On Compressive Sensing of Sparse Covariance Matrices Using Deterministic Sensing Matrices

Alihan Kaplan[1] Volker Pohl[1] Dae Gwan Lee[2]

[1]Institute of Theoretical Inform. Technology
Technical University of Munich
80333 München, Germany

[2]Faculty of Mathematics and Geography
Catholic University of Eichstätt-Ingolstadt
85071 Eichstätt, Germany

Overview

1. Motivation: Covariance matrix sketching
2. Some Simulations
3. Statistical Restricted Isometry property (StRIP)
4. Probabilistic Recovery Guarantee of StRIP
5. Construction Example

Motivation: Covariance matrix sketching

Given:

- $\mathbf{x} \in \mathbb{C}^N$, a vector of N independent zero-mean random variables
- covariance matrix $\mathbf{X} = \mathbb{E}\left[\mathbf{x}\mathbf{x}^*\right]$, sparse in most applications
- m linear measurements $\mathbf{y} = \mathbf{A}\mathbf{x}$, with measurement matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$

Determine $\mathbf{X}$ from $\mathbf{Y}$

$$\mathbf{Y} = \mathbb{E}\left[\mathbf{y}\mathbf{y}^*\right] = \mathbf{A}\mathbb{E}\left[\mathbf{x}\mathbf{x}^*\right]\mathbf{A}^* = \mathbf{A}\mathbf{X}\mathbf{A}^*$$

using vectorization...

$$\tilde{\mathbf{y}} = \left(\bar{\mathbf{A}} \otimes \mathbf{A}\right)\tilde{\mathbf{x}}$$

with $\tilde{\mathbf{y}} = \mathbf{vec}\left\{\mathbf{Y}\right\}$ and $\tilde{\mathbf{x}} = \mathbf{vec}\left\{\mathbf{X}\right\}$

$\Rightarrow$ Compressive Sensing setting!

Recap of relevant results in Compressive Sensing

Problem setting: $\mathbf{y} = \mathbf{A}\mathbf{x}$ with $\mathbf{A} \in \mathbb{C}^{m \times N}$, where $m \ll N$

Definition (Restricted Isometry Property)

$\mathbf{A} \in \mathbb{C}^{m \times N}$ is said to fulfill the k-th restricted isometry property (abbrv. RIP) with the restricted isometry constant δ_k (abbrv. RIC) if

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

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

Theorem

If $\mathbf{A}$ fulfills the $2k$-th RIP with RIC

$$\delta_{2k} < \frac{1}{3}$$

then every k-sparse $\mathbf{x}$ can be recovered uniquely by the ℓ_1-minimization (convex).

Covariance matrix sketching & Compressive Sensing

Covariance matrix sketching as compressive sensing problem:

$$\min \|\tilde{x}\|_0 \quad \text{subject to} \quad \tilde{\mathbf{y}} = \left(\bar{\mathbf{A}} \otimes \mathbf{A}\right)\tilde{\mathbf{x}}$$

with $\tilde{\mathbf{y}} = \mathbf{vec}\left\{\mathbf{Y}\right\}$ and $\tilde{\mathbf{x}} = \mathbf{vec}\left\{\mathbf{X}\right\}$

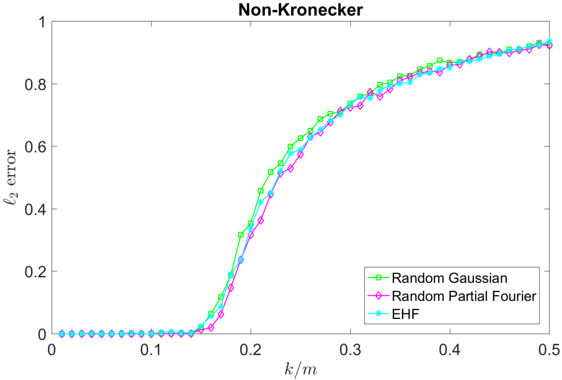
Question: Are there "good" (deterministic) matrices for compressive sensing with Kronecker structure?

$\rightarrow$ convex relaxation: ℓ_1-minimization instead of ℓ_0-minimization

• Result on RIC by Duarte and Baraniuk: $\delta_k\left(\overline{\mathbf{A}} \otimes \mathbf{A}\right) \geq \delta_k\left(\mathbf{A}\right)$

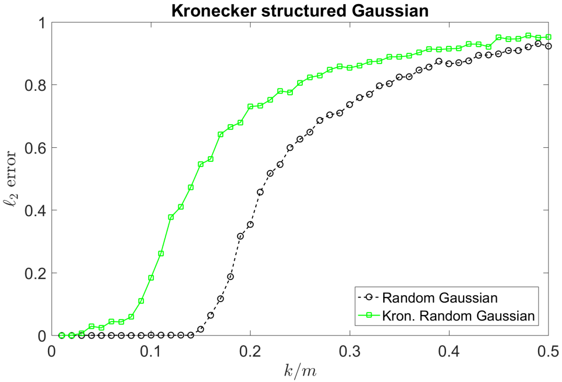
$\Rightarrow$ For fixed sparsity, RIC of the Kronecker structured matrix is lower bounded by the RIC of the non-Kronecker matrix despite having quadratically more measurements.

Simulations: non-Kronecker structured matrices



Random Gaussian: Each entry of the sensing matrix is a Gaussian random variable.
Random Partial Fourier: Rows of the DFT matrix are chosen at random to form the sensing matrix.
EHF (Equiangular harmonic frames): The sensing matrix is a "carefully" chosen minor of the DFT matrix.

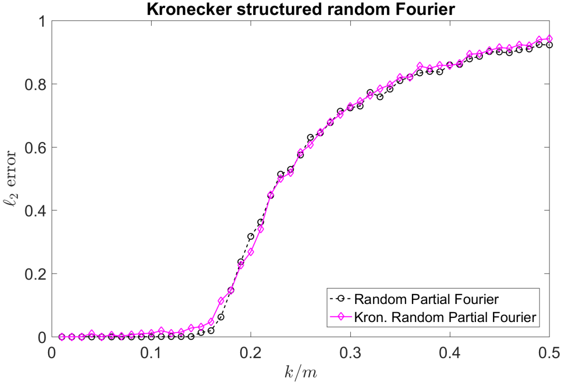
Simulation: Kronecker structured Gaussian



Kronecker structured Gaussian matrices seem to be "bad" for CS.

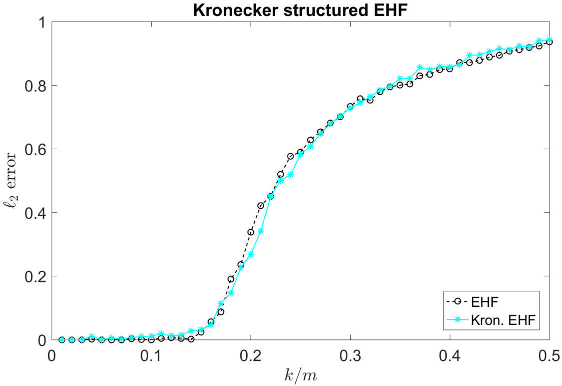
From previous slide: $\delta_k\left(\mathbf{A}\right) \leq \delta_k\left(\overline{\mathbf{A}} \otimes \mathbf{A}\right) \leq 2\delta_k\left(\mathbf{A}\right) + \delta_k\left(\mathbf{A}\right)^2$

Simulation: Kronecker structured partial Fourier



In contrast to Gaussian: random Fourier & Kronecker structured random Fourier matrices perform similarly.

Simulation: Kronecker structured EHF



Same observation as in the slide before: EHF & Kronecker structured EHF perform similarly.

$\rightarrow$ An attempt of explanation based on *Statistical Isometry Property* (Def. by Calderbank, Howard, Jafarpour, 2010).

Standard CS versus StRIP Approach

Standard CS

- random sensing matrices
- deterministic vectors $\mathbf{x}$
- recovery guarantee for all k-sparse vectors $\mathbf{x}$
- recovery guarantee with high probability for random $\mathbf{A}$

$\Rightarrow$ randomness in the choice of the sensing matrix $\mathbf{A}$

StRIP

- deterministic sensing matrix
- stochastic data vectors $\mathbf{x}$
- recovery guarantee with high probability for random choice of $\mathbf{x}$
- recovery guarantee for deterministic choice of $\mathbf{A}$

$\Rightarrow$ randomness in the choice of the data vector $\mathbf{x}$

Statistical Restricted Isometry Property (StRIP) - 1

Following definition of StRIP is by Calderbank, Howard and Jafarpour (2010).

Definition (StRIP & UStRIP)

- $\mathbf{A} = \frac{1}{\sqrt{m}}\Phi \in \mathbb{C}^{m \times N}$ has (k, δ, ϵ) - StRIP if

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

holds with probability $1-\epsilon$ for a random k-sparse vectors $\mathbf{x}$ (uniformly distributed over all k-sparse vectors).

- $\mathbf{A}$ is (k, δ, ϵ)-uniqueness-guaranteed StRIP (UStRIP) if

$$\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{z} \iff \mathbf{z} = \mathbf{x}\,, \quad \forall k\text{-sparse } \mathbf{z}$$

satisfied with probability $1-\epsilon$.

Statistical Restricted Isometry Property (StRIP) - 2

Definition (η-StRIP)

$\mathbf{A} = \frac{1}{\sqrt{m}}\Phi \in \mathbb{C}^{m \times N}$ with all entries of Φ having absolute value 1, is η-StRIP if St1 - St3 holds.

St1:
- rows of Φ are orthogonal
- sum of all elements in a row is equal to zero

St2:
- columns of Φ form a multiplicative group under pointwise multiplication

St3:
- $\exists \eta > 0$ s.t. $\left|\sum_{l=1}^{m} \phi_j\left[l\right]\right|^2 \leq m^{2-\eta} \quad \forall l$ apart from the identity column

Probabilistic recovery guarantee of StRIP

Theorem (Calderbank, Howard, Jafarpour)

Let $\mathbf{A} = \frac{1}{\sqrt{m}}\Phi \in \mathbb{C}^{m \times N}$ *be an* η*-StRIP matrix with* $\eta > 1/2$*. If* $k < 1 + (N-1)\delta$ *and*

$$m \geq \frac{(k \log N)}{\delta^2}$$

for some constant $c > 0$ *then* $\mathbf{A}$ *is* $(k, \delta, 2\epsilon)$*-UStRIP with* $\epsilon = 2\exp\left(-\left(\delta - \frac{k-1}{N-1}\right)^2 \frac{m^\eta}{8k}\right)$.

• Theorem connects structure of a deterministic CS matrix with probabilistic recovery guarantee.

• Easy applicability of StRIP on deterministic matrices (checking RIP is NP-Hard).

• linear scaling of the number of measurements m with the sparsity k

• **Main idea:** use this theorem for Kronecker structured matrices.

η-StRIP for Kronecker structured matrices

Theorem

Assume $\mathbf{A} \in \mathbb{C}^{n \times N}$ is η_A-StRIP and $\mathbf{B} \in \mathbb{C}^{m \times M}$ is η_B-StRIP, then the following holds.

(a) *$\overline{\mathbf{A}}$ is $\eta_{\overline{\mathbf{A}}}$-StRIP with $\eta_{\overline{\mathbf{A}}} = \eta_{\mathbf{A}}$.*

(b) *The matrix $\mathbf{C} = \mathbf{A} \otimes \mathbf{B} \in \mathbb{C}^{nm \times NM}$ is $\eta_{\mathbf{C}}$-StRIP with*

$$\eta_{\mathbf{C}} = \begin{cases} \eta_{\mathbf{A}} \frac{\ln(n)}{\ln(nm)} & \textit{if } n^{\eta_{\mathbf{A}}} \leq m^{\eta_{\mathbf{B}}} \\ \\ \eta_{\mathbf{B}} \frac{\ln(m)}{\ln(nm)} & \textit{if } n^{\eta_{\mathbf{A}}} > m^{\eta_{\mathbf{B}}} \end{cases}$$

Corollary

If $\mathbf{A} \in \mathbb{C}^{m \times N}$ is η-StRIP, then the matrix $\overline{\mathbf{A}} \otimes \mathbf{A} \in \mathbb{C}^{m^2 \times N^2}$ is $(\eta/2)$-StRIP.

• linear scaling of the number of measurements m^2 with the sparsity
$\rightarrow m^2 \geq ck \log N$

• search for deterministic matrices $\mathbf{A}$ s.t. $\eta_{\mathbf{A}} > 1$.

η-StRIP matrices with $\eta > 1$

- Coherence μ of a matrix $\mathbf{A}$ is defined by

$$\mu(\mathbf{A}) = \max_{i,j \text{ s.t. } i \neq j} |\langle \mathbf{a}_i, \mathbf{a}_j \rangle|$$

- using the Welch bound, for a matrix $\mathbf{A}$ fulfilling the η-StRIP definition follows:

$$\sqrt{\frac{N-m}{m(N-1)}} \leq \mu(\mathbf{A}) \leq \frac{1}{\sqrt{m^\eta}}$$

$\Rightarrow$ upper bound on η:

$$\eta \leq 1 + \ln\left(\frac{N-1}{N-m}\right) \frac{1}{\ln(m)}$$

$\Rightarrow$ any η-StRIP matrix coming very close to the Welch bound

Example (Equiangular Harmonic Frames)

EHFs are partial Fourier matrices with $\mu = \sqrt{\frac{N-m}{m(N-1)}}$

construction is based on difference sets

$\{0,1,3\}$ forms a $(7,3,1)$ diff. set in $\mathbb{Z}_7$ (integers modulo 7)

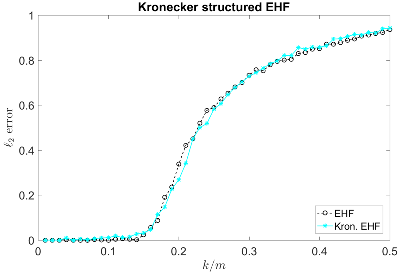
$$\mathbf{DFT}_7 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & \omega^1 & \omega^2 & \omega^3 & \omega^4 & \omega^5 & \omega^6 \\ 1 & \omega^2 & \omega^4 & \omega^6 & \omega^1 & \omega^3 & \omega^5 \\ 1 & \omega^3 & \omega^6 & \omega^2 & \omega^5 & \omega^1 & \omega^4 \\ 1 & \omega^4 & \omega^1 & \omega^5 & \omega^2 & \omega^6 & \omega^3 \\ 1 & \omega^5 & \omega^3 & \omega^1 & \omega^6 & \omega^4 & \omega^2 \\ 1 & \omega^6 & \omega^5 & \omega^4 & \omega^3 & \omega^2 & \omega^1 \end{bmatrix}$$

$$\mathbf{EHF}_3 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & \omega^1 & \omega^2 & \omega^3 & \omega^4 & \omega^5 & \omega^6 \\ 1 & \omega^3 & \omega^6 & \omega^2 & \omega^5 & \omega^1 & \omega^4 \end{bmatrix}$$

Summary

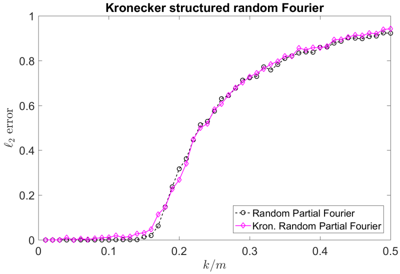
- Investigation of Kronecker structured sensing matrices for compressive sensing
- Used the StRIP approach for Kronecker structured sensing matrices
- Proved statistical recovery guaranties where the number of measurements scales linearly with the sparsity

Deterministic Matrices



StRIP approach

Random Partial Fourier



no explanation yet