Lecture 3. Random Fourier measurements

1. Sampling from Fourier matrices

2. Law of Large Numbers and its operator-valued versions

3. Frames. Rudelson's Selection Theorem

Sampling from Fourier matrices

- Our measurements $=$ random frequencies of a signal f.
- The frequencies are values of the (discrete) Fourier transform (DFT).
- DFT is a linear operator in $\mathbb{C}^d$, given by a $d \times d$ DFT matrix

$$\Psi_{\omega,t} = \frac{1}{\sqrt{d}} \exp(-i2\pi\omega t/d), \qquad \omega, t \in \{0, \ldots, d-1\}.$$

- More generally, Ψ is any orthogonal matrix with entries $O(\frac{1}{\sqrt{d}})$.
- The discrete Fourier transform of $f \in \mathbb{C}^d$ is

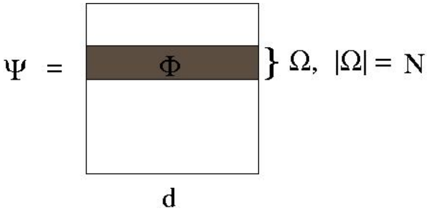
$$\widehat{f} = \Psi f.$$

Sampling from Fourier matrices

- Measurements of a signal f are N random frequencies

$$\{\widehat{f}(\omega),\ \omega \in \Omega\},$$

  where Ω is a random subset of $\{0, \dots, d-1\}$ of size N.
- Measurement matrix: $\Phi = N \times d$ minor of the DFT matrix Ψ with rows in the random subset Ω:



  Thus the measurements of f are given by Φf.
- Fourier Reconstruction Problem. When can we reconstruct any n-sparse signal f in $\mathbb{C}^d$ from the N Fourier measurements?

Reconstruction from Fourier measurements

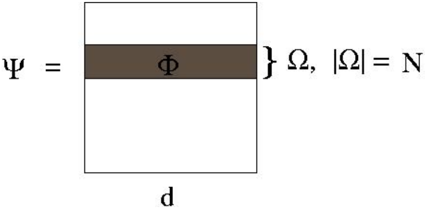
- Fourier Reconstruction Problem. Can we reconstruct an n-sparse signal f in $\mathbb{C}^d$ from the N Fourier measurements $\widehat{f^*}(\omega)$, $\omega \in \Omega$?
- Reconstruct using the convex optimization problem

$$\text{minimize } \|f^*\|_1 \text{ subject to } \widehat{f^*}(\omega) = \widehat{f}(\omega),\ \omega \in \Omega.$$

- How many measurements $N = N(n, d)$ are needed?
- Non-uniform result: For a given f, the random set of $N \sim n \log d$ frequencies is good with high probability: f is reconstructed correctly from these frequencies [Candes-Romberg-Tao]. Optimal.
- Uniform result: With high probability, the random set of $N \sim n \log^4 d$ frequencies is good for *every* f.
- Suffices to prove: a random minor of DFT is a restricted isometry.
- [Candes-Tao]: $N \sim n \log^6 d$; [Rudelson-V.]: $N \sim n \log^4 d$.
- Conjecture: $N \sim n \log d$.
  Improving beyond $N \sim n \log d \log \log d$ must be hard.
  (Related to Λ_1-conjecture; Bourgain-Talagrand).

Random sampling

- We sample N frequencies.
- Thus our measurement matrix Φ is a random sample (N rows) of the DFT matrix Ψ (total of d rows):



- We need to understand random sampling in matrices (operators).

Law of Large Numbers

- Fundamental principle of random sampling:
  Law of Large Numbers (LLN).
  It states that the empirical averages converge to true averages.
- Classical form: for reals rather than matrices.
  Let $X_1, \ldots, X_N$ be independent copies of a random variable X.
  Our goal: to estimate the mean $\mathbb{E}X$ using the sample $X_1, \ldots, X_N$.
- Approximate by the empirical mean

$$\frac{1}{N}\sum_{k=1}^{N} X_k \approx \mathbb{E}X.$$

- How good is this approximation?

Law of Large Numbers

Proposition (LLN)

$$E := \mathbb{E}\left|\frac{1}{N}\sum_{k=1}^{N} X_k - \mathbb{E}X\right| \lesssim \frac{\sigma(X)}{\sqrt{N}},$$

where $\sigma(X)$ is the standard deviation of X.

Proof.

Symmetrization: reducing general r.v.'s X_k to Bernoulli ε_k.
Let X_k' be independent copy of X_k. Then

$$E = \mathbb{E}\left|\frac{1}{N}\sum_{k=1}^{N} X_k - \frac{1}{N}\sum_{k=1}^{N} X_k'\right| = \mathbb{E}\left|\frac{1}{N}\sum_{k=1}^{N}(X_k - X_k')\right|.$$

$(X_k - X_k')$ is a symmetric random variable, thus distributed identically with $\varepsilon_k(X_k - X_k')$. Hence

$$E \lesssim \mathbb{E}\left|\frac{1}{N}\sum_{k=1}^{N}\varepsilon_k(X_k - X_k')\right| \le \frac{2}{N}\mathbb{E}\left|\sum_{k=1}^{N}\varepsilon_k X_k\right|.$$

We condition on (X_k); the randomness remains in the Bernoulli (ε_k).

Theorem (Khintchine's Inequality)

For reals (a_k) *and* $0 < p < \infty$,

$$A_p\Big(\sum_k |a_k|^2\Big)^{1/2} \le \Big(\mathbb{E}\Big|\sum_k \varepsilon_k a_k\Big|^p\Big)^{1/p} \le B_p\Big(\sum_k |a_k|^2\Big)^{1/2},$$

where $B_p = O(\sqrt{p})$. *The reverse inequality also holds.*

We condition on (X_k) and take expectation with respect to (ε_k).
By Khinchine's inequality with $p = 1$,

$$E \lesssim \frac{1}{N}\mathbb{E}\Big(\sum_k |X_k|^2\Big)^{1/2} \le \frac{1}{\sqrt{N}}\,(\mathbb{E}|X|^2)^{1/2} = \frac{\sigma(X)}{\sqrt{N}}. \qquad \square$$

Law of Large Numbers for operators

- We sample from matrices (linear operators) rather than reals. Need LLN for operators ("Non-commutative" version).
- Let $X_1, \ldots, X_N$ be independent copies of a random linear operator X on $\mathbb{R}^n$.
  Our goal: to learn the mean $\mathbb{E}X$ from the sample $X_1, \ldots, X_N$.
- Approximate by the empirical mean

$$\frac{1}{N}\sum_{k=1}^{N} X_k \approx \mathbb{E}X$$

  in the operator norm.
- Will the scalar argument carry over here? Symmetrization does. But is there a Khintchine's inequality for linear operators?
- Yes, but for the Schatten class norm rather than the operator norm.

Schatten class

- Schatten class C_p^n consists of all operators on $\mathbb{R}^n$ with the norm

$$\|A\|_{C_p^n} = \Big(\sum_{k=1}^{n} \lambda_j(A)^p \Big)^{1/p},$$

  where $\lambda_k(A)$ are the singular values (eigenvalues of $|A| = \sqrt{A^*A}$).
- $p = \infty$: operator norm.
  $p = 2$: Frobenius (Hilbert-Schmidt) norm.
- By Hölder's inequality, $\|A\| \leq \|A\|_{C_p}^n \leq n^{1/p}\|A\|$.
- Schatten norm is equivalent to the operator norm for $p = \log n$:

$$\|A\| \leq \|A\|_{C_p}^n \leq e\|A\|.$$

Non-commutative Khintchine's inequality

Theorem (Non-commutative Khintchine's [Lust-Picquard])

For self-adjoint operators (A_k) *on* $\mathbb{R}^n$ *and* $2 \le p < \infty$*,*

$$A_p \Big\| \Big(\sum_k A_k^2 \Big)^{1/2} \Big\|_{C_p^n} \le \Big(\mathbb{E} \Big\| \sum_k \varepsilon_k A_k \Big\|_{C_p^n}^p \Big)^{1/p} \le B_p \Big\| \Big(\sum_k A_k^2 \Big)^{1/2} \Big\|_{C_p^n},$$

where $B_p = O(\sqrt{p})$*. The reverse inequality also holds.*

- A version holds for $1 \le p \le 2$ [Lust-Picquard, Pisier].
- For the operator norm:
  use the equivalence of the norms (with $p = \log n$):

$$\mathbb{E} \Big\| \sum_k \varepsilon_k A_k \Big\| \lesssim \sqrt{\log n} \Big\| \Big(\sum_k A_k^2 \Big)^{1/2} \Big\|.$$

- Now we are ready to prove LLN for random operators:

Law of Large Numbers for operators

Theorem (LLN for operators)

Let $X_1, \ldots, X_N$ be independent copies of a random positive self-adjoint operator X on $\mathbb{R}^n$. Then

$$E := \mathbb{E}\Big\| \frac{1}{N} \sum_{k=1}^{N} X_k - \mathbb{E}X \Big\| \lesssim \sqrt{\log n} \cdot \frac{\sigma(X)}{\sqrt{N}},$$

where $\sigma(X) = \sqrt{\|X\|_\infty \|\mathbb{E}X\|}$.

Differences with the real-valued case:

- Factor $\sqrt{\log n}$ is needed and is optimal.
- Standard deviation $\sigma(X)$ replaced by a bound of $L^\infty \cdot L^1$ type. Typically larger.

Proof of LLN for random operators

Proof.

By the symmetrization and Khintchine's inequality,

$$E \lesssim \frac{1}{N}\mathbb{E}\Big\|\sum_{k=1}^{N} \varepsilon_k X_k\Big\| \lesssim \frac{\sqrt{\log n}}{N}\mathbb{E}\Big\|\Big(\sum_{k=1}^{n} X_k^2\Big)^{1/2}\Big\|.$$

Can' move the expectation inside the norm! Will relate the RHS to E:

$$\mathbb{E}\Big\|\Big(\sum_{k=1}^{n} X_k^2\Big)^{1/2}\Big\| = \mathbb{E}\Big\|\sum_{k=1}^{n} X_k^2\Big\|^{1/2} \leq \|X\|_\infty\Big(\mathbb{E}\Big\|\sum_{k=1}^{n} X_k\Big\|\Big)^{1/2}$$

$$\leq \sqrt{N}\|X\|_\infty\Big(\mathbb{E}\Big\|\frac{1}{N}\sum_{k=1}^{n} X_k\Big\|\Big)^{1/2} \leq \sqrt{N}\|X\|_\infty\big(E + \|\mathbb{E}X\|\big)^{1/2}.$$

Combining these two inequalities gives $E \lesssim \sqrt{\frac{\log n}{N}}\big(E + \|\mathbb{E}X\|\big)^{1/2}$.
Solving for E completes the proof. □

Frames

Definition

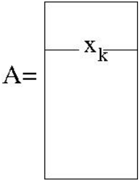
A system of vectors (x_k) in a Hilbert space is called a tight frame if it satisfies Parseval's identity:

$$\sum_k |\langle x_k, x \rangle|^2 = \|x\|^2 \qquad \text{for every vector } x.$$

- Examples:
  - Orthonormal bases
  - Orthonormal bases repeated 10 times and multiplied by $\frac{1}{\sqrt{10}}$
  - Orthogonal projections of orthonormal bases from a higher dimensional space (characterizes all frames)
- (x_k) is a tight frame $\Leftrightarrow \sum_k x_k \otimes x_k = I$.
  Recall: $x_k \otimes x_k$ is a rank-one linear operator that acts as $(x_k \otimes x_k)(x) = \langle x_k, x \rangle x_k$. In terms of matrices, $x_k \otimes x_k = x_k^T x_k$.
- ε-tight frames: the identities hold up to error ε.

Frames

- Duality for tight frames.
  The rows of a matrix A form a tight frame $\Leftrightarrow$ A is an isometry.



- Duality ε-tight frames.
  The rows of a matrix A form an ε-tight frame $\Leftrightarrow$ A is an ε-isometry:

$$(1-\varepsilon)\|x\| \le \|Ax\| \le (1+\varepsilon)\|x\| \qquad \text{for all } x.$$

Rudelson's Selection Theorem

- The number of frame elements must be at least the dimension n of the space.
- Conjecture. Every tight frame in dimension n has a subset of size $C(\varepsilon)n$, and which is an ε-tight frame (possibly after appropriate normalization).
- Rudelson's Selection Theorem: true for a *random* subset of size $C(\varepsilon)n \log n$.
- $\log n$ is needed for *random* subsets. (Ex.: repeated o.n.bases).
- Reduces the size of the frame to almost linear in the dimension, *independently of the original size* of the frame!

Rudelson's Selection Theorem

Theorem (Rudelson's Selection Theorem)

Let (x_k) be a tight frame in $\mathbb{C}^n$, and $\varepsilon > 0$. Choose a random N-element subset of (x_k) of by selecting each element independently with probability proportional to $\|x_k\|^2$. Then, with high probability, this (appropriately normalized) subset is an ε-tight frame of $\mathbb{C}^n$, provided

$$N \sim C(\varepsilon) n \log n.$$

- Normalization: $\sqrt{\frac{n}{N}} \cdot \frac{x_k}{\|x_k\|}$.

Proof of Rudelson's Selection Theorem

Proof.

Since (x_k) is a tight frame,

$$I = \sum_k x_k \otimes x_k. \qquad (*)$$

In particular, $\sum_k \|x_k\|^2 = n$ (take the trace).
We shall view $(*)$ as the expectation of a random operator.
With probabilities $p_k := \frac{\|x_k\|^2}{n}$ and normalization $\bar{x}_k = \frac{x_k}{\sqrt{p}}$, we rewrite

$$I = \sum_k p_k \bar{x}_k \otimes \bar{x}_k = \mathbb{E}X,$$

where X takes values $\bar{x}_k \otimes \bar{x}_k$ with probability p_k.

The probabilities and the normalization enforce even contributions of all values to the random operator X: $\|\bar{x}_k\| = \sqrt{n}$ for all k.

So, we have a random operator X, which takes values $\bar{x}_k \otimes \bar{x}_k$, and such that

$$\mathbb{E}X = I.$$

Apply LLN for random operators. First check that $\|X\|_\infty = \max_k \|\bar{x}_k\|^2 = n$, hence

$$\sigma(X) = \sqrt{\|X\|_\infty \|\mathbb{E}X\|} = \sqrt{n}.$$

Then LLN gives

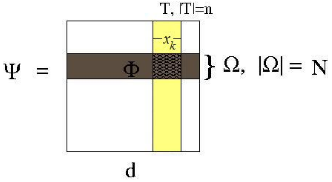
$$\mathbb{E}\left\| \frac{1}{N} \sum_{j=1}^{N} X_j - I \right\| \lesssim \sqrt{\log n} \cdot \frac{\sigma(X)}{\sqrt{N}} = \sqrt{\frac{n \log n}{N}} \leq \varepsilon, \qquad (*)$$

once $N \geq \varepsilon^{-2} n \log n$.

Fix a realization $X_j = \bar{x}_{k_j} \otimes \bar{x}_{k_j}$. By duality, $(*)$ means that the set (x_{k_j}) is an ε-tight frame (after an appropriate normalization). □

Application to Fourier measurements

- Our measurement matrix Φ is N random rows of the DFT matrix Ψ:



- Want to show that Φ is a restricted isometry: every minor Φ_T, $|T| = n$, is an ε-isometry ($\varepsilon = 0.2$).
- Fix T for now. By the frame duality, this condition is equivalent to the following: the rows of Φ_T form an ε-tight frame.
- These rows are sampled randomly from the rows of Ψ_T (amber matrix), which form a tight frame (again by the frame duality).
- By Rudelson's Selection Theorem, if $N \sim n \log n$ then the random sample is indeed an ε-tight frame. Restricted isometry proved.
- How to unfix T? Not easy. The probability for fixed T is insufficient to take the union bound over all T (there are $\binom{d}{n}$ of them).
- Have to handle all T at once. [Candes-Tao], [Rudelson-V.]