CS 395T: Sublinear Algorithms Fall 2016

Lecture 21 — November 8, 2016

Prof. Eric Price *Scribes: William Hoza and Vasileios-Orestis Papadigenopoulos*

1 Overview

Last time, we began a proof that a matrix consisting of a random small subset of the rows of a Fourier matrix has the RIP "in expectation." Today, we complete the proof of this theorem.

Let $F \in \mathbb{C}^{n\times n}$ be a Fourier matrix. For our purposes, the important properties of this matrix are that for all i, j, $|F_{ij}| \leq 1$, and

$$\langle F_i, F_j \rangle = \begin{cases} 0 & \text{if } i \neq j \\ n & \text{if } i = j. \end{cases}$$

Let Ω to be a multiset of m random rows of the Fourier matrix F. Observe that for any set $s \subseteq [n]$,

$$\mathbb{E}[F^T_{\Omega\times s} F_{s\times\Omega}] = m\mathbb{E}_{i\in[n]}[F_i^T F_i] = mI_n.$$

Define

$$\Delta \stackrel{\text{def}}{=} \mathbb{E}\left[\sup_{s:|s|\leq k} \left\| I_{k\times k} - \frac{1}{m} F^T_{\Omega\times s} F_{s\times\Omega} \right\|\right],$$

so that Δ is a measure of the expected RIP error of F_Ω. Our goal is to show that Δ is small.

2 Recap of last class

First we symmetrize and Gaussianize: to bound Δ, it suffices to bound

$$\mathop{\mathbb{E}}_{\Omega,g}\left[\left\|\sum_{i\in\Omega} g_i x_i^s x_i^{sT}\right\|\right], \tag{1}$$

where x_i is the column vector F_i^T and $g_1, \ldots, g_m$ are independent Gaussians. In fact, we will bound the expectation over g alone in Equation **??** for an *arbitrary* Ω. Let Σ_k be the set of k-sparse unit vectors. Then

$$(\textbf{??}) \leq \mathbb{E}\left[\sup_{y_i\in\Sigma_k}\left|\sum_{i\in\Omega} g_i\langle x_i, y\rangle^2\right|\right] \leq 2\mathbb{E}\left[\sup_{y_i\in\Sigma_k} \underbrace{\sum_{i\in\Omega} g_i\langle x_i, y\rangle^2}_{G_y}\right].$$

This is a Gaussian process. Last class, we showed that it can be bounded by Dudley's entropy integral, giving

$$m\Delta \lesssim \int_0^\infty \sqrt{\log N(\Sigma_k, d, u)}\, du,$$

where $d(y, z) = \mathbb{E}[(G_y - G_z)^2]^{\frac{1}{2}}$.

3 Bounding Dudley's entropy integral

3.1 Bounding the Gaussian process metric

By the definition of our Gaussian process,

$$G_y - G_z = \sum_{i \in \Omega} g_i(\langle x_i, y\rangle^2 - \langle x_i, z\rangle^2),$$

and hence

$$\begin{aligned}
d(y,z) &= (\sum_i (\langle x_i, y\rangle^2 - \langle x_i, z\rangle^2)^2)^{\frac{1}{2}} \\
&= (\sum_i (((\langle x_i, y+z\rangle \langle x_i, y-z\rangle)^2))^{\frac{1}{2}} \\
&\leq (\sum_i ((\langle x_i, y+z\rangle)^2)^{\frac{1}{2}}) \max_{i \in \Omega} |\langle x_i, y-z\rangle| \\
&= \|F_\Omega(y+z)\|_2 \|F_\Omega(y-z)\|_\infty \\
&\leq (2 \sup_{y' \in \Sigma_k} \|F_\Omega y'\|) \|F_\Omega(y-z)\|_\infty \\
&\leq 2\sqrt{(1+\Delta)m} \|F_\Omega(y-z)\|_\infty,
\end{aligned}$$

where the last inequality follows from the fact that:

$$\forall y \in \Sigma_k : y^T \left(I - \frac{1}{m} F_\Omega^T F_\Omega\right) y \leq 1 + \Delta$$
$$\Rightarrow \|F_\Omega - y\|_2^2 \leq m(1+\Delta) \Rightarrow \|F_\Omega - y\|_2 \leq \sqrt{m(1+\Delta)}.$$

In general, if d, d' are two metrics on a set X with $d \leq d'$ everywhere, $N(X, d, u) \leq N(X, d', u)$, just by the definition of N. Therefore,

$$\begin{aligned}
m\Delta &\lesssim \int_0^{+\infty} \sqrt{\log N(\Sigma_k, d, u)} du \\
&\leq \int_0^{\infty} \sqrt{\log N(\Sigma_k, \|F_\Omega\|_\infty, \frac{u}{2\sqrt{m(1+\Delta)}})}\, du \\
&= 2\sqrt{m(1+\Delta)} \int_0^{\infty} \sqrt{\log N(\Sigma_k, \|F_\Omega\|_\infty, u)}\, du.
\end{aligned}$$

3.2 Road map for the rest of the proof

Our goal, now, is to show something like

$$\int_0^{\infty} \sqrt{\log N(\Sigma_k, \|F_\Omega \cdot\|_\infty, u)}\, du \lesssim \sqrt{k}. \tag{2}$$

If we manage to do that, we'll be able to conclude that

$$\sqrt{m}\Delta \lesssim \sqrt{1+\Delta}\sqrt{k} \Rightarrow \Delta \lesssim \sqrt{1+\Delta}\sqrt{\frac{k}{m}}.$$

This would mean that for some $m \lesssim \frac{1}{\epsilon^2}k$, $\Delta \leq \epsilon\sqrt{1+\Delta}$, which implies $\Delta \lesssim \epsilon$ like we want. (In fact, we'll prove a slightly weaker bound than Equation **??**, leading to a slightly larger number of measurements m.)

3.3 Bounding $N(\Sigma_k, \|\cdot\|_{F,\infty}, u)$ using the $\|\cdot\|_1$ norm

Let $\|y\|_{F,\infty} = \|Fy\|_\infty$. By Cauchy-Schwartz, $\Sigma_k \subseteq B_1 \cdot \sqrt{k}$. Therefore,

$$\begin{aligned}\int_0^\infty \sqrt{\log N(\Sigma_k, \|F_\Omega \cdot\|_\infty, u)}\, du &\leq \int_0^\infty \sqrt{\log N(\Sigma_k, \|\cdot\|_{F,\infty}, u)}\, du \\ &\leq \sqrt{k}\int_0^\infty \sqrt{\log N(B_1, \|\cdot\|_{F,\infty}, u)}\, du.\end{aligned}$$

To bound $N(B_1, \|\cdot\|_{F,\infty}, u)$, observe that since $|F_{ij}| \leq 1$ for every i, j, $\|y\|_{F,\infty} \leq \|y\|_1$. Therefore,

$$N(B_1, \|\cdot\|_{F_\infty}, u) \leq N(B_1, \|\cdot\|_1, u).$$

Using volume ratios, we proved in a previous lecture that

$$N(B_1, \|\cdot\|_1, u) \leq \left(1 + \frac{2}{u}\right)^n. \tag{3}$$

Plugging this into our integral, we obtain

$$\int_0^1 \sqrt{\log N}\, du \leq \int_0^1 \sqrt{n}\sqrt{\log(1+\frac{2}{u})} \lesssim \sqrt{n},$$

which is not a very good bound. (Remember, we are shooting for a bound that is more like $\sqrt{k}$.)

3.4 Bounding $N(\Sigma_k, \|\cdot\|_{F,\infty}, u)$ using Maurey's empirical method

We are going to apply Maurey's Empirical Method to get a better bound. Choose any $y \in B_1$. Then y belongs to the convex hull of $\{\pm e_i, 0\}$. Randomly round y to some $z_r \in \{\pm e_i, 0\}$ in such a way that $\mathbb{E}[z_r] = y$. We do this for $r = 1 \ldots R$, independently each time. There is some sufficiently large R be such that

$$\mathbb{E}\left[\left\|\frac{1}{R}\sum_{r=1}^{\infty} z_r - y\right\|_{F,\infty}\right] \leq u, \forall y.$$

For such an R, we have $N \leq (2n+1)^R$, which is a much better bound on N as long as R is not too big. So now we turn to bounding R. Define $\sigma_R = \mathbb{E}\left[\left\|\frac{1}{R}\sum_{i=1}^R z_i - y\right\|_{F,\infty}\right]$.

To bound σ_R, we symmetrize. Draw $z'_1, z'_2, \dots z'_R \sim$ same distribution. Then:

$$\begin{aligned}
\sigma_R &= \mathbb{E}\left[\left\|\frac{1}{R}\sum_{i=1}^{R} z_i - y\right\|_{F,\infty}\right] \\
&= \mathbb{E}\left[\left\|\frac{1}{R}\sum_{r=1}^{R} z_r - \mathbb{E}\left[\frac{1}{R}\sum_{r=1}^{R} z'_r\right]\right\|_{F,\infty}\right] \\
&\leq \mathbb{E}\left[\left\|\frac{1}{R}\sum_{r=1}^{R} s_r(z_r - z'_r)\right\|\right] \quad (s_r \in \{\pm 1\}, \text{i.i.d}) \\
&\leq 2\mathbb{E}\left[\left\|\frac{1}{R}\sum_{r=1}^{R} s_r z_r\right\|\right]
\end{aligned}$$

For any particular coordinate i,

$$\langle \frac{1}{R}\sum_{r=1}^{R} s_r z_r, x_i\rangle = \frac{1}{R}\sum_{r=1}^{R} s_r \langle z_r, x_i\rangle$$

Notice that since $\langle z_r, x_i\rangle \in \{0,1\}$, this is ≤ 1 in magnitude. Hence, we can apply the Chernoff bound.

$$\begin{aligned}
\mathbb{P}\left[\langle \frac{1}{R}\sum z_r s_r, x\rangle \geq t\right] &\leq e^{\frac{-t^2 R}{2}} \\
\Rightarrow \mathbb{P}\left[\left\|\frac{1}{R}\sum z_r s_r\right\|_{F,\infty} \geq t\right] &\leq 2ne^{\frac{-t^2 R}{2}} \\
\Rightarrow \mathbb{E}\left[\left\|\frac{1}{R}\sum z_r s_r\right\|_{F,\infty}\right] &\lesssim \frac{1}{\sqrt{R}}\sqrt{\log n} \quad (= u \text{ for } (2u+1)^R)
\end{aligned}$$

Therefore setting $R \leftarrow \frac{\log n}{u^2}$ we get:

$$N(B_1, \|\cdot\|_{F,\infty}, u) \leq (2n+1)^{\mathcal{O}(\frac{\log n}{u^2})} \tag{4}$$

3.5 Combining the bounds

If we just tried to plug Equation **??** into our integral, we would get a bound of $\int_0^1 \frac{1}{u}\log n \, du$, which diverges. Instead, we plug in the minimum of the two bounds (Equation **??** and Equation **??**):

$$\int_0^1 \min(\frac{1}{u}\log n, \sqrt{n}\sqrt{\log\frac{1}{u}})du = \int_0^{\frac{1}{n}} \sqrt{n}\sqrt{\log\frac{1}{n}}du + \int_{\frac{1}{n}}^1 \frac{1}{u}\log n du$$

To conclude:

$$m\Delta \lesssim \sqrt{m(\Delta+1)}\sqrt{k}\log^2 n$$

To achieve $\Delta \leq \epsilon$, we choose $m \gtrsim \frac{1}{\epsilon^2} k \log^4 n$.

3.6 Conclusions

The proof we just finished is by Rudelson and Vershynin [?]. Recently, Haviv and Regev [?] improved the bound to $O(\frac{1}{\epsilon^2} k \log n \log^2 k)$. Even this improved bound is worse than the corresponding bound for Gaussian matrices, but Fourier matrices have the benefit of fast multiplication via the FFT.

We've merely bounded the RIP error "in expectation". One can show that if the number of measurements m is increased by a factor of C, the *probability* that the RIP fails is at most e^{-C^2}.

Recall that when we proved that Gaussian matrices have the RIP, we used the fact that they satisfy the JL condition. Krahmer and Ward [?] showed that there is a connection going the other direction (RIP $\implies$ JL.) In particular, suppose A satisfies the (k, ϵ) RIP. We can't hope for A itself to satisfy the JL condition, but let D be a diagonal matrix with i.i.d. uniform ± 1 entries on its diagonal. Krahmer and Ward showed that AD has the $(4\epsilon, e^{-\Omega(k)})$ JL condition.

To get a feel for this bound, recall that a Gaussian matrix with

$$m \lesssim \frac{1}{\epsilon^2} \log \frac{1}{\delta}$$

rows has the (ϵ, δ) JL condition. By our earlier proof, this shows that a Gaussian matrix with

$$m \lesssim \frac{1}{\epsilon^2} k \log(n/k)$$

rows has the (k, ϵ) RIP. By applying the Krahmer-Ward bound, this implies that a Gaussian matrix with

$$m \lesssim \frac{1}{\epsilon^2} \log \frac{1}{\delta} \log \frac{n}{\log(1/\delta)}$$

rows has the (ϵ, δ) JL condition. So we lose something, but not much.

By combining the Krahmer-Ward bound with the proof that Fourier matrices satisfy the RIP, we get a Fourier-based matrix which satisfies the (ϵ, δ) JL condition with

$$\frac{1}{\epsilon^2} \log \frac{1}{\delta} \log^2 \log \frac{1}{\delta} \log n$$

rows. Even with the diagonal sign flipping, this matrix still allows for $O(n \log n)$ embedding.

Naturally, this raises the question of whether we can get the "best of both worlds" for JL embeddings. That is, can we construct a JL matrix with $O(\frac{1}{\epsilon^2} \log(/\delta))$ rows which admits $O(n \log n)$ time embedding? In general, this is an open question, but if

$$m < \frac{\sqrt{n}}{\sqrt{\log n} \log \log \frac{1}{\delta}},$$

then yes, by simply composing a Fourier embedding with a Gaussian embedding.

References

[HR16] Ishay Haviv and Oded Regev. The restricted isometry property of subsampled Fourier matrices. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 288–297. SIAM, 2016.

[KW11] Felix Krahmer and Rachel Ward. New and improved Johnson-Lindenstrauss embeddings via the restricted isometry property. *SIAM Journal on Mathematical Analysis*, 43(3):1269–1281, 2011.

[RV06] Mark Rudelson and Roman Vershynin. Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements. In *2006 40th Annual Conference on Information Sciences and Systems*, pages 207–212. IEEE, 2006.