Data Reconstruction Algorithms: Stability

Yunbin Zhao[1]
(with H. Jiang, Cambridge; Z.-Q. Luo, CHUK (SZ))

[1]School of Mathematics, University of Birmingham, United Kingdom

Data sparse representation

Let $b \in \mathbb{R}^m$ be the given data which can be generated by known redundant bases:

$$a_i \in \mathbb{R}^m, \quad i = 1, \dots, n \quad (m \ll n).$$

There are many ways to represent the data:

$$a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = b,$$

i.e.,

$$Ax = b$$

where $A = [a_1, a_2, ..., a_n]$ is an $m \times n$ matrix. We look for the simplest one[2], i.e., the sparsest solution.

[2] "when we are facing a number of choices, the simplest choice is the best"

Under the sparsity assumption, many practical applications, for instance,

- compressed sensing
- signal and image processing
- statistical regression
- machine learning
- pattern recognition, etc.

can be formulated to the problem of seeking the sparse or the sparsest point of a certain set, leading to the so-called sparse optimization problem.[3]

[3] "Entities must not be multiplied beyond necessity" – William of Ockham (1287–1347)

Sparsity optimization Problem

ℓ_0-minimization:

$$\min\{\|x\|_0 : Ax = b\},$$

which seeks the simplest (sparsest) solution of linear system.[4]

In more general

$$\min\{\|x\|_0 : x \in P\}$$

where P is a convex set.

- ℓ_0**-problem is NP-hard** (Natarajan 1995).

[4] "Simplicity is the final achievement" – Frederic Chopin

How to solve the ℓ_0-problem?

- Continuous approximation (typically, concave minimization)
- Orthogonal matching pursuit
- Compressive sampling matching pursuit, or subspace pursuit
- Thresholding-type method
- ℓ_1-minimization
- Reweighted ℓ_1-methods
- LASSO and Dantzig selector

ℓ_1-minimizaton

$$\text{Minimize } \{\|x\|_1 : Ax = b\}$$

where

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|$$

Relevant models:

- Quadratically constrained basis pursuit:

$$\min\{\|x\|_1 : \|Ax - y\|_2 \le \varepsilon\}$$

- LASSO:

$$\min\{\|Ax - y\|_2 : \|x\|_1 \le \mu\} \text{ or } \min \lambda\|x\|_1 + \|Ax - y\|_2^2.$$

- The Dantzig selector:

$$\min\{\|x\|_1 : \|A^T(Ax - y)\|_\infty \le \varepsilon\}$$

Stability: Background

- The presence of noise in data is unavoidable due to **sensor imperfection**, **estimation inaccuracy**, **statistical or communication errors**, and **quantization errors**.
  For instance, signals might be contaminated by some form of random noise and the measurements of signals are subject to quantization error.
- A huge effort in sparse data recovery is made to ensure the recovery methods stable in the sense that recovery errors stay under control when the measurements are slightly inaccurate and when the data is not exactly sparse.

Definition. (Noise-free stability)

Let x^* be the solution of a reconstruction algorithm based on the accurate measurement

$$y := A\widetilde{x}$$

where $\widetilde{x}$ is the target k-sparse signal.

Definition. (Noise-free stability)

Let x^* be the solution of a reconstruction algorithm based on the accurate measurement

$$y := A\widetilde{x}$$

where $\widetilde{x}$ is the target k-sparse signal.

The algorithm is said to be stable (in noise-free cases) if x^* approximates any x satisfying $y = Ax$ with error

$$\|x - x^*\|_2 \leq C\sigma_k(x)_1, \tag{1}$$

where

$$\sigma_k(x)_1 = \min_{\|u\|_0 \leq k} \|x - u\|_1.$$

Definition. (Robust stability)

Let x^* be the solution of a reconstruction algorithm based on the inaccurate measurement

$$y = A\widetilde{x} + u$$

where $\|u\|_2 \leq \varepsilon$ and $\widetilde{x}$ is the target k-sparse signal.

The algorithm is said to be stable (or robustly stable) if the solution x^* produced by the algorithm approximates any x satisfying $\|y - Ax\|_2 \leq \varepsilon$ with error

$$\|x - x^*\|_2 \leq C_1 \sigma_k(x)_1 + C_2 \varepsilon. \tag{2}$$

Traditional Conditions for Stability

(a) **(RIP of order K)** [E. Candés & T. Tao (2005)]

The matrix A is said to satisfy the restricted isometry property of order K with constant $\delta_K \in (0,1)$ if

$$(1-\delta_K)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta_K)\|x\|_2^2$$

holds for all K-sparse vector $x \in \mathbb{R}^n$.

(b) **(NSP of order k)** [A. Cohen, W. Dahmen, R.DeVore (2009), Y. Zhang (2005)]

A is said to satisfy the null space property of order k if

$$\|v_S\|_1 < \|v_{\overline{S}}\|_1$$

holds for any nonzero $v \in \mathcal{N}(A)$ and any $S \subseteq \{1,\ldots,n\}$ with $|S| \leq k$.

Traditional Conditions (continued)

(c) **(Stable NSP of order k)** [A. Cohen et al. (2009), Foucart & Rauhut (2013)]

A is said to satisfy the stable null space property of order k with constant $\rho \in (0,1)$ if

$$\|v_S\|_1 \leq \rho \|v_{\overline{S}}\|_1$$

holds for any $v \in \mathcal{N}(A)$ and any $S \subseteq \{1, \ldots, n\}$ with $|S| \leq k$.

(d) **(Robust NSP of order k)** [A. Cohen et al. (2009), Foucart & Rauhut (2013)]

A is said to satisfy the robust null space property of order k with constants $\rho \in (0,1)$ and $\tau > 0$ if

$$\|v_S\|_1 \leq \rho \|v_{\overline{S}}\|_1 + \tau \|Av\|$$

holds for any $v \in \mathbb{R}^n$ and any $S \subseteq \{1, \ldots, n\}$ with $|S| \leq k$.

Disadvantage: sufficient but restrictive reconditions

- The standard CS stability theory does not assume the structure or other prior information about the locations of the nonzero entries of the signals.
- Recent study (B. Adcock, et al. (2014)) indicates that the standard sparsity assumption is actually hard to meet in practice since the signal is often structured or with prior information.
- The typical measurement matrices in practice is not Gaussian or Bernoulli, but one with a specific structures (such as likelihood of certain molecule configuration, minimum distance between sparse coefficients due to some repelling force, etc)

What is inspired from Optimality Conditions?

Theorem (Optimality Condition). $\widehat{x} \in \mathbb{R}^n$ is the optimal solution to the problem

$$\min\{\|z\|_1 : Az = A\widehat{x}\}$$

if and only if there is a vector $\eta \in \mathcal{R}(A^T)$ such that

$$\eta_i = 1 \text{ for } \widehat{x}_i > 0,$$

$$\eta_i = -1 \text{ for } \widehat{x}_i < 0,$$

$$|\eta_i| \leq 1 \text{ for } \widehat{x}_i = 0.$$

Weak RSP of order k of A^T [Zhao, Jiang & Luo (2017)]

The matrix A^T is said to satisfy the weak range space property of order k if for any disjoint subsets S_1, S_2 of $\{1, \ldots, n\}$ with $|S_1| + |S_2| \leq k$, there is a vector $\eta \in \mathcal{R}(A^T)$ satisfying that

$$\eta_i = 1 \text{ for } i \in S_1,$$

$$\eta_i = -1 \text{ for } i \in S_2,$$

$$|\eta_i| \leq 1 \text{ for } i \notin S_1 \cup S_2.$$

[5]

[5] "Everything should be made as simple as possible, but not simpler" –Albert Einstein

RSP of order k of A^T [Zhao (2013)]

The matrix A^T is said to satisfy the range space property of order k if for any disjoint subsets S_1, S_2 of $\{1, \ldots, n\}$ with $|S_1| + |S_2| \leq k$, there is a vector $\eta \in \mathcal{R}(A^T)$ satisfying that

$$\eta_i = 1 \text{ for } i \in S_1,$$

$$\eta_i = -1 \text{ for } i \in S_2,$$

$$|\eta_i| < 1 \text{ for } i \notin S_1 \cup S_2.$$

Existence of RSP matrix

- Let A be an $m \times n$ $(m \ll n)$ Gaussian random matrix. Then there exists a universal constant $C > 0$ such that

$$A^T/\sqrt{m}$$

satisfies the weak RSP of order k with probability at least $1 - \zeta$ provided

$$m \geq 2C\left(k + k\ln(n/k) + \ln(2\zeta^{-1})\right). \tag{3}$$

Compared with traditional conditions

$$\left.\begin{array}{r} \text{RIP of order } 2k \Rightarrow \\ \text{Stable NSP of order } k \Rightarrow \\ \text{Robust NSP of order } k \Rightarrow \\ \mu_1(k) + \mu_1(k-1) < 1 \Rightarrow \end{array}\right\} \begin{array}{lcl} \text{NSP of order } k & \Leftrightarrow & \text{RSP of order } k \text{ of } A^T \\ & \Rightarrow & \text{weak RSP of order } k \end{array}$$

The weak RSP is the mildest one among the above matrix conditions.

Necessity of RSP

Theorem [Zhao, Jiang and Luo (2017)].

Let $A \in \mathbb{R}^{m \times n}$ *be a given matrix with* $m < n$ *and* $\mathrm{rank}(A) = m$. *Suppose that for any given*

$$y \in \{Ax : \|x\|_0 \le k\},$$

the following holds: the solution x^* *of the* ℓ_1*-minimization problem*

$$\min_z \{\|z\|_1 : \; Az = y\}$$

approximates x *satisfying* $Ax = y$ *with error*

$$\|x - x^*\| \le C\sigma_k(x)_1.$$

Then A^T *must satisfy the weak RSP of order* k.

How to establish sufficiency?

Hoffman Lemma + solution set structure + RSP matrix property.

Lemma 2.4. [Hoffman (1952)] *Let $M' \in \mathbb{R}^{m \times q}$ and $M'' \in \mathbb{R}^{\ell \times q}$ be given matrices and*

$$\mathcal{F} = \{x \in \mathbb{R}^q : M'x \leq b, M''x = d\}.$$

For any vector x in $\mathbb{R}^q$, there is a point $x^* \in \mathcal{F}$ with

$$\|x - x^*\|_2 \leq \sigma_{\infty,2}(M', M'') \left\| \begin{bmatrix} (M'x - b)^+ \\ M''x - d \end{bmatrix} \right\|_1 .$$

Sufficiency of RSP

Theorem Let A be an $m \times n (m < n)$ matrix and $\mathrm{rank}(A) = m$. Let y be any given vector in $\mathbb{R}^m$. If A^T satisfies the weak RSP of order k, then for any $x \in \mathbb{R}^n$, the solution x^* of ℓ_1-minimization approximates x with

$$\|x - x^*\|_2 \leq 2\gamma\sigma_k(x)_1 + (1 + c)\gamma\|Ax - y\|_1,$$

where

$$c = \|(AA^T)^{-1}A\|_{\infty \to \infty},$$

and γ is the Robinson's constant depending on the problem data.

Moreover, for any x satisfying $Ax = y$, there is a solution x^* to the ℓ_1-problem such that

$$\|x - x^*\|_2 \leq 2\gamma\sigma_k(x)_1.$$

Stability Theorem for ℓ_1-Algorithm

Let $A \in \mathbb{R}^{m\times n}(m < n)$ be a matrix with $rank(A) = m$. Then the standard ℓ_1-minimization problem

$$\min\{\|x\|_1 : Ax = y\}$$

is stable in sparse data reconstruction for any given measurements

$$y \in \{Ax : \|x\|_0 \le k\}$$

if and only if A^T *satisfies the weak RSP of order k.*

Unified Stability Theorem for ℓ_1-Algorithm

Given (A, y) , if A admits one of the following properties:

(a) *RIP of order $2k$ with $\delta_{2k} < 1/\sqrt{2}$.*

(b) *ℓ_2-normalized A satisfying $\mu_1(k) + \mu_1(k-1) < 1$.*

(c) *Stable NSP of order k with constant $0 < \rho < 1$.*

(d) *Robust NSP of order k with $0 < \rho < 1$ and $\tau > 0$.*

(e) *NSP of order k.*

(f) *RSP of order k of A^T.*

Then the solution x^* of ℓ_1-minimization approximates $x \in \mathbb{R}^n$ with

$$\|x - x^*\|_2 \leq 2\gamma\sigma_k(x)_1 + \gamma(1 + c)\|Ax - y\|_1.$$

In particular, the solution x^* of ℓ_1-minimization approximates x satisfying $Ax = y$ with

$$\|x - x^*\|_2 \leq 2\gamma\sigma_k(x)_1.$$

General Dantzig Selector (DS)

$$\text{(GDS)} \qquad \min_x \{\|x\|_1 : \ \phi(M^T(Ax - y)) \le \varepsilon\}.$$

Consider the norm

$$\phi(\cdot) := \alpha\|\cdot\|_\infty + (1-\alpha)\|\cdot\|_1,$$

where $\alpha \in [0,1]$ is a fixed constant. The problem becomes

$$\min_x \{\|x\|_1 : \ \alpha\|M^T(Ax-y)\|_\infty + (1-\alpha)\|M^T(Ax-y)\|_1 \le \varepsilon\}, \quad (4)$$

Two spacial cases:

$$\text{(Standard DS)} \quad \min_x \{\|x\|_1 : \ \|A^T(Ax - y)\|_\infty \le \varepsilon\},$$

$$\min_x \{\|x\|_1 : \ \|A^T(Ax - y)\|_1 \le \varepsilon\}.$$

The standard DS was introduced by E. Candés and T. Tao in 2007.

Stability Theorem for DS

Let $A \in \mathbb{R}^{m\times n}$ and $M \in \mathbb{R}^{m\times q}$ be full-rank matrices with $m < n$ and$m \leq q$. If A^T satisfies the weak RSP of order k, then there is a solution x^* of the DS (4) approximating $x \in \mathbb{R}^n$ with error

$$\|x - x^*\|_2 \leq \gamma[\alpha\|M^T(Ax-y)\|_\infty + (1-\alpha)\|M^T(Ax-y)\|_1 - \varepsilon]^+ \\ + \gamma\left\{2\sigma_k(x)_1 + \widehat{c}\varepsilon + \widehat{c}\|M^T(Ax-y)\|_\infty\right\},$$

where γ is the Robinson's constant

$$\widehat{c} = \max_{G\subseteq\{1,\ldots,q\},|G|=m} \|M_G^{-1}(AA^T)^{-1}A\|_{\infty\to 1}.$$

In particular, for any x feasible to (4), then

$$\|x - x^*\|_2 \leq 2\gamma\left\{\sigma_k(x)_1 + \widehat{c}\varepsilon\right\}.$$

Special Cases: $M = A$ and $\alpha = 1$ in (4)

.

Theorem (Standard DS). Consider problem

$$\min_x \{\|x\|_1 : \ \|A^T(Ax - y)\|_\infty \leq \varepsilon\}.$$

Then the following two statements are equivalent:

(i) A^T satisfies the weak RSP of order k.

(ii) For any x satisfying $\|A^T(Ax - y)\|_\infty \leq \varepsilon$, there is a solution x^* of the standard DS approximating x with error

$$\|x - x^*\|_2 \leq 2\gamma \left\{\sigma_k(x)_1 + c_A\varepsilon\right\},$$

where γ is the Robinson's constant and

$$\widehat{c}_A = \max_{G \subseteq \{1,\ldots,q\}, |G| = m} \|A_G^{-1}(AA^T)^{-1}A\|_{\infty \to 1}.$$

Unified Stability for Dantzig Selector.

Given (A, y), if one of the following properties holds:

- *RIP of order $2k$ with $\delta_{2k} < 1/\sqrt{2}$.*
- *ℓ_2-normalized A satisfying $\mu_1(k) + \mu_1(k-1) < 1$.*
- *Stable NSP of order k with constant $0 < \rho < 1$.*
- *Robust NSP of order k with $0 < \rho < 1$ and $\tau > 0$.*
- *NSP of order k.*
- *RSP of order k of A^T.*

Then, for any x obeying $\|A^T(Ax - y)\|_\infty \leq \varepsilon$, there is a solution x^* of the standard DS satisfying

$$\begin{aligned} \|x - x^*\|_2 &\leq \gamma \left\{2\sigma_k(x)_1 + c_A\varepsilon + c_A\|A^T(Ax - y)\|_\infty\right\} \\ &\leq 2\gamma\left\{\sigma_k(x)_1 + c_A\varepsilon\right\}. \end{aligned}$$

Stability ℓ_1-minimization with ℓ_∞-norm constraint

$$\min\{\|x\|_1 : \ \|Ax - y\|_\infty \leq \varepsilon\}. \tag{5}$$

Theorem. Let A be an $m \times n$ matrix with $m < n$ and $\mathrm{rank}(A) = m$. If A^T satisfy the weak RSP of order k, then for any $x \in \mathbb{R}^n$, there is a solution x^* of (5) such that

$$\|x - x^*\|_2 \leq \gamma_1 \left\{2\sigma_k(x)_1 + (\|Ax - y\|_\infty - \varepsilon)^+ + c^*\varepsilon + c^*\|Ax - y\|_\infty\right\},$$

where c^* and γ_1 are constants dependent on the problem data. In particular, for any x with $\|Ax - y\|_\infty \leq \varepsilon$, there is a solution x^* of (5) such that

$$\|x - x^*\|_2 \leq 2\gamma_1\{\sigma_k(x)_1 + c^*\varepsilon\}.$$

Stability for ℓ_1-Minimization with ℓ_1-Norm Constraints

$$\min_x \{\|x\|_1 : \ \|Ax - y\|_1 \leq \varepsilon\}, \tag{6}$$

Theorem. Let $A \in \mathbb{R}^{m\times n}$ with $m < n$ and $\mathrm{rank}(A) = m$. Let A^T satisfy the weak RSP of order k. Then for any $x \in \mathbb{R}^n$, there is a solution x^* of (6) such that

$$\|x - x^*\|_2 \leq \gamma_2 \left\{2\sigma_k(x)_1 + (\|Ax - y\|_1 - \varepsilon)^+ + c^*(\varepsilon + \|Ax - y\|_1)\right\},$$

where c^* and γ_2 are the Robinson's constants determined by the problem data. In particular, for any x with $\|Ax - y\|_1 \leq \varepsilon$, there is a solution x^* of (6) such that

$$\|x - x^*\|_2 \leq 2\gamma_2\{\sigma_k(x)_1 + c^*\varepsilon\}.$$

Nonlinear Models: RSP Remains Sufficient for Stability

- Quadratically constrained basis pursuit:

$$\min\{\|x\|_1 : \|Ax - y\|_2 \leq \varepsilon\}$$

- LASSO (least absolute shrinkage and selection operator)

$$\min\{\|Ax - y\|_2 : \|x\|_1 \leq \mu\} \text{ or } \min \lambda\|x\|_1 + \|Ax - y\|_2^2.$$

- Nonlinear Dantzig selector:

$$\min\{\|x\|_1 : \phi(A^T(Ax - y)) \leq \varepsilon\}$$

where $\phi(\cdot)$ is a nonlinear norm.

Polytope Approximation of Convex Sets

Approximation of ℓ_2-ball:

$$\mathfrak{B} = \{z \in \mathbb{R}^m : \ \|z\|_2 \leq 1\}$$

Dudely's Theorem (1974). There exists a constant τ such that for every integer number $\kappa > m$ there is a polytope

$$\mathcal{P}_\kappa = \bigcap_{\|a^i\|_2=1, 1\leq i\leq \kappa} \left\{z \in \mathbb{R}^m : \ (a^i)^T z \leq 1\right\} \tag{7}$$

containing $\mathfrak{B}$ in $\mathbb{R}^m$ and satisfying

$$d^{\mathcal{H}}(\mathfrak{B}, \mathcal{P}_\kappa) \leq \frac{\tau}{\kappa^{2/(m-1)}}, \tag{8}$$

where $d^{\mathcal{H}}(\cdot, \cdot)$ is the Hausdorff metric.

Polytope Approximation of Convex Sets

Approximation of general ball:

$$\mathfrak{B}^{\phi} = \{z \in \mathbb{R}^q : \ \phi(z) \leq 1\}.$$

Barvinok's Theorem (2014) For any constant $\chi > \frac{e}{4\sqrt{2}} \approx 0.48$, there exists a $\epsilon_0 = \epsilon_0(\chi)$ such that for any $0 < \epsilon < \epsilon_0$ and for any symmetric convex body B in $\mathbb{R}^q$, there is a symmetric polytope in $\mathbb{R}^q$, denoted by P_ϵ, with N vertices such that

$$N \leq \left(\frac{\chi}{\sqrt{\epsilon}} \ln \frac{1}{\epsilon} \right)^q$$

and

$$P_\epsilon \subseteq B \subseteq (1+\epsilon) P_\epsilon.$$

Conclusions

- RSP is natural recovery conditions arising from strengthened KKT optimality condition.
- The RSP captures the intrinsic mechanism of signal reconstruction via convex optimization decoding methods.
- RSP is both necessary and sufficient to ensure stable reconstruction of sparse signals for L1-minimization and standard Dantzig Selector decoding algorithms.
- Hoffman'error bound provides a unified analytic tool for undertaking stability analysis of many data reconstruction algorithms under the RSP assumption.

Main References

- E. Candès and T. Tao, Decoding by linear programming, *IEEE Trans. Inform. Theory*, 51 (2005), pp. 4203–4215.
- A. Cohen, W. Dahmen and R. Devore, Compressed sensing and best k-term aproximation, *J. Amer. Math. Soc.*, 22 (2009), pp. 211–231.
- Y. Zhao and D. Li, Reweighted ℓ_1-minimization for sparse solutions to underdetermined linear systems, *SIAM Journal on Optimization*, 22 (2012), 1065-1088.
- Y. Zhao, RSP-Based analysis for sparest and least ℓ_1-norm solutions to underdetermined linear systems, *IEEE Transactions on Signal Processing*, 61 (2013), no.22, 5777-5788.
- Y. Zhao, H. Jiang and Z.-Q. Luo, Weak stability of l1-minimization methods in sparse data reconstruction, *Mathematics of Operations Research*, 44 (2019), no.1, 173-195.
- Y. Zhao, (Book) *"Sparse Optimization Theory and Methods"*, CRC Press, Taylor & Francis Group, 2018.