

Research Article

Method of Lagrange Multipliers for Normalized Zero Norm Minimization

Bamrung Tausiesakul 

Department of Electrical Engineering, Faculty of Engineering, Srinakharinwirot University, 63 Mu 7, Rangsit-Nakhonnayok Road, Canal 16, Ongkharak, Nakhonnayok 26120, Thailand

Correspondence should be addressed to Bamrung Tausiesakul; bamrungt@g.swu.ac.th

Received 4 September 2021; Accepted 5 November 2021; Published 20 January 2022

Academic Editor: Alessandro Formisano

Copyright © 2022 Bamrung Tausiesakul. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We present a normalization of the p -norm. A compressive sensing criterion is proposed using the normalized zero norm. Based on the method of Lagrange multipliers, we derive the solution of the proposed optimization framework. It turns out that the new solution is a limit case of the least fractional norm solution for $p = 0$, where its fixed-point iteration algorithm can readily follow an existing algorithm. The derivation of the minimal normalized zero norm solution herein gives a relation in the aspect of Lagrange multiplier method to existing works that invoke least fractional norm and least pseudo zero norm criteria.

1. Introduction

Various applications in science and engineering need to recover a desired signal $\mathbf{x} \in \mathbb{R}^{N \times 1}$ from a set of observed data or measured data $\mathbf{b} \in \mathbb{R}^{M \times 1}$ based on a modeling or measurement matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$, which either depends on the model or can be chosen beforehand, for $N \in \mathbb{N}^{1 \times 1}$ and $M \in \mathbb{N}^{1 \times 1}$. A linear system in Figure 1 can be represented by

$$\mathbf{Ax} + \boldsymbol{\delta}_b = \mathbf{b}, \quad (1)$$

where $\boldsymbol{\delta}_b \in \mathbb{R}^{M \times 1}$ is the amount of perturbation hidden in the output \mathbf{b} .

The signal can be recovered by solving an optimization problem related to linear least squares (LLS), i.e.,

$$\begin{aligned} \hat{\mathbf{x}}_{\text{LLS}} &= \arg \min_{\mathbf{x}} \|\mathbf{x}\|_2^2 \\ \text{s.t. } \mathbf{x} &\in \begin{cases} \{\mathbf{x} \in \mathbb{R}^{N \times 1} | \mathbf{Ax} = \mathbf{b}\}, & \text{noiseless,} \\ \{\mathbf{x} \in \mathbb{R}^{N \times 1} | \|\mathbf{Ax} - \mathbf{b}\|_2 < \varepsilon\}, & \text{noisy,} \end{cases} \end{aligned} \quad (2)$$

where ε is the square root of the maximal allowable noise power [1]. Usually, the matrix \mathbf{A} is of full rank, i.e.,

$$\text{rank}(\mathbf{A}) = \min(M, N), \quad (3)$$

where $\text{rank}(\cdot)$ is the rank of matrix. Whether the number of provided data M is greater than, equal to, or less than the number of unknown variables N , i.e., the size of the matrix \mathbf{A} , the LLS problem can be classified into

$$\hat{\mathbf{x}}_{\text{LLS}} = \begin{cases} \arg \min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|_2^2, & M > N, \\ \arg \mathbf{Ax} = \mathbf{b}, & M = N, \\ \arg \min_{\mathbf{x}} \|\mathbf{x}\|_2^2 \text{ s.t. } \mathbf{Ax} = \mathbf{b}, & M < N. \end{cases} \quad (4)$$

The solution of (4) is well known as the LLS estimate

$$\hat{\mathbf{x}}_{\text{LLS}} = \begin{cases} (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}, & M > N, \\ \mathbf{A}^{-1} \mathbf{b}, & M = N, \\ \mathbf{A}^T (\mathbf{A} \mathbf{A}^T)^{-1} \mathbf{b}, & M < N, \end{cases} \quad (5)$$

where T is the transpose of a vector or matrix, and \cdot^{-1} is the inverse of square matrix. The LLS estimate exists provided that the matrix $\mathbf{A}^T \mathbf{A}$ or \mathbf{A} or $\mathbf{A} \mathbf{A}^T$ is invertible, i.e., the

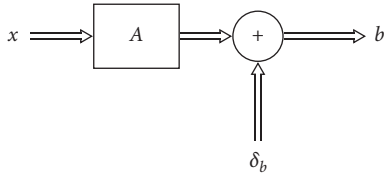


FIGURE 1: A noisy linear input-output system.

measurement matrix \mathbf{A} needs to be of full rank. The LLS estimate $\hat{\mathbf{x}}_{\text{LLS}}$ provides all nonzero entries to the desired signal \mathbf{x} .

Many works indicate that the desired signal often is subject to sparsity, i.e., the situation when a number of elements in \mathbf{x} are zeros. Even though the signal sometimes does not strictly entail the sparsity, it is efficient to keep an approximate value of the signal that contains only a sufficient amount of its largest components, from which we called compressible signal. The sparsity nature of the signal is usually hidden and can be exposed by discovering a sparse basis and its associated spanning coefficients. The decomposition of the coefficient vector can be seen as a superposition of dictionary elements with a remaining term [2]. Compressed sensing was an emerging field that spans many applications in science and engineering, e.g., imaging and vision [3], photonic mixer device [4], electronic defense [5], security and cryptosystem [6], radar [7, 8], earth observation [9], wireless networks [10, 11], biometric watermarking [12], and healthcare [13].

In compressive sensing, the ℓ_0 -norm is originally adopted to impose the zero elements in the solution. The optimization of ℓ_0 -norm criterion however appears to be a combinatorial nondeterministic polynomial-time hard (NP-hard) problem, which appears to be prohibitive. The performance of the above optimization problem can however be analyzed, e.g., in [14].

Instead of the ℓ_0 -norm, ℓ_1 -norm is often of interest because it is more convenient than ℓ_0 -norm optimization in terms of computability while its accuracy is comparable (see e.g., [15, 16]). A widely considered method designed for norm minimization of dictionary coefficients is known as matching pursuit [17]. Its variations are presented in terms of basis pursuit denoising [2], orthogonal matching pursuit [18], compressive sampling matching pursuit [19], stagewise orthogonal matching pursuit [20], gradient pursuits [21], etc. Most approaches based on the matching pursuit involve the ℓ_2 -norm, except for basis pursuit denoising, which considers the ℓ_1 -norm.

In this work, we point out that the zero norm that is mostly adopted in compressed sensing literature is not the actual zero norm, but rather a pseudo zero norm. We also show that the actual zero norm is unbounded and thus trivial. Later we present a normalized p -norm and apply its special case for $p = 0$ to be a new objective function. By using the method of Lagrange multipliers, the proposed constrained optimization is solved and the emerging solution is equal to the limit case of that given by the least fractional norm for $p = 0$.

This paper is organized as follows. In Section 2, we point out that the ℓ_0 norm is diverged or undefined, whereas the so-called zero norm adopted in compressive sensing is actually not a proper norm or is only a pseudonorm. In Section 3, we propose a normalized ℓ_p norm. It is later shown that for $p = 0$ the normalized zero norm is approximately a geometrical mean and unfortunately it does not hold the triangle inequality of the proper norm. In Section 4, we consider the compressive sensing model. The fractional norm for $p \in (0, 1)$ and its criterion presented in the past are revisited herein. The corresponding solution is found by the method of Lagrange multiplier. In Section 5, we propose an alternative criterion based on the normalized zero norm. We later derive its solution through the method of Lagrange multiplier. The solution is found in a closed form and unfortunately turns to be a limiting case to that of the least fractional norm criterion in the former works. In Section 6, numerical examples of the solution are provided in conjunction with other works. Concluding remarks of the paper are provided in Section 7.

2. Conventional Zero Norm

Let $\mathbf{z} \in \mathbb{C}^{N \times 1}$ be a complex-valued vector, expressed as

$$\mathbf{z} = [z_1 \ z_2 \ \cdots \ z_N]^T. \quad (6)$$

Let $p \in \mathbb{Z}_+$ be a positive integer. The ℓ_p -norm, or simply the p -norm, $\|\cdot\|_p: \mathbb{C}^{N \times 1} \mapsto \mathbb{R}^{1 \times 1}$, is given by

$$\|\mathbf{z}\|_p = (|z_1|^p + |z_2|^p + \cdots + |z_N|^p)^{1/p}, \quad (7)$$

where $|\cdot|$ is the absolute value of \cdot . Basic properties of the p -norm in (7) are as follows:

- (i) The p -norm is positive-definite, i.e.,

$$\|\mathbf{z}\|_p \geq 0. \quad (8)$$

- (ii) The p -norm is zero if and only if the vector \mathbf{z} is zero, i.e.,

$$\|\mathbf{z}\|_p = 0 \Leftrightarrow \mathbf{z} = \mathbf{0}, \quad (9)$$

where $\mathbf{0}$ is a zero vector whose all elements are zeros.

- (iii) The p -norm has a triangle inequality, i.e.,

$$\|\mathbf{z} + \boldsymbol{\beta}\|_p \leq \|\mathbf{z}\|_p + \|\boldsymbol{\beta}\|_p, \quad (10)$$

for $\mathbf{z} \in \mathbb{C}^{N \times 1}$ and $\boldsymbol{\beta} \in \mathbb{C}^{N \times 1}$.

- (iv) Furthermore, the p -norm has a scalability or homogeneity, which can be shown as

$$\|a\mathbf{z}\|_p = |a|\|\mathbf{z}\|_p, \quad (11)$$

for $a \in \mathbb{R}^{1 \times 1}$ and $\mathbf{z} \in \mathbb{C}^{N \times 1}$.

2.1. Zero Norm. When p is zero, the 0-norm of a vector can be determined from

$$\begin{aligned}\|\mathbf{z}\|_0 &= \lim_{p \rightarrow 0} \left(|z_1|^p + |z_2|^p + \dots + |z_N|^p \right)^{1/p} \\ &= \lim_{p \rightarrow 0} \left(e^{p \ln(|z_1|)} + e^{p \ln(|z_2|)} + \dots + e^{p \ln(|z_N|)} \right)^{1/p},\end{aligned}\quad (12)$$

where e^{\cdot} is the exponential function of \cdot . From the Taylor's series expansion, we can express

$$\begin{aligned}e^{p \ln(|z_n|)} &= \sum_{k=0}^{\infty} \frac{1}{k!} (p \ln(|z_n|))^k \\ &= \frac{1}{0!} (p \ln(|z_n|))^0 + \frac{1}{1!} p \ln(|z_n|) + \frac{1}{2!} (p \ln(|z_n|))^2 + \dots\end{aligned}\quad (13)$$

For a small value of p , i.e., $p \rightarrow 0$, we can approximate

$$\begin{aligned}e^{p \ln(|z_n|)} &= 1 + p \ln(|z_n|) + O(p^2 \ln^2(|z_n|)) \\ &\approx 1 + p \ln(|z_n|),\end{aligned}\quad (14)$$

where $O(\cdot)$ is the big 'oh' notation of Bachmann–Landau symbols, i.e.,

$$f(x) = O(g(x)); x \rightarrow a \Leftrightarrow \limsup_{x \rightarrow a} \frac{|f(x)|}{|g(x)|} < \infty. \quad (15)$$

Substituting (14) into (12), we obtain

$$\begin{aligned}\|\mathbf{z}\|_0 &\approx \lim_{p \rightarrow 0} \left(1 + p \ln(|z_1|) + 1 + p \ln(|z_2|) \right. \\ &\quad \left. + \dots + 1 + p \ln(|z_N|) \right)^{1/p} \\ &= \lim_{p \rightarrow 0} \left(N + p \sum_{n=1}^N \ln(|z_n|) \right)^{1/p}.\end{aligned}\quad (16)$$

By using a property of logarithm power, we can show that

$$\begin{aligned}\|\mathbf{z}\|_0 &= \lim_{p \rightarrow 0} N^{1/p} \left(1 + \frac{1}{N} p \ln \left(\prod_{n=1}^N |z_n| \right) \right)^{1/p} \\ &= \lim_{p \rightarrow 0} N^{1/p} \left(1 + p \ln \left(\left(\prod_{n=1}^N |z_n| \right)^{1/N} \right) \right)^{1/p} \\ &\approx \lim_{p \rightarrow 0} N^{1/p} e^{\ln \left(\sqrt[N]{\prod_{n=1}^N |z_n|} \right)} \\ &= \lim_{p \rightarrow 0} N^{1/p} \sqrt[N]{\prod_{n=1}^N |z_n|} \\ &\rightarrow \infty.\end{aligned}\quad (17)$$

It is obvious that the 0-norm is unbounded and thus trivial.

2.2. Pseudo Zero Norm. Most works, however, instead consider

$$\begin{aligned}\lim_{p \rightarrow 0} \|\mathbf{z}\|_p^p &= \lim_{p \rightarrow 0} \left(|z_1|^p + |z_2|^p + \dots + |z_N|^p \right)^{1/p} \\ &= |z_1|^0 + |z_2|^0 + \dots + |z_N|^0.\end{aligned}\quad (18)$$

By assigning 0^0 to 0, we have a convenient relation

$$\lim_{p \rightarrow 0} \|\mathbf{z}\|_p^p = |\{z_n | z_n \neq 0; n = 1, 2, \dots, N\}|, \quad (19)$$

where $|\cdot|$ is the cardinality of a set $\{\cdot\}$ or herein the number of nonzero members in a set $\{\cdot\}$. The pseudo zero norm counts the number of nonzero elements in \mathbf{z} . It is important to note that the pseudo zero norm $\lim_{p \rightarrow 0} \|\cdot\|_p^p$ actually is not the zero norm $\|\cdot\|_0$ and is not a proper norm, because it does not preserve the homogeneity. However, it is widely used to replace the zero norm $\|\cdot\|_0$ due to its simple computability.

3. Normalized Zero Norm

Let us introduce a normalization of the p -norm $\|\cdot\|_p$ as

$$\|\mathbf{z}\|_{\bar{p}} = \left(\frac{1}{N} (|z_1|^p + |z_2|^p + \dots + |z_N|^p) \right)^{1/p}. \quad (20)$$

In statistical analysis ([22], Ch. 3), the above quantity is known as p -mean, a generalized mean, Hölder mean, mean of degree p , power mean, etc. We can represent the relation between the normalized p -norm and the conventional p -norm by

$$\begin{aligned}\|\mathbf{z}\|_{\bar{p}} &= \left(\frac{1}{N} \right)^{1/p} (|z_1|^p + |z_2|^p + \dots + |z_N|^p)^{1/p} \\ &= \frac{1}{N^{1/p}} \|\mathbf{z}\|_p.\end{aligned}\quad (21)$$

When p is zero, we can derive

$$\begin{aligned}\|\mathbf{z}\|_0 &= \lim_{p \rightarrow 0} \frac{1}{N^{1/p}} (|z_1|^p + |z_2|^p + \dots + |z_N|^p)^{1/p} \\ &= \lim_{p \rightarrow 0} \frac{1}{N^{1/p}} \left(e^{p \ln(|z_1|)} + e^{p \ln(|z_2|)} + \dots + e^{p \ln(|z_N|)} \right)^{1/p}.\end{aligned}\quad (22)$$

Using the Taylor's series expansion of the exponential function, we have

$$\begin{aligned}\|\mathbf{z}\|_0 &= \lim_{p \rightarrow 0} \frac{1}{N^{1/p}} \left(\sum_{n=1}^N \sum_{k=0}^{\infty} \frac{1}{k!} (p \ln(|z_n|))^k \right)^{1/p} \\ &= \lim_{p \rightarrow 0} \frac{1}{N^{1/p}} \left(\sum_{k=0}^{\infty} \frac{1}{k!} p^k \sum_{n=1}^N \ln^k(|z_n|) \right)^{1/p}.\end{aligned}\quad (23)$$

It is hard to deal with the result in (23). We would rather consider an approximation of (22) associated with (14), which is given by

$$\begin{aligned} \|\mathbf{z}\|_{\bar{0}} \approx & \lim_{p \rightarrow 0} \frac{1}{N^{1/p}} \left(1 + p \ln(|z_1|) + 1 + p \ln(|z_2|) \right. \\ & \left. + \dots + 1 + p \ln(|z_N|) \right)^{1/p}. \end{aligned} \quad (24)$$

Under the same manipulation, we can show that

$$\begin{aligned} \|\mathbf{z}\|_{\bar{0}} + \lim_{p \rightarrow 0} \frac{1}{N^{1/p}} \left(N + p \sum_{n=1}^N \ln(|z_n|) \right)^{1/p} \\ = \lim_{p \rightarrow 0} \frac{1}{N^{1/p}} N^{1/p} \left(1 + \frac{1}{N} p \sum_{n=1}^N \ln(|z_n|) \right)^{1/p} \\ = \lim_{p \rightarrow 0} \left(1 + p \sum_{n=1}^N \ln(|z_n|^{1/N}) \right)^{1/p} \\ = \lim_{p \rightarrow 0} \left(1 + p \ln \left(\prod_{n=1}^N |z_n|^{1/N} \right) \right)^{1/p} \\ \approx \lim_{p \rightarrow 0} \left(e^{p \ln \left(\prod_{n=1}^N |z_n|^{1/N} \right)} \right)^{1/p} \\ = \lim_{p \rightarrow 0} \sqrt[p]{\prod_{n=1}^N |z_n|} \\ = \sqrt[p]{\prod_{n=1}^N |z_n|}. \end{aligned} \quad (25)$$

We can see that due to the first-order Taylor series approximation, $\|\mathbf{z}\|_{\bar{0}}$ can be approximated by the geometric mean of \mathbf{z} . The geometric mean $\sqrt[p]{\prod_{n=1}^N |z_n|}$ has the following properties:

(i) It is positive-definite for any \mathbf{z} , i.e.,

$$\left(|z_1| |z_2| \dots |z_N| \right)^{1/N} \geq 0. \quad (26)$$

(ii) It can simply be zero when only one of all entries in \mathbf{z} is zero, i.e.,

$$\left(|z_1| |z_2| \dots |z_N| \right)^{1/N} = 0 \Leftrightarrow z_n = 0; \quad n \in \{1, 2, \dots, N\}. \quad (27)$$

(iii) It does not hold the triangle inequality, e.g., for $\alpha \in \mathbb{C}^{1 \times 1}$ and $\beta \in \mathbb{C}^{1 \times 1}$,

$$\begin{aligned} \left\| \begin{bmatrix} \alpha \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \beta \end{bmatrix} \right\|_{\bar{0}} &= \left\| \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \right\|_{\bar{0}} \\ &= \sqrt{|\alpha| |\beta|}, \end{aligned} \quad (28)$$

$$\begin{aligned} \left\| \begin{bmatrix} \alpha \\ 0 \end{bmatrix} \right\|_{\bar{0}} + \left\| \begin{bmatrix} 0 \\ \beta \end{bmatrix} \right\|_{\bar{0}} &= 0 + 0 \\ &= 0, \end{aligned}$$

$$\text{which mean } \left\| \begin{bmatrix} \alpha \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \beta \end{bmatrix} \right\|_{\bar{0}} \not\leq \left\| \begin{bmatrix} \alpha \\ 0 \end{bmatrix} \right\|_{\bar{0}} + \left\| \begin{bmatrix} 0 \\ \beta \end{bmatrix} \right\|_{\bar{0}}.$$

(iv) It is homogeneous, i.e.,

$$\begin{aligned} \|\alpha \mathbf{z}\|_{\bar{0}} &= (|\alpha z_1| |\alpha z_2| \dots |\alpha z_N|)^{1/N} \\ &= (|\alpha| |z_1| |\alpha| |z_2| \dots |\alpha| |z_N|)^{1/N} \\ &= (|\alpha|^N |z_1| |z_2| \dots |z_N|)^{1/N} \\ &= |\alpha| (|z_1| |z_2| \dots |z_N|)^{1/N} \\ &= |\alpha| \|\mathbf{z}\|_{\bar{0}}. \end{aligned} \quad (29)$$

(v) It is concave for any \mathbf{z} (see, e.g., [23]) and a monotonically increasing function.

4. Compressive Sensing

Let us consider an underdetermined system where there are more unknown signal components than equations, i.e., $M < N$. In this case, there is infinite number of solutions for \mathbf{x} . Recent works indicate that the desired signal often is subject to sparsity, i.e., the situation when a number of elements in \mathbf{x} are zeros. Even though sometimes the signal does not strictly entail the sparsity, it is efficient to keep an approximate version that contains only a sufficient amount of its largest components, from which we called compressible signal. The sparsity nature of the signal is hidden and can be imposed by discovering a sparse basis and its associated spanning coefficients. The decomposition of the coefficient vector can be seen as a superposition of dictionary elements with a remaining term [2]. Let $K \in \mathbb{Z}_+^{1 \times 1}$ be the number of nonzero elements in \mathbf{x} . If \mathbf{x}_0 is the true value of \mathbf{x} , there is a relation

$$\lim_{p \rightarrow 0} \|\mathbf{x}_0\|_p^p = |\text{supp}(\mathbf{x}_0)| = K, \quad (30)$$

where $|\cdot|$ is the cardinality of set, and $\text{supp}(\mathbf{x}) = \{n | x_n \neq 0\}$ is the support set or the sparsity pattern of \mathbf{x} . The number K is often known as sparsity degree. Compressive sensing can be seen as a problem of finding a K -sparse signal \mathbf{x} [24]. Unfortunately, the solution in (5) does not preserve the inherent sparsity of the signal. A different kind of vector norms can be used to explore the signal sparsity.

The signal recovery can be formulated as an optimization problem, i.e.,

$$\begin{aligned} \hat{\mathbf{x}}_0 &= \arg \min_{\mathbf{x}} \lim_{p \rightarrow 0} \|\mathbf{x}\|_p^p \\ \text{s.t. } \mathbf{x} &\in \begin{cases} \{\mathbf{x} | \mathbf{A}\mathbf{x} = \mathbf{b}\}, & \text{noiseless,} \\ \{\mathbf{x} | \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 < \varepsilon\}, & \text{noisy.} \end{cases} \end{aligned} \quad (31)$$

Note that we do not express the norm in (31) as $\|\cdot\|_0$ similarly to most works, because the zero norm in most works is equal to the pseudo zero norm $\lim_{p \rightarrow 0} \|\cdot\|_p^p$ in this paper. In general, the objective function in terms of the modified zero norm $\lim_{p \rightarrow 0} \|\cdot\|_p^p$ is nonconvex. If \mathbf{A} is an identity matrix, an exact solution of (31) is a hard shrinkage of \mathbf{b} . For arbitrary matrix \mathbf{A} , one may resort to combinatorial optimization. Even an approximate value of the true minimum of the problem in (31) is nondeterministic polynomial-time hard or NP-hard, which appears to be prohibitive. The performance of the above optimization problem can, however, be analyzed, e.g., in [14].

4.1. Compressive Sensing by Fractional Norm. An alternative way is the consideration of $p \leq 1$ [24–26]. When p lies in the range $(0, 1)$,

- (i) the norm $\|\mathbf{x}\|_p$ accepts only a positive real-valued argument, i.e., $\mathbf{x} \in \mathbb{R}_+^{N \times 1}$,
- (ii) it is a concave function, and
- (iii) the fractional norm does not hold the triangle property, e.g.,

$$\begin{aligned} \left\| \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\|_p &= \left\| \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\|_p \\ &= (|1|^p + |1|^p)^{1/p} \\ &= 2^{1/p}, \end{aligned} \quad (32)$$

$$\begin{aligned} \left\| \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\|_p + \left\| \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\|_p &= (|1|^p)^{1/p} + (|1|^p)^{1/p} \\ &= 2, \end{aligned} \quad (33)$$

mean

$$\left\| \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\|_p \not\leq \left\| \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\|_p + \left\| \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\|_p. \quad (34)$$

The last one implies that the fractional norm is not a proper norm or is only a quasinorm. For $p \in (0, 1)$, the compressive sensing problem

$$\begin{aligned} \hat{\mathbf{x}}_p &= \arg \min_{\mathbf{x}} \|\mathbf{x}\|_p^p \\ \text{s.t. } \mathbf{x} &\in \begin{cases} \{\mathbf{x} | \mathbf{A}\mathbf{x} = \mathbf{b}\}, & \text{noiseless,} \\ \{\mathbf{x} | \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 < \varepsilon\}, & \text{noisy,} \end{cases} \end{aligned} \quad (35)$$

is nonconvex, nonsmooth, and non-Lipschitz. The fractional norm gives a closer approximation to the pseudo zero norm than the 1-norm, since the smaller the norm index p , the

sparser the solutions. It is shown in [27] that although a local minimum is found, exact reconstruction is possible with much sparser solution than that required by the 1-norm reconstruction. The case of $p = 1/2$ provides the sparsest solution for $p \in [1/2, 1)$, while the compressive sensing solution with $p \in (0, 1/2]$ has no significant difference from that with $p \in [1/2, 1)$ [28, 29].

4.2. Method of Lagrange Multipliers. Let $\boldsymbol{\lambda} \in \mathbb{C}^{M \times 1}$ be the Lagrange constants. The constrained optimization problem in (35) can be solved by [25], i.e.,

$$\{\hat{\mathbf{x}}_p, \hat{\boldsymbol{\lambda}}_p\} = \arg \min_{\{\mathbf{x}, \boldsymbol{\lambda}\}} L_p(\mathbf{x}, \boldsymbol{\lambda}), \quad (36)$$

where $L_p(\mathbf{x}, \boldsymbol{\lambda})$ is given by

$$L_p(\mathbf{x}, \boldsymbol{\lambda}) = \|\mathbf{x}\|_p^p + \boldsymbol{\lambda}^T (\mathbf{A}\mathbf{x} - \mathbf{b}). \quad (37)$$

The solution of (36) is given by [30]

$$\hat{\mathbf{x}}_p = \mathbf{D}(|\mathbf{x}|^{2-p}) \mathbf{A}^T (\mathbf{A} \mathbf{D}(|\mathbf{x}|^{2-p}) \mathbf{A}^T)^{-1} \mathbf{b}, \quad (38)$$

where $\mathbf{D}(\cdot)$ is the diagonal matrix whose diagonal is taken from vector \cdot , and $|\mathbf{x}|^a$ is given by

$$|\mathbf{x}|^a = \begin{bmatrix} |x_1|^a \\ |x_2|^a \\ \vdots \\ |x_N|^a \end{bmatrix}. \quad (39)$$

We can see that when p is equal to 2, the least fractional norm criterion provides the same result as that shown in (5), i.e.,

$$\begin{aligned} \lim_{p \rightarrow 2} \hat{\mathbf{x}}_p &= \hat{\mathbf{x}}_2 \\ &= \hat{\mathbf{x}}_{\text{LLS}}. \end{aligned} \quad (40)$$

Unlike the solution by the LLS criterion in (5), the solution by the fractional norm in (38) depends on the unknown variable \mathbf{x} , which later needs to be involved with an iterative computation. The computation procedure in an iterative way, which is known as FOCal Underdetermined System Solver (FOCUSS), is summarized in [30].

Algorithm 1 can suffer from many local minima. An alternative tries to avoid the NP-hard problem for $p \in (0, 1)$ by sequentially minimizing a smooth function.

Sometimes, the least fractional exponent norm criterion is called iteratively reweighted least squares (IRLS) in compressive sensing [31–33]. When the vector $\hat{\mathbf{x}}[i-1]$ in Algorithm 1 converges to the true value of \mathbf{x} , a number of elements in $\hat{\mathbf{x}}[i-1]$ may be close to zeros, which can cause an ill condition of the matrix $\mathbf{A} \mathbf{D}(|\hat{\mathbf{x}}[i-1]|^{2-p}) \mathbf{A}^T$. It is suggested in [31] that

$$\begin{aligned} \hat{\mathbf{x}}[i] &\leftarrow \mathbf{D} \left((\mathbf{x}^2 + \varepsilon_r \mathbf{1})^{-(1/2)p} \right) \mathbf{A}^T, \\ &\left(\mathbf{A} \mathbf{D} \left((\mathbf{x}^2 + \varepsilon_r \mathbf{1})^{-(1/2)p} \right) \mathbf{A}^T \right)^{-1} \mathbf{b} \Big|_{\mathbf{x}=\hat{\mathbf{x}}[i-1]}, \end{aligned} \quad (41)$$

```

Input:  $\mathbf{A} \in \mathbb{R}^{M \times N}$ ,  $\mathbf{b} \in \mathbb{R}^{M \times 1}$ ,  $p \in (0, 1]$ ,  $N_{\max} \in \mathbb{Z}_+^{1 \times 1}$ ,  $\varepsilon_{\min} \in \mathbb{R}_+^{1 \times 1}$ 
Output:  $\hat{\mathbf{x}}_p \in \mathbb{R}^{N \times 1}$ 
 $\hat{\mathbf{x}}[0] \leftarrow \mathbf{1}$ 
 $i \leftarrow 0$ 
 $\varepsilon_{\hat{\mathbf{x}}} \leftarrow \varepsilon_{\min} + 1$ 
while  $\varepsilon_{\hat{\mathbf{x}}} > \varepsilon_{\min} \wedge i \leq I_{\max}$  do
 $i \leftarrow i + 1$ 
 $\hat{\mathbf{x}}[i] \leftarrow \mathbf{D}(|\mathbf{x}|^{2-p})\mathbf{A}^T (\mathbf{A}\mathbf{D}(|\mathbf{x}|^{2-p})\mathbf{A}^T)^{-1}\mathbf{b}|_{\mathbf{x}=\hat{\mathbf{x}}[i-1]}$ 
 $\varepsilon_{\hat{\mathbf{x}}} \leftarrow \|\hat{\mathbf{x}}[i] - \hat{\mathbf{x}}[i-1]\|_2 / \|\hat{\mathbf{x}}[i-1]\|_2$ 
end while
return  $\hat{\mathbf{x}}[i]$ 

```

ALGORITHM 1: FOCal Underdetermined System Solver [30].

where $\varepsilon_r \in \mathbb{R}^{1 \times 1}$ is a regularization quantity that is large at the beginning of the iteration and gradually smaller when the iteration converges. In [32], the regularization quantity ε_r in (41) is replaced by its square ε_r^2 . Let $\ell(\mathbf{x})$ be a non-increasing set of all elements in \mathbf{x} , which can be represented by

$$\ell(\mathbf{x}) = \{l_{n_1} | l_{n_1} = x_{n_2}; n_1, n_2 \in \{1, 2, \dots, N\} \wedge |l_1| \geq |l_2| \geq \dots \geq |l_N|\}. \quad (42)$$

Let $L_k(\mathbf{x})$ be the k th element of the nonincreasing set $\ell(\mathbf{x})$, i.e., $L_k(\mathbf{x}) = l_k$.

Algorithm 2 is different from [32] in two aspects. First, the regularization parameter ε_r in [32] is computed from the updated $\hat{\mathbf{x}}[i]$ at the i th iteration, which is unavailable. Second, the procedure addressed in [32] considers only $p = 1$.

5. Compressive Sensing by Normalized Zero Norm

We propose a new criterion by using the normalized zero norm. Under a similar idea to (32), the usual compressive sensing problem can alternatively be formulated as

$$\begin{aligned} \hat{\mathbf{x}}_{\bar{0}} &= \arg \min_{\mathbf{x}} \|\mathbf{x}\|_{\bar{0}} \quad \text{s.t. } \mathbf{x} \in \begin{cases} \{\mathbf{x} | \mathbf{A}\mathbf{x} = \mathbf{b}\}, & \text{noiseless,} \\ \{\mathbf{x} | \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 < \varepsilon\}, & \text{noisy,} \end{cases} \\ &\approx \arg \min_{\mathbf{x}} \sqrt{\prod_{n=1}^N |x_n|} \quad \text{s.t. } \mathbf{x} \in \begin{cases} \{\mathbf{x} | \mathbf{A}\mathbf{x} = \mathbf{b}\}, & \text{noiseless,} \\ \{\mathbf{x} | \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 < \varepsilon\}, & \text{noisy.} \end{cases} \end{aligned} \quad (43)$$

Fortunately, when approximated as the geometrical mean in (25), the normalized zero norm makes the

optimization problem nonconvex under a concave objective function and an affine/convex constraint.

The Lagrange function can be expressed as

$$L_{\bar{0}}(\mathbf{x}, \boldsymbol{\lambda}) = \sqrt{\prod_{n=1}^N |x_n|} + \boldsymbol{\lambda}^T (\mathbf{A}\mathbf{x} - \mathbf{b}). \quad (44)$$

The derivative of the Lagrange function $L_{\bar{0}}(\mathbf{x}, \boldsymbol{\lambda})$ with respect to \mathbf{x} and $\boldsymbol{\lambda}$ can be written as

$$\begin{aligned} \begin{bmatrix} \frac{\partial}{\partial \mathbf{x}} \\ \frac{\partial}{\partial \boldsymbol{\lambda}} \end{bmatrix} L_{\bar{0}}(\mathbf{x}, \boldsymbol{\lambda}) &= \begin{bmatrix} \frac{\partial}{\partial \mathbf{x}} \sqrt{\prod_{n=1}^N |x_n|} + \frac{\partial}{\partial \mathbf{x}} \boldsymbol{\lambda}^T (\mathbf{A}\mathbf{x} - \mathbf{b}) \\ \left(\frac{\partial}{\partial \boldsymbol{\lambda}} \boldsymbol{\lambda}^T \right) (\mathbf{A}\mathbf{x} - \mathbf{b}) \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial}{\partial \mathbf{x}} \sqrt{\prod_{n=1}^N |x_n|} + \frac{\partial}{\partial \mathbf{x}} \boldsymbol{\lambda}^T \mathbf{A}^T \boldsymbol{\lambda} \\ \mathbf{A}\mathbf{x} - \mathbf{b} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{A}^T \boldsymbol{\lambda} + \frac{d}{d\mathbf{x}} \sqrt{\prod_{n=1}^N |x_n|} \\ \mathbf{A}\mathbf{x} - \mathbf{b} \end{bmatrix}. \end{aligned} \quad (45)$$

We can show that

```

Input:  $\mathbf{A} \in \mathbb{R}^{M \times N}$ ,  $\mathbf{b} \in \mathbb{R}^{M \times 1}$ ,  $p \in (0, 1]$ ,  $k \in \mathbb{Z}_+^{1 \times 1}$ 
Output:  $\hat{\mathbf{x}}_p \in \mathbb{R}^{N \times 1}$ 
 $\hat{\mathbf{x}}[0] \leftarrow \mathbf{1}$ 
 $\hat{\mathbf{w}}[0] \leftarrow \mathbf{1}$ 
 $\hat{\varepsilon}_r[0] \leftarrow \mathbf{1}$ 
 $i \leftarrow 0$ 
while  $\hat{\varepsilon}_r[i] \neq 0$  do
 $i \leftarrow i + 1$ 
 $\hat{\varepsilon}_r[i] \leftarrow \min(\hat{\varepsilon}_r[i - 1], 1/NL_{k+1}(\hat{\mathbf{x}}[i - 1]))$ 
 $\hat{\mathbf{x}}[i] \leftarrow \mathbf{D}(\hat{\mathbf{w}}[i - 1])\mathbf{A}^T(\mathbf{A}\mathbf{D}(\hat{\mathbf{w}}[i - 1])\mathbf{A}^T)^{-1}\mathbf{b}$ 
 $\hat{\mathbf{w}}[i] \leftarrow (\hat{\mathbf{x}}^2[i] + \hat{\varepsilon}_r^2[i]\mathbf{1})^{1 - (1/2)p}$ 
end while
return  $\hat{\mathbf{x}}[i]$ 
    
```

ALGORITHM 2: Iteratively reweighted least squares [31, 32].

$$\begin{aligned}
 \frac{d}{d\mathbf{x}} \sqrt{\prod_{n=1}^N |x_n|} &= \begin{bmatrix} \frac{d}{dx_1} \\ \frac{d}{dx_2} \\ \vdots \\ \frac{d}{dx_N} \end{bmatrix} \prod_{n=1}^N |x_n|^{1/N} \\
 &= \begin{bmatrix} \left(\frac{d}{dx_1} |x_1|^{1/N}\right) \prod_{n=2, \neq 1}^N |x_n|^{1/N} \\ \left(\frac{d}{dx_2} |x_2|^{1/N}\right) \prod_{n=1, \neq 2}^N |x_n|^{1/N} \\ \vdots \\ \left(\frac{d}{dx_N} |x_N|^{1/N}\right) \prod_{n=1, \neq N}^N |x_n|^{1/N} \end{bmatrix}. \quad (46)
 \end{aligned}$$

Using the chain rule for a real-valued variable \mathbf{x}_n , the derivative $d/dx_n \sum_{n=1}^N |x_n|^p$ can be written as

$$\begin{aligned}
 \frac{d}{dx_n} |x_n|^p &= \frac{d}{dx_n} \begin{cases} x_n^p, & x_n \geq 0, \\ (-x_n)^p, & x_n < 0, \end{cases} \\
 &= \begin{cases} px_n^{p-1}, & x_n \geq 0, \\ -p(-x_n)^{p-1}, & x_n < 0, \end{cases} \quad (47) \\
 &= \frac{x_n}{|x_n|} p |x_n|^{p-1} \\
 &= p |x_n|^{p-2} x_n.
 \end{aligned}$$

By using the result from (47), we can derive

$$\begin{aligned}
 \frac{d}{d\mathbf{x}} \sqrt{\prod_{n=1}^N |x_n|} &= \begin{bmatrix} \frac{1}{N} |x_1|^{(1/N)-2} x_1 \prod_{n=2, \neq 1}^N |x_n|^{1/N} \\ \frac{1}{N} |x_2|^{(1/N)-2} x_2 \prod_{n=1, \neq 2}^N |x_n|^{1/N} \\ \vdots \\ \frac{1}{N} |x_N|^{(1/N)-2} x_N \prod_{n=1, \neq N}^N |x_n|^{1/N} \end{bmatrix} \\
 &= \frac{1}{N} \begin{bmatrix} \frac{x_1}{|x_1|^2} \prod_{n=1}^N |x_n|^{1/N} \\ \frac{x_2}{|x_2|^2} \prod_{n=1}^N |x_n|^{1/N} \\ \vdots \\ \frac{x_N}{|x_N|^2} \prod_{n=1}^N |x_n|^{1/N} \end{bmatrix} \quad (48) \\
 &= \frac{1}{N} \sqrt{\prod_{n=1}^N |x_n|} D^{-1}(|\mathbf{x}|^2) \mathbf{x}.
 \end{aligned}$$

At the critical point $\partial/\partial \mathbf{x} L_{\bar{0}}(\mathbf{x}, \lambda) = \mathbf{0}$, we have

$$\mathbf{A}^T \boldsymbol{\lambda} + \frac{1}{N} \sqrt{\prod_{n=1}^N |x_n|} \mathbf{D}^{-1} (|\mathbf{x}|^2) \mathbf{x} = \mathbf{0} \Leftrightarrow \hat{\mathbf{x}}_0 = -\frac{1}{\sqrt{\prod_{n=1}^N |x_n|} \mathbf{N} \mathbf{D} (|\mathbf{x}|^2) \mathbf{A}^T \boldsymbol{\lambda}} \quad (49)$$

At the critical point $\partial/\partial \boldsymbol{\lambda}^* L_0(\hat{\mathbf{x}}_0, \boldsymbol{\lambda}) = \mathbf{0}$, we have

$$\begin{aligned} \mathbf{A}\mathbf{x} - \mathbf{b} &= \mathbf{0} \\ \Leftrightarrow -\frac{1}{\sqrt{\prod_{n=1}^N |x_n|}} \mathbf{N} \mathbf{A} \mathbf{D} (|\mathbf{x}|^2) \mathbf{A}^T \boldsymbol{\lambda} &= \mathbf{b} \\ \Leftrightarrow \hat{\boldsymbol{\lambda}}_0 &= -\frac{1}{N} \sqrt{\prod_{n=1}^N |x_n|} (\mathbf{A} \mathbf{D} (|\mathbf{x}|^2) \mathbf{A}^T)^{-1} \mathbf{b}. \end{aligned} \quad (50)$$

Substituting (50) into (49), we obtain

$$\hat{\mathbf{x}}_0 = \mathbf{D} (|\mathbf{x}|^2) \mathbf{A}^T (\mathbf{A} \mathbf{D} (|\mathbf{x}|^2) \mathbf{A}^T)^{-1} \mathbf{b}. \quad (51)$$

It should be noted that the result in (51) is equal to (38) with $p = 0$. Thus, the compressive sensing problem using the geometric mean is the limit case of the fractional norm problem for $p \in (0, 1)$ in (35), i.e.,

$$\lim_{p \rightarrow 0} \hat{\mathbf{x}}_p = \hat{\mathbf{x}}_0, \quad (52)$$

and tends to be the desired but complicated problem with the pseudo zero norm in (31). Although the solution of the minimization of the normalized zero norm by Lagrange multiplier method appears to be the same as the former works, it gives a relation in the Lagrange multiplier method point of view to existing works that invoke least fractional norm and least pseudo zero norm criteria.

6. Numerical Examples

All computer simulations in this work are conducted using Python language. The root-mean-squared relative error (RMSRE), denoted by

$$\sqrt{\mathcal{E}_{\mathbf{x}_0} \left\{ \frac{\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2}{\|\mathbf{x}_0\|_2^2} \right\}} = \sqrt{\mathcal{E}_{\mathbf{x}_0} \left\{ \left(\frac{\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2}{\|\mathbf{x}_0\|_2} \right)^2 \right\}}, \quad (53)$$

is the index for evaluating the performance of each algorithm. It is calculated by the square root of the probabilistic average of the square of the normalized estimation error, where the expectation $\mathcal{E}_{\mathbf{x}_0}\{\cdot\}$ is taken into account with respect to the randomization caused by

- (i) the true value of \mathbf{x}_0 , which is assumed to follow an identical and independent real-valued Gaussian distribution with zero mean and unit variance, i.e.,

$$\mathbf{x}_0 \sim \mathcal{N}_{\mathbb{R}}(\mathbf{0}, \mathbf{I}_N), \quad (54)$$

- (ii) the sparsity pattern of all K nonzero elements in \mathbf{x}_0 , which is assumed to have an equal probability for K locations on all N possible positions ($K < N$ for a sparse signal vector).

The algorithms intended to comparison include

- (i) ℓ_1 -norm, the problem in (4) whose Euclidean norm $\|\cdot\|_2$ is replaced by the ℓ_1 -norm, i.e., $\|\cdot\|_1$,
- (ii) ℓ_2 -norm, the problem in (4),
- (iii) FOCUSS, Algorithm 1 with $\varepsilon_{\min} = 10^{-7}$ and the maximum number of iterations of $N_{\max} = 30$,
- (iv) IRLS, Algorithm 2 with $k = K - 1$, and
- (v) Theoretical approximate normalized zero norm (TANZN); the best possibility of the first iteration of the fixed-point iteration in (52), calculated by substituting \mathbf{x} by the theoretical or true value $\mathbf{x} = \mathbf{x}_0$, i.e.,

$$\hat{\mathbf{x}}_0 = \mathbf{D} (|\mathbf{x}_0|^2) \mathbf{A}^T (\mathbf{A} \mathbf{D} (|\mathbf{x}_0|^2) \mathbf{A}^T)^{-1} \mathbf{b}. \quad (55)$$

The minimizations of the ℓ_1 -norm and ℓ_2 -norm subject to $\mathbf{A}\mathbf{x} = \mathbf{b}$ are conducted by an interior-point solver for convex optimization [34]. For both fixed-point iteration methods, such as the FOCUSS and the IRLS, the norm exponent is assumed to be $p = 0.8$. It should be noted that the solution in (55) is the ideal case of (51), because the true value \mathbf{x}_0 is unknown. The realistic implementations of (55) were addressed in the past, e.g., in terms of the FOCUSS, the IRLS, etc. The design of a more accurate algorithm to the fixed-point iteration required by (51) may remain open for a future work. The matrix inverse in (55) is usually subject to a large condition number, which can cause a numerical failure. One has to resolve this numerical instability by adding a tiny amount, e.g., 10^{-6} , to each diagonal element of $\mathbf{A} \mathbf{D} (|\mathbf{x}_0|^2) \mathbf{A}^T$ before its inverse operation.

In Figure 2, numerical computation is done with $M = 128$ and $N = 256$ from $N_R = 10,000$ independent runs for each value of K . One can see that the ℓ_1 -norm approach is very precise from $K = 2$ to $K = 32$. At this critical region, the error abruptly arises probably because it is beyond the

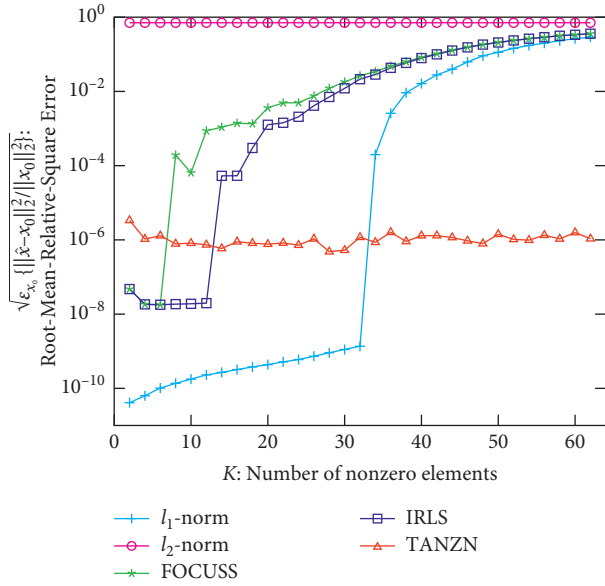


FIGURE 2: RMSRE as a function of K from $N_R = 10,000$ independent runs.

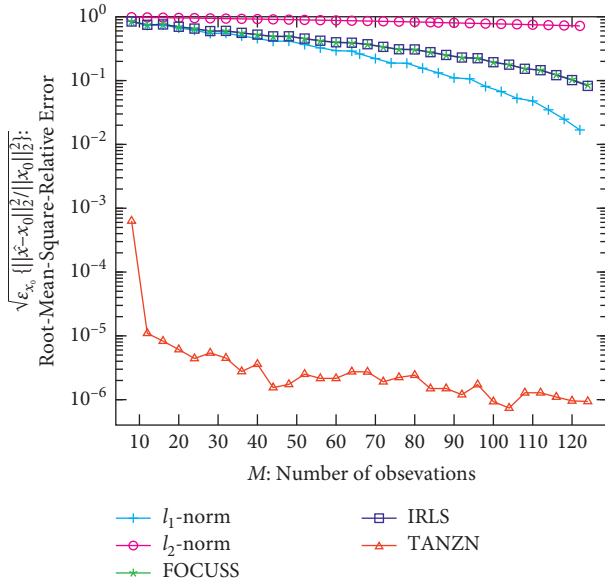


FIGURE 3: RMSRE as a function of M from $N_R = 10,000$ independent runs.

capability of the interior-point method in the convex optimization. The ℓ_2 -norm method does not explore the sparsity nature of the signal vector and thus performs worst. The IRLS performs almost identically to the FOCUSS, except for a little better performance during the transition region. The ideal TANZN method stays constant for any value of K . It is worse than its actual implementations, such as the FOCUSS and the IRLS, for $K < 10$. However, it is worth noting that both fixed-point iteration techniques each involve multiple iterations, e.g., maximal $N_{\max} = 30$ iterations in the FOCUSS, while the TANZN represents the best case for a single substitution or the first iteration.

In Figure 3, we assume that the length of the input elements is $N = 256$ and the number of nonzero elements is $K = \lfloor 0.3125M \rfloor$, which depends on the number output elements M , where $\lfloor \cdot \rfloor$ is the operator of rounding to the nearest lower integer. One can see that when more observed data are available, the RMSRE decreases or the signal acquisition is more precise from all the abovementioned methods. The FOCUSS approach performs identically to the IRLS algorithm. The ℓ_1 -norm minimization is the realistic method that provides the least amount of signal recovery error. The TANZAN technique indicates that if the desired signal \mathbf{x} reaches its true value, the possible acquisition error can be lower than that by the ℓ_1 -norm minimization.

7. Conclusion

A normalization of the p -norm, denoted by $\|\cdot\|_p$ is presented. A compressive sensing criterion using the normalized zero norm $\|\cdot\|_0$ is proposed. Based on the method of Lagrange multipliers, the solution of the proposed optimization framework, i.e., $\hat{\mathbf{x}}_0$ is derived. It turns out that the new solution is a limit case of the fractional norm solution for $p = 0$, where its fixed-point algorithm can readily follow the FOCUSS algorithm in [30]. In our companion works, we find that the minimization of the normalized zero norm by Tikhonov regularization method provides a different solution from that of the fractional norm [28, 29].

Data Availability

The data used to support the findings of this study are available upon request.

Conflicts of Interest

The author declares no conflicts of interest.

References

- [1] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3010–3022, 2005.
- [2] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.
- [3] V. M. Patel and R. Chellappa, *Sparse Representations and Compressive Sensing for Imaging and Vision*, Springer Science, New York, NY, USA, 2013.
- [4] M. H. Conde, Ed., *Compressive Sensing for the Photonic Mixer Device: Fundamentals, Methods and Results*, Springer Fachmedien, Wiesbaden, Germany, 2017.
- [5] A. K. Mishra and R. S. Verster, *Compressive Sensing Based Algorithms for Electronic Defence, Ser. Signals and Communication Technology*, W.-S. Gan, C.-C. J. Kuo, T. F. Zheng, and M. Barni, Eds., Springer International, Cham, Switzerland, 2017.
- [6] M. Testa, D. Valsesia, T. Bianchi, and E. Magli, *Compressed Sensing for Privacy-Preserving Data Processing, Ser. Signal Processing*, W.-S. Gan, C.-C. J. Kuo, T. F. Zheng, and M. Barni, Eds., Springer Nature, Singapore, 2019.

- [7] M. Amin, Ed., *Compressive Sensing for Urban Radar*, CRC Press, Boca Raton, FL, USA, 2015.
- [8] A. D. Maio, Y. C. Eldar, and A. M. Haimovich, Eds., *Compressed Sensing in Radar Signal Processing*, Cambridge University Press, Cambridge, UK, 2020.
- [9] C. Chen, *Compressive Sensing of Earth Observations, Ser. Signal and Image Processing of Earth Observations*, C. Chen, Ed., CRC Press, Boca Raton, FL, USA, 2017.
- [10] Z. Han, H. Li, and W. Yin, Eds., *Compressive Sensing for Wireless Networks*, Cambridge University Press, Cambridge, UK, 2013.
- [11] L. Kong, B. Wang, and G. Chen, *When Compressive Sensing Meets Mobile Crowdsensing*, Springer Nature, Singapore, 2019.
- [12] R. M. Thanki, V. J. Dwivedi, and K. R. Borisagar, Eds., *Multibiometric Watermarking with Compressive Sensing Theory: Techniques and Applications, Ser. Signals and Communication Technology*, Springer International, Cham, Switzerland, 2018.
- [13] M. Khosravy, N. Dey, and C. A. Duque, *Compressive Sensing in Healthcare, Ser. Advances in Ubiquitous Sensing Applications for Healthcare*, N. Dey, A. S. Ashour, and S. J. Fong, Eds., Academic Press, London, UK, 2020.
- [14] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 4036–4048, 2006.
- [15] M. Elad, "Optimized projections for compressed sensing," *IEEE Transactions on Signal Processing*, vol. 55, no. 12, pp. 5695–5702, 2007.
- [16] H. Shen, X. Li, L. Zhang, D. Tao, and C. Zeng, "Compressed sensing-based inpainting of aqua moderate resolution imaging spectroradiometer band 6 using adaptive spectrum-weighted sparse Bayesian dictionary learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 2, pp. 894–906, 2014.
- [17] S. G. Mallat and Z. Zhifeng Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.
- [18] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.
- [19] D. Needell and J. A. Tropp, "CoSaMP: iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.
- [20] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 1094–1121, 2012.
- [21] T. Blumensath and M. E. Davies, "Gradient pursuits," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2370–2382, 2008.
- [22] P. S. Bullen, "Means and their inequalities," *Mathematics and its Applications*, Vol. 560, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2nd edition, 2004.
- [23] S. Boyd and L. Vandenberghe, *Convex Optimization, Seventh Printing with Corrections 2009*, Cambridge University Press, New York, NY, USA, 2004.
- [24] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 1289–1306, 2006.
- [25] B. D. Rao and K. Kreutz-Delgado, "An affine scaling methodology for best basis selection," *IEEE Transactions on Signal Processing*, vol. 47, no. 1, pp. 187–200, 1999.
- [26] S. Foucart and M.-J. Lai, "Sparsest solutions of underdetermined linear systems via l_q -minimization for $0 < q \leq 1$," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 395–497, 2009.
- [27] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Processing Letters*, vol. 14, no. 10, pp. 707–710, 2007.
- [28] Z. Zongben Xu, X. Xiangyu Chang, F. Fengmin Xu, and H. Hai Zhang, " $L_{1/2}$ regularization: a thresholding representation theory and a fast solver," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 7, pp. 1013–1027, 2012.
- [29] J. Jinshan Zeng, S. Shaobo Lin, Y. Yao Wang, and Z. Zongben Xu, " $L_{1/2}$ regularization: convergence of iterative half thresholding algorithm," *IEEE Transactions on Signal Processing*, vol. 62, no. 9, pp. 2317–2329, 2014.
- [30] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 600–616, 1997.
- [31] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing 2008 (ICASSP 2008)*, pp. 3869–3872, Las Vegas, NV, USA, April 2008.
- [32] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Communications on Pure and Applied Mathematics*, vol. 63, no. 1, pp. 1–38, 2010.
- [33] M.-J. Lai, Y. Xu, and W. Yin, "Improved iteratively reweighted least squares for unconstrained smoothed l_q minimization," *SIAM Journal on Numerical Analysis*, vol. 51, no. 2, pp. 927–957, 2013.
- [34] S. Diamond and S. Boyd, "CVXPY: a python-embedded modeling language for convex optimization," *Journal of Machine Learning Research*, vol. 17, no. 83, pp. 2909–2913, 2016.