Limits of CSP Problems and Efficient Parameter Testing

Marek Karpinski* Roland Markó†

Abstract

We present a unified framework on the limits of *constraint satisfaction problems* (CSPs) and efficient parameter testing which depends only on array exchangeability and the method of cut decomposition without recourse to the weakly regular partitions. In particular, we formulate and prove a representation theorem for compact colored r-uniform directed hypergraph (r-graph) limits, and apply this to rCSP limits. We investigate the sample complexity of testable r-graph parameters, we discuss the generalized ground state energies and demonstrate that they are efficiently testable.

1 Introduction

We study the limits and efficient parameter testing properties of *Maximum Constraint Satisfaction Problems* of arity r (MAX-rCSP or rCSP for short), c.f. e.g., [2]. These two topics, limiting behavior and parameter estimation, are treated in the paper to a degree separately, as they require a different set of ideas and could be analyzed on their own right. The establishment of the underlying connection between convergence and testability is one of the main applications of the limit theory of dense discrete structures, see [8], [9].

In the first part of the paper we develop a general framework for the above CSP problems which depends only on the principles of the array exchangeability without a recourse to the weakly regular partitions used hitherto in the general graph and hypergraph settings. Those fundamental techniques and results were worked out in a series of papers by Borgs, Chayes, Lovász, Sós, Vesztergombi and Szegedy [8],[9],[26], and [29] for graphs including connections to statistical physics and complexity theory, and were subsequently extended to hypergraphs by Elek and Szegedy [13] via the ultralimit method. The central concept of r-graph convergence is defined through convergence of sub-r-graph densities, or equivalently through

*Dept. of Computer Science and the Hausdroff Center for Mathematics, University of Bonn. Supported in part by DFG grants, the Hausdorff grant EXC59-1. Research partly supported by Microsoft Research New England. E-mail: marek@cs.uni-bonn.de

†Hausdorff Center for Mathematics, University of Bonn. Supported in part by a Hausdorff scholarship. E-mail: roland.marko@hcm.uni-bonn.de

weak convergence of probability measures on the induced sub-r-graph yielded by uniform node sampling. Our line of work particularly relies on ideas presented in [12] by Diaconis and Janson, where the authors shed some light on the correspondence between combinatorial aspects (that is, graph limits via weak regularity) and the probabilistic viewpoint of sampling: Graph limits provide an infinite random graph model that has the property of exchangeability. The precise definitions, references and results will be given in Section 2, here we only formulate our main contribution informally: We prove a representation theorem for compact colored r-uniform directed hypergraph limits. This says that every limit object can be transformed into a measurable function on the (2^r-2)-dimensional unit cube that takes values from the probability distributions on the color palette, see Theorem 2.11. This extends the result of Diaconis and Janson [12], and of Lovász and Szegedy [28]. As an application, the description of the limit space of rCSPs is presented subsequent to the aforementioned theorem.

The second part of the paper, Sections 3 to 5, is dedicated to the introduction of a notion of efficient parameter testability of r-graphs and rCSP problems. We use the limit framework from the first part of the paper to formulate several results on it, which are proved with the aid of the cut decomposition method. We set our focus especially to parameters called *ground state energies* and study variants of them. These stand in close relationship to MAX-rCSP problems, our results can be regarded as continuous generalization of the former. We rely on the notion of property testing and sample complexity, that was introduced by Goldreich, Goldwasser, and Ron [14] and was employed in the graph limit theory in [9]. A graph parameter is testable in the sense of the last mentioned paper, when its value is estimable through a uniform sampling process, where the sample size only depends on the desired error gap, see Definition 3.1 below for the precise formulation. The characterization of these functions on the graph space was done in [9], the original motivation of the current paper was to provide an analogous characterization for efficiently testable parameters, that are parameters, whose required sample size for the estimation is at most polynomial in the multiplicative inverse of the error.

The investigation of such parameters has been an active area of research for the finite setting in complexity theory. The method of exhaustive sampling in order to approximately solve NP-hard problems was proposed by Arora, Karger and Karpinski [4], their upper bound on the required sample size is still logarithmic in the size of the problem. Subsequently the testability of MAX-CUT was shown in [14], explicit upper bounds for the sample complexity in the general boolean MAX-rCSP were given by Alon, F. de la Vega, Kannan and Karpinski [2] using cut decomposition of r-arrays and sampling, that was inspired by the introduction of weak regularity by Frieze and Kannan [17]. In [2] and [17], an the design of polynomial time approximation schemes (PTAS) was an important subject, we did not pursue the generalization regarding this aspect in the current work. The achievements of these two aforementioned papers took also a key role in the first elementary treatment of graph limits and in the definition of the $\delta_\square$-metric that defines an equivalent topology on the limit space as the subgraph density convergence.

The best currently known upper bound on the sample complexity of MAX-rCSP is

$O(\varepsilon^{-4})$, and was shown by Mathieu and Schudy [30], see also Alon, F. de la Vega, Kannan and Karpinski [2]. Unfortunately, their approach does not seem to have a natural counterpart in the continuous setting, although one can use their result on the *sample* to achieve an improved upper bound on the sample complexity. We mention that for the original problem we do not aim to produce an assignment for MAX-rCSP, or a partition for the ground state energy whose evaluation is nearly optimal as opposed to the above works, although we believe this could be done without serious difficulties.

Our contribution in the second part of the paper is the following. By employing a refined version of the proof of the main result of [2] adapted to the continuous setting we are able to prove the analogous efficient testability result for a general finite state space for ground state energies, see Theorem 4.4 for a precise formulation. Among the applications of this development we analyze the testability of the microcanonical version of ground state energies providing the first explicit upper bounds on efficiency. For the finite version a similar question was investigated by F. de la Vega, Kannan and Karpinski [15] by imposing additional global constraints (meaning a finite number of them with unbounded arity). Furthermore, the continuous version of the quadratic assignment problem is treated the first time in a sample complexity context, this subject is related to the recent contributions to topic of approximate graph isomorphism and homomorphism, see [24] and [6].

1.1 Outline of the paper

The organization of this paper is as follows. In Section 2 we develop the limit theory for $\mathcal{K}$-decorated r-uniform directed hypergraphs with reference to previously known special (and in some way generic) cases, and use the representation of the limit to describe the limit space of rCSP problems. In Section 3 the basic notion of efficiency in context of parameter testing is given with additional examples. The subsequent Section 4 contains the proof of Theorem 4.4 regarding ground state energies of r-graphons, and in the following Section 5 generalizations and special cases are examined, in particular microcanonical energies and the quadratic assignment problem. We summarize possible directions of further research in Section 6.

2 Limit theory and underlying notation

We will consider the objects called *rCSP formulas* that are used to define instances of the optimization problems called MAX-rCSP. In the current framework a formula consists of a variable set and a set of boolean or integer valued functions. Each of these functions is defined on the subset of the variables, and the sets of possible states of the variables are identical. Additionally, it will be required that each of the functions, which we will call *constraints* in what follows, depend exactly on r of the variables.

When considering the optimization problem MAX-rCSP, then the rCSP formulas are related to the objective function given by the instance, whose domain is the whole variable set instead of the collection of functions on small subsets. This viewpoint is of course highly

dependent on the optimization problem in question. The precise definitions will be provided next.

Let K be a finite set, and f be an r variable boolean-valued function $f\colon K^r \to \{0,1\}$ (or equivalently $f \subseteq K^r$). We call f a *constraint-type on K in r variables*, $\mathcal{C} = \mathcal{C}(K,r)$ denotes the set of all such objects.

Definition 2.1 (rCSP formula)**.** *Let $V = \{x_1, x_2, \ldots, x_n\}$ be the set of variables, $e = (i_1, \ldots, i_r) \in V^r$ and f a* ***constraint-type on K in r variables****. We call an n variable function $\omega = (f;e)\colon K^n \to \{0,1\}$ with $\omega(l_1, \ldots, l_n) = f(l_{i_1}, \ldots, l_{i_r})$ a* ***constraint on V in r variables*** *determined by an r-vector of constrained variables and a constraint type. We call a set F of constraints on $V(F)$ in r variables an rCSP formula.*

In the above definition the set of states K of the variables in $V(F)$ is not specified for each formula, it will be considered as fixed similar to the dimension r. We say that F is symmetric, if it contains only constraints with constraint-types which are invariant under the permutations of the constrained variables. When we relax the notion of the constraint types to be real- or $\mathcal{K}$-valued functions on K^r with $\mathcal{K}$ being a compact space, then we speak of weighted rCSP formulas.

The motivation for the name CSP is immediately clear from the notation used in Definition 2.1 if we consider constraints to be satisfied at some point in K^n, whenever they evaluate to 1 there. Most problems defined on these objects ask for parameters that are, in the language of real analysis, global or conditioned extreme values of the objective function given by an optimization problem and a formula (which can be regarded as an instance of the former). Such problems are for example MAX-rCSP, rCSP, MAX-CUT, MAX-BISECTION. They can also be viewed as directed r-graphs, whose edges are decorated with constraint types (perhaps with multiple types), and we will exploit this representation in our analysis.

Typically, we will not store and recourse to an rCSP formula F as it is given by its definition above, but we will only consider the r-array tuple $(F^z)_{z\in K^r}$, where $F^z(e) = \sum_{(f;e)\in F} f(z_1, \ldots, z_r)$ for each $e \in V^r$. The data set $(F^z)_{z\in K^r}$ is called the *evaluation representation of F*, or short eval(F). We impose a boundedness criteria on CSPs that will apply throughout the paper, that means we fix $d \geq 1$ for good, and require that $\|\mathrm{eval}(F)\|_\infty \leq d$ for all CSPs F in consideration.

Our main concern is parameter estimation via sampling: we pick a set of variables of fixed size at random from the constrained set $V(F)$ of an rCSP formula F defined on a large number of variables, and ask for all the constraints in which the sampled variables are involved and no other, that is, the induced subformula on the sample. Then we try to produce some quantitative statement about the parameter of the original formula by relying only on the estimation of the corresponding value of the parameter on a subformula, see Definition 3.1 below.

Let $\mathbb{G}(k,F)$ denote the random induced subformula of F on the set $S \subset V(F)$ that is chosen uniformly among the subsets of $V(F)$ of cardinality k.

2.1 Limits of $\mathcal{K}$-decorated r-uniform directed hypergraphs

Let $\mathcal{K}$ be a compact Polish space and $r \geq 1$ an integer. Recall that a space $\mathcal{K}$ is called Polish if it is a separable completely metrizable topological space. In what follows we will consider the limit space of $\mathcal{K}$-decorated r-uniform directed hypergraphs, or with different words r-arrays with non-diagonal entries from $\mathcal{K}$ with all the diagonal entries occupied by a special element which also can be in $\mathcal{K}$, but in general this does not have to be the case.

Our motivation for what follows originates from the aim of presenting a structural description of rCSP limits. The content of the current section will start with the more general setting given above, CSPs will be considered as a special case in this topic derived at the end. Some of the basic cases are already settled regarding the representation of the limits. Without claiming to provide a complete list of previously established results we refer to Lovász and Szegedy [26], [28], [25] for the $r = 2$, general $\mathcal{K}$, undirected case, to Elek and Szegedy [13] for the general r, $\mathcal{K} = \{0, 1\}$, undirected case; and Diaconis and Janson [12] for $r = 2$, $\mathcal{K} = \{0, 1\}$, directed and undirected case. These three approaches are fundamentally different in their proof methodology (they rely on weak regularity, ultralimits, and exchangeability principles respectively) and were further generalized or applied by Zhao [32] to general r; respectively by Aroskar [5] to the directed case; respectively by Austin [7] general r and by Janson [20] to the directed case where the graph induces a partial order on the vertex set.

It is worth to mention that when utilizing tools from exchangeability theory (as we will do in the following) with regard to the directedness, the results are mostly proven first for the directed case, and afterwards contain the analysis regarding in what form the symmetry emerges in the limit.

Let $\mathcal{C}$ denote space of continuous functionals on $\mathcal{K}$, and let $\mathcal{F} \subset \mathcal{C}$ be a countable generating set with $\|f\|_\infty \leq 1$ for each $f \in \mathcal{F}$, that is, the linear subspace generated by $\mathcal{F}$ is dense in $\mathcal{C}$ in the L^∞-norm.

Denote by $\Pi(S) = \Pi_r(S)$ the set of all unlabeled S-decorated directed r-uniform hypergraphs for some arbitrary set S, where we will suppress r in the notation, when it is clear which r is meant (alternatively, $\Pi(S)$ denotes the isomorphism classes of the node labeled respective objects).

Definition 2.2. *If $F \in \Pi(\mathcal{C})$ uniform directed graph with $V(F) = [k]$ and $G \in \Pi(\mathcal{K})$, then the homomorphism density of F into G is defined as*

$$t(F, G) = \frac{1}{|V(G)|^k} \sum_{\phi\colon [k]\to V(G)} \prod_{i_1,\dots,i_r=1}^{k} F_{i_1,\dots,i_r}(G_{\phi(i_1),\dots,\phi(i_r)}). \tag{2.1}$$

The injective homomorphism density $t_{\text{inj}}(F, G)$ is defined similarly, with the difference that the average of the products is taken over all injective ϕ maps.

Let the map τ be defined as $\tau(G) = (t(F, G))_{F\in\Pi(\mathcal{F})} \in [0, 1]^{\Pi(\mathcal{F})}$ for each $G \in \Pi(\mathcal{K})$. We set $\Pi(\mathcal{K})^* = \tau(\Pi(\mathcal{K})) \subset [0, 1]^{\Pi(\mathcal{F})}$, and $\overline{\Pi(\mathcal{K})^*}$ to the closure of $\Pi(\mathcal{K})^*$. Also, let $\Pi(\mathcal{K})^+ = \{(\tau(G), 1/|V(G)|) : G \in \Pi(\mathcal{K})\} \subset [0, 1]^{\Pi(\mathcal{F})} \times [0, 1]$, and let $\overline{\Pi(\mathcal{K})^+}$ be the closure of $\Pi(\mathcal{K})^+$.

The function $\tau^+(G) = (\tau(G), 1/|V(G)|)$ will be useful for our purposes, because, opposed to τ, it is injective, which can be verified easily. For any $F \in \Pi(\mathcal{F})$ the function $t(F,.)$ on $\Pi(\mathcal{K})$ can be uniquely continuously extended to a function $t(F,.)$ on $\overline{\Pi(\mathcal{K})^+}$, this is due to the compactness of $[0,1]^{\Pi(\mathcal{F})} \times [0,1]$. For an element $\Gamma \in \overline{\Pi(\mathcal{K})^+} \setminus \Pi(\mathcal{K})^+$, let $t(F,\Gamma)$ for $F \in \Pi(\mathcal{F})$ denote the real number in $[0,1]$ that is the coordinate of Γ corresponding to F.

The sets $\Pi_{\text{inj}}(\mathcal{K}) = \tau_{\text{inj}}(\Pi(\mathcal{K}))$ and $\Pi_{\text{inj}}(\mathcal{K})^+$, and the functions $\tau_{\text{inj}}(G)$ and $\tau^+_{\text{inj}}(G)$ are defined analogously.

It was shown in [26] that

$$|t_{\text{inj}}(F,G) - t(F,G)| \leq \frac{2|V(F)|\|F\|_\infty}{|V(G)|} \tag{2.2}$$

for any pair $F \in \Pi(\mathcal{C})$ and $G \in \Pi(\mathcal{K})$.

The correspondence analogous to that of Diaconis and Janson in [12] between elements of the limit space $\overline{\Pi(\mathcal{K})^+}$, that is compact, and extreme points of the space of random exchangeable infinite r-arrays with entries in $\mathcal{K}$ will now be established. These are arrays, whose distribution is invariant under finite permutations of the underlying index set.

Definition 2.3 (Exchangeable r-array)**.** *Let $(H_{i_1,\dots,i_r})_{1\leq i_1,\dots,i_r<\infty}$ be an infinite r-array of random entries taking values in a Polish space S. We call the random array separately exchangeable if*

$$(H_{i_1,\dots,i_r})_{1\leq i_1,\dots,i_r<\infty} \overset{\text{law}}{=} (H_{\rho_1(i_1),\dots,\rho_r(i_r)})_{1\leq i_1,\dots,i_r<\infty}$$

for any permutations $\rho_1,\dots,\rho_r \in S_{\mathbb{N}}$, and jointly exchangeable (or just exchangeable), if the former holds only for all $\rho_1 = \cdots = \rho_r \in S_{\mathbb{N}}$.

Let M be an arbitrary set and $k \geq 1$, and let $s(M,k)$ denote the set of non-empty subsets of M of cardinality at most k, furthermore let $s_0(M,k) = s(M,k) \cup \{\emptyset\}$. It is clear that if we consider a measurable function $f\colon [0,1]^{s_0([r],r)} \to S$, and independent random variables uniformly distributed on $[0,1]$ that are associated with each of the subsets of $\mathbb{N}$ of cardinality at most r, then by plugging in these random variables into f for every $e \in \binom{\mathbb{N}}{r}$ in the right way suggested by a fixed natural projection $p_e\colon e \to [r]$, the result will be an exchangeable random array. The shorthand $\text{Samp}(f)$ denotes this law of the infinite directed r-hypergraph model induced by f.

The next theorem, states that all exchangeable arrays with values in S arise from some f in the former way.

Theorem 2.4. *Let S be a Polish space. Every S-valued exchangeable r-array $(H_e)_{e\in\binom{\mathbb{N}}{r}}$ has law μ equal to $\text{Samp}(f)$ for some measurable $f\colon [0,1]^{s_0([r],r)} \to S$, that is, there exists a function f, so that if $(U_s)_{s\in s_0(\mathbb{N},r)}$ are independent uniform $[0,1]$ random variables, then*

$$H_e = f(U_\emptyset, U_{\{i_1\}}, U_{\{i_2\}}, \dots, U_{e\setminus\{i_r\}}, U_e) \tag{2.3}$$

for every $e = (i_1,\dots,i_r)$, where H_e are the entries of the infinite r-array and $(H_e)_{e\in\mathbb{N}^r}$ has law μ.

Theorem 2.4 was first proved by de Finetti [11] (in the case $S = \{0, 1\}$) and Hewitt and Savage [18] (in the case of general S) for $r = 1$, independently by Aldous [1] and Hoover [19] for $r = 2$, and by Kallenberg [21] for arbitrary $r \geq 3$. For equivalent formulations, proofs and further connections to related areas see the recent survey of Austin [7].

In general, there are no symmetry assumptions on f, in the directed case H_e might differ from $H_{e'}$, even if e and e' share a common base set. In this case these two entries do not have the property of conditional independence over a σ-algebra given by some lower dimensional structures, that means for instance for an exchangeable r-array with law $\mathrm{Samp}(f)$ given by a function f as above the independence over $\{U_\alpha : \alpha \subsetneq e\}$.

With the aid of Theorem 2.4 we will provide a form of representation of the limit space $\overline{\Pi(\mathcal{K})^+}$ through the points of the space of random infinite exchangeable r-arrays. The correspondence will be established through a sequence of theorems analogous to the ones stated and proved in Section 2 to 5 of [12], using also the compactification argument regarding the limit space from [28], see also Chapter 17.1 in [25] for the complete argumentation. The proofs in our case are mostly ported in a straightforward way, if not noted otherwise we refer the reader for the details in [12]. The definition of convergence will be given right after the next theorem. The expression $\mathbb{G}(k, G)$ denotes for some $G \in \Pi(\mathcal{K})$ the random induced sub-r-graph of G with the vertex set chosen uniformly among all subsets of $V(G)$ that have cardinality k.

Theorem 2.5. *Let $(G_n)_{n\geq 1}$ be a random sequence in $\Pi(\mathcal{K})$ with $|V(G_n)|$ tending to infinity in probability. Then the following are equivalent.*

(1) $\tau^+(G_n)$ converges in distribution in $\Pi(\mathcal{K})^+$.

(2) For every $F \in \Pi(\mathcal{F})$, the sequence $t(F, G_n)$ converges in distribution.

(3) For every $F \in \Pi(\mathcal{C})$, the sequence $t(F, G_n)$ converges in distribution.

(4) For every $k \geq 1$, the sequence $\mathbb{G}(k, G_n)$ of random elements of $\Pi(\mathcal{K})$ converges in distribution.

If any of the above apply, then the respective limits in (2) and (3) are $t(F, \Gamma)$ with Γ being a random element of $\overline{\Pi(\mathcal{K})^+}$ given by (1), and also $\Gamma \in \overline{\Pi(\mathcal{K})^+} \setminus \Pi(\mathcal{K})^+$, almost surely.

If $t(F, G_n)$ in (2) and (3) is replaced by $t_{\mathrm{inj}}(F, G_n)$, then the equivalence of the four statements still persists and the limits in (2) and (3) are $t(F, \Gamma)$.

If every G_n is concentrated on one element of $\Pi(\mathcal{K})$ (non-random case), then the equivalence holds with the sequences in (1), (2), and (3) being numerical, while (4) remains unchanged.

Proof. The equivalence of (1) and (2) is immediate. The implication from (3) to (2) is also clear by definition.

For showing that (2) implies (3), we consider first an arbitrary $F \in \Pi(\langle\mathcal{F}\rangle)$, where $\langle\mathcal{F}\rangle$ is the linear space generated by $\mathcal{F}$. Then there exist $F^1, \ldots, F^l \in \Pi(\mathcal{F})$ on the same vertex set as F, say $[k]$, and $\lambda_1, \ldots, \lambda_l \in \mathbb{R}$ such that for any non-random $G \in \Pi(\mathcal{K})$ and $\phi\colon [k] \to V(G)$

it holds that $\prod_{i_1,\ldots,i_r=1}^{k} F_{i_1,\ldots,i_r}(G_{\phi(i_1),\ldots,\phi(i_r)}) = \sum_{j=1}^{l} \lambda_j \prod_{i_1,\ldots,i_r=1}^{k} F^j_{i_1,\ldots,i_r}(G_{\phi(i_1),\ldots,\phi(i_r)})$. So therefore we can express $t(F,G) = \sum_{j=1}^{l} \lambda_j t(F^j,G)$. We return to the case when G_n is random. The weak convergence of $t(F,G_n)$ is equivalent to the convergence of each of its moments, its t'th moment can be written by linearity as a linear combination of a finite number of mixed moments of the densities corresponding to $F^1,\ldots,F^l \in \Pi(\mathcal{F})$. For an arbitrary vector of non-negative integers $\alpha = (\alpha_1,\ldots,\alpha_l)$, let F^α be the element of $\Pi(\mathcal{F})$ that is the disjoint union α_1 copies of F^1, α_2 copies of F^2, and so on. It holds that $t(F^1,G_n)^{\alpha_1}\ldots t(F^l,G_n)^{\alpha_l} = t(F^\alpha,G_n)$, and in particular the two random variables on the two sides are equal in expectation. Condition (2) implies that $\mathbf{E}[t(F^\alpha,G_n)]$ converges for each α, therefore the mixed moments of the $t(F^i,G_n)$'s and the moments of $t(F,G_n)$ also do. This implies that also $t(F,G_n)$ converges in distribution for any $F \in \Pi(\langle F\rangle)$. Now let $F' \in \Pi(\mathcal{C})$ and $\varepsilon > 0$ be arbitrary, and $F \in \Pi(\langle F\rangle)$ on the same vertex set $[k]$ as F' be such that its entries are at most ε-far in L^∞ from the corresponding entries of F'. Then $|t(F',G) - t(F,G)| \le k^r \varepsilon \max\{(\|F'\|_\infty + \varepsilon)^{k^r-1}, \|F'\|_\infty + \varepsilon\}$ for any $G \in \Pi(\mathcal{K})$ (random or non-random), which implies (3), as $\varepsilon > 0$ was chosen arbitrarily.

We turn to show the equivalence of (3) and (4). Let $\Pi^k(\mathcal{K}) \subset \Pi(\mathcal{K})$ the set of elements with vertex set of cardinality k. The sequence $(\mathbb{G}(k,G_n))_{n\ge1}$ converges in distribution exactly when for each $f \in C(\Pi^k(\mathcal{K}))$ continuous function on $\Pi^k(\mathcal{K})$ the expectation $\mathbf{E}[f(\mathbb{G}(k,G_n))]$ converges as $n \to \infty$. For each $F \in \Pi(\mathcal{C})$ and $\alpha \ge 1$, the function $t^\alpha_{\mathrm{inj}}(F,G)$ is continuous on $\Pi^{|V(F)|}(\mathcal{K})$ and $t_{\mathrm{inj}}(F,G) = t_{\mathrm{inj}}(F,\mathbb{G}(|V(F)|,G))$, so (3) follows form (4).

For the other direction let us fix $k \ge 1$. We claim that the linear function space $M = \langle t(F,.)|F \in \Pi(\mathcal{C})\rangle \subset C(\Pi^k(\mathcal{K}))$ is an algebra containing the constant function, and that it separates any two elements of $\Pi^k(\mathcal{K})$. It follows that $\langle t(F,.)|F \in \Pi(\mathcal{C})\rangle$ is L^∞-dense in $C(\Pi^k(\mathcal{K}))$ by the Stone-Weierstrass theorem, which implies by our assumptions that $\mathbf{E}[f(\mathbb{G}(k,G_n))]$ converges for any $f \in C(\Pi^k(\mathcal{K}))$, since we know that $\mathbf{E}[t_{\mathrm{inj}}(F,\mathbb{G}(k,G_n))] = \mathbf{E}[t_{\mathrm{inj}}(F,G_n)]$ whenever $|V(F)| \le k$. We will see in a moment that $t_{\mathrm{inj}}(F,.) \in M$, convergence of $\mathbf{E}[t_{\mathrm{inj}}(F,G_n)]$ follows from (2.2) and the requirement that $|V(G_n)|$ tends to infinity in probability.

Now we turn to show that our claim is indeed true. For two $F_1,F_2 \in \Pi(\mathcal{C})$ we have $t(F_1,G)t(F_2,G) = t(F_1F_2,G)$ for any $G \in \Pi^k(\mathcal{K})$, where the product F_1F_2 denotes the disjoint union of the two $\mathcal{C}$-colored graphs. Also, $t(F,G) = 1$ for the graph F on one node with a loop colored with the constant 1 function. Furthermore we have that $\hom(F,G) = k^{|V(F)|}t(F,G) \in M$ for $|V(G)| = k$, so therefore

$$\mathrm{inj}(F,G) = \sum_{P \text{ partition of } V(F)} (-1)^{|V(F)|-|P|} \prod_{S\in P} (|S|-1)!\, \hom(F/P,G) \quad \in M,$$

where $\mathrm{inj}(F,G) = t_{\mathrm{inj}}(F,G)k(k-1)\ldots(k-|V(F)|+1)$ and $F/P \in \Pi^{|P|}(\mathcal{C})$ with edges colored by the product of the colors of F on the edges between the respective classes. This equality is the consequence of the Mobius inversion formula, and that $\mathrm{inj}(F,G) =$

$\sum_{P \text{ partition of } V(F)} \hom(F/P, G)$. For G and F defined on the node set $[k]$ recall that

$$\mathrm{inj}(F, G) = \sum_{\phi \in S_k} \prod_{i_1,\ldots,i_r=1}^{k} F_{i_1,\ldots,i_r}(G_{\phi(i_1),\ldots,\phi(i_r)}). \tag{2.4}$$

Now fix $G_1, G_2 \in \Pi^k(\mathcal{K})$ and let $F \in \Pi^k(\mathcal{C})$ such that $\{F_{i_1,\ldots,i_r}((G_j)_{\phi(i_1),\ldots,\phi(i_r)})\}$ are algebraically independent elements of $\mathbb{R}$ (such an F exists, we require a finite number of algebraically independent reals, and can construct each entry of F by polynomial interpolation). If G_1 and G_2 are not isomorphic, than for any possible node-relabeling for G_2 there is a at least one term in $\mathrm{inj}(F, G_1) - \mathrm{inj}(F, G_2)$ that does not get canceled out, so therefore $\mathrm{inj}(F, G_1) \neq \mathrm{inj}(F, G_2)$.

We examine the remaining statements of the theorem. Clearly, $\Gamma \notin \Pi(\mathcal{K})^+$, because $|V(G_n)| \to \infty$ in probability. The results for the case where the map in (1) and the densities in (2) and (3) are replaced by the injective version are yielded by (2.2), the proof of the non-random case carries through in a completely identical fashion.

□

We are now ready to formulate the definition of convergence in $\Pi(\mathcal{K})$.

Definition 2.6. *If $(G_n)_{n\geq 1}$ is a sequence in $\Pi(\mathcal{K})$ with $|V(G_n)| \to \infty$ and any of the conditions above of Theorem 2.5 holds, then we say that $(G_n)_{n\geq 1}$ converges.*

We would like to add that, in the light of Theorem 2.5, the convergence notion is independent from the choice of the family $\mathcal{F}$.

The next lemma gives information about the limit behavior of the sequences where the vertex set cardinality is constant.

Lemma 2.7. *Let $(G_n)_{n\geq 1}$ be a random sequence in $\Pi^k(\mathcal{K})$, and additionally be such that for every $F \in \Pi(\mathcal{F})$ the sequences $t_{\mathrm{inj}}(F, G_n)$ converge in distribution. Then there exists a random $H \in \Pi^k(\mathcal{K})$, such that for every $F \in \Pi(\mathcal{F})$ we have $t(F, G_n) \to t(F, H)$ and $t_{\mathrm{inj}}(F, G_n) \to t_{\mathrm{inj}}(F, H)$ in distribution.*

Proof. We only sketch the proof. The distributional convergence of (G_n) follows the same way as in the proof of Theorem 2.5, the part about condition (2) implies (3) and (3) implies (4). The existence of a random H satisfying the statement of the lemma follows from the Riesz representation theorem. □

Let $\mathcal{L}_\infty$ denote the set of all node labeled countably infinite $\mathcal{K}$-decorated r-uniform directed hypergraphs. Set the common vertex set of the elements of $\mathcal{L}_\infty$ to $\mathbb{N}$, and define the set of $[n]$-labeled $\mathcal{K}$-decorated r-uniform directed hypergraphs as $\mathcal{L}_n$. Every $G \in \mathcal{L}_n$ can be viewed as an element of $\mathcal{L}_\infty$ simply by adding isolated vertices to G carrying the labels $\mathbb{N}\setminus[n]$, therefore we think about $\mathcal{L}_n$ as a subset of $\mathcal{L}_\infty$ (and also of $\mathcal{L}_m$ for every $m \geq n$). Conversely, if G is a (random) element of $\mathcal{L}_\infty$, then by restricting G to the vertices labeled by $[n]$, we get $G|_{[n]} \in \mathcal{L}_n$. If G is a labeled or unlabeled $\mathcal{K}$-decorated r-uniform directed hypergraph

(random or not) with vertex set of cardinality n, then let $\hat{G}$ stand for the random element of $\mathcal{L}_n$ (and also $\mathcal{L}_\infty$) which we obtain by first throwing away the labels of G (if there where any), and then apply a random labeling chosen uniformly from all possible ones with the label set $[n]$.

A random element of $\mathcal{L}_\infty$ is *exchangeable* if its distribution is invariant under any permutation of the vertex set $\mathbb{N}$ that only moves finitely many vertices, for example such infinite hypergraphs are whose edge-colors are independently identically distributed. An element of $\mathcal{L}_\infty$ can also be regarded as an infinite r-array whose diagonal elements are decorated with a special element ι that is not contained in $\mathcal{K}$, therefore the corresponding r-arrays will be $\mathcal{K} \cup \{\iota\}$-decorated.

The next theorem relates the elements of $\Pi(\mathcal{K})^+$ to exchangeable random elements of $\mathcal{L}_\infty$.

Theorem 2.8. *Let $(G_n)_{n\geq 1}$ be a random sequence in $\Pi(\mathcal{K})$ with $|V(G_n)|$ tending to infinity in probability. Then the following are equivalent.*

(1) $\tau^+(G_n) \to \Gamma$ in distribution for a random $\Gamma \in \overline{\Pi(\mathcal{K})^+} \setminus \Pi(\mathcal{K})^+$.

(2) $\hat{G}_n \to H$ in distribution in $\mathcal{L}_\infty$, where is H is a random element of $\mathcal{L}_\infty$.

If any of these hold true then $\mathbf{E}t(F,\Gamma) = \mathbf{E}t_{\mathrm{inj}}(F, H|_{[k]})$ for every $F \in \Pi(\mathcal{C})$ with $V(F) = [k]$, and also H is exchangeable.

Proof. If $G \in \Pi(\mathcal{K})$ and $F \in \Pi(\mathcal{F})$ with $V(F) = [k]$ with $|V(G)| \geq k$ then $\mathbf{E}t_{\mathrm{inj}}(F, \hat{G}|_{[k]}) = t_{\mathrm{inj}}(F, G)$, where the expectation is taken with respect to the random (re-)labeling $\hat{G}$ of G. For completeness we mention that for a labeled, finite G the quantity $t(F, G)$ is understood as $t(F, G')$ with G' being the unlabeled version of G. Therefore by our assumptions regarding $(G_n)_{n\geq 1}$ and the fact that $0 \leq t(F, G) \leq 1$ (as $\|F\|_\infty \leq 1$) we have that $|\mathbf{E}t_{\mathrm{inj}}(F, \hat{G}_n|_{[k]}) - \mathbf{E}t_{\mathrm{inj}}(F, G_n)| \leq \mathbf{P}(|V(G_n)| < k)$ for each $n \geq 1$ and $F \in \Pi^k(\mathcal{F})$. The expectation in the first term is taken with respect to the distribution of G_n and the random labeling and in the second only with respect to G_n, it will be meant also this way in the following if not noted otherwise. That implies, together with $\mathbf{P}(|V(G_n)| < k) \to 0$ and (1), that $\mathbf{E}t_{\mathrm{inj}}(F, \hat{G}_n|_{[k]}) \to \mathbf{E}t(F, \Gamma)$ (see Theorem 2.5). This implies that $\hat{G}_n|_{[k]} \to H_k$ in distribution for some random $H_k \in \mathcal{L}_k$ with $\mathbf{E}t_{\mathrm{inj}}(F, H_k) = \mathbf{E}t(F, \Gamma)$, see Lemma 2.7, furthermore, with appeal to the consistency of the H_k's, there exists a random $H \in \mathcal{L}_\infty$ such that $H|_{[k]} = H_k$ for each $k \geq 1$, so (1) yields (2).

Another consequence is that H is exchangeable: the exchangeability property is equivalent to the vertex permutation invariance of the distributions of $H|_{[k]}$ for each k. This is granted by the fact that $H|_{[k]} = H_k$ and H_k is the weak limit of a vertex permutation invariant random sequence, for each k.

For the converse direction we perform the above steps backwards using $|\mathbf{E}t(F, \hat{G}_n|_{[k]}) - \mathbf{E}t(F, G_n) \leq \mathbf{P}(|V(G_n)| < k)$ again that. Theorem 2.5 certifies now the existence of the suitable random $\Gamma \in \overline{\Pi(\mathcal{K})^+} \setminus \Pi(\mathcal{K})^+$, this shows that (2) implies (1).

□

We built up the framework in the preceding statements to formulate the following theorem, which is the crucial ingredient to the desired representation of limits.

Theorem 2.9. *There is a one-to-one correspondence between random elements of* $\overline{\Pi(\mathcal{K})^+} \setminus \Pi(\mathcal{K})^+$ *and random exchangeable elements of* $\mathcal{L}_\infty$*. Furthermore, there is a one-to-one correspondence between elements of* $\overline{\Pi(\mathcal{K})^+} \setminus \Pi(\mathcal{K})^+$ *and extreme points of the set of random exchangeable elements of* $\mathcal{L}_\infty$*. The relation is established via the equalities* $\mathbf{E}t(F,\Gamma) = \mathbf{E}t_{\text{inj}}(F, H|_{[k]})$ *for every* $F \in \Pi(\mathcal{C})$ *with* $V(F) = [k]$*.*

Proof. Let Γ a random element of $\overline{\Pi(\mathcal{K})^+} \setminus \Pi(\mathcal{K})^+$. Then by definition of $\Pi(\mathcal{K})^+$ there is a sequence $(G_n)_{n\geq 1}$ in $\Pi(\mathcal{K})$ with $|V(G_n)| \to \infty$ in probability such that $\tau^+(G_n) \to \Gamma$ in distribution in $\Pi(\mathcal{K})^+$. Then Theorem 2.8 implies that there exists a random $H \in \mathcal{L}_\infty$ so that $\hat{G}_n \to H$ in distribution in $\mathcal{L}_\infty$, and H is exchangeable. The distribution of $H|_{[k]}$ is determined by the numbers $\mathbf{E}t_{\text{inj}}(F, H|_{[k]})$, see Theorem 2.5,Lemma 2.7, and the claims inside their proofs, and these numbers are provided by the correspondence.

For the converse direction, let H be random exchangeable element of $\mathcal{L}_\infty$. Then let $G_n = H|_{[n]}$, we have $G_n \to H$ in distribution, and also $\hat{G}_n \to H$ in distribution by the vertex permutation invariance of G_n as a node labeled object. Now again we appeal to Theorem 2.8, so $\tau^+(G_n) \to \Gamma$ for a Γ random element of $\overline{\Pi(\mathcal{K})^+} \setminus \Pi(\mathcal{K})^+$, which is determined completely by $\mathbf{E}t(F,\Gamma)$ provided by the correspondence, see Theorem 2.8.

The second version of the relation between non-random Γ's and extreme points of exchangeable elements is proven similarly, the connection is given via $t(F,\Gamma) = \mathbf{E}t_{\text{inj}}(F, H|_{[k]})$ between the equivalent objects.

□

The characterization of the aforementioned extreme points in Theorem 2.9 was given [12], we state it next, but refrain from giving the proof here, as it is completely identical to that of Theorem 5.5. in [12].

Theorem 2.10. *The distribution of* H *that is exchangeable random element of* $\mathcal{L}_\infty$ *is exactly in then an extreme point of the set of exchangeable measures if the random objects* $H|_{[k]}$ *and* $H|_{\{k+1,\dots\}}$ *are independent for any* $k \geq 1$*. In this case the its representing function from Theorem 2.3 is not dependent on the variable corresponding to the empty set.*

Let $W\colon [0,1]^{s([r],r)} \to \mathcal{K}$ be measurable function, we will refer to such an object as a $(\mathcal{K}, r)$-graphon. We define the random exchangeable r-array H_W in $\mathcal{L}_\infty$ as the element that has law $\text{Samp}(W)$. Furthermore, we define $\Gamma_W \in \overline{\Pi(\mathcal{K})^+} \setminus \Pi(\mathcal{K})^+$ to be the element associated with H_W through Theorem 2.8.

Now we are able to formulate the representation theorem for $\mathcal{K}$-decorated r-uniform directed hypergraph limits using the representation of exchangeable arrays, see (2.3). It is an immediate consequence of Theorem 2.8 and Theorem 2.3 above.

Theorem 2.11. *Let* $(G_n)_{n\geq 1}$ *be a sequence in* $\Pi(\mathcal{K})$ *with* $|V(G_n)| \to \infty$ *such that for every* $F \in \Pi(\mathcal{F})$ *the sequence* $t(F, G_n)$ *converges. Then there exists a function* $W\colon [0,1]^{s([r],r)} \to \mathcal{K}$ *such that* $t(F, G_n) \to t(F, \Gamma_W)$ *for every* $F \in \Pi(\mathcal{F})$*.*

Alternatively we can also use the form $W\colon [0,1]^{s([r],r-1)} \to \mathcal{P}(\mathcal{K})$ for $(\mathcal{K},r)$-graphons whose values are probability measures, this was done in [28]. For $\mathcal{K}=\{0,1\}$ the set $\mathcal{P}(\mathcal{K})$ can be identified with $[0,1]$ to get the usual r-graphon form as a function $W\colon [0,1]^{2^r-2}\to[0,1]$.

In previous works, for example in [12], the limits of simple directed graphs without loops were represented by a 4-tuple of 2-graphons $(W_{00},W_{10},W_{01},W_{11})$ that satisfy $\sum_{i,j}W_{ij}(x,y)=1$ and $W_{10}(x,y)=W_{01}(y,x)$ for each $(x,y)\in[0,1]^2$. A generalization of this representation can be given in our case of the $\Pi(\mathcal{K})$ limits the following way. We only present here the case when $\mathcal{K}$ is infinite, the easier finite case can be established analogously. We have to fix a strictly positive measure μ on $\mathcal{K}$, we set this to be the uniform distribution if $\mathcal{K}\subset\mathbb{R}^d$ is a domain or $\mathcal{K}$ is finite. The limit space consists of collections of $(\mathbb{R},r)$-graphons $(W_\alpha)_{\alpha\in\mathcal{A}}$, where $\mathcal{A}$ is the set of all functions $\alpha\colon S_r\to\mathcal{K}$. Additionally the $(W_\alpha)_{\alpha\in\mathcal{A}}$ has to satisfy $\int_{\alpha\in\mathcal{A}}W_\alpha(x)\mathrm{d}\mu^{\otimes S_r}(\alpha)=1$ and $0\le W_{\phi\alpha}(x)=W_\alpha(x_\phi)$ for each $\phi\in S_r$ and $x\in[0,1]^{s([r],r-1)}$. As before, the action of ϕ on $[0,1]^{s([r],r-1)}$ is the induced coordinate permutation by ϕ, with the unit cubes coordinates indexed by non-trivial subsets of $[r]$. Without going into further details we state the connection between the limit form above and that in Theorem 2.11:

$$\int_A W_\alpha((u_S)_{S\in s([r],r-1)})\mathrm{d}\mu^{\otimes S_r}(\alpha)=\mathbf{P}[(W((u_{\phi(S)})_{S\in s([r],r-1)},Y)_{\phi\in S_r})\in A]$$

for every measurable $A\subset\mathcal{A}$ and $(u_S)_{S\in s([r],r-1)}\in[0,1]^{S\in s([r],r-1)}$, where Y uniform on $[0,1]$.

In several applications, among them some presented in the current paper, it is more convenient to use a naive form for the limit representation, from which the limit element in question is not retrievable. The naive limit space consists of *naive* $(\mathcal{K},r)$*-graphons* $\bar{W}\colon[0,1]^r\to\mathcal{P}(\mathcal{K})$, where now the arguments of W are indexed with elements of $[r]$. From a proper r-graphon $W\colon[0,1]^{s([r],r)}\to\mathcal{K}$ we get its naive counterpart by averaging, that is the $\mathcal{K}$-valued random variable $\mathbf{E}[W(x_1,\dots,x_r,(U_S)_{S\subsetneq[r],|S|>1},Y)|Y]$ has distribution $\bar{W}(x_1,\dots,x_r)$, where $(U_S)_{S\subsetneq[r],|S|>1}$ and Y are i.i.d. uniform on $[0,1]$. On a further note we introduce *averaged naive* $(\mathcal{K},r)$*-graphons* for the case, when $\mathcal{K}\subset\mathbb{R}$, these are of the form $\tilde{W}\colon[0,1]^r\to\mathbb{R}$ and are given by complete averaging, that is $\mathbf{E}[W(x_1,\dots,x_r,(U_S)_{S\subseteq[r],|S|>1}]=\tilde{W}(x_1,\dots,x_r)$, where $(U_S)_{S\subseteq[r],|S|>1}$ are i.i.d. uniform on $[0,1]$.

We introduce the naive $(\mathcal{K},r)$-graphon corresponding to a $G\in\Pi(\mathcal{K})$. The function W_G is constructed from the adjacency r-array of G in the natural way, that is by subdividing the unit cube into $|V(G)|^r$ congruent cubes and assigning the corresponding value of G to each of the subcubes as point measures, diagonals receive the point measure concentrated on the special element ι or 0 in the real case. The map W_G can be also viewed as a true r-graphon by adding coordinates to the domain that have no effect on the value it takes.

2.2 Representation of rCSPs as hypergraphs, the limit space

Recall Definition 2.1 for the way we look at rCSPs. Next we elaborate how homomorphism and sampling is meant in the CSP context.

Let $F[x_{i_1},\dots,x_{i_k}]$ be the induced subformula of F on the variable set $\{x_{i_1},\dots,x_{i_k}\}$, and let $\mathbb{G}(k,F)$ denote the random induced subformula $F[Y_1,\dots,Y_k]$ of F, where $Y_1,\dots,Y_k$ are

independently and uniformly chosen elements from $V(F)$. It is clear using this notation that the relation $\omega = (f; e) \in F[x_{i_1}, \dots, x_{i_k}]$ is equivalent to the relation $\phi(\omega) = (f_\phi, \phi(e)) \in F[x_{i_1}, \dots, x_{i_k}]$ for any permutation $\phi \in S_r$, where $f_\phi(l_1, \dots, l_r) = f(l_{\phi(1)}, \dots, l_{\phi(r)})$ and $\phi(e) = (e_{\phi(1)}, \dots, e_{\phi(r)})$. This emergence of symmetry will naturally be reflected in the limit, we will demonstrate this shortly. For $|K| = q$ we identify the set of rCSP formulas with the set of arrays whose entries are the sums of the evaluation tables of the constraints on r-tuples, that is F with $V(F) = [n]$ corresponds to a map $\mathrm{eval}(F)\colon [n] \times \cdots \times [n] \to \{0, 1, \dots, d\}^{([q]^r)}$. This will be the way throughout the paper we look at these objects. It seems that storing the whole structure of an rCSP formula does not provide any further insight, in fact splitting up constraints would produce non-identical formulas in that vaguely described complete structure representation, which does not seem sensible.

We denote the set $\{0, 1, \dots, d\}^{([q]^r)}$ by L for simplicity, which one could also interpret as the set of multisets whose base set is $[q]^r$ and whose elements have multiplicity at most d. This perspective allows us to treat rCSPs as directed r-uniform hypergraphs whose edges are decorated by the aforementioned elements of L, and leads to a representation of rCSP limits that is derived from the general representation of the limit set of $\Pi(L)$. We will show in a moment that the definition of convergence in the previous subsection is basically identical to the convergence via densities of sub-multi-hypergraphs.

The convergence of a sequence of rCSP formulas or equivalently of elements of $\Pi(L)$ is given in Definition 2.6. The map $\psi\colon H \to F$ is a homomorphism between two rCSPs H and F if it maps edges to edges of the same color from the color set $[q]^r$ and is consistent when considered as a mapping between vertex sets, $\psi'\colon V(H) \to V(F)$, for $r = 2$ this is a multigraph homomorphism. Let H be an rCSP, and let $\tilde{H}$ be the corresponding element in $C(L)$ on the same vertex set such that if an entry at the edge e of H is $((H_e)(i_1, \dots, i_r))_{1 \le i_1, \dots, i_r \le q}$ with matrix elements being non-negative integers, then the entry of $\tilde{H}$ at e is $\prod_{1 \le i_1, \dots, i_r \le q} x_{i_1, \dots, i_r}^{(H_e)(i_1, \dots, i_r)}$, more precisely, for an element $A \in L$ the value is given by $\tilde{H}_e(A) = \prod_{1 \le i_1, \dots, i_r \le q} A(i_1, \dots, i_r)^{(H_e)(i_1, \dots, i_r)}$. The linear space generated by the set $\tilde{L} = \{\prod_{1 \le i_1, \dots, i_r \le q} x_{i_1, \dots, i_r}^{d_{i_1, \dots, i_r}} : 0 \le d_{i_1, \dots, i_r} \le d\}$ forms a dense subset of $C(L)$, therefore Theorem 2.5 applies and for a sequence $(G_n)_{n \ge 1}$ requiring the convergence of $t(F, G_n)$ for all $F = \tilde{H}$ with $H \in \Pi(L)$ provides an equivalent formulation of the convergence of rCSPs.

2.3 The limit objects of rCSPs

The limit object will be given through Theorem 4.4 as the space of measurable functions $W\colon [0, 1]^{s([r], r)} \to L$, where, as in the general case, the coordinates of the domain of W are indexed by the non-empty subsets of $[r]$. In our case not every such W will serve as a limit of some sequence, the above mentioned symmetry of the finite object is inherited in the limit.

We state the rCSP version of Theorem 2.11 for completeness.

Theorem 2.12. *Let $(F_n)_{n \ge 1}$ be a sequence of rCSP formulas that evaluate to at most d on all r-tuples with $|V(F_n)| \to \infty$ such that for every H finite rCSP formula with the*

same boundedness property the sequence $t(\tilde{H}, F_n)$ converges. Then there exists a function $W\colon [0,1]^{s([r],r)} \to \{0,1,\dots,d\}^{([q]^r)}$ such that $t(\tilde{H}, F_n) \to t(\tilde{H}, \Gamma_W)$ for every H.

We should keep the notion of the naive and the weighted naive form of the representation of the limit object in mind, as seen above. They will be utilized in further sections, we will not write them out explicitly, as their structure and relationship with true rCSP limits is analogous to the general case.

We conclude the section with a remark that is motivated by the array representation of rCSPs. The next form presented seems to be the least redundant in some aspect, since no symmetry conditions has to be fulfilled.

Remark 2.13 (Exchangeable partition-indexed processes)**.** The most natural exchangeable infinite random object for the one-to-one correspondence of Theorem 2.8 with rCSP limits is the following process.

Definition 2.14. *Let $N_q^r = \{P = (p_1, \dots, p_q) :$ the sets $p_i \subset \mathbb{N}$ are pairwise disjoint and $\sum_{i=1}^q |p_i| = r\}$ be the set of directed partitions of r-subsets of $\mathbb{N}$. We call the random $(X_P)_{P\in N_q^r}$ that takes values in some compact Polish space $\mathcal{K}$ a partition indexed process. The process $(X_P)_{P\in N_q^r}$ has the exchangeability property if its distribution is invariant under the action induced by finite permutations of $\mathbb{N}$, i.e., $(X_P)_{P\in N_q^r} \overset{d}{=} (X_{\rho(P)})_{P\in N_q^r}$ for any $\rho \in \mathrm{Sym}_0(\mathbb{N})$.*

Unfortunately, the existence of a representation theorem for partition-indexed exchangeable processes that offers additional insight over the directed decorated r-array version is not established, and there is little hope in this direction. The reason for this again is the fact that there is no standard way of splitting up elements X_P and $X_{P'}$ non-trivially with P and P' having the same underlying vertex set but are different as partitions into two random parts with the first being identical for the both and the second ones being conditionally independent over the first.

3 Graph and graphon parameter testability

First we will recount the method of sampling from $\mathcal{K}$-valued r-graphs and r-graphons, as well as look into the metrics that will occur.

Let $(U_S)_{S\in s([k],r)}$ be an independent uniform sample from $[0,1]$. Then for an r-graph G, respectively an r-graphon W, the random r-graphs $\mathbb{G}(k,G)$ and $\mathbb{G}(k,W)$ have vertex set $[k]$, and edge weights $W_G((U_{p_e(S)})_{S\in s(\hat{e},r)})$, respectively $W((U_{p_e(S)})_{S\in s(\hat{e},r)})$, for the edge $e = (i_1,\dots,i_r)$ for every $1 \le i_1,\dots,i_r \le k$, where $\hat{e} = \{i_1,\dots,i_r\}$. Keep in mind that $\mathbb{G}(k,G) \neq \mathbb{G}(k,W_G)$, first term correspond to sampling without, the second with replacement, but it is true that $\mathbf{P}(\mathbb{G}(k,G) \neq \mathbb{G}(k,W_G)) \le \frac{r^2}{|V(G)|}$.

Additionally we define the averaged sampled r-graph for $\mathcal{K} \subset \mathbb{R}$ denoted by $\mathbb{H}(k,W)$, it has vertex set $[k]$, and the weight for the edge $e \in [k]^r$ without multiple entries is the conditional expectation $\mathbf{E}[W((U_{p_e(S)})_{S\in s(\hat{e},r)})|(U_S)_{S\in s(\hat{e},1)}]$, and therefore the random r-graph

is measurable with respect to $(U_S)_{S\in s([k],1)}$. The sampled random r-graphs for the naive r-graphons are defined analogously. Note that $\mathbb{H}(k,W_G)=\mathbb{G}(k,W_G)$ for some G, because its values are all point measures. We will sometimes use the notation U_i for $U_{\{i\}}$ for the elements of the sample indexed by singleton sets.

We also mention the definitions of the norms and distances that will play a important role in what follows, here each object is real-valued. The cut norm of an $n\times\cdots\times n$ r-array A is

$$\|A\|_\square=\max_{S_1,\dots,S_r\subset[n]}\left|A(S_1,\dots,S_r)\right|.$$

The cut distance of two labeled r-graphs or r-arrays F and G on the same vertex set $[n]$ is

$$d_\square(F,G)=\frac{1}{n^r}\|F-G\|_\square,$$

where $F(S_1,\dots,S_r)=\sum_{i_j\in S_j}F(i_1,\dots,i_r)$. The edit distance of the same pair is

$$d_1(F,G)=\frac{1}{n^r}\|F-G\|_1.$$

The cut norm of an r-graphon W is

$$||W||_\square=\max_{S_1,\dots,S_r\subset[0,1]}\left|\int\limits_{S_1\times\cdots\times S_r}W(x)\mathrm{d}x\right|,$$

the cut distance of two r-graphons W and U is

$$\delta_\square(W,U)=\inf_{\phi,\psi}||W^\phi-U^\psi||_\square,$$

where the infimum runs over all measure-preserving permutations of $[0,1]$, and the graphon W^ϕ is defined as $W^\phi(x_1,\dots,x_r)=W(\phi(x_1),\dots,\phi(x_r))$. The cut distance for arbitrary unlabeled r-graphs or r-arrays F and G is

$$\delta_\square(F,G)=\delta_\square(W_F,W_G).$$

We remark that the above definition of the cut norm and distance is not satisfactory in an important aspect for $r\geq 3$: not all sub-r-graph densities are continuous functions in the topology induced by this norm even in the most simple case, when $\mathcal{K}=\{0,1\}$. A norm that is sufficient regarding this issue could be given in a similar fashion, see [32], but it would be computationally inefficient to work with.

Originally, in [8], testability of $(\mathcal{K},r)$-graph parameters (which are real functions invariant under r-graph-isomorphisms) was defined as follows.

Definition 3.1. *An $(\mathcal{K},r)$-graph parameter f is testable, if for every $\varepsilon>0$ there exists a $k=k(\varepsilon)\in\mathbb{N}$ such that for every simple $(\mathcal{K},r)$-graph G on at least k vertices*

$$\mathbf{P}(|f(G)-f(\mathbb{G}(k,G))|>\varepsilon)<\varepsilon.$$

An $(\mathcal{K}, r)$-graphon parameter f is a functional on the space of r-graphons that is invariant under the action induced by measure preserving maps from $[0,1]$ to $[0,1]$, that is, $f(W) = f(W^\phi)$. Their testability is defined analogously to Definition 3.1.

A characterization of this property in terms of graph limits was developed in [8] for $\mathcal{K} = \{0,1\}$ in the undirected case.

Theorem 3.2. *[8] Let f be a simple graph parameter, then the following statements are equivalent.*

(i) f is testable.

(ii) For every convergent sequence $(G_n)_{n\geq 1}$ of simple graphs with $|V(G_n)| \to \infty$, $f(G_n)$ also converges.

(iii) For every $\varepsilon > 0$ there exist $\varepsilon' > 0$ and $n_0 \in \mathbb{N}$ such that for every pair G_1 and G_2 of simple graphs with $|V(G_1)|, |V(G_2)| \geq n_0$ and $\delta_\square(G_1, G_2) < \varepsilon'$ together imply $|f(G_1) - f(G_2)| < \varepsilon$.

(iv) There exists a $\delta_\square$-continuous functional f' on the space of graphons, so that $f(G_n) \to f'(W)$ whenever $G_n \to W$.

A closely related notion to parameter testing is property testing. A simple graph property $\mathcal{P}$ is characterized by the subset of the set of simple graphs containing the graphs which have the property, in what follows $\mathcal{P}$ will be identified with this subset.

Definition 3.3. *$\mathcal{P}$ is testable, if there exists another graph property $\mathcal{P}'$, such that*

(a) $\mathbf{P}(\mathbb{G}(k, G)) \in \mathcal{P}') \geq \frac{2}{3}$ for every $G \in \mathcal{P}$, and

(b) for every $\varepsilon > 0$ there is a $k(\varepsilon)$ such that for every $k \geq k(\varepsilon)$ and G with $d_1(G, \mathcal{P}) \geq \varepsilon$ we have that $\mathbf{P}(\mathbb{G}(k, G)) \in \mathcal{P}') \leq \frac{1}{3}$.

Note that $\frac{2}{3}$ and $\frac{1}{3}$ in the definition can be replaced by arbitrary constants $1 > a > b > 0$, this change may alter the corresponding certificate $\mathcal{P}'$.

Lemma 3.4. *[29] $\mathcal{P}$ is a testable graph property if and only if $d_1(., \mathcal{P})$ is a testable graph parameter.*

The link between the two notions can be proven by definition without serious difficulties. These concepts may be extended to the infinitary space of graphons, where a similar notion of sampling is available.

We list some simple consequences of Theorem 3.2.

Remark 3.5.

1. In the case $r = 2$, testability of a graphon parameter is equivalent to continuity in the $\delta_\square$ distance.

2. In the case $r \geq 3$, testability of a hypergraph parameter f implies that for any left convergent sequence $(H_n)_{n\geq 1}$ of hypergraphs $f(H_n)$ also converges. The other direction is open, as there is no analogous notion of distance to the 2-dimensional $\delta_\square$ in higher dimensions, since there is no standard way to compare a large hypergraph H_n to the random induced subgraph on a uniform sample of the vertices of $V(H_n)$ of fixed size.

3.1 Examples of efficiently testable parameters

We introduce now a notion of efficient parameter testability. Definition 3.1 of testability does not ask for a specific upper bound on $k(\varepsilon)$ in terms of ε, but in applications the order of magnitude of this may be an important issue. Therefore we introduce a more restrictive class of graph parameters, we refer to them as being efficiently testable.

Definition 3.6. *An r-graph parameter f is called β-testable for a family of measurable functions $\beta = \{\beta_i \colon \mathbb{R}^+ \to \mathbb{R}^+ : i \in I\}$, if there exists a β_i such that for every $\varepsilon > 0$ and r-graph G*

$$P(|f(G) - f(\mathbb{G}(\beta_i(\varepsilon), G))| > \varepsilon) < \varepsilon.$$

With slight abuse of notation we will also use the notion of β-testability for a family containing only a single function β. The term *efficient testability* will serve as shorthand for β-testability for some (family) of functions $\beta(\varepsilon)$ that are polynomial in $\frac{1}{\varepsilon}$. We refer to the smallest β also as the *sample complexity* of f for which it is β-testable.

We will list some examples of graph parameters, for which we have information about their sample complexity.

Example 3.7. One of the most basic testable graph parameters are subgraph densities $f_F(G) = t(F, G)$, where F is a simple graph. Lemma 4.4 from [8] states that

$$\mathbf{P}(|t(F, G) - t(F, \mathbb{G}(k, G))| > \varepsilon) < 2 \exp\left(-\frac{\varepsilon^2 k}{4|V(F)|^2}\right), \tag{3.1}$$

which implies that for any F, f_F is $O(\log(\frac{1}{\varepsilon})\varepsilon^{-2})$-testable. In the case of $(\mathcal{K}, r)$-graphs for $r \geq 3$ the same as (3.1) holds, no modifications in the proof of the above result are required.

Example 3.8. For $r = 2$, $q \in \mathbb{N}$, $J \in \mathbb{R}^{q \times q}$ and $h \in \mathbb{R}^q$ we consider

$$\hat{\mathcal{E}}(G, J, h) = \max_{\phi \colon V(G) \to [q]} \frac{1}{n^2} \sum_{1 \leq i,j \leq q} J_{ij} e(\phi^{-1}(i), \phi^{-1}(j)) + \frac{1}{n} \sum_{1 \leq i \leq q} h_i |\phi^{-1}(i)|,$$

the ground state energy of the graph G (cf. [9]), where $e(S, T)$ denotes the number of edges going form S to T. These graph functions originate from statistical physics, for the rigorous mathematical treatment of the topic see e.g. Sinai's book [31]). In the literature this notion is also often to be found with negative sign or different normalization.

This graph parameter can be expressed in the terminology applied for MAX-2CSP. Let the corresponding 2CSP formula to the pair (G,J) be F with domain $K = [q]$. F consists of the constraints $(g_0; (i, j))$ for every edge (i, j) of G, where g_0 is the constraint type whose evaluation table is J, and additionally it contains n copies of $(g_1; i)$ for every vertex i of G, where g_1 is the constraint type in one variable with evaluation vector h. Then the optimal value of the objective function of the MAX-2CSP problem of the instance F is equal to $\hat{\mathcal{E}}(G, J, h)$. Note that this correspondence is consistent with the sampling procedure, that is, to $(\mathbb{G}(k, G),J)$ corresponds the 2CSP $\mathbb{G}(k, F)$. Therefore for $q = 2$, $\hat{\mathcal{E}}(., J, h)$ has sample complexity $O(\frac{1}{\varepsilon^4})$(see [2],[30]).

These energies are directly connected to the number $\hom(G,H)$ of admissible vertex colorings of G by the colors $V(H)$ for a certain small weighted graph H. This was pointed out in [9], (2.16), namely

$$\frac{1}{|V(G)|^2}\ln \hom(G,H) = \hat{\mathcal{E}}(G,J,0) + O\left(\frac{1}{|V(G)|}\right), \tag{3.2}$$

where the edge weights of H are $\beta_{ij}(H) = \exp(J_{ij})$. The former line of thought of turning ground state energies into 2CSPs is also valid in the case of r-graphs.

The results on MAX-rCSP sample complexity for $q = 2$ can be extended beyond the case of simple hypergraphs, higher dimensional Hamiltonians are also expressible as rCSP formulas. The generalization for arbitrary q and to r-graphons will follow in the next subsection. Additionally we note, that an analogous statement to (3.2) on testability of coloring numbers does not follow immediately for $r \geq 3$.

On the other hand, with the notion of ground state energy available, we are able to rewrite the MAX-2CSP in a compact form as an energy problem. We will execute this task right away for limit objects. First, we introduce the ground state energy of a 2-graphon with respect to an interaction matrix J.

Definition 3.9. *Let $q \geq 1$, $J \in \mathbb{R}^{q\times q}$. Then the ground state energy of the 2-graphon W with respect to J is*

$$\mathcal{E}(W,J) = \max_{\phi} \sum_{z\in[q]^2} J_z \int\limits_{[0,1]^2} \prod_{m=1}^{2} \phi_{z_m}(x_m) W(x)\mathrm{d}x,$$

where ϕ runs over all q-fractional partitions of $[0,1]$.

Let $K = [q]$, $L = \{0,1,\ldots,d\}^{[q]^2}$ and $(F_n)_{n\geq 1}$ be a convergent sequence of 2CSPs. Consider the corresponding sequence of graphs $\mathrm{eval}(F_n) = (\tilde{F}_n^z)_{z\in[q]^2}$ for each n, and let $W = (W^z)_{z\in[q]^2}$ be the respective limit. Let f be the $(L,2)$-graph parameter so that $f(\mathrm{eval}(F))$ is equal to the density of the MAX-2CSP value for the instance F. Then it is not hard to see that f can be extended to the limit space the following way

$$f(W) = \max_{\phi} \sum_{i,j=1}^{q} \int\limits_{[0,1]^2} \phi_i(x)\phi_j(y) W^{i,j}(x,y)\mathrm{d}x\mathrm{d}y,$$

where ϕ runs over all q-fractional partitions of $[0,1]$. The formula is a special case of *layered* ground state energy with the interaction matrices defined by $J^{i,j}(k,l) = \delta_i(k)\delta_j(l)$ that is defined below.

Example 3.10. We will investigate the efficiency of graph parameters that obey some additional continuity condition in the $\delta_\square$ metric. Direct consequence of results from [8] will be presented in the next lemma.

Lemma 3.11. *Let f be a simple graph parameter that is α-Hölder-continuous in the $\delta_\square$ metric in the following sense: for every $\varepsilon > 0$ there exists $n_0(\varepsilon)$, such that there is a $C > 0$ so that for every G_1, G_2 with $|V(G_1)|, |V(G_2)| > n_0(\varepsilon)$ and $\delta_\square(G_1, G_2) < \varepsilon$, then $|f(G_1) - f(G_2)| < C\varepsilon^\alpha$. Then f is $\max\{2^{O\left(\frac{1}{\varepsilon^{2/\alpha}}\right)}, n_0(\varepsilon)\}$-testable.*

Proof. To see this, let us fix $\varepsilon > 0$. Then for an arbitrary simple graph G with $|V(G)| \geq n_0(\varepsilon)$ and $k \geq n_0(\varepsilon)$ we have

$$|f(G) - f(\mathbb{G}(k, G))| < C\left[\delta_\square(G, \mathbb{G}(k, G))\right]^\alpha < C\left(\frac{10}{\sqrt{\log_2 k}}\right)^\alpha, \tag{3.3}$$

with probability at least $1 - \exp(-\frac{k^2}{2\log_2 k})$. The last probability bound is the statement of Theorem 2.9 of [8]. We may rewrite (3.3) by setting $\varepsilon = C\left(\frac{10}{\sqrt{\log_2 k}}\right)^\alpha$, the substitution implies that f is $2^{O\left(\varepsilon^{-2/\alpha}\right)}$-testable, whenever $n_0(\varepsilon) < 2^{O(\varepsilon^{-2/\alpha})}$. □

This latter approach is hard to generalize to r-graphs for $r \geq 3$ because of the absence of a suitable metric.

4 Testability of the ground state energy

Next we introduce a generalization of the notion of the ground state energy of graphs from [9], see Definition 3.9, and restate the results of [2] on sample complexity of MAX-rCSP in that framework. The parameter derived from the maximal constraint satisfaction problem will also serve as an example for an efficiently testable colored hypergraph parameter. We will further generalize the main result of [2] in several directions.

First we provide the basic definition of the energy of a $(\mathcal{K}, r)$-graphon W with respect to some $q \geq 1$, an r-array $J \in C(\mathcal{K})^{q\times\cdots\times q}$, and a fractional partition $\phi = (\phi_1, \ldots, \phi_q)$. With slight abuse of notation, the graphons in the upcoming parts of the section assume both the $\mathcal{K}$-valued and the probability measure valued form, it will be clear from the context which one of them is meant.

A fractional partition satisfies the properties that the ϕ_i's are measurable functions from $[0, 1]$ to $[0, 1]$, and for each $x \in [0, 1]$ the equality $\sum_i \phi_i(x) = 1$ holds. Let

$$\mathcal{E}_\phi(W, J) = \sum_{z_1,\ldots,z_r=1}^{q} \int_{[0,1]^{s([r],r)}} J_{z_1,\ldots,z_r}(W(x)) \prod_{j=1}^{r} \phi_{z_j}(x_{\{j\}})\mathrm{d}x.$$

The value of the above integral can be determined by first integrating over the coordinates corresponding to subsets of $[r]$ with at least two elements, and then over the remaining ones. The first partial integral is not dependent on ϕ, so it can be calculated in advance.

When dealing with an integer partition $\phi = (\mathbb{I}_{T_1}, \dots, \mathbb{I}_{T_q})$, one is able to rewrite the former expression as

$$\mathcal{E}_\phi(W, J) = \sum_{z_1,\dots,z_r=1}^{q} \int_{p^{-1}_{s([r],1)}(T_{z_1}\times\cdots\times T_{z_r})} J_{z_1,\dots,z_r}(W(x))\mathrm{d}x,$$

where p_D stands for the projection to the coordinates contained in the set D.

The energy of a $\mathcal{K}$-valued r-graph G on k vertices with respect to $J \in C(\mathcal{K})^{q\times\cdots\times q}$ and q-fractional partition $x_n = (x_{n,1}, \dots, x_{n,q})$ for $n = 1, \dots, k$ (i. e., $x_{n,m} \in [0,1]$ and $\sum_m x_{n,m} = 1$) is defined as

$$\mathcal{E}_{\mathrm{x}}(G, J) = \frac{1}{k^r} \sum_{z_1,\dots,z_r=1}^{q} \sum_{n_1,\dots,n_r=1}^{k} J_{z_1,\dots,z_r}(G_{n_1,\dots,n_r}) \prod_{j=1}^{r} x_{n_j,z_j}. \tag{4.1}$$

Remark 4.1. Ground state energies and subgraph densities are Lipschitz continuous graph parameters in the sense of Lemma 3.11 ([8],[9]), but that result implies much weaker upper bounds on the sample complexity, than the best ones known to date. This is due to the fact, that $\delta_\square(G, \mathbb{G}(k, G))$ decreases logarithmically slowly in k, which is the result of the hardness of finding a near optimal overlay between two graphons through a measure preserving permutation of $[0,1]$ in order to calculate their $\delta_\square$ distance. On the other hand, if the sample size $k(\varepsilon)$ is exponentially large in $1/\varepsilon$, then the distance $\delta_\square(G, \mathbb{G}(k, G))$ is small enough with high probability, therefore all Hölder-continuous graph parameters can be estimated simultaneously with high success probability.

Next we introduce the layered version of the ground state energy. This is a generalized optimization problem where we wish to obtain the optimal value corresponding to fractional partitions of the sums of energies over a finite layer set.

Definition 4.2. *Let E be a finite layer set, $\mathcal{K}$ be a compact set, and $W = (W^e)_{e\in E}$ be a tuple of $(\mathcal{K}, r)$-graphons. Let q be a fixed positive integer and $J = (J^e)_{e\in E}$ with $J^e \in C(\mathcal{K})^{q\times\cdots\times q}$ for every $e \in E$. For a $\phi = (\phi_1, \dots, \phi_q)$ q-fractional partition of $[0,1]$ let*

$$\mathcal{E}_\phi(W, J) = \sum_{e\in E} \mathcal{E}_\phi(W^e, J^e)$$

and let

$$\mathcal{E}(W, J) = \max_\phi \mathcal{E}_\phi(W, J),$$

*denote the **layered ground state energy**, where the maximum runs over all q-fractional partitions of $[0,1]$. We define for $G = (G^e)_{e\in E}$ the energy $\mathcal{E}_{\mathrm{x}}(G, J)$ analogously as the energy sum over E, see (4.1) above, and $\hat{\mathcal{E}}(G, J) = \max_{\mathrm{x}} \mathcal{E}_{\mathrm{x}}(G, J)$ where the maximum runs over q-integer partitions ($x_{n,m} \in \{0, 1\}$), respectively $\mathcal{E}(G, J) = \max_{\mathrm{x}} \mathcal{E}_{\mathrm{x}}(G, J)$, where the maximum is taken over all q-fractional partitions* x.

Now we will rewrite the boolean limit MAX-rCSP as a layered ground state energy problem. Let $E = \{0,1\}^r$, $\mathcal{K} = \{0,1,\ldots,d\}$, $W = (W^z)_{z\in\{0,1\}^r}$ with W^z being $(\mathcal{K},r)$-graphons, and let

$$\alpha(W) = \max_{\phi} \sum_{z\in\{0,1\}^r} \int\limits_{[0,1]^{s([r],r)}} \prod_{j=1}^{r} \phi(x_{\{j\}})^{z_j}(1-\phi(x_{\{j\}}))^{1-z_j} W^z(x)\mathrm{d}x,$$

where the maximum is taken over all measurable functions $\phi\colon [0,1] \to [0,1]$. If F is a $(\mathcal{K}^{(\{0,1\}^r)},r)$-graph induced by an rCSP formula, then the finite integer version of α denoted by $\hat{\alpha}(F)$ (see $\hat{\mathcal{E}}(G,J)$ above) is equal to the density of the optimum of the MAX-rCSP problem of F.

We list the involved parameters in the layered ground state energy problem. These are the dimension r, the layer set E, the number of states q, the color set $\mathcal{K}$, the finite or limit case. Our main theorem will be a generalization of the following theorem on sample complexity of rCSPs with respect to these factors.

Theorem 4.3. *[2] Let F be a Boolean rCSP formula. Then for any $\varepsilon > 0$ and $\delta > 0$ we have that for every $k \geq 10^6 r^{12}\delta^{-5}\varepsilon^{-4}\log(\varepsilon^{-1})$ we have*

$$\mathbf{P}\left(|\hat{\alpha}(\mathrm{eval}(F)) - \hat{\alpha}(\mathbb{G}(k,\mathrm{eval}(F)))| > \varepsilon\right) < \delta.$$

We will see in what follows that also the infinitary version of the above statement is true. It will be stated in terms of generalized ground state energies of colored hypergraphs, and will settle the issue regarding the efficiency of testability in the greatest generality with respect to the previously highlighted aspects. In order to simplify the analysis we introduce the *canonical form* of the problem, that is ground state energies of $[q]^r$-layered $([-d,d],r)$-graphons with the special interaction r-arrays $\hat{J}^z$ for each $z \in [q]^r$, that have the identity function $f(x) = x$ as the $(z_1,\ldots,z_r)$ entry and the constant 0 function as the other entries.

Theorem 4.4. *Let E be a finite layer set, $\mathcal{K}$ a compact color set, $q \geq 1$, r-arrays $J = (J^e)_{e\in E}$ with $J^e \in C(\mathcal{K})^{q\times\cdots\times q}$, and $\varepsilon > 0$. Then we have that for any E-layered $(\mathcal{K},r)$-graphon $W = (W^e)_{e\in E}$ and $k \in O(\varepsilon^{-4}\log(\varepsilon^{-1}))$ that*

$$\mathbf{P}(|\mathcal{E}(W,J) - \hat{\mathcal{E}}(\mathbb{G}(k,W),J)| > \varepsilon|E|\|J\|_\infty\|W\|_\infty) < \varepsilon.$$

The proof of Theorem 4.4 will go along the lines of the proof of Theorem 4.3 from [2] with most of the required lemmas being refinements of the respective ones in the proof of that theorem. We will formulate and verify these auxiliary lemmas one after another, afterwards we will compile them to prove the main statement. The arguments made in [2] carry through adapted to our continuous setting with some modifications, and we will also draw on tools from [8] and [9].

The first lemma tells us that in the real-valued case the energy of the sample and that of the averaged sample do not differ by a large amount.

Lemma 4.5. *Let W be a $([-d,d],r)$-graphon, $q \geq 1$, $J \in \mathbb{R}^{q\times\cdots\times q}$. Then for every $k \geq 1$ there is coupling of $\mathbb{G}(k,W)$ and $\mathbb{H}(k,W)$ such that*

$$\mathbf{P}\left(|\hat{\mathcal{E}}(\mathbb{G}(k,W),J) - \hat{\mathcal{E}}(\mathbb{H}(k,W),J)| > \varepsilon\|J\|_\infty\|W\|_\infty\right) \leq 2\exp\left(-k\left(\frac{\varepsilon^2 k}{2} - \log q\right)\right)$$

Proof. Let us fix a q-integer partition x of $[k]$, and furthermore let the two random r-graphs be generated by the same sample $(U_S)_{S\in s([k],r)}$. Then

$$\hat{\mathcal{E}}_{\mathrm{x}}(\mathbb{G}(k,W),J) = \frac{1}{k^r}\sum_{z_1,\dots,z_r=1}^{q}\sum_{n_1,\dots,n_r=1}^{k} J_{z_1,\dots,z_r} W((U_S)_{S\in s(\{n_1,\dots,n_r\},r)})\prod_{j=1}^{r} x_{n_j,z_j},$$

and

$$\begin{aligned}&\hat{\mathcal{E}}_{\mathrm{x}}(\mathbb{H}(k,W),J)\\ &= \frac{1}{k^r}\sum_{z_1,\dots,z_r=1}^{q}\sum_{n_1,\dots,n_r=1}^{k} J_{z_1,\dots,z_r}\mathbf{E}[W((U_S)_{S\in s(\{n_1,\dots,n_r\},r)})|(U_S)_{S\in s(\{n_1,\dots,n_r\},1)}]\prod_{j=1}^{r} x_{n_j,z_j}.\end{aligned}$$

Let us enumerate the elements of $\binom{k}{2}$ as $e_1, e_2, \dots, e_{\binom{k}{2}}$, and define the martingale $Y_0 = \mathbf{E}[\hat{\mathcal{E}}_{\mathrm{x}}(\mathbb{G}(k,W),J)]\{U_j|j\in[k]\}]$, and $Y_t = \mathbf{E}[\hat{\mathcal{E}}_{\mathrm{x}}(\mathbb{H}(k,W),J)|\{U_j|j\in[k]\}\cup\cup_{j=1}^{t}\{U_S|e_j\subset S\}]$ for each $1\leq t\leq\binom{k}{2}$, so that $Y_0 = \hat{\mathcal{E}}_{\mathrm{x}}(\mathbb{H}(k,W),J)$ and $Y_{\binom{k}{2}} = \hat{\mathcal{E}}_{\mathrm{x}}(\mathbb{G}(k,W),J)$. for any $t\in\binom{k}{2}$ we can upper bound the difference, $|Y_{t-1}-Y_t| \leq \frac{1}{k^2}\|J\|_\infty\|W\|_\infty$. By the Azuma-Hoeffding inequality it follows that

$$\mathbf{P}(|Y_t - Y_0| \geq \rho) \leq 2\exp\left(-\frac{\rho^2 k^4}{2\binom{k}{2}\|J\|_\infty^2\|W\|_\infty^2}\right) \leq 2\exp\left(-\frac{\rho^2 k^2}{2\|J\|_\infty^2\|W\|_\infty^2}\right), \tag{4.2}$$

for any $\rho > 0$.

There are q^k distinct q-integer partitions of $[k]$, hence

$$\mathbf{P}\left(|\hat{\mathcal{E}}(\mathbb{G}(k,W),J) - \hat{\mathcal{E}}(\mathbb{H}(k,W),J)| > \varepsilon\|J\|_\infty\|W\|_\infty\right) \leq 2\exp\left(-k\left(\frac{\varepsilon^2 k}{2} - \log q\right)\right). \tag{4.3}$$

□

In the following lemmas every r-graph of graphon is meant as bounded real-valued and directed.

We would like to point out in the beginning that in the finite case we are able to shift from the integer optimization problem to the relaxed one with having a reasonably good upper bound on the difference of the optimal values of the two.

Lemma 4.6. *Let G be a real-valued r-graph on $[k]$ and $J \in \mathbb{R}^{q\times\cdots\times q}$. Then*

$$|\mathcal{E}(G,J) - \hat{\mathcal{E}}(G,J)| \leq r^2 \frac{1}{2k} ||G||_\infty ||J||_\infty.$$

Proof. Trivially we have $\mathcal{E}(G,J) \geq \hat{\mathcal{E}}(G,J)$. We define G' by setting all entries of G to 0 which have at least two coordinates which are the same (for $r = 2$ these are the diagonal entries). Thus, we get that

$$|\mathcal{E}(G,J) - \mathcal{E}(G',J)| \leq \binom{r}{2} \frac{1}{k} ||G||_\infty ||J||_\infty.$$

Now assume that we are given a fractional partition $\overline{x}$ for that the maximum $\mathcal{E}(G',J)$ is attained. We fix all the entries $\overline{x}_{n,1}, \ldots \overline{x}_{n,q}$ with $n = 2, \ldots, k$ and regard $\mathcal{E}_x(G',J)$ as a function of $x_{1,1}, \ldots, x_{1,q}$. This function will be linear in the variables $x_{1,1}, \ldots, x_{1,q}$, and with the additional condition $\sum_{j=1}^{r} x_{1,j} = 1$ we obtain a linear program. By standard arguments this program possesses an integer valued the optimal evaluation, so we are allowed to replace $\overline{x}_{1,1}, \ldots, \overline{x}_{1,q}$ by integers without letting $\mathcal{E}_x(G',J)$ decrease. We repeat this procedure for each $n \in [k]$, obtaining an integer optimum for $\mathcal{E}_x(G',J)$, which implies that $\mathcal{E}(G',J) = \hat{\mathcal{E}}(G',J)$. Hence, the claim follows. □

Next lemma is the continuous generalization of Theorem 4 from [2].

Lemma 4.7. *Let $\varepsilon > 0$ arbitrary. For any bounded measurable function $W\colon [0,1]^r \to \mathbb{R}$ there exist an $s \leq \frac{1}{\varepsilon^2}$, measurable sets $S_i^j \subset [0,1]$ with $i = 1, \ldots, s$, $j = 1, \ldots, r$, and real numbers $d_1, \ldots, d_s$ so that with $B = \sum_{i=1}^{s} d_i \mathbb{I}_{S_i^1 \times \cdots \times S_i^r}$ it holds that*

(i) $||W||_2 \geq ||W - B||_2$,

(ii) $||W - B||_\square < \varepsilon ||W||_2$, *and*

(iii) $\sum_{i=1}^{s} |d_i| \leq \frac{1}{\varepsilon} ||W||_2$.

Proof. We construct stepwise the required rectangles and the respective coefficients implicitly. Let $W^0 = W$, and suppose that after the t'th step of the construction we have already obtained every set $S_i^j \subset [0,1]$ with $i = 1, \ldots, t$, $j = 1, \ldots, r$, and the real numbers $d_1, \ldots, d_t$. Set $W^t = W - \sum_{i=1}^{t} d_i \mathbb{I}_{S_i^1 \times \cdots \times S_i^r}$. We proceed to the $(t+1)$'st step, where two possible situations can occur. The first case is when

$$||W^t||_\square \geq \varepsilon ||W||_2.$$

This implies by definition that there exist measurable subsets $S_{t+1}^1, \ldots, S_{t+1}^r$ of $[0,1]$ such that $|\int_{S_{t+1}^1 \times \cdots \times S_{t+1}^r} W^t(x) \mathrm{d}x| \geq \varepsilon ||W||_2$. We define d_{t+1} to be the average of W^t on the product set $S_{t+1}^1 \times \cdots \times S_{t+1}^r$, and proceed to the $(t+2)$'nd step. In the case of

$$||W^t||_\square < \varepsilon ||W||_2$$

we are ready with the construction and set $s = t$.

We analyze the first case to obtain an upper bound on the total number of steps required by the construction. Suppose that at some step we are not ready with the construction. Then

$$||W^t||_2^2 - ||W^{t+1}||_2^2 = \int\limits_{S_{t+1}^1 \times \cdots \times S_{t+1}^r} (W^t)^2(x)\mathrm{d}x - \int\limits_{S_{t+1}^1 \times \cdots \times S_{t+1}^r} (W^t(x) - d_{t+1})^2\mathrm{d}x$$
$$= d_{t+1}^2 \lambda(S_{t+1}^1) \dots \lambda(S_{t+1}^r) \geq \varepsilon^2 ||W||_2^2. \tag{4.4}$$

This means that the square of the 2-norm decreases in every step when the first case occurs in the construction by at least $\varepsilon^2||W||_2^2$, therefore it can happen only at most $\frac{1}{\varepsilon^2}$ times, with other words $s \leq \frac{1}{\varepsilon^2}$. It is also clear that the 2-norm decreases in each step, so we are left to verify the upper bound on the sum of the absolute values of the coefficients d_i. From (4.4) we get, that

$$||W||_2^2 = \sum_{t=1}^{s} ||W^{t-1}||_2^2 - ||W^t||_2^2 \geq \sum_{t=1}^{s} d_t^2 \lambda(S_t^1) \dots \lambda(S_t^r).$$

We also know for every $t \leq s$ that $|d_t|\lambda(S_t^1) \dots \lambda(S_t^r) \geq \varepsilon||W||_2$. Hence,

$$\sum_{t=1}^{s} |d_t|\varepsilon||W||_2 \leq \sum_{t=1}^{s} d_t^2 \lambda(S_t^1) \dots \lambda(S_t^r) \leq ||W||_2^2,$$

and therefore $\sum_{t=1}^{s} |d_t| \leq \frac{1}{\varepsilon}||W||_2$. □

Next we state that the cut approximation provided by Lemma 4.7 is invariant under sampling. This is a crucial point of the whole argument, and is the r-dimensional generalization of Lemma 4.6 from [8].

Lemma 4.8. *For any $\varepsilon > 0$ and bounded measurable function $W\colon [0,1]^r \to \mathbb{R}$ we have that*

$$\mathbf{P}\left(\left|\frac{1}{k^r}||\mathbb{H}(k,W)||_\square - ||W||_\square\right| > \varepsilon||W||_\infty\right) < \varepsilon,$$

where $k \geq \frac{c}{\varepsilon^4}$ for some c large enough.

Proof. Let us denote $\mathbb{H}(k,W)$ by G. We will need the following lemma from [2].

Lemma 4.9. *G is a real r-array on some finite product set $V_1 \times \cdots \times V_r$, where V_i are copies of V of cardinality k. Let $S_1 \subset V_1, \dots, S_r \subset V_r$ be fixed subsets and Q_1 a uniform random subset of $V_2 \times \cdots \times V_r$ of cardinality p. Then*

$$G(S_1, \dots, S_r) \leq E_{Q_1} G(P(Q_1 \cap S_2 \times \cdots \times S_r), S_2, \dots, S_r) + \frac{k^{r/2}}{\sqrt{p}}||G||_2,$$

where $P(Q_1) = P_G(Q_1) = \{x_1 \in V_1 : \sum_{(y_2,\dots y_r)\in Q_1} G(x_1, y_2, \dots, y_r) > 0\}$.

If apply Lemma 4.9 repeatedly r times to the r-arrays G and $-G$, then we arrive at an upper bound on $G(S_1,\dots,S_r)$ ($(-G)(S_1,\dots,S_r)$ respectively) for any collection of the $S_1,\dots,S_r$ which does not depend on the particular choice of these sets any more, so we get that

$$||G||_\square \leq E_{Q_1,\dots,Q_r} \max_{Q_i'\subset Q_i} \max\{G(P_G(Q_1'),\dots,P_G(Q_r')); (-G)(P_{-G}(Q_1'),\dots,P_{-G}(Q_r'))\} + \frac{rk^r}{\sqrt{p}}||G||_\infty, \tag{4.5}$$

since $||G||_2 \leq k^{r/2}||G||_\infty$.

Let us recall that G stands for the random $\mathbb{H}(k,W)$. We are interested in the expectation $\mathbf{E}$ of the left hand side of (4.5) over the sample that defines G. Now we proceed via the method of conditional expectation. We establish an upper bound on the expectation of right hand side of (4.5) over the sample $U_1,\dots,U_k$ for each choice of the tuple of sets $Q_1,\dots,Q_r$. This bound does not depend on the actual choice of the Q_i's, so if we take the average (over the Q_i's), that upper bound still remains valid.

In order to do this, let us fix $Q_1,\dots,Q_r$, set Q to be the set of elements of $V(G)$ which are contained by at least of the Q_i's, and fix also the sample points of U_i with $i\in Q$. Take the expectation only over the remaining U_i sample points.

To this end, by Fubini we have the estimate

$$\mathbf{E}_{U_{[k]}}||G||_\square \leq E_{Q_1,\dots Q_r}\mathbf{E}_{U_Q}[\mathbf{E}_{U_{Q^c}} \max_{Q_i'\subset Q_i} \max\{G(P_G(Q_1')\cap Q^c,\dots,P_G(Q_r')\cap Q^c); (-G)(P_{-G}(Q_1')\cap Q^c,\dots,P_{-G}(Q_r')\cap Q^c)\}] + \frac{rk^r}{\sqrt{p}}||G||_\infty + prk^{r-1}||G||_\infty, \tag{4.6}$$

where $U_S=\{U_i : i\in S\}$.

Our goal is to uniformly upper bound the expression in the brackets in (4.6) so that in the dependence on the particular $Q_1,\dots Q_r$ and the sample points from U_Q vanishes. To achieve this, we consider additionally a tuple of subsets $Q_i'\subset Q_i$, and introduce the random variable $Y(Q_1',\dots,Q_r') = G(P_G(Q_1')\cap Q^c,\dots,P_G(Q_r')\cap Q^c)$, where the randomness comes from U_{Q^c} exclusively. Let

$$T_i = \{x_i\in[0,1] : \sum_{(y_1,\dots,y_{i-1},y_{i+1},\dots y_r)\in Q_i'} W(U_{y_1},\dots,U_{y_{i-1}},x_i,U_{y_{i+1}},\dots U_{y_r})>0\}$$

for $i\in[r]$. Note that $t_i\in P_G(Q_i')$ is equivalent to $U_{t_i}\in T_i$. Then

$$\begin{aligned}\mathbf{E}_{U_{Q^c}}Y(Q_1',\dots,Q_r') &\leq \sum_{\substack{t_1,\dots,t_r\in Q^c\\ t_i\neq t_j}} \mathbf{E}_{U_{Q^c}}G(t_1,\dots,t_r)\mathbb{I}_{P_G(Q_1')}(t_1)\dots\mathbb{I}_{P_G(Q_r')}(t_r) + r^2k^{r-1}||W||_\infty \\ &\leq k^r \int_{T_1\times\dots\times T_r} W(x)\mathrm{d}x + r^2k^{r-1}||W||_\infty \leq k^r||W||_\square + r^2k^{r-1}||W||_\infty.\end{aligned}$$

By the Azuma-Hoeffding inequality we also have high concentration of the random variable $Y(Q'_1, \ldots, Q'_r)$ around its mean.

$$\mathbf{P}(Y(Q'_1, \ldots, Q'_r) \geq \mathbf{E}_{U_{Q^c}} Y(Q'_1, \ldots, Q'_r) + \rho k^r) < \exp(-\frac{\rho^2 k}{8r^2 ||W||_\infty}). \tag{4.7}$$

Analogous upper bounds on the expectation and the tail probability hold for each of the expressions $(-G)(P_{-G}(Q'_1), \ldots, P_G(Q'_r))$.

With regard to the maximum expression in (4.6) over the Q'_i's we have to this end either that the concentration event from (4.7) holds for each possible choice of the Q'_i's for both expressions, this has probability at least $1 - 2^{pr+1} \exp(-\frac{\rho^2 k}{8r^2 ||W||_\infty})$, or it fails for some choice. In the first case we can employ the upper bound $k^r ||W||_\square + r k^{r-1} ||W||_\infty + \rho k^r$. In the event of failure we still have the trivial upper bound of $k^r ||W||_\infty$. Eventually we presented an upper bound on the expectation that is not dependent on $Q_1, \ldots Q_r$, and the sample points from U_Q. Hence by taking expectation and assembling the terms, we have

$$\frac{1}{k^r} \mathbf{E}_{U_{[k]}} ||G||_\square \leq ||W||_\square + ||W||_\infty \left(\frac{r}{\sqrt{p}} + \frac{pr^3}{k} + \frac{r}{k} + 2^{pr+1} \exp\left(-\frac{\rho^2 k}{8r^2 ||W||_\infty} \right) \right) + \rho.$$

Let $k = c\varepsilon^{-4}$, $p = d\varepsilon^{-2}$ and $\rho = e\varepsilon ||W||_\infty$. With appropriate choice of the constants we have $\frac{1}{k^r} \mathbf{E} ||G||_\square \leq ||W||_\square + \varepsilon/2$.

The converse direction, $\frac{1}{k^r} \mathbf{E} ||G||_\square \geq ||W||_\square - \varepsilon/2$ follows from a standard sampling argument, the idea is that we can project each set $S \subset [0,1]$ to a set $\hat{S} \subset [k]$ through the sample, which will fulfill the desired conditions, we leave the details to the reader. Concentration follows by the Azuma-Hoeffding inequality. We conclude that

$$\begin{aligned} \mathbf{P}\left(\left| \frac{1}{k^r} ||G||_\square - ||W||_\square \right| > \varepsilon \right) &\leq \mathbf{P}\left(\left| \frac{1}{k^r} \mathbf{E} ||G||_\square - \frac{1}{k^r} ||G||_\square \right| > \varepsilon/2 \right) \\ &\leq 2 \exp\left(-\frac{\varepsilon^2 k}{16 r^2 ||W||_\infty} \right). \end{aligned}$$

□

Next we state a result on the relationship of a continuous LP and its randomly sampled finite subprogram. Measurability for all of the following functions is assumed.

Lemma 4.10. *Let $c_m \colon [0,1] \to \mathbb{R}$, $U_{i,m} \colon [0,1] \to \mathbb{R}$ for $i = 1, \ldots, s$, $m = 1, \ldots, q$, $u \in \mathbb{R}^{s \times q}$, $\alpha \in \mathbb{R}$. If the optimum of the linear program*

$$\begin{aligned} &\text{maximize} && \int_0^1 \sum_{m=1}^{q} f_m(t) c_m(t) \mathrm{d}t \\ &\text{subject to} && \int_0^1 f_m(t) U_{i,m}(t) \mathrm{d}t \leq u_{i,m} && \text{for } i \in [s] \text{ and } m \in [q] \end{aligned}$$

$$0 \leq f_m(t) \leq 1 \qquad \text{for } t \in [0,1] \text{ and } m \in [q]$$
$$\sum_{m=1}^{q} f_m(t) = 1 \qquad \text{for } t \in [0,1]$$

is less than α, then for any $\varepsilon > 0$ and $k \in \mathbb{N}$ and a uniform random sample $\{X_1, \ldots, X_k\}$ of $[0,1]^k$ the optimum of the sampled linear program

$$\begin{aligned}
\text{maximize} \quad & \sum_{1 \leq n \leq k} \sum_{m=1}^{q} \frac{1}{k} x_{n,m} c_m(X_n) \\
\text{subject to} \quad & \sum_{1 \leq n \leq k} \frac{1}{k} x_{n,m} U_{i,m}(X_n) \leq u_{i,m} - \varepsilon ||U||_\infty \qquad \text{for } i \in [s] \text{ and } m \in [q] \\
& 0 \leq x_{n,m} \leq 1 \qquad \text{for } n \in [k] \text{ and } m \in [q] \\
& \sum_{m=1}^{q} \frac{1}{k} x_{n,m} c_m(X_n) \qquad \text{for } n \in [k]
\end{aligned}$$

is less than $\alpha + \varepsilon ||c||_\infty$ with probability at least $1 - \exp(-\frac{\varepsilon^2 k}{2})$.

Proof. We require a continuous version of the Farkas lemma.

Claim 1. Let $(Af)_{i,m} = \int_0^1 A_{i,m}(t) f_m(t) \mathrm{d}t$ for the bounded measurable functions $A_{i,m}$ on $[0,1]$ for $i \in [s]$ and $m \in [q]$, and let $v \in \mathbb{R}^{sq}$. There is no q fractional partition solution $f = (f_1, \ldots, f_q)$ to $Af \leq v$ if and only if, there exists a non-zero $y \geq 0$ with $||y||_1 = 1$ such that there is no q- fractional partition solution f to $y^T(Af) \leq y^T v$.

For clarity we remark that in the current claim and the following one Af and v are indexed by a pair of parameters, but are regarded as 1-dimensional vectors in the multiplication operation.

Proof. One direction is trivial: if there is a solution f to $Af \leq v$, then it is also a solution to $y^T(Af) \leq y^T v$ for any $y \geq 0$.

We turn to show the opposite direction. Let $C = \{Af | f \colon [0,1] \to [0,1]\}$. The set C is a nonempty convex closed subset of $\mathbb{R}^{sq}$ containing 0. Let $B = \{x | x_{i,m} \leq v_{i,m}\} \subset \mathbb{R}^{sq}$, this set is also a nonempty convex closed set. The absence of a solution to $Af \leq v$ is equivalent to saying that $C \cap B$ is empty. It follows from the Separation Theorem for convex closed sets that there is a $0 \neq y' \in \mathbb{R}^{sq}$ such that $y'^T c < y'^T b$ for every $c \in C$ and $b \in B$. Additionally every coordinate $y'_{i,m}$ has to be non-positive. To see this suppose that $y'_{i_0,m_0} > 0$, we pick a $c \in C$ and $b \in B$, and send b_{i_0,m_0} to minus infinity leaving every other coordinate of the two points fixed (b will still be an element of B), for b_{i_0} small enough the inequality $y'^T c < y'^T b$ will be harmed eventually. We conclude that for any f we have $y'^T(Af) < y'^T v$, hence for $y = \frac{-y'}{||y'||_1}$ the inequality $y^T(Af) \leq y^T v$ has no solution. □

From this lemma the finitary version follows without any difficulties.

Claim 2. Let B a real $sq \times k$ matrix, and let $v \in \mathbb{R}^{sq}$. There is no q-fractional partition $x \in \mathbb{R}^{kq}$ so that $Bx \leq v$ if and only if, there is a non-zero $y \geq 0$ with $||y||_1 = 1$ such that there is no q-fractional partition $x \in \mathbb{R}^{kq}$ so that $y^T Bx \leq y^T v$.

Proof. Let $A_{i,m}(t) = \sum_{n=1}^{k} \frac{B_{(i,m),n}}{k} \mathbb{I}_{[\frac{n-1}{k}, \frac{n}{k})}(t)$ for $i = 1, \ldots, s$. The nonexistence of a q-fractional partition $x \in \mathbb{R}^{kq}$ so that $Bx \leq v$ is equivalent to nonexistence of a q-fractional partition f so that $Af \leq v$. For any nonzero $0 \leq y$, the nonexistence of a q-fractional partition $x \in \mathbb{R}^{kq}$ so that $y^T Bx \leq y^T v$ is equivalent to the nonexistence of a q-fractional partition f so that $y^T(Af) \leq y^T v$. Applying Claim 1 verifies the current claim. □

The assumption of the lemma is by Claim 1 equivalent to the statement that there exists a nonzero $0 \leq y \in \mathbb{R}^{n \times q}$ and $\beta \geq 0$ such that

$$\int_0^1 \sum_{i=1}^{s} \sum_{m=1}^{q} y_{i,m} U_{i,m}(t) f_m(t) \mathrm{d}t - \int_0^1 \beta \sum_{m=1}^{q} c_m(t) f_m(t) \leq \sum_{i=1}^{s} \sum_{m=1}^{q} y_{i,m} u_{i,m} - \beta\alpha$$

has no solution f among q-fractional partitions. This is equivalent to the condition

$$\int_0^1 h(t) \mathrm{d}t > A,$$

where $h(t) = \min_m [\sum_{i=1}^{s} y_{i,m} U_{i,m}(t) - \beta c_m(t)]$, and $A = \sum_{i=1}^{s} \sum_{m=1}^{q} y_{i,m} u_{i,m} - \beta\alpha$. By the Azuma-Hoeffding inequality it follows that with probability at least $1 - \exp(-\frac{k\varepsilon^2}{2})$ we have that

$$\frac{1}{k} \sum_{n=1}^{k} h(X_n) > A - \varepsilon(||h||\infty).$$

This last inequality implies the statement of our lemma by Claim 2. □

We start the proof of the main theorem in this topic.

Proof of Theorem 4.4. We make no specific restrictions on the color set $\mathcal{K}$ and on the set E of layers except for finiteness of the second, therefore it will be convenient to rewrite the layered energies $\mathcal{E}_\phi(W, J)$ into a more universal form as a sum of proper Hamiltonians in order to suppress the role of $\mathcal{K}$ and E. Let

$$\begin{aligned}
\mathcal{E}_\phi(W, J) &= \sum_{e \in E} \sum_{z_1, \ldots, z_r \in [q]} \int_{[0,1]^{s([r],r)}} \prod_{j \in [q]} \phi_{z_j}(x_{\{j\}}) J^e_{z_1, \ldots, z_r}(W^e(x)) \mathrm{d}x \\
&= \sum_{z_1, \ldots, z_r \in [q]} \int_{[0,1]^{s([r],r)}} \prod_{j \in [r]} \phi_{z_j}(x_{\{j\}}) \left[\sum_{e \in E} J^e_{z_1, \ldots, z_r}(W^e(x)) \right] \mathrm{d}x_{s([r],r)}.
\end{aligned}$$

Motivated by this reformulation we introduce for every (W, J) pair a special auxiliary instance of the ground state problem that is defined for a $[q]^r$-layered $([-d, d], r)$-graphon, where $d = |E|\|J\|_\infty$. For any $z \in [q]^r$ let $\tilde{W}^z(x) = \sum_{e \in E} J^e_{z_1,\dots,z_r}(W^e(x))$ for each $x \in [0,1]^{s([r],r)}$, and the interaction matrices $\tilde{J}^z$ of the canonical form. We obtain for any fractional partition ϕ of $[0,1]$ into q parts that $\mathcal{E}_\phi(W, J) = \mathcal{E}_\phi(\tilde{W}, \tilde{J})$, and also $\mathcal{E}_x(\mathbb{G}(k, W), J) = \mathcal{E}_x(\mathbb{G}(k, \tilde{W}), \tilde{J})$ for any fractional partition x, where the two random r-graphs are obtained via the same sample. Therefore, without lost of generality, we are able to restrict our attention to ground state energies of canonical form. In most of what follows we will drop the dependence on J in the energy function when it is clear that we mean the aforementioned $\tilde{J}$, and will employ the notion $\mathcal{E}_\phi(W)$, $\mathcal{E}(W)$, $\mathcal{E}_x(G)$ and $\mathcal{E}(G)$, where W and G are $[q]^r$-layered $[-d, d]$-edge colored objects.

The canonical reformulation allows us to employ first Lemma 4.5 to replace the energy $\hat{\mathcal{E}}(\mathbb{G}(k, W), J)$ by the energy of the averaged sample $\hat{\mathcal{E}}(\mathbb{H}(k, W), J)$ without altering the ground state energy of the sample substantially. From this point we may regard W as a $[q]^r$ layered $([-d, d], r)$-graphon without losing generality thanks to $\mathbb{H}(k, W)$ not relying on the structural additional value of a proper graphon. Subsequently, we apply Lemma 4.6 to change from the integer version of the energy $\hat{\mathcal{E}}(\mathbb{H}(k, W), J)$ to the relaxed one $\mathcal{E}(\mathbb{H}(k, W), J)$.

We are beginning the argument by showing that the ground state energy of the sample can not be substantially smaller than that of the original, formally $\mathcal{E}(\mathbb{H}(k, W)) \geq \mathcal{E}(W) - O(\frac{1}{k})||W||_\infty$ with high probability. Here here and in what follows $\mathbf{E}$ denotes the expectation with respect to the uniform independent random sample $(U_S)_{S \in s([k],r)}$ from $[0,1]$. To see the correctness of the inequality, we consider a fixed fractional partition ϕ of $[0,1]$, and define the random fractional partition of $[k]$ as $y_{n,m} = \phi_m(U_n)$ for every $n \in [k]$ and $m \in [q]$. Then we have that

$$\begin{aligned}
\mathbf{E}\mathcal{E}(\mathbb{H}(k, W)) &\geq \mathbf{E}\mathcal{E}_y(\mathbb{H}(k, W)) \\
&= \mathbf{E}\frac{1}{k^r} \sum_{z \in [q]^r} \sum_{n_1,\dots,n_r=1}^{k} W^z((U_S)_{S \in s(\{n_1,\dots,n_r\},r)}) \prod_{j=1}^{r} y_{n_j,z_j} \\
&\geq \frac{k!}{k^r(k-r)!} \sum_{z \in [q]^r} \int_{[0,1]^{s([r],r)}} W^z((t_S)_{S \in s([r],r)}) \prod_{j=1}^{r} \phi_{z_j}(t_j) \mathrm{d}t - \frac{r^2}{k}\|W\|_\infty \\
&\geq \mathcal{E}_\phi(W) - \frac{r^2}{k}\|W\|_\infty.
\end{aligned}$$

This argument proves the claim in expectation, concentration will be provided by standard martingale arguments. For convenience, we define a martingale by $Y_0 = \mathbf{E}\mathcal{E}(\mathbb{H}(k, W))$ and $Y_t = E[\mathcal{E}(\mathbb{H}(k, W))|(U_S)_{S \in s([t],r)}]$ for $1 \leq t \leq k$. The difference $|Y_i - Y_{i+1}| \leq \frac{2r}{k}||W||_\infty$ is bounded from above for any i, thus by the inequality of Azuma and Hoeffding it follows that

$$\mathbf{P}(\mathcal{E}(\mathbb{H}(k, W)) < \mathcal{E}(W) - \frac{2r^2}{k^{1/4}}||W||_\infty)$$

$$\leq \mathbf{P}(\mathcal{E}(\mathbb{H}(k,W)) < \mathbf{E}\mathcal{E}(\mathbb{H}(k,W)) - \frac{r^2}{k^{1/4}}||W||_\infty)$$
$$= \mathbf{P}(Y_k < Y_0 - \frac{r^2}{k^{1/4}}||W||_\infty) \leq \exp(-\frac{r^2\sqrt{k}}{8}). \tag{4.8}$$

So the lower bound on $\mathcal{E}(\mathbb{H}(k,W))$ is established for some $c > 0$ with $k(\varepsilon) \geq c\varepsilon^{-4}$.

Now we turn to prove that $\mathcal{E}(\mathbb{H}(k,W),J) < \mathcal{E}(W,J) + \varepsilon$ holds also with high probability for $k \geq c\varepsilon^{-4}\log(\varepsilon^{-1}))$ for some c large enough. Our two main tools will be a variant the Cut Decomposition Lemma from [2] (closely related to the Weak Regularity Lemma by Frieze and Kannan [17]), Lemma 4.7, and linear programming duality, in the form of Lemma 4.10. Recall the definition of the cut norm, for $W\colon [0,1]^r \to \mathbb{R}$ it is

$$||W||_\square = \max_{S^1,\ldots,S^r \subset [0,1]} \left| \int\limits_{S^1\times\cdots\times S^r} W(x)\mathrm{d}x \right|,$$

and for an r-array G it is

$$||G||_\square = \max_{S^1,\ldots,S^r \subset V(G)} \left|G(S^1,\ldots,S^r)\right|.$$

Before starting the second part of the technical proof, we present an informal outline. Our task is to certify that there is no evaluation of the variables on the sampled energy problem, which produces an overly large value relative to the ground state energy of the continuous problem. For this reason we build a cover of subsets over the set of fractional partitions of the variables of the finite problem, also build a cover of subsets over the fractional partitions of the original continuous energy problem, and establish an association between the elements of the two in such a way, that with high probability we can definitely say that the optimum on one particular set of the cover of the sampled energy problem does not exceed the optimal value of the original problem on the associated set of the other cover. To be able to do this, first we have to define these two covers, this is done with the aid of the cut decomposition, see Lemma 4.7. We will replace the original continuous problem by an auxiliary one, where the number of variables will be bounded uniformly in terms of our desired error ε. Lemma 4.8 makes it possible for us to replace the sampled energy problem by an auxiliary problem of the same number of variables as for the continuous problem. This latter will have a straightforward relationship to the former approximation of the original problem. We will produce the cover sets of the two problems by localizing the auxiliary problems, association happens through the aforementioned straightforward connection. Finally, we will linearize the local problems, and use the linear programming duality principle from Lemma 4.10 to verify that the local optimal value on the sample does not exceed the local optimal value on the original problem by an infeasible amount, with high probability.

Recall that for a $\phi = (\phi_1,\ldots,\phi_q)$ a q-fractional partition of $[0,1]$

$$\mathcal{E}_\phi(W) = \sum_{z\in[q]^r} \int\limits_{[0,1]^r} \prod_{j\in[r]} \phi_{z_j}(t_j) W^z(t)\mathrm{d}t, \tag{4.9}$$

and for an $x = (x_{1,1}, x_{1,2}, \dots, x_{1,q}, x_{2,1}, \dots, x_{k,q})$ a q-fractional partition of $[k]$

$$\mathcal{E}_x(\mathbb{H}(k,W)) = \sum_{z\in[q]^r} \frac{1}{k^r} \sum_{1\le t_1,\dots,t_r\le k} \prod_{j\in[r]} x_{t_j,z_j} W^z(U_{t_1},\dots,U_{t_r}). \tag{4.10}$$

We are going to establish a term-wise connection with respect to the parameter z in the previous formulas. Therefore we fix z for now and consider the function

$$\mathcal{E}^z_\phi(W^z) = \int_{[0,1]^r} \prod_{j\in[r]} \phi_{z_j}(t_j) W^z(t)\mathrm{d}t. \tag{4.11}$$

Lemma 4.7 delivers for any W^z an integer $s = s(z) \le \frac{4}{\varepsilon^2}$, measurable sets $S_i^j \subset [0,1]$ with $i = 1,\dots,s$, $j = 1,\dots,r$, and real numbers $d_1,\dots,d_s$ such that the conditions of the lemma are satisfied, namely $||W^z - \sum_{i=1}^s d_i \mathbb{I}_{S_i^1\times\cdots\times S_i^r}||_\square \le \varepsilon||W^z||_2$ and $\sum_i |d_i| \le \frac{1}{\varepsilon}||W^z||_2$. We also have that this cut function allows a sufficiently good approximation for $\mathcal{E}_\phi(W^z)$, for any ϕ. Let $D^z = \sum_{i=1}^{s_z} d_i \mathbb{I}_{S_i^1\times\cdots\times S_i^r}$.

$$\begin{aligned}|\mathcal{E}^z_\phi(W^z) - \mathcal{E}^z_\phi(D^z)| &= \left| \int_{[0,1]^r} \prod_{j\in[r]} \phi_{z_j}(t_j)\left[W^z(t) - D^z(t)\right]\mathrm{d}t \right| \\ &\le ||W^z - D^z||_\square \le \varepsilon||W^z||_2.\end{aligned}$$

We apply the cut approximation to W^z for every $z \in [q]^r$ obtaining a uniquely z-colored graphon $D = (D^z)_{z\in[q]^r}$. We define the "push-forward" of this approximation for the sample $\mathbb{H}(k,W^z)$. To do this we only need to define the subsets $[k] \supset \hat{S}_i^j = \{m : U_m \in S_i^j\}$. Let $D'^z = \sum_{i=1}^{s_z} d_i \mathbb{I}_{\hat{S}_i^1\times\cdots\times\hat{S}_i^r}$. First we condition on the event from Lemma 4.8, call this event E_1, that is $E_1 = \bigcap_z \left\{ \left| \frac{1}{k^r}||\mathbb{H}(k,W^z) - D'^z||_\square - ||W^z - D^z||_\square \right| < \varepsilon||W^z||_\infty \right\}$ .On E_1 it follows that for any $x \in [0,1]^k$

$$\begin{aligned}|\mathcal{E}^z_x(\mathbb{H}(k,W^z)) - \mathcal{E}^z_x(D'^z)| &\le \frac{1}{k^r}||\mathbb{H}(k,W^z) - D'^z||_\square \\ &\le ||W^z - D'^z||_\square + \varepsilon||W^z||_\infty.\end{aligned}$$

This implies

$$|\mathcal{E}_\phi(W) - \mathcal{E}_\phi(D)| \le \varepsilon q^r||W||_\infty \quad \text{and} \quad |\mathcal{E}_x(\mathbb{H}(k,W)) - \mathcal{E}_x(\mathbb{H}(k,D))| \le \varepsilon 2q^r||W||_\infty.$$

In what follows we refer to typical set that is involved in the definition of D as S_i^j, where we will not always stress to which particular D^z a it belongs to, let $\mathcal{S} = \{S_{z,i,j} : z \in [q]^r, 1 \le i \le s(z), 1 \le j \le r\}$ denote their set, $\mathcal{S}'$ the corresponding set on the sample. Note that $s' := |\mathcal{S}| \le rq^r \frac{1}{\varepsilon^2}$.Let $\eta > 0$.

$$I(b,\eta) = \{\phi : \forall z \in [q]^r, 1 \le i \le s(z), 1 \le j \le r : |\int\limits_{S_{z,i,j}} \phi_{z_j}(t)\mathrm{d}t - b_{z,i,j}| \le 2\eta\},$$

$$I'(b,\eta) = \{x : \forall z \in [q]^r, 1 \le i \le s(z), 1 \le j \le r : |\frac{1}{k}\sum_{U_n \in S_{z,i,j}} x_{n,z_j} - b_{z,i,j}| \le \eta\}.$$

We will use the grid points $\mathcal{A} = \{(b_{z,i,j})_{z,i,j} : \forall z,i,j : b_{z,i,j} \in [0,1] \cap \eta\mathbb{Z}\}$.

On every set $I(b,\eta)$ we can produce a linear approximation of $\mathcal{E}_\phi(D)$ (linear in the functions ϕ_m) which carries through to a linear approximation of $\mathcal{E}_x(\mathbb{H}(k,D))$ via sampling. The precise description of this is given in the next lemma.

Lemma 4.11 (Local linearization)**.** *Let $\varepsilon, \eta > 0$ be arbitrary. Then for every $b \in \mathcal{A}$ there exist $l_0 \in \mathbb{R}$ and a functions $l_1, l_2, \ldots, l_q \colon [0,1] \to \mathbb{R}$ such that for every $\phi \in I(b,\eta)$ it holds that $|\mathcal{E}_\phi(D) - l_0 - \int_0^1 \sum_{m=1}^q l_m(t)\phi_m(t)\mathrm{d}t| < \frac{\eta^2}{\varepsilon}||W||_\infty$, and for every $y \in I'(b,\eta)$ we have $|\mathcal{E}_y(\mathbb{H}(k,D)) - l_0 - \sum_{n=1}^k \sum_{m=1}^q \frac{1}{k} y_{n,m} l_m(U_i)| < \frac{\eta^2}{\varepsilon}||W||_\infty$. Additionally in the case when $\eta = \nu\varepsilon$ for some fixed real $\nu > 0|$, we have that $l_1, l_2, \ldots, l_q$ are bounded from above by $O(\frac{1}{\varepsilon})||W||_\infty$.*

Proof. First we decompose the energies as sums over $z \in [q]^r$. Recall that

$$\begin{aligned}\mathcal{E}_\phi^z(D^z) &= \sum_{i=1}^{s(z)} d_i \int\limits_{[0,1]^r} \prod_{j=1}^r \phi_{z_j}(x_j)\mathbb{I}_{S_{z,i,1}\times\cdots\times S_{z,i,r}}(x)\mathrm{d}x \\ &= \sum_{i=1}^{s(z)} d_i \int\limits_{[0,1]^r} \prod_{m=1}^q \prod_{\substack{j=1 \\ z_j=m}}^r \phi_m(x_j)\mathbb{I}_{S_{z,i,1}\times\cdots\times S_{z,i,r}}(x)\mathrm{d}x,\end{aligned}$$

and

$$\mathcal{E}_y^z(D'_z) = \sum_{i=1}^{s(z)} d_i \frac{1}{k^r} \prod_{m=1}^q \prod_{\substack{j=1 \\ z_j=m}}^r \sum_{n:U_n \in S_{z,i,j}} y_{n,m}.$$

In the previous formula we denoted the r-tuples of sets involved in the generation of D_z as the first $s(z)$ of the whole list of size $s = \frac{s'}{r}$. Of course this will be a different set of r-tuples for different z's, but it will always be clear what is meant. We will linearize and compare the functions $\mathcal{E}_\phi^z(D^z)$ and $\mathcal{E}_x^z(D'^z)$ termwise. In the end we will sum up the errors and deviations occurred at each term. Let us fix $\phi \in I(\eta,b)$. We are going to produce a linear approximation to each term of the expression $\mathcal{E}_\phi^z(D^z)$.

$$\prod_{m=1}^q \prod_{\substack{j=1 \\ z_j=m}}^r \left[\int_0^1 \phi_m(x_j)\mathbb{I}_{S_{z,i,j}}(x_j)\mathrm{d}x_j - b_{z,i,j} + b_{z,i,j}\right]$$

$$= \prod_{j=1}^{r} b_{z,i,j} + \sum_{m=1}^{q} \sum_{j=1,z_j=m}^{r} \left[\int_0^1 \phi_m(x_j) \mathbb{I}_{S_{z,i,j}}(x_j) \mathrm{d}x_j - b_{z,i,j} \right] \frac{B(z)}{b_{z,i,j}} + \Delta$$

$$= (1-r)B(z) + \sum_{m=1}^{q} \int_0^1 \phi_m(t) \left[\sum_{j=1,z_j=m}^{r} \mathbb{I}_{S_{z,i,j}}(t) \frac{B(z)}{b_{z,i,j}} \right] \mathrm{d}t + \Delta,$$

where $B(z)$ stands for $\prod_{j=1}^{r} b_{z,i,j}$, and $|\Delta| \leq \eta^2 q^r$. Analogously for a fixed element $y \in I'(b,\eta)$ and a term of $\mathcal{E}_y^z(D')$ we have

$$\prod_{m=1}^{q} \prod_{\substack{j=1 \\ z_j=m}}^{r} \left[\frac{1}{k} \sum_{n:U_n \in S_{z,i,j}} y_{n,m} - b_{z,i,j} + b_{z,i,j} \right]$$

$$= (1-r)B(z) + \sum_{m=1}^{q} \sum_{n=1}^{k} \frac{1}{k} y_{n,m} \left[\sum_{j=1,z_j=m}^{r} \mathbb{I}_{S_{z,i,j}}(U_n) \frac{B(z)}{b_{z,i,j}} \right] + \Delta',$$

where again $|\Delta'| \leq \eta^2 q^r$. We multiply these former expressions by the respective coefficient d_i and sum up over i, then we still have in both cases a linear approximation. As overall error we get in both cases at most $\sum_{i=1}^{s} |d_i| \eta^2 q^r \leq ||W||_\infty 2^r \frac{1}{\varepsilon} \eta^2 q^r \leq 2^r = O(\eta^2 \frac{1}{\varepsilon}) ||W||_\infty$. Now we turn to prove the upper bounds on $|l(t)|$. Assume that $||W||_\infty = 1$. Looking at the former formulas we could write out $l(t)$ explicitly, for the upper bound it is enough to note that $\left[\sum_{j=1,z_j=m}^{r} \mathbb{I}_{S_{z,i,j}}(t) \frac{B(z)}{b_{z,i,j}} \right]$ is at most r. So it follows that for any $t \in [0,1]$ it holds that

$$|l(t)| \leq \sum_{i=1}^{s} |d_i| r \leq O\left(\frac{1}{\varepsilon}\right).$$

□

For each $b \in \mathcal{A}$ we apply Lemma 4.11, so that we have for any $\phi \in I(b,\eta)$ and $x \in I'(b,\eta)$ that

$$\left| \mathcal{E}_\phi(W) - l_0 - \sum_{m=1}^{q} \int_0^1 \phi_m(t) l_m(t) \mathrm{d}t \right| = O(\frac{\eta^2}{\varepsilon}) ||W||_\infty,$$

$$\left| \mathcal{E}_x(\mathbb{H}(k,W)) - l_0 - \sum_{n=1}^{k} \frac{1}{k} x_{n,m} l_m(U_n) \right| = O(\frac{\eta^2}{\varepsilon}) ||W||_\infty.$$

Note that of course $l_0, l_1, \ldots, l_q$ depend on b. We introduce the event $E_2(b)$, which stands for the occurrence of the following implication:

If the linear program

$$\text{maximize} \qquad l_0 + \sum_{n=1}^{k} \sum_{m=1}^{q} \frac{1}{k} x_{n,m} l_m(U_n)$$

$$
\begin{aligned}
\text{subject to} \quad & x \in I'(b,\eta) \\
& 0 \le x_{n,m} \le 1 && \text{for } m = 1,\dots,k \text{ and } m = 1,\dots,q \\
& \sum_{m=1}^{q} x_{n,m} = 1 && \text{for } m = 1,\dots,q
\end{aligned}
$$

has optimal value α, then the continuous linear program

$$
\begin{aligned}
\text{maximize} \quad & l_0 + \int_0^1 \sum_{m=1}^{q} l_m(t)\phi_m(t)\mathrm{d}t \\
\text{subject to} \quad & \phi \in I(b,\eta) \\
& 0 \le \phi_m(t) \le 1 && \text{for } t \in [0,1] \text{ and } m = 1,\dots,q \\
& \sum_{m=1}^{q} \phi_m(t) = 1 && \text{for } t \in [0,1]
\end{aligned}
$$

has optimal value at least $\alpha - \eta$.

Let us set $\eta = \nu\varepsilon$. It follows from Lemma 4.10 that $P(E_2(b)) \ge 1 - \exp(-\frac{kO(\varepsilon)^2}{2})$. Denote E_2 the event that for each $b \in \mathcal{A}$ the event $E_2(b)$ occurs. Then $P(E_2) \ge 1 - (O(\frac{1}{\varepsilon}))^{\frac{2q^r}{\varepsilon^2}} \exp(-\frac{kO(\varepsilon)^2}{2})$. Therefore if $k \ge c\frac{1}{\varepsilon^4}\log\frac{1}{\varepsilon}$ with $c > 0$ large enough we have that $P(E_1 \cap E_2) \ge 1 - \varepsilon$. We only need to check that conditioned on E_1 and E_2 our requirements are fulfilled. For this consider an arbitrary $x = (x_{1,1}, x_{1,2}, \dots, x_{1,q}, x_{2,1}, \dots, x_{k,q})$ a q-fractional partition of $[k]$. For some $b \in \mathcal{A}$ we have that $x \in I'(b,\eta)$. By the argument we presented earlier it follows that there exists a $\phi \in I(b,\eta)$ such that on $E_1 \cap E_2$

$$\mathcal{E}_\phi(W) \ge \mathcal{E}_x(\mathbb{H}(k,W)) - O(\varepsilon)||W||_\infty.$$

This is what we wanted to show.

□

Remark 4.12. A simple investigation of the proof also exposes the role of q and r in the size of the required sample: there exists a $c'' > 0$ not dependent on q, r and ε, so that for $k \ge c''\log(q^r/\varepsilon)\frac{q^{4r}}{\varepsilon^4}$ so that the statement of the theorem is valid.

5 Applications of the ground state energy

We derive further testability results using the techniques employed in the proofs of the previous section, and apply Theorem 4.4 to quadratic programming problems.

5.1 Microcanonical version

Next we will state the microcanonical version of Theorem 4.4, that is the continuous generalization of the main result of [15]. To be able to do this, we require the microcanonical analog

of Lemma 4.6, that will be a generalization of Theorem 5.5 from [9] for arbitrary r-graphs (except for the fact that we are not dealing with node weights), and its proof will also follow the lines of the aforementioned theorem. Before stating the lemma, we recall some notation and state yet another auxiliary lemma.

Definition 5.1. *Let for* $\mathbf{a} = (a_1, \dots, a_q) \in \mathrm{Pd}_q$ *(that is,* $a_i \geq 0$ *for each* $i \in [q]$ *and* $\sum_i a_i = 1$*) denote*

$$\Omega_{\mathbf{a}} = \left\{ \phi \text{ fractional partition of } [0,1] : \int_0^1 \phi_i(t)\mathrm{d}t = a_i \text{ for } i \in [q] \right\},$$

$$\omega_{\mathbf{a}} = \left\{ \mathrm{x} \text{ frac. partition of } V(G) : \frac{1}{|V(G)|} \sum_{u \in V(G)} x_{u,i} = a_i \text{ for } i \in [q] \right\},$$

and

$$\hat{\omega}_{\mathbf{a}} = \left\{ \mathrm{x} \text{ int. partition of } V(G) : \left| \frac{\sum_{u \in V(G)} x_{u,i}}{|V(G)|} - a_i \right| \leq \frac{1}{|V(G)|} \text{ for } i \in [q] \right\}.$$

We will call the following expressions microcanonical ground state energies with respect to $\mathbf{a}$ *for* $(\mathcal{K}, r)$*-graphs and graphons and* $C(\mathcal{K})$*-valued* r*-arrays* J*, in the finite case we add the term fractional and integer respectively to the name. Denote*

$$\mathcal{E}_{\mathbf{a}}(W, J) = \max_{\phi \in \Omega_{\mathbf{a}}} \mathcal{E}_{\phi}(W, J), \quad \mathcal{E}_{\mathbf{a}}(G, J) = \max_{\mathrm{x} \in \omega_{\mathbf{a}}} \mathcal{E}_{\mathrm{x}}(G, J), \quad \hat{\mathcal{E}}_{\mathbf{a}}(G, J) = \max_{\mathrm{x} \in \hat{\omega}_{\mathbf{a}}} \mathcal{E}_{\mathrm{x}}(G, J).$$

The layered versions for a finite layer set E *are defined analogously.*

The requirements for an x to be discrete $\mathbf{a}$-fractional partition are rather strict and we are not able to guarantee with high probability that if we sample from an $\mathbf{a}$-fractional partition of $[0,1]$, that we will receive an $\mathbf{a}$-fractional partition on the sample, in fact this will not happen with probability 1. To tackle this problem we need to establish an upper bound on the difference of two microcanonical ground state energies with the same parameters. This was done in the two dimensional case in [9], we generalize that approach.

Lemma 5.2. *Let* E *be a finite layer set,* $\mathcal{K}$ *a compact color set,* $q \geq 1$*, and* r*-arrays* $J = (J^e)_{e \in E}$ *with* $J^e \in C(\mathcal{K})^{q \times \cdots \times q}$*. Then for any* E*-layered* $(\mathcal{K}, r)$*-graphon* $W = (W^e)_{e \in E}$*, and probability distributions* $\mathbf{a}, \mathbf{b} \in \mathrm{Pd}_q$ *we have*

$$|\mathcal{E}_{\mathbf{a}}(W, J) - \mathcal{E}_{\mathbf{b}}(W, J)| \leq r|E|\|J\|_{\infty}\|W\|_{\infty}\|\mathbf{a} - \mathbf{b}\|_1.$$

The analogous statement is true for an arbitrary E*-layered* $\mathcal{K}$*-valued* r*-graph* G*,*

$$|\mathcal{E}_{\mathbf{a}}(G, J) - \mathcal{E}_{\mathbf{b}}(G, J)| \leq r|E|\|J\|_{\infty}\|G\|_{\infty}\|\mathbf{a} - \mathbf{b}\|_1.$$

Proof. We can restrict our attention to the canonical form of the problem with W being a real-valued averaged naive graphon tuple, and $\|W\|_\infty = 1$. We will find for each $\mathbf{a}$-fractional partition ϕ a $\mathbf{b}$-fractional partition ϕ' and vice versa, so that the corresponding energies are as close to each other as in the statement. So let $\phi = (\phi_1, \dots, \phi_q)$ be an arbitrary $\mathbf{a}$-fractional partition, we define ϕ'_i so that the following holds: if $a_i \geq b_i$ then $\phi'_i(t) \leq \phi_i(t)$ for every $t \in [0,1]$, otherwise $\phi'_i(t) \geq \phi_i(t)$ for every $t \in [0,1]$. It is easy to see that such a $\phi' = (\phi'_1, \dots, \phi'_q)$ exists. Next we estimate the energy deviation.

$$\begin{aligned}
|\mathcal{E}_\phi(W) - \mathcal{E}_{\phi'}(W)| &\leq \sum_{z\in[q]^r} \left| \int\limits_{[0,1]^r} \phi_{z_1}(x_1)\dots\phi_{z_r}(x_r) - \phi'_{z_1}(x_1)\dots\phi'_{z_r}(x_r)\mathrm{d}x \right| \\
&\leq \sum_{z\in[q]^r} \sum_{m=1}^{r} \left| \int\limits_{[0,1]^r} (\phi_{z_m}(x_m) - \phi'_{z_m}(x_m)) \prod_{j<m} \phi_{z_j}(x_j) \prod_{j>m} \phi'_{z_j}(x_j)\mathrm{d}x \right| \\
&= \sum_{z\in[q]^r} \sum_{m=1}^{r} \int\limits_{[0,1]} \left|\phi_{z_m}(x_m) - \phi'_{z_m}(x_m)\right| \mathrm{d}x_m \prod_{j<m} a_{z_j} \prod_{j>m} b_{z_j} \\
&= \sum_{m=1}^{r} \sum_{j=1}^{q} \int\limits_{[0,1]} \left|\phi_j(t) - \phi'_j(t)\right| \mathrm{d}t (\sum_{j=1}^{q} a_j)^{m-1} (\sum_{j=1}^{q} b_j)^{r-m-1} \\
&= r||\mathbf{a} - \mathbf{b}||_1.
\end{aligned}$$

The same way we can find for any $\mathbf{b}$-fractional partition ϕ an $\mathbf{a}$-fractional partition ϕ' so that their respective energies differ at most by $r||\mathbf{a} - \mathbf{b}||_1$. This implies the first statement of the lemma. The finite case is proven in a completely analogous fashion.

□

We are ready to show that the difference of the fractional and the integer ground state energies is $o(|V(G)|)$ whenever all parameters are fixed.

Lemma 5.3. *Let $q, r \geq 1$ and let G be a $[q]^r$-layered real-valued r-graph on $[k]$. Let also be $J \in \mathbb{R}^{q\times\dots\times q}$ and $\mathbf{a} \in \mathrm{Pd}_q$. Then*

$$|\mathcal{E}_\mathbf{a}(G, J) - \hat{\mathcal{E}}_\mathbf{a}(G, J)| \leq \frac{1}{k} q^r \|J\|_\infty \|G\|_\infty 5^r q^{r+1}.$$

Proof. We may assume that J is the canonical interaction r-array. The inequality $\mathcal{E}_\mathbf{a}(G, J) \leq \hat{\mathcal{E}}_\mathbf{a}(G, J) + \frac{1}{k}\|J\|_\infty\|G\|_\infty 5^r q^{r+1}$ follows from Lemma 5.3. Indeed, in this direction one has a somewhat stronger bound,

$$\hat{\mathcal{E}}_\mathbf{a}(G, J) \leq \max_{\mathbf{b}:|b_i - a_i|\leq 1/k} \mathcal{E}_\mathbf{b}(G, J) \leq \mathcal{E}_\mathbf{a}(G, J) + r\frac{q^{r+1}}{k}\|J\|_\infty\|G\|_\infty.$$

Now we will show that $\hat{\mathcal{E}}_\mathbf{a}(G, J) \geq \mathcal{E}_\mathbf{a}(G, J) - \frac{1}{k}||G||_\infty 5^r q^{r+1}$. We consider an arbitrary $\mathbf{a}$-fractional partition x. A node i from $[n]$ is called bad in a fractional partition x, if the at

least two elements of $\{x_{i,1}, \dots, x_{i,q}\}$ are positive. We will reduce the number of fractional entries of the bad nodes of x step by step until we have at most q of them, and keep track of the cost of each conversion, at the end we round the corresponding fractional entries of the remaining bad nodes some certain way.

We will describe a step of the reduction of fractional entries. For now assume that we have at least $q+1$ bad nodes and select somehow a set S of cardinality $q+1$. To each element of S corresponds a q-tuple of entries and each of these q-tuples has at least two non-$\{0,1\}$ elements.

We will now reduce the number of fractional entries corresponding to S while not disrupting any entries outside of S. To do this we fix for each $i \in [q]$ the sums $\sum_{v \in S} x_{v,i}$ and for each $v \in S$ the sums $\sum_{i=1}^{q} x_{v,i}$ (these are naturally fixed), in total $2q+1$ linear equalities. We have at least $2q+2$ fractional entries corresponding to S, therefore there exists a subspace of solutions of dimension at least 1 for the $2q+1$ linear equalities. That is, there is a family of fractional partitions parametrized by $-t_1 \le t \le t_2$ for some $t_1, t_2 > 0$:$\mathrm{x}^t_{i,j} = x_{i,j} + t\beta_{i,j}$, where $\beta_{i,j} = 0$ if $i \notin S$ or $x_{i,j} \in \{0,1\}$, and $\beta_{i,j} \neq 0$ else, that obey our $2q+1$ fixed equalities. The boundaries $-t_1$ and t_2 are present, because eventually an entry corresponding to S would exceed 1 or would be less than 0 with t going to plus, respectively minus infinity. Therefore at these boundary points we still have an **a**-fractional partition that satisfies our selected equalities, but the number of fractional entries decreases by at least one. We will formalize how the energy behaves when applying this procedure.

$$\mathcal{E}_{\mathrm{x}^t}(G, J) = \mathcal{E}_{\mathrm{x}}(G, J) + c_1 t + \cdots + c_r t^r,$$

where for $l \in [r]$

$$c_l = \frac{1}{k^r} \sum_{z \in [q]^r} \sum_{\substack{u_1, \dots, u_l \in S \\ u_{l+1}, \dots, u_r \in V \setminus S \\ \pi}} \beta_{u_1, z_{\pi(1)}} \dots \beta_{u_l, z_{\pi(l)}} x_{u_{l+1}, z_{\pi(l+1)}} \dots x_{u_r, z_{\pi(r)}} G^z(u_{\pi(1)}, \dots, u_{\pi(r)}),$$

where the second sum runs over permutations π of $[k]$ that preserves the ordering of the elements of $\{1, \dots, l\}$ and $\{l+1, \dots, r\}$ respectively. We deform the entries corresponding to S through t in the direction so that $c_1 t \ge 0$ until we have eliminated at least one fractional entry. Note, that as x^t is a fractional partition, therefore $0 \le x_{i,j} + t\beta_{i,j} \le 1$, which implies that for $t\beta_{i,j} \le 0$ we have $|t\beta_{i,j}| \le x_{i,j}$. On the other hand, $\sum_j t\beta_{i,j} = 0$ for any t and i. Therefore $\sum_j |t\beta_{i,j}| = 2\sum_j |t\beta_{i,j}| \mathbb{I}_{\{t\beta_{i,j} \le 0\}} \le 2\sum_j x_{i,j} = 2$ for any $i \in [k]$. This simple fact enables us to upper bound the absolute value of the terms $c_l t^l$.

$$\begin{aligned}
|c_l t^l| &\le \frac{(k-q-1)^{r-l}}{k^r} \|G\|_\infty \sum_{z \in [q]^r} \sum_{\substack{u_1, \dots, u_l \in S \\ \pi}} |t\beta_{u_1, z_{\pi(1)}}| \dots |t\beta_{u_l, z_{\pi(l)}}| \\
&= \frac{(k-q-1)^{r-l}}{k^r} \|G\|_\infty \binom{r}{l} q^{r-l} \sum_{z \in [q]^l} \sum_{u_1, \dots, u_l \in S} |t\beta_{u_1, z_1}| \dots |t\beta_{u_l, z_l}| \\
&\le \frac{1}{k^l} \|G\|_\infty \binom{r}{l} q^{r-l} \left(\sum_{u \in S, j \in [q]} |t\beta_{u,j}| \right)^l \le \frac{1}{k^l} \|G\|_\infty \binom{r}{l} q^{r-l} (2q+2)^l.
\end{aligned}$$

It follows that in each step of elimination of a fractional entry of x we have to pay a price of at most

$$\sum_{l=2}^{r} |c_l t^l| \leq \frac{1}{k^2} \|G\|_\infty (3q+2)^r.$$

There are in total kq entries in x, as in each step the number of fractional entries is reduced by at least 1, we can upper bound the number of required steps by $k(q-1)$, and conclude that we to pay at most a price of $\frac{1}{k}\|G\|_\infty (q-1)(3q+2)^r$ to construct from x a fractional partition x$'$ with at most q fractional entries In the second stage we proceed as follows. Let $B = \{u_1, \ldots, u_r\}$ be the set of the remaining bad nodes of x$'$, with $r \leq q$. For $u_i \in B$ we set $x''_{u_i,j} = \mathbb{I}_{\{i=j\}}$, for the rest of the nodes we set x$'' =$ x$'$, obtaining an integer **a**-partition of $[k]$. Finally, we estimate the cost of this operation.

$$\mathcal{E}_{\mathrm{x}''}(G,J) \geq \mathcal{E}_{\mathrm{x}'}(G,J) - \frac{1}{k^r}\|G\|_\infty |B| k^{r-1} q^r.$$

The original **a**-fractional partition was arbitrary, therefore it follows that

$$\mathcal{E}_{\mathbf{a}}(G,J) - \hat{\mathcal{E}}_{\mathbf{a}}(G,J) \leq \frac{1}{k}\|G\|_\infty 5^r q^{r+1}.$$

□

We are ready state the generalization of Theorem 4.4 adapted to the microcanonical setting.

Corollary 5.4. *Let E be a finite layer set, $\mathcal{K}$ a compact color set, $q \geq 1$, $\mathbf{a} \in \mathrm{Pd}_q$ r-arrays $J = (J^e)_{e\in E}$ with $J^e \in C(\mathcal{K})^{q\times\cdots\times q}$, and $\varepsilon > 0$. Then we have that for any E-layered $(\mathcal{K}, r)$-graphon $W = (W^e)_{e\in E}$ and for $k \in O(\varepsilon^{-4}\log(\varepsilon^{-1}))$ that*

$$\mathbf{P}\left(|\mathcal{E}_{\mathbf{a}}(W,J) - \hat{\mathcal{E}}_{\mathbf{a}}(\mathbb{G}(k,W),J)| > \varepsilon |E| \|J\|_\infty \|W\|_\infty\right) < \varepsilon.$$

Proof. We start with pointing out that we are allowed to replace the quantity $\hat{\mathcal{E}}_{\mathbf{a}}(\mathbb{G}(k,W),J)$ by $\mathcal{E}_{\mathbf{a}}(\mathbb{G}(k,W),J)$ in the statement of the corollary by Lemma 5.3 and only introduce an initial error at most $\frac{1}{k}q^r\|J\|_\infty\|G\|_\infty 5^r q^{r+1}$. We can also require $W = (W^z)_z$ to be $[q]^r$ layered and real-valued with $\|W\|_\infty \leq 1$, furthermore J to have canonical form (see the proof of Theorem 4.4 for details).

The lower bound on $\mathcal{E}_{\mathbf{a}}(\mathbb{G}(k,W),J)$ is the result of standard sampling argument combined with Lemma 5.2. Let us consider a fixed **a**-partition ϕ of $[0,1]$, and define the random fractional partition of $[k]$ as $y_{n,m} = \phi_m(U_n)$ for every $n \in [k]$ and $m \in [q]$. Now y is not necessarily an **a**-fractional partition, but it cannot be very far from being one. For $m \in [q]$ it holds that

$$\mathbf{P}\left(\left|\frac{\sum_{n=1}^k y_{n,m}}{k} - a_m\right| \geq \varepsilon\right) \leq 2\exp(-\varepsilon^2 k/2),$$

therefore for $k \in O(\varepsilon^{-4}\log(\varepsilon^{-1}))$ the sizes of the partition classes obey $|\frac{1}{k}\sum_{n=1}^{k} y_{n,m} - a_m| < \varepsilon$ for every $m \in [q]$ with probability at least $1-\varepsilon$.
We apply Lemma 5.2 to arrive at

$$\begin{aligned}
\mathbf{E}\mathcal{E}(\mathbb{G}(k,W)) &\geq \mathbf{E}\mathcal{E}_y(\mathbb{G}(k,W)) - (q+1)\varepsilon \\
&= \mathbf{E}\frac{1}{k^r}\sum_{z\in[q]^r}\sum_{n_1,\ldots,n_r=1}^{k} W(U_{n_1},\ldots,U_{n_r})\prod_{j=1}^{r} y_{n_j,z_j} - (q+1)\varepsilon \\
&\geq \frac{k!}{k^r(k-r)!}\sum_{z\in[q]^r}\int_{[0,1]^r} W(t_1,\ldots,t_r)\prod_{j=1}^{r}\phi_{z_j}(t_j)\mathrm{d}t - \frac{r^2}{k} - (q+1)\varepsilon \\
&\geq \mathcal{E}_\phi(W) - \frac{r^2}{k} - (q+1)\varepsilon.
\end{aligned}$$

The concentration of the random variable $\mathcal{E}(\mathbb{G}(k,W))$ can be obtained through martingale arguments identical to the technique used in the proof of the lower bound in Theorem 4.4.

For the upper bound on $\mathcal{E}_{\mathbf{a}}(\mathbb{G}(k,W))$ we are going to use the cut decomposition and local linearization, the approach to approximate the energy of $\mathcal{E}_\phi(W)$ and $\mathcal{E}_x(\mathbb{G}(k,W))$ for certain partitions ϕ, respectively x is completely identical to the proof of Theorem 4.4, therefore we borrow all the notation from there, and we do not refer to again in what follows.

Now we consider a $b \in \mathcal{A}$ and define the event $E_3(b)$ that is occurrence the following implication.

If the linear program

$$\begin{aligned}
&\text{maximize} && l_0 + \sum_{n=1}^{k}\sum_{m=1}^{q}\frac{1}{k}x_{n,m}l_m(U_n) && \\
&\text{subject to} && x \in I'(b,\eta)\cap\omega_{\mathbf{a}} && \\
& && 0 \leq x_{n,m} \leq 1 && \text{for } n=1,\ldots,k \text{ and } m=1,\ldots,q \\
& && \sum_{m=1}^{q} x_{n,m} = 1 && \text{for } n=1,\ldots,k
\end{aligned}$$

has optimal value α, then the continuous linear program

$$\begin{aligned}
&\text{maximize} && l_0 + \int_0^1 \sum_{m=1}^{q} l_m(t)\phi_m(t)\mathrm{d}t && \\
&\text{subject to} && \phi \in I(b,\eta)\bigcap\left(\bigcup_{\mathbf{c}:|a_i-c_i|\leq\eta}\omega_{\mathbf{c}}\right) && \\
& && 0 \leq \phi_m(t) \leq 1 && \text{for } t\in[0,1] \text{ and } m=1,\ldots,q \\
& && \sum_{m=1}^{q}\phi_m(t) = 1 && \text{for } t\in[0,1]
\end{aligned}$$

has optimal value at least $\alpha - \eta$.

It follows by applying Lemma 4.10 with setting $\eta = \nu\varepsilon$ that $E_3(b)$ has probability at least $1 - \exp(-\frac{\varepsilon^2 k}{2})$. When conditioning on E_2 and $E_3 = \cap_{b\in\mathcal{A}} E_3(b)$ we conclude that

$$\mathcal{E}_{\mathbf{a}}(\mathbb{G}(k, W)) \leq \max_{\mathbf{c}:|a_i - c_i|\leq\nu\varepsilon} \mathcal{E}_{\mathbf{c}}(W) + O(\varepsilon) \leq \mathcal{E}_{\mathbf{a}}(W) + O(\varepsilon).$$

Also, like in Theorem 4.4, the probability of the required events to happen simultaneously is at least $1 - O(\varepsilon)$ for $k \geq c'\varepsilon^{-4}\log(\varepsilon^{-1})$ for a $c' > 0$ large enough.

□

Remark 5.5. The proof of Corollary 5.4 also reveals that there exists a positive constant c'' not depending neither on ε, q nor r, so that the statement of Corollary 5.4 remains valid for every sample size $k \geq c'' \log(q^r/\varepsilon)q^{4r}\varepsilon^{-4}$.

5.2 Quadratic assignment and maximum acyclic subgraph problem

The two optimization problems that are the subject of this subsection, the quadratic assignment problem (QAP) and maximum acyclic subgraph problem (AC), are known to be NP-hard, similarly to MAX-rCSP that was investigated above. The first polynomial time approximation schemes were designed for the QAP by Arora, Frieze and Kaplan [3]. Dealing with the QAP means informally that one aims to minimize the transportation cost of his enterprise that has n production locations and n types of production facilities. This is to be achieved by an optimal assignment of the facilities to the locations with respect to the distances (dependent on the location) and traffic (dependent on the type of the production). In formal, terms this means that we are given two real quadratic matrices of the same size, G and $J \in \mathbb{R}^{n\times n}$, and the objective is to calculate

$$\mathrm{Q}(G, J) = \frac{1}{n^2} \max_{\rho} \sum_{i,j=1}^{n} J_{i,j} G_{\rho(i),\rho(j)},$$

where ρ runs over all permutations of $[n]$. We speak of metric QAP, if the entries of J obey the triangle inequality, and d-dimensional geometric QAP if the rows and columns of J can be embedded in a d-dimensional L^p metric space so that distances of the images are equal to the entries of J.

The continuous analog of the problem is the following: given the measurable functions $W, J\colon [0,1]^2 \to \mathbb{R}$, we are interested in obtaining

$$\hat{\mathrm{Q}}_\rho(W, J) = \int\limits_{[0,1]^2} J(x, y)W(\rho(x), \rho(y))\mathrm{d}x\mathrm{d}y, \qquad \hat{\mathrm{Q}}(W, J) = \max_{\rho} \hat{\mathrm{Q}}_\rho(W, J),$$

where ρ in the previous formula runs over all measure preserving permutations of $[0, 1]$. In even greater generality we introduce the QAP with respect to fractional permutations of

$[0,1]$. A fractional permutation μ is a probability kernel, that is $\mu\colon [0,1]\times\mathcal{L}([0,1])\to[0,1]$ so that (i) for any $A\in\mathcal{L}([0,1])$ the function $\mu(.,A)$ is measurable, (ii) for any $x\in[0,1]$ the function $\mu(x,.)$ is a probability measure on $\mathcal{L}([0,1])$, and (iii) for any $A\in\mathcal{L}([0,1])$ $\int_0^1 \mathrm{d}\mu(x,A)=\lambda(A)$. Then

$$\mathrm{Q}_\mu(W,J)=\int\limits_{[0,1]^2}\int\limits_{[0,1]^2} J(i,j)W(x,y)\mathrm{d}\mu(i,x)\mathrm{d}\mu(j,y)\mathrm{d}i\mathrm{d}j,$$

and

$$\mathrm{Q}(W,J)=\max_\mu \mathrm{Q}_\mu(W,J),$$

where the maximum runs over all fractional permutations. For each measure preserving permutation ρ one can consider the fractional permutation μ with the probability measure $\mu(i,.)$ is defined as the atomic measure $\delta_{\rho(i)}$, for this choice of μ we have $\mathrm{Q}_\rho(W,J)=\mathrm{Q}_\mu(W,J)$.

The r-dimensional generalization of the problem for J and $W\colon[0,1]^r\to\mathbb{R}$ is

$$\mathrm{Q}(W,J)=\max_\mu\int\limits_{[0,1]^r}\int\limits_{[0,1]^r} J(j_1,\dots,j_r)W(x_1,\dots,x_r)\mathrm{d}\mu(j_1,x_1)\dots\mathrm{d}\mu(j_r,x_r)\mathrm{d}j_1\dots\mathrm{d}j_r,$$

where the maximum runs over all fractional permutations μ of $[0,1]$. The definition of the finitary case in r dimensions is analogous.

A similar problem to QAP is the maximum acyclic subgraph problem. Here we are given a weighted directed graph G with vertex set of cardinality n, and our aim is to determine the maximum of the total value of edge weights of a subgraph of G that contains no directed cycle. We can formalize this as follows. Let $G\in\mathbb{R}^{n\times n}$ be the input data, then the maximum acyclic subgraph density is

$$\mathrm{AC}(G)=\frac{1}{n^2}\max_\rho\sum_{i,j=1}^n G_{i,j}\mathbb{I}_{\{\rho(j)\geq\rho(i)\}},$$

where ρ runs over all permutations of $[n]$.

This is a special case of the QAP when J is the upper triangular $n\times n$ matrix with 0's on the diagonal and all nonzero entries being equal to 1. However in general AC cannot be reformulated as metric QAP. The continuous version of the problem $\hat{\mathrm{AC}}(W)$ for a function $W\colon[0,1]^2\to\mathbb{R}$ is defined analogous to the QAP, as well as the relaxation $\mathrm{AC}(W)$, where the maximum runs over probability kernels.

Both QAP and AC problems resemble the ground state energy problems that were investigated in previous parts of this paper. In fact, if the number of clusters of the distance matrix J in the QAP would be upper bounded independent of n, QAP would exactly turn into ground state energy with the number of states q equal to the number of clusters. By number of clusters we mean here the smallest number m such that there exists an $m\times m$ matrix J' so that J is $\lceil\frac{m}{n}\rceil$-fold equitable blow-up of J'. To establish an approximation to the solution of the QAP we will only need this condition approximately, and this will be shown in what follows.

Definition 5.6. *We call a measurable function $J\colon [0,1]^r \to \mathbb{R}$ ν-clustered for a non-increasing function $\nu\colon \mathbb{R}^+ \to \mathbb{R}^+$, if for any $\varepsilon > 0$ there exists another measurable function $J'\colon [0,1]^r \to \mathbb{R}$ that is a step function with $\nu(\varepsilon)$ steps and $||J - J'||_1 < \varepsilon||J||_\infty$.*

Note, that by the Weak Regularity Lemma ([17]) any J can be well approximated by a step function with $\nu(\varepsilon) = 2^{\frac{1}{\varepsilon^2}}$ steps in the cut norm. To see why it is likely that this cut norm approximation will not be sufficient for our purposes, consider a $J\colon [0,1]^r \to \mathbb{R}$. Suppose that we have an approximation in the cut norm of J in hand, namely J', and define the kernel $\mu_0(i,.) = \delta_i$ and the r-graphon $W_0 = J - J'$. In this case $|\mathrm{Q}_{\mu_0}(W,J) - \mathrm{Q}_{\mu_0}(W,J')| = ||J - J'||_2^2$. This 2-norm is not granted to be small in ε by any means.

In some special cases, for example if J is a d-dimensional metric or the array corresponding to the AC, we are able to require a smaller upper bounds on the number of steps required for the 1-norm approximation of J, than exponential in $\frac{1}{\varepsilon}$. By the aid of this fact we can achieve good approximation of the optimal value of the QAP via sampling. Next we state an application of Theorem 4.4 to the clustered QAP.

Lemma 5.7. *Let $J\colon [0,1]^r \to \mathbb{R}$ be a ν-clustered measurable function. Then there exists an absolute constant $c > 0$ so that for every $\varepsilon > 0$, every r-array W, and $k \geq c\log(\nu(\varepsilon)\varepsilon)(\frac{\nu(\varepsilon)^r}{\varepsilon})^4$ so that*

$$P(|\mathrm{Q}(W,J) - \mathrm{Q}(\mathbb{G}(k,W), \mathbb{G}'(k,J))| \geq \varepsilon||W||_\infty||J||_\infty) \leq \varepsilon,$$

where $\mathbb{G}(k,W)$ and $\mathbb{G}'(k,J)$ are generated by independent sampling.

Proof. Without loss of generality we may assume that $||J||_\infty \leq 1$. First we show that under the cluster condition we can introduce a microcanonical ground state energy problem whose optimum is close to $\mathrm{Q}(W,J)$, and the same holds for the sampled problem. Let $\varepsilon > 0$ be arbitrary and J' be an approximating step function with $q = \nu(\varepsilon)$ steps. We may assume that $||J'||_\infty \leq 1$ We set $\mathbf{a} = (a_1, \ldots, a_q)$ to be the vector of the sizes of the steps of J', and turn J' into a real r-array on $[q]$ the natural way by associating to each class of the partition of $[0,1]$ of J' a vertex of the vertex set $[q]$ (indexes should be chosen with regard to $\mathbf{a}$), and set the entries of the r-array corresponding to the value of the respective step of J'. We will call the resulting r-array J''. From the definitions it follows that

$$\mathrm{Q}(W,J') = \mathcal{E}_{\mathbf{a}}(W,J'')$$

for every r-graphon W. On the other hand we have that

$$\begin{aligned}
|\mathrm{Q}(W,J) - \mathrm{Q}(W,J')| &\leq \max_\mu |\mathrm{Q}_\mu(W,J) - \mathrm{Q}_\mu(W,J')| \\
&= \max_\mu |\int\limits_{[0,1]^r} (J - J')(j) \int\limits_{[0,1]^r} W(x)\mathrm{d}\mu(j_1,x_1)\ldots\mathrm{d}\mu(j_r,x_r)\mathrm{d}j_1\ldots\mathrm{d}j_r| \\
&\leq \max_\mu \int\limits_{[0,1]^r} |(J - J')(j)|||W||_\infty \mathrm{d}j_1\ldots\mathrm{d}j_r \\
&= ||J - J'||_1||W||_\infty \leq \varepsilon||W||_\infty.
\end{aligned}$$

Now we turn to the sampled version of the optimization problem. First we gain control over the difference of the QAPs, when we replace J by J'. $\mathbb{G}(k,W)$ is induced by the sample $U_1,\ldots,U_k$, $\mathbb{G}'(k,J)$ by the independent $Y_1,\ldots,Y_k$.

$$\begin{aligned}
&|\mathrm{Q}(\mathbb{G}(k,W),\mathbb{G}(k,J)) - \mathrm{Q}(\mathbb{G}(k,W),\mathbb{G}(k,J'))| \\
&\leq \max_{\rho} |\mathrm{Q}_\rho(\mathbb{G}(k,W),\mathbb{G}(k,J)) - \mathrm{Q}_\rho(\mathbb{G}(k,W),\mathbb{G}(k,J'))| \\
&= \max_{\rho} \frac{1}{k^r} |\sum_{i_1,\ldots,i_r=1}^{k} (J-J')(Y_{i_1},\ldots,Y_{i_r})|||W||_\infty. \qquad (5.1)
\end{aligned}$$

We analyze the random sum on the right hand side of (5.1) by first upper bounding its expectation.

$$\begin{aligned}
&\frac{1}{k^r}\mathbf{E}_Y \left| \sum_{i_1,\ldots,i_r=1}^{k} (J-J')(Y_{i_1},\ldots,Y_{i_r}) \right| \\
&\leq \frac{r^2}{k}||J||_\infty + ||J'||_\infty + \mathbf{E}_Y|(J-J')(Y_1,\ldots,Y_r)| = \frac{r^2}{k} + \varepsilon \leq 2\varepsilon
\end{aligned}$$

By the Azuma Hoeffding inequality the sum is also sufficiently small in probability.

$$P\left(\frac{1}{k^r} \left| \sum_{i_1,\ldots,i_r=1}^{k} (J-J')(Y_{i_1},\ldots,Y_{i_r}) \right| \geq 4\varepsilon \right) \leq \exp(-\varepsilon^2 k/2) \leq \varepsilon.$$

We obtain that

$$|\mathrm{Q}(\mathbb{G}(k,W),\mathbb{G}'(k,J)) - \mathrm{Q}(\mathbb{G}(k,W),\mathbb{G}'(k,J'))| \leq 4\varepsilon||W||_\infty$$

with probability at least $1-\varepsilon$, if k is such as in the statement of the lemma. Set $\mathbf{b} = (b_1,\ldots,b_q)$ to the probability distribution for that $b_i = \frac{1}{k}\sum_{j=1}^{k} \mathbb{I}_{\{Y_j \in S_i\}}$, where S_i is the i'th step of J' with $\lambda(S_i) = a_i$. Then we have that

$$\mathrm{Q}(\mathbb{G}(k,W),\mathbb{G}(k,J)) = \hat{\mathcal{E}}_{\mathbf{b}}(\mathbb{G}(k,W),J'').$$

It follows from the Azuma-Hoeffding inequality that for each $i \in [q]$ $P(|a_i - b_i| > \varepsilon) \leq 2\exp(-2\varepsilon^2 k)$ so we have that $|\mathbf{a}-\mathbf{b}|_1 < \varepsilon$ with probability at least $1-\varepsilon$. We can conclude that with probability at least $1-2\varepsilon$

$$\begin{aligned}
&|\mathrm{Q}(W,J) - \mathrm{Q}(\mathbb{G}(k,W),\mathbb{G}(k,J))| \leq |\mathrm{Q}(W,J) - \mathrm{Q}(W,J')| \\
&+ |\mathcal{E}_{\mathbf{a}}(W,J'') - \hat{\mathcal{E}}_{\mathbf{b}}(\mathbb{G}(k,W),J'')| + |\mathrm{Q}(\mathbb{G}(k,W),\mathbb{G}(k,J)) - \mathrm{Q}(\mathbb{G}(k,W),\mathbb{G}(k,J'))| \\
&\leq 6\varepsilon||W||_\infty + |\mathcal{E}_{\mathbf{a}}(W,J'') - \hat{\mathcal{E}}_{\mathbf{a}}(\mathbb{G}(k,W),J'')|.
\end{aligned}$$

By application of Corollary 5.4 the claim of the lemma is verified.

□

Next we present the application of Lemma 5.7 to two special cases of QAP.

Corollary 5.8. *The d-dimensional geometric QAP and the maximum acyclic subgraph problem are efficiently testable.*

Proof. We start with the continuous version of the d-dimensional geometric QAP given by the measurable function $J\colon [0,1]^2 \to \mathbb{R}^+$. Note, that d refers to the dimension corresponding to the embedding of the indices of J into an L^p a metric space, not the actual dimension of J. We are free to assume that $0 \leq J \leq 1$ simply by rescaling. There exists an embedding $\rho\colon [0,1] \to [0,1]^d$, so that $J(i,j) = ||\rho(i), \rho(j)||_p$ for every $(i,j) \in [0,1]^2$. Fix $\varepsilon > 0$ and consider the partition $\mathcal{P}' = (T_1, \ldots, T_\beta) = ([0, \frac{1}{\beta}), [\frac{1}{\beta}, \frac{2}{\beta}), \ldots, [\frac{\beta-1}{\beta}, 1])$ of the unit interval into $\beta = \lceil \frac{2\sqrt[p]{d}}{\varepsilon} \rceil$ classes. Define the partition $\mathcal{P} = (P_1, \ldots, P_q)$ of $[0,1]$ consisting of the classes $\rho^{-1}(T_{i_1} \times \cdots \times T_{i_d})$ for each $(i_1, \ldots, i_d) \in [\beta]^r$, where $|\mathcal{P}| = q = \beta^d = \frac{2^d d^{d/p}}{\varepsilon^d}$. We construct the approximating step function J' by averaging J on the steps determined by the partition classes of $\mathcal{P}$. It remains to show that this indeed is a sufficient approximation in the L^1-norm.

$$||J - J'||_1 = \int\limits_{[0,1]^2} |J(x) - J'(x)| \mathrm{d}x = \sum_{i,j=1}^{q} \int\limits_{P_i \times P_j} |J(x) - J'(x)| \mathrm{d}x \leq \sum_{i,j}^{q} \frac{1}{q^2} \varepsilon = \varepsilon.$$

By Lemma 5.7 and Corollary 5.4 it follows that the continuous d-dimensional metric QAP is $O(\log(\frac{1}{\varepsilon})\frac{1}{\varepsilon^{rd+4}})$-testable, and so is the discrete version of it.

Next we show that the AC is also efficiently testable given by the upper triangular matrix J whose entries above the diagonal are 1. Fix $\varepsilon > 0$ and consider the partition $\mathcal{P} = (P_1, \ldots, P_q)$ with $q = \frac{2}{\varepsilon}$, and set J' to 0 on every step $P_i \times P_j$ whenever $i \geq j$, and to 1 otherwise. This function is indeed approximating J in the L^1-norm.

$$||J - J'||_1 = \int\limits_{[0,1]^2} |J(x) - J'(x)| \mathrm{d}x = \sum_{i=1}^{q} \int\limits_{P_i \times P_i} |J(x) - J'(x)| \mathrm{d}x \leq \sum_{i=1}^{q} \frac{1}{2q^2} = \varepsilon.$$

Again, by Lemma 5.7 and Corollary 5.4 it follows that the AC is $O(\log(\frac{1}{\varepsilon})\frac{1}{\varepsilon^6})$-testable.

□

6 Further Research

Our framework based on exchangeability principles allows us to extend the notion of a limit to the case of unbounded hypergraphs and efficient testability of ground state energies in this setting. The notion of exchangeability is crucial here. The notion of efficient testability in an unbounded case could be of independent interest, perhaps the results on ground state energy carry through for the setting when the r-graphons (induced by r-graphs) are in an L^p space for some $p \geq 1$.

Another problem is to characterize more precisely the class of problems which are efficiently parameter testable as opposed to the hard ones. Improving the bounds in $\frac{1}{\varepsilon}$ for the efficiently testable problems is a another important problem.

Acknowledgement

We thank Jennifer Chayes, Christian Borgs and Tim Austin for a number of interesting and stimulating discussions and the relevant new ideas in the early stages of this research.

References

[1] D. J. Aldous: Representations for partially exchangeable arrays of random variables, J. Multivariate Anal.,11(4) (1981), 581-598.

[2] N. Alon, W. F. de la Vega, R. Kannan, and M. Karpinski: Random Sampling and Approximation of MAX-CSPs, Journal of Computer and System Sciences, 67 (2) (2003), 212-243.

[3] S. Arora, A. Frieze, and H. Kaplan: A new rounding procedure for the assignment problem with applications to dense graph arrangement problems, Mathematical Programming 92:1 (2002), 1-36.

[4] S. Arora, D. Karger, and M. Karpinski: Polynomial Time Approximation Schemes for Dense Instances of NP-Hard Problems, Journal of Computer and System Sciences 58 (1999), 193-210.

[5] A. Aroskar: Limits, Regularity and Removal for Relational and Weighted Structures (2012). Dissertations. Paper 144.

[6] V. Arvind, J. Köbler, S. Kuhnert, and Y. Vasudev: Approximate graph isomorphism, Proc. of ITCS, 2012, 100-111.

[7] T. Austin: On exchangeable random variables and the statistics of large graphs and hypergraphs, Probability Surveys Vol. 5 (2008), 80-145.

[8] C. Borgs, J. Chayes, L. Lovász, V.T. Sós, and K. Vesztergombi: Convergent Sequences of Dense Graphs I: Subgraph Frequencies, Metric Properties and Testing, Advances in Math. 219 (2008), 1801-1851.

[9] C. Borgs, J. Chayes, L. Lovász, V.T. Sós, and K. Vesztergombi: Convergent sequences of dense graphs II: Multiway cuts and statistical physics, Annals of Math. 176 (2012), 151-219.

[10] C. Borgs, J. Chayes, and L. Lovász: Moments of two-variable functions and the uniqueness of graph limits, GAFA Vol. 19 (2010), 1597-1619 .

[11] B. de Finetti: Fuzione caratteristica di un fenomeno aleatorio, Mem. R. Acc. Lincei, 4(6) (1930), 86-133.

[12] P. Diaconis and S. Janson: Graph Limits and Exchangeable Random Graphs, Rendiconti di Matematica, Serie VII 28, (2008), 33-61.

[13] G. Elek and B. Szegedy: A measure-theoretic approach to the theory of dense hypergraphs Advances in Mathematics 231 (2012), 1731-1772.

[14] O. Goldreich, S. Goldwasser, and D. Ron: Property Testing and Its Connection to Learning and Approximation, Proc. of FOCS (1996), 339-348.

[15] W. Fernandez de la Vega, R. Kannan, and M. Karpinski: Approximation of Global MAX-CSP Problems. Technical Report TR06-124, ECCC (2006).

[16] W. Fernandez de la Vega, R. Kannan, and M. Karpinski, S. Vempala: Tensor Decomposition and Approximation Schemes for Constraint Satisfaction Problems, Proc. of STOC, (2005) 747–754.

[17] A. Frieze and R. Kannan: Quick approximation to matrices and applications, Combinatorica, Vol. 19 (2) (1999), 175-220.

[18] E. Hewitt and L. J. Savage: Symmetric measures on Cartesian products, Trans. Amer. Math. Soc., 80 (1955), 470-501.

[19] D. N. Hoover: Relations on probability spaces and arrays of random variables, preprint (1979).

[20] S. Janson: Poset limits and exchangeable random posets, Combinatorica 31 (2011), 529-563.

[21] O. Kallenberg: Symmetries on random arrays and set-indexed processes. J. Theoret. Probab., 5(4)(1992), 727-765.

[22] M. Karpinski and W. Schudy: Approximation schemes for the betweenness problem in tournaments and related ranking problems, Proc. of APPROX and RANDOM (2011), 277-288.

[23] M. Karpinski and W. Schudy: Linear time approximation schemes for the Gale-Berlekamp game and related minimization problems, Proc. of STOC (2009), 313-322.

[24] M. Langberg, Y. Rabani, and C. Swamy: Approximation algorithms for graph homomorphism problems, Proc. of APPROX and RANDOM (2006), Vol. 4110, 176187.

[25] L. Lovász: Large Networks and Graph Limits, Colloquium Publications 60, American Mathematical Society 2012.

[26] L. Lovász and B. Szegedy: Limits of dense graph sequences, J. Comb. Theory B 96 (2006), 933-957.

[27] L. Lovász and B. Szegedy: Szemerédi's lemma for the analyst, GAFA Vol. 17 (1) (2007), 252-270.

[28] L. Lovász and B. Szegedy: Limits of compact decorated graphs (2010) arXiv: 1010.5155v1 (2010).

[29] L. Lovász and B. Szegedy: Testing properties of graphs and functions: Israel J. of Math. Vol. 178 (1) (2010), 113-156.

[30] C. Mathieu and W. Schudy: Yet Another Algorithm for Dense Max Cut: Go Greedy. Proc. of SODA (2008), 176-182.

[31] Ya. G. Sinai: Theory of phase transitions: rigorous results, Pergamon Press, Oxford (1982).

[32] Y. Zhao: Hypergraph limits: a regularity approach, Random Structures and Algorithms, to appear.