Notes on the complexity of coverings for Kronecker powers of symmetric matrices

Igor S. Sergeev*

Abstract

In the present note, we study a new method of constructing efficient coverings for Kronecker powers of matrices, recently proposed by J. Alman, Y. Guan, A. Padaki [1]. We provide an alternative proof for the case of symmetric matrices in a stronger form. As a consequence, the previously known upper bound on the depth-2 additive complexity of the boolean $N \times N$ Kneser-Sierpinski matrices is improved to $O(N^{1.251})$. This work can be viewed as a supplement to [3].

1 Introduction

Let us recall necessary concepts. See [3] for a more detailed introduction to the subject.

A *rectangle* of size $a \times b$ is an all-1s matrix with a rows and b columns. Further, depending on the context, sometimes under rectangle we will understand a rank-1 matrix, i.e. consisting of an all-1s submatrix and all 0s in other entries.

We define the *complexity*[1] of an $a \times b$ rectangle R as the sum of lengths of its two sides, $w(R) = a+b$. We introduce the characteristic of the narrowness of a rectangle as the ratio of the lengths of its larger and smaller sides, $\rho(R) = \frac{\max(a,b)}{\min(a,b)}$. The *spectral weight* of a rectangle is defined as $\sigma(R) = \sqrt{ab}$.

A set $F = \{R_1, \ldots, R_k\}$ of rectangles is a *covering* of a boolean matrix A, if

$$A = R_1 + \ldots + R_k. \tag{1}$$

(Here under R_i we mean rank-1 matrices.) If the operation "+" in (1) is an integer addition, then F is called SUM-covering. If "+" is a disjunction, then

*e-mail: isserg@gmail.com

[1]In [3], we used a term *weight* instead. Here we substitute it with complexity to avoid confusing with the spectral weight of a rectangle.

F is called OR-covering. In the case when "+" is a mod 2 addition, then we have an XOR-covering.

The complexity of a covering F is defined as $w(F) = w(R_1) + \ldots + w(R_k)$, and the spectral weight as $\sigma(F) = \sigma(R_1) + \ldots + \sigma(R_k)$. The L-*complexity* of a matrix A is defined for $\mathsf{L} \in \{\mathsf{SUM}, \mathsf{OR}, \mathsf{XOR}\}$ as the minimal complexity of its L-covering, denoted by $\mathsf{L}_2(A)$ (it means the complexity of computation of A by depth-2 linear circuits of the corresponding type).

► These notions may be extended to the case of matrices over an arbitrary semiring S. A rectangle over S is a matrix $(c_1, \ldots, c_a)^T \cdot (d_1, \ldots, d_b)$, where $c_i, d_j \in S \setminus \{0\}$. A covering of a matrix is conditioned by

$$A = e_1 R_1 + \ldots + e_k R_k, \qquad e_i \in S.$$

The results presented below may be applied also to the analogously defined measure of complexity of computation of matrices by algebraic linear circuits of depth 2. ◄

Let $\sigma(A)$ denote the minimal spectral weight of a matrix A. Since $w(R) \geq 2\sigma(R)$ for any rectangle R, spectral weight serves as a simple lower bound for complexity[2]: $\mathsf{L}_2(A) \succeq \sigma(A)$.

A convenient property of the spectral weight is its multiplicativity with respect to the Kronecker product. Recall that the *Kronecker product* of boolean matrices A, B is a matrix $A \otimes B$ obtained by replacing 1-entries of A by copies of B, and 0-entries by all-0s matrices of the same size.

Note that if F and G are coverings of matrices A and B, then $F \otimes G$ is a covering[3] of $A \otimes B$, and $\sigma(F \otimes G) = \sigma(F)\sigma(G)$. In particular, if we construct a covering of a matrix $A^{\otimes n}$ (a Kronecker power of A) by the product-of-coverings method above using appropriate coverings of A, then the complexity of a resulting covering H satisfies $w(H) \succeq \sigma(H) \succeq \sigma^n(A)$.

In [1], the authors actually pose a question: can we obtain upper bounds like $\mathsf{L}_2(A^{\otimes n}) \preceq \sigma^{n+o(n)}(A)$ or at least $\mathsf{L}_2(A^{\otimes n}) \preceq \sigma^{n+o(n)}(F)$ for some appropriate coverings F of a matrix A. In general, it is not possible, just consider an example $A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$.

A general obstacle for the desired bounds is the growing narrowness of the covering rectangles. Note that the complexity and the spectral weight of a rectangle R are related as[4] $w(R) \asymp \sqrt{\rho(R)}\sigma(R)$. Nevertheless, under some conditions, the authors of [1] were able to overcome the indicated obstacle

[2]Here and below, symbols $\asymp$, $\prec$, $\preceq$ denote the equality, strict and non-strict inequalities on the order of growth.

[3]The Kronecker product of sets of matrices is $F \otimes G = \{R \otimes R' \mid R \in F,\ R' \in G\}$.

[4]To be precise, $w(R) = \left(\sqrt{\rho(R)} + \sqrt{1/\rho(R)}\right) \sigma(R)$.

and to derive the desired bounds. The first situation is when the covering F satisfies some asymmetry criteria, the second one is when the matrix A is symmetric, and the covering F is one-sided (it means that all rectangles are stretched in the same direction).

Perhaps, the most challenging object (among boolean matrices) to apply the theory are the Kneser–Sierpinski (or disjointness) matrices. Recall that the boolean $N \times N$ *Kneser–Sierpinski matrix* D_N is defined for $N = 2^n$ as follows. Rows and columns of D_N are labeled by distinct subsets $u \in [n]$, and $D[u, v] = 1$ iff $u \cap v = \emptyset$. A matrix D_N also may be viewed as a Kronecker product

$$D_N = D_2^{\otimes n} = \underbrace{D_2 \otimes \ldots \otimes D_2}_{n}, \qquad D_2 = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}. \tag{2}$$

The problem of complexity of OR-coverings for D_N was almost closed in [2]. There was established that

$$N^{1.16} \prec \mathsf{OR}_2(D_N) \prec N^{1.17}$$

(the lower bound is from [3]). Moreover, the authors of [2] constructed a covering of almost minimal complexity, up to a factor of order $(\log N)^{O(1)}$.

The question about additive (SUM) complexity of matrices is less clear. In [3], we propose a simple way to show that

$$\mathsf{SUM}_2(D_N) \preceq \sigma^n(D_2) = \left(\sqrt{2} + 1\right)^n \prec N^{1.272}$$

relying on a trivial (and optimal) decomposition of the matrix D_2 into rectangles of size 1×1 and 1×2 (or 2×1). This approach was nontrivially generalized in [1]. Due to limitations inherent in the analysis of the proposed method, the efficient implementation of the Kneser–Sierpinski matrices is justified only with the basic coverings of matrices D_4 and D_8. In the latter case, the obtained bound [1] is

$$\mathsf{SUM}_2(D_N) \preceq \sigma^{n/3+o(n)}(D_8) = \left(\sqrt{8} + \sqrt{7} + 3\sqrt{3} + 3\right)^{n/3+o(n)} \prec N^{1.258}.$$

In the present note, we describe a version of this method for the (most interesting) case of symmetric matrices. The limits of applicability of the method are (comparatively) extended, and a more elementary proof is given. As a consequence, an upper bound for the complexity of the Kneser–Sierpinski matrices is reduced to $\mathsf{SUM}_2(D_N) \prec N^{1.251}$.

Note that the question about existence of substantially more efficient XOR-coverings for matrices D_N is still open.

2 The synthesis method

In this section, we provide a general method for constructing a covering of a symmetric matrix $A^{\otimes n}$. The method is essentially equivalent to the method [1]. By the way, we will follow an illustrating example where a non-trivial covering of D_N is obtained from coverings of the matrix D_4. This example doesn't require the method in its full generality.

Assume there exist two coverings of a symmetric matrix A, or to be precise, two pairs of coverings, if we count transposed ones. The first covering F (F^T) is supposed to be efficient in the terms of spectral weight. The second covering G is one-sided: the longest sides of all its non-square rectangles are parallel. Coverings G/G^T serve to compensate an imbalance caused by the use of F-type coverings.

► Fig. 1 shows appropriate coverings of D_4. The covering F_2 (on the left) has optimal spectral weight $\sigma(F_2) = 4 + \sqrt{3}$. The covering G_2 (on the right) has good correcting qualities. Its spectral weight is slightly higher, $\sigma(G_2) = 3 + 2\sqrt{2}$. ◄

F_2 G_2

Figure 1: Weight-minimal and compensating coverings of D_4

Let H be a covering of a matrix B. Then we can build a covering of a matrix $A \otimes B$ in the form $\{F_i \otimes R_i \mid R_i \in H\}$, where F_i are some coverings of A. This way we sequentially obtain coverings for matrices $A, A^{\otimes 2}, A^{\otimes 3}, \ldots$ In the main process, we choose $F_i \in \{F, F^T\}$. Precisely, we transform an $a \times b$ rectangle R into $F \otimes R$, if $a \leq b$, and into $F^T \otimes R$, otherwise.

Consider a covering $F = \{R_1, \ldots, R_s\}$ consisting of $a_i \times b_i$ rectangles R_i. The *characteristic function* of F is defined as

$$\chi_F(x) = \sigma(R_1) \left(\frac{a_1}{b_1}\right)^x + \ldots + \sigma(R_s) \left(\frac{a_s}{b_s}\right)^x - \sigma(F). \qquad (3)$$

As follows from the definition, $\chi_F(0) = 0$. We call a covering F *compact*, if $\chi_F(x)$ takes negative values on some negative arguments[5]. For a com-

[5]It is a weak analogue of imbalanced covering from [1].

pact covering F, let λ_F denote the minimal real root of χ_F in whose right semineighbourhood the function is negative[6].

Note that the condition $\chi'_F(0) = \sum_i \sigma(R_i)\ln(a_i/b_i) > 0$ is sufficient for F to be compact.

So, we require the compactness of a covering F for our algorithm of computation of a Kronecker power of a matrix A.

► The covering F_2 of D_4 is compact. Its characteristic function is $\chi_{F_2}(x) = 2 \cdot 4^x + \sqrt{3} \cdot 3^{-x} - 2 - \sqrt{3}$, and the minimal root is $\lambda_{F_2} \approx -0.305$. ◄

For compensation, we will use a compact *one-sided* covering G consisting of $a'_i \times b'_i$ rectangles R'_i satisfying $a'_i \geq b'_i$. The quality of such covering is characterized by the coefficient

$$\mu_G = \frac{1}{\sigma(G)} \sum_{R \in G} \frac{\sigma(R)}{\sqrt{\rho(R)}} = \frac{1}{\sigma(G)} \sum_i b'_i.$$

For $x > 1$, define the function

$$\pi_G(x) = \frac{1}{\sigma(G)} \sum_{R \in G} \sigma(R) \cdot x^{-\lfloor \log_x \rho(R) \rfloor / 2}. \tag{4}$$

It easily follows from the definition that $\pi_G(x) \geq \mu_G$, and $\pi_G(x) \to \mu_G$ as $x \to 1$.

► For the covering G_2, we have $\mu_{G_2} = \frac{4}{3+2\sqrt{2}}$. ◄

Theorem 1. *Let F be a compact L-covering, and G be a compact one-sided L-covering of a symmetric $r \times r$ matrix A, and $\sigma(G) \geq \sigma(F)$. If the condition*

$$\frac{\sigma(G)}{\sigma(F)} < \mu_G^{2\lambda_F} \tag{5}$$

is satisfied, then for $N = r^n$,

$$\mathsf{L}_2(A^{\otimes n}) \preceq N^{\log_r \sigma(F)}.$$

► The matrix D_4 and its coverings F_2, G_2 satisfy the conditions of the theorem, since $\lambda_{F_2} < -0.3$ and $\mu_{G_2} = \frac{4}{3+2\sqrt{2}}$. Hence, $\mathsf{SUM}_2(D_N) \preceq N^{\log_4(4+\sqrt{3})} \prec N^{1.26}$. ◄

[6] The compactness of a covering implies that $b_i < a_i$ for some i. Then, λ_F is correctly defined since $\chi_F(x) \to +\infty$ as $x \to -\infty$, and any exponential sum of the form (3) has a finite number of real zeros, see e.g. [4].

Proof. The proof strategy is the following. First, we analyze the evolution of rectangle sizes after multiple application of type-F coverings. Then, do the same for type-G coverings. Finally, we propose an appropriate combination of these two types of coverings.

I. By compactness of the covering F, for some (small enough) $\delta, \epsilon > 0$, and $\lambda = \lambda_F + \delta$ such that $\chi_F(\lambda) < -\epsilon$, the inequality (5) holds true after replacing λ_F by λ. Let us check that for all small enough $\tau > 1$,

$$\sigma(R_1)\tau^{\lambda\lfloor \log_\tau(a_1/b_1)\rfloor} + \ldots + \sigma(R_s)\tau^{\lambda\lfloor \log_\tau(a_s/b_s)\rfloor} \leq \sigma(F). \tag{6}$$

Indeed, the left side of (6) doesn't exceed

$$\tau^{-\lambda}(\chi_F(\lambda) + \sigma(F)) < \tau^{-\lambda}(\sigma(F) - \epsilon),$$

thus to satisfy (6), it is sufficient to require $\tau^{-\lambda} \leq 1 + \frac{\epsilon}{\sigma(F)}$. Therefore, any choice from the interval $1 < \tau \leq \left(1 + \frac{\epsilon}{\sigma(F)}\right)^{-1/\lambda}$ is suitable.

We assign to the parameter τ the meaning of a discretization step of changing the ratios between the rectangle's sides in the classification of rectangles. The final choice of τ will be decided later.

Let us introduce a classification on the set $\mathcal{R}$ of all rectangles depending on the ratio between the longer and the shorter sides. Set $\mathcal{R} = \bigcup_{k\geq 0} I_k$, where I_0 contains rectangles R satisfying $\rho(R) \leq r$, and for $k \geq 1$, the set I_k contains rectangles with ratios $r \cdot \tau^{k-1} < \rho(R) \leq r \cdot \tau^k$. Recall that r is the size of the matrix A.

► In the example with the covering F_2, we set $\tau = 4$. Then, $I_0 = \{R \mid \rho(R) \leq 4\}$, and $I_k = \{R \mid 4^k < \rho(R) \leq 4^{k+1}\}$ for $k > 0$. The function $\chi_{F_2}(x)$ is negative in the interval $(\lambda_{F_2}, 0)$, thus we are quite free in choosing λ. The specific value will be determined later. ◄

When performing compositions with coverings of A, we will track the distribution of the spectral weight of rectangles among the sets I_k. In doing so, we will be guided by the principle of *error to the right*. It means that: (a) we allow a rectangle to be placed into a set I_k with a higher index k, but not otherwise, and (b) we estimate the distribution of the weight of a covering of $A \otimes R$ only on the basis of information about assigning R to a certain set I_k. As a consequence, the estimated distribution of a covering obtained as a result of a series of iterations (with possible errors to the right) may differ from the true distribution only in that some rectangles appear in sets I_k with higher indices.

The redistribution of the spectral weight of a set of rectangles under the composition with type-F coverings (with possible errors to the right) may be estimated from the coefficients of the Laurent polynomial

$$P_F(x) = \sum_{i\in\mathbb{Z}} \beta_i x^i = \frac{\sigma(R_1)}{\sigma(F)} x^{\lfloor \log_\tau(a_1/b_1)\rfloor} + \ldots + \frac{\sigma(R_s)}{\sigma(F)} x^{\lfloor \log_\tau(a_s/b_s)\rfloor} \qquad (7)$$

obtained from (6). It means that for $R \in I_m$, the weight of rectangles from $(F/F^T) \otimes R$ is distributed so that the portion β_k of the total weight belongs to I_{m-k} in the case $m-k>0$, and to I_0, otherwise.

► For the covering F_2, and $\tau = 4$, we obtain $P_{F_2}(x) = \frac{2x+2+\sqrt{3}x^{-1}}{4+\sqrt{3}}$. The corresponding redistribution diagram is shown on Fig. 2. ◄

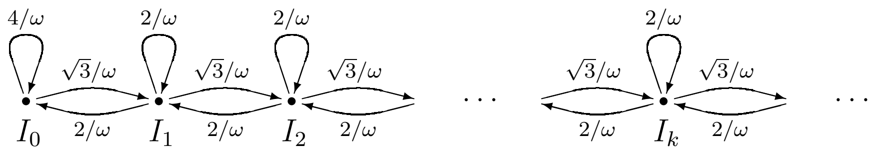

Figure 2: Diagram of spectral weight redistribution under the action of the composition with the covering F_2/F_2^T (here $\omega = \sigma(F_2) = 4+\sqrt{3}$)

Let $p_k(t)$ stand for the fraction of the spectral weight of the constructed covering of $A^{\otimes t}$ associated with the set I_k. In the beginning, one has $p_0(0) = 1$, and $p_k(0) = 0$ for all $k > 0$. Set $\nu = \tau^\lambda$. By (6), $P_F(\nu) \le 1$. Denote $d = \deg P_F = \max_{\beta_i>0} |i|$.

We are going to show that the distribution $\{p_k^*(t)\}$ with $p_k^*(t) = \nu^k$ for $k \le dt$, and $p_k^*(t) = 0$ for all $k > dt$, is a majorant for $\{p_k(t)\}$, meaning that the values $p_k^*(t)$ upper bound the components of some distribution $\{p_k'(t)\}$ obtained from $\{p_k(t)\}$ by a partial shift of the distribution to the right: from components with smaller indices to components with greater indices[7].

Obviously, in the moment $t = 0$, the majorization condition is fulfilled. Let us prove the induction step: apply the composition with F/F^T to a set of rectangles with the distribution $\{p_k^*(t)\}$. Rectangles from I_k, where $k > d(t+1)$, do not appear here. For $0 < k \le d(t+1)$, the weight of rectangles from I_k may be upper bounded as

$$\sum_{i\in\mathbb{Z}} \beta_i \nu^{k+i} = \nu^k P_F(\nu) \le \nu^k. \qquad (8)$$

In the case $k = 0$, this bound is, generally speaking, wrong, since an essential portion of the total weight remains in I_0.

[7]Though the distribution $\{p_k(t)\}$ is probabilistic, we don't require the same from the majorant $\{p_k^*(t)\}$ allowing the sum of its components be greater than 1. Our goal is just deriving upper bounds on $p_k(t)$.

However, if I_0 receives exceptional weight, then other sets $I_1, I_2, \ldots$ (in general) suffer from the weight deficit, as follows from (8). Therefore, we can redistribute the exceptional weight from I_0 to other sets, in accordance to the error-to-the-right principle. It is possible, since the composition preserves the conditional total weight, i.e. the sum of distribution components.

As a consequence, after t steps of composition with coverings F/F^T, any set I_k contains a portion at most ν^k of the total weight $\sigma^t(F)$ of the resulting covering of $A^{\otimes t}$, up to some errors to the right.

► In our example with D_4, we may choose $\nu = \sqrt{3}/2$ (it's a root of the polynomial $P_{F_2}(x)$). The choice allows us to avoid a redistribution of the weight from I_0. The distribution $\{1, \nu, \nu^2, \ldots, \nu^k, \ldots\}$ is stationary for the diagram on Fig. 2. Implicitly, we also chose $\lambda = \log_4 \nu > \lambda_{F_2}$. ◄

II. The redistribution of the spectral weight of a set of rectangles under the composition with type-G coverings (again, with possible errors to the right) may be described by the polynomial

$$P_G(x) = \alpha_l x^l + \ldots + \alpha_1 x + \alpha_0 = \frac{1}{\sigma(G)} \sum_{R \in G} \sigma(R) x^{\lfloor \log_\tau \rho(R) \rfloor}.$$

For $R \in I_m$, the spectral weight of rectangles from $(G/G^T) \otimes R$ is distributed so that the portion α_k belongs to I_{m-k} in the case $k < m$, and to I_0, otherwise.

► Actually, for the covering G_2, and $\tau = 4$, one can assign (assuming some errors to the right) $P_{G_2}(x) = (x+2)/3$. The weight redistribution under the action of coverings G_2/G_2^T is shown by the diagram on Fig. 3. ◄

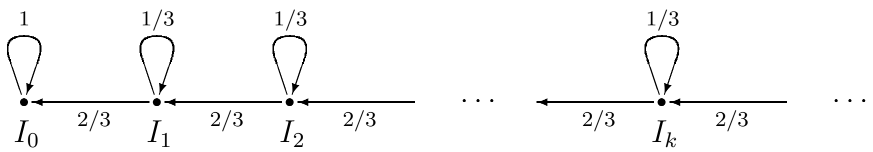

Figure 3: Diagram of spectral weight redistribution under the action of the composition with the covering G_2/G_2^T

Assuming that the initial weight distribution has a form $q_m(0) = 1$, $q_i(0) = 0$ for all $i \neq m$, that is, all rectangles are located in I_m, then after t steps of compositions with G-type coverings, we obtain a distribution[8]

$$q_{m-k}(t) = \sum_{\substack{k_1+\ldots+k_l \le t \\ k_1+2k_2+\ldots+lk_l=k}} C_t^{k_1,\ldots,k_l} \alpha_1^{k_1} \cdot \ldots \cdot \alpha_l^{k_l} \alpha_0^{t-(k_1+\ldots+k_l)} \tag{9}$$

for $0 \le k < m$. For a component associated with I_0, we use a trivial estimate $q_0(t) \le 1$. By consideration, if $\alpha_i = 0$, then $k_i = 0$, and the corresponding factor $\alpha_i^{k_i}$ in (9) should be replaced by 1. This remark will be implied in further calculations. The total weight of the (considered part of the) covering under construction will increase $\sigma^t(G)$ times in t steps.

Note that $P_G(1/\sqrt{\tau}) = \pi_G(\tau)$, see (4). Recall that $\pi_G(x) \to \mu_G$ as $x \to 1$. Our final choice of τ is such that the inequality (5) remains valid after replacing λ_F by λ, and μ_G by $\pi_G(\tau)$ (τ should be small enough).

Observe that our choice implies $\alpha_0 \ne 1$, since otherwise $\pi_G(\tau) \equiv 1$, and the inequality (5) cannot be satisfied. On the other hand, $\mu_G < 1$ due to the compactness of G: there should exist rectangles $R \in G$ with $\rho(R) > 1$.

III. Now we are ready to state the synthesis algorithm, namely the rule of combination of the coverings F and G. Choose γ satisfying the condition

$$-\frac{1}{\lambda} \log_\tau \frac{\sigma(G)}{\sigma(F)} < \gamma \le -2 \log_\tau \pi_G(\tau). \tag{10}$$

Such γ does exist, since the inequality between the left and the right sides of (10) is equivalent to (5), where λ_F is replaced by λ, and μ_G is replaced by $\pi_G(\tau)$ (just apply a base-τ logarithm to (5) and divide by λ).

To construct the required covering of $A^{\otimes n}$, we assign two sets of rectangles, $\mathcal{F}$ and $\mathcal{G}$.

(i) Before the start of the algorithm, the set $\mathcal{G}$ is empty, and the set $\mathcal{F}$ contains a 1×1 rectangle (a trivial covering of the matrix $A^{\otimes 0}$).

(ii) Then, perform n similar steps. A step t does the following:

— with the use of a suitable covering F or F^T, transform any rectangle $R \in \mathcal{F}$ into $(F/F^T) \otimes R$;

— with the use of a suitable covering G or G^T, transform any rectangle $R \in \mathcal{G}$ into $(G/G^T) \otimes R$;

— relocate from $\mathcal{F}$ to $\mathcal{G}$ all rectangles $R \in \mathcal{F}$ belonging to the sets I_m with $m \ge \gamma(n-t)$.

(iii) By construction, after any t steps, the set $\mathcal{F} \cup \mathcal{G}$ is a covering of the matrix $A^{\otimes t}$. In the end of the algorithm, the set $\mathcal{F}$ is empty, and $\mathcal{G}$ is a covering of $A^{\otimes n}$.

[8] Here $C_t^{k_1,\ldots,k_l}$ stands for the multinomial coefficient representing the number of ways to select from t elements l groups, with k_i elements in i-th group.

Essentially, this is the algorithm [1]. The parameter γ controls the switching between the two stages of the algorithm. Next, we are going to prove that $\mathcal{G}$ has the desired complexity.

For any m, we have to relocate rectangles belonging to $I_m \cap \mathcal{F}$ on at most $\frac{d}{\gamma}+2=O(1)$ (consecutive) steps, namely while $m-d<\gamma(n-t)\leq m$, and one more time, when $m-d\geq\gamma(n-t)$. On the subsequent steps, rectangles in I_m don't appear.

Let us turn to complexity bounds. The complexity of a rectangle $R\in I_m$ in relation to its spectral weight is estimated to be higher, when m is greater. Recall that $w(R)\asymp\sqrt{\rho(R)}\sigma(R)$. Therefore, an erroneous assignment of a rectangle to a set I_m with a higher index m (an error to the right) leads to overestimation of the complexity.

The complexity of the part of the covering $\mathcal{G}$ derived from rectangles that belonged to $I_m \cap \mathcal{F}$ at the moment of relocation is bounded from above as

$$L_m \preceq \sigma^n(F)\nu^m\left(\frac{\sigma(G)}{\sigma(F)}\right)^{\frac{m}{\gamma}}\cdot$$
$$\cdot\left(1+\sum_{k_1+\ldots+k_l\leq\frac{m}{\gamma}} C_{\frac{m}{\gamma}}^{k_1,\ldots,k_l}\alpha_1^{k_1}\cdot\ldots\cdot\alpha_l^{k_l}\alpha_0^{\frac{m}{\gamma}-(k_1+\ldots+k_l)}\tau^{\frac{m-(k_1+2k_2+\ldots+lk_l)}{2}}\right). \quad (11)$$

Here the first factor $\sigma^n(F)$ is the expected weight of the entire covering under the (optimistic) assumption that we apply only F-type coverings. The second factor ν^m is the upper bound for the weight portion of the covering $\mathcal{F}$ associated with I_m at the moment of relocation. The third factor reflects the weight increase caused by the application of G-type coverings instead of F/F^T on the last $m/\gamma-O(1)$ steps of the algorithm. In brackets, an additional factor taking into account the final distribution of the spectral weight is written. Namely, the term 1 represents the complexity of rectangles from I_0, and under the sum are written the products of the partial weight portions of rectangles from I_{m-k} with $k=k_1+2k_2+\ldots+lk_l$ (provided by (9)), and the estimate $\tau^{\frac{m-k}{2}}$ of the ratio between the complexity and the spectral weight of rectangles from I_{m-k}. The summands with $m\leq k$ are excess.

Multinomial coefficients satisfy the standard inequality[9]

$$C_n^{k_1,\ldots,k_l}\leq 2^{nH(k_1/n,\ldots,k_l/n)}, \qquad \text{where}$$
$$H(x_1,\ldots,x_l)=-\sum_{i=0}^{l}x_i\log_2 x_i, \qquad x_0=1-\sum_{i=1}^{l}x_i,$$

[9]Easily follows by induction on l from the well-known relation $C_n^k\leq 2^{H(k/n)}$.

is the binary entropy function defined on the l-dimensional simplex[10] $\mathbb{R}_+^l \cap \{x_1 + \ldots + x_l \le 1\}$.

It is easy to check that for any $c_i > 0$,

$$H(x_1, \ldots, x_l) + \sum_{i=1}^{l} x_i \log_2 c_i \le \log_2 \left(1 + \sum_{i=1}^{l} c_i\right). \tag{12}$$

Indeed, it immediately follows from the variant of the Hölder's inequality

$$\prod_{i=0}^{l} a_i^{b_i} \le \sum_{i=0}^{l} a_i b_i$$

that holds for $a_i, b_i > 0$, and $\sum b_i = 1$, see e.g. [5, Ch. V] (just assign $b_i = x_i$, $a_i = c_i/x_i$, where $c_0 = 1$, and take a logarithm; in the case $x_i = 0$ for some i, evaluate the limit).

By setting $k_i = \frac{x_i m}{\gamma}$ for $i = 1, \ldots, l$, and applying (12), we obtain

$$\log_2 \left(C_{\frac{m}{\gamma}}^{k_1,\ldots,k_l} \alpha_1^{k_1} \cdot \ldots \cdot \alpha_l^{k_l} \alpha_0^{\frac{m}{\gamma} - (k_1 + \ldots + k_l)} \tau^{\frac{m - (k_1 + 2k_2 + \ldots + lk_l)}{2}} \right) \le \\ \frac{m}{\gamma} \left(H(x_1, \ldots, x_l) + \sum_{i=1}^{l} x_i \log_2 \frac{\alpha_i}{\alpha_0 \tau^{i/2}} + \log_2 \left(\alpha_0 \tau^{\gamma/2}\right) \right) \le \\ \frac{m}{\gamma} \cdot \left(\log_2 \left(\sum_{i=0}^{l} \frac{\alpha_i}{\tau^{i/2}} \right) + \log_2 \left(\tau^{\gamma/2}\right) \right).$$

Thus, we continue (11) as

$$L_m \preceq \sigma^n(F) \left(C_0^{\frac{m}{\gamma}} + m^l C_1^{\frac{m}{\gamma}} \right), \quad \text{where}$$

$$C_0 = \frac{\sigma(G)\nu^\gamma}{\sigma(F)}, \qquad C_1 = C_0 \cdot P_G\left(1/\sqrt{\tau}\right) \cdot \tau^{\gamma/2}.$$

Here the power of C_0 corresponds to the contribution of rectangles from I_0, and the power of C_1 to the contribution of the remaining rectangles.

Let us check that $C_1 \le C_0 < 1$. Indeed, in the case $\sigma(G) = \sigma(F)$, the inequality $C_0 = \nu^\gamma < 1$ holds trivially. In the other case $\sigma(G) > \sigma(F)$, from the left part of (10), it follows that

$$C_0 = \frac{\sigma(G)}{\sigma(F)} \tau^{\lambda\gamma} < 1.$$

[10]On the boundary, the function is defined by continuity.

Further, the right part of (10) implies $\tau^{\gamma/2} \le 1/\pi_G(\tau)$, hence $C_1 \le C_0$.

► For our example $A = D_4$, choose $\gamma = 1/5$. Then $C_0 < 0.99$, and $C_1 < 0.95$. ◄

Finally, we conclude

$$\mathsf{L}_2(A^{\otimes n}) \preceq w(\mathcal{G}) = \sum_{m\ge 0} L_m \preceq \sigma^n(F) \sum_{m\ge 0} \left(C_0^{\frac{m}{\gamma}} + m^l C_1^{\frac{m}{\gamma}} \right) \asymp \sigma^n(F).$$

■

As a reserve for improving the method, one can suggest a more subtle combination of the two types of coverings involving reverse relocations of rectangles from $\mathcal{G}$ to $\mathcal{F}$.

3 Coverings of the Kneser–Sierpinski matrices

For a matrix D_r, $r = 2^t$, we propose a covering F_t consisting solely of rectangles of width 1. It generalizes the examples of Fig. 1 (for $t = 2$), and from [1] (for $t = 3$).

We exploit a simple gradient-fashion approach. First, put into F_t a column labeled by $\emptyset$, then add a row labeled by $\emptyset$ from the remaining part of the matrix. Next, we sequentially extract ones from all columns, and then from all rows labeled by the size-1 subsets of $[t]$. Then, we do the same with columns and rows labeled by the size-2 subsets, and go on until we reach $t/2$-size labels. At this point, all ones in D_r are covered.

Let $s(m, k)$ denote the binomial sum

$$s(m, k) = C_m^k + C_m^{k+1} + \ldots + C_m^m.$$

By construction, any rectangle corresponding to a column labeled by a size-k subset has height $s(t - k, k)$, and a rectangle corresponding to a row labeled by a size-k subset has length $s(t - k, k + 1)$. Hence,

$$\sigma(F_t) = \sum_{k=0}^{t/2} C_t^k \left(\sqrt{s(t-k,k)} + \sqrt{s(t-k,k+1)} \right).$$

Direct calculation shows that the quantity $\log_r \sigma(F_t)$ attains its minimum 1.2502... when $t = 18$, and, as easy to verify, it tends to the limit 1.259... as $t \to \infty$.

As a correcting covering G_t, we take a covering of all columns of D_r by individual rectangles, as shown on Fig. 1. Easy to see that $\sigma(G_t) = (\sqrt{2}+1)^t$, and $\mu_{G_t} = \left(\frac{2}{\sqrt{2}+1}\right)^t$.

With the use of coverings F_t, G_t, it is possible to satisfy the conditions of Theorem 1 only for $t \leq 15$. It can be directly verified that $\sigma(F_{15}) < 442412$, and $\lambda_{F_{15}} < -0.04$.

Corollary 1. $\mathsf{SUM}_2(D_N) \preceq N^{\log_2 15\, \sigma(F_{15})} \prec N^{1.251}$.

It's a kind of surprise, that coverings of matrices D_r by width-1 rectangles appear so efficient for the iterative procedure. However, they are not optimal in terms of spectral weight. At least starting from $t = 7$, one can construct better coverings via uniting common parts of columns or rows.

The author thanks Stasys Jukna for helpful comments.

References

[1] Alman J., Guan Y., Padaki A. Smaller low-depth circuits for Kronecker powers. 2022. arXiv:2211.05217v1.

[2] Chistikov D., Iván S., Lubiw A., Shallit J. Fractional coverings, greedy coverings, and rectifier networks. Proc. STACS (Hannover, 2017). LIPIcs. 2017. **66**, Art. 23.

[3] Jukna S., Sergeev I. Complexity of linear boolean operators. Foundations and Trends in Theoretical Computer Science. 2013. **9**(1), 1–123.

[4] Langer R. E. On the zeros of exponential sums and integrals. Bulletin AMS. 1931. **37**, 213–239.

[5] Mitrinović D. S., Pečarić J. E., Fink A. M. Classical and new inequalities in analysis. Dordrecht–Boston–London: Kluwer, 1993.