Living near the edge: A lower-bound on the phase transition of total variation minimization

Sajad Daei, Arash Amini, Farzan Haddadi

Abstract—This work is about the total variation (TV) minimization which is used for recovering gradient-sparse signals from compressed measurements. Recent studies indicate that TV minimization exhibits a phase transition behavior from failure to success as the number of measurements increases. In fact, in large dimensions, TV minimization succeeds in recovering the gradient-sparse signal with high probability when the number of measurements exceeds a certain threshold; otherwise, it fails almost certainly. Obtaining a closed-form expression that approximates this threshold is a major challenge in this field and has not been appropriately addressed yet. In this work, we derive a tight lower-bound on this threshold in case of any random measurement matrix whose null space is distributed uniformly with respect to the Haar measure. In contrast to the conventional TV phase transition results that depend on the simple gradient-sparserity level, our bound is highly affected by generalized notions of gradient-sparsity. Our proposed bound is very close to the true phase transition of TV minimization confirmed by simulation results.

Index Terms—Sample Complexity, Statistical Dimension, Total Variation Minimization

I. INTRODUCTION

COMPRESSED sensing (CS) has gained a lot of attention in the past decade as it provides a strategy to recover signals from undersampled measurements \cite{1,2}. In mathematical terms, the measured data about a signal \( x \in \mathbb{R}^n \) is given by a collection of linear projections

\[
y = Ax \in \mathbb{R}^m, \tag{1}
\]

where \( A \in \mathbb{R}^{m \times n} \) with \( m \ll n \) is the measurement matrix. Without having any prior knowledge about the structure of \( x \), it is impossible to reconstruct \( x \) from \( y \). The standard prior knowledge in CS is that the signal of interest is sparse in an orthonormal basis which means that it can be expressed as the linear combination of a few basis elements. However, this simple assumption often seems irrational in practical scenarios as we deal with signals that are sparse in some overcomplete dictionary \( D \in \mathbb{R}^{n \times N} \) with \( n \ll N \) which means that \( x \) can be described as \( x = D\alpha \) for some sparse \( \alpha \in \mathbb{R}^N \). This setting is known as the synthesis sparsity model since it describes a way to synthesize the signal \( x \). Then, for recovering \( \alpha \) from \( y \), it is common to use the so-called \( \ell_1 \) synthesis problem defined as

\[
\hat{\alpha} = \arg \min_{\alpha \in \mathbb{R}^N} \| D\alpha \|_1 \quad \text{s.t. } y = Ax. \tag{2}
\]

The reconstructed signal is then \( \hat{x} = D\hat{\alpha} \). Interestingly, the synthesis model has an analysis (and more general) counterpart which assumes that the signal of interest i.e. \( x \) is sparse after applying an operator called analysis operator \( \Omega \). The fact that \( \Omega x \) is sparse, motivates the optimization problem

\[
\hat{x} = \arg \min_{z \in \mathbb{R}^n} \| \Omega z \|_1 \quad \text{s.t. } y = Az, \tag{3}
\]

which is typically referred to as \( \ell_1 \) analysis problem. Although the two methods \( \ell_1 \) synthesis and \( \ell_1 \) analysis perform very differently on large families of signals, the numerical results of the works \cite{3,4} show that the analysis formulation outperforms its synthesis-based counterpart in many scenarios of interest.

An important and special case of \( \ell_1 \) analysis formulation \cite{5} is the case where \( \Omega \) is the finite difference matrix defined as

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

By replacing this \( \Omega \) in the problem \cite{5}, we reach the well-known total variation (TV) minimization:

\[
P_{TV} : \min_{z \in \mathbb{R}^n} \| \Omega z \|_1 \quad \text{s.t. } y_{m \times 1} = Az. \tag{4}
\]

TV minimization has been proven to be very effective in image processing \cite{6,7,8,9,10} and other fields \cite{11,12}. One of the important issues concerning the problem \( P_{TV} \) is the required number of measurements i.e. the minimum \( m \) that \( P_{TV} \) needs for successful recovery. A series of works studying convex geometry approaches have found valuable results regarding the required number of measurements in recovering structured signals \cite{13,14,15}. Specifically, it has been shown in \cite{13} that \( P_{TV} \) undergoes a phase transition from failure to success as the number of measurements increases. This means that there exists a certain curve \( \Psi(\Omega, x) \) (known as statistical dimension) in the boundary of failure and success that \( P_{TV} \) succeeds to recover the gradient-sparse vector with probability \( \frac{1}{2} \). Obtaining an expression that approximates \( \Psi(\Omega, x) \) is a fundamental challenge in image processing and has not been

\footnote{For example, sparse signals.}
exactly addressed yet. In this work, we obtain a tight lower-bound on \(\Psi(\Omega, x)\). As numerical results confirm, our proposed lower-bound follows the true phase transition curve very well.

A. Prior Works

In the last few years, great works have been established for obtaining the required number of measurements in \(P_{TV}\) (see e.g. [9], [10], [16]–[23]).

Needell et al. in [17], [24] obtain recovery guarantees for two-dimensional TV minimization. Their result is based on transforming gradient-sparse images into compressible signals in a Haar wavelet basis. Then, a modified restricted isometry property (RIP) is developed to guarantee the image recovery. Their approach does not hold for one-dimensional gradient-sparse signals. Besides, their sample complexity result is only designed for the asymptotic setting.

Nam et al. in [25] consider the original problem of minimizing the number of variations (NV) under some affine constraints defined as

\[
P_{NV} : \min_{z \in \mathbb{R}^n} \|z\|_{NV} := \|\Omega z\|_0 := \sum_{i=1}^{n-1} |z_{i+1} - z_i| > 0
\]

s.t. \(y_{m \times 1} = Az\),

(5)

where \(1_E\) denotes the indicator function of a set \(E\). Unfortunately, \(P_{NV}\) is known to be NP-complete [26, Section 4.1] and \(P_{TV}\) is the closest tractable problem to \(P_{NV}\). In fact, the TV norm sums the amplitudes of variations and in some sense, is to the NV function what the \(\ell_1\) norm is to the \(\ell_0\) function in the area of sparse recovery. For a \(s\)-gradient-sparse vector \(x \in \mathbb{R}^n\), it has been shown in [20] that

\[
m > 2\kappa_\Omega(s)
\]

(6)

measurements suffice for exact recovery via \(P_{NV}\). Here, \(\kappa_\Omega(s)\) is a special function describing the signal manifold dimension and is obtained via a combinatorial search over all \(s\)-gradient-sparse signals. This result matches with the well-known fact in conventional CS (in particular sparse recovery and matrix completion) where the number of needed measurements is proportional to the signal’s manifold dimension. However, the authors of [25] have numerically shown that when we are dealing with the recovery of gradient-sparse signals via \(P_{TV}\), the required number of measurements is not explained by the manifold dimension anymore. Somewhat interestingly, our proposed bound in Section III is consistent with this fact.

Donoho et al. in [21] obtain the asymptotic minimax mean square error (MSE) of a TV denoiser and using numerical simulations, conjecture that it matches with the phase transition curve of \(P_{TV}\). However, this conjecture is not proved. Moreover, the result holds only in the asymptotic case.

Cai et al. in [9] show that \(O(\sqrt{sn \log(n)})\) and \(O(\sqrt{sn})\) measurements are respectively sufficient and necessary for exact recovery of a given \(s\)-gradient-sparse vector via \(P_{TV}\). The order of their bounds seems to be optimal. However, their (necessary and sufficient) bounds do not provide a good prediction for sample complexity in the non-asymptotic case.

In [26, Theorem 4], an upper-bound is derived for the statistical dimension (in case of TV minimization) using the results of [27]. Their bound depends on the gradient-sparsity level and provides an inaccurate prediction for the required sample complexity in low-sparsity regimes.

Recently, the authors of [22] obtain a non-asymptotic sample complexity bound for \(P_{TV}\). In contrast to the previous bounds that would depend on the gradient-sparsity \(s\), their non-asymptotic bound is highly affected by the number of consecutive variations (adjacent pairs in the gradient support). Their proof approach is based on a refined analysis of [27]. While their bound outperforms the previous bounds, it is still far from the statistical dimension in low-sparsity regimes.

By using a more general analysis of [27], the authors in [19] obtain an explicit formula describing the required number of measurements. Their bound depends on the coherence structure of \(\Omega\). Again, the proposed bound does not provide an accurate prediction of the true sample complexity in low-sparsity regimes.

B. Contribution

In this work, we obtain a lower-bound on the statistical dimension that provides the necessary number of measurements that \(P_{TV}\) needs for exact recovery. Our bound is very close to the true sample complexity even in low-sparsity levels and numerical experiments in Section IV show that it is tight in the asymptotic case. Our bound depends on the number of the consecutive, individual and tail-end variations of the interested signal. We hope that our bound sheds more light on the effective parameters in the statistical dimension in case of TV minimization. It is worth mentioning that the only lower-bound on the statistical dimension in the literature is established in [9] Theorem 2.1 Part b), and therefore served as an object of comparison in Sections III and IV.

C. Outline of the paper

The paper is organized as follows. First, we review a few basic concepts in convex geometry. Section III amounts to our main result. Section IV is about numerical experiments that verify our theoretical findings. Finally, the paper is concluded in Section V.

D. Notation

Throughout the paper, scalars are denoted by lowercase letters, vectors by lowercase boldface letters, and matrices by uppercase boldface letters. The \(i\)th element of a vector \(x\) is given either by \(x(i)\) or \(x_i\). \((\cdot)^\dagger\) denotes pseudo inverse. We reserve calligraphic uppercase letters for sets (e.g. \(S\)). The cardinality of a set \(S\) is denoted by \(|S|\). The complement of a set \(S\) in \(\{1, \ldots, n\}\) is denoted by \(\overline{S}\). \(C^o\) denotes the polar of a cone \(C\). The symbol \(cone(\cdot)\) signifies the conic hull of a set. Null space of a matrix is shown by \(null(\cdot)\). For a matrix \(A\), \(A_{SC}\) means the matrix \(A\) restricted to the rows indexed by \(S\). We denote i.i.d. standard Gaussian random vector by \(g\). Lastly, \(\vee\) means logical “or ”.
II. CONVEX GEOMETRY

A. Statistical Dimension

Definition 1. Statistical Dimension [13]: Statistical dimension is a generalized concept of subspace dimension in the class of convex cones. Intuitively, it measures the size of a cone. More precisely, for a closed convex cone $C \subseteq \mathbb{R}^n$, statistical dimension of $C$ is defined as:

$$\delta(C) := \mathbb{E} \inf_{z \in C^\circ} \| g - z \|_2^2,$$  \hspace{1cm} (7)

where $g$ is a random vector in $\mathbb{R}^n$ chosen from Gaussian ensemble with independent entries.

B. Linear Inverse Problems and Sample Complexity

In [13], it is proved that any random convex optimization problem $P_f$:

$$P_f : \min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad Ax = b,$$  \hspace{1cm} (8)

undergoes a phase transition between success and failure as the number of measurements increases. The location of this transition (the boundary of success and failure) is specified by the statistical dimension of the descent cone of $f$ at $x \in \mathbb{R}^n$ i.e. $\delta(D(f,x))$ [13]. Here, $D(f,x)$ is defined as the set of directions toward which $f$ is decreased and is given by

$$D(f,x) := \bigcup_{t \geq 0} \{ z \in \mathbb{R}^n : f(x + tz) \leq f(x) \}. \hspace{1cm} (9)$$

There is also a well-known fact between the descent cone and the subdifferential (c.f. [28] Chapter 23) given by

$$D(f,x)^\circ = \text{cone}(\partial f(x)). \hspace{1cm} (10)$$

By using this fact and (7), one may write:

$$\delta(D(f,x)) = \mathbb{E} \inf_{z \in \text{cone}(\partial f(x))} \| g - z \|_2^2 = \mathbb{E} \inf_{t \geq 0} \inf_{z \in \partial f(x)} \| g - tz \|_2^2. \hspace{1cm} (11)$$

Calculating statistical dimension has been a difficult task in the literature and therefore, it is commonly approximated with the expression

$$B_u := \inf_{t \geq 0} \mathbb{E} \inf_{z \in \partial f(x)} \| g - tz \|_2^2, \hspace{1cm} (12)$$

which is first proposed by Stojnic [29] in the context of $\ell_1$ minimization. In [30] Theorem 2], it has been shown that $B_u$ approximates the statistical dimension well for a large class of structure inducing functions in particular $f = \| \cdot \|_{TV}$.

III. MAIN RESULT

Before stating our main result, we need to define some parameters regarding gradient-sparse signals which are required in our analysis.

Definition 2. (Consecutive, individual and tail-end variations) Consider a gradient-sparse signal $x \in \mathbb{R}^n$ with gradient $d := \Omega x$ and gradient support $S$. The consecutive variations correspond to the adjacent pairs in $S$ defined as

$$S_1 := \{ i \in [n - 1] : i \in S, \ i - 1 \notin S \}. \hspace{1cm} (13)$$

The set $S_1$ can be divided into two sets:

$$S_1^+ := \{ i \in [n - 1] : i \in S, \ i - 1 \notin S, \ sgn(d_i) sgn(d_{i-1}) > 0 \},$$

$$S_1^- := \{ i \in [n - 1] : i \in S, \ i - 1 \notin S, \ sgn(d_i) sgn(d_{i-1}) < 0 \}, \hspace{1cm} (14)$$

which are interpreted as the consecutive variations (adjacent pairs in $S$) with the same and opposite signs, respectively. We define the individual variations by the sets

$$S_2 := \{ i \in [n - 1] : i \in S, \ i - 1 \notin S \},$$

$$S'_2 := \{ i \in [n - 1] : i \in S, \ i - 1 \in S \}. \hspace{1cm} (15)$$

The variations in the left and right ends of the signal are called tail-end variations which are defined by the set

$$S_3 := \{ i \in [n - 1] : i \in S, \ i - 1 \notin [n - 1], \ i + 1 \notin [n - 1] \}. \hspace{1cm} (16)$$

We also show the number of consecutive, individual and tail-end variations respectively by

$$s_1 := |S_1| = |S_1^+| + |S_1^-| := s_1^+ + s_1^-,$$

$$s_2 := |S_2| + |S'_2|,$$

$$s_3 := |S_3|.$$  

As an illustrative example, in Figure 1 the discrete gradient $d := \Omega x$ of a typical gradient-sparse signal $x \in \mathbb{R}^n$ is depicted. There is one pair of elements in $d$ both of which belong to the gradient support (alternatively representing consecutive variations of $x$) and have the same sign. Thus, $s_1^+ = 1$ and $s_2 = 0$. Also, there are 10 elements for which $i \in S, i - 1 \notin S$ or $i \in S, i - 1 \notin S$ (alternatively representing individual variations in $x$). As a result, $s_2 = 10$. Lastly, there are no elements in the tail-ends of $d$. This means that the first and last elements of $x$ include no variations and $s_3 = 0$.

In the following theorem, we propose a lower-bound on $\delta(D(\| \cdot \|_{TV}, x))$. This lower-bound depends on the associated parameters in Definition 2.

Theorem 1. Let $x \in \mathbb{R}^n$ be a gradient-sparse signal with gradient $d := \Omega x$, and gradient support $S$. Assume that $x$ has $s_1 = s_1^+ + s_1^-$, $s_2$, and $s_3$ consecutive, individual and tail-end variations, respectively as defined in Definition 2. Let $A \in \mathbb{R}^{m \times n}$ be a random matrix whose null space

![Figure 1: This plot shows the discrete gradient of a typical gradient-sparse vector $x \in \mathbb{R}^n$ with parameters $s_1^+ = 1, s_1^- = 0, s_2 = 10$, and $s_3 = 0$.](image-url)
is uniformly distributed with respect to the Haar measure. Consider \( y_{m \times 1} = Ax \) as the vector of measurements. Then,
\[
\delta(D(\| \cdot \|_{TV}, x)) \geq \inf_{\epsilon \geq 0} \Psi_{\epsilon}(s_{1}^{+}, s_{1}^{-}, s_{2}, s_{3}) := \hat{m}_{TV},
\]
where
\[
\Psi_{\epsilon}(s_{1}^{+}, s_{1}^{-}, s_{2}, s_{3}) = s_{1}^{+} + s_{1}^{-} (1 + 4 \epsilon^{2}) + s_{2} \phi_{1}(t, t) + (n - 2 - s_{1}^{+} - s_{1}^{-} - s_{2}) \phi_{2}(2t) + s_{3} (1 + t^{2}) + (2 - s_{3}) \phi_{2}(t),
\]
and
\[
\phi_{1}(a, b) := \frac{1}{\sqrt{2\pi}} \int_{b}^{\infty} (u - b)^{2} e^{-\frac{(u - a)^{2}}{2}} + e^{-\frac{(u + a)^{2}}{2}} du,
\]
\[
\phi_{2}(z) := \frac{2}{\pi} \int_{z}^{\infty} (u - z)^{2} e^{-\frac{u^{2}}{2}} du,
\]
and thus if \( m \leq \hat{m}_{TV} \), then with probability at least \( 1 - 4e^{-\frac{(m_{TV} - m)^{2}}{10m_{TV}}} \), \( P_{TV} \) fails to recover \( x \).

**Proof sketch.** As discussed in [12], we intend to find a lower-bound for
\[
\inf_{\epsilon \geq 0} \mathbb{E} \inf_{z \in [\epsilon / \| TV(\epsilon) \|]} \| g - tz \|_{2}^{2}.
\]

The expression inside the latter infimum is formed of a few summands. Each summand separately has a unique minimizer over the set \( \mathbb{E} \| TV(\epsilon) \| \). By passing this infimum through each summand (this leads to a lower-bound) and then applying expectation, we reach a closed-form expression for each summand and thereby a strictly convex function with respect to \( t \) for the expression inside the former infimum.

**Discussion.** One of the key properties of Theorem 1 is that, unlike the most works in TV minimization, the widely-used concept of gradient sparsity \( s := |S| \) does not directly explain our proposed bound. In fact, it seems that the statistical dimension in case of TV minimization i.e. \( \delta(D(\| \cdot \|_{TV}, x)) \), depends on generalized concepts of gradient-sparsity: namely the number of consecutive, individual, and tail-end variations. To examine how these parameters affect our bound \( \hat{m}_{TV} \), we designed a numerical experiment in Table 1 and considered different values for \( s_{1}^{+}, s_{1}^{-}, s_{2}, s_{3} \). We observe from Table 1 that the number of consecutive variations with negative signs, i.e. \( s_{1}^{+} \), has the highest impact on \( \hat{m}_{TV} \). In other words, for a fixed number of variations \( s = |\Omega_2 x|_{0} \), recovering a highly oscillating signal is harder (alternatively needs more measurements) than the one with well-separated variations.

**Remark 1.** (Prior work) In [9] Theorem 2.1, it has been proved that
\[
\frac{9\sqrt{sn}}{50\pi} - \frac{12}{5\pi} \leq \delta(D(\| \cdot \|_{TV}, x)) \leq \sqrt{32} (2\sqrt{5} + \sqrt{10})^{2} \sqrt{ns \log(2n) + 1}.
\]

Their proof approach is based on estimating the statistical dimension of a certain set using a wavelet-based argument. Since the lower and upper-bound are in the same order (up to a log factor), their approach is optimal in the asymptotic case \( (n \to \infty) \). However, in the non-asymptotic case, it demonstrates poor prediction of the true sample complexity.

**IV. Simulations**

In this section, we evaluate how our proposed bound in Theorem 1 predicts the phase transition of \( P_{TV} \). For each \( m \) and \( s \), we repeat the following steps 50 times:

- Select a random subset \( S \subseteq \{1, \ldots, n - 1\} \) with \( |S| = s \).
- Generate a vector \( x \in \text{null}(\Omega_{2}) \) whose gradient is supported on \( S \).
- Construct the vector \( y = Ax \) where \( A \) is an i.i.d. Gaussian matrix (the null space of Gaussian matrices is distributed uniformly with respect to the Haar measure).
- Solve \( P_{TV} \) to obtain an estimate \( \hat{x} \).
- Declare success if \( \| x - \hat{x} \|_{2} \leq 10^{-6} \).

Figures 2 and 3 show the empirical probability of successful recovery obtained from 50 Monte Carlo iterations in case of \( n = 200 \) and \( n = 400 \), respectively. Notice that since the sample complexity bounds in [22] Theorem 1, [19] Theorem 1 and \( \hat{m}_{TV} \) do not directly depend on the gradient-sparsity \( s \), we depict the empirical mean of the bounds over 300 iterations, for each \( s \). As it turns out from Figures 2 and 3 our proposed bound in Theorem 1 predicts the statistical dimension well; the lower-bound [9] Theorem 2.1 does not seem to be exact; and the upper-bounds [22] Theorem 1 and [19] Theorem 6.8 fail to explain the true sample complexity in low-sparsity levels.
V. Conclusion

In this work, we obtained a lower-bound on the statistical dimension in case of TV minimization. This lower-bound provides the necessary number of measurements that $P_{TV}$ needs for successful recovery. Our bound depends on the number of consecutive, individual and tail-end variations of the signal and precisely captures the location of the TV phase transition. In fact, it seems that these quantities specify the effective parameters of the statistical dimension for gradient-sparse signals.

APPENDIX A
PROOF OF THEOREM [1]

Proof. Recall that the sets $S_i^+$, $S_i^-$, $S_2$, $S_2'$, $S_3$ are defined in Definition [2]. Moreover, define the set of adjacent pairs in $S$ as

$$S_4 := \{i \in [n - 1]: i \in S, \ i - 1 \in S\},$$  

which are used in the proof.

Since statistical dimension is approximately equal to $B_u$ in [12], we find a lower-bound for

$$B_u := \inf_{t \geq 0} \text{Edist}^2(g, t\partial) \cdot \|TV(x)\|.$$

(23)

The first step is to calculate $\partial\|TV(x)\$. From the chain rule lemma of subdifferential [28, Chapter 23], we have:

$$\partial\|TV(x)\| = \Omega^T \partial \|d\| = \Omega^T \left\{ z \in \mathbb{R}^{n-1}: z_i = \text{sgn}(d), \ i \in S \right\}.$$  

(24)

To calculate [23], regarding [24], we compute the distance of the dilated subdifferential of the descent cone of TV norm at $x \in \mathbb{R}^n$ from a standard Gaussian vector $g \in \mathbb{R}^n$ which is given by:

$$\text{dist}^2(g, t\partial) \cdot \|TV(x)\| = \inf_{x \in \mathbb{R}^n} \|g - tx\|^2 = \inf_{x \in \mathbb{R}^n} \|g - tx\|^2 = \inf_{x \in \mathbb{R}^n} \sum_{i=1}^{n} (g_i - t\Omega z_i)^2 = \inf_{x \in \mathbb{R}^n} \sum_{i=1}^{n} (g_i - t\sum_{j \in S} \Omega(j, i) \text{sgn}(d_j) - t\sum_{j \in S} \Omega(j, i) z(j))^2.$$  

(25)

Each row of the finite difference operator $\Omega$ includes a pair of $\{+1, -1\}$ and is zero elsewhere, i.e.,

$$\Omega(j, i) = \begin{cases} 1, & j = i \\ -1, & j = i - 1 \end{cases}$$  

(26)

By using the latter property, the relation [23] reads

$$\inf_{|x|, \|x\| \leq 1} \sum_{i=1}^{n} (g_i - t\text{sgn}(d_i)1_{i \in S} + t\text{sgn}(d_{i-1})1_{i-1 \in S} - tz(i)1_{i \in S} + tz(i-1)1_{i-1 \in S})^2.$$  

(27)

By passing the infimum through the summation, we have a lower-bound on [27] as follows:

$$\text{dist}^2(g, t\partial) \cdot \|TV(x)\| \geq \sum_{i \in S_i^+ \cup S_i^-} (g_i - t\text{sgn}(d_i) + t\text{sgn}(d_{i-1}))^2$$  

$$+ \sum_{i \in S_2} \inf_{|x| \leq 1} (g_i - t\text{sgn}(d_i) + t(i - 1))^2$$  

$$+ \sum_{i \in S_2'} \inf_{|x| \leq 1} (g_i - t\text{sgn}(d_i) - t(i))$$  

$$+ \sum_{i \in S_3} \inf_{|x| \leq 1} (g_i - t\text{sgn}(d_{i-1}) - t(i)) + t(i - 1))^2$$  

$$+ (g_i - t\text{sgn}(d_{i-1})1_{i \in S} + (g_i - t\text{sgn}(d_{i-1}))2_{i \in S} + (g_i - t\text{sgn}(d_{i-1}))2_{i \in S} + (g_i - t\text{sgn}(d_{i-1}))2_{i \in S} + (g_i - t\text{sgn}(d_{i-1}))2_{i \in S}.$$  

(28)

Now, we investigate the minimizations in [28], one by one. First, it holds that

$$\inf_{|x| \leq 1} (g_i - t\text{sgn}(d_i) + t(i - 1))^2$$  

$$= \inf_{|x| \leq 1} (g_i - t\text{sgn}(d_i) + t(i - 1))^2$$  

(29)

where the equality (1) is since the optimal objective function of the problem

$$\inf_{|x| \leq 1} (g_i - t\text{sgn}(d_i) + t(i - 1))^2,$$  

occurs either by the boundaries imposed by the feasible set $|x(i-1)| \leq 1$ or equals zero. With a similar reason, we have

$$\inf_{|x| \leq 1} (g_i + t\text{sgn}(d_{i-1}) - t(i))^2 = (|g_i + t\text{sgn}(d_{i-1})| - t)^2,$$  

(30)
and
\[
\inf_{\|z\|_\infty \leq 1} (g_i - tz(i) + tz(i - 1))^2 = \inf_{|z(i)| \leq 1, |z(i-1)| \leq 1} (g_i - tz(i) + tz(i - 1))^2 = (|g_i| - 2t)^2.
\]

Introduce the above expressions into (25) to reach
\[
\text{dist}^2(g, t\bar{\zeta}) \cdot \|\nu TV(x)\| \geq \sum_{i \in S_1^+ \cup S_1^-} (g_i - t\text{sgn}(d_i) + t\text{sgn}(d_{i-1}))^2
+ \sum_{i \in S_2^+} (\zeta_1 - t)^2 + \sum_{i \in S_2^-} (\zeta_2 - t)^2
+ \sum_{i \in S_3^+} (|g_i| - t - t)^2 + (g_i - t\text{sgn}(d_i))^2 1_{1_eS} +
(\|g_i\| - t)^2 1_{1_{-1}eS} + (g_i - t\text{sgn}(d_{i-1}))^2 1_{n-1eS} +
(\|g_i\| - t)^2 1_{n-1eS},
\] (32)

where \(\zeta_1 = |g_i - t\text{sgn}(d_i)|, \zeta_2 = |g_i + t\text{sgn}(d_{i-1})|\). By taking expectation from both sides, we reach
\[
\text{Edist}^2(g, t\bar{\zeta}) \cdot \|\nu TV(x)\| \geq \sum_{i \in S_1^+ \cup S_1^-} t^2(\text{sgn}(d_i) - \text{sgn}(d_{i-1}))^2
+ \sum_{i \in S_2^+} E(\zeta_1 - t)^2 + \sum_{i \in S_2^-} E(\zeta_2 - t)^2
+ \sum_{i \in S_3^+} E(|g_i| - t - t)^2 + (1 + t^2) 1_{1_eS}
E(|g_i| - t)^2 + (1 + t^2) 1_{n-1eS} + E(\|g_i\| - t)^2.
\] (33)

In what follows, we calculate the expressions within (33). First, consider \(E(\zeta_1 - t)^2\), which is calculated as follows:
\[
E(\zeta_1 - t)^2 = 2\int_0^\infty a\mathbb{P}(\zeta_1 \geq a + t)da
= 2\frac{1}{\sqrt{2\pi}} \int_0^\infty 2\pi a e^{-(a+u)^2/2} + e^{-(a+u+2)^2/2} du \, da
= 2\frac{1}{\sqrt{2\pi}} \int_0^\infty 2\pi a e^{-(a+u)^2/2} + e^{-(a+u+2)^2/2} du \, da \equiv \phi_1(t, t),
\] (34)

where in (34), the order of integration is changed together with a change of variable. Similarly, we have: \(E(\zeta_2 - t)^2 = \phi_2(t, t)\). Also,
\[
E(|g_i| - t)^2 = 2\int_0^\infty a\mathbb{P}(|g_i| \geq a + t)da
= 2\sqrt{\frac{2}{\pi}} \int_0^\infty 2\pi a e^{-\frac{a^2}{2}} du \, da
= 2\sqrt{\frac{2}{\pi}} \int_0^\infty a e^{-\frac{a^2}{2}} da \, du =: \phi_2(t),
\] (35)

where in the third line, the order of integration is changed. Thus, we have, \(E(|g_i| - t)^2 = \phi_2(t, t)\). As a consequence, (33) becomes:
\[
\text{Edist}^2(g, t\bar{\zeta}) \cdot \|\nu TV(x)\| \geq |S_1^+ \cup S_1^-| + \sum_{i \in S_1^+ \cup S_1^-} t^2(\text{sgn}(d_i) - \text{sgn}(d_{i-1}))^2
+ \sum_{i \in S_2^+} \phi_1(t, t) + \sum_{i \in S_2^-} \phi_1(t, t)
+ \sum_{i \in S_3^+} \phi_2(t, t) + (1 + t^2) 1_{1_eS} + \phi_2(t) 1_{1_eS}
+ (1 + t^2) 1_{n-1eS} + \phi_2(t) 1_{n-1eS} := \Psi_t(s_1^+, s_1^-, s_2, s_3).
\] (36)

Finally by setting \(s_3 = 1_{1_eS} + 1_{n-1eS}, s_2 = |S_2|, 2 - s_3 = 1_{1_eS} + 1_{n-1eS}, s_1^+ = |S_1^+|, s_1^- = |S_1^-|, \) and \(|S_4| = n - 2 - s_1^+ - s_1^- - s_2\), we reach
\[
\delta(D(\|\nu TV, x\|) \geq \inf_{t \geq 0} \Psi_t(s_1^+, s_1^-, s_2, s_3) = m_{TV}.\] (37)

We know from [13] Theorems 7.1, 6.1 that if
\[
m \leq \delta(D(\|\nu TV, x\|) := \delta,
\]
then,
\[
P[x] is the unique solution of \(P_{TV}\) \leq 4e^{-\frac{(x-m)^2}{4\delta}}.\] (38)

Since the function \(f : z \mapsto 4e^{-\frac{(x-m)^2}{4\delta}}\) is decreasing, and the fact that \(m \leq m_{TV} \leq \delta\), it holds that
\[
P[x] is the unique solution of \(P_{TV}\) \leq 4e^{-\frac{(x-m)^2}{4\delta}} \leq 4e^{-\frac{\delta}{4m_{TV}}}.
\] (39)