Les mathématiques du Compressive Sensing

—

une introduction

Simon Foucart
Drexel University

Laboratoire Paul Painlevé
Université Lille 1
21-22 Mars 2013

RÉSUMÉ

Ce mini-cours est une excursion dans l'élégante théorie du "Compressive Sensing". Son but est de donner un aperçu assez complet des aspects mathématiques fondamentaux.

Le contenu de ce mini-cours est en partie basé sur l'ouvrage

A Mathematical Introduction to Compressive Sensing

écrit en collaboration avec Holger Rauhut.

Leçon 4: Reconstruction parcimonieuse avec matrices aléatoires

Cette dernière leçon établit l'existence de matrices satisfaisant la propriété d'isométrie restreinte dans le régime optimal. Des résultats non-uniformes sont également donnés pour la minimisation ℓ_1. Les outils probabilistes sont essentiels dans cette leçon. Le cas le plus simple des matrices gaussiennes est traité en détail, puis le cas des matrices de Fourier partielles est aussi évoqué.

Concentration Inequality

Concentration Inequality

- Let $A \in \mathbb{R}^{m \times N}$ be a random matrix with entries

$$a_{i,j} = \frac{g_{i,j}}{\sqrt{m}} \qquad \text{where the } g_{i,j} \text{ are independent } \mathcal{N}(0,1).$$

Concentration Inequality

- Let $A \in \mathbb{R}^{m \times N}$ be a random matrix with entries

$$a_{i,j} = \frac{g_{i,j}}{\sqrt{m}} \qquad \text{where the } g_{i,j} \text{ are independent } \mathcal{N}(0,1).$$

- For a fixed $\mathbf{x} \in \mathbb{R}^N$, note that $(A\mathbf{x})_i = \sum_{j=1}^{N} a_{i,j} x_j$, hence

$$\mathbb{E}\big((A\mathbf{x})_i^2\big) = \mathbb{V}\big(\sum a_{i,j} x_j\big) = \sum x_j^2 \mathbb{V}(a_{i,j}) = \frac{\|\mathbf{x}\|_2^2}{m},$$
$$\mathbb{E}\big(\|A\mathbf{x}\|_2^2\big) = \|\mathbf{x}\|_2^2.$$

Concentration Inequality

- Let $A \in \mathbb{R}^{m \times N}$ be a random matrix with entries

$$a_{i,j} = \frac{g_{i,j}}{\sqrt{m}} \qquad \text{where the } g_{i,j} \text{ are independent } \mathcal{N}(0,1).$$

- For a fixed $\mathbf{x} \in \mathbb{R}^N$, note that $(A\mathbf{x})_i = \sum_{j=1}^N a_{i,j}x_j$, hence

$$\mathbb{E}\big((A\mathbf{x})_i^2\big) = \mathbb{V}\Big(\sum a_{i,j}x_j\Big) = \sum x_j^2 \mathbb{V}(a_{i,j}) = \frac{\|\mathbf{x}\|_2^2}{m},$$
$$\mathbb{E}\big(\|A\mathbf{x}\|_2^2\big) = \|\mathbf{x}\|_2^2.$$

- In fact, $\|A\mathbf{x}\|_2^2$ concentrates around its mean as, for $t \in (0,1)$,

$$\text{(CI)} \qquad \mathbb{P}\big(\big|\|A\mathbf{x}\|_2^2 - \|\mathbf{x}\|_2^2\big| > t\|\mathbf{x}\|_2^2\big) \le 2\exp(-ct^2 m).$$

Covering Arguments

Covering Arguments

Suppose that the random matrix $A \in \mathbb{R}^{m \times N}$ satisfies (CI).

Covering Arguments

Suppose that the random matrix $A \in \mathbb{R}^{m \times N}$ satisfies (CI).
Let $S \subseteq [N]$ with $\mathrm{card}(S) = s$.

Covering Arguments

Suppose that the random matrix $A \in \mathbb{R}^{m \times N}$ satisfies (CI).
Let $S \subseteq [N]$ with $\mathrm{card}(S) = s$. Then

$$\mathbb{P}\big(\|A_S^* A_S - \mathrm{Id}\|_{2\to 2} > \delta\big) \leq 2\exp(-c\delta^2 m)$$

Covering Arguments

Suppose that the random matrix $A \in \mathbb{R}^{m \times N}$ satisfies (CI).
Let $S \subseteq [N]$ with $\mathrm{card}(S) = s$. Then

$$\mathbb{P}\big(\|A_S^* A_S - \mathrm{Id}\|_{2\to 2} > \delta\big) \leq 2\exp(-c\delta^2 m)$$

provided

$$m \geq \frac{c'}{\delta^2} s.$$

Covering Arguments

Suppose that the random matrix $A \in \mathbb{R}^{m \times N}$ satisfies (CI).
Let $S \subseteq [N]$ with $\mathrm{card}(S) = s$. Then

$$\mathbb{P}\big(\|A_S^* A_S - \mathrm{Id}\|_{2\to 2} > \delta\big) \le 2\exp(-c\delta^2 m)$$

provided

$$m \ge \frac{c'}{\delta^2} s.$$

The argument relies on the following fact:

A subset U of the unit ball of $\mathbb{R}^n$ relative to a norm $\|\cdot\|$ has covering and packing numbers satisfying

$$\mathcal{N}(U, \|\cdot\|, t) \le \mathcal{P}(U, \|\cdot\|, t) \le \left(1 + \frac{2}{t}\right)^n.$$

Restricted Isometry Property

Restricted Isometry Property

- Suppose that the random matrix $A \in \mathbb{R}^{m \times N}$ satisfies (CI). Then
$$\mathbb{P}(\delta_s > \delta) \leq 2\exp(-c\delta^2 m)$$
provided
$$m \geq \frac{c'}{\delta^2} s \ln(eN/s).$$

Restricted Isometry Property

- Suppose that the random matrix $A \in \mathbb{R}^{m \times N}$ satisfies (CI). Then
$$\mathbb{P}(\delta_s > \delta) \leq 2\exp(-c\delta^2 m)$$
provided
$$m \geq \frac{c'}{\delta^2} s \ln(eN/s).$$
- The arguments are also valid for subgaussian matrices (e.g. Bernoulli), since these satisfy (CI), too.

Restricted Isometry Property

- Suppose that the random matrix $A \in \mathbb{R}^{m \times N}$ satisfies (CI). Then
$$\mathbb{P}(\delta_s > \delta) \leq 2\exp(-c\delta^2 m)$$
provided
$$m \geq \frac{c'}{\delta^2} s \ln(eN/s).$$
- The arguments are also valid for subgaussian matrices (e.g. Bernoulli), since these satisfy (CI), too.
- For Gaussian matrices, more powerful techniques can provide an explicit value for c'.

Nonuniform Recovery when $m \gtrsim 2s\ln(N/s)$

Nonuniform Recovery when $m \gtrsim 2s\ln(N/s)$

- A necessary and sufficient condition for the recovery of a fixed $\mathbf{x} \in \mathbb{C}^N$ supported on S via ℓ_1-minimization is

$$\left|\sum_{j \in S} \overline{\operatorname{sgn}(x_j)} v_j\right| < \|\mathbf{v}_{\overline{S}}\|_1 \qquad \text{for all } \mathbf{v}_{\neq \mathbf{0}} \in \ker A.$$

Nonuniform Recovery when $m \gtrsim 2s\ln(N/s)$

- A necessary and sufficient condition for the recovery of a fixed $\mathbf{x} \in \mathbb{C}^N$ supported on S via ℓ_1-minimization is

$$\Big|\sum_{j \in S} \overline{\mathrm{sgn}(x_j)} v_j\Big| < \|\mathbf{v}_{\overline{S}}\|_1 \qquad \text{for all } \mathbf{v}_{\neq \mathbf{0}} \in \ker A.$$

- This is implied by (in the real case, equivalent to) the injectivity of A_S and the existence of $\mathbf{h} \in \mathbb{C}^m$ such that

$$(A^*\mathbf{h})_j = \mathrm{sgn}(x_j),\ j \in S, \qquad \big|(A^*\mathbf{h})_\ell\big| < 1,\ \ell \in \overline{S}.$$

Nonuniform Recovery when $m \gtrsim 2s\ln(N/s)$

- A necessary and sufficient condition for the recovery of a fixed $\mathbf{x} \in \mathbb{C}^N$ supported on S via ℓ_1-minimization is

$$\left|\sum_{j\in S} \overline{\operatorname{sgn}(x_j)} v_j\right| < \|\mathbf{v}_{\overline{S}}\|_1 \qquad \text{for all } \mathbf{v}_{\neq \mathbf{0}} \in \ker A.$$

- This is implied by (in the real case, equivalent to) the injectivity of A_S and the existence of $\mathbf{h} \in \mathbb{C}^m$ such that

$$(A^*\mathbf{h})_j = \operatorname{sgn}(x_j),\ j \in S, \qquad \left|(A^*\mathbf{h})_\ell\right| < 1,\ \ell \in \overline{S}.$$

- The latter is in turn implied by

$$\left|\langle A_S^\dagger \mathbf{a}_\ell, \operatorname{sgn}(\mathbf{x}_S)\rangle\right| < 1 \qquad \text{for all } \ell \in \overline{S}.$$

Nonuniform Recovery when $m \gtrsim 2s\ln(N/s)$

- A necessary and sufficient condition for the recovery of a fixed $\mathbf{x} \in \mathbb{C}^N$ supported on S via ℓ_1-minimization is

$$\left| \sum_{j \in S} \overline{\mathrm{sgn}(x_j)} v_j \right| < \|\mathbf{v}_{\overline{S}}\|_1 \qquad \text{for all } \mathbf{v}_{\neq \mathbf{0}} \in \ker A.$$

- This is implied by (in the real case, equivalent to) the injectivity of A_S and the existence of $\mathbf{h} \in \mathbb{C}^m$ such that

$$(A^*\mathbf{h})_j = \mathrm{sgn}(x_j),\ j \in S, \qquad \left|(A^*\mathbf{h})_\ell\right| < 1,\ \ell \in \overline{S}.$$

- The latter is in turn implied by

$$\left|\langle A_S^\dagger \mathbf{a}_\ell, \mathrm{sgn}(\mathbf{x}_S)\rangle\right| < 1 \qquad \text{for all } \ell \in \overline{S}.$$

- For subgaussian matrices, this occurs with probability $\geq 1 - \varepsilon$ provided

$$m \gtrsim 2s\ln(2N/\varepsilon)$$

Nonuniform Recovery when $m \gtrsim 2s \ln(N/s)$

- A necessary and sufficient condition for the recovery of a fixed $\mathbf{x} \in \mathbb{C}^N$ supported on S via ℓ_1-minimization is

$$\left| \sum_{j \in S} \overline{\operatorname{sgn}(x_j)} v_j \right| < \|\mathbf{v}_{\overline{S}}\|_1 \qquad \text{for all } \mathbf{v}_{\neq \mathbf{0}} \in \ker A.$$

- This is implied by (in the real case, equivalent to) the injectivity of A_S and the existence of $\mathbf{h} \in \mathbb{C}^m$ such that

$$(A^* \mathbf{h})_j = \operatorname{sgn}(x_j),\ j \in S, \qquad \left|(A^* \mathbf{h})_\ell\right| < 1,\ \ell \in \overline{S}.$$

- The latter is in turn implied by

$$\left|\langle A_S^\dagger \mathbf{a}_\ell, \operatorname{sgn}(\mathbf{x}_S)\rangle\right| < 1 \qquad \text{for all } \ell \in \overline{S}.$$

- For subgaussian matrices, this occurs with probability $\geq 1 - \varepsilon$ provided

$$m \gtrsim 2s \ln(2N/\varepsilon)$$

(which is replaced by $m \gtrsim 2s\left(\sqrt{\ln(eN/s)} + \sqrt{\ln(1/\varepsilon)/s}\right)^2$ for Gaussian matrices using Gordon's *escape through the mesh*).

Bounded Orthonormal Systems

Bounded Orthonormal Systems

Let $\mathcal{D} \in \mathbb{R}^d$ be endowed with a probability measure ν. A bounded orthonormal system (BOS) with constant $K \geq 1$ is a system $(\phi_1, \dots, \phi_N)$ of function of $\mathcal{D}$ satisfying

$$\int_{\mathcal{D}} \phi_j(t)\overline{\phi_k(t)} d\nu(t) = \delta_{j,k},$$

$$\sup_{1 \leq j \leq N} \sup_{t \in \mathcal{D}} |\phi_j(t)| \leq K.$$

Bounded Orthonormal Systems

Let $\mathcal{D} \in \mathbb{R}^d$ be endowed with a probability measure ν. A bounded orthonormal system (BOS) with constant $K \geq 1$ is a system $(\phi_1, \dots, \phi_N)$ of function of $\mathcal{D}$ satisfying

$$\int_{\mathcal{D}} \phi_j(t)\overline{\phi_k(t)}d\nu(t) = \delta_{j,k},$$

$$\sup_{1\leq j\leq N} \sup_{t\in\mathcal{D}} |\phi_j(t)| \leq K.$$

1. trigonometric polynomials:
   take $\phi_k(t) = e^{i2\pi kt}$ for $t \in \mathcal{D} = [0,1]$, $d\nu(t) = dt$,

Bounded Orthonormal Systems

Let $\mathcal{D} \in \mathbb{R}^d$ be endowed with a probability measure ν. A bounded orthonormal system (BOS) with constant $K \geq 1$ is a system $(\phi_1, \ldots, \phi_N)$ of function of $\mathcal{D}$ satisfying

$$\int_{\mathcal{D}} \phi_j(t)\overline{\phi_k(t)}d\nu(t) = \delta_{j,k},$$

$$\sup_{1 \leq j \leq N} \sup_{t \in \mathcal{D}} |\phi_j(t)| \leq K.$$

1. trigonometric polynomials:
   take $\phi_k(t) = e^{i2\pi kt}$ for $t \in \mathcal{D} = [0,1]$, $d\nu(t) = dt$,
2. discrete orthonormal systems — let $U \in \mathbb{C}^{N\times N}$ be a unitary matrix with $\sqrt{N}|U_{k,j}| \leq K$ for all $k, j \in \{1, \ldots, N\}$ (e.g. Fourier or Hadamard matrix):
   take $\phi_k(t) = \sqrt{N}U_{k,t}$ for $t \in \{1, \ldots, N\}$, $\nu(B) = \dfrac{\text{card}(B)}{N}$.

Nonuniform Recovery for BOS

Nonuniform Recovery for BOS

For $f = \sum_{j=1}^{N} x_j \phi_j$, the samples of f at $t_1, \ldots, t_m \in \mathcal{D}$ are

$$f(t_k) = \sum_{j=1}^{N} x_j \phi_j(t_k) = (A\mathbf{x})_k,$$

where the sampling matrix $A \in \mathbb{C}^{m \times N}$ have entries

$$A_{k,j} = \phi_j(t_k). \tag{1}$$

Nonuniform Recovery for BOS

For $f = \sum_{j=1}^{N} x_j \phi_j$, the samples of f at $t_1, \ldots, t_m \in \mathcal{D}$ are

$$f(t_k) = \sum_{j=1}^{N} x_j \phi_j(t_k) = (A\mathbf{x})_k,$$

where the sampling matrix $A \in \mathbb{C}^{m \times N}$ have entries

$$A_{k,j} = \phi_j(t_k). \tag{1}$$

Let an s-sparse $\mathbf{x} \in \mathbb{C}^N$ be fixed.

Nonuniform Recovery for BOS

For $f = \sum_{j=1}^{N} x_j \phi_j$, the samples of f at $t_1, \ldots, t_m \in \mathcal{D}$ are

$$f(t_k) = \sum_{j=1}^{N} x_j \phi_j(t_k) = (A\mathbf{x})_k,$$

where the sampling matrix $A \in \mathbb{C}^{m \times N}$ have entries

$$A_{k,j} = \phi_j(t_k). \tag{1}$$

Let an s-sparse $\mathbf{x} \in \mathbb{C}^N$ be fixed. Let the sampling points $t_1, \ldots, t_m$ be randomly drawn according to ν, and let $A \in \mathbb{C}^{m \times N}$ be the sampling matrix (1).

Nonuniform Recovery for BOS

For $f = \sum_{j=1}^{N} x_j \phi_j$, the samples of f at $t_1, \ldots, t_m \in \mathcal{D}$ are

$$f(t_k) = \sum_{j=1}^{N} x_j \phi_j(t_k) = (A\mathbf{x})_k,$$

where the sampling matrix $A \in \mathbb{C}^{m \times N}$ have entries

$$A_{k,j} = \phi_j(t_k). \tag{1}$$

Let an s-sparse $\mathbf{x} \in \mathbb{C}^N$ be fixed. Let the sampling points $t_1, \ldots, t_m$ be randomly drawn according to ν, and let $A \in \mathbb{C}^{m \times N}$ be the sampling matrix (1). If

$$m \geq C\, K^2 s \ln(N) \ln(\varepsilon^{-1}),$$

then $\mathbf{x}$ is the unique minimizer of $\|\mathbf{z}\|_1$ subject to $A\mathbf{z} = A\mathbf{x}$ with probability at least $1 - \varepsilon$.

Nonuniform Recovery for BOS

For $f = \sum_{j=1}^{N} x_j \phi_j$, the samples of f at $t_1, \ldots, t_m \in \mathcal{D}$ are

$$f(t_k) = \sum_{j=1}^{N} x_j \phi_j(t_k) = (A\mathbf{x})_k,$$

where the sampling matrix $A \in \mathbb{C}^{m \times N}$ have entries

$$A_{k,j} = \phi_j(t_k). \tag{1}$$

Let an s-sparse $\mathbf{x} \in \mathbb{C}^N$ be fixed. Let the sampling points $t_1, \ldots, t_m$ be randomly drawn according to ν, and let $A \in \mathbb{C}^{m \times N}$ be the sampling matrix (1). If

$$m \geq C\, K^2 s \ln(N) \ln(\varepsilon^{-1}),$$

then $\mathbf{x}$ is the unique minimizer of $\|\mathbf{z}\|_1$ subject to $A\mathbf{z} = A\mathbf{x}$ with probability at least $1 - \varepsilon$.

(Proof based on the *golfing scheme*).

Restricted Isometry Property for BOS

Restricted Isometry Property for BOS

Let $A \in \mathbb{C}^{m \times N}$ be the random sampling matrix associated to a BOS with constant $K \geq 1$.

Restricted Isometry Property for BOS

Let $A \in \mathbb{C}^{m \times N}$ be the random sampling matrix associated to a BOS with constant $K \geq 1$. For $\delta \in (0, 1)$, if

$$m \geq C\, K^2 \delta^{-2} s \ln^4(N),$$

then with probability at least $1 - N^{-\ln^3(N)}$ the matrix $\frac{1}{\sqrt{m}} A$ has a restricted isometry constant satisfying $\delta_s \leq \delta$.

Restricted Isometry Property for BOS

Let $A \in \mathbb{C}^{m \times N}$ be the random sampling matrix associated to a BOS with constant $K \geq 1$. For $\delta \in (0, 1)$, if

$$m \geq C\, K^2 \delta^{-2} s \ln^4(N),$$

then with probability at least $1 - N^{-\ln^3(N)}$ the matrix $\frac{1}{\sqrt{m}} A$ has a restricted isometry constant satisfying $\delta_s \leq \delta$.

(Proof uses *Dudley's inequality*, *empirical method of Maurey*, etc.)