Quadrarithmic-Time Geometric Exact Boolean Fast Matrix Multiplication

Zhijia (Brian) Liu / 刘之佳
Email: bzliu94@gmail.com

September 4th, 2018

Abstract We use a geometric approach to solve Boolean matrix multiplication (hereafter MM) in $\mathcal{O}(n^2 \cdot \log^3(n))$ time. We use bounds based on how far n vectors on an n-dimensional hypercube can be from each other at worst and use always-successful two-distortion-ratio approximate nearest neighbor (ANN) with approximate bichromatic closest pairs and Prim's method via Chan [C98] to determine an approximate MST and topological ordering to memoize for an overcomplete dictionary's cut-out/expiry rectangles. Then, for query time, we use for each main query at most three subqueries based on dictionary. The main idea that makes the equidistance bound useful is that it decays as we add vectors. Then, we relate Euclidean and Manhattan distances. We assume perfect packing and n-dimensional hypercube surface area divided by $(n-1)$-dimensional hypercube volumes. Then, we use telescoping/prefixes (to deal with three-sided queries as opposed to having directly four-sided queries) and color carpentry on top of geometric configuration as from Kaplan [KRS08]. We draw ideas from Gupta [GJS95], as well. We use union-intersection arithmetic as from Kaplan. We have staircase pairs and introduce z-levels for rectangle arrangement. A trivial case we support is rectangular input matrices. Future ideas include number-of-non-zeroes-based sparsity exploitation (which is more ambitious than just having empty A rows or B columns, which we support), parallelism, cache exploitation. Our approach is cache-oblivious, deterministic. We compare to optimized brute-force/nearly-brute-force commonly-used approach library BLAS and observe cross-over for Boolean case at $n =$??. We note that Raz says, under certain assumptions, minimum time for MM is $\mathcal{O}(n^2 \cdot \log(n))$ [Raz02]. We acknowledge that the reason we are able to make progress on such a MM problem is that it is a special case and we are taking advantage of the concept of strength reduction.

Keywords linear algebra, matrix multiplication, complexity, computational geometry

1 Introduction

We were inspired by CS 61C machine architecture course matrix multiplication project at UC Berkeley from summer of 2011.

Also, for the future, we will have floating-point case s.t. we have exactness or numerical robustness even in the face of possible catastrophic cancellation (which we think is possible to handle via tiling and a persistent prefix tree and a secondary representation based on run-length encoding (RLE) as we can then efficiently handle binary overflow/underflow for floating-point addition or subtraction). Further, for floating-point case, we will have use of word-RAM and packing. We believe that it is possible that our Boolean MM approach can be used almost as a black box (i.e. we introduce word-packing for it and that means word-packed ANN) for floating-point case along with NTT, Zhang word-packed linear convolution [Z17], circular convolution in terms of linear convolution, Rader FFT as circular convolution [R68], bit spread/gather as via Nuetzi p-way q-bit bit interleaving [N13A, N13B], Parseval's theorem, assumption that $\log(m) = \mathcal{O}(\log(n))$. This means time for exact floating-point case of tentatively $\mathcal{O}(n^2 \cdot \log^3(n))$.

We also support semiring flavors. We are unsure as to whether our Boolean MM approach can be generalized for all semiring flavors. The time we have for Boolean case is assuming we don't use fractional cascading and lowest-level interval tree. For semiring flavors (min, $+$), (max, $+$), (min, max), (max, min), (min, $\leq$), (max, $\leq$), (min, $\times$), (max, $\times$), we use tower arrangement and critical min. distance range closest-pair query via Sharathkumar [SG07] and time is $\mathcal{O}(n^2 \cdot \log^2(n))$.

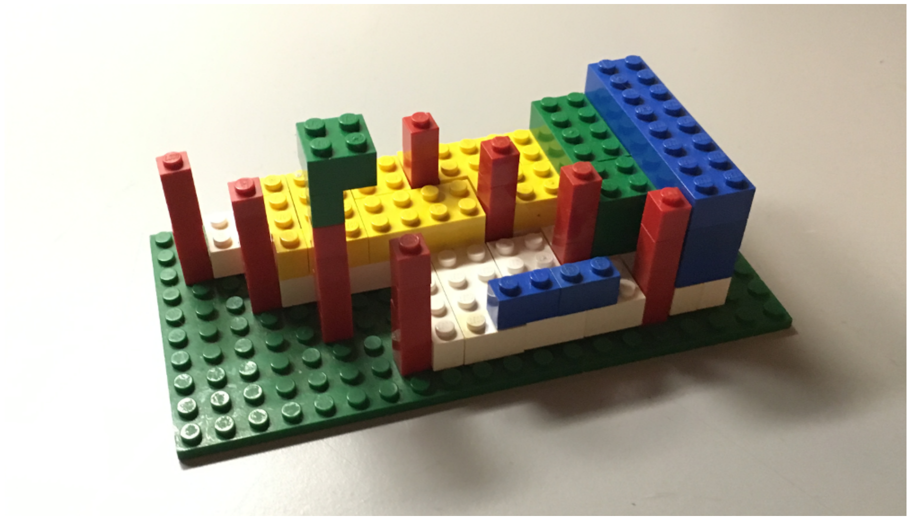

Figure 1: Pair of descending staircases with many colors for rectangle arrangement

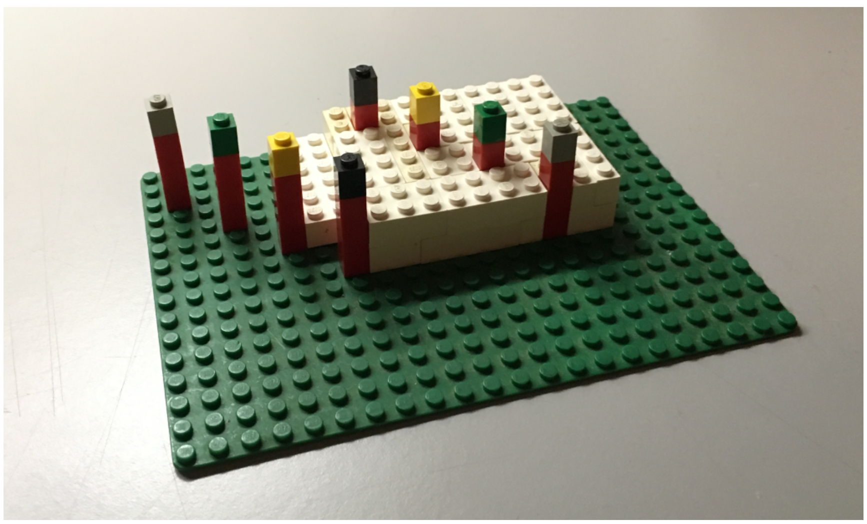

Figure 2: For fixed color, we have varying z-levels that facilitate using telescoping

2 References

Chan, Timothy M. 1998. "Approximate Nearest Neighbor Queries Revisited." *Discrete & Computational Geometry* 20 (3). Springer: 359–73.

Gupta, Prosenjit, Ravi Janardan, and Michiel Smid. 1995. "Further Results on Generalized Intersection Searching Problems: Counting, Reporting, and Dynamization." *Journal of Algorithms* 19 (2). Elsevier: 282–317.

Kaplan, Haim, Natan Rubin, Micha Sharir, and Elad Verbin. 2008. "Efficient Colored Orthogonal Range Counting." *SIAM Journal on Computing* 38 (3). SIAM: 982–1011.

Nuetzi, Gabriel. 2013a. "How to Compute a 3-d Morton Number (Interleave the Bits of 3 Ints)." https://stackoverflow.com/questions/1024754.

———. 2013b. "Produce Interleaving Bit Patterns (Morton Keys) for 32-Bit, 64-Bit, and 128-Bit." https://stackoverflow.com/questions/18529057.

Rader, Charles M. 1968. "Discrete Fourier Transforms When the Number of Data Samples Is Prime." *Proceedings of the IEEE* 56 (6). IEEE: 1107–8.

Raz, Ran. 2002. "On the Complexity of Matrix Product." In *Proceedings of the Thiry-Fourth Annual ACM Symposium on Theory of Computing*, 144–51. ACM.

Sharathkumar, R, and Prosenjit Gupta. 2007. "Range-Aggregate Proximity Queries." *IIIT Hyderabad, Telangana* 500032.

Zhang, Meng. 2017. "Fast Convolutions of Packed Strings and Pattern Matching with Wildcards." *International Journal of Foundations of Computer Science* 28 (03). World Scientific: 289–307.