DATA-DRIVEN TIGHT FRAME CONSTRUCTION FOR IMPULSIVE NOISE REMOVAL*

Yang Chen   and   Chunlin Wu[1]

School of Mathematical Science, Nankai University, Tianjin 300071, China
Email: yang_chen@mail.nankai.edu.cn, wucl@nankai.edu.cn

Abstract

The method of data-driven tight frame has been shown very useful in image restoration problems. We consider in this paper extending this important technique, by incorporating L_1 data fidelity into the original data-driven model, for removing impulsive noise which is a very common and basic type of noise in image data. The model contains three variables and can be solved through an efficient iterative alternating minimization algorithm in patch implementation, where the tight frame is dynamically updated. It constructs a tight frame system from the input corrupted image adaptively, and then removes impulsive noise by the derived system. We also show that the sequence generated by our algorithm converges globally to a stationary point of the optimization model. Numerical experiments and comparisons demonstrate that our approach performs well for various kinds of images. This benefits from its data-driven nature and the learned tight frames from input images capture richer image structures adaptively.

Mathematics subject classification: 68U10, 94A08.
Key words: Tight frame, Impulsive noise, Sparse approximation, Data-driven, Convergence analysis.

1. Introduction

Sparse approximation and representation are very powerful for many signal processing tasks, such as signal compression and image restoration. Sparse approximation models an image as a linear combination of a small number of elements of some system. Such system can be either a basis or an over-complete system, among which the wavelet tight frame [17] has been very successfully used in image restoration [8, 15, 18, 26]. Wavelet tight frames are able to approximate piecewise smooth signals efficiently with only few non-zero wavelet coefficients. Computationally, tight frames benefit from their efficient decomposition and reconstruction schemes. Numerous kinds of tight frames, including curvelets [12], ridgelts [11], framelets [17,36] and many others have been proposed for signal and image sparse representation. However, one certain tight frame cannot always perform well for all kinds of images. Naturally, it is better to design specific tight frame representation for a given image. The tight frame learned from the given image may achieve better performance for sparse approximation. Due to this nature, this kind of techniques are adaptive to input image data and hence work well for various types of images.

Cai et al. proposed in [9] a variational model under this idea, with an L_2-norm fidelity term and an L_0-norm regularization term. It is the first paper to propose data-driven tight

frame model for image restoration, which shows comparable performance and runs much faster than general dictionary learning methods like K-SVD [21]. Bao et al. showed the sub-sequence convergence of the iterative algorithm in [3], and gave a new globally convergent algorithm. Soon later, this data-driven tight frame model has been improved and applied to various problems [13, 22, 27, 32, 33, 37, 42, 44]. However, as far as we know, these data-driven models mainly use L_2 fidelity, which is not suitable for the basic and significant impulsive noise removal problems.

Impulsive noise is usually divided into salt-and-pepper noise and random-valued noise. It is often generated by electromagnetic interference as well as faults and defects in communication system. It may also occur when the electrical switches and relays in the communication system change state [7]. Unlike the additive Gaussian noise, which affects all pixels of an image, salt-and-pepper noise (or random-valued noise) corrupts a portion of the pixels with minimal or maximal intensities (or random-valued intensities) while keeping remaining pixels unchanged. Hence, additive Gaussian noise removal algorithms are not suitable for impulsive noise removal.

Many impulsive noise removal algorithms have been proposed in recent decades, most of which can be categorized into nonlinear digital filter based methods [14, 19, 24, 25, 39] and variational regularization based approaches [2, 23, 29, 31, 40, 41, 43]. Nonlinear digital filter based methods improve median filter through weighting and adaptive techniques. Variational regularization approaches use total variation and some fidelity terms considering the noise statistics. These methods can achieve good performance in most cases. However, nonlinear digital filters are less capable of distinguishing noisy pixels from non-noisy pixels in edges or textured regions. Variational regularization term may introduce undesirable stair-casing effects. Considering the great successes of data-driven tight frame techniques in denoising, in this paper we are interested in extending [9] for impulsive noise removal. This yields an L_0 balanced model [9] with L_1 data fidelity, which is convenient for optimization and implementation. We will give an efficient alternating minimization algorithm to optimize the objective function and prove its global convergence. The iterative algorithm constructs discrete tight frames adapted to input images, which are then applied to the noise removal step. We mention that, a very recent paper [28], although considered data-driven tight frames for mixed Gaussian and impulsive noise removal, actually did not formulate an optimization model for data-driven tight frame construction in the presence of impulsive noise. Instead, their method uses directly the data-driven tight frame construction in [9] (with L_2 fidelity) as an independent package combined with an analysis approach [10] for denoising. This combination, if iteratively performed, yields a complicated two-loop procedure.

The rest of the paper is organized as follows. In Section 2, we briefly review the concept of tight frame and the existing data-driven tight frame construction. We propose our optimization model for impulsive noise removal in Section 3. It is implemented in patch space and solved by a proximal alternating minimization procedure. In Section 4, we show that the iterates of the proposed algorithm converge globally to a stationary point of the minimization model. Some numerical experiments are provided in Section 5. We conclude the paper in Section 6.

2. A Brief Review of Tight Frame and Data-driven Tight Frame Construction

In this section, we briefly introduce the tight frames and the data-driven tight frames construction proposed in [9]. Interested readers are referred to [10, 38] for more details.

Suppose that $\mathcal{H}$ is a Hilbert space. Let $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ denote the standard inner product, a

norm of a Hilbert space $\mathcal{H}$, respectively. A sequence $\{x_n\} \subset \mathcal{H}$ is a tight frame for $\mathcal{H}$ if

$$\|x\|^2 = \sum_n |\langle x, x_n \rangle|^2, \quad \text{for any } x \in \mathcal{H}.$$

That is, given a tight frame $\{x_n\}$, we have $x = \sum_n \langle x, x_n \rangle x_n$, $x \in \mathcal{H}$. Thus, a tight frame $\{x_n\}$ can be regarded as generalization of orthonormal bases.

There are two operators associated with a tight frame: the analysis operator W defined by

$$W : \ \mathcal{H} \mapsto \ell_2(\mathbb{N}),$$
$$x \mapsto \{\langle x, x_n \rangle\},$$

and its adjoint operator W^{T} referred to the synthesis operator:

$$W^{\mathrm{T}} : \ \ell_2(\mathbb{N}) \mapsto \mathcal{H},$$
$$\{a_n\} \mapsto \sum_n a_n x_n.$$

The sequence $Wx := \{\langle x, x_n \rangle\}$ is called the canonical tight frame coefficient sequence. Then, the sequence $\{x_n\} \subset \mathcal{H}$ is a tight frame if and only if $W^{\mathrm{T}}W = I$, where I is the identity operator of $\mathcal{H}$. For simplicity, we will only discuss the single-level un-decimal wavelet tight frame system, which is generated by all integer shifts of a set of filters $\{a_i\}_{i=1}^r$.

For any given filter $a \in \ell_2(\mathbb{Z})$, define its associated convolution operator $\mathcal{S}_a : \ell_2(\mathbb{Z}) \to \ell_2(\mathbb{Z})$ by

$$[\mathcal{S}_a v](n) := [a * v](n) = \sum_{k \in \mathbb{Z}} a(n-k)v(k), \quad \forall v \in \ell_2(\mathbb{Z}).$$

Then, for a set of filters $\{a_i\}_{i=1}^r$, its analysis operator W can be defined by

$$W = [\mathcal{S}_{a_1(-\cdot)}^{\mathrm{T}}, \mathcal{S}_{a_2(-\cdot)}^{\mathrm{T}}, \cdots, \mathcal{S}_{a_r(-\cdot)}^{\mathrm{T}}]^{\mathrm{T}}. \tag{2.1}$$

Its corresponding synthesis operator reads

$$W^{\mathrm{T}} = [\mathcal{S}_{a_1}, \mathcal{S}_{a_2}, \cdots, \mathcal{S}_{a_r}]. \tag{2.2}$$

The rows of W form a tight frame for $\ell_2(\mathbb{Z})$ if and only if $W^{\mathrm{T}}W = I$.

The data-driven tight frame construction model proposed in [9] is

$$\min_{c, \{a_i\}_{i=1}^r} \|c - Wg\|_2^2 + \lambda \|c\|_0 \quad \text{s.t.} \quad W^{\mathrm{T}}W = I, \tag{2.3}$$

where g denotes an input image, c is the coefficient vector which sparsely approximates the canonical tight frame coefficients Wg, and $\|c\|_0$ stands for the number of non-zero elements of c. Note that we here use W to denote the analysis operator related to tight frame generated by filters $\{a_i\}_{i=1}^r$.

It is proved in [9, Proposition 2] that (2.3) is equivalent to

$$\min_{c, \{a_i\}_{i=1}^r} \|g - W^{\mathrm{T}}c\|_2^2 + \|(I - WW^{\mathrm{T}})c\|_2^2 + \lambda \|c\|_0 \quad \text{s.t.} \quad W^{\mathrm{T}}W = I, \tag{2.4}$$

which is a balanced approach model assembled with L_2 fidelity term.

3. The Proposed Model and Algorithm

3.1. The proposed model

According to the Bayes formula, the L_2 fidelity term in (2.4) is suitable for removing additive Gaussian white noise. In the case of impulsive noise removal, we adopt the L_1 fidelity term to reformulate the minimization model. It is shown that the L_1 fidelity term can accurately fit undamaged pixels and perfectly regularize corrupted pixels [30], in the context of variational regularization.

We therefore propose the following model:

$$\min_{c,\{a_i\}_{i=1}^r} \|g - W^{\mathrm{T}}c\|_1 + \tau\|(I - WW^{\mathrm{T}})c\|_2^2 + \lambda\|c\|_0, \quad \text{s.t.} \quad W^{\mathrm{T}}W = I.$$

It gives an approximate model:

$$\min_{c,\{a_i\}_{i=1}^r,f} \|f\|_1 + \tau\|(I - WW^{\mathrm{T}})c\|_2^2 + \lambda\|c\|_0 + \tau\|f - g + W^{\mathrm{T}}c\|_2^2, \quad \text{s.t.} \quad W^{\mathrm{T}}W = I, \quad (3.1)$$

which is, by the first statement in Proposition 3.1, equivalent to

$$\min_{c,\{a_i\}_{i=1}^r,f} \|f\|_1 + \tau\|c - W(g - f)\|_2^2 + \lambda\|c\|_0, \quad \text{s.t.} \quad W^{\mathrm{T}}W = I. \quad (3.2)$$

Thus the minimization (3.2) can be considered as a balanced approach as in (3.1). The following proposition, especially its second statement, is essential for us to design efficient numerical algorithm to solve our minimization model.

Proposition 3.1. *Let W denote a tight frame. Then*

1. $\|(I - WW^{\mathrm{T}})c\|_2^2 + \|f - g + W^{\mathrm{T}}c\|_2^2 = \|c - W(g - f)\|_2^2$;

2. $\arg\min_{f} \|f\|_1 + \tau\|c - W(g - f)\|_2^2 = \arg\min_{f} \|f\|_1 + \tau\|f - (g - W^{\mathrm{T}}c)\|_2^2.$

Proof. Part 1 is contained in [9, Proposition 2]. For Part 2, by the fact that $W^{\mathrm{T}}W = I$, we have

$$\arg\min_{f} \|f\|_1 + \tau\|c - W(g - f)\|_2^2$$

$$= \arg\min_{f} \|f\|_1 + \tau(c^{\mathrm{T}}c + (g - f)^{\mathrm{T}}W^{\mathrm{T}}W(g - f) - 2c^{\mathrm{T}}W(g - f))$$

$$= \arg\min_{f} \|f\|_1 + \tau(f^{\mathrm{T}}f - 2f^{\mathrm{T}}(g - W^{\mathrm{T}}c) + g^{\mathrm{T}}g + c^{\mathrm{T}}WW^{\mathrm{T}}c - 2g^{\mathrm{T}}W^{\mathrm{T}}c)$$

$$= \arg\min_{f} \|f\|_1 + \tau\|f - (g - W^{\mathrm{T}}c)\|_2^2,$$

where the second equality is due to that both W and c are constants in this minimization problem. $\square$

3.2. Model implementation and an iterative algorithm

The minimization problem with constraint condition $W^{\mathrm{T}}W = I$ is rather challenging to solve, if there is no special structural assumption for the tight frame filters. Thus we adopt the same way as [9, Proposition 3]. A certain special class of filters satisfying the orthogonality

constraints are taken into account. Hence the filters are $\{a_i\}_{i=1}^{m^2}$ with support on $\mathbb{Z}^2 \bigcap [1,m]^2$ satisfying the following constraints:

$$\langle a_i, a_j \rangle = \sum_{k \in \mathbb{Z}^2 \bigcap [1,m]^2} a_i(k) a_j(k) = \frac{1}{m^2} \delta_{i-j}, \quad 1 \le i, j \le m^2, \tag{3.3}$$

which ensure that $W^{\mathrm{T}} W = I$. Moreover, we implement our minimization model using image patches, like [9,21]. We partition the noisy image into patches with overlaps and then the model (3.2) is reformulated into an image patch based problem.

Let $\{\mathbf{g}_l\}_{l=1}^N \subset \mathbb{R}^{m^2}$ denote the set of all image patches of size $m \times m$ regularly sampled from the image g. Let $\mathbf{a}_j$ be the vector form of a_j by concatenating all of its columns. For each patch vector $\mathbf{g}_l$, let $\mathbf{c}_l$ denote the corresponding coefficient vector. Define four matrices as follows:

$$\begin{cases} G := \dfrac{1}{m}[\mathbf{g}_1, \mathbf{g}_2, \cdots, \mathbf{g}_N] \in \mathbb{R}^{m^2 \times N}, \\[2mm] F := \dfrac{1}{m}[\mathbf{f}_1, \mathbf{f}_2, \cdots, \mathbf{f}_N] \in \mathbb{R}^{m^2 \times N}, \\[2mm] D := m[\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_{m^2}] \in \mathbb{R}^{m^2 \times m^2}, \\[2mm] C := [\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_N] \in \mathbb{R}^{m^2 \times N}. \end{cases} \tag{3.4}$$

Similar to [9] and [3], the minimization (3.2) can be implemented as

$$\min_{D \in \mathbb{R}^{m^2 \times m^2}, \, C, F \in \mathbb{R}^{m^2 \times N}} \|F\|_{e_1} + \tau \|C - D^{\mathrm{T}}(G - F)\|^2 + \lambda \|C\|_0, \quad \text{s.t.} \quad D^{\mathrm{T}} D = I, \tag{3.5}$$

where, for any matrix X, $\|X\|_{e_1} = \sum_{i,j} |X_{ij}|$ and $\|X\|$ denotes the Frobenius norm of X.

In order to simplify the notations, we define $\chi = \{D \in \mathbb{R}^{m^2 \times m^2} : D^{\mathrm{T}} D = I_{m^2}\}$, $\Omega_{CF} = \mathbb{R}^{m^2 \times N}$, $\Omega_D = \mathbb{R}^{m^2 \times m^2}$, $Z = (C, D, F)$ and $\Omega_Z = (\Omega_{CF}, \Omega_D, \Omega_{CF})$. We also denote

$$r(C) = \lambda \|C\|_0, \ \ s(D) = I_\chi(D), \ \ t(F) = \|F\|_{e_1}, \ \ Q(C, D, F) = \tau \|C - D^{\mathrm{T}}(G - F)\|^2,$$

where $I_\chi(D) = 0$ for $D \in \chi$, and $I_\chi(D) = +\infty$ otherwise. Then, the minimization (3.5) can be written as

$$\min_{C, F \in \Omega_{CF}, D \in \Omega_D} L(C, D, F) := r(C) + s(D) + t(F) + Q(C, D, F). \tag{3.6}$$

Since the set χ is compact, it is not difficult to see that $L(C, D, F)$ in (3.6) is coercive. Together with the lower semi-continuity of $L(C, D, F)$, we claim that the minimization problem (3.6) always has a solution. As $L(C, D, F)$ is a non-convex function, the uniqueness of the solution cannot be guaranteed.

Then we consider how to solve (3.6). There are three unknowns in (3.6). The first one is the coefficient matrix C which sparsely approximates the canonical tight frame coefficient $D^{\mathrm{T}}(G - F)$, the second one is the filters matrix D that generates a tight frame, and the third one is a matrix F in image patch domain which can be considered as noise. A standard way to solve the model (3.6) is an iterative alternating minimization, which updates the estimation of C, D and F alternatively. To guarantee convergence, a proximal technique can be used.

Given (C_k, D_k, F_k), the next iteration updates it via the following scheme:

$$\begin{cases} C_{k+1} \in \arg\min_C \ L(C, D_k, F_k) + \alpha_k \|C - C_k\|^2, \\[2mm] D_{k+1} \in \arg\min_D \ L(C_{k+1}, D, F_k) + \beta_k \|D - D_k\|^2, \\[2mm] F_{k+1} \in \arg\min_F \ L(C_{k+1}, D_{k+1}, F) + \gamma_k \|F - F_k\|^2, \end{cases} \tag{3.7}$$

where $\alpha_k, \beta_k, \gamma_k \in (a, b)$ and a, b are positive constants for proximal terms. Specifically, we need to solve the following three sub-problems:

$$C_{k+1} \in \arg\min_C \ \tau\|C - D_k^T(G - F_k)\|^2 + \lambda\|C\|_0 + \alpha_k\|C - C_k\|^2; \tag{3.8a}$$

$$D_{k+1} \in \arg\min_D \ \tau\|C_{k+1} - D^T(G - F_k)\|^2 + \beta_k\|D - D_k\|^2, \quad \text{s.t.} \quad D^T D = I; \tag{3.8b}$$

$$F_{k+1} \in \arg\min_F \ \|F\|_{e_1} + \tau\|F - (G - D_{k+1}C_{k+1})\|^2 + \gamma_k\|F - F_k\|^2, \tag{3.8c}$$

where the third sub-problem is obtained according to Proposition 3.1. The first sub-problem of C_{k+1} is known to have a closed form solution via hard thresholding; the second about D_{k+1} can be solved through a Singular Value Decomposition (SVD) [9]; and the third one of F_{k+1} can be solved via soft thresholding. For the sake of completeness, we include some details here.

For the first sub-problem (3.8a), we let $T_\kappa : \mathbb{R} \to \mathbb{R}$ denote the hard thresholding operator which reads:

$$T_\kappa(u) = \begin{cases} u, & \text{if } |u| > \kappa; \\ 0, & \text{otherwise.} \end{cases}$$

When it acts on a matrix, it represents an entry by entry thresholding operation. Then (3.8a) has the following solution

$$C_{k+1} = T_{\sqrt{\lambda/(\tau+\alpha_k)}}\left(\frac{\tau D_k^T(G - F_k) + \alpha_k C_k}{\tau + \alpha_k}\right).$$

By [9, Theorem 4] and [45], the second sub-problem (3.8b) also has an explicit solution $D_{k+1} = U_k V_k^T$, where U_k and V_k are from the SVD decomposition of $\tau(G - F_k)C_{k+1}^T + \beta_k D_k$, i.e., $\tau(G - F_k)C_{k+1}^T + \beta_k D_k = U_k \Sigma_k V_k^T$.

As for the third sub-problem (3.8c), we let $\mathcal{T}_\kappa : \mathbb{R} \to \mathbb{R}$ be the soft thresholding operator [20] defined by

$$\mathcal{T}_\kappa(u) = \begin{cases} u + \kappa, & \text{if } u < -\kappa; \\ 0, & \text{if } |u| \le \kappa; \\ u - \kappa, & \text{if } u > \kappa. \end{cases}$$

If $\mathcal{T}_\kappa$ acts on a matrix, it means that the soft thresholding operates entry by entry. Then the unique solution F_{k+1} of (3.8c) is given by

$$F_{k+1} = \mathcal{T}_{1/(2(\tau+\gamma_k))}\left(\frac{\tau(G - D_{k+1}C_{k+1}) + \gamma_k F_k}{\tau + \gamma_k}\right).$$

We now summarize our iterative algorithm in Algorithm 3.1 for solving (3.6).

Algorithm 3.1. Proximal alternating iteration scheme for solving (3.6)

Input: Input image g;
Output: Adaptive filter set D, and patch matrix F;
  1. Construct the patch matrix G as (3.4).
  2. Set initial filter matrix D_0, coefficient matrix C_0 and noise matrix F_0.
  3. For $k = 0, 1, \cdots$
     Set $C_{k+1} = T_{\sqrt{\lambda/(\tau+\alpha_k)}}\left(\frac{\tau D_k^T(G-F_k)+\alpha_k C_k}{\tau+\alpha_k}\right)$;
     Compute the SVD of $\tau(G - F_k)C_{k+1}^T + \beta_k D_k = U_k \Sigma_k V_k^T$ and set $D_{k+1} = U_k V_k^T$;
     Set $F_{k+1} = \mathcal{T}_{1/(2(\tau+\gamma_k))}\left(\frac{\tau(G-D_{k+1}C_{k+1})+\gamma_k F_k}{\tau+\gamma_k}\right)$,
     until some termination conditions are satisfied.

---

4. Convergence Analysis

In this section, we establish the convergence of Algorithm 3.1 by showing that the iterative sequence is a Cauchy sequence and converges to a stationary point of (3.6). The arguments are based on the Kurdyka-Łojasiewicz property (see the appendix) and the following abstract theoretical framework for convergence analysis developed recently in [1].

Theorem 4.1 ([1, Theorem 2.9]). *For a proper lower semi-continuous function H, the sequence $\{X_k\}$ converges to $\bar{X}$ as k goes to infinity, $\bar{X}$ is a stationary point of H, and sequence $\{X_k\}$ has a finite length, i.e.*

$$\sum_{k=0}^{+\infty} \|X_{k+1} - X_k\| < +\infty,$$

as long as it satisfies the following conditions with positive constants μ and ν:

H1 (Sufficient decrease condition). *For each $k \in \mathbb{N}$,*

$$H(X_{k+1}) + \mu\|X_{k+1} - X_k\|^2 \leq H(X_k).$$

H2 (Relative error condition). *For each $k \in \mathbb{N}$, there exists $\omega^{k+1} \in \partial H(X_{k+1})$ such that*

$$\|\omega^{k+1}\| \leq \nu\|X_{k+1} - X_k\|,$$

where $\partial H(X_{k+1})$ is the limiting Fréchet sub-differential of H at X_{k+1}; see the appendix.

H3 (Continuity condition). *There exists a subsequence $\{X_{k_n}\}$ and a cluster point $\hat{X}$ such that*

$$X_{k_n} \to \hat{X} \text{ and } H(X_{k_n}) \to H(\hat{X}), \quad \text{as } n \to \infty.$$

H4 (Additional condition). *H has the Kurdyka-Łojasiewicz property at $\hat{X}$ specified in* **H3.**

We now check that our Algorithm 3.1 satisfies the four conditions **H1 H2 H3 H4** in Theorem 4.1. For convenience, we denote the sequence $\{(C_k, D_k, F_k)\}$ generated by Algorithm 3.1 as $\{Z_k\}$.

Lemma 4.1. *The sequence $\{Z_k = (C_k, D_k, F_k)\}$ satisfies:*

1. $L(Z_k) - L(Z_{k+1}) \geq a\|Z_{k+1} - Z_k\|^2$;

2. $\displaystyle\sum_{k=1}^{\infty} \|Z_k - Z_{k+1}\|^2 < \infty$;

3. $\{Z_k\}$ is bounded.

Proof. Some arguments are similar to those in [3]. For Part 1, by (3.7), we have

$$L(C_{k+1}, D_k, F_k) + \alpha_k\|C_{k+1} - C_k\|^2 \leq L(C_k, D_k, F_k),$$
$$L(C_{k+1}, D_{k+1}, F_k) + \beta_k\|D_{k+1} - D_k\|^2 \leq L(C_{k+1}, D_k, F_k),$$
$$L(C_{k+1}, D_{k+1}, F_{k+1}) + \gamma_k\|F_{k+1} - F_k\|^2 \leq L(C_{k+1}, D_{k+1}, F_k),$$

indicating that

$$L(Z_k) - L(Z_{k+1}) \geq \alpha_k \|C_{k+1} - C_k\|^2 + \beta_k \|D_{k+1} - D_k\|^2 + \gamma_k \|F_{k+1} - F_k\|^2$$
$$\geq a \|Z_{k+1} - Z_k\|^2 \geq 0. \tag{4.1}$$

For Part 2, the inequality $L(Z_k) - L(Z_{k+1}) \geq a \|Z_{k+1} - Z_k\|^2$ gives the monotonicity of $L(Z_k)$ and $L(Z_0) - L(Z_{k+1}) \geq \sum_{j=0}^{k} a \|Z_j - Z_{j+1}\|^2$. Together with the fact that $L(Z_k) \geq 0$, we know that $\{L(Z_k)\}$ converges. This further indicates that $\sum_{k=1}^{\infty} \|Z_k - Z_{k+1}\|^2 < \infty$. It then also holds that $\lim_{k \to \infty} \|Z_k - Z_{k+1}\| = 0$.

For Part 3, from (4.1), we have $L(Z_k) \leq L(Z_{k-1}) \leq \cdots \leq L(Z_0)$, which implies $\|F_k\|_{e_1} \leq L(Z_0)$ and $\|C_k\| - (\|D_k^T G\| + \|D_k^T F_k\|) \leq \|C_k\| - \|D_k^T (G - F_k)\| \leq \|C_k - D_k^T (G - F_k)\|_F \leq \sqrt{\frac{1}{\tau} L(Z_0)}$. Together with the fact that $D_k \in \chi$, where χ is a compact set, we have that $\{Z_k\}$ is bounded and it has at least one cluster point. $\qquad\square$

To verify the condition **H2**, we use the following relationship on the limiting sub-differential of $L(C, D, F)$:

$$\partial L(C, D, F)$$
$$= \{\partial r(C) + \nabla_C Q(C, D, F)\} \times \{\partial s(D) + \nabla_D Q(C, D, F)\} \times \{\partial t(F) + \nabla_F Q(C, D, F)\}$$
$$= \partial_C L(C, D, F) \times \partial_D L(C, D, F) \times \partial_F L(C, D, F), \tag{4.2}$$

for all $(C, D, F) \in domL = domr \times doms \times domt$. The reason of (4.2) is [35, Proposition 10.5].

Lemma 4.2. *For each $k \geq 1$, we define*

$$C_k^* = -2\alpha_{k-1}(C_k - C_{k-1}) - \nabla_C Q(C_k, D_{k-1}, F_{k-1}) + \nabla_C Q(C_k, D_k, F_k),$$
$$D_k^* = -2\beta_{k-1}(D_k - D_{k-1}) - \nabla_D Q(C_k, D_k, F_{k-1}) + \nabla_D Q(C_k, D_k, F_k),$$
$$F_k^* = -2\gamma_{k-1}(F_k - F_{k-1}).$$

Then

1. *$(C_k^*, D_k^*, F_k^*) \in \partial L(C_k, D_k, F_k)$;*

2. *there exists a constant ν such that $\|(C_k^*, D_k^*, F_k^*)\| \leq \nu \|Z_k - Z_{k-1}\|$.*

Proof. For Part 1, by the definition of C_k, we know $0 \in 2\alpha_{k-1}(C_k - C_{k-1}) + \partial_C L(C_k, D_{k-1}, F_{k-1})$, $\forall k \geq 1$. Similarly, $0 \in 2\beta_{k-1}(D_k - D_{k-1}) + \partial_D L(C_k, D_k, F_{k-1})$, $\forall k \geq 1$, and $0 \in 2\gamma_{k-1}(F_k - F_{k-1}) + \partial_F L(C_k, D_k, F_k)$, $\forall k \geq 1$. It follows from the representation (3.6) that

$$\partial_C L(C_k, D_{k-1}, F_{k-1}) = \partial r(C_k) + \nabla_C Q(C_k, D_{k-1}, F_{k-1}),$$
$$\partial_D L(C_k, D_k, F_{k-1}) = \partial s(D_k) + \nabla_D Q(C_k, D_k, F_{k-1}),$$
$$\partial_F L(C_k, D_k, F_k) = \partial t(F_k) + \nabla_F Q(C_k, D_k, F_k).$$

Thus we have

$$-2\alpha_{k-1}(C_k - C_{k-1}) - \nabla_C Q(C_k, D_{k-1}, F_{k-1}) + \nabla_C Q(C_k, D_k, F_k) \in \partial r(C_k) + \nabla_C Q(C_k, D_k, F_k),$$
$$-2\beta_{k-1}(D_k - D_{k-1}) - \nabla_D Q(C_k, D_k, F_{k-1}) + \nabla_D Q(C_k, D_k, F_k) \in \partial s(D_k) + \nabla_D Q(C_k, D_k, F_k),$$
$$-2\gamma_{k-1}(F_k - F_{k-1}) \in \partial t(F_k) + \nabla_F Q(C_k, D_k, F_k),$$

meaning that $(C_k^*, D_k^*, F_k^*) \in \partial L(C_k, D_k, F_k)$ by (4.2).

For Part 2, on one hand, $Q(C, D, F) = \tau \|C - D^\mathrm{T}(G - F)\|$ is a smooth function, thus $\nabla Q(C, D, F)$ is Lipschitz continuous on any bounded set. From Lemma 4.1, we know that $\{Z_k\}$ is bounded. Therefore there exists a constant M such that the following two inequalities are true:

$$\| (\nabla_C Q(C_k, D_k, F_{k-1}) - \nabla_C Q(C_k, D_k, F_k), \nabla_D Q(C_k, D_k, F_{k-1}) - \nabla_D Q(C_k, D_k, F_k), 0) \|$$
$$\leq M \|F_{k-1} - F_k\| \leq M \|Z_{k-1} - Z_k\|, \tag{4.3}$$

and

$$\| (\nabla_C Q(C_k, D_{k-1}, F_{k-1}) - \nabla_C Q(C_k, D_k, F_{k-1}), 0, 0) \|$$
$$\leq M \|D_{k-1} - D_k\| \leq M \|Z_{k-1} - Z_k\|. \tag{4.4}$$

On the other hand, there is

$$\| (2\alpha_{k-1}(C_k - C_{k-1}), 2\beta_{k-1}(D_k - D_{k-1}), 2\gamma_{k-1}(F_k - F_{k-1})) \|$$
$$\leq 2b \|Z_k - Z_{k-1}\|. \tag{4.5}$$

Combining (4.3), (4.4) and (4.5), we get

$$\|(C_k^*, D_k^*, F_k^*)\| \leq \nu \|Z_k - Z_{k-1}\|,$$

where $\nu = 2b + 2M$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Lemma 4.3. *The sequence $\{Z_k = (C_k, D_k, F_k)\}$ has at least one convergent sub-sequence. Each cluster point is a stationary point of* (3.6).

Proof. By Lemma 4.1, the sequence $\{Z_k\}$ is bounded. Hence it has at least one convergent sub-sequence. Suppose that $\hat{Z} := (\hat{C}, \hat{D}, \hat{F})$ is a cluster point of $\{Z_k\}$, i.e., the limit point of a sub-sequence $\{Z_{k_n}\}$. $\lim_{k \to \infty} \|Z_k - Z_{k+1}\| = 0$ from Lemma 4.1 gives immediately both $Z_{k_n - 1} \to \hat{Z}$ and $Z_{k_n + 1} \to \hat{Z}$ when $n \to \infty$.

We now show $\lim_{n \to \infty} L(Z_{k_n}) = L(\hat{Z})$. By (3.7), we have

$$Q(C_{k_n+1}, D_{k_n}, F_{k_n}) + \alpha_{k_n} \|C_{k_n+1} - C_{k_n}\|^2 + r(C_{k_n+1})$$
$$\leq Q(\hat{C}, D_{k_n}, F_{k_n}) + \alpha_{k_n} \|\hat{C} - C_{k_n}\|^2 + r(\hat{C}). \tag{4.6}$$

Taking limsup in the both sides of (4.6) results in $\limsup_{n \to \infty} r(C_{k_n+1}) \leq r(\hat{C})$. Together with the lower semi-continuity of $r(C) = \lambda \|C\|_0$, which means $\liminf_{n \to \infty} r(C_{k_n+1}) \geq r(\hat{C})$, we obtain $\lim_{n \to \infty} r(C_{k_n+1}) = r(\hat{C})$. Similarly, $\lim_{n \to \infty} r(C_{k_n}) = r(\hat{C})$. Notice that $D_k \in \chi$ for all k and χ is a compact set, thus $s(D_{k_n}) = s(D_{k_n+1}) = s(\hat{D}) = 0$ for all k_n. As Q and t are continuous functions, it then follows that

$$\lim_{n \to \infty} L(Z_{k_n}) = \lim_{n \to \infty} L(Z_{k_n+1}) = L(\hat{C}, \hat{D}, \hat{F}). \tag{4.7}$$

Then we prove that $\hat{Z}$ is a stationary point of (3.6). From $C_{k+1} \in \arg\min_C \ L(C, D_k, F_k) + \alpha_k \|C - C_k\|_F^2$, we have

$$L(C_{k_n+1}, D_{k_n}, F_{k_n}) + a \|C_{k_n+1} - C_{k_n}\|^2 \leq L(C, D_{k_n}, F_{k_n}) + b \|C - C_{k_n}\|^2, \quad \forall C \in \Omega_C.$$

Letting $n \to \infty$, with (4.1) and (4.7), we obtain

$$L(\hat{C}, \hat{D}, \hat{F}) \leq L(C, \hat{D}, \hat{F}) + b\|C - \hat{C}\|^2, \quad \forall C \in \Omega_C,$$

which means

$$Q(\hat{C}, \hat{D}, \hat{F}) + r(\hat{C}) \leq Q(C, \hat{D}, \hat{F}) + r(C) + b\|C - \hat{C}\|^2, \quad \forall C \in \Omega_C.$$

Thus $\hat{C} \in \arg\min_C Q(C, \hat{D}, \hat{F}) + r(C) + b\|C - \hat{C}\|^2$. With the first-order optimality conditions, $0 \in \nabla_C Q(\hat{C}, \hat{D}, \hat{F}) + \partial r(\hat{C})$. By similar arguments, we obtain for D and F that $0 \in \nabla_D Q(\hat{C}, \hat{D}, \hat{F}) + \partial s(\hat{D})$ and $0 \in \nabla_F Q(\hat{C}, \hat{D}, \hat{F}) + \partial t(\hat{F})$. By (4.2), we have $0 \in \partial L(\hat{Z})$, which means that $\hat{Z}$ is a stationary point. $\qquad\square$

To verify the KŁ property of our objective function $L(C, D, F)$, we show that $L(C, D, F)$ is a semi-algebraic function defined below, since any such function has KŁ property [1,5].

Definition 4.1 (Semi-algebraic function)[1]. *A subset S of $\mathbb{R}^n$ is a real semi-algebraic set if there exists a finite number of real polynomial functions $P_{ij}, Q_{ij} : \mathbb{R}^n \to \mathbb{R}$ such that*

$$S = \bigcup_{j=1}^{p} \bigcap_{i=1}^{q} \{x \in \mathbb{R}^n : P_{ij}(x) = 0, Q_{ij}(x) < 0\}.$$

A function $h : \mathbb{R}^n \to \mathbb{R} \bigcup \{+\infty\}$ (resp. a point-to-set mapping $H : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$) is called semi-algebraic if its graph $\{(x, \rho) \in \mathbb{R}^{n+1} : h(x) = \rho\}$ (resp. $\{x, y\} \in \mathbb{R}^{n+m} : y \in H(x)\}$ is a semi-algebraic subset of $\mathbb{R}^{n+1}$ (resp. $\mathbb{R}^{n+m}$).

Lemma 4.4. *$L(C, D, F)$ in (3.6) is a semi-algebraic function.*

Proof. Recall $L(C, D, F) = r(C) + s(D) + t(F) + Q(C, D, F)$. Referring to [3, Theorem 4.5], $r(C)$, $s(D)$ and $Q(C, D, F)$ are all semi-algebraic functions. We need only show that $t(F)$ is semi-algebraic, since finite sums of semi-algebraic functions are semi-algebraic [34].

For $t(F) = \|F\|_{e_1} = \sum_{j=1}^{N} \sum_{i=1}^{m^2} |F_{ij}|$, the graph of $|F_{ij}|$ is $S_{ij} = \{(F_{ij}, k) \in \mathbb{R}^2 : |F_{ij}| = k\} = \{F_{ij} - k = 0, \ -F_{ij} < 0\} \cup \{F_{ij} + k = 0, \ F_{ij} < 0\} \cup \{F_{ij} = 0\}$, which is a semi-algebraic set. Thus $|F_{ij}|$ is a semi-algebraic function. Hence $t(F)$ is semi-algebraic. $\qquad\square$

We are now ready to present our final convergence result.

Theorem 4.2. *The sequence $\{Z_k = (C_k, D_k, F_k)\}$ generated by (3.7) converges to a stationary point of (3.6).*

Proof. It is clear that $L(C, D, F)$ is a proper lower semi-continuous function. Combining Lemma 4.1, Lemma 4.2, Lemma 4.3 and Lemma 4.4, we can apply Theorem 4.1 to obtain that the sequence $\{Z_k = (C_k, D_k, F_k)\}$ converges to a stationary point of (3.6). $\qquad\square$

5. Experiments

Let $g = u + n(\sigma\%)$ denote some noisy observation of u, where $n(\sigma\%)$ is the impulsive noise with proportion of pollution $\sigma\%$. For an input noisy observation, Algorithm 3.1 generates an adaptive tight frame W, and a matrix F which somehow denotes the noise in patches. We then

de-noise the input image g by using the tight frame W and patch matrix F. We calculate, by hard thresholding, the following $\tilde{u}$

$$\tilde{u} = W^{\mathrm{T}}(T_{\tilde{\lambda}}(W(g - \tilde{f}))), \tag{5.1}$$

as our denoising result. Here $\tilde{f}$ represents the noise in image domain and $\tilde{\lambda}$ is a thresholding parameter. The $\tilde{f}$ is computed from F in Algorithm 3.1 by a reformulation from the patch domain to the image domain. According to the construction of the patch matrix G from the image g, we take an inverse procedure to get $\tilde{f}$ from F. For those pixels appearing for multiple times in patch domain, we average their values in the reformulation from F to $\tilde{f}$.

The test platform is a Window 10 Enterprise system equipped with Intel Core i7-4790 CPU at 3.60 GHz and 8.00GB memory. To study the performance of Algorithm 3.1 for images with different structures, we test the proposed algorithm using some images which contain both cartoon-type regions and texture regions. They are shown in Fig. 5.1.

(a) Chart 256×256 (b) Aerial 256×256 (c) Montage 256×256

(d) Baboon 512×512 (e) Barbara 512×512 (f) Airplane 512×512

Fig. 5.1. Six tested image selected from USC-SIPI image database.

We choose parameters according to the following basic rules, which were summarized from lots of experiments. The patch size is fixed as 16×16. The termination criterion is set either the maximum number of iterations $iter = 40$ or $\|D_k - D_{k+1}\| \leq 10^{-2}$. The parameters λ and $\tilde{\lambda}$ are suggested to be 11.2σ and 3.9σ respectively in general. The parameter τ is set to 0.02 for $\sigma = 10, 20$ and 0.01 for others. The $\lambda, \tilde{\lambda}, \tau$ may be tuned with some minor adjustments for some individual images. The proximal parameters α_k, β_k and γ_k during the iterations are all fixed to 0.01.

In order to show the influence of the initialization of tight frame, local discrete cosine transform (DCT) and un-decimal Haar wavelets [16] are selected as the initialization, respectively,

tested on the image "Montage" corrupted by random-valued noise and the image "Barbara" corrupted by salt-and-pepper noise. Results are listed in Table 5.1, which shows that the performances of two different initializations are very close in terms of PSNR value. In the following experiments, we simply choose local DCT as the initialization.

Table 5.1: Comparison of the PSNR values (dB) of the de-noising results by Algorithm 3.1 with different initializations on the tight frame.

Image	Noise Level	Local DCT	Haar Wavelet
Montage	10%	31.32	31.15
	20%	27.22	27.31
	30%	25.87	25.69
	40%	23.45	23.38
	50%	21.52	21.43
	60%	20.18	19.97
Barbara	10%	33.37	33.26
	20%	30.02	29.91
	30%	26.54	26.62
	40%	23.93	23.75
	50%	21.86	21.83
	60%	20.33	20.21

We now verify the convergence of our algorithm and study the influence of iteration numbers on de-noising results, by applying our Algorithm 3.1 to the image "Barbara" corrupted by 20% salt-and-pepper noise. We calculated the de-noising results by using the tight frame filters generated in Algorithm 3.1 at different iterations. Fig. 5.2(a) shows the PSNR values of these results along the iteration. As can be seen, the curve increases almost monotonically, indicating the progressive improvements of the de-noising effectiveness of our iterative algorithm. The convergence behavior of Algorithm 3.1 is shown in Fig. 5.2(b) by the Frobenius norm of $D_k - D_{k-1}$ along iteration. In Fig. 5.3, we illustrate the computed filters at the 1st, 3th, 15th and 30th iterations.

(a) PSNR value (dB) of de-noising result vs the iteration number.

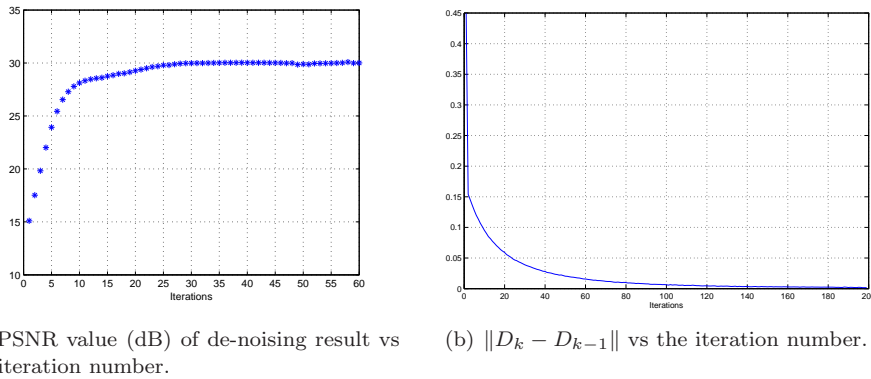(b) $\|D_k - D_{k-1}\|$ vs the iteration number.

Fig. 5.2. PSNR and convergence behavior along the iteration of Algorithm 3.1 on the image "Barbara".

Next, we compare the performance of Algorithm 3.1 with other methods including DWM filter [19], AK-SPR Algorithm [6] and ALM TV-L1 Algorithm [40], all of which ran on the same platform as mentioned before. For these three compared methods, we choose their suggested

(a) $k = 1$ (b) $k = 3$ (c) $k = 15$ (d) $k = 30$
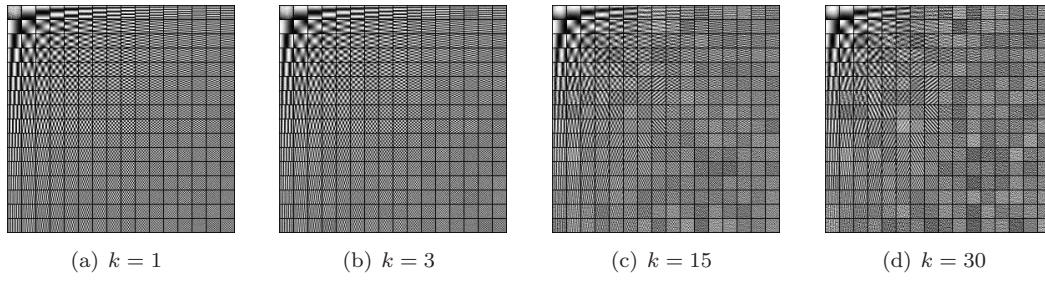
Fig. 5.3. Tight frame filters by Algorithm 3.1 after k iterations when restoring the image "Barbara".

(a) original image   (b) noisy image; 10.50 dB   (c) result of DWM; 24.25 dB

(d) result of AK-SPR; 23.65 dB   (e) result of TV-L1; 23.47 dB   (f) result of Alg. 1; 26.54 dB

Fig. 5.4. De-noising results for the image "Barbara" corrupted with salt-and-pepper noise.

parameter settings and also make minor adjustments to achieve good results. Fig. 5.4 and Fig. 5.5 show the noise removal results for image "Barbara" corrupted with 30% salt-and-pepper noise and image "Montage" with 20% random-valued noise by different methods. More quantitative comparisons are summarized in Table 5.3 and Table 5.4, which list the PSNR values (dB) of the restoration results by four algorithms for salt-and-pepper noise removal and random-valued noise removal problems, respectively. From these experiments, all of these methods work quite well in most cases. However, DWM erases slim structures like texts; see Fig. 5.5. AK-SPR keeps noise in textured regions. TV-L1 works well for cartoon-like images, but smoothes textures. Overall speaking, our Algorithm 3.1 seems to generate good results for different images with different structures. This is reasonable, because it learns tight frames for sparse regularization from input images adaptively and hence captures richer image structures. This advantage is particularly obvious in the "Montage" example; see Fig. 5.5. When the noise level is low, the proposed algorithm works especially well. We should admit that, this data adaptive advantage of our method weakens for high level noise removal problems, because in these problems the structure information of the input image is corrupted too much and the tight frame construction becomes inaccurate.

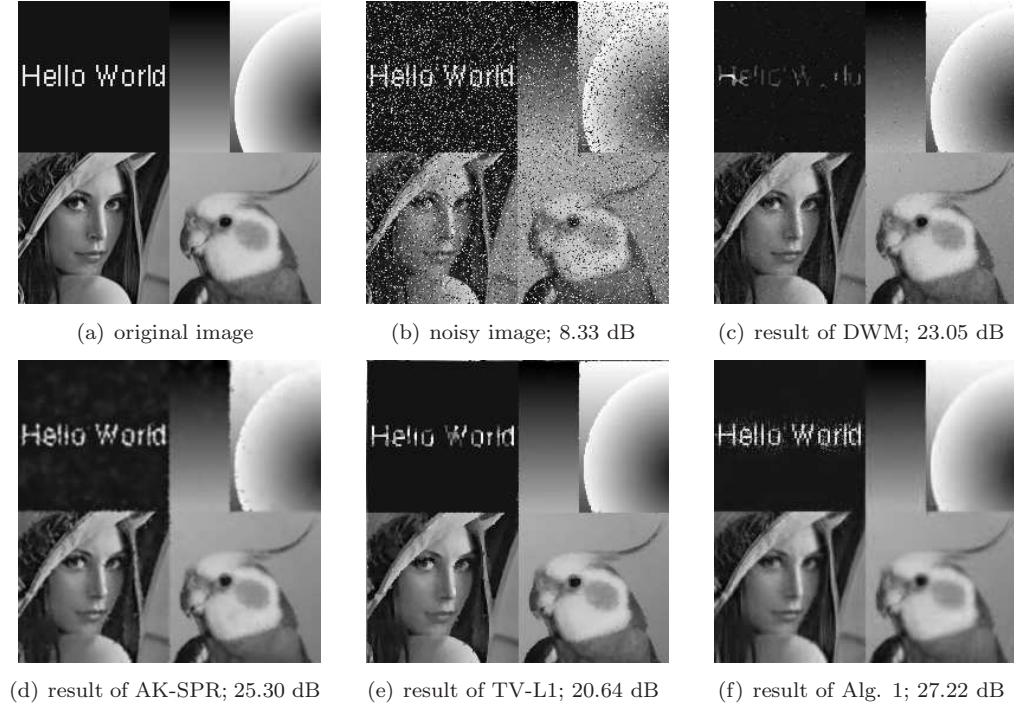| (a) original image | (b) noisy image; 8.33 dB | (c) result of DWM; 23.05 dB |
| (d) result of AK-SPR; 25.30 dB | (e) result of TV-L1; 20.64 dB | (f) result of Alg. 1; 27.22 dB |

Fig. 5.5. De-noising results for the image "Montage" corrupted with random-valued noise.

Table 5.2 lists the averaged CPU costs (in second) of different methods for salt-pepper noise removal problems with six noise levels. It can be seen that the proposed algorithm is considerably faster than DWM and AK-SPR but slower than TV-L1.

Table 5.2: Averaged CPU costs (in second) of different methods for salt-pepper noise removal problems with six noise levels.

Image / Method	Chart	Aerial	Montage	Baboon	Barbara	Airplane
DWM	29.52	29.67	28.35	116.36	117.26	114.88
AK-SPR	390.86	102.32	275.45	425.41	540.21	413.24
TV-L1	1.87	2.25	2.07	8.12	8.23	7.76
Alg. 1	14.15	13.85	14.28	54.05	53.35	53.72

Table 5.3: PSNR values (dB) of the de-noising results by DWM, AK-SPR, TV-L1 and Algorithm 3.1 for salt-and-pepper noise removal problems.

Image	Noise Level	DWM	AK-SPR	TV-L1	Alg. 1
Chart	10%	23.25	21.33	20.12	**24.29**
	20%	21.58	20.35	18.61	**22.90**
	30%	18.54	19.03	17.28	**20.39**
	40%	16.39	17.06	15.35	**18.57**
	50%	15.02	14.93	14.14	**16.23**
	60%	13.53	13.43	13.47	**15.11**
Aerial	10%	28.91	25.72	23.45	**29.09**
	20%	25.17	24.67	22.14	**25.35**
	30%	**24.25**	23.84	21.63	22.97
	40%	**22.47**	22.16	20.67	21.32
	50%	20.50	**20.65**	19.49	18.98
	60%	17.20	17.55	**18.84**	16.12
Montage	10%	23.58	26.57	23.03	**30.73**
	20%	22.85	26.17	21.54	**26.30**
	30%	21.12	23.83	20.76	**24.05**
	40%	20.91	22.24	19.98	**22.87**
	50%	19.89	20.07	19.44	**21.05**
	60%	18.79	18.41	18.05	**19.24**
Baboon	10%	26.06	24.46	24.12	**26.15**
	20%	24.03	23.55	22.61	**24.12**
	30%	**22.87**	22.01	21.57	22.13
	40%	**21.29**	21.10	20.33	20.65
	50%	19.99	**20.28**	20.02	19.87
	60%	19.13	19.01	**19.53**	18.48
Barbara	10%	28.88	27.45	26.95	**33.37**
	20%	25.96	26.09	24.93	**30.02**
	30%	24.25	23.65	23.47	**26.54**
	40%	23.38	23.31	23.36	**23.93**
	50%	22.25	21.73	**22.60**	21.86
	60%	20.80	20.02	**22.13**	20.33
Airplane	10%	31.91	33.12	32.52	**33.43**
	20%	29.94	**31.02**	29.84	30.19
	30%	**28.54**	27.50	27.53	28.01
	40%	**26.45**	25.02	25.80	25.97
	50%	24.22	23.34	**24.51**	23.45
	60%	21.85	21.03	**22.65**	21.33

Table 5.4: PSNR values (dB) of the de-noising results by DWM, AK-SPR, TV-L1 and Algorithm 3.1 for random-valued noise removal problems.

Image	Noise Level	DWM	AK-SPR	TV-L1	Alg. 1
Chart	10%	21.38	21.41	19.63	**25.96**
	20%	19.54	20.15	17.22	**23.14**
	30%	18.55	17.29	15.94	**20.86**
	40%	17.80	15.97	14.65	**19.07**
	50%	16.75	13.81	13.73	**17.63**
	60%	14.75	11.93	13.21	**15.89**
Aerial	10%	28.04	25.23	23.22	**29.57**
	20%	25.55	24.33	22.01	**25.72**
	30%	**24.11**	23.11	21.63	23.58
	40%	**22.72**	21.62	20.54	21.57
	50%	**21.43**	20.10	19.12	19.79
	60%	**20.03**	18.56	18.78	18.82
Montage	10%	23.97	26.52	23.12	**31.32**
	20%	23.05	25.30	20.64	**27.22**
	30%	22.77	23.24	20.15	**25.87**
	40%	22.48	20.86	19.34	**23.45**
	50%	**21.67**	18.14	19.25	21.52
	60%	**20.23**	17.98	18.30	20.18
Baboon	10%	27.02	24.62	24.41	**27.38**
	20%	23.92	23.81	22.49	**24.80**
	30%	22.70	22.53	21.31	**22.97**
	40%	21.51	**21.53**	20.42	21.45
	50%	**20.77**	20.45	19.87	20.54
	60%	**20.02**	19.72	19.18	19.74
Barbara	10%	29.26	27.02	26.77	**34.47**
	20%	26.71	26.64	25.09	**30.02**
	30%	24.50	23.79	23.47	**26.60**
	40%	23.61	22.50	22.38	**25.55**
	50%	**22.25**	21.50	21.32	22.04
	60%	**21.76**	19.54	18.98	21.67
Airplane	10%	32.75	33.42	32.54	**34.36**
	20%	29.83	**31.10**	29.74	29.90
	30%	**28.29**	27.19	27.31	27.74
	40%	**26.42**	24.52	24.84	25.38
	50%	**25.42**	21.82	21.25	23.41
	60%	**23.64**	19.01	17.22	22.33

6. Conclusion

Most of the existing data-driven tight frame denosing methods use L_2 fidelity, which does not fit the statistics of the very basic and widespread impulsive noise. In this paper we extended data-driven tight frame methods to impulsive noise removal problems by incorporating L_1 fidelity into the previous L_0 balanced type model. The proposed model inherits the convenience of optimization and implementation of the original data-driven approach. It is efficiently solved through an alternating iteration after partitioning the noisy image and variables into patches. We also showed that the sequence generated by our algorithm globally converges to a stationary

point of the energy function. Once an adaptive tight frame system is constructed from the input noisy image, it is used to remove impulsive noise. With the comparisons in the numerical experiments, advantages of our approach, especially for low level noise problems, were shown clearly for various images with diverse structures. This is due to the data-driven nature of our approach, which generates tight frames adaptive to varied structures of input images. In the future, we plan to design more powerful data-driven models for deblurring problems.

A. Sub-differential and Kurdyka-Łojasiewicz Property

Definition A.1 (Limiting Fréchet sub-differential)[35]. *A vector z is a Fréchet subgradient of a lower semi-continuous function $h : \mathbb{R}^n \to \mathbb{R} \bigcup \{+\infty\}$ at $x \in dom(h)$ if*

$$\liminf_{y \to x, y \neq x} \frac{h(y) - h(x) - \langle z, y - x \rangle}{\|y - x\|} \geq 0.$$

The set of Fréchet sub-gradient of h at x is called Fréchet sub-differential and denoted as $\hat{\partial}h(x)$. If $x \notin dom(h)$, then $\hat{\partial}h(x) = \emptyset$. The limiting Fréchet sub-differential is denoted by $\partial h(x)$ and defined as

$$\partial h(x) = \{z : \ there\ is\ x_m \to x,\ h(x_m) \to h(x)\ and\ z_m \in \hat{\partial}h(x_m)\ such\ that\ z_m \to z\}.$$

An $x \in dom(h)$ is called a stationary point of h if $0 \in \partial h(x)$.

Definition A.2 (Kurdyka-Łojasiewicz property)[34]. *The function $h : \mathbb{R}^n \to \mathbb{R} \bigcup \{+\infty\}$ is said to have the Kurdyka-Łojasiewicz property at $x^* \in dom(\partial h)$ if there exist $\eta \in (0, \infty]$, a neighborhood U of x^* and a continuous concave function $\varphi : [0, \eta) \to \mathbb{R}_+$ such that:*

1. *$\varphi(0) = 0$,*

2. *φ is C^1 on $(0, \eta)$,*

3. *for all $\zeta \in (0, \eta), \varphi'(\zeta) > 0$,*

4. *for all x in $U \bigcap \{x : h(x^*) < h(x) < h(x^*) + \eta\}$, the Kurdyka-Łojasiewicz inequality holds in the sense that*

$$\varphi'(h(x) - h(x^*))dist(0, \partial h(x)) \geq 1.$$

A proper lower semi-continuous function which satisfies the Kurdyka-Łojasiewicz inequality at each point of $dom(\partial h)$ is called a KŁ function.

References

[1] H. Attouch, J. Bolte, and B.F. Svaiter, Convergence of descent methods for semi-algebraic and tame problems: Proximal algorithms, forwardCbackward splitting, and regularized gauss-seidel methods, *Math. Program.*, **137**:1–2 (2013), 91–129.

[2] M. Bai, X. Zhang, and Q. Shao, Adaptive correction procedure for TVL1 image deblurring under impulse noise, *Inverse Probl.*, **32**:8 (2016), 085004.

[3] C. Bao, H. Ji, and Z. Shen, Convergence analysis for iterative data-driven tight frame construction scheme, *Appl. Comput. Harmon. Anal.*, **38**:3 (2015), 510–523.

[4] J. Bolte, S. Sabach, and M. Teboulle, Proximal alternating linearized minimization for nonconvex and nonsmooth problems, *Math. Program.*, **146**:1–2 (2014), 459–494.

[5] J. Bolte, A. Daniilidis, A. Lewis, and M. Shiota, Clarke subgradients of stratifiable functions, *SIAM J. Optim.*, **18**:2 (2007), 556–572.

[6] P. Bouboulis, K. Slavakis, and S. Theodoridis, Adaptive kernel-based image denoising employing semi-parametric regularization, *IEEE Trans. Image Process.*, **19**:6 (2010), 1465–1479.

[7] A.C. Bovik, Handbook of Image and Video Processing, Academic press, 2010.

[8] J.F. Cai, B. Dong, and Z. Shen, Image restoration: A wavelet frame based model for piecewise smooth functions and beyond, *Appl. Comput. Harmon. Anal.*, **41**:1 (2015), 94–138.

[9] J.F. Cai, H. Ji, Z. Shen, and G.B. Ye, Data-driven tight frame construction and image denoising, *Appl. Comput. Harmon. Anal.*, **37**:1 (2014), 89–105.

[10] J.F. Cai and Z. Shen, Image restoration: Total variation, wavelet frames, and beyond, *J. Am. Math. Soc.*, **25**:4 (2012), 1033–1089.

[11] E.J. Cands, Ridgelets: Estimating with ridge functions, *Ann. Stat.*, **31**:5 (2003), 1561–1599.

[12] E.J. Cands and D.L. Donoho, Recovering edges in ill-posed inverse problems: Optimality of curvelet frames, *Ann. Stat.*, **30**:3 (2002), 784–842.

[13] D.Q. Chen, Data-driven tight frame learning scheme based on local and non-local sparsity with application to image recovery, *J. Sci. Comput.*, **69**:2 (2016), 461–486.

[14] T. Chen and H.R. Wu, Adaptive impulse detection using center-weighted median filters, *IEEE Signal Process. Lett.*, **8**:1 (2001), 1–3.

[15] J.K. Choi, B. Dong, and X. Zhang, Limited tomography reconstruction via tight frame and simultaneous sinogram extrapolation, *J. Comput. Math.*, **34**:6 (2016), 575–589.

[16] I. Daubechies, Ten Lectures on Wavelets, volume 61, SIAM, 1992.

[17] I. Daubechies, B. Han, A. Ron, and Z. Shen, Framelets: MRA-based constructions of wavelet frames, *Appl. Comput. Harmon. Anal.*, **14**:1 (2003), 1–46.

[18] B. Dong, H. Ji, J. Li, Z. Shen, and Y. Xu, Wavelet frame based blind image inpainting, *Appl. Comput. Harmon. Anal.*, **32**:2 (2012), 268–279.

[19] Y. Dong and S. Xu, A new directional weighted median filter for removal of random-valued impulse noise, *IEEE Signal Process. Lett.*, **14**:3 (2007), 193–196.

[20] D.L. Donoho, De-noising by soft-thresholding, *IEEE Trans. Inf. Theory*, **41**:3 (1995), 613–627.

[21] M. Elad and M. Aharon, Image denoising via sparse and redundant representations over learned dictionaries, *IEEE Trans. Image Process.*, **15**:12 (2006), 3736–3745.

[22] X. Fan, Q. Lian, and B. Shi, Compressed-sensing MRI based on adaptive tight frame in gradient domain, *Appl. Magn. Reson.*, **49**:5 (2018), 465–477.

[23] G. Gu, S. Jiang, and J. Yang, A tvscad approach for image deblurring with impulsive noise, *Inverse Probl.*, **33**:12 (2017), 125008.

[24] H. Hwang and R.A. Haddad, Adaptive median filters: New algorithms and results, *IEEE Trans. Image Process.*, **4**:4 (1995), 499–502.

[25] S.J. Ko and H.L. Yong, Center weighted median filters and their applications to image enhancement, *IEEE Trans. Circuits Syst.*, **38**:9 (1991), 984–993.

[26] J. Liang, J. Li, Z. Shen, and X. Zhang, Wavelet frame based color image demosaicing, *Inverse Probl. Imaging*, **7**:3 (2013), 777–794.

[27] J. Liang, J. Ma, and X. Zhang, Seismic data restoration via data-driven tight frame, *Geophysics*, **79**:3 (2014), 65–74.

[28] D. Ma and W. Cong, Removal of mixed Gaussian and impulse noise using data-driven tight frames, *J. Eng. Sci. Tech. Rev.*, **11**:2 (2018), 26–31.

[29] L. Ma, J. Yu, M.K. Ng and T. Zeng, Efficient box-constrained TV-TYPE-l^1 algorithms for restoring images with impulsive noise, *J. Comput. Math.*, **31**:3 (2013), 249–270.

[30] M. Nikolova, Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers. *SIAM J. Numer. Anal.*, **40**:3 (2002), 965–994.

[31] M. Nikolova, A variational approach to remove outliers and impulse noise, *J. Math. Imaging Vis.*, **20**:1–2 (2004), 99–120.

[32] T. Pang, Q. Li, Z. Wen, and Z. Shen, Phase retrieval: A data-driven wavelet frame based approach, *Appl. Comput. Harmon. Anal.*, **49**:3 (2020), 971-1000.

[33] Y. Quan, H. Ji, and Z. Shen, Data-driven multi-scale non-local wavelet frame construction and image recovery, *J. Sci. Comput.*, **63**:2 (2015), 307–329.

[34] P. Redont and A. Soubeyran, Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality, *Math. Oper. Res.*, **35**:2 (2008), 438–457.

[35] R.T. Rockafellar and R.J.B. Wets, Variational Analysis, Springer, 2009.

[36] A. Ron and Z. Shen, Affine systems in $l_2(r^d)$: The analysis of the analysis operator, *J. Funct. Anal.*, **148**:2 (1995), 408–447.

[37] Y. Shen, B. Han, and E. Braverman. Adaptive frame-based color image denoising. *Appl. Comput. Harmon. Anal.*, **41**:1 (2016), 54–74.

[38] Z. Shen, Wavelet frames and image restorations, in *The International Congress of Mathematicians*, India, (2010), 2834–2863.

[39] T. Sun and Y. Neuvo, Detail-preserving median based filters in image processing, *Pattern Recognit. Lett.*, **15**:4 (1994), 341–347.

[40] C. Wu, J. Zhang, and X.C. Tai, Augmented lagrangian method for total variation restoration with non-quadratic fidelity, *Inverse Probl. Imaging*, **5**:1 (2011), 237–261.

[41] J. Yang, Y. Zhang, and W. Yin. An efficient TVL1 algorithm for deblurring multichannel images corrupted by impulsive noise, *SIAM J. Sci. Comput.*, **31**:4 (2009), 2842–2865.

[42] R. Zhan and B. Dong, CT image reconstruction by spatial-radon domain data-driven tight frame regularization, *SIAM J. Imaging Sci.*, **9**:3 (2016), 1063–1083.

[43] X. Zhang, M. Bai and M. K. Ng, Nonconvex-TV based image restoration with impulse noise removal, *SIAM J. Imaging Sci.*, **10**:3 (2017), 1627–1667.

[44] Y. Zhang and X. Zhang, Variational bimodal image fusion with data-driven tight frame, *Inf. Fusion*, **55** (2020), 164–172.

[45] H. Zou, T. Hastie, and R. Tibshirani, Sparse principal component analysis, *J. Comput. Graph. Stat.*, **15**:2 (2006), 265–286.