \text{sinon.}\end{cases}$$

On définit $Q_{a,b}^-$ de la même manière, et $Q_{a,b}=q^+Q_{a,b}^+ + q^-Q_{a,b}^-$. Par l'inégalité de Hoeffding, présentée au théorème 1.12, on a

$$\text{Prob}[|Q_{a,b}-\tilde{p}(a,b|x,y)|\ \geq\beta]\leq 2e^{\frac{T\beta^4}{2\nu^4}}$$

De plus, on a par construction $\sum_{a,b}Q_{a,b}^+=\sum_{a,b}Q_{a,b}^-=1$ et par conséquent $\sum_{a,b}Q_{a,b}=1$. On conclut, comme dans le cas précédent, en utilisant le lemme 3.7.

□

En utilisant ce théorème, on en déduit comment simuler la communication quantique dans le modèle de communication simultané. De plus, en utilisant les relations entre γ_2 et ν montrées dans la section 3.5, on peut tout exprimer en fonction de $Q_\varepsilon^{\text{ent}}$, la complexité du modèle le plus fort.

Corollaire 3.3. *Soit* $\mathbf{p}$ *une distribution définie sur* $X\times Y$ *et à valeurs dans* $A\times B$. *Supposons* $|X\times Y|\leq 2^{2n}$ *et pour* $\varepsilon<1/2$, $Q_\varepsilon^{\text{ent}}(\mathbf{p})\leq q$. *On a alors pour tout* $\delta<1/2$

- $R_{\varepsilon+\delta}^{||,\text{pub}}(\mathbf{p})\leq O(2^{4q\frac{(|A|\ |B|)^3}{\delta^2}}\ln^2\frac{|A||B|}{\delta})$,
- $Q_{\varepsilon+\delta}^{||}(\mathbf{p})\leq(2^{8q\frac{(|A||B|)^9}{\delta^4}}\ln\frac{|A||B|}{\delta}\log n)$.

Pour les fonctions booléennes, les résultats que nous avons établi sont plus précis. Les simulations correspondantes sont par conséquent plus fines.

Corollaire 3.4. *Soit* $f:\{0,1\}^n\times\{0,1\}^n\to\{0,1\}$. *Soit* $Q_\varepsilon^{\text{ent}}(f)\leq q$, *alors*

- $R_\delta^{||,\text{pub}}(f)\leq K_G^2.2^{2q+2}\ln(\frac{1}{\delta})\frac{1}{(1-2\varepsilon)^2}$,
- $Q_\delta^{||}(f)\leq O(\log n 2^{4q}\ln(\frac{1}{\delta})\frac{1}{(1-2\varepsilon)^4})$.

Dans cette section, nous avons montré comment utiliser les mesures ν et γ_2 pour prouver des bornes supérieures sur le coût en communication de la simulation approchée des distributions causales. Le modèle des distributions causales englobe en particulier la simulation quantique et les fonctions booléennes. La preuve de notre résultat général s'applique donc à la fois aux distributions quantiques et aux fonctions booléennes. La table 3.1 rappelle les résultats similaires précédemment connus pour ces cas particuliers.

Dans le cas des distributions quantiques, notre résultat implique que l'approximation peut être faite avec une communication constante, quel que soit le nombre de mesures possibles. Shi et Zhu avaient prouvé ce résultat en utilisant les normes tensorielles [SZ08]. Nous en avons donné une preuve alternative basée sur la géométrie des distributions quantiques. Dans le cas des fonctions booléennes, notre borne permet de retrouver les résultats de Shi et Zhu [SZ08], ainsi que ceux de Gavinsky, Kempe et de Wolf [GKdW06].

Auteurs	Distributions quantiques	Fonction booléennes
Shi, Zhu [SZ08]	$R_\varepsilon^{\parallel,\mathrm{pub}}(\mathbf{p}) = O(\ln \frac{1}{\varepsilon}/\varepsilon^2)$	$R_\varepsilon^{\parallel,\mathrm{pub}}(f) = \exp(O(Q_\varepsilon^{\mathrm{ent}}(f))$
Yao [Yao03]	-	$Q_\varepsilon^{\parallel}(f) = 2^{O(R_\varepsilon^{\parallel,\mathrm{pub}}(f))} \ln n$
Gavinsky, Kempe, de Wolf [GKdW06]	-	$Q_\varepsilon^{\parallel}(f) = 2^{O(Q_\varepsilon^{\mathrm{ent}}(f))} \log n$
Linial, Shraibman [LS08b]	-	$R_\varepsilon^{\parallel,\mathrm{pub}}(f) = O((\gamma_2^\infty(f))^2)$

TAB. 3.1 – Simulation des distributions et des protocoles dans le modèle simultané

3.7 Conclusion

Dans ce chapitre, nous avons étudié la simulation des distributions causales. Ceci nous a permis d'étudier dans un même modèle la simulation des distributions quantiques et la complexité de la communication des fonctions booléennes. L'avantage principal de cette approche est de donner de nouvelles preuves, dont l'intuition est basée sur la géométrie des distributions causales.

Nous avons donné des bornes inférieures sur la simulation exacte et approchée. Dans le cas général, ces bornes sont basées sur la décomposition d'une distribution en combinaison affine de distributions locales ou quantiques. Bien que n'étant pas une application directe du théorème de Randal et Foulis qui caractérise l'ensemble des distributions causales comme la clôture affine de l'ensemble des distributions locales [FR81], nous pensons que notre approche donne encore plus d'importance à ce résultat.

Les expressions duales de nos bornes ont également une interprétation intéressante en termes de violation d'inégalités de Bell et de Tsirelson. Dans le cas particulier des fonctions booléennes, nos bornes inférieures sont équivalentes à celles de Linial et Shraibman.

L'inégalité de Grothendieck joue un rôle important en informatique. Elle a été appliquée à des problèmes variés, notamment en informatique quantique. Celle-ci dit quelque chose de précis sur la structure géométrique des distributions causales, dans le cas particulier des distributions binaires à marginales uniformes. Nous avons étendu cette inégalité au cas général en donnant une relation entre γ_2 et ν dans le cas général. Ceci donne une indication de la distance maximale entre les ensembles $\mathcal{Q}$ et $\mathcal{L}$.

Enfin, nous avons donné des bornes supérieures dans ce modèle. En étudiant le cas général des distributions causales, nous avons des bornes qui s'appliquent à la fois à la simulation des distributions quantiques et au calcul des fonctions booléennes. Dans le cas des distributions causales à marginales uniformes, nous avons donné des nouvelles preuves des bornes supérieures en ν^2 et γ_2^2. Appliquées aux fonctions booléennes, cela permet de retrouver la borne supérieure de Linial et Shraibman.

Dans le cas général, nous avons également montré une borne générale en ν^2 sur le problème avec erreur, dans le modèle classique simultané avec aléa partagé, et dans le modèle quantique sans simultané, sans ressource partagée. En combinant cela avec la relation que nous avons prouvée entre ν et γ_2 et avec les bornes inférieures, cela montre comment simuler la communication quantique avec intrication dans les modèles sans interaction.

Chapitre 4

Complexité en boîtes non-locales

Dans cette section, nous étudions la complexité en boîte non-locale des fonctions booléennes. Dans ce modèle de calcul, deux joueurs reçoivent chacun une entrée et doivent calculer une fonction en parité en utilisant des boîtes non-locales mais pas de communication. Dans la section 4.1, nous définissons formellement le modèle de calcul et ses différentes variantes. Dans la section 4.2, nous étudions la complexité déterministe. Ceci nous permet d'analyser en détail l'algorithme de van Dam pour calculer toute fonction sans communication. Dans la section 4.3, nous étudions la complexité probabiliste. Ceci nous permet de donner des résultats plus précis sur les liens avec la communication simultanée. Dans la section 4.4, nous appliquons plus spécifiquement nos résultats à la simulation des corrélations quantiques et au delà. Enfin, dans la section 4.5, nous montrons les applications de nos résultats à l'évaluation sécurisée.

4.1 Introduction et définitions

4.1.1 boîtes non-locales

Dans ce chapitre, nous allons étudier une ressource appelée boîte non-locale. Celle-ci a été introduite par Popescu et Rohrlich en 1994 [PR94] pour étudier la non-localité. Même si on suppose que les boîtes non-locales n'existent pas dans la réalité, leur définition montre que d'un point de vue logique, la causalité est insuffisante pour caractériser la physique quantique. En effet, nous allons voir que la définition des boîtes non-locales implique qu'elles violent une inégalité de Tsirelson. Elles ne peuvent donc pas être reproduites par la physique quantique. Pourtant, elles respectent la causalité et d'un point de vue mathématique, la théorie des boîtes non-locales est parfaitement valide.

Dans les expériences de type EPR, deux joueurs reçoivent des entrées en fonction desquelles ils choisissent de faire une mesure. Les boîtes non-locales sont une abstraction de cette situation. Les entrées représentent le choix de la mesure et les sorties, les résultats de celles-ci.

La première raison qui nous amène à considérer les boîtes non-locales comme une ressource de calcul vient de la physique. Pour simuler les mesures sur des états maximalement intriqués, on utilise en général de la communication. Le principal défaut de cette approche est qu'on utilise pour quantifier la non-localité quantique, un processus de simulation qui n'est pas causal. Il est en effet évident qu'à l'issue d'un protocole de

communication, les joueurs ont reçu des informations sur la question posée à l'autre. Cela n'enlève pas toute pertinence à la communication comme échelle de mesure, mais ce point nécessite une attention particulière.

Si on veut mesurer la complexité d'un processus causal, il semble plus naturel de vouloir utiliser un autre processus causal comme unité de référence. L'idée est ici d'utiliser les boîtes non-locales comme unité de mesure de la non-localité [CGMP05]. On verra plus loin qu'on peut utiliser un bit de communication pour simuler une boîte non-locale. Cette ressource est donc plus faible que la communication.

Définition 4.1. *Soit $\mathbf{p}_{NL}$ la distribution définie sur $X \times Y$ et à valeurs dans $A \times B$, avec $X = Y = A = B = \{0, 1\}$ et définie par :*

$$p_{NL}(a, b|x, y) = \begin{cases} 1/2 & si\ a \oplus b = x \wedge y \\ 0 & sinon \end{cases}$$

Une boîte non-locale est une ressource qui prend en entrée des variables booléennes x et y et rend en sortie deux variables aléatoires booléennes a et b suivant la distribution $\mathbf{p}_{NL}$.

En sommant, on trouve immédiatement que pour tout x, y, a', b', $\sum_b p(a', b|x, y) = \sum_a p(a, b'|x, y) = 1/2$. Par conséquent, la distribution $\mathbf{p}_{NL}$ est bien causale.

On a défini, au chapitre 3, le jeu CHSH. Rappelons comment celui-ci est défini. Deux joueurs, Alice et Bob, reçoivent respectivement les entrées $x \in \{0, 1\}$ et $y \in \{0, 1\}$, suivant la distribution uniforme. Ils doivent répondre par deux sorties a et b dans $\{0, 1\}$ et gagent si $a \oplus b = x \wedge y$. On voit que par définition, si Alice et Bob partagent une boîte non-locale, ils peuvent gagner avec probabilité 1 le jeu CHSH quelle que soit la distribution sur les entrées. Pour ce jeu, Tsirelson a montré que la valeur quantique de 0,85 était optimale. Les boîtes non-locales permettent de gagner plus souvent que toutes les ressources quantiques. Enfin, d'après le lemme 3.3, page 56, il y a une bijection entre les jeux XOR et inégalités de Bell et de Tsirelson. Ainsi, le fait de gagner avec une probabilité supérieure à 0,85 au jeu CHSH en utilisant une boîte non-locale signifie que la distribution $\mathbf{p}_{NL}$ viole une inégalité de Tsirelson.

On peut introduire les boîtes non-locales d'une autre manière, basée sur la géométrie des distributions causales. Considérons l'ensemble $\mathcal{C}_{2,2}$ des distributions causales définies sur $\{0, 1\} \times \{0, 1\}$ et à valeurs dans $\{0, 1\} \times \{0, 1\}$. Comme on l'a fait dans le cas général, on peut plonger cet espace dans $\mathbb{R}^{|A \times B \times X \times Y|} = \mathbb{R}^{16}$ en identifiant chaque distribution à un vecteur. Comme dans le cas général, l'ensemble $\mathcal{C}_{2,2}$ est un polytope. Tout élément de cet ensemble peut s'écrire comme combinaison convexe de ses points extrémaux.

La boîte non-locale telle que nous l'avons définie est un point extrémal de ce polytope [BLM$^+$05]. Les autres points extrémaux sont soit des distributions locales, soit des distributions qui peuvent être simulées en composant la distribution $\mathbf{p}_{NL}$ avec des transformations locales et sans communication [CGMP05]. En utilisant des boîtes non-locales, on peut donc simuler toutes les distributions causales à deux entrées et deux sorties.

D'après sa définition, une boîte non-locale est une ressource qui respecte la causalité. Existe-t'il un principe physique que les boîtes non-locales ne vérifient pas, et qui par conséquent les empêche d'exister en réalité ? Van Dam a proposé une réponse à cette question. Il ne s'agit pas d'un principe physique, mais d'un principe informatique. Van Dam a en effet montré que l'utilisation de boîtes non-locales permettait de rendre tout

problème de communication trivial [vD05]. D'après lui, cette conséquence semble impliquer que le modèle de complexité en boîte non-locale n'est pas réaliste, et par conséquent que les boîtes non-locales n'existent pas.

Ce que nous voulons vérifier en étudiant la complexité en boîte non-locale, c'est que la conséquence improbable soulignée par van Dam vient bien du modèle de calcul et pas de l'algorithme utilisé. En effet, si le fait de pouvoir résoudre tout problème de communication avec un seul bit peut paraître improbable pour un informaticien, l'idée de voir se terminer un calcul qui consomme une quantité exponentielle de ressource paraît tout aussi improbable à un théoricien de la complexité.

Enfin, en calculant avec des boîtes non-locales et pas de communication, les joueurs ne reçoivent que des bits aléatoires uniformément distribués. Les bits d'Alice et de Bob sont corrélés mais sans communication, ils ne peuvent rien apprendre de la fonction tout seul. Dans notre modèle de calcul, la fonction est calculée en parité entre les joueurs. En particulier, lorsqu'on applique le protocole de van Dam, la réponse des joueurs est la parité des sorties de toutes les boîtes non-locales, qui sont des bits uniformément distribués. De sorte que même si Alice envoie à Bob sa réponse, Bob est capable de donner la valeur de la fonction, mais n'apprend aucune autre information sur l'entrée d'Alice. Ceci correspond à une tâche cryptographique qu'on appelle l'évaluation sécurisée. Nous avons là une troisième motivation pour étudier la complexité en boîte non-locale, qui est d'étudier l'évaluation sécurisée.

4.1.2 Modèles de complexité

Le modèle de complexité en boîte non-locale est similaire à celui de la simulation des distributions causales vu au chapitre précédent, mais sans communication. Les deux joueurs reçoivent des questions $x \in X$ et $y \in Y$, où X et Y sont des ensembles finis. Il utilisent ensuite des boîtes non-locales. A la fin, chaque joueurs doit répondre par un bit, a pour Alice et b pour Bob, telle que $a \oplus b = f(x, y)$, où f est une fonction fixée à l'avance. On verra par la suite que ce modèle permet bien de calculer toute fonction booléenne.

Comme dans le modèle de complexité de la communication, les joueurs utilisent les boîtes non-locales en suivant un protocole fixé à l'avance. Pour chaque boîte, on note le couple d'entrées de manière ordonnée. Par exemple, dire que l'entrée est (x, y) signifie qu'Alice a entré dans la boîte la valeur x et Bob la valeur y. De même, les sorties sont ordonnées, par exemple une sortie (a, b) sous-entend qu'Alice a reçu la valeur a et Bob la valeur b.

On rappelle que les sorties sont des variables aléatoires distribuées suivant la distribution $\mathbf{p}_{NL}$ conditionnée par les entrées. Pour simplifier l'écriture, si (a, b) est la sortie d'une boîte non-locale dont l'entrée est (x, y), on note $a \oplus b = x \wedge y$ pour signifier que cet l'événement est vrai avec probabilité 1.

Un protocole en boîte non-locale déterministe est défini par :

– un ensemble de fonctions décrivant les entrées de chaque boîte ; ces fonctions dépendent soit de l'entrée d'Alice, soit de celle de Bob. Par exemple, si le protocole utilise k boîtes non-locales, on pourra noter $p_1, \dots p_k$ les entrées d'Alice et $q_1, \dots q_k$ les entrées de Bob.

– Deux fonctions indiquant les réponses des joueurs en fonction des données en leur possession.

On note $\mathcal{P}(x, y) = (a, b)$ si en suivant le protocole $\mathcal{P}$, les joueurs répondent (a, b) lorsqu'ils ont reçu les questions (x, y). Même si les joueurs n'ont pas d'aléa, les sorties a et b sont des variables aléatoires. Cela vient de l'aléa intrinsèque des boîtes non-locales. Dans le cas déterministe, on impose que la fonction est calculée en parité sans erreur.

Définition 4.2. *Soit $f : X \times Y \to \{0, 1\}$, on note $NL(f)$ le plus petit t tel qu'il existe un protocole $\mathcal{P}$ utilisant t boîtes non-locales et tel que $\mathcal{P}(x, y) = (a, b)$ avec $a \oplus b = f(x, y)$ pour tout x, y.*

On définit également la complexité en boîte non-locale probabiliste. Dans un protocole en boîte non-locale probabiliste, les joueurs peuvent utiliser de l'aléa partagé en plus des boîtes non-locales. La définition suivante définit la correction pour un protocole probabiliste.

Définition 4.3. *Soit $f : X \times Y \to \{0, 1\}$, on note $NL_\varepsilon(f)$ le plus petit t tel qu'il existe un protocole probabiliste $\mathcal{P}$ utilisant t boîtes non-locales, et tel que* $\mathrm{Prob}[\mathcal{P}(x, y) = (a, b)$ *et* $a \oplus b = f(x, y)] \geq 1 - \varepsilon$.

Notons que dans la définition de la complexité probabiliste, la probabilité porte sur l'aléa du protocole mais également sur l'aléa intrinsèque des boîtes non-locales. En effet, étant donnée la définition des boîtes non-locales, on peut les utiliser pour produire de l'aléa. Du point de vue de la complexité, cet aléa n'est pas très avantageux étant donné qu'il est payant alors que l'aléa partagé est en général gratuit. Néanmoins, les joueurs pourraient utiliser des sorties des boîtes à la fois comme aléa et pour autre chose, et nous ne savons pas montrer qu'on peut empêcher une telle utilisation des boîtes non-locales.

On introduit ensuite certaines contraintes quant à l'utilisation des boîtes non-locales. On considerera par la suite les variations suivantes :

– $NL^{||}, NL_\varepsilon^{||}$ lorsque les boîtes sont utilisées simultanément, c'est-à-dire lorsque l'entrée d'une boîte ne dépend jamais de la sortie d'une autre.
– NL^g, NL_ε^g lorsque les joueurs sont contraints de répondre par une fonction g des sorties des boîtes non-locale. Si on note $a_1, \ldots a_t$ les sorties des boîtes du côté d'Alice, sa réponse sera donc $g(a_1, \ldots a_t)$.
– $NL^{\mathrm{ord}}, NL_\varepsilon^{\mathrm{ord}}$ lorsque les joueurs sont contraints à utiliser les boîtes non-locales dans le même ordre.

La définition d'une boîte non-locale est inspirée de l'expérience EPR et du jeu CHSH. Or, quand un joueur mesure un qubit d'une paire EPR, il obtient une réponse immédiatement, et sans attendre que l'autre joueur ne fasse sa mesure. De la même manière, on peut définir une boîte non-locale comme donnant sa réponse à un joueur sans attendre l'autre. En imposant que les deux joueurs utilisent les boîtes dans le même ordre, on peut supposer de plus que les joueurs agissent de manière simultanée. Cette question prendra toute son importance lorsqu'on abordera l'évaluation sécurisée. Dans ce cas, les joueurs ne se font pas confiance et Buhrman *et al.* ont montré qu'on pouvait utiliser ces propriétés temporelles des boîtes non-locales pour gagner de l'information [BCU$^+$06]. On ne sait pas dans quelle mesure cette hypothèse restreint le modèle général du point de vue de la complexité.

On peut également combiner les exposants $\|$ et "g" ou "ord" et "g". les exposants $\|$ et "ord" sont en revanche incompatibles, le premier étant un cas particulier du second. En effet, lorsque les joueurs utilisent toutes les boîtes simultanément, ils les utilisent dans le même sens.

4.2 Complexité déterministe

4.2.1 Rang d'une matrice sur un corps fini

L'analyse du protocole de van Dam ci-dessous va nous permettre de caractériser la complexité en boîte non-locale parallèle. Cette caractérisation utilise le rang de la matrice de communication. Avant de rentrer dans le sujet, nous commençons donc par rappeler quelques définitions et propositions sur celui-ci.

Définition 4.4. *Soit M une matrice à coefficients dans un corps $\mathbb{F}$ et $\mathbb{G}$ un corps contenant $\mathbb{F}$. Le rang de la matrice sur $\mathbb{G}$ est le plus petit nombre de lignes ou de colonnes linéairement indépendantes, pour des combinaisons linéaires dont les coefficients sont dans $\mathbb{G}$. On le note* $\mathrm{rang}_{\mathbb{G}}(M)$.

La définition précédente fait intervenir deux corps, inclus l'un dans l'autre. L'inclusion suffit à définir correctement le rang. Le lemme suivant, précise le cas particulier où l'un des corps est une extension de l'autre. On reprend ici la formulation de [BF92].

Lemme 4.1. *Soit $\mathbb{F}$ un sous-corps de $\mathbb{G}$ et M une matrice à coefficient dans $\mathbb{F}$. On a alors* $\mathrm{rang}_{\mathbb{F}}(M) = \mathrm{rang}_{\mathbb{G}}(M)$.

En particulier, si la matrice est à coefficients dans $\mathbb{Z}$, alors le rang sur un corps fini $\mathbb{K}$ ne dépend que de la caractéristique de celui ci. Si $\mathbb{F}$ est un corps fini de caractéristique p, on écrit rang_p au lieu de $\mathrm{rang}_{\mathbb{F}}$ pour désigner le rang sur $\mathbb{F}$.

En complexité de la communication, le rang joue un rôle capital. Il est bien connu que le logarithme du rang réel de la matrice est une borne inférieure sur la communication déterministe [MS82]. Depuis que cette borne inférieure est connue, la question de la borne supérieure est ouverte. On sait qu'il existe une différence non constante entre le logarithme du rang et la complexité de la communication [RS95]. Néanmoins, il est possible que cette différence soit au plus polynomiale, c'est-à-dire que la complexité de la communication soit bornée supérieurement par un polynôme en le logarithme du rang de la matrice.

En complexité en boîte non-locale, comme en complexité de la communication, on s'intéresse à des problèmes dont les entrées sont réparties entre les joueurs. Considérons une fonction f qui dépend de deux variables, x pour Alice, y pour Bob. Intuitivement, on aimerait pouvoir qualifier ces problèmes de communication ou de boîte non-locale comme des problèmes consistant à séparer la fonction f en une partie qui dépend de x et une qui dépend de y.

On a donné au chapitre 1 un théorème allant dans ce sens. En effet, le théorème 1.3, page 9 affirme qu'un protocole quantique induit une factorisation de la matrice de communication. Si la matrice de communication du problème est M_f, le théorème affirme qu'il existe une matrice proche de M_f qui se factorise sous la forme $A \cdot B$, avec des conditions supplémentaires sur les normes des matrices A et B. Rappelons que ce lemme est à

la base de toutes les bornes inférieures sur la complexité de la communication quantique que nous avons vu ici. On va voir ci-dessous que le rang peut être exprimé comme une propriété de factorisation similaire.

L'autre propriété intéressante du rang est que sa définition est compatible avec les relaxations du modèle de complexité de la communication. Ainsi, il est possible de relâcher la définition du rang, ou de la lisser, pour définir une notion de rang approché. Pour obtenir le rang approché de la matrice, il suffit de minimiser le rang sur une boule centrée sur la matrice de communication. Cette définition suffit pour prouver des bornes inférieures sur la complexité de la communication probabiliste et quantique [BdW01].

L'inconvénient du rang approché ainsi défini est qu'on ne sait pas le calculer efficacement. Il n'a pas été prouvé que ce calcul était NP-difficile, mais on sait en revanche que plusieurs problèmes de minimisation du rang similaires le sont. Même si l'argument n'a pas valeur de preuve, cela tend à faire penser que le calcul du rang approché est difficile. Récemment, on a néanmoins découvert un algorithme d'approximation pour ce problème [LS09]. L'idée de l'algorithme mérite d'être développée ici. Pour calculer le rang approché, les auteurs considèrent une relaxation semi-définie du problème. Celle-ci peut alors être calculée efficacement. Il leur reste à montrer que la valeur relâchée est proche de la valeur initiale. Nous avons déjà évoqué ici cette technique de conception d'algorithme d'approximation dans la section 3.5. Pour le rang approché, ce qu'on obtient avec cette technique de relaxation semi-définie est exactement la norme ν étudiée par Linial et Shraibman et qui a été développé au chapitre 3.

La complexité de la communication considère le rang sur $\mathbb{R}$. La structure algébrique des boîtes non-locales nous amène à considérer le rang sur $\mathbb{Z}/2\mathbb{Z}$. On montre maintenant que pour tout corps de base, le rang induit bien une factorisation de la fonction. Pour une fonction $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, on note $m_{\mathbb{K}}(f) = \min\{k \in \mathbb{N} : \exists p_i, q_i : \mathbb{K}^n \to \{0,1\}, i = 1, \ldots, k, f(x,y) = \sum_{i=1}^{k} p_i(x)q_i(y)\}$.

Lemme 4.2 ([BdW01])**.** *Soit $f : \{0,1\} \times \{0,1\} \to \{0,1\}$ et M_f la matrice de communication associée. Soit $\mathbb{K}$ un corps inclus dans $\mathbb{R}$ et contenant $\mathbb{Z}/2\mathbb{Z}$. On a $m_{\mathbb{K}}(f) = \mathrm{rang}_{\mathbb{K}}(M_f)$.*

Démonstration. Soit $m = m(f)$ et $f(x,y) = \sum_{i=1}^{m} p_i(x)q_i(y)$. Soit M_i la matrice définie par $M_i[x,y] = p_i(x)q_i(y)$. Chacune de ces matrices est bien de rang 1 sur $\mathbb{K}$. De plus $M_f = \sum_{i=1}^{m} M_i$. On a finalement $\mathrm{rang}(M_f) \leq \sum_i \mathrm{rang}(M_i) \leq m$.

Inversement, soit $\mathrm{rang}(M_f) = r$. Cela implique qu'il existe une factorisation $M_f = A \cdot B$ où A et B sont de dimensions respectivement $2^n \times r$ et $r \times 2^n$. En effet, soient $c_1, \ldots, c_r$ r colonnes indépendantes de M_f. A est alors la matrice composée de ces colonnes. La i-ème colonne de B est elle composée des coefficients $\alpha_1, \ldots, \alpha_r$ tels que pour la i-ème colonne de M_f, $col_i(M_f) = \sum \alpha_i c_i$. On a alors $f(x,y) = M_f[x,y] = \sum_i A[x,i]B[i,y]$. On définit donc les fonctions $p_i(x) = A[x,i]$ et $q_i(y) = B[i,y]$, et on a bien $f(x,y) = \sum_{i=1}^{r} p_i(x)q_i(y)$ et donc $m(f) \leq r$. □

4.2.2 Protocole de van Dam

Commençons par rappeler le fonctionnement du protocole de van Dam. Celui-ci permet de calculer une fonction quelconque en parité sans communication. La valeur exacte de la fonction peut ensuite être connue avec un seul bit de communication. Le protocole

permet de déduire une borne supérieure sur la complexité en boîte non-locale. L'auteur a noté que la complexité en boîte non-locale pouvait être exponentielle. Nous affinons ici la preuve pour donner la complexité exacte du protocole.

Proposition 4.1 ([vD05])**.** *Soit $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, on a $NL(f) \leq \mathrm{rang}_2(M_f)$.*

Démonstration. Soit $t = \mathrm{rang}_2(M_f)$. En appliquant le lemme 4.2 sur le corps $\mathbb{Z}/2\mathbb{Z}$, on en déduit qu'il existe une décomposition $f(x,y) = \bigoplus_{i=t}^{t} u_i(x)v_i(y)$ où les u_i et v_j sont des polynômes sur $\mathbb{Z}/2\mathbb{Z}$. Le protocole pour calculer f en utilisant t boîtes non-locales est alors le suivant :

- x et y étant fixés, Alice et Bob utilise une boîte non-locale pour chaque $(u_i(x), v_i(y))$. Chacune de celles-ci donnent en sortie (a_i, b_i) tel que $a_i \oplus b_i = u_i(x)v_i(y)$.
- Alice répond $\bigoplus_i a_i$, somme modulo 2 de toutes les sorties qu'elle a obtenues.
- Bob répond la valeur $\bigoplus_i b_i$.

Par définition des boîtes non-locales, la somme $a_i \oplus b_i$ est toujours égale à $u_i(x)v_i(y)$. En utilisant la commutativité, on obtient :

$$\begin{aligned} \bigoplus_i a_i \oplus \bigoplus_i b_i &= \bigoplus_i (a_i \oplus b_i), \\ &= \bigoplus_i u_i(x)v_i(y) = f(x,y). \end{aligned}$$

□

Ce protocole nous donne une borne sur la complexité. En appliquant le fait que $\mathrm{rang}_2(M_f) \leq 2^n$, on en déduit que pour toute fonction f, $NL(f) \leq 2^n$. L'objet de la section suivante est d'étudier l'optimalité de cette borne supérieure.

4.2.3 Caractérisation de la complexité en boîte non-locale parallèle

Si on étudie attentivement le protocole de van Dam, on voit que celui-ci satisfait clairement deux propriétés. La première est que toutes les boîtes sont utilisées simultanément. La seconde est que les joueurs répondent toujours par le XOR de toutes les sorties des boîtes non-locales. En se conformant aux notations présentées plus haut, on en déduit que la borne supérieure de la proposition 4.1 est une borne sur $NL^{\|,\oplus}$. On peut prouver assez facilement que celle-ci est optimale pour ce modèle.

Théorème 4.1. *Soit $f : X \times Y \to \{0,1\}$. On a $NL^{\|,\oplus}(f) = \mathrm{rang}_2(M_f)$.*

Démonstration. Le protocole de van Dam vérifie $NL^{\|,\oplus}(f) \leq \mathrm{rang}_2(M_f)$. Montrons maintenant que $NL^{\|,\oplus}(f) \geq \mathrm{rang}_2(M_f)$. Soit $NL^{\|,\oplus}(f) = t$. Soit $\mathcal{P}$ un protocole à t boîtes non-locales. Notons les entrées de celles-ci (p_i, q_i) et les sorties (a_i, b_i). La correction du protocole implique que $f(x,y) = (\bigoplus_i a_i) \oplus (\bigoplus_i b_i)$. On a ainsi $f(x,y) = \bigoplus_i^t p_i(x)q_i(y)$. Ainsi, d'après le lemme 4.2, on en déduit que $t \geq \mathrm{rang}_2(M_f)$, ce qui prouve le théorème. □

Naturellement, on veut maintenant savoir si on peut retirer les deux propriétés. Dans la suite de cette section, nous allons voir que la seconde est superflue. Précisément, nous allons montrer que le meilleur protocole consiste toujours à répondre par le XOR des sorties des boîtes non-locales. Ceci n'est valable que dans le cas de l'utilisation simultanée des boîtes.

Théorème 4.2. *Soit $f : X \times Y \to \{0,1\}$. On a $NL^{||}(f) \leq NL^{||,\oplus}(f) \leq NL^{||}(f) + 2$.*

La preuve du théorème se base sur le lemme suivant.

Lemme 4.3. *Soient a et b les sorties d'une boîte non-locale dont les entrées sont p et q. Soit F, G, H des coefficients indépendants de a et de b. Si pour tout couple (a, b), on a $aF \oplus bG \oplus H = 0$, alors $F = G$ et $H = pqG$.*

Démonstration. En posant $a = 0$ et $b = pq$, on a $Gpq = H$. D'autre part, en posant $a = 1$ et $b = 1 \oplus pq$, on a $F \oplus G \oplus Gqp \oplus H = 0$. En sommant les deux équations, on obtient $F = G$. □

Avant de prouver le théorème principal de cette section, nous allons préciser une propriété naturelle des protocoles optimaux. Ce qu'on propose de montrer en premier, c'est qu'on peut supposer que les entrées des boîtes non-locales sont linéairement indépendantes. Si ce n'est pas le cas, certaines boîtes sont inutiles.

Définition 4.5. *Soient $p_1, \ldots, p_n$ des fonctions booléennes. On dit que ces fonctions sont linéairement indépendantes si pour toutes constantes $c_1, \ldots, c_n \in \{0,1\}$, et fonctions α, β, on a pour tout (x, y) $\sum_{i=1}^{n} c_i p_i = \alpha(x) \oplus \beta(y) \Rightarrow c_i = 0 \ \forall i$ et $\alpha(x) = \beta(y)$.*

Lemme 4.4. *Soit f une fonction booléenne telle que $NL^{||}(f) = t$. Alors il existe un protocole utilisant t boîtes non-locales tel que les entrées des boîtes sont linéairement indépendantes.*

Démonstration. L'idée de la preuve est la suivante : si les entrées ne sont pas indépendantes, alors il existe une relation de linéarité entre une entrée (p_k, q_k) et les autres entrées. On se sert ensuite de cette relation pour supprimer la k-ième boîte, contredisant l'optimalité du protocole. Supposons qu'on a $\bigoplus_{i\in[t]} c_i p_i(x) q_i(y) = \alpha(x) \oplus \beta(y)$ avec $c_k = 1$. Alors $p_k(x)q_k(y) = \bigoplus_{i\in[t]\setminus k}(p_i(x) \oplus q_i(y)) \oplus \alpha(x) \oplus \beta(y)$. Dans le protocole initial, on a $a_k \oplus b_k = p_k(x)q_k(y)$, où (a_k, b_k) sont les sorties de la k-ème boîte. Au lieu d'utiliser cette boîte, Alice peut directement calculer $a_k = \alpha(x) \bigoplus_{i\in[t]\setminus k} p_i(x)$ et Bob $b_k = \beta(x) \bigoplus_{i\in[t]\setminus k} q_i(x)$. On obtient ainsi un protocole utilisant $t-1$ boîtes non-locales, ce qui contredit l'hypothèse. □

Démonstration du théorème. Par définition, on a $NL^{||}(f) \leq NL^{||,\oplus}(f)$. On doit donc maintenant prouver que $NL^{||,\oplus} \leq NL^{||}(f) + 2$. Soit $NL^{||}(f) = t$, et fixons un protocole déterministe $\mathcal{P}$ utilisant t boîtes non-locales. D'après le lemme 4.4, on peut supposer sans perte de généralité que les entrées des boîtes sont linéairement indépendantes.

On commence par fixer quelques notations. Sur l'entrée du problème (x, y), on note $(p_i(x), q_i(y))$ l'entrée de la i-ème boîte non-locale, et (a_i, b_i) les sorties de celle-ci. Notons $\mathbf{a} = (a_1, \ldots, a_t)$ et $\mathbf{b} = (b_1, \ldots, b_t)$. On décompose la sortie d'Alice en polynôme $A(x, \mathbf{a}) = \bigoplus_{S\neq[t]} A_S(x) a_S$, où $a_S = \prod_{i\in S} a_i$, $a_\emptyset = 1$ par convention et les A_S sont des polynômes en x. De même $B(y, \mathbf{b}) = \bigoplus_{S\in[t]} B_S(y) b_S$, où $b_s = \prod_{i\in S} b_i$ et les B_S sont des polynômes en y. Le protocole étant correct, on a pour tout $(x, y, \mathbf{a}) \in X \times Y \times \{0,1\}^t$ et $\mathbf{b}$ tel que $b_i = a_i \oplus p_i(x)q_i(y)$, $f(x, y) = A(x, \mathbf{a}) \oplus B(y, \mathbf{b})$.

Notons que $A(x, \mathbf{a}) = A_\emptyset(x) + \bigoplus_{\substack{S\subseteq[t] \\ S\neq\emptyset}} A_S(x) a_S$ et $B(y, \mathbf{b}) = B_\emptyset(y) + \bigoplus_{\substack{S\subseteq[t] \\ S\neq\emptyset}} B_S(y) b_S$. Ces deux termes peuvent être calculés en utilisant deux boîtes non-locales avec entrées

$(A_\emptyset(x), 1)$ et $(1, B_\emptyset(y))$. Par la suite, tous les sous ensembles de $[t]$ qu'on considère sont supposés non vides. La preuve découle ensuite des étapes suivantes.

1ère étape : Montrons que pour tout $(x, y, \mathbf{a})$ et $T \subseteq [t]$,

$$\bigoplus_{S:T\subseteq S} A_S(x)a_{S\setminus T} = \bigoplus_{S:T\subseteq S} B_S(y)b_{S\setminus T}. \tag{4.1}$$

On montre la validité de l'équation 4.1 par induction sur la taille de T. On a par définition $f(x, y) = A(x, \mathbf{a}) \oplus B(y, \mathbf{b})$. En factorisant a_k et b_k dans cette expression, on déduit que pour tout $(x, y, \mathbf{a})$,

$$f(x,y) = \bigoplus_{S:k\notin S} (A_S(x)a_S \oplus B_S(y)b_S) \oplus a_k \left(\bigoplus_{S:k\in S} A_S(x)a_{S\setminus\{k\}}\right) \oplus b_k \left(\bigoplus_{S:k\in S} B_S(y)b_{S\setminus\{k\}}\right).$$

En appliquant le lemme 4.3, on déduit que $\bigoplus_{S:k\in S} A_S(x)a_{S\setminus\{k\}} = \bigoplus B_S(y)b_{S\setminus\{k\}}$, ce qui prouve que l'équation 4.1 est vraie dans le cas $|T| = 1$. Supposons maintenant que l'équation est vraie pour tout ensemble de taille $n \leq t-1$, et fixons un ensemble T tel que $|T| = n$. Soit $k \notin T$. L'hypothèse d'induction assure

$$\bigoplus_{S:T\subseteq S} A_S(x)a_{S\setminus T} = \bigoplus_{S:T\subseteq S} B_S(y)b_{S\setminus T}.$$

En factorisant a_k et b_k, et en appliquant le lemme 4.3, on obtient directement

$$\bigoplus_{S:T\cup\{k\}\subseteq S} A_S(x)a_{S\setminus T\cup\{k\}} = \bigoplus_{S:T\cup\{k\}\subseteq S} B_S(y)b_{S\setminus T\cup\{k\}}.$$

En appliquant la même procédure pour tout T tel que $|T| = n$ et $k \notin T$, on prouve que l'équation 4.1 est vraie pour tout ensemble de taille $n+1$, ce qui conclut l'induction.

2ème étape : Montrons que pour tout (x, y) et tout $T \subseteq [t]$:

$$|T| > 1 \quad \Rightarrow \quad A_T(x) = B_T(y) = 0, \tag{4.2}$$
$$|T| = 1 \quad \Rightarrow \quad A_T(x) = B_T(y). \tag{4.3}$$

On prouve cet énoncé par induction descendante sur $|T|$. L'équation 4.1 appliquée avec $T = [t]$ assure que $A_{[t]}(x) = B_{[t]}(y)$. Ces deux fonctions sont égales pour tout (x, y), donc elles sont constantes. Posons $A_{[t]}(x) = B_{[t]}(y) = c_{[t]}$. Soit $k \in [t]$. Appliquons l'équation 4.1 à $T = [t]\setminus k$, et on obtient

$$A_{[t]\setminus k} \oplus c_{[t]}a_k = B_{[t]\setminus k} \oplus c_{[t]}b_k.$$

D'après le lemme 4.4, on a $c_{[t]} = 0$.

Soit $n \geq 2$. Supposons que pour tout S de taille supérieure ou égale à n, on a $A_S(x) = B_S(y) = 0$. Soit T tel que $|T| = n-1$. L'équation 4.1 appliquée à T donne

$$A_T(x) \oplus \bigoplus_{k\notin T} A_{T\cup\{k\}}(x)a_k = B_T(y) \oplus \bigoplus_{k\notin T} B_{T\cup\{k\}}(y)b_k.$$

En appliquant le lemme 4.3 successivement à tous les couples (a_k, b_k), on en déduit que $A_{T\cup\{k\}} = B_{T\cup\{k\}}$ pour tout $k \notin T$. Posons donc $c_{T\cup\{k\}} = A_{T\cup\{k\}}(x) = B_{T\cup\{k\}}(y)$. Par

suite $\bigoplus_{k\notin T} c_{T\cup\{k\}} p_k(x)q_k(y) = A_T(x) \oplus B_T(y)$. Par le lemme 4.4, on en déduit donc que $c_{T\cup\{k\}} = 0$ pour tout k et $A_T(x) = B_T(y)$. Ceci montre que les équations 4.2 et 4.3 sont vraies pour tout (x, y) et $T \subseteq [t]$.

Finalement, en utilisant le lemme 4.2, on en déduit bien que $NL^{\|}(f) \geq \text{rang}_2(f) = NL^{\|,\oplus}(f)$. □

Nous avons grâce au théorème précédent caractérisé exactement la complexité en boîte non-locale lorsque celles-ci sont utilisées simultanément. La différence entre le rang et la complexité est un écart additif d'au plus deux. Cela vient des termes locaux. En effet, il peut y avoir un terme dépendant uniquement de x, un autre uniquement de y, et ces deux termes entrent en compte dans le calcul du rang. Du point de vue de la complexité, ce sont en revanche des termes locaux qui peuvent être calculés sans boîtes non-locales.

Ce théorème nous permet de retirer l'hypothèse contraignant les joueurs à toujours répondre la parité des sorties des boîtes. Qu'en est-il de la seconde hypothèse ? Peut-on montrer que l'utilisation optimale des boîtes non-locales est toujours simultanée ? On va voir maintenant que c'est impossible. En utilisant les boîtes non-locales de manière séquentielle, on peut faire baisser la complexité. Dans la proposition suivante, on utilise le rang pour montrer que les fonctions EQ et $Disj$ ont une complexité en boîte non-locale simultanée exponentielle. Ensuite, on donne des protocoles utilisant les boîtes en série et de complexité linéaire pour ces fonctions.

Proposition 4.2. *Considérons les fonctions IP, EQ et $Disj$ définies sur $\{0,1\}^n \times \{0,1\}^n$ et à valeurs dans $\{0,1\}$. Ces fonctions vérifient :*
- $NL^{\|}(IP) = n$,
- $NL^{\|}(EQ) = 2^n$,
- $NL^{\|}(Disj) = 2^n$.

Démonstration. On va utiliser le théorème 4.2 et le lemme 4.2. Ceci nous permet de trouver la complexité en boîte non-locale en écrivant un polynôme interpolateur multilinéaire. En effet, si deux polynômes multilinéaires P et Q sur $\mathbb{Z}/2\mathbb{Z}$ interpolent la même fonction booléenne, alors $P \oplus Q$ est identiquement nul, donc $P = Q$. Comme les deux polynômes sont multilinéaires, ils ont exactement les même monômes et par conséquents, le même nombre de monômes.

Donnons maintenant les polynômes pour chacune des fonctions.
- $IP(x, y) = \bigoplus_i x_i y_i$.
- $EQ(x, y) = \prod_i (1 \oplus x_i \oplus y_i) = \bigoplus_{S\subseteq[n]} \prod_{i\in S} (x_i \oplus y_i)$. La matrice de communication de la fonction EQ est simplement la matrice identité.
- $Disj(x, y) = \prod(1 \oplus x_i y_i) = \bigoplus_{S\subseteq[n]} \prod_{i\in S} x_i y_i$.

□

Le lemme précédent montre que les complexités en boîte non-locale des fonctions EQ et $Disj$ sont exponentielles. Même si des boîtes non-locales existaient dans la nature, l'utilisation d'une quantité exponentielle de ressources serait certainement un obstacle à la mise en pratique de ces protocoles.

Nous allons maintenant présenter des protocoles de complexité linéaire pour les fonctions EQ et $Disj$. Ceci montre de plus qu'on ne peut retirer sans perte de généralité la contrainte sur l'utilisation simultanée des boîtes non-locales.

Proposition 4.3. *[LU08] On a $NL(EQ) = O(n)$ et $NL(Disj) = O(n)$.*

Démonstration. On présente le protocole pour la fonction $Disj$, celui pour EQ étant similaire. Le protocole est basé sur l'idée suivante. Fixons deux fonctions $f : X \times Y \to \{0,1\}$ et $g : X \times Y \to \{0,1\}$ et supposons que $f(x,y)$ et $g(x,y)$ sont donnés au joueur en parité. C'est à dire que chaque joueur connaît un bit tel que la parité des deux bits est égal à la valeur de la fonction. Alors les joueurs peuvent calculer le produit $f(x,y)g(x,y)$ avec deux boîtes non-locales. Soit $(a_f, b_f) \in \{0,1\}^2$ et $(a_g, b_g) \in \{0,1\}^2$ tels que $a_f \oplus b_f = f(x,y)$ et $a_g \oplus b_g = g(x,y)$. On peut alors calculer directement le produit $f(x,y)g(x,y) = (a_f \oplus b_f)(a_g \oplus b_g) = a_f a_g \oplus a_f b_g \oplus b_f a_g \oplus b_f b_g$. En utilisant une boîte pour calculer $a_f b_g$ et une pour $a_g b_f$, et en ajoutant au sorties les termes locaux $a_f a_g$ et $b_f b_g$, on a bien calculé le produit $f(x,y)g(x,y)$.

Reprenons l'expression de la fonction $Disj$ sous la forme $Disj(x,y) = \prod(1 \oplus x_i y_i)$. Chaque terme $1 \oplus x_i y_i$ peut être calculé en parité en utilisant une boîte non-locale. En appliquant n fois l'algorithme pour calculer le produit, on peut calculer $\prod_{i\in[n]}(1 \oplus x_i y_i)$. On utilise au total n boîtes pour calculer les monômes $1 \oplus x_i y_i$ en parité, puis $2n$ boîtes pour effectuer le produit. Le total est donc de $3n$ boîtes non-locales. □

4.2.4 Comparaison de la complexité en boîtes non-locales et de la communication

Les fonctions vues à la section précédente avaient toutes une complexité en boîte non-locale séquentielle du même ordre que la complexité de la communication. Dans le cas simultané, l'écart entre complexité de la communication et la complexité en boîte non-locale était, pour ces fonctions, au plus exponentiel. Nous allons dans cette section généraliser la comparaison entre ces deux ressources. Le résultat suivant est un corollaire du théorème 4.1.

Corollaire 4.1. *Pour toute fonction $f : X \times Y \to \{0,1\}$, $NL^{\|}(f) \leq NL^{\|,\oplus} \leq 2^{D(f)}$.*

Démonstration. On a d'après le théorème 4.1, $NL^{\|}(f) \leq NL^{\|,\oplus} = \text{rang}_2(M_f)$. Ici, on peut conclure en remarquant que $\text{rang}_2(M_f) \leq 2^{D(f)}$. En général, on donne comme borne inférieure sur $2^{D(f)}$ le rang sur $\mathbb{R}$, qui est plus grand. Néanmoins, il est facile de voir que la preuve pour $\mathbb{R}$ est vraie quel que soit le corps de base, donc en particulier pour $\mathbb{Z}/2\mathbb{Z}$ [MS82]. Néanmoins, nous allons donner une autre preuve, due à [LS08a]. Nous allons montrer que pour toute matrice M, on a $\text{rang}_2(M) \leq \text{rang}_{\mathbb{R}}(M)$. Comme d'une part, $\text{rang}_2(M_f) = NL^{\|,\oplus}(f)$, et d'autre part $\text{rang}_{\mathbb{R}}(M_f) \leq 2^{D(f)}$, on a bien $NL^{\|,\oplus} \leq 2^{D(f)}$.

Soit M une matrice à coefficients dans $\mathbb{Z}/2\mathbb{Z}$. En appliquant le lemme 4.1, on déduit $\text{rang}_{\mathbb{Q}}(M) \leq \text{rang}_{\mathbb{R}}(M)$, car $\mathbb{R}$ est une extension de $\mathbb{Q}$. Il suffit maintenant de prouver que $\text{rang}_2(M) \leq \text{rang}_{\mathbb{Q}}(M)$. Soit $k = \text{rang}_{\mathbb{Q}}(M)$. Alors pour tout ensemble $e_0, \ldots, e_k$ de $k+1$ lignes (ou colonnes) de la matrices M, il existe des coefficients $\lambda_0, \ldots, \lambda_k$ dans $\mathbb{Q}$ tel que $\sum_{i=0}^{k} \lambda_i e_i = 0$.

On va maintenant montrer que $e_0, \ldots e_k$ sont linéairement dépendants avec des coefficients sur $\mathbb{Z}/2\mathbb{Z}$. Chaque λ_i étant dans $\mathbb{Q}$, on peut l'exprimer comme une fraction irréductible $\lambda_i = p_i/q_i$. Notons $d = ppcm(q_i)$. Ainsi, on a pour tout i, $d\lambda_i \in \mathbb{Z}$. Finalement, notons $\lambda'_i = d\lambda_i \mod 2$. On vérifie facilement que $\sum_i \lambda'_i e_i = 0 \mod 2$, ce qui achève de prouver le corollaire. □

La proposition suivante donne une borne inférieure sur la complexité en boîte non-locale. L'idée est de simuler une boîte non-locale avec un bit de communication.

Proposition 4.4. *Pour toute fonction* $f : X \times Y \to \{0,1\}$, $D^{\to}(f) \leq NL(f)$.

Démonstration. Soit $\mathcal{P}$ un protocole pour f utilisant t boîtes non-locales. Au lieu d'utiliser les boîtes, Alice va envoyer ses t entrées $p_1, \ldots, p_t$ à Bob et considérer que toutes les boîtes non-locales ont sorti de son côté la valeur 0. De son côté, Bob considère que les boîtes ont sorti de son côté $p_i q_i$, où les q_i sont ses entrées. Comme le protocole $\mathcal{P}$ est déterministe, il est en particulier correct dans le cas où les sorties sont $(0, p_i q_i)$. Le protocole de communication est correct et on a $D(f) \leq t$. □

Ce dernier résultat nous permet de donner facilement des bornes inférieures sur la complexité en boîte non-locale déterministe. Cela montre en particulier que les protocoles définis à la proposition 4.3 sont optimaux. En effet, on sait que pour ces fonctions, la complexité de la communication déterministe est linéaire.

Nous n'avons parmi les exemples de fonction étudiée plus haut, aucun exemple de séparation entre la complexité en boîte non-locale (séquentielle) et la complexité de la communication. C'est un problème ouvert de trouver une séparation entre complexité de la communication et complexité en boîte non-locale, ou bien de trouver une fonction dont la complexité en boîte non-locale est super-linéaire. Rappelons qu'une fonction $g : \mathbb{N} \to \{0,1\}$ est dite super-linéaire si pour toute constance c, on a $g(n) \geq c.n$ pour tout n suffisamment grand.

4.3 Complexité probabiliste

4.3.1 Bornes sur la complexité en boîte non-locale

Nous allons maintenant nous intéresser au calcul probabiliste. Dans ce contexte, on autorise une petite erreur dans la probabilité de donner la bonne réponse. La plupart des bornes vues dans la section précédente sont montrées en transformant un protocole de communication en protocole en boîte non-locale. Ainsi, en utilisant le fait qu'un protocole probabiliste est une distribution de protocoles déterministes, on va pouvoir étendre certains résultats à la complexité en boîte non-locale probabiliste.

En étudiant le cas probabiliste, nous précisons encore la puissance calculatoire des boîtes non-locales. Sous cette forme relâchée, ce modèle tend à se rapprocher encore plus de la complexité de la communication. Nous allons borner la complexité en boîte non-locale par trois mesures différentes : le rang, la communication et les normes de factorisation.

Rang approché

Dans le cas déterministe, la complexité en boîte non-locale est caractérisée par le rang sur $\mathbb{Z}/2\mathbb{Z}$ de la matrice de communication. Ceci induit une factorisation de la matrice de communication. Dans le cas probabiliste, on a une factorisation similaire. Toutefois, cette factorisation n'est plus exacte mais approchée.

Comme dit plus haut, le logarithme du rang sur $\mathbb{R}$ de la matrice de communication M_f est une borne inférieure sur la complexité de la communication. On peut étendre cette borne inférieure au cas probabiliste relativement aisément. Il suffit pour cela de prendre le plus petit rang parmi les matrices proches de la matrice d'origine pour la norme infinie. Avec les boîtes non-locales, cette relaxation n'est pas aussi triviale. Le problème est qu'on considère maintenant le rang sur $\mathbb{Z}/2\mathbb{Z}$. Or une boule centrée sur la matrice M_f contient a priori des matrices dont les coefficients ne sont plus dans $\{0,1\}$. Le rang sur $\mathbb{Z}/2\mathbb{Z}$ de ces matrices n'est donc pas défini. Nous proposons ici une définition nouvelle du rang approché sur un corps fini.

Définition 4.6. *Soit $\mathcal{R}_t$ l'enveloppe convexe de l'ensemble des matrices booléennes de rang sur $\mathbb{Z}/2\mathbb{Z}$ au plus t. Pour une matrice booléenne A, le rang approché sur $\mathbb{Z}/2\mathbb{Z}$ est défini par $\varepsilon-rang_2(A) = \min\{t : \exists A' \in \mathcal{R}_t \text{ tel que } \| A - A' \|_\infty \leq \varepsilon \}$.*

Nous allons maintenant montrer que cette définition caractérise bien la complexité en boîte non-locale dans le cas simultané. La preuve est en deux temps. D'abord, nous donnons une expression alternative pour le rang, puis nous montrons que cette expression peut être interprétée comme la sortie d'un protocole probabiliste, et inversement.

Proposition 4.5. *$\varepsilon-\mathrm{rang}_2(A)$ est le plus petit t tel qu'il existe un ensemble de matrices booléennes $A_1, \ldots, A_N$ et une distribution de probabilité μ sur $[N]$ tels que :*
- *pour tout $n \in [N]$, $rang_2(A_n) \leq t$,*
- *pour tout (i,j), $\mathrm{Prob}_\mu[A[i,j] = A_n[i,j]] \geq 1 - \varepsilon$.*

Démonstration. Soient $\mathrm{rang}_2(A) = t$ et A' telle que $A' \in \mathcal{R}_t$ et $\| A - A' \|_\infty \leq \varepsilon$. Soient $A_1, \ldots, A_N$ les sommets de $\mathcal{R}_t$. Par définition, ces matrices sont à coefficients dans $\{0,1\}$ et vérifient $\mathrm{rang}_2(A_i) \leq t$. Par définition, A' est combinaison convexe des matrices $A_1, \ldots A_N$. Il existe des coefficients $\alpha_1, \ldots, \alpha_N \geq 0$ tels que $A' = \sum_n \alpha_n A_n$ et $\sum_n \alpha_n = 1$. Soit μ la distribution sur $[N]$ définie par $p(n) = \alpha_n$. On a évidemment $\mathbf{E}_\mu[A_n[x,y]] = A'[x,y]$, soit $\mathrm{Prob}_\mu[A[i,j] \neq A_n[i,j]] = |A[x,y] - \mathbf{E}_\mu[A_n[x,y]]| \leq 1 - \varepsilon$. Les matrices $A_1, \ldots A_N$ et la distribution μ ont donc les bonnes propriétés.

Inversement, supposons qu'il existe $A_1, \ldots, A_N$ vérifiant les bonnes propriétés. Soit A' la matrice définie par $A'[x,y] = \mathbf{E}_\mu[A_n[x,y]]$. On a $A' \in \mathcal{R}_t$ et $\| A - A' \|_\infty$, et donc $\varepsilon-\mathrm{rang}_2(A) \leq t$. □

Théorème 4.3. *Pour toute fonction booléenne $f : X \times Y \to \{0,1\}$, $NL_\varepsilon^{||,\oplus}(f) = \varepsilon-\mathrm{rang}_2(M_f)$.*

Démonstration. Soit $\mathcal{P}$ un protocole pour f utilisant t boîtes non-locales en parallèle et correct avec probabilité au moins $1 - \varepsilon$. Les entrées des boîtes dépendent de l'entrée (x,y) et de l'aléa partagé par les joueurs. Ecrivons explicitement l'aléa r choisi suivant une certaine distribution de probabilité dans un ensemble R. Notons les entrées des boîtes $(p_i(x,r), q_i(y,r))$, et les sorties (a_i, b_i). Les sorties des joueurs étant toujours la somme des sorties obtenues des boîtes, la fonction calculée est $g_r(x,y) = \bigoplus_{i=1}^t p_i(x,r) q_i(y,r)$. En particulier, on sait que pour chaque r, la matrice de communication M_{g_r} vérifie $\mathrm{rang}_2(M_{g_r}) \leq t$. Par ailleurs, le protocole étant ε-correct, on a pour tout (x,y), $\mathrm{Prob}_r[M_f[x,y] = M_{g_r}[x,y]] \geq 1 - \varepsilon$. Ainsi, on a bien $\varepsilon-\mathrm{rang}_2(M_f) \leq t$.

Réciproquement, supposons que $\varepsilon-\mathrm{rang}_2(M_f) = t$. D'après la proposition 4.5, il existe un ensemble de matrices booléennes $A_1, \ldots, A_N$ et une distribution de probabilité μ sur $[N]$

– pour tout $i \in [N]$, $rang_2(A_i) \leq t$,
– pour tout (x, y), $\text{Prob}_i[M_f[x, y] = A_i[x, y]] \geq 1 - \varepsilon$.

Si g_i est la fonction dont A_i est la matrice de communication, on a pour tout i, $NL^{\|}(g_i) \leq t$. De plus, en tirant une de ces fonctions au hasard suivant μ, on sait que $\text{Prob}_\mu[g_i[x, y] = f(x, y)] \geq 1 - \varepsilon$. On a donc prouvé $NL^{\oplus,\|}(f) \leq t$, ce qui clot la démonstration. □

Nous n'avons pas d'équivalent probabiliste au théorème 4.2 qui caractérise la complexité en boîte non-locale simultanée déterministe. Nous ne pouvons pas a priori enlever l'exposant $\oplus$, qui impose aux joueurs de répondre par la parité de toutes les sorties des boîtes non-locales. Si on voulait prouver que ce résultat est vrai dans le cas probabiliste, il semblerait naturel d'appliquer le raisonnement suivant :

1. Un protocole probabiliste est une combinaison convexe de protocoles déterministes.
2. Pour chaque protocole déterministe, on peut supposer sans perte de généralité que les joueurs répondent par la parité de toutes les sorties.
3. Par convexité, on peut également supposer que dans le cas probabiliste, les joueurs répondent par la parité de toutes les sorties.

Ce raisonnement serait correct si on l'appliquait à des protocoles de communication. Dans le cas des boîtes non-locales, la faille se situe dès la première étape. En effet, il n'est pas du tout assuré qu'on puisse décomposer un protocole probabiliste utilisant des boîtes non-locales en combinaison convexe de protocoles déterministes. Les boîtes non-locales introduisent une nouvelle source d'aléa et même si on fixe l'aléa partagé par les joueurs, il leur est toujours possible d'utiliser cette source là. Cet aléa est alors payant du point de vue de la complexité, mais pourrait éventuellement permettre de réduire le coût global en boîtes non-locales. Les protocoles obtenus en fixant l'aléa partagé par les joueurs n'ont a priori pas de raison d'être déterministes, et on ne peut pas leur appliquer individuellement le théorème 4.2. On ne sait donc pas dans le cas probabiliste si la meilleure stratégie pour les joueurs est de répondre par la parité de toutes les sorties obtenues des boîtes non-locales.

Communication

On a vu dans la section précédente qu'on ne savait pas en général si $NL_\varepsilon^{\|}(f) = NL_\varepsilon^{\|,\oplus}(f)$. Néanmoins, on peut étendre au cas probabiliste certains résultats établis dans le cas déterministe. Nous allons voir ici comment étendre les résultats sur la communication. Ceci est en général possible pour des résultats obtenus en transformant un protocole de communication en protocole utilisant des boîtes non-locales. L'idée est de décomposer un protocole de communication probabiliste en distribution sur des protocoles déterministes, et d'appliquer la même transformation sur chaque protocole de communication déterministe ainsi obtenu. Ceci définit un protocole probabiliste utilisant des boîtes non-locales. De plus, si la réduction déterministe réduit un protocole de communication sans erreur à un protocole utilisant des boîtes non-locales également sans erreur, alors dans le cas probabiliste, la réduction conserve la probabilité de succès.

Théorème 4.4. *Pour toute fonction* $f : X \times Y \to \{0, 1\}$, $R_\varepsilon(f) \leq NL_\varepsilon(f) \leq NL_\varepsilon^{\|,\oplus}(f) \leq 2^{R_\varepsilon(f)}$.

Démonstration. Pour simuler une boîte non-locale avec de la communication, Alice envoie son entrée p à Bob. La sortie d'Alice est simulée par un bit aléatoire partagé r. Bob, connaissant ce bit et les deux entrées p et q de la boîte non-locale peut simuler celle-ci en utilisant $r \oplus pq$. Par la même procédure, ils peuvent simuler t boîtes non-locales. Cette réduction conserve évidemment la probabilité de donner la bonne réponse. Ceci montre que $R_\varepsilon(f) \leq NL_\varepsilon(f)$. L'inégalité $NL_\varepsilon(f) \leq NL_\varepsilon^{\|,\oplus}(f)$ est évidente par définition.

Pour prouver la troisième inégalité, fixons une protocole de communication $\mathcal{P}$ de complexité t. Par définition, $\mathcal{P}$ est une combinaison de protocoles déterministes $\mathcal{P}^r$, où r est l'aléa des joueurs. La complexité de chacun de ces joueurs est au plus t. On note f_r la fonction calculée par le protocole $\mathcal{P}_r$. Pour chacune de ces fonctions, on a d'après le corollaire 4.1, $NL^{\|,\oplus}(f) \leq 2^t$. Fixons pour chaque fonction f_r un protocole en boîte non-locales. On en déduit un protocole pour f en prenant sur les protocoles en boîtes non-locales la même distribution que celle sur les $\mathcal{P}_r$. La complexité en boîte non-locale de ce protocole est 2^t et il est par définition ε-correct. On en déduit donc $NL^{\|,\oplus}(f) \leq 2^{R_\varepsilon(f)}$. □

Dans le cas général, où les boîtes non-locales peuvent être utilisées en série, on peut comparer la complexité en boîte non-locale avec un modèle de communication particulier. Il s'agit de la complexité simultanée, dans laquelle l'arbitre est contraint à appliquer la fonction Maj. Précisément, lorsqu'il reçoit les messages $\mathbf{a} = (a_1, \ldots, a_t)$ et $\mathbf{b} = (b_1, \ldots b_t)$, sa réponse est $Maj(a_1 \oplus b_1, \ldots, a_t \oplus b_t)$, c'est à dire 1 si la majorité des bits envoyés par les joueurs sont identiques, et 0 sinon.

Si on s'intéresse à ce modèle, c'est qu'il est naturel dans le sens où il est apparu plusieurs fois dans la littérature. Plusieurs bornes supérieures connues sur la communication sont en fait des bornes pour ce modèle, et donc a fortiori des bornes sur la complexité en boîte non-locale. La section suivante montre des exemples de telles bornes.

L'idée qu'on retrouve dans ces bornes supérieures est la suivante. Supposons qu'Alice et Bob peuvent calculer une fonction fixée avec une probabilité donnée avec un unique bit de communication. Ils peuvent alors augmenter la probabilité de succès en répétant le protocole plusieurs fois et en prenant la majorité. Cette amplification peut se faire dans le modèle simultané.

Théorème 4.5. *Pour toute fonction booléenne $f : X \times Y \to \{0,1\}$, on a*

$$NL_\varepsilon(f) \leq O(R_\varepsilon^{||,MAJ}(f)).$$

Démonstration. Fixons un protocole de communication simultané dans lequel l'arbitre reçoit deux messages a et b de longueur t, et répond $Maj(a_1 \oplus b_1, \ldots, a_t \oplus b_t)$. Fixons un circuit ET/OU pour la fonction $Maj : \{0,1\}^n \to \{0,1\}$. Il est connu qu'il existe un tel circuit utilisant $O(t)$ portes. Le ET logique correspond à la multiplication sur $\mathbb{Z}/2\mathbb{Z}$. Or, on a vu que la multiplication pouvait être calculée en parité en utilisant 2 boîtes non-locales. Les joueurs peuvent donc simuler le circuit en calculant en parité chacune des portes ET et OU. Ainsi, l'entrée du circuit est donnée en parité aux joueurs et chaque porte nécessite 2 boîtes non-locales. Au total, $O(t)$ boîtes non-locales sont suffisantes pour simuler le circuit et donc calculer la fonction en parité. □

Plus généralement, on peut borner la complexité en boîte non-locale par la complexité en circuit de la fonction calculée par l'arbitre. Dans la section suivante, nous allons donner quelques corollaires du théorème précédent. Ceci justifie d'avoir choisi la fonction Maj en particulier pour le présenter.

Norme de factorisation

On montre maintenant que la borne supérieure en $(\gamma_2^\infty)^2$ sur la communication, montrée à la section 3.6, est en fait une borne sur la complexité en boîte non-locale. Ceci découle du théorème 4.5. Cela nous permet de donner des bornes inférieures et supérieures sur la complexité en boîte non-locale.

Corollaire 4.2. *Pour toute fonction* $f : X \times Y \to \{0,1\}$, $\log(\gamma_2^\alpha(M_f)) \leq NL_\varepsilon(f) \leq O((\gamma_2^\infty(M_f))^2)$.

Démonstration. La borne inférieure est immédiate en utilisant d'une part le théorème 4.4, qui montre que $R_\varepsilon(f) \leq NL_\varepsilon(f)$ et d'autre part le théorème 3.6, qui montre qu'on a $2\log(\gamma_2^\alpha(M_f)) \leq R_\varepsilon(f)$. On a alors $2\log(\gamma_2^\alpha(M_f)) \leq NL_\varepsilon(f)$.

Pour prouver la borne supérieure, il suffit de remarquer que le protocole de communication en $O((\gamma_2^\infty)^2)$, donné dans la preuve du théorème 3.13, s'applique au modèle de communication simultané dans lequel l'arbitre applique la fonction Maj. En effet, rappelons le principe de ce protocole. On commence par changer la formulation du problème : chaque joueur reçoit maintenant un vecteur réel tel que le produit scalaire de ces deux vecteurs est égal à la valeur de la fonction, et le produit de leur norme est borné par $\gamma_2(M_f)$. On définit ensuite un protocole sans communication correct avec probabilité $1/2 + 1/\gamma_2^2(M_f)$. Comme il s'agit d'un protocole en parité, les joueurs peuvent le répéter plusieurs fois de manière à calculer la fonction avec probabilité constante, ce qui requière, d'après le corollaire 1.3, page 19, $\gamma_2(M_f)^2$ répétitions. Ils le répètent de plus dans le modèle simultané, c'est à dire en envoyant chaque fois leurs réponses à l'arbitre. L'arbitre applique donc bien la fonction majorité. Enfin, d'après le théorème 4.5, cette borne est bien une borne supérieure sur $NL_\varepsilon(f)$. □

Enfin, en utilisant le théorème 3.5, qui affirme que $\log \gamma_2$ est une borne inférieure sur la complexité de la communication quantique avec intrication, on en déduit le corollaire suivant.

Corollaire 4.3. *Pour toute fonction* $f : X \times Y \to \{0,1\}$, $NL_\varepsilon(f) \leq O(2^{2Q_\varepsilon^{ent}(f)})$.

4.3.2 Complexité probabiliste de la fonction $DISJ$

Dans le cas déterministe, nous avons vu que la complexité simultanée était très différente du modèle général. Nous avons même exhibé des fonctions, $Disj$ et EQ pour lesquelles la différence entre les deux modèles était exponentielle.

Nous allons voir maintenant que dans le modèle probabiliste, la fonction $Disj$ peut être calculée en utilisant simultanément $O(n)$ boîtes non-locales. On a donc dans le cas probabiliste une situation différente du cas déterministe, où la complexité était exponentielle. Non seulement nous n'avons plus de borne inférieure super-linéaire pour la complexité simultanée, mais nous avons également un protocole linéaire pour calculer la fonction $Disj$ en utilisant toutes les boîtes simultanément.

Proposition 4.6. $NL_\varepsilon^{||}(Disj) = O(n)$.

Démonstration. L'idée du protocole est de réduire la fonction $Disj$ au calcul d'un produit scalaire. Précisément, le protocole fonctionne comme suit. Alice et Bob partagent une

chaîne aléatoire r et vont utiliser les chaînes $x' = x \wedge r = (x_1 \wedge r_1, \ldots, x_n \wedge r_n)$ et $y' = y \wedge r$ défini de manière similaire. Ils calculent ensuite la fonction $IP(x', y')$ avec un protocole déterministe utilisant n boîtes non-locales et utilisent les sorties de ce protocole comme réponses. Analysons la probabilité de succès.

- si $Disj(x, y) = 0$, alors $IP(x', y') = 0$ et le protocole est toujours correct ;
- si $Disj(x, y) = 1$, alors $\text{Prob}_r[IP(x', y') = 1] = \frac{1}{2}$. C'est évident si on identifie les entrées x et y avec des ensembles de $[n]$ dont $i \mapsto x_i$ et $i \mapsto y_i$ sont les fonctions indicatrices. On voit alors que x' est la fonction indicatrice de $x \cap r$ et y' celle de $y \cap r$. La fonction $IP(x', y') = \bigoplus_i x'_i y'_i$ est alors égale à la parité de $x' \cap y' = x \cap y \cap r$, c'est-à-dire la parité d'un sous ensemble aléatoire de $x \cap y$. Comme $x \cap y$ est non vide, on a bien $\text{Prob}_r[IP(x', y') = 1] = \frac{1}{2}$.

Pour obtenir une probabilité de succès strictement plus grande que $1/2$, les joueurs utilisent le protocole précédent et obtiennent les sorties a et b. Ils modifient leurs sorties de la façon suivante. En utilisant l'aléa partagé, ils répondent par les valeurs suivantes : (a, b) avec probabilité $1-p$, $(0, 1)$ avec probabiltité $p/2$, $(1, 0)$ avec probabilité $p/2$. Ainsi, si $Disj(x, y) = 0$, alors la probabilité de succès est $1-p$ et si $Disj(x, y) = 1$, la probabilité de succès est $p + (1-p)/2 = (1+p)/2$. En prenant $p = 1 - 2\varepsilon$, on a une probabilité de succès de $1 - \varepsilon$. □

Les tables 4.1 et 4.2 rappellent tous les résultats obtenus dans cette section et la précédente.

	complexié simultanée	complexité sequentielle
déterministe	$NL^{\parallel}(f) = \text{rang}_2(M_f)$	$D^{\rightarrow}(f) \leq NL(f) \leq NL^{\parallel}(f) \leq 2^{D(f)}$
probabiliste	$NL_{\varepsilon}^{\parallel,\oplus}(f) = \varepsilon\text{−rang}_2(M_f)$	$\log(\gamma_2^{\varepsilon}(M_f)) \leq NL_{\varepsilon}(f) \leq O((\gamma_2^{\infty}(M_f))^2)$

Tab. 4.1 – Résultat de complexité en boîte non-locale

	déterministe		probabiliste	
	$NL^{\parallel}$	NL	$NL_{\varepsilon}^{\parallel}$	NL_{ε}
IP	n	n	?	$\Theta(n)$
EQ	2^n	$O(n)$	?	$O(1)$
$Disj$	2^n	$O(n)$	$O(n)$	$\Theta(n)$

Tab. 4.2 – Complexité de fonctions explicites

4.4 Simuler les corrélations quantiques et au delà

Les comparaisons entre la complexité en boîtes non-locales et la complexité de la communication s'appliquent au modèle traditionnel de la communication. Or dans le

modèle de complexité en boîte non-locale, la valeur de la fonction est la parité des réponses des joueurs. Dans cette section, nous allons comparer le modèle de communication en parité avec la complexité en boîte non-locale. Si on se contente de réutiliser les réductions précédentes, alors il faut ajouter un bit de communication pour passer du modèle en parité au modèle traditionnel, et ensuite simuler le protocole traditionnel. L'effet global de l'ajout d'un bit de communication est de doubler le nombre de boîtes non-locales dans le protocole. Nous allons montrer maintenant qu'il est inutile de repasser au modèle traditionnel et qu'on simuler directement le modèle de communication en parité. Ceci a pour conséquence de diviser par deux le nombre de boîtes non-locales utilisées.

Si on s'intéresse à simuler les protocoles de communication en parité, c'est en particulier dans le but d'utiliser les boîtes non-locales pour simuler les distributions causales. En effet, on a déjà parlé de simuler les distributions causales avec des boîtes non-locales. L'avantage d'un protocole en boîte non-locale est que contrairement à un protocole de communication, il respecte la causalité.

L'idée d'utiliser les boîtes non-locales pour simuler les distributions causales a déjà été envisagée [BP05]. La première application de cette idée est de simuler les corrélations quantiques [CGMP05]. Mais on peut utiliser les boîtes non-locales pour simuler les distributions causales non-quantiques. Nous avons déjà dit que les boîtes non-locales permettaient de simuler toutes les distributions causales binaires lorsque les entrées sont elles-mêmes binaires. Notre objectif ici est de simuler les distributions causales binaires avec marginales uniformes, pour une taille arbitraire de l'espace des entrées. L'ensemble de ces distributions est un polytope, et il suffit donc de simuler les points extrémaux. D'après le théorème 1.11, page 19, ces points sont exactement les distributions correspondant à des fonctions.

Dans le théorème suivant, on compare la complexité en boîte non-locale des distributions causales avec la complexité de la communication. La construction générale est inspirée de celle de Degorre, Laplante et Roland [DLR05] pour simuler les corrélations quantiques issues de mesure sur les qubits avec une boîte non-locale. On utilise de plus une astuce pour économiser une boîte non-locale qui peut être comparée à celle du protocole de Regev et Toner pour passer de 4 bits de communication à 3 [RT07].

Théorème 4.6. *Soit* $\mathbf{p}$ *une distribution causale binaire avec marginales uniformes définie sur* $X \times Y$*. Un protocole de communication de complexité* t *pour* $\mathbf{p}$ *peut être simulé en utilisant au plus* $2^t - 1$ *boîtes non-locales en parallèle.*

Démonstration. Toute distribution causale binaire à marginales uniformes peut s'écrire comme combinaison convexe de points extrémaux du polytope causale. Or, d'après le théorème 1.11, les points extrémaux sont exactement les distributions $\mathbf{p}_f$ correspondant aux fonctions booléennes $f : X \times Y \to \{0, 1\}$, définie à la section 1.3.

Une décomposition $\mathbf{p} = \sum_f \alpha_f \mathbf{p}_f$ implique un protocole pour $\mathbf{p}$. Il suffit de simuler $\mathbf{p}_f$ avec probabilité α_f, la probabilité étant choisie avec l'aléa partagé. Par conséquent, il est toujours possible d'exprimer $\mathbf{p}$ comme combinaison convexe de distributions de complexité au plus t. Il suffit donc de montrer que le théorème est vrai pour les distributions $\mathbf{p}_f$ correspondant aux fonctions booléennes.

Soit $\mathcal{P}$ un protocole de communication de complexité t pour $\mathbf{p}_f$. Si $\mathcal{P}$ est déterministe, on montre comment en déduire un protocole avec boîtes non-locales produisant les même sorties que $\mathcal{P}$. Par conséquent, si $\mathcal{P}$ est un protocole probabiliste, la réduction au proto-

cole en boîte non-locale produit la même distribution sur les sorties, et la correction est identique. On suppose maintenant que $\mathcal{P}$ est déterministe en fixant l'aléa s'il le faut.

Pour simplifier, nous allons prouver d'abord le théorème dans le cas où $\mathcal{P}$ est un protocole à sens unique. Supposons que le message va d'Alice vers Bob. Il y au plus 2^t transcriptions différentes. Notons les explicitement $T_0, \dots T_{2^t-1}$. Le protocole avec des boîtes non-locales est défini de la manière suivante :

- Les joueurs utilisent une boîte pour chaque transcription possible, exceptée T_0.
- Dans la i-ème boîte, Alice entre la fonction $\delta_{T_i}(T(x))$, c'est-à-dire 1 si son entrée donne lieu à la transcription T_i et 0 sinon.
- Dans la i-ème boîte, Bob entre la fonction $B(T_i, y) \oplus B(T_0, y)$, où $B(T, y)$ est sa sortie suivant $\mathcal{P}$ sur l'entrée y et le message T.
- Les boîtes donnent aux joueurs les valeurs (a_i, b_i)
- Alice répond par la valeur $\bigoplus_i a_i \oplus A(x)$, où $A(x)$ est la valeur qu'elle sort suivant $\mathcal{P}$ et l'entrée x.
- Bob répond $\bigoplus_i b_i \oplus B(T_0, y)$.

Ce protocole utilise bien $2^t - 1$ boîtes non-locales. Vérifions qu'il produit bien les même sorties que $\mathcal{P}$. Supposons $T(x) = T_i \neq T_0$, l'entrée de la i-ème boîte est alors $(1, B(T_i, y) \oplus B(T_0, y))$. Pour toutes les autres boîtes, l'entrée est $(0, B(T_1, y) \oplus B(T_j, y))$. On a donc $\sum_i (a_i \oplus b_i) \oplus A(x) \oplus B(T_1, y) = A(x) \oplus B(T_i, y)$.

Supposons maintenant que $T(x) = T_0$. Les entrées des boîtes sont toutes $(0, B(T_i, y) \oplus B(T_0, y))$. On a ainsi $a_i \oplus b_i = 0$ pour tout i. Finalement, on a $\sum_i (a_i \oplus b_i) \oplus A(x) \oplus B(T_0, y) = A(x) \oplus B(T_0, y)$, ce qui est la sortie attendue.

Dans le cas général, on va éliminer les bits de communication les uns après les autres. L'idée est d'examiner tous les scénarios de communication possibles, de t à 0 bits de communication. Par exemple, si on fixe le premier bit de communication, il y a deux scénarios possibles, suivant la valeur de celui-ci. Le nombre de scénarios double ainsi à chaque étape. On va donc considérer :

- des transcriptions de longueur k, notées $T^{(k)}$,
- 2^{t-k} protocoles de communications. On note les sorties de ceux-ci $A_i^{(k)}(T^{(k)}, x)$ et $B_i^{(k)}(T^{(k)}, y)$, où $i = 1, \dots 2^{t-k}$.

Au départ, on a un protocole qui calcule $f(x, y)$ en parité :

$$f(x, y) = A_0^{(t)}(T^{(t)}, x) \oplus B_0^{(t)}(T^{(t)}, y).$$

On va construire des protocoles tels que pour tout k allant de t à 0 :

$$f(x, y) \;=\; A_0^{(k)}(T^{(k)}, x) \oplus B_0^{(k)}(T^{(k)}, y) \oplus \bigoplus_{i=1}^{2^{t-k}-1} A_i^{(k)}(T^{(k)}, x) B_i^{(k)}(T^{(k)}, y), \quad (4.4)$$

ce qui peut être calculé avec k bits de communication et $2^{t-k} - 1$ boîtes non-locales. En particulier pour $k = 0$, on obtient :

$$f(x, y) \;=\; A_0^{(0)}(x) \oplus B_0^{(0)}(y) \oplus \bigoplus_{i=1}^{2^t-1} A_i^{(0)}(x) B_i^{(0)}(y),$$

ce qui peut être calculé avec $2^k - 1$ boîtes non-locales et pas de communication.

Montrons maintenant comment retirer un bit de communication. Supposons qu'on a 2^{t-k} protocoles de communication de complexité t et tels que l'équation 4.4 est vraie. Considérons les transcriptions $T^{(k)}$ dans lesquels le dernier bit de la communication va d'Alice vers Bob. Le dernier bit de $T^{(k)}$ étant calculé par Alice, notons le $c_i^{(k)} = c_i^{(k)}(T^{(k-1)}, x)$. On peut décomposer la sortie de Bob en fonction de la valeur de ce bit :

$$B_i^{(k)}(T^{(k)}, y) = B_i^{(k)}(T^{(k-1)}0, y) \oplus c_i^{(k)}[B_i^{(k)}(T^{(k-1)}0, y) \oplus B_i^{(k)}(T^{(k-1)}1, y), x].$$

On a donc pour $f(x, y)$:

$$\begin{aligned} f(x,y) &= A_0^{(k)}(T^{(k-1)}, x) \oplus B_0^{(k)}(T^{(k-1)}0, y) \\ &\quad \oplus c_0^{(k)}(T^{(k-1)}, x)[B_0^{(k)}(T^{(k-1)}0, y) \oplus B_0^{(k)}(T^{(k-1)}1, y)] \\ &\quad \oplus \bigoplus_{i=1}^{2^{t-k}-1} A_i^{(k)}(T^{(k-1)}, x) B_i^{(k)}(T^{(k-1)}0, y) \\ &\quad \oplus \bigoplus_{i=1}^{2^{t-k}-1} A_i^{(k)}(T^{(k-1)}, x) c_i^{(k)}[B_i^{(k)}(T^{(k-1)}0, y) \oplus B_i^{(k)}(T^{(k-1)}1, y)]. \end{aligned}$$

Pour $0 \leq i \leq 2^{t-k} - 1$, on définit les fonctions suivantes :

$$\begin{aligned} A_i^{(k-1)}(T^{(k-1)}, x) &= A_i^{(k)}(T^{(k-1)}, x) \\ B_i^{(k-1)}(T^{(k-1)}, y) &= B_i^{(k)}(T^{(k-1)}0, y) \\ B_{i+2^{t-k}}^{(k-1)}(T^{(k-1)}, y) &= B_i^{(k)}(T^{(k-1)}0, y) \oplus B_i^{(k)}(T^{(k-1)}1, y) \\ A_{2^{t-k}}^{(k-1)}(T^{(k-1)}, x) &= c_0^{(k)}(T^{(k-1)}, x), \\ A_{i+2^{t-k}}^{(k-1)}(T^{(k-1)}, x) &= A_i^{(k)}(T^{(k-1)}, x) c_i^{(k)} \quad (\text{si } i \neq 0), \end{aligned}$$

et les expressions similaires lorsque la communication va de Bob vers Alice. Enfin, ceci nous permet bien d'écrire :

$$\begin{aligned} f(x,y) &= A_0^{(k-1)}(T^{(k-1)}, x) \oplus B_0^{(k-1)}(T^{(k-1)}, y) \\ &\quad \oplus \bigoplus_{i=1}^{2^{t-k+1}-1} A_i^{(k-1)}(T^{(k-1)}, x) B_i^{(k-1)}(T^{(k-1)}, y) \end{aligned}$$

comme annoncé. □

En particulier, on peut combiner le résultat précédent avec le protocole de Regev et Toner [RT07] pour simuler les distributions de probabilité quantique.

Corollaire 4.4. *Soit* $\mathbf{p}$ *une distribution de probabilité booléenne à marginales uniformes. Supposons* $\mathbf{p} \in \mathcal{Q}$. *Alors* $NLB(\mathbf{p}) \leq 3$.

Démonstration. Rappelons le principe du protocole de Regev et Toner pour simuler les distributions quantiques. En utilisant le théorème de Tsirelson, le problème se réduit à celui-ci :

– Alice reçoit un vecteur unitaire $a \in \mathbb{R}^d$.

– Bob reçoit un vecteur unitaire $b \in \mathbb{R}^d$.
– Les joueurs doivent répondre par des valeurs $(\alpha, \beta) \in \{-1, 1\} \times \{-1, 1\}$ telles que $\mathbb{E}\alpha\beta = \langle a, b \rangle$.

Soit G une matrice de projection sur un sous espace aléatoire de dimension 3 de $\mathcal{R}^d$. Les joueurs commencent par appliquer une transformation C que nous ne détaillons pas ici, puis projettent leurs entrées en utilisant G. Soient a' et b' les vecteurs obtenus. Soit $\alpha_i = sgn(a'_i)$ pour $i = 1, 2, 3$, et $c_i = \alpha_1\alpha_i$ pour $i = 2, 3$. La sortie d'Alice est α_1 et elle envoie (c_2, c_3) à Bob. Intuitivement, elle dit à Bob dans quel orthant se trouve son vecteur.

Comme il y a 8 orthant, cela nécessite en principe 3 bits de communication, mais Alice peut économiser un bit de la manière suivante : au lieu de situer son vecteur parmi les 8 orthants possibles, elle se limite à 4 orthants, quitte à changer le signe de a'. Sa sortie indique précisément si elle a changé le signe ou non. Soit $z = (1, c_1, c_2)$. La sortie de Bob est $sgn(\langle b', z \rangle)$.

Sans la transformation C préliminaire, les corrélations calculées par ce protocole sont trop fortes. Cette transformation permet précisément de trouver les bonnes corrélations.

Ce protocole utilise 2 bits de communication. En appliquant le théorème 4.6, on en déduit que 3 boîtes non-locales sont suffisantes. □

4.5 Evaluation sécurisée

Comme nous l'avons dit plus haut, le calcul avec des boîtes non-locales permet de réaliser une tâche cryptographique appelée évaluation sécurisée. Ce problème de cryptographie a été introduit par Yao [Yao82]. L'auteur présente le problème de la manière suivante : deux millionnaires veulent comparer leurs fortunes pour savoir qui est le plus riche, mais aucun ne veut avoir à révéler à l'autre le montant de ses richesses.

Contrairement à la situation en complexité de la communication, les deux joueurs ne se font pas confiance. Il y a deux modèles de sécurité différents. Dans le modèle "Honnête mais curieux", on suppose que les deux joueurs suivent le protocole imposé. Toutefois, à l'issue du protocole, les joueurs doivent avoir appris aussi peu d'information que possible. Dans le modèle "malhonnête", on suppose que les joueurs peuvent tricher afin d'obtenir de l'information. De même, les joueurs doivent avoir appris à la fin du protocole, aussi peu d'information que possible. Aussi peu que possible signifie que le joueur qui donne la valeur de la fonction ne doit rien apprendre d'autre que ce qu'il peut déduire de cette valeur, et l'autre joueur n'apprend rien du tout.

Cette définition de la sécurité est très forte, et on sait que dans ce sens, il existe des fonctions qui ne peuvent être calculées de manière sécurisée [BOGW88, CCD88, CK91, Kus92]. L'évaluation devient en revanche possible si les joueurs ont accès à une primitive sécurisée telle que ET sécurisée, ou le "oblivious transfer" [GV88, Kil88]. Nous suivons ici l'approche de Beimel et Malkin qui consiste à quantifier le nombre d'utilisation de ces primitives [BM04]. Ceci permet de hiérarchiser les fonctions d'après le nombre d'appels à ces primitives nécessaire pour les évaluer de manière sécurisée. L'existence d'une telle hiérarchie est la conséquence d'un résultat de Beaver montrant qu'il existe des fonctions qui peuvent être évaluées avec k appels au ET sécurisé, mais pas avec $k-1$ appels[Bea96].

Voici les primitives que nous allons étudier dans la suite.

Définition 4.7. *Un ET sécurisé est une ressource de calcul dans laquelle Alice introduit un bit x et Bob un bit y et telle qu'à la fin, Alice reçoit un bit a tel que $a = xy$.*

Définition 4.8. *Un "oblivious transfer" (OT) est une ressource de calcul dans laquelle Alice entre deux bits x_0 et x_1, et Bob un bit i et telle qu'à la fin Bob reçoit un bit b tel que $b = x_i$.*

Bien entendu, ces deux ressources sont sécurisées dans le sens où le joueur qui reçoit de l'information n'apprend rien d'autre que celle-ci. Ces deux ressources ressemblent intuitivement à une boîte non-locale. Le ET sécurisé calcule, comme une boîte non-locale, le produit de deux bits. La différence est que le résultat n'est pas donné en parité, mais directement à Alice. Cette différence est importante car nous allons pouvoir caractériser exactement le nombre de ET sécurisés nécessaire et suffisants pour évaluer une fonction.

Le lien entre boîtes non-locales et cryptographie ont déjà été étudiés. Notamment, l'équivalence entre boîte non-locale et OT a été montrée par Wolf et Wullschleger [WW05]. Toutefois, les conditions de sécurité ne sont pas les mêmes, et pour qu'elles soient conservées, on peut supposer que les boîtes sont utilisées par les joueurs dans le même ordre. Sans cette hypothèse, la simulation de OT par des boîtes non-locales est plus complexe. Un protocole pour réaliser une boîte OT en utilisant des boîtes non-locales a été proposé par Buhrman et al. [BCU$^+$06]. Cette construction étend celle de Wolf et Wullschleger en tenant compte des propriétés temporelles des boîtes non-locales.

4.5.1 La primitive ET dans le modèle "Honnête mais curieux" déterministe

La première primitive qu'on considère est le ET sécurisé. On se place ici dans le modèle "honnête mais curieux". On suppose de plus que le calcul est déterministe. Pour une fonction f, on note $AND(f)$ le plus petit nombre de ET sécurisés permettant de calculer la fonction f dans ce modèle.

Rappelons que pour les boîtes non-locales, nous avons donné des bornes inférieures et supérieures, avec un écart exponentiel entre les deux. Notre résultat sur le nombre de ET résout un problème posé par Beimel et Malkin [BM04]. Ceux-ci avaient précédemment montré que pour une fonction $f : X \times Y \to \{0,1\}$, on a $AND(f) \leq |X|$.

Théorème 4.7. *Pour toute fonction booléenne $f : X \times Y \to \{0,1\}$, $AND(f) = 2^{D^{\to}(f)}$.*

Démonstration. Soit $\mathcal{P}$ un protocole de communication à sens unique de complexité t pour f. Sur l'entrée x, on note $T(x)$ le message envoyé par Alice et $A(x)$ sa sortie. On note $B(T(x), y)$ la sortie de Bob. On définit maintenant un protocole utilisant 2^t ET sécurisés. On utilise un ET pour chaque $m \in \{0,1\}^t$. Dans le m-ième ET, Alice entre 1 si $T(x) = m$ et 0 sinon. Bob entre $B(m, y)$. Soit a_m la sortie de la porte ET reçue par Alice. On a $a_{T(x)} = B(T(x), y)$, et $a_m = 0$ pour $m \neq T(x)$. Alice sort donc la valeur $A(x) \oplus a_{T(x)}$. La sortie est la même que dans le protocole de communication, donc ce protocole est correct. Comme Bob ne reçoit aucune donnée de la part d'Alice, la sécurité de l'entrée d'Alice est triviale. A la fin du protocole, Alice connaît seulement $B(T(x), y)$, ce qu'elle peut calculer en connaissant x et $f(x, y)$. La sécurité de l'entrée de Bob est donc également assurée.

Inversement, soit $\mathcal{P}$ un protocole pour f utilisant t ET. D'après Beimel et Malkin, on peut supposer sans perte de généralité qu'il n'y a aucune communication (cela contredirait l'hypothèse de sécurité). On note (p_i, q_i) les entrées des t ET et $a_i = p_i q_i$ les sorties. Bob ne recevant aucune information, ses entrées q_i ne dépendent que de y. Montrons qu'il en est de même pour Alice. Soit $\mathbf{a} = (a_1, \dots a_t)$ les sorties des portes AND. Pour Alice, l'hypothèse de sécurité est qu'elle n'apprend que la valeur de $f(x, y)$. Le protocole étant déterministe, x étant fixé, la valeur de $\mathbf{a}$ ne dépend que de la valeur de la fonction, sans quoi Alice apprendrait quelque chose de plus sur y, l'entrée de Bob. Notons $\mathbf{a}^0(x)$ la valeur de $\mathbf{a}$ pour $f(x, y) = 0$ et $\mathbf{a}^1(x)$ sa valeur pour $f(x, y) = 1$. Soit i le premier indice tel que $a_i^0(x) \neq a_i^1(x)$. Pour $j < i$, On a $a_j^0(x) = a_j^1(x)$ et Alice connaît donc ces valeurs. On peut donc supposer sans perte de généralité que $p_j = 0$ pour tout j de 1 à $i - 1$. De plus, dans la i-ème porte, l'entrée d'Alice est nécessairement 1, sinon la sortie serait constante égale à 0. La valeur de a_i lui permet donc de distinguer entre $\mathbf{a}^0(x)$ et $\mathbf{a}^1(x)$. En particulier, elle n'a pas besoin des valeurs a_j pour $j > i$. On peut donc supposer que $p_j = 0$ pour tout j de $i + 1$ à t.

On a transformé le protocole initial en un protocole dans lequel $p_i = 0$ pour tout sauf un seul indice i. Cet indice ne dépend que de x. Pour simuler ce protocole avec de la communication, il suffit à Alice d'envoyer à Bob l'indice i sur lequel p_j vaut 1, utilisant $\log t$ bits de communication. □

4.5.2 La primitive OT dans le modèle "malicieux" (probabiliste)

On va maintenant s'intéresser au modèle malicieux, où les joueurs peuvent tricher pour tenter d'obtenir des informations sur l'entrée de l'autre joueur. Dans ce modèle, les joueurs ne peuvent pas utiliser le ET sécurisé, car Alice n'aurait qu'à entrer 1 dans tous les ET pour obtenir de l'information sur l'entrée de Bob. On considère donc la primitive OT.

Par ailleurs, on sait que l'évaluation à la fois sécurisée et déterministe est impossible dans le modèle malicieux [DM99]. Le modèle est donc défini comme suit. Chaque joueur reçoit une entrée, respectivement x pour Alice et y pour Bob. Ils peuvent utiliser de l'aléa privé et public, une quantité illimitée de communication, et de boîtes OT. On suppose que c'est Bob qui donne la valeur de la fonction. Sa réponse doit être correcte avec probabilité $1 - \varepsilon$ où ε est une constante comprise entre 0 et 1/2. On note $OT_\varepsilon(f)$ le nombre d'utilisations de OT dans ce modèle.

On va d'abord prouver une borne supérieure sur le nombre d'utilisations de OT en fonction de la complexité en boîte non-locale. Pour que le protocole en boîte non-locale soit sécurisé, on ajoute une contrainte supplémentaire. On suppose en effet que les boîtes sont utilisées dans le même ordre par les deux joueurs. On note la complexité $NL_\varepsilon^{\text{ord}}$ comme cela a été défini à la section 4.1.2.

Théorème 4.8. *Pour toute fonction booléenne $f : X \times Y \to \{0, 1\}$ et $\varepsilon \geq 0$, $OT_\varepsilon(f) \leq NL_\varepsilon^{ord}(f)$.*

Démonstration. Soit $\mathcal{P}$ un protocole pour f utilisant t boîtes non-locales dans le même ordre pour les 2 joueurs. On note les entrées des joueurs dans les boîtes $(p_1, q_1), \dots, (p_t, q_t)$

et obtiennent les sorties $(a_1, b_1), \ldots, (a_t, b_t)$. Par définition, chaque entrée (p_i, q_i) ne peut dépendre que des entrées et sorties précédentes, de 1 à $i-1$, mais pas des suivantes.

On va maintenant remplacer chaque boîte non-locale par une boîte OT, en commençant par la première. Alice tire un bit aléatoire r_1 et entre dans la boîte OT r_1 et $r_1 \oplus p_1$. Bob entre q_1 et reçoit donc $(1 \oplus q_1)r_1 \oplus q_1(r_1 \oplus p_1) = r_1 \oplus p_1 q_1$. Les joueurs simulent ensuite le protocole utilisant des boîtes non-locales en posant $a_1 = r_1$ et $b_1 = r_1 \oplus p_1 q_1$. Ils procèdent ainsi pour toutes les boîtes non-locales du protocole.

Pour chaque boîte non-locale, Alice et Bob peuvent calculer les entrées p_i, q_i car celles-ci ne dépendent que des entrées et sorties précédentes. La distribution qu'on obtient à la fin est la même que dans le protocole original. La probabilité de calculer la valeur de f est donc inchangée.

Il reste à prouver que le protocole utilisant les boîtes OT est bien sécurisé. La sécurité de l'entrée de Bob est évidente car Bob n'envoie aucune information à Alice. La sécurité de l'entrée d'Alice vient du fait que les informations qu'elle donne sur sur son entrée sont à chaque étape masquées par un bit aléatoire. Les sorties que Bob obtient des boîtes OT sont donc uniquement des bits aléatoires indépendants. □

Avant de prouver la borne inférieure, détaillons les notations des différentes étapes du protocole. Chaque étape se déroule ainsi. Supposons qu'on est à l'étape i. Alice envoie un message A_i à Bob, et Bob envoie B_i à Alice. Ils utilisent ensuite une boîte OT. On note $S_i = (S_i^0, S_i^1)$ l'entrée d'Alice et T_i l'entrée de Bob. La sortie, du côté de Bob, est notée O_i. Lorsqu'on voudra noter la concaténation des données des i premieres rondes, on utilisera l'indice $[i]$.

La borne inférieure est en fait sur un modèle de sécurité plus fort que le modèle présenté plus haut. Commençons par définir le modèle formellement.

Définition 4.9. *Soit $f : X \times Y \to \{0,1\}$ un fonction booléenne. $\widehat{OT}_\varepsilon(f)$ est le nombre de boîtes OT nécessaires pour calculer f avec probabilité ε, sécurité parfaite, en utilisant une quantité illimitée de communication, dans le modèle malicieux et verifiant les conditions de sécurité suivantes :*

$$\mathrm{Prob}[A_i | A_{[i-1]}, B_{[i-1]}, S_{[i]}, x, r] = \mathrm{Prob}[A_i | A_{[i-1]}, B_{[i-1]}, r],$$

$$\mathrm{Prob}[B_i | A_{[i]}, B_{[i-1]}, T_{[i]}, O_{[i]}, y, r] = \mathrm{Prob}[B_i | A_{[i-1]}, B_{[i-1]}, r],$$

où r est l'aléa publique partagé par les joueurs.

Intuitivement, nous pensons que cette condition est vraie pour un protocole optimal. Nous n'avons pas pu prouver ce point, mais voyons plus précisément ce que signifie la condition. Du côté d'Alice, la condition s'interprète ainsi. Sachant la communication passée et l'aléa, le message A_i envoyé par Alice est indépendant de $(S_{[i]}, x)$. Si ce n'est pas le cas, alors Bob peut savoir avec une probabilité non nulle, en fonction du message reçu A_i, quel élément parmi $(S_{[i]}, x)$ et $(S'_{[i]}, x')$ Alice a utilisé.

Si $x \neq x'$, alors Bob a obtenu une information sur l'entrée d'Alice, ce qui est interdit par la définition de la correction. De même, si $S_{[i]} \neq S'_{[i]}$, alors Bob pourrait obtenir des informations sur les entrées des OT. En choisissant un boîte au hasard, et en entrant un bit au hasard, la probabilité pour Bob d'obtenir des informations n'est pas nulle. Ainsi, si dans un protocole optimal, Bob n'a pas le droit d'obtenir d'information sur les valeurs entrées par Alice dans les OT, alors cette définition de la sécurité est équivalente à la définition usuelle.

Théorème 4.9. *Pour toute fonction booléenne* $f : X \times Y \to \{0,1\}$, $\widehat{OT}_\varepsilon(f) \geq \Omega(R_\varepsilon(f))$

Démonstration. Soit $t = \widehat{OT}_\varepsilon(f)$. Le modèle autorisant une quantité illimitée de communication, on doit prouver qu'on peut réduire celle-ci à $O(t)$ bits. On prouve cette restriction en 4 étapes :

1. On commence par supprimer la communication allant d'Alice vers Bob. Ceci nous donne un protocole parfaitement sûr, utilisant t boîtes OT et où toute la communication va de Bob vers Alice.

2. On inverse les boîtes OT en introduisant t bits de communication allant d'Alice vers Bob. Ceci donne un protocole parfaitement sûr, où les sorties des t boîtes vont à Alice, t bits de communication vont d'Alice vers Bob et une quantité illimitée va de Bob vers Alice.

3. On réitère le premier point afin de supprimer cette fois la communication allant de Bob vers Alice. Il reste à la fin de cette étape t boîtes OT et t bits de communication allant d'Alice vers Bob.

4. Enfin, les joueurs simulent les boîtes OT en envoyant $2t$ bits de communication, donnant finalement un protocole avec en tout $3t$ bits.

On va détailler par la suite les étapes 1 et 3. La preuve pour ces 2 parties est la même. Cela nécessite qu'à la fin de l'étape 2, les conditions de sécurité soient les mêmes qu'avant l'étape 1. Pour l'étape 2, on utilise la construction de [WW06] qui utilise les boîtes nonlocales pour inverser le sens de OT. Celle ci permet en effet de préserver comme c'est nécessaire la sécurité de l'entrée de Bob. L'étape 4 en revanche ne pose aucune difficulté, les entrées sont simplement échangées entre les joueurs.

Détaillons maintenant la première étape. L'idée est de remplacer la communication d'Alice vers Bob par des bits aléatoires partagés. Les messages de Bob ne changent pas, ni les entrées des boîtes OT. L'aléa privé de Bob peut être fixé au début de l'exécution du protocole. Alice va commencer le protocole avec une distribution uniforme sur son aléa privé et après chaque étape, elle va actualiser sa distribution pour rester consistant avec l'historique du protocole. Commençons par préciser le déroulement et les notations du protocole :

- Alice et Bob choisissent leur aléa r_A et r_B.
- A la i-ème étape, Alice entre S_i et Bob T_i dans la boîte OT et Bob reçoit la sortie O_i. L'entrée d'Alice est une fonction de $(A_{[i-1]}, B_{[i-1]}, S_{[i-1]}, x, r, r_A)$ et celle de Bob, une fonction de $A_{[i-1]}, B_{[i-1]}, T_{[i-1]}, O_{[i-1]}, y, r, r_B)$.
- Alice envoie son message A_i à Bob. C'est une fonction de $(A_{[i-1]}, B_{[i-1]}, S_{[i]}, x, r, r_A)$.
- Bob envoie son message B_i à Alice. C'est une fonction de $(A_{[i]}, B_{[i-1]}, T_{[i]}, O_{[i]}, y, r, r_B)$.

Regardons maintenant la distribution que le protocole induit sur r_A, r_B, A, B, S, T, O

sachant x, y, r :

$$\begin{aligned}
&\text{Prob}[r_A, r_B, A, B, S, T, O | x, y, r] = \\
&\quad \text{Prob}[r_A].\text{Prob}[r_B].\prod_i (\text{Prob}[S_i | A_{[i-1]}, B_{[i-1]}, S_{[i-1]}, x, r, r_A]. \\
&\quad \text{Prob}[T_i | A_{[i-1]}, B_{[i-1]}, T_{[i-1]}, O_{[i-1]}, y, r, r_B].\text{Prob}[O_i | S_i, T_i]. \\
&\quad \text{Prob}[A_i | A_{[i-1]}, B_{[i-1]}, S_{[i]}, x, r, r_A]. \\
&\quad \text{Prob}[B_1 | A_{[i-1]}, B_{[i-1]}, T_{[i]}, O_{[i]}, y, r, r_B] \ ).
\end{aligned}$$

Voici maintenant le nouveau protocole. Celui-ci est conçu pour que la distribution reste identique, et pour y parvenir, Alice doit réactualiser son aléa en fonction des messages envoyés.

- Alice et Bob choisissent leur aléa r_A et r_B.
- A la i-ème étape, Alice choisit son entrée S_i dans la boîte suivant la distribution :

  $$\text{Prob}[S_i | A_{[i-1]}, B_{[i-1]}, S_{[i-1]}, x, r].$$

  Bob entre la même fonction de $(A_{[i-1]}, B_{[i-1]}, T_{[i-1]}, O_{[i-1]}, y, r, r_B)$ que dans le protocole original.
- En utilisant l'aléa partagé, les joueurs tirent un message suivant la distribution :

  $$\text{Prob}[A_i | A_{[i-1]}, B_{[i-1]}, r],$$

  c'est-à-dire un message consistant avec l'aléa partagé r et l'historique des messages, en moyenne sur l'aléa privé r_A d'Alice. L'idée est que d'après la définition 4.9, la distribution sur A_i ne dépend ni de x ni de r_A.
- Bob envoie le même message que dans le protocole original.
- Alice actualise son aléa r_A en tirant une nouvelle chaîne suivant la distribution $\text{Prob}[r_A | A_{[i]}, B_{[i]}, S_{[i]}, x, r]$.

Nous allons dans un premier temps vérifier que la distribution est la même que dans le protocole original. Dans un second temps, nous vérifierons que le protocole est bien défini. On a dans le nouveau protocole la distribution suivante :

$$\begin{aligned}
\text{Prob}[r_A, r_B, A, B, S, T, O | x, y, r] &= \text{Prob}[r_A].\text{Prob}[r_B].\prod_{i \in [t]} \text{Prob}[S_i | A_{[i-1]}, B_{[i-1]}, S_{[i-1]}, x, r] \\
&\text{Prob}[T_i | A_{[i-1]}, B_{[i-1]}, T_{[i-1]}, O_{[i-1]}, y, r, r_B].\text{Prob}[O_i | S_i, T_i].\text{Prob}[A_i | A_{[i-1]}, B_{[i-1]}, r] \\
&.\text{Prob}[B_i | A_{[i]}, B_{[i-1]}, T_{[i]}, O_{[i]}, y, r, r_B].\frac{\text{Prob}[r_A | A_{[i]}, B_{[i]}, S_{[i]}, x, r]}{\text{Prob}[r_A | A_{[i-1]}, B_{[i-1]}, S_{[i-1]}, x, r]}.
\end{aligned}$$

De plus après l étapes la distribution sur r_A est exactement :

$$\text{Prob}[r_A].\prod_{i \in [\ell]} \frac{\text{Prob}[r_A | A_{[i]}, B_{[i]}, S_{[i]}, x, r]}{\text{Prob}[r_A | A_{[i-1]}, B_{[i-1]}, S_{[i-1]}, x, r]} = \text{Prob}[r_A | A_{[\ell]}, B_{[\ell]}, S_{[\ell]}, x, r].$$

Pour montrer que les deux distributions sont les mêmes, il suffit maintenant de prouver l'énoncé suivant :

$$\begin{aligned}&\mathrm{Prob}[S_i|A_{[i-1]},B_{[i-1]},S_{[i-1]},x,r,r_A].\mathrm{Prob}[A_i|A_{[i-1]},B_{[i-1]},S_{[i]},x,r,r_A]=\\&\quad\mathrm{Prob}[S_i|A_{[i-1]},B_{[i-1]},S_{[i-1]},x,r].\mathrm{Prob}[A_i|A_{[i-1]},B_{[i-1]},r]\\&\quad.\frac{\mathrm{Prob}[r_A|A_{[i]},B_{[i]},S_{[i]},x,r]}{\mathrm{Prob}[r_A|A_{[i-1]},B_{[i-1]},S_{[i-1]},x,r]}.\end{aligned}$$

On a précisément :

$$\begin{aligned}&\mathrm{Prob}[S_i|A_{[i-1]},B_{[i-1]},S_{[i-1]},x,r,r_A]\cdot\mathrm{Prob}[A_i|A_{[i-1]},B_{[i-1]},S_{[i]},x,r,r_A]\\&=\ \mathrm{Prob}[S_i,A_i|A_{[i-1]},B_{[i-1]},S_{[i-1]},x,r,r_A]\\&=\ \frac{\mathrm{Prob}[S_i,A_i,r_A|A_{[i-1]},B_{[i-1]},S_{[i-1]},x,r]}{\mathrm{Prob}[r_A|A_{[i-1]},B_{[i-1]},S_{[i-1]},x,r]}\\&=\ \mathrm{Prob}[S_i,A_i|A_{[i-1]},B_{[i-1]},S_{[i-1]},x,r]\cdot\frac{\mathrm{Prob}[r_A|A_{[i]},B_{[i-1]},S_{[i]},x,r]}{\mathrm{Prob}[r_A|A_{[i-1]},B_{[i-1]},S_{[i-1]},x,r]}\\&=\ \mathrm{Prob}[S_i|A_{[i-1]},B_{[i-1]},S_{[i-1]},x,r]\cdot\mathrm{Prob}[A_i|A_{[i-1]},B_{[i-1]},r]\\&\quad.\frac{\mathrm{Prob}[r_A|A_{[i]},B_{[i]},S_{[i]},x,r]}{\mathrm{Prob}[r_A|A_{[i-1]},B_{[i-1]},S_{[i-1]},x,r]},\end{aligned}$$

où on a utilisé les deux propriétés suivantes :

– D'abord, $A_{[i]}$ et $S_{[i]}$ étant fixés, $B_{[i]}$ est indépendant de r_A. On a donc

$$\mathrm{Prob}[r_A|A_{[i]},B_{[i-1]},S_{[i]},x,r]=\mathrm{Prob}[r_A|A_{[i]},B_{[i]},S_{[i]},x,r].$$

– D'après la définition de sécurité 4.9, on a :

$$\mathrm{Prob}[A_i|A_{[i-1]},B_{[i-1]},S_{[i]},x,r]=\mathrm{Prob}[A_i|A_{[i-1]},B_{[i-1]},r].$$

Il reste à vérifier que le protocole est bien défini. Il s'agit, lorsqu'Alice actualise son aléa, d'être sûr que sa probabilité de trouver une chaîne consistante est non nulle. Cela découle de l'hypothèse de sécurité. Supposons en effet qu'après la $(i-1)$-ème étape, l'aléa d'Alice r_A est bien consistant avec l'historique. A la i-ème étape, Alice choisit S_i suivant toutes les entrées possibles S_i pour lesquels il existe une chaîne r_A consistante. Puis, A_i est choisi parmi tous les messages pour lesquels, l'aléa r étant choisi, il existe une question x et un ensemble d'entrées $S_{[i]}$, tel qu'il existe une chaîne r_A rendant A_i consistant avec $(A_{[i-1]},S_{[i]},r_A,r,x)$. La sécurité implique que s'il existe x et $S_{[i]}$ pour lesquels $A_{[i]}$ est consistant avec r_A, alors $A_{[i]}$ doit être consistant avec n'importe quel autre x,r_A et $S_{[i]}$. Sinon, cela permettrait à Bob d'apprendre quelque chose sur x ou $S_{[i]}$. Ainsi, il existe toujours un choix de r_A consistant avec le protocole. Ceci termine la preuve du point 1, et par conséquent de la preuve entière. □

4.6 Conclusion

Dans cette section, nous avons étudié le calcul des fonctions booléennes avec des boîtes non-locales. L'objectif de notre travail était de comprendre leur puissance calculatoire.

L'étude des différences entre le calcul quantique et le calcul utilisant des boîtes non-locales peut permettre de mieux comprendre le premier. A travers cette question, ce sont donc à nouveau les processus de traitement de l'information mis en oeuvre par la physique quantique qu'on chercher à étudier. Le calcul avec des boîtes non-locales présente deux autres avantages. D'abord, c'est un calcul qui respecte la causalité. De ce point de vue, il semble plus naturel d'utiliser des boîtes non-locales que de la communication pour simuler des processus causaux. Ensuite, le calcul avec des boîtes non-locales permet de réaliser une importante tâche cryptographique, l'évaluation sécurisée.

Dans le modèle déterministe, nous avons d'abord caractérisé la complexité en boîte non-locale simultanée en fonction du rang de la matrice de communication. Dans ce cas, on a des fonctions dont la complexité est exponentielle. Lorsqu'on autorise l'utilisation séquentielle des boîtes non-locales, les fonctions pour lesquelles on avait des bornes inférieures exponentielles deviennent linéaires. Nous n'avons pas pu identifier de fonction dont la complexité en boîte non-locale séquentielle est strictement plus que linéaire.

Dans le cas probabiliste, nous avons donné des bornes inférieures et supérieures. En particulier, nous avons montré que plusieurs bornes supérieures connues pour la complexité de la communication simultanée étaient en fait des bornes sur la complexité en boîte non-locale. Intuitivement, la différence entre un bit de communication et une boîte non-locale est que le premier permet de transmettre de l'information et pas le deuxième. Pourtant, notre approche n'a pas permis de montrer de séparation entre complexité en boîte non-locale et complexité de la communication.

Nous avons ensuite étudié l'évaluation sécurisée. Dans le modèle "honnête mais curieux", nous avons caractérisé exactement le nombre de ET sécurisés nécessaires et suffisants pour calculer une fonction de manière sécurisée. Comme une boîte non-locale, cette ressource calcule le produit des entrées. Toutefois, le résultat du calcul avec un ET est donné directement à un joueur, ce qui semble être une différence importante. Dans le cas probabiliste, nous avons donné des bornes inférieures et supérieures sur le nombre de boîtes OT nécessaires et suffisantes pour réaliser l'évaluation sécurisée. Les boîtes non-locales ont été utiles dans les deux cas. Pour la borne inférieure, elles ont permis de réaliser l'inversion des OT. Dans ce dernier cas, nous avons été obligé de faire des hypothèses supplémentaires sur la sécurité, hypothèse que nous n'avons pas pu simplifier.

Conclusion

Les frontières entre les sciences ont toujours quelque chose d'arbitraire. Toutes ont pour finalité d'améliorer la connaissance du monde qui nous entoure. Toutefois, on peut voir des différences dans les méthodologies. Ainsi, l'informatique et la physique ont beau avoir une origine commune, elles présentent néanmoins une importante différence dans leur rapport à la théorie. En physique, on élabore des théories d'abord, puis on conçoit des expérimentations pour tenter de les confirmer ou les infirmer. En informatique, on a tendance à considérer que les théories ont pour but de concevoir des procédés de calculs dont la finalité est avant tout dans ses applications. Il y a probablement d'autres différences entre ces deux disciplines. et d'ailleurs s'il n'en existait pas, on n'aurait pas de raison de vouloir les distinguer.

Au centre de notre travail se trouve la complexité de la communication. Ce modèle de calcul n'a probablement pas vocation à être réalisé expérimentalement, mais a néanmoins de nombreuses applications à d'autres modèles de calcul. En cherchant à améliorer cette discipline particulière, c'est toutes les disciplines connexes que nous avons entrepris d'enrichir. Ces disciplines sont des branches d'importance diverse de l'informatique, mais aussi de la physique. C'est un exemple de la beauté de la science de permettre de se placer au centre de problématiques informatiques, d'y appliquer des résultats mathématiques parfois très profonds, tout en gardant un oeil sur la physique.

Rappelons les thèmes que nous avons abordés ici. Le modèle de calcul sur lequel nous avons travaillé est la complexité de la communication. Ce modèle est très souple et admet de nombreuses variantes. L'une des questions qui a motivé notre travail est de chercher à savoir dans quel cas les ressources quantiques font baisser exponentiellement la complexité de la communication. Pour prouver une séparation entre la communication classique et la communication quantique, il faut être capable de prouver des bornes inférieures spécifiques au modèle probabiliste, c'est-à-dire plus grande que la complexité de la communication quantique.

Nous avons donc tenter de classifier les méthodes existantes afin de comprendre lesquels pouvaient permettre de prouver de telles séparations. Cette classification a permis de distinguer deux groupes de méthodes : les méthodes combinatoires d'une part, qui considèrent le modèle de la complexité de la communication comme un processus de traitement de l'information, et les méthodes algébriques d'autre part, qui opèrent sur les représentations mathématiques des problèmes.

Ces problèmes nous ont amené à considérer un nouveau modèle de calcul, la complexité en boîte non-locale. L'utilisation de cette ressource, plus forte que l'intrication mais néanmoins causale, rend trivial tout problème de communication. Nous avons cherché à comprendre plus profondément l'utilisation des boîtes non-locales comme ressource de calcul. Au centre de cette question se trouve celle, plus générale, des liens entre la phy-

sique et l'informatique.

Dans un premier temps, nous avons analysé la communication comme un problème combinatoire. Nous avons utilisé pour cela la complexité de Kolmogorov, un outil qui a pour but de donner une caractérisation algorithmique de l'aléatoire. Notre ambition était de donner une alternative combinatoire aux méthodes basées sur la théorie de l'information. Pour y parvenir, nous avons formellement étendu à la complexité de la communication la méthode dite d'incompressibilité, qui a servi à prouver des bornes inférieures dans divers modèles de complexité.

Notre méthode générale se décompose en trois points. D'abord, nous avons donné une borne inférieure générale, basée sur l'information mutuelle entre les entrées du problème et la transcription de la communication entre les joueurs. Puis, nous avons utilisé l'incompressibilité au sens de Kolmogorov pour spécifier les entrées sur lesquelles on considère le problème. Enfin, pour traiter le cas de la complexité probabiliste, nous avons donné une alternative au principe du min-max de Yao, également basée sur la complexité de Kolmogorov. Là encore, il s'agit de choisir l'aléa utilisé dans un protocole de communication de manière incompressible.

Dans un second temps, nous avons travaillé sur la structure algébrique sous-jacente à la complexité de la communication. Il s'agit d'une simple structure d'espace vectoriel, qui permet de représenter les problèmes de communication sous forme matricielle. Ceci nous a amené à considérer un modèle plus large, la simulation des distributions causales. D'une part, ce modèle permet de généraliser, en plus de la complexité de la communication traditionnelle, la simulation des mesures bipartites sur des états intriqués. D'autre part, c'est dans ce modèle que la structure matricielle définissant les problèmes de communication prend tout son sens, donnant une interprétation naturelle aux opérations algébriques comme la multiplication par un scalaire ou la combinaison convexe.

Nous avons introduit un nouveau modèle de complexité : la simulation des distributions causales. Les résultats obtenus l'ont été en étudiant la géométrie des structures dans l'espace de Banach sous-jacent. Les bornes inférieures que nous avons prouvées dans ce modèle s'interprètent naturellement comme violation maximale des inégalités de Bell et de Tsirleson. Il est intéressant de noter que la méthode de Linial et Shraibman, considérée comme importante en complexité de la communication, est en fait équivalente à celle qu'a utilisé Bell pour montrer l'existence de théories quantiques non-locales. Les modèles que nous avons traité sont la simulation avec et sans erreur des distributions, avec aléa partagé ou intrication partagée. Nous avons également donné des bornes supérieures. Dans ce cas, les bornes portent sur la simulation avec erreur et nos arguments ne peuvent a priori pas s'étendre au cas sans erreur. Pour prouver cette borne supérieure, nous avons montré une relation entre violation des inégalités de Bell et de Tsirelson. Il s'agit là d'un nouveau résultat géométrique qui borne la distance entre l'ensemble des distributions quantiques et classiques. Ceci nous a permis de montrer comment simuler les distributions causales dans le modèle quantique sans interaction et sans ressource partagée.

Enfin, dans une troisième partie, nous avons étudié la complexité en boîte non-locale, une ressource qui contrairement à la communication respecte la non-localité. Notre objectif a été d'étudier cette ressource du point de vue de la théorie de la complexité. La première raison est que cette ressource semble plus naturelle pour quantifier des phénomènes causaux, comme la simulation des mesures sur les états intriqués. La seconde est l'importance de celle-ci en physique théorique, et nous pensons que la puissance cal-

culatoire ce celle-ci pourrait jouer un rôle déterminant. Nous avons là encore donné des bornes inférieures et supérieures sur la complexité en boite non-locale, dans les modèles déterministes et probabiliste. Pour y parvenir, nous avons comparé les boites non-locales avec la communication, et relié celles-ci avec le rang de la matrice associée au problème. En particulier, nous avons caractérisé la complexité dans le cas de l'utilisation simultanée et sans aléa des boites non-locales comme étant égal au rang sur $\mathbb{Z}/2\mathbb{Z}$ de la matrice.

L'étude de la non-localité nous a permis de donner un nouveau point de vue sur un problème cryptographique, l'évaluation sécurisée. Par notre approche, nous avons caractérisé le cas déterministe, utilisant comme primitive le *ET sécurisé*. Dans le cas probabiliste, nous avons utilisé les boites non-locales pour donner des bornes inférieures et supérieures sur le nombre de primitives *oblivious transfer* nécessaire pour calculer une fonction de manière sécurisée.

Dans notre travail, nous avons cherché à opérer une distinction systématique entre méthodes algébriques et combinatoires. Les méthodes basées sur la combinatoire, telle la complexité de Kolmogorov, présentent l'avantage de pouvoir être utilisées pour prouver des séparations entre la communication classique et la communication quantique. Nous avons montré en effet que notre méthode permettait de prouver une borne inférieure en $\sqrt{n}$ sur le problème du couplage caché. Par ailleurs, on sait que les méthodes comme le γ_2 de Linial et Shraibman ne permettent pas de prouver de tels résultats. Le cas de la distribution de Deutsch-Josza est à ce titre éloquent. En effet, on sait que cette distribution nécessite une quantité linéaire de communication. De plus, ceci a été prouvé en utilisant la méthode des rectangles, dont nous avons montré qu'elle était généralisée par la complexité de Kolmogorov.

L'impossibilité de prouver des séparations en utilisant la méthode γ_2 était connue pour les fonctions booléennes, et nous avons étendu ce résultat à toutes les distributions causales. Si ceci semble à première vue négatif, on pourrait inversement utiliser cette approche pour montrer que, pour certaines classes de distributions, il n'y a pas de séparation possible. Le protocole de Regev et Toner, qui permet de simuler les distributions quantiques à marginales uniformes, utilise en effet l'inégalité de Grothendieck. Dans ce contexte, elle peut s'interpréter comme une majoration de la différence entre violations d'inégalité de Bell et de Tsirelson. Notre travail a montré que le calcul des fonctions booléennes était un problème similaire à la simulation des distributions causales. On pourrait donc imaginer que le résultat de Regev et Toner pourrait avoir des conséquences sur la simulation de la communication quantique par des ressources classiques.

En abordant la complexité de la communication du point de vue de la combinatoire, notre objectif était de proposer une alternative à l'utilisation de la théorie de l'information pour prouver des bornes inférieures. La complexité de Kolmogorov et la théorie de l'information ont plusieurs points en commun, et la formulation de notre borne inférieure est finalement assez similaire à celle qu'on peut obtenir en utilisant la théorie de l'information. En utilisant la complexité de Kolmogorov, on a une nouvelle intuition qui permet d'imaginer de nouvelles preuves de bornes inférieures. Plus important, la complexité de Kolmogorov permet de généraliser plusieurs méthodes de bornes inférieures. Ceci nous a permis d'initier une classification des méthodes de bornes inférieures existantes.

La théorie de l'information est appliquée à la complexité de la communication depuis plusieurs années. Les techniques qui en découlent ont été perfectionnées et leurs applications vont au-delà des bornes inférieures. L'une des applications les plus intéressantes de

ces techniques est de prouver des théorèmes de produit direct. L'objectif de ces théorèmes est de comparer la complexité de n instances d'un problème avec la complexité d'une seule. Les théorèmes de produit direct sont en général considérés comme un outil important, notamment parce qu'ils permettent de trouver des fonctions difficiles à calculer. Il serait intéressant de chercher à utiliser les propriétés structurelles de la complexité de Kolmogorov pour tenter de prouver ce type de théorème, ou d'autre théorèmes généraux sur la complexité de la communication.

L'une des questions que nous n'avons pas pu résoudre est la comparaison générale de notre méthode utilisant la complexité de Kolmogorov et des méthodes utilisant la théorie de l'information. Nous n'avons pas plus d'information sur la manière dont se comparent ces deux méthodes avec celle de Linial et Shraibman. D'un côté, les premières permettent de prouver des grandes séparations, là où la seconde montre au plus un écart constant. De l'autre, la méthode γ_2 est équivalente, pour les fonctions, à la méthode du rang, dont on conjecture qu'elle caractérise la complexité de la communication. Il ne fait pas de doute que des réponses à ces questions permettraient d'améliorer sensiblement la connaissance du modèle de complexité de la communication.

Enfin, sur les boîtes non-locales, nous n'avons fait que commencer l'étude de la complexité. Dans le modèle déterministe, nous avons pu montrer qu'il existait un protocole efficace pour des fonctions courantes, pour lesquelles les protocoles précédents donnaient une complexité exponentielle. La question est maintenant de savoir s'il existe une fonction dont la complexité est super-linéaire. Nous n'avons pas été en mesure d'identifier une telle fonction. Dans le modèle probabiliste, nous avons pu identifier un lien avec le rang approché. Auparavant, on connaissait le lien entre complexité de la communication et rang approché. Dans le cas de la complexité en boîte non-locale, le simple cas d'une fonction aléatoire n'est pas connue. La structure algébrique sous-jacente est cette fois complexe et mal connue. La réponse à cette question pourrait permettre de mieux comprendre les liens entre communication et boîtes non-locales.

En nous plaçant ainsi à l'intersection de différentes disciplines, ce sont toutes celles-ci que nous souhaitons voir progresser. La physique a probablement beaucoup apporté à l'informatique. Elle a produit la théorie des semi-conducteurs, qui a permis de construire les ordinateurs d'aujourd'hui. Elle fournira peut être demain les techniques nécessaires pour construire un ordinateur quantique. Toutefois, il serait ridicule de limiter à cela les relations entre ces deux disciplines. Rappelons-nous ici de l'aphorisme de Dijkstra pour qui "l'informatique n'est pas plus la science des ordinateurs que l'astronomie celle des télescopes". L'informatique est avant tout la science de l'information. Comprendre et expliquer la nature du point de vue du traitement de l'information est un problème fondamental et porteur de nouvelles interactions entre les différentes disciplines scientifiques. C'est à ce point d'intersection précis que nous avons cherché à nous placer, et ce sont ces interactions que nous avons voulu explorer.

Bibliographie

[ADR82] A. Aspect, J. Dalibard, and G. Roger. Experimental test of bell's inequalities using time-varying analyzers. *Physical Review Letters*, 49(25) :1804–1807, 1982.

[AGR81] A. Aspect, P. Grangier, and G. Roger. Experimental tests of realistic local theories via bell's theorem. *Physical Review Letters*, 47(7) :460–463, 1981.

[AGR82] A. Aspect, P. Grangier, and G. Roger. Experimental realization of einstein-podolsky-rosen-bohm gedankenexperiment : A new violation of bell's inequalities. *Physical Review Letters*, 49(2) :91–94, 1982.

[AKS04] M. Agrawal, N. Kayal, and N. Saxena. Primes is in P. *Annals of Mathematics*, 160 :781–793, 2004.

[ALM+98] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and hardness of approximation problems. *Journal of the ACM*, 43(3) :501–555, 1998.

[Amb02] A. Ambainis. Quantum lower bounds by quantum arguments. *Journal of Computer and System Sciences*, 64(4) :750–767, June 2002.

[AMS99] N. Alon, Y. Matias, and M. Szegedy. The space complexity of approximating the frequency moments. *Journal of Computer and System Sciences*, 58(1) :137–147, 1999.

[AN06] N. Alon and A. Naor. Approximating the cut-norm via Grothendieck's inequality. *SIAM Journal on Computing*, 35(4) :787–803, 2006.

[Bar07] J. Barrett. Information processing in generalized probabilistic theories. *Physical Review A*, 75(3) :032304, 2007.

[BB84] C.H. Bennett and G. Brassard. Quantum cryptography : Public key distribution and coin tossing. In *Proceedings of the IEEE International Conference on Computers, Systems, and Signal Processing*, page 175, 1984.

[BBC+93] C.H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W.K. Wootters. Teleporting an unknown quantum state via dual classical and epr channels. *Physical Review Letters*, 70 :1895–1899, 1993.

[BBC+01] R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. de Wolf. Quantum lower bounds by polynomials. *Journal of the ACM*, 48(4) :778–797, 2001.

[BBL+06] G. Brassard, H. Buhrman, N. Linden, A.A.Methot, A. Tapp, and F. Unger. Limit on nonlocality in any world in which communication complexity is not trivial. *Physical Review Letters*, 96(25) :250401, 2006.

[BBLW07] H. Barnum, J. Barrett, M. Leifer, and A. Wilce. Generalized no-broadcasting theorem. *Physical Review Letters*, 99(240501), 2007.

[BC97] H. Buhrman and R. Cleve. Substituting quantum entanglement for communication. *Physical Review A*, 56(2) :1201–1204, 1997.

[BCT99] G. Brassard, R. Cleve, and A. Tapp. Cost of exactly simulating quantum entanglement with classical communication. *Physical Review Letters*, 83 :1874–1877, 1999.

[BCU+06] H. Buhrman, M. Christandl, F. Unger, S. Wehner, and A. Winter. Implications of superstrong nonlocality for cryptography. *Proceedings of the Royal Society A*, 462(2071) :1919–1932, 2006.

[BCW98] H. Buhrman, R. Cleve, and A. Wigderson. Quantum vs. classical communication and computation. In *Proc. of the 30th Annual ACM Symposium on Theory of Computing*, pages 63–68, 1998.

[BCWdW01] H. Buhrman, R. Cleve, J. Watrous, and R. de Wolf. Quantum fingerprinting. *Physical Review Letters*, 87(16), 2001.

[BdW01] H. Buhrman and R. de Wolf. Communication complexity lower bounds by polynomials. In *Proc. of the 16th Annual IEEE Conference on Computational Complexity*, pages 120–130, 2001.

[Bea96] D. Beaver. Correlated pseudorandomness and the complexity of private computations. In *Proc. of the 28th Annual ACM Symposium on Theory of Computing*, pages 479–488, 1996.

[BEHW89] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth. Learnability and the vapnik-chervonenkis dimension. *J. of the ACM*, 36(4) :929–969, October 1989.

[Bel64] J.S. Bell. On the Einstein Podolsky Rosen paradox. *Physics*, 1 :195, 1964.

[BF92] L. Babai and P. Frankl. Linear algebra methods in combinatorics. Preliminary Version 2, September 1992.

[BJLV00] H. Buhrman, T. Jiang, M. Li, and P. Vitanyi. New applications of the incompressibility method : Part ii. *Theoretical Computer Science*, 235(1) :59–70, 2000.

[BK97] L. Babai and P.G. Kimmel. Randomized simultaneous messages : solution of a problem of yao incommunication complexity. In *Proc. of the 12th Annual IEEE Conference on Computational Complexity*, pages 239–246, 1997.

[BLM+05] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts. Non-local correlations as an information theoretic resource. *Physical Review A*, 71 :022101, 2005.

[BM04] A. Beimel and T. Malkin. A quantitative approach to reductions in secure computation. In *Proceedings of the First Theory of Cryptography Conference*, pages 238–257, 2004.

[BNS92] L. Babai, N. Nisan, and M. Szegedy. Multiparty protocols, pseudorandom generators for LOGSPACE, and time-space trade-offs. *Journal of Computer and System Sciences*, 45(2), 1992.

[BOGW88] M. Ben-Or, S. Goldwasser, and A. Wigderson. Completeness theorems for noncryptographic faulttolerant distributed computations. In *Proc. of the 20th Annual ACM Symposium on Theory of Computing*, pages 1–10, 1988.

[BP05] J. Barrett and S. Pironio. Popescu-Rohrlich correlations as a unit of nonlocality. *Physical Review Letters*, 95 :140401, 2005.

[BPSW06] P. Beame, T. Pitassi, N. Segerlind, and A. Wigderson. A strong direct product theorem for corruption and the multiparty communication complexity of disjointness. *Computational Complexity*, 15(4) :391–432, 2006.

[BS09] N. Brunner and P. Skrzypczyk. Non-locality distillation and post-quantum theories with trivial communication complexity. Technical Report arXiv :0901.4070v2, arXiv e-Print archive, 2009.

[BV04] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[BW92] C.H. Bennett and S. J. Wiesner. Communication via one- and two-particle operators on einstein-podolsky-rosen states. *Physical Review Letters*, 69(20) :2881–2884, 1992.

[BYJK08] Z. Bar-Yossef, T.S. Jayram, and I. Kerenidis. Exponential separation of quantum and classical one-way communication complexity. *SIAM Journal on Computing*, 38(1) :366–384, 2008.

[BYJKS02] Z. Bar-Yossef, T.S. Jayram, R. Kumar, and D. Sivakumar. Information theory methods in communication complexity. In *Proc. of the 17th Annual IEEE Conference on Computational Complexity*, pages 93–102. IEEE Computer Society, 2002.

[BYJKS04] Z. Bar-Yossef, T.S. Jayram, R. Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *Journal of Computer and System Sciences*, 68(4) :702–732, 2004.

[CCD88] D. Chaum, C. Crepeau, and I. Damgard. Multiparty unconditionally secure protocols. In *Proc. of the 20th Annual ACM Symposium on Theory of Computing*, pages 11–19, 1988.

[CG88] B. Chor and O. Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM Journal on Computing*, 17(2), April 1988.

[CGM00] N. Cerf, N. Gisin, and S. Massar. Classical teleportation of a quantum bit. *Physical Review Letters*, 84 :2521–2524, 2000.

[CGMP05] N. Cerf, N. Gisin, S. Massar, and S. Popescu. Simulating maximal quantum entanglement without communication. *Physical Review Letters*, 94(22) :220403, 2005.

[Cha69] G.J. Chaitin. On the length of programs for computing finite binary sequences : Statistical considerations. *Journal of the ACM*, 16(1) :145–169, 1969.

[Cha75] G.J. Chaitin. A theory of program size formally identical to information theory. *Journal of the ACM*, 22(3) :329–340, 1975.

[CHSH69] J.F. Clauser, M.A. Horne, A. Shimony, and R.A. Holt. Proposed experiment to test local hidden-variable theories. *Physical Review Letters*, 23 :880–884, 1969.

[CHTW04] R. Cleve, P. Høyer, B. Toner, and J. Watrous. Consequences and limits of nonlocal strategies. In *Proc. of the 19th Annual IEEE Conference on Computational Complexity*, pages 236–249, 2004.

[CK91] B. Chor and E. Kushilevitz. A zero-one law for boolean privacy. *SIAM Journal of Discrete Mathematics*, 4(1) :36–47, 1991.

[dGdW02] M. de Graaf and R. de Wolf. On quantum versions of the yao principle. In *Proc. of the 19th Annual Symposium on Theoretical Aspects of Computer Science*, pages 347–358, London, UK, 2002. Springer-Verlag.

[DH76] W. Diffie and M.E. Hellma. New directions in cryptography. *IEEE Transactions on Information Theory*, 22(6) :644–654, 1976.

[DKLR09] J. Degorre, M. Kaplan, S. Laplante, and J. Roland. The communication complexity of non-signaling distributions. In *Proc. of the 34th International Symposium on Mathematical Foundations of Computer Science*, 2009. à paraître.

[DLR05] J. Degorre, S. Laplante, and J. Roland. Simulating quantum correlations as a distributed sampling problem. *Physical Review A*, 72(062314), 2005.

[DLTW08] A.C. Doherty, Y. Liang, B. Toner, and S. Wehner. The quantum moment problem and bounds on entangled multi-prover games. In *Proc. of the 23rd Annual IEEE Conference on Computational Complexity*, pages 199 – 210, 2008.

[DM99] Y. Dodis and S. Micali. Lower bounds for oblivious transfer reductions. In *Proc. of EUROCRYPT '99*, pages 42–55, 1999.

[EPR35] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete ? *Physical Review*, 47 :777–780, 1935.

[FC72] S.J. Freedman and J.F. Clauser. Experimental test of local hidden-variable theories. *Physical Review Letters*, 28(14) :938–941, 1972.

[Fey82] R. Feynman. Simulating physics with computers. *International Journal of Theoretical Physics*, 21(6) :467–488, June 1982.

[FR81] D.J. Foulis and C.H. Randall. Empirical logic and tensor products. In *Interpretations and Foundations of Quantum Theory*, volume Interpretations and Foundations of Quantum Theory, pages 1–20. Wissenschaftsverlag, BibliographischesInstitut, 1981.

[FWW09] M. Forster, S. Winkler, and S. Wolf. Distilling nonlocality. *Physical Review Letters*, 102(120401), 2009.

[Gác74] P. Gács. On the symmetry of algorithmic information of algorithmic information. *Soviet Mathematics Doklady*, 15 :1477–1480, 1974.

[Gav09] D. Gavinsky. Classical interaction cannot replace quantum nonlocality. Technical Report arXiv :0901.0956v1, arXiv e-Print archive, 2009.

[GKdW06] D. Gavinsky, J. Kempe, and R. de Wolf. Strengths and weaknesses of quantum fingerprinting. In *Proc. of the 21st Annual IEEE Conference on Computational Complexity*, pages 288–295, 2006.

[GKK+07] D. Gavinsky, J. Kempe, I. Kerenidis, R. Raz, and R. de Wolf. Exponential separations for one-way quantum communication complexity, with applications to cryptography. In *Proc. of the 39th Annual ACM Symposium on Theory of Computing*, pages 516–525. ACM, 2007.

[Gro56] A. Grothendieck. Résumé de la théorie métrique des produits tensoriels topologiques. *Boletim Da Sociedade de Matemática de São Paolo*, 8 :1–79, 1956.

[Gro85] H.J. Groenewold. The elusive quantal individual. *Physics Reports*, 127(6) :379–401, 1985.

[Gro96] L.K. Grover. A fast quatum mechanical algorithm for database search. In *Proc. of the 28th Annual ACM Symposium on Theory of Computing*, pages 212–219, 1996.

[GV88] O. Goldreich and R. Vainish. How to solve any protocol problem - an efficiency improvement. In *Advances in Cryptology - CRYPTO '87*, pages 73–86, 1988.

[GW95] M.X. Goemans and D.P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, 42 :1115–1145, 1995.

[Hås01] J. Håstad. Some optimal innappoximability results. *Journal of the ACM*, 48(4) :798–859, 2001.

[Hol73] A.S. Holevo. Some estimates for the amount of information transmittable by a quantum communications channel. *Problemy Peredachi Informatsii*, 9(3) :3–11, 1973.

[IM02] K. Iwama and H. Morizumi. An explicit lower bound of 5n - o(n) for boolean circuits. In *Proc. of the 27th International Symposium on Mathematical Foundations of Computer Science*, volume 2420, pages 353–364, 2002.

[JKN08] R. Jain, H. Klauck, and A. Nayak. Direct product theorems for classical communication complexity via subdistribution bounds : extended abstract. In *Proc. of the 40th Annual ACM Symposium on Theory of Computing*, pages 599–608, 2008.

[JKS03] T. S. Jayram, R. Kumar, and D. Sivakumar. Two applications of information complexity. In *Proc. of the 35th Annual ACM Symposium on Theory of Computing*, pages 673–682, New York, NY, USA, 2003. ACM.

[JLR00] S. Janson, T. Luczak, and A. Ruci. *Random Graphs*. John Wiley and Sons, 2000.

[JM05] N.S. Jones and L. Masanes. Interconversion of nonlocal correlations. *Physical Review A*, 72 :052312, 2005.

[Jus72] J. Justesen. A class of constructive asymptotically good algebraic codes. *IEEE Transactions on Information Theory*, 18 :652–656, 1972.

[Kho02] S. Khot. On the power of unique 2-prover 1-round games. In *Proc. of the 35th Annual ACM Symposium on Theory of Computing*, pages 767–775, 2002.

[Kil88] J. Kilian. Basing cryptography on oblivious transfer. In *Proc. of the 20th Annual ACM Symposium on the Theory of Computing*, pages 20–31, 1988.

[Kit97] A. Y. Kitaev. Quantum computations : algorithms and error correction. *Russian Mathematical Survey*, 52 :1191–1249, 1997.

[KKLR09] M. Kaplan, I. Kerenidis, S. Laplante, and J. Roland. Non-local box complexity and secure function evaluation. Technical Report arXiv :0903.2179, arXiv e-Print archive, 2009.

[KKM+08] J. Kempe, H. Kobayashi, K. Matsumoto, B. Toner, and T. Vidick. Entangled games are hard to approximate. In *Proceedings of 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 447–456, 2008.

[KL09] M. Kaplan and S. Laplante. Kolmogorov complexity and combinatorial methods in communication complexity. In *Proc. of the 6th Conference on Theory and Applications of Models of Computation*, pages 261–270, 2009.

[Kla07] H. Klauck. Lower bounds for quantum communication complexity. *SIAM Journal on Computing*, 37(1) :20–46, 2007.

[KN97] E. Kushilevitz and N. Nisan. *Communication complexity*. Cambridge University Press, New York, 1997.

[KNR99] I. Kremer, N. Nisan, and D. Ron. On randomized one-round communication complexity. *Computational Complexity*, 8(1) :21–49, 1999.

[Kol65] A.N. Kolmogorov. Three approaches to the quantitative definition of information. *Problems of Information Transmission*, 1(1) :1–7, 1965.

[Kre95] I. Kremer. Quantum communication. Master's thesis, The Hebrew University of Jerusalem, 1995.

[KRF87] Matthias Kläy, Charles H. Randall, and David J. Foulis. Tensor products and probability weights. *International Journal of Theoretical Physics*, 26(3) :199–219, 1987.

[Kri79] J.L. Krivine. Constantes de Grothendieck et fonctions de type positif sur les spheres. *Advances in Math*, 31 :16–30, 1979.

[KS92] B. Kalyanasundaram and G. Schnitger. The probabilistic communication complexity of set intersection. *SIAM Journal on Discrete Mathematics*, 5(4) :545–557, 1992.

[Kus92] E. Kushilevitz. Privacy and communication complexity. *SIAM Journal of Discrete Mathematics*, 5(2) :273–284, 1992.

[KW90] M. Karchmer and A. Wigderson. Monotone circuits for connectity require super logarithmic depth. *SIAM Journal on Discrete Mathematics*, 3(2) :255–265, 1990.

[Len90] T. Lengauer. VLSI theory. *Handbook of theoretical computer science (vol. A) : algorithms and complexity*, A :835–866, 1990.

[Lev74] L.A. Levin. Laws of information conservation (non-growth) and aspects of the foundations of probability theory. *Problems of Information Transmission*, 10(3) :206–210, 1974.

[LM08] S. Laplante and F. Magniez. Lower bounds for randomized and quantum query complexity using kolmogorov arguments. *SIAM Journal on Computing*, 38(1) :46–62, 2008.

[LS93] L. Lovàcz and M. Saks. Lattices, Möbius functions, and communication complexity. *Journal of Computer and System Sciences*, 47 :322–349, 1993.

[LS08a] T. Lee and M. Saks. personal communication, 2008.

[LS08b] N. Linial and A. Shraibman. Lower bounds in communication complexity based on factorization norms. *Random Structures and Algorithms*, 2008.

[LS09] T. Lee and A. Shraibman. An approximation algorithm for approximation rank. In *Proc. of the 24th Annual IEEE Conference on Computational Complexity*, 2009. à paraître.

[LSŠ08] T. Lee, A. Shraibman, and R. Špalek. A direct product theorem for discrepancy. In *Proc. of the 23rd Annual IEEE Conference on Computational Complexity*, pages 71–80, 2008.

[LU08] T. Lee and F. Unger. personal communication, 2008.

[LV08] M. Li and P. Vitanyi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer-Verlag, 3rd edition, 2008.

[Mau92] T. Maudlin. Bell's inequality, information transmission, and prism models. In *Biennal Meeting of the Philosophy of Science Association*, pages 404–417, 1992.

[McD91] C. McDiarmid. Concentration. In *Probabilistic Methods for Algorithmic Discrete Mathematics*. Springer, 1991.

[Moo65] G.E. Moore. Cramming more components onto integrated circuits. *Electronics*, 38(8) :114–117, April 1965.

[MS82] K. Mehlhorn and E. Schmidt. Las vegas is better than determinism in vlsi and distributed computing. In *Proc. of the 14th Annual ACM Symposium on Theory of Computing*, pages 330–337, 1982.

[New91] I. Newman. Public vs. private coin flips in one round communication games. *Information Processing Letters*, 39(2) :67–71, 1991.

[NPA08] M. Navascues, S. Pironio, and A. Acín. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. *New Journal of Physics*, 10(7), 2008. 29 pages.

[Pir03] S. Pironio. Violations of bell inequalities as lower bounds on the communication cost of nonlocal correlations. *Physical Review A*, 68(6) :062102, 2003.

[PR94] S. Popescu and D. Rohrlich. Causality and nonlocality as axioms for quantum mechanics. *Foundations of Physics*, pages 379–385, 1994.

[PSS84] W.J. Paul, J. Seiferas, and J. Simon. An information theoretic approach to time bounds for on-line computation. *Journal of Computer and System Sciences*, 23(2) :108–126, 1984.

[Raz92] A. Razborov. On the distributional complexity of disjointness. *Theoretical Computer Science*, 106(2) :385–390, 1992.

[Raz95] R. Raz. Fourier analysis for probabilistic communication complexity. *Computational Complexity*, 5(3/4) :205–221, 1995.

[Raz99] R. Raz. Exponential separation of quantum and classical communication complexity. In *Proc. of the 31th Annual ACM Symposium on Theory of Computing*, pages 358–367, New York, NY, USA, 1999. ACM.

[Raz03] A. Razborov. Quantum communication complexity of symmetric predicates. *Izvestiya : Mathematics*, 67(1) :145–159, 2003.

[RF81] C.H. Randall and D.J. Foulis. Operational statistics and tensor products. In *Interpretations and Foundations of Quantum Theory*, pages 21–28. Wissenschaftsverlag, BibliographischesInstitut, 1981.

[Roc70] R.T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[RS95] R. Raz and B. Spieker. On the "log-rank" conjecture in communication complexity. *Combinatorica*, 15(4) :567–588, 1995.

[RT07] O. Regev and B. Toner. Simulating quantum correlations with finite communication. In *Proc. of the 48th Annual IEEE Symposium on Foundations of Computer Science*, pages 384–394, 2007.

[Sau72] N. Sauer. On the density of families of sets. *Journal of Combinatorial Theory (A)*, 12 :145–147, 1972.

[Sha48] C. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27 :379–423, 623–656, 1948.

[Sha49] C. Shannon. The synthesis of two-terminal switching circuits. *Bell System Technical Journal*, 28 :59–98, 1949.

[She08] A. Sherstov. Communication complexity under product and nonproduct distributions. In *Proc. of the 23rd Annual IEEE Conference on Computational Complexity*, pages 64–70. IEEE Computer Society, 2008.

[Sho97] P. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, 26 :1484–1509, 1997.

[Sol64] R.J. Solomonoff. A formal theory of inductive inference, part 1 and part 2. *Information and Control*, 7 :1–22, 224–254, 1964.

[Sol95] R.M. Solovay. Unpublished, 1995.

[SS77] R.M. Solovay and V. Strassen. A fast monte-carlo test for primality. *SIAM Journal on Computing*, 6(1) :84–85, 1977.

[Ste00] M. Steiner. Towards quantifying non-local information transfer : finite-bit non-locality. *Physical Letters A*, 270 :239–244, 2000.

[SZ08] Y. Shi and Y. Zhu. Tensor norms and the classical communication complexity of bipartite quantum measurements. *SIAM Journal on Computing*, 38(3) :753–766, 2008.

[TB03] B. Toner and D. Bacon. Communication cost of simulating Bell correlations. *Physical Review Letters*, 91 :187904, 2003.

[Tsi80] B.S. Tsirelson. Quantum generalizations of Bell's inequality. *Letters in Mathematical Physics*, 4(2) :93–100, 1980.

[Tsi85] B.S. Tsirelson. Problems of the theory probability distributions ix. *Zapiski Nauchnykh Seminarov Leningradskogo Otdeleniya Matematicheskogo Instituta*, 142 :174–194, 1985. English translation in Quantum analogues of the Bell inequalities. The case of two spatially separated domains, J. Soviet Math. 36, 557–570 (1987).

[Tur36] A. Turing. On computable numbers, with an application to the entscheidungsproblem. *Proceedings of the London Mathematical Society*, 2(42) :230–265, 1936.

[VB09] T. Vértisi and E. Bene. Lower bound on the communication cost of simulating bipartite quantum correlations. Technical Report arXiv :0904.1390v1, arXiv e-Print archive, 2009.

[vD05] W. van Dam. Implausible consequences of superstrong nonlocality. Technical Report quant-ph/0501159, arXiv e-Print archive, 2005.

[Wie83] S.J. Wiesner. Conjugate coding. *ACM SIGACT News - A special issue in cryptography*, pages 78–88, 1983.

[Wil92] A. Wilce. Tensor products in generalized measure theory. *International Journal of Theoretical Physics*, 31(11) :1915–1928, 1992.

[WW05] S. Wolf and J. Wullschleger. Oblivious transfer and quantum non-locality. In *Proceedings of the International Symposium on Information Theory*, pages 1745–1748, 2005.

[WW06] S. Wolf and J. Wullschleger. Oblivious transfer is symmetric. In *Proc. of EUROCRYPT '06*, pages 222–232, 2006.

[Yao77] A.C.C. Yao. Probabilistic computations : Toward a unified measure complexity (extended abstract). In *Proc. of the 18th Annual Symposium on Foudations of Computer Science*, pages 222–227, 1977.

[Yao79] A.C.C. Yao. Some complexity questions related to distributive computing. In *Proc. of the 11th Annual ACM Symposium on Theory of Computing*, pages 209–213, 1979.

[Yao82] A.C.C. Yao. Protocols for secure computations (extended abstract). In *proc. of the 23rd Annual Symposium of Foundations of Computer Science*, pages 160–164, 1982.

[Yao83] A.C.C. Yao. Lower bounds by probabilistic arguments (extended abstract). In *Proc. of the 24th Annual IEEE Symposium on Foundations of Computer Science*, pages 420–428, 1983.

[Yao93] A.C.C. Yao. Quantum circuit complexity. In *Proc. of the 34th Annual IEEE Symposium on Foundations of Computer Science*, pages 352–361, 1993.

[Yao03] A.C.C. Yao. On the power of quantum fingerprinting. In *Proc. of the 35th Annual ACM Symposium on Theory of Computing*, pages 77–81, 2003.