Quantum statistical mechanics, Kolmogorov complexity, and the asymptotic bound for error-correcting codes

Matilde Marcolli (joint work with Yuri Manin)

This talk is based on:

ManMar2 Yuri I. Manin, Matilde Marcolli, *Kolmogorov complexity and the asymptotic bound for error-correcting codes*, arXiv:1203.0653, to appear in Journal of Differential Geometry

ManMar1 Yuri I. Manin, Matilde Marcolli, *Error-correcting codes and phase transitions*, Mathematics in Computer Science (2011) 5:133–170.

Error-correcting codes

- *Alphabet*: finite set A with $\#A = q \geq 2$.
- *Code*: subset $C \subset A^n$, *length* $n = n(C) \geq 1$.
- *Code words*: elements $x = (a_1, \ldots, a_n) \in C$.
- *Code language*: $\mathcal{W}_C = \cup_{m \geq 1} \mathcal{W}_{C,m}$, words $w = x_1, \ldots, x_m$; $x_i \in C$.
- ω*-language*: Λ_C, infinite words $w = x_1, \ldots, x_m, \ldots$; $x_i \in C$.
- Special case: $A = \mathbb{F}_q$, *linear codes*: $C \subset \mathbb{F}_q^n$ linear subspace
- in general: *unstructured codes*

- $k = k(C) := \log_q \#C$ and $[k] = [k(C)]$ integer part of $k(C)$

$$q^{[k]} \leq \#C = q^k < q^{[k]+1}$$

- *Hamming distance*: $x = (a_i)$ and $y = (b_i)$ in C

$$d((a_i),(b_i)) := \#\{i \in (1,\dots,n) \mid a_i \neq b_i\}$$

- *Minimal distance* $d = d(C)$ of the code

$$d(C) := \min\{d(a,b) \mid a,b \in C, a \neq b\}$$

Code parameters

- $R = k/n =$ *transmission rate* of the code
- $\delta = d/n =$ *relative minimum distance* of the code

Small R: fewer code words, easier decoding, but longer encoding signal; small δ: too many code words close to received one, more difficult decoding. Optimization problem: increase R and δ... how good are codes?

The space of code parameters:

- $Codes_q$ = set of all codes C on an alphabet $\#A = q$
- function $cp : Codes_q \to [0,1]^2 \cap \mathbb{Q}^2$ to code parameters
$cp : C \mapsto (R(C), \delta(C))$
- the function $C \mapsto (R(C), \delta(C))$ is a *total recursive map*
- *Multiplicity* of a code point (R, δ) is $\#cp^{-1}(R, \delta)$

Spoiling operations on codes: C an $[n,k,d]_q$ code

- $C_1 := C *_i f \subset A^{n+1}$

$$(a_1,\ldots,a_{n+1}) \in C_1 \text{ iff } (a_1,\ldots,a_{i-1},a_{i+1},\ldots,a_n) \in C\,,$$

and $a_i = f(a_1,\ldots,a_{i-1},a_{i+1}\ldots,a_n)$
C_1 an $[n+1,k,d]_q$ code (f constant function)

- $C_2 := C*_i \subset A^{n-1}$

$$(a_1,\ldots,a_{n-1}) \in C_2 \text{ iff } \exists b \in A,\ (a_1,\ldots,a_{i-1},b,a_{i+1},\ldots,a_n) \in C.$$

C_2 an $[n-1,k,d]_q$ code

- $C_3 := C(a,i) \subset C \subset A^n$

$$(a_1,\ldots,a_n) \in C_3 \text{ iff } a_i = a.$$

C_3 an $[n-1, k-1 \leq k' < k, d' \geq d]_q$ code

Asymptotic bound [Man1]

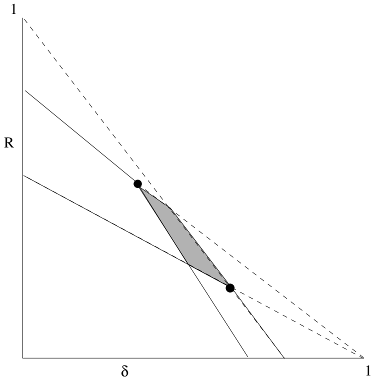
- $V_q \subset [0,1]^2$: all code points $(R,\delta) = cp(C)$, $C \in Codes_q$
- U_q: set of limit points of V_q
- Asymptotic bound: U_q all points below graph of a function

$$U_q = \{(R,\delta) \in [0,1]^2 \,|\, R \le \alpha_q(\delta)\}$$

- Isolated code points: $V_q \smallsetminus (V_q \cap U_q)$

[Man1] Yu.I.Manin, *What is the maximum number of points on a curve over $\mathbb{F}_2$?* J. Fac. Sci. Tokyo, IA, Vol. 28 (1981), 715–720.

Method: controlling quadrangles



$R = \alpha_q(\delta)$ continuous decreasing function with $\alpha_q(0) = 1$ and $\alpha_q(\delta) = 0$ for $\delta \in [\frac{q-1}{q}, 1]$; has inverse function on $[0, (q-1)/q]$; U_q union of all lower cones of points in $\Gamma_q = \{R = \alpha_q(\delta)\}$

Code points and multiplicities

Thm: [ManMar1] [Man2]

• Set of code points of infinite multiplicity $U_q \cap V_q = \{(R,\delta) \in [0,1]^2 \cap \mathbb{Q}^2 \,|\, R \leq \alpha_q(\delta)\}$ below the asymptotic bound

• Code points of finite multiplicity all above the asymptotic bound $V_q \smallsetminus (U_q \cap V_q)$ and isolated (open neighborhood containing (R,δ) as unique code point)

[Man2] Yu.I.Manin, *A computability challenge: asymptotic bounds and isolated error-correcting codes*, arXiv:1107.4246.

The computability question [Man2]

Other coarser bounds on codes:

- singleton bound: $R+\delta \leq 1$
- Gilbert–Varshamov line: $R=\frac{1}{2}(1-H_q(\delta))$

$$H_q(\delta)=\delta \log_q(q-1)-\delta \log_q \delta-(1-\delta) \log_q(1-\delta)$$

q-ary entropy (for linear codes GV line $R=1-H_q(\delta)$)

But no explicit expression for the asumptotic bound $R=\alpha_q(\delta)$:

- Is the function $R=\alpha_q(\delta)$ computable?
- Is there a characterization of good codes near or above the bound?

[ManMar2]: $R=\alpha_q(\delta)$ becomes computable with the help of an *oracle* that knows Kolmogorov complexity of codes...

Statistics of codes and the Gilbert–Varshamov bound

Known *statistical* approach to the GV bound: *random codes*

Shannon Random Code Ensemble: ω-language with alphabet A; uniform Bernoulli measure on Λ_A; choose code words of C as independent random variables in this measure

Volume estimate:

$$q^{(H_q(\delta)-o(1))n} \leq Vol_q(n, d = n\delta) = \sum_{j=0}^{d} \binom{n}{j} (q-1)^j \leq q^{H_q(\delta)n}$$

Gives probability of parameter δ for SRCE meets the GV bound with probability exponentially (in n) near 1: expectation

$$E \sim \binom{q^k}{2} Vol_q(n,d) q^{-n} \sim q^{n(H_q(\delta)-1+2R)+o(n)}$$

But... no good statistical description of the asymptotic bound

Kolmogorov complexity

X = *infinite constructive world*: have structural numbering computable bijections $\nu : \mathbb{Z}^+ \to X$ principal homogeneous space over group of total recursive permutations $\mathbb{Z}^+ \to \mathbb{Z}^+$

- *Ordering*: $x \in X$ is generated at the $\nu^{-1}(x)$-th step

Optimal partial recursive enumeration $u : \mathbb{Z}^+ \to X$ (Kolmogorov and Schnorr)

$$K_u(x) := \min\{k \in \mathbb{Z}^+ \mid u(k) = x\}$$

(exponential) Kolmogorov complexity

- changing $u : \mathbb{Z}^+ \to X$ changes $K_u(x)$ up to bounded (multiplicative) constants $c_1 K_v(x) \leq K_u(x) \leq c_2 K_v(x)$
- min length of program generating x (by Turing machine)

Warning: Kolmogorov complexity not a computable function

X, Y infinite constructive worlds, ν_X, ν_Y structural bijections, u, v optimal enumerations, K_u and K_v Kolmogorov complexities

• total recursive function $f : X \to Y \Rightarrow \forall y \in f(X)$, $\exists x \in X$, $y = f(x)$: $\exists$ computable $c = c(f, u, v, \nu_X, \nu_Y) > 0$

$$K_u(x) \leq c \cdot \nu_Y^{-1}(y)$$

Kolmogorov ordering
$\mathbf{K}_u(x)$ = order X by growing Kolmogorov complexity $K_u(x)$

$$c_1 K_u(x) \leq \mathbf{K}_u(x) \leq c_2 K_u(x)$$

So... if know how to generate elements of X in Kolmogorov ordering then can generate all elements of $f(X) \subset Y$ in their structural ordering

In fact... take $F(x) = (f(x), n(x))$ with

$$n(x) = \#\{x' \mid \nu_X^{-1}(x') \leq \nu_X^{-1}(x),\ f(x') = f(x)\}$$

total recursive function $\Rightarrow E = F(X) \subset Y \times \mathbb{Z}^+$ enumerable

- $X_m := \{x \in X \mid n(x) = m\}$ and $Y_m := f(X_m) \subset Y$ enumerable
- for $x \in X_1$ and $y = f(x)$: complexity $K_u(x) \leq c \cdot \nu_Y^{-1}(y)$ (using inequalities for complexity under composition)

Multiplicity: $mult(y) := \#f^{-1}(y)$

$$Y_\infty \subset \cdots f(X_{m+1}) \subset f(X_m) \subset \cdots \subset f(X_1) = f(X)$$

$Y_\infty = \cap_m f(X_m)$ and $Y_{fin} = f(X) \smallsetminus Y_\infty$

Prop: $y \in Y_\infty$ and $m \geq 1$: $\exists$ unique $x_m \in X$, $y = f(x_m)$, $n(x_m) = m$ and $c = c(f, u, v, \nu_X, \nu_Y) > 0$

$$K_u(x_m) \leq c \cdot \nu_Y^{-1}(y)\, m \log(\nu_Y^{-1}(y) m)$$

Oracle mediated recursive construction of Y_∞ and Y_{fin}

- Choose sequence (N_m, m), $m \geq 1$, $N_{m+1} > N_m$
- Step 1: $A_1 =$ list $y \in f(X)$ with $\nu_Y^{-1}(y) \leq N_1$; $B_1 = \emptyset$
- Step $m+1$: Given A_m and B_m, list $y \in f(X)$ with $\nu_Y^{-1}(y) \leq N_{m+1}$; $A_{m+1} =$ elements in this list for which $\exists\, x \in X$, $y = f(x)$, $n(x) = m+1$; $B_{m+1} =$ remaining elements in the list

- $A_m \cup B_m \subset A_{m+1} \cup B_{m+1}$, union is all $f(X)$; $B_m \subset B_{m+1}$ and $Y_{fin} = \cup_m B_m$, while $Y_\infty = \cup_{m\geq 1}(\cap_{n\geq 0} A_{m+n})$
- from A_m to A_{m+1} first add all new y with $N_m < \nu_Y^{-1}(y) \leq N_{m+1}$ then subtract those that have no more elements in the fiber $f^{-1}(y)$: these will be in B_{m+1}

Structural numbering for codes

- $X = Codes_q$, $Y = [0,1]^2 \cap \mathbb{Q}^2$ and $f : X \to Y$ is $cp : C \mapsto (R(C), \delta(C))$ code parameters map
- $A = \{0, \dots, q-1\}$ ordered, A^n lexicographically; computable total order ν_X:
(i) if $n_1 < n_2$ all $C \subset A^{n_1}$ before all $C' \subset A^{n_2}$;
(ii) $k_1 < k_2$ all $[n, k_1, d]_q$-codes before $[n, k_2, d']_q$-codes;
(iii) fixed n and q^k: lexicographic order of code words, concatenated into single word $w(C)$ (determines code): order all the $w(C)$ lexicographically
- total recursive map $cp : Codes_q \to [0,1]^2 \cap \mathbb{Q}^2$
- fixed enumeration ν_Y of rational points in $[0,1]^2$

Building the asymptotic bound

- C_m = set of code points with denominators dividing $m!$ (vertices of square lattice of size $1/m!$)
- Choose sequence N_m so that $\{y \,|\, \nu_Y^{-1}(y) \leq N_m\}$ contains C_m
- Plot points of $A_m \cap C_m$ and $B_m \cap C_m$
- *Saturated* subset of C_m: union of sets $S_{a,b} = \{(x,y) \,|\, x \leq a,\ y \leq b\}$, $(a,b) \in C_m$
- part of C_m below or on asymptotic bound is saturated
- D_m = maximal saturated subset of $A_m \cap C_m$
- upper boundary Γ_m of D_m is m-th step approximation to the asymptotic bound
- B_m is m-th step approximation of set of isolated code points
- points above Γ_m not in B_m sorted at subsequent step: end eventually either below asymptotic bound or in one of the B_{m+n}
- Question: is there a statistical view of this procedure?

Partition function for code complexity

$$Z(X,\beta) = \sum_{x \in X} K_u(x)^{-\beta}$$

weights elements in constructive world X by inverse complexity; $\beta =$ inverse temperature, thermodynamic parameter

- variant with prefix-free complexity $ZP(X,\beta) = \sum KP_v(x)^{-\beta}$
- prefix-free complexity: intrinsic characterization by Levin in terms of maximality for all probabilities enumerable from below $p : X \to \mathbb{R}_+ \cup \{\infty\}$,

$$\{(r,x)\,|\, r < p(x)\} \subset \mathbb{Q} \times X \quad \text{enumerable}$$

Convergence properties

• Kolmogorov complexity and Kolmogorov ordering

$$c_1\,\mathbf{K}_u(x) \leq K_u(x) \leq c_2\,\mathbf{K}_u(x)$$

• convergence of $Z(X,\beta)$ controlled by series

$$\sum_{x\in X} \mathbf{K}_u(x)^{-\beta} = \sum_{n\geq 1} n^{-\beta} = \zeta(\beta)$$

• Partition function $Z(X,\beta)$ convergence for $\beta > 1$; phase transition at pole $\beta = 1$

Asymptotic bound as a phase transition

- $X' \subset X$ infinite decidable subset of a constructive world
- $i : X' \hookrightarrow X$ total recursive function; also $j : X \to X'$ identity on X' constant on complement

$$K_u(i(x')) \leq c_1 K_v(x') \quad \text{and} \quad K_v(j(x)) \leq c_2 K_u(x)$$

- $\delta = \beta_q(R)$ inverse of $\alpha_q(\delta)$ on $R \in [0, 1 - 1/q]$
- Fix $R \in \mathbb{Q} \cap (0,1)$ and $\Delta \in \mathbb{Q} \cap (0,1)$

$$Z(R, \Delta; \beta) = \sum_{C : R(C) = R; 1-\Delta \leq \delta(C) \leq 1} K_u(C)^{-\beta + \delta(C) - 1}$$

Thm: Phase transition at the asymptotic bound

- $1 - \Delta > \beta_q(R)$: partition function $Z(R, \Delta; \beta)$ real analytic in β
- $1 - \Delta < \beta_q(R)$: partition function $Z(R, \Delta; \beta)$ real analytic for $\beta > \beta_q(R)$ and divergence for $\beta \to \beta_q(R)_+$

Classical statistical mechanical system on the space of codes

Partition function $Z(Codes_q, \beta) = \sum_{C \in Codes_q} K_u(C)^{-\beta}$ defines probability measure on $Codes_q$

$$\mathbb{P}_\beta(C) = \frac{K_u(C)^{-\beta}}{Z(Codes_q, \beta)}$$

Observables = computable functions; expectation values

$$\langle f \rangle_\beta = \int f(C)\, d\mathbb{P}_\beta(C) = \frac{1}{Z(Codes_q, \beta)} \sum_{C \in Codes_q} f(C)\, K_u(C)^{-\beta}$$

Measures and oracle aided plot of the asymptotic bound

Algorithm constructing A_m and B_m sets determines probability measures

$$\mathbb{P}_{B_m,\beta}(C) = \frac{K_u(C)^{-\beta}}{Z(cp^{-1}(B_m),\beta)}$$

$$\mathbb{P}_{E_{M,N},\beta}(C) = \frac{K_u(C)^{-\beta}}{Z(cp^{-1}(E_{M,N}),\beta)}$$

with $E_{M,N} = \cup_{m=1}^{M}(\cap_{n=0}^{N} A_{m+n})$, converging to

$$\mathbb{P}_{Y_{fin},\beta}(C) = \frac{K_u(C)^{-\beta}}{Z(cp^{-1}(Y_{fin}),\beta)}$$

$$\mathbb{P}_{Y_\infty,\beta}(C) = \frac{K_u(C)^{-\beta}}{Z(cp^{-1}(Y_\infty),\beta)}$$

Similarly get measures supported on Γ_m approximating measure on Γ asymptotic bound curve

Quantum statistical mechanical system on the space of codes

• Quantize the classical system: independent degrees of freedom
$\Rightarrow$ creation/annihilation operators

• for a single code C: code words are degrees of freedom

• Algebra of observable of a single code: Toeplitz algebra on code words

$$\mathcal{T}_C : \quad T_x, \quad x \in C, \quad T_x^* T_x = 1$$

$T_x T_x^*$ mutually orthogonal projectors

• Fock space representation $\mathcal{H}_C$ spanned by ϵ_w, words $w = x_1, \ldots, x_N$ in code language $\mathcal{W}_C$

$$T_x \, \epsilon_w = \epsilon_{xw}$$

QSM system of a single code

- algebra of observables $\mathcal{T}_C$; time evolution $\sigma : \mathbb{R} \to \mathrm{Aut}(\mathcal{T}_C)$

$$\sigma_t(T_x) = K_u(C)^{it}\, T_x$$

- Hamiltonian $\pi(\sigma_t(T)) = q^{itH}\pi(T)q^{-itH}$

$$H\,\epsilon_w = \ell(w)\, log_q K_u(C)\,\epsilon_w$$

in Fock representation, $\ell(w)$ length of word (# of code words)

- Partition function

$$Z(C,\sigma,\beta) = \mathrm{Tr}(e^{-\beta H}) = \sum_m (\# W_{C,m}) K_u(C)^{-\beta m}$$

$$= \sum_m q^{m(nR-\beta \log_q K_u(C))} = \frac{1}{1 - q^{nR} K_u(C)^{-\beta}}$$

- Convergence: $\beta > nr / \log_q K_u(C)$

QSM system at a code point (R, δ)

- Different codes $C \in cp^{-1}(R,\delta)$ as independent subsystems
- Tensor product of Toeplitz algebras $\mathcal{T}_{(R,\delta)} = \otimes_{C \in cp^{-1}(R,\delta)} \mathcal{T}_C$
- Shift on single code temperature so that

$$Z(C, \sigma, n(\beta - \delta + 1)) \leq (1 - K_u(C)^{-\beta})^{-1}$$

by *singleton bound* on codes $R + \delta - 1 \leq 0$

- Fock space $\mathcal{H}_{(R,\delta)} = \otimes \mathcal{H}_C$; time evolution $\sigma = \otimes \sigma^C$
- Partition function (variable temperature)

$$Z(cp^{-1}(R,\delta), \sigma; \beta) = \prod_{C \in cp^{-1}(R,\delta)} Z(C, \sigma, n(\beta - \delta + 1))$$

- Convergence controlled by $\prod_C (1 - K_u(C)^{-\beta})^{-1}$; in turned controlled by the classical zeta function $Z(cp^{-1}(R,\delta), \beta) = \sum_{C \in cp^{-1}(R,\delta)} K_u(C)^{-\beta}$

first versus second quantization

• Bosonic second quantization: example of primes p and integers $n \in \mathbb{N}$; independent degrees of freedom (primes) quantized by isometries $\tau_p^* \tau_p = 1$; tensor product of Toeplitz algebras $\otimes_p \mathcal{T}_p = C^*(\mathbb{N})$ semigroup algebra; $\sigma_t(\tau_p) = p^{it} \tau_p$, partition function $\zeta(\beta) = \prod_p (1 - p^{-\beta})^{-1}$ prod of partition functions individual systems

• Infinite tensor product: second quantization; finite tensor product: quantum mechanical (finitely many degrees of freedom) first quantization

• $(\mathcal{T}_{(R,\delta)}, \sigma)$ is quantum mechanical above the asymptotic bound; bosonic QFT below asymptotic bound

Asymptotic bound boundary between first and second quantization

Code parameters and Hausdorff dimensions [ManMar1]

• ω-language Λ_C of code C, infinite sequences of code words

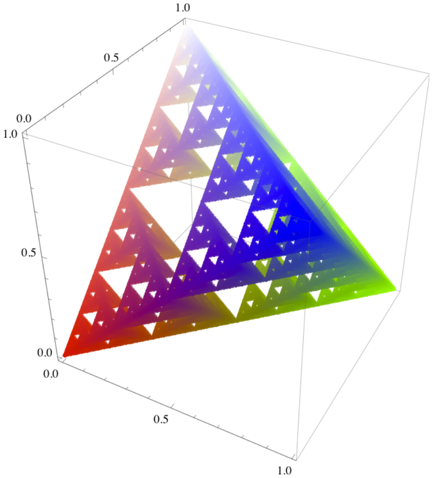
• Λ_C fractal in $[0,1]^n$ hypercube

• Hausdorff dimension $\dim_H(\Lambda_C) = R(C)$ rate of code

• min distance $d(C)$: threshold dim, lower dim slices (all directions parallel to coord axes) of Λ_C empty or singletons; higher dim some sections of positive Hausdorff dim

Example [MarPe]: unstructured $[3,2,2]_2$ code C, code words

$$\{(0,0,0),(0,1,1),(1,0,1),(1,1,0)\}$$



[MarPe] Matilde Marcolli, Christopher Perez, *Codes as fractals and noncommutative spaces*, arXiv:1107.5782.

Code algebra

- $\mathcal{T}_C$ Toeplitz algebra; quotient Cuntz algebra $\mathcal{O}_C$ by ideal $1 - P_C$, projector $P_C = \sum_x T_x T_x^*$
- $\mathcal{O}_C$ contains max abelian subalgebra $C(\Lambda_C)$; $\beta = \dim_H(\Lambda_C)$ unique inverse temperature for which KMS state for time evolution $\sigma_t(T_x) = q^{-itn} T_x$ on $\mathcal{O}_C$; KMS state Hausdorff measure on Λ_C
- $(\mathcal{T}_C, \sigma_t)$, $\sigma_t(T_x) = q^{-itn} T_x$, partition function

$$Z_C(\beta) = (1 - q^{(R-\beta)n})^{-1}$$

ω-language and complexity [ManMar1]

• Entropy of language $\mathcal{W}_C$, generating function:

$$s_C(m) = \#\mathcal{W}_{C,m}, \qquad G_C(t) = \sum_m s_C(m) t^m$$

Entropy: $\mathcal{S}_C = -\log_q \rho(G_C(t))$ with $\rho =$ radius of convergence

• $G_C(q^{-s}) = Z_C(s)$ partition function is generating function of language structure functions; entropy of language is code rate R

• complexity $K(w)$ of *words* in a language; for infinite words in ω-language Λ_C complexity $\kappa(w) = \liminf_{w_n \to w} K(w_n)/\ell(w_n)$

• Levin: $\kappa(w) = \liminf_{w_n \to w} \frac{-\log_q \mu_U(w_n)}{\ell(w_n)}$, universal enumerable semi-measure μ_U; bounds uniform Bernoulli measure on Λ_C so $\kappa(x) \leq \lim \frac{-\log_q \mu(w)}{\ell(w)} = R$ (achieved on full measure subset)

Asymptotic bound as a phase transition [ManMar1] (QSM point of view)

- Variable temperature partition function: $\mathcal{A} = \otimes_{s\in S}\mathcal{A}_s$, $\sigma = \otimes_s \sigma_s$; $\beta : S \to \mathbb{R}_+$; $Z(\mathcal{A},\sigma,\beta) = \prod_s Z(\mathcal{A}_s,\sigma_s,\beta(s))$
- fix a code point (R,δ); partition function (variable β)

$$Z((R,\delta),\sigma;\beta) = \prod_{C\in cp^{-1}(R,\delta)} (1 - q^{(R-\beta)n_C})^{-1}$$

- if (R,δ) above bound finite product; if below bound convergence governed by $\sum_C q^{(R-\beta)n_C}$, for $\beta > R$.
- change of behavior of the system at $R = \alpha_q(\delta)$ asymptotic bound