Introduction to Multi-Structured Signal Recovering
SCIE3500 Final Report

Haoyu Wu
Supervisor: Prof. Jianfeng Cai

Dec 2022

Abstract

In the applications of science and engineering, one often faces the problem of recovering structured models (e.g.., sparse vector, low-rank matrix, permutation matrix, low-rank tensor) through a few linear observations, often considered the output of a random matrix, acts on the original data, which has various applications in signal processing and machine learning. It's well-known that we can use the convex method, i.e. the l^1 norm, to recover the sparse signal, and nuclear norm to recover the low-rank signal. In section 2, we will introduce the problem and some general results, and the theorem first focuses on simultaneous structured (sparse and low-rank), and its corresponding methods: a combination of the two norms. However, we will conclude that applying multi norms to recover the signal would not work than the dominant norm in complexity. In section 3, we will introduce the low-rank case in tensor, which is considered tucker rank/decomposition. It's also can be considered a multi-structure signal. In section 4, We will introduce the famous atomic norm, which can be considered as the scalar of the convex hull of atoms. Atomic norm has more profound properties, which may imply the connection between these different recovering methods. We also want to gain deep insight into the atomic norms.

1 Introduction, Problem Setup

In signal recovering, the traditional convex method to recover x_0 is in the form:

$$\hat{x} = \text{argmin}_{x \in C} f(x) \quad s.t.\, Ax = Ax_0 \tag{1}$$

where $f(x)$ is a non-negative convex function, y is the data we obersved, C is the side information of x and A is always considered as a random matrix. We are trying to find the best f corresponding to the structure of x, such that we can recover x in how large of the gaussian measurement m, where $A \in \mathbb{R}^{n \times m}$, with a high probability.

More precisely, the f is a convex function w.s.t. some norms acting on x, and non-decreasing in each norm:

$$f(x) = h\left(||x||_{(1)}, ||x||_{(2)}, \cdots, ||x||_{(k)}\right)$$

where h is convex and non-decreasing in each norm. The subdifferential of f at x is defined as:

$$\partial f(x) = \{g \in C : f(z) \geq f(x) + \langle g, z - x \rangle\}$$

by well-known result, it's also can be written as

$$\partial f(x) = \{g \in C : \langle g, x \rangle = f(x), \sup_{f(u) \leq 1} \langle g, u \rangle \leq 1\}$$

and the decent cone of f at x, defined as $C(f, x)$:

$$c(f, x) = \text{cone}\{v : f(x + v) \leq f(x)\} = \text{cone}(\partial f(x))^\circ$$

where $\text{cone}(\partial f(x))^\circ$ is the polar cone of the subdifferential cone.
The Lipschitz constant of a norm is defined by

$$L = \sup_{x \neq y \in \mathbb{R}^{\times}} \frac{||x|| - ||y||}{||x - y||_2}$$

1.1 Property (Prop 2.1 of [CRPW12])

Given original data x_0, $\hat{x} = x_0$ is the unique optimal solution of (1) if and only if $\text{null}(A) \cap C(f, x_0) = \{0\}$.

Proof. We can rewrite $(*)$ in the form:

$$\min_z f(x_0 + z) \quad s.t.\, z \in \text{null}(A)$$

Suppose $\text{null}(A) \cap C(f, x_0) = \{0\}$. Notice $f(x_0 + z) \leq f(x_0)$ implies $z \in C(f, x_0)$, we have $f(x_0 + z) \geq f(x_0), \forall z \in \text{null}(A) \backslash \{0\}$. Conversely, x_0 is the unique optimal solution means $f(x_0 + z) \geq f(x_0), \forall z \in \text{null}(A) \backslash \{0\}$, implies the other direction.

□

1.2 Result Obtained of Gaussian Measurement for One-Structured Signal

List of Well-Known Result				
Model	Convex Method	Alternative No. of Measurement	Complexity	Degree of Freedom
s-Sparse vector in $\mathbb{R}^n$	l_1 norm	$2s \log(n/s) + 5s/4$	$O(s \log(n/s))$	s
$n \times n$ rank-r matrix	Nuclear norm	$3r(2n - r)$	$O(nr)$	$r(2n - r)$
$n \times n$ permutation matrix	Norm induced by Birkhoff polytope	$9n \log n$	$O(n \log n)$	$n!$
$n \times n$ orthogonal matrix	Spectral norm	$(3n^2 - n)/4$	$O(n^2)$	$n(n-1)/2$

2 Simultaneously Structured Models

Consider a signal x_0 with several low-dimensional structured $S_1, S_2, \cdots, S_d$, which correspond to norm $||\cdot||_{(1)}, ||\cdot||_{(2)}, \cdots, ||\cdot||_{(d)}$. Such an x_0 is defined as a simultaneously structured model.

Denote the convex recovery program:

$$\arg\min_{x \in C} f(x) = h\left(||x||_{(1)}, ||x||_{(2)}, \cdots, ||x||_{(d)}\right) \quad s.t.\, Ax = Ax_0 \tag{2}$$

where $h : \mathbb{R}_+ \to \mathbb{R}_+$ is convex and non-decreasing in each norm.

Theorem 2.1 (Deterministic Failure) (Theorem 3.1 of [OJF+15])

Suppose $C = \mathbb{R}^n$ and,

$$\inf_{g \in \partial f(x_0)} |\bar{g}^T \bar{x}_0| > \frac{||A\bar{x}_0||_2}{\sigma_{min}(A^T)}$$

Then x_0 is NOT a minimizer of (2).

Proof. Given a cone C, it's dual cone is define as $C^* = \{z : \langle z, v \rangle \geq 0, \forall v \in C\}$

From KKT condition of x_0 to be a minizer of (2), v, $z, g \in \partial f(x_0)$,s.t.

$$g - v - A^T z = 0, \quad x_0^T v = 0$$

Denote P as the projection from C to $\text{span}(\{x_0\})$.

$$P(g) = P(v + A^T z) = P(A^T z) = P(A^T) z \tag{3}$$

where $P(A^T) := P\dot{A}^T$, and

$$\sigma_{max}(P(A^T)) = ||\bar{x_0}\bar{x_0}^T A^T||_2 = ||Ax_0||_2 \tag{4}$$

Take l^2 norm on (3):

$$|\bar{x_0}^T g| = ||P(g)||_2 \leq \sigma_{max}(P(A^T))||z||_2 = ||A\bar{x_0}||_2||z||_2 \tag{5}$$

And

$$0 = ||P(-v)||_2 = ||P(A^T z - g)||_2 \Longrightarrow ||P(A^T z)||_2 \leq ||g||_2 \tag{6}$$

Hence

$$\frac{|\bar{x_0}^T g|}{||g||_2} \leq \frac{||A\bar{x}_0||_2||z||_2}{||g||_2} \leq \frac{||A\bar{x}_0||_2||z||_2}{||P(A^T z)||_2} \leq \frac{||A\bar{x}_0||_2}{\sigma_{min}(A^T)} \tag{7}$$

where the first inequlity is from (5) and the second is from (6). (7) violates the initial assumption □

Proposition 2.2

Let L_k be the lipschitz constant of k-th norm, and $\kappa_k := \frac{||\bar{x}_0||_{(k)}}{L_k}$, $1 \leq k \leq d$. Then

$$\inf_{g \in \partial f(x_0)} |\bar{g}^T \bar{x}_0| \geq \kappa_{min}^2$$

for all function f in (2), where $\kappa_{min} = \min_k\{\kappa_k\}$. Combinates the two results, we can deduce the Gaussian lower bound. Before introducing the result, we first define the Gaussian distance

Gaussian Distance

Let $h \in \mathbb{R}^n$ be a Gaussian random matrix. $\mathcal{M}$ is a convex subset of $\mathbb{R}^n$. We denote the Gaussian distance of $\mathcal{M}$,

$$\mathbf{D}(\mathcal{M}) = \mathbb{E}\left(\inf_{x \in \mathcal{M}} ||h - x||_2\right)$$

when $\mathcal{M}$ is a cone, by the result from random matrix, we have $0 \leq \mathbf{D}(\mathcal{M}) \leq \sqrt{n}$, denoting the normalized distance $\bar{\mathbf{D}}(\mathcal{M}) = \frac{\mathbf{D}(\mathcal{M})}{\sqrt{n}}$.

Theorem 2.3 (Gaussian Lower Bound) (Theorem 3.2 of [OJF+15]

Assume A is a random matrix, i.e. entries of A are independent and $\mathcal{N}(0,1)$. Then x_0 would not be minimizer of any f in (2) with probability at least $1 - 10\exp(-\frac{1}{16}\min\{m_{low}, (1 - \bar{\mathbf{D}}(C))^2)\}$, where

$$m_{low} := \frac{(1 - \bar{\mathbf{D}}(C))n\kappa_{min}^2}{100} \sim O(n\kappa_{min}^2)$$

Noticed $\bar{\mathbf{D}}(C) = 0$ when $C = \mathbb{R}^n$. The result conclude that $m_{low} = O(n\kappa_{min}^2)$. We can find that in this model:

Gaussian Lower Bound Method				
Model	the function f	lipschitz constant L	$\|\|\bar{x}_0\|\| \leq$	$n\kappa^2$
s-Sparse vector in $\mathbb{R}^n$	l_1 norm	$\sqrt{n}$	$\sqrt{s}$	s
r-rank matrix in $\mathbb{R}^{n\times n}$	Nuclear norm	$\sqrt{n}$	$\sqrt{r}$	nr
k,k-sparse, r-rank matrix	$h(\|\|\cdot\|\|_*, \|\|\cdot\|\|_1)$	-	-	$\min\{k^2, nr\}$

Noticed that when C is smaller, $\mathbf{D}(C)$ will be larger, m_{low} will be smaller. It means when the constraint, i.e. the domain of signal gets smaller, the m_{low} will also be smaller. In the meantime, it shows that $O(m_{low}) \geq O(dof)$, meaning if the gaussian measurement is lower than the degree of freedom in degree, we would have a high probability that all functions in (2) fail to recover the original signal. Combinates the above 3 theorems, we can handle the case for sparse and low-rank matrices. Actually, we see that it does not help to contain the domain in a cone, which is only different from a constant factor.

Key-points of Proof of Gaussian Lower Bound

It mostly applies theorem 2.1. and some result in random matrix. We focus on the case $C = \mathbb{R}^n$. By theorem from the random matrix(Cor 5.35 of [Ver]):

$$P(\sigma_{max}P(A^T)) \leq 1.5\sqrt{m} + 1.5\sqrt{1} \leq 2.5\sqrt{m}) = P(\sigma_{min}(A^T) \geq \sqrt{n} - 1.5\sqrt{m}) = 1 - 2e^{-m/8}$$

such that

$$\kappa_{min} \leq \frac{||A\bar{x}_0||_2}{\sigma_{min}A^T} = \frac{\sigma_{max}(P(A^T))}{\sigma_{min}A^T} \leq \frac{2.5\sqrt{m}}{\sqrt{n} - 1.5\sqrt{m}}$$

Recall $\inf_{g \in \partial f(x_0)} |\bar{g}^T \bar{x}_0| \geq \kappa_{min}$, hence

$$(\sqrt{n} - 1.5\sqrt{m})\kappa_{min} \leq 2.5\sqrt{m} \Longrightarrow c\sqrt{n}\kappa_{min} \leq \sqrt{m} \Longrightarrow m \geq c' n \kappa_{min}$$

Definition 2.4

Denote a matrix $\mathbf{X} \in \mathbb{R}^{\kappa \times \kappa \times \kappa}$ is a sparse and low-rank, $S\&L$, matrix if there is smallest square submatrix contains $\mathbf{X}$ is $k \times k$ and $\mathbf{X}$ is rank r. Denote $\mathbf{X}$ is $S\&L$ with (k, r). Then combinate the above theorem, we can deduce the gaussian lower bound for $S\&L$ matrix via the convex method.

Theorem 2.5 (Gaussian Lower Bound for $S\&L$ matrix)(Theorem 3.3 of [OJF+15]

(a) **No Constraint for X_0**

Denote $f(\mathbf{X}) = ||\mathbf{X}||_{1,2} + \lambda||\mathbf{X}||_*$ with $\lambda \geq 0$, $C = \mathbb{R}^{n \times n}$. Then (2) fail to recover $\mathbf{X}_0$ with probability $1 - \exp(-c_1 m_{low})$ when $m \leq c_2 m_{low}$, where $m_{low} = \min\{nk, 2nr\}$ and c_1, c_2 are some constant.

(b) **X_0 is PSD via $l_{1,2}$**

Denote $f(\mathbf{X}) = ||\mathbf{X}||_{1,2} + \lambda||\mathbf{X}||_*$ with $C = \mathbb{S}^d$, which is the PSD cone. Then (2) fail to recover $\mathbf{X}_0$ with probability $1 - \exp(c_1 nr)$ when $m \leq c_2 nr$, where c_1, c_2 are some constant.

(c) **X_0 is PSD via l_1**

Denote $f(\mathbf{X}) = ||\mathbf{X}||_1 + \lambda||\mathbf{X}||_*$ with $C = \mathbb{S}^d$, which is the PSD cone. Then (2) fail to recover $\mathbf{X}_0$ with probability $1 - \exp(c_1 m_{low})$ when $m \leq c_2 m_{low}$, where $m_{low} = \min\{||\bar{\mathbf{X}}||_1^2, n||\bar{\mathbf{X}}_0||_*^2\}$ and c_1, c_2 are some constant.

For the Special case that $\mathbf{X}_0 = vv^T$ where v is a k-sparse vector in $\mathbb{R}^n$, then $m_{low} = \min\{||\bar{v}||_1^4, n\}$. $m_{low} = \min\{k^2, n\}$ when the non-zero entries of v is ± 1.

The above results conclude that the use function in class (2), i.e. the convex method combined the l^1 norm and nuclear norm cannot recover the sparse and low-rank signal better than the dominant norm of the two. Also, there is the best function f_b in multi-structured signal:

Best Function (Lemma 6.1 of [OJF+15]

Consider the recovery program (2), and the best function:

$$f_b(x) = \max_{1 \leq k \leq d} \frac{||x||_{(k)}}{||x_0||_{(k)}}$$

if $\arg\min f_b$ cannot return x_0, then all the function in class (2) cannot return x_0. The "best" means the function is the best function to recover x_0, but not the most efficient function. Also, it's a form of max of the norms, which means the complexity only depends on the dominant norm.

Proof. Suppose $\arg\min f_b$ not return x_0, there exist x_1 such that $f_b(x_1) < f_b(x_0)$,

$$\frac{||x_1||_{(k)}}{||x_0||_{(k)}} \leq f_b(x_1) \leq f_b(x_0) = 1 \Longrightarrow ||x_1||_{(k)} < ||x_0||_{(k)}, 1 \leq k \leq d$$

Then recall that h is increasing in each norm, and x_1 is smaller than x_0 in all norms. Hence for any function f in the class, we will have $f(x_1) < f(x_0)$, which means any f would not return x_0.

□

3 Recovery of Low-rank Tensor

We now focus the case on tensor, only with low-rank structure. First, we need to denote what is the rank of a tensor. One is the CP rank for tensor k-way tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_k}$:

$$\text{rank}_{cp}(\mathcal{X}) = \min \left\{ r : \mathcal{X} = \sum_{i=1}^{r} a_1^{(i)} \circ a_2^{(i)} \circ \cdots \circ a_k^{(i)} \right\}$$

It's more like the traditional definition of rank. But the nuclear norm of a tensor in CP rank is intractable since it's NP-hard in general [HL09], out of the polynomial time. We focus on a numerical computation on another structure, recovering tensor with a small tucker rank. Consider the Tucker decomposition[Tuc66]:

$$\mathcal{X} = \mathcal{C} \times_1 \mathcal{U}_1 \times_2 \mathcal{U}_2 \times_3 \cdots \times_k \mathcal{U}_k$$

Unfold the tensor in different directions; we can obtain

$$\mathcal{X}_{(j)} = (U_j)\mathcal{C}_{[j]}(\mathcal{U}_k \otimes \cdots \mathcal{U}_{j+1} \otimes \mathcal{U}_{j-1} \otimes \cdots \otimes \mathcal{U}_1)^* \in \mathbb{R}^{n_j \times \prod_{i \neq j} n_i}$$

$$\text{rank}_{tc}(\mathcal{X}) := \big(\text{rank}(\mathcal{X}_{(1)}, \text{rank}(\mathcal{X}_{(2)}, \cdots, \text{rank}(\mathcal{X}_{(k)}\big)$$

we say tensor $\mathcal{X}$ is low-rank, i.e. tucker rank r means $\text{rank}(\mathcal{X}_{(j)}) \leq r$ for $1 \leq j \leq k$. For a low-rank tensor, actually, we have k structured on it. If we just use one of them by convex relaxation and use nuclear norm to recover it, we can recover the low-rank tensor at complexity $O(n^{k-1}r)$. It's the same as recovering a rank r, $n \times n^{k-1}$ matrix.

But only one structure is used in this method; we want to make some improvements. One way is to find a structure that includes all the k structures, i.e., a set that includes the intersection of the k corresponding set and make convex relaxation on the new set. We want to use nuclear norm to recover the signal, which depends on how large the matrix is. The method is to make the matrix more square.

Square Norm (Section 4 of [MHWG13]

For the tensor $\mathcal{X}$ with Tucker decomposition $\mathcal{X} = \mathcal{C} \times_1 \mathcal{U}_1 \times_2 \mathcal{U}_2 \times_3 \cdots \times_k \mathcal{U}_k \in \mathbb{R}^{n_1 \times n_2 \times, \cdots, \times n_k}$,

$$\mathcal{X}_{[j]} = (U_j \otimes U_{j-1} \otimes \cdots \otimes U_1)\mathcal{C}_{[1,2,\cdots,j]}(\mathcal{U}_k \otimes \cdots \mathcal{U}_{j+1})^* \in \mathbb{R}^{\prod_{i=1}^{j} n_i \times \prod_{i=j+1}^{k} n_i}$$

Recall $\text{rank}(A \otimes B) = \text{rank}(A)\ \text{rank}(B)$, $\text{rank}(\mathcal{X}_{[j]}) \leq \min\{\prod_{i=1}^{j} r_i, \prod_{i=j+1}^{k} r_i\}$. And we write $\mathcal{X}_\square = \mathcal{X}_{[[\frac{k}{2}]]}$ and $||\mathcal{X}_\square|| := ||\mathcal{X}_{[[\frac{k}{2}]]}||_*$. Since $\mathcal{X}_\square$ is a low-rank matrix, we recover it by nuclear norm,

$$\text{minimize}\, ||\mathcal{X}||_\square \, s.t. \, A\mathcal{X} = A\mathcal{X}_0$$

with $m \geq Cr^{\lceil \frac{k}{2} \rceil} n^{\lceil \frac{k}{2} \rceil}$, we can recover $\mathcal{X}$ with a high probability.

"Optimality" of the square norm method

We discuss the optimality of the square norm method. Assuming that k is an even number, it divides the k into two parts and makes convex relaxation for each part, i.e. the $\frac{k}{2}$ tucker rank structure becomes the rank of their tensor product is smaller than $r^{k/2}$. And noticed the two-part have the same convex relaxation. In the process, there are two steps that may loss information (make the set of convex relaxation too large). The first is dividing the structures into 2 parts, and make convex relaxation separately, and the second is the union of the two sets of convex relaxation. The square norm method will not enlarge in the second step, meaning somehow optimality.

Prospective

Consider the special case $k = 3$, then the square method does not work and the traditional convex nuclear norm method require complexity $O(rn^2)$, i.e. recover a rank r $n \times n^2$ matrix. But the non-convex method could do it better, in $O(rn^{1.5})$. We want to explore whether the convex method could do it either.

One hand is the best function, it shows directly use nuclear norm in tucker decomposition structures will recover the problem in high probability, but $O(rn^2)$. We want to make more improvement on the structures. Simply combinate the 3 structues would not help since 3 is odd number, respect to the square norm method.

One possible way is to use 2 of the 3 structures, i.e. the intersection of two structures, and make convex relaxation on the intersection set.

One possible way is try to reshape the $\mathbb{R}^{n_1 \times n_2 \times n_3}$ tensor in a $n^{1.5} \times n^{1.5}$, even rank r^2 matrix. Then apply nuclear norm, we can recover the matrix in complexity $O(r^2 n^{1.5})$. It's related to some fundamental properties of tensor theory.

Another way is to jump out the nuclear norm, to find out another way to recover low rank signal.

4 Atomic Norm

Consider a collection of atoms, name as an atom set $\mathcal{A}$, is a compact subset of $\mathbb{R}^p$. We can assume that element in $\mathcal{A}$ is all extreme point of $\text{conv}(\mathcal{A})$. Let $||x||_{\mathcal{A}}$ denote the gauge of $\mathcal{A}$:

$$||x||_{\mathcal{A}} = \inf\{t > 0 : x \in t\, conv(\mathcal{A})\}$$

where, WLOG, the centroid of $\text{conv}(\mathcal{A})$ is at origin. We assume more that the value of gauge is $+\infty$ when x does not lie in the affine hull of $\text{conv}(\mathcal{A})$. Furthermore, the gauge can be written as

$$||x||_{\mathcal{A}} = \inf\{\sum_{a \in \mathcal{A}} c_a : x = \sum_{a \in \mathcal{A}} c_a a, c_a \geq 0, \forall a \in \mathcal{A}\}$$

we have $||\cdot||_{\mathcal{A}}$ is a norm if $\mathcal{A}$ is centrally symmetry, called as atomic norm induced by $\mathcal{A}$.

Performance of Atomic norm in signal recovering

Given a structured signal, we can find a suitable atom set $\mathcal{A}$, usually consist of the simplest models, then use the corresponding atomic norm to recover it. In many cases, the atomic norm is the same as the traditional norm, the convex method we use to recover the signal.

Sparse Vector

Consider the atomic set $\mathcal{A} \subset \mathbb{R}^p$ consist of all the unit vectors $\{\pm e_i\}$, it's same as l^1 norm in $\mathbb{R}^p$

Low-Rank Matrices

Consider the set $\mathcal{A}$ of rank-one matrices with unit length under Euclidean-norm. Then $\text{conv}(\mathcal{A})$ is the nuclear norm ball, consisting of matrices with the sum of singular values, i.e. the length under nuclear norm, less than or equal to one.

And also there are many examples, see section 2.2. of [CRPW12].

4.1 Why atomic norm?

Smaller decent cone at x_0, easier to satisfy the empty intersection condition.
Points in $conv(\mathcal{A})$ with smaller decent cones corresponding to simpler models, the larger corresponds to the more complicated. Extreme points of $conv(\mathcal{A})$ is the simplest models.
The low-dimensional faces of $conv(\mathcal{A})$ consist of those elements of a linear combination of a few basic atoms of $\mathcal{A}$. It means points in low-dimensional faces have smaller decent cones than those points in high-dimensional faces.
The atomic norm is somehow the best possible convex method. Consider atom set $\mathcal{A}$ consist of all simplest models, any reasonable convex recovery method should have same value on all these points. These means on any other atom is preferred.
For any $a \in \mathcal{A}$, all $a' - a, a' \in \mathcal{A}$ is the decent directions. A best method should have smallest decent cone, and this is just the atomic norm.

Prospective

For a multi-structured signal, there's an atomic set corresponding to each norm (consist of the simplest models). In some cases, the intersection of these atomic sets will be the atomic set we want for the multi-structured signal, for example, the case of $2k$-sparse vector in $\mathbb{R}^{2p}$, consider strongly, it has 2 structures, k-sparse in first p entries and k-sparse in latter p entries. The corresponding atomic set for the first structure is:

$$(0, 0, \cdots, 0, \pm 1_{\text{(the k-th term)}}, 0, \cdots, 0_{\text{(the p-th term)}}, \mathbb{R}^p), 1 \leq k \leq p$$

and similar to the latter one. Or easily written as $e_k \oplus \mathbb{R}^p$ and $\mathbb{R}^p \oplus e_k$, where e_k is the standard bases of $\mathbb{R}^p$. Their intersection is just $\{\pm e_i \oplus \pm e_j\}_{i,j=1}^{2p}$, and the atomic norm is $\max\{||x_1||_{l^1}, ||x_2||_{l^1}\}$, $x = x_1 \oplus x_2$, it has a good performance in recovering a signal with the 2 structure. And it's a little similar to what we want for the $2k$-sparse structure in $\mathbb{R}^{2p}$. But we noticed that the 2 separated structures are slightly stronger than the original $2k$-sparse structure, hence the norm is larger than the l^1 norm, $conv(\mathcal{A})$ is contained in the unit ball w.s.t. l^1 norm. And more generally, this method will fail for some cases, like the case sparse and low-rank. This method just easily intersects the two atomic sets but does not distinguish which structure is dominant. We want to find out how and when this method would work. For example, how is the strictly definition of "dominant".

References

[CRPW12] Venkat Chandrasekaran, Benjamin Recht, Pablo A. Parrilo, and Alan S. Willsky. The convex geometry of linear inverse problems. *Foundations of Computational Mathematics*, 12(6):805–849, oct 2012.

[HL09] Christopher J. Hillar and Lek-Heng Lim. Most tensor problems are np hard. *CoRR*, 2009.

[MHWG13] Cun Mu, Bo Huang, John Wright, and Donald Goldfarb. Square deal: Lower bounds and improved relaxations for tensor recovery. 2013.

[OJF+15] Samet Oymak, Amin Jalali, Maryam Fazel, Yonina C. Eldar, and Babak Hassibi. Simultaneously structured models with application to sparse and low-rank matrices. *IEEE Transactions on Information Theory*, 61(5):2886–2908, 2015.

[Tuc66] L. Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3):279–311, 1966.

[Ver] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. arXiv:1011.3027v7.