

2 CSPs All Are Approximable Within a Constant Differential Factor

Jean-François Culus, Sophie Toulouse

► **To cite this version:**

Jean-François Culus, Sophie Toulouse. 2 CSPs All Are Approximable Within a Constant Differential Factor. International Symposium on Combinatorial Optimisation, Apr 2018, Marrakesh, Morocco. hal-02350144

HAL Id: hal-02350144

<https://hal.univ-antilles.fr/hal-02350144>

Submitted on 5 Nov 2019

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

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

2 CSPs All Are Approximable Within a Constant Differential Factor

Jean-François Culus¹ and Sophie Toulouse² ✉

¹ CEREGMIA, Université des Antilles, France
jean-francois-culus@espe-martinique.fr

² LIPN (UMR CNRS 7030), Institut Galilée, Université Paris 13, France
sophie.toulouse@lipn.univ-paris13.fr

Abstract. Only a few facts are known regarding the approximability of optimization CSPs with respect to the differential approximation measure, which compares the gain of a given solution over the worst solution value to the instance diameter. Notably, the question whether k CSP- q is approximable within any constant factor is open in case when $q \geq 3$ or $k \geq 4$. Given three integers $k \geq 2$, $p \geq k$ and $q > p$, we analyse the expansion of a precise reduction from k CSP- q to k CSP- p . We introduce a family of combinatorial designs from which we deduce a lower bound of $1/(q - p + k/2)^k$ for this expansion. When $p = k = 2$, this implies together with the result of Nesterov as regards 2 CSP-2 [?] that for all constant integers $q \geq 2$, 2 CSP- q is approximable within factor $(2 - \pi/2)/(q - 1)^2$.

Keywords: Differential approximation · Optimization constraint satisfaction problems · Combinatorial designs

1 Introduction

1.1 Optimization Constraint Satisfaction Problems

Thereafter, given a positive integer N , we use notation $[N]$ to refer to the discrete interval $\{1, \dots, N\}$. Optimization *Constraint Satisfaction Problems* (CSPs) over an alphabet Σ consider a set $\{x_1, \dots, x_n\}$ of variables and a set $\{C_1, \dots, C_m\}$ of constraints, where the variables have domain Σ , and the constraints consist of (non constant) predicates applied to tuples of variables. Most often, a positive weight is associated with each constraint C_i . The goal is then to optimize over Σ^n an objective function of the form

$$\sum_{i=1}^m w_i C_i = \sum_{i=1}^m w_i P_i(x_{J_i}) = \sum_{i=1}^m w_i P_i(x_{i_1}, \dots, x_{i_{k_i}})$$

where for all $i \in [m]$, $P_i : \Sigma^{k_i} \rightarrow \{0, 1\}$, $J_i = (i_1, \dots, i_{k_i}) \subseteq [n]$, and $w_i > 0$.

CSPs are specified by the functions that may occur in their constraints. For example, the *Satisfiability Problem* (Sat) is the boolean CSP where constraints are disjunctive clauses. In Lin- q , variables have domain \mathbb{Z}_q , and a constraint is a linear equation modulo q . In this paper, given two universal constant integers

$q, k \geq 2$, we restrict to the case when Σ has size q (we thus assume *w.l.o.g.* that $\Sigma = [q]$), and each constraint depends on at most k of the variables. Moreover, we allow the functions that occur in the constraints to take rational values. We denote by kCSP-q the corresponding unconstrained optimization problem.³

kCSP-q most often becomes harder as k or q grows. On the one hand, given two integers h, k with $k > h \geq 2$, hCSP-q is a special case of kCSP-q . On the other hand, given two integers p, q with $q > p \geq 2$, any surjective map from $[q]$ to $[p]$ can be used to convert a function on $[p]^k$ to a function on $[q]^k$. The alphabet size more accurately has a logarithmic impact on the constraint arity. Namely, let $\kappa = \lceil \log_p q \rceil$. Then any surjective map from $[p]^\kappa$ to $[q]$ similarly allows to interpret a function on $[q]^k$ as a function on $[p]^{\kappa k}$. As kCSP-q is **NP-hard** even in case when $q = k = 2$ [?], a major issue as regards kCSP-q consists in charactering its *approximation degree*.

1.2 Their Differential Approximability

Approximation algorithms aim at providing within polynomial time solution values proved to be relatively close to the optimum value, where the proximity to the optimum value is evaluated according to a specific measure. In this paper, we consider the *differential approximation measure* (see [?] for an introduction). Given an instance I of an optimization (generalized) CSP, we denote by $v(I, \cdot)$ its objective function, by $\text{opt}(I)$ and $\text{wor}(I)$ respectively the optimum and the worst solution values on I . Then the *differential ratio* reached at a given solution x on I is the ratio:

$$\frac{v(I, x) - \text{wor}(I)}{\text{opt}(I) - \text{wor}(I)}$$

Given $\rho \in]0, 1]$, we say that x is ρ -approximate when this ratio is at least ρ . Equivalently, x is ρ -approximate if either the goal on I is to maximize and $v(I, x) \geq \rho \times \text{opt}(I) + (1 - \rho) \times \text{wor}(I)$, or the goal on I is to minimize and $v(I, x) \leq \rho \times \text{opt}(I) + (1 - \rho) \times \text{wor}(I)$. Given an optimization CSP Π , a polynomial time algorithm \mathcal{A} is a ρ -approximation algorithm for Π if \mathcal{A} returns on every instance of Π a solution with differential ratio at least ρ . We say that Π is *approximable within factor ρ* whenever such an algorithm exists.

Only a few facts are known regarding the approximability of kCSP-q within a constant differential factor. The restrictions of **MinSat** and **MaxSat** to unweighted instances (*i.e.*, to instances in which weights w_i all are equal to 1) are not approximable within any constant factor unless **P = NP** [?]. For 2CSP-2 , the semidefinite programming based algorithm of Goemans and Williamson [?] produces solutions with expected differential ratio at least $2 - \pi/2 > 0.429$ [?], and the algorithm can be derandomized [?]. In between, the question whether kCSP-q is approximable within any constant factor remains open for all integers $k \geq 2, q \geq 2$ such that $q \geq 3$ or $k \geq 4$.

³ CSPs in which constraints can take non-boolean values are often called *generalized CSPs* in the literature. However, a constraint $P_i(x_{J_i})$ can be rewritten as the combination $\sum_{v \in \Sigma^{k_i}} P_i(v) \times (x_{J_i} = v)$ of constraints. Thus when k and q are universal constants, we may indifferently consider functions with boolean or rational values.

1.3 Outline

A common way to exhibit new approximability lower and upper bounds for a given optimization problem consists in *reducing* to or from another optimization problem for which approximability bounds are known:

Definition 1. *An optimization CSP Π D -reduces to another optimization CSP Π' if one can derive from any ρ -approximation algorithm \mathcal{A} for Π' a $\gamma \times \rho$ -approximation algorithm for Π , where γ is some positive quantity. When this occurs, γ is called the expansion of the reduction.*

For example, it is not hard to see that the approximability bound of $2 - \pi/2$ for 2 CSP-2 somehow extends by reduction to 3 CSP-2:

Proposition 1. *If E2 Lin-2 is approximable within factor ρ , then 3 CSP-2 is approximable within factor $\rho/2$.*

Proof (sketch). Let I be an instance of 3 CSP-2. First, interpret I as an instance J of 3 Lin-2 use the discrete Fourier transform. Then consider instance H of 2 Lin-2 that is obtained from J by removing its constraints with odd arity.

Any pair $x \in \{0, 1\}^n$, $\bar{x} = (1 - x_1, \dots, 1 - x_n)$ of solutions satisfy $v(I, x) = v(J, x) + N(I)$ and $v(J, x) + v(J, \bar{x}) = 2v(H, x) + N'(I)$, where $N(I)$ and $N'(I)$ are two quantities that only depend on I . If the goal on I is to maximize, then the latter equality notably implies inequalities $2 \times \text{opt}(H) \geq \text{opt}(J) + \text{wor}(J) - N'(I)$ and $2 \times \text{wor}(H) \geq 2 \times \text{wor}(J) - N'(I)$. Otherwise, the converse inequalities hold. In both cases, we deduce from these inequalities that, provided that x is ρ -approximate on H , a solution x or \bar{x} that performs the best objective value on J is $\rho/2$ -approximate on J and thus, on I . \square

We address the question whether given three integers $k \geq 2$, $p \geq 2$ and $q > p$, k CSP- q D -reduces to k CSP- p . We more specifically study the expansion of a specific reduction that basically consists in restricting a given instance I of k CSP- q to solution sets of the form T^n where T is a p -cardinality subset of $[q]$. The analysis we propose, though, requires to restrict to the case when $p \geq k$.

In the next section, we introduce a family of combinatorial designs (Definition 2) that provides some lower bound $\gamma(q, p, k)$ for the expansion of the reduction (Theorem 1). Section 3 is then dedicated to the exhibition of such combinatorial designs. Using a recursive construction for the case when $p = k$ (Theorem 2), we show that $1/(q - p + k/2)^k$ is a proper lower bound of for $\gamma(q, p, k)$. Therefore, we obtain the following conditional approximation result:

Corollary 1. *Given any three integers $k \geq 2$, $p \geq k$ and $q > p$, k CSP- q D -reduces to k CSP- p with an expansion of $1/(q - p + k/2)^k$ on the approximation guarantee. The reduction involves $O(q^p)$ instances of k CSP- p .*

The question whether k CSP- q is approximable within some constant factor consequently reduces to the consideration of integers k, q such that $k \geq q \geq 2$. Most importantly, it follows from Nesterov's result as regards 2 CSP-2 (we more specifically refer to Theorem 2.3, Theorem 3.3 and Corollary 3.4 of [?]) that for all integers $q \geq 2$, 2 CSP- q is approximable within a constant factor:

Corollary 2. *For all integers $q \geq 2$, $2\text{CSP-}q$ is approximable within factor $(2 - \pi/2)/(q - 1)^2$.*

2 Reducing the Alphabet Size of a CSP Instance

Let $k \geq 2$, $p \geq 2$ and $q > p$ be three integers, and I be an instance of $k\text{CSP-}q$. Thereafter, $\mathcal{P}_p([q])$ refers to the set of the p -cardinality subsets of $[q]$.

Given $S = (S_1, \dots, S_n) \in \mathcal{P}_p([q])^n$, any set $\{\pi_{S,j} : S_j \rightarrow [p] \mid j \in [n]\}$ of bijections allows to interpret the restriction of I to solutions in S as an instance of $k\text{CSP-}p$. A natural way to derive approximate solutions on I from a hypothetical algorithm \mathcal{A} for $k\text{CSP-}p$ therefore consists in restricting I to solution subsets $S \in \mathcal{P}_p([q])^n$. The *standard approximation measure* evaluates the performance of a given solution x by the ratio $v(I, x)/\text{opt}(I)$. In [?], the authors study the randomized reduction that consists in picking $S \in \mathcal{P}_p([q])^n$ uniformly at random, and then using \mathcal{A} to compute a solution $x \in S$. They show that, provided that the goal on I is to maximize and I is such that $w_i P_i \geq 0$, $i \in [m]$, the expected value of $\max_{x \in S} \{v(I, x)\}$ over all $S \in \mathcal{P}_p([q])^n$ is at least $(p/q)^k \times \text{opt}(I)$. Accordingly, picking $S \in \mathcal{P}_p([q])^n$ uniformly at random, and then computing a solution $x \in S$ with value at least $\rho \times \max_{x \in S} \{v(I, x)\}$, one gets a solution with expected value at least $(p/q)^k \rho \times \text{opt}(I)$. The reduction therefore preserves the expected standard ratio up to a multiplicative factor of $(p/q)^k$.

Given $T \in \mathcal{P}_p([q])$, let $I(T)$ refer to the restriction of I to solution set T^n . Then similarly to [?], we analyse the reduction that consists in using \mathcal{A} to compute for all $T \in \mathcal{P}_p([q])$ an approximate solution $x(T)$ on $I(T)$, and then returning a solution $x(T)$ that performs the best objective value.

2.1 Seeking Symmetries in the Solution Set

We may assume *w.l.o.g.* that the goal on I is to maximize (as otherwise, rather than I , consider instance I' obtained from I by substituting for each constraint $P_i(x_{J_i})$ constraint $-P_i(x_{J_i})$). When this occurs, the extremal values on I and on subinstances $I(T)$ trivially satisfy:

$$\text{opt}(I) \geq \text{opt}(I(T)) \geq \text{wor}(I(T)) \geq \text{wor}(I), \quad T \in \mathcal{P}_p([q]) \quad (1)$$

Now assume that for all $T \in \mathcal{P}_p([q])$, we are given a solution $x(T) \in T^n$ that is ρ -approximate on $I(T)$. Then for all $T^* \in \mathcal{P}_p([q])$, we have:

$$\begin{aligned} \max_{T \in \mathcal{P}_p([q])} \{v(I, x(T))\} &\geq v(I, x(T^*)) \\ &\geq \rho \times \text{opt}(I(T^*)) + (1 - \rho) \times \text{wor}(I(T^*)) \\ &\geq \rho \times \text{opt}(I(T^*)) + (1 - \rho) \times \text{wor}(I) \quad \text{by (1)} \end{aligned} \quad (2)$$

Eventually assume that T^* is a set in $\mathcal{P}_p([q])$ that contains a solution with optimal value over $\{T^n \mid T \in \mathcal{P}_p([q])\}$. Then provided that $\text{opt}(I(T^*))$ is δ -approximate on I , one gets a connection with $\text{opt}(I(T))$:

$$\begin{aligned} \max_{T \in \mathcal{P}_p([q])} \{v(I, x(T))\} - \text{wor}(I) &\geq \rho \times (\text{opt}(I(T^*)) - \text{wor}(I)) \quad (2) \\ &\geq \rho \times \delta \times (\text{opt}(I) - \text{wor}(I)) \quad (3) \end{aligned}$$

Hence, if we are able to compare — in a differential approximation manner — $\text{opt}(I(T^*))$ to $\text{opt}(I)$, then we can deduce from approximate solutions of subinstances $I(T)$ approximate solutions of I . We thus shall seek a lower bound for the differential ratio reached on I at $\text{opt}(I(T^*))$.

Let x^* be a solution with optimum value on I . Then one way to obtain such a lower bound consists in exhibiting two solution multisets $\mathcal{X} = (x^1, \dots, x^R)$ and $\mathcal{Y} = (y^1, \dots, y^R)$ of the same size R that satisfy the following conditions:

$$\mathcal{X} \subseteq \{T^n \mid T \in \mathcal{P}_p([q])\} \tag{4}$$

$$R^* \triangleq |\{r \in [R] \mid y_r = x^*\}| \geq 1 \tag{5}$$

$$|\{r \in [R] \mid x_{j_i}^r = v\}| = |\{r \in [R] \mid y_{j_i}^r = v\}|, v \in [q]^{J_i}, i \in [m] \tag{6}$$

Requirements (4), (5) and (6) respectively ensure that \mathcal{X} exclusively considers solutions of subinstances $I(T)$, x^* occurs at least once in \mathcal{Y} , and each constraint $P_i(x_{J_i})$ of I is evaluated on the same collection of $|J_i|$ -tuples over solution multisets \mathcal{X} and \mathcal{Y} . Requirement (6) thus ensures that the sum of solution values over \mathcal{X} and \mathcal{Y} are identical. Provided that such a pair $(\mathcal{X}, \mathcal{Y})$ exists, we have:

$$\begin{aligned} \text{opt}(I(T^*)) &\geq \sum_{r=1}^R \{v(I, x^r)\} / R \quad \text{by definition of } T^*, \text{ and (4)} \\ &= \sum_{r=1}^R \{v(I, y^r)\} / R \quad \text{by (6)} \\ &\geq R^* \times \text{opt}(I) / R + (R - R^*) \times \text{wor}(I) / R \quad \text{by (5)} \end{aligned} \tag{7}$$

Thus $\text{opt}(I(T^*))$ is R^*/R -approximate on I . Therefore, one shall seek such pairs of solution multisets on which the ratio R^*/R is as high as possible.

This is precisely what we do, and this is why we restrict our analysis to the case when $k \leq p$. Indeed, *e.g.* assume that $J_1 = (1, \dots, k)$ and $(x_1^*, \dots, x_k^*) = (1, \dots, k)$. Then by (6) and (5), \mathcal{X} shall contain at least $R^* > 0$ solutions x^r with $(x_1^r, \dots, x_k^r) = (1, \dots, k)$. If $k > p$, then such solutions violate condition (4). Hence, from now on, we assume $q > p \geq k$.

2.2 Partition-Based Solution Multisets

Solution x^* induces a partition of $[n]$ into q — possibly empty — subsets depending on the q possible values taken by its coordinates. Given $c \in [q]$, we denote by V_c the set of indices $j \in [n]$ such that $x_j^* = c$.

We restrict our solution multisets to vectors x that satisfy $x_j^* = x_h^* \Rightarrow x_j = x_h$, $j, h \in [n]$. It is thus possible to identify \mathcal{X} and \mathcal{Y} with two arrays Ψ and Φ with q columns, coefficients in $[q]$, and the same number R of rows. Each row $\Psi_r = (\Psi_r^1, \dots, \Psi_r^q)$ of Ψ gives rise in \mathcal{X} to the vector of $[q]^n$ that satisfies for every $c \in [q]$ that its coordinates with index in V_c all are equal to Ψ_r^c . \mathcal{Y} is derived from Φ in the exact same way. Formally, we define $\pi_{x^*} : [q]^q \rightarrow [q]^n$ by

$$\pi_{x^*}(u)_{V_c} = (u_c, \dots, u_c), \quad c \in [q]$$

and \mathcal{X}, \mathcal{Y} by $\mathcal{X} = (\pi_{x^*}(\Psi_r) \mid r \in [R])$ and $\mathcal{Y} = (\pi_{x^*}(\Phi_r) \mid r \in [R])$. Given $i \in [m]$, let $c_{i,1}, \dots, c_{i,h_i}$ refer to the distinct values taken by the coordinates of x^* with

Table 1. Pairs of arrays that achieve $\gamma(4, 3, 2)$ and $\gamma(5, 3, 2)$.

$\gamma(4, 3, 2) = 2/6 = 1/3$				$\gamma(5, 3, 2) = 1/6$													
Ψ^1	Ψ^2	Ψ^3	Ψ^4	Φ^1	Φ^2	Φ^3	Φ^4	Ψ^1	Ψ^2	Ψ^3	Ψ^4	Ψ^5	Φ^1	Φ^2	Φ^3	Φ^4	Φ^5
1	1	3	4	1	1	3	3	1	2	4	4	4	1	2	3	4	5
1	2	1	4	1	2	1	3	1	3	3	3	5	1	3	4	3	4
1	2	3	3	1	2	3	4	2	2	3	2	5	2	2	4	2	4
1	2	3	3	1	2	3	4	2	3	4	2	4	2	3	3	2	5
4	1	1	3	4	1	1	4	4	4	3	4	5	4	4	3	3	5
4	2	3	4	4	2	3	3	4	4	4	3	4	4	4	4	4	4

index in J_i , and let $H_i = (c_{i,1}, \dots, c_{i,h_i})$. Then solution multisets \mathcal{X} and \mathcal{Y} meet requirements (4), (5) and (6) of Section 2.1 *iff* arrays Ψ and Φ satisfy:

$$|\{\Psi_r^1, \dots, \Psi_r^q\}| \leq p, r \in [R] \quad (8)$$

$$R^* \triangleq |\{r \in [R] \mid \Phi_r = (1, \dots, q)\}| \geq 1 \quad (9)$$

$$|\{r \in [R] \mid \Psi_r^{H_i} = v\}| = |\{r \in [R] \mid \Phi_r^{H_i} = v\}|, v \in [q]^{|H_i|}, i \in [m] \quad (10)$$

Hence, if we are aware of such a pair of arrays, then we know by (7) that $\text{opt}(I(T^*))$ is R^*/R -approximate. These observations suggest to introduce the following families $\Gamma(R, R^*, q, p, k)$ of combinatorial designs and their associated numbers $\gamma(q, p, k)$ (see Table 1 for an illustration):

Definition 2. Let $k \geq 2$, $p \geq k$ and $q \geq p$ be three integers. Then given any two integers $R \geq 1$, $R^* \in [R]$, we define $\Gamma(R, R^*, q, p, k)$ as the (possibly empty) set of pairs (Ψ, Φ) of arrays with R rows, q columns, and coefficients in a set $\Sigma = \{\sigma_1, \dots, \sigma_q\}$ of q symbols, that satisfy the following:

1. the components of each row of Ψ take at most p distinct values;
2. $(\sigma_1, \dots, \sigma_q)$ occurs R^* times as a row in Φ ;
3. for all $J = \{c_1, \dots, c_k\} \subseteq [q]$ with $|J| = k$, subarrays $\Psi^J = (\Psi^{c_1}, \dots, \Psi^{c_k})$ and $\Phi^J = (\Phi^{c_1}, \dots, \Phi^{c_k})$ coincide up to the ordering of their rows.

Furthermore, we define $\gamma(q, p, k)$ as the greatest number $\gamma \in [0, 1]$ for which there exist two natural numbers R, R^* such that $R^*/R = \gamma$ and $\Gamma(R, R^*, q, p, k) \neq \emptyset$.

Since cardinalities $|H_i|$ may be at most $\min\{q, k\} = k$, by requirement 3. of Definition 2, a pair $(\Psi, \Phi) \in \Gamma(R, R^*, q, p, k)$ does satisfy (10) regardless of the precise instance I of kCSP-q and the precise solution x^* of I we consider. By (7), this implies that $\gamma(q, p, k)$ is a proper lower bound for the differential ratio reached on I at $\text{opt}(I(T^*))$. We thus have established the following:

Lemma 1. For all integers $k \geq 2$, $p \geq k$ and $q > p$, on any instance of kCSP-q , solutions that perform the best objective value among those whose coordinates take at most p distinct values are $\gamma(q, p, k)$ -approximate.

To conclude, according to inequality (3), Lemma 1 also establishes that $\gamma(q, p, k)$ is a proper lower bound for the expansion of our reduction:

Theorem 1. *For all integers $k \geq 2$, $p \geq k$ and $q > p$, k CSP- q D -reduces to k CSP- p with an expansion of $\gamma(q, p, k)$ on the approximation guarantee. The reduction involves $O(q^p)$ instances of k CSP- p .*

3 A Lower Bound for Numbers $\gamma(q, p, k)$

It remains us to exhibit lower bounds for numbers $\gamma(q, p, k)$. To do so, we mainly present a recursive construction for the case when $p = k$. But first, we mention a few combinatorial identities that are involved in the analysis of this construction.

We define:

$$T(a, b) \triangleq \sum_{r=0}^b \binom{a}{r} \binom{a-1-r}{b-r}, \quad a, b \in \mathbb{N}, a > b \quad (11)$$

$$S(a, b, c) \triangleq \sum_{r \geq 0} (-1)^r \binom{a}{r} \binom{b-r}{c-r}, \quad a, b, c \in \mathbb{N}, b \geq c \quad (12)$$

Numbers $T(a, b)$ and $S(a, b, c)$ satisfy the following identities:

Property 1. For all $a, b \in \mathbb{N}$ with $a > b \geq 1$, we have:

$$T(a, b) = 2^b \binom{a-1}{b} + T(a-1, b-1) \quad (13)$$

$$= 2^b \binom{a}{b} - T(a, b-1) \quad (14)$$

$$= 2 \sum_{c=b}^{a-1} T(c, b-1) + 1 \quad (15)$$

Proof (sketch). Recursions (13) and (14) are obtained using Pascal's rule on coefficients of the form respectively $\binom{a}{r}$ and $\binom{a-1-r}{b-r}$. Identity (15) can then be deduced from those recursions. \square

Property 2. For all $a, b, c \in \mathbb{N}$ with $b \geq \max\{a, c\}$, we have $S(a, b, c) = \binom{b-a}{c}$.

Proof (sketch). By induction on integer $b - a$. \square

3.1 A Recursive Construction for Families $\Gamma(R, 1, q, k, k)$

This section is dedicated to the proof of the following Theorem:

Theorem 2. *For all integers q and k with $q \geq k \geq 2$, we have $\gamma(q, k, k) = 1$ if $q = k$, and $\gamma(q, k, k) \geq 2/(T(q, k) + 1)$ otherwise.*

The case when $q = k$ is trivial, considering $\Psi = \Phi = \{(1, \dots, k)\}$. For greater integers $q - k$, the argument relies on the following Lemma:

Lemma 2. *Let $k \geq 2$, $q > k$, $R^* \geq 1$ and $R \geq R^*$ be four integers such that $\Gamma(R, R^*, q-1, k, k) \neq \emptyset$. Then $\Gamma(R+T(q-1, k-1), 1, q, k, k) \neq \emptyset$.*

Proof. Let $(\Psi, \Phi) \in \Gamma(R, R^*, q-1, k, k)$. We assume *w.l.o.g.* that $(1, \dots, q-1)$ occurs at row 1 in Φ . Our goal is to add in arrays Ψ and Φ a single new column and new rows so as to obtain a new pair $(\Psi, \Phi) \in \Gamma(R', R^{*'}, q, k, k)$ for some positive integers R' and $R^{*'}$ that ideally maximize the ratio $R^{*'}/R'$.

Table 2. Construction of a pair of arrays in $\Gamma(R + T(q - 1, k - 1), 1, q, k, k)$ starting with a pair $(\Psi, \Phi) \in \Gamma(R, R^*, q - 1, k, k)$ of arrays with $\Phi_1 = (1, \dots, q - 1)$.

1. Insert in Ψ and Φ new columns Ψ^q and Φ^q that are defined by:
 - (a) $\Psi_r^q = \Psi_r^1, r \in [R]$;
 - (b) $\Phi_1^q = q$ and $\Phi_r^q = \Phi_r^1, r \in \{2, \dots, R\}$.
2. For $h = k - 1$ down to 0:
 - (a) For all $H \subseteq [q - 1]$ with $|H| = h$:
 - i. let $\alpha(H) \in [q]^{q-1}$ be defined for all $c \in [q - 1]$ by $\alpha(H)_c = c$ if $c \in H$, and q otherwise;
 - ii. if $h \equiv k - 1 \pmod{2}$, then insert $\binom{q-2-h}{k-1-h}$ copies of row vectors $(\alpha(H), q)$ and $(\alpha(H), 1)$ in respectively Ψ and Φ ;
 - iii. otherwise, insert $\binom{q-2-h}{k-1-h}$ copies of row vectors $(\alpha(H), q)$ and $(\alpha(H), 1)$ in respectively Φ and Ψ .

We describe our construction in Table 2. Let us explain it before proving its rightness. Step 1. first inserts a q th column in the arrays. If we set Ψ^q to Ψ^1 and Φ^q to Φ^1 , then (Ψ, Φ) trivially fulfils requirements 1. and 3. of Definition 2. However, as $(1, \dots, q)$ must occur at least once as a row in Φ , we assign value q rather than Φ_1^1 to Φ_1^q . As a result, (Ψ, Φ) violates requirement 3. of Definition 2. Hence, in Step 2., we insert new rows in the arrays until they satisfy this requirement.

Let $J = \{c_1, \dots, c_{k-1}\}$ be a subset of $[q - 1]$ with size $k - 1$, and $u = (c_1, \dots, c_{k-1})$. After Step 1., row 1 is the single row of subarray (Φ^J, Φ^q) that coincides with (u, q) , while there is no such row in (Ψ^J, Ψ^q) . As $(\Phi_1^J, \Phi_1^q) = (u, 1)$ while $(\Phi, \Psi) \in \Gamma(R, R^*, q - 1, k, k)$, (Φ^J, Φ^q) symmetrically coincides with $(u, 1)$ on one less row than (Ψ^J, Ψ^q) does. Iteration $h = k - 1$ corrects this precise imbalance when it inserts row vectors $(\alpha(J), q)$ in Ψ and $(\alpha(J), 1)$ in Φ .

However, this iteration also introduces new violations of requirement 3. Notably, let $s \in [k - 1]$, and $v = (c_1, \dots, c_{s-1}, q, c_{s+1}, \dots, c_{k-1})$. Then iteration $h = k - 1$ inserts in each array a new row u with $u_J = v$ each time it selects a $(k - 1)$ -cardinality subset H of $[q - 1]$ with $c_1, \dots, c_{s-1}, c_{s+1}, \dots, c_{k-1} \in H$ and $c_s \notin H$. Since there are $q - 1 - (k - 1) = q - k$ such subsets, we deduce that at the end of this iteration, vectors (v, q) and $(v, 1)$ occur respectively $q - k$ and 0 times as a row in (Ψ^J, Ψ^q) , while the converse holds for (Φ^J, Φ^q) . During iteration $h = k - 2$, we correct this precise imbalance by inserting $q - k$ copies of row vectors $(\alpha(J \setminus \{c_s\}), 1)$ and $(\alpha(J \setminus \{c_s\}), q)$ in respectively Ψ and Φ .

More generally, for all $h \in \{0, \dots, k - 1\}$, iteration h ensures for all $v \in \{c_1, q\} \times \dots \times \{c_{k-1}, q\}$ with exactly h coordinates in $[q - 1]$ and all $a \in \{1, q\}$ that (v, a) occurs the same number of times as a row in (Ψ^J, Ψ^q) and (Φ^J, Φ^q) .

We now prove that, at the end of the process, $(\Psi, \Phi) \in \Gamma(R', 1, q, k, k)$ where $R' = R + T(q - 1, k - 1)$. By construction, the resulting arrays satisfy that:

- their number R' of rows is $R + \sum_{h=0}^{k-1} \binom{q-1}{h} \binom{q-2-h}{k-1-h} = R + T(q - 1, k - 1)$;
- in Φ , row 1 is the single row that coincides with $(1, \dots, q)$;
- in Ψ , the coefficients of every row take at most k distinct values.

It remains us to establish that (Ψ, Φ) fulfils requirement 3. of Definition 2. Let $J = (c_1, \dots, c_k)$ be a strictly increasing sequence of integers in $[q]$, and v be a vector of $[q]^k$. We shall show that subarrays Ψ^J and Φ^J coincide with v on the same number of rows. The case when $q \notin J$ is trivial. Thus assume that $c_k = q$. We consider two cases:

- $v \notin \{c_1, q\} \times \dots \times \{c_{k-1}, q\} \times \{1, q\}$. $\Psi_r^J = v$ or $\Phi_r^J = v$ may not occur unless $r \in [R]$. Let $r \in [R]$ and let $K = (c_1, \dots, c_{k-1}, 1)$. Then $\Psi_r^J = \Psi_r^K$, while $\Phi_r^J = \Phi_r^K$ unless $r = 1$, in which case $\Phi_1^J \neq v \neq \Phi_1^K$. Since the original pair of arrays belongs to $\Gamma(R, R^*, q-1, k, k)$, we deduce that Ψ^J and Φ^J indeed coincide with v on the same number of rows.

- $v \in \{c_1, q\} \times \dots \times \{c_{k-1}, q\} \times \{1, q\}$. If $(v_1, \dots, v_{k-1}) = (c_1, \dots, c_{k-1})$, then we already discussed the fact that iteration $h = k - 1$ of Step 2. corrects the imbalance induced by assignment $\Phi_1^q = q$. Otherwise, let L refer to the set of indices $c_s \in \{c_1, \dots, c_{k-1}\}$ such that $v_s = c_s$, and let $\ell = |L|$. As $\ell \leq k - 2$, $\Psi_r^J = v$ or $\Phi_r^J = v$ may not occur unless $r > R$. Thus consider an iteration $h \in \{0, \dots, k - 1\}$ of Step 2. For each $H \subseteq [q - 1]$ with $|H| = h$, $L \subseteq H$ and $H \setminus L \subseteq [q - 1] \setminus J$, this iteration generates $\binom{q-2-h}{k-1-h}$ rows u with $u_J = v$. On the one hand, if $v_k = q$ (*resp.*, $v_k = 1$), then these rows occur in Ψ (*resp.*, in Φ) iff h has the same parity as $k - 1$. On the other hand, there are $\binom{q-k}{h-\ell}$ such subsets H of $[q - 1]$. Hence, we have:

$$\begin{aligned} |\{r \in [R'] \mid \Psi_r^J = v\}| - |\{r \in [R'] \mid \Phi_r^J = v\}| &= \pm \sum_{h=\ell}^{k-1} (-1)^{k-1-h} \binom{q-2-h}{k-1-h} \binom{q-k}{h-\ell} \\ &= \pm \sum_{r=0}^{k-1-\ell} (-1)^r \binom{q-2-\ell-r}{k-1-\ell-r} \binom{q-k}{r} \end{aligned}$$

We recognise $S(q - k, q - 2 - \ell, k - 1 - \ell)$ which, according to Property 2, is equal to $\binom{k-2-\ell}{k-1-\ell} = 0$. We conclude that (Ψ, Φ) indeed satisfies requirement 3. of Definition 2. \square

The proof of Theorem 2 is straightforward from Lemma 2. Namely, given two integers $k \geq 2$ and $q \geq k$, we consider the following recursive construction:

1. Set $\Psi = \{(1, \dots, k)\}$ and $\Phi = \{(1, \dots, k)\}$.
2. For $a = k + 1$ to q , apply construction underlying Lemma 2 to (Ψ, Φ) .

Table 3 illustrates the construction when $k \in \{2, 3\}$. On the one hand, in view of Lemma 2, the resulting pair (Ψ, Φ) of arrays belongs to $\Gamma(R, 1, q, k, k)$ where

$$R = 1 + \sum_{a=k+1}^q T(a - 1, k - 1) = 1 + \sum_{a=k}^{q-1} T(a, k - 1)$$

On the other hand, by (15), we have:

$$1 + \sum_{a=k}^{q-1} T(a, k - 1) = 1 + (T(q, k) - 1)/2 = (T(q, k) + 1)/2$$

This completes the proof of Theorem 2.

3.2 Deduced Approximation Results

Let $k \geq 2$, $p \geq k$ and $q > p$ be three integers. If $p = k$, then Theorem 2 together with Theorem 1 provides a lower bound of $2/(T(q, k) + 1)$ for the expansion of our reduction. We seek an estimate of $2/(T(q, k) + 1)$.

Table 3. The recursive construction for families $\Gamma((T(5, 2) + 1)/2, 1, 5, 2, 2)$ and $\Gamma((T(4, 3) + 1)/2, 1, 4, 3, 3)$ of combinatorial designs.

$2/(T(5, 2) + 1) = 1/16$					$2/(T(4, 3) + 1) = 1/8$												
Ψ^1	Ψ^2	Ψ^3	Ψ^4	Ψ^5	Φ^1	Φ^2	Φ^3	Φ^4	Φ^5	Ψ^1	Ψ^2	Ψ^3	Ψ^4	Φ^1	Φ^2	Φ^3	Φ^4
1	2	1	1	1	1	2	3	4	5	1	2	3	1	1	2	3	4
1	3	3	1	1	1	3	1	1	1	1	2	4	4	1	2	4	1
3	2	3	3	3	3	2	1	3	3	1	4	3	4	1	4	3	1
3	3	1	3	3	3	3	3	3	3	4	2	3	4	4	2	3	1
1	4	4	4	1	1	4	4	1	1	1	4	4	1	1	4	4	4
4	2	4	4	4	4	2	4	1	4	4	2	4	1	4	2	4	4
4	4	3	4	4	4	4	3	1	4	4	4	3	1	4	4	3	4
4	4	4	1	4	4	4	4	4	4	4	4	4	4	4	4	4	1
4	4	4	1	4	4	4	4	4	4	4	4	4	4	4	4	4	1
1	5	5	5	5	1	5	5	5	1								
5	2	5	5	5	5	2	5	5	1								
5	5	3	5	5	5	5	3	5	1								
5	5	5	4	5	5	5	5	4	1								
5	5	5	5	1	5	5	5	5	5								
5	5	5	5	1	5	5	5	5	5								
5	5	5	5	1	5	5	5	5	5								

Property 3. For all $a, b \in \mathbb{N}$ with $a > b \geq 2$, we have $(T(a, b) + 1)/2 \leq (a - b/2)^b$.

Proof. Applying recursions first (14), and then (13), one gets equality:

$$T(a, b) + 1 = 2^b \binom{a}{b} - 2^{b-1} \binom{a-1}{b-1} - T(a-1, b-2) + 1 \quad (16)$$

On the one hand, we deduce again from (13) that $T(a-1, b-2) - 1 \geq T(a-b+1, 0) - 1 = 0$. On the other hand, we can rewrite $2^b \binom{a}{b} - 2^{b-1} \binom{a-1}{b-1}$ as:

$$2^b \binom{a}{b} - 2^{b-1} \binom{a-1}{b-1} = 2(a - b/2) \times 2^{b-1}/b! \times \prod_{i=0}^{b-2} (a - 1 - i)$$

Now $2^{b-1}/b! \leq 1$, while by the inequality of arithmetic and geometric means, we have $\prod_{i=0}^{b-2} (a - 1 - i) \leq (a - b/2)^{b-1}$. \square

We deduce from Property 3 that for all integers $k \geq 2$ and $q > k$, k CSP- q D -reduces to k CSP- k with an expansion of at least $(q - k/2)^k$ on the approximation guarantee. Corollary 2 thus holds. Now assume that $p > k$. Then observe:

$$\gamma(q, p, k) \geq \gamma(q - p + k, k, k) \geq 2/(T(q - p + k, k) + 1) \quad (17)$$

Indeed, let $a = q - p + k$, and assume that R and R^* are two integers such that $\Gamma(R, R^*, a, k, k) \neq \emptyset$. Let then $(\Psi, \Phi) \in \Gamma(R, R^*, a, k, k)$. Substituting for every row $u = (u_1, \dots, u_a)$ of Ψ and Φ row vector $u = (u_1, \dots, u_a, a + 1, \dots, q)$, one gets a new pair of arrays that trivially belongs to $\Gamma(R, R^*, q, p, k)$. Hence, combining Theorem 1, inequality (17) and Property 3, one obtains Corollary 1.

Table 4. Numbers $\gamma(q, p, k)$ and $\gamma_E(q, p, k)$ for some triples (q, p, k) . These values (and the underlying pairs of arrays) were calculated by computer.

		$\gamma(q, p, 2)$				$\gamma(q, p, 3)$			$\gamma_E(q, p, 2)$					$\gamma_E(q, p, 3)$			
		q				q			q					q			
		3	4	5	6	4	5	6	3	4	5	6	7	4	5	6	7
p	2	1/4	1/9	1/16	1/25	–	–	–	1/3	1/4	1/5	9/59	1/7	–	–	–	–
	3	–	1/3	1/6	1/10	1/8	1/25	1/56	–	1/2	2/5	4/13	2/7	1/4	1/11	$\frac{38425}{701342}$	
	4	–	–	4/9	1/4	–	1/5	2/27	–	–	3/5	7/15	3/7	–	1/3	1/6	5/52
	5	–	–	–	1/2	–	–	1/4	–	–	–	2/3	11/21	–	–	4/9	2/9
	6	–	–	–	–	–	–	–	–	–	–	–	5/7	–	–	–	1/2

4 Concluding Remarks

We make a few remarks as regards combinatorial designs of Definition 2. When $p = k$, we think that $2/(T(q, k) + 1)$ is the exact value of $\gamma(q, k, k)$. The question whether $2/(T(q, k) + 1)$ is optimal, though, still has to be settled. By contrast, when $p > k$, the only estimate of $\gamma(q, p, k)$ we are aware of is the trivial lower bound $\gamma(q - p + k, k, k)$. Yet, it the most likely holds given three integers $k \geq 2$, $p \geq k$ and $q > p$ that $\gamma(q + 1, p + 1, k) > \gamma(q, p, k)$. Table 4, in which we indicate the value of $\gamma(q, p, k)$ for a few triples (q, p, k) , illustrates this fact quite well.

According to Lemma 1, $\gamma(q, p, k)$ provides some lower bound on “how much we lose” on the optimal value of an instance of an optimization CSP when decreasing the size of its alphabet. This is a good motivation for studying these combinatorial designs in case when $p > k$. Likewise, let $k\text{CSP}(\mathcal{E}_q)$ refer to the optimization CSP over \mathbb{Z}_q where functions P_i that occur in the constraints have arity at most k , and satisfy:

$$P_i(y_1 + a, \dots, y_{k_i} + a) = P_i(y_1, \dots, y_{k_i}), \quad y \in \mathbb{Z}_q^{k_i}, \quad a \in \mathbb{Z}_q \quad (18)$$

$k\text{CSP}(\mathcal{E}_q)$ notably covers the restriction of $\text{Lin} - q$ to equations of the form $\alpha_1 y_1 + \dots + \alpha_{k-1} y_{k-1} - (\alpha_1 + \dots + \alpha_{k-1}) y_k \equiv \alpha_0 \pmod q$. Given an integer a , we denote by \mathbf{a} the vector (a, \dots, a) (whose dimension depends on the context). On an instance I of $k\text{CSP}(\mathcal{E}_q)$, any constraint C_i evaluates the same on any two entries x_{J_i} and $x_{J_i} + \mathbf{a}$. The objective function $v(I, \cdot)$ similarly evaluates the same on any two entries x and $x + \mathbf{a}$. This suggests to consider the slight relaxation $\Gamma_E(R, R^*, q, p, k)$ of families $\Gamma(R, R^*, q, p, k)$ where Ψ and Φ have coefficients in \mathbb{Z}_q and, rather than requirements 2. and 3., satisfy the two conditions below:

- 2'. $\Phi_r \in \{(a, 1 + a, \dots, q - 1 + a) \mid a \in \mathbb{Z}_q\}$ holds for R^* indices $r \in [R]$;
- 3'. for all $J \subseteq [q]$ with $|J| = k$ and all $v \in \{0\} \times \mathbb{Z}_q^{k-1}$, Ψ^J and Φ^J coincide with a vector in $\{(v_1 + a, \dots, v_k + a) \mid a \in \mathbb{Z}_q\}$ on the same number of rows.

We define numbers $\gamma_E(q, p, k)$ just as the same as numbers $\gamma(q, p, k)$. Table 5 pictures two such pairs of arrays, while Table 4 provides the value of $\gamma_E(q, p, k)$ for some triples (q, p, k) . Let $k \geq 2$, $p \geq k$ and $q > p$ be three integers. Using a

Table 5. Pairs of arrays that achieve $\gamma_E(5, 3, 2)$ and $\gamma_E(5, 4, 3)$.

$\gamma_E(5, 3, 2) = 4/10 = 2/5$					$\gamma_E(5, 4, 3) = 4/12 = 1/3$														
Ψ^0	Ψ^1	Ψ^2	Ψ^3	Ψ^4	Φ^0	Φ^1	Φ^2	Φ^3	Φ^4	Ψ^0	Ψ^1	Ψ^2	Ψ^3	Ψ^4	Φ^0	Φ^1	Φ^2	Φ^3	Φ^4
0	0	1	3	3	0	0	0	0	0	0	0	1	2	3	0	0	1	3	4
0	0	2	2	3	0	0	4	4	2	0	0	2	3	4	0	0	2	2	4
0	1	0	1	2	0	1	2	3	4	0	0	2	3	4	0	0	2	3	3
0	1	2	0	1	0	1	2	3	4	0	1	1	3	4	0	1	1	2	3
0	1	3	3	0	0	1	2	3	4	0	1	2	2	4	0	1	2	3	4
0	1	4	0	4	0	1	2	3	4	0	1	2	3	0	0	1	2	3	4
0	2	2	3	0	0	2	0	2	1	0	1	2	3	3	0	1	2	3	4
0	2	2	4	4	0	2	4	1	3	0	1	2	4	4	0	1	2	3	4
0	3	4	3	4	0	3	3	1	3	0	1	3	3	4	0	1	3	4	0
0	4	0	1	4	0	4	1	0	0	0	2	2	3	4	0	2	2	3	0
										0	2	2	3	4	0	2	2	4	4
										0	2	3	4	0	0	2	3	3	4

similar argument as for the general case, it is not hard to see that, when restricting to input instances of $\text{kCSP}(\mathcal{E}_q)$, the reduction we propose from $\text{kCSP}-q$ to $\text{kCSP}-p$ preserves the differential ratio up to a multiplicative factor of $\gamma_E(q, p, k)$. Notably, as $\gamma_E(q, 2, 2) = 1/q$, $q \in \{3, 4, 5, 7\}$, it follows from [?] that when $q \in \{3, 4, 5, 7\}$, $2\text{CSP}(\mathcal{E}_q)$ is approximable within factor $0.429/q$ (and not only $0.429/(q-1)^2$). Therefore, one also shall investigate families $\Gamma_E(R, R^*, q, p, k)$ of combinatorial designs, and this more specifically in case when $p = k = 2$.