Dimensionality Reduction — Notes 3

Jelani Nelson
minilek@seas.harvard.edu

August 13, 2015

1 Gordon's theorem

Let T be a finite subset of some normed vector space with norm $\|\cdot\|_X$. We say that a sequence $T_0 \subseteq T_1 \subseteq \ldots \subseteq T$ is *admissible* if $|T_0| = 1$ and $|T_r| \le 2^{2^r}$ for all $r \ge 1$, and $T_r = T$ for all $r \ge r_0$ for some r_0. We define the γ_2-functional

$$\gamma_2(T, \|\cdot\|_X) = \inf \sup_{x \in T} \sum_{r=1}^{\infty} 2^{r/2} \cdot d_X(x, T_r),$$

where the inf is taken over all admissible sequences. We also let $d_X(T)$ denote the diameter of T with respect to norm $\|\cdot\|_X$. For the remainder of this section we make the definitions $\pi_r x = \mathrm{argmin}_{y \in T_r} \|y - x\|_X$ and $\Delta_r x = \pi_r x - \pi_{r-1} x$.

Throughout this section we let $\|\cdot\|$ denote the $\ell_{2\to 2}$ operator norm in the case of matrix arguments, and the ℓ_2 norm in the case of vector arguments.

Krahmer, Mendelson, and Rauhut showed the following theorem [KMR14].

Theorem 1. *Let $\mathcal{A} \subset \mathbb{R}^{m\times n}$ be arbitrary. Let $\varepsilon_1, \ldots, \varepsilon_n$ be independent ± 1 random variables. Then*

$$\mathop{\mathbb{E}}_{\varepsilon} \sup_{A \in \mathcal{A}} \left| \|A\varepsilon\|^2 - \mathbb{E}\|A\varepsilon\|^2 \right| \lesssim \gamma_2^2(\mathcal{A}, \|\cdot\|) + \gamma_2(\mathcal{A}, \|\cdot\|) \cdot d_F(\mathcal{A}) + d_F(\mathcal{A}) \cdot d_{\ell_{2\to 2}}(\mathcal{A}).$$

The KMR theorem was actually more general, where the Rademacher variables could be replaced by subgaussian random variables. We present just the proof of the Rademacher case.

Proof. Without loss of generality we can assume $\mathcal{A}$ is finite (else apply the theorem to a sufficiently fine net, i.e. fine in $\ell_2 \to \ell_2$ operator norm). Define

$$E = \mathop{\mathbb{E}}_{\varepsilon} \sup_{A \in \mathcal{A}} \left| \|A\varepsilon\|^2 - \mathbb{E}\, \|A\varepsilon\|^2 \right|$$

and let A^i denote the ith column of A. Then by decoupling

$$\begin{aligned} E &= \mathop{\mathbb{E}}_{\varepsilon} \sup_{A \in \mathcal{A}} \left| \sum_{i \neq j} \varepsilon_i \varepsilon_j \left\langle A^i, A^j \right\rangle \right| \\ &\le 4 \cdot \mathop{\mathbb{E}}_{\varepsilon, \varepsilon'} \sup_{A \in \mathcal{A}} \left| \sum_{i,j} \varepsilon_i \varepsilon'_j \left\langle A^i, A^j \right\rangle \right| \\ &= 4 \cdot \mathop{\mathbb{E}}_{\varepsilon, \varepsilon'} \sup_{A \in \mathcal{A}} \left| \langle A\varepsilon, A\varepsilon' \rangle \right| . \end{aligned}$$

Let $\{T_r\}_{r=0}^{\infty}$ be admissible for $\mathcal{A}$. Direct computation shows

$$\langle A\varepsilon, A\varepsilon' \rangle = \langle (\pi_0 A)\varepsilon, (\pi_0 A)\varepsilon' \rangle + \sum_{r=1}^{\infty} \underbrace{\langle (\Delta_r A)\varepsilon, (\pi_{r-1} A)\varepsilon' \rangle}_{X_r(A)} + \sum_{r=1}^{\infty} \underbrace{\langle (\pi_r A)\varepsilon, (\Delta_r A)\varepsilon' \rangle}_{Y_r(A)} .$$

We have $T_0 = \{A_0\}$ for some $A_0 \in \mathcal{A}$. Thus $\mathbb{E}_{\varepsilon,\varepsilon'} \left| \langle (\pi_0 A)\varepsilon, (\pi_0 A)\varepsilon' \rangle \right|$ equals

$$\mathop{\mathbb{E}}_{\varepsilon, \varepsilon'} |\varepsilon^* A_0^* A_0 \varepsilon'| \le \left(\mathop{\mathbb{E}}_{\varepsilon, \varepsilon'} (\varepsilon^* A_0^* A_0 \varepsilon')^2 \right)^{1/2} = \|A_0^* A_0\|_F \le \|A_0\|_F \|A_0\| \le d_F(\mathcal{A}) \cdot d_{\ell_2 \to 2}(\mathcal{A}).$$

Thus,

$$\mathop{\mathbb{E}}_{\varepsilon, \varepsilon'} \sup_{A \in \mathcal{A}} |\langle A\varepsilon, A\varepsilon' \rangle| \le d_F(\mathcal{A}) \cdot d_{\ell_2 \to 2}(\mathcal{A}) + \mathop{\mathbb{E}}_{\varepsilon, \varepsilon'} \sup_{A \in \mathcal{A}} \sum_{r=1}^{\infty} |X_r(A)| + \mathop{\mathbb{E}}_{\varepsilon, \varepsilon'} \sup_{A \in \mathcal{A}} \sum_{r=1}^{\infty} |Y_r(A)|.$$

We focus on the second summand; handling the third summand is similar.

Note $X_r(A) = \langle (\Delta_r A)\varepsilon, (\pi_{r-1} A)\varepsilon' \rangle = \langle \varepsilon, (\Delta_r A)^* (\pi_{r-1} A)\varepsilon' \rangle$. Thus

$$\mathbb{P}(|X_r(A)| > t 2^{r/2} \cdot \|(\Delta_r A)^* (\pi_{r-1} A)\varepsilon'\|) \lesssim e^{-t^2 2^r / 2} \text{ (Khintchine).}$$

Let $\mathcal{E}(A)$ be the event that for all $r \ge 1$ simultaneously, $|X_r(A)| \le t 2^{r/2} \cdot \|\Delta_r A\| \cdot \sup_{A \in \mathcal{A}} \|A\varepsilon'\|$. Then

$$\mathbb{P}(\exists A \in \mathcal{A} \ s.t.\ \neg \mathcal{E}(A)) \lesssim \sum_{r=1}^{\infty} |T_r| \cdot |T_{r-1}| \cdot e^{-t^2 2^r / 2}$$

$$\leq \sum_{r=1}^{\infty} 2^{2^{r+1}} \cdot e^{-t^2 2^r/2}.$$

Therefore

$$\mathop{\mathbb{E}}_{\varepsilon,\varepsilon'} \sup_{A\in\mathcal{A}} \sum_{r=1}^{\infty} |X_r(A)| = \mathop{\mathbb{E}}_{\varepsilon'} \int_0^{\infty} \mathop{\mathbb{P}}_{\varepsilon}\left(\sup_{A\in\mathcal{A}} \sum_{r=1}^{\infty} |X_r(A)| > t\right) dt,$$

which by a change of variables is equal to

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\varepsilon'}\left(\sup_{A\in\mathcal{A}} \|A\varepsilon'\| \cdot \left(\sup_{A\in\mathcal{A}} \sum_{r=1}^{\infty} 2^{r/2} \|\Delta_r A\|\right)\right. \\
&\qquad \left.\times \cdot \int_0^{\infty} \mathop{\mathbb{P}}_{\varepsilon}\left(\sup_{A\in\mathcal{A}} \sum_{r=1}^{\infty} |X_r(A)| > t \sup_{A\in\mathcal{A}} 2^{r/2} \cdot \|\Delta_r A\| \cdot \sup_{A\in\mathcal{A}} \|A\varepsilon'\|\right) dt\right) \\
&\leq \left(\mathop{\mathbb{E}}_{\varepsilon'} \sup_{A\in\mathcal{A}} \|A\varepsilon'\|\right) \cdot \left(\sup_{A\in\mathcal{A}} \sum_{r=1}^{\infty} 2^{r/2} \|\Delta_r A\|\right) \cdot \left[3 + \sum_{r=1}^{\infty} \int_3^{\infty} 2^{2^{r+1}} e^{-t^2 2^r/2} dt\right] \\
&\lesssim \left(\mathop{\mathbb{E}}_{\varepsilon'} \sup_{A\in\mathcal{A}} \|A\varepsilon'\|\right) \cdot \sup_{A\in\mathcal{A}} \sum_{r=1}^{\infty} 2^{r/2} \|\Delta_r A\| \\
&\lesssim \left(\mathop{\mathbb{E}}_{\varepsilon'} \sup_{A\in\mathcal{A}} \|A\varepsilon'\|\right) \cdot \sup_{A\in\mathcal{A}} \sum_{r=1}^{\infty} 2^{r/2} \cdot d_{2\to 2}(A, T_r),
\end{aligned}$$

since $\|\Delta_r A\| \leq d_{2\to 2}(A, T_{r-1}) + d_{2\to 2}(A, T_r)$ via the triangle inequality. Choosing admissible $T_0 \subseteq T_1 \subseteq \ldots \subseteq T$ to minimize the above expression,

$$E \lesssim d_F(\mathcal{A}) \cdot d_{\ell_{2\to 2}}(\mathcal{A}) + \gamma_2(\mathcal{A}, \|\cdot\|) \cdot \mathop{\mathbb{E}}_{\varepsilon'} \sup_{A\in\mathcal{A}} \|A\varepsilon'\|.$$

Now observe

$$\begin{aligned}
\mathop{\mathbb{E}}_{\varepsilon'}\left(\sup_{A\in\mathcal{A}} \|A\varepsilon'\|\right) &\leq \left(\mathop{\mathbb{E}}_{\varepsilon'} \sup_{A\in\mathcal{A}} \|A\varepsilon'\|^2\right)^{1/2} \\
&\leq \left(\mathop{\mathbb{E}}_{\varepsilon'}\left(\sup_{A\in\mathcal{A}} \left| \|A\varepsilon'\|^2 - \mathop{\mathbb{E}}_{\varepsilon'} \|A\varepsilon'\|^2 \right| + \mathop{\mathbb{E}}_{\varepsilon'} \|A\varepsilon'\|^2\right)\right)^{1/2} \\
&= \left(\mathop{\mathbb{E}}_{\varepsilon'} \sup_{A\in\mathcal{A}} \left(\left| \|A\varepsilon'\|^2 - \mathop{\mathbb{E}}_{\varepsilon'} \|A\varepsilon'\|^2 \right| + \|A\|_F^2\right)\right)^{1/2}
\end{aligned}$$

$$\leq \sqrt{E} + d_F(\mathcal{A})$$

Thus in summary,

$$E \lesssim d_F(\mathcal{A}) \cdot d_{\ell_2 \to 2}(\mathcal{A}) + \gamma_2(\mathcal{A}, \|\cdot\|) \cdot (\sqrt{E} + d_F(\mathcal{A})).$$

This implies E is at most the square of the larger root of the associated quadratic equation, which gives the theorem. $\square$

Using the KMR theorem, we can recover Gordon's theorem [Gor88] (also see [KM05, MPTJ07, Dir14]). We again only discuss the Rademacher case. Note that in metric JL, we wish for the set of vectors X that

$$\forall x, y \in X, \ |\|\Pi(x-y)\|_2^2 - \|x-y\|_2^2\| < \varepsilon \|x-y\|_2^2.$$

If we define

$$T = \left\{ \frac{x-y}{\|x-y\|_2} : x, y \in X \right\},$$

then it is equivalent to have

$$\sup_{x \in T} |\|\Pi x\|_2^2 - 1| < \varepsilon.$$

Since Π is random, we will demand that this holds in expectation

$$\mathop{\mathbb{E}}_{\Pi} \sup_{x \in T} |\|\Pi x\|_2^2 - 1| < \varepsilon. \tag{1}$$

Theorem 2. *Let $T \subset \mathbb{R}^n$ be a set of vectors each of unit norm, and let $\varepsilon \in (0, 1/2)$ be arbitrary. Let $\Pi \in \mathbb{R}^{m \times n}$ be such that $\Pi_{i,j} = \sigma_{i,j}/\sqrt{m}$ for independent Rademacher $\sigma_{i,j}$, and where $m = \Omega((\gamma_2^2(T, \|\cdot\|) + 1)/\varepsilon^2)$. Then*

$$\mathbb{E} \sup_{x \in T} \left| \|\Pi x\|^2 - 1 \right| < \varepsilon.$$

Proof. For $x \in T$ let A_x denote the $m \times mn$ matrix defined as follows:

$$A_x = \frac{1}{\sqrt{m}} \cdot \begin{bmatrix} x_1 & \cdots & x_n & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \cdots & 0 & x_1 & \cdots & x_n & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & x_1 & \cdots & x_n \end{bmatrix}.$$

Then $\|\Pi x\|^2 = \|A_x \sigma\|^2$, so letting $\mathcal{A} = \{A_x : x \in T\}$,

$$\mathbb{E} \sup_{x\in T} \left| \|\Pi x\|^2 - 1 \right| = \mathbb{E} \sup_{A\in\mathcal{A}} \left| \|A\sigma\|^2 - \mathbb{E} \|A\sigma\|^2 \right| .$$

We have $d_F(\mathcal{A}) = 1$. Also $A_x^* A_x$ is a block-diagonal matrix, with m blocks each equal to xx^*/m, and thus the singular values of A_x are 0 and $\|x\|/\sqrt{m}$, implying $d_{\ell_{2\to 2}}(\mathcal{A}) = 1/\sqrt{m}$. Similarly, since $A_x - A_y = A_{x-y}$, for any vectors x, y we have $\|A_x - A_y\| = \|x - y\|$, and thus $\gamma_2(\mathcal{A}, \|\cdot\|) = \gamma_2(T, \|\cdot\|)/\sqrt{m}$. Thus by the KMR theorem we have

$$\mathbb{E} \sup_{x\in T} \left| \|\Pi x\|^2 - 1 \right| \lesssim \frac{\gamma_2^2(T, \|\cdot\|)}{m} + \frac{\gamma_2(T, \|\cdot\|)}{\sqrt{m}} + \frac{1}{\sqrt{m}},$$

which is at most ε for m as in the theorem statement. $\square$

Gordon's theorem was actually stated differently in [Gor88] in two ways: (1) Gordon actually only analyzed the case of Π having i.i.d. gaussian entries, and (2) the $\gamma_2(T, \|\cdot\|)$ terms in the theorem statement were written as the *gaussian mean width* $g(T) = \mathbb{E}_g \sup_{x\in T} \langle g, x\rangle$, where $g \in \mathbb{R}^n$ is a vector of i.i.d. standard normal random variables. For (1), the extension to arbitrary subgaussian random variables was shown first in [KM05]. Note the KMR theorem only bounds an expectation; thus if one wants to argue that the random variable in question is large with with probability at most δ, the most obvious way is Markov, which would introduce JL a poor $1/\delta^2$ dependence in m. One could remedy this by doing Markov on the pth moment; the tightest known p-norm bound is given in [Dir13, Theorem 6.5] (see also [Dir14, Theorem 4.8]).

For (2), Gordon actually wrote his paper before γ_2 was even defined! The definition of γ_2 given here is due to Talagrand, who also showed that for all sets of vectors $T \subset \mathbb{R}^n$, $g(T) \simeq \gamma_2(T, \|\cdot\|)$ [Tal14] (this is known as the "Majorizing Measures" theorem). In fact the upper bound $g(T) \lesssim \gamma_2(T, \|\cdot\|)$ was shown by Fernique [Fer75] (although γ_2 was not defined at that point; Talagrand later recast this upper bound in terms of his newly defined γ_2-functional).

We thus state the following corollary of the majorizing measures theorem and Theorem 2.

Corollary 1. *Let $T \subset \mathbb{R}^n$ be a set of vectors each of unit norm, and let $\varepsilon \in (0, 1/2)$ be arbitrary. Let $\Pi \in \mathbb{R}^{m\times n}$ be such that $\Pi_{i,j} = \sigma_{i,j}/\sqrt{m}$ for* *independent Rademacher $\sigma_{i,j}$, and where $m = \Omega((g^2(T)+1)/\varepsilon^2)$. Then*

$$\mathbb{E} \sup_{x \in T} \left| \|\Pi x\|^2 - 1 \right| < \varepsilon.$$

1.1 Application 1: numerical linear algebra

Consider, for example, the least squares regression problem. Given $A \in \mathbb{R}^{n \times d}, b \in \mathbb{R}^n$, $n \gg d$, the goal is to compute

$$x^* = \underset{x \in \mathbb{R}^n}{argmin} \|Ax - b\|_2. \tag{2}$$

It is standard that

$$x^* = (A^T A)^{-1} A^T b$$

when A has full column rank. Unfortunately, naively computing $A^T A$ takes time $\Theta(nd^2)$. We would like to speed this up.

Given our lectures on dimensionality reduction, one natural question is the following: if instead we compute

$$\tilde{x}^* = \underset{x \in \mathbb{R}^n}{argmin} \|\Pi Ax - \Pi b\|_2$$

for some JL map Π with few rows m, can we argue that $\tilde{x}^*$ is a good solution for (2)? The answer is yes.

Theorem 3. *Suppose* (1) *holds for T the unit vectors in the subspace spanned by b and the columns of A. Then*

$$\|A\tilde{x} - b\|_2^2 \leq \frac{1+\varepsilon}{1-\varepsilon} \cdot \|Ax^* - b\|_2^2$$

Proof.

$$(1-\varepsilon)\|A\tilde{x}^* - b\|_2^2 \leq \|\Pi A\tilde{x}^* - \Pi b\|_2^2 \leq \|\Pi Ax^* - \Pi b\|_2^2 \leq (1+\varepsilon)\|Ax^* - b\|_2^2.$$

The first and third inequalities hold since Π preserves $A\tilde{x}^* - b$ and $Ax^* - b$. The second inequality holds since $\tilde{x}^*$ is the optimal solution to the lower dimensional regression problem. □

Now we may ask ourselves: what is the number of rows m needed to preserve the vector T in Theorem 3? We apply Corollary 1. Note T is the set of unit vectors in a subspace of dimension $D \le d+1$. By rotational symmetry of the gaussian, we can assume this subspace equals $span\{e_1, \dots, e_D\}$. Then

$$\mathop{\mathbb{E}}_{g \in \mathbb{R}^D} \sup_{x \in \ell_2^D} \langle g, x \rangle = \mathop{\mathbb{E}}_{g \in \mathbb{R}^D} \|g\|_2 \le (\mathop{\mathbb{E}}_{g \in \mathbb{R}^D} \|g\|_2^2)^{1/2} = \sqrt{D}.$$

Thus it suffices for Π to have $m \gtrsim d/\varepsilon^2$ rows.

Unfortunately in the above, although solving the lower-dimensional regression problem is fast (since now ΠA has $O(d/\varepsilon^2)$ rows compared with the n rows of A), multiplying ΠA using dense random Π is actually *slower* than solving the original regression problem (2). This was remedied by Sarlós in [Sar06] by using a fast JL matrix as in Lecture 2; see [CNW15, Theorem 9] for the tightest analysis of this construction in this context. An alternative is to use a sparse Π. The first analysis of this approach was in [CW13]. The tightest known analyses are in [MM13, NN13, BDN15].

It is also the case that Π can be used more efficently to solve regression problems than simply requiring (1) for T as above. See for example [CW13, Theorem 7.7] in the full version of that paper for an iterative algorithm based on such Π which has running time dependence on ε equal to $O(\log(1/\varepsilon))$, instead of the $poly(1/\varepsilon)$ above. For further results on applying JL to problems in this domain, see the book [Woo14].

1.2 Application 2: compressed sensing

In *compressed sensing*, the goal is to (approximately) recover an (approximately) sparse signal $x \in \mathbb{R}^n$ from a few linear measurements. We will imagine that these m linear measurements are organized as the rows of a matrix $\Pi \in \mathbb{R}^{m \times n}$. Let T^k be the set of all k-sparse vectors in $\mathbb{R}^n$ of unit norm (i.e. the union of $\binom{n}{k}$ k-dimensional subspaces). One always has

$$\gamma_2(T, \|\cdot\|) \le \inf_{\{T_r\}} \sum_{r=1}^{\infty} 2^{r/2} \cdot \sup_{x \in T} d_{\ell_2}(x, T_r),$$

i.e. the sup can be moved inside the sum to obtain an upper bound. Minimizing the right hand side amounts to finding the best nets possible for T of some bounded size 2^{2^k} for each k. By doing this, which we do not discuss here, one can show that for our T^k, $\gamma_2(T^k, \|\cdot\|) \lesssim \sqrt{k \log(n/k)}$ so that one can obtain (1) with $m \simeq k \log(n/k)/\varepsilon^2$. A more direct net argument can also yield this bound (see [BDDW08] which suffered a $\log(1/\varepsilon)$ factor, and the removal of this factor in [FR13, Theorem 9.12]).

Now, any matrix Π preserving this T^k with distortion $1+\varepsilon$ is known as having the (k,ε)-*restricted isometry property (RIP)* [CT06]. We are ready to state a theorem of [CT06, Don06]. One can find a short proof in [Can08].

Theorem 4. *Suppose* Π *satisfies the* $(2k, \sqrt{2}-1)$*-RIP. Then given* $y = \Pi x$*, if one solves the linear program*

$$\begin{aligned} &\min \|z\|_1 \\ &s.t.\ \Pi z = y \end{aligned}$$

then the optimal solution $\tilde{x}$ *will satisfy*

$$\|x - \tilde{x}\|_2 = O(1/\sqrt{k}) \cdot \inf_{\substack{w \in \mathbb{R}^n \\ |\,supp(w)| \le k}} \|x - w\|_1.$$

References

[BDDW08] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constr. Approx.*, 28(3):253–263, 2008.

[BDN15] Jean Bourgain, Sjoerd Dirksen, and Jelani Nelson. Toward a unified theory of sparse dimensionality reduction in Euclidean space. *Geometric and Functional Analysis (GAFA), to appear*, 2015. Preliminary version in STOC 2015.

[Can08] Emmanuel Candès. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathematique*, 346(9-10):589–592, 2008.

[CNW15] Michael B. Cohen, Jelani Nelson, and David P. Woodruff. Optimal approximate matrix product in terms of stable rank. *CoRR*, abs/1507.02268, 2015.

[CT06] Emmanuel J. Candès and Terence Tao. Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inform. Theory*, 52:5406–5425, 2006.

[CW13] Kenneth L. Clarkson and David P. Woodruff. Low rank approximation and regression in input sparsity time. In *Proceedings of the 45th ACM Symposium on Theory of Computing (STOC)*, pages 81–90, 2013. Full version at `http://arxiv.org/pdf/1207.6365v4.pdf`.

[Dir13] Sjoerd Dirksen. Tail bounds via generic chaining. *CoRR*, abs/1309.3522v2, 2013.

[Dir14] Sjoerd Dirksen. Dimensionality reduction with subgaussian matrices: a unified theory. *CoRR*, abs/1402.3973, 2014.

[Don06] D. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.

[Fer75] Xavier Fernique. Regularité des trajectoires des fonctions aléatoires gaussiennes. *Lecture Notes in Math.*, 480:1–96, 1975.

[FR13] Simon Foucart and Holger Rauhut. *A Mathematical Introduction to Compressive Sensing*. Birkhaüser, Boston, 2013.

[Gor88] Yehoram Gordon. On Milman's inequality and random subspaces which escape through a mesh in $\mathbb{R}^n$. *Geometric Aspects of Functional Analysis*, pages 84–106, 1988.

[KM05] Bo'az Klartag and Shahar Mendelson. Empirical processes and random projections. *J. Funct. Anal.*, 225(1):229–245, 2005.

[KMR14] Felix Krahmer, Shahar Mendelson, and Holger Rauhut. Suprema of chaos processes and the restricted isometry property. *Comm. Pure Appl. Math.*, 67(11):1877–1904, 2014.

[MM13] Xiangrui Meng and Michael W. Mahoney. Low-distortion subspace embeddings in input-sparsity time and applications to robust linear regression. In *Proceedings of the 45th ACM Symposium on Theory of Computing (STOC)*, pages 91–100, 2013.

[MPTJ07] Shahar Mendelson, Alain Pajor, and Nicole Tomczak-Jaegermann. Reconstruction and subgaussian operators in asymptotic geometric analysis. *Geometric and Functional Analysis*, 17:1248–1282, 2007.

[NN13] Jelani Nelson and Huy L. Nguyễn. OSNAP: faster numerical linear algebra algorithms via sparser subspace embeddings. In *Proceedings of the 54th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 117–126, 2013.

[Sar06] Tamás Sarlós. Improved approximation algorithms for large matrices via random projections. In *47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006), 21-24 October 2006, Berkeley, California, USA, Proceedings*, pages 143–152, 2006.

[Tal14] Michel Talagrand. *Upper and lower bounds for stochastic processes: modern methods and classical problems*. Springer, 2014.

[Woo14] David P. Woodruff. Sketching as a tool for numerical linear algebra. *Foundations and Trends in Theoretical Computer Science*, 10(1-2):1–157, 2014.