From Linear Regression Without Correspondences to Homomorphic Sensing

by Liangzu Peng*

School of Information Science and Technology
ShanghaiTech University

May 10, 2021

Email: penglz@shanghaitech.edu.cn

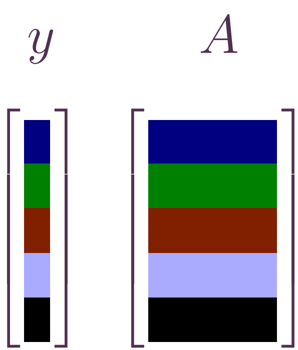
*. Since September 2017, Liangzu Peng has became a graduate student at ShanghaiTech. About one year later he received supervision from Dr. Manolis C. Tsakiris. This has lasted for at least 2.51 years. May 10, 2021 is the day when he presents his research over this period.

1. M. C. Tsakiris, **L. Peng**, A. Conca, L. Kneip, Y. Shi, and H. Choi, "An algebraic-geometric approach to linear regression without correspondences," in IEEE Transactions on Information Theory, vol. 66, no. 8, pp. 5130-5144, 2020.
2. **L. Peng** and M. C. Tsakiris, "Linear regression without correspondences via concave minimization," in IEEE Signal Processing Letters, vol. 27, pp. 1580-1584, 2020.
3. **L. Peng** and M. C. Tsakiris, "Homomorphic sensing of subspace arrangements," Applied and Computational Harmonic Analysis (under revision).
4. **L. Peng**, B. Wang, M. C. Tsakiris, "Homomorphic sensing: sparsity and noise," ICML 2021.
5. M. C. Tsakiris and **L. Peng**, "Homomorphic sensing," ICML 2019.
6. Y. Yao, **L. Peng**, and M. C. Tsakiris, "Unsigned matrix completion," ISIT 2021.
7. Y. Yao, **L. Peng**, and M. C. Tsakiris, "Unlabeled principal component analysis," arXiv:2101.09446v1 [cs.LG], 2021.
8. **L. Peng**, X. Song, M. C. Tsakiris, H. Choi, L. Kneip, and Y. Shi, "Algebraically-initialized expectation maximization for header-free communication," ICASSP 2019.

Linear Regression

$y = Ax_{\text{gt}} \in \mathbb{R}^m \quad x_{\text{gt}} \in \mathbb{R}^n$

Goal: find x_{gt} from A, y

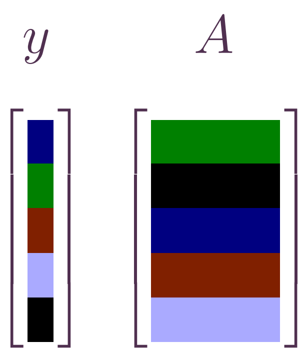


Linear Regression Without Correspondences[1]

unknown permutation

$y = \Pi_{\text{gt}} A x_{\text{gt}}$

Goal: find x_{gt} (and Π_{gt}) from A, y



1. J. Unnikrishnan, S. Haghighatshoar and M. Vetterli, "Unlabeled sensing: Solving a linear system with unordered measurements," Allerton 2015.

$y = \Pi_{\mathrm{gt}} A x_{\mathrm{gt}}$ $\qquad \Pi_{\mathrm{gt}} \in \mathcal{S}_m$ $\qquad A \in \mathbb{R}^{m \times n}$

Formulation: the set of $m \times m$ permutations

$$\min_{x \in \mathbb{R}^n, \Pi \in \mathcal{S}_m} \|y - \Pi A x\|_2 \tag{1}$$

If Π is known

(1) is a **least-squares** problem $\qquad \mathcal{O}(mn^2)$

If x is known

(1) is (reduced to) a **sorting** problem $\qquad \mathcal{O}(m \log m)$

$$\min_{x\in\mathbb{R}^n,\Pi\in\mathcal{S}_m} \|y-\Pi Ax\|_2 \tag{2}$$

If $n=1$

solve (2) by sorting $\mathcal{O}(m\log m)$

If $n>1$ (the case of interest)

solving (2) is NP-hard[2]

2. A. Pananjady, M. J. Wainwright and T. A. Courtade, "Linear regression with shuffled data: statistical and computational limits of permutation recovery," in IEEE Transactions on Information Theory, vol. 64, no. 5, pp. 3286-3300, 2018.

An Algebraic-Geometric Approach[3]

to

$$\min_{x \in \mathbb{R}^n, \Pi \in \mathcal{S}_m} \|y - \Pi A x\|_2$$

The **AI-EM** algorithm:

compute an (**a**lgebraic) **i**nitialization $x^{(0)}$

via solving a system of polynomial equations

expectation **m**aximization (alternating minimizaion)

3. M. C. Tsakiris, **L. Peng**, A. Conca, L. Kneip, Y. Shi, and H. Choi, “An algebraic-geometric approach to linear regression without correspondences", in IEEE Transactions on Information Theory, vol. 66, no. 8, pp. 5130-5144, 2020.

AI: Power-Sum Polynomials

$z = [z_1, \ldots, z_m]^\top$: a vector of variables of size m

$$\underbrace{\begin{array}{ccccc} z_1 & z_2 & z_3 & \cdots\cdots & z_m \\ \downarrow & \downarrow & \downarrow & \cdots\cdots & \downarrow \\ z_1^s & z_2^s & z_3^s & \cdots\cdots & z_m^s \end{array}}_{+}$$

power-sum polynomials[4] $p_s(z) = z_1^s + \cdots + z_m^s$

4. X. Song, H. Choi, and Y. Shi, "Permuted linear model for header-free communication via symmetric polynomials," ISIT 2018.

$p_s(z) = z_1^s + z_2^s + \cdots + z_m^s$ is permutation-invariant:

$$p_s(z) = p_s(\Pi z) \qquad \text{for any } \Pi \in \mathcal{S}_m$$

This allows to eliminate Π_{gt}:

$$p_s(y) = p_s(\Pi_{\text{gt}} A x_{\text{gt}}) = p_s(A x_{\text{gt}})$$

Hence $x_{\text{gt}} \in \mathbb{R}^n$ satisfies

$$p_s(Ax) - p_s(y) = 0 \qquad \forall s \in \{1, 2, \ldots, n\}$$

n variables

n equations

Theorem 1. *For $A \in \mathbb{R}^{m \times n}$ generic ($m > n$), the polynomial equations*

$$p_s(Ax) - p_s(y) = 0 \qquad \forall s \in \{1, 2, \ldots, n\}$$

have no more than $n!$ complex solutions, and one of them is x_{gt}.

With the noisy measurements $\tilde{y} = y + \epsilon \in \mathbb{R}^m$, we have:

Theorem 2. *For $A \in \mathbb{R}^{m \times n}$ generic ($m > n$), the polynomial equations*

$$p_s(Ax) - p_s(\tilde{y}) = 0 \qquad \forall s \in \{1, 2, \ldots, n\}$$

have no more than $n!$ complex solutions, $\boxed{\xi_1, \ldots, \xi_L}$.

Moreover, there is a solution whose real part has its distance to x_{gt} bounded above by a quantity polynomial in entries of ϵ.

In particular, this quantity is zero in the absence of noise.

AI: Extracting a good solution from $\xi_1, \ldots, \xi_L$

Approach 1:

"best" root   unknown!

$$\xi_{\mathrm{AI}}^* := \mathrm{argmin}_{\xi_i, i \in \{1,\ldots,L\}} \|x_{\mathrm{gt}} - (\xi_i)_{\mathbb{R}}\|_2$$

real part of ξ_i

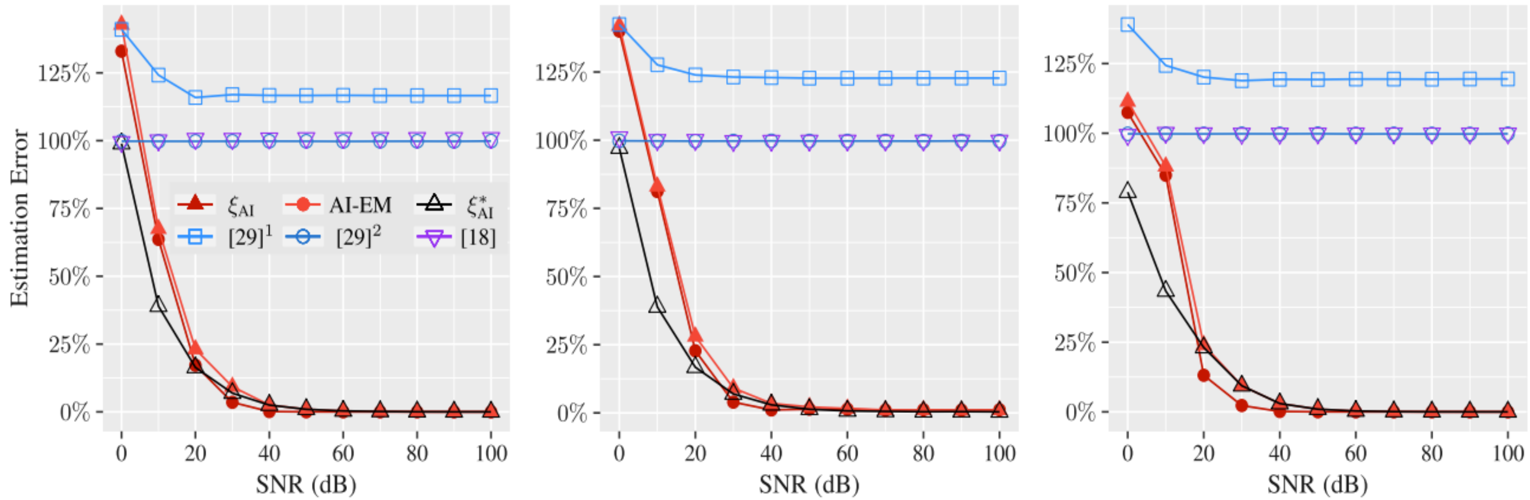
Approach 2:

$$\xi_{\mathrm{AI}} := \mathrm{argmin}_{\xi_i, i \in \{1,\ldots,L\}} \left\{ \min_{\Pi \in \mathcal{S}_m} \|\tilde{y} - \Pi A (\xi_i)_{\mathbb{R}}\|_2 \right\}$$

computable root   can be solved by sorting!

Use $(\xi_{\mathrm{AI}})_{\mathbb{R}}$ as initialization $\mathcal{O}(L \cdot m \log m)$

AI-EM: Experiments



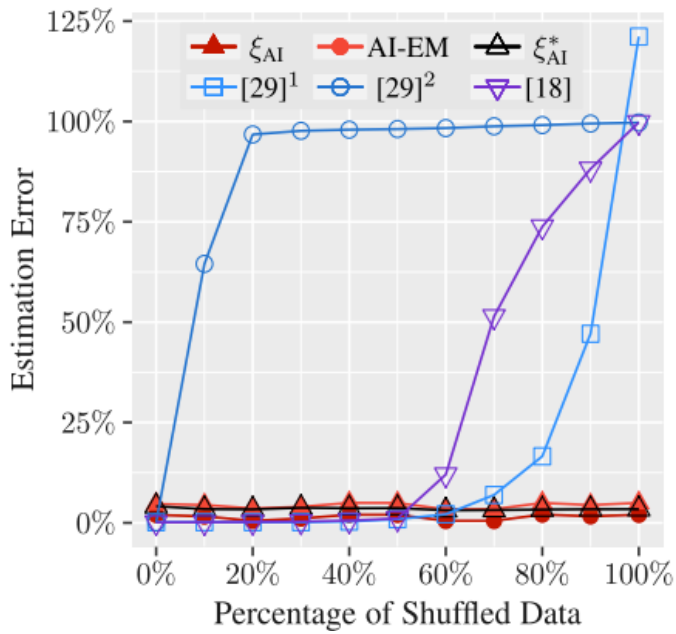
(a) $n = 3$ (b) $n = 4$ (c) $n = 5$

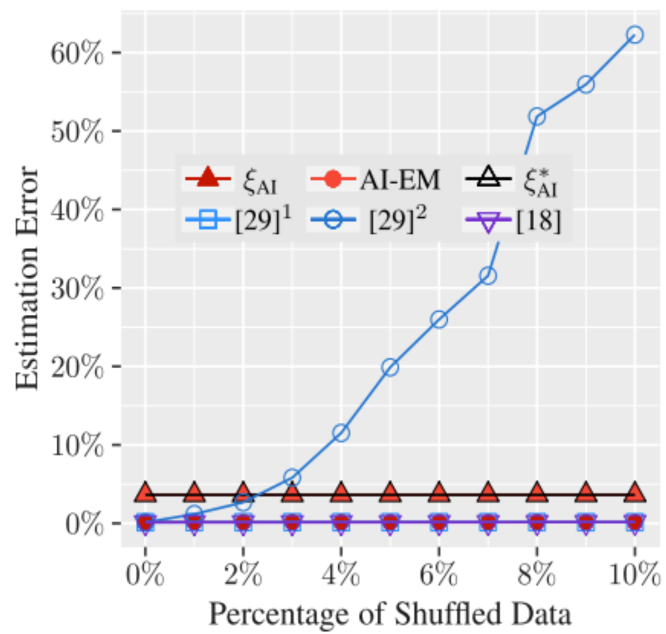
Estimation Error: $= \frac{\|x_{\mathrm{gt}} - x'\|_2}{\|x_{\mathrm{gt}}\|_2}$ $x' :=$ output of the algorithm

$[29]^1$: EM with least-squares initialization

[29] A. Abid and J. Zou, “A stochastic expectation-maximization approach to shuffled linear regression,” Allerton 2018. $[29]^2$: stochastic EM with least-squares initialization

[18] M. Slawski and E. Ben-David, “Linear regression with sparsely permuted data,” Electronic Journal of Statistics, vol. 13, no. 1, pp. 1–36, 2019. ℓ_1 robust regression





Percentage of Shuffled Data $\alpha :=$ (the number of elements Π_{gt} permutes)$/m$

EM

[29]^1: works for $\alpha \leqslant 60\%$

[18]: works for $\alpha \leqslant 50\%$

ℓ_1 robust regression

stochastic EM

[29]^2: works for $\alpha \leqslant 2\%$

AI-EM: works for all α

Running Time (milliseconds)

$m = 500$

	Running Time		Estimation Error		
n	$\tilde{\mathcal{P}}$	AM	ξ^*_{AI}	ξ_{AI}	AI-EM
3	0.7	6	2.3%	2.3%	0.1%
4	9	6	3.9%	3.9%	0.1%
5	45,157	7	1.4%	1.4%	0.1%
6	$\boxed{2,243,952}$	7	1.2%	1.2%	0.1%

>30min

exponential increase

$n = 4$

	Running Time		Estimation Error		
m	$\tilde{\mathcal{P}}$	AM	ξ^*_{AI}	ξ_{AI}	AI-EM
1,000	10	15	3.2%	3.3%	0.4%
5,000	21	105	3.5%	4.8%	2.5%
10,000	32	281	4.1%	5.9%	4.3%
25,000	66	357	3.4%	5.2%	4.4%
50,000	126	613	3.7%	6.0%	5.5%
100,000	268	1,271	4.1%	6.8%	6.5%

linear increase

For $n \geq 7$, the AI-EM approach is intractable.

A Concave Minimization Approach[5]

to

$$\min_{x \in \mathbb{R}^n, \Pi \in \mathcal{S}_m} \|\Pi \tilde{y} - Ax\|_2$$

5. **L. Peng** and M. C. Tsakiris, "Linear regression without correspondences via concave minimization," in IEEE Signal Processing Letters, vol. 27, pp. 1580-1584, 2020.

$K \in \mathbb{R}^{n \times m^2}$

After some algebraic manipulation. . . :

computable from data

$$\min_{x \in \mathbb{R}^n, \Pi \in \mathcal{S}_m} \|\Pi \tilde{y} - Ax\|_2 \Leftrightarrow \min_{\Pi \in \mathcal{S}_m} -\|K \operatorname{vec}(\Pi)\|_2^2$$

x eliminated!

Concave Minimization Reformulation

Proposition 3. *If*

$$\hat{\Pi} = \operatorname{argmin}_{\Pi \in \mathcal{S}_m} - \|K \operatorname{vec}(\Pi)\|_2^2,$$

then

$$\hat{\Pi} = \operatorname{argmin}_{B \in \operatorname{conv}(\mathcal{S}_m)} - \|K \operatorname{vec}(B)\|_2^2.$$

convex constraint

concave objective

$\operatorname{conv}(\mathcal{S}_m) :=$ the convex hull of $\mathcal{S}_m$, also known as

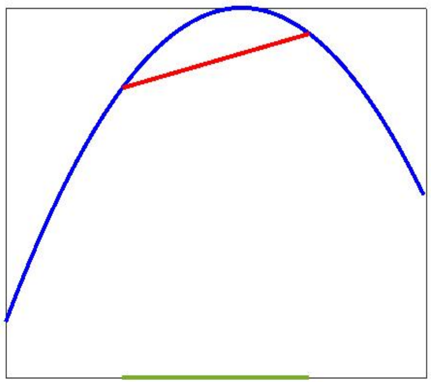
the set of doubly stochastic matrices

the Birkhoff polytope

in general NP hard

$$\min_{B \in \mathrm{conv}(\mathcal{S}_m)} -\|K\,\mathrm{vec}(B)\|_2^2$$

A concave function and its convex envelope over some sub-interval:



global optimum guarantee!

Branch & Bound:

branching over the smallest hyper-rectangle that contains $\mathrm{conv}(\mathcal{S}_m)$

tight lower bounds from minimizing its convex envelope

	Running Time			
n	[4]-A	[4]-B	[2]	CCV-Min
3	0.48sec	37sec	3msec	0.42sec
4	5sec	17min	7msec	2.43sec
5	> 12hr	> 12hr	43sec	7.16sec
6			37min	72.5sec
7			> 12hr	6min
8				40min
9				> 12hr

[4]: M. C. Tsakiris and **L. Peng**, "Homomorphic sensing," ICML 2019.

[4]-A: branch & bound

[4]-B: RANSAC type algorithm

[2]: the AI-EM algorithm

CCV-Min: the concave minimization approach

$$y = \Pi A x \qquad \Pi \in \mathcal{S}_m \qquad x \in \mathbb{R}^n \tag{3}$$

What we have done so far:

assume the existence of a solution x_{gt} (with Π_{gt}) to (3)

e.g., $y = \Pi_{\mathrm{gt}} A x_{\mathrm{gt}}$

eliminate the unknown $\Pi \implies$ AI-EM

eliminate the unknown $x \implies$ CCV-Min

What we have not discussed:

Is x_{gt} a **unique solution** to (3)?[6]

6. J. Unnikrishnan, S. Haghighatshoar, and M. Vetterli, "Unlabeled sensing: Solving a linear system with unordered measurements," IEEE Transactions on Information Theory, 2018.

$$y = TAx \qquad T \in \mathcal{T} \qquad x \in \mathbb{R}^n \tag{$*$}$$

a finite set of *arbitrary* matrices

Assume the existence of a solution x_{gt} (with T_{gt}) to $(*)$:

$$\text{e.g., } y = T_{\text{gt}} A x_{\text{gt}}$$

We ask for the **uniqueness**:

> **Problem 1. (Homomorphic Sensing)** Find conditions on $\mathcal{T}$ and A so that $(*)$ admits a unique solution x_{gt}.[7] [8] [9]

---

7. M. C. Tsakiris and **L. Peng**, "Homomorphic sensing," ICML 2019.

8. **L. Peng**, B. Wang, and M. C. Tsakiris, "Homomorphic sensing: sparsity and noise," ICML 2021.

9. **L. Peng** and M. C. Tsakiris, "Homomorphic sensing of subspace arrangements," Applied and Computational Harmonic Analysis (under revision).

missing signs | missing correspondences | missing entries

$\mathcal{T} = \mathcal{S}_m$ $\{m \times m$ permutation matrices$\}$

$\mathcal{T} = \mathcal{S}_{r,m}$ $\{r \times m$ selection matrices$\}$ unlabeled sensing

i.e., $m \times m$ permutations with $(m-r)$ rows removed

$\mathcal{T} = \mathcal{B}_m$ $\{m \times m$ sign matrices$\}$ real phase retrieval

i.e., $m \times m$ diagonal matrices with ± 1 on the diagonal

$\mathcal{T} = \mathcal{S}_{r,m}\mathcal{B}_m := \{SB \colon S \in \mathcal{S}_{r,m},\, B \in \mathcal{B}_m\}$ unsigned unlabeled sensing

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \in \mathcal{S}_{3,4} \qquad \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \in \mathcal{B}_3 \qquad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \end{bmatrix} \in \mathcal{S}_{3,4}\mathcal{B}_4$$

$T_1, T_2 \in \mathcal{T}$

z: a column vector of m variables

$\mathcal{Y}_{T_1,T_2}$: an algebraic variety of $\mathbb{C}^m$ consisting of

the common zeros of the 2×2 determinants of $[T_1 z \;\; T_2 z]$

i.e., $\mathcal{Y}_{T_1,T_2} := \{u : \operatorname{rank}[T_1 u \;\; T_2 u] \leqslant 1\}$

i.e., $u \in \mathcal{Y}_{T_1,T_2} \Leftrightarrow T_1 u$ and $T_2 u$ are linearly dependent

$\mathcal{Z}_{T_1,T_2} := \{u : T_1 u = T_2 u\}$, a linear subspace of $\mathbb{C}^m$

$\mathcal{Z}_{T_1,T_2} \subset \mathcal{Y}_{T_1,T_2}$

$$\mathcal{U}_{T_1,T_2} := \mathcal{Y}_{T_1,T_2} \backslash \mathcal{Z}_{T_1,T_2}$$

$$y = TAx \qquad T \in \mathcal{T} \qquad x \in \mathbb{R}^n \tag{$*$}$$

Theorem 4. *If every matrix $T \in \mathcal{T}$ satisfies the rank constraint*

$$\operatorname{rank}(T) \geqslant 2n$$

and if every pair of matrices $T_1, T_2 \in \mathcal{T}$ satisfies the quasi-variety constraint

$$\dim(\mathcal{U}_{T_1,T_2}) < m - n,$$

then, for $A \in \mathbb{R}^{m \times n}$ generic, x_{gt} is a ***unique*** *solution to $(*)$.*

$$y = TAx \qquad T \in \mathcal{T} \qquad x \in \mathbb{R}^n \qquad (*)$$

Corollary. The following holds for a generic matrix $A \in \mathbb{R}^{m\times n}$.

unlabeled sensing: x_{gt} is the unique solution to $(*)$ if $\mathcal{T} = \mathcal{S}_{r,m}$ and $r \geqslant 2n$.[10 11]

real phase retrieval:

$\pm x_{\text{gt}}$ is the unique solution to $(*)$ if $\mathcal{T} = \mathcal{B}_m$ and $m \geqslant 2n$.[12]

unsigned unlabeled sensing:

$\pm x_{\text{gt}}$ is the unique solution to $(*)$ if $\mathcal{T} = \mathcal{S}_{r,m}\mathcal{B}_m$ and $r \geqslant 2n$.[13]

---

10. J. Unnikrishnan, S. Haghighatshoar, and M. Vetterli, "Unlabeled sensing: Solving a linear system with unordered measurements," IEEE Transactions on Information Theory, 2018.

11. D. Han, F. Lv, W. Sun, "Recovery of signals from unordered partial frame coefficients", Applied and Computational Harmonic Analysis, 2018.

12. R. Balan, P. Casazza, D. Edidin, "On signal reconstruction without phase," Applied and Computational Harmonic Analysis, 2006.

13. F. Lv, W. Sun, "Real phase retrieval from unordered partial frame coefficients," Advances in Computational Mathematics, 2018.

Assume that x_{gt} is k-sparse and it is a feasible point of $(**)$.

$$\min_{x \in \mathbb{R}^n} \|x\|_0 \qquad \text{s.t.} \qquad y = TAx, \quad T \in \mathcal{T} \tag{$**$}$$

Again, we ask for the **uniqueness**:

> **Problem 2. (Sparse Homomorphic Sensing)** Find conditions on $\mathcal{T}$ and A under which $(**)$ admits x_{gt} as a unique solution.[14] [15]

14. **L. Peng**, B. Wang, and M. C. Tsakiris, “Homomorphic sensing: sparsity and noise,” ICML 2021.

15. **L. Peng** and M. C. Tsakiris, “Homomorphic sensing of subspace arrangements,” Applied and Computational Harmonic Analysis (under revision).

$$\min_{x\in\mathbb{R}^n} \|x\|_0 \qquad \text{s.t.} \qquad y = TAx, \quad T \in \mathcal{T} \tag{$**$}$$

Assume that x_{gt} is k-sparse and it is a feasible point of $(**)$.

Theorem 5. *If every matrix $T \in \mathcal{T}$ satisfies the rank constraint*

$$\mathrm{rank}(T) \geqslant 2k$$

and if every pair of matrices $T_1, T_2 \in \mathcal{T}$ satisfies the quasi-variety constraint

$$\dim(\mathcal{U}_{T_1,T_2}) < m - k,$$

*then, for $A \in \mathbb{R}^{m\times n}$ generic, x_{gt} is a **unique** solution to $(**)$.*

$$\min_{x \in \mathbb{R}^n} \|x\|_0 \qquad \text{s.t.} \qquad y = TAx, \quad T \in \mathcal{T} \qquad (**)$$

Corollary. The following holds for a generic matrix $A \in \mathbb{R}^{m \times n}$.

sparse unlabeled sensing:

x_{gt} is the unique solution to $(**)$ if $\mathcal{T} = \mathcal{S}_{r,m}$ and $r \geqslant 2k$.

sparse real phase retrieval:

$\pm x_{\mathrm{gt}}$ is the unique solution to $(**)$ if $\mathcal{T} = \mathcal{B}_m$ and $m \geqslant 2k$.[16] [17]

sparse unsigned unlabeled sensing:

$\pm x_{\mathrm{gt}}$ is the unique solution to $(**)$ if $\mathcal{T} = \mathcal{S}_{r,m}\mathcal{B}_m$ and $r \geqslant 2k$.

---

16. Y. Wang and Z. Xu, "Phase retrieval for sparse signals," Applied and Computational Harmonic Analysis, 2014.

17. M. Akcakaya, V. Tarokh, "New conditions for sparse phase retrieval," arXiv:1310.1351v2 [cs.IT], 2014.

Future Research

Towards Theory of Greater Generality:

e.g., homomorphic sensing of some manifold

Towards Algorithms of Greater Scalability and Guarantees:

e.g., first-order algorithms for linear regression without correspondences

Towards Applications of Greater Relevance:

e.g., simultaneous pose and correspondences

Towards Topics of Greater Interests (to my PhD advisor):

e.g., mathematics of deep dearning