Suprema of Chaos Processes and the RIP of structured random matrices

Holger Rauhut
Hausdorff Center for Mathematics
& Institute for Numerical Simulation
University of Bonn

Phenomena in high dimensions in geometric analysis, random matrices, and computational geometry
Roscoff, June 25, 2012

Joint work with Felix Krahmer and Shahar Mendelson

- Compressive Sensing
- Partial Random Circulant Matrices (Random Convolutions)
- Suprema of Chaos Processes
- Time-Frequency Structured Random Matrices

Reconstruct $\mathbf{x} \in \mathbb{C}^N$ from its vector $\mathbf{y}$ of m measurements

$$\mathbf{y} = A\mathbf{x}, \qquad A \in \mathbb{C}^{m \times N},$$

when $m \ll N$, under additional knowledge that $\mathbf{x}$ is s-sparse,

$$\|\mathbf{x}\|_0 := \#\{\ell : x_\ell \neq 0\} \leq s.$$

Reconstruct $\mathbf{x} \in \mathbb{C}^N$ from its vector $\mathbf{y}$ of m measurements

$$\mathbf{y} = A\mathbf{x}, \qquad A \in \mathbb{C}^{m \times N},$$

when $m \ll N$, under additional knowledge that $\mathbf{x}$ is s-sparse,

$$\|\mathbf{x}\|_0 := \#\{\ell : x_\ell \neq 0\} \leq s.$$

More generally, $\mathbf{x}$ is compressible in the sense that its best s-term approximation error $\sigma_s(\mathbf{x})_p$ decreases quickly in s, where

$$\sigma_s(\mathbf{x})_p = \inf_{\mathbf{z}:\|\mathbf{z}\|_0 \leq s} \|\mathbf{x} - \mathbf{z}\|_p.$$

Reconstruct $\mathbf{x} \in \mathbb{C}^N$ from its vector $\mathbf{y}$ of m measurements

$$\mathbf{y} = A\mathbf{x}, \qquad A \in \mathbb{C}^{m \times N},$$

when $m \ll N$, under additional knowledge that $\mathbf{x}$ is s-sparse,

$$\|\mathbf{x}\|_0 := \#\{\ell : x_\ell \neq 0\} \leq s.$$

More generally, $\mathbf{x}$ is compressible in the sense that its best s-term approximation error $\sigma_s(\mathbf{x})_p$ decreases quickly in s, where

$$\sigma_s(\mathbf{x})_p = \inf_{\mathbf{z}:\|\mathbf{z}\|_0 \leq s} \|\mathbf{x} - \mathbf{z}\|_p.$$

Preferably we would like to have a fast algorithm that performs the reconstruction.

ℓ_0-minimization:

$$\min_{\mathbf{x}\in\mathbb{C}^N} \|\mathbf{x}\|_0 \quad \text{subject to} \quad A\mathbf{x} = \mathbf{y}.$$

ℓ_0-minimization:

$$\min_{\mathbf{x} \in \mathbb{C}^N} \|\mathbf{x}\|_0 \quad \text{subject to} \quad A\mathbf{x} = \mathbf{y}.$$

Problem: ℓ_0-minimization is NP hard!

ℓ_1 minimization:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 = \sum_{j=1}^{N} |x_j| \quad \text{subject to} \quad A\mathbf{x} = \mathbf{y}$$

hausdorff CENTER FOR MATHEMATICS

ℓ_1 minimization:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 = \sum_{j=1}^N |x_j| \quad \text{subject to} \quad A\mathbf{x} = \mathbf{y}$$

Convex relaxation of ℓ_0-minimization problem.

Efficient minimization methods available.

ℓ_1 minimization:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 = \sum_{j=1}^{N} |x_j| \quad \text{subject to} \quad A\mathbf{x} = \mathbf{y}$$

Convex relaxation of ℓ_0-minimization problem.

Efficient minimization methods available.

Alternatives:
Greedy Algorithms (Matching Pursuits)
Iterative Algorithms

Recovery Conditions for ℓ_1-minimization

- Null space property: necessary and sufficient for exact recovery of all s-sparse vectors via ℓ_1-minimization with A.

  $\|\mathbf{v}_S\|_1 \leq \rho \|\mathbf{v}_{S^c}\|_1$ for all $\mathbf{v} \in \ker A, S \subset [N], |S| = s,$ some $\rho < 1$.

  Implies also stability of reconstruction.
- Restricted isometry property (RIP): sufficient for exact (and stable) recovery of all s-sparse vectors via ℓ_1-minimization and other recovery algorithms (see next slide).
- Dual certificates: Sufficient (and sometimes necessary) conditions on $\mathbf{x}$ and A for exact (and stable) recovery via ℓ_1-minimization.
  Leads to nonuniform recovery guarantees (see later).

hausdorff CENTER FOR MATHEMATICS

Definition

The restricted isometry constant δ_s of a matrix $A \in \mathbb{C}^{m \times N}$ is defined as the smallest δ_s such that

$$(1 - \delta_s)\|\mathbf{x}\|_2^2 \leq \|A\mathbf{x}\|_2^2 \leq (1 + \delta_s)\|\mathbf{x}\|_2^2$$

for all s-sparse $\mathbf{x} \in \mathbb{C}^N$.

Requires that all s-column submatrices of A are well-conditioned.

RIP implies recovery by ℓ_1-minimization

Theorem (Candès, Romberg, Tao 2004 – Candès 2008 – Foucart, Lai 2009 – Foucart 2009 – Li, Mo 2011)

Assume that the restricted isometry constant of $A \in \mathbb{C}^{m \times N}$ satisfies

$$\delta_{2s} < 0.4931.$$

Then ℓ_1-minimization reconstructs every s-sparse vector $\mathbf{x} \in \mathbb{C}^N$ from $y = Ax$.

hausdorff CENTER FOR MATHEMATICS

Theorem (Candès, Romberg, Tao 2004 – Candès 2008 – Foucart, Lai 2009 – Foucart 2009 – Li, Mo 2011)

Let $A \in \mathbb{C}^{m \times N}$ *with* $\delta_{2s} < 0.4931$*.Let* $x \in \mathbb{C}^N$*, and assume that noisy data are observed,* $y = Ax + \eta$ *with* $\|\eta\|_2 \leq \sigma$*. Let* $x^\#$ *by a solution of*

$$\min_z \|z\|_1 \quad \text{such that} \quad \|Az - y\|_2 \leq \sigma.$$

Then

$$\|x - x^\#\|_2 \leq C \frac{\sigma_s(x)_1}{\sqrt{s}} + D\sigma$$

and

$$\|x - x^\#\|_1 \leq C\sigma_s(x)_1 + D\sqrt{s}\sigma$$

for constants $C, D > 0$*, that depend only on* δ_{2s}*.*

hausdorff CENTER FOR MATHEMATICS

Open problem: Give explicit matrices $A \in \mathbb{C}^{m \times N}$ with small $\delta_{2s} \leq 0.49$ for "large" s.

Open problem: Give explicit matrices $A \in \mathbb{C}^{m \times N}$ with small $\delta_{2s} \leq 0.49$ for "large" s.

Goal: $\delta_s \leq \delta$, if

$$m \geq C_\delta s \ln^\alpha(N),$$

for constants C_δ and α.

Deterministic matrices known, for which $m \geq C_{\delta,k} s^2$ suffices if $N \leq m^k$.

Small improvement by Bourgain et al. (2010): $m \geq C_\delta s^{2-\epsilon}$, $\epsilon > 0$, under additional assumptions on m, N.

hausdorff CENTER FOR MATHEMATICS

Open problem: Give explicit matrices $A \in \mathbb{C}^{m \times N}$ with small $\delta_{2s} \leq 0.49$ for "large" s.

Goal: $\delta_s \leq \delta$, if

$$m \geq C_\delta s \ln^\alpha(N),$$

for constants C_δ and α.

Deterministic matrices known, for which $m \geq C_{\delta,k} s^2$ suffices if $N \leq m^k$.

Small improvement by Bourgain et al. (2010): $m \geq C_\delta s^{2-\epsilon}$, $\epsilon > 0$, under additional assumptions on m, N.

Way out: consider random matrices.

hausdorff
CENTER FOR MATHEMATICS

RIP for Gaussian and Bernoulli matrices

Gaussian: entries of A independent $\mathcal{N}(0,1)$ random variables
Bernoulli : entries of A independent Bernoulli ± 1 distributed rv

Theorem

Let $A \in \mathbb{R}^{m \times N}$ be a Gaussian or Bernoulli random matrix and assume

$$m \geq C\delta^{-2}(s \ln(eN/s) + \ln(2\varepsilon^{-1}))$$

for a universal constant $C > 0$. Then with probability at least $1 - \varepsilon$ the restricted isometry constant of $\frac{1}{\sqrt{m}}A$ satisfies $\delta_s \leq \delta$.

Consequence: Recovery via ℓ_1-minimization with probability exceeding $1 - e^{-cm}$ provided

$$m \geq Cs \ln(eN/s).$$

Bound is optimal as follows from lower bound for Gelfand widths of ℓ_p-balls, $0 < p \leq 1$. (Gluskin, Garnaev 1984 — Foucart, Pajor, Rauhut, Ullrich 2010)

Why structure?

- Applications impose structure due to physical constraints, limited freedom to inject randomness.
- Fast matrix vector multiplies (FFT) in recovery algorithms, unstructured random matrices impracticable for large scale applications.

Random Partial Fourier Matrices

$F \in \mathbb{C}^{N \times N}$ discrete Fourier matrix, $F_{jk} = e^{2\pi i jk/N}$.
$\Omega \subset [N]$ random subset of cardinality m (row indices).
$R_\Omega : \mathbb{C}^N \to \mathbb{C}^\Omega$: restriction operator.
$A = \frac{1}{\sqrt{m}} R_\Omega F$: random partial Fourier matrix, application corresponds to random subsampling of the Fourier transform.

$F \in \mathbb{C}^{N \times N}$ discrete Fourier matrix, $F_{jk} = e^{2\pi ijk/N}$.
$\Omega \subset [N]$ random subset of cardinality m (row indices).
$R_\Omega : \mathbb{C}^N \to \mathbb{C}^\Omega$: restriction operator.
$A = \frac{1}{\sqrt{m}} R_\Omega F$: random partial Fourier matrix, application corresponds to random subsampling of the Fourier transform.

A satisfies RIP with high probability (Candés, Romberg, Tao 2006 — Rudelson, Vershynin 2008 — Rauhut 2009):
$\delta_s = \delta_s(A) \leq \delta$ with probability at least $1 - N^{-\ln^3 s}$ provided

$$m \geq C\delta^{-2} s \ln^3(s) \ln(N).$$

Generalization to random sampling in bounded orthonormal systems, including Legendre/Jacobi polynomials (Rauhut, Ward 2010).

Circulant matrix: For $\mathbf{b} = (b_0, b_1, \ldots, b_{N-1}) \in \mathbb{C}^N$ let $\Phi = \Phi(\mathbf{b}) \in \mathbb{C}^{N \times N}$ be the matrix with entries $\Phi_{i,j} = b_{j-i \mod N}$,

$$\Phi(\mathbf{b}) = \begin{pmatrix} b_0 & b_1 & \cdots & \cdots & b_{N-1} \\ b_{N-1} & b_0 & b_1 & \cdots & b_{N-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ b_1 & b_2 & \cdots & b_{N-1} & b_0 \end{pmatrix}.$$

Partial random circulant matrices

Let $\Theta \subset [N]$ arbitrary of cardinality m.
R_Θ: operator that restricts a vector $\mathbf{x} \in \mathbb{C}^N$ to its entries in Θ.

Restrict $\Phi(\mathbf{b})$ to the rows indexed by Θ:
Partial circulant matrix: $\Phi^\Theta(\mathbf{b}) = R_\Theta \Phi(\mathbf{b}) \in \mathbb{C}^{m \times N}$

Let $\Theta \subset [N]$ arbitrary of cardinality m.
R_Θ: operator that restricts a vector $\mathbf{x} \in \mathbb{C}^N$ to its entries in Θ.

Restrict $\Phi(\mathbf{b})$ to the rows indexed by Θ:
Partial circulant matrix: $\Phi^\Theta(\mathbf{b}) = R_\Theta \Phi(\mathbf{b}) \in \mathbb{C}^{m \times N}$

Convolution followed by subsampling:
$\mathbf{y} = R_\Theta \Phi(\mathbf{b})\mathbf{x} = R_\Theta(\mathbf{b} * \mathbf{x})$

Partial random circulant matrices

Let $\Theta \subset [N]$ arbitrary of cardinality m.
R_Θ: operator that restricts a vector $\mathbf{x} \in \mathbb{C}^N$ to its entries in Θ.

Restrict $\Phi(\mathbf{b})$ to the rows indexed by Θ:
Partial circulant matrix: $\Phi^\Theta(\mathbf{b}) = R_\Theta \Phi(\mathbf{b}) \in \mathbb{C}^{m \times N}$

Convolution followed by subsampling:
$\mathbf{y} = R_\Theta \Phi(\mathbf{b})\mathbf{x} = R_\Theta(\mathbf{b} * \mathbf{x})$
Matrix vector multiplication via the FFT!

Partial random circulant matrices

Let $\Theta \subset [N]$ arbitrary of cardinality m.
R_Θ: operator that restricts a vector $\mathbf{x} \in \mathbb{C}^N$ to its entries in Θ.

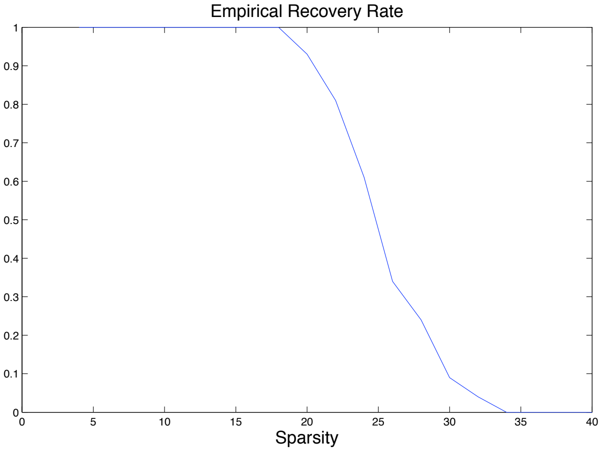
Restrict $\Phi(\mathbf{b})$ to the rows indexed by Θ:
Partial circulant matrix: $\Phi^\Theta(\mathbf{b}) = R_\Theta \Phi(\mathbf{b}) \in \mathbb{C}^{m \times N}$

Convolution followed by subsampling:
$\mathbf{y} = R_\Theta \Phi(\mathbf{b})\mathbf{x} = R_\Theta(\mathbf{b} * \mathbf{x})$
Matrix vector multiplication via the FFT!

We choose the vector $\mathbf{b} \in \mathbb{C}^N$ at random, in particular, as Rademacher sequence $\mathbf{b} = \epsilon$, that is, $\epsilon_\ell = \pm 1$.

Performance of $\Phi^\Theta(\epsilon)$ in compressive sensing?

Sparse recovery via ℓ_1-minimization with partial random circulant matrix $\Phi^{\Theta}(\epsilon) \in \mathbb{R}^{m \times N}$, $N = 500$, $m = 100$.



Nonuniform recovery result for circulant matrices

Theorem (Rauhut 2009)

Let $\Theta \subset [N]$ be an arbitrary (deterministic) set of cardinality m. Let $\mathbf{x} \in \mathbb{C}^N$ be s-sparse such that the signs of its non-zero entries form a Rademacher (or Steinhaus) sequence. Choose $\epsilon \in \mathbb{R}^N$ to be a Rademacher sequence. Let $\mathbf{y} = \Phi^{\Theta}(\epsilon)\mathbf{x} \in \mathbb{C}^m$. If

$$m \geq 57 s \ln^2(17N^2/\varepsilon)$$

then $\mathbf{x}$ can be recovered from $\mathbf{y}$ via ℓ_1-minimization with probability at least $1 - \varepsilon$.

RIP estimate for partial random circulant matrices

Theorem (Krahmer, Mendelson, Rauhut 2012)

Let $\Theta \subset [N]$ be an arbitrary (deterministic) set of cardinality m. Choose $\epsilon \in \mathbb{R}^N$ to be a Rademacher sequence. Assume that

$$m \geq C\delta^{-2} s \ln^2(s) \ln^2(N).$$

Then with probability at least $1 - N^{-\ln^2(s)\ln(N)}$ the restricted isometry constants of $\frac{1}{\sqrt{m}}\Phi^{\Theta}(\epsilon)$ satisfy $\delta_s \leq \delta$.

Generalizes to subgaussian generator.

Previous bounds:

Haupt, Bajwa, Raz (2008): $m \geq C_\delta s^2 \ln N$.

Rauhut, Romberg, Tropp (2010): $m \geq C_\delta s^{3/2} \ln^{3/2}(N)$.

Random sets Θ, Romberg (2009): $m \geq C_\delta s \ln^6 N$.

Consequence for Johnson-Lindenstrauss embeddings

Combination with result of F. Krahmer and R. Ward (2010) connecting the RIP and Johnson-Lindenstrauss embeddings yields:

Theorem

Fix $\eta, \delta \in (0,1)$, and consider a finite set $Q \subset \mathbb{R}^N$ of cardinality $|Q| = p$. Choose

$$m \geq C_\eta \delta^{-2} \log(p)(\log(\log p))^2(\log N)^2.$$

Let $\Phi \in \mathbb{C}^{m \times N}$ be a partial circulant matrix generated by a Rademacher vector. Furthermore, let $\epsilon \in \mathbb{R}^N$ be a Rademacher vector independent of Φ. Then with probability exceeding $1 - \eta$,

$$(1-\delta)\|\mathbf{x}\|_2^2 \leq \|\Phi D_\epsilon \mathbf{x}\|_2^2 \leq (1+\delta)\|\mathbf{x}\|_2^2 \quad \text{for all } \mathbf{x} \in Q.$$

Here D_ϵ denotes the diagonal matrix with diagonal ϵ.

Recall that δ_s is smallest constant such that

$$(1-\delta_s)\|\mathbf{x}\|_2^2 \leq \|A\mathbf{x}\|_2^2 \leq (1+\delta_s)\|\mathbf{x}\|_2^2 \quad \text{for all } s-\text{sparse } \mathbf{x}.$$

Equivalently, with $T_s = \{\mathbf{x} \in \mathbb{C}^N : \|\mathbf{x}\|_2 \leq 1, \|\mathbf{x}\|_0 \leq s\}$

$$\delta_s = \sup_{\mathbf{x} \in T_s} \left| \|A\mathbf{x}\|_2^2 - \|\mathbf{x}\|_2^2 \right|.$$

Recall that δ_s is smallest constant such that

$$(1-\delta_s)\|\mathbf{x}\|_2^2 \leq \|A\mathbf{x}\|_2^2 \leq (1+\delta_s)\|\mathbf{x}\|_2^2 \quad \text{for all } s-\text{sparse } \mathbf{x}.$$

Equivalently, with $T_s = \{\mathbf{x} \in \mathbb{C}^N : \|\mathbf{x}\|_2 \leq 1, \|\mathbf{x}\|_0 \leq s\}$

$$\delta_s = \sup_{\mathbf{x}\in T_s} \left| \|A\mathbf{x}\|_2^2 - \|\mathbf{x}\|_2^2 \right|.$$

For partial random circulant matrices,

$$A\mathbf{x} = \frac{1}{\sqrt{m}} R_\Omega(\epsilon * \mathbf{x}) = \frac{1}{\sqrt{m}} R_\Omega(\mathbf{x} * \epsilon) =: V_{\mathbf{x}}\epsilon$$

with appropriate $V_{\mathbf{x}} \in \mathbb{R}^{m\times N}$. Furthermore, $\mathbb{E}\|V_{\mathbf{x}}\epsilon\|_2^2 = \|\mathbf{x}\|_2^2$.
Therefore, δ_s is supremum of a chaos process,

$$\delta_s = \sup_{\mathbf{x}\in T_s} \left| \|V_{\mathbf{x}}\epsilon\|_2^2 - \mathbb{E}\|V_{\mathbf{x}}\epsilon\|_2^2 \right|.$$

Let $\mathcal{B}$ be a subset of a vector space with norm $\|\cdot\|$. Diameter

$$d_{\|\cdot\|}(\mathcal{B}) = \sup_{B \in \mathcal{B}} \|B\|.$$

Let $\mathcal{B}$ be a subset of a vector space with norm $\|\cdot\|$. Diameter

$$d_{\|\cdot\|}(\mathcal{B}) = \sup_{B\in\mathcal{B}} \|B\|.$$

A sequence of subsets $T_r \subset \mathcal{A}$, $r \in \mathbb{N}_0$, is called admissible if $|T_0| = 1$, $|T_r| \leq 2^{2^r}$, $r \geq 1$.
For $\alpha > 0$ define the γ_α-functional

$$\gamma_\alpha(\mathcal{B}, \|\cdot\|) = \inf \sup_{B\in\mathcal{B}} \sum_{r=0}^{\infty} 2^{r/\alpha} d(B, T_r), \quad d(B, T_r) = \inf_{B_r \in T_r} \|B - B_r\|,$$

where the infimum is over all admissible sequences (T_r).

Let $\mathcal{B}$ be a subset of a vector space with norm $\|\cdot\|$. Diameter

$$d_{\|\cdot\|}(\mathcal{B}) = \sup_{B \in \mathcal{B}} \|B\|.$$

A sequence of subsets $T_r \subset \mathcal{A}$, $r \in \mathbb{N}_0$, is called admissible if $|T_0| = 1$, $|T_r| \leq 2^{2^r}$, $r \geq 1$.
For $\alpha > 0$ define the γ_α-functional

$$\gamma_\alpha(\mathcal{B}, \|\cdot\|) = \inf \sup_{B \in \mathcal{B}} \sum_{r=0}^{\infty} 2^{r/\alpha} d(B, T_r), \quad d(B, T_r) = \inf_{B_r \in T_r} \|B - B_r\|,$$

where the infimum is over all admissible sequences (T_r).
Estimate by Dudley-type integral

$$\gamma_\alpha(\mathcal{B}, \|\cdot\|) \leq C \int_0^{d_{\|\cdot\|}(\mathcal{B})} \left(\log N(\mathcal{B}, \|\cdot\|, u)\right)^{1/\alpha} du,$$

where $N(\mathcal{B}, \|\cdot\|, u)$ denotes the smallest number of balls of radius u in the norm $\|\cdot\|$ required to cover $\mathcal{B}$.

Theorem (Krahmer, Mendelson, Rauhut 2012)

Let $\mathcal{B}$ be a symmetric set of matrices and ϵ a Rademacher vector. Then

$$\mathbb{E} \sup_{B \in \mathcal{B}} \left| \|B\epsilon\|_2^2 - \mathbb{E}\|B\epsilon\|_2^2 \right| \\ \leq C_1 \gamma_2(\mathcal{B}, \|\cdot\|_{2\to 2})^2 + C_2 d_{\|\cdot\|_F}(\mathcal{B}) \gamma_2(\mathcal{B}, \|\cdot\|_{2\to 2}).$$

Here, $\|\cdot\|_F = \sqrt{\mathrm{Tr}(B^* B)}$ denotes the Frobenius norm.

Tail bound

Theorem (Krahmer, Mendelson, Rauhut 2012)

Let $\mathcal{B} \subset \mathbb{C}^{m \times n}$ with $\mathcal{B} = -\mathcal{B}$ and ϵ be a Rademacher vector. Then

$$\mathbb{P}\left(\sup_{B \in \mathcal{B}} \left| \|B\varepsilon\|_2^2 - \mathbb{E}\|B\varepsilon\|_2^2 \right| \geq C_1 E + t\right) \leq 2\exp\left(-C_2 \min\left\{\frac{t^2}{V^2}, \frac{t}{U}\right\}\right),$$

where

$$E := \gamma_2(\mathcal{B}, \|\cdot\|_{2\to2})\left(\gamma_2(\mathcal{B}, \|\cdot\|_{2\to2}) + d_{\|\cdot\|_F}(\mathcal{B})\right)$$
$$V := d_{\|\cdot\|_{2\to2}}(\mathcal{B})(\gamma_2(\mathcal{B}, \|\cdot\|_{2\to2}) + d_{\|\cdot\|_F}(\mathcal{B})), \qquad U := d^2_{\|\cdot\|_{2\to2}}(\mathcal{B}).$$

Symmetry assumption $\mathcal{B} = -\mathcal{B}$ only made for convenience.
Generalizes to random vectors with independent mean-zero, variance 1, subgaussian entries.

Rewrite process as

$$\sup_{B\in\mathcal{B}} \left| \|B\epsilon\|_2^2 - \mathbb{E}\|B\epsilon\|_2^2 \right| = \sup_{B\in\mathcal{B}} \left| \sum_{j\neq k} \epsilon_j \epsilon_k (B^*B)_{j,k} \right|$$

Chaos process indexed by $\mathcal{D} = \{B^*B : B \in \mathcal{B}\}$.
General estimate (Talagrand, 1993)

$$\mathbb{E} \sup_{D\in\mathcal{D}} \Big| \sum_{j\neq k} \epsilon_j \epsilon_k D_{j,k} \Big| \leq C_1 \gamma_2(\mathcal{D}, \|\cdot\|_F) + C_2 \gamma_1(\mathcal{D}, \|\cdot\|_{2\to 2}).$$

Relation to previous estimate of Talagrand

Rewrite process as

$$\sup_{B \in \mathcal{B}} \left| \|B\epsilon\|_2^2 - \mathbb{E}\|B\epsilon\|_2^2 \right| = \sup_{B \in \mathcal{B}} \left| \sum_{j \neq k} \epsilon_j \epsilon_k (B^*B)_{j,k} \right|$$

Chaos process indexed by $\mathcal{D} = \{B^*B : B \in \mathcal{B}\}$.
General estimate (Talagrand, 1993)

$$\mathbb{E} \sup_{D \in \mathcal{D}} \left| \sum_{j \neq k} \epsilon_j \epsilon_k D_{j,k} \right| \leq C_1 \gamma_2(\mathcal{D}, \|\cdot\|_F) + C_2 \gamma_1(\mathcal{D}, \|\cdot\|_{2\to 2}).$$

This bound was used in the previous RIP estimate due to Rauhut, Romberg, Tropp (2010). The appearance of the γ_1-functional results in the exponent $3/2$.

Bound $\gamma_2(\mathcal{B}, \|\cdot\|_{2\to 2})$ for $\mathcal{B} = \{V_{\mathbf{x}}, \mathbf{x} \in T_s\}$.

Relate to the analysis of random partial Fourier matrices via the convolution theorem, and bound Dudley-type integral using covering number estimates due to Rudelson and Vershynin (2008).

Proof sketch for Bound of Chaos Processes

Decoupling (with ϵ' a Rademacher vector independent of ϵ)

$$\mathbb{E} C_{\mathcal{B}} := \mathbb{E} \sup_{B \in \mathcal{B}} \left| \|B\epsilon\|_2^2 - \mathbb{E}\|B\epsilon\|_2^2 \right| \leq 4\mathbb{E} \sup_{B \in \mathcal{B}} \left| \langle B\epsilon, B\epsilon' \rangle \right|$$

Proof sketch for Bound of Chaos Processes

Decoupling (with ϵ' a Rademacher vector independent of ϵ)

$$\mathbb{E} C_{\mathcal{B}} := \mathbb{E} \sup_{B \in \mathcal{B}} \left| \|B\epsilon\|_2^2 - \mathbb{E}\|B\epsilon\|_2^2 \right| \leq 4\mathbb{E} \sup_{B \in \mathcal{B}} \left| \langle B\epsilon, B\epsilon' \rangle \right|$$

Chaining: admissible sequence (T_r),
$\pi_r B = \arg\min_{A \in T_r} \|A - B\|_{2\to 2}$,

$$\sup_{A \in \mathcal{B}} \left| \langle B\epsilon, B\epsilon' \rangle \right| \leq \sum_{r=0}^{\infty} |\langle (\pi_{r+1}B - \pi_r B)\epsilon, \pi_{r+1}B\epsilon' \rangle|$$
$$+ \sum_{r=0}^{\infty} |\langle \pi_r B\epsilon, (\pi_{r+1}B - \pi_r B)\epsilon' \rangle| + |\langle \pi_0 B\epsilon, \pi_0 B\epsilon' \rangle|$$

Proof sketch for Bound of Chaos Processes

Decoupling (with ϵ' a Rademacher vector independent of ϵ)

$$\mathbb{E}C_{\mathcal{B}} := \mathbb{E}\sup_{B\in\mathcal{B}}\left|\|B\epsilon\|_2^2 - \mathbb{E}\|B\epsilon\|_2^2\right| \leq 4\mathbb{E}\sup_{B\in\mathcal{B}}\left|\langle B\epsilon, B\epsilon'\rangle\right|$$

Chaining: admissible sequence (T_r),
$\pi_r B = \arg\min_{A\in T_r}\|A - B\|_{2\to 2}$,

$$\sup_{A\in\mathcal{B}}\left|\langle B\epsilon, B\epsilon'\rangle\right| \leq \sum_{r=0}^{\infty}|\langle(\pi_{r+1}B - \pi_r B)\epsilon, \pi_{r+1}B\epsilon'\rangle| + \sum_{r=0}^{\infty}|\langle\pi_r B\epsilon, (\pi_{r+1}B - \pi_r B)\epsilon'\rangle| + |\langle\pi_0 B\epsilon, \pi_0 B\epsilon'\rangle|$$

Condition on ϵ' in the first term. Note that

$$|\langle(\pi_{r+1}B - \pi_r B)\epsilon, \pi_{r+1}B\epsilon'\rangle| = |\langle\epsilon, (\pi_{r+1}B - \pi_r B)^*(\pi_{r+1}B)\epsilon'\rangle|,$$
$$\|(\pi_{r+1}B - \pi_r B)^*(\pi_{r+1}B)\epsilon'\|_2 \leq \|\pi_{r+1}B - \pi_r B\|_{2\to 2}\|\pi_r B\epsilon'\|_2$$
$$\leq \|\pi_{r+1}B - \pi_r B\|_{2\to 2}N_{\mathcal{B}},$$

where $N_{\mathcal{B}} = \sup_{B\in\mathcal{B}}\|B\epsilon'\|_2$.

Proof sketch for Bound of Chaos Processes

Decoupling (with ϵ' a Rademacher vector independent of ϵ)

$$\mathbb{E} C_{\mathcal{B}} := \mathbb{E} \sup_{B\in\mathcal{B}} \left| \|B\epsilon\|_2^2 - \mathbb{E}\|B\epsilon\|_2^2 \right| \leq 4\mathbb{E} \sup_{B\in\mathcal{B}} \left| \langle B\epsilon, B\epsilon' \rangle \right|$$

Chaining: admissible sequence (T_r),
$\pi_r B = \arg\min_{A\in T_r} \|A - B\|_{2\to 2}$,

$$\sup_{A\in\mathcal{B}} \left| \langle B\epsilon, B\epsilon' \rangle \right| \leq \sum_{r=0}^{\infty} |\langle (\pi_{r+1}B - \pi_r B)\epsilon, \pi_{r+1}B\epsilon' \rangle| + \sum_{r=0}^{\infty} |\langle \pi_r B\epsilon, (\pi_{r+1}B - \pi_r B)\epsilon' \rangle| + |\langle \pi_0 B\epsilon, \pi_0 B\epsilon' \rangle|$$

Condition on ϵ' in the first term. Note that

$$|\langle (\pi_{r+1}B - \pi_r B)\epsilon, \pi_{r+1}B\epsilon' \rangle| = |\langle \epsilon, (\pi_{r+1}B - \pi_r B)^*(\pi_{r+1}B)\epsilon' \rangle|,$$

$$\|(\pi_{r+1}B - \pi_r B)^*(\pi_{r+1}B)\epsilon'\|_2 \leq \|\pi_{r+1}B - \pi_r B\|_{2\to 2} \|\pi_r B\epsilon'\|_2$$

$$\leq \|\pi_{r+1}B - \pi_r B\|_{2\to 2} N_{\mathcal{B}},$$

where $N_{\mathcal{B}} = \sup_{B\in\mathcal{B}} \|B\epsilon'\|_2$

Proof Ingredients II

With Hoeffding's inequality, a union bound over all elements of $T_r \times T_{r+1}$ and all levels r, and integration we obtain

$$\mathbb{E} C_{\mathcal{B}} \lesssim \gamma_2(\mathcal{B}, \|\cdot\|_{2\to 2}) \mathbb{E} N_{\mathcal{B}} + d_F(\mathcal{B}) d_{2\to 2}(\mathcal{B}).$$

Furthermore,

$$\begin{aligned}(\mathbb{E} N_{\mathcal{B}})^2 &\leq \mathbb{E} \sup_{B\in\mathcal{B}} \|B\epsilon\|_2^2 \leq \mathbb{E} \sup_{B\in\mathcal{B}} \left| \|B\epsilon\|_2^2 - \mathbb{E}\|B\epsilon\|_2^2 \right| + d_F^2(\mathcal{B}) \\ &= \mathbb{E} C_{\mathcal{B}} + d_F^2(\mathcal{B}) \lesssim \gamma_2(\mathcal{B}, \|\cdot\|_{2\to 2}) \mathbb{E} N_{\mathcal{B}} + d_F^2(\mathcal{B}).\end{aligned}$$

Theorefore,

$$\mathbb{E} N_{\mathcal{B}} \lesssim \gamma_2(\mathcal{B}, \|\cdot\|_{2\to 2}) + d_F(\mathcal{B})$$

and

$$\mathbb{E} C_{\mathcal{B}} \lesssim \gamma_2^2(\mathcal{B}, \|\cdot\|_{2\to 2}) + d_F(\mathcal{B}) \gamma_2(\mathcal{B}, \|\cdot\|_{2\to 2}) + d_F(\mathcal{B}) d_{2\to 2}(\mathcal{B}).$$

With Hoeffding's inequality, a union bound over all elements of $T_r \times T_{r+1}$ and all levels r, and integration we obtain

$$\mathbb{E} C_{\mathcal{B}} \lesssim \gamma_2(\mathcal{B}, \|\cdot\|_{2\to 2}) \mathbb{E} N_{\mathcal{B}} + d_F(\mathcal{B}) d_{2\to 2}(\mathcal{B}).$$

Furthermore,

$$(\mathbb{E} N_{\mathcal{B}})^2 \leq \mathbb{E} \sup_{B \in \mathcal{B}} \|B\epsilon\|_2^2 \leq \mathbb{E} \sup_{B \in \mathcal{B}} \left| \|B\epsilon\|_2^2 - \mathbb{E}\|B\epsilon\|_2^2 \right| + d_F^2(\mathcal{B})$$
$$= \mathbb{E} C_{\mathcal{B}} + d_F^2(\mathcal{B}) \lesssim \gamma_2(\mathcal{B}, \|\cdot\|_{2\to 2}) \mathbb{E} N_{\mathcal{B}} + d_F^2(\mathcal{B}).$$

Theorefore,

$$\mathbb{E} N_{\mathcal{B}} \lesssim \gamma_2(\mathcal{B}, \|\cdot\|_{2\to 2}) + d_F(\mathcal{B})$$

and

$$\mathbb{E} C_{\mathcal{B}} \lesssim \gamma_2^2(\mathcal{B}, \|\cdot\|_{2\to 2}) + d_F(\mathcal{B}) \gamma_2(\mathcal{B}, \|\cdot\|_{2\to 2}) + d_F(\mathcal{B}) d_{2\to 2}(\mathcal{B}).$$

Tail bound follows from moment estimate, for $p \geq 1$,

$$\left(\mathbb{E} C_{\mathcal{B}}^p\right)^{1/p} \lesssim \gamma_2(\mathcal{B}, \|\cdot\|_{2\to 2})\left(\gamma_2(\mathcal{B}, \|\cdot\|_{2\to 2}) + d_F(\mathcal{B})\right) + \sqrt{p} d_{2\to 2}(\mathcal{B})\left(\gamma_2(\mathcal{B}, \|\cdot\|_{2\to 2}) + d_F(\mathcal{B})\right) + p d_{2\to 2}^2(\mathcal{B}).$$

Gabor Systems in Finite Dimension

Translation and Modulation on $\mathbb{C}^m$

$$(T^k g)_j = g_{(j-k) \mod m} \quad \text{and} \quad (M^\ell g)_j = e^{2\pi i \ell j/m} g_j.$$

Time-frequency shifts

$$\pi(\lambda) = M^\ell T^k, \quad \lambda = (k, \ell) \in \{0, \ldots, m-1\}^2.$$

For $g \in \mathbb{C}^m$ define Gabor synthesis matrix ($\omega = e^{2\pi i/m}$)

$$\Psi_g = (\pi(\lambda)g)_{\lambda \in \{0,\ldots,m-1\}^2}$$

$$= \left(\begin{array}{cccc|ccc|cc} g_0 & g_{m-1} & \cdots & g_1 & g_0 & \cdots & g_1 & \cdots & g_1 \\ g_1 & g_0 & \cdots & g_2 & \omega g_1 & \cdots & \omega g_2 & \cdots & \omega^{m-1} g_2 \\ g_2 & g_1 & \cdots & g_3 & \omega^2 g_2 & \cdots & \omega^2 g_3 & \cdots & \omega^{2(m-1)} g_3 \\ g_3 & g_2 & \cdots & g_4 & \omega^3 g_3 & \cdots & \omega^3 g_4 & \cdots & \omega^{3(m-1)} g_4 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & & \vdots \\ g_{m-1} & g_{m-2} & \cdots & g_0 & \omega^{m-1} g_{m-1} & \cdots & \omega^{m-1} g_0 & \cdots & \omega^{(m-1)^2} g_0 \end{array} \right).$$

Performance of $\Psi_g \in \mathbb{C}^{m \times m^2}$ in compressive sensing?
Choice of g?

Translation and Modulation on $\mathbb{C}^m$

$$(T^k g)_j = g_{(j-k) \mod m} \quad \text{and} \quad (M^\ell g)_j = e^{2\pi i \ell j/m} g_j.$$

Time-frequency shifts

$$\pi(\lambda) = M^\ell T^k, \quad \lambda = (k,\ell) \in \{0,\dots,m-1\}^2.$$

For $g \in \mathbb{C}^m$ define Gabor synthesis matrix ($\omega = e^{2\pi i/m}$)

$$\Psi_g = (\pi(\lambda)g)_{\lambda \in \{0,\dots,m-1\}^2}$$

$$= \left(\begin{array}{cccc|ccc|cc}
g_0 & g_{m-1} & \cdots & g_1 & g_0 & \cdots & g_1 & \cdots & g_1 \\
g_1 & g_0 & \cdots & g_2 & \omega g_1 & \cdots & \omega g_2 & \cdots & \omega^{m-1} g_2 \\
g_2 & g_1 & \cdots & g_3 & \omega^2 g_2 & \cdots & \omega^2 g_3 & \cdots & \omega^{2(m-1)} g_3 \\
g_3 & g_2 & \cdots & g_4 & \omega^3 g_3 & \cdots & \omega^3 g_4 & \cdots & \omega^{3(m-1)} g_4 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & & \vdots \\
g_{m-1} & g_{m-2} & \cdots & g_0 & \omega^{m-1} g_{m-1} & \cdots & \omega^{m-1} g_0 & \cdots & \omega^{(m-1)^2} g_0
\end{array}\right).$$

Performance of $\Psi_g \in \mathbb{C}^{m \times m^2}$ in compressive sensing?
Choice of g?

Applications in: Radar, sonar, wireless communications.
Simplified discrete model for radar (Herman, Strohmer 2009)

Emitted signal: $g \in \mathbb{C}^m$.
Objects scatters g and radar device receives the contribution

$$x_\lambda \pi(\lambda) g = x_{k,\ell} M^\ell T^k g.$$

T^k corresponds to delay, i.e., distance of object
M^ℓ corresponds to Doppler shift, i.e., speed of the object
$x_{k,\ell}$ reflectivity of object

Received signal is superposition of contribution of all scatteres:

$$y = \sum_{\lambda \in \Lambda} x_\lambda \pi(\lambda) g = \Psi_g x.$$

Usually few scatterers so that $x \in \mathbb{C}^{m^2}$ can be assumed sparse.

Random choice of $g \in \mathbb{C}^m$,

$$g = \frac{1}{\sqrt{m}}\epsilon$$

where ϵ is either a Rademacher or Steinhaus vector,

- Steinhaus: entries ϵ_ℓ are independent and uniformly distributed on torus $\mathbb{T} = \{z \in \mathbb{C}, |z| = 1\}$.

Then Ψ_g is structured random matrix!

Random choice of $g \in \mathbb{C}^m$,

$$g = \frac{1}{\sqrt{m}}\epsilon$$

where ϵ is either a Rademacher or Steinhaus vector,

- Steinhaus: entries ϵ_ℓ are independent and uniformly distributed on torus $\mathbb{T} = \{z \in \mathbb{C}, |z| = 1\}$.

Then Ψ_g is structured random matrix!

Theorem (Pfander, Rauhut 2007)

Let $x \in \mathbb{C}^N$, $N = m^2$, be s-sparse. Choose $g \in \mathbb{C}^m$ as a random Steinhaus sequence. Assume that

$$m \geq Cs \log(N/\epsilon).$$

Then with probability at least $1 - \epsilon$ ℓ_1-minimization recovers x from $y = \Psi_g x$.

Theorem (Krahmer, Mendelson, Rauhut 2012)

Let $\Psi_g \in \mathbb{C}^{m \times N}$, $N = m^2$, be generated by a random draw of a Steinhaus or Rademacher vector. If, for $\delta \in (0,1)$,

$$m \geq C\delta^{-2} s \log^2 s \log^2 N,$$

then with probability at least $1 - N^{-\log N \log^2 s}$ the restricted isometry constant of Ψ_g satisfies $\delta_s \leq \delta$.

Improves previous estimate (Pfander, Rauhut, Tropp – 2011), which roughly requires

$$m \geq C_\delta s^{3/2} \log^3 N.$$

Congratulations Alain!