The springback penalty for robust signal recovery

Congpei An [a,*], Hao-Ning Wu [b], Xiaoming Yuan [b]

[a] *School of Mathematics, Southwestern University of Finance and Economics, Chengdu, China*
[b] *Department of Mathematics, The University of Hong Kong, Hong Kong, China*

A R T I C L E   I N F O

A B S T R A C T

We propose a new penalty, the springback penalty, for constructing models to recover an unknown signal from incomplete and inaccurate measurements. Mathematically, the springback penalty is a weakly convex function. It bears various theoretical and computational advantages of both the benchmark convex ℓ_1 penalty and many of its non-convex surrogates that have been well studied in the literature. We establish the exact and stable recovery theory for the recovery model using the springback penalty for both sparse and nearly sparse signals, respectively, and derive an easily implementable difference-of-convex algorithm. In particular, we show its theoretical superiority to some existing models with a sharper recovery bound for some scenarios where the level of measurement noise is large or the amount of measurements is limited. We also demonstrate its numerical robustness regardless of the varying coherence of the sensing matrix. The springback penalty is particularly favorable for the scenario where the incomplete and inaccurate measurements are collected by coherence-hidden or -static sensing hardware due to its theoretical guarantee of recovery with severe measurements, computational tractability, and numerical robustness for ill-conditioned sensing matrices.

© 2022 Elsevier Inc. All rights reserved.

1. Introduction

Signal recovery aims at recovering an unknown signal from its measurements, which are often incomplete and inaccurate due to technical, economical, or physical restrictions. Mathematically, a signal recovery problem can be expressed as estimating an unknown $\bar{x} \in \mathbb{R}^n$ from an underdetermined linear system

$$b = A\bar{x} + e, \tag{1.1}$$

where $A \in \mathbb{R}^{m \times n}$ is a full row-rank sensing matrix such as a projection or transformation matrix (see, e.g., [3,6,7]), $b \in \mathbb{R}^m \setminus \{0\}$ is a vector of measurements, $e \in \mathbb{R}^m$ is some unknown but bounded noise perturbation

in $\mathcal{B}(\tau) := \{e \in \mathbb{R}^m : \|e\|_2 \leq \tau\}$, and the number m of measurements is considerably smaller than the size n of the signal $\bar{x}$. The set $\mathcal{B}(\tau)$ encodes both the cases of noise-free ($\tau = 0$) and noisy ($\tau > 0$) measurements.

Physically, a signal of interest, or its coefficients under certain transformation, is often sparse (see, e.g., [3]). Hence, it is natural to seek a sparse solution to the underdetermined linear system (1.1), though it has infinitely many solutions. We say that $x \in \mathbb{R}^n$ is *s-sparse* if $\|x\|_0 \leq s$, where $\|x\|_0$ counts the number of nonzero entries of x. To find the sparsest solution to (1.1), one may consider solving the following minimization problem:

$$\min_{x \in \mathbb{R}^n} \|x\|_0 \quad \text{s.t.} \quad Ax - b \in \mathcal{B}(\tau), \tag{1.2}$$

in which $\|x\|_0$ serves as a penalty term of the sparsity, and it is referred to as the ℓ_0 penalty for convenience. Due to the discrete and discontinuous nature of the ℓ_0 penalty, the model (1.2) is NP-hard [3]. This means the model (1.2) is computationally intractable, and this difficulty has inspired many alternatives to the ℓ_0 penalty in the literature. A fundamental proxy of the model (1.2) is the basis pursuit (BP) problem proposed in [12]:

$$\min_{x \in \mathbb{R}^n} \|x\|_1 \quad \text{s.t.} \quad Ax - b \in \mathcal{B}(\tau). \tag{1.3}$$

In this convex model, $\|x\|_1 := \sum_{i=1}^{n} |x_i|$ and it is called the ℓ_1 penalty hereafter. Recall that $\|x\|_1$ is the convex envelope of $\|x\|_0$ (see, e.g., [35]), and it induces sparsity most efficiently among all convex penalties (see [3]). The BP problem (1.3) has been intensively studied in voluminous papers since the seminal works [5,6,13], in which various conditions have been comprehensively explored for the exact recovery via the convex model (1.3).

The BP problem (1.3) is fundamental for signal recovery, but its solution may be over-penalized because the ℓ_1 penalty tends to underestimate high-amplitude components of the solution, as analyzed in [15]. Hence, it is reasonable to consider non-convex alternatives to the ℓ_1 penalty and upgrade the model (1.3) to achieve a more accurate recovery. In the literature, some non-convex penalties have been well studied, such as the smoothly clipped absolute deviation (SCAD) [15], the capped ℓ_1 penalty [49], the transformed ℓ_1 penalty [29,48], and the ℓ_p penalty with $0 < p < 1$ [9,10,27]. Besides, one particular penalty is the minimax concave penalty (MCP) proposed in [46], and it has been widely shown to be effective in reducing the bias from the ℓ_1 penalty [46]. Moreover, the so-called ℓ_{1-2} penalty has been studied in the literature, e.g. [14,44,45], to mention a few. Some of these penalties will be summarized in Section 2. In a nutshell, convex penalties are more tractable in the senses of theoretical analysis and numerical computation, while they are less effective for achieving the desired sparsity (i.e., the approximation to the ℓ_0 penalty is less accurate). Non-convex penalties are generally the opposite.

Considering the pros and cons of various penalties, we are motivated to find a weakly convex penalty that can keep some favorable features from both the ℓ_1 penalty and its non-convex alternatives, and the resulting model for signal recovery is preferable in the senses of both theoretical analysis and numerical computation. More precisely, we propose the *springback* penalty

$$\mathcal{R}_\alpha^{\text{SPB}}(x) := \|x\|_1 - \frac{\alpha}{2}\|x\|_2^2, \tag{1.4}$$

where $\alpha > 0$ is a model parameter, and it should be chosen meticulously. We will show later that a larger α implies a tighter stable recovery bound. On the other hand, a too large α may lead to negative values of $\mathcal{R}_\alpha^{\text{SPB}}(x)$. Thus, a reasonable upper bound on α should be considered to ensure the well-definedness of the springback penalty (1.4). In the following, we will see that if the matrix A is well-conditioned (e.g., when A is drawn from a Gaussian matrix ensemble), then the requirement on α is quite loose; while if A is ill-conditioned (e.g., A is drawn from an oversampled partial DCT matrix ensemble), then generally

the upper bound on α should be better discerned for the sake of designing an algorithm with theoretically provable convergence. We refer to Theorem 3.1, Theorem 4.1, Section 5.2, and Section 6.2 for more detailed discussions on the determination of α for the springback penalty (1.4) theoretically and numerically. With the springback penalty (1.4), we propose the following model for signal recovery:

$$\min_{x \in \mathbb{R}^n} \ \mathcal{R}_{\alpha}^{\mathrm{SPB}}(x) \quad \text{s.t.} \quad Ax - b \in \mathcal{B}(\tau). \tag{1.5}$$

Mathematically, the springback penalty (1.4) is a weakly convex function, and thus the *springback-penalized model* (1.5) can be intuitively regarded as an "average" of the convex BP model (1.3) and the mentioned non-convex surrogates. Recall that a function $f : \mathbb{R}^n \to \mathbb{R}$ is α-*weakly convex* if $x \mapsto f(x) + \frac{\alpha}{2}\|x\|_2^2$ is convex. One advantage of the model (1.5) is that various results developed in the literature on weakly convex optimization problems (e.g., [21,31]) can be used for both theoretical analysis and algorithmic design. Indeed, the weak convexity of the springback penalty (1.4) enables us to derive sharper recovery results with fewer measurements and to design some efficient algorithms easily.

The rest of this paper is organized as follows. In the next section, we summarize some preliminaries for further analysis. In Sections 3 and 4, we establish the exact and stable recovery theory of the springback-penalized model (1.5) for sparse and nearly sparse signals, respectively. We also theoretically compare the springback penalty (1.4) with some other penalties in these two sections. In Section 5, we design a difference-of-convex algorithm (DCA) for the springback-penalized model (1.5) and study its convergence. Some numerical results are reported in Section 6 to verify our theoretical assertions, and some conclusions are drawn in Section 7.

2. Preliminaries

In this section, we summarize some preliminaries that will be used for further analysis.

2.1. Notations

For any $x, y \in \mathbb{R}^n$, let $\langle x, y \rangle = x^\mathrm{T} y$ be their inner product, and let $\mathrm{supp}(x) := \{1 \leq i \leq n : x_i \neq 0\}$ be the support of x. Let I be an identity matrix whose dimension is clear in accordance with the context. Let $\Lambda \subseteq \{1, 2, \ldots, n\}$ (or Λ with some super/subscripts) be an index set, and $|\Lambda|$ the cardinality of Λ. For $x \in \mathbb{R}^n$ and $A \in \mathbb{R}^{m \times n}$, let $x_\Lambda \in \mathbb{R}^n$ be the vector with the same entries as x on indices Λ and zero entries on indices Λ^c, and let $A_\Lambda \in \mathbb{R}^{m \times |\Lambda|}$ be the submatrix of A with column indices Λ. For $x \in \mathbb{R}$, $\mathrm{sgn}(x)$ is the sign function of x. For a convex function f, $\partial(f(x))$ denotes the subdifferential of f at x.

2.2. A glance at various penalties

In the literature, there are a variety of convex and non-convex penalties. Below we list six of the most important ones, with $x \in \mathbb{R}^n$.

⬦ The ℓ_1 penalty [3,12]:

$$\mathcal{R}^{\ell_1}(x) := \|x\|_1 = \sum_{i=1}^{n} |x_i|.$$

⬦ The elastic net penalty [50]:

$$\mathcal{R}^{\mathrm{EL}}(x) := \|x\|_1 + \frac{\alpha}{2}\|x\|_2^2 = \sum_{i=1}^{n} |x_i| + \frac{\alpha}{2}\sum_{i=1}^{n} |x_i|^2.$$

◇ The ℓ_p penalty with parameter $0 < p < 1$ [9,10]:

$$\mathcal{R}^{\ell_p}(x) := \|x\|_p^p = \sum_{i=1}^{n} |x_i|^p.$$

◇ The transformed ℓ_1 (TL1) with parameter $\beta > 0$ [29,48]:

$$\mathcal{R}_\beta^{\mathrm{TL1}}(x) := \sum_{i=1}^{n} \frac{(\beta + 1)|x_i|}{\beta + |x_i|}.$$

◇ The minimax concave penalty (MCP) with parameter $\mu > 0$ [46]:

$$\mathcal{R}_\mu^{\mathrm{MCP}}(x) := \sum_{i=1}^{n} \phi_\mu^{\mathrm{MCP}}(x_i), \tag{2.1}$$

where

$$\phi_\mu^{\mathrm{MCP}}(x_i) = \begin{cases} |x_i| - x_i^2/(2\mu), & |x_i| \le \mu, \\ \mu/2, & |x_i| \ge \mu. \end{cases}$$

◇ The ℓ_{1-2} penalty [14,45]:

$$\mathcal{R}^{\ell_{1-2}}(x) := \|x\|_1 - \|x\|_2 = \sum_{i=1}^{n} |x_i| - \sqrt{\sum_{i=1}^{n} |x_i|^2}.$$

Note that the ℓ_1 penalty is convex, the elastic net penalty is strongly convex, and the others are non-convex.

2.3. Relationship among various penalties

For any nonzero vector $x \in \mathbb{R}^n$ and $\alpha > 0$, the springback penalty $\mathcal{R}_\alpha^{\mathrm{SPB}}(x) \to \mathcal{R}^{\ell_1}(x)$ as $\alpha \to 0$. Besides, $\mathcal{R}_\alpha^{\mathrm{SPB}}(x)$ is reduced to the MCP in [46] within the ℓ_∞-ball $\{x \in \mathbb{R}^n : \|x\|_\infty \le \mu\}$ if $\alpha = 1/\mu$. The springback penalty appears to be a resemblance to the ℓ_{1-2} penalty, but their difference is many-sided. For instance, the gradient of $\|x\|_2$ is not defined at the origin.

Fig. 1 displays some scalar (one-dimensional) penalties, including the ℓ_1 penalty, the $\ell_{0.5}$ penalty, the transformed ℓ_1 penalty with $\beta = 1$, the MCP with $\mu = 0.75$, and the springback penalty with $\alpha = 1/\mu$ and $\alpha = 0.15$. The ℓ_{1-2} penalty is not plotted, as it is none other than zero in the one-dimensional case. To give a better visual comparison, we scale them to attain the point $(1, 1)$. It is shown in Fig. 1 that the springback penalty is close to the ℓ_1 penalty when $\alpha = 0.15$. The springback penalty with $\alpha = 1/\mu$ coincides with the MCP for $|x| \le \mu$ if we do not scale them. The behavior of the springback penalty for $|x| > \mu$ attracts our interest because it turns around and heads towards the x-axis. According to Fig. 1, this behavior is clearer in terms of the thresholding operator corresponding to the proximal mapping of the springback penalty, whose mathematical descriptions are given in Section 2.4.

As mentioned, the proposed springback penalty (1.4) balances the approximation quality of the ℓ_0 penalty and the tractability in analysis and computation, and it is in between the convex and non-convex penalties. More specifically, it is in between the ℓ_1 penalty and the MCP. For any $x \in \mathbb{R}^n$, we can always find a parameter μ for the MCP such that $\|x\|_\infty \le \mu$ with a resulting penalty in the form of $\|x\|_1 - \|x\|_2^2/(2\mu)$. This penalty inherits the approximation quality of the ℓ_0 penalty from the MCP and the analytical and computational advantages of the ℓ_1 penalty. Inasmuch as this penalty, we consider the more general penalty (1.4) in which $1/\mu$ is replaced by a more flexible parameter $\alpha > 0$.
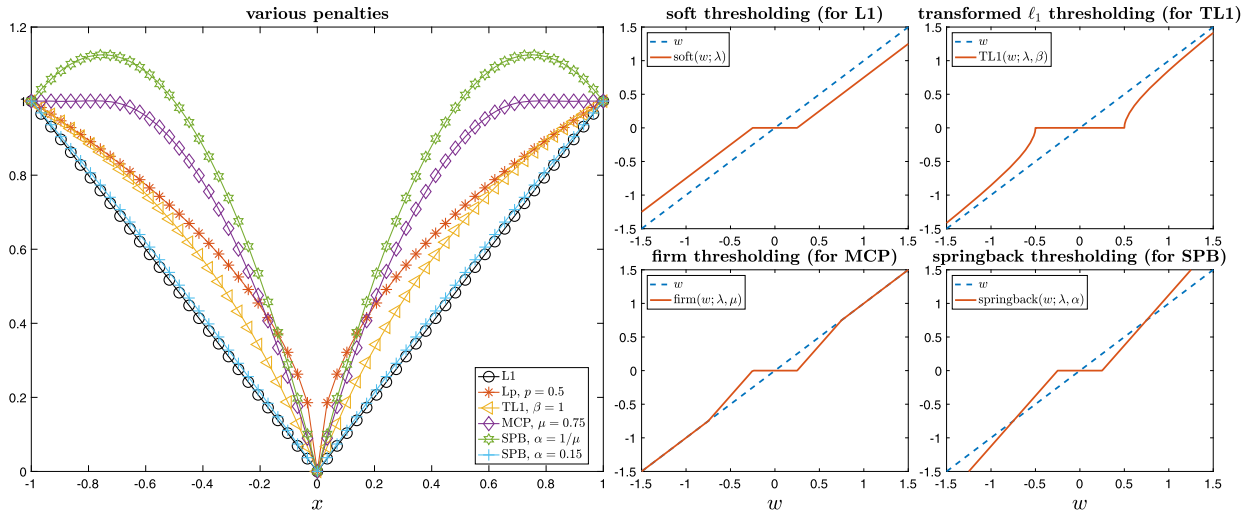
Fig. 1. Scalar penalties and corresponding thresholding operators (for representing proximal mappings with $\lambda = 0.25$): the ℓ_1 penalty and the soft thresholding operator; the ℓ_p penalty, whose proximal mapping has no closed-form expressions (hence no thresholding operator plotted); the transformed ℓ_1 penalty with $\beta = 1$, whose proximal mapping can be expressed explicitly by a thresholding operator given in [47]; the MCP with $\mu = 0.75$ and the firm thresholding operator; and two springback penalties with $\alpha = 1/\mu$ and $\alpha = 0.15$, and the springback thresholding operator.

2.4. Proximal mappings and thresholding operators

For a function $\mathcal{R} : \mathbb{R}^n \to \mathbb{R}$, as defined in [32], the proximal mapping of $\mathcal{R}$ is defined as

$$\text{prox}_\lambda \left[\mathcal{R} \right](x) := \arg \min_{y \in \mathbb{R}^n} \left\{ \lambda \mathcal{R}(y) + \frac{1}{2} \| y - x \|_2^2 \right\}, \tag{2.2}$$

where $\lambda > 0$ is a regularization parameter. In (2.2), we slightly abuse the notation "=". This mapping takes a vector $x \in \mathbb{R}^n$ and maps it into a subset of $\mathbb{R}^n$, which might be empty, a singleton, or a set with multiple vectors; and the image of y under this mapping is a singleton if the function $\mathcal{R}$ is proper closed and convex [1]. For a given optimization model, if the proximal mapping of its objective function has a closed-form expression, then usually it is important and necessary to consider how to take advantage of this feature for algorithmic design.

When the proximal mapping of a penalty can be represented explicitly, the closed-form representation is often called a *thresholding operator* or a shrinkage operator in the literature. For example, as analyzed in [47], with the *soft thresholding operator*

$$\text{soft}(w; \lambda) = \text{sgn}(w) \max\{ |w| - \lambda, 0 \},$$

which has been widely used in various areas such as compressed sensing and image processing, the proximal mapping (2.2) of the ℓ_1 penalty can be expressed explicitly by

$$\left[\text{prox}_\lambda \left[\mathcal{R}^{\ell_1} \right](x) \right]_i = \text{soft}(x_i; \lambda), \quad i = 1, \ldots, n.$$

The proximal mapping of a non-convex penalty, in general, does not have a closed-form expression; such cases include the ℓ_{1-2} penalty and the ℓ_p penalty with $0 < p < 1$. However, there are some particular non-convex penalties whose proximal mappings can still be represented explicitly. For instance, the transformed ℓ_1 penalty [47] and the MCP [46]. In particular, with the following *firm thresholding operator*

$$\text{firm}(w; \lambda, \mu) = \begin{cases} 0, & |w| \leq \lambda, \\ \text{sgn}(w)\frac{\mu(|w|-\lambda)}{\mu-\lambda}, & \lambda \leq |w| \leq \mu, \\ w, & |w| \geq \mu, \end{cases}$$

which was first proposed in [19], it was further studied in [46] that the proximal mapping (2.2) of the MCP can be expressed explicitly by a firm thresholding operator for the case of orthonormal designs. More specifically, the proximal mapping (2.2) of the MCP is

$$\left[\text{prox}_\lambda\left[\mathcal{R}_\mu^{\text{MCP}}\right](x)\right]_i = \text{firm}(x_i; \lambda, \mu), \quad i = 1, \dots, n.$$

Below, we show that for the springback penalty (1.4) with a well chosen α, its proximal mapping can also be expressed explicitly.

Definition 2.1. The *springback thresholding operator* is defined as

$$\text{springback}(w; \lambda, \alpha) = \begin{cases} 0, & |w| \leq \lambda, \\ \text{sgn}(w)\frac{|w|-\lambda}{1-\lambda\alpha}, & |w| > \lambda. \end{cases} \tag{2.3}$$

Proposition 2.1. *If* $1 - \lambda\alpha > 0$*, then the proximal mapping of the springback penalty* (1.4) *can be represented explicitly as*

$$\left[\text{prox}_\lambda\left[\mathcal{R}_\alpha^{\text{SPB}}\right](x)\right]_i = \text{springback}(x_i; \lambda, \alpha), \quad i = 1, \dots, n.$$

Proof. When $\mathcal{R}(x) = \mathcal{R}^{\ell_1}(x)$, it follows from (2.2) that, for any $z \in \mathbb{R}^n$ satisfying $0 \in z - y + \lambda\partial(\|z\|_1)$, there holds $z_i = \text{soft}(y_i; \lambda)$, i.e., $z = \text{prox}_\lambda\left[\mathcal{R}^{\ell_1}\right](y)$. The assumption $1 - \lambda\alpha > 0$ ensures $\nabla^2\left(\frac{1}{2}\|x-y\|_2^2 - \frac{\lambda\alpha}{2}\|x\|_2^2\right) = (1 - \lambda\alpha)I$ to be positive definite. Thus, the optimization problem occurred in (2.2) is convex. When $\mathcal{R}(x) = \mathcal{R}_\alpha^{\text{SPB}}(x)$ in (2.2), for any $z \in \mathbb{R}^n$ satisfying the condition $0 \in z - y + \lambda\partial(\|z\|_1) - \lambda\alpha z$, which is equivalent to

$$0 \in z - \frac{1}{1-\lambda\alpha}y + \frac{\lambda}{1-\lambda\alpha}\partial(\|z\|_1), \tag{2.4}$$

we have $z = \text{prox}_\lambda\left[\mathcal{R}_\alpha^{\text{SPB}}\right](y)$. It also follows from (2.4) that

$$z_i = \text{soft}\left(\frac{y_i}{1-\lambda\alpha}; \frac{\lambda}{1-\lambda\alpha}\right) = \text{springback}(y_i; \lambda, \alpha).$$

Hence, the assertion is proved. □

Recall that the springback penalty (1.4) is a weakly convex function. Its thresholding operator defined in (2.3) is also in between the soft and firm thresholding operators. As $\lim_{\mu\to\infty}\text{firm}(w; \lambda, \mu) = \text{soft}(w; \lambda)$, a compromising μ could be large enough such that $|w| \leq \mu$ and it reaches a certain compromise between the soft and firm thresholding operators. In this case, we have a particular springback thresholding operator

$$\text{springback}(w; \lambda, 1/\mu) = \begin{cases} 0, & |w| \leq \lambda, \\ \text{sgn}(w)\frac{\mu(|w|-\lambda)}{\mu-\lambda}, & |w| \geq \lambda. \end{cases}$$

If $1/\mu$ is replaced by a more general $\alpha > 0$, then the springback thresholding operator (2.3) is recovered.

2.5. Rationale of the name

Springback is a concept in applied mechanics (see, e.g., [41]). Fig. 1 gives more explanations for naming (1.4) *springback*. With $\lambda = 0.25$, Fig. 1 displays the thresholding operators for $w \in [-1.5, 1.5]$, including the soft thresholding operator, the transformed ℓ_1 thresholding operator with $\beta = 1$, the firm thresholding operator with $\mu = 0.75$, and the springback thresholding operator with $\alpha = 1/\mu$. The transformed ℓ_1 thresholding operator enforces w with $|w| \leq \lambda(\beta + 1)/\beta$ to be 0, and then its outputs approach to w as $|w|$ increases. All the other thresholding operators enforce w with $|w| \leq \lambda$ to be 0. For $w \geq \lambda$, the soft thresholding operator subtracts λ from $|w|$ and thus causes the ℓ_1 penalty to underestimate high-amplitude components; the firm thresholding operator's outputs jump from 0 to μ until $|w|$ exceeds μ, afterwards its output is w. For the springback thresholding operator, its outputs jump from 0 to μ until $|w|$ exceeds μ, and afterwards its outputs still keep going along the previous jumping trajectory.

In applied mechanics, spring is related to the process of bending some materials. When the bending process is done, the residual stresses cause the material to *spring back* towards its original shape, so the material must be *over-bent* to achieve the proper bending angle. Note that the soft thresholding operator always underestimates high-amplitude components, and the components $\|x\|_1$ and $-\frac{\alpha}{2}\|x\|_2^2$ in the springback penalty are decoupled. If we deem the soft thresholding operator as a process of over-bending, which stems for the component $\|x\|_1$, then the output of the soft thresholding operator will be sprung back toward w, which is achieved separately in consideration with the component $-\frac{\alpha}{2}\|x\|_2^2$. Such a springback process occurs for both $\lambda \leq |w| \leq \mu$ and $|w| \geq \mu$. The springback behavior is more obvious for those w with larger absolute values, and this coincides with the behavior of the springback penalty in Fig. 1. That is, once $|x|$ exceeds μ, the penalty turns around and heads towards the x-axis. This process may also be explained as a compensation of the loss of $|w|$ with $|w| \leq \lambda$.

3. Springback-penalized model for sparse signal recovery

In this section, we focus on the recovery of a sparse signal using the springback-penalized model (1.5). After reviewing some basic knowledge of compressed sensing, we identify some conditions for exact and robust recovery using the springback-penalized model (1.5), respectively.

3.1. Compressed sensing basics

In some seminal compressed sensing papers such as [4,13], recovery conditions have been established for the BP model (1.3). These conditions rely on the restricted isometry property (RIP) of the *sensing matrix* A, as proposed in [7].

Definition 3.1. For an index set $T \subset \{1, 2, \ldots, n\}$ and an integer s with $|T| \leq s$, the *s-restricted isometry constant* (RIC) of $A \in \mathbb{R}^{m \times n}$ is the smallest $\delta_s \in (0, 1)$ such that

$$(1 - \delta_s)\|x\|_2^2 \leq \|A_T x\|_2^2 \leq (1 + \delta_s)\|x\|_2^2$$

for all subsets T with $|T| \leq s$ and all $x \in \mathbb{R}^{|T|}$. The matrix A is said to satisfy the *s-restricted isometry property* (RIP) with δ_s.

Denoting by x^{opt} the minimizer of the BP problem (1.3), if A satisfies $\delta_{3s} < 3(1 - \delta_{4s}) - 1$, then for an s-sparse $\bar{x}$, one has

$$\|x^{\mathrm{opt}} - \bar{x}\|_2 \leq C_s \tau, \tag{3.1}$$

where C_s is a constant which may only depend on δ_{4s}. We refer to [5,6] for more details. If the measurements are noise-free, i.e., $\tau = 0$, then the error bound (3.1) implies *exact recovery*. Exact recovery is guaranteed only in the idealized situation where $\bar{x}$ is s-sparse and the measurements are noise-free. If the measurements are perturbed by some noise, then the bound (3.1) is usually referred to as the *robust recovery* result with respect to the measurement noise. In more realistic scenarios, we can only claim that $\bar{x}$ is close to an s-sparse vector, and the measurements may also be contaminated. In such cases, we can recover $\bar{x}$ with an error controlled by its distance to s-sparse vectors, and it was proved in [5] that

$$\|x^{\mathrm{opt}} - \bar{x}\|_2 \leq C_{1,s}\tau + C_{2,s}\frac{\|\bar{x} - \bar{x}_s\|_1}{\sqrt{s}}, \tag{3.2}$$

where $\bar{x}_s$ is the truncated vector corresponding to the s largest values of $\bar{x}$ (in absolute value), and $C_{1,s}$ and $C_{2,s}$ are two constants which may only depend on δ_{4s}. The bound (3.2) is usually referred to as the *stable recovery* results. Recovery conditions for other models with different penalties are usually not as extensive as the BP model (1.3). Under the framework of the RIP or some generalized versions, recovery theory for the BP model (1.3) has been generalized to the ℓ_p-penalized model in [9,17]. With the *unique representation property* of A, stable recovery results for the MCP-penalized model were derived in [43] and an upper bound for $\sum_{i=1}^{n} \phi_\mu^{\mathrm{MCP}}(x_i^{\mathrm{opt}} - \bar{x}_i)$, but not for $\|x^{\mathrm{opt}} - \bar{x}\|_2$, was obtained. We recommend the monograph [18] for a more comprehensive and detailed exhibition on compressed sensing.

3.2. Recovery guarantee using the springback-penalized model

Still denoting by x^{opt} the minimizer of the springback-penalized model (1.5), we have the following exact and robust recovery results of the model (1.5) for an s-sparse $\bar{x}$.

Theorem 3.1 (Recovery of sparse signals). *Let $\bar{x} \in \mathbb{R}^n$ be an unknown s-sparse vector to be recovered. For a given sensing matrix $A \in \mathbb{R}^{m \times n}$, let $b \in \mathbb{R}^m$ be a vector of measurements from $b = A\bar{x} + e$ with $\|e\|_2 \leq \tau$, and let δ_{3s} and δ_{4s} be the $3s$- and $4s$-RIC's of A, respectively. Suppose A satisfies $\delta_{3s} < 3(1 - \delta_{4s}) - 1$ and α satisfies*

$$\alpha \leq \frac{\sqrt{1-\delta_{4s}}\sqrt{3s} - \sqrt{1+\delta_{3s}}\sqrt{s}}{(\sqrt{1-\delta_{4s}} + \sqrt{1+\delta_{3s}})\|x^{\mathrm{opt}}\|_2}, \tag{3.3}$$

then the minimizer x^{opt} of the problem (1.5) satisfies $x^{\mathrm{opt}} = \bar{x}$ when $\tau = 0$; and it satisfies

$$\|x^{\mathrm{opt}} - \bar{x}\|_2 \leq \frac{\sqrt{2}}{\sqrt{D_1}}\sqrt{\tau} \tag{3.4}$$

when $\tau \geq 0$, where

$$D_1 = \frac{\alpha}{2}\frac{\sqrt{1-\delta_{4s}} + \sqrt{1+\delta_{3s}}}{\sqrt{3s} + \sqrt{s}}. \tag{3.5}$$

Proof. Let $x^{\mathrm{opt}} = \bar{x} + v$, and Λ_0 be the support of $\bar{x}$. It is clear that $v_{\Lambda_0} = x_{\Lambda_0}^{\mathrm{opt}} - \bar{x}$ and $v_{\Lambda_0^c} = x_{\Lambda_0^c}^{\mathrm{opt}}$. On the one hand, we know that

$$\|x^{\mathrm{opt}}\|_1 - \frac{\alpha}{2}\|x^{\mathrm{opt}}\|_2^2 \leq \|\bar{x}\|_1 - \frac{\alpha}{2}\|\bar{x}\|_2^2.$$

On the other hand, it holds that

$$
\begin{aligned}
\|x^{\mathrm{opt}}\|_1 - \frac{\alpha}{2}\|x^{\mathrm{opt}}\|_2^2 &= \|\bar{x} + v_{\Lambda_0}\|_1 + \|v_{\Lambda_0^c}\|_1 - \frac{\alpha}{2}\|\bar{x} + v\|_2^2 \\
&\geq \|\bar{x}\|_1 - \|v_{\Lambda_0}\|_1 + \|v_{\Lambda_0^c}\|_1 - \frac{\alpha}{2}\left(\|\bar{x}\|_2^2 + 2\langle \bar{x}, v\rangle + \|v\|_2^2\right).
\end{aligned}
$$

Then, we have that

$$
\|v_{\Lambda_0^c}\|_1 \leq \|v_{\Lambda_0}\|_1 - \frac{\alpha}{2}\|v\|_2^2 + \alpha\|v\|_2^2 + \alpha\langle \bar{x}, v\rangle = \|v_{\Lambda_0}\|_1 - \frac{\alpha}{2}\|v\|_2^2 + \alpha\langle x^{\mathrm{opt}}, v\rangle.
$$

We continue by arranging the indices in Λ_0^c in order of decreasing magnitudes (in absolute value) of $v_{\Lambda_0^c}$, and then dividing Λ_0^c into subsets of size $3s$. Set $\Lambda_0^c = \Lambda_1 \bigcup \Lambda_2 \bigcup \cdots \bigcup \Lambda_\ell$, i.e., Λ_1 contains the indices of the $3s$ largest entries (in absolute value) of $v_{\Lambda_0^c}$, Λ_2 contains the indices of the next $3s$ largest entries (in absolute value) of $v_{\Lambda_0^c}$, and so on. The cardinal number of Λ_ℓ may be less than $3s$. Denoting $\Lambda_{01} = \Lambda_0 \bigcup \Lambda_1$ and using the RIP of A, we have

$$
\|Av\|_2 \geq \|A_{\Lambda_{01}} v_{\Lambda_{01}}\|_2 - \left\|\sum_{i=2}^{\ell} A_{\Lambda_i} v_{\Lambda_i}\right\|_2 \geq \sqrt{1-\delta_{4s}}\|v_{\Lambda_{01}}\|_2 - \sqrt{1+\delta_{3s}}\sum_{i=2}^{\ell}\|v_{\Lambda_i}\|_2.
$$

As the magnitude of every v_t indexed by $t \in \Lambda_{i+1}$ is less than the average of magnitudes of v_t indexed by $t \in \Lambda_i$, there holds $|v_t| \leq \frac{\|v_{\Lambda_i}\|_1}{3s}$, where $t \in \Lambda_{i+1}$. Then, we have

$$
\|v_{\Lambda_{i+1}}\|_2^2 \leq 3s\frac{\|v_{\Lambda_i}\|_1^2}{(3s)^2} = \frac{\|v_{\Lambda_i}\|_1^2}{3s}.
$$

Together with $\|v_{\Lambda_0}\|_1 \leq \sqrt{s}\|v_{\Lambda_0}\|_2 \leq \sqrt{s}\|v_{\Lambda_{01}}\|_2$, we have

$$
\sum_{i=2}^{\ell}\|v_{\Lambda_i}\|_2 \leq \sum_{i=1}^{\ell-1}\frac{\|v_{\Lambda_i}\|_1}{\sqrt{3s}} \leq \frac{1}{\sqrt{3s}}\|v_{\Lambda_0^c}\|_1 \leq \frac{1}{\sqrt{3s}}\left(\sqrt{s}\|v_{\Lambda_{01}}\|_2 - \frac{\alpha}{2}\|v\|_2^2 + \alpha\langle x^{\mathrm{opt}}, v\rangle\right).
$$

Thus, it holds that

$$
\|Av\|_2 \geq \left(\sqrt{1-\delta_{4s}} - \frac{\sqrt{1+\delta_{3s}}\sqrt{s}}{\sqrt{3s}}\right)\|v_{\Lambda_{01}}\|_2 + \frac{\alpha\sqrt{1+\delta_{3s}}}{2\sqrt{3s}}\|v\|_2^2 - \frac{\alpha\sqrt{1+\delta_{3s}}}{\sqrt{3s}}\langle x^{\mathrm{opt}}, v\rangle. \tag{3.6}
$$

Note that

$$
\|v\|_2 \leq \|v_{\Lambda_{01}}\|_2 + \sum_{i=2}^{\ell}\|v_{\Lambda_i}\|_2 \leq \left(1 + \frac{\sqrt{s}}{\sqrt{3s}}\right)\|v_{\Lambda_{01}}\|_2 - \frac{\alpha}{2\sqrt{3s}}\|v\|_2^2 + \frac{\alpha}{\sqrt{3s}}\langle x^{\mathrm{opt}}, v\rangle,
$$

and it can be written as

$$
\|v_{\Lambda_{01}}\|_2 \geq \frac{\sqrt{3s}}{\sqrt{3s}+\sqrt{s}}\left(\frac{\alpha}{2\sqrt{3s}}\|v\|_2^2 + \|v\|_2 - \frac{\alpha}{\sqrt{3s}}\langle x^{\mathrm{opt}}, v\rangle\right).
$$

With the assumption $\delta_{3s} < 3(1 - \delta_{4s}) - 1$ on A, the coefficient of $\|v_{\Lambda_{01}}\|_2$ in (3.6) is positive and thus we have

$$
\begin{aligned}
\|Av\|_2 \geq & \frac{\sqrt{1-\delta_{4s}}\sqrt{3s} - \sqrt{1+\delta_{3s}}\sqrt{s}}{\sqrt{3s}+\sqrt{s}} \left(\frac{\alpha}{2\sqrt{3s}}\|v\|_2^2 + \|v\|_2 - \frac{\alpha}{\sqrt{3s}}\left\langle x^{\mathrm{opt}}, v \right\rangle \right) \\
& + \frac{\alpha\sqrt{1+\delta_{3s}}}{2\sqrt{3s}}\|v\|_2^2 - \frac{\alpha\sqrt{1+\delta_{3s}}}{\sqrt{3s}}\left\langle x^{\mathrm{opt}}, v \right\rangle \\
= & \frac{\alpha}{2}\left(\frac{\sqrt{1-\delta_{4s}} + \sqrt{1+\delta_{3s}}}{\sqrt{3s}+\sqrt{s}} \right)\|v\|_2^2 + \frac{\sqrt{1-\delta_{4s}}\sqrt{3s} - \sqrt{1+\delta_{3s}}\sqrt{s}}{\sqrt{3s}+\sqrt{s}}\|v\|_2 \\
& - \alpha\left(\frac{\sqrt{1-\delta_{4s}} + \sqrt{1+\delta_{3s}}}{\sqrt{3s}+\sqrt{s}} \right)\left\langle x^{\mathrm{opt}}, v \right\rangle .
\end{aligned}
\tag{3.7}
$$

If $\left\langle x^{\mathrm{opt}}, v \right\rangle \leq 0$, then $\|Av\|_2 \geq D_1\|v\|_2^2$. If $\left\langle x^{\mathrm{opt}}, v \right\rangle > 0$, then the condition (3.3) on α guarantees

$$
\begin{aligned}
& \frac{\sqrt{1-\delta_{4s}}\sqrt{3s} - \sqrt{1+\delta_{3s}}\sqrt{s}}{\sqrt{3s}+\sqrt{s}}\|v\|_2 - \alpha\left(\frac{\sqrt{1-\delta_{4s}} + \sqrt{1+\delta_{3s}}}{\sqrt{3s}+\sqrt{s}} \right)\left\langle x^{\mathrm{opt}}, v \right\rangle \\
\geq & \frac{\sqrt{1-\delta_{4s}}\sqrt{3s} - \sqrt{1+\delta_{3s}}\sqrt{s}}{\sqrt{3s}+\sqrt{s}}\left(\|v\|_2 - \left\langle \frac{x^{\mathrm{opt}}}{\|x^{\mathrm{opt}}\|_2}, v \right\rangle \right) \geq 0,
\end{aligned}
$$

where we use the Cauchy–Schwarz inequality. Hence we also have $\|Av\|_2 \geq D_1\|v\|_2^2$.

When $\tau = 0$, the inequality $Av = A(x^{\mathrm{opt}} - \bar{x}) = b - b = 0$ renders $0 = \|Av\|_2 \geq D_1\|v\|_2^2$, which implies $\|v\|_2 = 0$. Thus $x^{\mathrm{opt}} = \bar{x}$. When $\tau > 0$, the inequality

$$
\|Av\|_2 = \|Ax^{\mathrm{opt}} - A\bar{x}\|_2 \leq \|Ax^{\mathrm{opt}} - b\|_2 + \|A\bar{x} - b\|_2 \leq 2\tau
$$

leads to $2\tau \geq D_1\|v\|_2^2$, which implies (3.4). $\quad\square$

In analysis of signal recovery models with various convex and non-convex penalties, such as the ℓ_1 penalty [6,9] and the ℓ_{1-2} penalty [44,45], a linear lower bound for $\|A(x^{\mathrm{opt}} - \bar{x})\|_2$ is derived somehow. The proof of Theorem 3.1 mainly follows the idea of [6], but we derive a quadratic lower bound for the term $\|A(x^{\mathrm{opt}} - \bar{x})\|_2$. Thus, it is worthy noting that our results cannot be reduced to the result of the BP model (1.3) as $\alpha \to 0$. Indeed, the quadratic bound (3.6) in our proof is reduced to a linear bound as $\alpha \to 0$, which then leads to the same results as the BP model (1.3). However, we handle our final quadratic bound by removing its linear and constant terms and hence the obtained result cannot be reduced to the result of the BP model (1.3) as $\alpha \to 0$.

Besides, the condition (3.3) on α is required for the springback-penalized model (1.5). It is impossible to choose an α satisfying (3.3) unless we have *a priori* estimation on $\|x^{\mathrm{opt}}\|_2$ before solving the problem (1.5). Thus, the condition (3.3) then can be interpreted as a *posterior verification* in the sense that it can be verified once x^{opt} is obtained by solving the problem (1.5).

Remark 3.1 *(Posterior verification).* In practice, we solve the springback-penalized model (1.5) numerically and thus obtain an approximate solution, denoted by x^*, subject to a preset accuracy $\epsilon > 0$. That is, $\|x^{\mathrm{opt}} - x^*\|_2 \leq \epsilon$. Then, the posterior verification (3.3) is guaranteed if

$$
\alpha \leq \frac{\sqrt{1-\delta_{4s}}\sqrt{3s} - \sqrt{1+\delta_{3s}}\sqrt{s}}{(\sqrt{1-\delta_{4s}} + \sqrt{1+\delta_{3s}})(\|x^*\|_2 + \epsilon)}.
$$

Table 1
Exact recovery conditions recovery models with various penalties.

Penalty	RIP condition
ℓ_1 [6]	$\delta_{3s} < 3(1 - \delta_{4s}) - 1$
ℓ_p $(0 < p < 1)$ [9]	$\delta_{3s} < 3^{(2-p)/p}(1 - \delta_{4s}) - 1$
transformed ℓ_1 [48]	$\delta_{3s} < \left(\frac{\beta}{\beta+1}\right)^2 3(1 - \delta_{4s}) - 1$
ℓ_{1-2} [44,45]	$\delta_{3s} < a(s)(1 - \delta_{4s}) - 1$
springback	$\delta_{3s} < 3(1 - \delta_{4s}) - 1$

3.3. On the exact and robust recovery

In Theorem 3.1, we establish conditions for exact and robust recovery using the springback-penalized model (1.5). Table 1 lists the exact recovery conditions for five other popular models in the literature. In particular, the springback-penalized model (1.5) and the ℓ_1-penalized model, i.e., the BP model (1.3), have the same RIP condition. This condition is more stringent than that of the ℓ_p-penalized model $(0 < p < 1)$ but weaker than those of the transformed ℓ_1- and ℓ_{1-2}-penalized models. Beside the RIP condition, there is an additional assumption $a(s) > 1$ for the ℓ_{1-2}-penalized model, where $a(s)$ was first derived in [45] and slightly improved in [44] as

$$a(s) = \left(\frac{3s - 1}{\sqrt{3s} + \sqrt{4s - 1}}\right)^2.$$

Note that $a(s) < 3$ was shown in [44,45] for both the cases.

We then discuss robust recovery results. If $\alpha \to 0$, then the result (3.4) cannot provide any information as $\frac{\sqrt{2}}{\sqrt{D_1}} \to \infty$. However, for an appropriate α, the bound (3.4) is informative and attractive. The robust recovery results of the ℓ_1-, ℓ_p-, transformed ℓ_1- and ℓ_{1-2}-penalized models were shown to be linear with respect to the level of noise τ [6,9,44,45,48], in the sense of

$$\|x^{\text{opt}} - \bar{x}\|_2 \leq C_s \tau, \tag{3.8}$$

where C_s is some constant. Thus, under the conditions of Theorem 3.1, the bound (3.4) for the springback-penalized model (1.5) is tighter than (3.8) in the sense of

$$\frac{\sqrt{2}}{\sqrt{D_1}} \sqrt{\tau} \leq C_s \tau \tag{3.9}$$

if the level of noise τ satisfies

$$\tau > \frac{2}{D_1 C_s^2}. \tag{3.10}$$

Assume that the recovery conditions listed in Table 1 are satisfied for each model, respectively. Then, we can summarize their corresponding ranges of τ in Table 2 such that the robust recovery bound (3.4) of the springback-penalized model (1.5) is tighter than all the others in the sense of (3.9).

These ranges on τ look complicated. To have a better idea, we consider a toy example with $s = 20$, $\delta_{3s} = 1/4$, $\delta_{4s} = 1/3$, $\alpha = 1$ for the springback penalty (1.4), and $\beta = 1$ for the transformed ℓ_1 penalty. Then, the springback-penalized model (1.5) would give a tighter bound in the sense of (3.9) than the ℓ_1-, $\ell_{0.2}$-, $\ell_{0.5}$-, $\ell_{0.999}$-, transformed ℓ_1-, and ℓ_{1-2}-penalized models if $\tau > 0.1385, 0.0271, 0.2333, 0.1391, 0.0807$, and 2.8652×10^{-4}, respectively.

Table 2
Ranges of the level of noise such that the springback bound (3.4) is tighter than the bound (3.8) in the sense of (3.9).

Penalty	When the springback bound (3.4) is tighter than the bound (3.8)
ℓ_1 [5,6]	$\tau > \dfrac{(\sqrt{3s}+\sqrt{s})(\sqrt{3}\sqrt{1-\delta_{4s}}-\sqrt{1+\delta_{3s}})^2}{4\alpha(\sqrt{1+\delta_{3s}}+\sqrt{1-\delta_{4s}})}$
ℓ_p $(0 < p < 1)$ [37]	$\tau > \dfrac{(\sqrt{3s}+\sqrt{s})\left((1-\delta_{4s})^{p/2}-(1+\delta_{3s})^{p/2}3^{p/2-1}\right)^{2/p}}{\alpha(\sqrt{1-\delta_{4s}}+\sqrt{1+\delta_{3s}})\left(1+\frac{1}{(2/p-1)3^{2/p-1}}\right)}$
transformed ℓ_1 [48]	$\tau > \dfrac{4(\sqrt{3s}+\sqrt{s})(1-\delta_{3s})\left(\frac{\beta}{\beta+1}\sqrt{3}\sqrt{1-\delta_{4s}}-\sqrt{1+\delta_{3s}}\right)^2}{\alpha(\sqrt{1-\delta_{4s}}+\sqrt{1+\delta_{3s}})\left(\frac{\beta}{\beta+1}\sqrt{3}\sqrt{1-\delta_{4s}}-\sqrt{1+\delta_{3s}}+\sqrt{3s}\sqrt{1-\delta_{3s}}\right)^2}$
ℓ_{1-2} [44]	$\tau > \dfrac{(\sqrt{3s}+\sqrt{s})(\sqrt{a(s)(1-\delta_{4s})}-\sqrt{1+\delta_{3s}})^2}{\alpha(\sqrt{1-\delta_{4s}}+\sqrt{1+\delta_{3s}})(\sqrt{3s}-\sqrt{s\cdot a(s)})^2}$

Can we further improve the robust recovery result (3.4) in Theorem 3.1? The following proposition suggests a potential improvement. Moreover, without any requirement on α, this proposition also means, even if the posterior verification (3.3) is violated sometimes, the springback-penalized model (1.5) may still give a good recovery. Note that this proposition is only of conceptual sense, because its assumption $\langle x^{\mathrm{opt}}, x^{\mathrm{opt}} - \bar{x} \rangle \leq 0$ is not verifiable. Nevertheless, it helps us discern a possibility of achieving a better recovery bound than (3.4).

Proposition 3.1. *Let $\bar{x} \in \mathbb{R}^n$ be an unknown s-sparse vector to be recovered. For a given sensing matrix $A \in \mathbb{R}^{m \times n}$, let $b \in \mathbb{R}^m$ be a vector of measurements from $b = A\bar{x} + e$ with $\|e\|_2 \leq \tau$, and let δ_{3s} and δ_{4s} be the 3s- and 4s-RIC's of A, respectively. Let x^{opt} be the minimizer of the problem (1.5) and assume $\langle x^{\mathrm{opt}}, x^{\mathrm{opt}} - \bar{x} \rangle \leq 0$. Suppose A satisfies $\delta_{3s} < 3(1 - \delta_{4s}) - 1$, then $x^{\mathrm{opt}} = \bar{x}$ when $\tau = 0$; and x^{opt} satisfies*

$$\|x^{\mathrm{opt}} - \bar{x}\|_2 \leq \sqrt{\frac{D_2^2}{4D_1^2} + \frac{2}{D_1}\tau} - \frac{D_2}{2D_1} \tag{3.11}$$

when $\tau \geq 0$, where D_1 is the constant (3.5) given in Theorem 3.1 and

$$D_2 = \frac{\sqrt{3}\sqrt{1-\delta_{4s}} - \sqrt{1+\delta_{3s}}}{\sqrt{3}+1}. \tag{3.12}$$

Proof. In the case of $\langle x^{\mathrm{opt}}, v \rangle \leq 0$, it follows straightforwardly from (3.7) that

$$\|Av\|_2 \geq \frac{\alpha}{2}\left(\frac{\sqrt{1-\delta_{4s}} + \sqrt{1+\delta_{3s}}}{\sqrt{3s}+\sqrt{s}}\right)\|v\|_2^2 + \frac{\sqrt{1-\delta_{4s}}\sqrt{3s} - \sqrt{1+\delta_{3s}}\sqrt{s}}{\sqrt{3s}+\sqrt{s}}\|v\|_2$$
$$:= D_1\|v\|_2^2 + D_2\|v\|_2.$$

The assumption $\delta_{3s} < 3(1 - \delta_{4s}) - 1$ guarantees $D_2 > 0$. Hence, when $\tau = 0$, as $Av = A(x^{\mathrm{opt}} - \bar{x}) = 0$, we have $0 = \|Av\|_2 \geq D_1\|v\|_2^2 + D_2\|v\|_2$, which implies $\|v\|_2 = 0$. When $\tau > 0$, the inequality

$$\|Av\|_2 = \|Ax^{\mathrm{opt}} - A\bar{x}\|_2 \leq \|Ax^{\mathrm{opt}} - b\|_2 + \|A\bar{x} - b\|_2 \leq 2\tau$$

implies

$$\|v\|_2 \leq \frac{\sqrt{D_2^2 + 8D_1\tau} - D_2}{2D_1}.$$

The assertion is proved. □

Remark 3.2. The robust recovery result (3.11) is always better than (3.4) in Theorem 3.1 due to the sub-additivity of the square root function. Under the conditions of Proposition 3.1, the bound (3.11) for the springback-penalized model (1.5) is tighter than (3.8) in the sense of

$$\sqrt{\frac{D_2^2}{4D_1^2} + \frac{2}{D_1}\tau} - \frac{D_2}{2D_1} < C_s\tau,$$

if the level of noise τ satisfies

$$\tau > \frac{2 - D_2 C_s}{D_1 C_s^2} = \left(1 - \frac{D_2 C_s}{2}\right)\frac{2}{D_1 C_s^2}.$$

Comparing with (3.10), this improvement enlarges the value range of τ. For example, if C_s is the coefficient in the result (3.1) of the BP model (1.3), then $1 - D_2 C_s/2$ is approximately 0.2679.

4. Springback-penalized model for nearly sparse signal recovery

We then study the stable recovery of the springback-penalized model (1.5) when $\bar{x}$ is nearly sparse and the measurements are noisy.

4.1. Recovery guarantee using the springback-penalized model

If the signal $\bar{x}$ to be recovered is nearly s-sparse, then we have the following stable recovery theorem for the springback-penalized model (1.5).

Theorem 4.1 (*Recovery of nearly sparse signals*). *Let $\bar{x} \in \mathbb{R}^n$ be an unknown vector to be recovered. For a given sensing matrix $A \in \mathbb{R}^{m \times n}$, let $b \in \mathbb{R}^m$ be a vector of measurements from $b = A\bar{x} + e$ with $\|e\|_2 \leq \tau$, and let δ_{3s} and δ_{4s} be the 3s- and 4s-RIC's of A, respectively. Let $\bar{x}_s \in \mathbb{R}^n$ be the truncated vector corresponding to the s largest values of $\bar{x}$ (in absolute value). Suppose A satisfies $\delta_{3s} < 3(1 - \delta_{4s}) - 1$ and α satisfies (3.3), then the minimizer x^{opt} of the problem (1.5) satisfies*

$$\|x^{\mathrm{opt}} - \bar{x}\|_2 \leq \sqrt{\frac{2}{D_1}\tau + \frac{4}{\alpha}\|\bar{x} - \bar{x}_s\|_1}, \tag{4.1}$$

where D_1 is the constant (3.5) given in Theorem 3.1.

Proof. Let $x^{\mathrm{opt}} = \bar{x} + v$, and Λ_0 be the support of $\bar{x}_s$. It is clear that $v_{\Lambda_0} = x_{\Lambda_0}^{\mathrm{opt}} - \bar{x}_s$ and $v_{\Lambda_0^c} = x_{\Lambda_0^c}^{\mathrm{opt}} - \bar{x}_{\Lambda_0^c}$. We know that

$$\|x^{\mathrm{opt}}\|_1 - \frac{\alpha}{2}\|x^{\mathrm{opt}}\|_2^2 \leq \|\bar{x}\|_1 - \frac{\alpha}{2}\|\bar{x}\|_2^2 = \|\bar{x}_s\|_1 + \|\bar{x}_{\Lambda_0^c}\|_1 - \frac{\alpha}{2}\|\bar{x}\|_2^2.$$

On the other hand, it holds that

$$\|x^{\mathrm{opt}}\|_1 - \frac{\alpha}{2}\|x^{\mathrm{opt}}\|_2^2 = \|\bar{x}_s + v_{\Lambda_0}\|_1 + \|\bar{x}_{\Lambda_0^c} + v_{\Lambda_0^c}\|_1 - \frac{\alpha}{2}\|\bar{x} + v\|_2^2$$

$$\geq \|\bar{x}_s\|_1 - \|v_{\Lambda_0}\|_1 + \|v_{\Lambda_0^c}\|_1 - \|\bar{x}_{\Lambda_0^c}\|_1 - \frac{\alpha}{2}\left(\|\bar{x}\|_2^2 + 2\langle\bar{x}, v\rangle + \|v\|_2^2\right).$$

Then, v satisfies the following estimation:

$$\|v_{\Lambda_0^c}\|_1 \leq \|v_{\Lambda_0}\|_1 + 2\|\bar{x} - \bar{x}_s\|_1 - \frac{\alpha}{2}\|v\|_2^2 + \alpha\|v\|_2^2 + \alpha\langle\bar{x}, v\rangle$$

$$= \|v_{\Lambda_0}\|_1 + 2\|\bar{x} - \bar{x}_s\|_1 - \frac{\alpha}{2}\|v\|_2^2 + \alpha\langle x^{\mathrm{opt}}, v\rangle.$$

We divide Λ_0^c into subsets of size $3s$, $\Lambda_0^c = \Lambda_1 \bigcup \Lambda_2 \bigcup \cdots \bigcup \Lambda_\ell$, in terms of decreasing order of magnitudes (in absolute value) of $v_{\Lambda_0^c}$. Denoting $\Lambda_{01} = \Lambda_0 \bigcup \Lambda_1$ and using the RIP of A, we have

$$\|Av\|_2 \geq \|A_{\Lambda_{01}} v_{\Lambda_{01}}\|_2 - \left\| \sum_{i=2}^{\ell} A_{\Lambda_i} v_{\Lambda_i} \right\|_2 \geq \sqrt{1-\delta_{4s}} \|v_{\Lambda_{01}}\|_2 - \sqrt{1+\delta_{3s}} \sum_{i=2}^{\ell} \|v_{\Lambda_i}\|_2.$$

As proved for Theorem 3.1, we have $\sum_{i=2}^{\ell} \|v_{\Lambda_i}\|_2 \leq \|v_{\Lambda_0^c}\|_1 / \sqrt{3}$ and $\|v_{\Lambda_0}\|_1 \leq \sqrt{s} \|v_{\Lambda_{01}}\|_2$. Thus, we obtain

$$\sum_{i=2}^{\ell} \|v_{\Lambda_i}\|_2 \leq \frac{1}{\sqrt{3s}} \left(\sqrt{s} \|v_{\Lambda_{01}}\|_2 + 2\|\bar{x} - \bar{x}_s\|_1 - \frac{\alpha}{2} \|v\|_2^2 + \alpha \left\langle x^{\mathrm{opt}}, v \right\rangle \right).$$

Furthermore, it holds that

$$
\begin{aligned}
\|Av\|_2 &\geq \left(\sqrt{1-\delta_{4s}} - \frac{\sqrt{1+\delta_{3s}}\sqrt{s}}{\sqrt{3s}} \right) \|v_{\Lambda_{01}}\|_2 - \frac{2\sqrt{1+\delta_{3s}}}{\sqrt{3s}} \|\bar{x} - \bar{x}_s\|_1 \\
&\quad + \frac{\alpha\sqrt{1+\delta_{3s}}}{2\sqrt{3s}} \|v\|_2^2 - \frac{\alpha\sqrt{1+\delta_{3s}}}{\sqrt{3s}} \left\langle x^{\mathrm{opt}}, v \right\rangle.
\end{aligned}
\tag{4.2}
$$

As

$$
\begin{aligned}
\|v\|_2 &\leq \|v_{\Lambda_{01}}\|_2 + \sum_{i=2}^{\ell} \|v_{\Lambda_i}\|_2 \\
&\leq \left(1 + \frac{\sqrt{s}}{\sqrt{3s}} \right) \|v_{\Lambda_{01}}\|_2 + \frac{2}{\sqrt{3s}} \|\bar{x} - \bar{x}_s\|_1 - \frac{\alpha}{2\sqrt{3s}} \|v\|_2^2 + \frac{\alpha}{\sqrt{3s}} \left\langle x^{\mathrm{opt}}, v \right\rangle,
\end{aligned}
$$

we have

$$\|v_{\Lambda_{01}}\|_2 \geq \frac{\sqrt{3s}}{\sqrt{3s} + \sqrt{s}} \left(\frac{\alpha}{2\sqrt{3s}} \|v\|_2^2 + \|v\|_2 - \frac{\alpha}{\sqrt{3s}} \left\langle x^{\mathrm{opt}}, v \right\rangle - \frac{2}{\sqrt{3s}} \|\bar{x} - \bar{x}_s\|_1 \right).$$

Recall the assumption $\delta_{3s} < 3(1-\delta_{4s}) - 1$. The coefficient of $\|v_{\Lambda_{01}}\|_2$ in (4.2) is positive, and it follows that

$$
\begin{aligned}
\|Av\|_2 &\geq \frac{\sqrt{1-\delta_{4s}}\sqrt{3s} - \sqrt{1+\delta_{3s}}\sqrt{s}}{\sqrt{3s} + \sqrt{s}} \left(\frac{\alpha}{2\sqrt{3s}} \|v\|_2^2 + \|v\|_2 - \frac{\alpha}{\sqrt{3s}} \left\langle x^{\mathrm{opt}}, v \right\rangle - \frac{2}{\sqrt{3s}} \|\bar{x} - \bar{x}_s\|_1 \right) \\
&\quad + \frac{\alpha\sqrt{1+\delta_{3s}}}{2\sqrt{3s}} \|v\|_2^2 - \frac{\alpha\sqrt{1+\delta_{3s}}}{\sqrt{3s}} \left\langle x^{\mathrm{opt}}, v \right\rangle - \frac{2\sqrt{1+\delta_{3s}}}{\sqrt{3s}} \|\bar{x} - \bar{x}_s\|_1 \\
&= \frac{\alpha}{2} \left(\frac{\sqrt{1-\delta_{4s}} + \sqrt{1+\delta_{3s}}}{\sqrt{3s} + \sqrt{s}} \right) \|v\|_2^2 + \frac{\sqrt{1-\delta_{4s}}\sqrt{3s} - \sqrt{1+\delta_{3s}}\sqrt{s}}{\sqrt{3s} + \sqrt{s}} \|v\|_2 \\
&\quad - \alpha \left(\frac{\sqrt{1-\delta_{4s}} + \sqrt{1+\delta_{3s}}}{\sqrt{3s} + \sqrt{s}} \right) \left\langle x^{\mathrm{opt}}, v \right\rangle - 2 \left(\frac{\sqrt{1-\delta_{4s}} + \sqrt{1+\delta_{3s}}}{\sqrt{3s} + \sqrt{s}} \right) \|\bar{x} - \bar{x}_s\|_1.
\end{aligned}
\tag{4.3}
$$

If $\left\langle x^{\mathrm{opt}}, v \right\rangle \leq 0$, then $\|Av\|_2 \geq D_1 \|v\|_2^2 - \frac{4}{\alpha} D_1 \|\bar{x} - \bar{x}_s\|_1$. If $\left\langle x^{\mathrm{opt}}, v \right\rangle > 0$, then the condition (3.3) on α guarantees

$$\frac{\sqrt{1-\delta_{4s}}\sqrt{3s} - \sqrt{1+\delta_{3s}}\sqrt{s}}{\sqrt{3s} + \sqrt{s}} \|v\|_2 - \alpha \left(\frac{\sqrt{1-\delta_{4s}} + \sqrt{1+\delta_{3s}}}{\sqrt{3s} + \sqrt{s}} \right) \left\langle x^{\mathrm{opt}}, v \right\rangle \geq 0,$$

which is shown in the proof of Theorem 3.1. Hence, we also have $\|Av\|_2 \geq D_1 \|v\|_2^2 - \frac{4}{\alpha} D_1 \|\bar{x} - \bar{x}_s\|_1$. As $\|Av\|_2 = \|Ax^{\mathrm{opt}} - A\bar{x}\|_2 \leq \|Ax^{\mathrm{opt}} - b\|_2 + \|A\bar{x} - b\|_2 \leq 2\tau$, we have

$$2\tau \geq D_1\|v\|_2^2 - \frac{4}{\alpha}D_1\|\bar{x} - \bar{x}_s\|_1,$$

which implies (4.1).   □

Similar to the improvement in Proposition 3.1, the above stable recovery result can be improved as follows.

Proposition 4.1. *Let $\bar{x} \in \mathbb{R}^n$ be an unknown vector to be recovered. For a given sensing matrix $A \in \mathbb{R}^{m \times n}$, let $b \in \mathbb{R}^m$ be a vector of measurements from $b = A\bar{x} + e$ with $\|e\|_2 \leq \tau$, and let δ_{3s} and δ_{4s} be the 3s- and 4s-RIC's of A, respectively. Let x^{opt} be the minimizer of the problem (1.5) and assume $\langle x^{\mathrm{opt}}, x^{\mathrm{opt}} - \bar{x}\rangle \leq 0$. Let $\bar{x}_s \in \mathbb{R}^n$ be the truncated vector corresponding to the s largest values of $\bar{x}$ (in absolute value). Suppose A satisfies $\delta_{3s} < 3(1 - \delta_{4s}) - 1$, then x^{opt} satisfies*

$$\|x^{\mathrm{opt}} - \bar{x}\|_2 \leq \sqrt{\frac{D_2^2}{4D_1^2} + \frac{2}{D_1}\tau + \frac{4}{\alpha}\|\bar{x} - \bar{x}_s\|_1} - \frac{D_2}{2D_1},$$

where D_1 and D_2 are the constants (3.5) and (3.12) given in Theorem 3.1 and Proposition 3.1, respectively.

Proof. In the case of $\langle x^{\mathrm{opt}}, v\rangle \leq 0$, it follows straightforwardly from the estimation (4.3) that

$$\|Av\|_2 \geq D_1\|v\|_2^2 + D_2\|v\|_2 - \frac{4}{\alpha}D_1\|\bar{x} - \bar{x}_s\|_1.$$

The assumption $\delta_{3s} < 3(1 - \delta_{4s}) - 1$ guarantees $D_2 > 0$. Therefore, it follows from the triangle inequality that

$$\|Av\|_2 = \|Ax^{\mathrm{opt}} - A\bar{x}\|_2 \leq \|Ax^{\mathrm{opt}} - b\|_2 + \|A\bar{x} - b\|_2 \leq 2\tau.$$

We thus have

$$D_1\|v\|_2^2 + D_2\|v\|_2 - \frac{4}{\alpha}D_1\|\bar{x} - \bar{x}_s\|_1 \leq 2\tau, \tag{4.4}$$

which gives the improved result by solving the system of inequalities (4.4) and $\|v\|_2 \geq 0$.   □

4.2. On the stable recovery

If $\bar{x}$ is known to be s-sparse, then the estimation (4.1) in Theorem 4.1 is reduced to (3.4) in Theorem 3.1; and if the measurements are additionally noise-free, then both the estimations (3.4) and (4.1) imply exact recovery of the signal $\bar{x}$. We compare the estimation (4.1) with the estimation (3.2) for the BP model (1.3). The following comparison is based on theoretical error bounds. We are interested in the case where the estimation (4.1) is tighter than the estimation (3.2) in the sense of

$$\sqrt{\frac{2}{D_1}\tau + \frac{4}{\alpha}\|\bar{x} - \bar{x}_s\|_1} \leq C_{1,s}\tau + C_{2,s}\frac{\|\bar{x} - \bar{x}_s\|_1}{\sqrt{s}}, \tag{4.5}$$

which is equivalent to

$$\frac{s^{1/4}}{\sqrt{\alpha}}\sqrt{\frac{4(\sqrt{3}+1)}{\sqrt{1-\delta_{4s}} + \sqrt{1+\delta_{3s}}}\tau + \frac{4\|\bar{x} - \bar{x}_s\|_1}{\sqrt{s}}} \leq C_{1,s}\tau + C_{2,s}\frac{\|\bar{x} - \bar{x}_s\|_1}{\sqrt{s}}. \tag{4.6}$$

Note that s takes values among $\{1, 2, \ldots, n\}$ and the right-hand side of (4.6) decreases as s increases. If the left-hand side of (4.6) is smaller than the right-hand side of (4.6) for $s = 1$ and the left-hand side is larger than the right-hand side for $s = n$, then there must exist a constant C such that the inequality (4.5) holds for $s \leq C$. Besides, if $\bar{x}$ is known to be s-sparse, then $\|\bar{x} - \bar{x}_s\|_1 = 0$ and thus (4.6) implies the existence of C without any assumption. Therefore, we have the following corollary.

Corollary 4.1. *If $\bar{x}$ is s-sparse, then there exists a constant C such that the inequality (4.5) holds for $s \leq C$, where*

$$C = \alpha^2 C_{1,s}^4 \tau^2 \left(\frac{\sqrt{1 - \delta_{4s}} + \sqrt{1 + \delta_{3s}}}{4(\sqrt{3} + 1)} \right)^2. \tag{4.7}$$

When no information of the sparsity of $\bar{x}$ is known, if α satisfies

$$\frac{\frac{4(\sqrt{3}+1)}{\sqrt{1-\delta_4}+\sqrt{1+\delta_3}}\tau + 4\|\bar{x} - \bar{x}_1\|_1}{(C_{1,1}\tau + C_{2,1}\|\bar{x} - \bar{x}_1\|_1)^2} \leq \alpha \leq \frac{1}{C_{1,n}^2 \tau} \frac{4(\sqrt{3} + 1)\sqrt{n}}{\sqrt{1 - \delta_{4n}} + \sqrt{1 + \delta_{3n}}},$$

then there exists a constant C such that the inequality (4.5) holds for $s \leq C$, where C depends on α, $\bar{x}$, τ, δ_{3s}, and δ_{4s}.

In virtue of random matrix theory, we give two examples to show that the condition $\delta_{3s} < 3(1 - \delta_{4s}) - 1$ on A in Theorems 3.1 and 4.1 holds.

- *Random Gaussian matrices*: the entries of A are i.i.d. Gaussian with mean zero and variance $1/m$. It was shown in [6,7] that the condition $\delta_{3s} < 3(1 - \delta_{4s}) - 1$ holds with overwhelming probability when $s \leq C'm/\log(n/m)$, where C' is a constant. Similar results were extended to sub-gaussian matrices in [30].
- *Fourier ensemble*: A is obtained by selecting m rows from the $n \times n$ discrete Fourier transform and renormalizing the columns so that they are unit-normed. If the rows are selected at random, the condition $\delta_{3s} < 3(1 - \delta_{4s}) - 1$ holds with overwhelming probability for $s \leq C'm/(\log(n))^4$, where C' is a constant. This was initially considered in [8] and then improved in [36].

Remark 4.1. Assume that α satisfies the conditions in Theorem 4.1 and Corollary 4.1. For a random Gaussian sensing matrix A, if $s \leq C'm/\log(n/m)$, then the RIP condition $\delta_{3s} < 3(1 - \delta_{4s}) - 1$ on A holds with high probability; and additionally if, $C'm/\log(n/m) \leq C$, i.e.,

$$m \exp \left(\frac{C'}{C} m \right) \leq n,$$

then the estimation (4.1) is tighter than the estimation (3.2) in the sense of (4.5). For a randomly subsampled Fourier sensing matrix A, if $s \leq C'm/(\log(n))^4$, then the RIP condition on A holds with overwhelming probability; and additionally if $C'm/(\log(n))^4 \leq C$, i.e.,

$$m \leq \frac{C}{C'} (\log(n))^4,$$

then the estimation (4.1) is tighter than the estimation (3.2) in the sense of (4.5). In a nutshell, for a sensing matrix satisfying the RIP condition, if the number m of observation data is limited, where "limited" can be characterized as the fact that m is less than some constant depending on n, C, and C', then the stable recovery using the springback-penalized model (1.5) is guaranteed by a tighter bound than that of BP model

(1.3) in the sense of (4.5). These results can be extended to general orthogonal sensing matrices [8]. Similar comparative results with other recovery models may also be derived if the recovery error bounds of these models are linear to τ and $\|\bar{x} - \bar{x}_s\|_1$, e.g., the ℓ_{1-2}-penalized model [44].

5. Computational aspects of the springback-penalized model

Now we focus on computational aspects for the springback-penalized model (1.5). We first design an algorithm for solving (1.5) in Section 5.1, and then discuss its convergence in Section 5.2 and elaborate on how to solve its subproblems in Section 5.3.

5.1. DCA-springback: an algorithm for the springback penalized model

Some well-developed algorithms for solving difference-of-convex (DC) optimization problems can be easily implemented to solve the springback-penalized model (1.5). We focus on the simplest DCA in [39,40] without any line-search step, which has been shown to be efficient for solving signal recovery problems, see, e.g., [24,45,48].

Recall a standard DC optimization problem

$$\min_{x \in \mathbb{R}^n} \; f(x) := g(x) - h(x), \tag{5.1}$$

where g and h are lower semicontinuous proper convex functions on $\mathbb{R}^n$. Here, f is called a DC function, and $g - h$ is a DC decomposition of f. At each iteration, the DCA replaces the concave part $-h$ with a linear majorant and solves the resulting convex problem. That is, the DCA generates a sequence $\{x^k\}$ by solving the following subproblem iteratively:

$$x^{k+1} \in \arg \min_{x \in \mathbb{R}^n} \; \left\{ g(x) - \langle x - x^k, \xi^k \rangle \right\},$$

where $\xi^k \in \partial(h(x^k))$. Note that the springback-penalized model (1.5) can be written as

$$\arg \min_{x \in \mathbb{R}^n} F(x) := (\|x\|_1 + \chi_\Omega(x)) - \frac{\alpha}{2} \|x\|_2^2, \tag{5.2}$$

where $\Omega := \{x \in \mathbb{R}^n : \|Ax - b\|_2 \leq \tau\}$ and

$$\chi_\Omega(x) := \begin{cases} 0, & x \in \Omega, \\ +\infty, & x \notin \Omega, \end{cases}$$

is the indictor function of the set Ω. Thus, the DCA iterate scheme for solving (5.2) reads as

$$x^{k+1} \in \arg \min_x \left\{ (\|x\|_1 + \chi_\Omega(x)) - \langle x - x^k, \xi^k \rangle \right\}$$

$$= \arg \min_x \; \left\{ \|x\|_1 - \langle x - x^k, \xi^k \rangle \quad \text{s.t.} \quad x \in \Omega \right\}.$$

More specifically, the resulting DCA is listed in Algorithm 1, where $\epsilon_{\text{outer}} > 0$ is the preset tolerance for iterations, and "MaxIt" means the maximal number of iterations set beforehand.

Algorithm 1: DCA-springback: Solving the constrained springback model (1.5) via DCA.

Input: Model parameters: $\alpha > 0$ satisfying the condition (5.6);
Stopping criterion: $\epsilon_{\text{outer}} > 0$, MaxIt > 0;
Initialization: $k = 0$, x^0 satisfying $\|Ax - b\|_2 \leq \tau$;

1 while $k < $ MaxIt and $\min\left\{\|x^{k+1} - x^k\|_2, \|x^{k+1} - x^k\|_2 / \|x^k\|_2\right\} > \epsilon_{\text{outer}}$ **do**

2 $\xi^k = \alpha x^k$;

3 $x^{k+1} \in \arg\min\limits_x \left\{\|x\|_1 - \left\langle x - x^k, \xi^k \right\rangle \text{ s.t. } \|Ax - b\|_2 \leq \tau\right\}$;

4 $k \leftarrow k + 1$;

5 end

5.2. Convergence

Recall that the *modulus of strong convexity* of a convex function f on $\mathbb{R}^n$, denoted by $d(f)$, is defined as $d(f) := \sup\{\nu > 0 : f(\cdot) - \frac{\nu}{2}\|\cdot\|_2^2 \text{ is convex on } \mathbb{R}^n\}$. Then, according to [40, Proposition A.1], for a general DC function $f = g - h$, any sequence $\{x^k\}$ generated by the DCA satisfies

$$f(x^k) - f(x^{k+1}) \geq \frac{d(g) + d(h)}{2}\|x^{k+1} - x^k\|_2^2, \tag{5.3}$$

which immediately implies the decreasing property of $\{f(x^k)\}$ if at least one of g and h is strongly convex. Note that $\frac{\alpha}{2}\|x\|_2^2$ is strongly convex with modulus α. Thus, starting with a feasible x^0, we have the decreasing property

$$F(x^k) - F(x^{k+1}) \geq \frac{\alpha}{2}\|x^{k+1} - x^k\|_2^2, \tag{5.4}$$

where F is defined as (5.2). However, the decreasing property (5.4) of F is not sufficient to ensure the convergence of DCA-springback. The function F could be negative if α is inappropriately large. Note that for any x^k, we have

$$\|Ax^k\|_2 - \|b\|_2 \leq \|Ax^k - b\|_2 \leq \tau.$$

Moreover, as A is assumed to be full rank, we have $\sigma_{\min}(A) > 0$. It follows from the geometric interpretation of the SVD [42, Lecture 4] that $\|Ax\|_2 \geq \sigma_{\min}(A)$ for any $x \in \mathbb{R}^n$ on the unit sphere $\{x \in \mathbb{R}^n : \|x\|_2 = 1\}$. Thus, it holds that

$$0 < \sigma_{\min}(A) \leq \min_{x \in \mathbb{R}^n} \frac{\|Ax\|_2}{\|x\|_2} = \min_{\|x\|_2 = 1} \|Ax\|_2,$$

and we have

$$\|x^k\|_2 \leq \frac{\|b\|_2 + \tau}{\sigma_{\min}(A)}. \tag{5.5}$$

Note that $\|x\|_1 - \frac{\alpha}{2}\|x\|_2^2 \geq 0$ and hence F is non-negative if $\|x\|_2 \leq 2/\alpha$. Clearly, if

$$\alpha \leq \frac{2\sigma_{\min}(A)}{\|b\|_2 + \tau}, \tag{5.6}$$

then $F(x^k) \geq 0$ for any $k \geq 0$ because all iterates x^k satisfy (5.5). Together with the decreasing property (5.4), we can establish the convergence of DCA-springback easily by following the analytical framework in [39,40]. Moreover, it follows the convergence of $\{F(x^k)\}$ and (5.4) that $\|x^{k+1} - x^k\|_2 \to 0$ as $k \to \infty$.

Remark 5.1. Note that the condition (3.3) depends on the RIP condition of A, and (5.6) depends on the conditioning of A. It is easy to deduce that if

$$\frac{\sqrt{1 - \delta_{4s}}\sqrt{3s} - \sqrt{1 + \delta_{3s}}\sqrt{s}}{\sqrt{1 - \delta_{4s}} + \sqrt{1 + \delta_{3s}}} \leq \frac{2\sigma_{\min}(A)\|x^{\mathrm{opt}}\|_2}{\|b\|_2 + \tau}, \tag{5.7}$$

then the condition (5.6) is implied by (3.3). Otherwise, it can be verified that the condition (3.3) is implied by (5.6).

5.3. Solving the subproblem of DCA-springback

For the proposed DCA-springback, its subproblem at each iteration is

$$\min_x \ \|x\|_1 - \langle x - x^k, \xi^k \rangle \quad \text{s.t.} \quad \|Ax - b\|_2 \leq \tau. \tag{5.8}$$

This problem can be easily solved by, e.g., the ADMM, which was originally proposed in [20] and had been well developed in the literature such as [11,22]. Some details are given for completeness. Note that the subproblem (5.8) can be reformulated as

$$\min_{x,y,z} \ \|y\|_1 - \langle x - x^k, \xi^k \rangle$$
$$\text{s.t. } y = x,$$
$$z = Ax - b,$$
$$z \in \mathcal{B}(\tau),$$

where $y, z \in \mathbb{R}^n$ are two auxiliary variables. With some trivial details skipped, the iterative scheme of the (scaled) ADMM for the subproblem (5.8) reads as

$$\begin{cases} x^{j+1} = (\rho A^{\mathrm{T}} A + \zeta I)^{-1} \left(\rho A^{\mathrm{T}}(b + z^j - \eta^j) + \xi^k + \zeta(y^j - u^j) \right), \\ y_i^{j+1} = \mathrm{soft}(x_i^{j+1} + u_i^j; 1/\zeta) \text{ for } i = 1, \ldots, n, \\ z^{j+1} = \mathcal{P}_{\mathcal{B}(\tau)}(Ax^{j+1} - b + \eta^j), \\ u^{j+1} = u^j + x^{j+1} - y^{j+1}, \\ \eta^{j+1} = \eta^{j+1} + Ax^{j+1} - b - z^{j+1}, \end{cases} \tag{5.9}$$

where $u \in \mathbb{R}^n$ and $\eta \in \mathbb{R}^m$ are the Lagrange multipliers, $\zeta > 0$ and $\rho > 0$ are penalty parameters, and $\mathcal{P}_{\mathcal{B}(\tau)}(\cdot)$ is the projection operator onto the ball $\mathcal{B}(\tau)$. If the measurement process is noise-free, i.e., $\tau = 0$, then z^j is always set as zero and the projection of the z-subproblem in (5.9) is not necessary.

6. Numerical experiments

In this section, we implement the DCA-springback to the constrained springback-penalized model (1.5) with simulated data. All codes were written by MATLAB R2022a, and all numerical experiments were conducted on a laptop (16 GB RAM, Intel® CoreTM i7-9750H Processor) with macOS Monterey 12.4.

We mainly show the effectiveness of the model (1.5) for some specific scenarios and demonstrate the efficiency of the DCA-springback. Several state-of-the-art signal recovery solvers listed below are also tested for comparison.

1) The accelerated iterative hard thresholding (AIHT) algorithm in [2]: solving the constrained model

$$\min_{x \in \mathbb{R}^n} \ \|Ax - b\|_2^2 \quad \text{s.t.} \quad \|x\|_0 \le s$$

by the accelerated iterative hard thresholding, where s is set beforehand to estimate the sparsity of x. For simplicity, we only choose the fundamental AIHT in [2], and refer to, e.g., [16,23,25,26,33,34], for various other more sophisticated algorithms.

2) ADMM-ℓ_1 [20]: solving the unconstrained ℓ_1-penalized problem by the ADMM.

3) IRLS-ℓ_p ($0 < p < 1$) [28]: smoothing the unconstrained ℓ_p-penalized model as

$$\min_{x \in \mathbb{R}^n} \ \frac{1}{2}\|Ax - b\|_2^2 + \lambda \|x\|_{p,\epsilon}^p \quad \text{with} \quad \|x\|_{p,\epsilon}^p := \sum_{j=1}^{n} (x_j^2 + \epsilon^2)^{p/2},$$

where $\epsilon > 0$, and implementing the iteratively reweighted least squares (IRLS) algorithm.

4) DCA-TL1 [48]: solving the unconstrained transformed ℓ_1-penalized model with parameter β by DCA and implementing the ADMM for its subproblems.

5) DCA-ℓ_{1-2} [45]: solving the unconstrained ℓ_{1-2}-penalized model by DCA and implementing the ADMM for its subproblems.

6) DCA-MCP [38]: solving the unconstrained MCP-penalized model by DCA and implementing the ADMM for its subproblems (the authors in [38] consider the ℓ_1-norm data fidelity term instead of the ℓ_2 norm, but the implementation of the MCP term is similar).

Note that the AIHT solves the ℓ_0-penalized model directly; the ADMM-ℓ_1 solves a convex surrogate model, and the others solve different non-convex approximate models.

6.1. Setup

We consider both incoherent and coherent sensing matrices to generate synthetic data for simulation. In the incoherent regime, we use random Gaussian matrices and random partial discrete cosine transform (DCT) matrices. For the former kind, its columns are generated by

$$A_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_m/m), \quad i = 1, \ldots, n,$$

where $\mathcal{N}(0, I_m/m)$ is the multivariate Gaussian distribution with location 0 and covariance I_m/m. For the latter kind, its columns are generated by

$$A_i = \frac{1}{\sqrt{m}} \cos(2i\pi \chi_i), \quad i = 1, \ldots, n,$$

where $\chi_i \in \mathbb{R}^m \overset{\text{i.i.d.}}{\sim} \mathcal{U}([0,1]^m)$ is uniformly and independently sampled from $[0,1]$. Note that both kinds of matrices have small RIP constants with high probability. The coherent regime consists of more ill-conditioned sensing matrices with higher coherence, and it is represented by the randomly oversampled partial DCT matrix in our experiments. A randomly oversampled partial DCT matrix is defined as

$$A_i = \frac{1}{\sqrt{m}} \cos(2i\chi_i/\mathcal{F}), \quad i = 1, \ldots, n,$$

where $\mathcal{F} \in \mathbb{N}$ is the *refinement factor*. As $\mathcal{F}$ increases, A becomes more coherent. A matrix sampled in this way cannot satisfy an RIP, and the sparse recovery with such a matrix is possible only if the non-zero

elements of the ground-truth $\bar{x}$ are sufficiently separated. Technically, we select the elements of supp$(\bar{x})$ such that $\min_{j,k\in\mathrm{supp}(\bar{x})}|j-k| \geq L$, where L is characterized as the *minimum separation*.

We generate a ground-truth vector $\bar{x} \in \mathbb{R}^n$ with sparsity s supported on a random index set (for incoherent matrices) or an index set satisfying the required minimum separation (for coherent matrices) with non-zero entries i.i.d. drawn from the normal distribution. We then compute $b = A\bar{x}$ as the measurements, and apply each solver to produce a reconstruction vector x^* of $\bar{x}$. A reconstruction is considered successful if the relative error satisfies $\|x^* - \bar{x}\|_2/\|\bar{x}\|_2 < 10^{-3}$. We test some cases with different sparsity s of $\bar{x}$, different levels of noise, or different numbers of measurements. We run 100 times independently for each scenario and report the success rate, which is the ratio of the number of successful trials over 100. All experiments are run in parallel with the MATLAB Parallel Computing Toolbox.

The initial guess for all tested algorithms is $x^0 = 0$. The choice of the parameter α in the springback penalty is discussed in Section 6.2. For outer iterates of the DCA-springback, we set $\rho = 10^5$, MaxIt = 10, and $\epsilon_{\mathrm{outer}} = 10^{-5}$ (for noise-free measurements) or 10^{-3} (for noisy measurements). To implement the ADMM (5.9) for subproblems, we set $\zeta = 10^{-5}$, $\tau = \|A\bar{x} - b\|_2$, and the stopping criterion as either $\|x^{j+1} - x^j\|_2/\max\{\|x^{j+1}\|_2, \|x^j\|_2\} < 10^{-5}$ or the iteration number exceeds 500. The DCA-TL1, the DCA-ℓ_{1-2}, and the DCA-MCP are solved by DCA and their subproblems are also solved by the ADMM. We thus set the regularization parameter $\lambda = 10^{-6}$ and adopt the same parameters of the rest and stopping criterion as the DCA-springback. In particular, the parameter β in the transformed ℓ_1 penalty is set as 1 for the DCA-TL1, following [48], and the parameter μ in the MCP is set as $1/\alpha$ for the DCA-MCP. For the AIHT, we set all parameters as [2]. For the ADMM-ℓ_1, we set $\lambda = 10^{-6}$, $\zeta = 10^{-5}$, $\epsilon_{\mathrm{outer}} = 10^{-5}$ (for noise-free measurements) and 10^{-3} (for noisy measurements), and MaxIt = 5000. For IRLS-ℓ_p, we set $p = 0.5$, $\lambda = 10^{-6}$, $\epsilon_{\mathrm{outer}} = 10^{-8}$, and MaxIt = 1000.

6.2. A subroutine for choosing the model parameter α

Let us focus on the parameter α of the springback penalty (1.4). For an 128×512 random Gaussian matrix, we test the DCA-springback with different α varying among $\{0.2, 0.4, 0.6, 0.8, 1\}$, and different levels of sparsity s among $\{25, 27, \ldots, 65\}$. The DCA-springback with $\alpha = 0.6$ or 0.8, indicated by success rates in Fig. 2, has the best performance. For small α such as 0.2 and 0.4, the DCA-springback is not satisfactory because the springback penalty performs similarly to the ℓ_1 penalty. For $\alpha = 1$, its performance is also inferior since the convergence condition of the DCA-springback or the posterior verification (3.3) can be easily violated with a large α. We refer to the latter reason as the "violating behavior" of the DCA-springback. An "unsuccessful" trial is recognized due to unsatisfactory (but reasonable) recovery or violating behavior. Thus, success rates cannot fully reflect "violating behavior," and we also plot the relative errors in Fig. 2. Indeed, the "violating behavior" often occurs when s becomes large. Performance of $\alpha = 0.8$ and 1 is generally inferior, and also there are few such cases when $\alpha = 0.6$. Thus, we adopt a `safeguard` for $\alpha = 0.7$, a compromise between 0.6 and 0.8. If $\alpha = 0.7$ violates the condition (5.6), then we replace 0.7 with the largest constant complying with this condition (5.6). That is, we choose $\alpha = \min\{0.7, 2\sigma_{\min}(A)/(\|b\|_2 + \tau)\}$. Success rates and relative errors with `safeguarded` $\alpha = 0.7$ are also displayed in Fig. 2, indicating that there is no violating behavior.

Though a reasonable upper bound of α is needed, behaviors for $\alpha = 0.2$ and 0.4 suggest that a lower bound for α should be taken to maintain the satisfactory performance of the DCA-springback in terms of success rates. Especially if A is ill-conditioned in the sense that its singular values lie within a wide range of values, i.e., $\sigma_{\min}(A)$ could be very small, then the condition on α could be pretty stringent. To maintain the success rates of the DCA-springback, we adopt an `efficiency detection` step as follows. If the condition number $\mathrm{cond}(A) := \sigma_{\max}(A)/\sigma_{\min}(A)$ is greater than 5 (or other values set by the user), then we start an `efficiency detection` to enforce α to be greater than an *efficiency detection factor* ω. Thus, we suggest choosing α as the following subroutine:
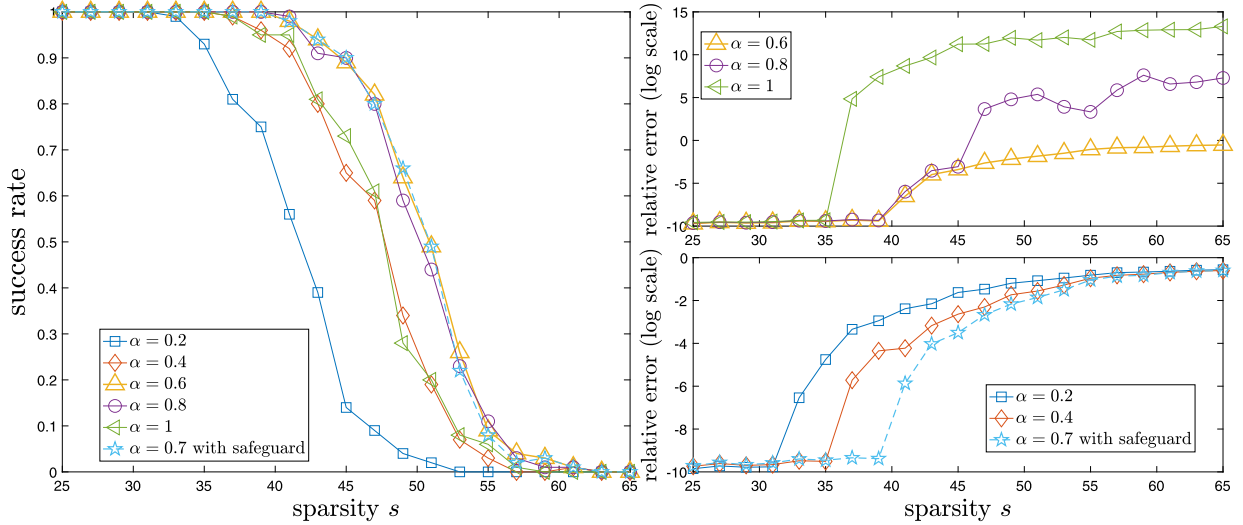
Fig. 2. Success rates and relative errors in a natural logarithmic scale of recovery using DCA-springback under 128×512 random Gaussian sensing matrix, with various α.

$$\alpha = \begin{cases} \min\left\{0.7, 2\sigma_{\min}(A)/(\|b\|_2 + \tau)\right\}, & \text{if } \operatorname{cond}(A) \leq 5, \\ \max\left\{\omega, \min\{0.7, 2\sigma_{\min}(A)/(\|b\|_2 + \tau)\}\right\}, & \text{otherwise.} \end{cases} \tag{6.1}$$

In short, the `safeguard` step suffices to guarantee convergence of the DCA-springback; and the `efficiency detection` step is adopted to maintain the success rates of the DCA-springback for ill-conditioned sensing matrices.

6.3. Exact recovery of sparse vectors

We first compare the DCA-springback with some state-of-the-art solvers mentioned above for noise-free measurements. We consider both the incoherent and coherent sensing matrices, respectively.

Tests on incoherent matrices. We first consider a ground-truth vector and display its reconstructions by the ADMM-ℓ_1, the DCA-TL1, the DCA-ℓ_{1-2}, the DCA-MCP, and the DCA-springback. Let the sensing matrix $A \in \mathbb{R}^{m \times n}$ be a random Gaussian matrix with $(m, n) = (64, 250)$, and the ground-truth $\bar{x} \in \mathbb{R}^{250}$ be a 22-sparse vector with nonzero entries drawn from the standard normal distribution and set the efficiency detection factor as $\omega = 0.5$. The ground-truth and its reconstructions are displayed in Fig. 3. We see that the DCA-springback, the DCA-MCP, and the DCA-TL1 produce better reconstructions than the ADMM-ℓ_1 and the DCA-ℓ_{1-2}.

We then conduct a more comprehensive study and involve more solvers. We choose the sensing matrix $A \in \mathbb{R}^{m \times n}$ as a random Gaussian matrix and random partial DCT matrices with $(m, n) = (64, 160)$, $(64, 320)$, and $(64, 640)$, and set the efficiency detection factor as $\omega = 0.5$. Different levels of sparsity s varying among $\{6, 8, 10, \ldots, 40\}$ are tested. The success rates of each solver are plotted in Fig. 4. For both the Gaussian and partial DCT matrices, the IRLS-ℓ_p with $p = 0.5$ has the best performance, followed by the DCA-TL1, the DCA-MCP, and the DCA-springback. In particular, the performances of the DCA-MCP and the DCA-springback are very close because we let the parameter μ in the MCP be $1/\alpha$. The DCA-ℓ_{1-2} performs moderately well, outperforming both the ADMM-ℓ_1 and the AIHT. Our numerical results are consistent with some observations in the literature (e.g., [45,48]).

Tests on coherent matrices. Now, we choose the sensing matrix $A \in \mathbb{R}^{100 \times 1500}$ as a randomly oversampled partial DCT matrix with various refinement factors $\mathcal{F} = 4, 6, 8, 10, 12, 16$ and minimum separation $L = 2\mathcal{F}$, with the sparsity s varying among $\{5, 7, 9, \ldots, 35\}$. The efficiency detection factor is set as $\omega = 0.5$. The
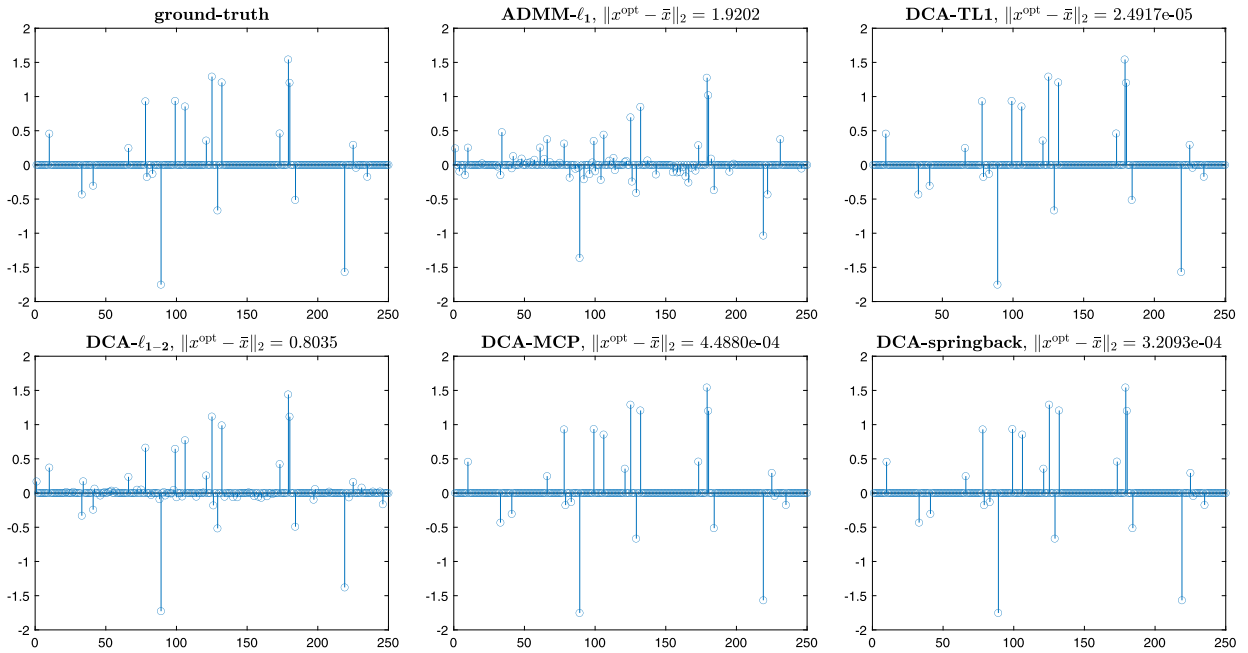
Fig. 3. A ground-truth and its reconstructions using random Gaussian sensing matrices and noise-free measurements.
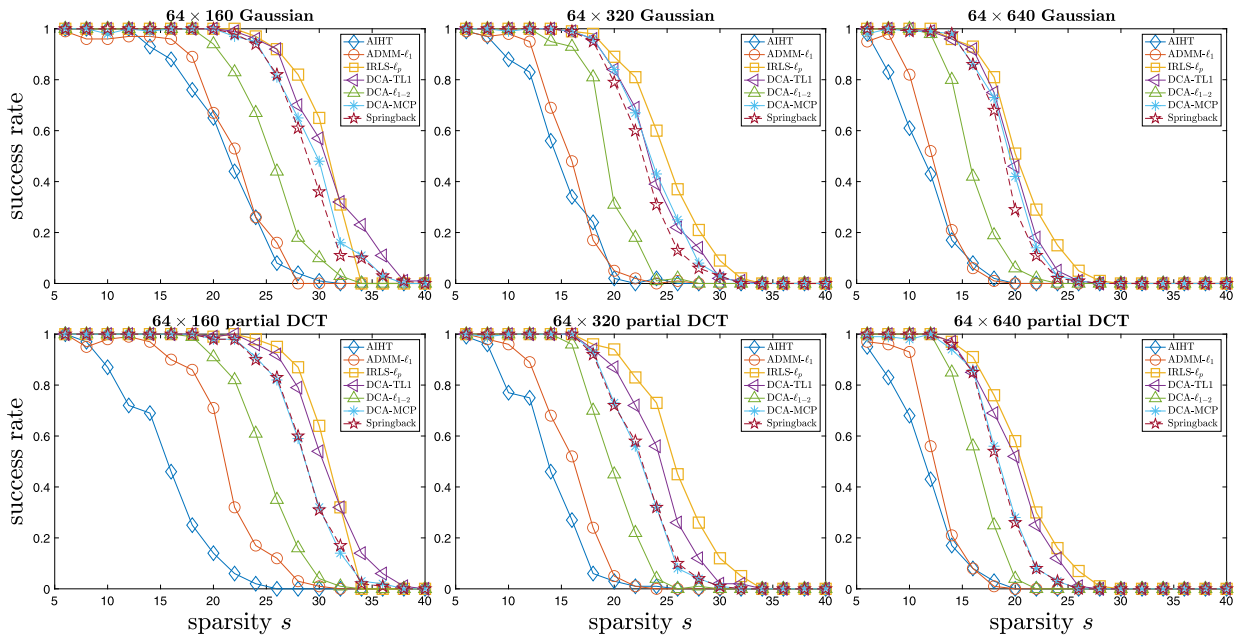

Fig. 4. Success rates using random Gaussian and partial DCT sensing matrices.

success rates of each solver are plotted in Fig. 5. This figure suggests that the DCA-TL1, the DCA-MCP, and the DCA-springback are robust regardless of the varying coherence of sensing matrix A. Moreover, when the coherence of A is modest, e.g. $\mathcal{F} = 6, 8$, the DCA-MCP and the DCA-springback perform better than others. In the coherent regime, the DCA-springback is comparable with the DCA-ℓ_{1-2}, and it outperforms the DCA-TL1, the ADMM-ℓ_1, the IRLS-ℓ_p, and the AIHT. However, the best-performance solver IRLS-ℓ_p in the incoherent regime becomes inefficient as A becomes coherent.
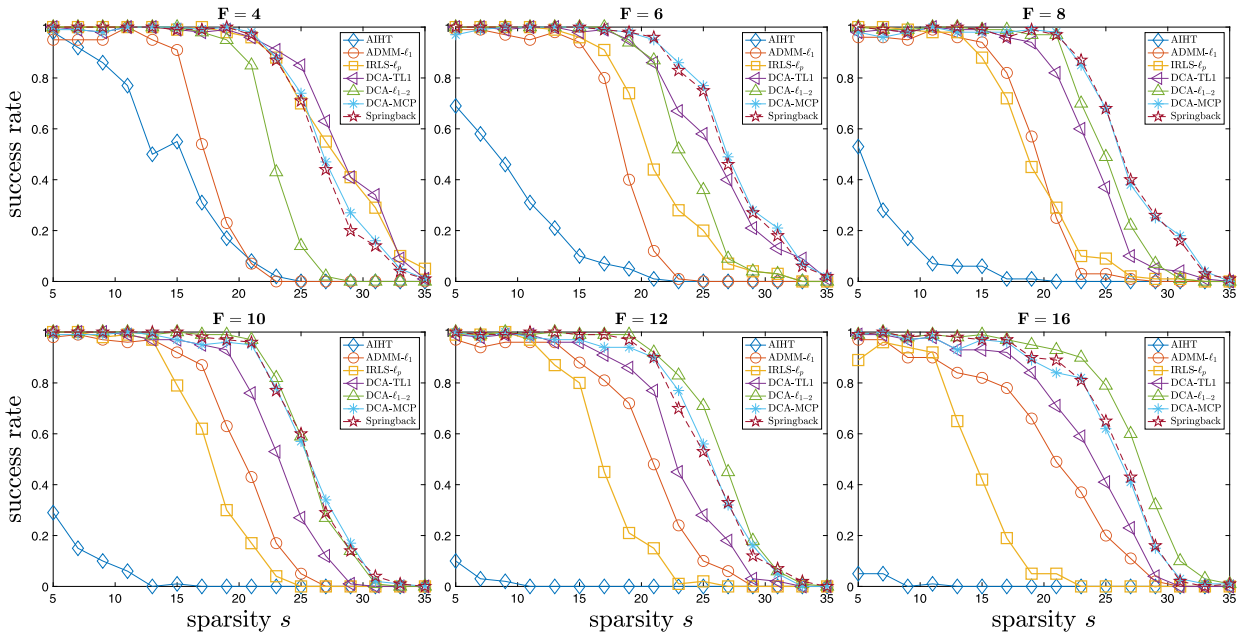
Fig. 5. Success rates using randomly oversampled partial DCT matrices in $\mathbb{R}^{100 \times 1500}$.

6.4. Robust recovery in the presence of noise

We then consider noisy measurements. The noisy measurements b are obtained by `b = awgn(A`$\bar{x}$`,snr)`, a subroutine of the MATLAB Communication Toolbox, where `snr` corresponds to the value of signal-to-noise ratio (SNR) measured in dB. The larger the value of SNR is, the lighter the noise is added on.

We first consider a ground-truth vector with noisy measurements and display its reconstructions by the ADMM-ℓ_1, the DCA-TL1, the DCA-ℓ_{1-2}, the DCA-MCP, and the DCA-springback. Let the sensing matrix $A \in \mathbb{R}^{m \times n}$ be a random Gaussian matrix with $(m, n) = (64, 250)$, and the ground-truth $\bar{x} \in \mathbb{R}^{250}$ be a 20-sparse vector with nonzero entries drawn from the standard normal distribution and set the efficiency detection factor as $\omega = 0.4$. The measurement vector $b = A\bar{x}$ is contaminated by 30 dB noise. The ground-truth and its reconstructions are displayed in Fig. 6. In particular, we see that the DCA-springback works better on small perturbations than the other solvers.

We test both the random Gaussian matrix and the randomly oversampled partial DCT matrix with different levels of noise in dB. For Gaussian measurements, we choose $n = 64$, $m = 128$, and $s = 25$. For the oversampled partial DCT measurements, we test $n = 1500$, $m = 128$, $s = 30$, and $\mathcal{F} = 8$. We run 100 times for each scenario and record the average errors. The efficiency detection factor is set as $\omega = 0.4$.

Once we adopt the `efficiency detection` step, a single "violating behavior" could lift the mean error to a pretty large level. To overcome this computational myopia, we only reserve the accepted results, where a result of the DCA-springback is considered "accepted" if the absolute error $\|x^* - \bar{x}\|_2$ is ten times less than the absolute error of the ADMM-ℓ_1. In addition to errors displayed in Fig. 7, we report the *acceptance rates* of the DCA-springback, which are ratios of the number of accepted trials over 100.

According to our experiments, there are no "violating behaviors" with the Gaussian measurements. However, there are a few cases with the oversampled partial DCT measurements when the noise level is relatively large. To illustrate the necessity of the `efficiency detection` step and to validate the convergence condition (5.6), we test the DCA-springback without the `efficiency detection` for the randomly oversampled partial DCT measurements, and we do not remove unaccepted trials. The results are labeled as "DCA-springback w/o effcy det." in Fig. 7, as we see that the DCA-springback only performs slightly better than the ADMM-ℓ_1.
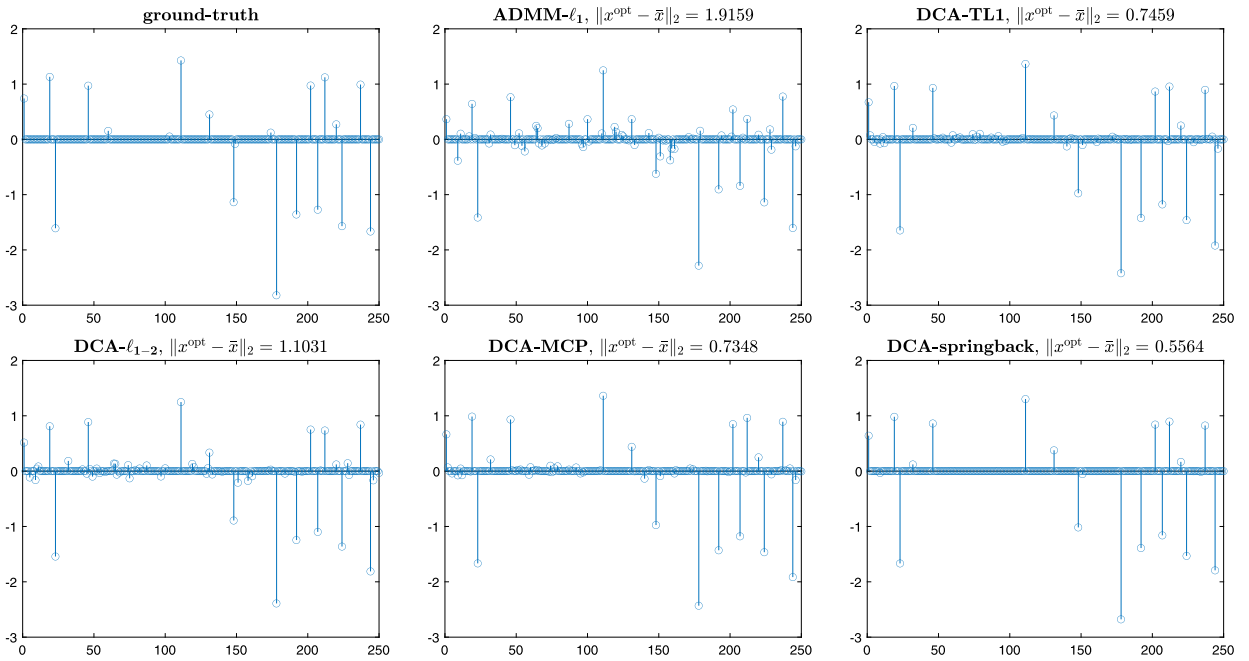
Fig. 6. A ground-truth and its reconstructions using random Gaussian sensing matrices and noisy measurements.
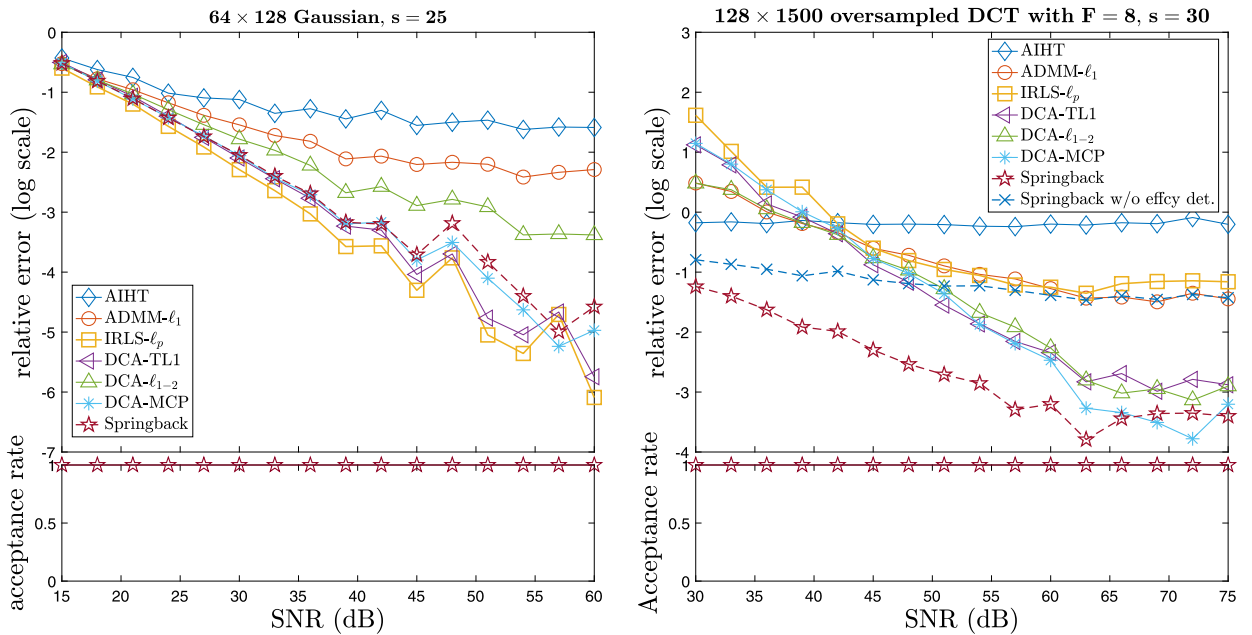

Fig. 7. Robust recovery results with randomly Gaussian and oversampled partial DCT measurements.

Fig. 7 shows that the DCA-ℓ_{1-2} and the IRLS-ℓ_p are still sensitive to the coherence of A. For Gaussian measurements, the IRLS-ℓ_p with $p = 0.5$ has the best performance, followed by the DCA-TL1, the DCA-MCP, the DCA-springback, the DCA-ℓ_{1-2}, and the ADMM-ℓ_1. For oversampled DCT measurements, the DCA-springback appears to be the best solver, followed by the DCA-MCP, the DCA-ℓ_{1-2}, and the DCA-TL1, because the noise level is considered in solving the subproblems of the DCA-springback. In both cases, the DCA-springback consistently performs better than the ADMM-ℓ_1 and the DCA-ℓ_{1-2}. AIHT appears not
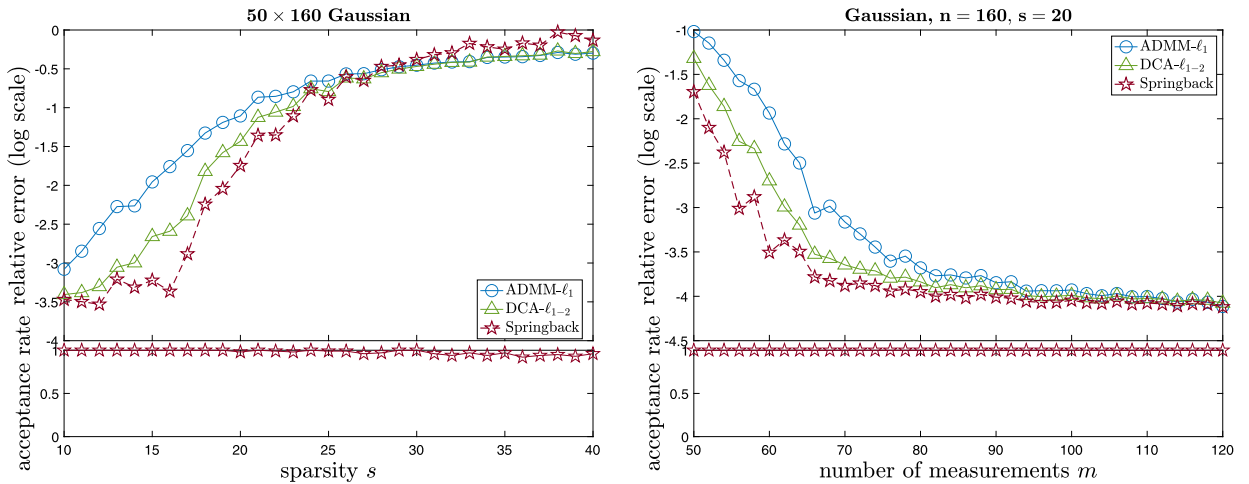
Fig. 8. Numerical validation of theoretical results in Section 4.2.

to perform well for both matrices. According to the plots of the DCA-springback and the DCA-springback without the `efficiency detection`, the model parameter α matters for the same solver.

We also validate some theoretical results proved in Section 4.2, with Gaussian measurements perturbed by 45 dB noise. We first study $m = 50$, $n = 160$, and s varying among $\{10, 11, \dots, 40\}$, and then consider $n = 160$, $s = 20$, and m varying among $\{50, 51, \dots, 120\}$. Errors of the ADMM-ℓ_1, the DCA-ℓ_{1-2}, and the DCA-springback are plotted in Fig. 8, and the acceptance rates of the DCA-springback are also displayed. According to our analysis in Section 4.2, for an RIP sensing matrix A and an s-sparse $\bar{x}$, when $s \leq C$ (C is given in (4.7)) or m is limited by some constant, the estimation (4.1) of the springback-penalized model is tighter than the estimation (3.2) of the ℓ_1- and ℓ_{1-2}-penalized models in the sense of (4.5). We see in the left plot of Fig. 8 that the error of the DCA-springback is less than the others for small s, and it becomes larger than the others when s exceeds some constant. The right plot also indicates that the error of the DCA-springback is less than the others when m is relatively small.

6.5. Remarks on numerical results

As observed in the literature, recovery results by different models may vary for different scenarios, and no one can unanimously outperform all the others for all scenarios. For instance, the IRLS-ℓ_p prevails in the incoherent regime but quickly fails in the coherent regime, see [28,45]. For incoherent sensing matrices, the IRLS-ℓ_p and the DCA-TL1 perform better than the DCA-ℓ_{1-2} and the ADMM-ℓ_1, while the DCA-ℓ_{1-2} performs the best for coherent sensing matrices; see [45,48]. The DCA-TL1 is robust, and it performs well for both incoherent and coherent sensing matrices, while it is less efficient than either the IRLS-ℓ_p in the incoherent regime or the DCA-ℓ_{1-2} in the coherent regime.

Together with these known facts and our numerical observations, we have the following remarks on the numerical performance of the DCA-springback.

- For an incoherent sensing matrix: the DCA-springback performs slightly worse than the IRLS-ℓ_p and the DCA-TL1;
- For a coherent sensing matrix: the DCA-springback performs slightly worse than the DCA-ℓ_{1-2} but better than the DCA-TL1.
- For a sensing matrix with modest coherence: the DCA-springback performs comparably with the DCA-MCP, and they perform better than the others.

Similar comparison results are also observed when the measurements are contaminated by some noise. For all the three scenarios, the DCA-springback and the DCA-MCP perform comparably if the parameter μ of the MCP is set as $1/\alpha$, and their performances with well-tuned parameters are also comparable. Moreover, we see that only the DCA-springback, the DCA-MCP, and the DCA-TL1 are robust with respect to the coherence of the sensing matrix. The DCA-springback and the DCA-MCP perform better than the DCA-TL1 in the coherent regime but worse in the incoherent regime. When the coherence of the sensing matrix is unknown, for example, when the sensing hardware cannot be modified or upgraded, coherence-robust algorithms such as the DCA-springback and the DCA-MCP are preferred for signal recovery.

7. Conclusion

We proposed a weakly convex penalty, named the springback penalty, for signal recovery from incomplete and inaccurate measurements. The springback penalty inherits major theoretical and numerical advantages from the convex ℓ_1 penalty and its various non-convex alternatives. We established exact and stable recovery results for the springback-penalized model (1.5) under the same RIP condition as the BP model (1.3); both the sparse and nearly sparse signals are considered. The springback-penalized model (1.5) is particularly suitable for signal recovery with a large level of noise or a limited number of measurements. We verified the effectiveness of the model and its computational tractability. The springback penalty provides a new tool to construct effective models for various sparsity-driven recovery problems arising in many areas such as compressed sensing, signal processing, image processing, and least-squares approximation.

References

[1] A. Beck, First-Order Methods in Optimization, SIAM, Philadelphia, 2017, Mathematical Optimization Society, Philadelphia.
[2] T. Blumensath, M.E. Davies, Iterative hard thresholding for compressed sensing, Appl. Comput. Harmon. Anal. 27 (2009) 265–274.
[3] A.M. Bruckstein, D.L. Donoho, M. Elad, From sparse solutions of systems of equations to sparse modeling of signals and images, SIAM Rev. 51 (2009) 34–81.
[4] E.J. Candès, J. Romberg, T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inf. Theory 52 (2006) 489–509.
[5] E.J. Candès, J.K. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements, Commun. Pure Appl. Math. 59 (2006) 1207–1223.
[6] E.J. Candès, M. Rudelson, T. Tao, R. Vershynin, Error correction via linear programming, in: 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05), IEEE, 2005, pp. 668–681.
[7] E.J. Candès, T. Tao, Decoding by linear programming, IEEE Trans. Inf. Theory 51 (2005) 4203–4215.
[8] E.J. Candès, T. Tao, Near-optimal signal recovery from random projections: universal encoding strategies?, IEEE Trans. Inf. Theory 52 (2006) 5406–5425.
[9] R. Chartrand, Exact reconstruction of sparse signals via nonconvex minimization, IEEE Signal Process. Lett. 14 (2007) 707–710.
[10] R. Chartrand, V. Staneva, Restricted isometry properties and nonconvex compressive sensing, Inverse Probl. 24 (2008) 035020.
[11] C. Chen, B. He, Y. Ye, X. Yuan, The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent, Math. Program. 155 (2016) 57–79.
[12] S.S. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit, SIAM Rev. 43 (2001) 129–159.
[13] D.L. Donoho, Compressed sensing, IEEE Trans. Inf. Theory 52 (2006) 1289–1306.
[14] E. Esser, Y. Lou, J. Xin, A method for finding structured sparse solutions to nonnegative least squares problems with applications, SIAM J. Imaging Sci. 6 (2013) 2010–2046.
[15] J. Fan, R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, J. Am. Stat. Assoc. 96 (2001) 1348–1360.
[16] S. Foucart, Hard thresholding pursuit: an algorithm for compressive sensing, SIAM J. Numer. Anal. 49 (2011) 2543–2563.

[17] S. Foucart, M.-J. Lai, Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leq 1$, Appl. Comput. Harmon. Anal. 26 (2009) 395–407.

[18] S. Foucart, H. Rauhut, A Mathematical Introduction to Compressive Sensing, Applied and Numerical Harmonic Analysis, Birkhäuser, Basel, 2013.

[19] H.-Y. Gao, A.G. Bruce, WaveShrink with firm shrinkage, Stat. Sin. (1997) 855–874.

[20] R. Glowinski, A. Marrocco, Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité, d'une classe de problèmes de Dirichlet non linéaires, Rev. Fr. Autom. Inform. Rech. Opér., Anal. Numér. 9 (1975) 41–76.

[21] K. Guo, D. Han, X. Yuan, Convergence analysis of Douglas–Rachford splitting method for "strongly+weakly" convex programming, SIAM J. Numer. Anal. 55 (2017) 1549–1577.

[22] B. He, X. Yuan, On the $O(1/n)$ convergence rate of the Douglas–Rachford alternating direction method, SIAM J. Numer. Anal. 50 (2012) 700–709.

[23] J. Huang, Y. Jiao, Y. Liu, X. Lu, A constructive approach to L_0 penalized regression, J. Mach. Learn. Res. 19 (2018) 403–439.

[24] M. Huang, M.-J. Lai, A. Varghese, Z. Xu, On DC based methods for phase retrieval, in: Approximation Theory XVI, Springer, Cham, 2021, pp. 87–121.

[25] Y. Jiao, B. Jin, X. Lu, A primal dual active set with continuation algorithm for the ℓ^0-regularized optimization problem, Appl. Comput. Harmon. Anal. 39 (2015) 400–426.

[26] Y. Jiao, B. Jin, X. Lu, Iterative soft/hard thresholding with homotopy continuation for sparse recovery, IEEE Signal Process. Lett. 24 (2017) 784–788.

[27] M.-J. Lai, J. Wang, An unconstrained ℓ_q minimization with $0 < q \leq 1$ for sparse solution of underdetermined linear systems, SIAM J. Optim. 21 (2011) 82–101.

[28] M.-J. Lai, Y. Xu, W. Yin, Improved iteratively reweighted least squares for unconstrained smoothed ℓ_q minimization, SIAM J. Numer. Anal. 51 (2013) 927–957.

[29] J. Lv, Y. Fan, A unified approach to model selection and sparse recovery using regularized least squares, Ann. Stat. 37 (2009) 3498–3528.

[30] S. Mendelson, A. Pajor, N. Tomczak-Jaegermann, Reconstruction and subgaussian operators in asymptotic geometric analysis, Geom. Funct. Anal. 17 (2007) 1248–1282.

[31] T. Möllenhoff, E. Strekalovskiy, M. Moeller, D. Cremers, The primal-dual hybrid gradient method for semiconvex splittings, SIAM J. Imaging Sci. 8 (2015) 827–857.

[32] J.-J. Moreau, Proximité et dualité dans un espace hilbertien, Bull. Soc. Math. Fr. 93 (1965) 273–299.

[33] D. Needell, J.A. Tropp, CoSaMP: iterative signal recovery from incomplete and inaccurate samples, Appl. Comput. Harmon. Anal. 26 (2009) 301–321.

[34] Y.C. Pati, R. Rezaiifar, P.S. Krishnaprasad, Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition, in: Proceedings of 27th Asilomar Conference on Signals, Systems and Computers, IEEE, 1993, pp. 40–44.

[35] R.T. Rockafellar, Convex Analysis, Princeton Mathematical Series, vol. 28, Princeton University Press, Princeton, 1970.

[36] M. Rudelson, R. Vershynin, On sparse reconstruction from Fourier and Gaussian measurements, Commun. Pure Appl. Math. 61 (2008) 1025–1045.

[37] R. Saab, R. Chartrand, O. Yilmaz, Stable sparse approximations via nonconvex optimization, in: 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2008, pp. 3885–3888.

[38] Y. Sun, H. Chen, J. Tao, Sparse signal recovery via minimax-concave penalty and ℓ_1-norm loss function, IET Signal Process. 12 (2018) 1091–1098.

[39] P.D. Tao, L.T.H. An, Convex analysis approach to DC programming: theory, algorithms and applications, Acta Math. Vietnam. 22 (1997) 289–355.

[40] P.D. Tao, L.T.H. An, A DC optimization algorithm for solving the trust-region subproblem, SIAM J. Optim. 8 (1998) 476–505.

[41] R.H. Todd, D.K. Allen, L. Alting, Manufacturing Processes Reference Guide, Industrial Press, Inc., New York, 1994.

[42] L.N. Trefethen, D. Bau III, Numerical Linear Algebra, SIAM, Philadelphia, 1997.

[43] J. Woodworth, R. Chartrand, Compressed sensing recovery via nonconvex shrinkage penalties, Inverse Probl. 32 (2016) 075004.

[44] L. Yan, Y. Shin, D. Xiu, Sparse approximation using $\ell_1 - \ell_2$ minimization and its application to stochastic collocation, SIAM J. Sci. Comput. 39 (2017) A229–A254.

[45] P. Yin, Y. Lou, Q. He, J. Xin, Minimization of ℓ_{1-2} for compressed sensing, SIAM J. Sci. Comput. 37 (2015) A536–A563.

[46] C.-H. Zhang, Nearly unbiased variable selection under minimax concave penalty, Ann. Stat. 38 (2010) 894–942.

[47] S. Zhang, J. Xin, Minimization of transformed L_1 penalty: closed form representation and iterative thresholding algorithms, Commun. Math. Sci. 15 (2017) 511–537.

[48] S. Zhang, J. Xin, Minimization of transformed L_1 penalty: theory, difference of convex function algorithm, and robust application in compressed sensing, Math. Program. 169 (2018) 307–336.

[49] T. Zhang, Analysis of multi-stage convex relaxation for sparse regularization, J. Mach. Learn. Res. 11 (2010) 1081–1107.

[50] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, J. R. Stat. Soc., Ser. B, Stat. Methodol. 67 (2005) 301–320.