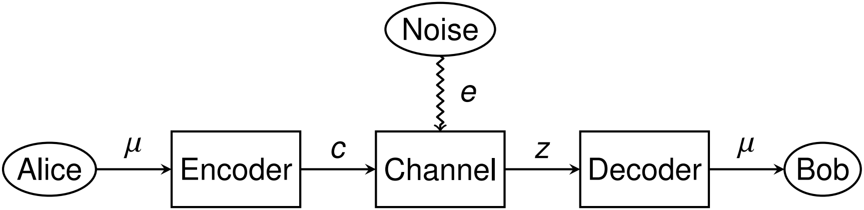
Evolutionary Strategies for the Design of Binary Linear Error-Correcting Codes

Claude Carlet, **Luca Mariot**, Luca Manzoni, Stjepan Picek

l.mariot@utwente.nl

EvoCOP 2023 – Brno, April 13, 2023

Error Correction Problem



- $\mu \in \{0,1\}^k$: message
- $c \in \{0,1\}^n$: codeword ($n > k$)
- $e \in \{0,1\}^n$: error pattern
- $z = c \oplus e$ (received word)

Error-Correcting Codes

Hamming Distance (HD) of $x, y \in \{0,1\}^n$: number of positions where x and y differ

Definition

(n, d_C) Binary (unrestricted) code of length n and minimum distance d_C: subset $C \subseteq \{0,1\}^n$ such that for all $c_1, c_2 \in C$

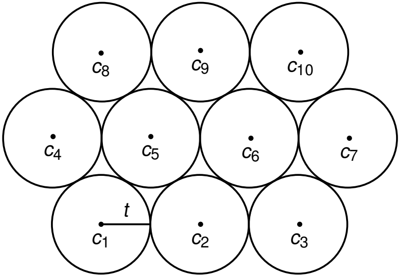
$$HD(c_1, c_2) \geq d_C$$

Example: a $(4,2)$ code $C \subseteq \{0,1\}^4$

0000	1001
0011	1010
0101	1100
0110	1111

Conflicting Requirements on Codes

- **High minimum distance d_C**
- **High number of codewords $c \in C$**



- Sphere of $c \in C \Leftrightarrow$ $S_c = \{z \in \mathbb{F}_2^n : \ d_H(z, c) \leq t\}$
- $t = \left\lfloor \frac{d-1}{2} \right\rfloor \Leftrightarrow$ Error-correction capability of C

Linear Codes

Notation:

- $\mathbb{F}_2 = \{0, 1\}$: finite field of order 2
- $\mathbb{F}_2^n = \{0, 1\}^n$: n-dimensional vector space over $\mathbb{F}_2$

Definition

A (n, k, d) binary linear code C: A (n, d) code C that is also a k-dimensional subspace of $\mathbb{F}_2^n$

$$g_1, \cdots, g_k \in \mathbb{F}_2^n \text{ basis of } C \Leftrightarrow G = \begin{pmatrix} g_1 \\ \vdots \\ g_k \end{pmatrix} k \times n \text{ generator matrix of } C$$

Encoding: vector-matrix multiplication

$$\mu \mapsto c = \mu G$$

Motivation

- Usual construction: **Algebraic Methods**
- **Metaheuristics** used, but only for:
  - Unrestricted codes [D90, M98, M12]
  - Similar combinatorial designs [K18, M18, M22b]

Research Question (the usual one)

Find $A(n,d)$: the max number of codewords for a given n and $d \Rightarrow$ instance of MAX-CLIQUE problem

- What about **Evolutionary Algorithms** (EA) for linear codes?

Research Question (ours)

Can EA discover *new* optimal linear codes?

Solutions Encoding and Search Space

- **Genotype**: a $k \times n$ binary matrix of full rank k

$$G = \begin{pmatrix} g_{1,1} & g_{1,2} & \cdots & g_{1,n} \\ g_{2,1} & g_{2,2} & \cdots & g_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ g_{k,1} & g_{k,2} & \cdots & g_{k,n} \end{pmatrix}$$

- **Phenotype**: subspace $C \subseteq \mathbb{F}_2^n$ spanned by G

$$C = \{c \in \mathbb{F}_2^n : c = x \cdot G,\ x \in \mathbb{F}_2^k\}$$

- **Search Space:** *Grassmannian* [M13] $\mathcal{S}_{n,k} = Gr(\mathbb{F}_2^n, k)$

$$|\mathcal{S}_{n,k}| = \binom{n}{k}_2 = \frac{(2^n-1)(2^{n-1}-1)\cdots(2^{n-k+1}-1)}{(2^k-1)(2^{k-1}-1)\cdots(2^{k-(k-2)}-1)}$$

Rank-preserving Mutation

Question: how do we preserve the rank with random mutations?

$$G = \begin{pmatrix} g_{1,1} & g_{1,2} & \cdots & g_{1,n} \\ g_{2,1} & g_{2,2} & \cdots & g_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ g_{k,1} & g_{k,2} & \cdots & g_{k,n} \end{pmatrix} \rightsquigarrow G' = \begin{pmatrix} g'_{1,1} & g'_{1,2} & \cdots & g'_{1,n} \\ g'_{2,1} & g'_{2,2} & \cdots & g'_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ g'_{k,1} & g'_{k,2} & \cdots & g'_{k,n} \end{pmatrix}$$

Idea: mutate at the *row* level [M22a]

1. Remove the i-th row of G
2. Span the subspace of the reduced matrix
3. Pick a random vector in the *complement* of the span
4. Insert the random vector in row i

Rank-preserving Crossover

Question: how to generalize this idea to crossover?

$$G = \begin{pmatrix} g_{1,1} & g_{1,2} & \cdots & g_{1,n} \\ g_{2,1} & g_{2,2} & \cdots & g_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ g_{k,1} & g_{k,2} & \cdots & g_{k,n} \end{pmatrix} \chi \begin{pmatrix} h'_{1,1} & h'_{1,2} & \cdots & h'_{1,n} \\ h'_{2,1} & h'_{2,2} & \cdots & h'_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ h'_{k,1} & h'_{k,2} & \cdots & h'_{k,n} \end{pmatrix} = H$$

1. Merge the rows of G and H
2. Randomly shuffle the rows
3. Select a subset of k linearly independent vectors

Fitness Function – Boolean Functions

- *n*-variable *Boolean function*: mapping $f : \mathbb{F}_2^n \to \mathbb{F}_2$
- Common representation: *truth table* Ω_f
- Other representation: *Algebraic Normal Form* (ANF) [C21]

Example with $n = 3$:

(x_1, x_2, x_3)	000	001	010	011	100	101	110	111
Ω_f	0	1	1	0	1	1	1	0
a_I	0	1	1	0	1	1	1	1

$$\Downarrow$$

$$f(x_1, x_2, x_3) = x_1 \oplus x_2 \oplus x_3 \oplus x_1 x_2 \oplus x_1 x_3 \oplus x_1 x_2 x_3$$

Degree of a monomial: # of factors

Fitness Function from ANF

Boolean function $f_C : \mathbb{F}_2^n \to \mathbb{F}_2$ associated to a code $C \subseteq \mathbb{F}_2^n$:

1. Initialize the truth table Ω_f to all zeros
2. For each $c \in C$, set $f(c) = 1$

Theorem ([C22])

The minimum distance of a linear (n, k, d) code $C \subseteq \mathbb{F}_2^n$ is:

$$d = \min\{|I| \in 2^{[n]} : a_I = 0\} \ ,$$

where a_I are the ANF coefficients of f_C

Fitness Function: maximize

$$fit(C) = \{I \in 2^{[n]} : |I| < d, \ a_I \neq 0\}| \ .$$

Experimental Settings

Optimization techniques:

- (μ,λ) and $(\mu+\lambda)$ Evolutionary Strategies
- Augmentation with crossover $(+\chi)$

Common parameters:

- (n,k,d): $(12,6,4)$, $(13,6,4)$, $(14,7,4)$, $(15,7,5)$, $(16,8,5)$
- $\lambda = n$, $\mu = \lfloor n/3 \rfloor$, $p_m ut = 1/n$
- Fitness budget: 20 000 generations
- Repetitions: 100

(n,k,d)	$\#\mathcal{S}_{n,k}$	$fit^*_{n,d}$	λ	μ	p_{mut}
$(12,6,4)$	$2.31 \cdot 10^{11}$	299	12	4	0.083
$(13,6,4)$	$1.49 \cdot 10^{13}$	378	13	4	0.077
$(14,7,4)$	$1.92 \cdot 10^{15}$	470	14	4	0.071
$(15,7,5)$	$2.47 \cdot 10^{17}$	1941	15	5	0.067
$(16,8,5)$	$6.34 \cdot 10^{19}$	2517	16	5	0.063

Results – Success Rate

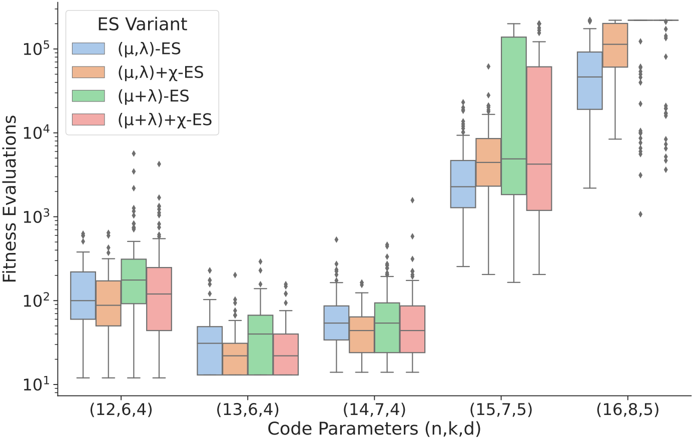
(n,k,d)	(μ,λ)-ES	(μ,λ)+χ-ES	$(\mu+\lambda)$-ES	$(\mu+\lambda)$+χ-ES
$(12,6,4)$	100	100	100	100
$(13,6,4)$	100	100	100	100
$(14,7,4)$	100	100	100	100
$(15,7,5)$	100	100	77	81
$(16,8,5)$	92	76	18	17

Table: Success rates (over 100 runs) of the four considered ES variants.

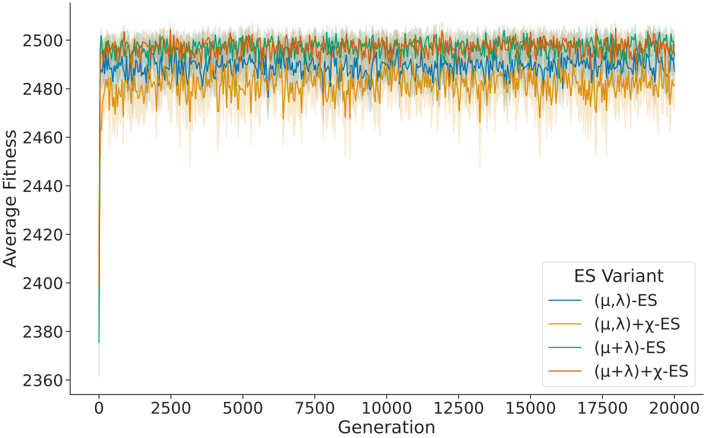
Main Findings:

- Easy for any variant up to $n = 14$
- Steep increase in difficulty from $n = 15$
- The simplest (μ,λ)-ES is the best one

Results – Fitness Evaluations



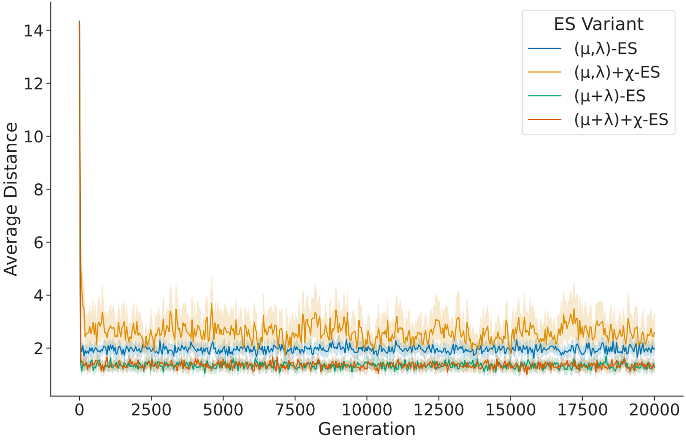
Results – Average Fitness



Solution Diversity – Distance

Distance between two subspaces $A, B \subseteq \mathbb{F}_2^n$:

$$d(A,B) = dim(A) + dim(B) - 2dim(A \cap B) = 2(k - dim(A \cap B))$$



Solutions Diversity – Equivalence

BKLC: Best Known Linear Code (found by MAGMA)

(n,k,d)	(μ,λ)-ES		$(\mu,\lambda)+\chi$-ES		$(\mu+\lambda)$-ES		$(\mu+\lambda)+\chi$-ES	
	#non-iso	#eq	#non-iso	#eq	#non-iso	#eq	#non-iso	#eq
$(12,6,4)$	100	23	100	22	100	22	100	22
$(13,6,4)$	100	85	100	81	100	78	100	79
$(14,7,4)$	100	89	100	94	100	95	100	93
$(15,7,5)$	72	5	63	6	51	5	44	5
$(16,8,5)$	0	1	0	1	0	1	0	1

Table: Number of non-isomorphic codes to the BKLC (#non-iso) and equivalence classes (#eq) found by the four considered ES variants.

Main Findings:

- For $d = 4$: all found codes are non-isomorphic to the BKLC
- For $n = 16$: situation reversed!

Conclusions and Future Works

Conclusions:

- Evolutionary Strategies are able to find many new unknown codes for small lengths
- There is a steep increase in difficulty of the problem from $n = 14$ to $n = 15$

Future work:

- Investigate why for $(16, 8, 5)$ all codes are equivalent
- Explore other fitness functions (ANF characterization is computationally cumbersome) [M18, M20]

References

[C22]Carlet, C.: Expressing the minimum distance, weight distribution and covering radius of codes by means of the algebraic and numerical normal forms of their indicators. Adv. Math. Commun. 16(4): 693–707 (2022)

[C21] Carlet, C.: Boolean functions for cryptography and coding theory. Cambridge University Press (2021)

[D90] Dontas, K., Jong, K.A.D.: Discovery of maximal distance codes using genetic algorithms. In: Proceedings of IEEE TAI 1990, pp. 805–811 (1990)

[K18] Knezevic, K., Picek, S., Mariot, L., Jakobovic, D., Leporati, A.: The design of (almost) disjunct matrices by evolutionary algorithms. In: Fagan, D., Mart'in-Vide C., O'Neill, M., Vega-Rodríguez, M.A. (eds.): TPNC 2018. LNCS vol. 11324, pp. 152–163. Springer (2018)

[M20] Manzoni, L., Mariot, L., Tuba, E.: Balanced crossover operators in Genetic Algorithms. Swarm Evol. Comput. 54: 100646 (2020)

[M22a] Mariot, L., Saletta, M., Leporati, A., Manzoni, L.: Heuristic search of (semi-)bent functions based on cellular automata. Nat. Comput. 21(3): 377–391 (2022)

[M22b]Mariot, L., Picek, S., Jakobovic, D., Djurasevic, M., Leporati, A.: On the difficulty of evolving permutation codes. In: Proceedings of EvoApplications 2022, LNCS vol. 13244, pp. 141–156 (2022)

[M18] Mariot, L., Picek, S., Jakobovic, D., Leporati, A.: Evolutionary Search of Binary Orthogonal Arrays. In: Auger, A., Fonseca, C.M., Lourenço, N., Machado, P., Paquete, L., Whitley, D. (eds.): PPSN 2018 (I). LNCS vol. 11101, pp. 121–133. Springer (2018)

[M12] McCarney, D.E., Houghten, S.K., Ross, B.J.: Evolutionary approaches to the generation of optimal error correcting codes. In: Proceedings of GECCO'12, pp. 1135–1142 (2012)

[M98] McGuire, K.M., Sabin, R.E.: Using a genetic algorithm to find good linear error-correcting codes. In: Proceedings of SAC'98, pp. 332–337 (1998)

[M13] Mullen, G.L., Panario, D.: Handbook of Finite Fields. CRC Press (2013)