

Compressed Matrix Multiplication

Rasmus Pagh
IT University of Copenhagen

ITCS, January 10, 2012

Compressed

Matrix Multiplication

Rasmus Pagh
IT University of Copenhagen

ITCS, January 10, 2012

Outline

- Algorithm and analysis
- Related work
- Case study: Correlations
- Open problems

Informal problem statement

- **Input:** n -by- n matrices A and B , parameter b .
- **Output:** Approximation of AB that is good if AB is dominated by its b largest entries (“compressible”).

Basic algorithm

1. Take hash functions $s_1, s_2: [n] \rightarrow \{-1, 1\}$ and $h_1, h_2: [n] \rightarrow [b]$.

2. Compute the polynomial

$$\sum_i c_i x^i = \sum_{k=1}^n \left(\sum_{i=1}^n A_{ik} s_1(i) x^{h_1(i)} \right) \left(\sum_{j=1}^n B_{kj} s_2(j) x^{h_2(j)} \right)$$

3. Extract unbiased estimator

$$(AB)_{ij} \approx s_1(i) s_2(j) c_{h_1(i)+h_2(j)}$$

Basic algorithm

1. Take hash functions $s_1, s_2: [n] \rightarrow \{-1, 1\}$ and $h_1, h_2: [n] \rightarrow [b]$.

2. Compute the polynomial

$$\sum_i c_i x^i = \sum_{k=1}^n \left(\sum_{i=1}^n A_{ik} s_1(i) x^{h_1(i)} \right) \left(\sum_{j=1}^n B_{kj} s_2(j) x^{h_2(j)} \right)$$

3. Extract unbiased estimator

$$(AB)_{ij} \approx s_1(i) s_2(j) c_{h_1(i)+h_2(j)}$$

Observation: Each coefficient c_i is a sum of entries of AB with random signs

Why unbiased?

Lemma: If s_1 and s_2 are pairwise independent,

$$E[s_1(i_1)s_1(i_2)s_2(j_1)s_2(j_2)] = \begin{cases} 1 & \text{if } i_1 = i_2 \text{ and } j_1 = j_2 \\ 0 & \text{otherwise} \end{cases}$$

Using lemma, expected value of $s_1(i)s_2(j) \sum_i c_i x^i$ is:

$$\begin{aligned} & E \left[s_1(i)s_2(j) \sum_{k=1}^n \left(\sum_{i=1}^n A_{ik} s_1(i) x^{h_1(i)} \right) \left(\sum_{j=1}^n B_{kj} s_2(j) x^{h_2(j)} \right) \right] \\ &= \sum_{k=1}^n A_{ik} s_1^2(i) A_{ik} x^{h_1(i)} s_2^2(j) B_{kj} x^{h_2(j)} \\ &= (AB)_{ij} x^{h_1(i)+h_2(j)} \end{aligned}$$

What is the variance?

- Consider the “noise” in estimator caused by $(AB)_{i'j'}$:

$$X_{i'j'} = \begin{cases} s_1(i')s_2(j')(AB)_{i'j'} & \text{if } h_1(i) + h_2(j) = h_1(i') + h_2(j') \\ 0 & \text{otherwise} \end{cases}$$

- If h_1, h_2 are 3-wise independent, these random variables are uncorrelated, so:

$$\begin{aligned} \mathbf{Var} \left(\sum_{i',j'} X_{i'j'} \right) &= \sum_{i',j'} \mathbf{Var} (X_{i'j'}) = \sum_{i',j'} E[X_{i'j'}^2] \\ &\leq \sum_{i',j'} (AB)_{i'j'}^2 / b = \|AB\|_F^2 / b \end{aligned}$$

Sparse outputs

- Suppose AB has at most $b/3$ nonzero entries.
- Then with probability $2/3$ there is no noise in a given estimator.
- Repeat $O(\log n)$ times and take median estimate, to get exact result whp.

Time analysis

- Construct $2n$ degree b polynomials: $O(n^2+nb)$.
- Multiply n pairs of degree b polynomials, using FFT: $O(nb \log b)$.
- Extracting estimates: $O(n^2)$.

Total time: $O(n^2+nb \log b)$.

Background

- The polynomial computed is in fact a Count-Sketch [Charikar et al. '04], an early *compressed sensing* method.

Background

- The polynomial computed is in fact a Count-Sketch [Charikar et al. '04], an early *compressed sensing* method.
- Polynomial multiplication combines Count-Sketches of column vector of A and a row vector of B into a Count-Sketch for their *outer product*.

Background

- The polynomial computed is in fact a Count-Sketch [Charikar et al. '04], an early *compressed sensing* method.
- Polynomial multiplication combines Count-Sketches of column vector of A and a row vector of B into a Count-Sketch for their *outer product*.
- Add up outer product sketches to get a sketch for AB .

Some related results

- **Folklore:** Computing AB with b nonzeros in time $O(nb)$ if there are *no cancellations*.

Some related results

- **Folklore:** Computing AB with b nonzeros in time $O(nb)$ if there are *no cancellations*.
- **Cohen and Lewis '99:** For *nonnegative* matrices, estimate AB with low *relative* error.

Some related results

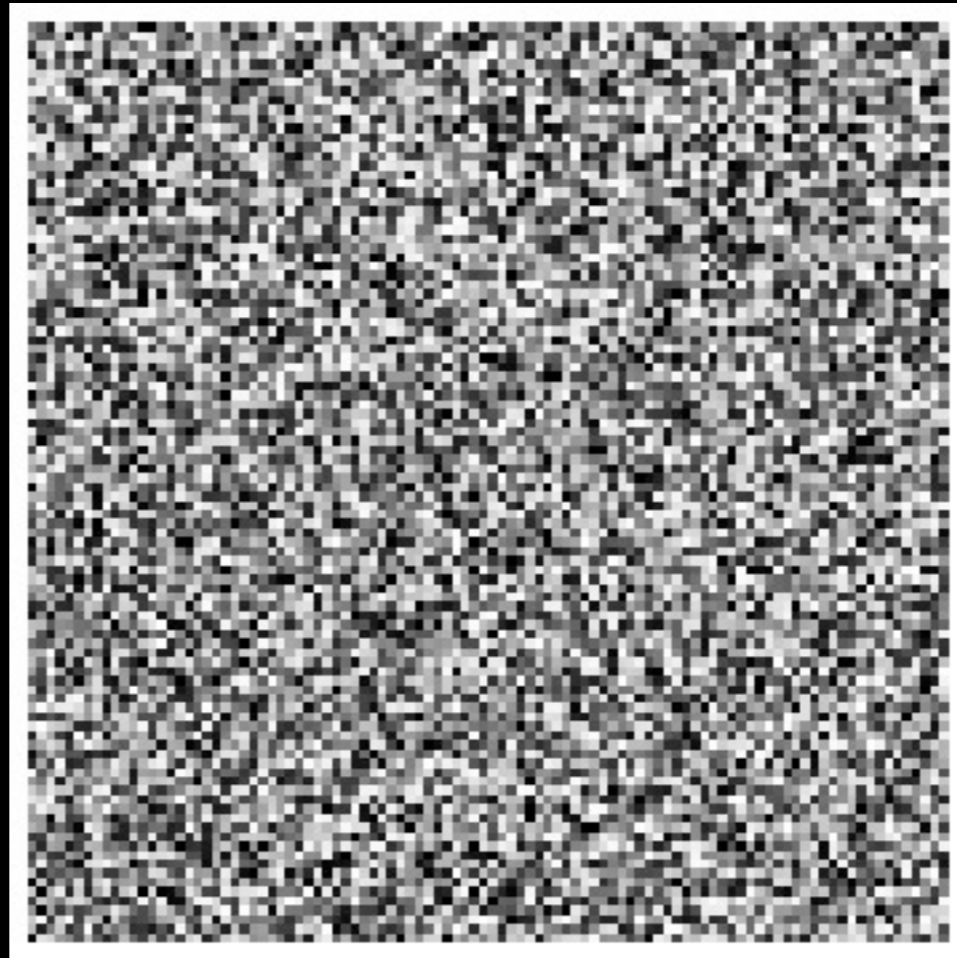
- **Folklore:** Computing AB with b nonzeros in time $O(nb)$ if there are *no cancellations*.
- **Cohen and Lewis '99:** For *nonnegative* matrices, estimate AB with low *relative* error.
- **Iwen and Spencer '09:** Computing AB with $\leq b/n$ nonzeros in *each column* in time $\tilde{O}(nb)$.

Some related results

- **Folklore:** Computing AB with b nonzeros in time $O(nb)$ if there are *no cancellations*.
- **Cohen and Lewis '99:** For *nonnegative* matrices, estimate AB with low *relative* error.
- **Iwen and Spencer '09:** Computing AB with $\leq b/n$ nonzeros in *each column* in time $\tilde{O}(nb)$.
- **Drineas, Kannan, Mahoney '06; Sarlós '06:** Computing AB with low *total* error in terms of $\|A\|_F$ and $\|B\|_F$.

Case study: Correlations

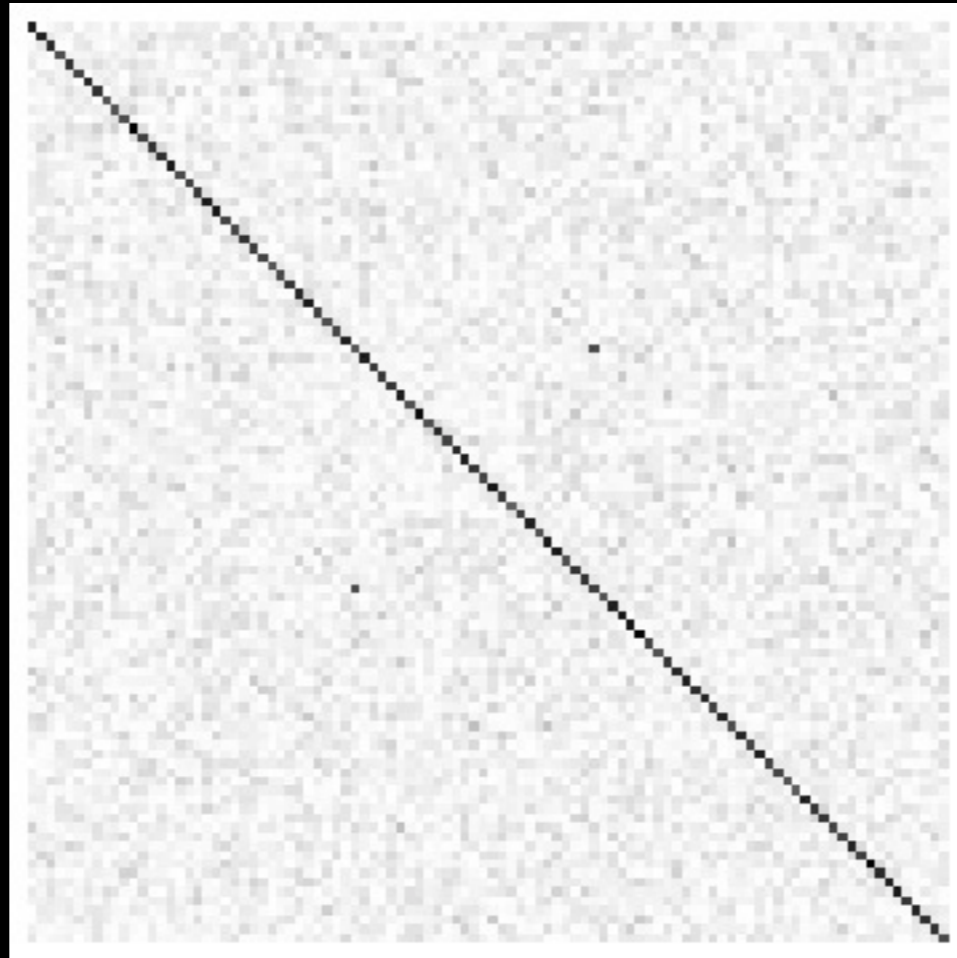
$A =$



Two rows of A are correlated. Which ones?

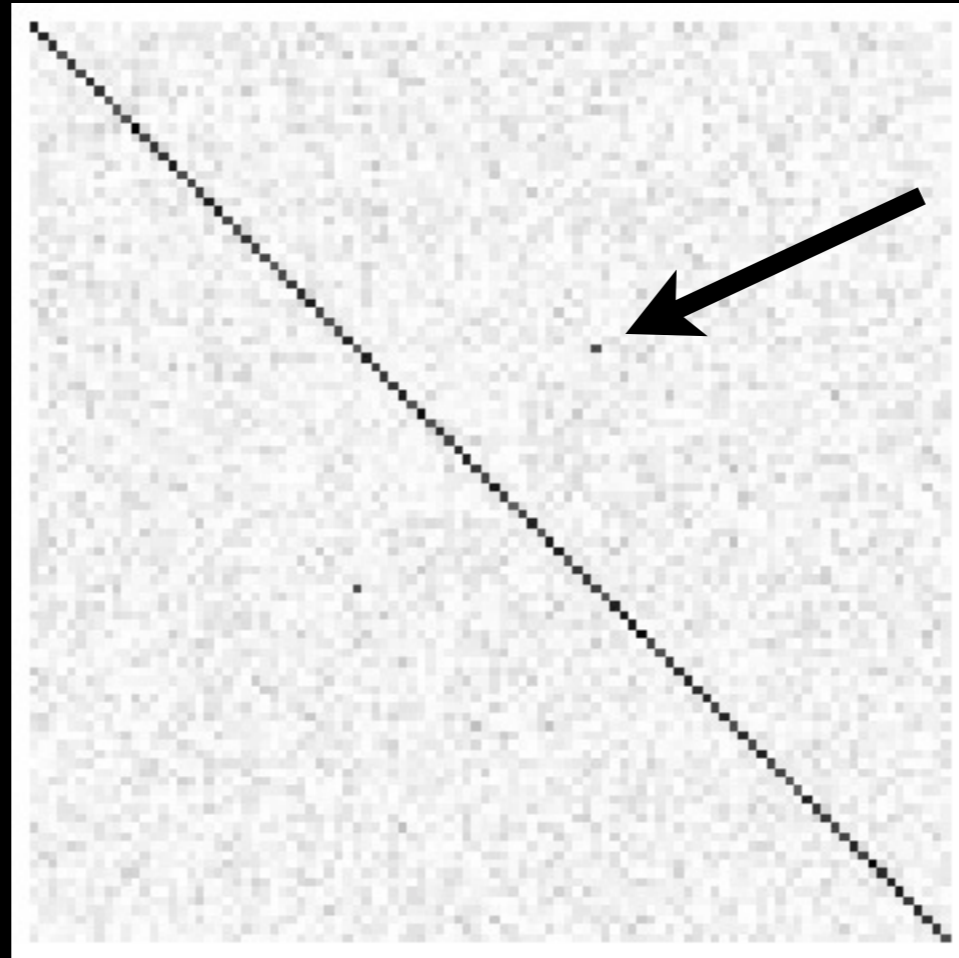
Sample covariance matrix

$$AA^T =$$



Sample covariance matrix

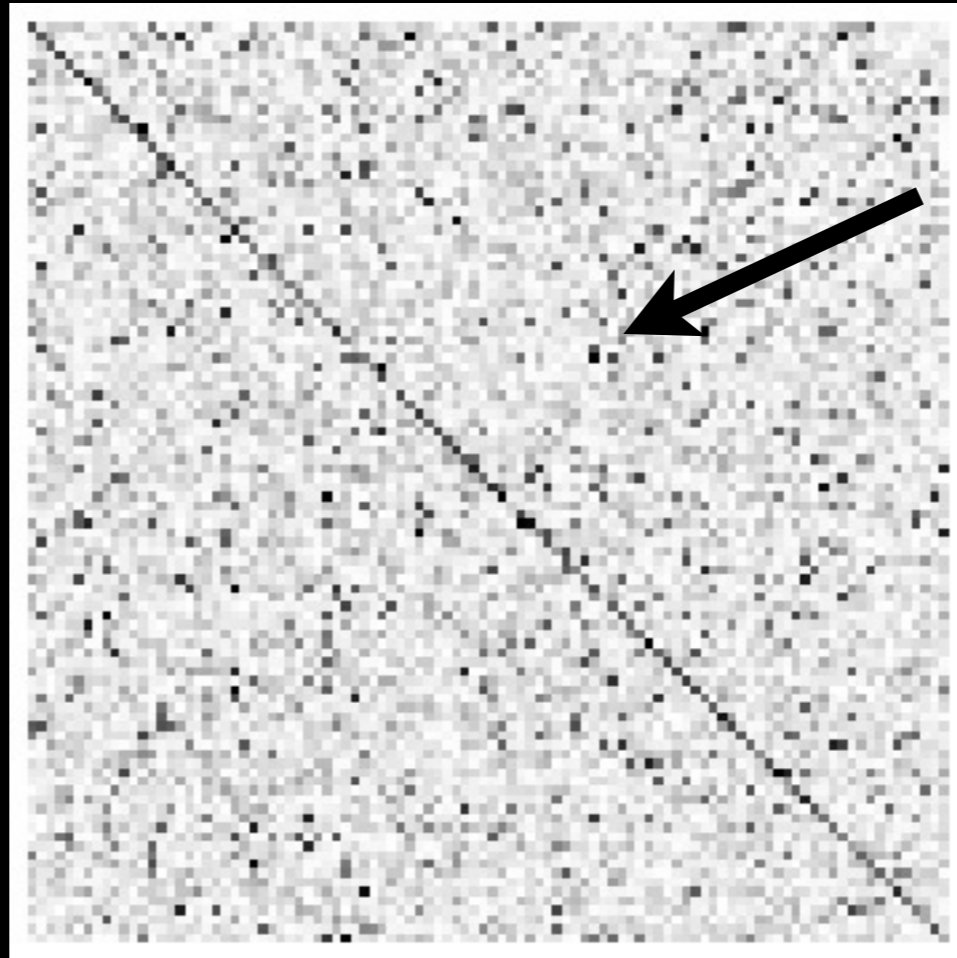
$$AA^T =$$



Sample covariance matrix

estimated using compressed matrix multiplication

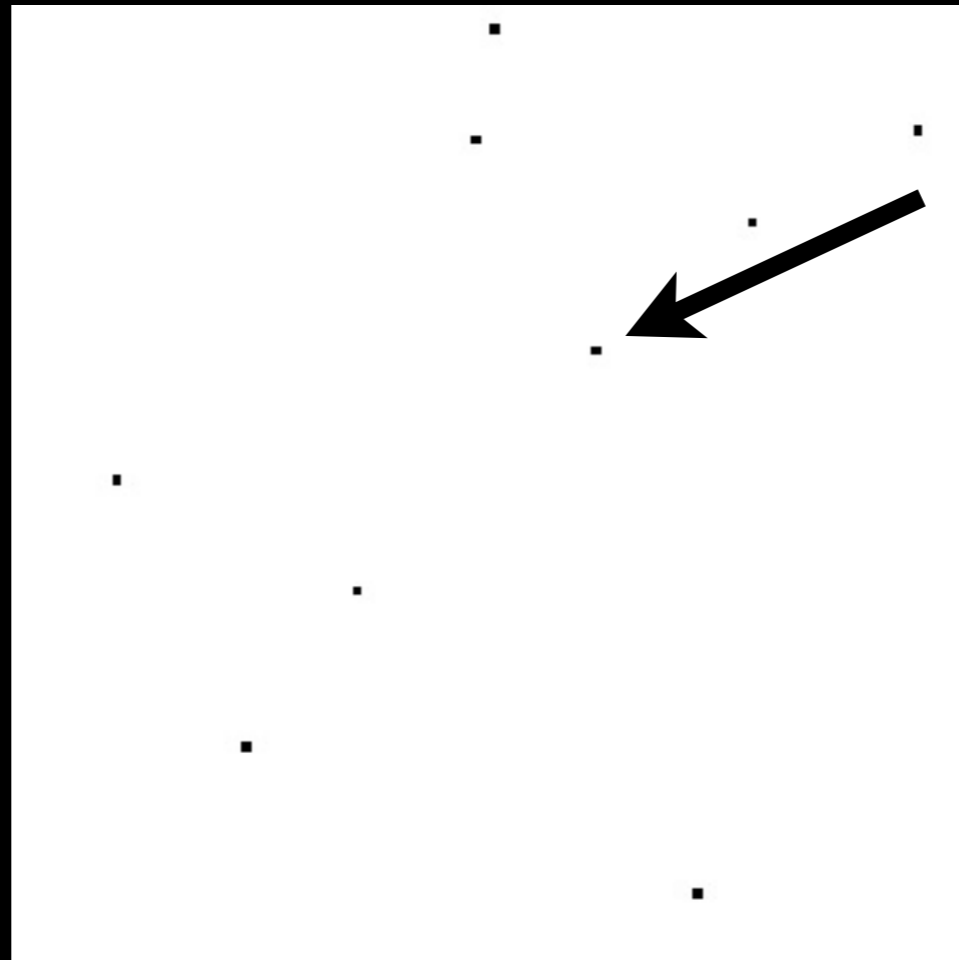
$$AA^T \approx$$



Sample covariance matrix

estimated using compressed matrix multiplication

$$f(AA^T) =$$



Showing large values not explained by hash collisions.

Some open problems

- Can other problems with “sparse solutions” be solved efficiently using compressed sensing techniques?
 - Matrix inversion?
 - Linear systems with a sparse solution?
 - Sparse transitive closure of a graph?
 - Product of > 2 matrices?



Discussion:

Combinatorial algorithms

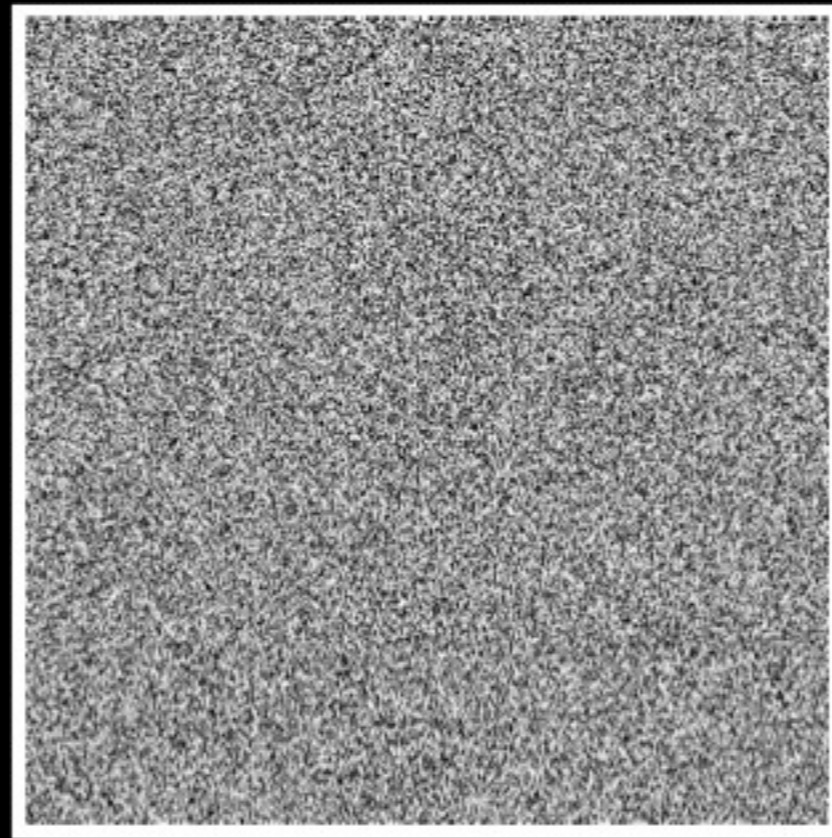
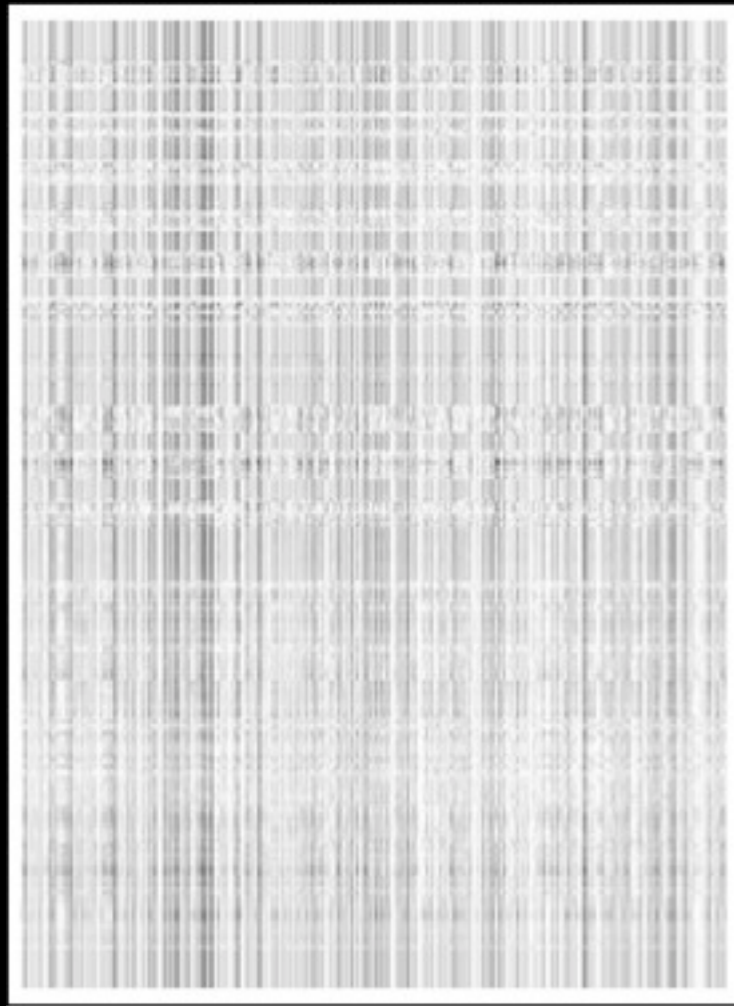
- Compressed MM can be considered “combinatorial”.
- Another view: No large hidden constants
(in contrast to “algebraic” approaches leading to $\omega < 2.3727$).

Discussion:

Combinatorial algorithms

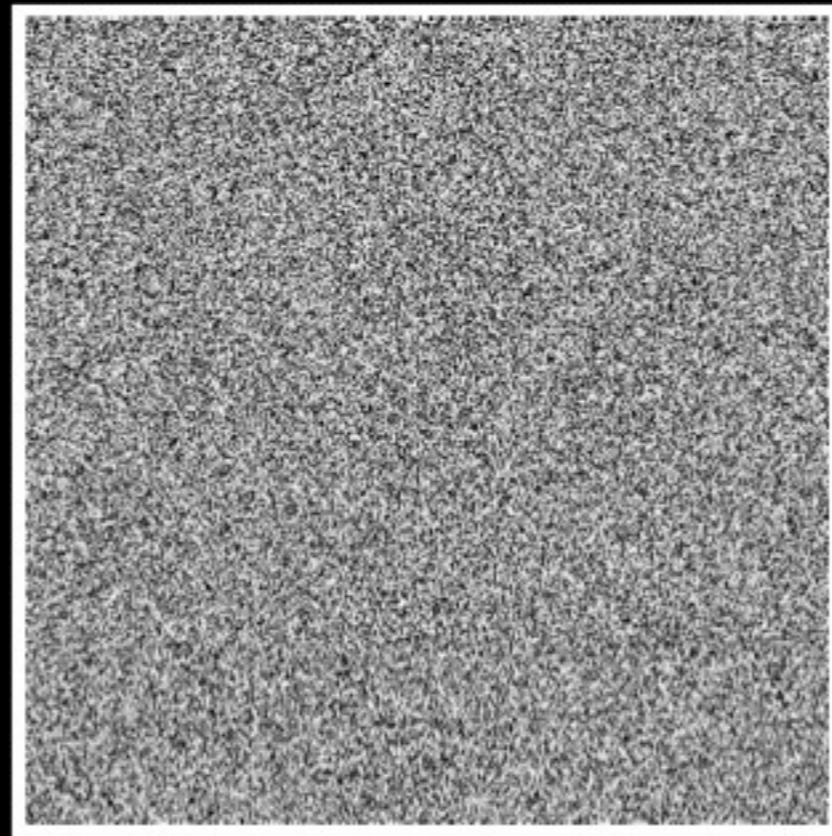
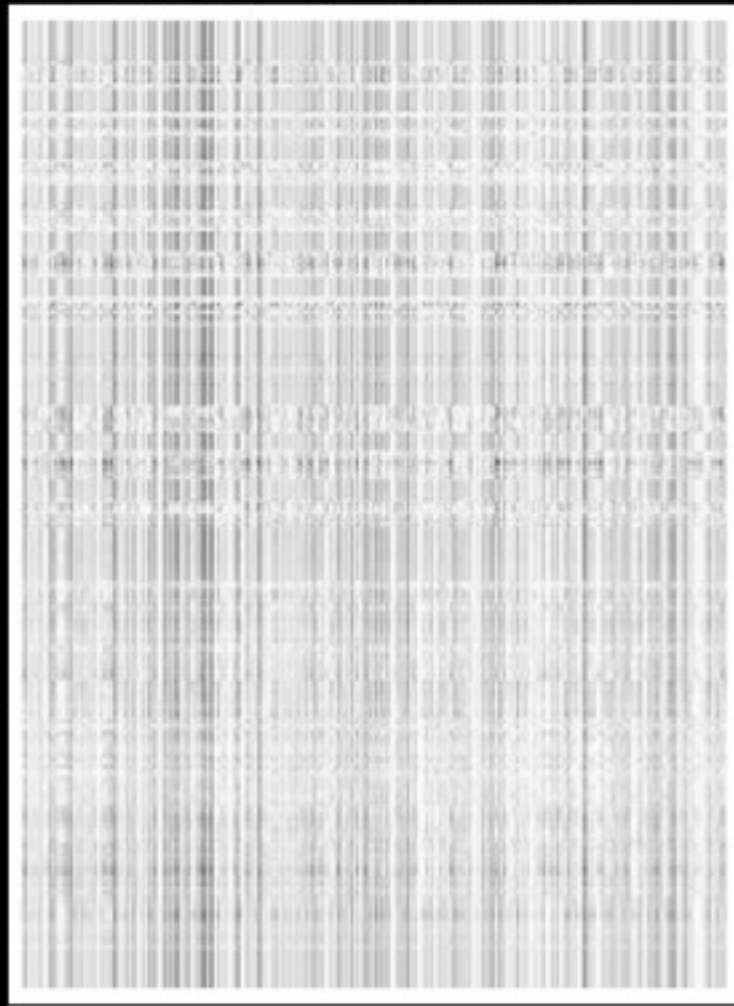
- Compressed MM can be considered “combinatorial”.
- Another view: No large hidden constants
(in contrast to “algebraic” approaches leading to $\omega < 2.3727$).
- It is interesting to consider what other subclasses of matrix products can be computed in time, say, $n^{2+\epsilon}$, using algorithms with these properties.

Hidden slide: Extra application

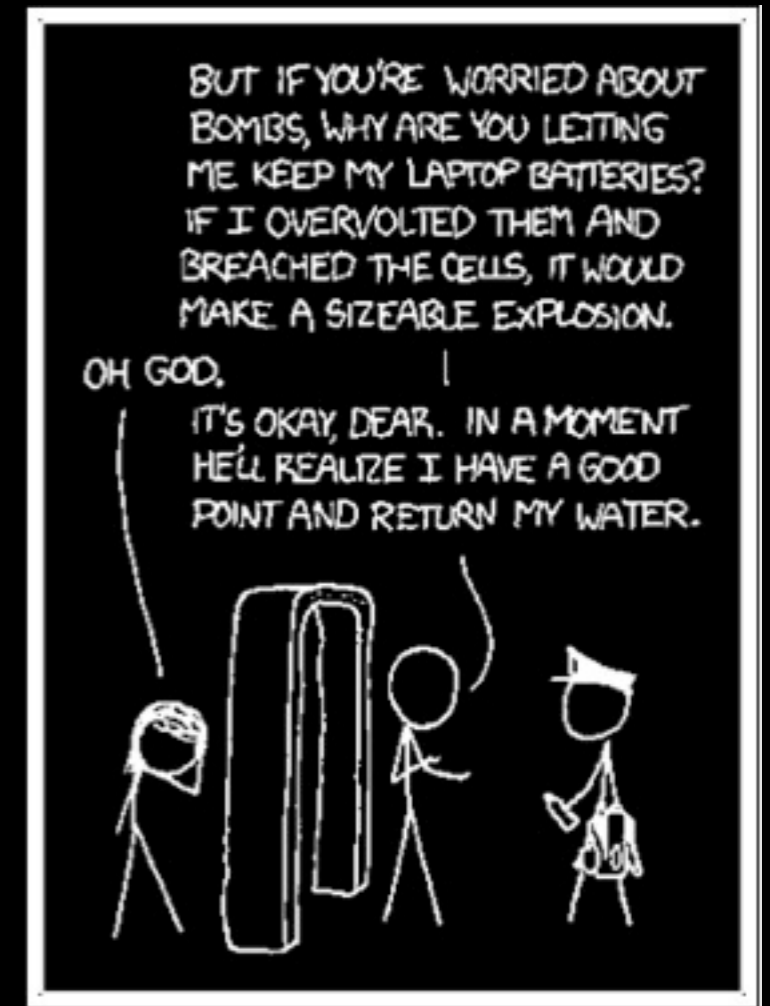


=

Hidden slide: Extra application



=



<http://xkcd.com/651/>