Two added structures in sparse recovery: nonnegativity and disjointedness

Simon Foucart
University of Georgia

Semester Program on "High-Dimensional Approximation"
ICERM
7 October 2014

Part I:
Nonnegative Sparse Recovery

(joint work with D. Koslicki)

Motivation from Metagenomics

Motivation from Metagenomics

- $\mathbf{x} \in \mathbb{R}^N$ ($N = 273,727$): concentrations of known bacteria in a given environmental sample.

Motivation from Metagenomics

- $\mathbf{x} \in \mathbb{R}^N$ ($N = 273,727$): concentrations of known bacteria in a given environmental sample. Sparsity assumption is realistic.

Motivation from Metagenomics

- $\mathbf{x} \in \mathbb{R}^N$ ($N = 273,727$): concentrations of known bacteria in a given environmental sample. Sparsity assumption is realistic. Note also that $\mathbf{x} \geq \mathbf{0}$ and $\sum_j x_j = 1$.

Motivation from Metagenomics

- $\mathbf{x} \in \mathbb{R}^N$ ($N = 273,727$): concentrations of known bacteria in a given environmental sample. Sparsity assumption is realistic. Note also that $\mathbf{x} \geq \mathbf{0}$ and $\sum_j x_j = 1$.
- $\mathbf{y} \in \mathbb{R}^m$ ($m = 4^6 = 4,096$): frequencies of length-6 subwords (in 16S rRNA gene reads or in whole-genome shotgun reads)

Motivation from Metagenomics

- $\mathbf{x} \in \mathbb{R}^N$ ($N = 273,727$): concentrations of known bacteria in a given environmental sample. Sparsity assumption is realistic. Note also that $\mathbf{x} \geq \mathbf{0}$ and $\sum_j x_j = 1$.
- $\mathbf{y} \in \mathbb{R}^m$ ($m = 4^6 = 4,096$): frequencies of length-6 subwords (in 16S rRNA gene reads or in whole-genome shotgun reads)
- $\mathbf{A} \in \mathbb{R}^{m \times N}$: frequencies of length-6 subwords in all known (i.e., sequenced) bacteria.

Motivation from Metagenomics

- $\mathbf{x} \in \mathbb{R}^N$ ($N = 273,727$): concentrations of known bacteria in a given environmental sample. Sparsity assumption is realistic. Note also that $\mathbf{x} \geq \mathbf{0}$ and $\sum_j x_j = 1$.
- $\mathbf{y} \in \mathbb{R}^m$ ($m = 4^6 = 4,096$): frequencies of length-6 subwords (in 16S rRNA gene reads or in whole-genome shotgun reads)
- $\mathbf{A} \in \mathbb{R}^{m \times N}$: frequencies of length-6 subwords in all known (i.e., sequenced) bacteria. It is a frequency matrix, that is,

Motivation from Metagenomics

- $\mathbf{x} \in \mathbb{R}^N$ $(N = 273,727)$: concentrations of known bacteria in a given environmental sample. Sparsity assumption is realistic. Note also that $\mathbf{x} \geq \mathbf{0}$ and $\sum_j x_j = 1$.
- $\mathbf{y} \in \mathbb{R}^m$ $(m = 4^6 = 4,096)$: frequencies of length-6 subwords (in 16S rRNA gene reads or in whole-genome shotgun reads)
- $\mathbf{A} \in \mathbb{R}^{m \times N}$: frequencies of length-6 subwords in all known (i.e., sequenced) bacteria. It is a frequency matrix, that is,

$$A_{i,j} \geq 0 \qquad \text{and} \qquad \sum_{i=1}^m A_{i,j} = 1.$$

Motivation from Metagenomics

- $\mathbf{x} \in \mathbb{R}^N$ ($N = 273,727$): concentrations of known bacteria in a given environmental sample. Sparsity assumption is realistic. Note also that $\mathbf{x} \geq \mathbf{0}$ and $\sum_j x_j = 1$.
- $\mathbf{y} \in \mathbb{R}^m$ ($m = 4^6 = 4,096$): frequencies of length-6 subwords (in 16S rRNA gene reads or in whole-genome shotgun reads)
- $\mathbf{A} \in \mathbb{R}^{m \times N}$: frequencies of length-6 subwords in all known (i.e., sequenced) bacteria. It is a frequency matrix, that is,

$$A_{i,j} \geq 0 \qquad \text{and} \qquad \sum_{i=1}^m A_{i,j} = 1.$$

- **Quikr** improves on traditional read-by-read methods, especially in terms of speed.

Motivation from Metagenomics

- $\mathbf{x} \in \mathbb{R}^N$ ($N = 273,727$): concentrations of known bacteria in a given environmental sample. Sparsity assumption is realistic. Note also that $\mathbf{x} \geq \mathbf{0}$ and $\sum_j x_j = 1$.
- $\mathbf{y} \in \mathbb{R}^m$ ($m = 4^6 = 4,096$): frequencies of length-6 subwords (in 16S rRNA gene reads or in whole-genome shotgun reads)
- $\mathbf{A} \in \mathbb{R}^{m \times N}$: frequencies of length-6 subwords in all known (i.e., sequenced) bacteria. It is a frequency matrix, that is,

$$A_{i,j} \geq 0 \qquad \text{and} \qquad \sum_{i=1}^m A_{i,j} = 1.$$

- **Quikr** improves on traditional read-by-read methods, especially in terms of speed.
- Codes available at
  `sourceforge.net/projects/quikr/`
  `sourceforge.net/projects/wgsquikr/`

Exact Measurements

Exact Measurements

Let $\mathbf{x} \in \mathbb{R}^N$ be a nonnegative vector with support S.

Exact Measurements

Let $\mathbf{x} \in \mathbb{R}^N$ be a nonnegative vector with support S.

- $\mathbf{x}$ is the unique minimizer of $\|\mathbf{z}\|_1$ s.to $\mathbf{A}\mathbf{z} = \mathbf{y}$ iff

$$\text{(BP)} \qquad \text{for all } \mathbf{v} \in \ker \mathbf{A} \setminus \{\mathbf{0}\}, \quad \left| \sum_{j \in S} v_j \right| < \sum_{\ell \in \overline{S}} |v_\ell|.$$

Exact Measurements

Let $\mathbf{x} \in \mathbb{R}^N$ be a nonnegative vector with support S.

- $\mathbf{x}$ is the unique minimizer of $\|\mathbf{z}\|_1$ s.to $\mathbf{Az} = \mathbf{y}$ iff

$$\text{(BP)} \qquad \text{for all } \mathbf{v} \in \ker \mathbf{A} \setminus \{\mathbf{0}\}, \quad \left| \sum_{j \in S} v_j \right| < \sum_{\ell \in \overline{S}} |v_\ell|.$$

- $\mathbf{x}$ is the unique minimizer of $\|\mathbf{z}\|_1$ s.to $\mathbf{Az} = \mathbf{y}$ and $\mathbf{z} \geq \mathbf{0}$ iff

$$\text{(NNBP)} \quad \text{for all } \mathbf{v} \in \ker \mathbf{A} \setminus \{\mathbf{0}\}, \quad \mathbf{v}_{\overline{S}} \geq \mathbf{0} \Rightarrow \sum_{i=1}^N v_i > 0.$$

Exact Measurements

Let $\mathbf{x} \in \mathbb{R}^N$ be a nonnegative vector with support S.

- $\mathbf{x}$ is the unique minimizer of $\|\mathbf{z}\|_1$ s.to $\mathbf{Az} = \mathbf{y}$ iff

(BP) for all $\mathbf{v} \in \ker \mathbf{A} \setminus \{\mathbf{0}\}$, $\left| \sum_{j \in S} v_j \right| < \sum_{\ell \in \overline{S}} |v_\ell|$.

- $\mathbf{x}$ is the unique minimizer of $\|\mathbf{z}\|_1$ s.to $\mathbf{Az} = \mathbf{y}$ and $\mathbf{z} \geq \mathbf{0}$ iff

(NNBP) for all $\mathbf{v} \in \ker \mathbf{A} \setminus \{\mathbf{0}\}$, $\mathbf{v}_{\overline{S}} \geq \mathbf{0} \Rightarrow \sum_{i=1}^N v_i > 0$.

- $\mathbf{x}$ is the unique $\mathbf{z} \geq \mathbf{0}$ s.to $\mathbf{Az} = \mathbf{y}$ iff

(F) for all $\mathbf{v} \in \ker \mathbf{A} \setminus \{\mathbf{0}\}$, $\mathbf{v}_{\overline{S}} \geq \mathbf{0}$ is impossible.

Exact Measurements

Let $\mathbf{x} \in \mathbb{R}^N$ be a nonnegative vector with support S.

- $\mathbf{x}$ is the unique minimizer of $\|\mathbf{z}\|_1$ s.to $\mathbf{A}\mathbf{z} = \mathbf{y}$ iff

  (BP) for all $\mathbf{v} \in \ker \mathbf{A} \setminus \{\mathbf{0}\}$, $\left|\sum_{j \in S} v_j\right| < \sum_{\ell \in \overline{S}} |v_\ell|$.

- $\mathbf{x}$ is the unique minimizer of $\|\mathbf{z}\|_1$ s.to $\mathbf{A}\mathbf{z} = \mathbf{y}$ and $\mathbf{z} \geq \mathbf{0}$ iff

  (NNBP) for all $\mathbf{v} \in \ker \mathbf{A} \setminus \{\mathbf{0}\}$, $\mathbf{v}_{\overline{S}} \geq \mathbf{0} \Rightarrow \sum_{i=1}^N v_i > 0$.

- $\mathbf{x}$ is the unique $\mathbf{z} \geq \mathbf{0}$ s.to $\mathbf{A}\mathbf{z} = \mathbf{y}$ iff

  (F) for all $\mathbf{v} \in \ker \mathbf{A} \setminus \{\mathbf{0}\}$, $\mathbf{v}_{\overline{S}} \geq \mathbf{0}$ is impossible.

In general, (F)$\Rightarrow$(NNBP) and (BP)$\Rightarrow$(NNBP).

Exact Measurements

Let $\mathbf{x} \in \mathbb{R}^N$ be a nonnegative vector with support S.

- $\mathbf{x}$ is the unique minimizer of $\|\mathbf{z}\|_1$ s.to $\mathbf{Az} = \mathbf{y}$ iff

(BP) for all $\mathbf{v} \in \ker \mathbf{A} \setminus \{\mathbf{0}\}$, $\left|\sum_{j \in S} v_j\right| < \sum_{\ell \in \overline{S}} |v_\ell|$.

- $\mathbf{x}$ is the unique minimizer of $\|\mathbf{z}\|_1$ s.to $\mathbf{Az} = \mathbf{y}$ and $\mathbf{z} \geq \mathbf{0}$ iff

(NNBP) for all $\mathbf{v} \in \ker \mathbf{A} \setminus \{\mathbf{0}\}$, $\mathbf{v}_{\overline{S}} \geq \mathbf{0} \Rightarrow \sum_{i=1}^N v_i > 0$.

- $\mathbf{x}$ is the unique $\mathbf{z} \geq \mathbf{0}$ s.to $\mathbf{Az} = \mathbf{y}$ iff

(F) for all $\mathbf{v} \in \ker \mathbf{A} \setminus \{\mathbf{0}\}$, $\mathbf{v}_{\overline{S}} \geq \mathbf{0}$ is impossible.

In general, (F)$\Rightarrow$(NNBP) and (BP)$\Rightarrow$(NNBP). If $\mathbf{1} \in \mathrm{im}(\mathbf{A}^\top)$ (e.g. if $\mathbf{A}$ is a frequency matrix), then (NNBP)$\Rightarrow$(F)$\Rightarrow$(BP).

Exact Measurements

Let $\mathbf{x} \in \mathbb{R}^N$ be a nonnegative vector with support S.

- $\mathbf{x}$ is the unique minimizer of $\|\mathbf{z}\|_1$ s.to $\mathbf{A}\mathbf{z} = \mathbf{y}$ iff

$$\text{(BP)} \qquad \text{for all } \mathbf{v} \in \ker \mathbf{A} \setminus \{\mathbf{0}\}, \quad \Big| \sum_{j \in S} v_j \Big| < \sum_{\ell \in \overline{S}} |v_\ell|.$$

- $\mathbf{x}$ is the unique minimizer of $\|\mathbf{z}\|_1$ s.to $\mathbf{A}\mathbf{z} = \mathbf{y}$ and $\mathbf{z} \geq \mathbf{0}$ iff

$$\text{(NNBP)} \quad \text{for all } \mathbf{v} \in \ker \mathbf{A} \setminus \{\mathbf{0}\}, \quad \mathbf{v}_{\overline{S}} \geq \mathbf{0} \Rightarrow \sum_{i=1}^N v_i > 0.$$

- $\mathbf{x}$ is the unique $\mathbf{z} \geq \mathbf{0}$ s.to $\mathbf{A}\mathbf{z} = \mathbf{y}$ iff

$$\text{(F)} \qquad \text{for all } \mathbf{v} \in \ker \mathbf{A} \setminus \{\mathbf{0}\}, \quad \mathbf{v}_{\overline{S}} \geq \mathbf{0} \text{ is impossible.}$$

In general, (F)$\Rightarrow$(NNBP) and (BP)$\Rightarrow$(NNBP). If $\mathbf{1} \in \mathrm{im}(\mathbf{A}^\top)$ (e.g. if $\mathbf{A}$ is a frequency matrix), then (NNBP)$\Rightarrow$(F)$\Rightarrow$(BP).

Morale: ℓ_1-minimization not suited for nonnegative sparse recovery.

Nonnegative Least Squares

- To solve the feasibility problem, one may consider

$$\underset{\mathbf{z}\in\mathbb{R}^N}{\text{minimize}}\ \|\mathbf{y}-\mathbf{A}\mathbf{z}\|_2^2 \qquad \text{subject to} \quad \mathbf{z}\geq\mathbf{0}.$$

Nonnegative Least Squares

- To solve the feasibility problem, one may consider

$$\underset{\mathbf{z} \in \mathbb{R}^N}{\text{minimize}} \; \|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2^2 \qquad \text{subject to} \quad \mathbf{z} \geq \mathbf{0}.$$

- MATLAB's **lsqnonneg** implements [Lawson–Hanson 74].

Nonnegative Least Squares

- To solve the feasibility problem, one may consider

$$\underset{\mathbf{z}\in\mathbb{R}^N}{\text{minimize}} \; \|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2^2 \qquad \text{subject to} \quad \mathbf{z} \geq \mathbf{0}.$$

- MATLAB's **lsqnonneg** implements [Lawson–Hanson 74].
- This algorithm iterates the scheme

$$S^{n+1} = S^n \cup \left\{ j^{n+1} = \operatorname{argmax}_j \big(\mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^n)\big)_j \right\},$$
$$\mathbf{x}^{n+1} = \operatorname{argmin} \left\{ \|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2, \; \operatorname{supp}(\mathbf{z}) \subseteq S^{n+1} \right\},$$

Nonnegative Least Squares

- To solve the feasibility problem, one may consider

$$\underset{\mathbf{z}\in\mathbb{R}^N}{\text{minimize}}\ \|\mathbf{y}-\mathbf{A}\mathbf{z}\|_2^2 \qquad \text{subject to} \quad \mathbf{z} \geq \mathbf{0}.$$

- MATLAB's **lsqnonneg** implements [Lawson–Hanson 74].
- This algorithm iterates the scheme

$$S^{n+1} = S^n \cup \{j^{n+1} = \text{argmax}_j \big(\mathbf{A}^*(\mathbf{y}-\mathbf{A}\mathbf{x}^n)\big)_j\},$$
$$\mathbf{x}^{n+1} = \text{argmin}\,\{\|\mathbf{y}-\mathbf{A}\mathbf{z}\|_2,\ \text{supp}(\mathbf{z}) \subseteq S^{n+1}\},$$

and inner loop to make sure that $\mathbf{x}^{n+1} \geq \mathbf{0}$.

Nonnegative Least Squares

- To solve the feasibility problem, one may consider

$$\underset{\mathbf{z} \in \mathbb{R}^N}{\text{minimize}} \, \|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2^2 \qquad \text{subject to} \quad \mathbf{z} \geq \mathbf{0}.$$

- MATLAB's **lsqnonneg** implements [Lawson–Hanson 74].
- This algorithm iterates the scheme

$$S^{n+1} = S^n \cup \left\{ j^{n+1} = \operatorname{argmax}_j \left(\mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^n) \right)_j \right\},$$
$$\mathbf{x}^{n+1} = \operatorname{argmin} \left\{ \|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2, \; \operatorname{supp}(\mathbf{z}) \subseteq S^{n+1} \right\},$$
$$\text{and inner loop to make sure that } \mathbf{x}^{n+1} \geq \mathbf{0}.$$

- Connection with OMP explains suitability for sparse recovery.

Inaccurate Measurements

Inaccurate Measurements

- When $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ with $\mathbf{e} \neq \mathbf{0}$, a classical strategy consists in solving the ℓ_1-regularization

$$\underset{\mathbf{z}\in\mathbb{R}^N}{\text{minimize}}\ \|\mathbf{z}\|_1 + \nu\|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2^2 \qquad \text{subject to} \quad \mathbf{z} \geq \mathbf{0}.$$

Inaccurate Measurements

- When $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ with $\mathbf{e} \neq \mathbf{0}$, a classical strategy consists in solving the ℓ_1-regularization

$$\underset{\mathbf{z} \in \mathbb{R}^N}{\text{minimize}} \ \|\mathbf{z}\|_1 + \nu \|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2^2 \qquad \text{subject to} \quad \mathbf{z} \geq \mathbf{0}.$$

- We prefer the ℓ_1-squared regularization

$$\underset{\mathbf{z} \in \mathbb{R}^N}{\text{minimize}} \ \|\mathbf{z}\|_1^2 + \lambda^2 \|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2^2 \qquad \text{subject to} \quad \mathbf{z} \geq \mathbf{0},$$

Inaccurate Measurements

- When $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ with $\mathbf{e} \neq \mathbf{0}$, a classical strategy consists in solving the ℓ_1-regularization

$$\underset{\mathbf{z}\in\mathbb{R}^N}{\text{minimize}} \ \|\mathbf{z}\|_1 + \nu\|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2^2 \qquad \text{subject to} \quad \mathbf{z} \geq \mathbf{0}.$$

- We prefer the ℓ_1-squared regularization

$$\underset{\mathbf{z}\in\mathbb{R}^N}{\text{minimize}} \ \|\mathbf{z}\|_1^2 + \lambda^2\|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2^2 \qquad \text{subject to} \quad \mathbf{z} \geq \mathbf{0},$$

because it is recast as the Nonnegative Least Squares problem

$$\underset{\mathbf{z}\in\mathbb{R}^N}{\text{minimize}} \ \|\widetilde{\mathbf{y}} - \widetilde{\mathbf{A}}\mathbf{z}\|_2^2 \qquad \text{subject to} \quad \mathbf{z} \geq \mathbf{0},$$

where $\widetilde{\mathbf{A}} = \left[\begin{array}{ccc} 1 & \cdots & 1 \\ \hline & \lambda\mathbf{A} & \end{array}\right]$ and $\widetilde{\mathbf{y}} = \left[\begin{array}{c} 0 \\ \hline \lambda\mathbf{y} \end{array}\right]$.

Inaccurate Measurements

- When $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ with $\mathbf{e} \neq \mathbf{0}$, a classical strategy consists in solving the ℓ_1-regularization

$$\underset{\mathbf{z} \in \mathbb{R}^N}{\text{minimize}}\ \|\mathbf{z}\|_1 + \nu \|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2^2 \qquad \text{subject to} \quad \mathbf{z} \geq \mathbf{0}.$$

- We prefer the ℓ_1-squared regularization

$$\underset{\mathbf{z} \in \mathbb{R}^N}{\text{minimize}}\ \|\mathbf{z}\|_1^2 + \lambda^2 \|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2^2 \qquad \text{subject to} \quad \mathbf{z} \geq \mathbf{0},$$

  because it is recast as the Nonnegative Least Squares problem

$$\underset{\mathbf{z} \in \mathbb{R}^N}{\text{minimize}}\ \|\widetilde{\mathbf{y}} - \widetilde{\mathbf{A}}\mathbf{z}\|_2^2 \qquad \text{subject to} \quad \mathbf{z} \geq \mathbf{0},$$

  where $\widetilde{\mathbf{A}} = \left[\begin{array}{ccc} 1 & \cdots & 1 \\ \hline & \lambda \mathbf{A} & \end{array}\right]$ and $\widetilde{\mathbf{y}} = \left[\begin{array}{c} 0 \\ \hline \lambda \mathbf{y} \end{array}\right]$.

- For frequency matrices, as $\lambda \to \infty$, the minimizer $\mathbf{x}_\lambda$ tends to the minimizer of $\|\mathbf{z}\|_1$ subject to $\mathbf{A}\mathbf{z} = \mathbf{y}$ and $\mathbf{z} \geq \mathbf{0}$.

Extension: Sparse Recovery via NNLS

Extension: Sparse Recovery via NNLS

- Decompose vectors $\mathbf{z} \in \mathbb{R}^N$ as $\mathbf{z} = \mathbf{z}_+ - \mathbf{z}_-$ with $\mathbf{z}_+, \mathbf{z}_- \in \mathbb{R}_+^N$.

Extension: Sparse Recovery via NNLS

- Decompose vectors $\mathbf{z} \in \mathbb{R}^N$ as $\mathbf{z} = \mathbf{z}_+ - \mathbf{z}_-$ with $\mathbf{z}_+, \mathbf{z}_- \in \mathbb{R}^N_+$.
- The ℓ_1-squared regularization

(REG) $$\text{minimize} \ \ \|\mathbf{z}\|_1^2 + \lambda^2 \|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2^2$$

is recast as the Nonnegative Least Squares problem

$$\text{minimize} \ \ \|\widetilde{\mathbf{y}} - \widetilde{\mathbf{A}}\widetilde{\mathbf{z}}\|_2^2 \quad \text{subject to } \widetilde{\mathbf{z}} \geq \mathbf{0},$$

where $\widetilde{\mathbf{y}} = \left[\begin{array}{c} 0 \\ \hline \lambda \mathbf{y} \end{array}\right], \widetilde{\mathbf{A}} = \left[\begin{array}{ccc|ccc} 1 & \cdots & 1 & 1 & \cdots & 1 \\ \hline & \lambda \mathbf{A} & & & -\lambda \mathbf{A} & \end{array}\right], \widetilde{\mathbf{z}} = \left[\begin{array}{c} \mathbf{z}_+ \\ \hline \mathbf{z}_- \end{array}\right].$

Extension: Sparse Recovery via NNLS

- Decompose vectors $\mathbf{z} \in \mathbb{R}^N$ as $\mathbf{z} = \mathbf{z}_+ - \mathbf{z}_-$ with $\mathbf{z}_+, \mathbf{z}_- \in \mathbb{R}^N_+$.
- The ℓ_1-squared regularization

$$\text{(REG)} \qquad \text{minimize } \|\mathbf{z}\|_1^2 + \lambda^2 \|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2^2$$

is recast as the Nonnegative Least Squares problem

$$\text{minimize } \|\widetilde{\mathbf{y}} - \widetilde{\mathbf{A}}\widetilde{\mathbf{z}}\|_2^2 \quad \text{subject to } \widetilde{\mathbf{z}} \geq \mathbf{0},$$

where $\widetilde{\mathbf{y}} = \left[\begin{array}{c} 0 \\ \hline \lambda \mathbf{y} \end{array}\right], \widetilde{\mathbf{A}} = \left[\begin{array}{ccc|ccc} 1 & \cdots & 1 & 1 & \cdots & 1 \\ \hline & \lambda \mathbf{A} & & & -\lambda \mathbf{A} & \end{array}\right], \widetilde{\mathbf{z}} = \left[\begin{array}{c} \mathbf{z}_+ \\ \hline \mathbf{z}_- \end{array}\right].$

- For Gaussian matrices (RNSP and QP hold), the solutions $\mathbf{x}_\lambda$ of (REG) with $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ obey, for all $\mathbf{x} \in \mathbb{R}^N$ and $\mathbf{e} \in \mathbb{R}^m$,

$$\|\mathbf{x} - \mathbf{x}_\lambda\|_1 \leq C\, \sigma_s(\mathbf{x})_1 + D\sqrt{s}\, \|\mathbf{e}\|_2 + \frac{E\, s}{\lambda^2} \|\mathbf{x}\|_1.$$

Part II:
Disjointed Sparse Recovery

(joint work with M. Minner and T. Needham)

Motivation from Radar

Motivation from Radar

- $\mathbf{x} \in \mathbb{R}^N$: positions of airplanes relative to a discretized grid.

Motivation from Radar

- $\mathbf{x} \in \mathbb{R}^N$: positions of airplanes relative to a discretized grid.
- Few airplanes that are not too close to one another: sparsity and disjointedness.

Motivation from Radar

- $\mathbf{x} \in \mathbb{R}^N$: positions of airplanes relative to a discretized grid.
- Few airplanes that are not too close to one another: sparsity and disjointedness.
- Disjointedness is also relevant to model neural spike trains. [Hedge–Duarte–Cevher 09]

Motivation from Radar

- $\mathbf{x} \in \mathbb{R}^N$: positions of airplanes relative to a discretized grid.
- Few airplanes that are not too close to one another: sparsity and disjointedness.
- Disjointedness is also relevant to model neural spike trains. [Hedge–Duarte–Cevher 09]
- We say that $\mathbf{x} \in \mathbb{R}^N$ is s-sparse and d-disjointed if

Motivation from Radar

- $\mathbf{x} \in \mathbb{R}^N$: positions of airplanes relative to a discretized grid.
- Few airplanes that are not too close to one another: sparsity and disjointedness.
- Disjointedness is also relevant to model neural spike trains. [Hedge–Duarte–Cevher 09]
- We say that $\mathbf{x} \in \mathbb{R}^N$ is s-sparse and d-disjointed if
  - $\mathbf{x}$ has no more than s nonzero entries,

Motivation from Radar

- $\mathbf{x} \in \mathbb{R}^N$: positions of airplanes relative to a discretized grid.
- Few airplanes that are not too close to one another: sparsity and disjointedness.
- Disjointedness is also relevant to model neural spike trains. [Hedge–Duarte–Cevher 09]
- We say that $\mathbf{x} \in \mathbb{R}^N$ is s-sparse and d-disjointed if
  - $\mathbf{x}$ has no more than s nonzero entries,
  - there are $\geq d$ zero entries between two nonzero entries.

Resolution of the Fundamental Question

Resolution of the Fundamental Question

- The minimal number of linear measurements for the recovery of all s-sparse vectors is

$$m_{\mathrm{spa}} \asymp s \ln\left(e\frac{N}{s}\right).$$

Resolution of the Fundamental Question

- The minimal number of linear measurements for the recovery of all s-sparse vectors is
$$m_{\mathrm{spa}} \asymp s \ln\left(e\frac{N}{s}\right).$$
- The minimal number of linear measurements for the recovery of all d-disjointed vectors is [Candès–Fernandez-Granda 14]
$$m_{\mathrm{dis}} \asymp \frac{N}{d}.$$

Resolution of the Fundamental Question

- The minimal number of linear measurements for the recovery of all s-sparse vectors is
$$m_{\mathrm{spa}} \asymp s \ln\left(e\frac{N}{s}\right).$$
- The minimal number of linear measurements for the recovery of all d-disjointed vectors is [Candès–Fernandez-Granda 14]
$$m_{\mathrm{dis}} \asymp \frac{N}{d}.$$
- What is the minimal number of linear measurements needed for the recovery of all s-sparse d-disjointed vectors?

Resolution of the Fundamental Question

- The minimal number of linear measurements for the recovery of all s-sparse vectors is
$$m_{\mathrm{spa}} \asymp s \ln\left(e\frac{N}{s}\right).$$
- The minimal number of linear measurements for the recovery of all d-disjointed vectors is [Candès–Fernandez-Granda 14]
$$m_{\mathrm{dis}} \asymp \frac{N}{d}.$$
- What is the minimal number of linear measurements needed for the recovery of all s-sparse d-disjointed vectors? Answer:
$$m_{\mathrm{spa\&dis}} \asymp s \ln\left(e\frac{N-d(s-1)}{s}\right).$$

Resolution of the Fundamental Question

- The minimal number of linear measurements for the recovery of all s-sparse vectors is
$$m_{\rm spa} \asymp s \ln\left(e\frac{N}{s}\right).$$
- The minimal number of linear measurements for the recovery of all d-disjointed vectors is [Candès–Fernandez-Granda 14]
$$m_{\rm dis} \asymp \frac{N}{d}.$$
- What is the minimal number of linear measurements needed for the recovery of all s-sparse d-disjointed vectors? Answer:
$$m_{\rm spa\&dis} \asymp s \ln\left(e\frac{N-d(s-1)}{s}\right).$$
- There is no benefit in knowing the simultaneity of sparsity and disjointedness over knowing only one of the structures, since
$$m_{\rm spa\&dis} \asymp \min\left\{m_{\rm spa}, m_{\rm dis}\right\}.$$

Sparse Disjointed Supports

Sparse Disjointed Supports

There are

$$\binom{N-d(s-1)}{s} \leq \left(e\frac{N-d(s-1)}{s}\right)^s$$

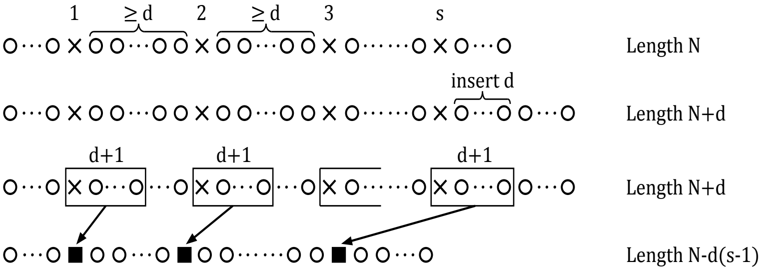
d-disjointed subsets of $[\![1:N]\!]$ with size s.

Sparse Disjointed Supports

There are

$$\binom{N-d(s-1)}{s} \leq \left(e\frac{N-d(s-1)}{s}\right)^s$$

d-disjointed subsets of $[\![1:N]\!]$ with size s.



Sufficient Number of Measurements via IHT

Sufficient Number of Measurements via IHT

- The adaptation of iterative hard thresholding is

$$\mathbf{x}^{n+1} = \mathbf{P}_{s,d}(\mathbf{x}^n + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^n)),$$

  where $\mathbf{P}_{s,d}$ is the projection onto s-sparse d-disjointed vectors.

Sufficient Number of Measurements via IHT

- The adaptation of iterative hard thresholding is

$$\mathbf{x}^{n+1} = \mathbf{P}_{s,d}(\mathbf{x}^n + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^n)),$$

  where $\mathbf{P}_{s,d}$ is the projection onto s-sparse d-disjointed vectors.
- For any s-sparse d-disjointed $\mathbf{x} \in \mathbb{R}^N$ and any $\mathbf{e} \in \mathbb{R}^m$,

$$\|\mathbf{x} - \lim_{n\to\infty} \mathbf{x}^n\|_2 \leq D\|\mathbf{e}\|_2$$

  as soon as the RI-like property

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

  holds with $\delta < 1/2$ for all s-sparse d-disjointed $\mathbf{z}, \mathbf{z}', \mathbf{z}'' \in \mathbb{R}^N$.

Sufficient Number of Measurements via IHT

- The adaptation of iterative hard thresholding is

$$\mathbf{x}^{n+1} = \mathbf{P}_{s,d}(\mathbf{x}^n + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^n)),$$

  where $\mathbf{P}_{s,d}$ is the projection onto s-sparse d-disjointed vectors.
- For any s-sparse d-disjointed $\mathbf{x} \in \mathbb{R}^N$ and any $\mathbf{e} \in \mathbb{R}^m$,

$$\|\mathbf{x} - \lim_{n\to\infty} \mathbf{x}^n\|_2 \leq D\|\mathbf{e}\|_2$$

  as soon as the RI-like property

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

  holds with $\delta < 1/2$ for all s-sparse d-disjointed $\mathbf{z}, \mathbf{z}', \mathbf{z}'' \in \mathbb{R}^N$.
- The latter occurs w/hp for $m \geq C\delta^{-2} \ln(e(N - d(s-1))/s)$.

Sufficient Number of Measurements via IHT

- The adaptation of iterative hard thresholding is

$$\mathbf{x}^{n+1} = \mathbf{P}_{s,d}(\mathbf{x}^n + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^n)),$$

  where $\mathbf{P}_{s,d}$ is the projection onto s-sparse d-disjointed vectors.
- For any s-sparse d-disjointed $\mathbf{x} \in \mathbb{R}^N$ and any $\mathbf{e} \in \mathbb{R}^m$,

$$\|\mathbf{x} - \lim_{n\to\infty} \mathbf{x}^n\|_2 \leq D\|\mathbf{e}\|_2$$

  as soon as the RI-like property

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

  holds with $\delta < 1/2$ for all s-sparse d-disjointed $\mathbf{z}, \mathbf{z}', \mathbf{z}'' \in \mathbb{R}^N$.
- The latter occurs w/hp for $m \geq C\delta^{-2}\ln(e(N - d(s-1))/s)$.
- Similar results obtained earlier for the adaptation of CoSaMP.

Computing the Projection $\mathbf{P}_{s,d}$

Computing the Projection $\mathbf{P}_{s,d}$

- [Hedge–Duarte–Cevher 09] propose an integer program relaxed to a linear program that is solved in $\mathcal{O}(N^{3.5})$ operations.

Computing the Projection $\mathbf{P}_{s,d}$

- [Hedge–Duarte–Cevher 09] propose an integer program relaxed to a linear program that is solved in $\mathcal{O}(N^{3.5})$ operations.
- A dynamic program can be solved in $\mathcal{O}(N^2)$ operations.

Computing the Projection $\mathbf{P}_{s,d}$

- [Hedge–Duarte–Cevher 09] propose an integer program relaxed to a linear program that is solved in $\mathcal{O}(N^{3.5})$ operations.
- A dynamic program can be solved in $\mathcal{O}(N^2)$ operations.
- Determine $F(N,s)$, where

$$F(n,r) := \min\left\{\sum_{j=1}^{n} |x_j - z_j|^2 : \mathbf{z} \in \mathbb{C}^n \text{ } r\text{-sparse } d\text{-disjointed}\right\}$$

$$= \min\left\{\begin{array}{c} F(n-1,r) + |x_n|^p, \\ F(n-d-1,r-1) + \sum\limits_{j=n-d}^{n-1} |x_j|^p. \end{array}\right.$$

Computing the Projection $\mathbf{P}_{s,d}$, ctd.

Dynamic program for $\mathbf{x} = (1, 0, 1, 2^{1/4}, 1, 0, 2^{-1/2})$, $s = 3$, $d = 1$.

$\mathbf{x}$
1
0
1
1.1892
1
0
0.7071

$F(n,r)$	$r=0$	$r=1$	$r=2$	$r=3$
$n=1$	*1*	*0*	*0*	*0*
$n=2$	*1*	*0*	*0*	*0*
$n=3$	*2*	1	0	0
$n=4$	*3.4142*	2	1	1
$n=5$	*4.4142*	3	2	1.4142
$n=6$	*4.4142*	3	2	1.4142
$n=7$	*4.9142*	3.5	2.5	1.9142

Necessary Number of Measurements

Necessary Number of Measurements

- Noninflating measurements relative to our model:

$$\|\mathbf{Az}\|_2 \leq c\|\mathbf{z}\|_2 \quad \text{whenever } \mathbf{z} \text{ is } s\text{-sparse } d\text{-disjointed.}$$

Necessary Number of Measurements

- Noninflating measurements relative to our model:

$$\|\mathbf{Az}\|_2 \leq c\|\mathbf{z}\|_2 \quad \text{whenever } \mathbf{z} \text{ is } s\text{-sparse } d\text{-disjointed.}$$

- Δ reconstruction map providing the robust estimate

$$\|\mathbf{x} - \Delta(\mathbf{Ax} + \mathbf{e})\|_2 \leq D\|\mathbf{e}\|_2,$$

valid for all s-sparse d-disjointed $\mathbf{x}$ and for all $\mathbf{e}$.

Necessary Number of Measurements

- Noninflating measurements relative to our model:

$$\|\mathbf{Az}\|_2 \le c\|\mathbf{z}\|_2 \quad \text{whenever } \mathbf{z} \text{ is } s\text{-sparse } d\text{-disjointed.}$$

- Δ reconstruction map providing the robust estimate

$$\|\mathbf{x} - \Delta(\mathbf{Ax} + \mathbf{e})\|_2 \le D\|\mathbf{e}\|_2,$$

  valid for all s-sparse d-disjointed $\mathbf{x}$ and for all $\mathbf{e}$.
- Then

$$m \ge C\, s \ln\left(e\frac{N - d(s-1)}{s}\right).$$

Key Combinatorial Lemma

Key Combinatorial Lemma

There exist

$$n \geq \left(\frac{N - d(s-1)}{c_1 s} \right)^{c_2 s}$$

d-disjointed subsets $S_1, \ldots, S_n$ of $[\![1 : N]\!]$ such that

$$\mathrm{card}(S_i) = s \quad \text{for all } i, \qquad \mathrm{card}(S_i \cap S_j) < \frac{s}{2} \quad \text{for all } i \neq j.$$

Key Combinatorial Lemma

There exist

$$n \geq \left(\frac{N - d(s-1)}{c_1 s} \right)^{c_2 s}$$

d-disjointed subsets $S_1, \ldots, S_n$ of $[\![1 : N]\!]$ such that

$$\mathrm{card}(S_i) = s \quad \text{for all } i, \qquad \mathrm{card}(S_i \cap S_j) < \frac{s}{2} \quad \text{for all } i \neq j.$$

This extends a crucial result known for $d = 0$ (sparse vectors),

Key Combinatorial Lemma

There exist

$$n \geq \left(\frac{N - d(s-1)}{c_1 s} \right)^{c_2 s}$$

d-disjointed subsets $S_1, \ldots, S_n$ of $[\![1 : N]\!]$ such that

$$\mathrm{card}(S_i) = s \quad \text{for all } i, \qquad \mathrm{card}(S_i \cap S_j) < \frac{s}{2} \quad \text{for all } i \neq j.$$

This extends a crucial result known for $d = 0$ (sparse vectors), but the counting argument must be somewhat refined.