L_1-norm Regularization for Short-and-sparse Blind Deconvolution: Point Source Separability and Region Selection*

Weixi Wang[†], Ji Li[‡], and Hui Ji[§]

Abstract. Blind deconvolution is about estimating both the convolution kernel and the latent signal from their convolution. Many blind deconvolution problems have a short-and-sparse (SaS) structure, *i.e.* the signal (or its gradient) is sparse and the kernel size is much smaller than the signal size. While ℓ_1-norm relating regularizations have been widely used for solving SaS blind deconvolution problems, the so-called region/edge selection technique brings great empirical improvement to such ℓ_1-norm relating regularizations in image deblurring. The essence of region/edge selection is during an alternative iterative scheme of SaS blind deconvolution, one estimates the kernel on an estimate of the latent image with well-separated image edges instead of the one with the least fitting error. In this paper, we first examines the validity and soundness of ℓ_1-norm relating regularization in the setting of 1D SaS blind deconvolution. The analysis reveals the importance of the separation of non-zero signal entries toward the soundness of such a regularization. The studies laid out the foundation of region selection technique, *i.e.*, during the iteration, an estimate of the latent image with well-separated edges is a better candidate for estimating the kernel than the one with least fitting error. Based on the studies conducted in this paper, an alternating iterative scheme with region selection model is developed for SaS blind deconvolution, which is then applied on blind motion deblurring. The experiments showed its effectiveness over many existing ℓ_1-norm relating approaches.

Key words. Blind deconvolution, L_1 regularization, Sparse-and-short structure

AMS subject classifications. 68U10, 94A08

1. Introduction.

One often-seen signal degradation in practice is blurring, which attenuates or erases high frequencies of signal during acquisition. The relation between the recorded signal $\boldsymbol{b}$ and the true signal $\boldsymbol{x}_0$ usually is modeled by a convolution process:

$$\boldsymbol{b} = \boldsymbol{a}_0 \otimes \boldsymbol{x}_0 + \boldsymbol{n}, \tag{1.1}$$

where $\boldsymbol{a}_0$ denotes a smoothing kernel (low-pass filter), $\boldsymbol{n}$ denotes measurement noise, and $\otimes$ denotes discrete circular convolution operator. If the kernel $\boldsymbol{a}_0$ is known in advance, solving (1.1) is called *non-blind deconvolution*. If both the kernel $\boldsymbol{a}_0$ and the signal $\boldsymbol{x}_0$ are unknown, solving (1.1) is called *blind deconvolution*. While non-blind deconvolution focuses on how to suppress noise amplification when deconvolving the signal, blind deconvolution focuses on how to estimate smoothing kernel $\boldsymbol{a}_0$. Once the kernel $\boldsymbol{a}_0$ is estimated, the problem of blind deconvolution is reduced to the problem of non-blind deconvolution.

Blind deconvolution is an important problem seen in a wide range of applications, including astronomical imaging [16, 9, 32], microscopy imaging [14, 10], and digital image photography [12, 3, 30, 21]. In general, blind deconvolution is an ill-posed problem with many

*Submitted to the editors February 8, 2022.
Funding: This work was funded by xx.
[†]Department of Mathematics, National University of Singapore, Singapore (wangweixi@u.nus.edu).
[‡]Department of Mathematics, National University of Singapore, Singapore (matliji@nus.edu.sg).
[§]Department of Mathematics, National University of Singapore, Singapore (matjh@nus.edu.sg).

solutions. Certain priors need to be imposed on both $\boldsymbol{a}$ and $\boldsymbol{x}$ to address possible degenerate solution. Once the priors in image/kernel are determined, blind deconvolution can then be formulated as solving an optimization problem:

$$\min_{\boldsymbol{a}\in\Omega,\boldsymbol{x}} \phi(\boldsymbol{x}) + \lambda\psi(\boldsymbol{a}), \quad \text{subject to} \quad f(\boldsymbol{b} - \boldsymbol{a}\otimes\boldsymbol{x}) \le \epsilon, \tag{1.2}$$

or its regularized variational form

$$\min_{\boldsymbol{a}\in\Omega,\boldsymbol{x}} f(\boldsymbol{b} - \boldsymbol{a}\otimes\boldsymbol{x}) + \lambda_1\phi(\boldsymbol{x}) + \lambda_2\psi(\boldsymbol{a}),$$

where Ω denotes the feasible set for the kernel, and the function $f(\cdot)$ denotes the fidelity term determined by noise $\boldsymbol{n}$, *i.e.* $f(\cdot) = \|\cdot\|_1$ or $\|\cdot\|_2^2$. In the optimization models above, there are two terms $\phi(\cdot)$ and $\psi(\cdot)$, the regularization terms on $\boldsymbol{x}$ and $\boldsymbol{a}$ derived from their corresponding priors. The feasible set Ω for kernel comes from the physics of signal acquisition systems. Taking optical imaging systems for example, the set Ω is often defined as the follows

$$\Omega = \{\boldsymbol{a}\in\mathbb{R}^n \mid \boldsymbol{a}\ge 0, \|\boldsymbol{a}\|_1 = 1\}. \tag{1.3}$$

The model (1.2) or its regularized variational form is widely used in many existing blind image deblurring methods; see *e.g.* [23, 22, 4, 33, 27, 35] for more details.

This paper concerns blind deconvolution of signals with the *short-and-sparse* (SaS) structure (see *e.g.* [19, 20]):

- The effective size of true kernel $\boldsymbol{a}_0$ is much less than that of true signal $\boldsymbol{x}_0$.
- true signal $\boldsymbol{x}_0$ is sparse with most entries being zero or close to zero.

Such an SaS structure exists in many blind deconvolution problems, including blind image deblurring in digital photography. In such application, the blurring is modeled by $\boldsymbol{B} = \boldsymbol{a}\otimes\boldsymbol{I}$, where $\boldsymbol{I}/\boldsymbol{B}$ are the clean/blurred images respectively. In certain cases, the natural image $\boldsymbol{I}$ itself is not sparse, but its gradient $\nabla\boldsymbol{I}$ is sparse. Recall that the discrete implementation of $\nabla\boldsymbol{I}$ can be formulated as the convolutions between $\boldsymbol{I}$ and the high-pass filter $[1,-1]$. By the commutative property of convolution, we have $\boldsymbol{b} = \nabla\boldsymbol{B} = \boldsymbol{a}\otimes(\nabla\boldsymbol{I})$. As the aim of blind deconvolution is for kernel estimation, one can recast the problem to the one with SaS structure by solving the problem in the gradient domain, *i.e.*

$$\min_{\boldsymbol{a}\in\Omega,\nabla\boldsymbol{I}} f(\nabla\boldsymbol{B} - \boldsymbol{a}\otimes\nabla\boldsymbol{I}) + \lambda_1\phi(\nabla\boldsymbol{I}) + \lambda_2\psi(\boldsymbol{a}).$$

Once $\boldsymbol{a}$ is determined by solving the problem above, one can then switch back to estimate $\boldsymbol{I}$ by deconvolving $\boldsymbol{B}$ in image domain. Such a practice is widely used in many existing blind image deblurring methods.

In last decade, motivated by practical needs, there has been rapid progress on the development of blind deconvolution methods, especially in the domain of blind motion deblurring. See *e.g.* [23, 22, 4, 33, 27, 35]. Most of these methods require solving a non-convex problem. While these methods demonstrated good performance and empirical stability in practice, theoretical understanding and mathematical soundness of these methods are scant in existing literature. For example, it is not clear under what condition, the true kernel/signal is indeed one global minimum of the model (1.2). This paper aims to analyze one type of ℓ_1-norm relating regularization model (1.2) to provide a clearer picture of its soundness when being used for solving blind SaS deconvolution problems. Consider a blind SaS problem where the signal $\boldsymbol{x} \in \mathbb{R}^n$. Let $\mathbb{R}^n_k$ denote the subspace of $\mathbb{R}^n$ where all vectors are supporting on the first k entries. We assume the kernel $\boldsymbol{a} \in \mathbb{R}^n_k$. The following model is considered in this paper for solving (1.1):

$$\min_{(\boldsymbol{a},\boldsymbol{x})\in\mathcal{S}^{\epsilon}_{\boldsymbol{b}}} \quad \|\boldsymbol{x}\|_1 + \nu \|\boldsymbol{a}\|_2^2, \tag{1.4}$$

where

$$\mathcal{S}^{\epsilon}_{\boldsymbol{b}} = \{(\boldsymbol{a} \in \mathbb{R}^n_k, \boldsymbol{x} \in \mathbb{R}^n) \mid \|\boldsymbol{a} \otimes \boldsymbol{x} - \boldsymbol{b}\|_1 \leq \epsilon, \boldsymbol{a} \geq 0, \|\boldsymbol{a}\|_1 = 1\}. \tag{1.5}$$

The bound ϵ is determined by measurement noise $\boldsymbol{n}$, *i.e.* $\epsilon \geq \|\boldsymbol{n}\|_1$.

In the model (1.4), the ℓ_1-norm regularization for the signal $\boldsymbol{x}$, $\phi(\boldsymbol{x}) = \mu\|\boldsymbol{x}\|_1$, is a widely used convex function for prompting sparsity of signal; see *e.g.* [7, 5, 29, 36]. There are two terms related to the kernel $\boldsymbol{a}$. One comes from the feasible set $\mathcal{S}^{\epsilon}_{\boldsymbol{b}}$, which is determined by the physics of many signal acquisition systems, especially optics-based imaging. Two most general physical constraints in these signal acquisition systems are (1) non-negativity constraint $\boldsymbol{a} \geq 0$; and (2) normalization constraint $\|\boldsymbol{a}\|_1 = 1$. The squared ℓ_2-norm relating regularization for smoothing kernel, $\psi(\boldsymbol{a}) = \nu\|\boldsymbol{a}\|_2^2$, comes from the fact that $\boldsymbol{a}$ is a band-limited filter supported only in low-frequency domain. The term $\psi(\boldsymbol{a}) = \nu\|\boldsymbol{a}\|_2^2$ is also critical to avoid scenario where the trivial no-blur pair $(\boldsymbol{\delta}, \boldsymbol{a}_0 \otimes \boldsymbol{x}_0)$ has a lower cost than that of the true pair $(\boldsymbol{a}_0, \boldsymbol{x}_0)$, where $\boldsymbol{\delta}$ denotes Dirac delta. Note that the no-blur pair $(\boldsymbol{\delta}, \boldsymbol{a}_0 \otimes \boldsymbol{x}_0)$ satisfies $\boldsymbol{\delta} \otimes (\boldsymbol{a}_0 \otimes \boldsymbol{x}_0) = \boldsymbol{a}_0 \otimes \boldsymbol{x}_0$, and for a non-negative kernel $\boldsymbol{a}$ with $\|\boldsymbol{a}\|_1 = 1$,

$$\|\boldsymbol{a} \otimes \boldsymbol{x}\|_1 \leq \|\boldsymbol{a}\|_1 \|\boldsymbol{x}\|_1 = \|\boldsymbol{x}\|_1.$$

In other words, the true pair is unlikely to be a global minimum of the model (1.4) with $\nu = 0$. It is shown in [27, 33] that, without such regularization term $\psi(\boldsymbol{a})$, the solution of the model is biased to the kernel closer to the Dirac Delta $\boldsymbol{\delta}$. The ℓ_1 norm is used as the fidelity metric in this paper, as it is widely used in practical image deblurring methods for its robustness to outliers. Before proceeding, we briefly introduce some notations used in the discussions.

Notations. We use bold font to denote vector and, without of specification, the indices of vector $\boldsymbol{a} \in \mathbb{R}^n$ is $\{0, 1, \ldots, n-1\}$. The space $\mathbb{R}^n_k = \{\boldsymbol{a} \in \mathbb{R}^n \mid \boldsymbol{a}[j] = 0, k \leq j \leq n-1\}$ denotes the space of all n-dimensional vectors with their support on the first k entries. Given a vector $\boldsymbol{v} \in \mathbb{R}^n$, we denote $\mathcal{S}_\tau(\boldsymbol{v})$ the cyclic shift of the vector $\boldsymbol{v}$ by τ entries: $\mathcal{S}_\tau(\boldsymbol{v})(i) = \boldsymbol{v}([i-\tau]_n)$, here n is the length of the vector, $[i]_n$ denotes the modulo with respect to n. Taking sign$(\cdot)$ on a vector means taking sign$(\cdot)$ point-wisely on this vector.

Organization. We review the related works in Section 1.1. We present the main results of this paper in Section 1.2. The analysis of the model (1.4) with detailed proofs is presented in Section 2. The model (1.4) is discussed in Section 2.1 and 2.2. Finally Section 2.3 is devoted to a novel region selection approach motivated by our theoretical results and algorithm for blind SaS deconvolution with application to blind image deblurring. Section 3 shows the experiments to validate the effectiveness of our region selector within the proximal alternative minimization framework for blind image deblurring. Section 4 provides the conclusion.

1.1. Related works. The optimization problem (1.4) is a non-convex problem with a complicated optimization landscape. Some non-critical solution ambiguities are addressed in the feasible set $\mathcal{S}_{\boldsymbol{b}}^{\epsilon}$ in (1.4). For instance, there is the so-called *shift ambiguity*: The pair $(\mathcal{S}_{-\tau}(\boldsymbol{a_0}), \mathcal{S}_{\tau}(\boldsymbol{x}_0))$ has the same convolution as the true pair $(\boldsymbol{a}_0, \boldsymbol{x}_0)$. There is also the so-called *scale ambiguity*: The pair $(s\boldsymbol{a_0}, \frac{1}{s}\boldsymbol{x}_0)$ has the same convolution as $(\boldsymbol{a}_0, \boldsymbol{x}_0)$. The shift ambiguity is addressed by $\boldsymbol{a} \in \mathbb{R}_k^n$ and the scale ambiguity is addressed by $\|a\|_1 = 1$. These types of ambiguities are not critical, shift ambiguity only causes the resulting signal be a shifted version of the true signal, and the scale ambiguity does not modify the pattern of the signal. Besides these non-critical ambiguities, in general, there are additional global minima and a variety of critical points of the non-convex problem (1.4), which are away from the truth.

Recently, there have been quite a few impressive works on studying provable algorithms for blind deconvolution. Ahmed *et al.* [1] recast the problem of blind deconvolution to a linear inverse problem on rank-1 matrix [1, 25, 24], as an extension to the convexification for phase retrieval [6]. Consider two vectors $\boldsymbol{x}_0$ and $\boldsymbol{a}_0$, its outer product $\boldsymbol{a}_0\boldsymbol{x}_0^*$ is a rank-1 matrix, and their convolution $\boldsymbol{b}$ can be expressed as $\mathcal{A}(\boldsymbol{a}_0\boldsymbol{x}_0^*)$ for some linear mapping $\mathcal{A}$. It is shown in [1] that such a nuclear-norm-based convex model can exactly recover the pair $(\boldsymbol{a}_0, \boldsymbol{x}_0)$ (up to a scale), if the pairs follow the following configuration: The signal $\boldsymbol{x}_0$ is drawn from a random subspace, and $\boldsymbol{a}_0$ is a vector in a subspace whose basis vectors are "flat" in its frequency domain. Based the same lifting-based formulation, Li *et al.* [25] examined various configurations of $\boldsymbol{a}_0$ and $\boldsymbol{x}_0$ for identifiability and stability of the problem, including subspace constraints for $\boldsymbol{x}_0, \boldsymbol{a}_0$, sparse constraints and the mixture of both. It is shown in [25] that, up to a set of zero measure, the pair $(\boldsymbol{a}_0, \boldsymbol{x}_0)$ is identifiable up to a scale. Despite its theoretical soundness, the lifting scheme does not scale well as the dimension of the matrix for recovery is the multiplication of signal dimension and kernel dimension.

Li *et al.* [24] considered a slightly different configuration, where the signal $\boldsymbol{x}$ is drawn from a random subspace spanned by the columns of a Gaussian matrix, and the kernel $\boldsymbol{a}$ is short and has small correlation (coherence). Under such a configuration, a provable regularized gradient descent algorithm is proposed [24] for blind deconvolution with the provable convergence to the true pair with a large probability. Nevertheless, the assumptions imposed on signals and kernels in [24] do not hold for many practical scenarios, especially blind deblurring of natural images. The performance of such a method is also not competitive to many existing regularization methods for blind image deblurring. Several works [38, 37, 20] insists the sparsity of signal is w.r.t. the natural basis, while they considered the ℓ_2 normalization constraint on $\boldsymbol{a}$ for its smoothness over ℓ_1 normalization constraint. Zhang *et al.* [38] studied the geometry of ℓ_1-norm regularization model over ℓ_2 sphere in the case where the signal $\boldsymbol{x}_0 = \delta$. It is shown in [38] that for ℓ_1-norm relating regularization model, all strict local minima of the model are close to signed shift truncations of $\boldsymbol{a}_0$. The same structured local minima can be obtained by replacing ℓ_1-norm by another sparsity-prompting function $-\|\cdot\|_4^4$, as shown in [37]. In the figuration that the convolution erases little information of signal, *i.e.*, $\|\boldsymbol{a}_0 \otimes \boldsymbol{x_0}\|_2 \approx \|\boldsymbol{x}_0\|_2$, Kuo *et al.* [20] presented an algorithm which can guarantee exact recovery of an incoherent kernel $\boldsymbol{a}_0$ and sparse $\boldsymbol{x}_0$. While the works discussed above provide good insights for ℓ_1-norm relating method for blind SaS deconvolution and computational algorithm with recovery guarantee, the discussions focus on noiseless observation and the configuration where the smoothing effect is relative little, *i.e.* kernel is close to Dirac Delta or $\|\boldsymbol{a}_0 \otimes \boldsymbol{x_0}\|_2 \approx \|\boldsymbol{x}_0\|_2$.

1.2. Main results. Different from the models and assumptions of the existing works, this paper studies the validity of model (1.4) for blind SaS deconvolution where there might be significant frequency attenuation on true signal during convolution. In the setting of this paper, the input measurement $\boldsymbol{b} \in \mathbb{R}^n$ is formulated as

$$\boldsymbol{b} = \boldsymbol{a}_0 \otimes \boldsymbol{x}_0 + \boldsymbol{n}, \quad \text{where } \boldsymbol{x}_0 \in \mathbb{R}^n,\ \boldsymbol{a}_0 \in \mathbb{R}^n_k, \boldsymbol{a}_0 \geq 0, \|\boldsymbol{a}_0\|_1 = 1, \tag{1.6}$$

where $\boldsymbol{x}_0 \in \mathbb{R}^n$ denotes true sparse signal and $\boldsymbol{a}_0 \in \mathbb{R}^n_k$, the convolution is defined as the circular convolution

$$(\boldsymbol{a} \otimes \boldsymbol{x})[j] = \sum_{i=1}^{n} \boldsymbol{a}[i]\boldsymbol{x}[(j-i) \bmod n].$$

The model (1.4) is used for estimating the kernel $\boldsymbol{a}$ from (1.6). Recall that model (1.4) is different from what has been studied for SaS blind deconvolution [38, 37, 20] on two parts related to $\boldsymbol{a}$:

- The term $\nu\|\boldsymbol{a}\|_2^2$ for the corresponding minimum biased to a band-limited filter.
- The non-negativity constraint, $\boldsymbol{a} \geq 0$, which holds true for many signal acquisition systems.

The problem (1.4) is an optimization problem with convex objective function and non-convex constraints. As the focus of blind SaS deconvolution is on the estimation of the kernel $\boldsymbol{a}_0$, this paper aims at analyzing the soundness and well-posedness of the model (1.4) for estimating $\boldsymbol{a}_0$. In our study, instead of using ℓ_2-norm for error measurement of kernel estimation, we measures the estimation error by the correlation between the true kernel $\boldsymbol{a}_0$ and the estimation $\boldsymbol{a}$ up to a circled translation [15, 38]:

$$C(\boldsymbol{a}_0, \boldsymbol{a}) = \max_{0 \leq i \leq k} \frac{|\langle \boldsymbol{a}_0, \mathcal{S}_i(\boldsymbol{a})\rangle|}{\|\boldsymbol{a}_0\|_2 \, \|\boldsymbol{a}\|_2}, \quad \text{where } \mathcal{S}_i(\cdot) \text{ is the translation operator.} \tag{1.7}$$

Such a metric handles the translation ambiguity of kernel estimation which does not impact the information of the recovered signal. The soundness of the model (1.4) for estimating $\boldsymbol{a}_0$ is closely related to the separation of non-zero entries of the sparse signal $\boldsymbol{x}_0$.

Definition 1.1. *A signal $\boldsymbol{x} \in \mathbb{R}^n$ is* k-separable, *if its support satisfies*

$$\min_{i \neq j; i,j \in \operatorname{supp}(\boldsymbol{x})} |(i-j) \bmod n| \geq k, \tag{1.8}$$

where $\operatorname{supp}(\boldsymbol{x}) = \{\ell : \boldsymbol{x}[\ell] \neq 0\}$.

For a blind SaS deconvolution problem where the signal is k-separable and the measurement is noise-free, we have the following results regarding the model (1.4) used for solving the SaS blind deconvolution problem:

Theorem 1.2 (k-separable signal with noise-free measurement). *Consider a non-zero signal $\boldsymbol{x}_0$ and its noise-free measurement $\boldsymbol{b}$ defined by $\boldsymbol{b} = \boldsymbol{a}_0 \otimes \boldsymbol{x}_0$. If $\boldsymbol{x}_0$ is k-separable, then the true pair $(\boldsymbol{a}_0, \boldsymbol{x}_0)$ is a global minimum of the model (1.4) with the feasible set $\mathcal{S}^0_{\boldsymbol{b}}$ defined by (1.5) with $\epsilon = 0$, for $0 < \nu \leq \frac{\|\boldsymbol{x}_0\|_1}{2\|\boldsymbol{a}_0\|_2^2}$.*

It can be seen from Theorem 1.2 that as long as the minimal distance between two non-zero entries of the signal is no less than the support size of the kernel, the model admits the truth, up to a shift, as its one global minimizer. However, it is not necessarily the unique global minimizer. A stronger separation condition is needed for the model (1.4) to admit the truth, up to a shift, as the unique global minimizer.

In the next, we establish a sufficient condition for guaranteeing the truth, up to a shift, is the unique minimizer of the model (1.4). For a sparse signal $\boldsymbol{x}$, decompose it to the summation of two sparse signals:

$$\boldsymbol{x} = \boldsymbol{x}^+ + \boldsymbol{x}^-,$$

where

$$(1.9) \qquad \boldsymbol{x}^+[k] = \begin{cases} \boldsymbol{x}[k], & \boldsymbol{x}[k] > 0 \\ 0, & \boldsymbol{x}[k] \leq 0; \end{cases} \qquad \boldsymbol{x}^-[k] = \begin{cases} \boldsymbol{x}[k], & \boldsymbol{x}[k] < 0 \\ 0, & \boldsymbol{x}[k] \geq 0; \end{cases} \qquad \text{for} \quad k = 1, 2, \ldots, n.$$

Definition 1.3. *A signal* $\boldsymbol{x} \in \mathbb{R}^n$ *is* signed $2k$-separable*, if* $\boldsymbol{x}$ *is* k*-separable and* $\boldsymbol{x}^+, \boldsymbol{x}^-$ *defined by* (1.9) *are* $2k$ *separable.*

For a signed $2k$-separable signal, we have then

Theorem 1.4 (Signed $2k$-separable signal with noise-free measurement). *Consider a non-zero signal* $\boldsymbol{x}_0$ *and its non-zero noise-free measurement* $\boldsymbol{b}$ *defined by* $\boldsymbol{b} = \boldsymbol{a}_0 \otimes \boldsymbol{x}_0$*. If* $\boldsymbol{x}_0$ *is* k*-separable and* $\boldsymbol{x}_0^+, \boldsymbol{x}_0^-$ *defined by* (1.9) *are* $2k$*-separable, the set of all global minimum of the model* (1.4) *with* $0 < \nu \leq \frac{\|\boldsymbol{x}_0\|_1}{2\|\boldsymbol{a}_0\|_2^2}$ *is then* $\{(\mathcal{S}_i(\boldsymbol{a}_0), \mathcal{S}_{-i}(\boldsymbol{x}_0)) | -k < i < k, \mathcal{S}_i(\boldsymbol{a}_0) \in \mathbb{R}^n_k\}$.

For guaranteeing the soundness of the model (1.4), the signed $2k$-separation condition on the signal $\boldsymbol{x}_0$ stated in Theorem 1.4 appears to be very strong. While many signals in practice, *e.g.* natural images, do not satisfy such separation condition as a whole, one often can find certain regions of natural image where the signal can be well approximated by a signed $2k$-separable signal. The question is then if a sparse signal $\boldsymbol{x}$ can be well-approximated by a signed $2k$-separable signal, will the truth remains to be close to the global minimizer of the model (1.4) (up to a cyclic shift)?

Theorem 1.5 (Approximate signed $2k$-separable signal with noisy measurement). *Consider a non-zero sparse signal* $\boldsymbol{x}_0$ *and its noisy measurement* $\boldsymbol{b} = \boldsymbol{a_0} \otimes \boldsymbol{x}_0 + \boldsymbol{n}$ *with non-zero* $\boldsymbol{a_0} \otimes \boldsymbol{x}_0$ *and* $\|\boldsymbol{n}\|_1 \leq \epsilon$*. Suppose* $\boldsymbol{x}_{0,s2k}$ *is a nonzero signed 2k-separable approximation to* $\boldsymbol{x}_0$ *such that*

$$\boldsymbol{x}_0 = \boldsymbol{x}_{0,s2k} + \Delta \boldsymbol{x}_0.$$

Let $(\boldsymbol{a}^*, \boldsymbol{x}^*)$ *denote an optimal solution to the problem* (1.4) *with the feasible set* $\mathcal{S}_{\boldsymbol{b}}^{\epsilon}$*. Then, assume* $\frac{\|\Delta \boldsymbol{x}_0\|_1 + \epsilon}{\|\boldsymbol{x}_0\|_1} < 0.21$*, by setting* $\nu = \frac{\|\boldsymbol{x}_0\|_1}{2\|\boldsymbol{a}_0\|_2^2}$*, we have*

$$C(\boldsymbol{a}_0, \boldsymbol{a}^*) \geq 1 - 2\frac{\|\Delta \boldsymbol{x}_0\|_1}{\|\boldsymbol{x}_0\|_1} - \frac{\|\boldsymbol{a}_0\|_\infty}{\|\boldsymbol{a}_0\|_2^2} \frac{2\epsilon}{\|\boldsymbol{x}_0\|_1}.$$

It can be seen from Theorem 1.5 that, as long as noise is not significant and the signed $2k$ approximation residual to the signal $\boldsymbol{x}_0$, $\Delta \boldsymbol{x}_0$, is small, the global minimum of the model (1.4)

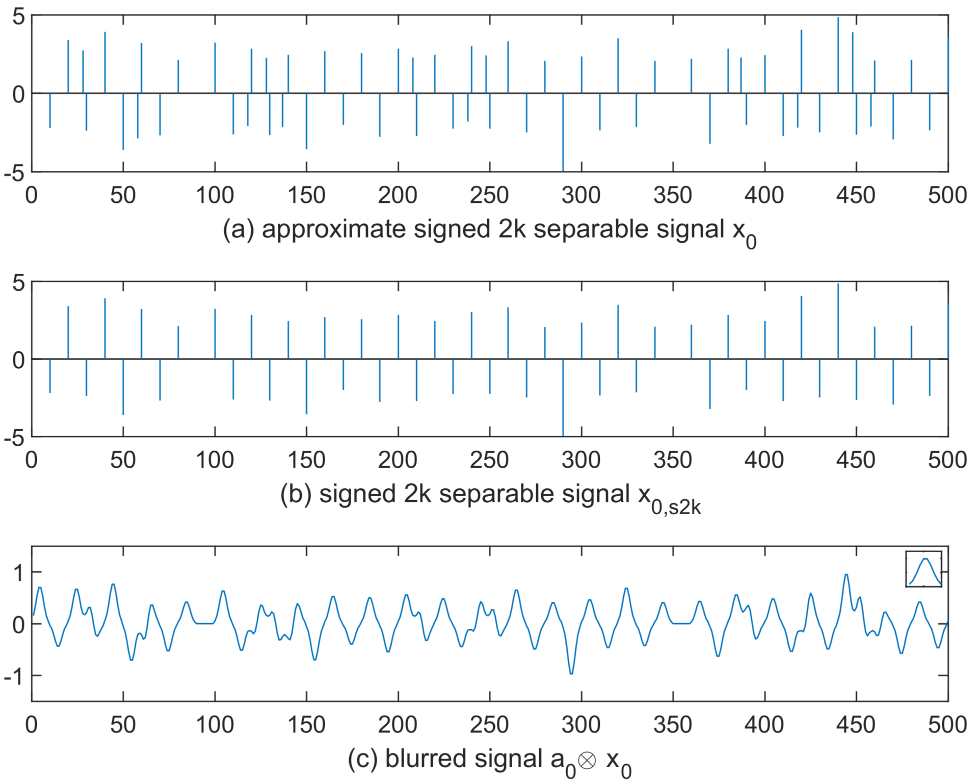


Figure 1. *A demonstration of approximate signed* $2k$*-separable signal and its blurred observation by a Gaussian kernel. (a) An approximately signed* 20*-separable signal* $\boldsymbol{x}_0$*; (b) The signed* 20*-separable approximation of* $\boldsymbol{x}_0$ *with approximation error:* $\frac{\|\Delta \boldsymbol{x}_0\|_1}{\|\boldsymbol{x}_0\|_1} \approx 0.16$*; (c) The blurred signal* $\boldsymbol{x}_0 \otimes \boldsymbol{a}_0$ *where* $\boldsymbol{a}_0$ *is a Gaussian blur kernel of size* 10 *shown in the top right corner of the graph.*

is close to the true kernel $\boldsymbol{a}_0$ (up to a cyclic shift). In other words, the model (1.4) is sound for an approximately signed $2k$-separable signal. The hyper-parameter ν plays an important role when using the model (1.4) to solve SaS blind deconvolution problem. The setting of ν in the above theorem depends on $\|\boldsymbol{a}_0\|_2^2$ and $\|\boldsymbol{x}_0\|_1$.

Remark 1.6. In Theorem 1.5, we assume $\boldsymbol{x}_0$ is non-zero. If $\boldsymbol{x}_0$ is zero, then in both noiseless and noisy cases, for $\nu > 0$, the optimal solution is $\boldsymbol{a}^* = [1/k, 1/k, ..., 1/k]$, $\boldsymbol{x}^* = [0, 0, ..., 0]$; for $\nu = 0$, the optimal solution is $\boldsymbol{x}^* = [0, 0, ..., 0]$, $\boldsymbol{a}^*$ can be any positive kernel satisfying $\|\boldsymbol{a}^*\|_1 = 1$.

Remark 1.7. For a signal which can be well-approximated by a signed $2k$-separable signal, one possible concern is whether the kernel with support size k can be trivially found from the blurred signal. See Figure 1 for an illustration. It can be seen that the local shapes of the signal are not consistent with the true kernel, a Gaussian kernel.

1.3. Region selection for blind SaS deconvolution. From the results shown in Section 1.2, the signed $2k$ separability condition on a sparse signal $\boldsymbol{x}$ is very strong, even in the approximate sense, for guaranteeing the soundness of the model (1.4). The distribution of non-zero entries of a sparse signal is not necessarily uniform. It is possible that the non-zero entries of a sparse signal are dense in some small regions. In other words, the model (1.4) can not guarantee a good estimation to the kernel $\boldsymbol{a}_0$ for general sparse signals. However, recall that the focus of blind SaS deconvolution is about kernel estimation, the recovery of the image in the model (1.4) is for the purpose of estimating blur kernel. One solution is then selecting the regions with small $2k$-separable approximation residual and using them to estimate the kernel.

The model (1.4) is usually solved via an iterative scheme that alternatively estimates the image and the kernel for the blind image deblurring. Motivated by Theorem 1.5, the deconvolution of enough separable signal is more faithful than a general sparse signal. There are two approaches to take such advantages into consideration in deconvolution algorithms. (1) During each iteration, once the estimation of the signal $\boldsymbol{x}$ is updated, we select those image regions $\{\boldsymbol{x}_k^{(t)}\}_k$ with small relative $2k$-separable approximation residual $\frac{\|\Delta \boldsymbol{x}_k\|_1}{\|\boldsymbol{x}_k\|_1}$. Then, these image regions are used as the image in the model (1.4) to estimate the kernel $\boldsymbol{a}$. (2) We first select a good region of the input blurred image, we say a region is good if the selected region is with the most possible enough separation. Comparing the two approaches, the latter approach is easier to implement and the size reduction of deblurring problem saves computational time for kernel estimation. To select several regions during iteration, we should take care of the boundary effect of possible region overlapping. In our experiments in this paper, we follow the second approach to estimate kernel with a selected good region from blurred image. Once a faithful $\boldsymbol{a}$ is obtained, we then deconvolve the whole image $\boldsymbol{x}$ using updated $\boldsymbol{a}$.

The region selection technique for blind SaS deconvolution is not completely new. It has been adopted in quite a few methods for blind image deblurring with different empirical strategies. Xu and Jia [34] demonstrated that edges of smaller size than the blur kernel may have adverse effect on kernel estimation, which coincides with our theoretical analysis. Hu and Yang [15] proposed a learning based region selection method. This method needs two stages: At stage one, they collect some blurred images with corresponding ground true kernel, then they separate each blurred image into several patches and estimate the blurring kernel using each patch. They use the correlation $C(\cdot,\cdot)$ to measure the similarity between the estimated kernel and the ground truth, and evaluate each patch using the kernel similarity. At stage two, they train a logistic function to predict the good region using the data prepared at stage one. Compared with them, this paper provides a region selection technique, which is easier to use and more computationally efficient, with strong mathematical motivations.

2. Main body.

In this section, we first prove Theorem 1.2 and Theorem 1.4, and construct some cases to show that the separation condition assumed on signal is necessary to guarantee the soundness of the model in these cases. Then we present a detailed proof of Theorem 1.5 and discuss its implification. At last, we present the region selection method inspired from our theorems.

2.1. Model (1.4) for noisy-free measurement. First we give a proof of Theorem 1.2 and 1.4.

Proof of Theorem 1.2 and 1.4. Suppose $(\boldsymbol{a}^*, \boldsymbol{x}^*)$ is an optimal solution to the model (1.4). We show that ground true pair $(\boldsymbol{a}_0, \boldsymbol{x}_0)$ attains the same minimum of (1.4) given by $\|\boldsymbol{x}^*\|_1 + \nu\|\boldsymbol{a}^*\|_2^2$. Then $(\boldsymbol{a}_0, \boldsymbol{x}_0)$ obviously belongs to the optimal solution set. As $\boldsymbol{x}_0$ is k-separable, we have

$$\|\boldsymbol{b}\|_1 = \|\boldsymbol{a}_0 \otimes \boldsymbol{x}_0\|_1 = \|\boldsymbol{a}_0\|_1 \|\boldsymbol{x}_0\|_1 = \|\boldsymbol{x}_0\|_1 .$$

From the inequality

$$\|\boldsymbol{a}^* \otimes \boldsymbol{x}^*\|_1 \le \|\boldsymbol{a}^*\|_1 \|\boldsymbol{x}^*\|_1 ,$$

we have $\|\boldsymbol{x}^*\|_1 \ge \|\boldsymbol{x}_0\|_1$. By the optimality of $(\boldsymbol{a}^*, \boldsymbol{x}^*)$, we have

$$\nu\|\boldsymbol{a}^*\|_2^2 + \|\boldsymbol{x}^*\|_1 \le \nu\|\boldsymbol{a}_0\|_2^2 + \|\boldsymbol{x}_0\|_1 ,$$

which implies $\|\boldsymbol{a}^*\|_2 \le \|\boldsymbol{a}_0\|_2$. Taking an inner product between $\boldsymbol{a}_0 \otimes \operatorname{sign}(\boldsymbol{x}_0)$ and the both sides of $\boldsymbol{a}^* \otimes \boldsymbol{x}^* = \boldsymbol{a}_0 \otimes \boldsymbol{x}_0$, we have

$$\langle \boldsymbol{a}_0 \otimes \operatorname{sign}(\boldsymbol{x}_0), \boldsymbol{a}^* \otimes \boldsymbol{x}^* \rangle = \langle \boldsymbol{a}_0 \otimes \operatorname{sign}(\boldsymbol{x}_0), \boldsymbol{a}_0 \otimes \boldsymbol{x}_0 \rangle.$$

Writing $\boldsymbol{a}^* \otimes \boldsymbol{x}^*$ as $\sum_i \boldsymbol{x}_i^* \mathcal{S}_i(\boldsymbol{a}^*)$, we have

$$\langle \boldsymbol{a}_0 \otimes \operatorname{sign}(\boldsymbol{x}_0), \boldsymbol{a}^* \otimes \boldsymbol{x}^* \rangle = \sum_i \boldsymbol{x}_i^* \langle \boldsymbol{a}_0 \otimes \operatorname{sign}(\boldsymbol{x}_0), \mathcal{S}_i(\boldsymbol{a}) \rangle \le \sum_i |\boldsymbol{x}_i^*| |\langle \boldsymbol{a}_0 \otimes \operatorname{sign}(\boldsymbol{x}_0), \mathcal{S}_i(\boldsymbol{a}) \rangle|.$$

When $\boldsymbol{x}_0$ is k-separable, we have the inequality

$$|\langle \boldsymbol{a}_0 \otimes \operatorname{sign}(\boldsymbol{x}_0), \mathcal{S}_i(\boldsymbol{a}^*) \rangle| \le \|\boldsymbol{a}^*\|_2 \|\boldsymbol{a}_0\|_2 ,$$

and the equality

$$\langle \boldsymbol{a}_0 \otimes \operatorname{sign}(\boldsymbol{x}_0), \boldsymbol{a}_0 \otimes \boldsymbol{x}_0 \rangle = \|\boldsymbol{a}_0\|_2^2 \|\boldsymbol{x}_0\|_1 .$$

Therefore,

$$\|\boldsymbol{a}^*\|_2 \|\boldsymbol{a}_0\|_2 \|\boldsymbol{x}^*\|_1 \ge \|\boldsymbol{a}_0\|_2^2 \|\boldsymbol{x}_0\|_1 ,$$

which means $\|\boldsymbol{x}^*\|_1 \ge \frac{\|\boldsymbol{a}_0\|_2 \|\boldsymbol{x}_0\|_1}{\|\boldsymbol{a}^*\|_2}$. Consider $\|\boldsymbol{x}^*\|_1 + \nu\|\boldsymbol{a}^*\|_2^2 - \|\boldsymbol{x}_0\|_1 - \nu\|\boldsymbol{a}_0\|_2^2$ with $0 < \nu \le \frac{\|\boldsymbol{x}_0\|_1}{2\|\boldsymbol{a}_0\|_2^2}$, we have

$$\begin{aligned}
\|\boldsymbol{x}^*\|_1 + \nu\|\boldsymbol{a}^*\|_2^2 - \|\boldsymbol{x}_0\|_1 - \nu\|\boldsymbol{a}_0\|_2^2 &\ge \|\boldsymbol{x}_0\|_1 \left(\frac{\|\boldsymbol{a}_0\|_2}{\|\boldsymbol{a}^*\|_2} - 1 \right) - \nu(\|\boldsymbol{a}_0\|_2^2 - \|\boldsymbol{a}^*\|_2^2) \\
&= \|\boldsymbol{x}_0\|_1 \frac{\|\boldsymbol{a}_0\|_2 - \|\boldsymbol{a}^*\|_2}{\|\boldsymbol{a}^*\|_2} - \nu\left(\|\boldsymbol{a}_0\|_2^2 - \|\boldsymbol{a}^*\|_2^2\right) \\
&= (\|\boldsymbol{a}_0\|_2 - \|\boldsymbol{a}^*\|_2)\left(\frac{\|\boldsymbol{x}_0\|_1}{\|\boldsymbol{a}^*\|_2} - \nu(\|\boldsymbol{a}^*\|_2 + \|\boldsymbol{a}_0\|_2) \right) \\
&\ge (\|\boldsymbol{a}_0\|_2 - \|\boldsymbol{a}^*\|_2)\left(\frac{\|\boldsymbol{x}_0\|_1}{\|\boldsymbol{a}^*\|_2} - 2\nu\|\boldsymbol{a}_0\|_2 \right) \\
&\ge (\|\boldsymbol{a}_0\|_2 - \|\boldsymbol{a}^*\|_2)\left(\frac{\|\boldsymbol{x}_0\|_1}{\|\boldsymbol{a}^*\|_2} - \frac{\|\boldsymbol{x}_0\|_1}{\|\boldsymbol{a}^*\|_2} \right) = 0.
\end{aligned}$$

Now we have

$$\|\boldsymbol{x}^*\|_1 + \nu \|\boldsymbol{a}^*\|_2^2 \geq \|\boldsymbol{x}_0\|_1 + \nu \|\boldsymbol{a}_0\|_2^2 .$$

By the fact $(\boldsymbol{a}^*, \boldsymbol{x}^*)$ is one optimal solution to model (1.4), we have $\|\boldsymbol{x}^*\|_1 + \nu \|\boldsymbol{a}^*\|_2^2 = \|\boldsymbol{x}_0\|_1 + \nu \|\boldsymbol{a}_0\|_2^2$. So $(\boldsymbol{a}_0, \boldsymbol{x}_0)$ is also an optimal solution and all the above inequalities should be equalities. Therefore, we have for all i such that $\boldsymbol{x}_i^* \neq 0$, $|\langle \boldsymbol{a}_0 \otimes \operatorname{sign}(\boldsymbol{x}_0), \mathcal{S}_i(\boldsymbol{a}^*)\rangle| = \|\boldsymbol{a}^*\|_2 \|\boldsymbol{a}_0\|_2$. If $\boldsymbol{x}_0$ is signed $2k$-separable, then

$$|\langle \boldsymbol{a}_0 \otimes \operatorname{sign}(\boldsymbol{x}_0), \mathcal{S}_i(\boldsymbol{a}^*)\rangle| \leq |\langle \mathcal{S}_j(\boldsymbol{a}_0), \mathcal{S}_i(\boldsymbol{a}^*)\rangle|,$$

for some j being the index of the support set of $\boldsymbol{a}_0$. So we must have $|\langle \mathcal{S}_j(\boldsymbol{a}_0), \mathcal{S}_i(\boldsymbol{a}^*)\rangle| = \|\boldsymbol{a}^*\|_2 \|\boldsymbol{a}_0\|_2$, $(\boldsymbol{a}^*, \boldsymbol{x}^*) = (\boldsymbol{a}_0, \boldsymbol{x}_0)$ up to a cyclic shift. ■

In a quick glance, the k-separation condition on the signal $\boldsymbol{x}_0$ is a very strong condition to ensure that the truth is one of the global minima of the model (1.4). The next example shows that it is indeed tight to ensure the soundness of the model.

Example 2.1 (The necessity of k-separation). Consider a measurement $\boldsymbol{b} = \boldsymbol{a}_0 \otimes \boldsymbol{x}_0$ where the pair $(\boldsymbol{a}_0, \boldsymbol{x}_0)$ is defined by

$$\begin{aligned} \boldsymbol{a}_0 &= [1/18, 1/9, 1/9, 1/9, 1/9, 1/9, 1/9, 1/9, 1/9, 1/18] \in \mathbb{R}^{10}, \\ \boldsymbol{x}_0 &= [\underbrace{\boldsymbol{b}, \ldots, \boldsymbol{b}}_{9}, \underbrace{-\boldsymbol{b}, \ldots, -\boldsymbol{b}}_{9}] \in \mathbb{R}^{162}, \quad \text{where } \boldsymbol{b} = [1, \underbrace{0, \ldots, 0}_{8}]. \end{aligned}$$

Then, the kernel size $k = 10$ and the signal is $k-1$ separable. Consider another pair $(\boldsymbol{a}_1, \boldsymbol{x}_1)$ defined by

$$\begin{aligned} \boldsymbol{a}_1 &= [1/10, 1/10, 1/10, 1/0, 1/10, 1/10, 1/10, 1/10, 1/10, 1/10] \in \mathbb{R}^{10}, \\ \boldsymbol{x}_1 &= [0, \underbrace{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_1}_{8}, 0, \underbrace{-\boldsymbol{b}_1, \ldots, -\boldsymbol{b}_1}_{8}] \in \mathbb{R}^{162}, \quad \text{where } \boldsymbol{b}_1 = [10/9, \underbrace{0, \ldots, 0}_{9}]. \end{aligned}$$

We have that $(\boldsymbol{a}_1, \boldsymbol{x}_1) \in \mathcal{S}_{\boldsymbol{b}}^0$. By direct calculation, we have $\|\boldsymbol{a}_0\|_2 = \sqrt{\frac{17}{162}}$, $\|\boldsymbol{a}_1\|_2 = \sqrt{\frac{1}{10}}$, $\|\boldsymbol{x}_0\|_1 = 18$ and $\|\boldsymbol{x}_1\|_1 = 160/9$. Thus, for any $\nu > 0$,

$$\|\boldsymbol{x}_1\|_1 + \nu \|\boldsymbol{a}_1\|_2^2 < \|\boldsymbol{x}_0\|_1 + \nu \|\boldsymbol{a}_0\|_2^2 .$$

In other words, for the true pair $(\boldsymbol{a}_0, \boldsymbol{x}_0)$ where the signal is $(k-1)$-separable, neither it nor its variations with cyclic translations is a global minimum of (1.4).

While the k-separation condition is sufficient to guarantee that the truth is one of the global minimum of the model (1.4), it is not sufficient to guarantee that the truth is an unique one up to a cyclic shift.

Example 2.2 (Insufficiency of k-separation for unique global minimum). Consider the measurement $\boldsymbol{b} = \boldsymbol{a}_0 \otimes \boldsymbol{x}_0$ defined by

$$\begin{aligned} \boldsymbol{a}_0 &= [1, 1, 0, 0, 0, 0, 2]/4 \in \mathbb{R}^7, \\ \boldsymbol{x}_0 &= [\bar{\boldsymbol{x}}_0, \bar{\boldsymbol{x}}_0, \bar{\boldsymbol{x}}_0] \in \mathbb{R}^{24} \quad \text{where } \bar{\boldsymbol{x}}_0 = [1, \underbrace{0, \ldots, 0}_{7}], \end{aligned}$$

Then, the pair $(\boldsymbol{a}_1, \boldsymbol{x}_1) \in \mathcal{S}_{\boldsymbol{b}}^0$ given by

$$\begin{aligned} \boldsymbol{a}_1 &= [2,0,1,1,0,0,0]/4 \in \mathbb{R}^7, \\ \boldsymbol{x}_1 &= [\bar{\boldsymbol{x}}_1, \bar{\boldsymbol{x}}_1, \bar{\boldsymbol{x}}_1] \in \mathbb{R}^{24} \quad \text{where } \bar{\boldsymbol{x}}_1 = [\underbrace{0,\dots,0}_{6},1,0], \end{aligned}$$

is also a global minimum of (1.4). Note that $\boldsymbol{a}_1$ is not any shift of $\boldsymbol{a}_0$.

2.2. Model (1.4) for approximately signed $2k$-separable with measurement noise. In this subsection, we consider the estimation error of the kernel for general sparse signals when using the model (1.4) in the presence of measurement noise. Consider a noisy measurement $\boldsymbol{b} = \boldsymbol{a}_0 \otimes \boldsymbol{x}_0 + \boldsymbol{n}$. Let $\|\boldsymbol{n}\|_1 \le \epsilon$. Before proving Theorem 1.5, we first establish the following lemmas.

Lemma 2.3. *Suppose $\boldsymbol{x}_{0,s2k}$ is nonzero and signed $2k$-separated and $\boldsymbol{x}_0 = \boldsymbol{x}_{0,s2k} + \Delta\boldsymbol{x}_0$. Suppose $(\boldsymbol{a}^*, \boldsymbol{x}^*)$ is an optimal solution to model (1.4) with feasibility set $\mathcal{S}_{\boldsymbol{b}}^{\epsilon}$, then*

$$C(\boldsymbol{a}_0, \boldsymbol{a}^*) \ge (1 - 2\frac{\|\Delta\boldsymbol{x}_0\|_1}{\|\boldsymbol{x}_0\|_1} - \frac{\|\boldsymbol{a}_0\|_\infty}{\|\boldsymbol{a}_0\|_2^2}\frac{2\epsilon}{\|\boldsymbol{x}_0\|_1})\frac{\|\boldsymbol{a}_0\|_2 / \|\boldsymbol{a}^*\|_2}{1 + \frac{\nu}{\|\boldsymbol{x}_0\|_1}(\|\boldsymbol{a}_0\|_2^2 - \|\boldsymbol{a}^*\|_2^2)}.$$

Proof. Since $(\boldsymbol{a}^*, \boldsymbol{x}^*)$ belongs to the feasible set, there exists a $\boldsymbol{z}$ such that $\boldsymbol{a}^* \otimes \boldsymbol{x}^* = \boldsymbol{a}_0 \otimes (\boldsymbol{x}_{0,s2k} + \Delta\boldsymbol{x}_0) + \boldsymbol{z}$, $\|\boldsymbol{z}\|_1 \le 2\epsilon$. We have

$$\begin{aligned} &\langle \boldsymbol{a}_0 \otimes \operatorname{sign}(\boldsymbol{x}_{0,s2k}), \boldsymbol{a}^* \otimes \boldsymbol{x}^* \rangle \\ &= \langle \boldsymbol{a}_0 \otimes \operatorname{sign}(\boldsymbol{x}_{0,s2k}), \boldsymbol{a}_0 \otimes \boldsymbol{x}_{0,s2k} \rangle + \langle \boldsymbol{a}_0 \otimes \operatorname{sign}(\boldsymbol{x}_{0,s2k}), \boldsymbol{a}_0 \otimes \Delta\boldsymbol{x}_0 \rangle + \langle \boldsymbol{a}_0 \otimes \operatorname{sign}(\boldsymbol{x}_{0,s2k}), \boldsymbol{z} \rangle, \end{aligned}$$

which means

$$\sum_{i=0}^{n-1} |\boldsymbol{x}_i^*| \cdot |\langle \boldsymbol{a}_0 \otimes \operatorname{sign}(\boldsymbol{x}_{0,s2k}), \mathcal{S}_i(\boldsymbol{a}^*) \rangle| \ge \|\boldsymbol{a}_0\|_2^2 (\|\boldsymbol{x}_{0,s2k}\|_1 - \|\Delta\boldsymbol{x}_0\|_1) - \|\boldsymbol{a}_0\|_\infty \|\boldsymbol{z}\|_1,$$

Define

$$A := \max_{\tau \in [n]} \{ |\langle \boldsymbol{a}_0 \otimes \operatorname{sign}(\boldsymbol{x}_{0,s2k}), \mathcal{S}_\tau(\boldsymbol{a}^*) \rangle| \}.$$

As $\boldsymbol{x}_{0,s2k}$ is signed 2k-separable, we also have $A \le \max_{\tau \in [n]} \{ |\langle \boldsymbol{a}_0, \mathcal{S}_\tau(\boldsymbol{a}^*) \rangle| \}$. Then

$$A \|\boldsymbol{x}^*\|_1 \ge \|\boldsymbol{a}_0\|_2^2 (\|\boldsymbol{x}_{0,s2k}\|_1 - \|\Delta\boldsymbol{x}_0\|_1 - 2\frac{\|\boldsymbol{a}_0\|_\infty}{\|\boldsymbol{a}_0\|_2^2}\epsilon).$$

On the other hand, $(\boldsymbol{a}^*, \boldsymbol{x}^*)$ is the minimizer of the problem, so we have

$$\nu \|\boldsymbol{a}^*\|_2^2 + \|\boldsymbol{x}^*\|_1 \le \nu \|\boldsymbol{a}_0\|_2^2 + \|\boldsymbol{x}_0\|_1.$$

The Combination of the two inequalities gives

$$\frac{\|\boldsymbol{a}_0\|_2^2 (\|\boldsymbol{x}_0\|_1 - 2\|\Delta\boldsymbol{x}_0\|_1 - 2\frac{\|\boldsymbol{a}_0\|_\infty}{\|\boldsymbol{a}_0\|_2^2}\epsilon)}{A} \le \frac{\|\boldsymbol{a}_0\|_2^2 (\|\boldsymbol{x}_{0,s2k}\|_1 - \|\Delta\boldsymbol{x}_0\|_1 - 2\frac{\|\boldsymbol{a}_0\|_\infty}{\|\boldsymbol{a}_0\|_2^2}\epsilon)}{A}$$

$$\leq \nu(\|\boldsymbol{a}_0\|_2^2 - \|\boldsymbol{a}^*\|_2^2) + \|\boldsymbol{x}_0\|_1 .$$

Hence we have an estimation about A:

$$A \geq \frac{\|\boldsymbol{a}_0\|_2^2 \left(\|\boldsymbol{x}_0\|_1 - 2\|\Delta \boldsymbol{x}_0\|_1 - 2\frac{\|\boldsymbol{a}_0\|_\infty}{\|\boldsymbol{a}_0\|_2^2}\epsilon\right)}{\nu(\|\boldsymbol{a}_0\|_2^2 - \|\boldsymbol{a}^*\|_2^2) + \|\boldsymbol{x}_0\|_1},$$

which leads to

$$\frac{A}{\|\boldsymbol{a}_0\|_2^2} \geq \frac{\|\boldsymbol{x}_0\|_1 - 2\|\Delta \boldsymbol{x}_0\|_1 - 2\frac{\|\boldsymbol{a}_0\|_\infty}{\|\boldsymbol{a}_0\|_2^2}\epsilon}{\nu(\|\boldsymbol{a}_0\|_2^2 - \|\boldsymbol{a}^*\|_2^2) + \|\boldsymbol{x}_0\|_1} = \frac{1 - 2\frac{\|\Delta \boldsymbol{x}_0\|_1}{\|\boldsymbol{x}_0\|_1} - \frac{\|\boldsymbol{a}_0\|_\infty}{\|\boldsymbol{a}_0\|_2^2}\frac{2\epsilon}{\|\boldsymbol{x}_0\|_1}}{1 + \frac{\nu}{\|\boldsymbol{x}_0\|_1}(\|\boldsymbol{a}_0\|_2^2 - \|\boldsymbol{a}^*\|_2^2)}.$$

Thus, we have

$$\begin{aligned} C(\boldsymbol{a}_0, \boldsymbol{a}^*) &\geq \frac{A}{\|\boldsymbol{a}_0\|_2 \|\boldsymbol{a}^*\|_2} = \frac{A}{\|\boldsymbol{a}_0\|_2^2}\frac{\|\boldsymbol{a}_0\|_2}{\|\boldsymbol{a}^*\|_2} \\ &\geq \left(1 - 2\frac{\|\Delta \boldsymbol{x}_0\|_1}{\|\boldsymbol{x}_0\|_1} - \frac{\|\boldsymbol{a}_0\|_\infty}{\|\boldsymbol{a}_0\|_2^2}\frac{2\epsilon}{\|\boldsymbol{x}_0\|_1}\right)\frac{\|\boldsymbol{a}_0\|_2 / \|\boldsymbol{a}^*\|_2}{1 + \frac{\nu}{\|\boldsymbol{x}_0\|_1}(\|\boldsymbol{a}_0\|_2^2 - \|\boldsymbol{a}^*\|_2^2)}. \end{aligned}$$

The proof is done. ■

It can be seen that the estimation error is also related to the term

$$R_\nu(\|\boldsymbol{a}^*\|_2, \|\boldsymbol{a}_0\|_2) := \frac{\|\boldsymbol{a}_0\|_2 / \|\boldsymbol{a}^*\|_2}{1 + \frac{\nu}{\|\boldsymbol{x}_0\|_1}(\|\boldsymbol{a}_0\|_2^2 - \|\boldsymbol{a}^*\|_2^2)},$$

which depends on the value ν. In the next lemma, we give a lower bound of such an estimator for a specific value of ν.

Lemma 2.4. *Under the same assumptions as Lemma 2.3, setting $\nu = \frac{\|\boldsymbol{x}_0\|_1}{2\|\boldsymbol{a}_0\|_2^2}$, assuming $\frac{\|\Delta \boldsymbol{x}_0\|_1 + \epsilon}{\|\boldsymbol{x}_0\|_1} < \frac{\sqrt{3}-1}{2\sqrt{3}}$, we have*

$$R_\nu(\|\boldsymbol{a}^*\|_2, \|\boldsymbol{a}_0\|_2) \geq 1.$$

Proof. There are two cases for the relation between $\|\boldsymbol{a}^*\|_2$ and $\|\boldsymbol{a}_0\|_2$.

Case 1: $\|\boldsymbol{a}^*\|_2 \geq \|\boldsymbol{a}_0\|_2$.
In this case, from the definition of $(\boldsymbol{a}^*, \boldsymbol{x}^*)$, we have $\boldsymbol{a}^* \otimes \boldsymbol{x}^* = \boldsymbol{a}_0 \otimes \boldsymbol{x}_0 + \boldsymbol{z}$ with $\|\boldsymbol{z}\|_1 \leq 2\epsilon$, and

$$\nu \|\boldsymbol{a}^*\|_2^2 + \|\boldsymbol{x}^*\|_1 \leq \nu \|\boldsymbol{a}_0\|_2^2 + \|\boldsymbol{x}_0\|_1 .$$

Moreover,

$$\|\boldsymbol{x}^*\|_1 \geq \|\boldsymbol{a}^* \otimes \boldsymbol{x}^*\|_1 = \|\boldsymbol{a}_0 \otimes \boldsymbol{x}_0 + \boldsymbol{z}\|_1 \geq \|\boldsymbol{x}_0\|_1 - 2\|\Delta \boldsymbol{x}_0\|_1 - 2\epsilon.$$

and thus

$$\nu \|\boldsymbol{a}^*\|_2^2 + \|\boldsymbol{x}_0\|_1 - 2\|\Delta \boldsymbol{x}_0\|_1 - 2\epsilon \leq \nu \|\boldsymbol{a}^*\|_2^2 + \|\boldsymbol{x}^*\|_1 \leq \nu \|\boldsymbol{a}_0\|_2^2 + \|\boldsymbol{x}_0\|_1 ,$$

which implies

$$\nu(\|\boldsymbol{a}^*\|_2^2 - \|\boldsymbol{a}_0\|_2^2) \leq 2\|\Delta\boldsymbol{x}_0\|_1 + 2\epsilon.$$

Together with $(\|\boldsymbol{a}^*\|_2^2 - \|\boldsymbol{a}_0\|_2^2) = (\|\boldsymbol{a}^*\|_2 + \|\boldsymbol{a}_0\|_2)(\|\boldsymbol{a}^*\|_2 - \|\boldsymbol{a}_0\|_2) \geq 2\|\boldsymbol{a}_0\|_2((\|\boldsymbol{a}^*\|_2 - \|\boldsymbol{a}_0\|_2)$, we have

$$(\|\boldsymbol{a}^*\|_2 - \|\boldsymbol{a}_0\|_2) \leq \frac{\|\Delta\boldsymbol{x}_0\|_1}{\nu\|\boldsymbol{a}_0\|_2} + \frac{\epsilon}{\nu\|\boldsymbol{a}_0\|_2}.$$

Thus,

$$\frac{\|\boldsymbol{a}_0\|_2}{\|\boldsymbol{a}^*\|_2} \geq 1 - \frac{\|\Delta\boldsymbol{x}_0\|_1}{\nu\|\boldsymbol{a}_0\|_2\|\boldsymbol{a}^*\|_2} - \frac{\epsilon}{\nu\|\boldsymbol{a}_0\|_2\|\boldsymbol{a}^*\|_2} \geq 1 - \frac{\|\Delta\boldsymbol{x}_0\|_1 + \epsilon}{\nu\|\boldsymbol{a}_0\|_2^2},$$

when $\nu = \frac{\|\boldsymbol{x}_0\|_1}{2\|\boldsymbol{a}_0\|_2^2}$, we have

$$\frac{\|\boldsymbol{a}_0\|_2}{\|\boldsymbol{a}^*\|_2} \geq 1 - 2\frac{\|\Delta\boldsymbol{x}_0\|_1 + \epsilon}{\|\boldsymbol{x}_0\|_1}.$$

Therefore, as long as $1 - 2\frac{\|\Delta\boldsymbol{x}_0\|_1+\epsilon}{\|\boldsymbol{x}_0\|_1} > \frac{1}{\sqrt{3}}$, we have the denominator $1 + \frac{\nu}{\|\boldsymbol{x}_0\|_1}(\|\boldsymbol{a}_0\|_2^2 - \|\boldsymbol{a}^*\|_2^2) > 0$, for $\nu \leq \frac{\|\boldsymbol{x}_0\|_1}{2\|\boldsymbol{a}_0\|_2^2}$. The following inequality holds:

$$\begin{aligned}
\frac{\|\boldsymbol{a}_0\|_2 / \|\boldsymbol{a}^*\|_2}{1 + \frac{\nu}{\|\boldsymbol{x}_0\|_1}(\|\boldsymbol{a}_0\|_2^2 - \|\boldsymbol{a}^*\|_2^2)} &\geq (1 + \frac{\nu}{\|\boldsymbol{x}_0\|_1}(\|\boldsymbol{a}^*\|_2^2 - \|\boldsymbol{a}_0\|_2^2))\frac{\|\boldsymbol{a}_0\|_2}{\|\boldsymbol{a}^*\|_2} \\
&= (1 + \frac{\nu}{\|\boldsymbol{x}_0\|_1}(\|\boldsymbol{a}^*\|_2 + \|\boldsymbol{a}_0\|_2)(\|\boldsymbol{a}^*\|_2 - \|\boldsymbol{a}_0\|_2))\frac{\|\boldsymbol{a}_0\|_2}{\|\boldsymbol{a}^*\|_2} \\
&\geq (1 + \frac{2\nu\|\boldsymbol{a}_0\|_2}{\|\boldsymbol{x}_0\|_1}(\|\boldsymbol{a}^*\|_2 - \|\boldsymbol{a}_0\|_2))\frac{\|\boldsymbol{a}_0\|_2}{\|\boldsymbol{a}^*\|_2} \\
&= \frac{2\nu\|\boldsymbol{a}_0\|_2^2}{\|\boldsymbol{x}_0\|_1} + (1 - \frac{2\nu\|\boldsymbol{a}_0\|_2^2}{\|\boldsymbol{x}_0\|_1})\frac{\|\boldsymbol{a}_0\|_2}{\|\boldsymbol{a}^*\|_2}.
\end{aligned}$$

Case 2: $\|\boldsymbol{a}^*\|_2 \leq \|\boldsymbol{a}_0\|_2$.

$$\begin{aligned}
\frac{\|\boldsymbol{a}_0\|_2 / \|\boldsymbol{a}^*\|_2}{1 + \frac{\nu}{\|\boldsymbol{x}_0\|_1}(\|\boldsymbol{a}_0\|_2^2 - \|\boldsymbol{a}^*\|_2^2)} &\geq (1 - \frac{\nu}{\|\boldsymbol{x}_0\|_1}(\|\boldsymbol{a}_0\|_2^2 - \|\boldsymbol{a}^*\|_2^2))\frac{\|\boldsymbol{a}_0\|_2}{\|\boldsymbol{a}^*\|_2} \\
&= (1 - \frac{\nu}{\|\boldsymbol{x}_0\|_1}(\|\boldsymbol{a}_0\|_2 + \|\boldsymbol{a}^*\|_2)(\|\boldsymbol{a}_0\|_2 - \|\boldsymbol{a}^*\|_2))\frac{\|\boldsymbol{a}_0\|_2}{\|\boldsymbol{a}^*\|_2} \\
&\geq (1 - \frac{2\nu\|\boldsymbol{a}_0\|_2}{\|\boldsymbol{x}_0\|_1}(\|\boldsymbol{a}_0\|_2 - \|\boldsymbol{a}^*\|_2))\frac{\|\boldsymbol{a}_0\|_2}{\|\boldsymbol{a}^*\|_2} \\
&= \frac{2\nu\|\boldsymbol{a}_0\|_2^2}{\|\boldsymbol{x}_0\|_1} + (1 - \frac{2\nu\|\boldsymbol{a}_0\|_2^2}{\|\boldsymbol{x}_0\|_1})\frac{\|\boldsymbol{a}_0\|_2}{\|\boldsymbol{a}^*\|_2}.
\end{aligned}$$

Based on the discussions on the two cases above, we have the following inequality:

$$R_\nu(\|\boldsymbol{a}^*\|_2, \|\boldsymbol{a}_0\|_2) \geq \frac{2\nu\|\boldsymbol{a}_0\|_2^2}{\|\boldsymbol{x}_0\|_1} + (1 - \frac{2\nu\|\boldsymbol{a}_0\|_2^2}{\|\boldsymbol{x}_0\|_1})\frac{\|\boldsymbol{a}_0\|_2}{\|\boldsymbol{a}^*\|_2}.$$

When $\nu = \frac{\|\boldsymbol{x}_0\|_1}{2\|\boldsymbol{a}_0\|_2^2}$, we have

$$R_\nu(\|\boldsymbol{a}^*\|_2, \|\boldsymbol{a}_0\|_2) \geq 1.$$

The proof is done. ■

Remark 2.5. In practice, $\|\boldsymbol{a}_0\|_2$ and $\|\boldsymbol{x}_0\|_1$ are not accessible. Thus, we use $\nu \approx \frac{\sqrt{k}\|\boldsymbol{b}\|_1}{2}$ as the initial value ν. Then, we use the the estimation of these two quantities to update the value of ν during the iteration.

Proof of Theorem 1.5. The proof is done by combining Lemma 2.3 and Lemma 2.4, and the fact $0.21 < \frac{\sqrt{3}-1}{2\sqrt{3}}$. ■

In the case where the noise $\boldsymbol{n}$ is negligible and the kernel erases a significant portion of the information of $\boldsymbol{x}$ in terms of the energy of the measurement: $\|\boldsymbol{a_0} \otimes \boldsymbol{x}_0\| < \|\boldsymbol{x}_0\| - c_0$, where c_0 is a non-negligible positive constant. Our theorem shows that the recovery of the kernel can be robust to such a loss of information. When $\|\boldsymbol{a}_0 \otimes \Delta\boldsymbol{x}_0\|_1 \leq \epsilon$, we can treat $\boldsymbol{a}_0 \otimes \Delta\boldsymbol{x}_0$ as noise and the above theorem can be applied.

Corollary 2.6. *Suppose that the measurement $\boldsymbol{b} = \boldsymbol{a_0} \otimes \boldsymbol{x}_0$ is noise-free, $\boldsymbol{x}_{0,s2k}$ is a nonzero signed 2k-separable approximation to $\boldsymbol{x}_0$ such that*

$$\boldsymbol{x}_0 = \boldsymbol{x}_{0,s2k} + \Delta\boldsymbol{x}_0.$$

Moreover, assume $\|\boldsymbol{a}_0 \otimes \Delta\boldsymbol{x}_0\|_1 \leq \epsilon$, $\frac{\epsilon}{\|\boldsymbol{x}_{0,s2k}\|_1} \leq 0.21$. Let $(\boldsymbol{a}^*, \boldsymbol{x}^*)$ denote an optimal solution to the problem (1.4) with the feasible set $\mathcal{S}_{\boldsymbol{b}}^\epsilon$. Then, by setting $\nu = \frac{\|\boldsymbol{x}_0\|_1}{2\|\boldsymbol{a}_0\|_2^2}$, we have

$$C(\boldsymbol{a}_0, \boldsymbol{a}^*) \geq 1 - \frac{\|\boldsymbol{a}_0\|_\infty}{\|\boldsymbol{a}_0\|_2^2} \frac{2\epsilon}{\|\boldsymbol{x}_{0,s2k}\|_1}.$$

Proof. Replace $\boldsymbol{x}_0$ and $\boldsymbol{n}$ in Theorem 1.5 by $\boldsymbol{x}_{0,s2k}$ and $\boldsymbol{a}_0 \otimes \Delta\boldsymbol{x}_0$. ■

2.3. Region selection for blind SaS deconvolution and its application in blind image deblurring. Theorem 1.5 shows that the error between the true kernel and the global minimum of the model (1.4) depends on $\|\Delta\boldsymbol{x}\|_1/\|\boldsymbol{x}\|_1$, measuring how close the signal is to a signed $2k$-sparse signal. In other words, as long as the entries of the signal with significant magnitude are well separated, the model (1.4) is good for estimating the kernel. In general, For a sparse signal, there are the areas containing non-zero entries with sufficient separation and the areas contains dense non-zero entries. For blind SaS problem, the signal size is much larger than kernel size. Thus, one might only select certain parts of the signal which only contains well-separated non-zero entries, and use these parts to estimate the blur kernel. After the blur kernel is accurately estimated. Then, a non-blind deconvolution method is called to deconvolve the whole signal.

In the blind deblurring application, the blurring is modeled by $\boldsymbol{B} = \boldsymbol{a} \otimes \boldsymbol{I}$, where $\boldsymbol{I}/\boldsymbol{B}$ are the clean/blurred images respectively. In certain cases, the gradient image $\nabla\boldsymbol{I}$ is assumed with SaS structure. Thus when we estimate kernel $\boldsymbol{a}$, we regard the measurement model in gradient domain, *i.e.*, $\boldsymbol{b} = \nabla\boldsymbol{B} = \boldsymbol{a} \otimes \boldsymbol{x}$ where $\boldsymbol{x}$ is the gradient of $\boldsymbol{I}$. It is empirically observed that the intermediate estimate $\boldsymbol{x}^t$ with least fitting error is indeed not a good candidate for refining the estimation of the kernel. Many empirical techniques are proposed for processing the intermediate result to facilitate the refinement of the kernel; see *e.g.* [11, 34, 26, 15, 13, 35]. For example, Cho and Lee [11] propose to modify the estimated image by shock filter before being used for refining the kernel estimation. Xu and Jia [34] proposed to run a salient edge selection scheme to erase certain edges of the intermediate result. These methods post-process the intermediate image recovery with heuristic strategies to promote the separation of remained edges. There exists another approach to obtain more accurate kernel estimation from a good region extracted from the input blurred image. In this direction, Hu and Yang [15] proposed a learning method to learn which region is selected for kernel estimation. In this paper, we follow the region selection approach to estimate the kernel. Considering a motion-blurred natural image, it usually contains both cartoon regions and texture regions and image gradients are usually well-separated in cartoon regions. Let $\boldsymbol{x}$ denote image gradient $\nabla \boldsymbol{I}$ and the whole image is divided into several overlapping regions $\{\boldsymbol{x}_i\}$. For better kernel estimation, we use the good regions from $\{\boldsymbol{x}_i\}$. Here the good regions means the regions whose residual component $\|\Delta\boldsymbol{x}_i\|_1/\|\boldsymbol{x}_i\|_1$ is sufficiently small. In other words, based on our analysis shown in Section 2.2, only these regions should be used for estimating kernel, not the whole image.

Based on the analysis conducted in Section 2.1 and 2.2, we proposed a computational scheme to identify such region. Instead of attempting to identifying the regions with approximate signed $2k$ separability, we identify a subset of such regions which can be well approximated by $2k$ separable signals, *i.e.*, we consider

$$\boldsymbol{x}_0 = \boldsymbol{x}_{0,2k} + \Delta\boldsymbol{x}_0,$$

where $\boldsymbol{x}_{0,2k}$ is a nonzero $2k$-separable signal. It is noted that the $2k$-separable signals are a subset of signed $2k$-separable signals defined in Definition 1.3.

Corollary 2.7. *Under the same assumption as Theorem 1.5, let $\boldsymbol{x}_{0,2k}$ denote a non-zero $2k$-separable approximation to $\boldsymbol{x}$:*

$$\boldsymbol{x}_0 = \boldsymbol{x}_{0,2k} + \Delta\boldsymbol{x}_0.$$

Let $(\boldsymbol{a}^*, \boldsymbol{x}^*)$ denote an optimal solution to the problem (1.4) with the feasible set $\mathcal{S}_{\boldsymbol{b}}^{\epsilon}$, we have

$$C(\boldsymbol{a}_0, \boldsymbol{a}^*) \geq 1 - 2\frac{\|\Delta\boldsymbol{x}_0\|_1}{\|\boldsymbol{x}_0\|_1} - \frac{\|\boldsymbol{a}_0\|_\infty}{\|\boldsymbol{a}_0\|_2^2}\frac{2\epsilon}{\|\boldsymbol{x}_0\|_1}.$$

Proof. By Definition 1.3, a $2k$-separable signal is also a signed $2k$-separable. By directly calling Theorem 1.5, we have the conclusion. ∎

The main idea to identify the regions with approximate $2k$-separability is based on the following observation. Notice that the convolution between a k-separated signal $\boldsymbol{y}_0$ and a normalized non-negative kernel $\boldsymbol{a}_0$ with size up to k does not change the ℓ_1 norm of signal, *i.e.*, $\|\boldsymbol{y}_0\|_1 = \|\boldsymbol{a}_0 \otimes \boldsymbol{y}_0\|_1$. Then, consider a normalized Gaussian smooth kernel $\boldsymbol{g}$ with size $\leq k+1$. The kernel $\boldsymbol{g} \otimes \boldsymbol{a}_0$ has size $\leq 2k$. Assume $\boldsymbol{x}_0 = \boldsymbol{x}_{0,2k} + \Delta\boldsymbol{x}_0$ with $\boldsymbol{x}_{0,2k}$ a $2k$-separable signal. Each entry of $\Delta\boldsymbol{x}_0$ follows some i.i.d. with zero expectation. Then we have

$$\|\boldsymbol{a}_0 \otimes \boldsymbol{x}_{0,2k}\|_1 = \|\mathbb{E}_{\Delta\boldsymbol{x}_0}[\boldsymbol{a}_0 \otimes \boldsymbol{x}_0]\|_1 \leq \mathbb{E}_{\Delta\boldsymbol{x}_0}[\|\boldsymbol{a}_0 \otimes \boldsymbol{x}_0\|_1] \leq \|\boldsymbol{a}_0 \otimes \boldsymbol{x}_{0,2k}\|_1 + \mathbb{E}_{\Delta\boldsymbol{x}_0}[\|\boldsymbol{a}_0 \otimes \Delta\boldsymbol{x}_0\|_1].$$

Replacing $\boldsymbol{a}_0$ by $\boldsymbol{g} \otimes \boldsymbol{a}_0$, we have

$$\|\boldsymbol{a}_0 \otimes \boldsymbol{x}_{0,2k}\|_1 = \|\boldsymbol{g} \otimes \boldsymbol{a}_0 \otimes \boldsymbol{x}_{0,2k}\|_1 = \|\mathbb{E}_{\Delta \boldsymbol{x}_0}[\boldsymbol{g} \otimes \boldsymbol{a}_0 \otimes \boldsymbol{x}_0]\|_1 \leq \mathbb{E}_{\Delta \boldsymbol{x}_0}[\|\boldsymbol{g} \otimes \boldsymbol{a}_0 \otimes \boldsymbol{x}_0\|_1].$$

Combine the above two inequalities and the fact $\|\boldsymbol{g} \otimes \boldsymbol{a}_0 \otimes \boldsymbol{x}_0\|_1 \leq \|\boldsymbol{a}_0 \otimes \boldsymbol{x}_0\|_1$, we have

$$0 \leq \mathbb{E}_{\Delta \boldsymbol{x}_0}[\|\boldsymbol{a}_0 \otimes \boldsymbol{x}_0\|_1 - \|\boldsymbol{g} \otimes \boldsymbol{a}_0 \otimes \boldsymbol{x}_0\|_1] \leq \mathbb{E}_{\Delta \boldsymbol{x}_0}[\|\boldsymbol{a}_0 \otimes \Delta \boldsymbol{x}_0\|_1] \leq \mathbb{E}_{\Delta \boldsymbol{x}_0}[\|\Delta \boldsymbol{x}_0\|_1].$$

Therefore, statistically, the smaller the $\|\Delta \boldsymbol{x}_0\|_1$ is, the smaller the gap between $\|\boldsymbol{a}_0 \otimes \boldsymbol{x}_0\|_1$ and $\|\boldsymbol{g} \otimes \boldsymbol{a}_0 \otimes \boldsymbol{x}_0\|_1$ is.

Such property can be used to detecting approximately $2k$-separable regions, *e.g.*, the regions with well-separated prominent gradients are likely to be the ones whose ℓ_1-norm change less after smoothed by a Gaussian smooth kernel. By Corollary 2.7, the model can recover the kernel with good accuracy in these regions. As our goal is to find the regions with small relatively residual component, we restricted the regions to be selected in the set of the ones with sufficient large $\|\nabla \boldsymbol{I}\|_1$. In the implementation, we first erase the regions whose ℓ_1-norm of gradients are smaller than a pre-defined threshold. Then, we select the regions whose ℓ_1-norm of gradients change relative little after smoothing the image by a Gaussian filter. See Algorithm 2.1 for the outline of such a region selection scheme. To facilitate the iteration and save computational time, we only select just one region to predict kernel. The extension to multiple regions is straightforward with careful boundary management.

Algorithm 2.1 Region selector for kernel estimation in blind image deblurring

Input: The input blurred image $\boldsymbol{B}$, the kernel size $[k_1, k_2]$, predefined region size $[m, n]$

Output: A good region patch.

1: Pre-processing the blurred image by erasing the small edges following [34], obtain horizontal and vertical gradient image $\boldsymbol{b} = [\boldsymbol{b}_x, \boldsymbol{b}_y]$

2: Computing the re-blurred image feature $\boldsymbol{b}_g = [\boldsymbol{b}_x \otimes \boldsymbol{g}, \boldsymbol{b}_y \otimes \boldsymbol{g}]$ with a Gaussian blur kernel $\boldsymbol{g}$

3: Computing the change map of reblurring: $\boldsymbol{c} = |\boldsymbol{b}| - |\boldsymbol{b}_g|$

4: Computing the feature map $\boldsymbol{b}_f$ of the whole image with box filter $\boldsymbol{f} \in \mathbb{R}^{m,n}$: $\boldsymbol{b}_f = |\boldsymbol{b}_x \otimes \boldsymbol{f}| + |\boldsymbol{b}_y \otimes \boldsymbol{f}|$

5: Masking out the region using mask $M = (\boldsymbol{b}_f >= 0.95 \max(\boldsymbol{b}_f))$, remained element is set to Inf

6: Selecting the center index of region: $index = \operatorname{argmin}(\boldsymbol{c} \odot M)$

7: Outputting the selected region

Initialization of kernel estimation. In addition to region selection, a multi-scale coarse-to-fine strategy is implemented for providing a good initialization which is close to the true kernel. The basic idea is that after downsampling a blurred image , the resulting blur kernel will has a smaller size too. The smaller the size of the kernel, the easier the problem becomes. Thus, one can first estimate a blur kernel from a down-sampled version of the input image and then up-sampled the kernel to provide a good initialization to the input image. Such a strategy can be recursively used to provide a good initialization of the kernel. In the implementation, a pyramid of the input image is constructed with downsampling rate 2 between two consecutive scales, and the image in the coarsest scale has a kernel size no larger than 3×3. Then, starting with the coarsest scale, the kernel is initialized by a 3×3 constant kernels. The kernel estimated in one scale is then be used as the initial kernel used in the finer scale, after being upsampled by the factor 2 using linear interpolation.

Built on the selected region, we propose using the proximal alternating iterative minimization to solve the variational formulation of the model (1.4) for blind image deblurring. For blind image deblurring, we first estimate the blur kernel $\boldsymbol{a}$ using (2.1) in the domain of image gradients. In the presence of Gaussian white noise, the resulting variational optimization reads as

$$\min_{\boldsymbol{a}\in\Omega,\boldsymbol{x}} \quad \frac{1}{2}\left\|\boldsymbol{b}-\boldsymbol{a}\otimes\boldsymbol{x}\right\|_2^2+\lambda(\left\|\boldsymbol{x}\right\|_1+\nu\left\|\boldsymbol{a}\right\|_2^2) \tag{2.1}$$

where Ω denotes the feasible set for the kernel and $\boldsymbol{b}$ denotes the gradient of the input blurred image $\nabla\boldsymbol{B}$ or its selected region. The model (2.1) is a challenging nonconvex problem. In this paper, the proximal alternating minimization scheme alternatively solves the following two convex sub-problems:

$$\begin{aligned} &\boldsymbol{x}\text{-subproblem}: \boldsymbol{x}_{k+1}=\operatorname*{argmin}_{\boldsymbol{x}} \quad \frac{1}{2}\left\|\boldsymbol{b}-\boldsymbol{a}_k\otimes\boldsymbol{x}\right\|_2^2+\lambda\left\|\boldsymbol{x}\right\|_1+\frac{1}{2\lambda_k}\left\|\boldsymbol{x}-\boldsymbol{x}_k\right\|_2^2;\\ &\boldsymbol{a}\text{-subproblem}: \boldsymbol{a}_{k+1}=\operatorname*{argmin}_{\boldsymbol{a}\in\Omega} \quad \frac{1}{2}\left\|\boldsymbol{b}-\boldsymbol{a}\otimes\boldsymbol{x}_{k+1}\right\|_2^2+\lambda\nu\left\|\boldsymbol{a}\right\|_2^2+\frac{1}{2\mu_k}\left\|\boldsymbol{a}-\boldsymbol{a}_k\right\|_2^2, \end{aligned} \tag{2.2}$$

where λ_k, μ_k denote the step sizes at the k-th iteration.

The two sub-problems in (2.1) are convex and can be solved efficiently by the primal dual hybrid gradient algorithm (Chambolle-Pock algorithm) [8]. For the $\boldsymbol{a}$-subproblem, we first solve it using the primal dual hybrid gradient algorithm without considering the feasible set Ω. Then the solution is projected to the feasible set Ω. Such a strategy is widely used in blind deblurring [27, 17] with satisfactory empirical performance. For the $\boldsymbol{x}$-sub-problem, we simply call the same primal dual hybrid gradient algorithm. Briefly, these two sub-problems without feasible set constraints can be expressed as the following standard composite optimization with proximable function:

$$\min_{\boldsymbol{z}} \quad F(\boldsymbol{A}\boldsymbol{z})+G(\boldsymbol{z}), \tag{2.3}$$

where $F(\cdot)$ denote the differentiable fidelity term, and $G(\cdot)$ denotes the regularization term. The minimization (2.3) is then solved by the following iterative scheme:

$$\begin{aligned} \boldsymbol{z}^{k+1} &= \operatorname*{prox}_{\sigma F}(\boldsymbol{z}^k-\sigma\boldsymbol{A}^*\boldsymbol{y}^k)\\ \boldsymbol{y}^{k+\frac{1}{2}} &= \boldsymbol{y}^k+\sigma\boldsymbol{A}(2\boldsymbol{z}^{k+1}-\boldsymbol{z}^k)\\ \boldsymbol{y}^{k+1} &= \boldsymbol{y}^{k+\frac{1}{2}}-\sigma\operatorname*{prox}_{\sigma^{-1}G}(\sigma^{-1}\boldsymbol{y}^{k+\frac{1}{2}}). \end{aligned}$$

where the stepsize satisfies $0<\sigma<1/\left\|\boldsymbol{A}\right\|_2$. For the terms $\left\|\boldsymbol{x}\right\|_1$ and $\left\|\boldsymbol{a}\right\|_2^2$, their related proximities can be efficiently computed.

As the problem (2.1) is nonconvex, the proximal alternating minimization scheme (2.2) can not guarantee the convergence to one global minimizer of (2.2). Nevertheless, suppose that two sub-problems in (2.2) are exactly solved during the iteration. Then, one can show that by [2, Theorem 9], the sequence generated by the scheme (2.2) converges to a critical point of the problem

$$\Psi(\boldsymbol{a},\boldsymbol{x}) := \frac{1}{2}\|\boldsymbol{b}-\boldsymbol{a}\otimes\boldsymbol{x}\|_2^2 + \lambda(\|\boldsymbol{x}\|_1 + \nu\,\|\boldsymbol{a}\|_2^2) + \delta_\Omega(\boldsymbol{a}), \tag{2.4}$$

where δ_Ω is the indicator function of the feasible set Ω of kernel $\boldsymbol{a}$. For the completeness of the paper, we provide a sketch of the proof.

Definition 2.8 (Subdifferential [28]). *Consider a proper low semi-continuous function $f : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$. Denote its domain by $dom f := \{\boldsymbol{x}|f(\boldsymbol{x}) < +\infty\}$. Then, the Fréchet subdifferential of f at $\boldsymbol{x}\in dom f$, written as $\hat{\partial} f(\boldsymbol{x})$, is the set of vectors $\boldsymbol{v}\in\mathbb{R}^n$ satisfying*

$$\liminf_{\boldsymbol{x}\neq\boldsymbol{y},\boldsymbol{y}\to\boldsymbol{x}} \frac{1}{\|\boldsymbol{x}-\boldsymbol{y}\|}[f(\boldsymbol{y}) - f(\boldsymbol{x}) - \langle \boldsymbol{v},\boldsymbol{x}\rangle] \geq 0.$$

If $\boldsymbol{x}\notin dom f$, then $\hat{\partial} f(\boldsymbol{x}) = \emptyset$. he subdifferential of f at $\boldsymbol{x}\in dom f$, written as $\partial f(\boldsymbol{x})$, is defined as

$$\partial f(\boldsymbol{x}) := \{\boldsymbol{v}\in\mathbb{R}^n : \exists \boldsymbol{x}_n \to \boldsymbol{x}, f(\boldsymbol{x}_n)\to f(\boldsymbol{x}), \boldsymbol{v}_n \in \hat{\partial} f(\boldsymbol{x}_n) \to \boldsymbol{v}\}.$$

Definition 2.9 (Critical Point). *A point $\boldsymbol{x}$ is called critical point of f if $0\in\partial f(\boldsymbol{x})$, where $\partial f(\boldsymbol{x})$ denotes the subdifferential of f.*

It is noted that $0\in\partial f(\boldsymbol{x})$ is only a necessary condition for x being a local minimizer of f. It is shown in [2, Theorem 9] that, for the sequence $\{(\boldsymbol{a}_k,\boldsymbol{x}_k)\}_k$ generated from (2.2), either $\|(\boldsymbol{a}_k,\boldsymbol{x}_k)\|$ tends to infinity or converges to a critical point of (2.4), if the objective function ψ satisfies the following properties:

$$\text{(K-L)}: \quad \Psi \text{ satisfies the so-called Kurdyka-Lojasiewic property [2];}$$

$$(\mathcal{H}): \begin{cases} \Psi(\boldsymbol{a},\boldsymbol{x}) = f(\boldsymbol{a}) + Q(\boldsymbol{a},\boldsymbol{x}) + g(\boldsymbol{x}), \\ f:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}, g:\mathbb{R}^m\to\mathbb{R}\cup\{+\infty\} \text{ are proper lower semicontinuous,} \\ Q:\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R} \text{ is a } \mathcal{C}^1 \text{ function,} \\ \nabla Q \text{ is Lipschitz continuous on bounded subsets of } \mathbb{R}^n\times\mathbb{R}^m; \end{cases}$$

$$(\mathcal{H}_1) \begin{cases} \inf_{\mathbb{R}^n\times\mathbb{R}^m}\Psi > -\infty, \\ \text{The function } \Psi(\cdot,\boldsymbol{x}_0) \text{ is proper,} \\ \text{There exsit constants } 0<\lambda_- <\lambda_+ \text{ such that } \lambda_- < \lambda_k,\mu_k < \lambda_+, \text{ for all } k\geq 0. \end{cases}$$

For the objective function Ψ of the problem (2.4), let $Q(\boldsymbol{a},\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{a}\otimes\boldsymbol{x}-\boldsymbol{b}\|_2^2$, $f(\boldsymbol{a}) = \lambda\nu\|\boldsymbol{a}\|_2^2 + \delta_\Omega(\boldsymbol{a})$, $g(\boldsymbol{x}) = \lambda\|\boldsymbol{x}\|_1$. It can be seen that the function Ψ satisfies both Condition $\mathcal{H}$ and Condition $\mathcal{H}_1$. Also, as Q, g, $\lambda\mu\|\boldsymbol{a}\|_2^2$ are all semi-algebraic functions, and Ω is a semi-algebraic set, the function Ψ is thus a semi-algebraic function, which satisfy the K-L property [2]. In other words, the objective function Ψ defined by (2.4) satisfies all conditions assume in [2, Theorem 9]. Furthermore, Lemma 5 in [2] shows that $\Psi(\boldsymbol{a}_k,\boldsymbol{x}_k)\leq\Psi(\boldsymbol{a}_{k-1},\boldsymbol{x}_{k-1})\leq\Psi(\boldsymbol{a}_0,\boldsymbol{x}_0)$, therefore, $\lambda(\|\boldsymbol{x}_k\|_1 + \nu\,\|\boldsymbol{a}_k\|_2^2) \leq \Psi(\boldsymbol{a}_k,\boldsymbol{x}_k)\leq\Psi(\boldsymbol{a}_0,\boldsymbol{x}_0)$. The initial kernel $\boldsymbol{a}_0$ is chosen to be in the feasible set, so $\Psi(\boldsymbol{a}_0, \boldsymbol{x}_0) < +\infty$. Therefore, $\{(\boldsymbol{a}_k, \boldsymbol{x}_k)\}$ is bounded, by [2, Theorem 9], the sequence generated by (2.2) converges to a critical point of Ψ.

Furthermore, as long as the initialization is close enough to one global minima, the proximal alternating minimization scheme will converge to such a global minima; See [2, Theorem 10]. In practice, a good initialization is possible using the coarse-to-fine strategy. For better empirical performance, we use the continuation decreasing technique of ν to reduce the regularization effect of $\|\boldsymbol{a}\|_2^2$ and shift correction to address the shift ambiguity. Once we obtain the kernel $\boldsymbol{a}$, we solve the non-blind deblurring

$$\boldsymbol{I} = \operatorname*{argmin}_{\boldsymbol{I}} \quad \|\boldsymbol{B} - \boldsymbol{a} \otimes \boldsymbol{I}\|_1 + \lambda \|\nabla \boldsymbol{I}\|_1 \tag{2.5}$$

to produce the deblurred image using the resolved blur kernel $\boldsymbol{a}$. We use ℓ_1 relating data fidelity for its robustness to outlier. This problem can be reformulated as an ℓ_1 minimization problem. There are many numerical solvers for solving such convex problem, and we use iteratively reweighted least squares method in the implementation. See Algorithm 2.2 for the outline for the blind deblurring algorithm using model (2.1) and good region selector. While there is no guarantee on finding one global minima using the proximal alternating minimization scheme, it is likely to converge to a solution close to the truth kernel

3. Experiments.

The analysis conducted in Section 2 reveals the importance of the separation of significant non-zero entries, when using ℓ_1-norm relating regularization for kernel estimation. As the problem (1.4) is nonconvex, the proximal alternating minimization algorithm cannot guarantee the convergence to a global minimizer. In this section, we run some experiments on the synthesized 1D sparse signals to examine the landscape of the variational regularization formulation of problem (1.4). In the second part of this section, the proposed algorithm with region selection is applied to solve the problem of blind motion deblurring. The experiments are conducted on two popular benchmark datasets, and the proposed method is compared to other existing related techniques seen in blind motion deblurring. The results show the effectiveness of the proposed method which is inspired by the analysis conducted in this paper.

3.1. Landscape visualization on 1D blind SaS deconvolution.

As the proposed algorithm cannot guarantee finding a global minimizer of the non-convex problem (1.4), it can be a concern on the convergence of the iterative scheme adopted in this paper to a local minimizer far away from the truth. In this section, we visualize the landscape of the problem (1.4) with respect to different values of the ν in the 1D case. In the experiment, the problem (1.4) is solved by considering the following regularized form:

$$\min_{\boldsymbol{a} \in \Omega, \boldsymbol{x}} \quad \frac{1}{2} \|\boldsymbol{b} - \boldsymbol{a} \otimes \boldsymbol{x}\|_2^2 + \lambda(\|\boldsymbol{x}\|_1 + \nu \|\boldsymbol{a}\|_2^2), \tag{3.1}$$

where $\Omega = \{\boldsymbol{a} \in \mathbb{R}_k^n \,|\, \boldsymbol{a} \geq 0, \sum a_i = 1\}$, and λ is an appropriate parameter.

While the loss function in (3.1) is of two variables, blind SaS deconvolution focuses on the kernel estimation. Thus, we consider the alternating minimization iteration which marginalizes the variable $\boldsymbol{x}$ by solving a convex Lasso related to $\boldsymbol{x}$, and the approximately optimal

Algorithm 2.2 Coarse-to-fine SaS Blind Deconvolution with Region Selection

Input: Observation blurred $\boldsymbol{B}$, regularization parameter λ, ν_0 and the minimum value $\nu_{\min}$, the stepsizes for the proximal alternating minimization r_1, r_2, continuation parameter $\beta_\nu > 1$, shift correction step I_c, maximum iteration $\text{iter}_{\max}$.
Output: Kernel estimation $\boldsymbol{a}_k$, and the recovered image $\boldsymbol{I}$

```
Call Algorithm 2.1 to infer a good region B_good
Construct blurred image pyramid B^s_good, s = 1, ..., S from fine-to-coarse
for s = S : −1 : 1 do
  Set k = 1, if s = S, generate random initialization a_1 and ν_1 ← ν_0, otherwise a_1 ← resize(a_{iter_max}, s) and ν_1 ← ν_{iter_max}
  repeat
    Solve x-subproblem. Set a = a_k, x_k = ∇B_good and λ_k = r_1, solve x_{k+1} = arg min_x 1/2 ||∇B_good − a ⊗ x||_2^2 + λ ||x||_1 + 1/(2λ_k) ||x − x_k||_2^2
    Solve a-subproblem using post-projection. Set ν = ν_k, μ_k = r_2 and solving problem a_{k+1} = arg min_{a∈Ω} 1/2 ||∇B_good − a ⊗ x_{k+1}||_2^2 + λν ||a||_2^2 + 1/(2μ_k) ||a − a_k||_2^2
    if k ∈ I_c then
      Kernel shift correction
    end if
    Set ν_{k+1} = max{ν_k/β_ν, ν_min}
    k ← k + 1
  until k > iter_max
end for
(Optional): Using the kernel a_k as an initialization, solve a few alternating minimization to refine the kernel using the whole deblurred image
Non-blind deblurring: Using the estimated kernel a_k, solve I = argmin_I ||B − a_k ⊗ I||_1 + λ ||∇I||_1
```

solution

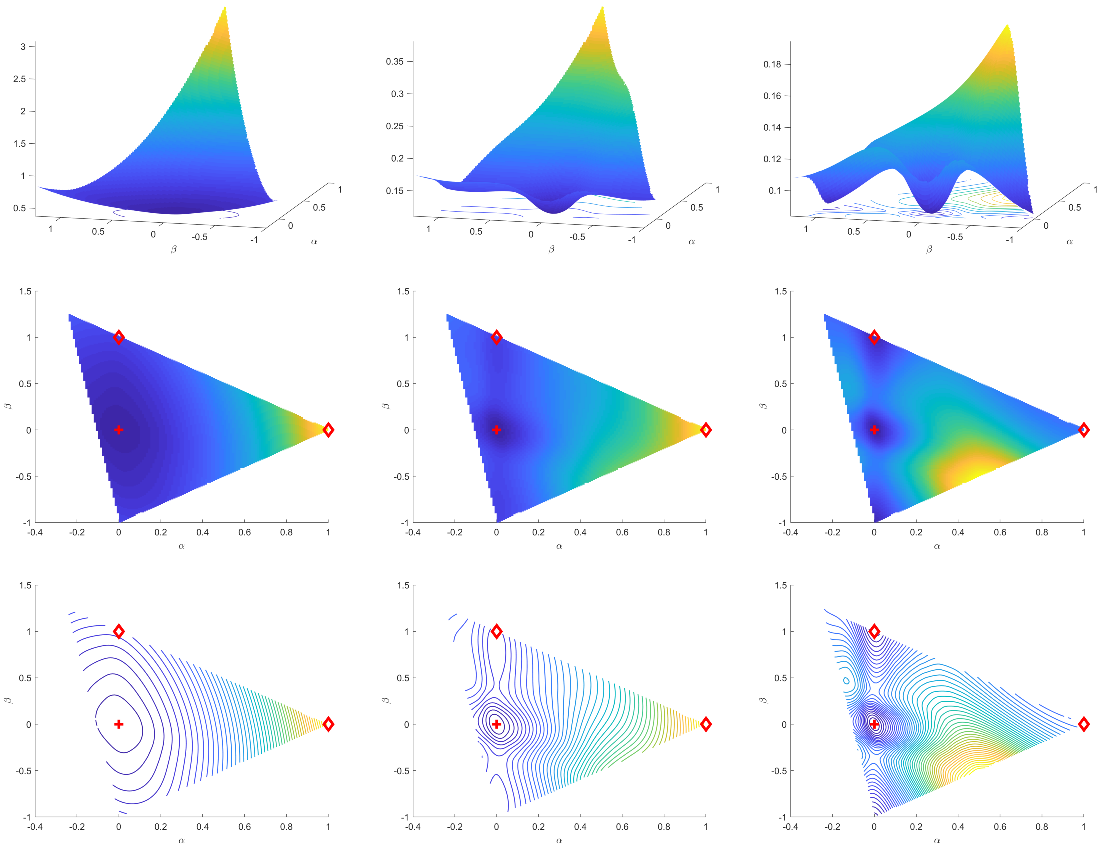
$$\boldsymbol{a}^* = \underset{\boldsymbol{a}\in\Omega}{\operatorname{argmin}}\, \phi_\nu(\boldsymbol{a}) = \underset{\boldsymbol{a}\in\Omega}{\operatorname{argmin}} \left\{ \min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{b} - \boldsymbol{a} \otimes \boldsymbol{x}\|_2^2 + \lambda(\|\boldsymbol{x}\|_1 + \nu \|\boldsymbol{a}\|_2^2) \right\}, \tag{3.2}$$

where $\lambda = 0.03$ in the experiments. Then, we visualize the landscape of the function $\phi_\nu(\boldsymbol{a})$ to see the distribution of local minimizers of the problem.

In the experiments, the 1D data with SaS structure is synthesized as follows. The kernel $\boldsymbol{a}_0 = [1, 1, \ldots, 1, 1]/10 \in \mathbb{R}^{10}$ is used in the experiment. A sparse signal $\boldsymbol{x}_0 \in \mathbb{R}^{1000}_{20}$ is generated with totally 40 nonzero Gaussian random elements which are separated by at least 20 entries. In this configuration, Theorem 1.2 states that the ground truth is a global minimizer. To visualize its landscape of the function with 10-dimensional unknown vector, we project it onto 2-dimensional plane defined by three points $\boldsymbol{a}_0$, $\boldsymbol{a}_1 = [1, 0, \ldots, 0] \in \mathbb{R}^{10}$ and $\boldsymbol{a}_2 = [1, 1, 1, 1, 1, 0, \ldots, 0]/5 \in \mathbb{R}^{10}$. The 2-dimensional function is defined as $\phi_\nu(\alpha, \beta) = \phi_\nu(\boldsymbol{a}_0 + \alpha(\boldsymbol{a}_1 - \boldsymbol{a}_0) + \beta(\boldsymbol{a}_2 - \boldsymbol{a}_0))$. To fulfill the constraint on $\boldsymbol{a} := \boldsymbol{a}_0 + \alpha(\boldsymbol{a}_1 - \boldsymbol{a}_0) + \beta(\boldsymbol{a}_2 - \boldsymbol{a}_0)$, we have that

$$1 + 9\alpha + \beta \geq 0, 1 - \alpha - \beta \geq 0, 1 - \alpha + \beta \geq 0, -1/4 \leq \alpha \leq 1, -1 \leq \beta \leq 5/4.$$

See Figure 2 for landscape visualization of $\phi_\nu(\alpha, \beta)$ with 3 different ν: *i.e.* $\nu \in \{1, 0.1, 0.01\}$. It can be seen that the landscape of the function is impacted by different values of ν. A larger ν leads to a large cost at the Dirac kernel, which makes the algorithm more likely to be away from the Delta Dirac kernel $\boldsymbol{\delta}$. When the ν is set to 1 and 0.1, in a large region around the true kernel, there is no local minimizer. In other words, as long as ν is set to a sufficiently large value, an alternating iterative scheme is likely to converge to the global minimizer, provided the initialization is reasonably close to the truth.



(a) (b) (c)

Figure 2. *2D Geometry of the function $\phi_\nu(\boldsymbol{a})$ with varying ν. (a) $\nu = 1$; (b) $\nu = 0.1$; (c) $\nu = 0.01$. The true solution $\boldsymbol{a}_0$ is marked in red $+$, and another two points $\boldsymbol{a}_1, \boldsymbol{a}_2$ are marked in red diamonds. The last row shows the contour of function* (3.2)*.*

3.2. Application on blind motion deblurring. In this section, we applied the proposed iterative blind SaS deconvolution algorithm with region selection, Alg. 2.2, to solve the problem of blind motion deblurring. We consider two datasets: Hu *et al.*'s dataset [15] and Sun *et al.*'s dataset [31]. There are 120 burred images synthesized from 10 clean image with 12 kernels in Hu *et al.*'s dataset[1]. No true images are provided in Hu *et al.*'s dataset. There are 640 images in Sun's dataset [31], which are synthesized from 80 clean images convolved by 8 kernels from [21] and contaminated by 1% Gaussian noise[2]. Both true images and true kernels are available in Sun *et al.*'s dataset. See Figure 3 for the 8 blur kernels and some sample images from Sun *et al.*'s dataset. The parameters of Alg. 2.2 are set as follows. The maximum iteration is set to 1000, λ is set to 0.03, the stepsizes for the proximal alternating minimization are set as $r_1 = r_2 = 1e-3$, ν_0 is set to 0.1, $\nu_{\min} = 0.01$ and the kernel shift correction step is set to 200. Two experiments are conducted in this section. The first is to examine the effectiveness of region selection technique described in Alg. 2.1 for blind motion deblurring, and the second is for quantitative performance comparison of Alg. 2.2 and the other very related methods on Sun *et al.*'s dataset.

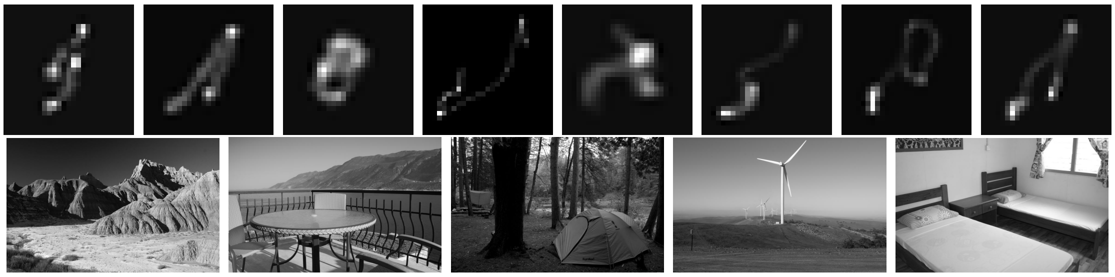


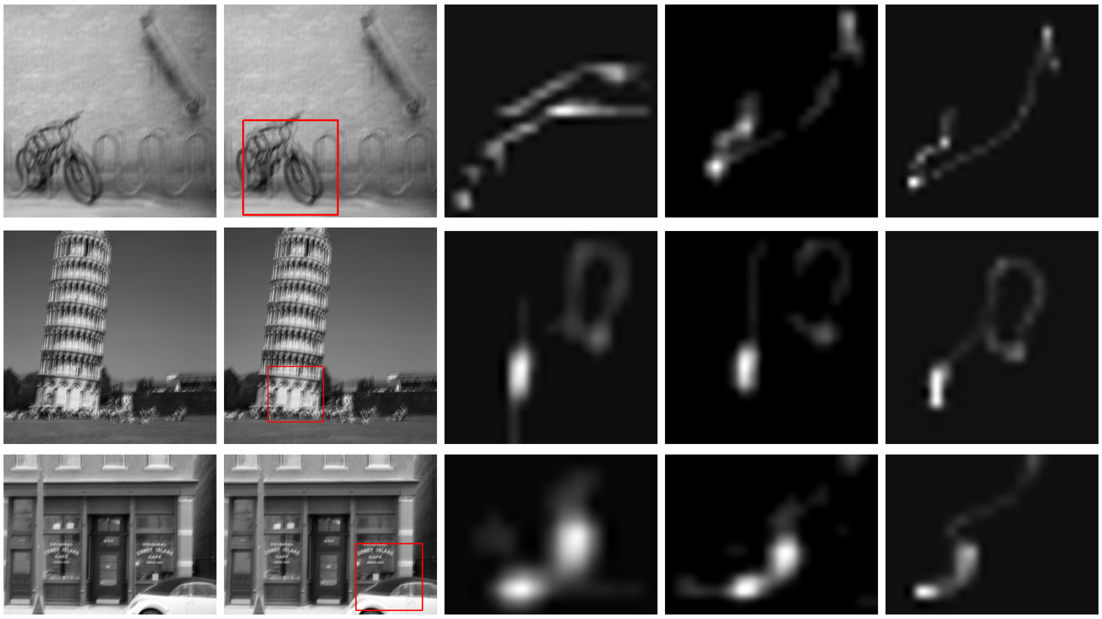
Figure 3. *The blur kernels (top row) and sample images (bottom row) from Sun* et al.*'s dataset [31].*

3.2.1. Performance evaluation of region selection. The main contribution in Alg. 2.2 lies in the introduction of Alg. 2.1-based region selection. Such a motivation comes from the analysis which shows the importance of sufficient separation of image edges in the region. To show the benefit of such a region selection to kernel estimation, we run the experiments on Hu *et al.*'s dataset [15]. Following the same setting as [15], a region of size 200×200 is extracted by Alg. 2.1 where the smoothing Gaussian kernel $\boldsymbol{g}$ used in Alg. 2.1 is of the same size as the true kernel. The kernels estimated using Alg. 2.2 on the extracted region are compared to their counterparts estimated using Alg. 2.2 on the whole image (without regions selection). See Figure 4 for the visual comparison of the estimated kernels on some sample images. It can be seen that the kernels estimated on the regions selected by Alg. 2.1, the rectangular regions bounded by red box, are certainly closer to the true kernels.

To show the quality improvement of the deblurred results brought by region selection via Alg. 2.1. An experiment is conducted to show the comparison of the visual quality of the image deblurred by the kernel estimated by Alg. 2.2. The results from the proposed method is compared to several existing methods which also call some edge/region selection module during the iteration, including Fergus *et al.* [12], Xu *et al.* [34] and Hu and Yang [15]. Fergus *et al.* selected the regions with the highest edge energy. Xu *et al.* [34] removes the edges with small magnitude. Hu and Yang [15] learned a discrimination map and selected the region

[1] The dataset is available from https://eng.ucmerced.edu/people/zhu/ECCV12_dataset.zip.

[2] The dataset is available from https://cs.brown.edu/people/lbsun/deblur2013/all_deblurred_results_SunICCP2013.zip.



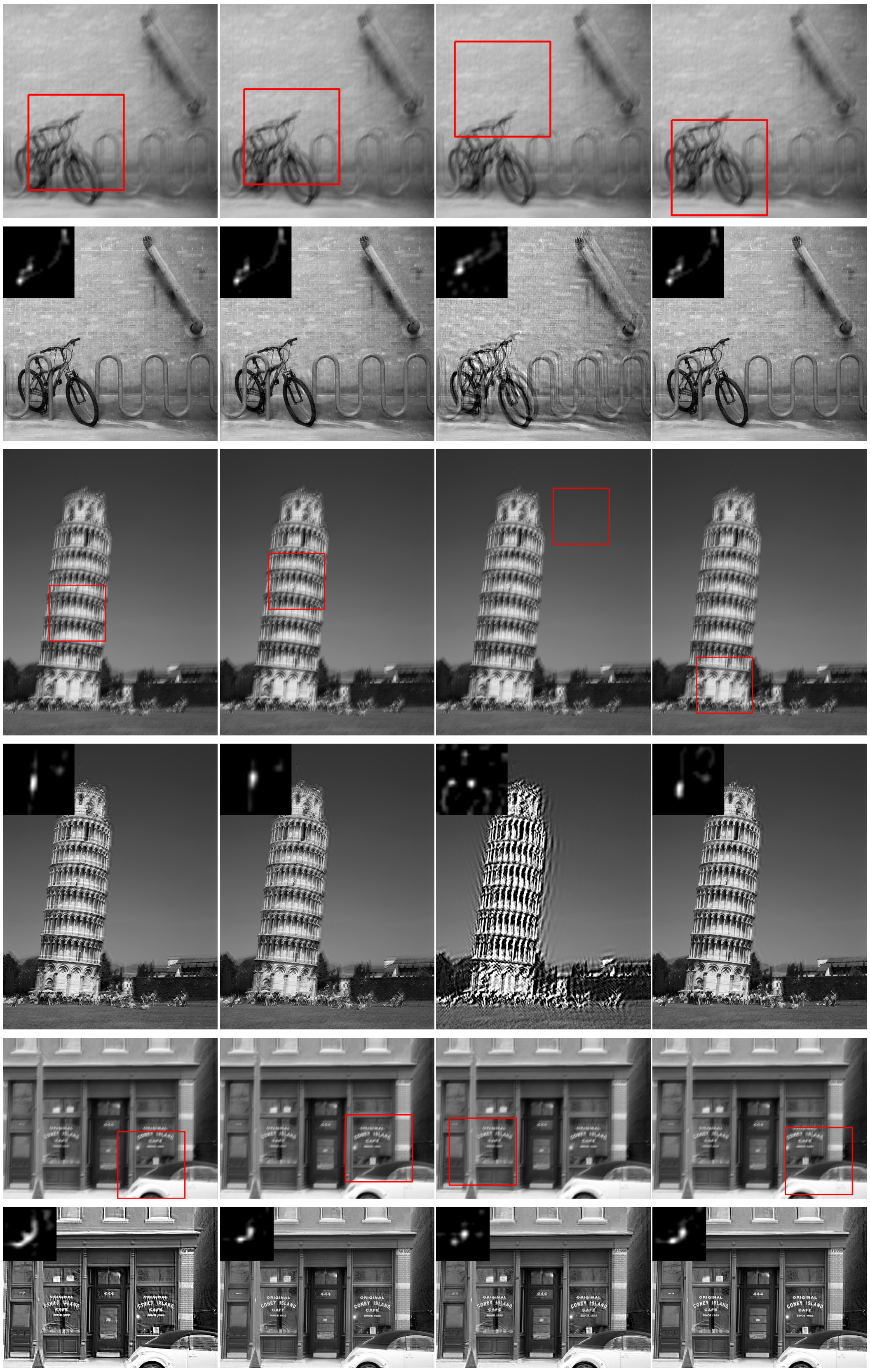
(a) (b) (c) (d) (e)

Figure 4. *Comparison between the kernels estimated by Alg 2.2 on the regions selected by Alg. 2.1 and the ones estimated on the whole image. (a) Input blurred image from Hu* et al.*'s dataset; (b) The regions selected by Alg. 2.1, bounded in red box; (c) The kernel estimated on the whole image; (d) The kernels estimated on selected regions; (e) The true kernel.*

with the highest score of kernel recovery quality. See Figure 5 for visual comparison of the deblurred results from the methods with different edge/region selection techniques. It can be seen that overall, the proposed method perform consistently on the sample images and yields the results with the best visual quality.

3.2.2. Quantitative evaluation for blind motion deblurring . In this section, a quantitative evaluation of the proposed method is conducted on Sun *et al.*'s dataset [31]. Following the common practice, the measurement on the accuracy of the kernel estimation is done by examining the quality of the image recovered by the estimated kernel using some widely-used non-blind deblurring method. Following most works, the robust non-blind deconvolution method [34] with default parameter setting is called for deblurring the image using the estimated kernel. Three metrics are used to quantitatively measure the quality of the deblurred images: mean PSNR, mean SSIM and mean of error ratios [21]. The last metric accounts for the difficulty of the non-blind deconvolution step considered. An error ratio larger than 3 is deemed visually unacceptable as [21] did.

In experiment, we compare Alg 2.2 against other existing approaches, including Krishnan *et al.* [18], Cho & Lee [11], Levin *et al.* [22], Xu & Jia [34] and Sun *et al.* [31]. Except Krishnan *et al.* [18] which used a normalized TV regularization on images for blind deblurring, all other methods methods contain edge/region selection for more accurate blind deblurring. See Table 1 for quantitative comparison of different methods on the Sun *et al.*'s dataset. Note that the failed cases in Table 1 refers to the number of the cases where relative error ratio



(a) Fergus *et al.* [12] (b) Xu & Jia [34] (c) Hu & Yang [15] (d) Proposed

Figure 5. *Visual comparison of the results, de-blurred images and estimated kernels, using the methods with different region selection techniques. For each two rows, the first row shows the input and selected regions, and the second row shows the de-blurred images and the estimated kernels on the top left of the images.*

is larger than 3. From Table 1, the proposed Alg 2.2 is the best in terms of PSNR and No. of failing cases, and is the second best in terms of SSIM and Error Ratio. Our algorithm outperforms other methods in terms of failed cases. For the comparison of visual quality, See Figure 6 for the comparison of different methods on some sample images. It can be seen that overall, the proposed one yields the results with best visual quality, with more details and less artifacts. See Figure 7 for the curve of cumulative error ratio, and the proposed one is the top performer.

In summary, for blind motion deblurring, edge/region selection plays an important role in blind motion deblurring. Among all edge/region selection technique, the proposed one achieves the best performance. This experiments show the effectiveness of region selection built on the analysis conducted in this paper, i.e. the relationship between the separation of non-zero entries of the signal and the estimation accuracy of the kernel.

Table 1
Average PSNR/SSIM and Error Ratio for Sun et al.*'s dataset [31].*

Different approaches		-	Edge selection				Region selection	
	Known kernel	Krishnan *et al.* [18]	Levin *et al.* [22]	Cho & Lee [11]	Xu & Jia [34]	Sun *et al.* [31]	Hu & Yang [15]	Ours
PSNR	32.4204	23.8157	25.6754	26.9552	28.2503	29.4993	28.5242	**29.5939**
SSIM	0.9491	0.8188	0.8657	0.8775	0.9157	**0.9232**	0.8959	0.9205
Error Ratio	1	2.9344	2.2561	2.2671	1.7327	**1.3546**	1.822	1.4252
Failed cases	-	196	73	135	37	23	60	**14**

4. Conclusion.

This paper studys an ℓ_1-norm relating regularization model for solving the problem of blind SaS deconvolution. It is shown that the model is sound when the sparse signal is sufficiently separated among non-zero entries. In the presence of noise, the analysis also reveals that the global minimizers of the studied model remain a good approximation to the truth, when the signal can be well approximated by a well-separated sparse signal. Such a study inspires a region selection technique for blind SaS deconvolution. The experiments on blind image deblurring show the benefit of region selection. In future, we plan to study the soundness of other widely-used and efficient models to see whether we could further improve the theoretical results. We would also like to investigate the landscape of the ℓ_1-norm relating regularization models and the provable algorithms for blind SaS deconvolution.

REFERENCES

[1] A. Ahmed, B. Recht, and J. Romberg, *Blind deconvolution using convex programming*, IEEE Transactions on Information Theory, 60 (2013), pp. 1711–1732.
[2] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, *Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the kurdyka-łojasiewicz inequality*, Mathematics of Operations Research, (2010), pp. 438–457.
[3] J.-F. Cai, H. Ji, C. Liu, and Z. Shen, *Blind motion deblurring from a single image using sparse approximation*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 104–111.

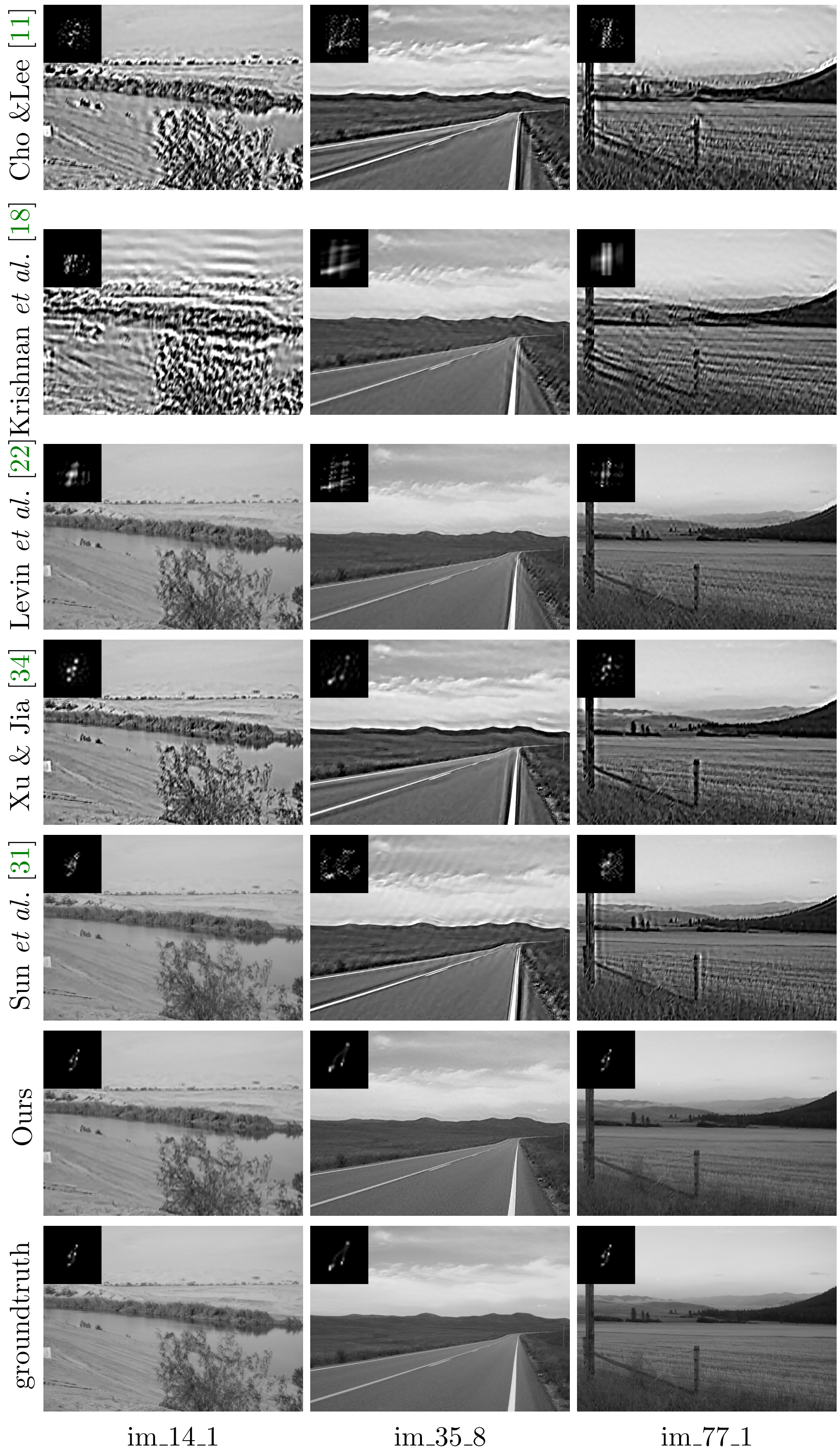


Figure 6. *Visual comparison of the results from different methods on the images from [31]. Zoom-in for better visual inspection.*

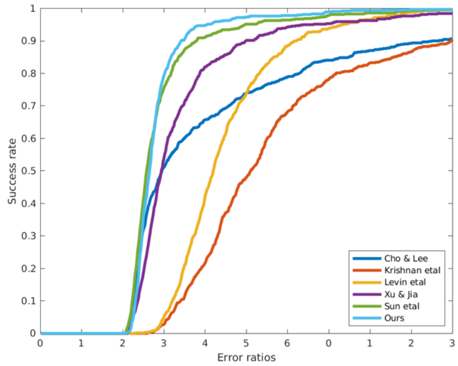


Figure 7. *Cumulative error ratios of different methods on Sun* et al.*'s dataset [31].*

[4] J.-F. Cai, H. Ji, C. Liu, and Z. Shen, *Framelet-based blind motion deblurring from a single image*, IEEE Transactions on Image Processing, 21 (2011), pp. 562–572.

[5] E. Candes and J. Romberg, *Sparsity and incoherence in compressive sampling*, Inverse Problems, 23 (2007), p. 969.

[6] E. J. Candes, Y. C. Eldar, T. Strohmer, and V. Voroninski, *Phase retrieval via matrix completion*, SIAM Review, 57 (2015), pp. 225–251.

[7] E. J. Candès and M. B. Wakin, *An introduction to compressive sampling*, IEEE Signal Processing Magazine, 25 (2008), pp. 21–30.

[8] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, Journal of Mathematical Imaging and Vision, 40 (2011), pp. 120–145.

[9] T. F. Chan and C.-K. Wong, *Total variation blind deconvolution*, IEEE Transactions on Image Processing, 7 (1998), pp. 370–375.

[10] J. Chen, R. Lin, H. Wang, J. Meng, H. Zheng, and L. Song, *Blind-deconvolution optical-resolution photoacoustic microscopy in vivo*, Optics Express, 21 (2013), pp. 7316–7327.

[11] S. Cho and S. Lee, *Fast motion deblurring*, ACM Transactions on Graphics (TOG), 28 (2009), pp. 1–8.

[12] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, *Removing camera shake from a single photograph*, ACM Transactions on Graphics, 25 (2006), pp. 787–794.

[13] D. Gong, M. Tan, Y. Zhang, A. V. D. Hengel, and Q. Shi, *Blind Image Deconvolution by Automatic Gradient Activation*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1827–1836.

[14] T. J. Holmes, *Blind deconvolution of quantum-limited incoherent imagery: maximum-likelihood approach*, JOSA A, 9 (1992), pp. 1052–1061.

[15] Z. Hu and M.-H. Yang, *Good regions to deblur*, in European Conference on Computer Vision, Springer, 2012, pp. 59–72.

[16] S. M. Jefferies and J. C. Christou, *Restoration of astronomical images by iterative blind deconvolution*, The Astrophysical Journal, 415 (1993), p. 862.

[17] M. Jin, S. Roth, and P. Favaro, *Normalized blind deconvolution*, in Proceedings of the European Conference on Computer Vision, 2018, pp. 668–684.

[18] D. Krishnan, T. Tay, and R. Fergus, *Blind deconvolution using a normalized sparsity measure*, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, 2011, pp. 233–240.

[19] D. Kundur and D. Hatzinakos, *Blind image deconvolution*, IEEE Signal Processing Magazine, 13 (1996), pp. 43–64.

[20] H.-W. Kuo, Y. Lau, Y. Zhang, and J. Wright, *Geometry and symmetry in short-and-sparse deconvolution*, in International Conference on Machine Learning, PMLR, 2019, pp. 3570–3580.

[21] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, *Understanding and evaluating blind deconvolu-*

tion algorithms, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 1964–1971.

[22] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, *Efficient marginal likelihood optimization in blind deconvolution*, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, 2011, pp. 2657–2664.

[23] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, *Understanding blind deconvolution algorithms*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 33 (2011), pp. 2354–2367.

[24] X. Li, S. Ling, T. Strohmer, and K. Wei, *Rapid, robust, and reliable blind deconvolution via nonconvex optimization*, Applied and Computational Harmonic Analysis, 47 (2019), pp. 893–934.

[25] Y. Li, K. Lee, and Y. Bresler, *Identifiability and stability in blind deconvolution under minimal assumptions*, IEEE Transactions on Information Theory, 63 (2017), pp. 4619–4633.

[26] J. Pan, R. Liu, Z. Su, and X. Gu, *Kernel estimation from salient structure for robust motion deblurring*, Signal Processing: Image Communication, 28 (2013), pp. 1156–1170.

[27] D. Perrone and P. Favaro, *A clearer picture of total variation blind deconvolution*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 38 (2015), pp. 1041–1055.

[28] R. T. Rockafellar and R. J.-B. Wets, *Variational analysis*, vol. 317, Springer Science & Business Media, 2009.

[29] J. Romberg, *Imaging via compressive sampling*, IEEE Signal Processing Magazine, 25 (2008), pp. 14–20.

[30] Q. Shan, J. Jia, and A. Agarwala, *High-quality motion deblurring from a single image*, ACM Transactions on Graphics, 27 (2008), pp. 1–10.

[31] L. Sun, S. Cho, J. Wang, and J. Hays, *Edge-based blur kernel estimation using patch priors*, in Proceedings of the IEEE International Conference on Computational Photography, 2013, pp. 1–8.

[32] S. Vorontsov, V. Strakhov, S. Jefferies, and K. Borelli, *Deconvolution of astronomical images using sor with adaptive relaxation*, Optics Express, 19 (2011), pp. 13509–13524.

[33] D. Wipf and H. Zhang, *Revisiting bayesian blind deconvolution*, Journal of Machine Learning Research, 15 (2014), pp. 3775–3814.

[34] L. Xu and J. Jia, *Two-phase kernel estimation for robust motion deblurring*, in European Conference on Computer Vision, Springer, 2010, pp. 157–170.

[35] L. Yang and H. Ji, *A variational EM framework with adaptive edge selection for blind motion deblurring*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 10167–10176.

[36] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, *Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing*, SIAM Journal on Imaging Sciences, 1 (2008), pp. 143–168.

[37] Y. Zhang, H.-W. Kuo, and J. Wright, *Structured local minima in sparse blind deconvolution*, in Proceedings of Advances in Neural Information Processing Systems, 2018, pp. 2322–2331.

[38] Y. Zhang, Y. Lau, H.-w. Kuo, S. Cheung, A. Pasupathy, and J. Wright, *On the global geometry of sphere-constrained sparse blind deconvolution*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4894–4902.