Non-convex Compressed Sensing with the Sum-of-Squares Method

Tasuku Soma* Yuichi Yoshida†

Abstract

We consider stable signal recovery in ℓ_q quasi-norm for $0 < q \leq 1$. In this problem, given a measurement vector $\boldsymbol{y} = A\boldsymbol{x}$ for some unknown signal vector $\boldsymbol{x} \in \mathbb{R}^n$ and a known matrix $A \in \mathbb{R}^{m \times n}$, we want to recover $\boldsymbol{z} \in \mathbb{R}^n$ with $\|\boldsymbol{x} - \boldsymbol{z}\|_q = O(\|\boldsymbol{x} - \boldsymbol{x}^*\|_q)$ from a measurement vector, where $\boldsymbol{x}^*$ is the s-sparse vector closest to $\boldsymbol{x}$ in ℓ_q quasi-norm.

Although a small value of q is favorable for measuring the distance to sparse vectors, previous methods for $q < 1$ involve ℓ_q quasi-norm minimization which is computationally intractable. In this paper, we overcome this issue by using the sum-of-squares method, and give the first polynomial-time stable recovery scheme for a large class of matrices A in ℓ_q quasi-norm for any fixed constant $0 < q \leq 1$.

1 Introduction

Compressed sensing, the subject of recovering a sparse signal vector $\boldsymbol{x} \in \mathbb{R}^n$ from its measurement vector $\boldsymbol{y} = A\boldsymbol{x}$ for some known matrix $A \in \mathbb{R}^{m \times n}$ with $m \ll n$, has been intensively studied in the last decade [10, 22, 23]. It is well known that, if A is a random Gaussian or Rademacher matrix with $m = \Theta(s \log(n/s))$, then with high probability[1] we can exactly recover any *s-sparse* signal vector, that is, a vector with the number of non-zero elements at most s [13]. This means that we can significantly compress signal vectors if they are sparse (in an appropriate choice of basis). In a realistic scenario, however, we can assume only that the signal vector $\boldsymbol{x}$ is not exactly sparse but close to a sparse vector. In such cases, we want to recover $\boldsymbol{x}$ with an error controlled by its distance to s-sparse vectors, and this feature is called *stability* [23] or *instance optimality* [20] of a recovery scheme. For $q > 0$, an integer $s \in \mathbb{Z}_+$, and a vector $\boldsymbol{x} \in \mathbb{R}^n$, we define

$$\sigma_s(\boldsymbol{x})_q = \min\{\|\boldsymbol{x} - \boldsymbol{x}'\|_q \mid \boldsymbol{x}' \in \mathbb{R}^n, \|\boldsymbol{x}'\|_0 \leq s\}.$$

That is, $\sigma_s(\boldsymbol{x})_q$ is the distance of $\boldsymbol{x}$ to the closest s-sparse vector in ℓ_q (quasi-)norm. The objective of stable signal recovery in ℓ_q quasi-norm is, given a measurement vector $\boldsymbol{y} = A\boldsymbol{x}$ generated from an unknown signal vector $\boldsymbol{x} \in \mathbb{R}^n$, to recover a vector $\boldsymbol{z} \in \mathbb{R}^n$ such that $\|\boldsymbol{z} - \boldsymbol{x}\|_q = O(\sigma_s(\boldsymbol{x})_q)$. It is desirable to take a small q, for example, in the sparse noise model; that is, $\boldsymbol{x} = \boldsymbol{x}' + \boldsymbol{e}$, where $\boldsymbol{x}'$ is the best s-sparse approximation to $\boldsymbol{x}$ and $\boldsymbol{e}$ is an s'-sparse noise with $s' = O(1)$. To see this, let us compare the cases $q = 1/2$ and $q = 1$ as an illustrative example. Note that $\|\boldsymbol{z} - \boldsymbol{x}\|_1 \leq \|\boldsymbol{z} - \boldsymbol{x}\|_{1/2} \leq n\|\boldsymbol{z} - \boldsymbol{x}\|_1$ and $\sigma_s(\boldsymbol{x})_{1/2} = \|\boldsymbol{e}\|_{1/2} \leq s'\|\boldsymbol{e}\|_1 = s'\sigma_s(\boldsymbol{x})_1$. Hence, $\|\boldsymbol{z} - \boldsymbol{x}\|_{1/2} = O(\sigma_s(\boldsymbol{x})_{1/2})$ implies $\|\boldsymbol{z} - \boldsymbol{x}\|_1 = O(\sigma_s(\boldsymbol{x})_1)$, whereas $\|\boldsymbol{z} - \boldsymbol{x}\|_1 = O(\sigma_s(\boldsymbol{x})_1)$ only implies a weaker bound $\|\boldsymbol{z} - \boldsymbol{x}\|_{1/2} = O(n\sigma_s(\boldsymbol{x})_{1/2})$. It is reported that the noise vector $\boldsymbol{e}$ is often sparse in practice [18, 29]. Intuitively speaking, smaller q works better in the sparse noise model because $\|\boldsymbol{z} - \boldsymbol{x}\|_q^q \to \|\boldsymbol{z} - \boldsymbol{x}\|_0$ and $\sigma_s(\boldsymbol{x})_q \to s'$ as $q \to 0$, and hence $\boldsymbol{z}$ converges to an $(s + O(s'))$-sparse vector.

A common approach to stable signal recovery is relaxing the ℓ_0 quasi-norm by an ℓ_q quasi-norm for $0 < q \leq 1$:

$$(P_q) \qquad \begin{array}{ll} \text{minimize} & \|\boldsymbol{z}\|_q \\ \text{subject to} & A\boldsymbol{z} = \boldsymbol{y} \text{ and } \boldsymbol{z} \in \mathbb{R}^n. \end{array}$$

For $q = 1$, this approach is called *basis pursuit*, a well-studied and central algorithm in compressed sensing. If A satisfies some technical condition, called the *stable null space property*, then the solution $\boldsymbol{z}$ to (P_q) satisfies $\|\boldsymbol{z} - \boldsymbol{x}\|_q = O(\sigma_s(\boldsymbol{x})_q)$. It is folklore that the following result can be obtained by extending the results in [11, 13], which handle the $q = 1$ case.

THEOREM 1.1. *Let $q \in (0, 1]$ and $s, n \in \mathbb{Z}_+$. Then, a random Gaussian or Rademacher matrix $A \in \mathbb{R}^{m \times n}$ with $m = \Theta(s \log(n/s))$ satisfies the following property with high probability: For any vectors $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{y} = A\boldsymbol{x}$, the solution $\boldsymbol{z}$ to (P_q) satisfies $\|\boldsymbol{z} - \boldsymbol{x}\|_q = O(\sigma_s(\boldsymbol{x})_q)$.*

When $q = 1$, i.e., in the basis pursuit, the program (P_q) can be represented as a linear program and we can find

---

*Graduate School of Information Science and Technology, The University of Tokyo (`tasuku_soma@mist.i.u-tokyo.ac.jp`). Supported by JSPS Grant-in-Aid for JSPS Fellows.

†National Institute of Informatics and Preferred Infrastructure, Inc. (`yyoshida@nii.ac.jp`). Supported by MEXT Grant-in-Aid for Scientific Research on Innovative Areas (24106001), JST, CREST, Foundations of Innovative Algorithms for Big Data, and JST, ERATO, Kawarabayashi Large Graph Project.

[1] Here and henceforth, "with high probability" means with a constant probability sufficiently close to one.

its solution in polynomial time. When $q < 1$, however, solving (P_q) is NP-hard in general and Theorem 1.1 is currently only of theoretical interest.

The main contribution of this paper is showing that, even when $q < 1$, there is a recovery scheme whose theoretical guarantee almost matches that of Theorem 1.1:

THEOREM 1.2. (MAIN THEOREM, INFORMAL VERSION) *Let $q \in (0, 1]$ be a constant of the form $q = 2^{-k}$ for some $k \in \mathbb{Z}_+$, $s, n \in \mathbb{Z}_+$, and $\epsilon > 0$. Then, a random (normalized) Rademacher matrix $A \in \mathbb{R}^{m \times n}$ with $m = O(s^{2/q} \log n)$ satisfies the following property with high probability: For any vector $\boldsymbol{y} = A\boldsymbol{x}$, where $\boldsymbol{x} \in [-1, 1]^n$ is an unknown vector, we can recover $\boldsymbol{z}$ with $\|\boldsymbol{z} - \boldsymbol{x}\|_q \leq O(\sigma_s(\boldsymbol{x})_q) + \epsilon$ in polynomial time.*

To our knowledge, our algorithm is the first polynomial-time stable recovery scheme in ℓ_q quasi-norm for $q < 1$. Our algorithm is based on the sum-of-squares method, which we describe next. The assumption on $\boldsymbol{x}$ can be easily relaxed to $\boldsymbol{x} \in [-R, R]$ for a polynomial R in n, keeping the running time polynomial in n.

1.1 Sum of Squares Method As we cannot exactly solve the non-convex program (P_q) when $q < 1$, we use the *sum-of-squares* (SoS) semidefinite programming hierarchy [30, 31, 33, 39], or simply the SoS method. Here, we adopt the description of the SoS method introduced in [3, 4]. The SoS method attempts to solve a *polynomial optimization problem* (POP), via convex relaxation called the *moment* relaxation. We specify an even integer d called the *degree* or the *level* as a parameter. The SoS method will find a linear operator $\mathcal{L}$ from the set of polynomials of degree at most d to $\mathbb{R}$, which shares some properties of expectation operators and satisfies the system of polynomials. Such linear operators are called *pseudoexpectations*.

DEFINITION 1.1. (PSEUDOEXPECTATION) *Let $\mathbb{R}[\boldsymbol{z}]$ denote the polynomial ring over the reals in variables $\boldsymbol{z} = (z(1), \ldots, z(n))$ and let $\mathbb{R}[\boldsymbol{z}]_d$ denote the set of polynomials in $\mathbb{R}[\boldsymbol{z}]$ of degree at most d. A degree-d pseudoexpectation operator for $\mathbb{R}[\boldsymbol{z}]$ is a linear operator $\widetilde{\mathbf{E}} : \mathbb{R}[\boldsymbol{z}]_d \to \mathbb{R}$ satisfying that $\widetilde{\mathbf{E}}(1) = 1$ and $\widetilde{\mathbf{E}}(P^2) \geq 0$ for every polynomial P of degree at most $d/2$.*

For notational simplicity, it is convenient to regard a pseudoexpectation operator as the "expectation" operator of the so-called *pseudodistribution*[2] $\mathcal{D}$. Note that it is possible that no actual distribution realizes a pseudoexpectation operator as the expectation operator (see, e.g. [19, 35]). Also note that we cannot sample from a pseudodistribution; only its low-degree moments are available. We use the subscript $\widetilde{\mathbf{E}}_{\mathcal{D}}$ to emphasize the underlying pseudodistribution. Also we write $\widetilde{\mathbf{E}}_{\boldsymbol{z} \sim \mathcal{D}} P(\boldsymbol{z})$ to emphasize the variable $\boldsymbol{z}$. We say that a degree-d pseudodistribution $\mathcal{D}$ satisfies the constraint $\{P = 0\}$ if $\widetilde{\mathbf{E}} PQ = 0$ for all Q of degree at most $d - \deg P$. We say that $\mathcal{D}$ satisfies $\{P \geq 0\}$ if it satisfies the constraint $\{P - S = 0\}$ for some SoS polynomial $S \in \mathbb{R}_d[\boldsymbol{z}]$.

A pseudoexpectation can be compactly represented by its moment matrix, an $n^{O(d/2)} \times n^{O(d/2)}$ symmetric matrix M containing all the values for monomials of degree at most d. The above condition $\widetilde{\mathbf{E}} P^2 \geq 0$ is equivalent to requiring that M is positive semidefinite, and the other conditions can be represented by linear constraints among entries of M. The following theorem states that we can efficiently optimize over pseudodistributions, basically via semidefinite programming.

THEOREM 1.3. (THE SOS METHOD [30, 31, 33, 39]) *For every $\epsilon > 0$, $d, n, m, M \in \mathbb{N}$ and n-variate polynomials $P_1, \ldots, P_m$ in $\mathbb{R}_d[\boldsymbol{z}]$, whose coefficients are in $0, \ldots, M$, if there exists a degree-d $\mathcal{D}$ for $\boldsymbol{z}$ satisfying the constraint $P_i = 0$ for every $i \in [m]$, then we can find in $(n \cdot \operatorname{poly}\log(M/\epsilon))^{O(d)}$ time a degree-d pseudodistribution $\mathcal{D}'$ for $\boldsymbol{z}$ satisfying $P_i \leq \epsilon$ and $P_i \geq -\epsilon$ for every $i \in [m]$.*

By using binary search, we can also handle objective functions. We ignore numerical issues since it will never affect our arguments (at least from a theoretical perspective). We simply assume that the SoS method finds an optimal degree-d pseudodistribution in $\operatorname{poly}(n^{O(d)})$ time.

The dual of the SoS method corresponds to the *SoS proof system* [27], an automizable and restricted proof system for refuting a system of polynomials. The SoS proof system are built on several deep results of real algebraic geometry, such as *positivestellensatz* and its variants [34, 37]. Roughly speaking, any fact that is provable in the (bounded-degree) SoS proof system is also reflected to pseudodistributions. This approach lies at the heart of recent surprising results [3, 32]; they imported known proofs for integrality gap instances into the SoS proof system and showed that (low-degree) SoS hierarchy solves the integrality gap instances.

1.2 Proof Outline Now we give a proof outline of Theorem 1.2. First we recall basic facts in stable signal recovery using the non-convex program (P_q). For a set $S \subseteq [n]$, we denote by $\overline{S}$ the complement set $[n] \setminus S$. For a vector $\boldsymbol{v} \in \mathbb{R}^n$ and a set $S \subseteq [n]$, $\boldsymbol{v}_S$ denotes the vector restricted to S. We abuse this notation and regard $\boldsymbol{v}_S$ as the vector in $\mathbb{R}^S$ with $v_S(i) = v(i)$ for $i \in S$ as well as a vector in $\mathbb{R}^n$ with $v_S(i) = v(i)$ for $i \in S$ and $v_S(i) = 0$

[2] It is also called *locally consistent distribution* [19, 35].

for $i \in \overline{S}$. The following notion is important in stable signal recovery.

DEFINITION 1.2. (ROBUST NULL SPACE PROPERTY) *Let $q \in (0,1]$, $\rho, \tau > 0$, and $s \in \mathbb{Z}_+$. A matrix $A \in \mathbb{R}^{m \times n}$ is said to satisfy the* (ρ, τ, s)-robust null space property *(RNSP) in ℓ_q quasi-norm if*

$$\|\boldsymbol{v}_S\|_q^q \leq \rho \|\boldsymbol{v}_{\overline{S}}\|_q^q + \tau \|A\boldsymbol{v}\|_2^q \tag{1.1}$$

holds for any $\boldsymbol{v} \in \mathbb{R}^n$ and $S \subseteq [n]$ with $|S| \leq s$.

If A satisfies the (ρ, τ, s)-robust null space property for $\rho < 1$, then the solution to the program (P_q) satisfies $\|\boldsymbol{z} - \boldsymbol{x}\|_q = O(\sigma_s(\boldsymbol{x})_q)$ and we establish Theorem 1.1. Indeed, a random Gaussian or Rademacher matrix will do as long as $m = \Omega(s \log n/s)$ [13].

The main idea of our algorithm is that we can convert the proof of Theorem 1.1 into a *sum of squares (SoS) proof*, a proof such that all the polynomials appearing there have bounded degree and all the inequalities involved are simply implied by the SoS partial order. Due to the duality between SoS proofs and the SoS semidefinite programming hierarchy (see, e.g. [32]), any consequence of a SoS proof derived from a given set of polynomial constraints is satisfied by the pseudodistribution obtained by the SoS method on these constraints.

Let us consider the following SoS version of (P_q).

$$(\tilde{P}_q) \qquad \begin{array}{ll} \text{minimize} & \widetilde{\mathbf{E}}_{\boldsymbol{z} \sim \mathcal{D}} \|\boldsymbol{z}\|_q^q \\ \text{subject to} & \mathcal{D} \text{ satisfies } A\boldsymbol{z} = \boldsymbol{y}, \end{array}$$

where $\mathcal{D}$ is over pseudodistributions for the variable $\boldsymbol{z}$. In our algorithm, we solve $(\tilde{P}_q)$ with technical additional constraints using the SoS method, and then recover a vector from the obtained pseudoexpectation operator. A subtle issue here is that we need to deal with terms of the form $|\boldsymbol{z}(i)|^q$ for fractional q, which is not polynomial in $\boldsymbol{z}$. A remedy for this is adding auxiliary variables and constraints among them (details are explained later), and for now we assume that we can exactly solve the program $(\tilde{P}_q)$.

As we can only execute SoS proofs by using the SoS method, we need an SoS version of the robust null space property:

DEFINITION 1.3. (INFORMAL VERSION) *Let $q \in (0,1]$, $\rho, \tau > 0$, and $s \in \mathbb{Z}_+$. A matrix $A \in \mathbb{R}^{m \times n}$ is said to satisfy the* (ρ, τ, s)-pseudo robust null space property *in ℓ_q quasi-norm (ℓ_q-(ρ, τ, s)-PRNSP for short) if*

$$\widetilde{\mathbf{E}}_{\boldsymbol{v} \sim \mathcal{D}} \|\boldsymbol{v}_S\|_q^q \leq \rho \widetilde{\mathbf{E}}_{\boldsymbol{v} \sim \mathcal{D}} \|\boldsymbol{v}_{\bar{S}}\|_q^q + \tau \left(\widetilde{\mathbf{E}}_{\boldsymbol{v} \sim \mathcal{D}} \|A\boldsymbol{v}\|_2^2 \right)^{q/2} \tag{1.2}$$

holds for any pseudodistribution $\mathcal{D}$ of degree at least $2/q$ and $S \subseteq [n]$ with $|S| \leq s$.

In the formal definition of PRNSP, we allow to impose some other technical constraints. Note that PRNSP is a stronger condition than the robust null space property because the latter considers only actual distributions consisting of single vectors.

If the measurement matrix $A \in \mathbb{R}^{m \times n}$ satisfies the PRNSP, then the resulting pseudodistribution $\mathcal{D}$ satisfies an SoS version of stable signal recovery. Moreover, we can round $\mathcal{D}$ to a vector close to the signal vector, as shown in the following theorem.

THEOREM 1.4. (INFORMAL VERSION) *Let $q \in (0,1]$ be a constant of the form $q = 2^{-k}$ for some $k \in \mathbb{Z}_+$, $0 < \rho < 1$, $\tau > 0$, $s \in \mathbb{Z}_+$, and $\epsilon > 0$. If A satisfies the ℓ_q-(ρ, τ, s)-PRNSP, then, from the pseudodistribution $\mathcal{D}$ obtained by solving $(\tilde{P}_q)$ with additional constraints depending on ϵ, we can compute a vector $\bar{\boldsymbol{z}} \in \mathbb{R}^n$ such that*

$$\|\bar{\boldsymbol{z}} - \boldsymbol{x}\|_q \leq \frac{2 + 2\rho}{1 - \rho} \sigma_s(\boldsymbol{x})_q + \epsilon$$

in polynomial time.

We then show that there exists a family of matrices satisfying PRNSP, namely the family of Rademacher matrices.

THEOREM 1.5. (INFORMAL VERSION) *Let $q \in (0,1]$ and $s \in \mathbb{Z}_+$. For some $0 < \rho < 1$ and $\tau = O(s)$, a (normalized) Rademacher matrix $A \in \mathbb{R}^{m \times n}$ with $m = \Omega(s^{2/q} \log n)$ satisfies the ℓ_q-(ρ, τ, s)-PRNSP with high probability.*

Theorem 1.2 is a consequence of Theorems 1.4 and 1.5.

1.3 Technical Contributions In order to prove Theorem 1.4, we would like to follow the proof of Theorem 1.1 in the SoS proof system. However, many steps of the proof requires non-trivial modifications for the SoS proof system to go through. Let us explain in more detail below.

Handling ℓ_q quasi-norm The first obvious obstacle is that the program $(\tilde{P}_q)$ contains non-polynomial terms such as $|z(i)|^q$ in the objective function. This is easily resolved by *lifting*; i.e., we introduce a variable $t_{z(i)}$, which is supposed to represent $|z(i)|^q$, and impose the constraints $t_{z(i)}^{2/q} = z(i)^2$ and $t_{z(i)} \geq 0$ (recall that q is of the form 2^{-k} for some $k \in \mathbb{Z}_+$).

Imposing the triangle inequality for ℓ_q^q metric The second problem is that the added variable $t_{z(i)}$ does not always behave as expected. When proving Theorem 1.1, we frequently use the triangle inequality $\|\boldsymbol{u} + \boldsymbol{v}\|_q^q \leq \|\boldsymbol{u}\|_q^q + \|\boldsymbol{v}\|_q^q$ for vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, or in other words, $\sum_i |u(i) + v(i)|^q \leq \sum_i |u(i)|^q + \sum_i |v(i)|^q$. Unfortunately, we cannot derive its SoS version $\widetilde{\mathbf{E}} \sum_i t_{u(i)+v(i)} \leq \widetilde{\mathbf{E}} \sum_i t_{u(i)} + \widetilde{\mathbf{E}} \sum_i t_{v(i)}$ in the SoS proof system. This is because, even if we impose the constraints $t_{u(i)}^{2/q} = u(i)^2$, $t_{v(i)}^{2/q} = v(i)^2$, and $t_{u(i)+v(i)}^{2/q} = (u(i) + v(i))^2$, these constraints do not involve any constraint among $t_{u(i)}$, $t_{v(i)}$, and $t_{u(i)+v(i)}$. (Instead, we can derive $\left(\widetilde{\mathbf{E}}\sum_i t_{u(i)+v(i)}^{2/q}\right)^{q/2} \leq \left(\widetilde{\mathbf{E}}\sum_i t_{u(i)}^{2/q}\right)^{q/2} + \left(\widetilde{\mathbf{E}}\sum_i t_{v(i)}^{2/q}\right)^{q/2}$ in the SoS proof system, which is not very useful.)

To overcome this issue, we use the fact that we only need the triangle inequality among $\boldsymbol{z} - \boldsymbol{x}$, $\boldsymbol{z}$, and $\boldsymbol{x}$, where $\boldsymbol{z}$ is a variable in $(\tilde{P}_q)$ and $\boldsymbol{x}$ is the signal vector. Of course we do not know $\boldsymbol{x}$ in advance; we only know that $\boldsymbol{x}$ lies in $[-1, 1]^n$. We will discretize the region $[-1, 1]^n$ using a small $\delta \ll 1/n^{2/q}$ and explicitly add the constraints $\widetilde{\mathbf{E}}\, t_{z(i)-b} \leq \widetilde{\mathbf{E}}\, t_{z(i)} + b$ for each $i \in [n]$ and multiple $b \in [-1, 1]$ of δ. Instead of $\boldsymbol{x}$, we use an approximated vector $\boldsymbol{x}^L$ each coordinate of which is a multiple of δ with $\|\boldsymbol{x} - \boldsymbol{x}^L\|$ being small. The SoS proof system augmented with these triangle inequalities can now follow the proof of Theorem 1.1.

Rounding Rounding a pseudodistribution to a feasible solution is a challenging task in general. Fortunately, we can easily round the pseudodistribution $\mathcal{D}$ obtained by solving the program $(\tilde{P}_q)$ to a vector $\bar{\boldsymbol{z}}$ close to the signal vector by fixing each coordinate $\bar{z}(i)$ to $\arg\min_b \widetilde{\mathbf{E}}_{\mathcal{D}}\, |z(i) - b|^q$, where $b \in [-1, 1]$ runs over multiples of δ.

Imposing the PRNSP In order to prove Theorem 1.5, we show that a (normalized) Rademacher matrix with a small coherence satisfies PRNSP. Here, the *coherence* of a matrix $A \in \mathbb{R}^{m\times n}$ with columns $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m$ is defined as $\mu = \max_{i,j} |\langle \boldsymbol{a}_i, \boldsymbol{a}_j\rangle|$. Again our approach is the following: rewrite a proof that matrices with a small coherence satisfy the robust null space property, into an SoS proof. Here is the last obstacle; it seems inevitable for us to add *exponentially many* extra variables and constraints, in order to convert known proofs. However, we will show that the number of extra variables is polynomially bounded and that the extra constraints admit a separation oracle, if A is a Rademacher matrix. Thanks to the ellipsoid method, we can obtain the desired SoS proof for Rademacher matrices in polynomial time.

1.4 Related work The notion of stable signal recovery in ℓ_1 norm is introduced in [11, 12] and it is shown that basis pursuit is a stable recover scheme for a random Gaussian matrix $A \in \mathbb{R}^{m\times n}$ provided that $m = \Omega(s \log(n/s))$. The notion of stable signal recovery is extended to ℓ_q quasi-norm in [20, 36]. Stable recovery schemes with almost linear encoding time and recovery time have been proposed by using sparse matrices [8, 9, 28]. See [23, 26] for a survey of this area.

Besides the fact that $\|\cdot\|_q^q$ converges to the ℓ_0 norm as $q \to 0$, an advantage of using the ℓ_q quasi-norm is that we can use a smaller measurement matrix for signal recovery (at least for the exact case). Chartrand and Staneva [16] showed that ℓ_q quasi-norm minimization is a stable recovery scheme for a certain matrix $A \in \mathbb{R}^{m\times n}$ with $m = \Omega(s + qs \log(n/s))$.[3] Hence, as $q \to 0$, the required dimension of the measurement vector scales like $O(s)$. As stated in the beginning of this section, the ℓ_q quasi-norm minimization is effective for the sparse noise model. Other methods (e.g., Prony's method [21]) are also applicable to the sparse noise model, but they require a priori information on the sparsity of noise.

In order to implement the ℓ_q quasi-norm minimization for $q < 1$, several heuristic methods have been proposed, including the gradient method [14], the iteratively reweighed least squares method (IRLS) [17], and an operator-splitting algorithm [15], and a method that computes a local minimum [24].

Recently, the sum-of-squares method has been found to be useful in several unsupervised learning tasks such as the planted sparse vector problem [4], dictionary learning [5], and tensor decompositions [5, 6, 25, 40]. Surveys [19, 35] provide detailed information on the Lasserre and other hierarchies. A handbook [2] contains various articles on the semidefinite programming, conic programming, and the SoS method, from both the theoretical and the practical point of view.

1.5 Organization In Section 2, we review basic facts on the SoS proof system. In Section 3, we show that stable signal recovery is achieved if the measurement matrix A satisfies PRNSP. Then, we show that a Rademacher matrix satisfies PRNSP with high probability in Section 4.

2 Preliminaries

For an integer $k \in \mathbb{N}$, we denote by $[k]$ the set $\{1, 2, \ldots, k\}$. For a matrix A, $\sigma_{\max}(A)$ denotes its largest singular value. We write $P \succeq 0$ if P is a sum-of-squares polynomial, and similarly we write $P \succeq Q$ if $P - Q \succeq 0$.

We now see several useful lemmas for pseudodistributions. Most of these results were proved in the series of works by Barak and his coauthors [3, 4, 5, 6]. We basically follow the terminology and notations in the recent survey by Barak and Steurer [7].

LEMMA 2.1. (PSEUDO CAUCHY-SCHWARTZ [3]) *Let*

[3] Although they do not give this result explicitly, it can be shown by simple adaptation of their result to a proof for stability of ℓ_1 minimization.

$d \in \mathbb{Z}_+$ *be an integer. Let* $P, Q \in \mathbb{R}[\boldsymbol{x}]_d$ *be polynomials of degree* d *and* $\mathcal{D}$ *be a pseudodistribution of degree at least* $2d$*. Then, we have*

$$\widetilde{\mathbf{E}}_{\boldsymbol{x}\sim\mathcal{D}} P(\boldsymbol{x})Q(\boldsymbol{x}) \leq \left(\widetilde{\mathbf{E}}_{\boldsymbol{x}\sim\mathcal{D}} P(\boldsymbol{x})^2\right)^{1/2} \left(\widetilde{\mathbf{E}}_{\boldsymbol{x}\sim\mathcal{D}} Q(\boldsymbol{x})^2\right)^{1/2}.$$

LEMMA 2.2. ([3]) *Let* $d \in \mathbb{Z}_+$ *be an integer. Let* $\boldsymbol{u}$ *and* $\boldsymbol{v}$ *be vectors whose entries are polynomials in* $\mathbb{R}[\boldsymbol{x}]_d$*. Then for any pseudodistribution* $\mathcal{D}$ *of degree at least* $2d$*, we have*

$$\widetilde{\mathbf{E}}_{\boldsymbol{x}\sim\mathcal{D}} \langle \boldsymbol{u}, \boldsymbol{v}\rangle \leq \left(\widetilde{\mathbf{E}}_{\boldsymbol{x}\sim\mathcal{D}} \|\boldsymbol{u}\|_2^2\right)^{1/2} \left(\widetilde{\mathbf{E}}_{\boldsymbol{x}\sim\mathcal{D}} \|\boldsymbol{v}\|_2^2\right)^{1/2}.$$

LEMMA 2.3. ([3]) *Let* $\boldsymbol{u}$*,* $\boldsymbol{v}$*, and* $\mathcal{D}$ *be the same as above. Then, we have*

$$\left(\widetilde{\mathbf{E}}_{\boldsymbol{x}\sim\mathcal{D}} \|\boldsymbol{u}+\boldsymbol{v}\|_2^2\right)^{1/2} \leq \left(\widetilde{\mathbf{E}}_{\boldsymbol{x}\sim\mathcal{D}} \|\boldsymbol{u}\|_2^2\right)^{1/2} + \left(\widetilde{\mathbf{E}}_{\boldsymbol{x}\sim\mathcal{D}} \|\boldsymbol{v}\|_2^2\right)^{1/2}.$$

LEMMA 2.4. *For any pseudodistribution* $\mathcal{D}$ *of degree at least* 4*, we have*

$$\widetilde{\mathbf{E}}_{(\boldsymbol{u},\boldsymbol{v})\sim\mathcal{D}} \langle \boldsymbol{u}, \boldsymbol{v}\rangle^2 \leq \widetilde{\mathbf{E}}_{(\boldsymbol{u},\boldsymbol{v})\sim\mathcal{D}} \|\boldsymbol{u}\|_2^2 \|\boldsymbol{v}\|_2^2.$$

Proof. By the Lagrange identity, we have

$$\|\boldsymbol{u}\|_2^2\|\boldsymbol{v}\|_2^2 - \langle \boldsymbol{u}, \boldsymbol{v}\rangle^2 = \sum_{i<j} (u(i)v(j) - u(j)v(i))^2.$$

Hence, the statement has an SoS proof of degree 4.

3 Signal Recovery Based on Pseudo Robust Null Space Property

In this section, we prove Theorem 1.4. The formal statement of Theorem 1.4 is provided in Theorem 3.1. In the rest of this section, we fix $q \in (0, 1]$ and an error parameter $\epsilon > 0$. We also fix a matrix $A \in \mathbb{R}^{m\times n}$ and the unknown signal vector $\boldsymbol{x} \in [-1, 1]^n$, and hence the measurement vector $\boldsymbol{y} = A\boldsymbol{x}$.

Let $\delta > 0$ be a parameter chosen later, and L be a set of multiples of δ in $[-1, 1]$. Clearly $|L| = O(1/\delta)$, and there exists $\bar{x} \in L$ with $|x - \bar{x}| \leq \delta$ for any $x \in [-1, 1]$. Let $\boldsymbol{x}^L \in \mathbb{R}^n$ be the point in L^n closest to $\boldsymbol{x}$. Our strategy is recovering $\boldsymbol{x}^L$ instead of $\boldsymbol{x}$. As we know that $\boldsymbol{x}^L \in L^n$ and each coordinate of $\boldsymbol{x}^L$ is discretized, we can add several useful constraints to the moment relaxation, such as triangle inequality.

For the measurement vector $\boldsymbol{y} = A\boldsymbol{x}$, we have

$$\begin{aligned}\|A\boldsymbol{x}^L - \boldsymbol{y}\|_2 &= \|A(\boldsymbol{x}^L - \boldsymbol{x})\|_2 \\ &\leq \sigma_{\max}(A)\|\boldsymbol{x}^L - \boldsymbol{x}\|_2 \\ &\leq \sigma_{\max}(A)\sqrt{s}\delta.\end{aligned}$$

To recover a vector close to $\boldsymbol{x}^L$ in ℓ_q quasi-norm from $\boldsymbol{y}$, we consider the following non-convex program:

$$(P_{q,\eta}) \qquad \begin{array}{ll} \text{minimize} & \|\boldsymbol{z}\|_q^q \\ \text{subject to} & \|A\boldsymbol{z} - \boldsymbol{y}\|_2^2 \leq \eta^2, \end{array}$$

where $\eta = \sigma_{\max}(A)\sqrt{s}\delta$. Note that the desired vector $\boldsymbol{x}^L$ is a feasible solution of this program.

In Section 3.1, we show how to formulate the non-convex program $(P_{q,\eta})$ as a POP. Although the form of the POP depends on ϵ, we can solve its moment relaxation using the SoS method. Then, we show that, if A satisfies a certain property, which we call pseudo robust null space property, then the vector $\boldsymbol{z}$ "sampled" from the pseudodistribution obtained by the SoS method is close to the vector $\boldsymbol{x}^L$ in ℓ_q quasi-norm. Then in Section 3.2, we show how to round the pseudodistribution to an actual vector $\bar{\boldsymbol{z}}$ so that $\|\bar{\boldsymbol{z}} - \boldsymbol{x}\|_q = O(\sigma_s(\boldsymbol{x})_q) + \epsilon$ for small $\epsilon > 0$.

3.1 Formulating ℓ_q quasi-norm minimization as a POP To formulate $(P_{q,\eta})$ as a POP, we introduce additional variables. Formally, we define $\|\boldsymbol{z}\|_q^q$ as a shorthand for $\sum_{i=1}^n t_{z(i)}$, where $t_{z(i)}$ for $i \in [n]$ is an additional variable corresponding to $|z(i)|^q$. We also impose constraints $t_{z(i)} \geq 0$ and $t_{z(i)}^{2/q} = z(i)^2$ for each $i \in [n]$.

Unfortunately, these constraints are not sufficient to derive formula connecting $\widetilde{\mathbf{E}}\, t_{z(i)}$ and $z(i)$ in the SoS proof system because $\widetilde{\mathbf{E}}\, t_{z(i)}$ and $\widetilde{\mathbf{E}}\, t_{z(i)}^{2/q}$ are unrelated in general. To address this issue, we add several other variables and constraints. Specifically, for each $i \in [n]$ and $b \in L$, we add a new variable $t_{z(i)-b}$ along with the following constraints:

- $t_{z(i)-b} \geq 0$,
- $t_{z(i)-b}^{2/q} = (z(i) - b)^2$,
- $t_{z(i)-b} \leq t_{z(i)-b'} + |b - b'|^q \quad (b' \in L)$,
- $|b - b'|^q \leq t_{z(i)-b} + t_{z(i)-b'} \quad (b' \in L)$.

Intuitively, $t_{z(i)-b}$ represents $|z(i) - b|^q$. The last two constraints encode triangle inequalities; for example, the first one corresponds to $|z(i) - b|^q \leq |z(i) - b'|^q + |b - b'|^q$. This increases the number of variables and constraints by $O(n/\delta)$ and $O(n/\delta^2)$, respectively. For the sake of notational simplicity, we will denote $t_{z(i)-b}$ by $|z(i) - b|^q$ for $i \in [n]$ and $b \in L$ in POPs. Note that these constraints are valid in the sense that the desired solution $\boldsymbol{x}^L$ satisfies all of them.

Thus we arrive at the following POP:

$$\begin{aligned} &\text{minimize} && \|\boldsymbol{z}\|_q^q \\ &\text{subject to} && \boldsymbol{z} \text{ satisfies the system} \\ &&& \{\|A\boldsymbol{z}-\boldsymbol{y}\|_2^2 \le \eta^2\} \cup \mathcal{P}, \end{aligned}$$

where $\mathcal{P}$ is a system of constraints. Here, $\mathcal{P}$ contains all the constraints on $t_{z(i)-b}$ mentioned above but not limited to. Indeed, in Section 4, we add other (valid) constraints to $\mathcal{P}$ so that the matrix A satisfies the pseudo robust null space property. From now on, we also fix $\mathcal{P}$.

Now, we consider the following moment relaxation:

$$(\tilde{P}_{q,\eta}) \quad \begin{aligned} &\text{minimize} && \widetilde{\mathbf{E}}_{\mathcal{D}}\|\boldsymbol{z}\|_q^q \\ &\text{subject to} && \mathcal{D} \text{ satisfies the system} \\ &&& \{\|A\boldsymbol{z}-\boldsymbol{y}\|_2^2 \le \eta^2\} \cup \mathcal{P}, \end{aligned}$$

PROPOSITION 3.1. *Let $n_{\mathcal{P}}$ be the number of variables in $\mathcal{P}$ and $d_{\mathcal{P}}$ be the maximum degree that appears in the system $\mathcal{P}$. If there is a polynomial-time separation oracle for constraints in $\mathcal{P}$, then we can solve the moment relaxation $(\tilde{P}_{q,\eta})$ in $O(n/\delta+n_{\mathcal{P}})^{O(2/q+d_{\mathcal{P}})}$ time.*

Proof. Immediate from Theorem 1.3.

Since $\boldsymbol{x}^L \in L^n$, the value $\widetilde{\mathbf{E}}\|\boldsymbol{z}-\boldsymbol{x}^L\|_q^q$ is well defined. Furthermore, $\mathcal{D}$ satisfies the following triangle inequalities:

$$\begin{aligned} \widetilde{\mathbf{E}}\|\boldsymbol{z}\|_q^q &\le \widetilde{\mathbf{E}}\|\boldsymbol{z}-\boldsymbol{x}^L\|_q^q + \|\boldsymbol{x}^L\|_q^q, \\ \widetilde{\mathbf{E}}\|\boldsymbol{z}-\boldsymbol{x}^L\|_q^q &\le \widetilde{\mathbf{E}}\|\boldsymbol{z}\|_q^q + \|\boldsymbol{x}^L\|_q^q, \\ \|\boldsymbol{x}^L\|_q^q &\le \widetilde{\mathbf{E}}\|\boldsymbol{z}-\boldsymbol{x}^L\|_q^q + \widetilde{\mathbf{E}}\|\boldsymbol{z}\|_q^q. \end{aligned}$$

Now we formally define pseudo robust null space property.

DEFINITION 3.1. *Let $\rho, \tau > 0$ and $s \in \mathbb{Z}_+$. A matrix $A \in \mathbb{R}^{m\times n}$ is said to satisfy the (ρ,τ,s)* pseudo robust null space property *in ℓ_q quasi-norm (ℓ_q-(ρ,τ,s)-PRNSP for short) with respect to the system $\mathcal{P}$ if*

$$\widetilde{\mathbf{E}}_{\mathcal{D}}\|\boldsymbol{v}_S\|_q^q \le \rho\,\widetilde{\mathbf{E}}_{\mathcal{D}}\|\boldsymbol{v}_{\overline{S}}\|_q^q + \tau\left(\widetilde{\mathbf{E}}_{\mathcal{D}}\|A\boldsymbol{v}\|_2^2\right)^{q/2}$$

holds for any pseudodistribution $\mathcal{D}$ of degree at least $2/q$ satisfying $\mathcal{P}$ and all $\boldsymbol{v}$ of the form $\boldsymbol{z}-\boldsymbol{b}$ ($\boldsymbol{b} \in L^n$), and for any $S \subseteq [n]$ with $|S| = s$.

LEMMA 3.1. *Let $0 < \rho < 1$, $\tau > 0$, and $s \in \mathbb{Z}_+$. If the matrix $A \in \mathbb{R}^{m\times n}$ satisfies the ℓ_q-(ρ,τ,s)-PRNSP with respect to the system $\mathcal{P}$, then we have*

$$\begin{aligned} \widetilde{\mathbf{E}}\|\boldsymbol{z}-\boldsymbol{x}^L\|_q^q \le{}& \frac{1+\rho}{1-\rho}\left(\widetilde{\mathbf{E}}\|\boldsymbol{z}\|_q^q - \|\boldsymbol{x}^L\|_q^q + 2\|\boldsymbol{x}^L_{\overline{S}}\|_q^q\right) \\ &+ \frac{2\tau}{1-\rho}\left(\widetilde{\mathbf{E}}\|A(\boldsymbol{z}-\boldsymbol{x}^L)\|_2^2\right)^{q/2} \end{aligned}$$

for any $S \subseteq [n]$ with $|S| \le s$.

Proof. Let $\boldsymbol{v} := \boldsymbol{z}-\boldsymbol{x}^L$. Then from the PRNSP of A, we have

$$\widetilde{\mathbf{E}}\|\boldsymbol{v}_S\|_q^q \le \rho\,\widetilde{\mathbf{E}}\|\boldsymbol{v}_{\overline{S}}\|_q^q + \tau\left(\widetilde{\mathbf{E}}\|A\boldsymbol{v}\|_2^2\right)^{q/2}. \tag{3.3}$$

Note that

$$\begin{aligned} \|\boldsymbol{x}^L\|_q^q &= \|\boldsymbol{x}^L_S\|_q^q + \|\boldsymbol{x}^L_{\overline{S}}\|_q^q \\ &\le \widetilde{\mathbf{E}}\|(\boldsymbol{z}-\boldsymbol{x}^L)_S\|_q^q + \widetilde{\mathbf{E}}\|\boldsymbol{z}_S\|_q^q + \|\boldsymbol{x}^L_{\overline{S}}\|_q^q \\ &= \widetilde{\mathbf{E}}\|\boldsymbol{v}_S\|_q^q + \widetilde{\mathbf{E}}\|\boldsymbol{z}_S\|_q^q + \|\boldsymbol{x}^L_{\overline{S}}\|_q^q, \\ \widetilde{\mathbf{E}}\|\boldsymbol{v}_{\overline{S}}\|_q^q &= \widetilde{\mathbf{E}}\|(\boldsymbol{z}-\boldsymbol{x}^L)_{\overline{S}}\|_q^q \le \|\boldsymbol{x}^L_{\overline{S}}\|_q^q + \widetilde{\mathbf{E}}\|\boldsymbol{z}_{\overline{S}}\|_q^q. \end{aligned}$$

Summing up these two inequalities, we get

$$\widetilde{\mathbf{E}}\|\boldsymbol{v}_{\overline{S}}\|_q^q \le \widetilde{\mathbf{E}}\|\boldsymbol{z}\|_q^q - \|\boldsymbol{x}^L\|_q^q + \widetilde{\mathbf{E}}\|\boldsymbol{v}_S\|_q^q + 2\|\boldsymbol{x}^L_{\overline{S}}\|_q^q. \tag{3.4}$$

Combining (3.3) and (3.4), we get

$$\begin{aligned} \widetilde{\mathbf{E}}\|\boldsymbol{v}_{\overline{S}}\|_q^q \le{}& \frac{1}{1-\rho}\left(\widetilde{\mathbf{E}}\|\boldsymbol{z}\|_q^q - \|\boldsymbol{x}^L\|_q^q + 2\|\boldsymbol{x}^L_{\overline{S}}\|_q^q\right) \\ &+ \frac{\tau}{1-\rho}\left(\widetilde{\mathbf{E}}\|A\boldsymbol{v}\|_2^2\right)^{q/2}. \end{aligned}$$

Therefore,

$$\begin{aligned} \widetilde{\mathbf{E}}\|\boldsymbol{v}\|_q^q &= \widetilde{\mathbf{E}}\|\boldsymbol{v}_S\|_q^q + \widetilde{\mathbf{E}}\|\boldsymbol{v}_{\overline{S}}\|_q^q \\ &\le (1+\rho)\,\widetilde{\mathbf{E}}\|\boldsymbol{v}_{\overline{S}}\|_q^q + \tau\left(\widetilde{\mathbf{E}}\|A\boldsymbol{v}\|_2^2\right)^{q/2} \\ &\le \frac{1+\rho}{1-\rho}\left(\widetilde{\mathbf{E}}\|\boldsymbol{z}\|_q^q - \|\boldsymbol{x}^L\|_q^q + 2\|\boldsymbol{x}^L_{\overline{S}}\|_q^q\right) \\ &\quad + \frac{2\tau}{1-\rho}\left(\widetilde{\mathbf{E}}\|A\boldsymbol{v}\|_2^2\right)^{q/2}. \end{aligned}$$

COROLLARY 3.1. *Let $0 < \rho < 1$, $\tau > 0$, and $s \in \mathbb{Z}_+$. If the matrix $A \in \mathbb{R}^{m\times n}$ satisfies the ℓ_q-(ρ,τ,s)-PRNSP with respect to the system $\mathcal{P}$, then we have*

$$\widetilde{\mathbf{E}}\|\boldsymbol{z}-\boldsymbol{x}^L\|_q^q \le \frac{2(1+\rho)}{1-\rho}\sigma_s(\boldsymbol{x}^L)_q^q + \frac{2^{1+q}\tau}{1-\rho}\eta^q.$$

Proof. Note that $\widetilde{\mathbf{E}}\|\boldsymbol{z}\|_q^q - \|\boldsymbol{x}^L\|_q^q \le 0$ and $\left(\widetilde{\mathbf{E}}\|A(\boldsymbol{z}-\boldsymbol{x}^L)\|_2^2\right)^{q/2} \le \left(\widetilde{\mathbf{E}}\|A\boldsymbol{z}-\boldsymbol{y}\|_2^2\right)^{q/2} + \left(\|\boldsymbol{y}-A\boldsymbol{x}^L\|_2^2\right)^{q/2} \le 2\eta^q$ by Lemma 2.3 and the concavity of $|\cdot|^q$. The statement is immediate from the previous lemma with S being an index set of top-s largest absolute entries of $\boldsymbol{x}$.

3.2 Rounding Suppose that we have obtained a pseudodistribution $\mathcal{D}$ as a solution to $(\tilde{P}_{q,\eta})$. Then, we use the following simple rounding scheme to extract an actual vector $\bar{\boldsymbol{z}} \in \mathbb{R}^n$: For each $i \in [n]$, find $b \in L$ that minimizes $\widetilde{\mathbf{E}}_{\boldsymbol{z}\sim\mathcal{D}}|z(i)-b|^q$ and set $\bar{z}(i) = b$. In other words, $\bar{\boldsymbol{z}}$ is a minimizer of $\widetilde{\mathbf{E}}_{\boldsymbol{z}\sim\mathcal{D}}\|\boldsymbol{z}-\boldsymbol{b}\|_q^q$, where $\boldsymbol{b}$ is over vectors in L^n. Clearly, we can find $\bar{\boldsymbol{z}}$ in $O(n/\delta)$ time. Also, we have the following guarantee on its distance to $\boldsymbol{x}^L$.

LEMMA 3.2. *Let $0 < \rho < 1$, $\tau > 0$, and $s \in \mathbb{Z}_+$. If the matrix $A \in \mathbb{R}^{m\times n}$ satisfies the ℓ_q-(ρ,τ,s)-PRNSP with respect to the system $\mathcal{P}$, then the vector $\bar{\boldsymbol{z}} \in L^n$ satisfies*

$$\|\bar{\boldsymbol{z}} - \boldsymbol{x}^L\|_q^q \le 2\left(\frac{1+\rho}{1-\rho}\sigma_s(\boldsymbol{x}^L)_q^q + \frac{2^{1+q}\tau}{1-\rho}\eta^q\right). \tag{3.5}$$

Proof. By Corollary 3.1, we have

$$\widetilde{\mathbf{E}}\|\boldsymbol{z} - \boldsymbol{x}^L\|_q^q \le \frac{1+\rho}{1-\rho}\sigma_s(\boldsymbol{x}^L)_q^q + \frac{2^{1+q}\tau}{1-\rho}\eta^q.$$

Since $\boldsymbol{x}^L \in L^n$ holds and $\bar{\boldsymbol{z}}$ is a minimizer of $\widetilde{\mathbf{E}}\|\boldsymbol{z}-\boldsymbol{b}\|_q^q$, we also have

$$\widetilde{\mathbf{E}}\|\boldsymbol{z} - \bar{\boldsymbol{z}}\|_q^q \le \frac{1+\rho}{1-\rho}\sigma_s(\boldsymbol{x}^L)_q^q + \frac{2^{1+q}\tau}{1-\rho}\eta^q.$$

Then by the triangle inequality,

$$\begin{aligned}\|\bar{\boldsymbol{z}} - \boldsymbol{x}^L\|_q^q &\le \widetilde{\mathbf{E}}\|\boldsymbol{z} - \bar{\boldsymbol{z}}\|_q^q + \widetilde{\mathbf{E}}\|\boldsymbol{z} - \boldsymbol{x}^L\|_q^q \\ &\le 2\left(\frac{1+\rho}{1-\rho}\sigma_s(\boldsymbol{x}^L)_q^q + \frac{2^{1+q}\tau}{1-\rho}\eta^q\right).\end{aligned}$$

Then, we show that $\bar{\boldsymbol{z}}$ is close to the signal vector $\boldsymbol{x}$ in ℓ_q^q metric.

LEMMA 3.3. *Let $0 < \rho < 1$, $\tau > 0$, and $s \in \mathbb{Z}_+$. If the matrix $A \in \mathbb{R}^{m\times n}$ satisfies the ℓ_q-(ρ,τ,s)-PRNSP with respect to the system $\mathcal{P}$, then the vector $\bar{\boldsymbol{z}}$ satisfies*

$$\|\bar{\boldsymbol{z}} - \boldsymbol{x}\|_q^q \le \frac{2(1+\rho)}{1-\rho}\sigma_s(\boldsymbol{x})_q^q + O(\epsilon)$$

by choosing

$$\delta = \min\{\left((1-\rho)\epsilon/\tau\right)^{1/q}/\left(\sigma_{\max}(A)\sqrt{s}\right), ((1-\rho)\epsilon/n)^{1/q}\}.$$

Proof. From the choice of δ, we have $\eta = \sigma_{\max}(A)\sqrt{s}\delta \le \left(\epsilon(1-\rho)/\tau\right)^{1/q}$. Then from Lemma 3.2, we get

$$\|\bar{\boldsymbol{z}} - \boldsymbol{x}^L\|_q^q \le \frac{2(1+\rho)}{1-\rho}\sigma_s(\boldsymbol{x}^L)_q^q + O(\epsilon).$$

Note that we have $\|\boldsymbol{x}^L - \boldsymbol{x}\|_q^q \le n\delta^q$ and $|\sigma_s(\boldsymbol{x})_q^q - \sigma_s(\boldsymbol{x}^L)_q^q| = O(n\delta^q)$. Then, we have

$$\begin{aligned}\|\bar{\boldsymbol{z}} - \boldsymbol{x}\|_q^q &\le \|\bar{\boldsymbol{z}} - \boldsymbol{x}^L\|_q^q + \|\boldsymbol{x}^L - \boldsymbol{x}\|_q^q \\ &\le \frac{2(1+\rho)}{1-\rho}(\sigma_s(\boldsymbol{x})_q^q + n\delta^q) + O(\epsilon) + n\delta^q \\ &= \frac{2(1+\rho)}{1-\rho}\sigma_s(\boldsymbol{x})_q^q + O\Big(\epsilon + \frac{n\delta^q}{1-\rho}\Big) \\ &= \frac{2(1+\rho)}{1-\rho}\sigma_s(\boldsymbol{x})_q^q + O(\epsilon).\end{aligned}$$

Finally, we show that $\bar{\boldsymbol{z}}$ is close to $\boldsymbol{x}$ in ℓ_q quasi-norm.

COROLLARY 3.2. *Let $0 < \rho < 1$, $\tau > 0$, and $s \in \mathbb{Z}_+$. If the matrix $A \in \mathbb{R}^{m\times n}$ satisfies the ℓ_q-(ρ,τ,s)-PRNSP with respect to the system $\mathcal{P}$, then the vector $\bar{\boldsymbol{z}}$ satisfies*

$$\|\bar{\boldsymbol{z}} - \boldsymbol{x}\|_q \le \frac{2(1+\rho)}{1-\rho}\sigma_s(\boldsymbol{x})_q + O(\epsilon)$$

by choosing $\delta = O\Big(\big(\frac{\epsilon(1-\rho)^{1/q+1}}{\tau n^{1/q}}\big)^{1/q}/\big(\sigma_{\max}(A)\sqrt{s}\big)\Big)$.

Proof. Invoke Lemma 3.3 with

$$\epsilon' = \epsilon\cdot\Big(\frac{1-\rho}{2(1+\rho)\sigma_s(\boldsymbol{x})_q^q}\Big)^{1/q-1} = \Omega\Big(\epsilon\big(\frac{1-\rho}{n}\big)^{1/q}\Big).$$

Then, we get a vector $\bar{\boldsymbol{z}}$ such that

$$\begin{aligned}\|\bar{\boldsymbol{z}} - \boldsymbol{x}\|_q &\le \Big(\frac{2(1+\rho)}{1-\rho}\sigma_s(\boldsymbol{x})_q^q + \epsilon'\Big)^{1/q} \\ &\le \frac{2(1+\rho)}{1-\rho}\sigma_s(\boldsymbol{x})_q + O\Big(\epsilon'\Big(\frac{2(1+\rho)}{1-\rho}\sigma_s(\boldsymbol{x})_q^q\Big)^{1/q-1}\Big) \\ &\le \frac{2(1+\rho)}{1-\rho}\sigma_s(\boldsymbol{x})_q + O(\epsilon).\end{aligned}$$

Note that the required δ is

$$\begin{aligned}&\Omega\Big(\min\Big\{\Big(\frac{(1-\rho)\epsilon'}{\tau}\Big)^{1/q}/\big(\sigma_{\max}(A)\sqrt{s}\big), \Big(\frac{(1-\rho)\epsilon'}{n}\Big)^{1/q}\Big\}\Big) \\ &= \Omega\Big(\big(\frac{\epsilon(1-\rho)^{1/q+1}}{\tau n^{1/q}}\big)^{1/q}/\big(\sigma_{\max}(A)\sqrt{s}\big)\Big).\end{aligned}$$

Combining Proposition 3.1 and Corollary 3.2, we get the following theorem.

THEOREM 3.1. (FORMAL VERSION OF THEOREM 1.4) *Let $0 < \rho < 1$, $\tau > 0$, and $s \in \mathbb{Z}_+$. If the matrix $A \in \mathbb{R}^{m\times n}$ satisfies the ℓ_q-(ρ,τ,s)-PRNSP with respect to the system $\mathcal{P}$ and $\mathcal{P}$ has a polynomial-time separation oracle, then we can compute a vector $\bar{\boldsymbol{z}} \in \mathbb{R}^n$ such that*

$$\|\bar{\boldsymbol{z}} - \boldsymbol{x}\|_q \le \frac{2(1+\rho)}{1-\rho}\sigma_s(\boldsymbol{x})_q + \epsilon.$$

The running time is $(n/\delta + n_{\mathcal{P}})^{O(2/q+d_{\mathcal{P}})}$, where $n_{\mathcal{P}}$ is the number of variables in $\mathcal{P}$, $d_{\mathcal{P}}$ is the maximum degree that appears in constraints of $\mathcal{P}$, and $\delta = \Omega\Big(\big(\frac{\epsilon(1-\rho)^{1/q+1}}{\tau n^{1/q}}\big)^{1/q}/\big(\sigma_{\max}(A)\sqrt{s}\big)\Big)$.

4 Imposing Pseudo Robust Null Space Property

A *normalized Rademacher matrix* $A \in \mathbb{R}^{m\times n}$ is a random matrix whose each entry is i.i.d. sampled from $\{\pm 1/\sqrt{m}\}$. In this section, we show that we can impose the PRNSP to the normalized Rademacher matrix by introducing some additional variables and constraints to pseudodistributions. Now we introduce coherence and review properties of a normalized Rademacher matrix.

Definition 4.1. (Coherence) *Let $A \in \mathbb{R}^{m\times n}$ be a matrix whose columns $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ are ℓ^2-normalized. The ℓ_q-*coherence function *of A is defined as* $\mu_q^q(s) := \max_{i\in[n]} \max\left\{\sum_{j\in S} |\langle \boldsymbol{a}_i, \boldsymbol{a}_j\rangle|^q : S \subseteq [n], |S| = s, i \notin S\right\}$ *for* $s = 1, \ldots, n-1$.

Now we review properties of a normalized Rademacher matrix.

Lemma 4.1. ([1]) *Let $n \in \mathbb{Z}_+$ and $\epsilon > 0$. If $m = \Omega(\log n/\epsilon^2)$, then a normalized Rademacher matrix $A \in \mathbb{R}^{m\times n}$ with columns $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ satisfies $|\langle \boldsymbol{a}_i, \boldsymbol{a}_j\rangle| \le \epsilon$ for all $i \ne j$ with high probability.*

Corollary 4.1. *Let $s, n \in \mathbb{Z}_+$, $\epsilon > 0$. If $m = \Omega(s^{2/q} \log n/\epsilon^{2/q})$, then a normalized Rademacher matrix $A \in \mathbb{R}^{m\times n}$ with columns $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ satisfies $\mu_q^q(s) \le \epsilon$ with high probability.*

Lemma 4.2. ([38]) *A normalized Rademacher matrix $A \in \mathbb{R}^{m\times n}$ with $m \le n$ satisfies $\sigma_{\max}(A) = O(\frac{n\log m}{m})$.*

In what follows, we fix $q \in (0, 1]$, $s \in \mathbb{Z}_+$, $\epsilon > 0$, and a matrix $A \in \mathbb{R}^{m\times n}$ that satisfies the properties of Corollary 4.1 and Lemma 4.2. Let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ be the columns of A.

For $\boldsymbol{b} \in L^n$, we define $\boldsymbol{v^b} = \boldsymbol{z} - \boldsymbol{b}$, where $\boldsymbol{z}$ is a variable in POP. To impose the PRNSP to A, for each $i \in [n]$ and $\boldsymbol{b} \in L^n$, we add a variable $t_{\langle A\boldsymbol{v^b}, \boldsymbol{a}_i\rangle}$ and constraints $t^{2/q}_{\langle A\boldsymbol{v^b}, \boldsymbol{a}_i\rangle} = \langle A\boldsymbol{v^b}, \boldsymbol{a}_i\rangle^2$ and $t_{\langle A\boldsymbol{v^b}, \boldsymbol{a}_i\rangle} \ge 0$. Intuitively, $t_{\langle A\boldsymbol{v^b}, \boldsymbol{a}_i\rangle}$ represents $|\langle A\boldsymbol{v^b}, \boldsymbol{a}_i\rangle|^q$, and as with the previous section, we will denote $t_{\langle A\boldsymbol{v^b}, \boldsymbol{a}_i\rangle}$ by $|\langle A\boldsymbol{v^b}, \boldsymbol{a}_i\rangle|^q$ in POPs for notational simplicity.

Furthermore, we add the following constraints for each $i \in [n]$ and $\boldsymbol{b} \in L^n$,

$$|v^{\boldsymbol{b}}(i)|^q \le |\langle A\boldsymbol{v^b}, \boldsymbol{a}_i\rangle|^q + \sum_{j\ne i} |v^{\boldsymbol{b}}(j)|^q |\langle \boldsymbol{a}_j, \boldsymbol{a}_i\rangle|^q. \tag{4.6}$$

These constraints are valid since $v^{\boldsymbol{b}}(i)\boldsymbol{a}_i = A\boldsymbol{v^b} - \sum_{j\ne i} v^{\boldsymbol{b}}(j)\boldsymbol{a}_j$ and $\langle \boldsymbol{a}_i, \boldsymbol{a}_i\rangle = 1$.

At first sight, the resulting moment relaxation is intractable since there are exponentially many new variables. However, this is not the case. To see this, note that $\langle A\boldsymbol{v^b}, \boldsymbol{a}_i\rangle = \sum_{j=1}^n z(i)\langle \boldsymbol{a}_j, \boldsymbol{a}_i\rangle - \sum_{j=1}^n b(j)\langle \boldsymbol{a}_j, \boldsymbol{a}_i\rangle$. Also, the number of possible values of $\sum_{j=1}^n b(j)\langle \boldsymbol{a}_j, \boldsymbol{a}_i\rangle$ is $O(nm/\delta)$ because A is $\pm 1/\sqrt{m}$-valued and $b(j) \in [-1, 1]$ is a multiple of δ. Hence, identifying variables $t_{\langle A\boldsymbol{v^b}, \boldsymbol{a}_i\rangle}$ and $t_{\langle A\boldsymbol{v}_{\boldsymbol{b}'}, \boldsymbol{a}_i\rangle}$ if $\sum_{j=1}^n b(j)\langle \boldsymbol{a}_j, \boldsymbol{a}_i\rangle = \sum_{j=1}^n b'(j)\langle \boldsymbol{a}_j, \boldsymbol{a}_i\rangle$, we only have to add $O(nm/\delta)$ variables.

Another issue is that we have added exponentially many constraints. The following lemma says that we can design a separation oracle for (4.6), and hence we can use the ellipsoid method to solve SoS relaxations involving these constraints.

Lemma 4.3. *There is a polynomial time algorithm that validates whether* (4.6) *is satisfied by a given pseudodistribution $\mathcal{D}$, and if not, it finds $i \in [n]$ and $\boldsymbol{b} \in L^n$ such that*

$$\widetilde{\mathbf{E}}|v^{\boldsymbol{b}}(i)|^q > \widetilde{\mathbf{E}}|\langle A\boldsymbol{v^b}, \boldsymbol{a}_i\rangle|^q + \sum_{j\ne i} \widetilde{\mathbf{E}}|v^{\boldsymbol{b}}(j)|^q |\langle \boldsymbol{a}_j, \boldsymbol{a}_i\rangle|^q.$$

Proof. Since there are only polynomially many choices for the values of i, $b(i)$, and $\sum_{j=1}^n b(j)\langle \boldsymbol{a}_j, \boldsymbol{a}_i\rangle$, we can exhaustively check all of them. Hence, in what follows, we fix $i \in [n]$, $b_i := b(i)$, and $c := \sum_{j=1}^n b(j)\langle \boldsymbol{a}_j, \boldsymbol{a}_i\rangle$. Note that, by fixing c, we have also fixed the value of $\widetilde{\mathbf{E}}|\langle A\boldsymbol{v^b}, \boldsymbol{a}_i\rangle|^q$ to $c' := \widetilde{\mathbf{E}}\,|\langle A\boldsymbol{z}, \boldsymbol{a}_i\rangle - c|^q$.

Now we want to check whether there exists $\boldsymbol{b} \in L^n$ with $b(i) = b_i$ and $\sum_{j=1}^n b(j)\langle \boldsymbol{a}_j, \boldsymbol{a}_i\rangle = c$ such that $\widetilde{\mathbf{E}}|z(i) - b_i|^q > c' + \sum_{j\ne i} \widetilde{\mathbf{E}}|z(j) - b(j)|^q |\langle \boldsymbol{a}_j, \boldsymbol{a}_i\rangle|^q$. We can solve this problem by dynamic programming using a table $T[\cdot,\cdot]$, where $T[k, a]$ stores the minimum value of $\sum_{j\in[k], j\ne i} \widetilde{\mathbf{E}}|z(j) - b(j)|^q |\langle \boldsymbol{a}_j, \boldsymbol{a}_i\rangle|^q$ conditioned on $\sum_{j\in[k], j\ne i} b(j)\langle \boldsymbol{a}_j, \boldsymbol{a}_i\rangle = a$. Since the size of T is $O(n \cdot nm/\delta)$, the dynamic programming can be solved in polynomial time.

Let $\mathcal{P}'$ be the system of all the added constraints on $t_{z(i)-b}$ (given in Section 3) and constraints on $t_{\langle A\boldsymbol{v^b}, \boldsymbol{a}_i\rangle}$. Then, A satisfies the PRNSP by imposing the system $\mathcal{P}'$.

Theorem 4.1. (formal version of Theorem 1.5) *The matrix A satisfies the ℓ_q-(ρ, τ, s)-PRNSP with respect to $\mathcal{P}'$ with $\rho = \frac{\mu_q^q(s)}{1-\mu_q^q(s-1)} = O(\epsilon)$ and $\tau = \frac{s}{1-\mu_q^q(s-1)} = O(s)$.*

Proof. The proof basically follows the standard proof in compressed sensing (see [23]). Let us fix $S \subseteq [n]$ of size s and $\boldsymbol{b} \in L^n$. Then for any pseudodistribution $\mathcal{D}$ satisfying the above constraints, we have

$$\begin{aligned}\widetilde{\mathbf{E}}|v^{\boldsymbol{b}}(i)|^q \le \widetilde{\mathbf{E}}|\langle A\boldsymbol{v^b}, \boldsymbol{a}_i\rangle|^q + \widetilde{\mathbf{E}}\sum_{l\in\bar{S}} |v^{\boldsymbol{b}}(l)|^q |\langle \boldsymbol{a}_l, \boldsymbol{a}_i\rangle|^q \\ + \widetilde{\mathbf{E}}\sum_{j\in S: j\ne i} |v^{\boldsymbol{b}}(j)|^q |\langle \boldsymbol{a}_j, \boldsymbol{a}_i\rangle|^q\end{aligned}$$

for $i \in S$. Using the pseudo Cauchy Schwartz $k = -\log_2 q$ times, we get

$$\left(\widetilde{\mathbf{E}}|\langle A\boldsymbol{v^b}, \boldsymbol{a}_i\rangle|^q\right)^{2/q} \le \widetilde{\mathbf{E}}\langle A\boldsymbol{v^b}, \boldsymbol{a}_i\rangle^2.$$

By Lemma 2.4, $\widetilde{\mathbf{E}}\langle A\boldsymbol{v}^{\boldsymbol{b}}, \boldsymbol{a}_i\rangle^2 \leq \widetilde{\mathbf{E}}\|A\boldsymbol{v}^{\boldsymbol{b}}\|_2^2\|\boldsymbol{a}_i\|_2^2 = \widetilde{\mathbf{E}}\|A\boldsymbol{v}^{\boldsymbol{b}}\|_2^2$. Thus we have

$$\widetilde{\mathbf{E}}|v^{\boldsymbol{b}}(i)|^q \leq \left(\widetilde{\mathbf{E}}\|A\boldsymbol{v}^{\boldsymbol{b}}\|_2^2\right)^{q/2} + \widetilde{\mathbf{E}}\sum_{l\in\bar{S}}|v^{\boldsymbol{b}}(l)|^q|\langle \boldsymbol{a}_l, \boldsymbol{a}_i\rangle|^q + \widetilde{\mathbf{E}}\sum_{j\in S:j\neq i}|v^{\boldsymbol{b}}(j)|^q|\langle \boldsymbol{a}_j, \boldsymbol{a}_i\rangle|^q.$$

Summing up these inequalities yields

$$\begin{aligned}\widetilde{\mathbf{E}}\|\boldsymbol{v}_S^{\boldsymbol{b}}\|_q^q &\leq s\left(\widetilde{\mathbf{E}}\|A\boldsymbol{v}^{\boldsymbol{b}}\|_2^2\right)^{q/2} + \sum_{l\in\bar{S}}\widetilde{\mathbf{E}}|v^{\boldsymbol{b}}(l)|^q\sum_{i\in S}|\langle \boldsymbol{a}_l, \boldsymbol{a}_i\rangle|^q \\ &\quad + \sum_{j\in S}\widetilde{\mathbf{E}}|v^{\boldsymbol{b}}(j)|^q\sum_{i\in S:i\neq j}|\langle \boldsymbol{a}_j, \boldsymbol{a}_i\rangle|^q \\ &\leq s\left(\widetilde{\mathbf{E}}\|A\boldsymbol{v}^{\boldsymbol{b}}\|_2^2\right)^{q/2} + \mu_q^q(s)\,\widetilde{\mathbf{E}}\|\boldsymbol{v}_{\bar{S}}^{\boldsymbol{b}}\|_q^q \\ &\quad + \mu_q^q(s-1)\,\widetilde{\mathbf{E}}\|\boldsymbol{v}_S^{\boldsymbol{b}}\|_q^q.\end{aligned}$$

Rearranging this inequality finishes the proof.

Combining Lemma 4.2 and Theorem 4.1, we immediately have the following.

Corollary 4.2. *The matrix A satisfies the ℓ_q-(ρ, τ, s)-PRNSP with respect to $\mathcal{P}'$ with $\rho = O(\epsilon)$ and $\tau = O(s)$. Moreover, $\sigma_{\max}(A) = O((n\log m)/m)$.*

Now we establish our main theorem.

Theorem 4.2. (formal version of Theorem 1.2) *Let $q \in (0, 1]$ be a constant of the form $q = 2^{-k}$ for some $k \in \mathbb{Z}_+$, $s, n \in \mathbb{Z}_+$, and $\epsilon > 0$. Then, a normalized Rademacher matrix $A \in \mathbb{R}^{m\times n}$ with $m = O(s^{2/q}\log n)$ satisfies the following property with high probability: For any vector $\boldsymbol{y} = A\boldsymbol{x}$, where $\boldsymbol{x} \in [-1, 1]^n$ is an unknown vector, we can recover $\bar{\boldsymbol{z}}$ with $\|\bar{\boldsymbol{z}} - \boldsymbol{x}\|_q \leq O(\sigma_s(\boldsymbol{x})_q) + \epsilon$. The running time is $(nm/\delta)^{O(2/q)}$, where $\delta = O\Big(m\big(\frac{\epsilon}{sn^{1/q}}\big)^{1/q}/\big(\sqrt{s}n\log m\big)\Big)$.*

Proof. Solve the SoS relaxation $(\tilde{P}_{q,\eta})$ with the system $\mathcal{P}'$. Then, with high probability, the matrix A satisfies the ℓ_q-(ρ, τ, s)-PRNSP with respect to $\mathcal{P}'$ for $\rho = O(1)$ and $\tau = O(s)$, and $\sigma_{\max}(A) = O((n\log m)/m)$ with high probability by Corollary 4.2. Then, we can recover a vector $\bar{\boldsymbol{z}} \in \mathbb{R}^n$ with the desired property by Theorem 3.1. Note that the number $n_{\mathcal{P}'}$ of added variables in the system $\mathcal{P}'$ is $O(nm/\delta)$ and the maximum degree $d_{\mathcal{P}'}$ that appears in a constraint of $\mathcal{P}'$ is $2/q$. Hence, the running time becomes as stated.

Acknowledgements

The authors thank Satoru Iwata and Yuji Nakatsukasa for reading an earlier version of this paper.

References

[1] D. Achlioptas. Database-friendly random projections: Johnson-Lindenstrauss with binary coins. *Journal of Computer and System Sciences*, 66(4):671–687, 2003.

[2] M. F. Anjos and J. B. Lasserre. *Handbook on Semidefinite, Conic and Polynomial Optimization.* Springer Science & Business Media, 2011.

[3] B. Barak, F. G. S. L. Brandao, A. W. Harrow, J. Kelner, D. Steurer, and Y. Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *Proceedings of the 44th Annual ACM Symposium on Theory of Computing (STOC)*, pages 307–326, 2012.

[4] B. Barak, J. A. Kelner, and D. Steurer. Rounding sum-of-squares relaxations. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing (STOC)*, pages 31–40, 2014.

[5] B. Barak, J. A. Kelner, and D. Steurer. Dictionary learning and tensor decomposition via the sum-of-squares method. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing (STOC)*, pages 143–151, 2015.

[6] B. Barak and A. Moitra. Tensor prediction, Rademacher complexity and random 3-XOR. *CoRR*, abs/1501.06521, 2015.

[7] B. Barak and D. Steurer. Sum-of-squares proofs and the quest toward optimal algorithms. In *Proceedings of International Congress of Mathematicians (ICM)*, pages 509–533, 2014.

[8] R. Berinde and P. Indyk. Sequential sparse matching pursuit. In *Proceedings of the 47th Annual Allerton Conference on Communication, Control, and Computing*, pages 36–43, 2009.

[9] R. Berinde, P. Indyk, and M. Ruzic. Practical near-optimal sparse recovery in the L1 norm. In *Proceedings of the 46th Annual Allerton Conference on Communication, Control, and Computing*, pages 198–205, 2008.

[10] A. M. Bruckstein, D. L. Donoho, and M. Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Review*, 51(1):34–81, 2009.

[11] E. J. Candès. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathematique*, 346(9-10):589–592, 2008.

[12] E. J. Candès, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006.

[13] E. J. Candès and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, 2005.

[14] R. Chartrand. Exact reconstruction of sparse signals via nonconvex minimization. *Signal Processing Letters*, 14(10):707–710.

[15] R. Chartrand. Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data. In *Proceedings of the IEEE International Symposium*

on Biomedical Imaging: From Nano to Macro (ISBI), pages 262–265, 2009.

[16] R. Chartrand and V. Staneva. Restricted isometry properties and nonconvex compressive sensing. *Inverse Problems*, 24(3), 2008.

[17] R. Chartrand and W. Yin. Iteratively reweighted algorithms for compressive sensing. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3869–3872, 2008.

[18] A. Cherian, S. Sra, and N. Papanikolopoulos. Denoising sparse noise via online dictionary learning. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2060–2063, 2011.

[19] E. Chlamtac and M. Tulsiani. Convex relaxations and integrality gaps. In *Handbook on Semidefinite, Conic and Polynomial Optimization*. 2012.

[20] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k-term approximation. *Journal of the American Mathematical Society*, 22(1):211–231, 2009.

[21] B. G. R. de Prony. Essay experimental et analytique: sur les lois de la dilatabilite de fluides elastique et sur celles de la force expansive de la vapeur de l'alcool, a differentes temperatures. *Journal de l'ecole Polytechnique*, pages 24–76, 1795.

[22] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[23] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. Springer, 2013.

[24] D. Ge, X. Jiang, and Y. Ye. A note on the complexity of L_p minimization. *Mathematical Programming*, 129(2):285–299, 2011.

[25] R. Ge and T. Ma. Decomposing overcomplete 3rd order tensors using sum-of-squares algorithms. *CoRR*, abs/1504.05287, 2015.

[26] A. Gilbert and P. Indyk. Sparse recovery using sparse matrices. *Proceedings of the IEEE*, 6:937–947, 2010.

[27] D. Grigoriev and N. Vorobjov. Complexity of null-and positivstellensatz proofs. *Annals of Pure and Applied Logic*, 113(1):153–160, 2001.

[28] P. Indyk and M. Ružić. Near-optimal sparse recovery in the L1 norm. In *Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 199–207, 2008.

[29] J. N. Laska, M. A. Davenport, and R. G. Baraniuk. Exact signal recovery from sparsely corrupted measurements through the pursuit of justice. *Conference Record of the 43rd Asilomar Conference on Signals, Systems and Computers (ACSSC)*, pages 1556–1560, 2009.

[30] J. B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2006.

[31] Y. Nesterov. Squared functional systems and optimization problems. In *High Performance Optimization*, pages 405–440. Springer, 2000.

[32] R. O'Donnell and Y. Zhou. Approximability and proof complexity. In *Proceedings of the 24th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1537–1556, 2013.

[33] P. A. Parrilo. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*. PhD thesis, California Institute of Technology, 2000.

[34] M. Putinar. Positive polynomials on compact semi-algebraic sets. *Indiana University Mathematics Journal*, 42(3):969–984, 1993.

[35] T. Rothvoß. The Lasserre hierarchy in approximation algorithms. *Lecture Notes for the MAPSP*, pages 1–25, 2013.

[36] R. Saab, R. Chartrand, and O. Yilmaz. Stable sparse approximations via nonconvex optimization. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3885–3888, 2008.

[37] K. Schmüdgen. The k-moment problem for compact semi-algebraic sets. *Mathematische Annalen*, 289(1):203–206, 1991.

[38] Y. Seginer. The expected norm of random matrices. *Combinatorics, Probability and Computing*, 9(2):149–166, 2000.

[39] N. Z. Shor. An approach to obtaining global extremums in polynomial mathematical programming problems. *Cybernetics*, 23(5):695–700, 1987.

[40] G. Tang and P. Shah. Guaranteed tensor decomposition: A moment approach. *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, pages 1491–1500, 2015.