Reconstruction From Anisotropic Random Measurements

Mark Rudelson and Shuheng Zhou

March 26, 2014
High-dimensional sparse models

\[ y = X\beta^* + w, \quad y \in \mathbb{R}^n, X \in \mathbb{R}^{n \times p}, w \sim (0, \sigma^2 I_{n \times n}), p >> n \]

Assumption of exact sparsity

\[ S(\beta^*) := \{ j \in \{1, \ldots, p\} | \beta_j^* \neq 0 \} \]

Problem reduces to: Find \( \hat{\beta} \) close to \( \beta^* \) such that \( \|\beta\|_0 \leq s \)

Convex relaxation: Use \( \ell_1 \)-norm along with different estimators

**Basis pursuit:** \( \hat{\beta} \in \arg\min_{\beta \in \mathbb{R}^p} \|\beta\|_1 \quad \text{such that} \quad X\beta = y \)

**Lasso:** \( \hat{\beta} \in \arg\min_{\beta \in \mathbb{R}^p} \{ \|y - X\beta\|_2^2 + \lambda\|\beta\|_1 \} \)

Under what conditions on matrix \( X \) can we recover \( \hat{\beta} \)?
• $e_1, \ldots, e_p$ are the canonical basis of $\mathbb{R}^p$

• For a set $J \subset \{1, \ldots, p\}$ denote $E_J = \text{span}\{e_j : j \in J\}$

• For a set $V \subset \mathbb{R}^p$, $\text{conv}(V)$ - convex hull of $V$ and $\text{absconv}(V)$ - absolute convex hull of $V$

• $B^p_2$ - unit Euclidean ball, $S^{p-1}$ - Unit sphere

• For a vector $u \in \mathbb{R}^p$, $T_0$ denotes the location of the $s_0$ largest coefficients of $u$ in absolute values, $u_{T_0}$ - subvector of $u$ confined to index locations given by set $T_0$

• $C(s_0, k_0) := \{x \in \mathbb{R}^p | \exists I \in [1, p], |I| = s_0 \text{ s.t. } \|x_{I^c}\|_1 \leq k_0 \|x_I\|_1\}$, $k_0 = 1$ for Dantzig and $k_0 = 3$ for Lasso

• $A_{q \times p}$ satisfies $RE(s_0, k_0, A)$ condition with parameter $K(s_0, k_0, A)$ if for any $v \neq 0$

\[
\frac{1}{K(s_0, k_0, A)} := \min_{J \subset \{1, \ldots, p\}, |J| \leq s_0} \min_{J \subset \{1, \ldots, p\}, |J| \leq s_0} \frac{\min \|v_{J^c}\|_1 \leq k_0 \|v_J\|_1}{\|Av\|_2} > 0
\]
Main Result - Reduction Principle

- Define: \( X = \Psi A \)

**Reduction principle condition**

Let \( 1/5 > \delta > 0 \). Let \( 0 < s_0 < p \) and \( k_0 > 0 \). Let \( A \) be a \( q \times p \) matrix such that \( RE(s_0, 3k_0, A) \) holds for \( 0 < K(s_0, 3k_0, A) < \infty \). Set

\[
d = s_0 + s_0 \max_j \| Ae_j \|_2^2 \times \frac{16K^2(s_0, 3k_0, A)(3k_0)^2(3k_0 + 1)}{\delta^2}
\]

Let \( E = \bigcup_{|J| = d} E_J \) for \( d < p \) and \( E \) denotes \( \mathbb{R}^p \) otherwise. Let \( \tilde{\Psi} \) be a matrix s.t.

\[
\forall x \in AE \quad (1 - \delta)\|x\|_2 \leq \|\tilde{\Psi}x\|_2 \leq (1 + \delta)\|x\|_2
\]

**Theorem 3**

Under the reduction principle condition \( RE(s_0, k_0, \tilde{\Psi}A) \) condition holds with

\[
0 < K(s_0, k_0, \tilde{\Psi}A) \leq K(s_0, k_0, A)/(1 - 5\delta)
\]
Theorem 10

Under the reduction principle condition for any $x \in A(C(s_0, k_0) \cap S^{q-1}$

$$(1 - 5\delta) \leq \|\tilde{\Psi}x\|_2 \leq (1 + 3\delta)$$

Proof:

- $RE(s_0, k_0, A)$ condition holds for $A$. Therefore for any $u \in C(s_0, k_0)$

  $$\|Au\|_2 \geq \frac{\|u_T\|_2}{K(s_0, k_0, A)} > 0$$

- If condition of Theorem 10 is satisfied

  $$\|\tilde{\Psi}Au\|_2 \geq (1 - 5\delta)\|Au\|_2 \geq (1 - 5\delta)\frac{\|u_T\|_2}{K(s_0, k_0, A)} > 0$$
Lemma 14

Let $1 > \delta > 0$. Let $0 < s_0 < p$ and $k_0 > 0$. Let $A$ be a $q \times p$ matrix such that $RE(s_0, k_0, A)$ condition holds for $0 < K(s_0, k_0, A) < \infty$. Define

$$d = d(k_0, A) = s_0 + s_0 \max_j \|Ae_j\|_2^2 \times \frac{16K^2(s_0, k_0, A)k_0^2(k_0 + 1)}{\delta^2}$$

Then

$$A(C(s_0, k_0)) \cap S^{q-1} \subset (1 - \delta)^{-1} \text{conv} \left( \bigcup_{|J| \leq d} AE_J \cap S^{q-1} \right)$$

- Lemma 14 is vacuously true for $d > p$
- Consider a set $V$

$$V := \{ x = x_{T_0} + x_{T_0^c} \in x_{T_0} + k_0 \|x_{T_0}\|_1 \text{absconv}(e_j | j \in T_0^c) | x \in C(s_0, k_0) \cap S^{p-1} \}$$

- Define function $F(v)$ for any $v \in \mathbb{R}^p$ such that $\|Av\|_2 \neq 0$

$$F(v) = \frac{Av}{\|Av\|_2}$$

- Then $AC(s_0, k_0) \cap S^{q-1} = F(C(s_0, k_0) \setminus \{0\}) = F(V)$
By duality, Lemma 14 can be derived from the fact that the supremum of any linear functional over l.h.s does not exceed the supremum over the r.h.s.

To prove that: For any $\theta \in S^{q-1}$, $\exists z' \in \mathbb{R}^p \setminus \{0\}$ s.t. $|\text{supp}(z')| \leq d$ and $F(z')$ is well defined and satisfies

$$z = \max_{v \in V} \langle F(v), \theta \rangle \leq (1 - \delta)^{-1} \langle F(z'), \theta \rangle$$

There exists $I \subset \{1, ...., p\}$ such that $|I| = s_0$, and for some $\epsilon_j \in \{1, -1\}$

$$z = z_I + \|z_I\|_1 k_0 \sum_{j \in I^c} \alpha_j \epsilon_j e_j$$

where $\alpha_j \in [0, 1)$ for all $i \in I^c$.

Set $\alpha_{p+1} = 1 - \sum_{j \in I^c} \alpha_j$ and $e_{p+1} = \vec{0}$

$$y := \|z_I\|_1 k_0 \sum_{j \in I^c \cup \{p+1\}} \alpha_j \epsilon_j e_j$$
Lemma 11 - Maurey’s empirical approximation argument

Let $u_1, \ldots, u_M \in \mathbb{R}^q$. Let $y \in \text{conv}(u_1, \ldots, u_M)$. Then, there exists a set $L \subset \{1, 2, \ldots, M\}$ such that

$$|L| \leq m = \frac{4 \max_{j \in \{1, \ldots, M\}} \|u_j\|_2^2}{\epsilon^2}$$

and a vector $y' \in \text{conv}(u_j, j \in L)$ such that

$$\|y' - y\|_2 \leq \epsilon$$

Following from the previous slide denote

$M := \{j \in I^c \cup \{p + 1\} : \alpha_j > 0\}$ and let $\epsilon > 0$ to be defined later

- $u_j = k_0 \|z_I\|_1 \epsilon_j A e_j$ for $j \in M$

- Construct a set $J' \subset M$ satisfying

$$\|J'\| \leq m := \frac{4 \max_{j \in I^c} k_0^2 \|z_I\|_1^2 \|A e_j\|_2^2}{\epsilon^2} \leq \frac{4 k_0^2 s_0 \max_{j \in I^c} \|A e_j\|_2^2}{\epsilon^2}$$

and a vector $y' = k_0 \|z_I\|_1 \sum_{j \in J'} \beta_j \epsilon_j A e_j$, $\beta_j \in [0, 1]$ and $\sum_{j \in J'} \beta_j = 1$
• Set $z' = z_I + y'$ and $\|Az - Az'\|_2 \leq \epsilon$. By construction $Az' \in AE_j$

• Consider the vector

$$z + \lambda(z' - z) = z_I + k_0\|z_I\|_1 \sum_{j \in I^c \cup \{p+1\}} [(1 - \lambda)\alpha_j + \lambda\beta_j]e_j$$

where $\sum_{j \in I^c \cup \{p+1\}} [(1 - \lambda)\alpha_j + \lambda\beta_j] = 1$ and $\exists \delta_0 > 0$ s.t. $\forall j \in I^c \cup \{p + 1\}, (1 - \lambda)\alpha_j + \lambda\beta_j \in [0, 1]$ if $|\lambda| < \delta_0$

• Therefore $z + \lambda(z' - z) \in V$ whenever $|\lambda| < \delta_0$

• Consider a function $\phi : (-\delta_0, \delta_0) \rightarrow \mathbb{R}$

$$\phi(\lambda) := \langle F(z + \lambda(z' - z)), \theta \rangle = \frac{\langle Az + \lambda(Az' - Az), \theta \rangle}{\|Az + \lambda(Az' - Az)\|_2}$$

• $\phi(\lambda)$ attains the local maxima at 0
Lemma 13

Let \( u, \theta, x \in \mathbb{R}^q \) be vectors such

1) \( \|\theta\|_2 = 1 \)
2) \( \langle x, \theta \rangle \neq 0 \)
3) Vector \( u \) is not parallel to \( x \).

Define \( \phi : \mathbb{R} \to \mathbb{R} \) by

\[
\phi(\lambda) = \frac{\langle x + \lambda u, \theta \rangle}{\|x + \lambda u\|_2}
\]

Assume \( \phi(\lambda) \) has a local maximum at 0; then

\[
\frac{\langle x + u, \theta \rangle}{\langle x, \theta \rangle} \geq 1 - \frac{\|u\|_2}{\|x\|_2}
\]

- Applying the above lemma after setting \( \epsilon = \frac{\delta}{2\sqrt{1+k_0K(s_0,k_0,a)}} \) and after simplifications

\[
\frac{\langle F(z'), \theta \rangle}{\langle F(z), \theta \rangle} \geq 1 - \delta
\]

and

\[
m \leq s_0 \max_{j \in I_c} \|Ae_j\|_2 \left( \frac{16K^2(s_0, k_0, A)k_0^2(k_0 + 1)}{\delta^2} \right)
\]
The upper bound follows naturally from Lemma 14. For any vector \( x \in A(C(s_0, 3k_0)) \cap S^{q-1} \)

\[
\|\tilde{\Psi}x\|_2 \leq (1 + \delta)(1 - \delta)^{-1} \leq 1 + 3\delta, \text{ for } \delta < 1/3
\]

For the lower bound:

- Assume \( x \in C(s_0, k_0) \cap S^{p-1} \) and \( x = x_I + x_{I^c} \)
- Construct a vector \( d \)-sparse vector \( y = x_I + u \), such that \( \|u\|_1 = \|y_{I^c}\|_1 = \|x_{I^c}\|_1 \), \( y \in C(S_0, k_0) \) and \( \|Ax - Ay\|_2 \leq \epsilon \)
- If \( \epsilon \) is chosen such that \( y \) is \( d \)-sparse then \( \left\| \frac{\tilde{\Psi}Ay}{\|Ay\|} \right\| \geq 1 - \delta \)
- Choose \( v \) such that \( y = \frac{1}{2}(x + v) \), \( v \in C(s_0, k_0) \)
- Comparison of upper estimate for \( v \) with the lower estimate of \( y \) will yield the result on \( x \) as

\[
\left\| \frac{\tilde{\Psi}Ax}{\|Ax\|} \right\|_2 \geq 1 - 5\delta \text{ for } \delta < 1/5
\]
Random matrix decompositions

- Apply reduction principle to different classes of random design matrices
- Analysis reduces to checking the almost isometry property holds for all vectors from some low-dimensional subspaces
- Consider random matrix $\Psi$ whose rows are independent isotropic vectors with sub-Gaussian marginals
  - A random vector $Y \in \mathbb{R}^p$ is called isotropic if for every $y \in \mathbb{R}^p$
    $$\mathbb{E}|\langle Y, y \rangle|^2 = \|y\|^2$$
  - $Y$ is $\psi_2$ with constant $\alpha$ if for every $y \in \mathbb{R}^p$
    $$\|\langle Y, y \rangle\|_{\psi_2} := \inf\{t : \mathbb{E}\exp(\langle Y, y \rangle^2 / t^2) \leq 2\} \leq \alpha\|y\|_2$$
- Random vector $Y$ with i.i.d $N(0, 1)$ random coordinates is an isotropic random vector
- Any sub-Gaussian design matrix $X$ can be expressed as $X = \Psi \Sigma^{1/2}$
- For any random vector $Y$, $\Psi = \Sigma^{-1/2} Y$ is an isotropic random vector
Sub-Gaussian condition

Set $0 < \delta < 1$, $k_0 > 0$, and $0 < s_0 < p$. Let $A$ be a $q \times p$ matrix satisfying the RE($s_0, 3k_0, A$) condition. Let $d$ be as defined earlier, and let $m = \min(d, p)$. Let $\Psi$ be a $n \times q$ matrix whose rows are independent isotropic $\psi_2$ random vectors in $\mathbb{R}^q$ with constant $\alpha$. Suppose the sample size satisfies

$$n \geq \frac{2000m\alpha^4}{\delta^2} \log \left( \frac{60ep}{m\delta} \right)$$

Theorem 6

Under the condition above with probability at least $1 - 2\exp(-\delta^2 n/2000\alpha^4)$, the RE($s_0, k_0, \frac{1}{\sqrt{n}} \Psi A$) condition holds for matrix $\frac{1}{\sqrt{n}} \Psi A$ with

$$0 < K \left( s_0, k_0, \frac{1}{\sqrt{n}} \Psi A \right) \leq \frac{K(s_0, k_0, A)}{1 - \delta}$$
Theorem 16

Under the sub-gaussian condition above with probability at least \( 1 - 2\exp(\delta^2 n/2000\alpha^4) \), for all \( v \in C(s_0, k_0) \) s.t. \( v \neq 0 \), we have

\[
(1 - \delta) \leq \frac{1}{\sqrt{n}} \frac{\| \Psi Av \|_2}{\| Av \|_2} \leq 1 + \delta
\]

Theorem 17

Set \( 0 < \delta < 1 \). Let \( A \) be a \( q \times p \) matrix, and let \( \Psi \) be an \( n \times q \) matrix whose rows are independent \( \psi_2 \) random vectors in \( \mathbb{R}^q \) with constant \( \alpha \). For \( m \leq p \),

\[
n \geq \frac{80m\alpha^4}{\tau^2} \log \left( \frac{12ep}{m\tau} \right)
\]

Then with prob. at least \( 1 - 2\exp(-\tau^2 n/80\alpha^4) \), for all \( m \)-sparse vectors \( u \) in \( \mathbb{R}^p \)

\[
(1 - \tau)\| Au \|_2 \leq \frac{1}{\sqrt{n}} \| \Psi Au \|_2 \leq (1 + \tau)\| Au \|_2
\]

- Theorem 6 follows from Theorem 16 and Theorem 16 follows from Theorem 17 by Theorem 10.
Lemma 20

Given $m \geq 1$ and $\epsilon > 0$. There exists an $\epsilon$-net $\Pi \subset B_2^m$ of $B_2^m$ with respect to the Euclidean metric such that $B_2^m \subset (1 - \epsilon)^{-1} \text{conv}(\Pi)$ and $|\Pi| \leq (1 + 2/\epsilon)^m$. Similarly there exists an $\epsilon$-net of the sphere $S^{m-1}$, $\Pi' \subset S^{m-1}$ such that $|\Pi'| \leq (1 + 2/\epsilon)^m$.

- For a set $J \subset \{1, \ldots, p\}$, denote $E_J = \text{span}\{e_j : j \in J\}$, and set $F_J = AE_J$.
- Covering number for set $F_J \cap S^{q-1}$: $|\Pi_J| \leq (1 + 2/\epsilon)^m$.
- If $\Pi = \bigcup_{|J|=m} \Pi_J$
  \[
  |\Pi| = (3/\epsilon)^m \binom{p}{m} \leq \left( \frac{3ep}{m\epsilon} \right)^m = \exp \left( m \log \left( \frac{3ep}{m\epsilon} \right) \right)
  \]
- For $y \in S^{q-1} \cup F_J$, let $\pi(y)$ be one of the closest point in the $\epsilon$-cover $\Pi_J$. Then
  \[
  \frac{y - \pi(y)}{\|y - \pi(y)\|_2} \in F_J \cup S^{q-1}, \text{ where } \|y - \pi(y)\|_2 \leq \epsilon
  \]
Lemma 21

Let $Y_1, \ldots, Y_n$ be independent random variables such that $\mathbb{E} Y_j^2 = 1$ and $\|Y_j\| \leq \alpha$ for all $j = 1, \ldots, n$. Then for any $\theta \in (0, 1)$

$$\mathbb{P} \left( \left| \frac{1}{n} \sum_{j=1}^{n} Y_j^2 - 1 \right| > \theta \right) \leq 2 \exp \left( - \frac{\theta^2 n}{10\alpha^4} \right)$$

Let $\Gamma = n^{-1/2} \Psi$ and let $x \in S^{q-1}$

$$\mathbb{P} \left( \left| \| \Gamma x \|_2^2 - 1 \right| > \theta \right) = \mathbb{P} \left( \left| \frac{1}{n} \sum_{i=1}^{n} \langle \Psi, x \rangle^2 - 1 \right| > \theta \right) \leq 2 \exp \left( - \frac{n\theta^2}{10\alpha^4} \right)$$

Union bound implies

$$\mathbb{P} \left( \exists x \in \Pi \text{s.t.} \left| \| \Gamma x \|_2^2 \right| > \theta \right) \leq 2 |\Pi| \exp \left( - \frac{n\theta^2}{10\alpha^4} \right)$$

Bound over entire $S^{q-1} \cap F_J$ is obtained by approximation

$$\left( 1 - 2\theta \right) \| Au \|_2 \leq \| \Gamma Au \|_2 \leq \left( 1 + 2\theta \right) \| Au \|_2$$

Taking $\tau = \theta/2$ proves Theorem 17
Condition for random matrices with bounded entries

Let $0 < \delta < 1$ and $0 < s_0 < p$. Let $Y \in \mathbb{R}^p$ be a random vector such that $\|Y\|_\infty \leq M$ a.s. and denote $\Sigma = \mathbb{E}YY^T$. Let $X$ be an $n \times p$ matrix, whose rows $X_1, \ldots, X_n$ are independent copies of $Y$. Let $\Sigma$ satisfy $RE(s_0, 3k_0, \Sigma^{1/2})$ condition. Set as before with $A$ replaced by $\Sigma^{1/2}$. Assume that $d \leq p$ and $\rho = \rho_{\min}(d, \Sigma^{1/2}) > 0$. If for some absolute constant $C$

$$n \geq \frac{CM^2 d \log p}{\rho \delta^2} \log^3 \left( \frac{CM^2 d \log p}{\rho \delta^2} \right)$$

Theorem 8

If the above condition holds then with probability at least

$$1 - \exp(-\delta \rho n/(6M^2 d)),$$

$RE(s_0, k_0, X)$ condition holds for matrix $\frac{1}{\sqrt{n}}X$ with

$$0 < K \left( s_0, k_0, \frac{1}{\sqrt{n}}X \right) \leq \frac{K(s_0, k_0, \Sigma^{1/2})}{1 - \delta}$$
Theorem 22

Under the conditions mentioned in the previous slide with probability as least $1 - \exp(-\delta\rho n/(6M^2d))$, all vectors $u \in C(s_0, k_0)$ satisfy

$$(1 - \delta)\|u\|_2 \leq \frac{\|Xu\|_2}{\sqrt{n}} \leq (1 + \delta)\|u\|_2$$

Theorem 23

Under the above condition with probability at least $1 - 2\exp\left(-\frac{\epsilon\rho n}{6M^2m}\right)$, all $m$-sparse vectors $u$ satisfy

$$1 - \delta \frac{1}{\sqrt{n}} \left\| \frac{Xu}{\|\Sigma^{1/2}u\|_2} \right\|_2 \leq 1 + \delta$$

- Consider $F = \bigcup_{|J|=m} \Sigma^{1/2} E_J \cap S^{p-1}$, $y \in F$
- Estimate $\Delta := E \sup_{y \in F} \left| 1 - \frac{1}{n} \sum_{j=1}^{n} \langle \psi_j, y \rangle^2 \right|$
- Use Talagrand’s measure concentration theorem for empirical processes to derive large deviation estimate
Concluding remarks

- The reduction principle can be used for any matrix $X = \Psi A$. Examples include random vectors with heavy-tailed marginals, random vectors with log-concave densities.

- For sub-Gaussian design matrices the theorem does not involve any condition on $\rho_{\text{max}}(s_0, A)$ nor any of the global parameters of the $A$ and $\Psi$ matrix.

- The estimate of Theorem 23 contains the minimal sparse singular value $\rho$, which cannot be avoided.
Thank you