SPARSE GRAPH BASED SKETCHING FOR FAST NUMERICAL LINEAR ALGEBRA

Dong Hu[*]   *Shashanka Ubaru*[†]   *Alex Gittens*[*]   *Kenneth L. Clarkson*[#]   *Lior Horesh*[†]   *Vassilis Kalantzis*[†]

[*]Rensselaer Polytechnic Institute, Troy, NY
[†] IBM Research, Yorktown Heights, NY. [#]IBM Research, San Jose, CA

ABSTRACT

In recent years, a variety of randomized constructions of *sketching matrices* have been devised, that have been used in fast algorithms for numerical linear algebra problems, such as least squares regression, low-rank approximation, and the approximation of leverage scores. A key property of sketching matrices is that of *subspace embedding*. In this paper, we study sketching matrices that are obtained from bipartite graphs that are sparse, i.e., have left degree s that is small. In particular, we explore two popular classes of sparse graphs, namely, *expander graphs* and *magical graphs*. For a given subspace $\mathcal{U} \subseteq \mathbb{R}^n$ of dimension k, we show that the magical graph with left degree $s = 2$ yields a $(1 \pm \epsilon)$ ℓ_2-subspace embedding for $\mathcal{U}$, if the number of right vertices (the sketch size) $m = \mathcal{O}(k^2/\epsilon^2)$. The expander graph with $s = \mathcal{O}(\log k/\epsilon)$ yields a subspace embedding for $m = \mathcal{O}(k \log k/\epsilon^2)$. We also discuss the construction of sparse sketching matrices with reduced randomness using expanders based on error-correcting codes. Empirical results on various synthetic and real datasets show that these sparse graph sketching matrices work very well in practice.

Index Terms— Randomized sketching, expander graphs, low-rank approximation, least squares regression.

1. INTRODUCTION

Many modern applications involving large-dimensional data resort to data reduction techniques such as matrix approximation [14, 33] or compression [29] in order to lower the amount of data processed, and to speed up computations. In recent years, randomized sketching techniques have attracted much attention in the literature for computing such matrix approximations, due to their high efficiency and well-established theoretical guarantees [19, 14, 34]. Sketching methods have been used to speed up numerical linear algebra problems such as least squares regression, low-rank approximation, matrix multiplication, and approximating leverage scores [28, 11, 14, 34, 32]. These primitives as well as improved matrix algorithms have been developed based on sketching for various tasks in machine learning [5, 37], signal processing [9, 26], scientific computing [33], and optimization [25, 23].

Given a large matrix $\boldsymbol{A} \in \mathbb{R}^{n \times d}$, sketching techniques consider a random "sketching" matrix $\boldsymbol{S} \in \mathbb{R}^{m \times n}$ with $m \ll n$, where the matrix product $\boldsymbol{S} \cdot \boldsymbol{A}$ is called the sketch of $\boldsymbol{A}$. Then, the sketch (smaller matrix) substitutes the original matrix in applications. For example, in over-constrained least-squares regression, where design matrix $\boldsymbol{A} \in \mathbb{R}^{n \times d}$ has $n \gg d$, and with response vector $\boldsymbol{b} \in \mathbb{R}^n$, we wish to solve $\min_{\boldsymbol{x} \in \mathbb{R}^d} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2$ inexpensively, and obtain an approximate solution $\tilde{\boldsymbol{x}}$ such that $\|\boldsymbol{A}\tilde{\boldsymbol{x}} - \boldsymbol{b}\|_2 \leq (1+\epsilon)\|\boldsymbol{A}\boldsymbol{x}^* - \boldsymbol{b}\|_2$, where $\boldsymbol{x}^* = (\boldsymbol{A}^\top \boldsymbol{A})^{-1}\boldsymbol{A}^\top \boldsymbol{b}$ is the exact solution and ϵ is an error tolerance. Using sketching, we can obtain a fast approximate solution, by solving $\tilde{\boldsymbol{x}} = \arg\min_{\boldsymbol{x} \in \mathbb{R}^d} \|\boldsymbol{S}\boldsymbol{A}\boldsymbol{x} - \boldsymbol{S}\boldsymbol{b}\|_2$. Similarly, in low-rank approximation, given a large matrix $\boldsymbol{A} \in \mathbb{R}^{n \times d}$ and a rank k, we wish to compute a rank-k matrix $\tilde{\boldsymbol{A}}_k$ for which $\|\boldsymbol{A} - \tilde{\boldsymbol{A}}_k\| \leq (1+\epsilon)\|\boldsymbol{A} - \boldsymbol{A}_k\|$, where $\boldsymbol{A}_k$ is the best rank-k approximation with respect to the matrix norm. Again, the matrix $\tilde{\boldsymbol{A}}_k$ can be computed inexpensively using the sketch $\boldsymbol{S}\boldsymbol{A}$, see section 4. In both cases, the $(1 + \epsilon)$-approximation error bound is obtained if the sketching matrix $\boldsymbol{S}$ is a $(1 \pm \epsilon)$ ℓ_2-*subspace embedding* [28, 34] (defined in section 3).

Different variants of sketching matrices have been proposed in the literature [2, 24, 34]. Random matrices with i.i.d. Gaussian entries [14] and random sign matrices [10] were proposed, which for a rank-k approximation require a sketch size of only $m = \mathcal{O}\left(\frac{k}{\epsilon}\right)$. However, these matrices are dense, require many i.i.d random numbers to be generated, and the cost of computing the sketch $\boldsymbol{S}\boldsymbol{A}$ will be $\mathcal{O}(ndm)$. Structured random matrices, like subsampled random Fourier transform (SRFT), Hadamard transform (SRHT) and error-correcting code (ECC) matrices have been used to overcome the randomness and sketch cost issues [35, 3, 32]. For dense $\boldsymbol{A}$, the sketch cost will be $\mathcal{O}(nd \log(m))$. However, the sketch size will need to be $m = \mathcal{O}\left(\frac{k \log k}{\epsilon}\right)$.

Sparse sketching matrices have also been explored [11, 22] for sparse input matrices. *Countsketch* matrices [11], which has just one nonzero per column of $\boldsymbol{S}$, have been shown to perform well for sketching. For sparse $\boldsymbol{A}$, the sketching cost will be $\mathcal{O}(\text{nnz}(\boldsymbol{A}))$, that is, linear in the number of nonzeros of $\boldsymbol{A}$. But, for a good rank-k approximation, the sketch size will have to be $m = \mathcal{O}\left(\frac{k^2}{\epsilon^2}\right)$ [34]. Applications involving low-rank approximation of large sparse matrices include Latent Semantic Indexing (LSI) [20], matrix completion problems [15], subspace tracking in signal processing [36] and
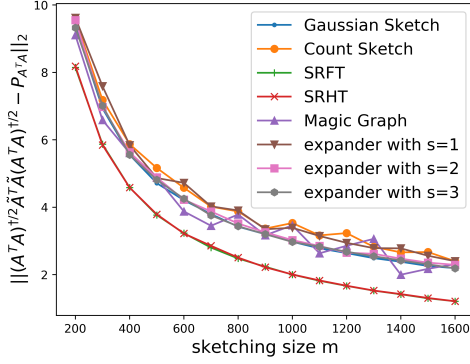
Fig. 1. Various sketching methods applied on a Gaussian matrix $A \in \mathbf{R}^{4096 \times 1000}$ with entries $A_{i,j} \sim \mathcal{N}(0,1)$. The distortion (see Sec. 4) is plotted as a function of the sketch size m. Distortion decreases as m is increased, and all the sketching methods perform similarly on this synthetic data.

others [33].

In this study, we explore sparse sketching matrices that are obtained from bipartite graphs. Our study is motivated in part by the numerical results we obtained while developing a Matrix Sketching toolkit. In particular, we observed that a sparse matrix obtained from so-called *magical graphs*, used in [8] for fast rank estimation, performed very well as a sketching matrix in practice. Figure 1 presents one such result, where we plot the distortion [21] (see, section 4 for definition) between the input matrix A and its sketch SA when we applied different sketching matrices. We observe that the matrices from sparse graphs such as magical graph and expander graphs perform as well as known sketching matrices. In this paper, we show that such matrices based on sparse graphs are indeed $(1 \pm \epsilon)$ ℓ_2-subspace embeddings. Specifically, we present theoretical results which show that the magical graph matrix ($s = 2$) of [8] will be a subspace embedding for $m = \mathcal{O}\left(\frac{k^2}{\epsilon^2}\right)$. Matrices obtained from expander graphs with degree $s = \mathcal{O}\left(\frac{\log k}{\epsilon}\right)$ will be subspace embeddings for a sketch size $m = \mathcal{O}\left(\frac{k \log k}{\epsilon^2}\right)$. We also discuss how we can construct sparse sketching matrices with reduced randomness using expander graphs based on error correcting codes [13]. These can be viewed as alternate constructions for the sparse sketch matrices with limited randomness discussed in [22]. Finally, we present numerical results that illustrate the performance of the various sketching matrices for low-rank approximation.

2. SPARSE GRAPHS

First, we describe a class of sparse graphs known as *expander graphs*, and variants called magical graphs.

Definition 1 (Expander Graph). *A bipartite graph* $\mathcal{G} = (L, R; E)$ *with* $|L| = n, |R| = m$*, and regular left degree* s*, is*
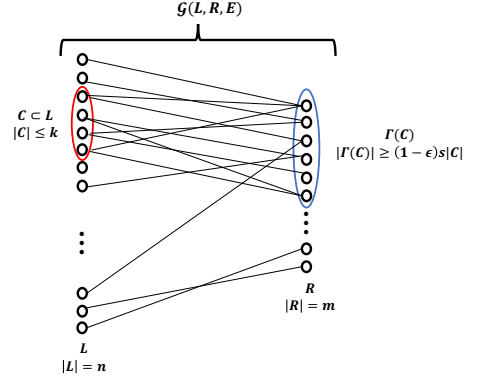
Fig. 2. $(k, 1 - \epsilon)$ -Expander graph

a $(k, 1 - \epsilon)$*-expander graph if for any* $C \subseteq L$ *with* $|C| \leq k$*, the set of neighbors* $\Gamma(C)$ *of* C *has size* $|\Gamma(C)| > (1 - \epsilon)s|C|$*.*

Figure 2 gives an illustration of an expander graph. These expander graphs have been used in a number of applications including coding theory [30], compressed sensing [17], group testing [7], multi-label classification [31], and others [16]. A number of random and deterministic constructions of expander graphs exist. A uniformly random bipartite graph is an expander graph with high probability when $s = O\left(\frac{\log n/k}{\epsilon}\right)$ and $m = O\left(\frac{k \log(n/k)}{\epsilon^2}\right)$ [16, 7, 31]. That is, for each left vertex, we choose s right vertices uniformly and independently at random.

Explicit constructions of expander graphs are also known [4, 27], and there are many constructions of expander graphs based on error-correcting codes [13, 1, 7]. Such constructions can yield sparse sketching matrices with reduced randomness. A few constructions of subspace-embedding sketching matrices based on error-correcting codes are known [32].

Magical graphs are variants of expander graphs that have been used to construct super-concentrators [16]. Cheung et. al [8] showed that such graphs (defined below) can be used to develop fast algorithms for matrix rank estimation.

Definition 2 (Magical Graph). *A random bipartite graph* $\mathcal{G} = (L, R; E)$ *with* $|L| = n, |R| = m$ *is* (k, δ)*-magical if for every* $C \subseteq L$ *with* $|C| = k$*, the probability that there is a matching in* $\mathcal{G}$ *in which every vertex in* C *is matched is at least* $1 - \delta$*.*

Article [8] also presented a random bipartite graph construction with $s = 2$ that is $(k, O(1/k))$-magical, when $m \geq 11k$. We can show that this construction is also a $(k, 1 - \epsilon)$-expander, if $m = \mathcal{O}(k^{1/\epsilon})$.

Here we consider the incidence/adjacency matrix S corresponding to such a bipartite graph as our sketching matrix, with the nonzero entries (edges) randomly set to $\pm 1/\sqrt{s}$. Note that, with the sketch matrix based on an expander graph, the sketch $S \cdot A$ can be computed in $\mathcal{O}(\text{nnz}(A) \log k/\epsilon)$ time, and with the magical graph ($s = 2$), $S \cdot A$ can be computed in $\mathcal{O}(\text{nnz}(A))$ time.

3. THEORY

In this section, we present the theoretical results that show that the sketching matrices obtained from sparse graphs such as expanders are subspace embeddings under certain conditions. First, we give the definition of subspace embedding. We use the notation $a = (1 \pm \epsilon)b$ to mean $a \in [(1-\epsilon)b, (1+\epsilon)b]$.

Definition 3 (Subspace embedding). *A sketching matrix $\boldsymbol{S} \in \mathbb{R}^{m \times n}$ is a $(1 \pm \epsilon)$ ℓ_2-subspace embedding for a subspace $\mathcal{U} \subseteq \mathbb{R}^n$ of dimension k, if for an orthonormal basis $\boldsymbol{U} \in \mathbb{R}^{n \times k}$ of $\mathcal{U}$ and for all $\boldsymbol{x} \in \mathbb{R}^k$, we have*

$$\|\boldsymbol{S}\boldsymbol{U}\boldsymbol{x}\|_2^2 = (1 \pm \epsilon)\|\boldsymbol{U}\boldsymbol{x}\|_2^2 = (1 \pm \epsilon)\|\boldsymbol{x}\|_2^2.$$

This means that $\boldsymbol{S}$ preserves the norm of any vectors from the k-dimensional subspace. It also implies the singular values of $\boldsymbol{S}\boldsymbol{U}$ are bounded as $|1 - \sigma_i(\boldsymbol{S}\boldsymbol{U})| \leq \epsilon$ $i \in [1, \ldots, k]$.

Theorem 1 (Magical graph). *A sketching matrix $\boldsymbol{S} \in \mathbb{R}^{m \times n}$ obtained from a random bipartite graph with $s = 2$ is $(1 \pm \epsilon)$ ℓ_2-subspace embedding with probability $1 - \delta$ if $m = O\left(\frac{k^2}{\delta \epsilon^2}\right)$.*

Proof. We briefly sketch the proof here. First, we consider the JL moment property defined in [18]. A sketch matrix $\boldsymbol{S} \in \mathbb{R}^{m \times n}$ has the (ϵ, δ, ρ)-JL moment property, if for all $\boldsymbol{x} \in \mathbb{R}^n$, with $\|\boldsymbol{x}\|_2 = 1$, we have

$$\mathbb{E}|\|\boldsymbol{S}\boldsymbol{x}\|_2^2 - 1|^\rho \leq \epsilon^\rho \delta.$$

We next show that our graph sketch matrix $\boldsymbol{S}$ with $s = 2$ satisfies the $(\epsilon_0, \delta, 2)$-JL moment property for $m = O\left(\frac{1}{\delta \epsilon_0^2}\right)$. For this, we adapt the proof of [18, Theorem 15], where the (ϵ, δ, ρ)-JL moment property is proved for a random bipartite graph (where the edges for each left vertex and the signs ± 1 are assigned randomly via. a hash function) with m right vertices and left degree s. From that proof, by setting $Y = \|\boldsymbol{S}\boldsymbol{x}\|_2^2 - 1$, we have

$$\mathbb{E}|Y^\rho| \leq \frac{1}{s^\rho} \cdot 2^{O(\rho)} \cdot \sum_{q=1}^{\rho/2} \rho^\rho \cdot \left(\frac{s^2}{qm}\right)^q.$$

For the magical graph of [8], we have $s = 2$, and if we set $\rho = 2$ and $m = \frac{16}{\delta \epsilon_0^2}$, we obtain the $(\epsilon_0, \delta, 2)$-JL moment property. We then use the proof of [34, Theorem 9] to establish the subspace embedding property. In particular, [34, Thm 13] shows any sketch matrix $\boldsymbol{S}$ with (ϵ, δ, ρ)-JL moment property also satisfies, for many matrices $\boldsymbol{A}, \boldsymbol{B}$,

$$Pr[\|\boldsymbol{A}^\top \boldsymbol{B} - \boldsymbol{A}^\top \boldsymbol{S}^\top \boldsymbol{S}\boldsymbol{B}\|_F > \epsilon\|\boldsymbol{A}\|_F\|\boldsymbol{B}\|_F] \leq \delta.$$

By setting, the two matrices to be $\boldsymbol{U}$ and $\epsilon = \epsilon_0/k$, we get the subspace embedding property. □

These results and the sketch size order match that of Countsketch [11] that has $s = 1$. However, in practice, we note that sparse graph sketch matrices with $s = 2$ perform slightly better than Countsketch, see section 4. Our next result shows that the expander graphs can achieve improved sketch sizes.

Theorem 2 (Expander). *A sketching matrix $\boldsymbol{S} \in \mathbb{R}^{m \times n}$ obtained from an expander graph with $s = O\left(\frac{\log(k/\delta\epsilon)}{\epsilon}\right)$ is $(1 \pm \epsilon)$ ℓ_2-subspace embedding with probability $1 - \delta$ if $m = O\left(\frac{k \log(k/\delta\epsilon)}{\epsilon^2}\right)$.*

Proof. We first use the result from [18] to show that the graph sketch matrix $\boldsymbol{S}$ is JLT. A sketch matrix is JLT(ϵ, δ), if for any $\boldsymbol{x} \in \mathbb{R}^n$, with $\|\boldsymbol{x}\|_2 = 1$, we have

$$Pr[|\|\boldsymbol{S}\boldsymbol{x}\|_2^2 - 1| > \epsilon] < \delta.$$

Results in [18] show that sketch matrix from a random bipartite graph (where again the edges for each left vertex and the signs ± 1 are assigned randomly via. a hash function) with degree $s = \mathcal{O}(\log(1/\delta)/\epsilon)$ and $m = \mathcal{O}(\log(1/\delta)/\epsilon^2)$ is JLT. Note that, such random bipartite graphs are also expanders with appropriate parameters [16]. We next use the result by Sarlos in [28, Corollary 11], that showed that any JLT with $m = O\left(\frac{k \log(k/\delta\epsilon)}{\epsilon^2}\right)$ satisfies

$$Pr[\forall i \in [1..k] : |1 - \sigma_i(\boldsymbol{S}\boldsymbol{U})| \leq \epsilon] \geq 1 - \delta.$$

A similar result was proved using a different approach in [22] for sparse sketch matrices. □

Using the relations in [22, 18], we can show that the above results for expanders hold even when the random signs $\pm$ assignment is just $\mathcal{O}(\log k)$-wise independent, as well as the s edges assignment for each left vertex (the column permutations) can also be $\mathcal{O}(\log k)$-wise independent. This implies, we can use the code based expanders [13, 1, 7] to construct sparse sketching matrices with reduced randomness (almost deterministic). A code matrix obtained from a coding scheme with dual distance γ will have columns that are γ-wise independent [32] (the γ-wise independence property of codes). Hence, an expander with the above s, m values constructed from a code with dual distance $\mathcal{O}(\log k)$ will yield a sparse sketching matrix that is subspace embedding.

4. NUMERICAL RESULTS

In this section, we present numerical results that illustrate the performance of the expander (random bipartite) graph based sketching matrices on three real datasets, and compare them with two of the most commonly used sketching matrices, viz., Countsketch [11] and Gaussian sketch [34]. The three datasets are: Movielens 100k dataset [6], which is a sparse matrix with size 1682×943; Jester dataset [12], which is a dense
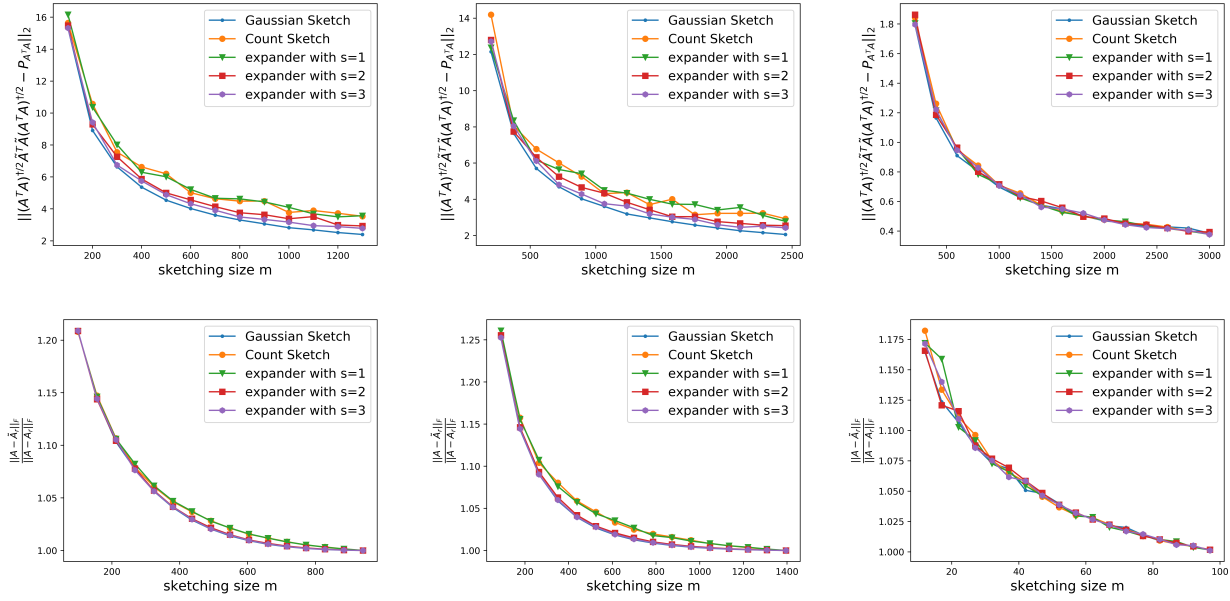
Fig. 3. The top three images plot the distortion against sketching size m and the bottom 3 plots are the relative low-rank approximation error against sketching size. The first pair of plots (top and bottom left) are for the Movielens100K dataset, the second pair of plots are for the CRAN (LSI) dataset, and the last pair is for Jester dataset. Results averaged over 10 trials.

dataset with size 7200×100 (both from the matrix completion application); and CRAN [20], a Latent Semantic Indexing (LSI) dataset, which is sparse with size 5204×1398. The distortion metric used in our results is defined as:

Definition 4 (Distortion of $\tilde{A}$). *The distortion η of a sketched matrix $\tilde{A} = SA$ with respect to the original matrix A is calculated by* $\eta = \left\| I - \left(A^\top A \right)^{-1/2} \tilde{A}^\top \tilde{A} \left(A^\top A \right)^{-1/2} \right\|_2$.

For further details, we refer to page 3 of [21]. Figure 3 presents numerical results. The performance of the expander graph based matrices is comparable to the popular Countsketch and Gaussian sketch matrices, and the performance slightly improves as we increase the degree s. The performance is similar for the low-rank approximation application as well. Note that the magical graph construction in [8] is the same as the expander with $s = 2$ in our figure. Since SRFT and SRHT also yield similar results on these datasets, we elected to exclude them for visual clarity sake. The sketch matrices will have different size n as per the original data matrix.

For the low-rank approximation, we apply the following steps after sketching: WLOG, we sketch the row space of the input matrix A, and we get $Y = SA$, and we compute QR factorization to get $Q, R = \text{qr}(Y^\top)$. Next, we construct $B = AQ$, compute rank-k SVD of B, $[U_k, \Sigma_k, V_k] = SVD(B, k)$ and set $V_k = QV_k$. And then the low-rank approximation is calculated from the formula $\|A - AV_kV_k^\top\|_F$, and we denote $\tilde{A}_k = AV_kV_k^\top$ for the y axis in the bottom 3 plots, and A_k is the true rank-k approximation.

Conclusions and future work

In this study, we proposed sparse graph based sketching matrices for fast numerical linear algebra computations. We presented theoretical results that show that under certain conditions, these matrices satisfy the subspace embedding property. Numerical results illustrated the performance of different sketching matrices in applications. We have made our implementation publicly available as a *Python Matrix-Sketching toolkit*: `https://github.com/jurohd/Matrix-Sketching-Toolkits`. As part of our future work, we plan to explore randomized sketching techniques for nonlinear operators, with applications to deep learning.

Acknowledgements

5. REFERENCES

[1] N. Alon and Y. Roichman. Random Cayley graphs and expanders. *Random Structures & Algorithms*, 5(2):271–284, 1994.

[2] H. Avron, V. Sindhwani, and D. Woodruff. Sketching structured matrices for faster nonlinear regression. In *Advances in neural information processing systems*, pages 2994–3002, 2013.

[3] C. Boutsidis and A. Gittens. Improved matrix algorithms via the

subsampled randomized Hadamard transform. *SIAM Journal on Matrix Analysis and Applications*, 34(3):1301–1340, 2013.

[4] M. Capalbo, O. Reingold, S. Vadhan, and A. Wigderson. Randomness conductors and constant-degree lossless expanders. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 659–668, 2002.

[5] A. Chatalic, R. Gribonval, and N. Keriven. Large-scale high-dimensional clustering with fast sketching. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4714–4718. IEEE, 2018.

[6] Y. Chen, F. M. Harper, J. Konstan, and S. X. Li. Social comparisons and contributions to online communities: A field experiment on movielens. *American Economic Review*, 100(4):1358–98, 2010.

[7] M. Cheraghchi. Derandomization and group testing. In *2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 991–997. IEEE.

[8] H. Y. Cheung, T. C. Kwok, and L. C. Lau. Fast matrix rank algorithms and applications. *Journal of the ACM (JACM)*, 60(5):1–25, 2013.

[9] H. Choi, T. Jiang, and Y. Shi. Large-scale beamforming for massive MIMO via randomized sketching. *arXiv preprint arXiv:1903.05904*, 2019.

[10] K. L. Clarkson and D. P. Woodruff. Numerical linear algebra in the streaming model. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 205–214, 2009.

[11] K. L. Clarkson and D. P. Woodruff. Low-rank approximation and regression in input sparsity time. *Journal of the ACM (JACM)*, 63(6):1–45, 2017.

[12] K. Goldberg, T. Roeder, D. Gupta, and C. Perkins. Eigentaste: A constant time collaborative filtering algorithm. *information retrieval*, 4(2):133–151, 2001.

[13] V. Guruswami, C. Umans, and S. Vadhan. Unbalanced expanders and randomness extractors from parvaresh–vardy codes. *Journal of the ACM (JACM)*, 56(4):1–34, 2009.

[14] N. Halko, P.-G. Martinsson, and J. A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM review*, 53(2):217–288, 2011.

[15] T. Hastie, R. Mazumder, J. D. Lee, and R. Zadeh. Matrix completion and low-rank SVD via fast alternating least squares. *The Journal of Machine Learning Research*, 16(1):3367–3402, 2015.

[16] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–561, 2006.

[17] S. Jafarpour, W. Xu, B. Hassibi, and R. Calderbank. Efficient and robust compressed sensing using optimized expander graphs. *IEEE Transactions on Information Theory*, 55(9):4299–4308, 2009.

[18] D. M. Kane and J. Nelson. Sparser Johnson-Lindenstrauss transforms. *Journal of the ACM (JACM)*, 61(1):1–23, 2014.

[19] R. Kannan and S. Vempala. Randomized algorithms in numerical linear algebra. *Acta Numerica*, 26:95, 2017.

[20] E. Kokiopoulou and Y. Saad. Polynomial filtering in latent semantic indexing for information retrieval. In *ACM SIGIR conference on Research and development in information retrieval*, pages 104–111, 2004.

[21] M. Magdon-Ismail and A. Gittens. Fast fixed dimension L2-subspace embeddings of arbitrary accuracy, with application to l1 and l2 tasks. *arXiv preprint arXiv:1909.12580*, 2019.

[22] J. Nelson and H. L. Nguyên. OSNAP: Faster numerical linear algebra algorithms via sparser subspace embeddings. In *2013 ieee 54th annual symposium on foundations of computer science*, pages 117–126. IEEE, 2013.

[23] I. K. Ozaslan, M. Pilanci, and O. Arikan. Iterative hessian sketch with momentum. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7470–7474. IEEE, 2019.

[24] R. Pagh. Compressed matrix multiplication. *ACM Transactions on Computation Theory (TOCT)*, 5(3):1–17, 2013.

[25] M. Pilanci and M. J. Wainwright. Newton sketch: A near linear-time optimization algorithm with linear-quadratic convergence. *SIAM Journal on Optimization*, 27(1):205–245, 2017.

[26] M. Rahmani, A. Beckus, A. Karimian, and G. K. Atia. Scalable and robust community detection with randomized sketching. *IEEE Transactions on Signal Processing*, 68:962–977, 2020.

[27] O. Reingold, S. Vadhan, and A. Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders and extractors. In *Proceedings 41st Annual Symposium on Foundations of Computer Science*, pages 3–13. IEEE, 2000.

[28] T. Sarlos. Improved approximation algorithms for large matrices via random projections. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pages 143–152. IEEE, 2006.

[29] K. Sayood. *Introduction to data compression*. Morgan Kaufmann, 2017.

[30] M. Sipser and D. A. Spielman. Expander codes. *IEEE transactions on Information Theory*, 42(6):1710–1722, 1996.

[31] S. Ubaru and A. Mazumdar. Multilabel classification with group testing and codes. In *International Conference on Machine Learning*, pages 3492–3501, 2017.

[32] S. Ubaru, A. Mazumdar, and Y. Saad. Low rank approximation and decomposition of large matrices using error correcting codes. *IEEE Transactions on Information Theory*, 63(9):5544–5558, 2017.

[33] S. Ubaru and Y. Saad. Sampling and multilevel coarsening algorithms for fast matrix approximations. *Numerical Linear Algebra with Applications*, 26(3):e2234, 2019.

[34] D. P. Woodruff. Sketching as a tool for numerical linear algebra. *Theoretical Computer Science*, 10(1-2):1–157, 2014.

[35] F. Woolfe, E. Liberty, V. Rokhlin, and M. Tygert. A fast randomized algorithm for the approximation of matrices. *Applied and Computational Harmonic Analysis*, 25(3):335–366, 2008.

[36] B. Yang. Projection approximation subspace tracking. *IEEE Transactions on Signal processing*, 43(1):95–107, 1995.

[37] X. Zhang and S. Liao. Incremental randomized sketching for online kernel learning. In *International Conference on Machine Learning*, pages 7394–7403, 2019.