Local Distribution and the Symmetry Gap: Approximability of Multiway Partitioning Problems

Alina Ene* Jan Vondrák[†] Yi Wu[‡]

Abstract

We study the approximability of multiway partitioning problems, examples of which include Multiway Cut, Node-weighted Multiway Cut, and Hypergraph Multiway Cut. We investigate these problems from the point of view of two possible generalizations: as Min-CSPs, and as Submodular Multiway Partition problems. These two generalizations lead to two natural relaxations that we call respectively the *Local Distribution LP*, and the *Lovász relaxation*. The Local Distribution LP is generally stronger than the Lovász relaxation, but applicable only to Min-CSP with predicates of constant size. The relaxations coincide in some cases such as Multiway Cut where they are both equivalent to the CKR relaxation.

We show that the Lovász relaxation gives a $(2 - 2/k)$-approximation for Submodular Multiway Partition with k terminals, improving a recent 2-approximation [2]. We prove that this factor is optimal in two senses: (1) A $(2-2/k-\epsilon)$-approximation for Submodular Multiway Partition with k terminals would require exponentially many value queries (in the oracle model), or imply $NP = RP$ (for certain explicit submodular functions). (2) For Hypergraph Multiway Cut and Node-weighted Multiway Cut with k terminals, both special cases of Submodular Multiway Partition, we prove that a $(2-2/k-\epsilon)$-approximation is NP-hard, assuming the Unique Games Conjecture.

Both our hardness results are more general: (1) We show that the notion of *symmetry gap*, previously used for submodular maximization problems [19, 6], also implies hardness results for submodular minimization problems. (2) Assuming the Unique Games Conjecture, we show that the Local Distribution LP gives an optimal approximation for every Min-CSP that includes the Not-Equal predicate.

Finally, we connect the two hardness techniques by proving that the *integrality gap* of the Local Distribution LP coincides with the *symmetry gap* of the multilinear relaxation (for a related instance). This shows that the appearance of the same hardness threshold for a Min-CSP and the related submodular minimization problem is not a coincidence.

1 Introduction

In this paper, we study the approximability of *multiway cut/partitioning problems*, where a ground set V should be partitioned into k parts while minimizing a certain objective function. Classical examples of such problems are Multiway Cut (that we abbreviate by GRAPH-MC), Node-weighted Multiway Cut (NODE-WT-MC) and Hypergraph Multiway Cut (HYPERGRAPH-MC). These problems are NP-hard but admit constant-factor approximations.

Multiway Cut (GRAPH-MC): *Given a graph $G = (V, E)$ with weights on the edges and k terminals $t_1, t_2, \ldots, t_k \in V$, remove a minimum-weight set of edges so that every two terminals are disconnected.*

Node-weighted Multiway Cut (NODE-WT-MC): *Given a graph $G = (V, E)$ with weights on the nodes and k terminals $t_1, t_2, \ldots, t_k \in V$, remove a minimum-weight set of vertices so that every two terminals are disconnected.*

Hypergraph Multiway Cut (HYPERGRAPH-MC): *Given a hypergraph $H = (V, E)$ with weights on the hyperedges and k terminals $t_1, t_2, \ldots, t_k \in V$, remove a minimum-weight set of hyperedges so that every two terminals are disconnected.*

Although the problems above are formulated as vertex/edge/hyperedge removal problems, GRAPH-MC and HYPERGRAPH-MC can be also viewed as partitioning problems where vertices are assigned to terminals and we pay for each edge/hyperedge that is *cut* between different terminals. The NODE-WT-MC problem can also be stated in this form, and in fact shown to be approximation-equivalent to HYPERGRAPH-MC (although the reduction is more complicated, see [16]). Given this point of view, and the fact that the cut function in graphs/hypergraphs is submodular, the following generalization of these problems was proposed in [21] and recently studied in [3, 2].

Submodular Multiway Partition (SUB-MP): *Given a submodular set function $f : 2^V \to \mathbb{R}_+$ and k terminals $t_1, t_2, \ldots, t_k \in V$, find a partition of V into $A_1, \ldots, A_k$ such that $t_i \in A_i$ and $\sum_{i=1}^{k} f(A_i)$ is minimized.*

---

*Department of Computer Science, University of Illinois at Urbana-Champaign. Supported in part by NSF grants CCF-0728782 and CCF-1016684. Part of this work was done while the author was visiting the IBM Almaden Research Center.

[†] IBM Almaden Research Center, San Jose, CA

[‡] Department of Computer Science, Purdue University. Part of this work was done while the author was a postdoc at the IBM Almaden Research Center.

This problem captures the problems Graph-MC, Node-Wt-MC and Hypergraph-MC as special cases.[1] One of the useful aspects of viewing the problems in this more general framework is that non-trivial linear programs, discovered on a case-by-case basis in the past, can be viewed in a unified way: by considering the *Lovász extension* of a submodular function (see Section 2 for details). In particular, [3] rederives the geometric CKR relaxation for Multiway Cut [1] in this way. It was shown in [2] that Sub-MP is not significantly more difficult than its special cases: Sub-MP admits a 2-approximation in general, and an improved $(3/2 - 1/k)$-approximation when the number of terminals is k and f is a symmetric submodular function (in the sense that $f(S) = f(\bar{S})$; we denote this special case as Sub-MP-Sym). We remark that Graph-MC is a special case of Sub-MP-Sym, while Node-Wt-MC and Hypergraph-MC are not.

This compares to the approximability of the classical problems as follows: Graph-MC admits a 1.3438-approximation [9], which is an improvement of an earlier breakthrough result, a $(3/2-1/k)$-approximation using the CKR relaxation [1]. Note that the latter approximation matches the result of [2] for Sub-MP-Sym. On the hardness side, it is known that the CKR relaxation gives the optimal approximation for Graph-MC, assuming the Unique Games Conjecture [14]. However, the actual approximation factor is not known: it is only known that it is between 8/7 and 1.3438, and in the case of 3 terminals it is known to be 12/11 [14].

The problems Node-Wt-MC and Hypergraph-MC are known to be approximation-equivalent, and both admit a $(2 - 2/k)$-approximation for k terminals [16, 7]. It is known that a $(2 - \epsilon)$-approximation independent of k would imply a $(2 - \epsilon)$-approximation for Vertex Cover, which would refute the Unique Games Conjecture. Therefore, we do not expect an approximation better than 2 for Node-Wt-MC and Hypergraph-MC when the number of terminals is large. Nevertheless, this reduction does not give any hardness for a constant number of terminals k, and the optimal approximation for a given k was not known.

Our contribution. We study these partitioning problems from two points of view: (a) the submodular generalization Sub-MP of the problem, and a natural convex program for the problem based on the Lovász extension of a submodular function; we refer to this convex program as the *Lovász relaxation*; (b) regarding them as Min-CSP (constraint satisfaction problems) leads to another natural relaxation that we dub the *Local distribution LP*. Our concrete results are as follows.

Concrete results:

- We give a $(2 - 2/k)$-approximation for the Sub-MP problem with k terminals using the Lovász relaxation. We also show that this is optimal, in two different senses.
- We prove that any $(2 - 2/k - \epsilon)$-approximation for Sub-MP (for a constant number of terminals k) requires exponentially many value queries (in the oracle model), or it implies $NP = RP$ (for certain succinctly represented submodular functions, using the technique of [6]).
- We prove that for Node-Wt-MC, a special case of Sub-MP, it is Unique-Games-hard to achieve a $(2-2/k-\epsilon)$-approximation (for a constant number of terminals k).

Since Hypergraph-MC is approximation-equivalent to Node-Wt-MC, we determine the approximability of all three problems, Sub-MP, Hypergraph-MC and Node-Wt-MC, to be exactly $2 - 2/k$ (assuming the Unique Games Conjecture in the case of Hypergraph-MC and Node-Wt-MC).

We also present a generic LP rounding algorithm that matches the integrality gap of the Local Distribution LP. A similar rounding algorithm also applies to the *Lovász relaxation* of Sub-MP, achieving a ratio arbitrarily close to its integrality gap. This rounding algorithm is inspired by the technique of [11].

Unique Games-hardness vs. NP-hardness more generally: Our hardness proofs in fact lead to more general results, revealing an interesting relationship between the Unique Games-hardness of Min-CSP problems and NP-hardness of their natural submodular generalizations.

- We show a Unique Games-based hardness result for Min-CSP problems, generalizing the machinery of [14]. Roughly speaking, we show that for every Min-CSP problem that includes the Not-Equal predicate, the integrality gap of the Local Distribution LP can be translated to a Unique Games-hardness result.
- We show how the *symmetry gap* technique, previously developed for submodular maximization problems [19, 6], applies to submodular minimization problems. In particular, we prove that a

---

[1] We point out that in the case of Hypergraph-MC, the reduction is not as direct as one might expect - this is due to the fact that we want to count each cut hyperedge only once, independently of how many terminals share it. Consequently, the arising submodular function is *not* symmetric, although the cut function in a hypergraph is.

1.26-approximation for SUB-MP-SYM implies that $NP = RP$.

Finally, we present a connection between the two approaches, proving that the *integrality gap* of the Local Distribution LP coincides with the *symmetry gap* of the multilinear relaxation (see the discussion below and Section 6 for more details).

Discussion. The interplay of different relaxations and different techniques to prove related results is in our opinion the most interesting aspect of our work. Let us comment on some connections that we observed here.
Integrality gap vs. symmetry gap. While UNIQUE GAMES-hardness results typically start from an integrality gap instance, hardness results for submodular functions often start from the *multilinear relaxation* of a problem, exhibiting a certain *symmetry gap* (see [19, 6]). This is a somewhat different concept, where instead of integral vs. fractional solutions, we compare symmetric vs. asymmetric solutions. In this paper, we clarify the relationship between the two: For any integrality gap Min-CSP instance of the Local Distribution LP, there is a related Min-CSP instance that exhibits the same symmetry gap in its multilinear relaxation. Conversely, for any symmetry gap instance of the multilinear relaxation of a Min-CSP instance, there is a related Min-CSP instance whose Local Distribution LP has the same integrality gap (see Section 6). Therefore, the two concepts are in some sense equivalent (at least for Min-CSP problems). This explains why the UNIQUE GAMES-hardness threshold for HYPERGRAPH-MC and the NP-hardness threshold for its submodular generalization SUB-MP are the same.
Lovász vs. multilinear relaxation. The fact that the symmetry gap technique gives optimal results for a submodular *minimization* problem is interesting: The symmetry gap technique is intimately tied to the notion of a *multilinear extension* of a submodular function, which has recently found numerous applications in maximization of submodular functions [18, 19, 10, 12, 4]. Nevertheless, it has been common wisdom that the *Lovász extension* is the relevant extension for submodular minimization [13, 8, 3, 2]. Here, we obtain a positive result using the Lovász extension, and a matching hardness result using the multilinear extension.

Organization. The rest of the paper is organized as follows. In Section 2, we discuss the Lovász relaxation, and show how it yields a $(2-2/k)$-approximation for the SUB-MP problem. In Section 3, we present the symmetry gap technique for submodular minimization problems, and show how it implies our hardness results in the value oracle model. In Section 4, we present our hardness result for Min-CSP, and show how it implies the hardness result for HYPERGRAPH-MC. In section 5, we discuss the relationship of the Lovász relaxation and the Local Distribution LP. In Section 6, we discuss the relationship of integrality gaps and symmetry gaps.

2 Approximation for Submodular Multiway Partition

In this section, we revisit the convex relaxation proposed by Chekuri and Ene [3], and provide an improved analysis that gives the following result.

THEOREM 2.1. *There is a polynomial-time $(2-2/k)$-approximation for the* SUB-MP *problem with k terminals, where k is given as input and the objective set function is given by a value oracle.*

The Lovász relaxation. The following is the convex relaxation that has been used by Chekuri and Ene:

$$\begin{array}{rl} (\text{SUBMP-REL}) & \min \sum_{i=1}^{k} \hat{f}(\mathbf{x}_i) : \\ \forall j \in V; & \sum_{i=1}^{k} x_{i,j} = 1; \\ \forall i \in [k]; & x_{i,t_i} = 1; \\ \forall i, j; & x_{i,j} \geq 0. \end{array}$$

Here, $\hat{f}(\mathbf{x}_i)$ denotes the Lovász extension of a submodular function. The function $\hat{f}$ can be defined in several equivalent ways (see [3, 2]). One definition is based on the following rounding strategy. We choose a uniformly random $\theta \in [0,1]$ and define $A_i(\theta) = \{j : x_{ij} > \theta\}$. Then $\hat{f}(\mathbf{x}_i) = \mathbf{E}[f(A_i(\theta))]$. Equivalently (for submodular functions), $\hat{f}$ is the convex closure of f on $[0,1]^V$. The second definition shows that the relaxation SUBMP-REL is a convex program and therefore it can be solved in polynomial time.

Given a fractional solution, we use the following randomized rounding technique, a slight modification of one proposed by Chekuri and Ene:

Randomized rounding for the Lovász relaxation.

- Choose $\theta \in (\frac{1}{2}, 1]$ uniformly at random and define $A_i(\theta) = \{j : x_{ij} > \theta\}$.
- Define $U(\theta) = V \setminus \bigcup_{i=1}^{k} A_i(\theta) = \{j : \max_i x_{ij} \leq \theta\}$.
- Allocate each $A_i(\theta)$ to terminal i, and in addition allocate $U(\theta)$ to a terminal i' chosen uniformly at random.

Each terminal t_i is allocated to itself with probability one. Moreover, the sets $A_i(\theta)$ are disjoint by construction, and therefore the rounding constructs a feasible solution. The only difference from Chekuri and Ene's rounding [2] is that we assign the "unallocated set"

$U(\theta)$ to a random terminal rather than a fixed terminal. (However, taking advantage of this in the analysis is not straightforward.) We prove the following.

THEOREM 2.2. *The above rounding gives a feasible solution of expected value at most* $(2-\frac{2}{k})\sum_{i=1}^{k}\hat{f}(\mathbf{x}_i)$.

This clearly implies Theorem 2.1. In the following, we prove Theorem 2.2. We assume that $f(\emptyset)=0$. This is without loss of generality, as the value of the empty set can be decreased without violating submodularity and this does not affect the problem (since terminals are always assigned to themselves).

We start by defining several sets, parameterized by θ, that will be important in the analysis.

- $A_i(\theta)=\{j : x_{ij}>\theta\}$
- $A(\theta)=\bigcup_{i=1}^{k}A_i(\theta)=\{j : \max_i x_{ij}>\theta\}$
- $U(\theta)=V\setminus A(\theta)=\{j : \max_i x_{ij}\le\theta\}$.
- $B(\theta)=U(1-\theta)=\{j : 1-\max_i x_{ij}\ge\theta\}$.

We can express the LP cost and the cost of the rounded solution in terms of these sets as follows. The following lemma follows immediately from the definition of the Lovász extension.

LEMMA 2.1. *The cost of the LP solution is*

$$LP=\sum_{i=1}^{k}\int_{0}^{1}f(A_i(\theta))d\theta.$$

The next lemma gives an expression for the expected value achieved by the algorithm.

LEMMA 2.2. *The expected cost of the rounded solution is*

$$\begin{aligned}ALG &= (2-\frac{2}{k})\sum_{i=1}^{k}\int_{1/2}^{1}f(A_i(\theta))d\theta\\ &\quad+\frac{2}{k}\sum_{i=1}^{k}\int_{0}^{1/2}f(A_i(\theta)\cup B(\theta))d\theta.\end{aligned}$$

Proof. The set allocated to terminal i is $A_i(\theta)$ with probability $1-1/k$, and $A_i(\theta)\cup U(\theta)$ with probability $1/k$. We are choosing θ uniformly between $\frac{1}{2}$ and 1. This gives the expression

$$\begin{aligned}ALG &= (2-\frac{2}{k})\sum_{i=1}^{k}\int_{1/2}^{1}f(A_i(\theta))d\theta\\ &+ \frac{2}{k}\sum_{i=1}^{k}\int_{1/2}^{1}f(A_i(\theta)\cup U(\theta))d\theta.\end{aligned}$$

We claim that for $\theta\in[\frac{1}{2},1]$, $A_i(\theta)\cup U(\theta)$ can be written equivalently as $A_i(1-\theta)\cup B(1-\theta)$. We consider three cases for each element j:

- If $x_{ij}>\frac{1}{2}$, then $j\in A_i(\theta)\cup U(\theta)$ for every $\theta\in[\frac{1}{2},1]$, because $x_{i'j}<\frac{1}{2}$ for every other $i'\neq i$ and hence j cannot be allocated to any other terminal. Similarly, $j\in A_i(1-\theta)\cup B(1-\theta)$ for every $\theta\in[\frac{1}{2},1]$, because $1-\theta\le\frac{1}{2}$ and so $j\in A_i(1-\theta)$.
- If $x_{ij}\le\frac{1}{2}$ and $x_{ij}=\max_{i'}x_{i'j}$, then again $j\in A_i(\theta)\cup U(\theta)$ for every $\theta\in[\frac{1}{2},1]$, because j is always in the unallocated set $U(\theta)$. Also, $j\in A_i(1-\theta)\cup B(1-\theta)$, because $B(1-\theta)=U(\theta)$.
- If $x_{ij}\le\frac{1}{2}$ and $x_{ij}<\max_{i'}x_{i'j}$, then $j\in A_i(\theta)\cup U(\theta)$ if and only if $j\in U(\theta)=B(1-\theta)$. Also, we have $x_{ij}=1-\sum_{i'\neq i}x_{i'j}\le 1-\max_{i'}x_{i'j}$, and therefore $j\in A_i(1-\theta)\cup B(1-\theta)$ if and only if $j\in B(1-\theta)$.

To summarize, for every $\theta\in[\frac{1}{2},1]$, $j\in A_i(\theta)\cup U(\theta)$ if and only if $j\in A_i(1-\theta)\cup B(1-\theta)$. Therefore, the total expected cost can be written as

$$\begin{aligned}ALG &= (2-\frac{2}{k})\sum_{i=1}^{k}\int_{1/2}^{1}f(A_i(\theta))d\theta\\ &\quad+\frac{2}{k}\sum_{i=1}^{k}\int_{1/2}^{1}f(A_i(\theta)\cup U(\theta))d\theta\\ &= (2-\frac{2}{k})\sum_{i=1}^{k}\int_{1/2}^{1}f(A_i(\theta))d\theta\\ &\quad+\frac{2}{k}\sum_{i=1}^{k}\int_{1/2}^{1}f(A_i(1-\theta)\cup B(1-\theta))d\theta\\ &= (2-\frac{2}{k})\sum_{i=1}^{k}\int_{1/2}^{1}f(A_i(\theta))d\theta\\ &\quad+\frac{2}{k}\sum_{i=1}^{k}\int_{0}^{1/2}f(A_i(\theta)\cup B(\theta))d\theta.\end{aligned}$$

In the rest of the analysis, we prove several inequalities that relate the LP cost to the ALG cost. Note that the integrals $\int_{1/2}^{1}f(A_i(\theta))d\theta$ appear in both LP and ALG. The interesting part is how to relate $\int_{0}^{1/2}f(A_i(\theta))d\theta$ to $\int_{0}^{1/2}f(A_i(\theta)\cup B(\theta))d\theta$.

The following statement was proved in [2]; we give a simplified new proof in the process of our analysis.

LEMMA 2.3. (THEOREM 1.5 IN [2]) *Let $f\ge 0$ be submodular, $f(\emptyset)=0$, and $\mathbf{x}$ a feasible solution to* SUBMP-REL. *For $\theta\in[0,1]$ let $A_i(\theta)=\{v\mid x_{v,i}>\theta\}$,*

$A(\theta) = \cup_{i=1}^k A_i(\theta)$ and $U(\theta) = V \setminus A(\theta)$. For any $\delta \in [\frac{1}{2}, 1]$ the following holds:

$$\sum_{i=1}^{k} \int_0^{\delta} f(A_i(\theta))d\theta \geq \int_0^{\delta} f(A(\theta))d\theta + \int_0^{1} f(U(\theta))d\theta.$$

In the following, we assume the conditions of Lemma 2.3 without repeating them. First, we prove the following inequality.

Lemma 2.4. *For any $\delta \in [\frac{1}{2}, 1]$,*

$$\sum_{i=1}^{k-1} \int_0^{\delta} f((A_1(\theta) \cup \cdots A_i(\theta)) \cap A_{i+1}(\theta))d\theta \geq \int_0^{1} f(U(\theta))d\theta.$$

Proof. First consider $\delta = 1$. We can view the value $\int_0^1 f(A_1(\theta) \cup \cdots \cup A_i(\theta)) \cap A_{i+1}(\theta)d\theta$ as the Lovász extension evaluated on a suitable vector, $\mathbf{y}_i = (\mathbf{x}_1 \vee \cdots \vee \mathbf{x}_i) \wedge \mathbf{x}_{i+1}$. Note that v is in $(A_1(\theta) \cup \cdots \cup A_i(\theta)) \cap A_{i+1}(\theta)$ if and only if $y_{v,i} \geq \theta$. Therefore

$$\int_0^1 f((A_1(\theta) \cup \cdots \cup A_i(\theta)) \cap A_{i+1}(\theta))d\theta = \hat{f}(\mathbf{y}_i).$$

We can also view $f(U(\theta))$ as follows: Let $\boldsymbol{u} = \sum_{i=1}^{k-1} \mathbf{y}_i = \mathbf{1} - (\mathbf{x}_1 \vee \cdots \vee \mathbf{x}_k)$. (This holds because $\sum_{i=1}^{k-1} \mathbf{y}_i + (\mathbf{x}_1 \vee \cdots \vee \mathbf{x}_k) = \sum_{i=1}^{k-1}((\mathbf{x}_1 \vee \cdots \vee \mathbf{x}_i) \wedge \mathbf{x}_{i+1}) + (\mathbf{x}_1 \vee \cdots \vee \mathbf{x}_k) = \sum_{i=1}^{k} \mathbf{x}_i$, which can be proved by repeated use of the rule $(\boldsymbol{u} \wedge \mathbf{v}) + (\boldsymbol{u} \vee \mathbf{v}) = \boldsymbol{u} + \mathbf{v}$, and finally $\sum_{i=1}^{k} \mathbf{x}_i = \mathbf{1}$.) Therefore

$$\begin{aligned} \frac{1}{k-1} \sum_{i=1}^{k-1} \hat{f}(\mathbf{y}_i) &\geq \hat{f}\left(\frac{1}{k-1} \sum_{i=1}^{k-1} \mathbf{y}_i\right) \qquad (\hat{f} \text{ is convex}) \\ &= \hat{f}\left(\frac{1}{k-1} \boldsymbol{u}\right) = \frac{1}{k-1} \hat{f}(\boldsymbol{u}) \end{aligned}$$

where we also used the fact that $\hat{f}(\alpha \mathbf{x}) = \alpha \hat{f}(\mathbf{x})$ for any $\alpha \in [0,1]$ ($\hat{f}(\mathbf{x})$ is linear along any line through the origin). Equivalently,

$$\sum_{i=1}^{k-1} \int_0^{1} f((A_1(\theta) \cup \cdots A_i(\theta)) \cap A_{i+1}(\theta))d\theta \geq \int_0^{1} f(U(\theta))d\theta.$$

Now note that, if $\theta > \delta$, the sets $(A_1(\theta) \cup \cdots \cup A_i(\theta)) \cap A_{i+1}(\theta)$ are empty, since $\sum_{i=1}^{k} \mathbf{x}_i = \mathbf{1}$ and hence two vectors $\mathbf{x}_i, \mathbf{x}_j$ cannot have the same coordinate larger than $\theta > \delta \geq \frac{1}{2}$. We also assumed that $f(\emptyset) = 0$, so we proved in fact

$$\sum_{i=1}^{k-1} \int_0^{\delta} f((A_1(\theta) \cup \cdots A_i(\theta)) \cap A_{i+1}(\theta))d\theta \geq \int_0^{1} f(U(\theta))d\theta$$

as desired.

Given this inequality, Lemma 2.3 follows easily:

Proof. [Lemma 2.3] By applying submodularity inductively to the sets $A_1(\theta) \cup \cdots \cup A_i(\theta)$ and $A_{i+1}(\theta)$, we get

$$\begin{aligned} & \sum_{i=1}^{k} f(A_i(\theta)) \\ \geq\ & \sum_{i=1}^{k-1} f((A_1(\theta) \cup \cdots \cup A_i(\theta)) \cap A_{i+1}(\theta)) \\ & + f(A_1(\theta) \cup \cdots \cup A_k(\theta)) \\ =\ & \sum_{i=1}^{k-1} f((A_1(\theta) \cup \cdots \cup A_i(\theta)) \cap A_{i+1}(\theta)) \\ & + f(A(\theta)). \end{aligned}$$

Integrating from 0 to δ and using Lemma 2.4, we obtain

$$\begin{aligned} & \sum_{i=1}^{k} \int_0^{\delta} f(A_i(\theta))d\theta \\ \geq\ & \sum_{i=1}^{k-1} \int_0^{\delta} f((A_1(\theta) \cup \cdots \cup A_i(\theta)) \cap A_{i+1}(\theta))d\theta \\ & + \int_0^{\delta} f(A(\theta))d\theta \\ \geq\ & \int_0^{1} f(U(\theta))d\theta + \int_0^{\delta} A(\theta)d\theta. \end{aligned}$$

A corollary of Lemma 2.3 is the following inequality.

Lemma 2.5.

$$\sum_{i=1}^{k} \int_0^{1/2} f(A_i(\theta))d\theta \geq \int_0^{1/2} f(B(\theta))d\theta.$$

Proof. Considering Lemma 2.3, we simply note that $U(\theta) = B(1-\theta)$. We discard the contribution of $f(A(\theta))$ and keep only one half of the integral involving $B(1-\theta)$.

We combine this bound with the following lemma which is a new contribution of this paper.

Lemma 2.6.

$$\begin{aligned} \sum_{i=1}^{k} \int_0^{1/2} f(A_i(\theta))d\theta \ \geq\ & \sum_{i=1}^{k} \int_0^{1/2} f(A_i(\theta) \cup B(\theta))d\theta \\ & -(k-2) \int_0^{1/2} f(B(\theta))d\theta. \end{aligned}$$

Proof. For simplicity of notation, we drop the explicit dependence on θ, keeping in mind that all the sets depend on θ. By submodularity, we have $f(A_i)+f(B) \geq f(A_i \cup B) + f(A_i \cap B)$. Therefore,

$$\sum_{i=1}^{k} f(A_i) \geq \sum_{i=1}^{k} (f(A_i \cup B) + f(A_i \cap B) - f(B))$$

$$= \sum_{i=1}^{k} f(A_i \cup B) + \sum_{i=1}^{k} f(A_i \cap B) - k \cdot f(B).$$

This would already prove the lemma with k instead of $k-2$; however, we use $\sum_{i=1}^{k} f(A_i \cap B)$ to save the additional terms. We apply a sequence of inequalities using submodularity, starting with $f(A_1 \cap B) + f(A_2 \cap B) \geq f(A_1 \cap A_2 \cap B) + f((A_1 \cup A_2) \cap B)$, then $f((A_1 \cup A_2) \cap B) + f(A_3 \cap B) \geq f((A_1 \cup A_2) \cap A_3 \cap B) + f((A_1 \cup A_2 \cup A_3) \cap B)$, etc. until we obtain

$$\begin{aligned} \sum_{i=1}^{k} f(A_i \cap B) \quad &\geq \quad \sum_{i=1}^{k-1} f((A_1 \cup \ldots \cup A_i) \cap A_{i+1} \cap B) \\ &\quad + f((A_1 \cup \ldots \cup A_k) \cap B). \end{aligned}$$

The last term is equal to $f(A \cap B)$. Moreover, we observe that for every element j, at most one variable x_{ij} can be larger than $1 - \max_{i'} x_{i'j}$ (because otherwise the two variables would sum up to more than 1). Therefore for every i, $(A_1 \cup \ldots \cup A_i) \cap A_{i+1} \subseteq B$. So we get

$$\sum_{i=1}^{k} f(A_i \cap B) \geq \sum_{i=1}^{k-1} f((A_1 \cup \ldots \cup A_i) \cap A_{i+1}) + f(A \cap B).$$

Integrating from 0 to 1/2, we get

$$\begin{aligned} &\sum_{i=1}^{k} \int_0^{1/2} f(A_i \cap B) d\theta \\ \geq\ &\sum_{i=1}^{k-1} \int_0^{1/2} f((A_1 \cup \ldots \cup A_i) \cap A_{i+1}) d\theta \\ &+ \int_0^{1/2} f(A \cap B) d\theta. \end{aligned}$$

By Lemma 2.4 (recalling that $A_i = A_i(\theta)$), we obtain

$$\begin{aligned} &\sum_{i=1}^{k} \int_0^{1/2} f(A_i \cap B) d\theta \\ \geq\ &\int_0^{1} f(U) d\theta + \int_0^{1/2} f(A \cap B) d\theta. \end{aligned}$$

Using $B(\theta) = U(1-\theta)$, submodularity, and the fact that U is the complement of A, we obtain

$$\begin{aligned} &\sum_{i=1}^{k} \int_0^{1/2} f(A_i \cap B) d\theta \\ \geq\ &\int_0^{1/2} f(B) d\theta + \int_0^{1/2} f(U) d\theta + \int_0^{1/2} f(A \cap B) d\theta \\ \geq\ &\int_0^{1/2} f(B) d\theta + \int_0^{1/2} f(U \cup (A \cap B)) d\theta \\ =\ &\int_0^{1/2} f(B) d\theta + \int_0^{1/2} f(U \cup B) d\theta. \end{aligned}$$

Finally, for $\theta \in [0, \frac{1}{2}]$, we claim that $U \cup B = B$. This is because if $\max_i x_{ij} > \frac{1}{2}$, then $j \notin U$, and hence the membership on both sides depends only on $j \in B$. If $\max_i x_{ij} \leq \frac{1}{2}$, then $j \in B$ and hence also $j \in U \cup B$. We conclude that

$$\sum_{i=1}^{k} \int_0^{1/2} f(A_i \cap B) d\theta \quad \geq \quad 2 \int_0^{1/2} f(B) d\theta$$

and

$$\begin{aligned} &\sum_{i=1}^{k} \int_0^{1/2} f(A_i) d\theta \\ \geq\ &\sum_{i=1}^{k} \int_0^{1/2} (f(A_i \cup B) + f(A_i \cap B) - f(B)) d\theta \\ \geq\ &\sum_{i=1}^{k} \int_0^{1/2} f(A_i \cup B) d\theta - (k-2) \int_0^{1/2} f(B) d\theta \end{aligned}$$

which finishes the proof.

A combination of Lemma 2.5 and Lemma 2.6 relates $\sum_{i=1}^{k} \int_0^{1/2} f(A_i(\theta)) d\theta$ to $\sum_{i=1}^{k} \int_0^{1/2} f(A_i(\theta) \cup B(\theta)) d\theta$, and finishes the analysis.

Proof. [Proof of Theorem 2.2] Add up $\frac{k-2}{k-1} \times$ Lemma 2.5 $+ \frac{1}{k-1} \times$ Lemma 2.6:

$$\sum_{i=1}^{k} \int_0^{1/2} f(A_i(\theta)) d\theta \geq \frac{1}{k-1} \sum_{i=1}^{k} \int_0^{1/2} f(A_i(\theta) \cup B(\theta)) d\theta.$$

Add $\sum_{i=1}^{k} \int_{1/2}^{1} f(A_i(\theta)) d\theta$ to both sides: we obtain

$$\begin{aligned} \sum_{i=1}^{k} \int_0^{1} f(A_i(\theta)) d\theta \quad &\geq \quad \sum_{i=1}^{k} \int_{1/2}^{1} f(A_i(\theta)) d\theta \\ &\quad + \frac{1}{k-1} \cdot \sum_{i=1}^{k} \int_0^{1/2} f(A_i(\theta) \cup B(\theta)) d\theta. \end{aligned}$$

The left-hand side is equal to LP, while the right-hand side is equal to $\frac{ALG}{2-2/k}$ (see Lemma 2.2).

3 Optimality from symmetry gap

Here we show how the symmetry gap technique of [19] applies to submodular minimization problems. We remark that while the technique was presented in [19] for submodular maximization problems, it applies to submodular minimization problems practically without any change. Rather than repeating the entire construction of [19], we summarize the main components of the proof and point out the important differences. Finally, we mention that the recent techniques of [6] turn a query-complexity hardness result into a computational hardness result. First, we show the result for general Submodular Multiway Partition, which is technically simpler.

Remark. The main general theorem in [19] contains an error which was recently corrected in [20]. The applications in [19] still hold true. The error affects the hardness result of [19] for general feasibility constraints. This does not affect our exposition here, since we discuss only concrete problems and do not formulate a general hardness result.

3.1 Hardness of Sub-MP

Here we show that the $(2 - 2/k)$-approximation is optimal for Submodular Multiway Partition in the value oracle model. More precisely, we prove the following.

THEOREM 3.1. *For any fixed $k > 2$ and $\epsilon > 0$, a $(2 - 2/k - \epsilon)$-approximation for the Submodular Multiway Partition problem with k terminals in the value oracle model requires exponentially many value queries (or it implies that $NP = RP$ for certain succinctly represented submodular functions).*

We note that the computational hardness part (assuming $NP \neq RP$) follows from the oracle hardness proof, using the techniques of [6]. We defer the details to the full version and focus here on the oracle hardness part.

A starting point of the hardness construction is a particular instance of the problem which exhibits a certain *symmetry gap*, a gap between symmetric and asymmetric solutions of the multilinear relaxation. We propose the following instance (which is somewhat related to the gadget used in [5] to prove the APX-hardness of Multiway Cut). The instance is in fact an instance of Hypergraph Multiway Cut (which is a special case of Submodular Multiway Partition). As in other cases, we should keep in mind that this does not mean that we prove a hardness result for Hypergraph Multiway Cut, since the instance gets modified in the process.

The symmetric instance. Let the vertex set be $V = [k] \times [k]$ and let the terminals be $t_i = (i, i)$. We consider $2k$ hyperedges: the "rows" $R_i = \{(i, j) : 1 \leq j \leq k\}$ and the "columns" $C_j = \{(i, j) : 1 \leq i \leq k\}$. The submodular function $f : 2^V \to \mathbb{R}_+$ is the following function: for each set S, $f(S) = \sum_{i=1}^{k} \phi(|S \cap R_i|) + \sum_{j=1}^{k} \phi(|S \cap C_j|)$, where $\phi(t) = t/k$ if $t < k$ and $\phi(t) = 0$ if $t = k$.

Since ϕ is a concave function, it follows easily that f is submodular. Further, if a hyperedge is assigned completely to one terminal, it does not contribute to the objective function, while if it is partitioned among different terminals, it contributes t/k to each terminal containing t of its k vertices, and hence 1 altogether. Therefore, $\sum_{i=1}^{k} f(S_i)$ captures exactly the number of hyperedges cut by a partition $(S_1, \ldots, S_k)$.

The multilinear relaxation. Along the lines of [19], we want to compare symmetric and asymmetric solutions of the *multilinear relaxation* of the problem, where we allocate vertices fractionally and the objective function $f : 2^V \to \mathbb{R}_+$ is replaced by its multilinear extension $F : [0,1]^V \to \mathbb{R}_+$; we have $F(\mathbf{x}) = \mathbf{E}[f(\hat{\mathbf{x}})]$, where $\hat{\mathbf{x}}$ is the integral vector obtained from $\mathbf{x}$ by rounding each coordinate independently. The multilinear relaxation of the problem has variables x_{ij}^{ℓ} corresponding to allocating (i, j) to terminal t_ℓ.

$$\begin{aligned} \min\Big\{ \sum_{\ell=1}^{k} F(\mathbf{x}^\ell) \quad &: \quad \forall i, j \in [k]; \sum_{\ell=1}^{k} x_{ij}^{\ell} = 1, \\ &\quad\ \forall i \in [k]; x_{ii}^{i} = 1, \\ &\quad\ \forall i, j, \ell \in [k]; x_{ij}^{\ell} \geq 0 \Big\}. \end{aligned}$$

In fact, this formulation is equivalent to the discrete problem, since any fractional solution can be rounded by assigning each vertex (i, j) independently with probabilities x_{ij}^{ℓ}, and the expected cost of this solution is by definition $\sum_{\ell=1}^{k} F(\mathbf{x}^\ell)$.

Computing the symmetry gap. What is the symmetry gap of this instance? It is quite easy to see that there is a symmetry between the rows and the columns, i.e., we can exchange the role of rows and columns and the instance remains the same. Formally, the instance is invariant under a group $\mathcal{G}$ of permutations of V, where $\mathcal{G}$ consists of the identity and the transposition of rows and columns. A symmetric solution is one invariant under this transposition, i.e., such that the vertices (i, j) and (j, i) are allocated in the same manner, or $x_{ij}^{\ell} = x_{ji}^{\ell}$. For a fractional solution $\mathbf{x}$, we define the symmetrized solution as $\bar{\mathbf{x}} = \frac{1}{2}(\mathbf{x} + \mathbf{x}^T)$ where $\mathbf{x}^T$ is the transposed solution $(x^T)_{ij}^{\ell} = x_{ji}^{\ell}$.

There are two optimal solutions to this problem: one that assigns vertices based on rows, and one that assigns vertices based on columns. The first one can be written as follows: $x^\ell_{ij} = 1$ iff $i = \ell$ and 0 otherwise. (One can recognize this as a "dictator" function, one that copies the first coordinate.) The cost of this solution is k, because we cut all the column hyperedges and none of the rows. In fact, we must cut at least k hyperedges, because for any $i \neq j$, we must cut either row R_i or column C_j. Therefore, $OPT = k$.

Next, we want to find the optimal symmetric solution. As we observed, there is a symmetry between rows and columns and hence we want to consider only solutions satisfying $x^\ell_{ij} = x^\ell_{ji}$ for all i, j. Again, we claim that it is enough to consider integer (symmetric) solutions. This is for the following reason: we can assign each pair of vertices (i, j) and (j, i) in a coordinated fashion to the same random terminal: we assign (i, j) and (j, i) to the terminal t_ℓ with probability $x^\ell_{ij} = x^\ell_{ji}$. Since these two vertices never participate in the same hyperedge, the expected cost of this correlated randomized rounding is equal to the cost of independent randomized rounding, where each vertex is assigned independently. Hence the expected cost of the rounded symmetric solution is exactly $\sum_{\ell=1}^{k} F(\mathbf{x}^\ell)$.

Considering integer symmetric instances yields the following optimal solution: We can assign all vertices (except the terminals themselves) to the same terminal, let's say t_1. This will cut all hyperedges except 2 (the row R_1 and the column C_1). This is in fact the minimum-cost symmetric solution, because once we have any monochromatic row (where monochromatic means assigned to the same terminal), the respective column is also monochromatic. But this row and column intersect all other rows and columns, and hence no other row or column can be monochromatic (recall that the terminals are on the diagonal and by definition are assigned to themselves). Hence, a symmetric solution can have at most 2 hyperedges that are not cut. Therefore, the symmetric optimum is $\overline{OPT} = 2k - 2$ and the symmetry gap is $\gamma = (2k - 2)/k = 2 - 2/k$.

The hardness proof. We appeal now to a technical lemma from [20], which serves to produce blown-up instances from the initial symmetric instance.

LEMMA 3.1. *Consider a function $f : 2^V \to \mathbb{R}$ that is invariant under a group of permutations $\mathcal{G}$ on the ground set X. Let $F(\mathbf{x}) = \mathbf{E}[f(\hat{\mathbf{x}})]$, $\bar{\mathbf{x}} = \mathbf{E}_{\sigma \in \mathcal{G}}[\sigma(\mathbf{x})]$, and fix any $\epsilon > 0$. Then there exists $\delta > 0$ and functions $\hat{F}, \hat{G} : [0, 1]^V \to \mathbb{R}_+$ (which are also symmetric with respect to $\mathcal{G}$), satisfying:*

1. *For all $\mathbf{x} \in [0, 1]^V$, $\hat{G}(\mathbf{x}) = \hat{F}(\bar{\mathbf{x}})$.*
2. *For all $\mathbf{x} \in [0, 1]^V$, $|\hat{F}(\mathbf{x}) - F(\mathbf{x})| \leq \epsilon$.*
3. *Whenever $||\mathbf{x} - \bar{\mathbf{x}}||^2 \leq \delta$, $\hat{F}(\mathbf{x}) = \hat{G}(\mathbf{x})$ and the value depends only on $\bar{\mathbf{x}}$.*
4. *The first partial derivatives of $\hat{F}, \hat{G}$ are absolutely continuous.*
5. *If f is monotone, then $\frac{\partial \hat{F}}{\partial x_i} \geq 0$ and $\frac{\partial \hat{G}}{\partial x_i} \geq 0$ everywhere.*
6. *If f is submodular, then $\frac{\partial^2 \hat{F}}{\partial x_i \partial x_j} \leq 0$ and $\frac{\partial^2 \hat{G}}{\partial x_i \partial x_j} \leq 0$ almost everywhere.*

We apply Lemma 3.1 to the function f from the symmetric instance. This will produce continuous functions $\hat{F}, \hat{G} : 2^V \to \mathbb{R}_+$. Next, we use the following lemma from [20] to discretize the continuous function $\hat{F}, \hat{G}$ and obtain instances of SUB-MP.

LEMMA 3.2. *Let $F : [0, 1]^V \to \mathbb{R}$ be a function with absolutely continuous[2] first partial derivatives. Let $N = [n]$, $n \geq 1$, and define $f : N \times V \to \mathbb{R}$ so that $f(S) = F(\mathbf{x})$ where $x_i = \frac{1}{n}|S \cap (N \times \{i\})|$. Then*

1. *If $\frac{\partial F}{\partial x_i} \geq 0$ everywhere for each i, then f is monotone.*
2. *If $\frac{\partial^2 F}{\partial x_i \partial x_j} \leq 0$ almost everywhere for all i, j, then f is submodular.*

Using Lemma 3.2, we define blown-up instances on a ground set $X = N \times V$ as follows: For each $i \in N$, choose independently a random permutation $\sigma \in \mathcal{G}$ on V, which is either the identity or the transposition of rows and columns. Then for a set $S \subseteq N \times V$, we define $\xi(S) \in [0, 1]^V$ as follows:

$$\xi_j(S) = \frac{1}{n} \left| \{ i \in N : (i, \sigma^{(i)}(j)) \in S \} \right|.$$

We define two functions $\hat{f}, \hat{g} : 2^V \to \mathbb{R}_+$, where

$$\hat{f}(S) = \hat{F}(\xi(S)), \qquad \hat{g}(S) = \hat{G}(\xi(S)).$$

By Lemma 3.2, $\hat{f}, \hat{g}$ are submodular functions. We consider the following instances of SUB-MP:

$$\max \Big\{ \sum_{\ell=1}^{k} \hat{f}(S_\ell) : (S_1, \ldots, S_k) \text{ partition of } X \\ \& \ \forall i \in N; \forall \ell \in [k]; (i, t_\ell) \in S_\ell \Big\},$$

$$\max \Big\{ \sum_{\ell=1}^{k} \hat{g}(S_\ell) : (S_1, \ldots, S_k) \text{ partition of } X \\ \& \ \forall i \in N; \forall \ell \in [k]; (i, t_\ell) \in S_\ell \Big\}.$$

---

[2] A function $F : [0, 1]^V \to \mathbb{R}$ is absolutely continuous, if $\forall \epsilon > 0; \exists \delta > 0; \sum_{i=1}^{t} ||\mathbf{x}_i - \mathbf{y}_i|| < \delta \Rightarrow \sum_{i=1}^{t} |F(\mathbf{x}_i) - F(\mathbf{y}_i)| < \epsilon$.

Note that in these instances, multiple vertices are pre-labeled as assigned to a certain terminal (n vertices for each terminal). However, this can be still viewed as a Submodular Multiway Partition problem; if desired, the set of pre-labeled vertices $T_\ell = N \times \{t_\ell\}$ for each terminal can be contracted into one vertex.

Finally, we appeal to the following lemma in [20].

Lemma 3.3. *Let $\hat{F}, \hat{G}$ be the two functions provided by Lemma 3.1. For a parameter $n \in \mathbb{Z}_+$ and $N = [n]$, define two discrete functions $\hat{f}, \hat{g} : 2^{N \times V} \to \mathbb{R}_+$ as follows: Let $\sigma^{(i)}$ be an arbitrary permutation in $\mathcal{G}$ for each $i \in N$. For every set $S \subseteq N \times V$, we define a vector $\xi(S) \in [0,1]^V$ by*

$$\xi_j(S) = \frac{1}{n} \left| \{ i \in N : (i, \sigma^{(i)}(j)) \in S \} \right|.$$

Let us define: $\hat{f}(S) = \hat{F}(\xi(S))$, $\hat{g}(S) = \hat{G}(\xi(S))$. Then deciding whether a function given by a value oracle is $\hat{f}$ or $\hat{g}$ (even using a randomized algorithm with any constant probability of success) requires an exponential number of queries.

Lemma 3.3 implies that distinguishing these pairs of objective functions requires an exponential number of queries. We need to make one additional argument, that the knowledge of the terminal sets $T_\ell = N \times \{t_\ell\}$ (which is part of the instance) does not help in distinguishing the two objective functions. This is because given oracle access to $\hat{f}$ or $\hat{g}$, we are in fact able to identify the sets T_ℓ, if we just modify the contribution of each row/column pair R_ℓ, C_ℓ by a factor of $1 + \epsilon_\ell$, where ϵ_ℓ is some arbitrary small parameter. This does not change the optimal values significantly, but it allows an algorithm to distinguish the sets T_ℓ easily by checking marginal values. Note that then we can also determine sets such as $T_{i,j} = N \times \{(i,j),(j,i)\}$, but we cannot distinguish the two symmetric parts of $T_{i,j}$, which is the whole point of the symmetry argument. In summary, revealing the sets T_ℓ does not give any information that the algorithm cannot determine from the value oracles for $\hat{f}, \hat{g}$, and given this oracle access, $\hat{f}$ and $\hat{g}$ cannot be distinguished.

It remains to compare the optima of the two optimization problems. The problem with the objective function $\hat{f}$ corresponds to the multilinear relaxation with objective $\hat{F}$, and admits the "dictatorship" solution $S_i = \{(i,j) : 1 \leq j \leq k\}$ for each $i \in [k]$, which has a value close to k. On the other hand, any solution of the problem with objective function $\hat{g}$ corresponds to a fractional solution of the symmetrized multilinear relaxation of the problem with objective $\hat{G}$, which as we argued has a value close to $2k - 2$. Therefore, achieving a $(2 - 2/k - \epsilon)$-approximation for any fixed $\epsilon > 0$ requires an exponential number of value queries.

3.2 Hardness of Symmetric Submodular Multiway Partition

Here we state a result for the Sub-MP-Sym problem. The proof is essentially identical to the previous section, however the symmetric instance is different due to the requirement that the submodular function itself be symmetric (in the sense that $f(S) = f(\bar{S})$). The analysis of the symmetry gap in this case is technically more involved than in the previous section. The result that we obtain is as follows; we defer the proof to the full version.

Theorem 3.2. *For any fixed k sufficienly large, a 1.26-approximation for the* Sub-MP-Sym *problem with k terminals requires exponentially many value queries (or it implies that $NP = RP$ for certain succinctly represented submodular functions).*

4 Hypergraph Multiway Cut and Min-CSPs

In this section, we formulate our general hardness result for Min-CSP problems, and in particular we show how it implies the hardness result for Hypergraph Multiway Cut (Hypergraph-MC). The Min-CSPs we consider consist of a set of variables and a set of predicates (or cost functions) with constant arity over the variables. The goal is to assign a value from some finite domain to each variable so as to minimize the total cost of an assignment. Alternatively, we can view these variables as vertices of a hypergraph and the predicates being evaluated on the hyperedges of the hypergraph.

Definition 4.1. (Min-CSP) *A ß-CSP $\mathcal{I}(V, E, k, \{L_v | v \in V\}, h)$ is defined over a weighted h-uniform multi-hypergraph (V, E). Here ß is a set of predicate functions that take at most h inputs in $[k]$ and output a value in $[0,1]$. There is a candidate list of labels $L_v \subseteq [k]$ for each vertex v. For every hyperedge $e = (v_{i_1}, v_{i_2}, .., v_{i_j}) \in E$ where $j \leq h$, there is an associated predicate function $\Psi_e \in$ ß $: [k]^j \to [0,1]$ and a positive weight w_e. The goal is to assign a value $l_i \in L_v$ to each vertex v_i so as to minimize*

$$\sum_{e=(v_{i_1},..v_{i_j}) \in E} w_e \cdot \Psi_e(l_{i_1}, \ldots, l_{i_j}).$$

One may notice that a candidate list for a vertex can be replaced by a unary constraint on the vertex with a large cost on choosing labels outside the candidate list. We still include the candidate list in the definition mainly because for many important Min-CSPs, it is more natural. For example, the Hypergraph-MC problem can be viewed as a ß-CSP with the following predicate function: for every edge $e = (v_1, v_2, \ldots, v_j)$, $\Psi_e = \mathbf{NAE}_j(x_1, x_2, .., x_j) : [k]^j \to \{0, 1\}$, where $\mathbf{NAE}_j$ is defined to be equal to 0 if and only if $x_1 = x_2 = \ldots = x_j$ and is equal to 1 otherwise. So ß is $\{\mathbf{NAE}_j \mid j \leq h\}$. In addition, there are k terminal vertices $v_1, v_2, \ldots, v_k$ such that, for each vertex v_i, its candidate label list is $L_{v_i} = \{i\}$.

Given a ß-CSP instance, we can write down the following linear program. We remark that this LP can be seen as a generalization of the Earthmover LP from [14]; this turns out to be the "right formulation" to generalize, rather than the geometric CKR relaxation for Multiway Cut.

The Local Distribution LP. There are variables $x_{e,\alpha}$ for every hyperedge $e \in E$ and assignment $\alpha \in [k]^{|e|}$, and variables $x_{v,j}$ for every vertex $v \in V$ and $j \in [k]$. The objective function is of the following form:

$$LP(\mathcal{I}) = \min_{x_{e,\alpha}, x_{v,i}} \sum_e w_e \cdot \sum_{\alpha \in [k]^{|e|}} x_{e,\alpha} \cdot \Psi_e(\alpha)$$

under the constraint that, for every $v \in V$,

$$\sum_{i=1}^{k} x_{v,i} = 1$$

where $0 \leq x_{v,i} \leq 1$ for every $v, i \in L_v$ and $x_{v,i'} = 0$ for every i', v such that $i' \notin L_v$. We also have constraints that for every edge $e = (v_1, v_2, \ldots, v_j)$ and $i \in S$,

$$x_{v_i,k} = \sum_{\alpha_i = k} x_{e,\alpha}.$$

where $0 \leq x_{v_i,\alpha} \leq 1$ for every $x_{v_i,\alpha}$.

To see why this is a relaxation, one should think of $x_{v,i}$ as the probability of assigning label i to vertex v. For every edge $e = (v_1, v_2, \ldots, v_j)$ and labeling $\alpha = (l_1, l_2, \ldots, l_j)$ for the vertices of e, $x_{e,\alpha}$ is the probability of labeling the vertices of e according to α (that is, vertex v_i receives label l_i). For every edge e, we define $\mathcal{P}_e$ as the distribution that assigns probability $x_{e,\alpha}$ to each $\alpha \in [k]^{|e|}$.

4.1 The Min-CSP hardness theorem

THEOREM 4.1. *Suppose we have a ß-CSP instance $\mathcal{I}(V, E, k, \{L_v | v \in V\}, h)$ with LP value $LP(\mathcal{I}) = c$, optimum value $\mathrm{OPT}(\mathcal{I}) = s$, and ß contains the Not-Equal predicate function $\mathbf{NAE}_2(x, y) = \mathbf{1}[x \neq y]$. Then assuming the Unique Games Conjecture, it is NP-hard to distinguish the following two cases for a given instance of ß-CSP:*

- $\mathrm{OPT} \geq s - o(1)$*;*
- $\mathrm{OPT} \leq c + o(1)$.

This implies that it is NP-hard to get an approximation ration that is better than s/c. Such a hardness result holds even for instances that contain only the $\mathbf{NAE}_2$ functions and predicate functions and candidate lists from $\mathcal{I}$.

The proof appears in Section 4.2. To complement the UNIQUE GAMES-hardness result based on the integrality gap $\alpha_ß$ of the Local Distribution LP, we give a polynomial-time algorithm based on the above LP with a $(1+\epsilon)\alpha_ß$-approximation guarantee. The running time is dependent on ϵ: it is $k^{(k^2/\epsilon)^k} \cdot \mathrm{poly}(|V|, k)$. This result is inspired by the rounding technique of [11];

THEOREM 4.2. *For any ß-CSP with worst-case integrality gap $\alpha_ß$, there is an $(\alpha_ß(1 + \epsilon))$-approximation algorithm for any $\epsilon > 0$.*

The proof appears in Section 4.3.

As a corollary of Theorem 4.1, we obtain a hardness result for Hypergraph Multiway Cut. This follows from a known integrality gap example for Hypergraph Multiway Cut, reformulated for the Local Distribution LP.

COROLLARY 4.1. *The* HYPERGRAPH-MC *problem with k terminals is* UNIQUE GAMES*-hard to approximate within $(2 - \frac{2}{k} - \epsilon)$ for any fixed $\epsilon > 0$. This holds even for k-uniform hypergraphs.*

Proof. Assuming the correctness of Theorem 4.1, consider the following HYPERGRAPH-MC instance $\mathcal{H}_k$: there are $k(k+1)/2$ vertices indexed by (i, j) for $1 \leq i \leq j \leq k$. We have k hyperedges of size k: for every $i \in [k]$ edge e_i contains all the vertices indexed by (i_1, i_2) such that either $i_1 = i$ or $i_2 = i$. We define the k terminals as those vertices indexed by (i, i) for every $i \in [k]$ with its label fixed to be i.

We claim that $\mathrm{OPT}(\mathcal{H}_k) \geq (k - 1)$; i.e., there is no assignment with cost 0 on more than two edges. Without loss of generality, suppose the optimal solution has cost 0 on edge e_1; i.e, assign label 1 to every vertex indexed by $(1, i)$ for $i \in [k]$. Then we can not have cost 0 for any of the remaining $(k - 1)$-hyperedges since to satisfy e_i, we need $(1, i)$ to be labelled by i.

On the other hand, $LP(\mathcal{H}_k) \leq k/2$. Following is a fractional solution: for every vertex $v = (i, j)$, $x_{v,i} = 1/2$ and $x_{v,j} = 1/2$. All the other $x_{v,k'}$ is 0 for $k' \neq i$ and $k' \neq j$. For every edge e_i with its vertices ordered by $(1, i), (2, i), \ldots, (i, i), (i, i + 1), \ldots (i, k)$, we have $x_{e,(i,i,\ldots,i)} = 1/2$ and $x_{e,(1,2,\ldots,k)} = 1/2$.

If we want to apply Theorem 4.1 with $I = H_k$, we also need to verify that we indeed a hardness result for

Hypergraph-MC instance with k terminals. This is true as the instance our hardness result hold have the same type of constraints and $\mathbf{NAE}_2$ which is already in the predicate function class of Hypergraph-MC. Also the candidate lists is the same as the one in H_k. It is possible that we have multiple vertices being the same terminal; i.e., with the same candidate list $\{i\}$ for some $i \in [k]$. One can merge them by create a new vertex of candidate list $\{i\}$ and replace all the other terminals of the same type by this new vertex in every hyper-edge (predicate function). Therefore, applying Theorem 4.1, we get that it is Unique Games-hard to approximate Hypergraph-MC with k terminals beyond the factor $\frac{k-1}{k/2} = 2 - \frac{2}{k}$.

4.2 Proof of Theorem 4.1 Here we give a proof of our general hardness result, which is an extension of the technique of [14].

We describe a reduction from the Unique Games problem to ß-CSP where ß contains the $\mathbf{NAE}_2$ predicate.

Definition 4.2. *A* unique-games *instance $\mathcal{U}(V, E, \{\pi_{u,v}\}_{(u,v)\in E}, R)$ consists of a regular graph $G(V, E)$ and for each edge $e = (u, v) \in E$ a permutation $\pi_{uv} : [R] \mapsto [R]$. The algorithm needs to assign a label from $[R]$ to every vertex $u \in V$. For a given labeling $L : V \mapsto [R]$, an edge $e = (u, v) \in E$ is said to be satisfied if $L(v) = \pi_{u,v}(L(u))$. The goal is to find a labeling which satisfies the maximum fraction of edges possible.*

Conjecture 4.1. (The Unique Games Conjecture) *Given a* unique-games *instance $\mathcal{U}$, let $OPT(\mathcal{U})$ denote the maximum fraction of edges satisfiable by any assignment. Then, for every $\epsilon > 0$, there is a large enough label size n so that, given a unique games instance $\mathcal{U}\left(V, E, [n], \{\pi_{u,v}\}_{(u,v)\in E}\right)$, it is NP-hard to distinguish between the following two cases:*

- *$OPT(\mathcal{U}) \geq 1 - \epsilon$;*
- *$OPT(\mathcal{U}) \leq \epsilon$.*

Now we are ready to prove Theorem 4.1. Without loss of generality, let us assume that for the integrality gap instance, the sum of the weights over all the hyperedges is normalized to 1; thus we can view the weight w_e as the probability on edge e. The reduction takes the Unique Games instance $\mathcal{U}(V, E, \{\pi_e \mid e \in E\}, R)$ and it maps it to a ß-CSP instance $\mathcal{I}(V', E, k, \{L_v | v \in V'\}, h)$. Suppose that $|V'| \leq m$ and $\text{OPT}(\mathcal{I}') = c$ and $LP(\mathcal{I}') = s$. The reduction produces a new ß-CSP instance $\mathcal{M}$ with the following properties:

1. Completeness property: if $\text{OPT}(\mathcal{U}) \geq 1 - \epsilon$, then $\text{OPT}(\mathcal{M}) \leq c \cdot \eta + O(\eta/m)$;
2. Soundness property: if $\text{OPT}(\mathcal{U}) \leq \epsilon$, then $\text{OPT}(\mathcal{M}) \geq s \cdot \eta - O(\eta/m)$.

Here $\mathcal{M}$'s vertices are $V \times V' \times [k]^R$. For every vertex (v, v', x), its candidate label list is L_v. The predicate functions that appear in $\mathcal{M}$ are the $\mathbf{NAE}_2$ function as well as all the predicate functions that appear in $\mathcal{I}$. The sum of the weights in $\mathcal{M}$ is equal to one. If we can obtain such a reduction, then we can multiply all the weights of the instance by $1/\eta$ and choose a large enough m to get Theorem 4.1.

Below is a description of the reduction with the following choice of parameters: $\epsilon \leq 1/m^{80}, \eta = 1/m^{39}, \delta = 1/m^{40}$ and $k^{10} \leq m$. We will specify the values of m and ϵ later.

Reduction from Unique Games to ß-CSP.

The weight w_e of a constraint Ψ_e is the probability it is generated by the following:

- (Edge test) With probability η, we pick an edge $e = (v_1', v_2', \dots, v_j')$ from E'. Then we randomly pick a vertex v from V and randomly pick j of its neighbors $v_1, v_2, \dots, v_j$. We generate $x^1, x^2 \dots, x^j \in [k]^R$ according to $\mathcal{P}_e^R$. Let us use the notation $\pi(x)$ to indicate a vector in $[k]^R$ such that $\pi(x)_i = x_{\pi(i)}$. Output a constraint between $(v_1, v_1', \pi_{v_1,v}(x^1)), (v_2, v_2', \pi_{v_2,v}(x^2))$ $\dots, (v_j, v_j', \pi_{v_j,v}(x^j))$ with the predicate function Ψ_e.
- (Vertex test) With probability $(1 - \eta)$, we pick a vertex v from V and two of its neighbors v_1, v_2. We randomly pick a vertex v' from V'. Then we generate $x, y \in \{-1, 1\}^n$ with $1 - \delta$ correlation and output a constraint between $(v_1, v', \pi_{v_1,v}(x))$ and $(v_2, v', \pi_{v_2,v}(y))$ with the predicate function $\mathbf{NAE}_2$.

A function $f : V \times V' \times [k]^R \to [k]$ corresponds to an assignment of the variables. Let us use $\mathbf{Val}(f)$ to denote the cost of f and $\mathbf{Val}_{edge}$ and $\mathbf{Val}_{vertex}$ to denote the cost of f on the edge test and the vertex test. Also let us use the notation $f_{v,v'} : [k]^R \to [k]$ to denote the restriction of f for a fixed pair v and v', where $v \in V$ and $v' \in V'$.

We know that $\mathbf{Val}(f) = (1 - \eta) \cdot \mathbf{Val}_{vertex}(f) + \eta \cdot \mathbf{Val}_{edge}(f)$. Now we prove the completeness and soundness property of the reduction.

Lemma 4.1. (completeness) *If* $\mathrm{OPT}(\mathcal{U}) \geq 1-\epsilon$, *then* $\mathrm{OPT}(\mathcal{M}) \leq c/m^{39} + O(1/m^{40})$.

Proof. Suppose that a labelling $\Lambda : V \to [R]$ that satisfies $(1-\epsilon)$-fraction of the edges in the Unique Games instance, we can take the Dictator labelling function $f_{v,v'}(x) = x_{\Lambda(v)}$ for any $x \in [k]^R$. Notice that $f_{v,v'}(x) \in L'_v$ as when $(x^1, x^2, \ldots, x^j)$ is generated by $\mathcal{P}_e$ for $e = (v_1, v_2, \ldots, v_j)$, we must have that for $x^i_t \in L_{v_t}$ for every $t \in [R]$.

Since $(1-\epsilon)$ fraction of the edges in the Unique Games can be satisfied, by a averaging argument and the regularity of the graph, we know that for at least $(1-\sqrt{\epsilon})$-fraction of the vertices, we have that $(1-\sqrt{\epsilon})$-fraction of its neighbors is satisfied by Λ. Therefore, for the vertex test, with probability at most $1-3\sqrt{\epsilon}$ we have $\pi_{v_1,v}(\Lambda(v_1)) = \pi_{v_2,v}(\Lambda(v_2))$. When this is the case, the cost of the dictator labelling is at most δ. Overall, we have that the vertex test cost is at most $(1-3\sqrt{\epsilon})\cdot\delta+3\sqrt{\epsilon}$

As for the edge test, we know that with probability at least $1-(m+1)\sqrt{\epsilon}$, we have $\pi_{v_1,v}(\Lambda(v_1)) = \pi_{v_2,v}(\Lambda(v_2)) \ldots = \pi_{v_j,v}(\Lambda(v_j))$. When this is the case, since $x^1_{\pi_{v_1,v}(\Lambda(v_1))}, \ldots, x^j_{\pi_{v_j,v}(\Lambda(v_j))} \sim \mathcal{P}_e$, we get

$$\mathop{\mathbf{E}}_{e,v_1,v_2,\ldots,v_j}[\Psi_e(x^1_{\pi_{v_1,v}(\Lambda(v_1))}, \ldots, x^j_{\pi_{v_j,v}(\Lambda(v_k))})] = \sum_e w_e \sum_{\alpha\in[k]^e} p_{e,\alpha}\Psi_e(\alpha) = c$$

Therefore, the cost of the edge test is at most $(m+1)\sqrt{\epsilon} + (1-(m+1)\sqrt{\epsilon})\cdot c$.

Overall, the cost for the dictator function is then

$$\eta \cdot \big((m+1)\sqrt{\epsilon} + (1-(m+1)\sqrt{\epsilon})\cdot c\big) + (1-\eta)\cdot((1-3\sqrt{\epsilon})\cdot\delta + 3\sqrt{\epsilon}).$$

By the choice of parameters, we know that $\mathbf{Val}(f) = c/m^{39} + O(1/m^{40})$.

It remains to prove the following soundness property.

Lemma 4.2. (soundness) *If* $\mathrm{OPT}(\mathcal{U}) \leq \epsilon$, *then* $\mathrm{OPT}(\mathcal{M}) \geq s/m^{39} - O(1/m^{40})$.

The proof of above lemma is more involved. Below is an overview of our proof strategy which is analogous to the proof of [14].

Overview of the proof We consider a solution of $\mathcal{M}$ described by the labeling functions $f_{v,v'} : [k]^R \to [k]$ for each $v \in V, v' \in V'$.

- We classify all the functions $f_{v,v'}$ into three categories: (i) dictator function; (ii) constant function; (iii) low influence non-constant function.
- In order to achieve a low cost, most $f_{v,v'}$ functions must be either close to a dictator or to a constant function. Otherwise, the cost of the vertex test is overwhelming, due to Theorem 4.3 (Majority Is Stablest).
- If, for some v', most of the $f_{v,v'}$ functions are close to a dictator, then we can assign a label to each v from the influential coordinates of $f_{v,v'}$. Such an assignment gives a good solution to the Unique Games which leads to a contradiction. Therefore the fraction of $f_{v,v'}$ functions that are close to a dictator cannot be too large.
- In conclusion, most $f_{v,v'}$ functions are close to constant functions, and f can be interpreted as an integral solution of the original instance. Therefore the total cost corresponds to the integral optimum of the original instance.

Below are more details of the proof. In Section 4.2.1, we first review some standard definitions from the analysis of boolean function. Then in Section 4.2.2, we show how to use these analytical tools to prove Lemma 4.2.

4.2.1 Harmonic Analysis and Unique Games Conjecture

We will be considering functions of the form $f : [q]^n \to \mathbb{R}^k$, where $q, n, k \in \mathbb{N}$. The set of all functions $f : [q]^n \to \mathbb{R}^k$ forms an inner product space with inner product

$$\langle f, g\rangle = \mathop{\mathbf{E}}_{x\sim[q]^n}[\langle f(x), g(x)\rangle];$$

here we mean that x is uniformly random and the $\langle\cdot,\cdot\rangle$ inside the expectation is the usual inner product in $\mathbb{R}^k$. We also write $\|f\| = \sqrt{\langle f, f\rangle}$ as usual.

Definition 4.3. *For $x, y \in [q]^n$, we say that y is ρ-correlated with x if we generate each $y_i = x_i$ with probability ρ and set it to be random in $[q]$ with probability $1-\rho$.*

For $0 \leq \rho \leq 1$, we define T_ρ to be the linear operator on this inner product space given by

$$T_\rho f(x) = \mathop{\mathbf{E}}_y[f(y)],$$

where y is a random string in $[q]^n$ which is ρ-correlated to x. We define the noise stability of f at ρ *to be*

$$\mathbf{Stab}_\rho[f] = \langle f, T_\rho f\rangle.$$

Definition 4.4. *For $i \in [n]$, we define the* influence of i on $f : [q]^n \to \mathbb{R}^k$ *to be*

$$\mathbf{Inf}_i[f] = \mathop{\mathbf{E}}_{x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n\sim[q]}\left[\mathop{\mathbf{Var}}_{x_i\sim[q]}[f(x)]\right],$$

where $\mathbf{Var}[f]$ *is defined to be* $\mathbf{E}[\|f\|^2] - \|\mathbf{E}[f]\|^2$. *More generally, for* $0 \le \delta \le 1$ *we define the* δ-noisy-influence of i on f *to be*

$$\mathbf{Inf}_i^{(1-\delta)}[f] = \mathbf{Inf}_i[T_{1-\delta}f].$$

One may observe that

$$\mathbf{Inf}_i^{(1-\delta)}[f] = \sum_{j=1}^{k}\mathbf{Inf}_i^{(1-\delta)}[f_j],$$

where $f_j : [q]^n \to \mathbb{R}$ denotes the jth-coordinate output function of f.

We need the following easy "convexity of noisy-influences" fact:

Fact 4.1. *Let $f^{(1)},\ldots,f^{(t)}$ be a collection of functions $[q]^n \to \mathbb{R}^k$. Then*

$$\mathbf{Inf}_i^{(1-\delta)}\left[\mathop{\mathrm{avg}}_{j\in[t]}\left\{f^{(j)}\right\}\right] \le \mathop{\mathrm{avg}}_{j\in[t]}\left\{\mathbf{Inf}_i^{(1-\delta)}[f^{(j)}]\right\}.$$

Here for any $c_1, c_2, \ldots c_t \in \mathbb{R}$ (or $\mathbb{R}^m$), we use the notation $\mathrm{avg}(c_1,\ldots,c_t)$ *to denote their average:*

$$\frac{\sum_{j=1}^{t} c_j}{t}.$$

For (randomized) functions with discrete output, $f' : [q]^n \to [k]$, we can view it as a function defined as $f : [q]^n \to \Delta_k$ where Δ_k is the $(k-1)$-dimensional standard simplex. The i-th coordinate indicates the probability that the function f' output i. Following fact is also well known.

Fact 4.2. *For any $f : [q]^n \to \Delta_k$, $\sum_{i=1}^{n}\mathbf{Inf}_i^{1-\delta} f \le 1/\delta$.*

One tool we need is the Majority Is Stablest Theorem from [15]. We state here a slightly modified version [17] using a small "noisy-influence" assumption rather than a small "low degree influence" assumption.

Theorem 4.3. *(Majority is the stablest) Suppose $f : [q]^n \to [0,1]$ has $\mathbf{Inf}_i^{(1-\delta)}[f] \le \tau$, then*

$$\mathbf{Stab}_{1-\delta}[f] \le \Gamma_{1-\delta}(\mu) + err(\tau, q, \delta)$$

where for any fixed τ and q, $err(\tau, q, \eta)$ goes to 0 for when τ goes to 0. Here the quantity $\Gamma_{1-\delta}(\mu)$ is defined to be $\mathbf{Pr}[x, y \le t]$ when (x, y) are joint standard Gaussian with covariance $1-\delta$ and t is defined by $\mathbf{Pr}[x \le t] = \mu$.

We will use the following estimate for $T_{1-\delta}$.

Lemma 4.3. *If $\delta^c \le \mu \le 1-\delta^c$ for some $0 < c < 1$, then we have that $\mu - T_{1-\delta}(\mu) \ge \Omega(\delta^{0.5+2c})$.*

A stronger version of the Lemma appears in [14]; for completeness we give a proof in Appendix A.

4.2.2 Proof of Lemma 4.2

Proof. We will prove this by contradiction. Assuming that there is an assignment f such that $\mathbf{Val}(f) \le s/m^{39} - O(1/m^{40})$. We know the cost of the vertex test is:

$$(4.1)\quad \mathbf{Val}_{vertex}(f) = \mathop{\mathbf{E}}_{v_1,v_2,v',x,y}[\mathbf{1}(f_{v_1,v'}(\pi_{v_1,v}(x)) \neq f_{v_2,v'}(\pi_{v_2,v}(y)))]$$

$$= 1 - \sum_{i=1}^{k}\mathop{\mathbf{E}}_{v_1,v_2,v',x,y}[f^i_{v_1,v'}(\pi_{v_1,v}(x))\cdot f^i_{v_2,v'}(\pi_{v_2,v}(y))]$$

$$= 1 - \sum_{i=1}^{k}\mathop{\mathbf{E}}_{v,v,x,y}\left[\mathop{\mathbf{E}}_{v_1\sim v}[f^i_{v_1,v'}(\pi_{v_1,v}(x))]\mathop{\mathbf{E}}_{v_2\sim v}[f^i_{v_2,v'}(\pi_{v_2,v}(y))]\right]$$

where we write $f^i_{v,v'}$ as the indicator function of whether $f_{v,v'} = i$. Here $v_i \sim v$ means v_i is a random neighbor of v.

If we define $g^i_{v,v'}$ as $\mathbf{E}_{u\sim v}[f^i_{u,v'}(\pi_{u,v}(x)]$ we have that

$$(4.2)\quad (4.1) = 1 - \sum_{i=1}^{k}\mathop{\mathbf{E}}_{v,v'x,y}\left[g^i_{v,v'}(x)\cdot g^i_{v,v'}(y))\right]$$

$$= \mathop{\mathbf{E}}_{v,v'}[1 - \sum_{i=1}^{k}\mathbf{Stab}_{1-\delta}(g^i_{v,v'})] \le \Theta(1/m^{39})$$

For $g_{v,v'} = (g^1_{v,v'}, g^2_{v,v'}, \ldots, g^k_{v,v'})$ and it is easy to verify $g_{v,v'} \in \Delta_k$

For every $g_{v,v'}$, we classify it into the following three categories : 1) dictator function: $g_{v,v'}$ has a coordinate with its δ-noisy-influence influence above τ (with τ being specified later) 2) constant function: there exists some i with $\mathbf{E}[g^i_{v,v'}] \ge 1-\delta^{0.1}$ 3) all the other $g_{v,v'}$ not in category 1 and 2.

The main idea of the remaining proof is to show that for every $v' \in V'$, in order to bound the cost on vertex test bounded by $O(1/m^{39})$, we must have at least $1-O(1/m^2)$ fraction of v, $g_{v,v'}$ is in category (2). Then we argue that this will have a big cost on the edge test.

1. Bound the fraction of category 3: If for a given v', an α-fraction $g_{v,v'}$ is in category 3; i.e.,

$$\max_{i\in R}\mathbf{Inf}_i^{1-\delta} g_{v,v'} \le \tau$$

and

$$\max_{i\in[k]} \mathop{\mathbf{E}}_{x\in[k]^R}[g^i_{v,v'}(x)] \le 1-\delta^{0.1}.$$

Then by Theorem 4.3 and setting τ to make $err(\tau,k,\delta) \le 1/m^{30}$, we have that

$$\sum_{i=1}^{k} \mathbf{Stab}(g^i_{v,v'}) \le \sum_{i=1}^{k} T_{1-\delta}(\mu^i_{v,v'}) + k\cdot err(\tau,q,\delta).$$

where $\mu^i_{v,v'} = \mathbf{E}[g^i_{v,v'}]$. It is easy to check that $\sum \mu^i_{v,v'} = 1$. Using Lemma 4.3 and notice that $\delta^{0.1} \le 1/k \le \max_{i\in[k]} \mu^i_{v,v'} \le 1-\delta^{0.1}$. Suppose that $\mu^t_{v,v'}$ has the biggest value among $\mu^1_{v,v'}, \mu^2_{v,v'}, \dots, \mu^k_{v,v'}$, then

$$\begin{aligned} 1 - \sum_{i=1}^{k} &\mathbf{Stab}_{1-\delta}(g^i_{v,v'}) \\ &\ge \mu^t_{v,v'} - T_{1-\delta}(\mu^t_{v,v'}) - k\cdot 1/m^{30} \\ &\ge \Omega(\delta^{0.7}) - 0(1/m^{29}) = \Omega(1/m^{28}). \end{aligned}$$

Overall, we know each particular v' is picked with probability at least $1/m$ and if α fraction of the $g_{v',v}$ is in category 3, this will have a cost of

$$\alpha \cdot 1/m \cdot \Omega(1/m^{28}) \le O(1/m^{39})$$

and hence $\alpha \le O(1/m^2)$.

2. Bound the fraction of category 1: For a given $v' \in V'$, if for β for fraction of the $v \in V$ such that $g_{v,v'}$ has a coordinate with γ-noisy influence above τ. Then consider the following labelling for the Unique Games instance: for each $v \in V$, we can just assign randomly a label from the candidate list as follows:

   $$\begin{aligned} L_v = \{i \mid i\in\mathbb{R}, &\mathbf{Inf}^{1-\delta}_i g_{v,v'} \ge \tau\} \\ &\cup \{i \mid i\in\mathbb{R}, \mathbf{Inf}^{1-\delta}_i f_{v,v'} \ge \tau/2\}. \end{aligned}$$

   Then by Lemma 4.2, since the sum of the influence over all coordinate is $O(1/\delta)$ (either for $\mathbf{Inf}^{1-\delta}_i f^i_{v,v'}$ or $\mathbf{Inf}^{1-\delta}_i g_{v,v'}$), we know $|L_v| = O(1/\delta\tau)$.

   By Fact 4.1, for every i, $\mathbf{Inf}^{1-\delta}_i g_{v,v'} \ge \mathbf{E}_{u\sim v,x}[\mathbf{Inf}^{1-\delta}_i f_{u,v'}(\pi_{v,u}(x))(\pi_{v,u}(x))] \ge \tau$, by an averaging argument, at least $\tau/2$ fraction of its neighbors u has a coordinate j such that $\mathbf{Inf}^{1-\delta}_j f_{u,v'} \ge \tau/2$ and $\pi_{v,u}(i) = j$. Therefore, such a labelling will satisfies at least $\Omega(\beta\tau^3\delta^2)$-fraction of the edges of the Unique Games. Since we can take ϵ to be arbitrarily small, say $\min(\tau^3 \cdot 1/m^2 \cdot \delta^2, 1/m^{80})$ Then we can also conclude that $\beta = O(1/m^2)$.

Therefore, we can assume that for every $v' \in \mathcal{I}$, $1 - O(1/m^2)$ fraction of the $g_{v,v'}$ is in category 2; i.e., there exist some i such that $\mathbf{E}[g^i_{v,v'}] = \mathbf{E}_{u\sim v}[f^i_{u,v'}] \ge 1-\delta^{0.1} = 1-1/m^4$. Below we will show for such f, it have a large cost on the edge test.

By an average argument, for at least $1-1/m^2$ fraction of $v's$ neighbor u, we have that $\mathbf{E}_{u,v'}[f^i_{u,v}] \ge 1-1/m^2$. Overall by the regularity of the graph of Unique Games, we know that for every $v'$and for at least $1-O(1/m^2)$ fraction of the vertices $u \in V$, we have that $\mathbf{E}_{u,v'}[f^i_{u,v'}] \ge 1-1/m^2$.

By a union bound, we have that $1-O(1/m)$ fraction of the $v \in V$, $\max_i \mathbf{E}[f^i_{v,v'}] \ge 1-1/m^2$ for every $v' \in V'$. Let us call these v good.

Given (v,v') fixed, let us just consider the labelling of (v,v',x) by $\arg\max_i \mathbf{E}[f^i_{v,v'}]$ for every x. This labelling has a cost at least s (conditioned on (v,v') fixed) as it assigns a label just depending only on (v,v') which can be viewed as the cost of a labelling for the gap instance $\mathcal{I}'$.

Given a good v, f $(v_1, v_1', \pi_{v_1,v}(x^1)), (v_2, v_2', \pi_{v_2,v}(x^2)), \dots, (v_j, v_j', \pi_j(x^j))$ in the reduction has the same cost as the above labelling with probability $(1-1/m)$. Therefore, we have that $\mathbf{Val}(f) \ge \eta(1-O(1/m))^2 \cdot s = \eta\cdot s - O(\eta/m)$ which leads to a contradiction.

4.3 Proof of Theorem 4.2

Proof. For any instance $\mathcal{I}(V,E,k,\{L_v|v\in V\},h)$, given the solution of Local Distribution LP, we round the fractional solution $x_v = (x_{v,1}, x_{v,2}, \dots, x_{v,k})$ of v by the algorithm in Figure 1.

Let us first verify that it is indeed in polynomial time. Without loss of generality, let us assume that 1 is a multiple of ϵ and therefore $1/k$ is a multiple of ϵ/k^2. We know x''_{v,i_v} is also a multiple of ϵ/k^2 as $\sum_{i=1}^{k} x''_{v,i} = 1$ and every other $x''_{v,i}$ is a multiple of ϵ/k^2.

Since each coordinate of x''_v is a multiple of ϵ/k^2, there are at most $(k^2/\epsilon)^k$ possible x''_v. The total number of vertices in V' is at most $(k^2/\epsilon)^k$. The running time of the fourth step of the algorithm is $k^{(k^2/\epsilon)^k} \cdot poly(|V|)$. For other steps in the algorithm, it is clearly polynomial with respect to k and $|V|$.

Below, we prove that above rounding algorithm always achieve a ratio of $\alpha_{\ss}(1+\epsilon)$. Let us denote the cost of above algorithm's assignment by $ALG(\mathcal{I})$. Since the algorithm find an optimal assignment on $\mathcal{I}'$,

Figure 1: LP rounding algorithm

LP Rounding Algorithm

1. for every $v \in V, i \in [k]$, set each $x'_{v,i}$ to be the largest multiple of ϵ/k^2 less than $x_{v,i}$; i.e,
$$x'_{v,i} = \max_{n \in \mathbb{Z}, n \geq 0, n \cdot \epsilon/k^2 \leq x_{v,i}} n \cdot \epsilon/k^2.$$
2. for every $v \in V, i \in [k]$ and $x'_v = (x'_{v,1}, x'_{v,2}, \ldots, x'_{v,k})$, suppose that $i_v = \arg\max_{i \in [k]} x'_{v_i}$, set
$$x''_{v,i_v} = x'_{v,i_v} + (1 - \sum_{i=1}^{k} x'_{v,i})$$
3. Define a new ß-CSP instance $\mathcal{I}'(V', E', k, \{L_v | v \in V'\}, h)$ by merging all variables $v \in V$ with the same x''_v; i.e., replace all the variables with the same x''_v by a single variable in $\mathcal{I}$.
4. Search the optimal solution in $\mathcal{I}'$ by trying all the $|k|^{|V'|}$ possible assignments and get a labelling $L : V' \to [k]$.
5. Project the optimal assignment in $\mathcal{I}'$ back to $\mathcal{I}$: for every $v' \in \mathcal{I}$ that is merged by $v_1, v_2, \ldots, v_j$ in V, assign the same label for $v_1, v_2, \ldots, v_j$ as $L(v')$.

it has the same cost as the assignment on $\mathcal{I}$. We have that $ALG(\mathcal{I}) = \text{OPT}(\mathcal{I}')$. We will show $\text{OPT}(I') \leq \alpha_{\text{ß}}(1+\epsilon) \cdot LP(\mathcal{I})$.

First notice that if we change the equality constraint in the LP to
$$x_{v_i,k} \leq \sum_{\alpha_i = k} x_{e,\alpha}$$
and
$$\sum_{i=1}^{k} x_{v,i} \geq 1.$$
Above modification gives the same LP.

Let us define $x_V = \{x_v \, | v \in V\}$. For a fixed x_V, we define $LP(\mathcal{I}, x_V)$ to be the optimum of the LP with x_V fixed and the we only optimize over variables of the form $x_{e,\alpha}$. We also ignore the constraint that $\sum_{i=1}^{k} x_{v,i} \geq 1$, allowing any $x_V \in \mathbb{R}^{k|V|}$ in the definition of $LP(\mathcal{I}, x_V)$.

Below are the two simple properties of $LP(\mathcal{I}, x_V)$:

- Linearity: for any $\alpha > 0$ $LP(\mathcal{I}, \alpha \cdot x_V) = \alpha \cdot LP(\mathcal{I}, x_V)$. To see the linearity, notice that scaling the solution of the optimization problem with x_V by a factor of α will give a solution for the optimization with $\alpha \cdot x_V$.
- Monotonicity: if for two x_V, y_V, we have that $x_{v,i} \leq y_{v,i}$ for every $i \in [k], v \in V$, then $LP(\mathcal{I}, x_V) \leq LP(\mathcal{I}, y_V)$. To prove the monotonicity, one just need to notice that the feasible solution for the LP with x_V fixed is also a feasible solution given y_V fixed.

We know
$$LP(\mathcal{I}, x_V) \geq LP(\mathcal{I}, x'_V).$$
by the monotonicity. Since $x'_{v,i_v} \geq 1/k$ and $1 - \sum x'_{v_i} \leq \epsilon/k^2 \cdot k \leq \epsilon/k$, we have $x''_{v,i_v} \leq (1+\epsilon) \cdot x'_{v,i_v}$ and for $i \neq i_v$, $x'_{v,i} = x''_{v,i}$.

Therefore,
$$\begin{aligned} LP(\mathcal{I}, x''_V) &\leq LP(\mathcal{I}, (1+\epsilon)x'_V) && (\textit{monotonicity}) \\ &= (1+\epsilon)LP(\mathcal{I}, x'_V) && (\textit{linearity}) \\ &\leq (1+\epsilon)LP(\mathcal{I}, x_V) && (\textit{monotonicity}) \\ &= (1+\epsilon)LP(\mathcal{I}). \end{aligned}$$

Lastly, using the definition of the integrality gap on $\mathcal{I}$, we know that $\text{OPT}(I') \leq \alpha_{\text{ß}} \cdot LP(I')$. Overall,
$$\begin{aligned} ALG(I) &= \text{OPT}(\mathcal{I}') \\ &\leq \alpha_{\text{ß}} \cdot LP(I') \\ &= \alpha_{\text{ß}} \cdot LP(I, x''_V) \\ &\leq \alpha_{\text{ß}}(1+\epsilon) \cdot LP(\mathcal{I}) \end{aligned}$$

5 Lovász versus Local Distribution

Here we compare the Local Distribution LP and the Lovász LP, for problems that can be phrased both as a Min-CSP and as a Submodular Multiway Partition problem. That is, we consider a ß-Min-CSP where each predicate $\Psi_e \in ß$ is of the form

$$\Psi_e(i_1, \ldots, i_l) = \sum_{i=1}^{k} f_e(\{j : i_j = i\})$$

and each $f_e : 2^{[l]} \to \mathbb{R}_+$ is a submodular function. (Recall that Graph-MC, Hypergraph-MC and Node-Wt-MC fall in this category.) Then in both relaxations, we replace the labeling $x_v \in [k]$ by variables $y_{v,i} \geq 0, i \in [k]$ such that $\sum_{i=1}^{k} y_{v,i} = 1$. We also impose the list-coloring constraints $y_{v,i} = 0$ for $i \notin L_v$ in both cases. The objective function is defined in different ways in the two relaxations.

The Lovász relaxation. We minimize $\sum_{e \in E} \sum_{i=1}^{k} \hat{f}_e(y_{v,i} : v \in e)$, where $\hat{f}$ is the Lovász extension of f, $\hat{f}(\mathbf{y}) = \mathbf{E}_{\theta \in [0,1]}[f(A_{\mathbf{y}}(\theta))]$ where $A_{\mathbf{y}}(\theta) = \{i : y_i > \theta\}$.

The Local Distribution LP. We minimize $\sum_{e \in E} \sum_{i_1, \ldots, i_l \in [k]} y_{e,i_1,\ldots,i_k} \Psi_e(i_1, \ldots, i_l)$ subject to the consistency constraints $\sum_{i_1,\ldots,i_{j-1},i_{j+1},\ldots,i_l} y_{e,i_1,\ldots,i_l} = y_{v,i_j}$ where v is the j-th vertex of e.

Lemma 5.1. *The value of the Lovász relaxation is at most the value of the Local Distribution LP.*

Proof. Given a fractional solution of the Local Distribution LP, with variables $y_{v,i}$ for vertices $v \in V$ and $y_{e,i_1,\ldots,i_l}$ for hyperedges $e \in E$, each hyperedge contributes $\sum_{i_1,\ldots,i_l \in [k]} y_{e,i_1,\ldots,i_k} \Psi_e(i_1, \ldots, i_l)$, where $\Psi_e(i_1, \ldots, i_l) = \sum_{i=1}^{k} f_e(\{j : i_j = i\})$. In other words, each assignment $(i_1, \ldots, i_l)$ contributes $\sum_{i=1}^{k} f(S_i)$ where S_i is the set of coordinates labeled by i. Aggregating all the contributions of a given set S, we can define $z_{e,i,S}$ as the sum of $y_{e,i_1,\ldots,i_l}$ over all choices where the coordinates labeled by i are exactly S. Then, the contribution of hyperedge e becomes $\sum_{i=1}^{k} \sum_{S \subseteq [l]} z_{e,i,S} f(S)$. Moreover, the variables $z_{e,i,S}$ are consistent with $y_{v,i}$ for $v \in e$ in the sense that $y_{v,i} = \sum_{S: v \in S} z_{e,i,S}$.

On the other hand, the contribution of a hyperedge e in the Lovász relaxation is given by the Lovász extension $\sum_{i=1}^{k} \hat{f}_e(y_{v,i} : v \in e)$. It is known that the Lovász extension is the convex closure of a submodular function, i.e. $\hat{f}_e(y_{v,i} : v \in e)$ is the minimum possible value of $\sum_{S \subseteq [l]} z_{e,i,S} f(S)$ subject to the consistency constraints $y_{v,i} = \sum_{S: v \in S} z_{e,i,S}$, and $z_{e,i,S} \geq 0$. Therefore, the contribution of e to the Lovász relaxation is always at most the contribution in the Local Distribution LP.

This means that the Local Distribution LP is potentially a *tighter* relaxation, since its optimum is closer to the integer optimum, and also its fractional solution carries potentially more information than a fractional solution of the Lovász relaxation. However, in some cases the two LPs coincide: We remark that for the Graph-MC problem, both relaxations are known to be equivalent to the CKR relaxation (see [3] for a discussion of the Lovász relaxation, and [14] for a discussion of the "earthmover LP", identical to our Local Distribution LP). Our results also imply that for the Hypergraph-MC problem, both relaxations have the same integrality gap, $2 - 2/k$. However, for certain problems the Local Distribution LP can be strictly more powerful than the Lovász relaxation.

Hypergraph Multiway Partition. *Given a hypergraph $H = (V, E)$ and k terminals $t_1, \ldots, t_k \in V$, find a partition $(A_1, \ldots, A_k)$ of the vertices, so that $\sum_{i=1}^{k} f(A_i)$ is minimized, where $f(A)$ is the number of hyperedges cut by $(A, \bar{A})$.*

This is a special case of Sub-MP-Sym, because f here is the cut function in a hypergraph, a symmetric submodular function. The difference from Hypergraph Multiway Cut is that in Hypergraph Multiway Partition, each cut hyperedge contributes 1 for each terminal that gets assigned some of its vertices (unlike in Hypergraph-MC where such a hyperedge contributes only 1 overall).

Lemma 5.2. *There is an instance of Hypergraph Multiway Partition where the Lovász relaxation has a strictly lower optimum than the Local Distribution LP.*

Proof. The instance is the following: We have a ground set $V = \{t_1, t_2, t_3, t_4, t_5, a_{12}, a_{23}, a_{34}, a_{45}, a_{51}\}$. The hyperedges are $\{t_1, t_2, a_{12}\}$, $\{t_2, t_3, a_{23}\}$, $\{t_3, t_4, a_{34}\}$, $\{t_4, t_5, a_{45}\}$, $\{t_5, t_1, a_{51}\}$ with unit weight, and $\{a_{12}, a_{23}, a_{34}, a_{45}, a_{51}\}$ with weight $\epsilon = 0.001$.

The idea is that fractionally, each non-terminal $a_{i,i+1}$ must be assigned half to t_i and half to t_{i+1}, otherwise the cost of cutting the triple-edges is prohibitive. Then, the cost of the 5-edge $\{a_{12}, a_{23}, a_{34}, a_{45}, a_{51}\}$ is strictly higher in the Local Distribution LP, since there is no good distribution consistent with the vertex variables (while the Lovász relaxation is not sensitive to this). We verified by an LP solver that the two LPs have indeed distinct optimal values.

6 The equivalence of integrality gap and symmetry gap

In this section, we prove that for any ß-CSP (specified by allowed predicates and candidate lists), the worst-case integrality gap of its Local Distribution LP and the worst-case symmetry gap of its multilinear relaxation are the same. As we have seen, if we consider a ß-Min-CSP problem and ß includes the Not-Equal predicate, then the integrality gap of the Local Distribution LP implies a matching inapproximability result assuming UGC. Similarly, if the objective function can be viewed as $\sum_{i=1}^{k} f(S_i)$ where f is submodular and S_i is the set of vertices labeled i, then the symmetry gap implies an inapproximability result for this "submodular generalization" of the ß-CSP problem. In fact, the two hardness threshold often coincide as we have seen in the case of Hypergraph-MC and Sub-MP, where they are equal to $2 - 2/k$. Here, we show that it is not a coincidence that the two hardness thresholds are the same.

We recall that the symmetry gap of a Min-CSP is the ratio s/c between the optimal fractional solution c and the optimal symmetric fractional solution s of the *multilinear relaxation.* The objective function in the multilinear relaxation is $F(\mathbf{x}) = \mathbf{E}[f(\hat{\mathbf{x}})]$, where $\hat{\mathbf{x}}$ is an integer solution obtained by independently labeling each vertex v by i with probability $x_{v,i}$. The notion of symmetry here is that $\mathcal{I}'$ on a ground set V' is invariant under a group $\mathcal{G}$ of permutations of V'. A fractional solution is symmetric if for any permutation $\sigma \in \mathcal{G}$ and $v' \in V'$, v' and $\sigma(v')$ have the same fractional assignment.

The following theorem states that an integrality gap instance can be converted into a symmetry gap instance.

Theorem 6.1. *For any ß-CSP instance $\mathcal{I}(V, E, k, L_v, h)$ whose Local Distribution LP has optimum $LP(\mathcal{I}) = c$, and the integer optimum is $\mathrm{OPT}(\mathcal{I}) = s$, there is a symmetric ß-CSP instance $\mathcal{I}'(V', E', k, L_v', h)$ whose symmetry gap is s/c.*

Proof. Given an optimal solution $\mathbf{x}$ of the Local Distribution LP for instance $\mathcal{I}$, without loss of generality, let us assume that all the variables in the solution have a rational value and there exists some M such that the values of all the variables, i.e, $x_{v,i}$ and $x_{e,\alpha}$, in the LP solution become integers if multiplied by M. For every vertex $v \in V$, we define

$$\begin{aligned} S_v \;=\; & \{y \in [k]^M : \text{for every } i \in [k]; \\ & x_{v,i} \text{ fraction of } y\text{'s coordinates have value } i\}. \end{aligned}$$

In other words, S_v is the collection of strings in $[k]^M$ such that the portion of appearances of i in each string is exactly $x_{v,i}$.

We define a new instance on a ground set $V' = \{(v, y) : v \in V, y \in S_v\}$. For every vertex (v, y) in V', its candidate list is L_v, the candidate list of vertex v in instance $\mathcal{I}$. Given an edge $e = (v_1, v_2, \ldots, v_{|e|}) \in E$ and its local distribution $\mathcal{P}_e$ over $[k]^{|e|}$ (implied by the fractional solution $x_{e,\alpha}$), we call $(y^1, y^2, \ldots, y^{|e|}) \in ([k]^M)^{|e|}$ "consistent with e" if for every $\alpha \in [k]^{|e|}$, the fraction of $i \in [M]$ such that $(y_i^1, y_i^2, \ldots, y_i^{|e|}) = \alpha$ is exactly $x_{e,\alpha}$. It is easy to check (using the LP consistency constraints) that each y^i must be in S_{v_i}. We define S_e to be the collection of all possible $(y^1, y^2, \ldots, y^{|e|})$ that are consistent with edge e.

For every edge $e = (v_1, v_2, \ldots, v_{|e|})$ and every $(y^1, y^2, \ldots, y^{|e|})$ consistent with e, we add a constraint Ψ_e over $(v_1, y^1), \ldots, (v_j, y^{|e|})$ with the same predicate function as edge e. We assign equal weights to all these copies of constraint Ψ_e, so that their total weight is equal to w_e.

Let $\mathcal{G}$ be the group of all permutations $\pi : [M] \to [M]$. For any $\pi \in \mathcal{G}$, we also use the notation $\pi(y)$ to indicate $(y_{\pi(1)}, y_{\pi(2)}, \ldots, y_{\pi(M)})$ for any $y \in [k]^M$. Let us also think of $\mathcal{G}$ as a permutation on V': for any $(v, y) \in V'$, we map it to $(v, \pi(y))$. Then it is easy to check that $\mathcal{I}'$ is invariant under any permutation $\sigma \in \mathcal{G}$. Also, for any $(v, y), (v, y') \in V'$, since y and y' contain the same number of occurences for each $i \in [M]$ (consistent with $x_{v,i}$), we know there exists some $\pi \in \mathcal{G}$ such that $\pi(y) = y'$. Therefore, a fractional solution of $\mathcal{I}'$ is symmetric with respect to $\mathcal{G}$ if and only if (v, y) has the same fractional assignment as (v, y'), for all $y, y' \in S_v$.

Let us also assume that both $\mathcal{I}$ and $\mathcal{I}'$ are normalized with all their weights summing to 1; i.e, there is a distribution over the constraints with the probability being the weights. Essentially, $\mathcal{I}'$ is constructed by randomly picking a constraint $\Psi_e(v_1, v_2, \ldots, v_{|e|})$ in $\mathcal{I}$ and then randomly picking $(y^1, y^2, \ldots, y^{|e|})$ in S_e. Then we add a constraint Ψ_e on $(v_1, y^1), (v_2, y^2), \ldots, (v_{|e|}, y^{|e|})$ in $\mathcal{I}'$.

Below, we prove that assuming there is a gap of s versus c between integer and fractional solutions for the Local Distribution LP of $\mathcal{I}$, there is a gap of at least s versus c between the best symmetric fractional solution and the best asymmetric fractional solution for the multilinear relaxation of $\mathcal{I}'$.

Some (asymmetric) solution of $\mathcal{I}'$ has cost at most c: Consider the following (random) assignment for $\mathcal{I}'$: we first choose a random $i \in [M]$ and then assign to each (v, y) the label y_i. We know $y_i \in L_v$, as it is a valid assignment for v by the definition of S_v. Such an assignment has expected cost

$$\begin{aligned} \mathop{\mathbf{E}}_{i\in[M],e,y^1,\ldots,y^{|e|}}[\Psi_e(y_i^1, \ldots, y_i^{|e|})] &= \mathop{\mathbf{E}}_{e}[\sum_{\alpha\in[k]^{|e|}} x_{e,\alpha}\Psi(\alpha)] \\ &= LP(\mathcal{I}) = c. \end{aligned}$$

Therefore there is also some deterministic assignment of cost at most c. Our assignment is in fact integral, so the multilinear relaxation does not play a role here.

Every symmetric fractional solution of $\mathcal{I}'$ has cost at least s: We know that a symmetric solution gives the same fractional assignment $x_v = (x_{v,1}, x_{v,2}, \ldots, x_{v,k})$ to vertex (v, y) for every $y \in S_v$. Also by definition of the multilinear relaxation, the value of this fractional solution is equal to the expected cost of independently choosing the label of each (v, y) from the distribution $\mathcal{P}_v$, which means i with probability $x_{v,i}$. Therefore, given a symmetric solution of $\mathcal{I}'$ specified by $\{x_v \mid v \in V\}$, it has cost

(6.3)
$$\mathop{\mathbf{E}}_{e=(v_1,\ldots,v_{|e|})\in E,(y^1,\ldots,y^{|e|})\in S_e,l_i\sim\mathcal{P}_{v_i}}[\Psi_e(l_1, l_2, \ldots, l_{|e|})]$$

Since $l_1, l_2, \ldots, l_{|e|}$ is independent of $(y^1, y^2, \ldots, y^{|e|})$, we have

$$(6.3) = \mathop{\mathbf{E}}_{e=(v_1,v_2,\ldots,v_{|e|})\in E,l_1,l_2,\ldots,l_{|e|}}[\Psi_e(l_1, l_2, \ldots, l_{|e|})]$$

where l_i is chosen independently from the distribution $\mathcal{P}_{v_i}$ for the respective vertex $v_i \in e$. This is exactly the cost on the original instance $\mathcal{I}$ if we independently label each v by sampling from distribution $\mathcal{P}_v$. Since the integer optimum of $\mathcal{I}$ is at least s, we have that (6.3) is also at least s.

In the other direction, we have the following.

Theorem 6.2. *For any symmetric ß-CSP instance $\mathcal{I}(V, E, k, L_v, h)$ whose multilinear relaxation has symmetry gap γ, and for any $\epsilon > 0$ there is a ß-CSP instance $\mathcal{I}'(V', E', k, L'_v, h)$ whose Local Distribution LP has integrality gap at least $(1 - \epsilon)\gamma$.*

Proof. Assume that $\mathcal{I}$ is an instance symmetric under a group of permutations $\mathcal{G}$ on the vertices V. For each $v \in V$, we denote by $\omega(v)$ the *orbit* of v, $\omega(v) = \{\sigma(v) : \sigma \in \mathcal{G}\}$. We produce a new instance $\mathcal{I}'$ by *folding* the orbits, i.e. we identify all the elements in a given orbit.

First, let us assume for simplicity that no constraint (edge) in $\mathcal{I}$ operates on more than one variable from each orbit. Then, we just fold each orbit $\omega(v)$ into a single vertex. I.e., $V' = \{\omega(v) : v \in V\}$. (We abuse notation here and use $\omega(v)$ also to denote the vertex of V' corresponding to the orbit $\omega(v)$.) We also define E' to be edges corresponding one-to-one to the edges in E, with the same predicates; i.e. each constraint $\Psi_e(x_1, \ldots, x_{|e|})$ in $\mathcal{I}$ becomes $\Psi_e(\omega(x_1), \ldots, \omega(x_{|e|}))$ in $\mathcal{I}'$. The candidate list for $\omega(v)$ is identical to the candidate list of v (and hence also to the candidate list for any other $w \in \omega(v)$, by the symmetry condition on $\mathcal{I}$).

Now consider an optimal fractional solution of the multilinear relaxation of $\mathcal{I}$, with variables $x_{v,i}$. We define a fractional solution of the Local Distribution LP for $\mathcal{I}'$, by setting $x_{\omega(v),i} = \frac{1}{|\omega(v)|}\sum_{w\in\omega(v)} x_{w,i}$. We claim that the edge variables $x_{e,\alpha}$ can be assigned values consistent with $x_{\omega(v),i}$ in such a way that the objective value of the Local Distribution LP $\mathcal{I}'$ is equal to the value of the multilinear relaxation for $\mathcal{I}$. This is because the value of the multilinear relaxation is obtained by independently labeling each vertex v by i with probability $x_{v,i}$. We can then define $x_{e,\alpha}$ as the probability that edge e is labeled α under this random labeling. This shows that the optimum of the Local Distribution LP for $\mathcal{I}'$ is at most the optimum of the multilinear relaxation of $\mathcal{I}$ (if we have a minimization problem; otherwise all inequalities are reversed).

Now, consider an integer solution of the instance $\mathcal{I}'$, i.e. a labeling of the orbits $\omega(v)$ by labels in $[k]$. This induces a natural symmetric labeling of the instance $\mathcal{I}$, where all the vertices in each orbit receive the same label. The values of the two labelings of $\mathcal{I}$ and $\mathcal{I}'$ are the same. (Both labelings are integer, so the multilinear relaxation does not play a role here.) This shows that the symmetric optimum of the multilinear relaxation of $\mathcal{I}$ is at most the integer optimum of $\mathcal{I}'$. Together with the previous paragraph, we obtain that the gap between integer and fractional solutions of the Local Distribution LP of $\mathcal{I}'$ is at least the gap between symmetric and asymmetric solutions of the multilinear relaxation of $\mathcal{I}$. This completes the proof in the special case where no predicate takes two variables from the same orbit.

Now, let us consider the general case, in which a constraint of $\mathcal{I}$ can contain multiple variables from the same orbit. Note that the proof above doesn't work here, because it would produce predicates in $\mathcal{I}'$ that operate on the same variable multiple times, which we do not allow. (As an instructive example, consider a symmetric instance on two elements $\{1, 2\}$ with the constraint $l_1 \neq l_2$, which serves as a starting point in proving the hardness of $(1/2 + \epsilon)$-approximating the maximum of a nonnegative submodular function. This instance is symmetric with respect to switching 1 and 2, and hence the folded instance would contain only one element and the constraint $l'_1 \neq l'_1$, which is not a meaningful integrality gap instance.)

Instead, we replace each orbit by a large cluster of identical elements, and we produce copies of each constraint of $\mathcal{I}$, using distinct elements from the respective clusters. As an example, consider the 2-element instance above. We would first fold the elements $\{1, 2\}$ into one, and then replace it by a large cluster C of identical elements. The original constraint $l_1 \neq l_2$ will be replaced by the same constraint for every pair of distinct elements $\{i, j\} \subset C$. In other words, the new instance is a Max Cut instance on a complete graph and we consider the Local Distribution LP for this instance. The symmetry gap of the original instance is 2, and the integrality gap of the Local Distribution LP is arbitrarily close to 2 (for $|C| \to \infty$).

Now let us describe the general reduction. For each orbit $\omega(v)$ of $\mathcal{I}$, we produce a disjoint cluster of elements $C_{\omega(v)}$. The candidate list for each vertex in $C_{\omega(v)}$ is the same as the candidate list for any vertex in $\omega(v)$ (which must be the same by the symmetry assumption for $\mathcal{I}$). The ground set of $\mathcal{I}'$ is $V' = \bigcup_{v \in V} C_{\omega(v)}$. For each edge $e = (v_1, \dots, v_{|e|})$ of $\mathcal{I}$, we produce a number of copies by considering all possible edges $e' = (v'_1, \dots, v'_{|e|})$ where $v'_i \in C_{\omega(v_i)}$ and $v'_1, \dots, v'_{|e|}$ are distinct. We use the same predicate, $\Psi_{e'} = \Psi_e$. The edge weights $w_{e'}$ are defined so that they are equal and add up to w_e. Let us assume that $\mathcal{I}, \mathcal{I}'$ are minimization problems. Let c be value of the optimal solution of the multilinear relaxation of $\mathcal{I}$, and let s be the the value of the optimal symmetric solution of the multilinear relaxation of $\mathcal{I}$.

The fractional optimum of $\mathcal{I}'$ is at most c: Let $x_{v,i}$ be an solution of the multilinear relaxation of value c. We define a fractional solution of the Local Distribution LP for $\mathcal{I}'$ to be equal to $x_{v',i} = \frac{1}{|\omega(v)|} \sum_{w \in \omega(v)} x_{w,i}$ for each $v' \in C_{\omega(v)}$. Similar to the construction in the simpler case above, we define the edge variables $x_{e',\alpha}$ to simulate the independent rounding that defines the value of the multilinear extension of $\mathcal{I}$; specifically, we define $x_{e',\alpha}$ for $e' = (\omega(v_1), \omega(v_2), \dots, \omega(v_{|e|}))$ to be the probability that $e = (v_1, \dots, v_{|e|})$ receives assignment α, where e is the edge that produced e' in our reduction. By construction, the value of this fractional solution for $\mathcal{I}'$ is equal to the value of the multilinear relaxation for $\mathcal{I}$ which is c.

The integer optimum of $\mathcal{I}'$ is at least s: Here, we consider any integer solution of $\mathcal{I}'$ and we produce a symmetric fractional solution for $\mathcal{I}'$. If the integer solution for $\mathcal{I}'$ is denoted by $l(v')$, we consider each cluster $C_{\omega(v)}$ in $\mathcal{I}'$ and we define $x_{v,i}$ to be the fraction of vertices in $C_{\omega(v)}$ that are labeled i. Note that by definition, $x_{v,i} = x_{w,i}$ for all $w \in \omega(v)$ and hence $x_{v,i}$ depends only on the orbit of v. This means that $x_{v,i}$ is a symmetric fractional solution. Also, it respects the candidate lists of $\mathcal{I}$ because $\mathcal{I}'$ has the same candidate lists and hence the fractional solution uses only the allowed labels for each vertex v. The value of this symmetric fractional solution is given by the multilinear extension, i.e. the expected value of a random labeling where each label is chosen independently with probabilities $x_{v,i}$. The final observation is that in the limit as $|C_{\omega(v)}| \to \infty$, this is very close to the value of the integer solution $l(v')$ for $\mathcal{I}'$: this is because the edges e' in $\mathcal{I}'$ are defined by all possible choices of distinct elements in the respective clusters. In the limit, this is arbitrarily close to sampling independently random vertices from the respective clusters, which is what the multilinear relaxation corresponds to. Therefore, the optimal symmetric fractional solution to $\mathcal{I}$ cannot be worse than the integer optimum of $\mathcal{I}'$ in the limit.

In conclusion, the integrality gap of the Local Distribution LP for $\mathcal{I}'$ can be made arbitrarily close to the symmetry gap of $\mathcal{I}$, s/c.

References

[1] Gruia Calinescu, Howard J. Karloff, and Yuval Rabani. An improved approximation algorithm for multiway cut. *Journal of Computer and System Sciences*, 60(3):564–574, 2000. Preliminary version in STOC 1998.

[2] Chandra Chekuri and Alina Ene. Approximation algorithms for submodular multiway partition. In *IEEE Symposium on Foundations of Computer Science (FOCS)*, 2011.

[3] Chandra Chekuri and Alina Ene. Submodular cost allocation problem and applications. In *Automata, Languages and Programming (ICALP)*, pages 354–366, 2011. A longer version is available on the arXiv, abs/1105.2040.

[4] Chandra Chekuri, Jan Vondrák, and Rico Zenklusen. Submodular function maximization via the multilinear extension and contention resolution schemes. In *ACM Symposium on Theory of Computing (STOC)*, pages 783–792, 2011.

[5] Elias Dahlhaus, David S. Johnson, Christos H. Papadimitriou, Paul D. Seymour, and Mihalis Yannakakis. The complexity of multiterminal cuts. *SIAM Journal on Computing*, 23(4):864–894, 1994. Preliminary version in STOC 1992.

[6] Shahar Dobzinski and Jan Vondrák. From query complexity to computational complexity. In *ACM Symposium on Theory of Computing (STOC)*, 2012.

[7] Naveen Garg, Vijay V. Vazirani, and Mihalis Yannakakis. Multiway cuts in node weighted graphs. *Journal of Algorithms*, 50(1):49–61, 2004. Preliminary version in ICALP 1994.

[8] Satoru Iwata and Kiyohito Nagano. Submodular function minimization under covering constraints. In *IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 671–680, 2009.

[9] David R. Karger, Philip N. Klein, Clifford Stein, Mikkel Thorup, and Neal E. Young. Rounding algorithms for a geometric embedding of minimum multiway cut. *Mathematics of Operations Research*, 29(3):436–461, 2004. Preliminary version in STOC 1999.

[10] Ariel Kulik, Hadas Shachnai, and Tami Tamir. Maximizing submodular set functions subject to multiple

linear constraints. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 545–554, 2009.

[11] Amit Kumar, Rajsekar Manokaran, Madhur Tulsiani, and Nisheeth K. Vishnoi. On LP-based approximability for strict CSPs. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1560–1573, 2011.

[12] Jon Lee, Vahab S. Mirrokni, Viswanath Nagarajan, and Maxim Sviridenko. Maximizing nonmonotone submodular functions under matroid or knapsack constraints. *SIAM Journal on Discrete Mathematics*, 23(4):2053–2078, 2010.

[13] László Lovász. Submodular functions and convexity. In *Mathematical programming: the state of the art*, pages 235–257. Springer, 1983.

[14] Rajsekar Manokaran, Joseph S. Naor, Prasad Raghavendra, and Roy Schwartz. SDP gaps and UGC hardness for multiway cut, 0-extension, and metric labeling. In *ACM Symposium on Theory of Computing (STOC)*, pages 11–20, 2008.

[15] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. In *IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 21–30, 2005.

[16] Kazumasa Okumoto, Takuro Fukunaga, and Hiroshi Nagamochi. Divide-and-conquer algorithms for partitioning hypergraphs and submodular systems. *Algorithmica*, pages 1–20, 2010. Preliminary version in ISAAC 2009.

[17] Prasad Raghavendra. *Approximating NP-hard problems: efficient algorithms and their limits.* PhD thesis, University of Washington, 2009.

[18] Jan Vondrák. Optimal approximation for the submodular welfare problem in the value oracle model. In *ACM Symposium on Theory of Computing (STOC)*, pages 67–74, 2008.

[19] Jan Vondrák. Symmetry and approximability of submodular maximization problems. In *IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 651–670, 2009.

[20] Jan Vondrák. Symmetry and approximability of submodular maximization problems. Available on the arXiv, abs/1110.4860, 2011.

[21] Liang Zhao, Hiroshi Nagamochi, and Toshihide Ibaraki. Greedy splitting algorithms for approximating multiway partition problems. *Mathematical Programming*, 102(1):167–183, 2005.

A Proof of Lemma 4.3

Proof. Suppose that we have $\mathbf{Pr}_{x\sim N(0,1)}(x \le t) = \mu$ for some t then

$$\begin{aligned} T_{1-\delta}(\mu) &= \mathbf{Pr}(x \le t, y \le t) \\ &= \mathbf{Pr}((1-\delta)x + \sqrt{2\delta-\delta^2}z \le t | x \le t)\cdot \mu \end{aligned} \tag{A.1}$$

where x, z is independently generated from Gaussian Distribution. For $t' = t - \sqrt{2\delta-\delta^2}$, we have that

$$\begin{aligned} &\mathbf{Pr}((1-\sqrt{\delta})x + \sqrt{2\delta-\delta^2}z \le t | x \le t) = \mathbf{Pr}(x \ge t' | \, x \le t) \\ &\cdot \mathbf{Pr}\left((1-\delta)x + \sqrt{2\delta-\delta^2}z \le t \mid t' \le x \le t)\right) + \\ &\mathbf{Pr}(x \le t' \mid x \le t)\cdot \mathbf{Pr}\left((1-\delta)x + \sqrt{2\delta-\delta^2}z \le t | x \le t'\right) \\ &\le \mathbf{Pr}(x \ge t' | \, x \le t)\cdot \mathbf{Pr}(z < \frac{t-(1-\delta)t'}{\sqrt{2\delta-\delta^2}}) + \\ &\mathbf{Pr}(x \le t' | x \le t) \\ &\le \mathbf{Pr}(x \ge t' | \, x \le t)\cdot \mathbf{Pr}(z < \sqrt{\delta}t + (1-\delta)) \\ &+ \mathbf{Pr}(x \le t' | x \le t) \\ &= p_1 \cdot (1-p_2) + (1-p_1) = 1 - p_1 p_2 \end{aligned}$$

where $p_1 = \mathbf{Pr}(x \ge t' | \, x \le t)$ and $p_2 = \mathbf{Pr}(z \ge \sqrt{\delta}t + (1-\delta))$. We now show both p_1 and p_2 are not too small. Following fact is well know for Gaussian Distribution.

Fact A.1. *For every $s > 0$*

$$\frac{e^{-s^2/2}\cdot s}{\sqrt{2\pi}(s^2+1)} \le \mathop{\mathbf{Pr}}_{g\sim N(0,1)}(g < s) \le \frac{e^{-s^2/2}}{\sqrt{2\pi}s}$$

From above fact, we know that $\frac{e^{-t^2/2}}{|t|\sqrt{2\pi}} \ge \gamma$ and therefore, $|t| \le \sqrt{2\log(1/\gamma)} = O(\sqrt{\log(1/\delta)})$ and this will imply $p_2 \ge \mathbf{Pr}(z \ge 1 + \sqrt{\delta}\cdot(\sqrt{\log 1/\delta}) \ge \mathbf{Pr}(z \ge 2) \ge 0.1$. As for p_1, we have that

$$p_1 = \mathbf{Pr}(t' < x < t | x < t) \ge \mathbf{Pr}(t' \le x \le t) \tag{A.2}$$

$$\ge \frac{e^{-(\sqrt{2\log(1/\gamma)}+\sqrt{\delta})^2/2}}{\sqrt{2\pi}}\cdot\sqrt{2\delta-\delta^2} \ge \Omega(\gamma\cdot\sqrt{\delta}) \tag{A.3}$$

Overall, we have that $\mu - T_{1-\delta}(\mu) \ge \Omega(\delta^{0.5+2c})$ and this finishes the proof of Lemma 4.3.