

# Continuous-variable nonlocality and contextuality

Rui Soares Barbosa

Department of Computer Science  
University of Oxford

rui.soares.barbosa@cs.ox.ac.uk

Tom Douce

Laboratoire d'Informatique de Paris 6  
CNRS and Sorbonne Université

School of Informatics  
University of Edinburgh

doucetom@hotmail.fr

Pierre-Emmanuel Emeriau

Laboratoire d'Informatique de Paris 6  
CNRS and Sorbonne Université

pierre-emmanuel.emeriau@lip6.fr

Elham Kashefi

Laboratoire d'Informatique de Paris 6  
CNRS and Sorbonne Université

School of Informatics  
University of Edinburgh

elham.kashefi@lip6.fr

Shane Mansfield

Laboratoire d'Informatique de Paris 6  
CNRS and Sorbonne Université

shane.mansfield@lip6.fr

Contextuality is a non-classical behaviour that can be exhibited by quantum systems. It is increasingly studied for its relationship to quantum-over-classical advantages in informatic tasks. To date, it has largely been studied in discrete variable scenarios, where observables take values in discrete and usually finite sets. Practically, on the other hand, continuous-variable scenarios offer some of the most promising candidates for implementing quantum computations and informatic protocols. Here we set out a framework for treating contextuality in continuous-variable scenarios. It is shown that the Fine–Abramsky–Brandenburger theorem extends to this setting, an important consequence of which is that nonlocality can be viewed as a special case of contextuality, as in the discrete case. The contextual fraction, a quantifiable measure of contextuality that bears a precise relationship to Bell inequality violations and quantum advantages, can also be defined in this setting. It is shown to be a non-increasing monotone with respect to classical operations that include binning to discretise data. Finally, we consider how the contextual fraction can be formulated as an infinite linear program, and calculated with increasing accuracy using semi-definite programming approximations.

## Introduction

Contextuality is one of the principal non-classical behaviours that can be exhibited by quantum systems. The Heisenberg uncertainty principle identified that certain pairs of quantum observables are incompatible, e.g. position and momentum. In operational terms, observing one will disturb the outcome statistics of the other. Imprudent commentators will sometimes cite this as evidence that position and momentum cannot simultaneously be assigned definite values. However, this is not quite right and a more careful conclusion is that we simply cannot observe these values simultaneously. To make a stronger statement requires contextuality. Roughly speaking, the latter is present whenever the behaviour of a system is inconsistent with the basic assumptions that (i) all of its observable properties may be assigned definite values at all times, and (ii) jointly performing *compatible* observables does not disturb the global value assignment. Aside from its foundational importance, today contextuality is increasingly studied as the essential ingredient for enabling a range of quantum-over-classical advantages in informatic tasks, which include the onset of universal quantum computing in certain computational models [65, 44, 6, 20, 7].

It is notable that to date the study of contextuality has largely focused on discrete variable scenarios and that the main frameworks and approaches to contextuality are tailored to modelling these, e.g. [8, 26, 13, 33]. In such scenarios, observables can only take values in discrete, and usually finite, sets. Discrete variable scenarios typically arise in finite-dimensional quantum mechanics, e.g. when dealing with quantum registers in the form of systems of multiple qubits as is common in quantum information and computation theory.

Yet, from a practical perspective, continuous-variable quantum systems are emerging as some of the most promising candidates for implementing quantum informational and computational tasks [25, 74]. The main reason for this is that they offer unrivalled possibilities for deterministic generation of large-scale resource states [76] and for highly-efficient measurements of certain observables. Together these cover many of the basic operations required in the one-way or measurement-based model of quantum computing [67] for example. Typical implementations are in optical systems where the continuous variables correspond to the position-like and momentum-like quadratures of the quantised modes of an electromagnetic field. Indeed position and momentum as mentioned previously in relation to the uncertainty principle are the prototypical examples of continuous variables in quantum mechanics.

Since quantum mechanics itself is infinite dimensional, it also makes sense from a foundational perspective to extend analyses of the key concept of contextuality to the continuous-variable setting. Furthermore, continuous variables can be useful when dealing with iteration, even when attention is restricted to finite-variable actions at discrete time steps, as is traditional in informatics. An interesting question, for example, is whether contextuality arises and is of interest in such situations as the infinite behaviour of quantum random walks.

The main contributions of this article are the following:

- We present a robust framework for contextuality in continuous-variable scenarios that follows along the lines of the discrete-variable framework introduced by Abramsky and Brandenburger [8] (Section 3).
- We show that the Fine–Abramsky–Brandenburger theorem [36, 8] extends to continuous variables (Section 4). This establishes that noncontextuality of an empirical behaviour, originally characterised by the existence of deterministic hidden-variable models [19, 50], can equivalently be characterised by the existence of a factorisable hidden-variable models, and that ultimately both of these are subsumed by a canonical form of hidden-variable model – a *global section* in the sheaf-theoretic perspective. An important consequence is that nonlocality may be viewed as a special case of contextuality in continuous-variable scenarios just as for discrete-variable scenarios.
- The contextual fraction, a quantifiable measure of contextuality that bears a precise relationship to Bell inequality violations and quantum advantages [6], can also be defined in this setting using infinite linear programming (Section 5). It is shown to be a non-increasing monotone with respect to the free operations of a resource theory for contextuality [6, 4]. Crucially, these include the common operation of binning to discretise data. A consequence is that any witness of contextuality on discretised empirical data also witnesses and gives a lower bound on genuine continuous-variable contextuality.
- While the infinite linear programs are of theoretical importance and capture exactly the quantity and inequalities we are interested in, they are not directly useful for actual numerical computations. To get around this limitation, we introduce a hierarchy of semi-definite programs which are relaxations of the original problem, and whose values converge monotonically to the contextual fraction (Section 6).

**Related work.** Note that we will specifically be interested in scenarios involving observables with continuous spectra, or in more operational language, measurements with continuous outcome spaces. We will still consider scenarios featuring only discrete sets of observables or measurements, as is typical in continuous-variable quantum computing. The possibility of considering contextuality in settings with continuous measurement spaces has also been evoked in [30]. We also note that several prior works have explicitly considered contextuality in continuous-variable systems [63, 40, 59, 73, 14, 52, 48]. Our approach is distinct from these in that it provides a genuinely continuous-variable treatment of contextuality itself as opposed to embedding discrete variable contextuality arguments into, or extracting them from, continuous-variable systems.

# 1 Continuous-variable behaviours

In this section we provide a brief motivational example for the kind of continuous-variable empirical behaviour we are interested in analysing. Suppose that we can interact with a system by performing measurements on it and observing their outcomes. A feature of quantum systems is that not all observables commute, so that certain combinations of measurements may be incompatible. At best we obtain empirical observational data for contexts in which only compatible measurements are performed, which can be collected by running the experiment repeatedly. As we shall make more precise in Sections 3 and 4, contextuality arises when the empirical data obtained is inconsistent with the assumption that for each run of the experiment the system has a global and context-independent assignment of values to all of its observable properties.

To take an operational perspective, a typical example of an experimental setup or scenario that we can consider is the one depicted in Figure 1 [left]. In this scenario, a system is prepared in some fixed bipartite state, following which parties  $A$  and  $B$  may each choose between two measurement settings,  $m_A \in \{a, a'\}$  for  $A$  and  $m_B \in \{b, b'\}$  for  $B$ . We will assume that outcomes of each measurement live in  $\mathbf{R}$ , which typically would be a bounded measurable subspace of the real numbers  $\mathbb{R}$ . Depending on which choices of inputs were made, the empirical data might for example be distributed according to one of the four hypothetical probability density plots in  $\mathbf{R}^2$  depicted in Figure 1 [right]. This scenario and hypothetical empirical behaviour has been considered elsewhere [48] as a continuous-variable version of the Popescu–Rohrlich (PR) box [64].

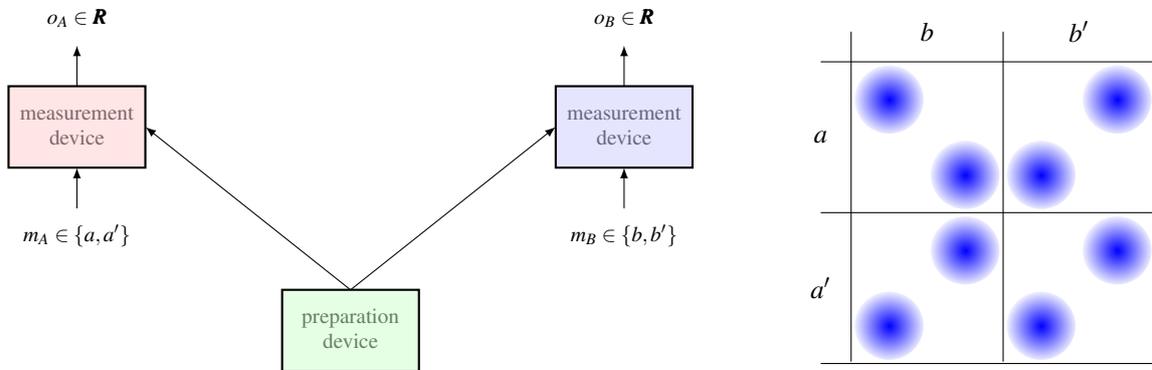


Figure 1: [Left] operational depiction of a typical bipartite experimental scenario. [Right] Hypothetical probability density plots for empirical data arising from such an experiment. Cf. the discrete-variable probability tables of [57, 55].

# 2 Preliminaries on measures and probability

In order to properly treat probability on continuous-variable spaces, it is necessary to introduce a modicum of measure theory. This section serves to recall some basic ideas and fix notation. The reader may choose to skip the section and consult it as reference for the remainder of the article.

A **measurable space** is a pair  $\mathbf{X} = \langle X, \mathcal{F} \rangle$  consisting of a set  $X$  and a  $\sigma$ -algebra (or  $\sigma$ -field)  $\mathcal{F}$  on  $X$ , i.e. a family of subsets of  $X$  containing the empty set and closed under complementation and countable unions. In some sense, this specifies the subsets of  $X$  that can be assigned a ‘size’, and which are therefore called the **measurable sets** of  $\mathbf{X}$ . Throughout this paper, we follow the convention of using boldface to denote the measurable space and the same symbol in regular face for its underlying set.

A trivial example of a  $\sigma$ -algebra over any set  $X$  is its powerset  $\mathcal{P}(X)$ , which gives the discrete measurable space  $\langle X, \mathcal{P}(X) \rangle$ , where every set is measurable. This is typically used when  $X$  is countable (finite or countably infinite), in which case this discrete  $\sigma$ -algebra is generated by the singletons. Another example, of central importance in measure theory, is  $\langle \mathbb{R}, \mathcal{B}_{\mathbb{R}} \rangle$ , where  $\mathcal{B}_{\mathbb{R}}$  is the  $\sigma$ -algebra generated from the open sets of  $\mathbb{R}$ , whose elements are called the Borel sets. Working with Borel sets avoids the problems that would arise if we naively attempted to measure or assign probabilities to points in the continuum. More generally, any topological space gives rise to a Borel measurable space in this fashion.

A **measurable function** between measurable spaces  $\mathbf{X} = \langle X, \mathcal{F}_X \rangle$  and  $\mathbf{Y} = \langle Y, \mathcal{F}_Y \rangle$  is a function  $f: X \rightarrow Y$  between the underlying sets whose preimage preserves measurable sets, i.e. such that  $f^{-1}(E) \in \mathcal{F}_X$  for any  $E \in \mathcal{F}_Y$ . This is analogous to the definition of a continuous function between topological spaces. Clearly, the identity function is measurable and measurable functions compose. We will denote by **Meas** the category whose objects are measurable spaces and whose morphisms are measurable functions.

The **product** of measurable spaces  $\mathbf{X}_1 = \langle X_1, \mathcal{F}_1 \rangle$  and  $\mathbf{X}_2 = \langle X_2, \mathcal{F}_2 \rangle$  is the measurable space

$$\mathbf{X}_1 \times \mathbf{X}_2 = \langle X_1 \times X_2, \mathcal{F}_1 \otimes \mathcal{F}_2 \rangle,$$

where the so-called tensor-product  $\sigma$ -algebra  $\mathcal{F}_1 \otimes \mathcal{F}_2$  is the  $\sigma$ -algebra on the Cartesian product  $X_1 \times X_2$  generated by the ‘rectangles’, the subsets of the form  $E_1 \times E_2$  with  $E_1 \in \mathcal{F}_1$  and  $E_2 \in \mathcal{F}_2$ . This is the categorical product in Meas.

A **measure** on a measurable space  $\mathbf{X} = \langle X, \mathcal{F} \rangle$  is a function  $\mu: \mathcal{F} \rightarrow \overline{\mathbb{R}}$  from the  $\sigma$ -algebra to the extended real numbers  $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$  satisfying:

- (i) [nonnegativity]  $\mu(E) \geq 0$  for all  $E \in \mathcal{F}$ ;
- (ii) [null empty set]  $\mu(\emptyset) = 0$ ;
- (iii) [ $\sigma$ -additivity] for any countable family  $(E_i)_{i=1}^{\infty}$  of pairwise disjoint measurable sets,  $\mu(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \mu(E_i)$ .

A measure on  $\mathbf{X}$  allows one to **integrate** well-behaved<sup>1</sup> measurable functions  $f: \mathbf{X} \rightarrow \langle \mathbb{R}, \mathcal{B}_{\mathbb{R}} \rangle$  to obtain a real value, denoted  $\int_{\mathbf{X}} f \, d\mu$  or  $\int_{x \in \mathbf{X}} f(x) \, d\mu(x)$ . The simplest example of such a measurable function is the **indicator function** of a measurable set  $E \in \mathcal{F}$ :

$$\chi_E(x) := \begin{cases} 1 & \text{if } x \in E \\ 0 & \text{if } x \notin E. \end{cases}$$

For any measure  $\mu$  on  $\mathbf{X}$ , its integral is

$$\int_{\mathbf{X}} \chi_E \, d\mu = \mu(E). \quad (1)$$

A measure is finite if  $\mu(X) < \infty$  and in particular it is a **probability measure** if  $\mu(X) = 1$ . We will denote by  $\mathbb{M}(\mathbf{X})$  and  $\mathbb{P}(\mathbf{X})$ , respectively, the sets of measures and probability measures on the measurable space  $\mathbf{X}$ .

A measurable function  $f: \mathbf{X} \rightarrow \mathbf{Y}$  carries any measure  $\mu$  on  $\mathbf{X}$  to a measure  $f_*\mu$  on  $\mathbf{Y}$ . This **push-forward** measure is given by  $f_*\mu(E) = \mu(f^{-1}(E))$  for any set  $E$  measurable in  $\mathbf{Y}$ . An important use of push-forward measures is that for any integrable function  $g: \mathbf{Y} \rightarrow \langle \mathbb{R}, \mathcal{B}_{\mathbb{R}} \rangle$ , it allows us to write the following change-of-variables formula

$$\int_{\mathbf{Y}} g \, df_*\mu = \int_{\mathbf{X}} g \circ f \, d\mu. \quad (2)$$

The push-forward operation preserves the total measure, hence it takes  $\mathbb{P}(\mathbf{X})$  to  $\mathbb{P}(\mathbf{Y})$ . A case that will be of particular interest to us is the push-forward of a measure  $\mu$  on a product space  $\mathbf{X}_1 \times \mathbf{X}_2$  along a projection  $\pi_i: \mathbf{X}_1 \times \mathbf{X}_2 \rightarrow \mathbf{X}_i$ : this yields the **marginal measure**  $\mu|_{\mathbf{X}_i} = \pi_{i*}\mu$ , where e.g. for  $E$  measurable in  $\mathbf{X}_1$ ,  $\mu|_{\mathbf{X}_1}(E) = \mu(\pi_1^{-1}(E)) = \mu(E \times X_2)$ . In the opposite direction, given measures  $\mu_1$  on  $\mathbf{X}_1$  and  $\mu_2$  on  $\mathbf{X}_2$ , a **product measure**  $\mu_1 \times \mu_2$  is a measure on the product measurable space  $\mathbf{X}_1 \times \mathbf{X}_2$  satisfying  $(\mu_1 \times \mu_2)(E_1 \times E_2) = \mu_1(E_1)\mu_2(E_2)$  for all  $E_1 \in \mathcal{F}_1$  and  $E_2 \in \mathcal{F}_2$ . For probability measures, there is a uniquely determined product measure.<sup>2</sup>

We can view  $\mathbb{M}$  as a map that takes a measurable space to the set of measures on that space, and similarly for  $\mathbb{P}$ . These become functors  $\text{Meas} \rightarrow \text{Set}$  if we define the action on morphisms to be the push-forward operation. Explicitly, for  $f: \mathbf{X} \rightarrow \mathbf{Y}$  a measurable function, we set  $\mathbb{M}(f) := f_*: \mathbb{M}(\mathbf{X}) \rightarrow \mathbb{M}(\mathbf{Y}) :: \mu \mapsto f_*\mu$ , and similarly for  $\mathbb{P}$ . Remarkably, the set  $\mathbb{P}(\mathbf{X})$  of probability measures on  $\mathbf{X}$  can itself be made into a measurable space by equipping it with the least  $\sigma$ -algebra that makes the evaluation functions  $\text{ev}_E: \mathbb{P}(\mathbf{X}) \rightarrow [0, 1] :: \mu \mapsto \mu(E)$  measurable for all  $E \in \mathcal{F}_X$ .<sup>3</sup> This yields an endofunctor  $\mathbb{P}: \text{Meas} \rightarrow \text{Meas}$ , which moreover has the structure of a monad, called the **Giry monad** [39]. The unit of this monad is given by  $\eta_{\mathbf{X}}: \mathbf{X} \rightarrow \mathbb{P}(\mathbf{X}) :: x \mapsto \delta_x$  where  $\delta_x$  is the **Dirac measure**, or point mass, at  $x$  given by  $\delta_x(E) := \chi_E(x)$ . Multiplication of the monad is given by  $\mu_{\mathbf{X}}: \mathbb{P}(\mathbb{P}(\mathbf{X})) \rightarrow \mathbb{P}(\mathbf{X})$  which takes a probability measure  $P$  on  $\mathbb{P}(\mathbf{X})$  to its ‘average’, a probability measure  $\mu_{\mathbf{X}}(P): \mathcal{F}_X \rightarrow [0, 1]$  on  $\mathbf{X}$  whose value on a measurable set  $E \in \mathcal{F}_X$  is given by  $\mu_{\mathbf{X}}(P)(E) := \int_{\mathbb{P}(\mathbf{X})} \text{ev}_E \, dP$ .

The Kleisli category of this monad is the category of Markov kernels, which represent continuous-variable probabilistic maps and generalise the discrete notion of stochastic matrix. Concretely, a **Markov kernel** between measurable spaces  $\mathbf{X} = \langle X, \mathcal{F}_X \rangle$  and  $\mathbf{Y} = \langle Y, \mathcal{F}_Y \rangle$  is a function  $k: X \times \mathcal{F}_Y \rightarrow [0, 1]$  such that:

- (i) for all  $E \in \mathcal{F}_Y$ ,  $k(-, E): X \rightarrow [0, 1]$  is a measurable function;<sup>4</sup>
- (ii) for all  $x \in X$ ,  $k(x, -): \mathcal{F}_Y \rightarrow [0, 1]$  is a probability measure.

### 3 Framework

In this section we will follow closely the discrete-variable framework of [8] in more formally describing the kinds of experimental scenarios in which we are interested and the empirical behaviours that arise on these, although some extra care is required for continuous variables.

<sup>1</sup>For a comprehensive treatment we refer the reader to e.g. [21], or for a beautiful and more concise introduction aimed particularly at computer scientists to [62].

<sup>2</sup>In fact, this holds more generally for  $\sigma$ -finite measures, i.e. when  $X$  is a countable union of sets of finite measure.

<sup>3</sup>More concretely, it is the  $\sigma$ -algebra generated by the sets  $\text{ev}_E^{-1}([0, r]) = \{\mu \in \mathbb{P}(X) \mid \mu(E) < r\}$  with  $E \in \mathcal{F}_X$  and  $r \in [0, 1]$ .

<sup>4</sup>The space  $[0, 1]$  is assumed to be equipped with its Borel  $\sigma$ -algebra.

## Measurement scenarios

**Definition 1.** A **measurement scenario** is a triple  $\langle X, \mathcal{M}, \mathbf{O} \rangle$  whose elements are specified as follows.

- $X$  is a finite set of **measurement labels**.
- $\mathcal{M}$  is a covering family of subsets of  $X$ , i.e. such that  $\bigcup \mathcal{M} = X$ . The elements  $C \in \mathcal{M}$  are called **maximal contexts** and represent maximal sets of compatible observables. We therefore require that  $\mathcal{M}$  be an anti-chain with respect to subset inclusion, i.e. that no element of this family is a proper subset of another. Any subset of a maximal context also represents a set of compatible measurements, and we refer to elements of  $\mathcal{U} := \{U \subseteq C \mid C \in \mathcal{M}\}$  as **contexts**.<sup>5</sup>
- $\mathbf{O} = (\mathbf{O}_x)_{x \in X}$  specifies for each measurement  $x \in X$  a measurable **space of outcomes**  $\mathbf{O}_x = \langle \mathcal{O}_x, \mathcal{F}_x \rangle$ .

Measurement scenarios can be understood as providing a concise description of the kind of experimental setup that is being considered. For example, the setup represented in Figure 1 is described by the measurement scenario:

$$X = \{a, a', b, b'\}, \quad \mathcal{M} = \{\{a, b\}, \{a, b'\}, \{a', b\}, \{a', b'\}\}, \quad \mathbf{O}_x = \mathbf{R}, \quad (3)$$

where  $\mathbf{R}$  is a bounded measurable subspace of  $\mathbb{R}$ .

If some set of measurements  $U \subseteq X$  is considered together, there is a joint outcome space given by the product of the respective outcome spaces:

$$\mathbf{O}_U := \prod_{x \in U} \mathbf{O}_x = \langle \mathcal{O}_U, \mathcal{F}_U \rangle = \left\langle \prod_{x \in U} \mathcal{O}_x, \bigotimes_{x \in U} \mathcal{F}_x \right\rangle.$$

The map  $\mathcal{E}$  that maps  $U \subseteq X$  to  $\mathcal{E}(U) = \mathbf{O}_U$  is called the **event sheaf** as concretely it assigns to any set of measurements information about the outcome events that could result from jointly performing them. Note that as well as applying the map to valid contexts  $U \in \mathcal{U}$  we will see that it can also be of interest to consider hypothetical outcome spaces for sets of measurements that do not necessarily form valid contexts, in particular  $\mathcal{E}(X) = \mathbf{O}_X$ , the joint outcome space for *all* measurements. Moreover, as we will briefly discuss, this map satisfies the conditions to be a sheaf  $\mathcal{E}: \mathcal{P}(X)^{\text{op}} \rightarrow \text{Meas}$  where  $\mathcal{P}(X)$  denotes the powerset of  $X$ , similarly to its discrete-variable analogue in [8].

## The language of sheaves

Sheaves are widely used in modern mathematics. They might roughly be thought of as providing a means of assigning information to the open sets of a topological space in such a way that information can be restricted to subsets and consistent information can be ‘glued’ on unions<sup>6</sup>. In this work we are concerned with discrete topological spaces whose points represent measurements, and the information that we are interested in assigning has to do with outcome spaces for these measurements and probability measures on these outcome spaces. Sheaves can be defined concisely in category-theoretic terms as contravariant functors (presheaves) satisfying an additional gluing condition, though in what follows we will also give a more concrete description in terms of restriction maps. Categorically, the event sheaf is a functor  $\mathcal{E}: \mathcal{P}(X)^{\text{op}} \rightarrow \text{Meas}$  where  $\mathcal{P}(X)$  is viewed as a category in the standard way for partial orders, with morphisms corresponding to subset inclusions.

Sheaves come with a notion of **restriction**. In our example restriction arises in the following way: whenever  $U \subseteq V$  we have an obvious restriction map  $\rho_U^V: \mathcal{E}(V) \rightarrow \mathcal{E}(U)$  which simply projects from the product outcome space for  $V$  to that for  $U$ . Note that  $\rho_U^U$  is the identity map and that if  $U \subseteq V \subseteq W$  then  $\rho_U^V \circ \rho_V^W = \rho_U^W$ . Already this is enough to show that  $\mathcal{E}$  is a *presheaf*. In categorical terms it establishes functoriality. For  $U \subseteq V$  and  $\mathbf{o} \in \mathcal{O}_V$  it is often more convenient to use the notation  $\mathbf{o}|_U$  to denote  $\rho_U^V(\mathbf{o})$ . Our map assigns outcome spaces  $\mathcal{E}(U) = \mathbf{O}_U$  to sets of measurements  $U$ , and in sheaf and presheaf terminology elements of these outcome spaces are called **sections** over  $U$ . Sections over  $X$  are called **global sections**.

Additionally, the unique gluing property holds for  $\mathcal{E}$ . Suppose we have a family of sections  $\{\mathbf{o}_U \in \mathcal{O}_U\}_{U \in \mathcal{N}}$  that is compatible in the sense that its assignments agree on restrictions, i.e.  $\mathbf{o}_U|_{U \cap V} = \mathbf{o}_V|_{U \cap V}$  for all  $U, V \in \mathcal{N}$ . Then these sections can always be ‘glued’ together in a unique fashion to obtain a section  $\mathbf{o}_N$  over  $N := \bigcup \mathcal{N}$  such that  $\mathbf{o}_N|_U = \mathbf{o}_U$  for all  $U \in \mathcal{N}$ . This makes  $\mathcal{E}$  a **sheaf**.

We will primarily be concerned with probability measures on outcome spaces. For this, we recall that the Giry monad  $\mathbb{P}: \text{Meas} \rightarrow \text{Meas}$  takes a measurable space and returns the probability measures over that space. Composing it with the event sheaf yields the map  $\mathbb{P} \circ \mathcal{E}$  that takes any context and returns the probability measures on its joint outcome space. In fact, this is a presheaf  $\mathbb{P} \circ \mathcal{E}: \mathcal{P}(X)^{\text{op}} \rightarrow \text{Meas}$ , where restriction on sections is given by marginalisation of probability measures. Note that marginalisation simply corresponds to the push-forward of a measure along projections to a component of the product space, which are precisely the restriction maps of  $\mathcal{E}$ . Note, however, that this presheaf does not satisfy the gluing condition and thus it crucially is not a sheaf.

<sup>5</sup>While it is more convenient to specify  $\mathcal{M}$ , note that the set of contexts  $\mathcal{U}$  carries exactly the same information. It forms an abstract simplicial complex whose simplices are the contexts and whose facets are the maximal context. This combinatorial topological structure is emphasised in some presentations [15, 16, 29, 47, 4].

<sup>6</sup>For a comprehensive reference on sheaf theory see e.g. [54].

## Empirical models

**Definition 2.** An **empirical model** on a measurement scenario  $\langle X, \mathcal{M}, \mathbf{O} \rangle$  is a compatible family for the presheaf  $\mathbb{P} \circ \mathcal{E}$  on the cover  $\mathcal{M}$ . Concretely, it is a family  $e = \{e_C\}_{C \in \mathcal{M}}$ , where  $e_C$  is a probability measure on the space  $\mathcal{E}(C) = \mathbf{O}_C$  for each maximal context  $C \in \mathcal{M}$ , which satisfies the compatibility condition:

$$e_C|_{C \cap C'} = e_{C'}|_{C \cap C'}.$$

Empirical models capture in a precise way the probabilistic **behaviours** that may arise upon performing measurements on physical systems. The compatibility condition ensures that the empirical behaviour of a given measurement or compatible subset of measurements is independent of which other compatible measurements might be performed along with them. This is sometimes referred to as the no-disturbance condition. A special case is **no-signalling**, which applies in multi-party or Bell scenarios such as that of Figure 1 and Eq. (3). In that case, contexts consist of measurements that are supposed to occur in space-like separated locations, and compatibility ensures for instance that the choice of performing  $a$  or  $a'$  at the first location does not affect the empirical behaviour at the second location, i.e.  $e_{\{a,b\}}|_{\{b\}} = e_{\{a',b\}}|_{\{b\}}$ .

Note also that while empirical models may arise from the predictions of quantum theory, their definition is theory-independent. This means that empirical models can just as well describe hypothetical behaviours beyond what can be achieved by quantum mechanics such as the well-studied Popescu–Rohrlich box [64]. This can be useful in probing the limits of quantum theory, and for singling out what distinguishes and characterises quantum theory within larger spaces of probabilistic theories, both established lines of research in quantum foundations.

**Sheaf-theoretically.** An empirical model is a compatible family of sections for the presheaf  $\mathbb{P} \circ \mathcal{E}$  indexed by the maximal contexts of the measurement scenario. A natural question that may occur at this point is whether these sections can be glued to form a global section, and this is what we address next.

## Extendability and contextuality

**Definition 3.** An empirical model  $e$  on a scenario  $\langle X, \mathcal{M}, \mathbf{O} \rangle$  is **extendable** (or **noncontextual**) if there is a probability measure  $\mu$  on the space  $\mathcal{E}(X) = \mathbf{O}_X$  such that  $\mu|_C = e_C$  for every  $C \in \mathcal{M}$ .<sup>7</sup>

Recall that  $\mathbf{O}_X$  is the global outcome space, whose elements correspond to global assignments of outcomes to all measurements. Of course, it is not always the case that  $X$  is a valid context, and if it were then  $\mu = e_X$  would trivially extend the empirical model. The question of the existence of such a probability measure that recovers the context-wise empirical content of  $e$  is particularly significant. When it exists, it amounts to a way of modelling the behaviour as arising stochastically from the behaviours of underlying states, identified with the elements of  $\mathbf{O}_X$ , each of which deterministically assigns outcomes to *all* the measurements in  $X$  independently of the context that is actually performed. If an empirical model is not extendable it is said to be **contextual**.

**Sheaf-theoretically.** A contextual empirical model is a compatible family of sections for the presheaf  $\mathbb{P} \circ \mathcal{E}$  over the contexts of the measurement scenario that cannot be glued to form a global section. Contextuality thus arises as the tension between local consistency and global inconsistency.

## 4 A FAB theorem

Quantum theory presents a number of non-intuitive features. For instance, Einstein, Podolsky and Rosen (EPR) identified early on that if the quantum description of the world is taken as fundamental then entanglement poses a problem of “spooky action at a distance” [35]. Their conclusion was that quantum theory should be consistent with a deeper or more complete description of the physical world, in which such problems would disappear. The import of seminal foundational results like the Bell [18] and Bell–Kochen–Specker [19, 50] theorems is that they identify such non-intuitive behaviours and then rule out the possibility of finding *any* underlying model for them that would not suffer from the same issues. Incidentally, we note that the EPR paradox was originally presented in terms of continuous variables, whereas Bell’s theorem addressed a discrete variable analogue of it.

In the previous section, we characterised contextuality of an empirical model by the absence of a global section for that empirical model. We also saw that global sections capture one kind of underlying model for the behaviour, namely via deterministic global states that assign predefined outcomes to all measurements, which is precisely the kind of model referred to in the Kochen–Specker theorem [50]. Bell’s theorem, on the other hand, related to a different kind of hidden-variable model, where the salient feature – Bell locality – was a kind of factorisability rather than determinism. Fine [36] showed that in one important measurement scenario (that of the concrete example from Fig. 1) the existence of one kind of model is equivalent to existence of the other. Abramsky and Brandenburger [8] proved in full generality that

<sup>7</sup>Notions of partial extendability have also been considered in the discrete setting in [56, 70].

this existential equivalence holds for any measurement scenario, and that global sections of  $\mathbb{P} \circ \mathcal{E}$  provide a canonical form of hidden-variable model.

In this section, we prove a Fine–Abramsky–Brandenburger theorem in the continuous-variable setting. It establishes that in this setting there is also an unambiguous, unified description of locality and noncontextuality, which is captured in a canonical way by the notion of extendability.

We will begin by introducing hidden-variable models in a more precise way. The idea is that there exists some space  $\Lambda$  of hidden variables, which determine the empirical behaviour. However, elements of  $\Lambda$  may not be directly empirically accessible themselves, so we allow that we might only have probabilistic information about them in the form of a probability measure  $p$  on  $\Lambda$ . The empirically observable behaviour should then arise as an average over the hidden-variable behaviours.

**Definition 4.** Let  $\langle X, \mathcal{M}, \mathcal{O} \rangle$  be a measurement scenario. A **hidden-variable model**<sup>8</sup> on this scenario consists of the following ingredients:

- A measurable space  $\Lambda = \langle \Lambda, \mathcal{F}_\Lambda \rangle$  of **hidden variables**.
- A probability measure  $p$  on  $\Lambda$ .
- For each maximal context  $C \in \mathcal{M}$ , a probability kernel  $k_C: \Lambda \rightarrow \mathcal{E}(C)$ , satisfying the following compatibility condition:<sup>9</sup>

$$\forall \lambda \in \Lambda. \quad k_C(\lambda, -)|_{C \cap C'} = k_{C'}(\lambda, -)|_{C \cap C'}. \quad (4)$$

**Remark 5.** Equivalently, we can regard Eq. (4) as defining a function  $\underline{k}$  from  $\Lambda$  to the set of empirical models over  $\langle X, \mathcal{M}, \mathcal{O} \rangle$ . The function assigns to each  $\lambda \in \Lambda$  the empirical model  $\underline{k}(\lambda) := (\underline{k}(\lambda)_C)_{C \in \mathcal{M}}$ , where the correspondence with the definition above is via  $\underline{k}(\lambda)_C = k_C(\lambda, -)$ . This function must be ‘measurable’ in  $\Lambda$  in the sense that  $\underline{k}(-)_C(B): \Lambda \rightarrow [0, 1]$  is a measurable function for all  $C \in \mathcal{M}$  and  $B \in \mathcal{F}_C$ .

**Definition 6.** Let  $\langle X, \mathcal{M}, \mathcal{O} \rangle$  be a measurement scenario and  $\langle \Lambda, p, k \rangle$  be a hidden-variable model. Then the corresponding empirical model  $e$  is given by

$$e_C(B) = \int_{\Lambda} k_C(-, B) \, dp = \int_{\lambda \in \Lambda} k_C(\lambda, B) \, dp(\lambda).$$

Note that our definition of hidden-variable model assumes the properties known as  $\lambda$  independence [31] and parameter-independence [45, 69]. The former corresponds to the fact that the probability measure  $p$  on the hidden-variable space is independent of the measurement context to be performed, while the latter corresponds to the compatibility condition (4), which also ensures that the corresponding empirical model is no-signalling [23]. We refer the reader to [24] for a detailed discussion of these and other properties of hidden-variable models specifically in the case of multi-party Bell scenarios.

The idea behind the introduction of hidden variables is that they could *explain away* some of the more non-intuitive aspects of the empirical predictions of quantum mechanics, which would then arise as resulting from an incomplete knowledge of the true state of a system rather than being a fundamental feature. There is some precedent for this in physical theories: for instance, statistical mechanics – a probabilistic theory – admits a deeper, albeit usually unwieldily complex, description in terms of classical mechanics which is purely deterministic. Therefore it is desirable to impose conditions on hidden-variable models which amount to requiring that they behave in some sense *classically* when conditioned on each particular value of the hidden variable  $\lambda$ . This motivates the notions of deterministic and of factorisable hidden-variable models.

**Definition 7.** A hidden-variable model  $\langle \Lambda, p, k \rangle$  is **deterministic** if  $k_C(\lambda, -): \mathcal{F}_C \rightarrow [0, 1]$  is a Dirac measure for every  $\lambda \in \Lambda$  and for every maximal context  $C \in \mathcal{M}$ ; in other words, there is an assignment  $\mathbf{o} \in \mathcal{O}_C$  such that  $k_C(\lambda, -) = \delta_{\mathbf{o}}$ .

In general discussions on hidden-variable models (e.g. [24]), the condition above, requiring that each hidden variable determines a unique joint outcome for each measurement context, is sometimes referred to as weak determinism. This is contraposed to strong determinism, which demands not only that each hidden variable fix a deterministic outcome to each individual measurement, but that this outcome be independent of the context in which the measurement is performed. Note, however, that since our definition of hidden-variable models assumes the compatibility condition of (4) (i.e. parameter-independence), both notions of determinism coincide [23].

**Definition 8.** A hidden-variable model  $\langle \Lambda, p, k \rangle$  is **factorisable** if  $k_C(\lambda, -): \mathcal{F}_C \rightarrow [0, 1]$  factorises as a product measure for every  $\lambda \in \Lambda$  and for every maximal context  $C \in \mathcal{M}$ . That is, for any family of measurable sets  $(B_x \in \mathcal{F}_x)_{x \in C}$ ,

$$k_C(\lambda, \prod_{x \in C} B_x) = \prod_{x \in C} k_C|_{\{x\}}(\lambda, B_x)$$

where  $k_C|_{\{x\}}(\lambda, -)$  is the marginal of the probability measure  $k_C|_{\{x\}}(\lambda, -)$  on  $\mathcal{O}_C = \prod_{x \in C} \mathcal{O}_x$  to the space  $\mathcal{O}_{\{x\}} = \mathcal{O}_x$ .<sup>10</sup>

<sup>8</sup>The alternative term **ontological model** [72] has become widely used in quantum foundations in recent years. It indicates that the hidden variable, sometimes referred to as the ontic state, is supposed to provide an underlying description of the physical world at perhaps a more fundamental level than the empirical-level description, via the quantum state for example.

<sup>9</sup>Recall from Section 2 that a probability kernel  $k_C: \Lambda \rightarrow \mathcal{E}(C)$  is a function  $k_C: \Lambda \times \mathcal{F}_C \rightarrow [0, 1]$  which is a measurable function in the first argument and a probability measure in the second argument.

**Remark 9.** In other words, if we consider elements of  $\Lambda$  as inaccessible ‘empirical’ models – i.e. if we use the alternative definition of hidden-variable models using the map  $\underline{k}$ , see Remark 5 – then factorisability is the requirement that each of these be factorisable in the sense that

$$\underline{k}_C(\lambda) \left( \prod_{x \in C} B_x \right) = \prod_{x \in C} \underline{k}_C(\lambda)|_{\{x\}}(B_x)$$

where  $\underline{k}_C|_{\{x\}}(\lambda)$  is the marginal of the probability measure  $\underline{k}_C(\lambda)$  on  $\mathbf{O}_C = \prod_{x \in C} \mathbf{O}_x$  to the space  $\mathbf{O}_x$ .

We now prove the continuous-variable analogue of the theorem proved in the discrete probability setting Abramsky and Brandenburger [8, Proposition 3.1 and Theorem 8.1], generalising the result of Fine [36] to arbitrary measurement scenarios.

**Theorem 10.** *Let  $e$  be an empirical model on a measurement scenario  $\langle X, \mathcal{M}, \mathbf{O} \rangle$ . The following are equivalent:*

- (1)  $e$  is extendable;
- (2)  $e$  admits a realisation by a deterministic hidden-variable model;
- (3)  $e$  admits a realisation by a factorisable hidden-variable model.

*Proof.* We prove the sequence of implications (1)  $\Rightarrow$  (2)  $\Rightarrow$  (3)  $\Rightarrow$  (1).

(1)  $\Rightarrow$  (2). The idea is that  $\mathcal{E}(X) = \mathbf{O}_X$  provides a canonical deterministic hidden-variable space. Suppose that  $e$  is extendable to a global probability measure  $\mu$ . Let us set  $\mathbf{\Lambda} := \mathbf{O}_X$ , set  $p := \mu$ , and set  $k_C(\mathbf{g}, -) := \delta_{\mathbf{g}|_C}$  for all global outcome assignments  $\mathbf{g} \in \mathbf{O}_X$ . This is by construction a deterministic hidden-variable model, which we claim gives rise to the empirical model  $e$ .

Let  $C \in \mathcal{M}$  and write  $\rho : \mathbf{O}_X \rightarrow \mathbf{O}_C$  for the measurable projection which, in the event sheaf, is the restriction map  $\rho_C^X = \mathcal{E}(C \subseteq X) : \mathcal{E}(X) \rightarrow \mathcal{E}(C)$ .

For any  $E \in \mathcal{F}_C$ , we have

$$k_C(\mathbf{g}, E) = \delta_{\mathbf{g}|_C}(E) = \delta_{\rho(\mathbf{g})}(E) = \chi_E(\rho(\mathbf{g})) = (\chi_E \circ \rho)(\mathbf{g}) \quad (5)$$

and therefore, as required,

$$\begin{aligned} & \int_{\Lambda} k_C(-, E) \, dp & & = & & \{ \text{marginalisation for probability measures} \} \\ & = & \{ \text{Eq. (5)} \} & & & \int_{\mathbf{O}_C} \chi_E \, d\mu|_C \\ & \int_{\mathbf{O}_X} \chi_E \circ \rho \, d\mu & & = & & \{ \text{integral of indicator function, Eq. (1)} \} \\ & = & \{ \text{change of variables, Eq. (2)} \} & & & \mu|_C(E) \\ & \int_{\mathbf{O}_C} \chi_E \, d\rho_*\mu & & = & & \{ \mu \text{ extends the empirical model } e \} \\ & & & & & e_C(E). \end{aligned}$$

(2)  $\Rightarrow$  (3). It is enough to show that if a hidden-variable model  $\langle \mathbf{\Lambda}, p, k \rangle$  is deterministic then it is also factorisable. For this, it is sufficient to notice that a Dirac measure  $\delta_{\mathbf{o}}$  with  $\mathbf{o} \in \mathbf{O}_C$  on a product space  $\mathbf{O}_C = \prod_{x \in C} \mathbf{O}_x$  factorises as the product of Dirac measures

$$\delta_{\mathbf{o}} = \prod_{x \in C} \delta_{\mathbf{o}|_{\{x\}}} = \prod_{x \in C} \delta_{\mathbf{o}(x)}$$

(3)  $\Rightarrow$  (1). Suppose that  $e$  is realised by a factorisable hidden-variable model  $\langle \mathbf{\Lambda}, p, k \rangle$ . Write  $k_x$  for  $k_C|_{\{x\}}$  as in the definition of factorisability. Define  $\mu$  on  $\mathbf{O}_X$  as follows: for any family of measurable sets  $(E_x \in \mathcal{F}_x)_{x \in X}$ , the value of  $\mu$  on the corresponding rectangle is given by

$$\mu \left( \prod_{x \in X} E_x \right) = \int_{\Lambda} \left( \prod_{x \in X} k_x(-, E_x) \right) \, dp \quad (6)$$

Next we will show that this is a global section for the empirical probabilities. Let  $C \in \mathcal{M}$  and consider a family of measurable sets  $\{F_x \in \mathcal{F}_x\}_{x \in C}$  and let  $F = \prod_{x \in C} F_x \in \mathcal{F}_C$  be the corresponding rectangle. Then

<sup>10</sup> Note that, due to the assumption of parameter independence (Eq. (4)), we can unambiguously write  $k_x(\lambda, -)$  for  $k_C|_{\{x\}}$ , as this marginal is independent of the context  $C$  from which one is restricting.

$$\begin{aligned}
& \mu|_C(F) \\
= & \quad \{ \text{definition of marginalisation} \} \\
& \mu(F \times O_{X \setminus C}) \\
= & \quad \{ \text{definition of } F \text{ and } O_U \} \\
& \mu\left(\prod_{x \in C} F_x \times \prod_{x \in X \setminus C} O_x\right) \\
= & \quad \{ \text{definition of } \mu, \text{ Eq. (6)} \} \\
& \int_{\Lambda} \left( \prod_{x \in C} k_x(-, F_x) \right) \left( \prod_{x \in X \setminus C} k_x(-, O_x) \right) \mathrm{d}p \\
& \quad \left| \right. \\
= & \quad \{ k_x(\lambda, -) \text{ probability measure so } k_x(\lambda, O_x) = 1 \} \\
& \int_{\Lambda} \left( \prod_{x \in C} k_x(-, F_x) \right) \mathrm{d}p \\
= & \quad \{ \text{factorisability of the hidden-variable model} \} \\
& \int_{\Lambda} k_C(-, \prod_{x \in C} F_x) \mathrm{d}p \\
= & \quad \{ \text{definition of } F \} \\
& \int_{\Lambda} k_C(-, F) \mathrm{d}p \\
= & \quad \{ e, \text{ empirical model corresponding to } \langle \Lambda, p, k \rangle \} \\
& e_C(F)
\end{aligned}$$

Since the  $\sigma$ -algebra  $\mathcal{F}_C$  of  $O_C$  is generated by the rectangles (such as  $F$  above) and we have seen that  $\mu|_C$  agrees with  $e_C$  on these sets, we conclude that  $\mu|_C = e_C$  as required.  $\square$

## 5 Quantifying contextuality

Beyond questioning whether a given empirical behaviour is contextual or not, it is also interesting to ask to what *degree* it is contextual. In discrete-variable scenarios, a very natural measure of contextuality is the contextual fraction [8]. This measure was shown in [6] to have a number of very desirable properties. It can be calculated using linear programming, an approach that subsumes the more traditional approach to quantifying nonlocality and contextuality using Bell and noncontextual inequalities in the sense that we can understand the (dual) linear program as optimising over *all* such inequalities for the scenario in question and returning the maximum normalised violation of *any* Bell or noncontextuality inequality achieved by the given empirical model. Crucially, the contextual fraction was also shown to *quantifiably* relate to quantum-over-classical advantages in specific informatic tasks [6, 58, 75]. Moreover it has been demonstrated to be a monotone with respect to the free operations of resource theories for contextuality [6, 32, 4].

In this section, we consider how to carry those ideas to the continuous-variable setting. The formulation as a linear optimisation problem and the attendant correspondence with Bell inequality violations requires special care as one needs to use infinite linear programming, necessitating some extra assumptions on the outcome measurable spaces.

### The contextual fraction

Asking whether a given behaviour is noncontextual amounts to asking whether the empirical model is extendable, or in other words whether it admits a deterministic hidden-variable model. However, a more refined question to pose is *what fraction of the behaviour admits a deterministic hidden-variable model?* This quantity is what we call the noncontextual fraction. Similarly, the fraction of the behaviour that is left over and that can thus be considered *irreducibly* contextual is what we call the contextual fraction.

**Definition 11.** Let  $e$  be an empirical model on the scenario  $\langle X, \mathcal{M}, \mathbf{O} \rangle$ . The **noncontextual fraction** of  $e$ , written  $\text{NCF}(e)$ , is defined as

$$\sup \{ \mu(O_X) \mid \mu \in \mathbb{M}(\mathbf{O}_X), \forall C \in \mathcal{M}. \mu|_C \leq e_C \} .$$

Note that since  $e_C \in \mathbb{P}(\mathbf{O}_C)$  for all  $C \in \mathcal{M}$  it follows that  $\text{NCF}(e) \in [0, 1]$ . The **contextual fraction** of  $e$ , written  $\text{CF}(e)$ , is given by  $\text{CF}(e) := 1 - \text{NCF}(e)$ .

### Monotonicity under free operations including binning

In the discrete-variable setting, the contextual fraction was shown to be a monotone under a number of natural classical operations that transform and combine empirical models and control their use as resources, therefore constituting the ‘free’ operations of a resource theory of contextuality [6, 32, 4].

All of the operations defined for discrete variables in [6] – viz. translations of measurements, transformation of outcomes, probabilistic mixing, product, and choice – carry almost verbatim to our current setting. One detail is that one must insist that the coarse-graining of outcomes be achieved by (a family of) measurable functions. A particular example of practical importance is **binning**, which is widely used in continuous-variable quantum information as a method of discretising data by partitioning the outcome space  $\mathbf{O}_x$  for each measurement  $x \in X$  into a finite number of ‘bins’, i.e. measurable sets. Note that a binned empirical model is obtained by pushing forward along a family

$(t_x)_{x \in X}$  of outcome translations  $t_x: \mathcal{O}_x \rightarrow \mathcal{O}'_x$  where  $\mathcal{O}'_x$  is finite for all  $x \in X$ . For the conditional measurement operation introduced in [4], which allows for adaptive measurement protocols such as those used in measurement-based quantum computation [67], one must similarly insist that the map determining the next measurement to perform based on the observed outcome of a previous measurement be a measurable function. Since we are, for the moment, only considering scenarios with a finite number of measurements, this amounts to a partition into measurable subsets of the outcome space  $\mathcal{O}_x$  of the first measurement,  $x$ , indexed by measurements compatible with  $x$ , indicating which will be subsequently performed depending on the outcome observed for  $x$ .

The inequalities establishing monotonicity from [6, Theorem 2] will also hold for continuous variables. There is a caveat for the equality formula for the product of two empirical models,  $\text{NCF}(e_1 \otimes e_2) = \text{NCF}(e_1)\text{NCF}(e_2)$ . Whereas the inequality establishing monotonicity ( $\geq$ ) stills holds in general, the proof establishing the other direction ( $\leq$ ) makes use of duality of linear programs. Therefore, it will only hold under the assumptions we will impose in the remainder of this section.

**Proposition 12.** *If  $e$  is an empirical model, and  $e^{\text{bin}}$  is any discrete-variable empirical model obtained from  $e$  by binning, then contextuality of  $e^{\text{bin}}$  witnesses contextuality of  $e$ , and quantifiably gives a lower bound  $\text{CF}(e^{\text{bin}}) \leq \text{CF}(e)$ .*

### Assumptions on the outcome spaces

In order to phrase the problem of contextuality as an (infinite) linear programming problem and establish the connection with violations of Bell inequalities, we need to impose some conditions on the measurable spaces of outcomes. From now on, we restrict attention to the case where the outcome space  $\mathcal{O}_x$  for each measurement  $x \in X$  is the Borel measurable space for a second-countable locally compact Hausdorff space, i.e. that the set  $\mathcal{O}_x$  is equipped with a second-countable locally compact Hausdorff topology and  $\mathcal{F}_x$  is the  $\sigma$ -algebra generated by its open sets, written  $\mathcal{B}(\mathcal{O}_x)$ . Note that this includes most situations of interest in practice. In particular, it includes the case of measurements with outcomes in  $\mathbb{R}$  or  $\mathbb{R}^n$  or a bounded subset of these.

Second countability and Hausdorffness of two spaces  $Y$  and  $Z$  suffice to show that  $\mathcal{B}(Y \times Z) = \mathcal{B}(Y) \otimes \mathcal{B}(Z)$ , i.e. the Borel  $\sigma$ -algebra of the product topology is the tensor product of the Borel  $\sigma$ -algebras [22, Lemma 6.4.2 (Vol. 2)]. Hence, these assumptions guarantee that  $\mathcal{O}_U$  is the Borel  $\sigma$ -algebra of the product topology on  $O_U = \prod_{x \in U} \mathcal{O}_x$ . These spaces are also second-countable, locally compact, and Hausdorff as all three properties are preserved by finite products.

In order to phrase the problem as an infinite linear program, we need to work with vector spaces. However, probability measures, or even finite or arbitrary measures, do not form one. We will therefore consider the set  $\mathbb{M}_{\pm}(\mathbf{Y})$  of **finite signed measures** (a.k.a. real measures) on a measurable space  $\mathbf{Y} = \langle Y, \mathcal{F}_Y \rangle$ . These are functions  $\mu: \mathcal{F}_Y \rightarrow \mathbb{R}$  such that  $\mu(\emptyset) = 0$  and  $\mu$  is  $\sigma$ -additive. In comparison to the definition of a measure, one drops the nonnegativity requirement, but insists that the values be finite. The set  $\mathbb{M}_{\pm}(\mathbf{Y})$  forms a real vector space which includes the probability measures  $\mathbb{P}(\mathbf{Y})$ , and total variation gives a norm on this space. When  $Y$  is a second-countable locally compact Hausdorff space and  $\mathbf{Y} = \langle Y, \mathcal{B}(Y) \rangle$ , the Riesz–Markov–Kakutani representation theorem [46] shows that  $\mathbb{M}_{\pm}(\mathbf{Y})$  is a concrete realisation of the topological dual space of  $C_0(Y, \mathbb{R})$ , the space of continuous real-valued functions on  $Y$  that vanish at infinity.<sup>11</sup> The duality is given by  $\langle \mu, f \rangle := \int_Y f \, d\mu$  for  $\mu \in \mathbb{M}_{\pm}(\mathbf{Y})$  and  $f \in C_0(Y, \mathbb{R})$ .<sup>12</sup>

### Linear programming

Calculation of the noncontextual fraction of an empirical model  $e = \{e_C\}_{C \in \mathcal{M}}$  can be expressed as an infinite linear programming problem, (P). This is our primal linear program, which also has a dual linear program given by (D). We will see how to derive the dual and show that the optimal solutions of both programs coincide in what follows. We also refer the interested reader to Appendix A where the programs are expressed in the standard form for infinite linear programming [17].

$$\begin{array}{l}
 \text{(P)} \left\{ \begin{array}{l} \text{Find} \quad \mu \in \mathbb{M}_{\pm}(\mathcal{O}_X) \\ \text{maximising} \quad \mu(\mathcal{O}_X) \\ \text{subject to} \quad \forall C \in \mathcal{M}. \mu|_C \leq e_C \\ \text{and} \quad \mu \geq 0. \end{array} \right.
 \end{array}
 \qquad
 \begin{array}{l}
 \text{(D)} \left\{ \begin{array}{l} \text{Find} \quad (f_C)_{C \in \mathcal{M}} \in \prod_{C \in \mathcal{M}} C_0(\mathcal{O}_C, \mathbb{R}) \\ \text{minimising} \quad \sum_{C \in \mathcal{M}} \int_{\mathcal{O}_C} f_C \, de_C \\ \text{subject to} \quad \sum_{C \in \mathcal{M}} f_C \circ \rho_C^X \geq \mathbf{1} \text{ on } \mathcal{O}_X \\ \text{and} \quad \forall C \in \mathcal{M}. f_C \geq 0 \text{ on } \mathcal{O}_C. \end{array} \right.
 \end{array}$$

We have written  $\rho_C^X$  for the projection  $\mathcal{O}_X \rightarrow \mathcal{O}_C$  as before, and  $\mathbf{1}$  for the constant function that assigns the number 1 to all  $\mathbf{g} \in \mathcal{O}_X$ .<sup>13</sup> We denote the optimal values of problems (P) and (D), respectively, as  $\sup(\text{P}) = \text{NCF}(e)$  and  $\inf(\text{D})$ .

<sup>11</sup>A function  $f: Y \rightarrow \mathbb{R}$  on a locally compact space  $Y$  is said to **vanish at infinity** if the set  $\{y \in Y \mid \|f(y)\| \geq \varepsilon\}$  is compact for all  $\varepsilon > 0$ .

<sup>12</sup>Note that this theorem holds more generally for locally compact Hausdorff spaces if one considers only (finite signed) Radon measures, which are measures that play well with the underlying topology. However, second-countability, together with local compactness and Hausdorffness, guarantees that every Borel measure is Radon [37, Theorem 7.8].

<sup>13</sup>This is just a simplified notation for the indicator function on  $\mathcal{O}_X$ ; i.e.  $\mathbf{1} = \chi_{\mathcal{O}_X}$ .

Analogues of these programs have been studied in the discrete-variable setting [6]. Note however that, in general, these continuous-variable linear programs are over infinite-dimensional spaces and thus not practical to compute directly. For this reason, in Section 6 we will introduce a hierarchy of finite-dimensional semi-definite programs that approximate the solution of (P) to arbitrary precision.

### Bell inequalities and the dual program

The dual program is of particular interest in its own right. As we will now show, it can essentially be understood as computing a continuous-variable ‘Bell inequality’ that is optimised to the empirical model. Making the change of variables  $\beta_C := |\mathcal{M}|^{-1} f_C$  for each  $C \in \mathcal{M}$ , the dual program (D) transforms to the following.

$$(B) \begin{cases} \text{Find} & (\beta_C)_{C \in \mathcal{M}} \in \prod_{C \in \mathcal{M}} C_0(O_C, \mathbb{R}) \\ \text{maximising} & \sum_{C \in \mathcal{M}} \int_{O_C} \beta_C \, d e_C \\ \text{subject to} & \sum_{C \in \mathcal{M}} \beta_C \circ \rho_C^X \leq 0 \text{ on } O_X \\ \text{and} & \forall C \in \mathcal{M}. \beta_C \leq |\mathcal{M}|^{-1} \mathbf{1} \text{ on } O_C. \end{cases}$$

Here  $(\beta_C)_{C \in \mathcal{M}}$  is a family of continuous functions  $\beta_C \in C_0(O_C, \mathbb{R})$ , one for the outcome space of each context. The program maximises, subject to constraints, the combined value obtained by integrating these functionals context-wise against the empirical model in question. The first set of constraints ensures that, for noncontextual empirical models, the value of the program will be at most 0, since any such model extends to a measure  $\mu$  on  $O_X$  such that  $\mu(O_X) = 1$ . The final set of constraints act as a normalisation condition on the value of the program, ensuring that it takes values in the interval  $[0, 1]$  for any empirical model. Any family of functions  $\beta$  satisfying the constraints will thus result in what can be regarded as a generalised Bell inequality,

$$\sum_{C \in \mathcal{M}} \int_{O_C} \beta_C \, d e_C \leq 0,$$

which is satisfied by all noncontextual empirical models.

**Definition 13.** A **generalised Bell inequality**  $(\beta, R)$  on a measurement scenario  $\langle X, \mathcal{M}, \mathbf{O} \rangle$  is a family  $\beta = (\beta_C)_{C \in \mathcal{M}}$  with  $\beta_C \in C_0(O_C, \mathbb{R})$  for all  $C \in \mathcal{M}$ , together with a bound  $R \in \mathbb{R}$ , such that for all noncontextual empirical models  $e$  on  $\langle X, \mathcal{M}, \mathbf{O} \rangle$  it holds that  $\langle \beta, e \rangle_2 := \sum_{C \in \mathcal{M}} \int_{O_C} \beta_C \, d e_C \leq R$ . The **normalised violation** of a generalised Bell inequality  $(\beta, R)$  by an empirical model  $e$  is  $\max\{0, \langle \beta, e \rangle_2\} / (\|\beta\| - R)$  where  $\|\beta\| := \sum_{C \in \mathcal{M}} \|\beta_C\| = \sum_{C \in \mathcal{M}} \sup\{f(\mathbf{o}) \mid \mathbf{o} \in O_C\}$ .<sup>14</sup>

The above definition restricts to the usual notions of Bell inequality and noncontextual inequality in the discrete-variable case and is particularly close to the presentation in [6]. The following theorem also generalises to continuous variables the main result of [6].

**Theorem 14.** *Let  $e$  be an empirical model. (i) The normalised violation by  $e$  of any Bell inequality is at most  $CF(e)$ ; (ii) if  $CF(e) > 0$  then for every  $\varepsilon > 0$  there exists a Bell inequality whose normalised violation by  $e$  is at least  $CF(e) - \varepsilon$ .*

*Proof.* The proof follows directly from the definitions of the linear programs, and from strong duality, i.e. the fact that their optimal values coincide (Proposition 15 below).  $\square$

### Deriving the dual via the Lagrangian

We now give an explicit derivation of (D) as the dual of (P) via the Lagrangian method. To simplify notation, we set  $\tilde{E}_1 := \mathbb{M}_{\pm}(O_X)$  and  $F_2 := \prod_{C \in \mathcal{M}} C_0(O_C, \mathbb{R})$ . This matches the standard form notation for infinite linear programming of [17], in which we present our programs in Appendix A. We do not take into account positivity constraints as they translate directly from primal to dual. Hence we introduce  $|\mathcal{M}|$  dual variables, one continuous map  $f_C \in C_0(O_C, \mathbb{R})$  for each  $C \in \mathcal{M}$ , to account for the constraints  $\mu|_C \leq e_C$ . From (P), we then define the Lagrangian  $\mathcal{L} : \tilde{E}_1 \times F_2 \rightarrow \mathbb{R}$  as

$$\mathcal{L}(\mu, (f_C)) := \underbrace{\mu(O_X)}_{\text{objective}} + \underbrace{\sum_{C \in \mathcal{M}} \int_{O_C} f_C \, d(e_C - \mu|_C)}_{\text{constraints}}. \quad (7)$$

The primal program (P) corresponds to

$$\sup_{\mu \in \tilde{E}_1} \inf_{(f_C) \in F_2} \mathcal{L}(\mu, (f_C)), \quad (8)$$

<sup>14</sup>The notation  $\langle \cdot, \cdot \rangle_2$  is further discussed and explained to be a canonical duality in Appendix A.

as the infimum here imposes the constraints that  $\mu|_C \leq e_C$  for all  $C \in \mathcal{M}$ , for otherwise the Lagrangian diverges. If these constraints are satisfied, then because of the infimum, the second term of the Lagrangian vanishes yielding the objective of the primal problem. To express the dual, which amounts to permuting the infimum and the supremum, we need to rewrite the Lagrangian:

$$\begin{aligned}
\mathcal{L}(\mu, (f_C)) &= \mu(O_X) + \sum_{C \in \mathcal{M}} \int_{O_C} f_C d(e_C - \mu|_C) \\
&= \int_{O_X} \mathbf{1} d\mu + \sum_{C \in \mathcal{M}} \int_{O_C} f_C de_C - \sum_{C \in \mathcal{M}} \int_{O_C} f_C d\mu|_C \\
&= \int_{O_X} \mathbf{1} d\mu + \sum_{C \in \mathcal{M}} \int_{O_C} f_C de_C - \sum_{C \in \mathcal{M}} \int_{O_X} f_C \circ \rho_C^X d\mu \\
&= \int_{O_X} \mathbf{1} d\mu + \sum_{C \in \mathcal{M}} \int_{O_C} f_C de_C - \int_{O_X} \left( \sum_{C \in \mathcal{M}} f_C \circ \rho_C^X \right) d\mu \\
&= \sum_{C \in \mathcal{M}} \int_{O_C} f_C de_C + \int_{O_X} \left( \mathbf{1} - \sum_{C \in \mathcal{M}} f_C \circ \rho_C^X \right) d\mu.
\end{aligned}$$

The dual (D) indeed corresponds to

$$\inf_{(f_C) \in F_2} \sup_{\mu \in \tilde{E}_1} \mathcal{L}(\mu, (f_C)). \quad (9)$$

The supremum imposes that  $\sum_{C \in \mathcal{M}} f_C \circ \rho_C^X \geq \mathbf{1}$  on  $O_X$ , since otherwise the Lagrangian diverges. If this constraint is satisfied, then the supremum makes the second term vanish yielding the objective of the dual problem (D).

### Zero duality gap

A key result about the noncontextual fraction, which is essential in establishing the connection to Bell inequality violations, is that (P) and (D) are strongly dual, in the sense that no gap exists between their optimal values. Strong duality always holds in finite linear programming, but it does not hold in general for the infinite case.

**Proposition 15.** *Problems (P) and (D) have zero duality gap and their optimal values satisfy:*

$$\sup(\text{P}) = \inf(\text{D}) = \text{NCF}(e) \quad (10)$$

*Proof.* This proof relies on [17, Theorem 7.2]. The complete proof is provided in Appendix B. Here, we only provide a brief outline. Let  $E_1 := \mathbb{M}_\pm(O_X) \times \prod_{C \in \mathcal{M}} \mathbb{M}_\pm(O_C)$  and  $E_2 := \prod_{C \in \mathcal{M}} \mathbb{M}_\pm(O_C)$ . Strong duality between (P) and (D) amounts to showing that the cone

$$\mathcal{K} = \{((\mu|_C + \nu_C)_{C \in \mathcal{M}}, \mu(O_X)) \mid (\mu, (\nu_C)_{C \in \mathcal{M}}) \in E_{1+}\}$$

is weakly closed in  $E_2 \oplus \mathbb{R}$ , where  $E_{1+} := \{(\mu, (\nu_C)_{C \in \mathcal{M}}) \in E_1 \mid \mu \geq 0 \text{ and } \forall C \in \mathcal{M}. \nu_C \geq 0\} \subset E_1$ . We do so by considering a sequence  $(\mu^k, (\nu_C^k)_{C \in \mathcal{M}})_{k \in \mathbb{N}}$  in  $E_{1+}$  and showing that the accumulation point

$$\lim_{k \rightarrow \infty} \left( (\mu^k|_C + \nu_C^k)_{C \in \mathcal{M}}, \mu^k(O_X) \right)$$

belongs to  $\mathcal{K}$ . □

## 6 Approximating the contextual fraction with SDPs

In Section 5, we presented the problem of computing the noncontextual fraction as an infinite linear program. Although this is of theoretical importance, it does not allow one to directly perform the actual, numerical computation of this quantity. Here we exploit the link between measures and their sequence of moments to express a hierarchy of truncated semi-definite programming problems which are relaxations of the original problem, in the particular case when the outcome spaces of measurements are certain subsets of  $R^n$ . These finite problems can actually be implemented numerically and have the crucial feature that their optimal values converge monotonically to the noncontextual fraction.

This section makes use of the global optimisation techniques developed by Lasserre and others [51, 42].

### Notation and terminology

Let  $\mathbb{R}[x]$  denote the ring of real polynomials in the variables  $x \in \mathbb{R}^d$ , and let  $\mathbb{R}[x]_k \subset \mathbb{R}[x]$  contain those polynomials of total degree at most  $k$ . The latter forms a vector space of dimension  $s(k) := \binom{d+k}{k}$ , with a canonical basis consisting of

monomials  $x^\alpha = x_1^{\alpha_1} \cdots x_d^{\alpha_d}$  indexed by the set  $\mathbb{N}_k^d := \{\alpha \in \mathbb{N}^d \mid |\alpha| \leq k\}$  where  $|\alpha| := \sum_{i=1}^d \alpha_i$ . Any  $p \in \mathbb{R}[x]_k$  can be expanded in this basis as  $p(x) = \sum_{\alpha \in \mathbb{N}_k^d} p_\alpha x^\alpha$  and we write  $\mathbf{p} := (p_\alpha) \in \mathbb{R}^{s(k)}$  for the resulting vector of coefficients.

Given a sequence  $\mathbf{y} = (y_\alpha)_{\alpha \in \mathbb{N}^d}$  with  $y_\alpha \in \mathbb{R}$ , we define the linear functional  $L_{\mathbf{y}} : \mathbb{R}[x] \rightarrow \mathbb{R}$  by  $L_{\mathbf{y}}(p) := \sum_{\alpha \in \mathbb{N}^d} p_\alpha y_\alpha$ . Let  $\mathbf{K}$  be a Borel measurable subspace of  $\mathbb{R}^n$ . Given a measure  $\mu \in \mathbb{M}_\pm(\mathbf{K})$ , its **moment sequence**  $\mathbf{y} = (y_\alpha)$  is given by

$$y_\alpha := \int_{\mathbf{K}} x^\alpha \, d\mu(x).$$

The linear functional  $L_{\mathbf{y}}$  then gives integration of polynomials with respect to  $\mu$ : for any  $p \in \mathbb{R}[x]$ ,

$$L_{\mathbf{y}}(p) = \sum_{\alpha \in \mathbb{N}^d} p_\alpha y_\alpha = \sum_{\alpha \in \mathbb{N}^d} p_\alpha \int_{\mathbf{K}} x^\alpha \, d\mu(x) = \int_{\mathbf{K}} \sum_{\alpha \in \mathbb{N}^d} p_\alpha x^\alpha \, d\mu(x) = \int_{\mathbf{K}} p(x) \, d\mu(x) = \int_{\mathbf{K}} p \, d\mu.$$

For each  $k \in \mathbb{N}$ , the **moment matrix**  $M_k(\mathbf{y}) \in \text{Mat}_{s(k)}(\mathbb{R})$  of  $\mu$  is then the symmetric matrix with rows and columns indexed by  $\mathbb{N}_k^d$  (i.e. by the canonical basis for  $\mathbb{R}[x]_k$ ) defined by: for any  $\alpha, \beta \in \mathbb{N}_k^d$ ,

$$(M_k(\mathbf{y}))_{\alpha\beta} := L_{\mathbf{y}}(x^{\alpha+\beta}) = y_{\alpha+\beta}.$$

Moreover, given a polynomial  $p \in \mathbb{R}[x]$ , the **localising matrix**  $M_k(p\mathbf{y}) \in \text{Mat}_{s(k)}(\mathbb{R})$  is defined by: for all  $\alpha, \beta \in \mathbb{N}_k^d$ ,

$$(M_k(p\mathbf{y}))_{\alpha\beta} := L_{\mathbf{y}}(p(x)x^{\alpha+\beta}) = \sum_{\gamma \in \mathbb{N}^d} p_\gamma y_{\alpha+\beta+\gamma}.$$

Note that moment matrices are positive semidefinite (see Appendix C).

A polynomial  $p$  is a **sums-of-squares (SOS) polynomial** if there exist polynomials  $\{q_i\}$  such that  $p = \sum_i q_i^2$ . SOS polynomials are widely used in convex optimisation. We will denote by  $\Sigma^2\mathbb{R}[x] \subset \mathbb{R}[x]$  the set of SOS polynomials, and  $\Sigma^2\mathbb{R}[x]_k \subset \Sigma^2\mathbb{R}[x]$  the set of SOS polynomials of degree at most  $2k$ . Finally, the **quadratic module**  $Q((q_j)) \in \mathbb{R}[x]$  generated by a sequence of polynomials  $(q_j)_{j \in \{1, \dots, m\}}$  is defined as  $Q((q_j)) := \left\{ \sigma_0 + \sum_{j=1}^m \sigma_j q_j \mid (\sigma_j)_{j \in \{1, \dots, m\}} \subset \Sigma^2\mathbb{R}[x] \right\}$ .

**Assumptions.** In what follows, we will operate under the assumption that the set of global sections of the event sheaf  $O_X$  is a compact basic semi-algebraic set, i.e. that it can be described by polynomial inequalities:  $O_X = \{x \in \mathbb{R}^d \mid \forall j = 1, \dots, m. p_j(x) \geq 0\}$  for some polynomials  $p_j \in \mathbb{R}[x]$  for which we will write  $r_j := \lfloor \frac{\deg(p_j)}{2} \rfloor$ . We will also assume that there exists an  $a \in \mathbb{R}$  such that the quadratic polynomial  $x \mapsto a^2 - \|x\|^2$  belongs to the quadratic module  $Q((p_j))$ . This amounts to requiring that the set  $O_X$  be bounded. These are standard assumptions for semi-definite programming and in particular for Theorem 17 to hold. Note that both assumptions can be imposed on each set  $O_x$  of outcomes for a single measurement, since these conditions are preserved by products.

### Hierarchy of semidefinite relaxations for computing the noncontextual fraction

As a prerequisite, we first need to compute the sequences of moments associated with all measures derived from the empirical model. For  $C \in \mathcal{M}$ , let  $\mathbf{y}^{e,C} = (y_\alpha^{e,C})_{\alpha \in \mathbb{N}^d}$  be the sequence of all moments of  $e_C$ . We will build a hierarchy of finite semidefinite programs ( $\text{SP}_k$ ) which converge to the optimal solution of the primal (P), i.e. the noncontextual fraction. For a given  $k \in \mathbb{N}$ , we only need to compute a finite number  $s(k)$  of moments. As  $k$  increases, the approximation becomes more precise and the hierarchy of SDPs provides a monotonically decreasing sequence of upper bounds on the noncontextual fraction that converges to its value.

We consider the following semidefinite programs in which  $\mathbf{y}$  is interpreted as corresponding to the moment sequence of a measure  $\mu \in \mathbb{M}_\pm(O_X)$ .

$$(\text{SP}_k) \begin{cases} \sup_{\mathbf{y} \in \mathbb{R}^{s(k)}} y_0 (= \mu(O_X)) \\ \text{s.t. } \forall C \in \mathcal{M}. M_k(\mathbf{y}^{e,C} - \mathbf{y}|_C) \succeq 0 \\ M_k(\mathbf{y}) \succeq 0 \\ \forall j \in \{1, \dots, m\}. M_{k-r_j}(P_j \mathbf{y}) \succeq 0. \end{cases} \quad (\text{SD}_k) \begin{cases} \inf_{\substack{(f_C) \subset \Sigma^2\mathbb{R}[x]_k \\ (\sigma_j) \subset \Sigma^2\mathbb{R}[x]_{k-r_j}}} \sum_{C \in \mathcal{M}} \int_{O_C} f_C \, de_C \\ \text{s.t. } \sum_{C \in \mathcal{M}} f_C - \mathbf{1} = \sigma_0 + \sum_{j=1}^m \sigma_j P_j \end{cases}$$

These problems are dual (see Appendix E). We will denote the optimal values of these programs by  $\sup(\text{SP}_k)$  and  $\inf(\text{SD}_k)$ , respectively.

**Theorem 16.** *The optimal values of the hierarchy of semidefinite programs ( $\text{SD}_k$ ) provide monotonically decreasing upper bounds on the optimal solution of the linear program (D) that converge to its value  $\text{NCF}(e)$ . That is,*

$$\inf(\text{SD}_k) \downarrow \inf(\text{D}) = \text{NCF}(e) \quad \text{as } k \rightarrow \infty. \quad (11)$$

*Proof.* We want to show that we can approximate the problem (D) to arbitrary precision by the problem  $(SD_k)$  by choosing  $k$  sufficiently large. We first use the Stone–Weierstrass theorem to approximate the continuous functions that appear in (D) by polynomials. Then, because of the positivity constraints and the additional assumption on closure of quadratic modules, we can use the SOS based representation (see theorem 17 in Appendix D) to rewrite these polynomials as SOS polynomials. We can thus approximate (D) by  $(SD_k)$  to arbitrary precision. It holds that  $(\inf (SD_k))_{k \in \mathbb{N}}$  decreases monotonically because for all  $k \in \mathbb{N}$ ,  $(SD_k)$  is included in  $(SD_{k+1})$ .  $\square$

Because  $(SD_k)$  is a relaxation of (D) and because problems  $(SP_k)$  and  $(SD_k)$  are dual, we have, respectively:

$$\text{NCF}(e) = \sup(\text{P}) \underset{\substack{\text{strong} \\ \text{duality}}}{=} \inf(\text{D}) \leq \inf(SD_k) \quad (12) \qquad \sup(\text{P}) \leq \sup(SP_k) \leq \inf(SD_k) \quad (13)$$

Thus, Theorem 16 also holds for the primal SDP.

## Outlook

*Logical* forms of contextuality, which are present at the level of the *possibilistic* rather than *probabilistic* information contained in an empirical model, remain to be considered (e.g. [38, 8, 2, 57]). In the discrete setting, these can be treated by analysing ‘possibilistic’ empirical models obtained by considering the supports of the discrete-variable probability distributions [8], which indicate the elements of an outcome space that occur with non-zero probability. In general, the notion of support of a measure is not as straightforward, and the naïve approach is not viable since typically all singletons have measure 0. Nevertheless, supports can be defined in the setting of Borel measurable spaces, for instance, which in any case are the kind of spaces in which we are practically interested, in Sections 5 and 6.

Approaches to contextuality that characterise obstructions to global sections using cohomology have had some success [11, 5, 27, 28, 66, 68, 60, 29, 61] and typically apply to logical forms of contextuality. An interesting prospect is to explore how the present framework may be employed to these ends, and to see whether the continuous-variable setting can open the door to new techniques that can be applied, or whether qualitatively new forms of contextual behaviour may be uncovered. A related direction to be developed is to understand how our treatment of contextuality can be further extended to continuous measurement spaces as proposed in [30].

Another direction to be explored is how our continuous-variable framework for contextuality can be extended to apply to more general notions of contextuality that relate not only to measurement contexts but also more broadly to contexts of preparations and transformations as well [72, 58], noting that these also admit quantifiable relationships to quantum advantage [58, 41].

Indeed, a major motivation to study contextuality is for its connections to quantum-over-classical advantages in informatic tasks. An important line of questioning is to ask what further connections can be found in the continuous-variable setting, and whether continuous-variable contextuality might offer advantages that outstrip those achievable with discrete-variable contextual resources. Note that it is known that infinite-dimensional quantum systems can offer certain additional advantages beyond finite-dimensional ones [71], though the empirical model that arises in that example is still a discrete-variable one in our sense.

The present work sets the theoretical basis for computational exploration of continuous-variable contextuality in quantum-mechanical empirical models. This, we hope, can provide new insights and inform all other avenues to be developed in future work. It can also be useful in verifying the non-classicality of empirical models. Numerical implementation of the programs of Section 6 is of particular interest. The hierarchy of semi-definite programs can be used numerically to witness contextuality in continuous-variable experiments. Even if the time-complexity of the semi-definite program may increase drastically with its degree, a low-degree program can already provide a first witness of contextual behaviour.

Since our framework for continuous-variable contextuality is independent of quantum theory itself, it can equally be applied to ‘empirical models’ that arise in other, non-physical settings. The discrete-variable framework of [8] has led to a number of surprising connections and cross-fertilisations with other fields [3], including natural language [12], relational databases [1, 16], logic [10, 5, 49], constraint satisfaction [9, 7] and social systems [34]. It may be hoped that similar connections and applications can be found for the present framework to fields in which continuous-variable data is of central importance. For instance probability kernels of the kind we have used are also widely employed in machine learning (e.g. [43]), inviting intriguing questions about how our framework might be used or what advantages contextuality may confer in that setting.

## Acknowledgements

The authors thank Robert Booth, Ulysse Chabaud, Marcelo Terra Cunha, and Antoine Oustry for helpful comments and discussions.

Financial support from the following is gratefully acknowledged: the Engineering and Physical Sciences Research Council, EP/N018745/1, ‘Contextuality as a Resource in Quantum Computation’ (RSB); the European Union’s Horizon

2020 Research and Innovation Programme under the Marie Skłodowska-Curie Grant Agreement No. 750523, ‘Resource Sensitive Quantum Computing’ (SM).

## Appendices

### A Linear programs in standard form

This appendix may be of particular interest to readers familiar with global optimisation. We express the problem (P) in the standard form of infinite linear programming [17, IV–(7.1)].

We first write (P) with an equality constraint. Adding slack variables in the form of complementary measures  $\nu_C$  representing the contextual parts of the empirical model, we can express (P) as:

$$(P_C) : \begin{cases} \text{Find} & \mu \in \mathbb{M}_{\pm}(\mathcal{O}_X), (\nu_C \in \mathbb{M}_{\pm}(\mathcal{O}_C))_{C \in \mathcal{M}} \\ \text{maximising} & \mu(\mathcal{O}_X) \\ \text{subject to} & \forall C \in \mathcal{M}. \mu|_C + \nu_C = e_C \\ \text{and} & \mu \geq 0, \forall C \in \mathcal{M}. \nu_C \geq 0 \end{cases}$$

Problems (P) (or (P<sub>C</sub>)) and (D) are indeed a infinite linear programs as both the objective and the constraints are linear with respect to the unknown measure  $\mu \in \mathbb{M}_{\pm}(\mathcal{O}_X)$ . To write (P<sub>C</sub>) in the standard form [17], we introduce the following spaces:

- $E_1 := \mathbb{M}_{\pm}(\mathcal{O}_X) \times \prod_{C \in \mathcal{M}} \mathbb{M}_{\pm}(\mathcal{O}_C)$ , and also  $\tilde{E}_1 := \mathbb{M}_{\pm}(\mathcal{O}_X)$  when considering the problem (P).
- $F_1 := C_0(\mathcal{O}_X, \mathbb{R}) \times \prod_{C \in \mathcal{M}} C_0(\mathcal{O}_C, \mathbb{R})$ , the dual space of  $E_1$ .
- $E_2 := \prod_{C \in \mathcal{M}} \mathbb{M}_{\pm}(\mathcal{O}_C)$ .
- $F_2 := \prod_{C \in \mathcal{M}} C_0(\mathcal{O}_C, \mathbb{R})$ , the dual space of  $E_2$ .

The dualities  $\langle -, - \rangle_1 : E_1 \times F_1 \rightarrow \mathbb{R}$  and  $\langle -, - \rangle_2 : E_2 \times F_2 \rightarrow \mathbb{R}$  are defined as follows:

$$\begin{aligned} \forall \omega = (\mu, (\nu_C)_{C \in \mathcal{M}}) \in E_1, \forall F = (f, (f_C)_{C \in \mathcal{M}}) \in F_1, \quad \langle \omega, F \rangle_1 &:= \int_{\mathcal{O}_X} f \, d\mu + \sum_{C \in \mathcal{M}} \int_{\mathcal{O}_C} f_C \, d\nu_C \\ \forall \omega = ((\nu_C)_{C \in \mathcal{M}}) \in E_2, \forall F = ((f_C)_{C \in \mathcal{M}}) \in F_2, \quad \langle \omega, F \rangle_2 &:= \sum_{C \in \mathcal{M}} \int_{\mathcal{O}_C} f_C \, d\nu_C. \end{aligned}$$

Let  $A : E_1 \rightarrow E_2$  be the following linear transformation:

$$\forall \omega = (\mu, (\nu_C)_{C \in \mathcal{M}}) \in E_1, \quad A(\omega) := (\mu|_C + \nu_C)_{C \in \mathcal{M}}.$$

We also define  $A^* : F_2 \rightarrow F_1$  as:

$$\forall F = ((f_C)_{C \in \mathcal{M}}) \in F_2, \quad A^*(F) := \left( \sum_{C \in \mathcal{M}} f_C, (f_C)_{C \in \mathcal{M}} \right).$$

We can verify that  $A^*$  is the dual transformation of  $A$ : for all  $\omega = (\mu, (\nu_C)_{C \in \mathcal{M}}) \in E_1$  and for all  $F = ((f_C)_{C \in \mathcal{M}}) \in F_2$ , we have

$$\begin{aligned} & \langle A(\omega), F \rangle_2 \\ &= \{ \text{rewriting with } (\mu, (\nu_C)) \text{ and } (f_C) \text{ and definition of operator } A \} \\ & \langle (\mu|_C + \nu_C)_{C \in \mathcal{M}}, (f_C)_{C \in \mathcal{M}} \rangle_2 \\ &= \{ \text{definition of } \langle -, - \rangle_2 \} \\ & \sum_{C \in \mathcal{M}} \int_{\mathcal{O}_C} f_C \, d(\mu|_C + \nu_C) \\ &= \{ \text{linearity of the integral and definition of the marginalisation of } \mu \} \\ & \int_{\mathcal{O}_X} \sum_{C \in \mathcal{M}} f_C \, d\mu + \sum_{C \in \mathcal{M}} \int_{\mathcal{O}_C} f_C \, d\nu_C \end{aligned}$$

$$\begin{aligned}
&= \{ \text{definition of } \langle -, - \rangle_1 \} \\
&\quad \langle (\mu, (v_C)_{C \in \mathcal{M}}), (\sum_{C \in \mathcal{M}} f_C, (f_C)_{C \in \mathcal{M}}) \rangle_1 \\
&= \{ \text{rewriting with } \omega \text{ and } F \} \\
&\quad \langle \omega, A^*(F) \rangle_1.
\end{aligned}$$

We can now rewrite problem (P<sub>C</sub>). The vector function in the objective is  $c = (\mathbf{1}, 0) \in E_1$  and we also choose to set  $b = ((e_C)_{C \in \mathcal{M}}) \in E_2$  for the constraints. The standard form in the sense of [17] can then be written as follows.

$$\left\{ \begin{array}{l} \text{Find} \quad \gamma = \sup_{\omega \in E_1} \langle \Omega, c \rangle_1 \\ \text{subject to} \quad A(\omega) = b \\ \text{and} \quad \omega \geq 0 \end{array} \right. \quad (14)$$

Indeed, because  $c$  is of the specific form  $(\mathbf{1}, 0)$ , it holds that

$$\sup_{\omega \in E_1} \langle \omega, c \rangle_1 = \sup_{\mu \in \tilde{E}_1} \langle (\mu, 0), (\mathbf{1}, 0) \rangle = \sup_{\mu \in \tilde{E}_1} \mu(\mathbf{O}_X),$$

and the constraints  $A(\omega) = b$  and  $\omega \geq 0$  are equivalent to  $\mathcal{M}, \mu|_C + v_C = e_C$  for all  $C \in \mathcal{M}$ , and  $\mu \geq 0$ . This is exactly our primal (P<sub>C</sub>). One can note that the primal program amounts to optimising on  $\tilde{E}_1$  with an inequality constraint while the problem (P<sub>C</sub>) amounts to optimising on  $E_1$  with an equality constraint. From [17], the standard form of the dual of problem (14) can be expressed as follows.

$$\left\{ \begin{array}{l} \text{Find} \quad \beta = \inf_{F \in F_2} \langle b, F \rangle_2 \\ \text{subject to} \quad A^*(F) \geq c \\ \text{and} \quad F \geq 0 \end{array} \right. \quad (15)$$

In our case the objective is

$$\inf_{F \in F_2} \langle b, F \rangle_2 = \inf_{(f_C)_{C \in \mathcal{M}} \in F_2} \sum_{C \in \mathcal{M}} \int_{O_C} f_C \, d e_C,$$

while the constraints  $A^*(F) \geq c$  and  $F \geq 0$  can be expressed as  $\sum_{C \in \mathcal{M}} f_C \geq \mathbf{1}$  on  $O_X$ , and  $f_C \geq 0$  on  $O_C$  for all  $C \in \mathcal{M}$ . This is exactly problem (D).

## B Proof of Proposition 15: zero duality gap

In this appendix we give a full proof of Proposition 15; i.e. that strong duality holds between problems (P) and (D).

*Proof.* To show strong duality, we rely on [17, Theorem 7.2]. We define:

$$E_{1+} = \{ (\mu, (v_C)_{C \in \mathcal{M}}) \in E_1 \mid \mu \geq 0 \text{ and } \forall C \in \mathcal{M}, v_C \geq 0 \} \subset E_1.$$

$E_{1+}$  is a positive convex cone in  $E_1$ . Since the linear program of (5) is consistent with finite value (because  $\mu = 0$  is feasible for (A)), it suffices to show that the following cone

$$\mathcal{H} = \{ (A(\omega), \langle \omega, c \rangle_1) : \omega \in E_{1+} \} = \{ ((\mu|_C + v_C)_C, \mu(\mathbf{O}_X)) : (\mu, (v_C)_C) \in E_{1+} \}$$

is weakly closed in  $E_2 \oplus \mathbb{R}$  (i.e. closed in the weak topology of  $E_{1+}$ ).

We first notice that  $A$  is a bounded linear operator. Boundedness comes from the fact that, for all  $\omega = (\mu, (v_C)_C) \in E_{1+}$ ,

$$\begin{aligned}
\|A(\omega)\|_{E_2} &= \|(\mu|_C + v_C)_C\|_{E_2} \\
&= \sum_{C \in \mathcal{M}} \|\mu|_C + v_C\| \\
&\leq \sum_{C \in \mathcal{M}} (\|\mu|_C\| + \|v_C\|) \\
&\leq |\mathcal{M}| \|\mu\| + \sum_{C \in \mathcal{M}} \|v_C\| \\
&\leq |\mathcal{M}| \|(\mu, (v_C)_C)\|_{E_{1+}}
\end{aligned}$$

$$= \|\cdot\|_{\mathcal{M}} \|\omega\|_{E_{1+}},$$

where we take the following norm on finite Borel measures over a measurable space  $\mathbf{X} = \langle X, \mathcal{F} \rangle$ :

$$\|\mu\|_{\mathbf{X}} = \sup_{\sigma \in \mathcal{F}} \mu(\sigma) = \mu(Y) \text{ for } Y = \cup_{\sigma \in \mathcal{F}} \sigma.$$

Secondly, we consider a sequence  $(\omega^k)_{k \in \mathbb{N}} = (\mu^k, (v_C^k)_{C \in \mathcal{M}})_{k \in \mathbb{N}}$  in  $E_{1+}$  and we want to show that the accumulation point  $((\Theta_C)_C, \lambda) = \lim_{k \rightarrow \infty} (A(\omega^k), \langle \omega^k, c \rangle_1)$  belongs to  $\mathcal{K}$ , where  $\Theta \in E_2$  and  $\lambda \in \mathbb{R}$ . The sequence  $(\Omega^k)_k$  is bounded because  $A(\omega^k) = (\mu_C^k + v_C^k)_C \rightarrow (\Theta_C)_C$  as  $k \rightarrow \infty$ . Next, by weak-\* compactness of the unit ball (Alaoglu's theorem [53]), there exists a subsequence  $(\omega_{k_i})_{k_i}$  that converges weakly to an element  $\omega \in E_{1+}$ . By continuity of  $A$ , it yields that the accumulation point is such that  $((\Theta_C)_C, \lambda) = (A(\omega), \langle \omega, c \rangle_1) \in \mathcal{K}$ .  $\square$

## C Moment matrices are positive semi-definite

For well-defined moment sequences, i.e. sequences that have a representing finite Borel measure, moment matrices are indeed positive semi-definite which provides insight on the reason why problem (P) features positive semi-definiteness constraints.

Let  $\mathbf{y} = (y_\alpha)$  the moment sequence of a given Borel measure  $\mu$  on  $O_X$  (and similarly by marginalisation on every  $O_C$  for  $C \in \mathcal{M}$ ). For a given integer  $k$ , we construct the moment matrix  $M_k(\mathbf{y})$ . Then for any vector  $V \in \mathbb{R}^{s(k)}$  (noting that  $V$  is canonically associated with a polynomial  $v \in \mathbb{R}[x]_k$  with its basis  $(x_\alpha)$ ):

$$\begin{aligned} V^T M_k(\mathbf{y}) V &= \sum_{\alpha, \beta \in \mathbb{N}_k^d} v_\alpha y_{\alpha+\beta} v_\beta \\ &= \sum_{\alpha, \beta \in \mathbb{N}_k^d} v_\alpha v_\beta \int_{O_X} x^{\alpha+\beta} d\mu \\ &= \int_{O_X} \left( \sum_{\alpha \in \mathbb{N}_k^d} v_\alpha x^\alpha \right) d\mu \\ &= \int_{O_X} v^2(x) d\mu \geq 0. \end{aligned} \tag{16}$$

Thus  $M_k(\mathbf{y}) \succeq 0$  and similarly we can prove that the matrices  $(M_k(P_j \mathbf{y}))_{j=1 \dots m}$  are positive semi-definite. Indeed, it holds that, for all  $V \in \mathbb{R}^{s(k)}$  and for all  $j = 1 \dots m$ ,

$$V^T M_k(P_j \mathbf{y}) V = \int_{O_X} v^2(x) P_j(x) d\mu \geq 0, \tag{17}$$

since  $O_X = \{x \in \mathbb{R}^n \mid P_j(x) \geq 0, j = 1, \dots, m\}$ .

## D Useful results on SOS polynomials

The following theorem is used to prove the convergence of the optimal values of  $(SD_k)$  to the noncontextual fraction.

**Theorem 17** (Putinar's Positivstellensatz [42]). *If  $f \in \mathbb{R}[x]$  is strictly positive on  $O_X$  then  $f \in Q((P_j))$  i.e. for some SOS polynomials  $(\sigma_j)_{j=1 \dots m} \subset \Sigma^2 \mathbb{R}[x]$ :*

$$f = \sigma_0 + \sum_{j=1 \dots m} \sigma_j g_j$$

The following theorem [51] is used in showing duality between problems  $(SD_k)$  and  $(SP_k)$ .

**Theorem 18.** *A polynomial  $p$  belongs to  $\Sigma^2 \mathbb{R}[x]_{2k}$  if and only if there exists a positive semidefinite matrix  $Q \in \mathcal{M}_{s(k)}(\mathbb{R})$  such that  $p(x) = \mathbf{z}(x) Q \mathbf{z}(x)^T$  for  $x \in \mathbb{R}^{2k}$ , where  $\mathbf{z}(x) \in \mathbb{R}^{s(k)}$  is the vector of monomials of degree at most  $s(k)$   $\mathbb{R}[x]_k$  ( $z(x)_\alpha = x^\alpha$ ).*

## E Duality between $(SP_k)$ and $(SD_k)$

This appendix deals with the proof of duality between the semidefinite programs  $(SP_k)$  and  $(SD_k)$ .

We first rewrite  $M_k(\mathbf{y})$  as  $\sum_{\alpha \in \mathbb{N}_k^d} y_\alpha A_\alpha$  and  $M_{k-r_j}(P_j \mathbf{y})$  as  $\sum_{\alpha \in \mathbb{N}_k^d} y_\alpha B_\alpha^j$  for  $1 \leq j \leq m$  and for appropriate real symmetric matrices  $A_\alpha$  and  $(B_\alpha^j)_j$ . For instance, in the basis  $(x^\alpha)$ :

$$(A_\alpha)_{xy} = \begin{pmatrix} 1 & \text{if } x+y = \alpha \\ 0 & \text{otherwise} \end{pmatrix}_{xy}.$$

From  $A_\alpha$ , we also extract  $A_\alpha^C$  for  $C \in \mathcal{M}$  in order to rewrite  $M_k(\mathbf{y}|_C)$  as  $\sum_{\alpha \in \mathbb{N}_k^d} y_\alpha A_\alpha^C$ . This amounts to identifying which matrices  $(A_\alpha)$  contribute for a given context  $C \in \mathcal{M}$ . Then the dual (SD<sub>k</sub>) can be rewritten as:

$$(\text{SP}_k) \left\{ \begin{array}{l} \sup_{\mathbf{y} \in \mathbb{R}^{s(k)}} y_0 (= \mu(S)) \\ \text{s.t. } M_k(\mathbf{y}^{e,C}) - \sum_{\alpha \in \mathbb{N}_k^d} y_\alpha A_\alpha^C \succeq 0, \quad \forall C \in \mathcal{M} \\ \sum_{\alpha \in \mathbb{N}_k^d} y_\alpha A_\alpha \succeq 0 \\ \sum_{\alpha \in \mathbb{N}_k^d} y_\alpha B_\alpha^j \succeq 0, \quad \forall j = 1 \dots m. \end{array} \right. \quad (18)$$

**Lagrangian method for deriving the dual** We introduce one conjugate variable for each constraint:  $X^C$  for the  $|\mathcal{M}|$  first constraints,  $Y$  for the middle one,  $Z_j$  for the  $m$  last ones. The associated Lagrangian reads as follows.

$$\begin{aligned} \mathcal{L}(\mathbf{y}, (X_C), Y, (Z_j)) = & \underbrace{y_0}_{\text{objective}} + \underbrace{\sum_{C \in \mathcal{M}} \left( \text{Tr}(M_k(\mathbf{y}^{e,C}) X^C) - \sum_{\alpha \in \mathbb{N}_k^d} y_\alpha \text{Tr}(A_\alpha^C X^C) \right)}_{\text{1st } |\mathcal{M}| \text{ constraints}} \\ & + \underbrace{\sum_{\alpha \in \mathbb{N}_k^d} y_\alpha \text{Tr}(A_\alpha Y)}_{\text{middle constraint}} + \underbrace{\sum_{j=1 \dots m} \sum_{\alpha \in \mathbb{N}_k^d} y_\alpha \text{Tr}(B_\alpha^j Z_j)}_{\text{last } m \text{ constraints}}. \end{aligned} \quad (19)$$

The primal indeed corresponds to the following equation:

$$\sup_{\mathbf{y} \in \mathbb{R}^{s(k)}} \inf_{\substack{(X_C), Y, (Z_j) \\ \text{SDP matrices}}} \mathcal{L}(\mathbf{y}, (X_C), Y, (Z_j)). \quad (20)$$

To obtain the dual, we need to permute the infimum and the supremum and we thus rewrite the Lagrangian as:

$$\begin{aligned} \mathcal{L}(\mathbf{y}, (X_C), Y, (Z_j)) = & \sum_{C \in \mathcal{M}} \text{Tr}(M_k(\mathbf{y}^{e,C}) X^C) \\ & + \sum_{\alpha \in \mathbb{N}_k^d} y_\alpha \left( \delta_{\alpha 0} - \sum_{C \in \mathcal{M}} \text{Tr}(A_\alpha^C X^C) + \text{Tr}(A_\alpha Y) + \sum_{j=1 \dots m} \text{Tr}(B_\alpha^j Z_j) \right). \end{aligned} \quad (21)$$

The dual then reads as follows.

$$\left\{ \begin{array}{l} \inf_{\substack{X_C, Y, Z_j \\ \text{SDP matrices}}} \text{Tr}(M_k(\mathbf{y}^{e,C}) X^C) \\ \text{s.t. } \sum_{C \in \mathcal{M}} \text{Tr}(A_\alpha^C X^C) - \text{Tr}(A_\alpha Y) - \sum_{j=1 \dots m} \text{Tr}(B_\alpha^j Z_j) = \delta_{\alpha 0} \\ X^C, Y, Z_j \succeq 0, \quad \forall C \in \mathcal{M}, \forall j = 1 \dots m \end{array} \right. \quad (22)$$

From Theorem 18, the dual can be reformulated as (SD<sub>k</sub>).

$$(\text{SD}_k) \left\{ \begin{array}{l} \inf_{\substack{(g_C) \subset \Sigma^2 \mathbb{R}[x]_k \\ (\sigma_j) \subset \Sigma^2 \mathbb{R}[x]_{k-r_j}}} \sum_{C \in \mathcal{M}} \int_{O_C} g_C \, d e_C \\ \text{s.t. } \sum_{C \in \mathcal{M}} g_C - 1 = \sigma_0 + \sum_{j=1 \dots m} \sigma_j P_j \end{array} \right.$$

## References

- [1] Samson Abramsky (2013): **Relational databases and Bell's theorem**. In Val Tannen, Limsoon Wong, Leonid Libkin, Wenfei Fan, Wang-Chiew Tan & Michael Fourman, editors: *In search of elegance in the theory and practice of computation: Essays dedicated to Peter Buneman, Lecture Notes in Computer Science* 8000, Springer, pp. 13–35, doi:[10.1007/978-3-642-41660-6\\_2](https://doi.org/10.1007/978-3-642-41660-6_2).
- [2] Samson Abramsky (2013): **Relational hidden variables and non-locality**. *Studia Logica* 101(2), pp. 411–452, doi:[10.1007/s11225-013-9477-4](https://doi.org/10.1007/s11225-013-9477-4).
- [3] Samson Abramsky (2015): **Contextual semantics: From quantum mechanics to logic, databases, constraints, and complexity**. In Ehtibar Dzhafarov, Scott Jordan, Ru Zhang & Victor Cervantes, editors: *Contextuality from Quantum Physics to Psychology, Advanced Series on Mathematical Psychology* 6, World Scientific, pp. 23–50, doi:[10.1142/9789814730617\\_0002](https://doi.org/10.1142/9789814730617_0002).
- [4] Samson Abramsky, Rui Soares Barbosa, Martti Karvonen & Shane Mansfield (2019): **A comonadic view of simulation and quantum resources**. To appear in Proceedings of the 34th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS 2019).
- [5] Samson Abramsky, Rui Soares Barbosa, Kohei Kishida, Raymond Lal & Shane Mansfield (2015): **Contextuality, cohomology and paradox**. In Stephan Kreutzer, editor: *24th EACSL Annual Conference on Computer Science Logic (CSL 2015), Leibniz International Proceedings in Informatics (LIPIcs)* 41, Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, pp. 211–228, doi:[10.4230/LIPIcs.CSL.2015.211](https://doi.org/10.4230/LIPIcs.CSL.2015.211).
- [6] Samson Abramsky, Rui Soares Barbosa & Shane Mansfield (2017): **Contextual fraction as a measure of contextuality**. *Physical Review Letters* 119(5), p. 050504, doi:[10.1103/PhysRevLett.119.050504](https://doi.org/10.1103/PhysRevLett.119.050504).
- [7] Samson Abramsky, Rui Soares Barbosa, Nadish de Silva & Octavio Zapata (2017): **The quantum monad on relational structures**. In Kim G Larsen, Hans L Bodlaender & Jean-François Raskin, editors: *42nd International Symposium on Mathematical Foundations of Computer Science (MFCS 2017), Leibniz International Proceedings in Informatics (LIPIcs)* 83, Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, pp. 35:1–35:19, doi:[10.4230/LIPIcs.MFCS.2017.35](https://doi.org/10.4230/LIPIcs.MFCS.2017.35).
- [8] Samson Abramsky & Adam Brandenburger (2011): **The sheaf-theoretic structure of non-locality and contextuality**. *New Journal of Physics* 13(11), p. 113036, doi:[10.1088/1367-2630/13/11/113036](https://doi.org/10.1088/1367-2630/13/11/113036).
- [9] Samson Abramsky, Georg Gottlob & Phokion G. Kolaitis (2013): **Robust constraint satisfaction and local hidden variables in quantum mechanics**. In Francesca Rossi, editor: *23rd International Joint Conference on Artificial Intelligence (IJCAI 2013)*, AAAI Press, pp. 440–446.
- [10] Samson Abramsky & Lucien Hardy (2012): **Logical Bell inequalities**. *Physical Review A* 85(6), p. 062114, doi:[10.1103/PhysRevA.85.062114](https://doi.org/10.1103/PhysRevA.85.062114).
- [11] Samson Abramsky, Shane Mansfield & Rui Soares Barbosa (2012): **The cohomology of non-locality and contextuality**. In Bart Jacobs, Peter Selinger & Bas Spitters, editors: *Proceedings of 8th International Workshop on Quantum Physics and Logic (QPL 2011), Electronic Proceedings in Theoretical Computer Science* 95, Open Publishing Association, pp. 1–14, doi:[10.4204/EPTCS.95.1](https://doi.org/10.4204/EPTCS.95.1).
- [12] Samson Abramsky & Mehrnoosh Sadrzadeh (2014): **Semantic unification: A sheaf-theoretic approach to natural language**. In Claudia Casadio, Bob Coecke, Michael Moortgat & Philip Scott, editors: *Categories and Types in Logic, Language, and Physics: Essays dedicated to Jim Lambek on the occasion of his 90th birthday*, Springer, pp. 1–13, doi:[10.1007/978-3-642-54789-8\\_1](https://doi.org/10.1007/978-3-642-54789-8_1).
- [13] Antonio Acín, Tobias Fritz, Anthony Leverrier & Ana Belén Sainz (2015): **A combinatorial approach to nonlocality and contextuality**. *Communications in Mathematical Physics* 334(2), pp. 533–628, doi:[10.1007/s00220-014-2260-1](https://doi.org/10.1007/s00220-014-2260-1).
- [14] Ali Asadian, Costantino Budroni, Frank ES Steinhoff, Peter Rabl & Otfried Gühne (2015): **Contextuality in phase space**. *Physical Review Letters* 114(25), p. 250403, doi:[10.1103/PhysRevLett.114.250403](https://doi.org/10.1103/PhysRevLett.114.250403).
- [15] Rui Soares Barbosa (2014): **On monogamy of non-locality and macroscopic averages: examples and preliminary results**. In Bob Coecke, Ichiro Hasuo & Prakash Panangaden, editors: *11th Workshop on Quantum Physics and Logic (QPL 2014), Electronic Proceedings in Theoretical Computer Science* 172, Open Publishing Association, pp. 36–55, doi:[10.4204/EPTCS.172.4](https://doi.org/10.4204/EPTCS.172.4).
- [16] Rui Soares Barbosa (2015): **Contextuality in quantum mechanics and beyond**. DPhil thesis, University of Oxford.
- [17] Alexander Barvinok (2002): **A course in convexity**. *Graduate Studies in Mathematics* 54, American Mathematical Society, doi:[10.1090/gsm/054](https://doi.org/10.1090/gsm/054).
- [18] John S Bell (1964): **On the Einstein Podolsky Rosen paradox**. *Physics Physique Fizika* 1(3), p. 195, doi:[10.1103/PhysicsPhysiqueFizika.1.195](https://doi.org/10.1103/PhysicsPhysiqueFizika.1.195).
- [19] John S Bell (1966): **On the problem of hidden variables in quantum mechanics**. *Reviews of Modern Physics* 38(3), pp. 447–452, doi:[10.1103/RevModPhys.38.447](https://doi.org/10.1103/RevModPhys.38.447).
- [20] Juan Bermejo-Vega, Nicolas Delfosse, Dan E Browne, Cihan Okay & Robert Raussendorf (2017): **Contextuality as a resource for models of quantum computation with qubits**. *Physical Review Letters* 119(12), p. 120505, doi:[10.1103/PhysRevLett.119.120505](https://doi.org/10.1103/PhysRevLett.119.120505).
- [21] Patrick Billingsley (1979): **Probability and Measure**. Wiley Series in Probability and Mathematical Statistics, Wiley.
- [22] Vladimir I Bogachev (2007): **Measure theory**. Springer, doi:[10.1007/978-3-540-34514-5](https://doi.org/10.1007/978-3-540-34514-5).
- [23] Adam Brandenburger & H Jerome Keisler (2013): **Use of a canonical hidden-variable space in quantum mechanics**. In Bob Coecke, Luke Ong & Prakash Panangaden, editors: *Computation, Logic, Games, and Quantum Foundations. The Many*

- Facets of Samson Abramsky: Essays dedicated to Samson Abramsky on the occasion of his 60th birthday*, Springer, pp. 1–6, doi:[10.1007/978-3-642-38164-5\\_1](https://doi.org/10.1007/978-3-642-38164-5_1).
- [24] Adam Brandenburger & Noson Yanofsky (2008): **A classification of hidden-variable properties**. *Journal of Physics A: Mathematical and Theoretical* 41(42), p. 425302, doi:[10.1088/1751-8113/41/42/425302](https://doi.org/10.1088/1751-8113/41/42/425302).
- [25] Samuel L Braunstein & Peter Van Loock (2005): **Quantum information with continuous variables**. *Reviews of Modern Physics* 77(2), p. 513, doi:[10.1103/RevModPhys.77.513](https://doi.org/10.1103/RevModPhys.77.513).
- [26] Adán Cabello, Simone Severini & Andreas Winter (2014): **Graph-theoretic approach to quantum correlations**. *Physical Review Letters* 112(4), p. 040401, doi:[10.1103/PhysRevLett.112.040401](https://doi.org/10.1103/PhysRevLett.112.040401).
- [27] Giovanni Carù (2015): **Detecting contextuality: Sheaf cohomology and All vs Nothing arguments**. Master's thesis, University of Oxford. Available at <http://www.cs.ox.ac.uk/files/7608/Dissertation.pdf>.
- [28] Giovanni Carù (2017): **On the cohomology of contextuality**. In Ross Duncan & Chris Heunen, editors: *13th International Conference on Quantum Physics and Logic (QPL 2016)*, *Electronic Proceedings in Theoretical Computer Science* 236, Open Publishing Association, pp. 21–39, doi:[10.4204/EPTCS.236.2](https://doi.org/10.4204/EPTCS.236.2).
- [29] Giovanni Carù (2018): **Towards a complete cohomology invariant for non-locality and contextuality**. Preprint [arXiv:1807.04203](https://arxiv.org/abs/1807.04203) [quant-ph].
- [30] Marcelo Terra Cunha (2019): **On measures and measurements: a fibre bundle approach to contextuality**. Preprint [arXiv:1903.08819](https://arxiv.org/abs/1903.08819) [math.PR].
- [31] W. Michael Dickson (1998): **Quantum chance and non-locality: Probability and non-locality in the interpretations of quantum mechanics**. Cambridge University Press, doi:[10.1017/CBO9780511524738](https://doi.org/10.1017/CBO9780511524738).
- [32] Cristhiano Duarte & Barbara Amaral (2018): **Resource theory of contextuality for arbitrary prepare-and-measure experiments**. *Journal of Mathematical Physics* 59(6), p. 062202, doi:[10.1063/1.5018582](https://doi.org/10.1063/1.5018582).
- [33] Ehtibar N Dzhafarov, Janne V Kujala & Victor H Cervantes (2015): **Contextuality-by-Default: A brief overview of ideas, concepts, and terminology**. In Harald Atmanspacher, Thomas Filk & Emmanuel Pothos, editors: *9th International Conference on Quantum Interaction (QI 2015)*, *Lecture Notes in Computer Science* 9535, Springer, pp. 12–23, doi:[10.1007/978-3-319-28675-4\\_2](https://doi.org/10.1007/978-3-319-28675-4_2).
- [34] Ehtibar N Dzhafarov, Ru Zhang & Janne Kujala (2016): **Is there contextuality in behavioural and social systems?** Theme issue on Quantum probability and the mathematical modelling of decision making, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374(2058), p. 20150099, doi:[10.1098/rsta.2015.0099](https://doi.org/10.1098/rsta.2015.0099).
- [35] Albert Einstein, Boris Podolsky & Nathan Rosen (1935): **Can quantum-mechanical description of physical reality be considered complete?** *Physical Review* 47(10), p. 777, doi:[10.1103/PhysRev.47.777](https://doi.org/10.1103/PhysRev.47.777).
- [36] Arthur Fine (1982): **Hidden variables, joint probability, and the Bell inequalities**. *Physical Review Letters* 48(5), p. 291, doi:[10.1103/PhysRevLett.48.291](https://doi.org/10.1103/PhysRevLett.48.291).
- [37] Gerald B Folland (1984): **Real Analysis: Modern techniques and their applications**. Pure and Applied Mathematics, John Wiley & Sons. 2nd Edition (1999).
- [38] Tobias Fritz (2009): **Possibilistic Physics**. Available at <https://fqxi.org/community/forum/topic/569>. FQXI Essay Contest 2009.
- [39] Michèle Giry (1982): **A categorical approach to probability theory**. In B Banaschewski, editor: *Categorical aspects of topology and analysis*, *Lecture Notes in Mathematics* 915, Springer, pp. 68–85, doi:[10.1007/BFb0092872](https://doi.org/10.1007/BFb0092872).
- [40] Qiong-Yi He, Eric G Cavalcanti, Margaret D Reid & Peter D Drummond (2010): **Bell inequalities for continuous-variable measurements**. *Physical Review A* 81(6), p. 062106, doi:[10.1103/PhysRevA.81.062106](https://doi.org/10.1103/PhysRevA.81.062106).
- [41] Luciana Henaut, Lorenzo Catani, Dan E Browne, Shane Mansfield & Anna Pappa (2018): **Tsirelson's bound and Landauer's principle in a single-system game**. *Physical Review A* 98(6), p. 060302, doi:[10.1103/PhysRevA.98.060302](https://doi.org/10.1103/PhysRevA.98.060302).
- [42] Didier Henrion & Milan Korda (2014): **Convex computation of the region of attraction of polynomial control systems**. *IEEE Transactions on Automatic Control* 59(2), pp. 297–312, doi:[10.1109/TAC.2013.2283095](https://doi.org/10.1109/TAC.2013.2283095).
- [43] Thomas Hofmann, Bernhard Schölkopf & Alexander J Smola (2008): **Kernel methods in machine learning**. *The Annals of Statistics* 36(3), pp. 1171–1220, doi:[10.1214/009053607000000677](https://doi.org/10.1214/009053607000000677).
- [44] Mark Howard, Joel Wallman, Victor Veitch & Joseph Emerson (2014): **Contextuality supplies the ‘magic’ for quantum computation**. *Nature* 510(7505), pp. 351–355, doi:[10.1038/nature13460](https://doi.org/10.1038/nature13460).
- [45] Jon P. Jarrett (1984): **On the physical significance of the locality conditions in the Bell arguments**. Special issue on the Foundations of Quantum Mechanics, *Noûs* 18(4), pp. 569–589, doi:[10.2307/2214878](https://doi.org/10.2307/2214878).
- [46] Shizuo Kakutani (1941): **Concrete representation of abstract ( $M$ )-spaces (A characterization of the space of continuous functions)**. *Annals of Mathematics* 42(4), pp. 994–1024, doi:[10.2307/1968778](https://doi.org/10.2307/1968778).
- [47] Martti Karvonen (2019): **Categories of Empirical Models**. In Peter Selinger & Giulio Chiribella, editors: *15th International Conference on Quantum Physics and Logic (QPL 2018)*, *Electronic Proceedings in Theoretical Computer Science* 287, Open Publishing Association, pp. 239–252, doi:[10.4204/EPTCS.287.14](https://doi.org/10.4204/EPTCS.287.14).
- [48] Andreas Ketterer, Adrien Laversanne-Finot & Leandro Aolita (2018): **Continuous-variable supraquantum nonlocality**. *Physical Review A* 97(1), p. 012133, doi:[10.1103/PhysRevA.97.012133](https://doi.org/10.1103/PhysRevA.97.012133).

- [49] Kohei Kishida (2016): **Logic of local inference for contextuality in quantum physics and beyond**. In Ioannis Chatzigiannakis, Michael Mitzenmacher, Yuval Rabani & Davide Sangiorgi, editors: *43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016)*, *Leibniz International Proceedings in Informatics (LIPIcs)* 55, Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, pp. 113:1–113:14, doi:[10.4230/LIPIcs.ICALP.2016.113](https://doi.org/10.4230/LIPIcs.ICALP.2016.113).
- [50] Simon Kochen & Ernst P Specker (1967): **The problem of hidden variables in quantum mechanics**. *Journal of Mathematics and Mechanics* 17(1), pp. 59–87.
- [51] Jean-Bernard Lasserre (2009): **Moments, positive polynomials and their applications**. *Series on Optimization and Its Applications* 1, Imperial College Press, doi:[10.1142/p665](https://doi.org/10.1142/p665).
- [52] Adrien Laversanne-Finot, Andreas Ketterer, Mariana R Barros, Stephen P Walborn, Thomas Coudreau, Arne Keller & Perola Milman (2017): **General conditions for maximal violation of non-contextuality in discrete and continuous variables**. *Journal of Physics A: Mathematical and Theoretical* 50(15), p. 155304, doi:[10.1088/1751-8121/aa6016](https://doi.org/10.1088/1751-8121/aa6016).
- [53] David G. Luenberger (1997): **Optimization by vector space methods**. John Wiley & Sons.
- [54] Saunders MacLane & Ieke Moerdijk (1992): **Sheaves in geometry and logic: A first introduction to topos theory**. Universitext, Springer, doi:[10.1007/978-1-4612-0927-0](https://doi.org/10.1007/978-1-4612-0927-0).
- [55] Shane Mansfield (2017): **Consequences and applications of the completeness of Hardy’s nonlocality**. *Physical Review A* 95(2), p. 022122, doi:[10.1103/PhysRevA.95.022122](https://doi.org/10.1103/PhysRevA.95.022122).
- [56] Shane Mansfield & Rui Soares Barbosa (2014): **Extendability in the sheaf-theoretic approach: Construction of Bell models from Kochen–Specker models**. Preprint [arXiv:1402.4827](https://arxiv.org/abs/1402.4827) [quant-ph].
- [57] Shane Mansfield & Tobias Fritz (2012): **Hardys non-locality paradox and possibilistic conditions for non-locality**. *Foundations of Physics* 42(5), pp. 709–719, doi:[10.1007/s10701-012-9640-1](https://doi.org/10.1007/s10701-012-9640-1).
- [58] Shane Mansfield & Elham Kashefi (2018): **Quantum advantage from sequential-transformation contextuality**. *Physical Review Letters* 121(23), p. 230401, doi:[10.1103/PhysRevLett.121.230401](https://doi.org/10.1103/PhysRevLett.121.230401).
- [59] Gerard McKeown, Matteo GA Paris & Mauro Paternostro (2011): **Testing quantum contextuality of continuous-variable states**. *Physical Review A* 83(6), p. 062105, doi:[10.1103/PhysRevA.83.062105](https://doi.org/10.1103/PhysRevA.83.062105).
- [60] Cihan Okay, Sam Roberts, Stephen D Bartlett & Robert Raussendorf (2017): **Topological proofs of contextuality in quantum mechanics**. *Quantum Information and Computation* 17(13–14), pp. 1135–1166, doi:[10.26421/QIC17.13-14](https://doi.org/10.26421/QIC17.13-14).
- [61] Cihan Okay, Emily Tyhurst & Robert Raussendorf (2018): **The cohomological and the resource-theoretic perspective on quantum contextuality: common ground through the contextual fraction**. Preprint [arXiv:1806.04657](https://arxiv.org/abs/1806.04657) [quant-ph].
- [62] Prakash Panangaden (2009): **Labelled Markov Processes**. Imperial College Press, doi:[10.1142/p595](https://doi.org/10.1142/p595).
- [63] Ángel R Plastino & Adán Cabello (2010): **State-independent quantum contextuality for continuous variables**. *Physical Review A* 82(2), p. 022114, doi:[10.1103/PhysRevA.82.022114](https://doi.org/10.1103/PhysRevA.82.022114).
- [64] Sandu Popescu & Daniel Rohrlich (1994): **Quantum nonlocality as an axiom**. *Foundations of Physics* 24(3), pp. 379–385, doi:[10.1007/BF02058098](https://doi.org/10.1007/BF02058098).
- [65] Robert Raussendorf (2013): **Contextuality in measurement-based quantum computation**. *Physical Review A* 88(2), p. 022322, doi:[10.1103/PhysRevA.88.022322](https://doi.org/10.1103/PhysRevA.88.022322).
- [66] Robert Raussendorf (2016): **Cohomological framework for contextual quantum computations**. Preprint <https://arxiv.org/abs/1602.04155> arXiv:1602.04155 [quant-ph].
- [67] Robert Raussendorf & Hans J Briegel (2001): **A one-way quantum computer**. *Physical Review Letters* 86(22), p. 5188, doi:[10.1103/PhysRevLett.86.5188](https://doi.org/10.1103/PhysRevLett.86.5188).
- [68] Frank Roumen (2017): **Cohomology of effect algebras**. In Ross Duncan & Chris Heunen, editors: *13th International Conference on Quantum Physics and Logic (QPL 2016)*, *Electronic Proceedings in Theoretical Computer Science* 236, Open Publishing Association, pp. 174–201, doi:[10.4204/EPTCS.236.12](https://doi.org/10.4204/EPTCS.236.12).
- [69] Abner Shimony (1986): **Events and processes in the quantum world**. In Roger Penrose & Chris J. Isham, editors: *Quantum concepts in space and time*, Oxford University Press, pp. 182–203, doi:[10.1017/CBO9781139172196.011](https://doi.org/10.1017/CBO9781139172196.011).
- [70] Andrew W Simmons (2017): **How (maximally) contextual is quantum mechanics?** Preprint [arXiv:1712.03766](https://arxiv.org/abs/1712.03766) [quant-ph].
- [71] William Slofstra (2016): **Tsirelson’s problem and an embedding theorem for groups arising from non-local games**. Preprint [arXiv:1606.03140](https://arxiv.org/abs/1606.03140) [quant-ph].
- [72] Robert W Spekkens (2005): **Contextuality for preparations, transformations, and unsharp measurements**. *Physical Review A* 71(5), p. 052108, doi:[10.1103/PhysRevA.71.052108](https://doi.org/10.1103/PhysRevA.71.052108).
- [73] Hong-Yi Su, Jing-Ling Chen, Chunfeng Wu, Sixia Yu & CH Oh (2012): **Quantum contextuality in a one-dimensional quantum harmonic oscillator**. *Physical Review A* 85(5), p. 052126, doi:[10.1103/PhysRevA.85.052126](https://doi.org/10.1103/PhysRevA.85.052126).
- [74] Christian Weedbrook, Stefano Pirandola, Raúl García-Patrón, Nicolas J Cerf, Timothy C Ralph, Jeffrey H Shapiro & Seth Lloyd (2012): **Gaussian quantum information**. *Reviews of Modern Physics* 84(2), p. 621, doi:[10.1103/RevModPhys.84.621](https://doi.org/10.1103/RevModPhys.84.621).
- [75] Linde Wester (2018): **Classical and quantum structures of computation**. DPhil thesis, University of Oxford.
- [76] Jun-ichi Yoshikawa, Shota Yokoyama, Toshiyuki Kaji, Chanond Sornphiphatpong, Yu Shiozawa, Kenzo Makino & Akira Furusawa (2016): **Invited Article: Generation of one-million-mode continuous-variable cluster state by unlimited time-domain multiplexing**. *APL Photonics* 1(6), p. 060801, doi:[10.1063/1.4962732](https://doi.org/10.1063/1.4962732).