Optimizing Alphabet Reduction Pairs of Arrays

CODESCO'24

Jean-François Culus[1] Sophie Toulouse[2]

[1]MEMIAD, Université des Antilles and CReC Saint-Cyr, France

[2]LIPN – Institut Galilée, Université Paris 13, France

Sevilla, July 11, 2024

Definition

Object: a pair (Q, P) of arrays with q columns on symbol set $\Sigma_q := \{0, \ldots, q-1\}$

Constraints:

Definition

Object: a pair (Q, P) of arrays with q columns on symbol set $\Sigma_q := \{0, \ldots, q-1\}$

Constraints:

- (Γ_Q) $(0, 1, \ldots, q-1)$ occurs at least once in Q

Definition

Object: a pair (Q, P) of arrays with q columns on symbol set $\Sigma_q := \{0, \ldots, q-1\}$

Constraints:

- (Γ_Q) $(0, 1, \ldots, q-1)$ occurs at least once in Q
- (Γ_P) each row of P involves at most $p < q$ symbols of Σ_q

Definition

Object: a pair (Q, P) of arrays with q columns on symbol set $\Sigma_q := \{0, \ldots, q-1\}$

Constraints:

- (Γ_Q) $(0, 1, \ldots, q-1)$ occurs at least once in Q
- (Γ_P) each row of P involves at most $p < q$ symbols of Σ_q
- $(k_=)$ Q and P are "*k-wise equivalent*" :=
  for all $J = (j_1, \ldots, j_k) \in \Sigma_q^k$, subarrays Q^J and P^J are the same collection of rows

(NB by $(k_=)$, the number of rows in P is the same as in Q.)

Parameters:

- q the alphabet size
- p a positive number $\leq q$
- $k \in \{1, \ldots, p\}$ the strength

Illustration when $(q, p, k) = (5, 4, 3)$

Q^0	Q^1	Q^2	Q^3	Q^4	P^0	P^1	P^2	P^3	P^4
0	0	1	3	4	0	0	1	0	3
0	0	2	0	4	0	0	2	3	4
0	0	2	3	3	0	0	2	3	4
0	1	1	0	4	0	1	1	3	4
0	1	1	3	3	0	1	1	3	4
0	1	2	0	3	0	1	2	0	4
0	1	2	3	4	0	1	2	0	4
0	1	2	3	4	0	1	2	3	3
0	1	2	3	4	0	1	2	3	3
4	0	1	0	3	4	0	1	0	4
4	0	1	0	3	4	0	1	3	3
4	0	2	3	4	4	0	2	0	3
4	1	1	3	4	4	1	1	0	3
4	1	2	0	4	4	1	2	3	4
4	1	2	3	3	4	1	2	3	4

$(Q, P) \in \Gamma(5, 4, 3)$?

(Γ_Q): $(0, 1, 2, 3, 4)$ occurs $3 > 1$ times in Q

(Γ_P): P_1 uses the $3 < 4$ symbols $0, 1, 3$, P_2 uses the $4 \leq 4$ symbols $0, 2, 3, 4$, ...

Illustration when $(q, p, k) = (5, 4, 3)$

Q^0	Q^1	Q^2	Q^3	Q^4
0	0	1	3	4
0	0	2	0	4
0	0	2	3	3
0	1	1	0	4
0	1	1	3	3
0	1	2	0	3
0	1	2	3	4
0	1	2	3	4
0	1	2	3	4
4	0	1	0	3
4	0	1	0	3
4	0	2	3	4
4	1	1	3	4
4	1	2	0	4
4	1	2	3	3

P^0	P^1	P^2	P^3	P^4
0	0	1	0	3
0	0	2	3	4
0	0	2	3	4
0	1	1	3	4
0	1	1	3	4
0	1	2	0	4
0	1	2	0	4
0	1	2	3	3
0	1	2	3	3
4	0	1	0	4
4	0	1	3	3
4	0	2	0	3
4	1	1	0	3
4	1	2	3	4
4	1	2	3	4

$(Q, P) \in \Gamma(5, 4, 3)$?

(Γ_Q): $(0, 1, 2, 3, 4)$ occurs $3 > 1$ times in Q

(Γ_P): P_1 uses the $3 < 4$ symbols $0, 1, 3$, P_2 uses the $4 \leq 4$ symbols $0, 2, 3, 4$, ...

$(k_=)$:
for $J = \{2, 3, 4\}$ and $w = (1, 3, 4)$,
w occurs as many times in P^J as in Q^J;

Illustration when $(q, p, k) = (5, 4, 3)$

Q^0	Q^1	Q^2	Q^3	Q^4
0	0	1	3	4
0	0	2	0	4
0	0	2	3	3
0	1	1	0	4
0	1	1	3	3
0	1	2	0	3
0	1	2	3	4
0	1	2	3	4
0	1	2	3	4
4	0	1	0	3
4	0	1	0	3
4	0	2	3	4
4	1	1	3	4
4	1	2	0	4
4	1	2	3	3

P^0	P^1	P^2	P^3	P^4
0	0	1	0	3
0	0	2	3	4
0	0	2	3	4
0	1	1	3	4
0	1	1	3	4
0	1	2	0	4
0	1	2	0	4
0	1	2	3	3
0	1	2	3	3
4	0	1	0	4
4	0	1	3	3
4	0	2	0	3
4	1	1	0	3
4	1	2	3	4
4	1	2	3	4

$(Q, P) \in \Gamma(5, 4, 3)$?

(Γ_Q): $(0, 1, 2, 3, 4)$ occurs $3 > 1$ times in Q

(Γ_P): P_1 uses the $3 < 4$ symbols $0, 1, 3$, P_2 uses the $4 \leq 4$ symbols $0, 2, 3, 4$, ...

$(k_=)$:
for $J = \{2, 3, 4\}$ and $w = (1, 3, 4)$,
w occurs as many times in P^J as in Q^J;
for $J = \{2, 3, 4\}$ and $w = (2, 0, 4)$,
w occurs as many times in P^J as in Q^J;

Connection to CSPs with bounded constraint arity [CT18]

Motivation: reducing an instance I of a k CSP over Σ_q to k CSPs over Σ_p

- we use Q to model solutions of the initial instance I of a k CSP over Σ_q
  in particular: $(0, 1, \ldots, q-1)$ models an optimum solution of I
- we use P to model solutions of k CSPs over Σ_p

We define:

- $R^*(Q, P)$: the number of times $(0, 1, \ldots, q-1)$ occurs in Q
- $R(Q, P)$: the number of rows in P (or, by ($k_=$), in Q)
- $\Gamma(q, p, k)$: the set of the ARPAs with parameters (q, p, k)
- $\gamma(q, p, k)$: the greatest ratio $R^*(Q, P)/R(Q, P)$ over $\Gamma(q, p, k)$

Back to CSPs:

- $\gamma(q, p, k)$ is a lower bound for the best approximation ratio reached on I by a solution whose coordinates take at most p distinct values
- it also is a lower bound for the expansion of the reduction

Illustration when $(q, p, k) = (5, 4, 3)$

Q^0	Q^1	Q^2	Q^3	Q^4		P^0	P^1	P^2	P^3	P^4
0	0	1	3	4		0	0	1	0	3
0	0	2	0	4		0	0	2	3	4
0	0	2	3	3		0	0	2	3	4
0	1	1	0	4		0	1	1	3	4
0	1	1	3	3		0	1	1	3	4
0	1	2	0	3		0	1	2	0	4
0	1	2	3	4		0	1	2	0	4
0	1	2	3	4		0	1	2	3	3
0	1	2	3	4		0	1	2	3	3
4	0	1	0	3		4	0	1	0	4
4	0	1	0	3		4	0	1	3	3
4	0	2	3	4		4	0	2	0	3
4	1	1	3	4		4	1	1	0	3
4	1	2	0	4		4	1	2	3	4
4	1	2	3	3		4	1	2	3	4

$$(Q, P) \in \Gamma(5, 4, 3)$$

$$\begin{aligned} R^*(Q, P) &= 3 \\ R(Q, P) &= 15 \\ R^*(Q, P)/R(Q, P) &= 3/15 \\ \Rightarrow \gamma(5, 4, 3) &\geq 1/5 \end{aligned}$$

Illustration when $(q, p, k) = (5, 4, 3)$

Q^0	Q^1	Q^2	Q^3	Q^4		P^0	P^1	P^2	P^3	P^4
0	0	1	3	4		0	0	1	0	3
0	0	2	0	4		0	0	2	3	4
0	0	2	3	3		0	0	2	3	4
0	1	1	0	4		0	1	1	3	4
0	1	1	3	3		0	1	1	3	4
0	1	2	0	3		0	1	2	0	4
0	1	2	3	4		0	1	2	0	4
0	1	2	3	4		0	1	2	3	3
0	1	2	3	4		0	1	2	3	3
4	0	1	0	3		4	0	1	0	4
4	0	1	0	3		4	0	1	3	3
4	0	2	3	4		4	0	2	0	3
4	1	1	3	4		4	1	1	0	3
4	1	2	0	4		4	1	2	3	4
4	1	2	3	3		4	1	2	3	4

$$(Q, P) \in \Gamma(5, 4, 3)$$

$$\begin{aligned} R^*(Q, P) &= 3 \\ R(Q, P) &= 15 \\ R^*(Q, P)/R(Q, P) &= 3/15 \\ \Rightarrow \gamma(5, 4, 3) &\geq 1/5 \end{aligned}$$

$\rightarrow$ *for* 3 *CSPs over* Σ_5*, the best solutions among those whose coordinates take at most* 4 *distinct values are* $1/5$*-approximate*

Facts that are already known about ARPAs

Property 1

$$\gamma(q,q,k) = 1, \qquad q \geq k \geq 1 \qquad (1)$$

$$\gamma(q+1,p+1,k) \geq \gamma(q,p,k), \qquad q \geq p \geq k > 0 \qquad (2)$$

Proof of (2).

Let $(Q,P) \in \Gamma(q,p,k)$. Consider e.g.:

$$\begin{array}{cccc|c} Q^0 & Q^1 & \dots & Q^{q-1} & Q^q \\ \hline Q^0_0 & Q^1_0 & \dots & Q^{q-1}_0 & q \\ Q^0_1 & Q^1_1 & \dots & Q^{q-1}_1 & q \\ \vdots & \vdots & \dots & \vdots & \vdots \\ Q^0_R & Q^1_R & \dots & Q^{q-1}_R & q \end{array} \quad \text{or} \quad \begin{array}{cccc|c} Q^0 & Q^1 & \dots & Q^{q-1} & Q^q \\ \hline Q^0_0 & Q^1_0 & \dots & Q^{q-1}_0 & Q^0_0+q \\ Q^0_1 & Q^1_1 & \dots & Q^{q-1}_1 & Q^0_1+q \\ \vdots & \vdots & \dots & \vdots & \vdots \\ Q^0_R & Q^1_R & \dots & Q^{q-1}_R & Q^0_R+q \end{array}$$

(for the latter taking the addition modulo $(q+1)$) □

Already known facts

We define:

$$T(q,k) := \sum_{i=0}^{k} \binom{q}{i}\binom{q-1-i}{k-i}, \qquad q > k \geq 0 \qquad (3)$$

Proposition 2 ([CT18])

For all integers $k > 0$ and $q > k$, there exists $(Q,P) \in \Gamma(q,k,k)$ such that $R^*(Q,P) = 1$ and $R(Q,P) = (T(q,k)+1)/2$.

Proof (sketch).

Recursive construction starting with $P = Q = (0, 1, \ldots, k-1)$. □

Consequence (combining Proposition 2 and (2)):

$$\gamma(q,p,k) \geq 2/\left(T(q-p+k,k)+1\right), \qquad q > p \geq k \geq 1 \qquad (4)$$

Illustration when $(k, q) = (3, 5)$

Q^0	Q^1	Q^2	Q^3	Q^4
0	1	2	3	4
0	1	3	0	0
0	3	2	0	0
3	1	2	0	3
0	3	3	3	0
3	1	3	3	3
3	3	2	3	3
3	3	3	0	3
0	1	4	4	0
0	4	2	4	0
0	4	4	3	0
4	1	2	4	0
4	1	4	3	0
4	4	2	3	0
0	4	4	4	4
0	4	4	4	4
4	1	4	4	4
4	1	4	4	4
4	4	2	4	4
4	4	2	4	4
4	4	4	3	4
4	4	4	3	4
4	4	4	4	0
4	4	4	4	0
4	4	4	4	0

P^0	P^1	P^2	P^3	P^4
0	1	2	0	0
0	1	3	3	0
0	3	2	3	0
3	1	2	3	3
0	3	3	0	0
3	1	3	0	3
3	3	2	0	3
3	3	3	3	3
0	1	4	4	4
0	4	2	4	4
0	4	4	3	4
4	1	2	4	4
4	1	4	3	4
4	4	2	3	4
0	4	4	4	0
0	4	4	4	0
4	1	4	4	0
4	1	4	4	0
4	4	2	4	0
4	4	2	4	0
4	4	4	3	0
4	4	4	3	0
4	4	4	4	4
4	4	4	4	4
4	4	4	4	4

Questions addressed

1. Is the bound of $2/\left(T(q,k)+1\right)$ for $\gamma(q,k,k)$ tight, $q > k > 0$?
2. Can we find better bounds for $\gamma(q,p,k)$, $q > p > k > 0$?

Can we simplify the problem?

Intuition behind $\gamma(q, p, k)$:

- We want to "cover" *as many as possible* occurrences of the word of q symbols $(0, 1, \ldots, q-1)$ by *as few as possible* words of at most p symbols
- The most critical aspect of a coefficient in Q and P is whether it matches its column index or not

Can we simplify the problem?

Q^0	Q^1	Q^2	Q^3	Q^4
0	?	?	3	4
0	?	2	?	4
0	?	2	3	?
0	1	?	?	4
0	1	?	3	?
0	1	2	?	?
0	1	2	3	4
0	1	2	3	4
0	1	2	3	4
?	?	?	?	?
?	?	?	?	?
?	?	2	3	4
?	1	?	3	4
?	1	2	?	4
?	1	2	3	?

P^0	P^1	P^2	P^3	P^4
0	?	?	?	?
0	?	2	3	4
0	?	2	3	4
0	1	?	3	4
0	1	?	3	4
0	1	2	?	4
0	1	2	?	4
0	1	2	3	?
0	1	2	3	?
?	?	?	?	4
?	?	?	3	?
?	?	2	?	?
?	1	?	?	?
?	1	2	3	4
?	1	2	3	4

(Q, P) is a partially defined solution:

- the coefficients that coincide with their column index are fixed,
- the other coefficients (with value '?') still must be defined,
- Q and P are k-wise equivalent

Question: can we replace each symbol '?' by a value *distinct from its column index* in such a way that (Q, P) stills satisfies $(k_=)$, but also (Γ_P)?

Cover pairs of arrays: definition

$\Rightarrow$ new (Boolean) object: *cover pairs of arrays*

Object: a pair (N, D) of arrays with n columns on symbol set $\{0, 1\}$

Constraints:

- (Δ_N) the row of all-ones occurs at least once in N
- (Δ_D) each row of D has at most $d < n$ non-zero coefficients
- $(k_=)$ N and D are "*k-wise equivalent*" :=
  for all $J = (j_1, \ldots, j_k) \in [n]^k$, subarrays D^J and N^J are the same collection of rows

Parameters:

- n the dimension
- d a positive number $\leq n$
- $k \in \{1, \ldots, d\}$ the strength

Illustration when $(n, d, k) = (5, 4, 3)$

N^1	N^2	N^3	N^4	N^5	D^1	D^2	D^3	D^4	D^5
1	0	0	1	1	1	0	0	0	0
1	0	1	0	1	1	0	1	1	1
1	0	1	1	0	1	0	1	1	1
1	1	0	0	1	1	1	0	1	1
1	1	0	1	0	1	1	0	1	1
1	1	1	0	0	1	1	1	0	1
1	1	1	1	1	1	1	1	0	1
1	1	1	1	1	1	1	1	1	0
1	1	1	1	1	1	1	1	1	0
0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	1	0
0	0	1	1	1	0	0	1	0	0
0	1	0	1	1	0	1	0	0	0
0	1	1	0	1	0	1	1	1	1
0	1	1	1	0	0	1	1	1	1

$(1, 1, \ldots, 1)$ occurs $3 > 1$ times in N

Optimal CPAs

Notations:

- $R^*(N, D)$: the number of times $(1, 1, \ldots, 1)$ occurs in N
- $R(N, D)$: the number of rows in D and N
- $\Delta(n, d, k)$: the set of the CPAs with paramaters (n, d, k)

Quantity of interest: $\delta(n, d, k)$: the greatest ratio $R^*(N, D)/R(N, D)$ over $\Delta(n, d, k)$

$\rightarrow$ *we call optimal the CPAs that achieve* $\delta(n, d, k)$

Connection to optimal ARPAs:

- since CPAs model partially defined ARPAs, we have: $\delta(n, d, k) \geq \gamma(n, d, k)$
- $\rightarrow$ *question: what about the reverse inequality?*

Definition

Weight of a boolean word: the number of its non-zero coordinates

Definition 3 (Regular CPAs)

CPAs in which the words of a given weight all occur the same number of times, in the same array.

Illustration when $(n, d, k) = (5, 4, 3)$

N^1	N^2	N^3	N^4	N^5
1	0	0	1	1
1	0	1	0	1
1	0	1	1	0
1	1	0	0	1
1	1	0	1	0
1	1	1	0	0
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
0	0	0	0	0
0	0	0	0	0
0	0	1	1	1
0	1	0	1	1
0	1	1	0	1
0	1	1	1	0

D^1	D^2	D^3	D^4	D^5
1	0	0	0	0
1	0	1	1	1
1	0	1	1	1
1	1	0	1	1
1	1	0	1	1
1	1	1	0	1
1	1	1	0	1
1	1	1	1	0
1	1	1	1	0
0	0	0	0	1
0	0	0	1	0
0	0	1	0	0
0	1	0	0	0
0	1	1	1	1
0	1	1	1	1

the words of weight 3 occur once in N
the word of weight 5 occurs 3 times in N
the word of weight 0 occurs twice in N

the words of weight 1 occur once in D
the words of weight 4 occur twice in D

Property

Property 4

Among the CPAs (N, D) that realize $\delta(n, d, k)$, there exist a regular one

Proof (sketch).

Permute the coefficients of each row of N an D by each permutation on $\{1, \ldots, n\}$. □

Deriving ARPAs from regular CPAs

Data: $(N, D) \in \Delta(n, d, k)$, $r :=$ the greatest weight $< n$ of a word in $N \cup D$

Theorem 5 ([CT24])

We can derive from (N, D) an ARPA $(Q, P) \in \Gamma(n, d', k)$ with the same ratio R^*/R as (N, D), where $d' \leq d + 2$. In particular, $d' = d$ provided that $r = d$ and the words occurring in D have weight $\neq d - 1$.

Proof (sketch for the zero coefficients).

1. translate (N, D) into a partially defined ARPA (Q, P)
2. for each row u of weight r that occurs in N or D, map its zero coefficients to the column index of its leftmost coefficient initially equal to 1
3. "propagate" these assignments to words of smaller weight

□

Consequence: $\gamma(q, p, k) \geq \delta(q, d + 2, k)$

Illustration when $(n, d, k) = (5, 4, 3)$ and $r = d = d'$

$(N, D) \in \Delta(5, 4, 3)$

N^1	N^2	N^3	N^4	N^5	D^1	D^2	D^3	D^4	D^5
1	1	1	1	1	1	1	1	1	0
1	1	1	1	1	1	1	1	1	0
1	1	1	1	1	1	1	1	0	1
1	1	1	0	0	1	1	1	0	1
1	1	0	1	0	1	1	0	1	1
1	1	0	0	1	1	1	0	1	1
1	0	1	1	0	1	0	1	1	1
1	0	1	0	1	1	0	1	1	1
1	0	0	1	1	0	1	1	1	1
0	1	1	1	0	0	1	1	1	1
0	1	1	0	1	1	0	0	0	0
0	1	0	1	1	0	1	0	0	0
0	0	1	1	1	0	0	1	0	0
0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	1

$(Q, P) \in \Gamma(5, 4, 3)$

Q^0	Q^1	Q^2	Q^3	Q^4	P^0	P^1	P^2	P^3	P^4
0	1	2	3	4	0	1	2	3	0
0	1	2	3	4	0	1	2	3	0
0	1	2	3	4	0	1	2	0	4
0	1	2	0	0	0	1	2	0	4
0	1	0	3	0	0	1	0	3	4
0	1	0	0	4	0	1	0	3	4
0	0	2	3	0	0	0	2	3	4
0	0	2	0	4	0	0	2	3	4
0	0	0	3	4	1	1	2	3	4
1	1	2	3	0	1	1	2	3	4
1	1	2	0	4	0	0	0	0	0
1	1	0	3	4	1	1	0	0	0
1	0	2	3	4	1	0	2	0	0
1	0	0	0	0	1	0	0	3	0
1	0	0	0	0	1	0	0	0	4

Modelling regular CPAs

Variables:

- For $i \in \{0, \dots, n\}$, y_i: the number of times the words of weight i occur in N
- For $i \in \{0, \dots, d\}$, x_i: the number of times the words of weight i occur in D

$\Rightarrow$ y_n represents $R^*(N, D)$, while $\sum_{i=0}^{d} \binom{n}{i} x_i$ and $\sum_{i=0}^{n} \binom{n}{i} y_i$ both represent $R(N, D)$

Constraints:

- For (Δ_N): $y_n \geq 1$
- For $(k_=)$: for $J \subseteq \{1, \dots, n\}$ with $|J| = k$ and $w \in \{0,1\}^k$, the numbers of rows u of N and D satisfying $u_J = w$ only depends on the number of the non-zero coordinates of w

$\rightarrow$ we consider for $(k_=)$ the constraints:

$$\sum_{i=h}^{d} \binom{n-k}{i-h} x_i = \sum_{i=h}^{n-k+h} \binom{n-k}{i-h} y_i, \qquad h \in \{0, \dots, k\} \tag{5}$$

(Where $\binom{n-k}{i-h}$ counts the number of words $u \in \{0,1\}^n$ of weight i verifying $u_J = w$.)

Linear program

We denote by $LP_{n,d,k}$ the linear program in continuous variables below:

$$\left\{ \begin{array}{lrll} 2/\delta(n,d,k) - 1 = & \min \; \sum_{i=0}^{n-1} \binom{n}{i} y_i + \sum_{i=0}^{d} \binom{n}{i} x_i & & \\ & s.t. \; \sum_{i=k}^{d} \binom{n-k}{i-k} x_i - \sum_{i=k}^{n-1} \binom{n-k}{i-k} y_i & = 1 & \\ & \sum_{i=h}^{d} \binom{n-k}{i-h} x_i - \sum_{i=h}^{n-k+h} \binom{n-k}{i-h} y_i & = 0, & h \in \{0, \ldots, k-1\} \\ & y_0, \ldots, y_{n-1}, x_0, \ldots, x_d & \geq 0 & \end{array} \right.$$

NB:

- It only requires $\Theta(n)$ variables and $\Theta(k)$ constraints to model the restriction of $\Delta(n,d,k)$ to regular designs
- (while it requires $\Theta(n^n)$ variables and $\Theta(\binom{n}{k} \times n^k)$ constraints to model $\Gamma(n,d,k)$, and still $\Theta(2^n)$ variables and $\Theta(\binom{n}{k} \times 2^k)$ constraints to model $\Delta(n,d,k)$)

Optimal regular CPAs

Theorem 6 ([CT24])

For each choice of $k+2$ word weights

$$i_{k+1} = n > i_k = d > i_{k-1} > \ldots > i_1 > i_0 = 0$$

that occur alternately in N and D, there exists a regular CPA $(N, D) \in \Delta(n, d, k)$ with ratio $R^*(N, D)/R(N, D)$ equal to:

$$2/\left(1 + \sum_{r=0}^{k} \prod_{s \in \{0,\ldots,k\} \setminus \{r\}} \frac{n - i_s}{|i_r - i_s|}\right) \tag{6}$$

The best such CPA realizes $\delta(n, d, k)$.

Proof (sketch).

- we characterize the feasible bases of $LP_{n,d,k}$
- we give necessary conditions for a feasible base of $LP_{n,d,k}$ to be optimal

□

Optimal regular CPAs

Corollary 7

$$
\begin{aligned}
\gamma(q,p,k) &= \delta(q,p,k), && q \geq p \geq k > 0 \quad (7)\\
\gamma(q,k,k) &= 2/\left(T(q,k)+1\right), && q > k \geq 1 \quad (8)\\
\gamma(q,p,1) &= p/q, && q \geq p \geq 1 \quad (9)\\
\gamma(q,p,2) &= \lceil p/2\rceil \lfloor p/2\rfloor / \left((q-\lceil p/2\rceil)(q-\lfloor p/2\rfloor)\right), && q \geq p \geq 2 \quad (10)
\end{aligned}
$$

Proof (sketch).

- we deduce from Theorem 6 the analytic expression of $\delta(n,d,k)$ in case when $k \in \{1,2,p\}$
- by Theorem 5, we can derive from optimal CPAs of Theorem 6 ARPAs with the same set of parameters and the same ratio of R^*/R

□

Summary of the facts exposed

As regards $\gamma(q, p, k)$:

- computing optimal ARPAs reduces to compute optimal CPAs
- we are aware of optimal ARPAs (and CPAs) in case when $p \in \{q, k\}$ or $k \in \{1, 2\}$
- for the other cases, we somehow know how to derive the expression of optimal solutions
- we know, however, how to construct (suboptimal) ARPAs (and CPAs) for all set (q, p, k) of parameters (many ways)

Direction for further researches:

- providing the analytic expression of $\gamma(q, p, k)$ (and $\delta(q, p, k)$) for other cases (i.e., when $q > p > k > 2$)
- studying the case where repeated rows are not allowed
- ddesigning (optimal) solutions using only a few rows

Notice that CPAs (and thus, ARPAs) have another connection to CSPs: $\delta(n, d, k)$ is a lower bound on approximation guarantee reached on Hamming balls of radius k.

Illustration: two ARPAs achieving $\gamma(5,3,2)$

Q^0	Q^1	Q^2	Q^3	Q^4
0	1	2	3	4
1	2	2	1	4
3	3	2	2	4
3	3	3	3	3
1	1	3	1	3
0	2	3	2	3

P^0	P^1	P^2	P^3	P^4
3	3	2	3	4
1	1	2	1	4
0	2	2	2	4
0	1	3	3	3
1	2	3	1	3
3	3	3	2	3

Q^0	Q^1	Q^2	Q^3	Q^4
0	1	2	3	4
0	1	2	3	4
0	1	2	3	4
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
1	1	0	0	0
1	1	0	0	0
1	1	1	1	1
1	0	2	0	0
1	0	2	0	0
2	2	2	1	1
1	0	0	3	0
1	0	0	3	0
2	2	1	3	1
1	0	0	0	4
1	0	0	0	4
2	2	1	1	4

P^0	P^1	P^2	P^3	P^4
0	1	2	0	0
0	1	0	3	0
0	1	0	0	4
0	0	2	3	0
0	0	2	0	4
0	0	0	3	4
1	1	2	3	1
1	1	2	1	4
1	1	1	3	4
2	2	2	3	4
1	0	0	0	0
1	0	0	0	0
1	0	0	0	0
1	0	0	0	0
1	0	0	0	0
1	0	0	0	0
2	2	1	1	1
2	2	1	1	1

Illustration: alternate constructions

Proposition 8

For all integers $q \geq 3$, there exists $(Q, P) \in \Gamma(q, q - \lfloor q/3 \rfloor, 2)$ such that $R^*(Q, P) = 1$ and $R(Q, P) = 4$.

Proof.

- Partition Σ_q into any three subsets A, B, C of cardinality in $\{\lfloor q/3 \rfloor, \lceil q/3 \rceil\}$.
- Pick any two symbols $x \in A$ and $y \in B \cup C$.
- Consider then the pair (Q, P) below

Q^A	Q^B	Q^C	P^A	P^B	P^C
A	B	C	A	B	$x \dots x$
A	$x \dots x$	$x \dots x$	A	$x \dots x$	C
$y \dots y$	B	$x \dots x$	$y \dots y$	B	C
$y \dots y$	$x \dots x$	C	$y \dots y$	$x \dots x$	$x \dots x$

(NB by (10), the construction is optimal if q is a multiple of 3.) □

ARPAs maximizing R^*/R or minimizing R

We define $R(q, p, k)$:= the smallest number of rows over $\Gamma(q, p, k)$.

For $\gamma(q, p, k)$, we indicate the ratio $R^*(Q, P)/R(Q, P)$ on ARPAs (Q, P) that minimize $R(Q, P)$ among those which realize $\gamma(q, p, k)$.

	q :	3		4		5		6		7	
k	p	γ	R	γ	R	γ	R	γ	R	γ	R
2	2	$1/4^*$	4^*	$1/9^*$	9^*	$1/16^*$	16^*	$1/25^*$	25^*	$1/36^*$	36^*
	3	—	—	$2/6^*$	4^*	$1/6^*$	6^*	$1/10$	10	$1/15$	15
	4	—	—	—	—	$8/18^*$	4^*	$1/4^*$	4^*		
	5	—	—	—	—	—	—	$7/14$	4^*	$3/10$	4^*
	6	—	—	—	—	—	—	—	—	$9/16$	
3	3	—	—	$1/8^*$	8^*	$1/25^*$	25^*	$1/56^*$	56^*	$1/105^*$	105^*
	4	—	—	—	—	$3/15$	8	$4/54$			
	5	—	—	—	—	—	—	$6/24$			

* mark: cases for which we know how to construct a design that realizes the corresponding value (the other values have been calculated by computer).

blue color: cases where a regular design achieves the corresponding number

Relaxed ARPAs

In brief: almost the same thing as ARPAs, but any two words $(w_0, w_1, \ldots, w_{n-1})$ and $(w_0+a, w_1+a, \ldots, w_{n-1}+a)$ are considered equivalent $(\sim_q)$

Notations:

- $R^*(Q,P)$: the number of rows u of Q satisfying $u \sim_q (0, 1, \ldots, q-1)$
- $R(Q,P)$: the number of rows in P (or, by $(k_\sim)$, in Q)
- $\Gamma_E(q,p,k)$: the set of the relaxed ARPAs with parameters (q,p,k)
- $\gamma_E(q,p,k)$: the greatest ratio $R^*(Q,P)/R(Q,P)$ over $\Gamma_E(q,p,k)$ (NB of course, we have $\gamma_E(q,p,k) \geq \gamma(q,p,k)$)

Motivation [CT18]:

- k CSP($\mathcal{E}_\mathbf{q}$): k CSPs over Σ_q in which the constraints are stable under the shift by a same quantity of all their entries
- $\Rightarrow$ the same as for ARPAs, but reducing k CSP($\mathcal{E}_\mathbf{q}$) to k CSPs over Σ_p

Similarly to the case of ARPAs, we can seek and find bounds and constructions for $\gamma_E(q,p,k)$ and $\Gamma_E(q,p,k)$

Illustration when $(q, p, k) = (5, 4, 3)$

Q^0	Q^1	Q^2	Q^3	Q^4
0	0	1	3	4
0	0	2	2	4
0	0	2	3	3
0	1	1	2	3
0	1	2	3	4
0	1	2	3	4
1	2	3	4	0
2	3	4	0	1
0	1	3	4	0
0	2	2	3	0
0	2	2	4	4
0	2	3	3	4

P^0	P^1	P^2	P^3	P^4
0	0	1	2	3
0	0	2	3	4
0	0	2	3	4
0	1	1	3	4
0	1	2	2	4
0	1	2	3	0
0	1	2	3	3
0	1	2	4	4
0	1	3	3	4
0	2	2	3	4
0	2	2	3	4
0	2	3	4	0

$$R^*(Q,P) = 4$$
$$R(Q,P) = 12$$
$$R^*(Q,P)/R(Q,P) = 4/12$$

$(k_\sim)$?

Illustration when $(q, p, k) = (5, 4, 3)$

Q^0	Q^1	Q^2	Q^3	Q^4
0	0	1	3	4
0	0	2	2	4
0	0	2	3	3
0	1	1	2	3
0	1	2	3	4
0	1	2	3	4
1	2	3	4	0
2	3	4	0	1
0	1	3	4	0
0	2	2	3	0
0	2	2	4	4
0	2	3	3	4

P^0	P^1	P^2	P^3	P^4
0	0	1	2	3
0	0	2	3	4
0	0	2	3	4
0	1	1	3	4
0	1	2	2	4
0	1	2	3	0
0	1	2	3	3
0	1	2	4	4
0	1	3	3	4
0	2	2	3	4
0	2	2	3	4
0	2	3	4	0

$$\begin{aligned} R^*(Q,P) &= 4 \\ R(Q,P) &= 12 \\ R^*(Q,P)/R(Q,P) &= 4/12 \end{aligned}$$

$(k_\sim)$?

- for $J = \{1, 2, 3\}$ and $w = (0, 1, 3)$, $P^J_r \sim_q w$ occurs as many often as $Q^J_r \sim_q w$;

Illustration when $(q, p, k) = (5, 4, 3)$

Q^0	Q^1	Q^2	Q^3	Q^4
0	0	1	3	4
0	0	2	2	4
0	0	2	3	3
0	1	1	2	3
0	1	2	3	4
0	1	2	3	4
1	2	3	4	0
2	3	4	0	1
0	1	3	4	0
0	2	2	3	0
0	2	2	4	4
0	2	3	3	4

P^0	P^1	P^2	P^3	P^4
0	0	1	2	3
0	0	2	3	4
0	0	2	3	4
0	1	1	3	4
0	1	2	2	4
0	1	2	3	0
0	1	2	3	3
0	1	2	4	4
0	1	3	3	4
0	2	2	3	4
0	2	2	3	4
0	2	3	4	0

$$R^*(Q,P) = 4$$
$$R(Q,P) = 12$$
$$R^*(Q,P)/R(Q,P) = 4/12$$

$(k_\sim)$?

- for $J = \{1,2,3\}$ and $w = (0,1,3)$, $P_r^J \sim_q w$ occurs as many often as $Q_r^J \sim_q w$;
- for $J = \{1,2,3\}$ and $w = (0,2,2)$, $P_r^J \sim_q w$ occurs as many often as $Q_r^J \sim_q w$;

Illustration when $(q, p, k) = (5, 4, 3)$

Q^0	Q^1	Q^2	Q^3	Q^4
0	0	1	3	4
0	0	2	2	4
0	0	2	3	3
0	1	1	2	3
0	1	2	3	4
0	1	2	3	4
1	2	3	4	0
2	3	4	0	1
0	1	3	4	0
0	2	2	3	0
0	2	2	4	4
0	2	3	3	4

P^0	P^1	P^2	P^3	P^4
0	0	1	2	3
0	0	2	3	4
0	0	2	3	4
0	1	1	3	4
0	1	2	2	4
0	1	2	3	0
0	1	2	3	3
0	1	2	4	4
0	1	3	3	4
0	2	2	3	4
0	2	2	3	4
0	2	3	4	0

$$R^*(Q,P) = 4$$
$$R(Q,P) = 12$$
$$R^*(Q,P)/R(Q,P) = 4/12$$

$(k_\sim)$?

- for $J = \{1, 2, 3\}$ and $w = (0, 1, 3)$, $P_r^J \sim_q w$ occurs as many often as $Q_r^J \sim_q w$;
- for $J = \{1, 2, 3\}$ and $w = (0, 2, 2)$, $P_r^J \sim_q w$ occurs as many often as $Q_r^J \sim_q w$;
- for $J = \{1, 2, 3\}$ and $w = (0, 2, 3)$, $P_r^J \sim_q w$ occurs as many often as $Q_r^J \sim_q w$;

Illustration when $(q, p, k) = (5, 4, 3)$

Q^0	Q^1	Q^2	Q^3	Q^4
0	0	1	3	4
0	0	2	2	4
0	0	2	3	3
0	1	1	2	3
0	1	2	3	4
0	1	2	3	4
1	2	3	4	0
2	3	4	0	1
0	1	3	4	0
0	2	2	3	0
0	2	2	4	4
0	2	3	3	4

P^0	P^1	P^2	P^3	P^4
0	0	1	2	3
0	0	2	3	4
0	0	2	3	4
0	1	1	3	4
0	1	2	2	4
0	1	2	3	0
0	1	2	3	3
0	1	2	4	4
0	1	3	3	4
0	2	2	3	4
0	2	2	3	4
0	2	3	4	0

$$R^*(Q,P) = 4$$
$$R(Q,P) = 12$$
$$R^*(Q,P)/R(Q,P) = 4/12$$

$(k_\sim)$?
- for $J = \{1,2,3\}$ and $w = (0,1,3)$,
  $P_r^J \sim_q w$ occurs as many often as $Q_r^J \sim_q w$;
- for $J = \{1,2,3\}$ and $w = (0,2,2)$,
  $P_r^J \sim_q w$ occurs as many often as $Q_r^J \sim_q w$;
- for $J = \{1,2,3\}$ and $w = (0,2,3)$,
  $P_r^J \sim_q w$ occurs as many often as $Q_r^J \sim_q w$;

...

$\Rightarrow \gamma_E(5,4,3) \geq 1/3$

Relaxed ARPAs maximizing R^*/R or minimizing R

We define $R_E(q,p,k) :=$ the smallest number of rows over $\Gamma_E(q,p,k)$.

For $\gamma_E(q,p,k)$, we indicate the ratio $R^*(Q,P)/R(Q,P)$ in $(Q,P) \in \Gamma_E(q,p,k)$ that minimize $R(Q,P)$ among those which realize $\gamma_E(q,p,k)$

	q :	3		4		5		6		7	
k	p	γ_E	R_E	γ_E	R_E	γ_E	R_E	γ_E	R_E	γ_E	R_E
2	2	$1/3^*$	3^*	$1/4^*$	4^*	$2/10$	7	$9/59$	8	$3/21$	
	3	—	—	$6/12$	3^*	$4/10$	4^*	$8/26$	5	$48/168$	
	4	—	—	—	—	$6/10$	3^*	$7/15$	3^*	$9/21$	
	5	—	—	—	—	—	—	$40/60$	3^*	$11/21$	3^*
	6	—	—	—	—	—	—	—	—	$15/21$	3^*
3	3	—	—	$1/4^*$	4^*	$5/55$	14	$153700/2805368$			
	4	—	—	—	—	$4/12$	8	$1/6$	6		
	5	—	—	—	—	—	—	$8/18$	4^*		
	6	—	—	—	—	—	—	—	—	$14/28$	
4	4	—	—	—	—	$4/44$	15^*				
	5	—	—	—	—	—	—	$44/264$			
5	5	—	—	—	—	—	—	$1/16$	16		
	6	—	—	—	—	—	—	—	—	$6/60$	

* mark: cases for which we know how to construct a design that realizes the corresponding value (the other values have been calculated by computer).

blue color: cases that meet a lower bound we have established

Jean-François Culus and Sophie Toulouse.
2 csps all are approximable within a constant differential factor.
In Jon Lee, Giovanni Rinaldi, and A. Ridha Mahjoub, editors, *Combinatorial Optimization*, volume 10856 of *Lecture Notes in Computer Science*, pages 389–401, Cham, 2018. Springer International Publishing.

Jean-François Culus and Sophie Toulouse.
Optimizing alphabet reduction pairs of arrays, 2024.
preprint arXiv 2406.10930.