

Article

Error Estimations for Total Variation Type Regularization

Kuan Li ^{1,2,*} , Chun Huang ³ and Ziyang Yuan ⁴¹ School of Cyberspace Security, Dongguan University of Technology, Dongguan 523808, China² Guangdong Key Laboratory of Intelligent Information Processing, Shenzhen 518060, China³ College of Computer Science and Technology, National University of Defense Technology, Changsha 410073, China; chunhuang@nudt.edu.cn⁴ Department of Mathematics, National University of Defense Technology, Changsha 410073, China; yuanziyang11@nudt.edu.cn

* Correspondence: likuan@dgut.edu.cn

Abstract: This paper provides several error estimations for total variation (TV) type regularization, which arises in a series of areas, for instance, signal and imaging processing, machine learning, etc. In this paper, some basic properties of the minimizer for the TV regularization problem such as stability, consistency and convergence rate are fully investigated. Both a priori and a posteriori rules are considered in this paper. Furthermore, an improved convergence rate is given based on the sparsity assumption. The problem under the condition of non-sparsity, which is common in practice, is also discussed; the results of the corresponding convergence rate are also presented under certain mild conditions.

Keywords: total variation; regularization; inverse problem



Citation: Li, K.; Huang, C.; Yuan, Z. Error Estimations for Total Variation Type Regularization. *Mathematics* **2021**, *9*, 1373. <https://doi.org/10.3390/math9121373>

Academic Editor: Christine Böckmann

Received: 1 June 2021

Accepted: 11 June 2021

Published: 13 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Compressed sensing [1,2] has gained increasing attention in recent years; it plays an important role in signal processing [3,4], imaging science [5,6] and machine learning [7]. Compressed sensing focusses on signals with sparse presentation. Let \mathcal{H}_1 be a Hilbert space, and $\{e_i \in \mathcal{H}_1 | i \in \mathbb{N}\}$ be the orthonormal basis of \mathcal{H}_1 . For any $x \in \mathcal{H}_1$, let $x_i := \langle x, e_i \rangle$. Given some operators K satisfy certain conditions, it is possible to recover a sparse $x^\dagger \in \mathbb{C}^n$ signal with length n by Basis Pursuit (BP) [8], i.e.,

$$\min \|x\|_1 \text{ s.t. } y^\dagger = Kx,$$

from the samples $y^\dagger = Kx^\dagger$, even K is ill-posed [2,9,10]. However, in most cases, noise is inevitable. The literatures has turned to studying the noised BP model

$$\min \|x\|_1 \text{ s.t. } \|Kx - y^\dagger\|_2 \leq \delta,$$

where δ is the allowed error. Actually, the unconstrained form of the noised BP model, i.e., sparse regularization which is the focus in [11–16] is more attractive. While the success of compressed sensing greatly inspired the development of sparse regularization, it is interesting to see that sparse regularization appeared much earlier than compressed sensing [11,12]. As an inverse problem, the error theory of sparse regularization is well studied in the literature [17–19].

In practical terms, a large crowd of signals is not sparse unless being transformed by some operators (maybe ill posed). Thus, many studies have been proposed to analyze the regularized optimization problem [20]. A typical example of them is signal with a sparse gradient which arises frequently from imaging processing (nature images are usually piecewise constant, i.e., they have a sparse gradient). The Total Variation (TV) has been used extensively in the literature for decades in imaging sciences and a series of techniques have

been dedicated to researching its choice of regularization parameter [21–31]; others [32,33] are developed based on this observation. Similar to [34], Total Variation can also smooth the signal of interest. Let \mathcal{H}_2 be another Hilbert space. For any $x \in \mathcal{H}_1$, define that $T : \mathcal{H}_1 \mapsto \mathcal{H}_1$ satisfies

$$(Tx)_i := x_i - x_{i+1}.$$

Under the above definition, T is an ill-posed linear operator. Given a linear map $K : \mathcal{H}_1 \mapsto \mathcal{H}_2$ and $y^\delta \in \mathcal{H}_2$, the total variation regularization problem can be represented as

$$\Psi_\alpha(x) = \frac{1}{2} \|Kx - y^\delta\|_2^2 + \alpha \sum_i |(Tx)_i|,$$

where $\alpha > 0$ is the regularization parameter. The regularization term $\sum_i |(Tx)_i|$ is the right total variation (TV) of x . The TV type regularization has a similar form to the sparse regularization. However, the perfect reconstruction result established in sparse regularization can not be applied to the TV type directly, especially when T is ill-posed (T has a nontrivial null space).

So in this paper, firstly, we discuss the stability and consistency of the minimizers of Ψ_α . Besides basic properties, we are also interested in the convergence rate to solve the TV problem. Then, under the source conditions [19,35,36], convergence rates get obtained for both a priori and a posteriori parameter choice rules. However, the linear convergence rate requires K to be injective, which is strict usually. In the latter part, the linear convergence rate can also be derived under the sparsity assumption on Tx^\dagger and some suitable conditions for K . This requirement of deduction does not depend on the injectivity of K . Meanwhile, this paper also considers the case when the sparsity assumption on Tx^\dagger fails. Last, based on some recent works [37–39], which also assume the Tx^\dagger is not sparse, a convergence rate is also given in this case.

The rest of this paper is organized as follows. Section 2 provides a brief summary of the notations. Section 3 presents some basic properties and gives the convergence rate of the minimizer. Section 4 proves the improved convergence rate. Finally, Section 5 concludes the whole paper.

2. Notation

The notations described in this section are adopted throughout this paper. Let $\mathcal{H}_1, \mathcal{H}_2$ be two Hilbert spaces and $\{e_i \in \mathcal{H}_1 | i \in \mathbb{N}\}, \{\xi_i \in \mathcal{H}_2 | i \in \mathbb{N}\}$ be the orthonormal basis of \mathcal{H}_1 and \mathcal{H}_2 , respectively. For any $x \in \mathcal{H}_1$ and $y \in \mathcal{H}_2$, $x_i := \langle x, e_i \rangle$ and $y_i := \langle y, \xi_i \rangle$. The ℓ_1 and ℓ_2 norms of x and y are denoted by $\|x\|_{\ell_1} := \sum_i |x_i|$, $\|x\|_{\ell_2} := (\sum_i |x_i|^2)^{\frac{1}{2}}$ and $\|y\|_1 := \sum_i |y_i|$, $\|y\|_2 := (\sum_i |y_i|^2)^{\frac{1}{2}}$, respectively. In this paper, if not specified, for any $x \in \mathcal{H}_1$ and $y \in \mathcal{H}_2$, we assume that $x, y \in L^2$, i.e., $\|x\|_{\ell_2} < +\infty$ and $\|y\|_2 < +\infty$. $x^n \rightharpoonup x$ means that x^n converges weakly to x , while $x^n \rightarrow x$ means x^n converges strongly to x . The operator norm of the linear operator $K : \mathcal{H}_1 \mapsto \mathcal{H}_2$ is defined as

$$\|K\| := \max_{\|x\|_{\ell_2}=1} \|Kx\|_2.$$

Through the paper, x^\dagger means the signal of interest; $y^\dagger := Kx^\dagger$ are the measurements. y^δ denotes an element in \mathcal{H}_2 satisfying $\|y^\dagger - y^\delta\|_2 \leq \delta$. Under these notations, the TV regularization can be expressed as

$$\Psi_\alpha(x) = \frac{1}{2} \|Kx - y^\delta\|_2^2 + \alpha \|Tx\|_{\ell_1}.$$

Denote that x_α^δ is one of the minimizers of Ψ_α .

Remark 1. Considering the set $L = \{x^n\}_{n=1,2,\dots} \subseteq \mathcal{H}_1$, where

$$x_i^n := \begin{cases} 1/\sqrt{n} & \text{if } i \leq n, \\ 0 & \text{if } i > n. \end{cases}$$

Obviously, for any n , $x^n \in L^2$ and $T(x^n) = 1/\sqrt{n}$. As $n \rightarrow +\infty$, $\|x^n\|_{\ell^2} = 1$ and $T(x^n) \rightarrow 0$. That means T is ill posed.

Remark 2. Let $D = T - I_d$, where I_d is the identical operator over \mathcal{H}_1 . Then, $(Dx)_i = -x_{i+1}$ for any $i \in \mathbb{N}$. It is easy to verify that D is continuous. Then, T is continuous over \mathcal{H}_1 and

$$\|T - I_d\| = \|D\| \leq 1. \tag{1}$$

In practice, The ill condition of T brings trouble to the analysis. To overcome this problem, we consider a condition which plays an important role in the deduction.

Condition 1. There exist two constants $c, m > 0$ such that

$$c\|Kx\|_2 + m\|Tx\|_{\ell^2} \geq \|x\|_{\ell^2}$$

for any $x \in \mathcal{H}_1$.

We present a finite-dimensional understanding of this condition. Let $\dim(\mathcal{H}_1) = M$ and $\dim(\mathcal{H}_2) = N$. Then, $K \in \mathcal{R}^{M \times N}$ satisfies $\text{null}(K) \neq \mathbf{0}$ and $T \in \mathcal{R}^{(N-1) \times N}$. In the finite dimension case, T has the form

$$\begin{pmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & \ddots & \ddots & \\ & & & 1 & -1 \end{pmatrix}_{(N-1) \times N}.$$

The definition of T gives that $\text{null}(T) = \text{span}(\vec{\mathbf{1}})$. If $K\vec{\mathbf{1}} \neq \mathbf{0}$. Then, $\text{null}(K) \cap \text{null}(T) = \mathbf{0}$, we have that $\text{null}\begin{pmatrix} \hat{c}K \\ \hat{m}T \end{pmatrix} = \mathbf{0}$, where $\hat{c}, \hat{m} > 0$. Hence, for any $x \in \mathcal{R}^n$ and some $\iota > 0$, $\iota\left\|\begin{pmatrix} \hat{c}K \\ \hat{m}T \end{pmatrix}x\right\|_2 \geq \|x\|_2$. Note that $\left\|\begin{pmatrix} \hat{c}K \\ \hat{m}T \end{pmatrix}x\right\|_2 \leq \hat{c}\|Kx\|_2 + \hat{m}\|Tx\|_2$; we then have $\hat{c}\|Kx\|_2 + \hat{m}\|Tx\|_2 \geq \|x\|_2$.

3. Basic Error Estimations

The properties of TV type regularization are investigated in this section. First, a lemma is introduced which is used in this section frequently.

Lemma 1. Let y^δ be bounded, α be fixed and $\{x^n\}_{n=1,2,\dots}$ be a sequence. Assume that Condition 1 holds and $\{\Psi_\alpha(x^n)\}_{n=1,2,\dots}$ is bounded. Then, $\{x^n\}_{n=1,2,\dots}$ is also bounded.

Proof. It is trivial to prove $\{\|Kx^n - y^\delta\|_2\}_{n=1,2,\dots}$ and $\{\|Tx^n\|_{\ell^1}\}_{n=1,2,\dots}$ are bounded. Note that

$$\|Tx^n\|_{\ell^2} \leq \|Tx^n\|_{\ell^1} \text{ and } \|Kx^n\|_2 \leq \|Kx^n - y^\delta\|_2 + \|y^\delta\|_2,$$

which implies $\{\|Kx^n\|_2\}_{n=1,2,\dots}$ and $\{\|Tx^n\|_{\ell^2}\}_{n=1,2,\dots}$ are bounded. From Condition 1, we derive that

$$\|x^n\|_{\ell^2} \leq c\|Kx^n\|_2 + m\|Tx^n\|_{\ell^2}$$

implies the boundedness of $\{x^n\}_{n=1,2,\dots}$. \square

3.1. Stability

In this subsection, we investigate the performance of $x_{\hat{\alpha}}^{\delta}$ as $\hat{\alpha} \rightarrow \alpha$, when y^{δ} is fixed. A lemma is introduced which arises in convex optimization.

Lemma 2 ([40,41]). *Let χ^* be the solution set of the convex minimization problem*

$$\min_x \Psi_{\alpha}(x).$$

Then, Kx and $\|Tx\|_1$ is constant over χ^ .*

Theorem 1. *Assume that K, T satisfies Condition 1. For any fixed $\alpha > 0$ and $y^{\delta} \in \mathcal{H}_2$, we have*

$$\lim_{\alpha_n \rightarrow \alpha} Kx_{\alpha_n}^{\delta} = Kx_{\alpha}^{\delta}. \tag{2}$$

Proof. The minimizing property of $x_{\alpha_n}^{\delta}$ gives that $\frac{1}{2}\|Kx_{\alpha_n}^{\delta} - y^{\delta}\|_2^2 + \alpha\|Tx_{\alpha_n}^{\delta}\|_1 \leq \Psi_{\alpha_n}(0)$. Then, Lemma 1 indicates that there exists a subsequence of $\{x_{\alpha_n}^{\delta}\}$ converging weakly to some $x^* \in \ell^2$. For simplicity, we also denote this subsequence as $\{x_{\alpha_n}^{\delta}\}$. By the weak lower continuity of the norms, we have

$$\|Kx^* - y^{\delta}\|_2 \leq \liminf_n \|Kx_{\alpha_n}^{\delta} - y^{\delta}\|_2 \text{ and } \|Tx^*\|_{\ell^1} \leq \liminf_n \|Tx_{\alpha_n}^{\delta}\|_{\ell^1}. \tag{3}$$

Therefore, we have that

$$\begin{aligned} \Psi_{\alpha}(x^*) &= \frac{1}{2}\|Kx^* - y^{\delta}\|_2^2 + \alpha\|Tx^*\|_1 \\ &\leq \liminf_n \left\{ \frac{1}{2}\|Kx_{\alpha_n}^{\delta} - y^{\delta}\|_2^2 + \alpha_n\|Tx_{\alpha_n}^{\delta}\|_{\ell^1} \right\} \\ &= \liminf_n \Psi_{\alpha_n}(x_{\alpha_n}^{\delta}). \end{aligned}$$

On the other hand, by the minimizing property of $x_{\alpha_n}^{\delta}$,

$$\limsup_n \Psi_{\alpha_n}(x_{\alpha_n}^{\delta}) \leq \limsup_n \Psi_{\alpha_n}(x_{\alpha}^{\delta}) = \lim_n \Psi_{\alpha_n}(x_{\alpha}^{\delta}) = \Psi_{\alpha}(x_{\alpha}^{\delta}).$$

Obviously, it holds that

$$\limsup_n \Psi_{\alpha_n}(x_{\alpha_n}^{\delta}) \leq \Psi_{\alpha}(x_{\alpha}^{\delta}) \leq \Psi_{\alpha}(x^*) \leq \liminf_n \Psi_{\alpha_n}(x_{\alpha_n}^{\delta}).$$

That means x^* minimizes $\Psi_{\alpha}(x)$. From Lemma 2, $Kx^* = Kx_{\alpha}^{\delta}$ and $\|Tx^*\|_{\ell^1} = \|Tx_{\alpha}^{\delta}\|_{\ell^1}$. Consequently, we have $Kx_{\alpha_n}^{\delta} \rightarrow Kx_{\alpha}^{\delta}$, $\Psi_{\alpha_n}(x_{\alpha_n}^{\delta}) \rightarrow \Psi_{\alpha}(x_{\alpha}^{\delta})$ and $\|Tx_{\alpha}^{\delta}\|_{\ell^1} \leq \liminf_n \|Tx_{\alpha_n}^{\delta}\|_{\ell^1}$. In the following, we present the proof by the mean of contradiction. Assume that $t := \limsup_n \|Kx_{\alpha_n}^{\delta} - y^{\delta}\|_2 > \|Kx_{\alpha}^{\delta} - y^{\delta}\|_2$. We can obtain that

$$\begin{aligned} \alpha\|Tx_{\alpha}^{\delta}\|_{\ell^1} &\leq \liminf_n \{ \alpha\|Tx_{\alpha_n}^{\delta}\|_{\ell^1} \} \\ &= \liminf_n \{ \Psi_{\alpha_n}(x_{\alpha_n}^{\delta}) - \|Kx_{\alpha_n}^{\delta} - y^{\delta}\|_2^2 \} \\ &= \Psi_{\alpha}(x_{\alpha}^{\delta}) - \limsup_n \|Kx_{\alpha_n}^{\delta} - y^{\delta}\|_2^2 \\ &= \alpha\|Tx_{\alpha}^{\delta}\|_{\ell^1} + (\|Kx_{\alpha}^{\delta} - y^{\delta}\|_2 - t) \\ &< \alpha\|Tx_{\alpha}^{\delta}\|_{\ell^1}. \end{aligned} \tag{4}$$

This is a contradiction. Then, we have

$$\limsup_n \|Kx_{\alpha_n}^{\delta} - y^{\delta}\|_2 \leq \|Kx_{\alpha}^{\delta} - y^{\delta}\|_2.$$

From relations (3), we can obtain that $Kx_{\alpha_n}^\delta \rightarrow Kx_\alpha^\delta$. \square

If K is injective, we can further have that $\lim_{\alpha_n \rightarrow \alpha} x_{\alpha_n}^\delta = x_\alpha^\delta$. The theorem above indicates that $\Psi_\alpha(x_\alpha^\delta)$ and $\|Tx_\alpha^\delta\|_1$ are continuous at α . In fact, we can obtain a stronger result; the value function is differentiable at α .

Theorem 2. Let $F(\alpha) := \Psi_\alpha(x_\alpha^\delta)$; then, $F(\alpha)$ is differentiable with respect to α , and $F'(\alpha) = \|Tx_\alpha^\delta\|_1$.

Proof. For $\alpha > \hat{\alpha}$, we have

$$\begin{aligned} F(\alpha) - F(\hat{\alpha}) &= \frac{1}{2} \|Kx_\alpha^\delta - y^\delta\|_2^2 + \alpha \|Tx_\alpha^\delta\|_{\ell^1} - \frac{1}{2} \|Kx_{\hat{\alpha}}^\delta - y^\delta\|_2^2 \\ &\quad - \alpha \|Tx_{\hat{\alpha}}^\delta\|_{\ell^1} + (\alpha - \hat{\alpha}) \|Tx_{\hat{\alpha}}^\delta\|_{\ell^1}. \end{aligned}$$

Due to that x_α^δ minimizing Ψ_α , we have

$$\frac{1}{2} \|Kx_\alpha^\delta - y^\delta\|_2^2 + \alpha \|Tx_\alpha^\delta\|_{\ell^1} - \frac{1}{2} \|Kx_{\hat{\alpha}}^\delta - y^\delta\|_2^2 - \alpha \|Tx_{\hat{\alpha}}^\delta\|_{\ell^1} \leq 0.$$

It follows that $F(\alpha) - F(\hat{\alpha}) \leq (\alpha - \hat{\alpha}) \|Tx_{\hat{\alpha}}^\delta\|_{\ell^1}$. On the other hand, $F(\alpha) - F(\hat{\alpha})$ can be written as

$$\begin{aligned} F(\alpha) - F(\hat{\alpha}) &= \frac{1}{2} \|Kx_\alpha^\delta - y^\delta\|_2^2 + \hat{\alpha} \|Tx_\alpha^\delta\|_{\ell^1} - \frac{1}{2} \|Kx_{\hat{\alpha}}^\delta - y^\delta\|_2^2 \\ &\quad - \hat{\alpha} \|Tx_{\hat{\alpha}}^\delta\|_{\ell^1} + (\alpha - \hat{\alpha}) \|Tx_\alpha^\delta\|_{\ell^1}. \end{aligned}$$

Similarly, we have $F(\alpha) - F(\hat{\alpha}) \geq (\alpha - \hat{\alpha}) \|Tx_\alpha^\delta\|_{\ell^1}$. Combining the two inequalities above, we have

$$\|Tx_\alpha^\delta\|_{\ell^1} \leq \frac{F(\alpha) - F(\hat{\alpha})}{\alpha - \hat{\alpha}} \leq \|Tx_{\hat{\alpha}}^\delta\|_{\ell^1}.$$

When $\alpha < \hat{\alpha}$, similar results can be also obtained. The continuity of $\|Tx_\alpha^\delta\|_{\ell^1}$ at α gives that $\frac{dF(\alpha)}{d\alpha} = \|Tx_\alpha^\delta\|_{\ell^1}$. \square

3.2. Consistency

The performance of x_α^δ is investigated under a prior parameter choice as $\delta \rightarrow 0$. In the analysis, we assume that the following conditions hold.

Condition 2. For any $x \in \mathcal{H}_1$ obeying $Kx = y^\dagger$, x^\dagger satisfies that

$$\|Tx^\dagger\|_{\ell^1} \leq \|Tx\|_{\ell^1}.$$

The equality holds if and only if $x = x^\dagger$.

Lemma 3. Let $\{x^n\}_n \rightarrow x^*$, $\|Kx^n - y^\dagger\|_2^2 \rightarrow \|Kx^* - y^\dagger\|_2^2$ and $\|Tx^n\|_{\ell^1} \rightarrow \|Tx^*\|_{\ell^1}$. Then, we have $\|T(x^n - x^*)\|_{\ell^1} \rightarrow 0$ and $\|K(x^n - x^*)\|_2 \rightarrow 0$.

Proof. We can obtain that

$$\begin{aligned} \limsup_n \|T(x^n - x^*)\|_{\ell^1} &= \limsup_n (\|Tx^n\|_{\ell^1} + \|Tx^*\|_{\ell^1}) \\ &\quad - (\|Tx^n\|_{\ell^1} + \|Tx^*\|_{\ell^1}) + \|T(x^n - x^*)\|_{\ell^1}) \\ &= 2\|Tx^*\|_{\ell^1} - \liminf_n (\|Tx^n\|_{\ell^1} + \|Tx^*\|_{\ell^1}) \\ &\quad - \|T(x^n - x^*)\|_{\ell^1}. \end{aligned}$$

The triangle inequality gives that $(\|Tx^n\|_{\ell^1} + \|Tx^*\|_{\ell^1}) - \|T(x^n - x^*)\|_{\ell^1} \geq 0$. The Fatou’s lemma gives that

$$\begin{aligned} &\liminf_n (\|Tx^n\|_{\ell^1} + \|Tx^*\|_{\ell^1} - \|T(x^n - x^*)\|_{\ell^1}) \\ &= \liminf_n \left(\sum_i |(Tx^n)_i| + |(Tx^*)_i| - |(T(x^n - x^*))_i| \right) \\ &\leq \sum_i \liminf_n (|(Tx^n)_i| + |(Tx^*)_i| - |(T(x^n - x^*))_i|). \end{aligned}$$

Note that $x^n - x^* \rightarrow \mathbf{0}$; then, $T(x^n - x^*) \rightarrow T\mathbf{0} = \mathbf{0}$. Hence, $[T(x^n - x^*)]_i \rightarrow 0$. Similarly, we can obtain $(Tx^n)_i \rightarrow (Tx^*)_i$. Therefore,

$$\sum_i \liminf_n (|(Tx^n)_i| + |(Tx^*)_i| - |(T(x^n - x^*))_i|) = \sum_i 2|(Tx^*)_i| = 2\|Tx^*\|_{\ell^1}.$$

Thus, we have

$$\limsup_n \|T(x^n - x^*)\|_{\ell^1} = 0.$$

By the same method, we also can obtain that $\|K(x^n - x^*)\|_2 \rightarrow 0$. \square

Theorem 3. Assume that K, T satisfies Condition 1 and Lemma 1. Let the parameters satisfy that

$$\alpha(\delta), \frac{\delta^2}{\alpha(\delta)} \rightarrow 0 \text{ as } \delta \rightarrow 0.$$

Then the sequence $\{x_\alpha^\delta\}_\delta \rightarrow x^\dagger$.

Proof. By the definition of x_α^δ , we have

$$\begin{aligned} \frac{1}{2} \|Kx_\alpha^\delta - y^\delta\|_2^2 + \alpha \|Tx_\alpha^\delta\|_{\ell^1} &\leq \frac{1}{2} \|Kx^\dagger - y^\delta\|_2^2 + \alpha \|Tx^\dagger\|_{\ell^1} \\ &\leq \frac{1}{2} \delta^2 + \alpha \|Tx^\dagger\|_{\ell^1}. \end{aligned}$$

From the parameters’ choice rule of α and δ , we can see that $\{\Psi_\alpha(x_\alpha^\delta)\}$ are bounded. Then, from Lemma 1, there exists a subsequence also denoted by $\{x_\alpha^\delta\}_\delta$ and some point x^* such that $x_\alpha^\delta \rightarrow x^*$. We can have that

$$\begin{aligned} \|Kx^* - y^\dagger\|_2^2 &\leq \liminf_\delta \|Kx_\alpha^\delta - y^\dagger\|_2^2 \\ &\leq 2 \liminf_\delta (\|Kx_\alpha^\delta - y^\delta\|_2^2 + \|y^\delta - y^\dagger\|_2^2) \\ &\leq \liminf_\delta (\delta^2 + 2\alpha(\delta) \|Tx^\dagger\|_{\ell^1} + \delta^2) = 0. \end{aligned}$$

This means $Kx^* = y^\dagger$. It is easy to see that $\lim_\delta \|Kx_\alpha^\delta - y^\dagger\|_2^2 = 0$. On the other hand, we can obtain that

$$\|Tx^*\|_{\ell^1} \leq \liminf_\delta \left\{ \|Tx_\alpha^\delta\|_{\ell^1} + \frac{\delta^2}{2\alpha(\delta)} \right\} = \|Tx^\dagger\|_{\ell^1}.$$

Condition 2 gives that $x^* = x^\dagger$. From the inequality above, we see that $\lim_\delta \|Tx_\alpha^\delta\|_{\ell^1} = \|Tx^\dagger\|_{\ell^1}$. By Lemma 3, we have $\|T(x_\alpha^\delta - x^*)\|_{\ell^1} \rightarrow 0$ and $\|K(x_\alpha^\delta - x^*)\|_2 \rightarrow 0$. Consequently, from Condition 1, it holds that

$$\begin{aligned} \lim_\delta \|x_\alpha^\delta - x^*\|_{\ell^2} &\leq \lim_\delta m \|T(x_\alpha^\delta - x^*)\|_{\ell^2} + c \|K(x_\alpha^\delta - x^*)\|_2 \\ &\leq \lim_\delta m \|T(x_\alpha^\delta - x^*)\|_{\ell^1} + c \|K(x_\alpha^\delta - x^*)\|_2 = 0. \end{aligned}$$

□

3.3. Convergence Rate

This subsection concerns the convergence rate under different parameter choice rules (a priori and a posteriori). First, we discuss the a priori one. Like the classical Tikhonov regularization method [19,35,36], we introduce a source condition.

Condition 3. Let x^\dagger satisfy the source condition

$$\exists w : K^*w \in T^*\partial\|Tx^\dagger\|_{\ell^1}.$$

Theorem 4. If x^\dagger satisfies the source condition, it holds that

$$\|Kx_\alpha^\delta - y^\delta\|_2 \leq 2\alpha\|w\|_2 + \delta.$$

If K is injective, there exists $\gamma > 0$ such that

$$\|x_\alpha^\delta - x^\dagger\|_{\ell^2} \leq 2\gamma\alpha\|w\|_2 + 2\gamma\delta.$$

Proof. The definition of x_α^δ gives that

$$\frac{1}{2}\|Kx_\alpha^\delta - y^\delta\|_2^2 + \alpha\|Tx_\alpha^\delta\|_{\ell^1} \leq \frac{1}{2}\|Kx^\dagger - y^\delta\|_2^2 + \alpha\|Tx^\dagger\|_{\ell^1}.$$

Using the notation $C(x) = \|Tx\|_{\ell^1}$, we obtain that

$$\frac{1}{2}\|Kx_\alpha^\delta - y^\delta\|_2^2 + \alpha C(x_\alpha^\delta) \leq \frac{1}{2}\|Kx^\dagger - y^\delta\|_2^2 + \alpha C(x^\dagger).$$

For any $v \in \partial C(x^\dagger)$, the convexity of C indicates $C(x_\alpha^\delta) \geq C(x^\dagger) + \langle v, x_\alpha^\delta - x^\dagger \rangle$. Then, we have that

$$\begin{aligned} \frac{1}{2}\|Kx_\alpha^\delta - y^\delta\|_2^2 + \alpha C(x^\dagger) &+ \alpha \langle v, x_\alpha^\delta - x^\dagger \rangle \\ &\leq \frac{1}{2}\|Kx_\alpha^\delta - y^\delta\|_2^2 + \alpha C(x_\alpha^\delta) \\ &\leq \frac{1}{2}\|Kx^\dagger - y^\delta\|_2^2 + \alpha C(x^\dagger). \end{aligned}$$

Choose $v = K^*w$ in the source condition; after simplification, we derive that

$$\frac{1}{2}\|Kx_\alpha^\delta - y^\delta\|_2^2 + \alpha \langle w, Kx_\alpha^\delta - y^\delta \rangle \leq \frac{1}{2}\|Kx^\dagger - y^\delta\|_2^2 + \alpha \langle w, Kx^\dagger - y^\delta \rangle.$$

By adding both sides with $\frac{\alpha^2\|w\|_2^2}{2}$, we obtain that

$$\|Kx_\alpha^\delta - y^\delta + \alpha w\|_2 \leq \|Kx^\dagger - y^\delta + \alpha w\|_2.$$

This means

$$\|Kx_\alpha^\delta - y^\delta\|_2 \leq 2\alpha\|w\|_2 + \|Kx^\dagger - y^\delta\|_2 \leq 2\alpha\|w\|_2 + \delta.$$

If K is injective, there exists $\gamma > 0$ such that $\|x\|_{\ell^2} \leq \gamma\|Kx\|_{\ell^2}$. Then, we derive that

$$\begin{aligned} \|x_\alpha^\delta - x^\dagger\|_2 &\leq \gamma\|Kx_\alpha^\delta - y^\dagger\|_2 \\ &\leq \gamma(\|Kx_\alpha^\delta - y^\delta\|_2 + \|y^\delta - y^\dagger\|_2) \\ &\leq \gamma(2\alpha\|w\|_2 + 2\delta). \end{aligned}$$

□

Remark 3. In fact, the first result in Theorem 4 has been proved by [42] for general convex regularization. The proof here is for the completeness.

The following part investigates the a posteriori parameter choice rule. The analysis is motivated by the work in [43,44]. For simplicity of presentation, the parameter α is chosen as

$$\|Kx_\alpha^\delta - y^\delta\|_2 = \delta. \tag{5}$$

Theorem 5. Assume that α is chosen as rule (5), and x^\dagger satisfies Condition 2. It then holds that

$$\lim_{\delta \rightarrow 0} x_\alpha^\delta = x^\dagger.$$

If K is injective, there exists $\theta > 0$ such that

$$\|x_\alpha^\delta - x^*\|_{\ell^2} \leq 2\theta\delta.$$

Proof. It is trivial to prove that

$$\frac{1}{2} \|Kx_\alpha^\delta - y^\delta\|_2^2 + \alpha \|Tx_\alpha^\delta\|_{\ell^1} \leq \frac{1}{2} \|Kx^\dagger - y^\delta\|_2^2 + \alpha \|Tx^\dagger\|_{\ell^1}.$$

Lemma 2 indicates that $\{x_\alpha^\delta\}_\delta$ is bounded. Note that $\|Kx_\alpha^\delta - y^\delta\|_2 = \delta$ and $\|Kx^\dagger - y^\delta\|_2 \leq \delta$. It then follows that

$$\|Tx_\alpha^\delta\|_{\ell^1} \leq \|Tx^\dagger\|_{\ell^1}. \tag{6}$$

Then, the sequence has a sub-sequence also denoted by $\{x_\alpha^\delta\}_\delta$ converging weakly to some x^* . We can easily see that

$$\begin{aligned} \|Kx^* - y^\dagger\|_2 &\leq \liminf_{\delta \rightarrow 0} \|Kx_\alpha^\delta - y^\dagger\|_2 \\ &\leq \liminf_{\delta \rightarrow 0} (\|Kx_\alpha^\delta - y^\delta\|_2 + \|y^\delta - y^\dagger\|_2) \\ &\leq \liminf_{\delta \rightarrow 0} 2\delta = 0. \end{aligned}$$

That is actually to say that $Kx^* = y^\dagger$. Moreover, it is easy to see that $\lim_{\delta} \|Kx_\alpha^\delta - y^\dagger\|_2^2 = 0$. Using relation (6), we have that

$$\|Tx^*\|_{\ell^1} \leq \liminf_{\delta \rightarrow 0} \|Tx_\alpha^\delta\|_{\ell^1} \leq \|Tx^\dagger\|_{\ell^1}.$$

Condition 2 gives that $x^* = x^\dagger$; hence, the whole sequence converges weakly to x^\dagger and

$$\|Tx^\dagger\|_{\ell^1} \leq \liminf_{\delta \rightarrow 0} (\|Tx_\alpha^\delta\|_{\ell^1}) \leq \limsup_{\delta \rightarrow 0} (\|Tx_\alpha^\delta\|_{\ell^1}) \leq \|Tx^\dagger\|_{\ell^1}.$$

Thus, we have $\|Tx_\alpha^\delta\|_{\ell^1} \rightarrow \|Tx^\dagger\|_{\ell^1}$. From Lemma 3, we have $\|T(x_\alpha^\delta - x^*)\|_{\ell^1} \rightarrow 0$ and $\|K(x_\alpha^\delta - x^*)\|_2 \rightarrow 0$ which leads to

$$\begin{aligned} \|x_\alpha^\delta - x^*\|_{\ell^2} &\leq m \|T(x_\alpha^\delta - x^*)\|_{\ell^2} + c \|K(x_\alpha^\delta - x^*)\|_2 \\ &\leq m \|T(x_\alpha^\delta - x^*)\|_{\ell^1} + c \|K(x_\alpha^\delta - x^*)\|_2 \rightarrow 0. \end{aligned} \tag{7}$$

If K is injective, there exists $\theta > 0$ such that $\|x\|_{\ell^2} \leq \theta \|Kx\|_{\ell^2}$. Then, we derive that

$$\begin{aligned} \|x_\alpha^\delta - x^\dagger\|_2 &\leq \theta \|Kx_\alpha^\delta - y^\dagger\|_2 \\ &\leq \theta (\|Kx_\alpha^\delta - y^\delta\|_2 + \|y^\delta - y^\dagger\|_2) \leq 2\theta\delta. \end{aligned}$$

□

4. Improved Convergence Rate

In this section, we investigate the convergence rate when K may be not injective. The first part presents the analysis under the sparse assumption while the second one deals with the case when the sparsity assumption fails.

4.1. Performance under Sparsity Assumption

The analysis in this subsection assumes that Tx^\dagger is sparse. To prove the convergence rate we need the finite injectivity property [45].

Condition 4. *The operator K satisfies the uniformly finite injectivity property, i.e., for any finite subset $S \subseteq \mathbb{N}$, $K|_S$ is injective.*

Remark 4. *In the finite dimension case, if $\#supp(S)$ is small, it is easy to find that the finite injectivity property is actually the restrict isometry property [2,46].*

Let $z := Tx$ and $z^\dagger := Tx^\dagger$. Denote S as the set $S := \{i \in \mathbb{N} : |v_i| > \frac{1}{2}\}$, where $v \in \partial\|z^\dagger\|_{\ell^1}$ satisfies the source condition. Let $m = \sup_{i \notin S} \{|v_i|\}$. Due to that $v \in \ell^2$, S is finite and it contains the support of z^\dagger . Let P be the identical projection onto S and P^\perp be the one onto $\mathbb{N} \setminus S$. From Condition 4, there exists some $d > 0$ such that

$$d\|KPz\|_2 \geq \|Pz\|_{\ell^2}.$$

Lemma 4. *Assume that x^\dagger satisfies the source condition and Condition 1 holds. If $md\|K\| < 1$, there exist $c_1 > 0$ and $c_2 > 0$ such that*

$$\|Tx\|_{\ell^1} - \|Tx^\dagger\|_{\ell^1} \geq c_1\|x - x^\dagger\|_{\ell^2} - c_2\|T(x - x^\dagger)\|_2.$$

Proof. Assume the conditions in Lemma 4 are held. Then, we can obtain that

$$\begin{aligned} \|z - z^\dagger\|_{\ell^2} &\leq \|P(z - z^\dagger)\|_{\ell^2} + \|P^\perp(z - z^\dagger)\|_{\ell^2} \\ &\leq d\|KP(z - z^\dagger)\|_2 + \|P^\perp z\|_{\ell^2} \\ &\leq d\|K(z - z^\dagger)\|_2 + (1 + d\|K\|)\|P^\perp z\|_{\ell^2}. \end{aligned}$$

Hence, we derive that

$$\begin{aligned} \|K(z - z^\dagger)\|_2 &= \|KT(x - x^\dagger)\|_2 \\ &\leq \|K(x - x^\dagger)\|_2 + \|K(T - I_d)(x - x^\dagger)\|_2 \\ &\leq \|K(x - x^\dagger)\|_2 + \|K\| \cdot \|T - I_d\| \cdot \|x - x^\dagger\|_{\ell^2} \\ &\leq \|K(x - x^\dagger)\|_2 + \|K\| \cdot \|x - x^\dagger\|_{\ell^2}. \end{aligned}$$

We now turn to estimating $\|P^\perp z\|_{\ell^2}$. Let $m = \sup_{i \notin S} \{|v_i|\}$. Obviously, $m \leq \frac{1}{2}$. We then have that

$$\begin{aligned} \|P^\perp z\|_{\ell^2} &\leq \sum_{i \notin S} |z_i| \leq 2 \sum_{i \notin S} (1 - m)|z_i| \leq 2 \sum_{i \notin S} (|z_i| - v_i z_i) \\ &\leq 2 \sum_{i \notin S} (|z_i| - |z_i^\dagger| - v_i(z_i - z_i^\dagger)) \\ &\leq 2(\|z\|_{\ell^1} - \|z^\dagger\|_{\ell^1} - \langle v, z - z^\dagger \rangle) \\ &= 2(\|Tx\|_{\ell^1} - \|Tx^\dagger\|_{\ell^1} - \langle v, z - z^\dagger \rangle). \end{aligned}$$

The source condition gives that

$$\begin{aligned} -\langle v, z - z^\dagger \rangle &= -\langle v, Tx - Tx^\dagger \rangle = -\langle T^*v, x - x^\dagger \rangle \\ &= -\langle K^*w, x - x^\dagger \rangle = -\langle w, Kx - Kx^\dagger \rangle \\ &\leq \|w\|_2 \cdot \|Kx - Kx^\dagger\|_2. \end{aligned}$$

Therefore, we have that

$$\begin{aligned} \|z - z^\dagger\|_{\ell^2} &\leq d\|K\| \cdot \|x - x^\dagger\|_{\ell^2} + 2(1 + d\|K\|)(\|Tx\|_{\ell^1} - \|Tx^\dagger\|_{\ell^1}) \\ &\quad + (2\|w\|_2 + 2d\|K\|\|w\|_2 + d)\|K(x - x^\dagger)\|_2. \end{aligned}$$

From Condition 1, we have that

$$\begin{aligned} \|x - x^\dagger\|_{\ell^2} &= c\|K(x - x^\dagger)\|_2 + m\|Tx - Tx^\dagger\|_{\ell^2} \\ &\leq c\|K(x - x^\dagger)\|_2 + m\|z - z^\dagger\|_{\ell^2} \\ &\leq md\|K\| \cdot \|x - x^\dagger\|_{\ell^2} + 2(m + md\|K\|)(\|Tx\|_1 - \|Tx^\dagger\|_1) \\ &\quad + (2m\|w\|_2 + 2dm\|K\|\|w\|_2 + md + c)\|K(x - x^\dagger)\|_2. \end{aligned}$$

Note that $md\|K\| < 1$; let $q = \frac{1}{1 - md\|K\|}$; we have that

$$\begin{aligned} \|x - x^\dagger\|_{\ell^2} &\leq 2q(m + md\|K\|)(\|Tx\|_1 - \|Tx^\dagger\|_1) \\ &\quad + q(2m\|w\|_2 + 2md\|K\|\|w\|_2 + md + c)\|K(x - x^\dagger)\|_2. \end{aligned}$$

□

With the lemma above, we can obtain the following result. The proofs can be found in [44,47–49].

Theorem 6. *Let the regularization parameter be chosen a priori as $\alpha(\delta) = \mathcal{O}(\delta)$ or a posteriori as $\alpha(\delta)$ according to the strong discrepancy principle (5) Then we have the convergence rate*

$$\|x_\alpha^\delta - x^\dagger\|_{\ell^2} = \mathcal{O}(\delta).$$

4.2. Performance if Sparsity Assumption Fails

In this subsection, we focus on the case where Tx^\dagger is not sparse. As presented in the last section, lemma 4 is critical for the convergence rate analysis. In this part, a similar lemma will be proposed. Then, the convergence rate will be proved. The first lemma is motivated by [37].

Lemma 5. *For any $x \in \mathcal{H}_1$ and $n \in \mathbb{N}$, it holds that*

$$\|T(x - x^\dagger)\|_{\ell^1} - \|Tx\|_{\ell^1} + \|Tx^\dagger\|_{\ell^1} \leq 2\left(\sum_{k=n+1}^{\infty} |(Tx^\dagger)_k| + \sum_{k=1}^n |(Tx)_k - (Tx^\dagger)_k|\right).$$

Proof. Denote the projection $P_n(x) = (x_1, x_2, \dots, x_n, 0, \dots)$ for any $x \in \mathcal{H}_1$. Hence, we have

$$\|Tx\|_{\ell^1} = \|P_nTx\|_{\ell^1} + \|(I_d - P_n)Tx\|_{\ell^1}.$$

Algebra computation gives that

$$\begin{aligned} \|T(x - x^\dagger)\|_{\ell^1} &- \|Tx\|_{\ell^1} + \|Tx^\dagger\|_{\ell^1} = \|P_nT(x - x^\dagger)\|_{\ell^1} + \|(I_d - P_n)Tx^\dagger\|_{\ell^1} \\ &\quad + \|(I_d - P_n)(Tx - Tx^\dagger)\|_{\ell^1} - \|(I_d - P_n)Tx\|_{\ell^1} \\ &\quad + \|P_nTx^\dagger\|_{\ell^1} - \|P_nTx\|_{\ell^1}. \end{aligned}$$

Note that

$$\|(I_d - P_n)(Tx - Tx^\dagger)\|_{\ell^1} \leq \|(I_d - P_n)Tx\|_{\ell^1} + \|(I_d - P_n)Tx^\dagger\|_{\ell^1}$$

and

$$\|P_nTx^\dagger\|_{\ell^1} \leq \|P_nT(x - x^\dagger)\|_{\ell^1} + \|P_nTx\|_{\ell^1}.$$

Combining the equations above, we obtain

$$\|T(x - x^\dagger)\|_1 - \|Tx\|_{\ell^1} + \|Tx^\dagger\|_{\ell^1} \leq 2\|P_nT(x - x^\dagger)\|_{\ell^1} + 2\|(I_d - P_n)Tx^\dagger\|_{\ell^1}.$$

□

Condition 5. For all $k \in \mathbb{N}$ there exists $f_k \in \mathcal{H}_2$ such that $T^*e_k = K^*f_k$ and $\lim_{k \rightarrow \infty} \|f_k\|_2 \rightarrow +\infty$.

Lemma 6. Let $\varphi(t) := 2 \inf_n \{ \sum_{k=n+1}^\infty |(Tx^\dagger)_k| + t \sum_{k=1}^n \|f_k\| \}$; $\varphi(t)$ is concave index function. Assume that x^\dagger satisfies the source condition and Conditions 1 and 5 hold. If $c\|K\| < 1$, it holds that

$$\|x - x^\dagger\|_{\ell^2} \leq c_1\|Tx\|_{\ell^1} - c_2\|Tx^\dagger\|_{\ell^1} + c_3\varphi(\|K(x - x^\dagger)\|_2)$$

for some positive c_1, c_2, c_3 .

Proof. φ is concave and upper semi-continuous since it is an infimum of affine functions. For any $t \geq 0$, φ is finite and continuous. Note that $\varphi(0) = 0$; the upper semi-continuity at $t = 0$ gives the continuity of φ at $t = 0$. We turn to the strict monotonicity of φ . Condition 5 means the infimum of $\varphi(t)$ is attained at some $n \in \mathbb{N}$. Considering $0 < t_1 < t_2 < +\infty$, we have

$$\begin{aligned} \varphi(t_1) &= 2\left(\sum_{k=n_1+1}^\infty |(Tx^\dagger)_k| + t_1 \sum_{k=1}^{n_1} \|f_k\|_2\right) \\ &\leq 2\left(\sum_{k=n_2+1}^\infty |(Tx^\dagger)_k| + t_1 \sum_{k=1}^{n_2} \|f_k\|_2\right) \\ &< 2\left(\sum_{k=n_2+1}^\infty |(Tx^\dagger)_k| + t_2 \sum_{k=1}^{n_2} \|f_k\|_2\right) = \varphi(t_2). \end{aligned}$$

From Condition 5, we have that

$$\begin{aligned} \sum_{k=1}^n |(Tx - Tx^\dagger)_k| &= \sum_{k=1}^n |\langle Tx - Tx^\dagger, e_k \rangle| \\ &= \sum_{k=1}^n |\langle x - x^\dagger, T^*e_k \rangle| \\ &= \sum_{k=1}^n |\langle x - x^\dagger, K^*f_k \rangle| \\ &\leq \|K(x - x^\dagger)\|_2 \sum_{k=1}^n \|f_k\|_2. \end{aligned} \tag{8}$$

Therefore, we obtain that

$$\|T(x - x^\dagger)\|_{\ell^2} \leq \|Tx\|_{\ell^1} - \|Tx^\dagger\|_{\ell^1} + 2\varphi(\|K(x - x^\dagger)\|_2).$$

From Condition 1, we have that

$$\begin{aligned}\|x - x^\dagger\|_{\ell^2} &\leq c\|K(x - x^\dagger)\|_{\ell^2} + m\|T(x - x^\dagger)\|_2 \\ &\leq c\|K\|\|x - x^\dagger\|_{\ell^2} + m\|Tx\|_{\ell^1} - m\|Tx^\dagger\|_{\ell^1} \\ &\quad + 2m\varphi(\|K(x - x^\dagger)\|_2).\end{aligned}$$

Let $q = \frac{1}{1-c\|K\|}$, we obtain that

$$\|x - x^\dagger\|_{\ell^2} \leq qm\|Tx\|_{\ell^1} - qm\|Tx^\dagger\|_{\ell^1} + 2qm\varphi(\|K(x - x^\dagger)\|_2).$$

□

Theorem 7. Let the regularization parameter be chosen a priori as $\alpha(\delta) = \mathcal{O}(\frac{\delta^2}{\varphi(\delta)})$ or a posteriori as $\alpha(\delta)$ according to the strong discrepancy principle (5). Then we have the convergence rate

$$\|x_\alpha^\delta - x^\dagger\|_{\ell^2} = \mathcal{O}(\varphi(\delta)).$$

5. Conclusions

In this paper, we study some problems in total variation type regularization. While owning a familiar form as the sparse regularization, the TV type is hard to investigate for the ill condition of T . A group of regularization conditions has been given in this paper. Under these conditions, we study several theoretical properties such as stability, consistency and convergence rates of the minimizer of the TV type regularization. These analyses are deepened for the convergence rate under the assumption of sparsity. In the non-sparse case, we also present a conservative result based on some recent works. Now, the regularizers learned from the data are all the rage in research. So, in future work, we will make the error estimations for this type of regularization problem.

Author Contributions: Conceptualization, K.L. and Z.Y.; methodology, C.H.; validation, K.L., C.H. and Z.Y.; formal analysis, K.L.; writing—original draft preparation, K.L. and Z.Y.; writing—review and editing, K.L. and Z.Y.; supervision, C.H.; project administration, Z.Y.; funding acquisition, K.L. and C.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China under Grant 2020YFA0709803, 173 Program under Grant 2020-JCJQ-ZD-029, the Science Challenge Project under Grant TZ2016002, and Dongguan Science and Technology of Social Development Program under Grant 2020507140146.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Donoho, D.L. Compressed sensing. *IEEE T. Inform. Theory* **2006**, *52*, 1289–1306. [[CrossRef](#)]
2. Candes, E.J.; Tao, T. Decoding by linear programming. *IEEE T. Inform. Theory* **2005**, *51*, 4203–4215. [[CrossRef](#)]
3. Lustig, M.; Donoho, D.L.; Santos, J.M.; Pauly, J.M. Compressed Sensing MRI. *IEEE Signal Proc. Mag.* **2008**, *25*, 72–82. [[CrossRef](#)]
4. Haupt, J.; Bajwa, W.U.; Rabbat, M.; Nowak, R. Compressed Sensing for Networked Data. *IEEE Signal Proc. Mag.* **2008**, *25*, 92–101. [[CrossRef](#)]
5. Yin, W.; Osher, S.; Goldfarb, D.; Darbon, J. Bregman iterative algorithms for ℓ_1 -minimization with applications to compressed sensing. *SIAM J. Imaging Sci.* **2008**, *1*, 143–168. [[CrossRef](#)]
6. Cai, J.F.; Osher, S.; Shen, Z. Split Bregman methods and frame based image restoration. *Multiscale Model. Sim.* **2009**, *8*, 337–369. [[CrossRef](#)]
7. Van den Berg, E.; Friedlander, M.P. Probing the Pareto Frontier for Basis Pursuit Solutions. *SIAM J. Sci. Comput.* **2008**, *31*, 890–912. [[CrossRef](#)]
8. Chen, S.S.; Donoho, D.L.; Saunders, M.A. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.* **1998**, *20*, 33–61. [[CrossRef](#)]
9. Candès, E.J.; Romberg, J.; Tao, T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory* **2006**, *52*, 489–509. [[CrossRef](#)]
10. Candès, E.J. The restricted isometry property and its implications for compressed sensing. *CR Math.* **2008**, *346*, 589–592.

11. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. B* **1996**, *58*, 267–288. [[CrossRef](#)]
12. Daubechies, I.; Defrise, M.; Mol, C.D. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Commun. Pure Appl. Math.* **2004**, *57*, 1413–1457. [[CrossRef](#)]
13. Daubechies, I.; Devore, R.; Fornasier, M.; Güntürk, C.S. Iteratively reweighted least squares minimization for sparse recovery. *Commun. Pure Appl. Math.* **2010**, *38*, 1–38. [[CrossRef](#)]
14. Bot, R.I.; Hofmann, B. The impact of a curious type of smoothness conditions on convergence rates in ℓ^1 -regularization. *Eur. J. Math. Comput. Appl.* **2013**, *1*, 29–40.
15. Jin, B.; Maass, P. Sparsity regularization for parameter identification problems. *Inverse Probl.* **2012**, *28*, 123001. [[CrossRef](#)]
16. Grasmair, M.; Haltmeier, M.; Scherzer, O. Sparse regularization with ℓ^q penalty term. *Inverse Probl.* **2008**, *24*, 055020. [[CrossRef](#)]
17. Hans, E.; Raasch, T. Global convergence of damped semismooth Newton methods for 1 Tikhonov regularization. *Inverse Probl.* **2015**, *31*, 025005. [[CrossRef](#)]
18. Lorenz, D.A.; Schiffler, S.; Trede, D. Beyond convergence rates: Exact recovery with the Tikhonov regularization with sparsity constraints. *Inverse Probl.* **2011**, *27*, 085009. [[CrossRef](#)]
19. Lorenz, D.A. Convergence rates and source conditions for Tikhonov regularization with sparsity constraints. *J. Inverse III-Pose P* **2008**, *16*, 463–478. [[CrossRef](#)]
20. Rubinov, A.M.; Yang, X.Q.; Bagirov, A.M. Penalty functions with a small penalty parameter. *Optim. Methods Softw.* **2002**, *17*, 931–964. [[CrossRef](#)]
21. Cai, J.F.; Dong, B.; Osher, S.; Shen, A.Z. Image Restoration: Total Variation, Wavelet Frames and Beyond. *J. Am. Math. Soc.* **2012**, *25*, 1033–1089. [[CrossRef](#)]
22. Rudin, L.I.; Osher, S.; Fatemi, E. Nonlinear total variation based noise removal algorithms. *Phys. D* **1992**, *60*, 259–268. [[CrossRef](#)]
23. Jun, L.; Huang, T.Z.; Gang, L.; Wang, S.; Lv, X.-G. Total variation with overlapping group sparsity for speckle noise reduction. *Neurocomputing* **2019**, *216*, 502–513.
24. Van den Berg, P.M.; Kleinman, R.E. A total variation enhanced modified gradient algorithm for profile reconstruction. *Inverse Probl.* **1995**, *11*, L5. [[CrossRef](#)]
25. Clason, C.; Jin, B.; Kunisch, K. A Duality-Based Splitting Method for ℓ^1 -TV Image Restoration with Automatic Regularization Parameter Choice. *SIAM J. Sci. Comput.* **2010**, *32*, 1484–1505. [[CrossRef](#)]
26. Cai, J.F.; Xu, W.Y. Guarantees of total variation minimization for signal recovery. *Inform. Infer.* **2015**, *4*, 328–353. [[CrossRef](#)]
27. Needell, D.; Ward, R. Stable Image Reconstruction Using Total Variation Minimization. *SIAM J. Imaging Sci.* **2012**, *6*, 1035–1058. [[CrossRef](#)]
28. Sun, T.; Yin, P.; Cheng, L.; Jiang, H. Alternating direction method of multipliers with difference of convex functions. *Adv. Comput. Math.* **2018**, *44*, 723–744. [[CrossRef](#)]
29. Sun, T.; Jiang, H.; Cheng, L.; Zhu, W. Iteratively Linearized Reweighted Alternating Direction Method of Multipliers for a Class of Nonconvex Problems. *IEEE Trans. Signal Proc.* **2018**, *66*, 5380–5391. [[CrossRef](#)]
30. Leonov, A.S. On the total-variation convergence of regularizing algorithms for ill-posed problems. *Comput. Math. Math. Phys.* **2007**, *47*, 732–747. [[CrossRef](#)]
31. Tikhonov, A.; Leonov, A.; Yagola, A. *Nonlinear Ill-Posed Problems*; De Gruyter: Berlin, Germany, 2011; pp. 505–512. [[CrossRef](#)]
32. Hào, D.N.; Quyen, T.N.T. Convergence rates for total variation regularization of coefficient identification problems in elliptic equations I. *Inverse Probl.* **2011**, *27*, 075008. [[CrossRef](#)]
33. Hào, D.N.; Quyen, T.N.T. Convergence rates for total variation regularization of coefficient identification problems in elliptic equations II. *J. Math. Anal. Appl.* **2012**, *388*, 593–616. [[CrossRef](#)]
34. Ciegis, R.; Sev, A.J. Nonlinear Diffusion Problems In Image Smoothing. *Math. Model. Anal.* **2005**, *1*, 381–388.
35. Grasmair, M.; Scherzer, O.; Haltmeier, M. Necessary and sufficient conditions for linear convergence of ℓ^1 -regularization. *Commun. Pure Appl. Math.* **2011**, *64*, 161–182. [[CrossRef](#)]
36. Burger, M.; Osher, S. Convergence rates of convex variational regularization. *Inverse Probl.* **2004**, *20*, 1411. [[CrossRef](#)]
37. Burger, M.; Flemming, J.; Hofmann, B. Convergence rates in ℓ^1 -regularization if the sparsity assumption fails. *Inverse Probl.* **2013**, *29*, 025013. [[CrossRef](#)]
38. Flemming, J.; Hofmann, B.; Veselić, I. A unified approach to convergence rates for ℓ^1 -regularization and lacking sparsity. *J. Inverse III-Pose P* **2015**, *24*, 139–148. [[CrossRef](#)]
39. Anzengruber, S.W.; Hofmann, B.; Mathé, P. Regularization properties of the sequential discrepancy principle for Tikhonov regularization in Banach spaces. *Appl. Anal.* **2014**, *93*, 1382–1400. [[CrossRef](#)]
40. Luo, Z.Q.; Tseng, P. On the convergence of the coordinate descent method for convex differentiable minimization. *J. Optim. Theory Appl.* **1992**, *72*, 7–35. [[CrossRef](#)]
41. Luo, Z.Q.; Tseng, P. Error bound and convergence analysis of matrix splitting algorithms for the affine variational inequality problem. *SIAM J. Optim.* **1992**, *2*, 43–54. [[CrossRef](#)]
42. Jin, B.; Lorenz, D.A. Heuristic parameter-choice rules for convex variational regularization based on error estimates. *SIAM J. Numer. Anal.* **2010**, *48*, 1208–1229. [[CrossRef](#)]
43. Jin, B.; Zou, J. Iterative parameter choice by discrepancy principle. *IMA J. Numer. Anal.* **2012**, *32*, 1714–1732. [[CrossRef](#)]
44. Jin, B.; Lorenz, D.A.; Schiffler, S. Elastic-Net Regularization: Error estimates and Active Set Methods. *Inverse Probl.* **2009**, *25*, 115022. [[CrossRef](#)]

-
45. Bredies, K.; Lorenz, D.A. Linear convergence of iterative soft-thresholding. *J. Fourier Anal. Appl.* **2008**, *14*, 813–837. [[CrossRef](#)]
 46. Foucart, S.; Rauhut, H. *A Mathematical Introduction to Compressive Sensing*; Springer: New York, NY, USA, 2013.
 47. Anzengruber, S.W.; Ramlau, R. Convergence rates for Morozov’s discrepancy principle using variational inequalities. *Inverse Probl.* **2011**, *27*, 105007–105024. [[CrossRef](#)]
 48. Flemming, J. *Generalized Tikhonov Regularization and Modern Convergence Rate Theory in Banach Spaces*; Shaker Verlag GmbH: Düren, Germany, 2012.
 49. Hofmann, B.; Mathé, P. Parameter choice in Banach space regularization under variational inequalities. *Inverse Probl.* **2012**, *6*, 1035–1058. [[CrossRef](#)]