Sample Complexity of Total Variation Minimization

Sajad Daei, Farzan Haddadi, Arash Amini

***Abstract*—This work considers the use of Total Variation (TV) minimization in the recovery of a given gradient sparse vector from Gaussian linear measurements. It has been shown in recent studies that there exists a sharp phase transition behavior in TV minimization for the number of measurements necessary to recover the signal in asymptotic regimes. The phase transition curve specifies the boundary of success and failure of TV minimization for large number of measurements. It is a challenging task to obtain a theoretical bound that reflects this curve. In this work, we present a novel upper-bound that suitably approximates this curve and is asymptotically sharp. Numerical results show that our bound is closer to the empirical TV phase transition curve than the previously known bound obtained by Kabanava.**

***Index Terms*—sample complexity, total variation minimization, phase transition.**

I. INTRODUCTION

COMPRESSED Sensing (CS) is a method to recover a sparse vector $\boldsymbol{x} \in \mathbb{R}^n$ from a few linear measurements $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \in \mathbb{R}^m$ where $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is the measurement matrix. In most cases in practice, the signal $\boldsymbol{x}$ is not sparse itself but there exists a dictionary such that $\boldsymbol{x} = \boldsymbol{D}\boldsymbol{\alpha}$ for some sparse $\boldsymbol{\alpha}$. This is known as synthesis sparsity and the following problem called ℓ_1 minimization in the synthesis form is considered for recovering $\boldsymbol{x}$:

$$\min_{\boldsymbol{z}\in\mathbb{R}^n} \|\boldsymbol{z}\|_1 \text{ s.t. } \boldsymbol{y} = \boldsymbol{A}\boldsymbol{D}\boldsymbol{z}. \tag{1}$$

In [1]–[3], recovery guarantees of this problem are studied. In general, one may not be able to correctly estimate $\boldsymbol{\alpha}$ from (1), but can hope for a good approximation of $\boldsymbol{x} = \boldsymbol{D}\boldsymbol{\alpha}$ [2]. The second approach to deal with such cases, is to focus on signals that are sparse after the application of an operator called analysis operator $\boldsymbol{\Omega}$ (See e.g. [3]–[5]). In the literature this is known as cosparsity or analysis sparsity. The following problem called ℓ_1 minimization in the analysis form is studied to estimate the signal $\boldsymbol{x}$:

$$\min_{\boldsymbol{z}\in\mathbb{R}^n} \|\boldsymbol{\Omega}\boldsymbol{z}\|_1 \text{ s.t. } \boldsymbol{y} = \boldsymbol{A}\boldsymbol{z}. \tag{2}$$

A special case of this problem that has great importance in a variaty of applications including image processing [1] is the case where $\boldsymbol{\Omega}$ is the one- or two-dimensional difference operator that leads to the total variation (TV) minimization problem which we call $\mathsf{P}_{\rm TV}$ from this point on.

Although many results in the CS literature have been established via Restricted Isometry Property (RIP) and Null Space Property (NSP) conditions (e.g. in [1], [6]–[9]), they fail to address gradient sparse [2] vectors (the rows of the difference matrix do not form a dictionary).

In a separate field of study, it is shown that the problem (2) undergoes a transition from failure to success (known as phase transition) as the number of measurements increases (e.g. see [10], [11]). Namely, there exist a curve $m = \Psi(s, \boldsymbol{\Omega})$ that the problem (2) succeeds to recover a gradient s-sparse vector with probability $\frac{1}{2}$. Obtaining a bound that approximates this curve has been an important and challenging task in recent years as it specifies the required number of measurements in problem (2) (See for example [11], [12]). This work revolves around this challenge. Specifically, we propose an upper-bound on $\Psi(s, \boldsymbol{\Omega})$ in the case of one dimensional difference operator

$$\boldsymbol{\Omega} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ & \ddots & \ddots & \ddots & \\ 0 & \cdots & \cdots & 1 & -1 \end{bmatrix} \in \mathbb{R}^{n-1\times n}.$$

A. Related Works

Despite the great importance of TV minimization in imaging sciences, few works have been established to find explicit formula for the number of measurements required for $\mathsf{P}_{\rm TV}$ to succeed [11]–[14]. In [14], Needle et al. transformed two-dimensional signals with low variations into those with compressible Haar wavelet coefficients. Then a modified RIP is considered for $\boldsymbol{A}$ to guarantee stable recovery. However, their proof does not hold for one-dimensional gradient sparse signals. In [13], a geometric approach based on "escape through a mesh lemma" is used to recover gradient s-sparse vectors from Gaussian measurements. Recently, in [12], Krahmer et al. provided an overview on the number of Gaussian linear measurements in TV minimization based on the mean empirical width [15], [16]. It is not evident from [13], [14] whether the obtained lower-bound on the number of measurements is sharp. The authors in [11] conjectured that the phase transition for the TV Approximate Message Passing (AMP) algorithm coincides with the number of measurements that $\mathsf{P}_{\rm TV}$ needs. However, this conjecture has not been proved yet. Furthermore, it is not simple nor practical to know the required number for TV AMP algorithm in advance. In [5], an upper-bound on $\Psi(s, \boldsymbol{\Omega})$ is proposed. The approach is based on generalizing the proofs of [17, Proposition 1] to TV minimization. In [18], a Monte Calro method is proposed to replace the involved expectation operators with empirical means to obtain $\Psi(s, \boldsymbol{\Omega})$; due to the numerical nature of the method, the computational cost might not be feasible in certain settings. Overall, [5] and [13] seem to be the only available bounds

S. Daei and F. Haddadi are with the School of Electrical Engineering, Iran University of Science & Technology. A. Amini is with EE department, Sharif University of Technology.

[1] Piecewise constant images are modeled as low variational functions.

[2] Low variational signal.

with closed-form expressions that approximate $\Psi(s, \boldsymbol{\Omega})$; we will discuss them in detail in Section III.

B. Outline of the paper

The paper is organized as follows. Section II provides a brief review of some concepts from convex geometry. Section III discusses our main contribution which determines an upper-bound on the sufficient number of Gaussian measurements for $\mathsf{P}_{\rm TV}$ to succeed. In Section IV, numerical experiments are presented to verify our theoretical bound. Finally, the paper is concluded in Section V.

C. Notation

Throughout the paper, scalars are denoted by lowercase letters, vectors by lowercase boldface letters, and matrices by uppercase boldface letters. The ith element of a vector $\boldsymbol{x}$ is shown either by $x(i)$ or x_i. $(\cdot)^\dagger$ denotes the pseudo inverse operation. We reserve calligraphic uppercase letters for sets (e.g. $\mathcal{S}$). The cardinality of a set $\mathcal{S}$ is shown by $|\mathcal{S}|$. $[n]$ refers to the set $\{1, ..., n\}$. Furthermore, we write $\bar{\mathcal{S}}$ for the complement $[n]\setminus\mathcal{S}$ of a set $\mathcal{S}$ in $[n]$. For a matrix $\boldsymbol{X} \in \mathbb{R}^{m\times n}$ and a subset $\mathcal{S} \subseteq [n]$, the notation $\boldsymbol{X}_\mathcal{S}$ is a matrix in $\mathbb{R}^{m\times n}$ consisting of the rows of $\boldsymbol{X}$ indexed by $\mathcal{S}$ and zero elsewhere. Similarly, for $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{x}_\mathcal{S}$ is the vector in $\mathbb{R}^n$ which coincides with $\boldsymbol{x}$ on the entries in $\mathcal{S}$ and is zero on the entries outside $\mathcal{S}$. Lastly, the polar $\mathcal{K}^\circ$ of a cone $\mathcal{K} \subset \mathbb{R}^n$ is the set of vectors forming non-acute angles with every vector in $\mathcal{K}$, i.e.

$$\mathcal{K}^\circ = \{\boldsymbol{v} \in \mathbb{R}^n : \langle \boldsymbol{v}, \boldsymbol{z}\rangle \leq 0 \ \forall \boldsymbol{z} \in \mathcal{K}\}. \tag{3}$$

II. Convex Geometry

In this section, basic concepts of convex geometry are reviewed.

A. Descent Cones

The descent cone of a proper convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ at point $\boldsymbol{x} \in \mathbb{R}^n$ is the set of directions from $\boldsymbol{x}$ in which f does not increase:

$$\mathcal{D}(f, \boldsymbol{x}) = \bigcup_{t\geq 0} \Big\{\boldsymbol{z} \in \mathbb{R}^n : f(\boldsymbol{x} + t\boldsymbol{z}) \leq f(\boldsymbol{x})\Big\}. \tag{4}$$

The descent cone of a convex function is a convex set. There is a famous duality result [19, Ch. 23] between the decent cone and the subdifferential of a convex function given by:

$$\mathcal{D}^\circ(f, \boldsymbol{x}) = \mathrm{cone}(\partial f(\boldsymbol{x})) := \bigcup_{t\geq 0} t.\partial f(\boldsymbol{x}), \tag{5}$$

where $\partial f(\boldsymbol{x})$ denotes the subdifferential of the function f at the point $\boldsymbol{x}$.

B. Statistical Dimension

Definition 1. (Statistical Dimension [10]). Let $\mathcal{C} \subseteq \mathbb{R}^n$ be a convex closed cone. Also, assume that $\boldsymbol{g} \in \mathbb{R}^n$ is a random vector with i.i.d standard normal entries. The statistical dimension of $\mathcal{C}$ is defined as:

$$\delta(\mathcal{C}) := \mathbb{E}\|\mathcal{P}_\mathcal{C}(\boldsymbol{g})\|_2^2 = \mathbb{E}\mathrm{dist}^2(\boldsymbol{g}, \mathcal{C}^\circ), \tag{6}$$

where, $\mathcal{P}_\mathcal{C}(\boldsymbol{x})$ is the projection of $\boldsymbol{x} \in \mathbb{R}^n$ onto the set $\mathcal{C}$ defined by: $\mathcal{P}_\mathcal{C}(\boldsymbol{x}) = \underset{\boldsymbol{z}\in\mathcal{C}}{\arg\min}\|\boldsymbol{z} - \boldsymbol{x}\|_2$.

The statistical dimension generalizes the concept of dimension for subspaces to the class of convex cones. Let f be a function that promotes some low-dimensional structure of $\boldsymbol{x}$. Then, $\delta(\mathcal{D}(f, \boldsymbol{x}))$ specifies the required number of Gaussian measurements that the optimization problem

$$\min_{\boldsymbol{z}\in\mathbb{R}^n} f(\boldsymbol{z}) \ \text{ s.t. } \ \boldsymbol{y} = \boldsymbol{A}\boldsymbol{z}, \tag{7}$$

needs for successful recovery [10, Theorem 2].

III. Main result

In this work, we provide an upper-bound for the required number of Gaussian measurements for $\mathsf{P}_{\rm TV}$ to succeed.

Theorem 1. *Let $\boldsymbol{x} \in \mathbb{R}^n$ be an arbitrary gradient s-sparse vector with gradient support $\mathcal{S}$. Let $\boldsymbol{A}$ be an $m \times n$ matrix whose rows are independent random vectors drawn from $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)$. Then, if $m \gtrapprox \widehat{m}_{\rm TV}$ with*

$$\widehat{m}_{\rm TV} := \inf_{t\geq 0} \mathbb{E}\mathrm{dist}^2(\boldsymbol{g}, t\partial\|\cdot\|_{\rm TV}(\boldsymbol{x})) \leq n - \frac{3(n-1-s)^2}{\pi(2n+s-4)}, \tag{8}$$

($\boldsymbol{g} \in \mathbb{R}^n$ is a random vector with i.i.d standard normal entries) $\mathsf{P}_{\rm TV}$ recovers $\boldsymbol{x}$ from $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ with probability at least $\frac{1}{2}$. In the noisy case of $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{e} \in \mathbb{R}^m$, where $\|\boldsymbol{e}\|_2 \leq \eta$, the outcome of $\mathsf{P}_{\rm TVn}$

$$\mathsf{P}_{\rm TVn}: \quad \min_{\boldsymbol{z}\in\mathbb{R}^n} \|\boldsymbol{\Omega}\boldsymbol{z}\|_1 \text{ s.t. } \|\boldsymbol{A}\boldsymbol{z} - \boldsymbol{y}\|_2 \leq \eta \tag{9}$$

satisfies

$$\|\widehat{\boldsymbol{x}}_\eta - \boldsymbol{x}\|_2 \leq \frac{2\eta}{\tau}, \tag{10}$$

with probability at least $1 - e^{-\frac{t^2}{2}}$, given that

$$m > \left(\sqrt{n - \frac{3(n-1-s)^2}{\pi(2n+s-4)}} + t + \tau\right)^2 + 1. \tag{11}$$

Proof sketch . The left-hand side of (8), besides the infimum over t, implicitly includes an infimum inside the expectation over the set $\partial\|\cdot\|_{\rm TV}(\boldsymbol{x})$ because of the definition of "dist". Instead of this latter infimum, we choose a vector in the set $\partial\|\cdot\|_{\rm TV}(\boldsymbol{x})$ that leads to an upper bound for $\mathrm{dist}^2(\boldsymbol{g}, t\partial\|\cdot\|_{\rm TV}(\boldsymbol{x}))$. This results in a strictly convex function of t. Then, by finding infimum over t, we obtain the desired upper-bound (8). The bound (11) is the consequence of the fact that $\delta(\mathcal{D}(\|\cdot\|_{\rm TV}, \boldsymbol{x}))$ appears in the number of measurements required for stable recovery of $\boldsymbol{x}$ [15, Corollarry 3.5] and $\widehat{m}_{\rm TV}$ provides an upper bound for $\delta(\mathcal{D}(\|\cdot\|_{\rm TV}, \boldsymbol{x}))$.

See Appendix A for details.

A. Comparison

In [5, Lemma 1], the following upper-bound is derived for $\delta(\mathcal{D}(\|\cdot\|_{\rm TV}, \boldsymbol{x}))$:

$$\widehat{m}_{\rm TV} \leq n - \frac{(n-1-s)^2}{n\pi}. \tag{12}$$

This bound is rather loose in low sparsity regimes. The main ideas in the proof of this bound are drawn from [17, Proposition 1]. In [17, Proposition 1], an upper-bound is derived for $\delta(\mathcal{D}(f,\boldsymbol{x}))$ where f is a decomposable norm function[3] that promotes a low-dimensional structure. This upper-bound does not approach the phase transition curve in the low-dimensional structured regimes. The problem arises from a redundant maximization in the proof that increases the number of required measurements (See section IV).

In [18], it is shown that $f = \|\cdot\|_{\rm TV}$ satisfies the weak decomposability assumption in [17, Equation 27]. Then, it is inferred that ([18, Equation 12]):

$$\widehat{m}_{\rm TV} \leq \delta(\mathcal{D}(\|\cdot\|_{\rm TV},\boldsymbol{x})) + 6. \tag{13}$$

While the validity of the above result needs more clarification, the value of $\widehat{m}_{\rm TV}$ describes $\delta\big(\mathcal{D}(\|\cdot\|_{\rm TV},\boldsymbol{x})\big)$ asymptotically well. To compute $\widehat{m}_{\rm TV}$, a numerical approach is devised in [18]: for each sparsity level, the minimization problem inside the expectation is numerically performed using a gradient method for an arbitrary $t \geq 0$. The outcome for 10^4 random realizations of $\boldsymbol{g}$ are averaged to replace the expectation. Finally, a numerical search determines the optimal t over a large interval.

In [13, Theorem 3.3], unlike [18], a closed-form upper-bound is [implicitly] derived for $\delta\big(\mathcal{D}(\|\cdot\|_{\rm TV},\boldsymbol{x})\big)$:

$$\delta(\mathcal{D}(\|\cdot\|_{\rm TV},\boldsymbol{x})) \leq \sqrt{32}(2\sqrt{5}+\sqrt{10})^2\sqrt{ns}\log(2n) + 1. \tag{14}$$

The scaling $\mathcal{O}(\sqrt{n}\log(n))$ of this bound, which is optimal up to a logarithmic factor, is obviously superior to our $\mathcal{O}(n)$ bound in (8). However, the effect of this scaling order difference becomes evident only for $n \geq 10^9$. On one hand, there are nowadays very rare applications for which n exceeds 10^9, and on the other hand, for $n < 10^9$ our bound in (8) is much better than (14). Therefore, we believe that (8) is still a useful bound.

In Theorem 1, we propose a tight upper-bound that leads to a reduction in the required number of Gaussian measurements in TV minimization. This upper-bound better follows the empirical TV phase transition curve. The upper-bound and the proof approach are completely new and differ from [5] and [17]. Our bound only depends on the sparsity level s and the special properties of the difference operator $\boldsymbol{\Omega}$. It also tends to the empirical TV phase transition curve at large values of m. In addition to TV, our approach can be applied to other low dimensional structures. For instance, the result in Theorem 1 can be easily extended to two dimensional images. Compared with [5, Theorem 5], the reduction of the required number of measurements, would be more evident in that case.

IV. Numerical Experiments

In this section, we evaluate how the number of Gaussian measurements scales with gradient sparsity. For each m and s, we repeat the following procedure 50 times in the cases $n = 50$, $n = 200$ and $n = 400$:

[3]See [20, Section 2.1] for more explanations.

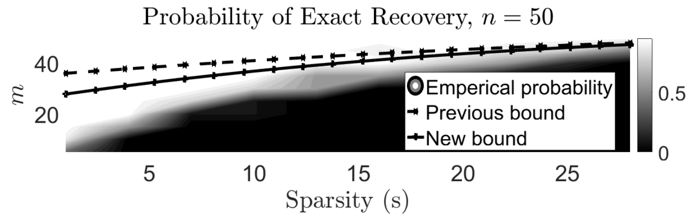

Fig. 1. Phase transition of $\mathsf{P}_{\rm TV}$ in the case of $n = 50$. The empirical probability is computed over 50 trials (black=0%, white=100%). The previous and new bounds come from (12) and (8), respectively.

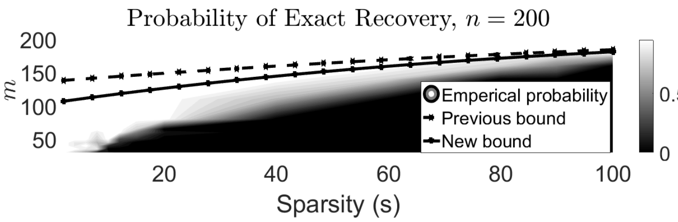

Fig. 2. Phase transition of $\mathsf{P}_{\rm TV}$ in the case of $n = 200$. The empirical probability is computed over 50 trials (black=0%, white=100%). The previous and new bounds come from (12) and (8), respectively.

- Generate a vector $\boldsymbol{x} \in \mathbb{R}^n$ that its discrete gradient has s non-zero entries. The locations of the non-zeros are selected at random.
- Observe the vector $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, where $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is a random matrix whose elements are drawn from an i.i.d standard Gaussian distribution.
- Obtain an estimate $\widehat{\boldsymbol{x}}$ by solving $\mathsf{P}_{\rm TV}$.
- Declare success if $\|\boldsymbol{x} - \widehat{\boldsymbol{x}}\|_2 \leq 10^{-3}$.

Figs. 1 2 and 3 show the empirical probability of success for this procedure. As shown in Figs. 1 2 and 3, our new bound better describes $\delta(\mathcal{D}(\|\cdot\|_{\rm TV},\boldsymbol{x}))$ in particular in low sparsity regimes. As sparsity increases, the difference between our bound and the bound (12) gets less. When the dimension of the true signal i.e. n, increases, the difference between our bound and (12) enhances (See Figs. 1, 2 and 3). In the asymptotic case, it seems that our bound reaches the empirical TV phase transition curve.

V. Conclusion

We have investigated the nonuniform recovery of gradient sparse signals from Gaussian random measurements. Obtaining a bound that suitably describes the precise behavior of TV minimization from failure to success, is left as an unanswered question. In this work, we derived an upper-bound for the required number of measurements that approximately estimates this behavior. Also, this bound is close to the empirical TV phase transition curve and seems to be asymptotically sharp.

Appendix

A. *Proof of Theorem 1*

Proof. Fix $\boldsymbol{g} \in \mathbb{R}^n$. Define

$$s_1 = \big|\{i \in \{2,...,n-1\}:\ i \in \mathcal{S}, i-1 \in \mathcal{S}\}\big|,$$
$$s_2 = \big|\{i \in \{2,...,n-1\}:\ i \in \bar{\mathcal{S}}, i-1 \in \bar{\mathcal{S}}\}\big|.$$

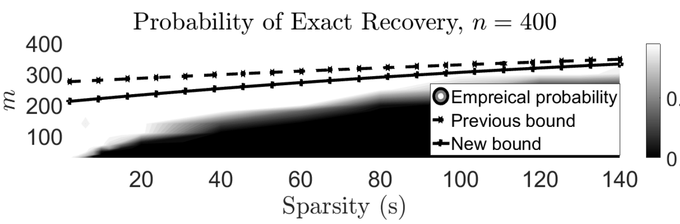

Fig. 3. Phase transition of $\mathsf{P}_{\rm TV}$ in the case of $n = 500$. The empirical probability is computed over 50 trials (black=0%, white=100%). The previous and new bounds come from (12) and (8), respectively.

As indicated above, s_1 and s_2 stand for the number of adjacent pairs in $\mathcal{S}$ and $\bar{\mathcal{S}}$, respectively, that are used in the sum of differences (TV operator). Since $\partial\|\cdot\|_1(\boldsymbol{\Omega x})$ is a compact set,

$$\boldsymbol{z}_0 = \underset{\boldsymbol{z}\in\partial\|\cdot\|_1(\boldsymbol{\Omega x})}{\text{argmax}}\ \langle \boldsymbol{g}, \boldsymbol{\Omega}^T\boldsymbol{z}\rangle$$

is well defined. In addition, we have that

$$\boldsymbol{z}_0 = \underset{\boldsymbol{z}\in\partial\|\cdot\|_1(\boldsymbol{\Omega x})}{\text{argmax}}\ \langle \boldsymbol{\Omega g}, \boldsymbol{z}\rangle = \underset{\|\boldsymbol{z}\|_\infty\le 1}{\text{argmax}}\langle \boldsymbol{\Omega g}\,,\,\text{sgn}(\boldsymbol{\Omega x})_{\mathcal{S}} + \boldsymbol{z}_{\bar{\mathcal{S}}}\rangle$$
$$= \text{sgn}(\boldsymbol{\Omega x})_{\mathcal{S}} + \underset{\|\boldsymbol{z}\|_\infty\le 1}{\text{argmax}}\langle \boldsymbol{\Omega g}\,,\,\boldsymbol{z}_{\bar{\mathcal{S}}}\rangle = \text{sgn}(\boldsymbol{\Omega x})_{\mathcal{S}} + \text{sgn}(\boldsymbol{\Omega g})_{\bar{\mathcal{S}}}$$

The above choice of $\boldsymbol{z}_0$ helps us to make the below upper-bounds more sharp:

$$\text{dist}^2(\boldsymbol{g}, t\boldsymbol{\Omega}^T\partial\|\cdot\|_1(\boldsymbol{\Omega x})) \overset{\text{(I)}}{\le} \|\boldsymbol{g} - t\boldsymbol{\Omega}^T\boldsymbol{z}_0\|_2^2 =$$
$$\|\boldsymbol{g} - t\boldsymbol{\Omega}_{\mathcal{S}}^T\text{sgn}(\boldsymbol{\Omega x})_{\mathcal{S}} - t\boldsymbol{\Omega}_{\bar{\mathcal{S}}}^T\text{sgn}(\boldsymbol{\Omega g})_{\bar{\mathcal{S}}}\|_2^2 = \|\boldsymbol{g}\|_2^2 +$$
$$t^2\|\boldsymbol{\Omega}_{\mathcal{S}}^T\text{sgn}(\boldsymbol{\Omega x})_{\mathcal{S}}\|_2^2 + t^2\|\boldsymbol{\Omega}_{\bar{\mathcal{S}}}^T\text{sgn}(\boldsymbol{\Omega g})_{\bar{\mathcal{S}}}\|_2^2$$
$$-2t\langle \boldsymbol{g}, \boldsymbol{\Omega}_{\bar{\mathcal{S}}}^T\text{sgn}(\boldsymbol{\Omega g})_{\bar{\mathcal{S}}}\rangle + 2t^2\langle \boldsymbol{\Omega}_{\mathcal{S}}^T\text{sgn}(\boldsymbol{\Omega x})_{\mathcal{S}}, \boldsymbol{\Omega}_{\bar{\mathcal{S}}}^T\text{sgn}(\boldsymbol{\Omega g})_{\bar{\mathcal{S}}}\rangle,$$

where for (I), we used the chain rule lemma of subdifferential [19, Theorem 23.9] $\partial\|\cdot\|_{\rm TV}(\boldsymbol{x}) = \boldsymbol{\Omega}^T\partial\|\cdot\|_1(\boldsymbol{\Omega x})$. The role of $\boldsymbol{z}_0$ is to minimize the upper-bound in the inequality (I). By taking expectation from both sides, we have:

$$\mathbb{E}\|\boldsymbol{g} - t\boldsymbol{\Omega}_{\mathcal{S}}^T\text{sgn}(\boldsymbol{\Omega x})_{\mathcal{S}} - t\boldsymbol{\Omega}_{\bar{\mathcal{S}}}^T\text{sgn}(\boldsymbol{\Omega g})_{\bar{\mathcal{S}}}\|_2^2 \overset{\text{(I)}}{=} n - 2t\sqrt{\frac{2}{\pi}}\sum_{i\in\bar{\mathcal{S}}}$$
$$\|\boldsymbol{\omega}_i\|_2 + t^2\Big[\sum_{j\in\mathcal{S}}\sum_{k\in\mathcal{S}}\boldsymbol{\omega}_j^T\boldsymbol{\omega}_k\text{sgn}(\boldsymbol{\Omega x})_j\text{sgn}(\boldsymbol{\Omega x})_k\Big] + t^2\mathbb{E}$$
$$\Big[\sum_{j\in\bar{\mathcal{S}}}\sum_{k\in\bar{\mathcal{S}}}\boldsymbol{\omega}_j^T\boldsymbol{\omega}_k\text{sgn}(\boldsymbol{\Omega g})_j\text{sgn}(\boldsymbol{\Omega g})_k\Big] \overset{\text{(II)}}{\le} n - 2t\sqrt{\frac{2}{\pi}}\sum_{i\in\bar{\mathcal{S}}}\|\boldsymbol{\omega}_i\|_2$$
$$+ t^2\Big[\sum_{j\in\mathcal{S}}\sum_{k\in\mathcal{S}}\boldsymbol{\omega}_j^T\boldsymbol{\omega}_k\text{sgn}(\boldsymbol{\Omega x})_j\text{sgn}(\boldsymbol{\Omega x})_k\Big] +$$
$$t^2\Big[\sum_{j\in\bar{\mathcal{S}}}\sum_{k\in\bar{\mathcal{S}}}\boldsymbol{\omega}_j^T\boldsymbol{\omega}_k\frac{2}{\pi}\sin^{-1}\frac{\boldsymbol{\omega}_j^T\boldsymbol{\omega}_k}{\|\boldsymbol{\omega}_j\|_2\|\boldsymbol{\omega}_k\|_2}\Big] \overset{\text{(III)}}{\le} n - \frac{4t}{\sqrt{\pi}}\bar{s} + t^2[2s$$
$$+ 2s_1 + 2\bar{s} + \frac{2s_2}{3}] \overset{\text{(IV)}}{\le} n - \frac{4t}{\sqrt{\pi}}\bar{s} + t^2[4s - 2 + \frac{8}{3}\bar{s} - \frac{2}{3}], \tag{15}$$

where in the equality (I) in (15), we used the facts $\mathbb{E}\|\boldsymbol{g}\|_2^2 = n$ and

$$\mathbb{E}\langle \boldsymbol{g}, \boldsymbol{\Omega}_{\bar{\mathcal{S}}}^T\text{sgn}(\boldsymbol{\Omega g})_{\bar{\mathcal{S}}}\rangle = \mathbb{E}\langle \boldsymbol{\Omega}_{\bar{\mathcal{S}}}\boldsymbol{g}, \text{sgn}(\boldsymbol{\Omega g})_{\bar{\mathcal{S}}}\rangle = \mathbb{E}\|(\boldsymbol{\Omega g})_{\bar{\mathcal{S}}}\|_1$$
$$\sum_{i\in\bar{\mathcal{S}}}\mathbb{E}|\boldsymbol{\omega}_i^T\boldsymbol{g}| = \sqrt{\frac{2}{\pi}}\sum_{i\in\bar{\mathcal{S}}}\|\boldsymbol{\omega}_i\|_2,$$

where the last equality above comes from the mean of a folded normal distribution. The inequality (II) follows from the following lemma and $\boldsymbol{\Omega} := [\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, ..., \boldsymbol{\omega}_p]^T$.

Lemma 1. *Let $\boldsymbol{g}\in\mathbb{R}^n$ be a standard random Gaussian i.i.d vector and $\boldsymbol{\Omega}\in\mathbb{R}^{p\times n}$ be an analysis operator. Then,*

$$\mathbb{E}\{\text{sgn}(\boldsymbol{\Omega g})_j\text{sgn}(\boldsymbol{\Omega g})_k\} = \frac{2}{\pi}\sin^{-1}\frac{\boldsymbol{\omega}_j^T\boldsymbol{\omega}_k}{\|\boldsymbol{\omega}_j\|_2\|\boldsymbol{\omega}_k\|_2}.$$

Proof. see Appendix B.

The inequality (III) is the result of the following properties of the difference operator.

$$\boldsymbol{\omega}_j^T\boldsymbol{\omega}_k = \left\{\begin{array}{ll} -1, & |j-k| = 1 \\ 0, & \text{o.w.}\end{array}\right\},$$
$$\|\boldsymbol{\omega}_i\|_2 = \sqrt{2}\ :\ \forall i\in 1,...,n-1.$$

Also, $\bar{s} = n - 1 - s$. The inequality (IV) comes from the facts

$$s_1 \le s - 1,\quad s_2 \le \bar{s} - 1.$$

Now, by minimizing (15) with respect to t, we reach (8). Due to [15, Corollary 3.5], if

$$m > (\sqrt{\delta(\mathcal{D}(\|\cdot\|_{\rm TV}, \boldsymbol{x}))} + t + \tau)^2 + 1, \tag{16}$$

then, with probability $1 - e^{-\frac{t^2}{2}}$,

$$\|\boldsymbol{x} - \widehat{\boldsymbol{x}}\|_2 \le \frac{2\eta}{\tau}$$

A good upper-bound for $\delta(\mathcal{D}(\|\cdot\|_{\rm TV}, \boldsymbol{x}))$, is given by (8) and thus, the claim is proved. ■

B. Proof of Lemma 1

Proof. Consider $\boldsymbol{\Omega} := [\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, ..., \boldsymbol{\omega}_p]^T$. Define

$$h_j = \frac{\boldsymbol{\omega}_j^T\boldsymbol{g}}{\|\boldsymbol{\omega}_j\|_2},\quad h_k = \frac{\boldsymbol{\omega}_k^T\boldsymbol{g}}{\|\boldsymbol{\omega}_k\|_2}.$$

We have:

$$\mathbb{E}\{\text{sgn}(\boldsymbol{\Omega g})_j\text{sgn}(\boldsymbol{\Omega g})_k\} = \mathbb{E}\{\text{sgn}(h_j)\text{sgn}(h_k)\} =$$
$$1 - 2\mathbb{P}\left\{\frac{h_j}{h_k} < 0\right\} = 1 - 2\left(\frac{1}{2} - \frac{1}{\pi}\sin^{-1}\left(\frac{\mathbb{E}\{\boldsymbol{\omega}_k^T\boldsymbol{g}\boldsymbol{\omega}_j^T\boldsymbol{g}\}}{\|\boldsymbol{\omega}_j\|_2\|\boldsymbol{\omega}_k\|_2}\right)\right)$$
$$= \frac{2}{\pi}\sin^{-1}\frac{\boldsymbol{\omega}_j^T\boldsymbol{\omega}_k}{\|\boldsymbol{\omega}_j\|_2\|\boldsymbol{\omega}_k\|_2} \tag{17}$$

where the second equality comes from total probability theorem, the third equality comes from the fact that $\frac{h_j}{h_k}$ is a Cauchy random variable. ■

References

[1] H. Rauhut, K. Schnass, and P. Vandergheynst, "Compressed sensing and redundant dictionaries," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 2210–2219, 2008.

[2] E. J. Candes, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Applied and Computational Harmonic Analysis*, vol. 31, no. 1, pp. 59–73, 2011.

[3] M. Kabanava and H. Rauhut, "Analysis ℓ_1-recovery with frames and gaussian measurements," *Acta Applicandae Mathematicae*, vol. 140, no. 1, pp. 173–195, 2015.

[4] S. Nam, M. E. Davies, M. Elad, and R. Gribonval, "The cosparse analysis model and algorithms," *Applied and Computational Harmonic Analysis*, vol. 34, no. 1, pp. 30–56, 2013.

[5] M. Kabanava, H. Rauhut, and H. Zhang, "Robust analysis ℓ_1-recovery from gaussian measurements and total variation minimization," *European Journal of Applied Mathematics*, vol. 26, no. 6, pp. 917–929, 2015.

[6] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE transactions on information theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[7] D. L. Donoho, "High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension," *Discrete & Computational Geometry*, vol. 35, no. 4, pp. 617–652, 2006.

[8] D. L. Donoho, A. Maleki, and A. Montanari, "The noise-sensitivity phase transition in compressed sensing," *IEEE Transactions on Information Theory*, vol. 57, no. 10, pp. 6920–6941, 2011.

[9] W. Xu and B. Hassibi, "Precise stability phase transitions for ℓ_1 minimization: A unified geometric framework," *IEEE transactions on information theory*, vol. 57, no. 10, pp. 6894–6919, 2011.

[10] D. Amelunxen, M. Lotz, M. B. McCoy, and J. A. Tropp, "Living on the edge: Phase transitions in convex programs with random data," *Information and Inference: A Journal of the IMA*, vol. 3, no. 3, pp. 224–294, 2014.

[11] D. L. Donoho, I. Johnstone, and A. Montanari, "Accurate prediction of phase transitions in compressed sensing via a connection to minimax denoising," *IEEE transactions on information theory*, vol. 59, no. 6, pp. 3396–3433, 2013.

[12] F. Krahmer, C. Kruschel, and M. Sandbichler, "Total variation minimization in compressed sensing," *arXiv preprint arXiv:1704.02105*, 2017.

[13] J.-F. Cai and W. Xu, "Guarantees of total variation minimization for signal recovery," *Information and Inference*, vol. 4, no. 4, pp. 328–353, 2015.

[14] D. Needell and R. Ward, "Stable image reconstruction using total variation minimization," *SIAM Journal on Imaging Sciences*, vol. 6, no. 2, pp. 1035–1058, 2013.

[15] J. A. Tropp, "Convex recovery of a structured signal from independent random linear measurements," in *Sampling Theory, a Renaissance*, pp. 67–101, Springer, 2015.

[16] R. Vershynin, "Estimation in high dimensions: a geometric perspective," in *Sampling theory, a renaissance*, pp. 3–66, Springer, 2015.

[17] R. Foygel and L. Mackey, "Corrupted sensing: Novel guarantees for separating structured signals," *IEEE Transactions on Information Theory*, vol. 60, no. 2, pp. 1223–1247, 2014.

[18] B. Zhang, W. Xu, J.-F. Cai, and L. Lai, "Precise phase transition of total variation minimization," in *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*, pp. 4518–4522, IEEE, 2016.

[19] R. T. Rockafellar, *Convex analysis*. Princeton university press, 2015.

[20] E. Candes and B. Recht, "Simple bounds for recovering low-complexity models," *Mathematical Programming*, vol. 141, no. 1-2, pp. 577–589, 2013.