CrossMark

Cauchy Noise Removal by Nonconvex ADMM with Convergence Guarantees

Jin-Jin Mei[1,2] · Yiqiu Dong[2] · Ting-Zhu Huang[1] · Wotao Yin[3]

Abstract Image restoration is one of the essential tasks in image processing. In order to restore images from blurs and noise while also preserving their edges, one often applies total variation (TV) minimization. Cauchy noise, which frequently appears in engineering applications, is a kind of impulsive and non-Gaussian noise. Removing Cauchy noise can be achieved by solving a *nonconvex* TV minimization problem, which is difficult due to its nonconvexity and nonsmoothness. In this paper, we adapt recent results in the literature and develop a specific alternating direction method of multiplier to solve this problem. Theoretically, we establish the convergence of our method to a stationary point. Experimental results demonstrate that the proposed method is competitive with other methods in visual and quantitative measures. In particular, our method achieves higher PSNRs for 0.5 dB on average.

✉ Yiqiu Dong
yido@dtu.dk

Jin-Jin Mei
meijinjin666@126.com

Ting-Zhu Huang
tingzhuhuang@126.com

Wotao Yin
wotaoyin@math.ucla.edu

1 School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu 611731, People's Republic of China

2 Department of Applied Mathematics and Computer Science, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

3 Department of Mathematics, University of California, Los Angeles, CA 90025, USA

1 Introduction

In many imaging applications, images inevitably contain natural non-Gaussian noises, such
as impulse noise, Poisson noise, multiplicative noise, and Cauchy noise. At the same time, the
images may have been blurred by the point spread function (PSF) during their acquisition.
Therefore, the image restoration problem is an essential task. Researchers have proposed
many methods to deblur and denoise images; see [12,16,17,27,35,36,41,54] and references
therein. In this paper, we focus on recovering images corrupted by blurring and Cauchy noise.
Cauchy noise usually arises in echo of radar, in the presence of low-frequency atmospheric
noise, and in underwater acoustic signals [26,31,40]. According to [44,45], it follows Cauchy
distribution and is impulsive.

We assume that the original gray-scale image u is defined on a connected bounded domain
$\Omega \subset \mathbb{R}^2$ with a compacted Lipschitz boundary. The observed image with blurs and Cauchy
noise is given as follows:

$$f = Ku + \eta, \tag{1}$$

where $f \in L^2(\Omega)$ denotes the observed image, $K \in \mathcal{L}(L^1(\Omega), L^2(\Omega))$ represents a known
linear and continuous blurring (or convolution) operator, and $\eta \in L^2(\Omega)$ denotes Cauchy
noise. Our goal is to recover u from the observed image f.

In recent years, much attention has been given to Cauchy noise removal, and several
methods have been proposed. In [13], the authors applied a recursive algorithm based on
the Markov random field to reconstruct images and retain sharp edges. In 2005, Achim and
Kuruoğlu utilized a bivariate maximum a posteriori estimator (BMAP) to propose a new sta-
tistical model in the complex wavelet domain for removing Cauchy noise [1]. In [34], Loza et
al. proposed a statistical approach based on non-Gaussian distributions in the wavelet domain
for tackling the image fusion problem. Their method achieved a significant improvement in
fusion quality and noise reduction. In [46], Wan et al. developed a novel segmentation method
for RGB images that are corrupted by Cauchy noise. They combined statistical methods with
denoising techniques and obtained a satisfactory performance. Since TV regularization is
able to preserve edges effectively while still suppressing noise satisfactorily [21], Sciacchi-
tano et al. proposed a convex TV-based variational method for recovering images corrupted
by Cauchy noise in [42]. The variational model in this method is as follows:

$$\min_{u \in BV(\Omega)} \int_\Omega |Du| + \frac{\lambda}{2} \left(\int_\Omega \log \left(\gamma^2 + (u - f)^2 \right) dx + \alpha \|u - \tilde{u}\|_2^2 \right), \tag{2}$$

where $\gamma > 0$ is the scale parameter of Cauchy distribution, and $BV(\Omega)$ is the space of
functions of bounded variation. Here, $u \in BV(\Omega)$ if $u \in L^1(\Omega)$ and its total variation (TV)

$$\int_\Omega |Du| \triangleq \sup \left\{ \int_\Omega u \, \text{div}\mathbf{v} \, dx : \mathbf{v} \in \left(C_0^\infty(\Omega) \right)^2, \|\mathbf{v}\|_\infty \leq 1 \right\}$$

is finite, where $(C_0^\infty(\Omega))^2$ is the space of vector-valued functions with compact support in
Ω. The space $BV(\Omega)$ endowed with the norm $\|u\|_{BV(\Omega)} = \|u\|_{L^1(\Omega)} + \int_\Omega |Du|$ is a Banach
space; see, e.g., [21]. In (2), λ denotes the positive regularization parameter, which controls
the trade-off between TV regularization and the fitting to f and $\tilde{u}$, $\tilde{u}$ is the result obtained by
the median filter, and α is a positive penalty parameter. Note that if $8\alpha\gamma^2 \geq 1$, the objective

functional in (2) is strictly convex and leads to a unique solution. Because of strict convexity, the model avoids the common issues of nonconvex optimization: the solutions depend on the numerical methods and how they are initialized. But the last term in (2) in fact pushes the solution close to the median filter result, and the median filter does not always provide satisfactory removals of Cauchy noise. Hence, in this paper we turn our focus back to a nonconvex model.

Recently, researchers have discovered some useful convergence properties of the optimization algorithms for solving nonconvex minimization problems [24,47,48,53]. In particular, the paper [48] established the global convergence (to a stationary point) of the alternating direction method of multipliers (ADMM) for nonconvex nonsmooth optimization with linear constraints. To take advantages of the recent results, in this paper we develop the ADMM algorithm to solve the following nonconvex variational model directly for denoising and deblurring simultaneously:

$$\min_{u \in BV(\Omega)} \int_\Omega |Du| + \frac{\lambda}{2} \int_\Omega \log \left(\gamma^2 + (Ku - f)^2 \right) dx. \tag{3}$$

We prove that our algorithm starting from any initialization is globally convergent to a stationary point under certain conditions. Furthermore, we compare our proposed method to the state-of-the-art method proposed in [42] and show the effectiveness of our method in terms of restoration quality and noise reduction.

The outline of the paper is summarized as follows. In the next section, we analyse some fundamental properties of Gaussian distribution, Laplace distribution and Cauchy distribution. In Sect. 3, we illustrate the nonconvex variational model for denoising and deblurring, and prove the existence and uniqueness of the solution. In Sect. 4, we develop our algorithm for the proposed nonconvex model and present the convergence results. In Sect. 5, we demonstrate the performance of our algorithm by comparing with other existing algorithms. Finally, we conclude the paper with some remarks in Sect. 6.

2 Statistical Properties for Cauchy Distribution

The Cauchy distribution is a special kind of the α-stable distributions with $\alpha = 1$ and is important as a canonical example of the "pathological" case [3,15,29]. It is closed under linear fractional transformations with real coefficients [30]. However, different from most α-stable distributions, it possesses a probability density function that can be expressed analytically [19,28] as:

$$P_C(x) = \frac{\gamma}{\pi \left((x - \mu)^2 + \gamma^2 \right)},$$

where the parameter μ specifies the location of the peak and the parameter $\gamma > 0$ decides the half-width at half-maximum. Here, we let $\mathcal{C}(\mu, \gamma)$ denote the Cauchy distribution. Its mode and median are both μ while the mean, variance, and higher moments are undefined. In addition, the Cauchy distribution is infinitely divisible, that is, for every positive integer n, there exist n independent identically distributed (i.i.d.) random variables $X_{n1}, X_{n2}, \ldots X_{nn}$ such that $X_{n1} + X_{n2} + \cdots + X_{nn}$ follows the Cauchy distribution. Due to their infinite divisibility, random variables following the Cauchy distribution obey the generalized central limit theorem [37].

The Cauchy distribution is closely related to some other probability distributions. The Cauchy distribution is heavy-tailed, and its tail's heaviness is determined by the scale parameter γ. In particularly, if X and Y are two independent Gaussian random variables with mean

0 and variance 1, then the ratio X/Y follows the standard Cauchy distribution $\mathcal{C}(0, 1)$ [6,38]. In Sect. 5, we will apply this property to simulate images corrupted by Cauchy noise.

Further to show the statistical properties of the Cauchy distribution, we compare it with two most commonly used probability distributions: the Gaussian distribution ($\mathcal{N}(\mu, \sigma^2)$ with mean μ and variance σ^2) and the Laplace distribution $\mathcal{L}(\mu, b)$ with mean μ and variance $2b^2$. Since the Gaussian and Cauchy distributions are α-stable distributions with $\alpha = 2$ and $\alpha = 1$, respectively, they are both bell-shaped. Moreover, we can easily obtain the following relation between them at $x = 0$.

Proposition 2.1 *Let X_1 and X_2 be two independent random variables. Assume that $X_1 \sim \mathcal{N}(0, 1)$ and $X_2 \sim \mathcal{C}(0, \sqrt{\frac{2}{\pi}})$. Then the values of their probability density functions (PDFs) at $x = 0$ are equal.*

In addition, both the Laplace and Cauchy distributions are heavy-tailed distributions. We demonstrate their relation by the tails of their distribution curves in the following proposition.

Proposition 2.2 *Let P_G, P_L and P_C denote the PDFs for $\mathcal{N}(0, \sigma^2)$, $\mathcal{L}(0, b)$ and $\mathcal{C}(0, \sqrt{\frac{2}{\pi}})$, respectively. Then, the followings hold:*

1. *At $x = \sigma = b = \gamma$, the ratio of P_G, P_L and P_C is $1 : \sqrt{\frac{\pi}{2e}} : \sqrt{\frac{e}{2\pi}}$;*
2. *At $x = 3\sigma = 3b = 3\gamma$, the ratio of P_G, P_L and P_C is $1 : \sqrt{\frac{\pi}{2}}e^{\frac{3}{2}} : \sqrt{\frac{1}{50\pi}}e^{\frac{9}{2}}$.*

Based on Proposition 2.2, we can see that the probability density value of the Gaussian distribution at a rather small x, saying $x = \sigma = b = \gamma$, is the largest, which shows that the additive Gaussian noise tends to mainly produce small perturbations. However, at larger x, saying $x = 3\sigma = 3b = 3\gamma$, the density of the Laplace distribution is more than 5 times that of the Gaussian distribution, and the density of the Cauchy distribution is even more than 7 times. Hence, the Laplace and Cauchy distributed additive noise tend to corrupt images with high perturbations.

Figure 1 depicts the PDFs of the Gaussian, Laplace, and Cauchy distributions. From Fig. 1a, we see that these three distributions have different behaviours at the peaks and tails. See the details in the zoom-ins. Figure 1b depicts the portion around the peaks of the three distributions. The Gaussian distribution has the same peak as the Cauchy distribution while the density of the Gaussian distribution is slightly higher on both sides of the peak. Figure 1c depicts the portion around the tails of the three distributions. The tail of the Laplace
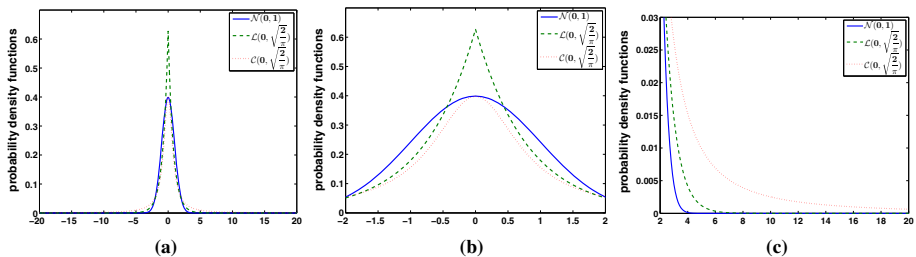

Fig. 1 Comparison for probability density functions of $\mathcal{N}(0, 1)$, $\mathcal{L}(0, \sqrt{\frac{2}{\pi}})$ and $\mathcal{C}(0, \sqrt{\frac{2}{\pi}})$. **a** The plots of three distributions, **b** the zoomed-in portion of the *curves* around the peaks, **c** the zoomed-in portion of the *curves* around the tails

distribution is closer to that of the Cauchy distribution than Gaussian distribution, but there still exists a big gap between the densities of the Laplace and Cauchy distributions. Therefore, the Cauchy distribution cannot be simply replaced with the Gaussian or Laplace distribution during image restoration.

3 Nonconvex Variational Model

This section describes our model of deblurring and denoising. In [42], a variational model for denoising was proposed. To make our exposition self-contained, we deduce a similar nonconvex variational model for deblurring and denoising based on the maximum a posteriori (MAP) estimator and Bayes' rule.

3.1 Nonconvex Variational Model Via MAP Estimator

We consider $f(x)$ and $u(x)$ as random variables for each $x \in \Omega$. The MAP estimator of u is the most likely value of u given f, i.e., $u^* = \arg\max_u P(u|f)$. Based on Bayes' rule and the independence of $u(x)$ and $f(x)$ for all $x \in \Omega$, we obtain

$$
\begin{aligned}
\arg\max_u P(u|f) &= \arg\max_u \frac{P(f|u)P(u)}{P(f)} \\
&= \arg\min_u -\log P(f|u) - \log P(u) \\
&= \arg\min_u -\int_\Omega \log P(f(x)|u(x))\, dx - \log P(u),
\end{aligned}
\tag{4}
$$

where the term $\log P(f(x)|u(x))$ describes the degradation process that produces f from u based on (1), and $\log P(u)$ is the prior on u. Since $\eta(x)$ follows $\mathcal{C}(0, \gamma)$ for each $x \in \Omega$, we have

$$
P(f(x)|u(x)) = \frac{\gamma}{\pi\left(((Ku)(x) - f(x))^2 + \gamma^2\right)}.
$$

In addition, we use the prior $P(u) = \exp(-\frac{2}{\lambda} \int_\Omega |Du|)$. Then, we arrive at the variational model for deblurring and denoising:

$$
\min_{u \in BV(\Omega)} \frac{\lambda}{2} \int_\Omega \log\left((Ku - f)^2 + \gamma^2\right)\, dx + \int_\Omega |Du|,
\tag{5}
$$

where $\lambda > 0$ is the regularization parameter. Although $\int_\Omega |Du|$ is convex, due to the logarithm in the data-fitting term, $\int_\Omega \log\left(\gamma^2 + (Ku - f)^2\right)\, dx$ is nonconvex. Therefore, the numerical solution of (5) depends on the numerical approach and how it is initialized.

3.2 Solution Existence and Uniqueness of the Model (5)

According to the properties of the total variation, we prove that there exists at least one solution for the nonconvex variational problem in the BV space.

Theorem 3.1 *Assume that Ω is a connected bounded set with compacted Lipschitz boundary and $f \in L^2(\Omega)$. Suppose that $K \in \mathcal{L}(L^1(\Omega), L^2(\Omega))$ is nonnegative and linear with $K\mathbf{1} \neq 0$. Then the model (5) has at least one solution $u^* \in BV(\Omega)$.*

Proof Let $E(u) = \int_\Omega |Du| + \frac{\lambda}{2} \int_\Omega \log \left(\gamma^2 + (Ku - f)^2\right) \, dx$. Obviously, $E(u)$ is bounded from below. For a minimizing sequence $\{u^k\}$, we know that $E(u^k)$ is bounded, so both $\left\{\int_\Omega |Du^k|\right\}$ and $\int_\Omega \log \left(\gamma^2 + (Ku^k - f)^2\right) \, dx$ are bounded.

Now we apply proof by contradiction to show that $\{Ku^k\}$ is bounded in $L^2(\Omega)$ and therefore also bounded in $L^1(\Omega)$. Assume that $\|Ku^k\|_2 = +\infty$, so there exists a set $E \subset \Omega$, whose measure is not zero, such that for any $x \in E$ we have $Ku^k(x) = +\infty$. Then, with $f \in L^2(\Omega)$ we will also have $\log \left(\gamma^2 + (Ku^k(x) - f(x))^2\right) = +\infty$ for all $x \in E$, which derives a contradiction with $\int_\Omega \log \left(\gamma^2 + (Ku^k - f)^2\right) \, dx < +\infty$.

Based on $\left\{\int_\Omega |Du^k|\right\}$ being bounded, by the Poincaré inequality [2], we have

$$\|u^k - m_\Omega(u^k)\|_2 \leq C \int_\Omega \left| D\left(u^k - m_\Omega(u^k)\right) \right| = C \int_\Omega |Du^k|, \tag{6}$$

where $m_\Omega(u^k) = \frac{1}{|\Omega|} \int_\Omega u^k dx$, C is a positive constant, and $|\Omega|$ represents the measure of Ω. As Ω is bounded, $\|u^k - m_\Omega(u^k)\|_2$ and $\|u^k - m_\Omega(u^k)\|_1$ are bounded for each k. Because $K \in \mathcal{L}(L^1(\Omega), L^2(\Omega))$ is continuous, we have that $\{K(u^k - m_\Omega(u^k))\}$ is bounded in $L^2(\Omega)$ and $L^1(\Omega)$. Thus, we conclude

$$|m_\Omega(u^k)| \cdot \|K\mathbf{1}\|_1 \leq \left\| K\left(u^k - m_\Omega(u^k)\right) \right\|_1 + \|Ku^k\|_1.$$

Due to $K\mathbf{1} \neq 0$, $m_\Omega(u^k)$ is uniformly bounded. Combining with (6), this gives that the sequence $\{u^k\}$ is bounded in $L^2(\Omega)$ and in $L^1(\Omega)$. Recalling that $\left\{\int_\Omega |Du^k|\right\}$ is bounded, we obtain the boundedness of $\{u^k\}$ in $BV(\Omega)$.

Therefore, there exists a subsequence $\{u^{n_k}\}$ in $BV(\Omega)$ that converges strongly in $L^1(\Omega)$ to some $u^* \in BV(\Omega)$ as $k \to \infty$, while $\{Du^{n_k}\}$ converges weakly as a measure to Du^*. Since K is linear and continuous, $\{Ku^{n_k}\}$ converges strongly to Ku^* in $L^2(\Omega)$. By the lower semicontinuity of total variation and Fatou's lemma, we conclude that u^* is a solution of the model (5). $\square$

Although the objective function in (5) is nonconvex, we are still able to obtain a result on the uniqueness of the solution.

Theorem 3.2 *Assume that $f \in L^2(\Omega)$ and K is injective. Then, the model (5) has a unique solution u^* in $\Omega_U := \{u \in BV(\Omega) : f(x) - \gamma < (Ku)(x) < f(x) + \gamma \text{ for all } x \in \Omega\}$.*

Proof For each fixed $x \in \Omega$, we define a function $g : \mathbb{R} \to \mathbb{R}$:

$$g(t) = \log \left(\gamma^2 + (t - f(x))^2\right).$$

Since the second order derivative of g:

$$g''(t) = \frac{2\left(\gamma^2 - (t - f(x))^2\right)}{\left(\gamma^2 + (t - f(x))^2\right)^2}$$

is positive when $f(x) - \gamma < t < f(x) + \gamma$, g is strictly convex in this case. Since K is injective, we have that, if $f(x) - \gamma < (Ku)(x) < f(x) + \gamma$, $g((Ku)(x))$ is strictly convex. By the convexity of TV and linearity of K, the objective function of the model (5) is strictly convex in Ω_U. Hence, there exists a unique solution for the model (5) in Ω_U. $\square$

Note that Cauchy noise is so impulsive that, even with a small γ, many points in f are still heavily corrupted and thus some impulsive noise is still left in the images in Ω_U. If we also take the smoothing property of K into account, then the unique solution in Ω_U will not be satisfactory. In Sect. 5.1, we will demonstrate this point numerically.

4 Proposed ADMM Algorithm

Due to the nonconvexity of the variational model (5), different numerical algorithms and initializations may yield different solutions. Taking advantage of the recent result in [48], in this section we apply the ADMM algorithm to the minimization problem (5), which restores images degraded by blurring and Cauchy noise. Then, we prove that the proposed algorithm is globally convergent to a stationary point.

4.1 The ADMM Algorithm for Nonconvex and Nonsmooth Problem

We briefly review the ADMM algorithm and its recent convergence result under nonconvexity and nonsmoothness.

Let $\mathbf{x} = [x_1^\top, \cdots, x_s^\top]^\top \in \mathbb{R}^N$ and $\mathbf{A} = [A_1, \cdots, A_s] \in \mathbb{R}^{M \times N}$ where $x_i \in \mathbb{R}^{n_i}$, $A_i \in \mathbb{R}^{M \times n_i}$, $\sum_{i=1}^s n_i = N$. We consider the minimization problem formulated as:

$$\min_{\mathbf{x}, y} \mathcal{F}(\mathbf{x}) + \mathcal{G}(y) \tag{7}$$

$$\text{s.t. } \mathbf{A}\mathbf{x} + By = 0,$$

where $\mathcal{F}(\mathbf{x})$ is a continuous function, $\mathcal{G}(y)$ is a differentiable, and $y \in \mathbb{R}^L$ is a variable with the corresponding coefficient $B \in \mathbb{R}^{M \times L}$. In general, $\mathcal{F}$ can be nonsmooth and nonconvex, and $\mathcal{G}$ can be nonconvex (but is differentiable as stated). By introducing a Lagrangian multiplier $w \in \mathbb{R}^M$ for the linear constraint $\mathbf{A}\mathbf{x} + By = 0$, we obtain the augmented Lagrangian:

$$\mathcal{L}_\beta(\mathbf{x}, y; w) = \mathcal{F}(\mathbf{x}) + \mathcal{G}(y) + w^\top(\mathbf{A}\mathbf{x} + By) + \frac{\beta}{2}\|\mathbf{A}\mathbf{x} + By\|_2^2,$$

where $\beta > 0$ is a penalty parameter.

Extending from the classic ADMM [7,20], the multi-block ADMM generates the iterates $(\mathbf{x}^{k+1}, y^{k+1})$ by

$$\begin{cases} x_1^{k+1} \in \arg\min_{x_1} \mathcal{L}_\beta\left(x_1, x_2^k, \cdots, x_s^k, y^k; w^k\right) \\[2mm] \quad\vdots \quad \in \quad\vdots \quad\vdots \\[2mm] x_s^{k+1} \in \arg\min_{x_s} \mathcal{L}_\beta\left(x_1^{k+1}, x_2^{k+1}, \cdots, x_s, y^k; w^k\right) \\[2mm] y^{k+1} \in \arg\min_{y} \mathcal{L}_\beta\left(\mathbf{x}^{k+1}, y; w^k\right) \\[2mm] w^{k+1} = w^k + \beta\left(\mathbf{A}\mathbf{x}^{k+1} + By^{k+1}\right), \end{cases} \tag{8}$$

where we use $\in \arg\min$ when minimizers are not necessarily unique (in which case, any minimizer is fine). The general assumption is that all subproblems have minimizers. The convergence result of the ADMM algorithm under nonconvexity and nonsmoothness is summarized as follows [48]. We present the conditions that are simplified to fit our need yet more restrictive than those in [48].

Theorem 4.1 *Let $\mathcal{D} = \{(\mathbf{x}, y) \in \mathbb{R}^{N+L} : \mathbf{A}\mathbf{x} + By = 0\}$ be a nonempty feasible set. Assume $\mathcal{F}(\mathbf{x}) + \mathcal{G}(y)$ is $\mathcal{D}$-coercive, that is, for $(\mathbf{x}, y) \in \mathcal{D}$, $\mathcal{F}(\mathbf{x}) + \mathcal{G}(y) \to \infty$ as $\|(\mathbf{x}, y)\| \to \infty$. Also, assume that $\mathbf{A}$, B have full column rank[1] and $Im(\mathbf{A}) \subset Im(B)$. Further assume that*

---

[1] The full column rank assumption can be weakened to the following assumption: for the general matrix $\mathbf{A}$ and B, there exists two Lipschitz continuous maps such that $\mathcal{H}_1(u) \in \arg\min_x\{\mathcal{F}(\mathbf{x}) : \mathbf{A}\mathbf{x} = u\}$ and $\mathcal{H}_2(v) \in \arg\min_y\{\mathcal{G}(y) : By = v\}$.

$\mathcal{F}(\mathbf{x})$ *is either restricted prox-regular[2] or piecewise linear, and $\mathcal{G}(y)$ is Lipschitz differentiable with the constant $L_{\nabla \mathcal{G}} > 0$. Then, for any β larger than a certain constant β_0 and starting from any initialization $(\mathbf{x}^0, y^0, w^0)$, ADMM (8) produces a sequence of iterates that has a convergent subsequence, whose limit is a stationary point $(\mathbf{x}^*, y^*, w^*)$ of the augmented Lagrangian $\mathcal{L}_\beta(\mathbf{x}, y; w)$. If in addition $\mathcal{L}_\beta$ satisfies the Kurdyka–Łojasiewicz (KL) inequality [4,8,32,33], then the result improves to global convergence to that stationary point.*

4.2 The ADMM Algorithm for Solving (5)

Taking advantage of the ADMM convergence result, we apply it to solve the nonconvex variational model in (5) for simultaneous denoising and deblurring. Hereafter, we switch to the discrete form, but, for the sake of simplicity, we still use the same letters defined in the continuous context. We assume that the discrete image domain Ω contains $n \times n$ pixels. The discrete minimization nonconvex model of (5) is formulated as follows:

$$\min_{u \in \mathbb{R}^{n^2}} \|\nabla u\|_1 + \frac{\lambda}{2} \left\langle \log \left(\gamma^2 + (Ku - f)^2 \right), \mathbf{1} \right\rangle, \tag{9}$$

where $f \in \mathbb{R}^{n^2}$ is obtained by stacking the columns of the corresponding $n \times n$ gray-scale image, and $K \in \mathbb{R}^{n^2 \times n^2}$. The TV regularization $\|\nabla u\|_1$ is defined as:

$$\|\nabla u\|_1 = \sum_{i=1}^{n^2} \sqrt{(\nabla_x u)_i^2 + (\nabla_y u)_i^2},$$

where $\nabla_x \in \mathbb{R}^{n^2 \times n^2}$ and $\nabla_y \in \mathbb{R}^{n^2 \times n^2}$ are the discrete first order forward differences in the x- and y-directions, respectively. The discrete gradient of u, ∇u, is defined as $\nabla u = [(\nabla_x u)^\top, (\nabla_y u)^\top]^\top \in \mathbb{R}^{2n^2}$.

To derive the ADMM algorithm for our model, we introduce a new auxiliary variable $v \in \mathbb{R}^{n^2}$ and obtain the following constrained nonconvex minimization problem:

$$\min_{u,v \in \mathbb{R}^{n^2}} \|\nabla u\|_1 + \frac{\lambda}{2} \left\langle \log \left(\gamma^2 + (v - f)^2 \right), \mathbf{1} \right\rangle \tag{10}$$
$$\text{s.t. } Ku = v.$$

Let $w \in \mathbb{R}^{n^2}$ be the Lagrangian multiplier for the constraint $Ku = v$. Then we have the corresponding augmented Lagrangian:

$$\mathcal{L}_\beta(u, v, w) = \|\nabla u\|_1 + \frac{\lambda}{2} \left\langle \log \left(\gamma^2 + (v - f)^2 \right), \mathbf{1} \right\rangle + \langle w, Ku - v \rangle + \frac{\beta}{2} \|Ku - v\|_2^2,$$

where $\beta > 0$ is a penalty parameter. The whole algorithm for restoring the blurred images corrupted by Cauchy noise is given in Algorithm 1.

In Algorithm 1, the dominant computation is the steps to solve the two minimization subproblems in (11) and (12). The u-subproblem (11) can be efficiently solved by many

---

[2] A function $h : \mathbb{R}^N \to \mathbb{R}$ is restricted prox-regular, if for any sufficiently large $M \in \mathbb{R}_+$ and any bounded set $T \subset \mathbb{R}^N$, there exists $\tau > 0$ such that

$$h(y) + \frac{\tau}{2} \|x - y\|^2 \geq h(x) + \langle d, y - x \rangle, \text{ for all } x, y \in T \setminus S_M, d \in \partial h(x), \|d\| \leq M$$

where $S_M := \{d \in dom(\partial h) : \|d\| > M \text{ for all } d \in \partial h\}$ is the exclusion. When $\mathbf{x}$ has multiple subblocks $x_1, \ldots, x_n$, on the first block x_1, the function $\mathcal{F}$ is only required to be proper and lower semi-continuous.

Algorithm 1 ADMM algorithm for solving (9)

1: Initialize u^0, v^0, w^0; set λ, β.
2: For $k = 1, 2, \ldots$, calculate u^{k+1}, v^{k+1}, w^{k+1} by

$$u^{k+1} \in \arg\min_{u} \|\nabla u\|_1 + \frac{\beta}{2}\left\|Ku - v^k + \frac{w^k}{\beta}\right\|_2^2 \tag{11}$$

$$v^{k+1} = \arg\min_{v} \frac{\lambda}{2}\left\langle \log\left(\gamma^2 + (v - f)^2\right), \mathbf{1} \right\rangle + \frac{\beta}{2}\left\|Ku^{k+1} - v + \frac{w^k}{\beta}\right\|_2^2 \tag{12}$$

$$w^{k+1} = w^k + \beta(Ku^{k+1} - v^{k+1}) \tag{13}$$

3: If u^{k+1} satisfies the stopping criteria, return u^{k+1} and stop.

---

methods, for instance, the dual algorithm [10], the split-Bregman algorithm [9,23,43,50], the primal-dual algorithm [11,18], the infeasible primal-dual algorithm of semi-smooth Newton-type [25], the ADMM algorithm [14,52], as well as the max-flow algorithm [22]. Here, we apply the dual algorithm proposed in [10]. Since the objective function in (12) is twice continuously differentiable, we can utilize the Newton method to solve it efficiently. Inspired by [48], as a special case of (7), we have the following convergence result for Algorithm 1. In addition, taking some specific properties of the variational model (9) into account, we provide a relatively simple proof.

Theorem 4.2 *Let* (u^0, v^0, w^0) *be any initial point and* $\{(u^k, v^k, w^k)\}$ *be the sequence of iterates generated by Algorithm 1. Then, if* $\beta > \frac{\lambda}{\gamma^2}$ *and* K *has full column rank, the sequence* $\{(u^k, v^k, w^k)\}$ *converges globally to a point* (u^*, v^*, w^*)*, which is a stationary point of* $\mathcal{L}_\beta$*.*

In order to prove Theorem 4.2, based on the model in (7), we define the following functions:

$$\mathcal{F} : \mathbb{R}^{n^2} \to \mathbb{R}, \qquad \mathcal{F}(u) = \|\nabla u\|_1$$

$$\mathcal{G} : \mathbb{R}^{n^2} \to \mathbb{R}, \qquad \mathcal{G}(v) = \frac{\lambda}{2}\left\langle \log\left(\gamma^2 + (v - f)^2\right), \mathbf{1} \right\rangle. \tag{14}$$

The feasible set is $\Omega_F = \{(u, v) \in \mathbb{R}^{n^2} \otimes \mathbb{R}^{n^2} : Ku - v = 0\}$. First, we give some useful lemmas that will be used in the main proof.

Lemma 4.1 *The iterates of Algorithm 1 satisfy:*

1. for all $k \in \mathbb{N}$*,* $\nabla\mathcal{G}(v^k) = w^k$*;*
2. $\|w^k - w^{k+1}\| \leq \frac{\lambda}{\gamma^2}\|v^k - v^{k+1}\|$*.*

Proof Substituting (13) on w^k into the first-order optimality condition of the v-subproblem on v^k: $\nabla\mathcal{G}(v^k) - w^{k-1} + \beta(v^k - Ku^k) = 0$, we have $\nabla\mathcal{G}(v^k) = w^k$ for all $k \in \mathbb{N}$.

Since $\mathcal{G}$ is smooth, we can calculate its second derivative

$$\nabla^2_{v_i}\mathcal{G} = \lambda\frac{\gamma^2 - (v_i - f)^2}{(\gamma^2 + (v_i - f)^2)^2} \leq \frac{\lambda}{\gamma^2}, \tag{15}$$

and thus $L_{\nabla\mathcal{G}} = \frac{\lambda}{\gamma^2}$ is a Lipschitz constant for $\nabla\mathcal{G}$. Consequently, we obtain the bound

$$\|w^k - w^{k+1}\| = \|\nabla\mathcal{G}(v^k) - \nabla\mathcal{G}(v^{k+1})\| \leq L_{\nabla\mathcal{G}}\|v^k - v^{k+1}\|.$$

$\square$

Lemma 4.2 *Let $\{(u^k, v^k, w^k)\}$ be the sequence of iterates generated by Algorithm 1. If $\beta > \frac{\lambda}{\gamma^2}$, then $\{(u^k, v^k, w^k)\}$ satisfies:*

1. $\mathcal{L}_\beta(u^k, v^k, w^k)$ *is lower bounded and nonincreasing for all $k \in \mathbb{N}$;*
2. $\{(u^k, v^k, w^k)\}$ *is bounded.*

Proof According to the optimality condition of the u-subproblem (11), we define

$$d^{k+1} := -\left(K^\top w^k + \beta K^\top (Ku^{k+1} - v^k)\right) \in \partial \mathcal{F}(u^{k+1}). \tag{16}$$

From (11) and the definition of subgradient $\mathcal{F}$, it follows

$$\mathcal{L}_\beta(u^k, v^k, w^k) - \mathcal{L}_\beta(u^{k+1}, v^k, w^k)$$

$$= \mathcal{F}(u^k) - \mathcal{F}(u^{k+1}) + \langle w^k, Ku^k - Ku^{k+1}\rangle + \frac{\beta}{2}\|Ku^k - v^k\|^2 - \frac{\beta}{2}\|Ku^{k+1} - v^k\|^2$$

$$= \mathcal{F}(u^k) - \mathcal{F}(u^{k+1}) + \langle K^\top w^k + \beta K^\top (Ku^{k+1} - v^k), u^k - u^{k+1}\rangle + \frac{\beta}{2}\|Ku^k - Ku^{k+1}\|^2$$

$$= \mathcal{F}(u^k) - \mathcal{F}(u^{k+1}) - \langle d^{k+1}, u^k - u^{k+1}\rangle + \frac{\beta}{2}\|Ku^k - Ku^{k+1}\|^2$$

$$\geq \frac{\beta}{2}\|Ku^k - Ku^{k+1}\|^2 \tag{17}$$

where the second equality follows from the cosine rule: $\|b + c\|^2 - \|a + c\|^2 = \|b - a\|^2 + 2\langle a + c, b - a\rangle$ and the last inequality follows from the convexity of $\mathcal{F}(u)$.

For the updates of v^{k+1}, w^{k+1}, by the cosine rule and Lemma 4.1, we have

$$\mathcal{L}_\beta(u^{k+1}, v^k, w^k) - \mathcal{L}_\beta(u^{k+1}, v^{k+1}, w^{k+1})$$

$$= \mathcal{G}(v^k) - \mathcal{G}(v^{k+1}) + \langle w^{k+1}, v^{k+1} - v^k\rangle - \beta\|Ku^{k+1} - v^{k+1}\|^2 + \frac{\beta}{2}\|v^k - v^{k+1}\|^2$$

$$= \mathcal{G}(v^k) - \mathcal{G}(v^{k+1}) - \langle \nabla\mathcal{G}(v^{k+1}), v^k - v^{k+1}\rangle - \frac{1}{\beta}\|w^k - w^{k+1}\|^2 + \frac{\beta}{2}\|v^k - v^{k+1}\|^2$$

$$\geq -\frac{L_{\nabla\mathcal{G}}}{\beta}\|v^k - v^{k+1}\|^2 + \frac{\beta}{2}\|v^k - v^{k+1}\|^2$$

$$= C\|v^k - v^{k+1}\|^2 \tag{18}$$

where we have applied the inequality

$$\mathcal{G}(v) - \mathcal{G}(v') - \langle\nabla\mathcal{G}(v'), v - v'\rangle \geq 0$$

and used the constant $C = \frac{\beta}{2} - \frac{L_{\nabla\mathcal{G}}}{2}$. In order to ensure $C > 0$, we need the penalty parameter β to satisfy:

$$\beta > L_{\nabla\mathcal{G}} \overset{(15)}{=} \frac{\lambda}{\gamma^2}.$$

According to (17) and (18), we have

$$\mathcal{L}_\beta(u^k, v^k, w^k) - \mathcal{L}_\beta(u^{k+1}, v^{k+1}, w^{k+1}) \geq C\|v^k - v^{k+1}\|^2 + \frac{\beta}{2}\|Ku^k - Ku^{k+1}\|^2.$$

This means that $\mathcal{L}_\beta(u^k, v^k, w^k)$ is nonincreasing in $k \in \mathbb{N}$.

As K has full column rank, there exists $\hat{v}$ such that $Ku^k - \hat{v} = 0$. Therefore, we have

$$\mathcal{F}(u^k) + \mathcal{G}(\hat{v}) \geq \min_{u,v}\{\mathcal{F}(u) + \mathcal{G}(v) : Ku - v = 0\} > -\infty.$$

Thus, we arrive at

$$\begin{aligned}
\mathcal{L}_\beta(u^k, v^k, w^k) &= \mathcal{F}(u^k) + \mathcal{G}(v^k) + \langle w^k, Ku^k - v^k \rangle + \frac{\beta}{2}\|Ku^k - v^k\|^2 \\
&= \mathcal{F}(u^k) + \mathcal{G}(v^k) + \langle \nabla\mathcal{G}(v^k), \hat{v} - v^k \rangle + \frac{\beta}{2}\|Ku^k - v^k\|^2 \\
&\geq \mathcal{F}(u^k) + \mathcal{G}(\hat{v}) - \frac{L_{\nabla\mathcal{G}}}{2}\|\hat{v} - v^k\|^2 + \frac{\beta}{2}\|Ku^k - v^k\|^2 \\
&= \mathcal{F}(u^k) + \mathcal{G}(\hat{v}) + \frac{\beta - L_{\nabla\mathcal{G}}}{2}\|Ku^k - v^k\|^2 > -\infty.
\end{aligned}$$

Since $\mathcal{L}_\beta(u^k, v^k, w^k)$ is upper bounded by $\mathcal{L}_\beta(u^0, v^0, w^0)$ and obviously $\mathcal{F}(u) + \mathcal{G}(v)$ is coercive over Ω_F, we conclude that $\{u^k\}$ and $\{v^k\}$ are bounded. By Lemma 4.1, $\{w^k\}$ is also bounded. $\qquad\square$

Lemma 4.3 *Let* $\partial\mathcal{L}(u^{k+1}, v^{k+1}, w^{k+1}) = (\partial_u\mathcal{L}, \nabla_v\mathcal{L}, \nabla_w\mathcal{L})$. *Then, there exists a constant* $C_1 > 0$ *such that, for all* $k \geq 1$, *for some* $p^{k+1} \in \partial\mathcal{L}(u^{k+1}, v^{k+1}, w^{k+1})$ *we have* $\|p^{k+1}\| \leq C_1\|v^k - v^{k+1}\|$.

Proof Because $\nabla_w\mathcal{L} = Ku^{k+1} - v^{k+1} = \frac{1}{\beta}(w^{k+1} - w^k)$ and $\nabla_v\mathcal{L} = w^{k+1} - w^k$, based on Lemma 4.1, we have

$$\|\nabla_w\mathcal{L}\| \leq \frac{L_{\nabla\mathcal{G}}}{\beta}\|v^k - v^{k+1}\|, \tag{19}$$

$$\|\nabla_v\mathcal{L}\| \leq L_{\nabla\mathcal{G}}\|v^k - v^{k+1}\|. \tag{20}$$

By the definition of the subgradient, we have

$$\begin{aligned}
\partial_u\mathcal{L} &= \partial\mathcal{F}(u^{k+1}) + K^\top w^{k+1} + \beta K^\top(Ku^{k+1} - v^{k+1}) \\
&= \partial\mathcal{F}(u^{k+1}) + K^\top w^k + \beta K^\top(Ku^{k+1} - v^k) \\
&\quad + K^\top(w^{k+1} - w^k) + \beta K^\top(v^k - v^{k+1}).
\end{aligned} \tag{21}$$

Thus, according to the optimal condition $0 \in \partial\mathcal{F}(u^{k+1}) + K^\top w^k + \beta K^\top(Ku^{k+1} - v^k)$, we have $K^\top(w^{k+1} - w^k) + \beta K^\top(v^k - v^{k+1}) \in \partial_u\mathcal{L}$. Letting

$$p^{k+1} \triangleq \left(K^\top(w^{k+1} - w^k) + \beta K^\top(v^k - v^{k+1}), \frac{1}{\beta}(w^{k+1} - w^k), w^{k+1} - w^k\right)$$

and combining (19), (20), (21), and Lemma 4.1, we arrive at

$$\begin{aligned}
\|p^{k+1}\| &\leq \left(L_{\nabla\mathcal{G}}\left(1 + \frac{1}{\beta} + \|K\|\right) + \beta\|K\|\right)\|v^k - v^{k+1}\| \\
&= C_1\|v^k - v^{k+1}\|
\end{aligned}$$

where $C_1 = \left(L_{\nabla\mathcal{G}}(1 + \frac{1}{\beta} + \|K\|) + \beta\|K\|\right)$. $\qquad\square$

Now we give the proof to our main convergence theorem.

Proof of Theorem 4.2 As K has full column rank, the feasible set Ω_F is nonempty. By Lemma 4.2, the iterative sequence $\{(u^k, v^k, w^k)\}$ is bounded, so there exists a convergent subsequence $\{(u^{n_k}, v^{n_k}, w^{n_k})\}$, i.e., $(u^{n_k}, v^{n_k}, w^{n_k})$ converges to (u^*, v^*, w^*) as k goes to infinity. Since $\mathcal{L}_\beta(u^k, v^k, w^k)$ is nonincreasing and lower-bounded, we have $\|K(u^k - u^{k+1})\| \to 0$ and $\|v^k - v^{k+1}\| \to 0$ as $k \to \infty$. According to Lemma 4.3, there exists $p^k \in \partial \mathcal{L}_\beta(u^k, v^k, w^k)$ such that $\|p^k\| \to 0$. Further, this leads to $\|p^{n_k}\| \to 0$ as $k \to \infty$. Based on the definition of the general subgradient [39], we obtain that $0 \in \partial \mathcal{L}_\beta(u^*, v^*, w^*)$, i.e., (u^*, v^*, w^*) is a stationary point.

Referring to [47,51], the function $\mathcal{F}(u)$ is semi-algebraic, and $\mathcal{G}(v)$ is a real analytic function. Thus, we conclude that $\mathcal{L}_\beta$ satisfies the KL inequality [8]. Then, as in the proof of Theorem 2.9 in [5], we can deduce that the iterative sequence $\{(u^k, v^k, w^k)\}$ is globally convergent to (u^*, v^*, w^*). □

Remark 1 In Theorem 4.2 we need K to have full column rank. Since K is a blurring matrix in our problem, this requirement is satisfied.

5 Numerical Experiments

In this section, we present the results of several numerical experiments to demonstrate the performance of the proposed method for restoring images corrupted by blurs and Cauchy noise. Here, we use ten 8-bit 256-by-256 gray-scale test images, see Fig. 2. All numerical results are performed under Windows 10 and Matlab Version 7.10 (R2012a) running on a Lenovo laptop with a 1.7 GHz Intel Core CPU and 4GB RAM.

We utilize the peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM) [49] as performance measures, which are respectively defined as

$$\text{PSNR} = 20 \log 10 \left(\frac{255n}{\|\tilde{u} - u\|_2} \right), \quad \text{SSIM} = \frac{2\mu_{\tilde{u}}\mu_u(2\sigma + c_2)}{(\mu_{\tilde{u}}^2 + \mu_u^2 + c_1)(\sigma_{\tilde{u}}^2 + \sigma_u^2 + c_2)},$$

where $\tilde{u}$ is the restored image, u is the original image, $\mu_{\tilde{u}}$ and μ_u denote their respective means, $\sigma_{\tilde{u}}^2$ and σ_u^2 represent their respective variances, σ is the covariance of $\tilde{u}$ and u,
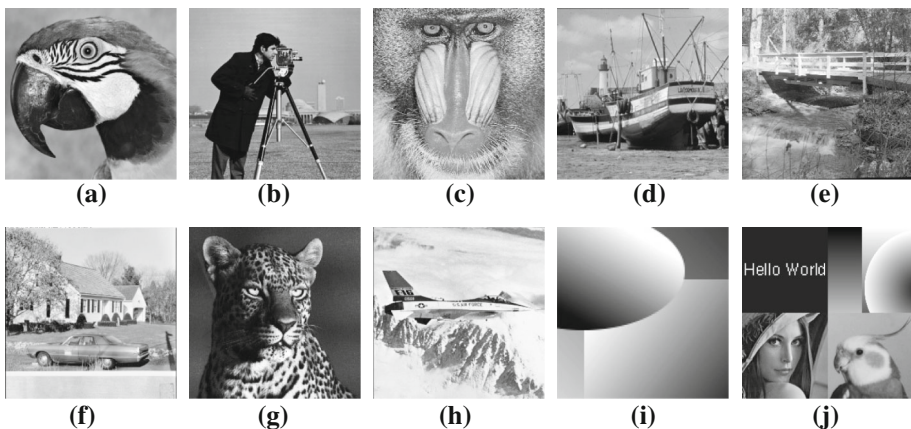

Fig. 2 Original images **a** Parrot, **b** Cameraman, **c** Baboon, **d** Boat, **e** Bridge, **f** House, **g** Leopard, **h** Plane, **i** Test, **j** Montage

Table 1 PSNR and SSIM for the test images "Parrot" and "Cameraman" with different initial values

Noise	Condition	Parrot		Cameraman	
		PSNR	SSIM	PSNR	SSIM
$\gamma = 5$	I	29.06	0.8729	28.72	0.8520
	II	27.83	0.8505	26.31	0.8500
	III	24.88	0.7730	23.75	0.7275
$\gamma = 10$	I	27.12	0.8268	26.67	0.7949
	II	26.68	0.8218	25.60	0.8093
	III	22.87	0.6895	22.43	0.6653

(a)				(b)				(c)				(d)
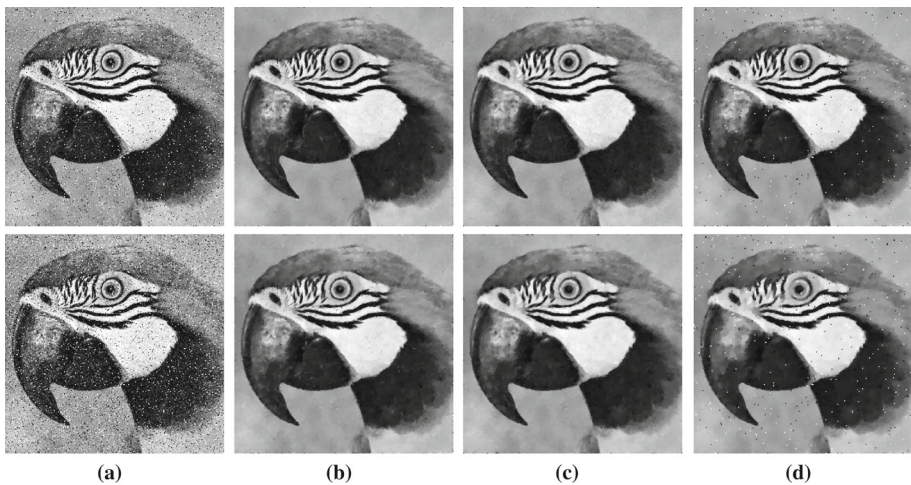
Fig. 3 Comparison of different initial values for removing Cauchy noise in the image "Parrot", with $\gamma = 5$ (in the 1st row) and 10 (in the 2nd row). **a** Noisy images, **b** restored images of (I), **c** restored images of (II), **d** restored images of (III)

and $c_1, c_2 > 0$ are constants. PSNR is a good measure of the human subjective sensation, and a higher PSNR implies better quality of the restored image. SSIM conforms with the quality perception of the human visual system (HVS). If the SSIM value is closer to 1, the characteristic (edges and textures) of the restored image is more similar to the original image.

In our method, we set the stopping condition based on the following relative improvement inequality:

$$\left| \frac{E(u^{k+1}) - E(u^k)}{E(u^{k+1})} \right| < \epsilon, \tag{22}$$

where E is the objective function in (9) and $\epsilon = 5 \times 10^{-5}$. In addition, since the regularization parameter λ balances the trade-off between fitting f and TV, we manually tune it in order to obtain the highest PSNRs of the restored images. The selection method of λ is out of the scope in this paper. The parameter β in Algorithm 1 affects the convergent speed. Based on Theorem 4.2, we round $\beta > \frac{\lambda}{\gamma^2}$ up to the nearest value with two digits after the decimal point as β. In addition, we set the iteration number for the Newton method while solving the
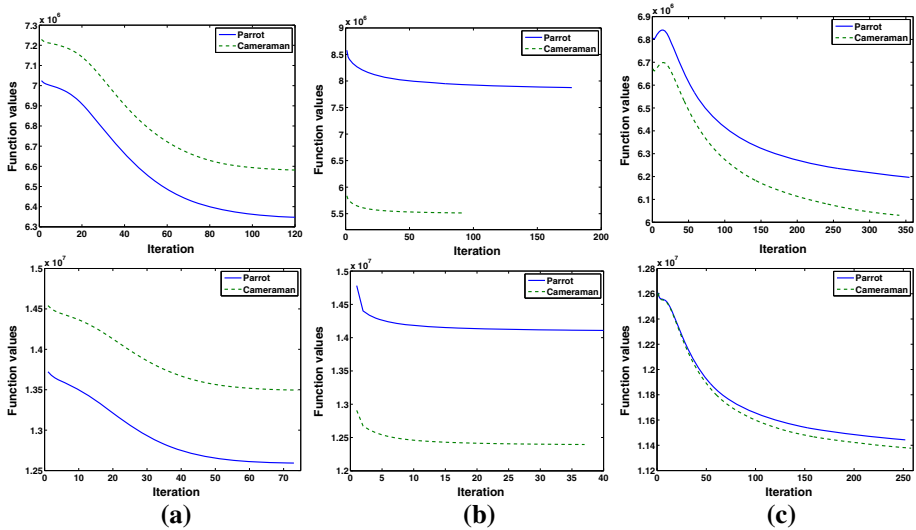
Fig. 4 Plots of the objective function values versus iterations for the noisy images "Parrot" and "Cameraman" with $\gamma = 5$ (in the 1st row) and 10 (in the 2nd row). **a** I **b** II **c** III

Table 2 PSNR and SSIM for the noisy images and the restored images by applying different methods ($\gamma = 5$)

Image	Noisy	PSNR			SSIM		
		Median	conRE	Ours	Median	conRE	Ours
Parrot	19.20	27.18	27.19	**29.06**	0.8341	0.8465	**0.8729**
Cameraman	18.98	25.94	26.51	**28.72**	0.7996	0.8225	**0.8520**
Baboon	17.74	19.18	21.18	**22.56**	0.5069	0.7178	**0.7781**
Boat	18.01	25.94	27.03	**27.94**	0.7779	0.8165	**0.8541**
Bridge	19.13	22.63	24.32	**25.25**	0.6312	0.7857	**0.8097**
House	17.94	24.06	25.25	**26.26**	0.7510	0.7774	**0.8308**
Leopard	19.07	25.34	26.54	**27.51**	0.7787	0.7861	**0.8309**
Plane	17.37	25.09	25.83	**27.23**	0.8235	0.8354	**0.8611**
Test	19.19	34.79	39.55	**40.77**	0.8922	0.9726	**0.9864**
Montage	19.14	27.52	27.94	**31.08**	0.8772	**0.9180**	0.9135

The largest values are given in bold

v-subproblem as 3. The iteration number for solving the u-subproblem equals 5 in denoising and 10 in simultaneous deblurring and denoising.

5.1 Different Initializations

Since our model (9) is nonconvex, though we are able to prove that the ADMM algorithm converges globally to a stationary point from any given starting point (u^0, v^0, w^0), the local minimizers that we obtained may still depend on the initial points. To study the influence of initializations and obtain better restorations, in this section we test three different choices of u^0 in denoising:

Table 3 PSNR and SSIM for the noisy images and the restored images by applying different methods ($\gamma = 10$)

Image	Noisy	PSNR			SSIM		
		Median	conRE	Ours	Median	conRE	Ours
Parrot	16.35	25.51	26.74	**27.12**	0.7254	0.8202	**0.8268**
Cameraman	16.06	24.68	25.68	**26.67**	0.6801	0.7896	**0.7981**
Baboon	14.87	18.79	20.27	**20.96**	0.4671	0.6336	**0.6634**
Boat	15.11	24.39	25.71	**25.79**	0.6843	0.7710	**0.7769**
Bridge	16.30	21.94	23.37	**23.58**	0.5870	**0.7123**	0.7033
House	15.01	22.91	24.24	**24.37**	0.6631	0.7465	**0.7582**
Leopard	16.16	24.16	25.40	**25.60**	0.6981	0.7649	**0.7800**
Plane	14.49	23.64	24.85	**25.25**	0.7104	0.8085	**0.8161**
Test	16.29	30.45	37.38	**38.01**	0.7078	0.9607	**0.9793**
Montage	16.27	26.10	27.16	**28.89**	0.7451	**0.8922**	0.8850

The largest values are given in bold

$$(I) : u^0 = \max(0, \min(f, 255));$$
$$(II) : u^0 = medfile2(f);$$
$$(III) : u^0 = f,$$

where $medfile2(f)$ denotes the result from the median filter with window size 3. Note that due to the impulsive feature of Cauchy noise, the median filter usually provides fairly good results. In addition, based on Theorem 3.2 with u^0 in case (III), we obtain the unique solution in Ω_U.

In Table 1, we list PSNRs and SSIMs for different initial points for the test images "Parrot" and "Cameraman" at the noise levels $\gamma = 5$ and 10. The noisy images are obtained via $f = u + \gamma \frac{\eta_1}{\eta_2}$, where η_1 and η_2 are independent random variables following the Gaussian distribution with mean 0 and variance 1. It is obvious that both PSNRs and SSIMs are highest in the case (I), and are lowest in the case (III), which shows that the unique solution in Ω_U is not a satisfactory local minimizer.

Figure 3 depicts the restored "Parrot" images in order to compare the visual performance due to different initial points. Figure 3d shows the unique solution in Ω_U, and we can see that there is still some noise left in the restored images. The reason is that Cauchy noise is so impulsive that by corrections in a small range, $[-\gamma, \gamma]$, it is not enough to remove all noise. Compared with the results from (II), the ones from (I) include clearer features and less noise, especially in the region around the eye and black stripes of "Parrot". Hence, we choose (I) as initialization in our remaining numerical experiments.

Theorem 4.2 demonstrate that with any given initial points, Algorithm 1 converges globally to a stationary point. Figure 4 depicts the plots of the objective function values in (9) versus the number of iteration in order to observe the convergence of our method. It is clear that the objective function value keeps decreasing over the iterations. Furthermore, our method converges very fast except in case (III), which does not provide good restorations.
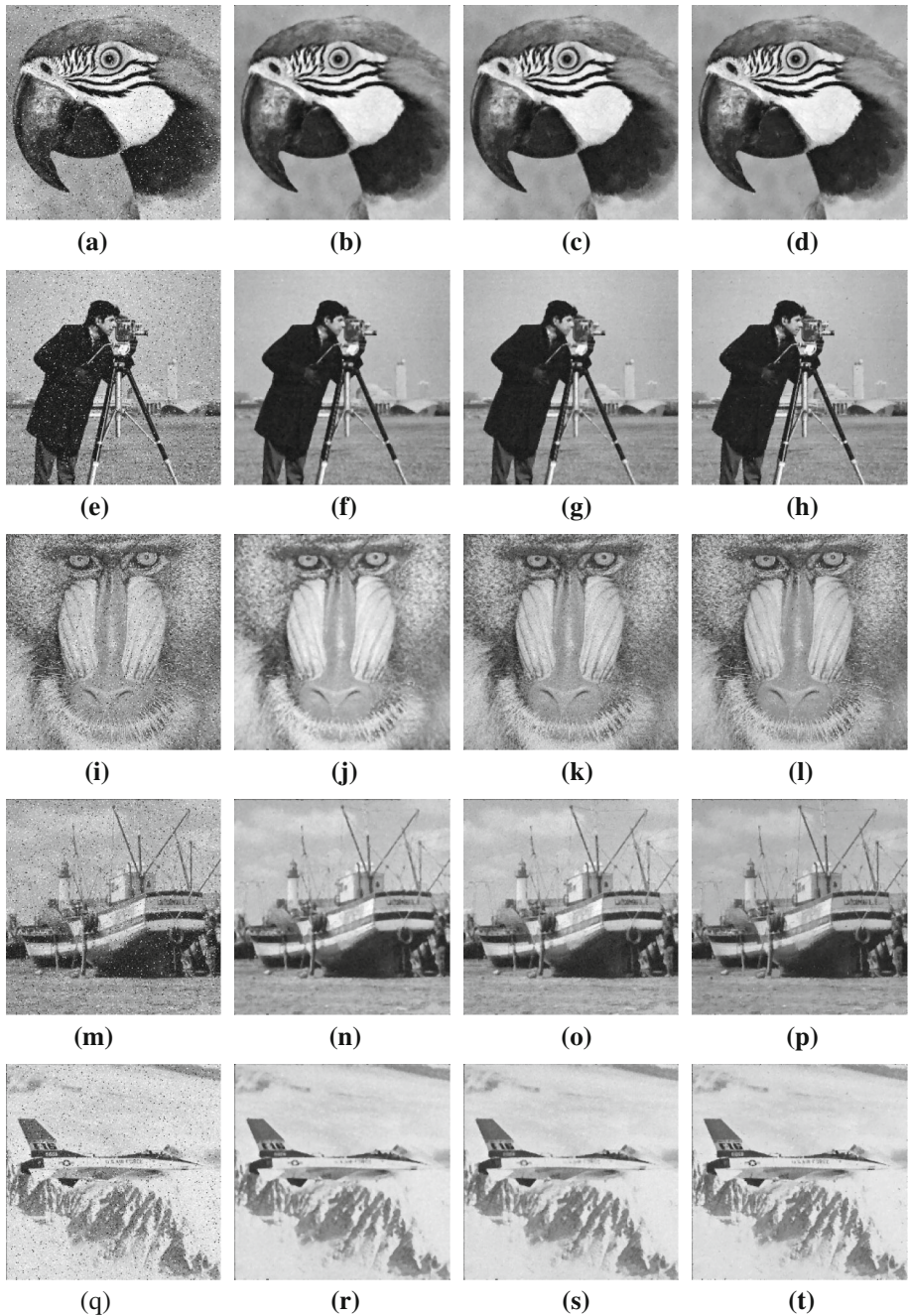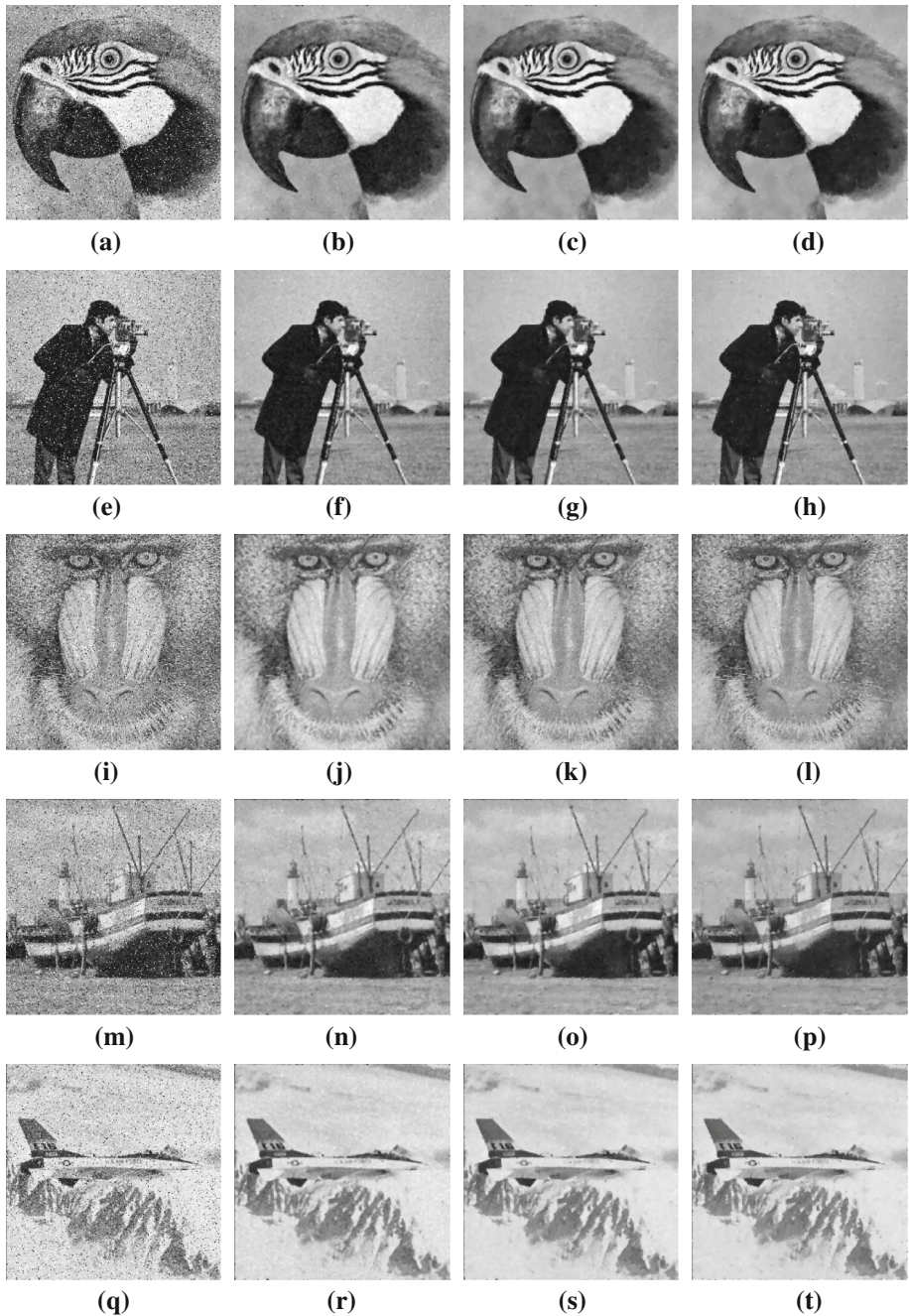
Fig. 5 Comparison of different methods for removing Cauchy noise, with $\gamma = 5$. **a** Noisy image:19.20, **b** Median:27.18, **c** conRE:27.91, **d** Ours:29.06, **e** Noisy image:18.98, **f** Median:25.94, **g** conRE:26.51, **h** Ours:28.72, **i** Noisy image:17.74, **j** Median:19.18, **k** conRE:21.18, **l** Ours:22.56, **m** Noisy image:18.01, **n** Median:25.94, **o** conRE:27.03, **p** Ours:27.94, **q** Noisy image:17.37, **r** Median:25.09, **s** conRE:25.83, **t** Ours:27.25
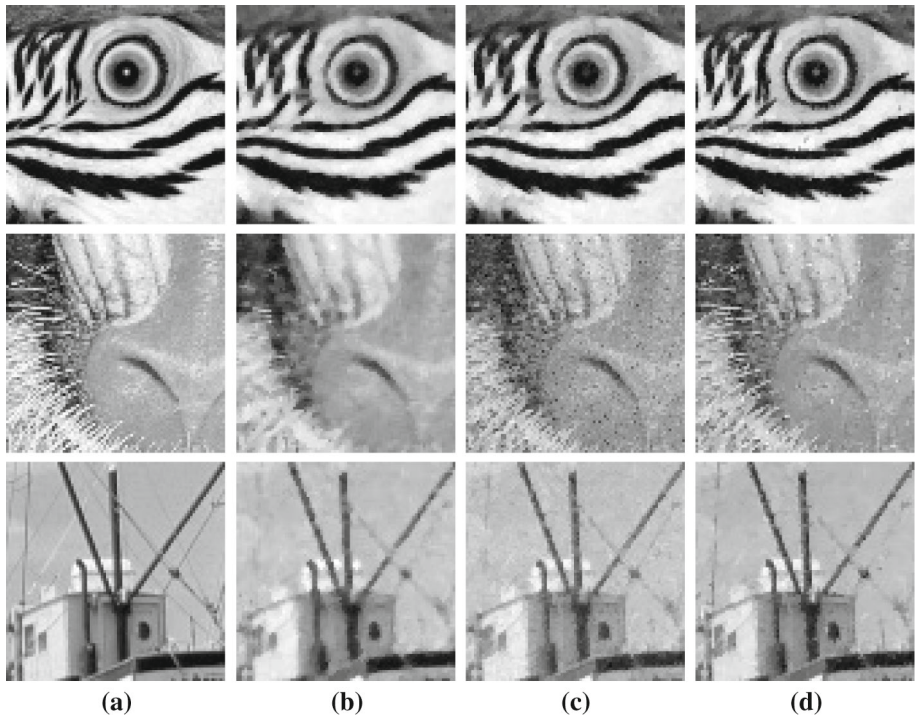
Fig. 6 Comparison of different methods for removing Cauchy noise, with $\gamma = 10$. **a** Noisy image:16.35, **b** Median:25.51, **c** conRE:26.74, **d** Ours:27.12, **e** Noisy image:16.06, **f** Median:24.68, **g** conRE:25.68, **h** Ours:26.67, **i** Noisy image:14.87, **j** Median:18.79, **k** conRE:20.27, **l** Ours:20.96, **m** Noisy image:15.11 **n** Median:24.39, **o** conRE:25.71, **p** Ours:25.79, **q** Noisy image:14.49, **r** Median:23.64, **s** conRE:24.85, **t** Ours:25.25

Fig. 7 Zoomed version of the restored images in Fig. 5. **a** original images; **b** the median filter; **c** the "conRe" model; **d** our method

5.2 Comparisons of Image Deblurring and Denoising

In order to demonstrate the superior performance of our proposed method, we compare it with two other well-known methods: the median filter (matlab function '*medfilt2*') with window size 3 and the convex variational method in [42] ("conRE" for short). For fair comparison, we use the same stopping rule in the convex variational method and adjust the two parameters in the model for highest PSNRs.

First, we compare the three methods for Cauchy noise removal, i.e., by setting K as the identity matrix. Tables 2 and 3 list the PSNRs and SSIMs of the restored images at the noise levels $\gamma = 5$ and $\gamma = 10$, respectively. Obviously, comparing to the two variational methods, the median filter provides the worst PSNRs and SSIMs. Our method always yields the highest PSNRs. Especially at the lower noise level ($\gamma = 5$), our PSNRs are about 1dB higher than the convex method [42]. Furthermore, in most cases, our SSIMs are also higher than others.

In Figs. 5 and 6, we present the results from different methods for removing Cauchy noise from the images "Parrot", "Cameraman", "Baboon", "Boat" and "Plane". Although the median filter effectively removes Cauchy noise, it also oversmooths the edges and destroys many details. It is obvious that two variational methods outperform the median filter. Comparing to the convex method, our nonconvex method can provide better balance between preserving detail and removing noise. To further illustrate the performance of our method, we show the zoomed regions of the restored images "Parrot", "Baboon" and "Boat" in Figs. 7 and 8, where we can clearly see the difference among the results from the three methods,
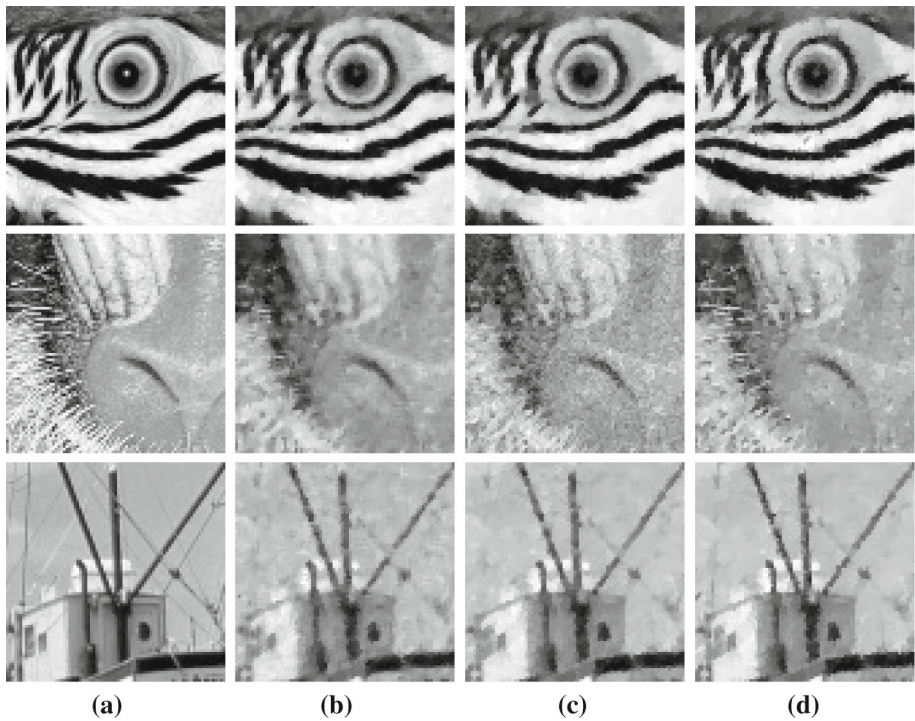
Fig. 8 Zoomed version of the restored images in Fig. 6. **a** Original images; **b** the median filter; **c** the "conRe" model; **d** our method

e.g., the stripes around the eye in "Parrot", the nose and whiskers of "Baboon", and the ropes and iron pillars of "Boat".

In the following experiments, we compare the three methods on recovering images corrupted by blurs and Cauchy noise. Here, we consider the Gaussian blur with size 7 and standard deviation 3, and the out-of-focus blur with size 5. Further, Cauchy noise with $\gamma = 5$ is added into the blurry images. Tables 4 and 5 list the PSNRs and SSIMs by applying different methods to the images "Parrot", "Cameraman", "Plane" and "Test". Figures 9 and 10 show the restored images.

Table 4 PSNR and SSIM for the images degraded by Gaussian blur and Cauchy noise ($\gamma = 5$) and the restored images by different methods

Image	Noisy	PSNR			SSIM		
		Median	conRE	Ours	Median	conRE	Ours
Parrot	17.16	21.13	23.73	**24.45**	0.6698	0.7174	**0.7649**
Cameraman	17.17	21.61	23.51	**24.10**	0.6397	0.6846	**0.7490**
Plane	15.74	20.52	22.11	**22.57**	0.6508	0.6666	**0.7521**
Test	18.86	29.60	32.31	**36.71**	0.8713	0.8263	**0.9647**

The largest values are given in bold

Table 5 PSNR and SSIM for the images degraded by the out-of-focus blur and Cauchy noise ($\gamma = 5$) and the restored images by different methods

Image	Noisy	PSNR			SSIM		
		Median	conRE	Ours	Median	conRE	Ours
Parrot	17.59	22.26	25.02	**25.45**	0.7157	0.7496	**0.8005**
Cameraman	17.42	22.36	24.62	**25.08**	0.6777	0.7125	**0.7735**
Plane	15.98	21.31	23.41	**23.77**	0.6956	0.7210	**0.7854**
Test	18.94	30.61	33.68	**36.97**	0.8784	0.8553	**0.9560**

The largest values are given in bold

(a)   (b)   (c)   (d)
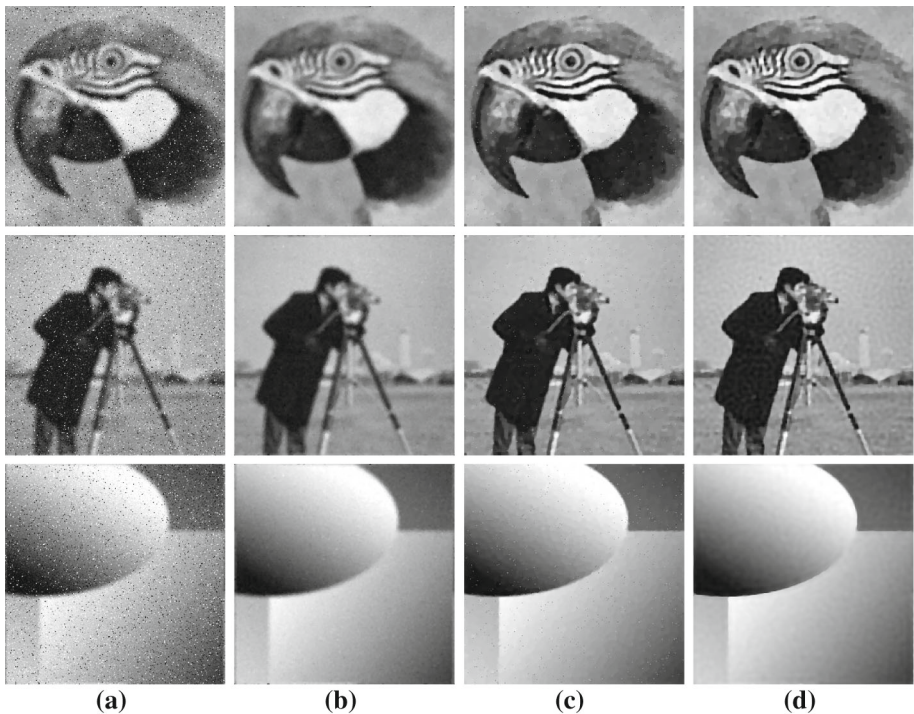
Fig. 9 Comparison of the restored results by applying different methods for deblurring and denoising the images degraded by a Gaussian blur ($G, 7, 3$) and Cauchy noise ($\gamma = 5$). **a** Degraded images; **b** the median filter; **c** the "conRe" model; **d** our method

From Tables 4 to 5, we find that our method provides the highest PSNRs and SSIMs. Comparing with the convex method, our method can improve by at least 0.36dB on PSNR. In Figs. 9 and 10, it is easy to see that the restored images by the median filter are oversmoothing as the median filter does not deblur. The convex method can recover edges and textures, but some noise is not removed. However, our method not only preserves the fine features but also effectively removes Cauchy noise, which can be clearly seen in the zoomed regions in Fig. 11.
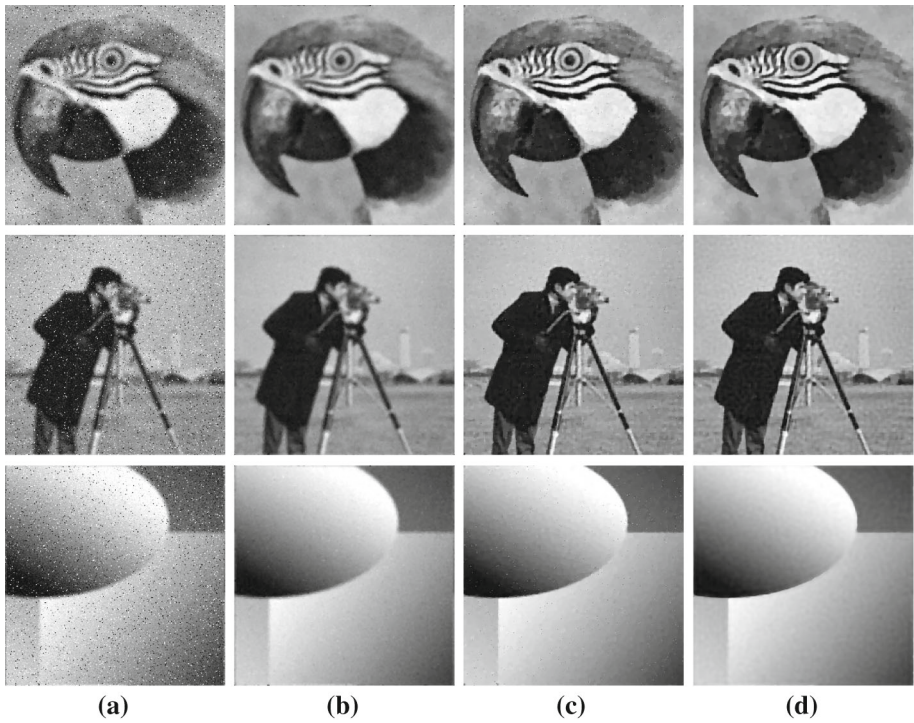
Fig. 10 Comparison of the restored results by applying different methods for deblurring and denoising the images degraded by the out-of-focus blur $(A, 5)$ and Cauchy noise $(\gamma = 5)$. **a** Degraded images; **b** the median filter; **c** the "conRe" model; **d** our method
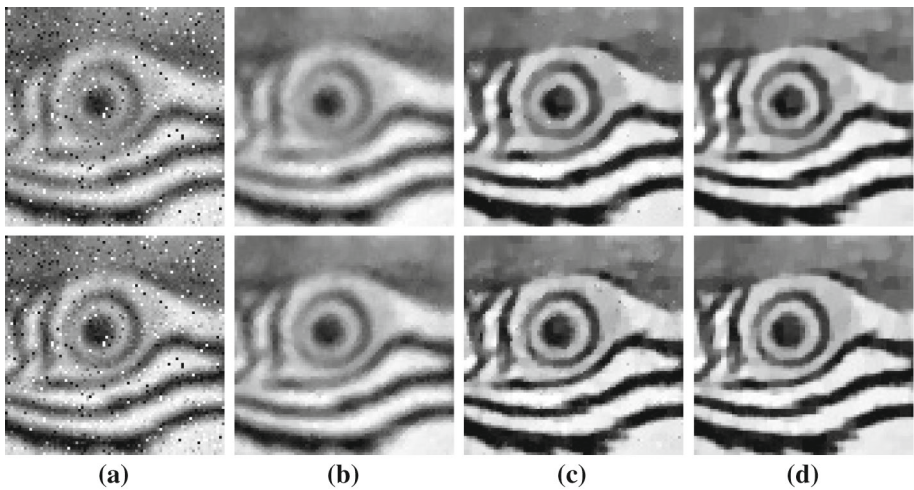

Fig. 11 Zoomed version of the restored results for the image "Parrot" degraded by the Gaussian blur (in the 1st row) and the out-of-focus blur (in the 2nd row), respectively. **a** Degraded images; **b** the median filter; **c** the "conRe" model; **d** our method

6 Conclusion

In this paper, we have reviewed and analyzed the statistic properties of the Cauchy distribution by comparing it with the Gaussian and Laplace distributions. Based on the MAP estimator, we have developed a nonconvex variational model for restoring images degraded by blurs and Cauchy noise. Taking advantage of a recent result in [48], the alternating direction method of multiplier (ADMM) algorithm is applied to solve the nonconvex variational optimization problem with a convergence guarantee. Numerical experiments show that the proposed method outperforms two well-known methods in both qualitative and quantitative comparisons.

References

1. Achim, A., Kuruoğlu, E.E.: Image denoising using bivariate α-stable distributions in the complex wavelet domain. IEEE Signal Process. Lett. **12**, 17–20 (2005)
2. Ambrosio, L., Fusco, N., Pallara, D.: Functions of Bounded Variation and Free Discontinuity Problems, vol. 254. Oxford University Press, Oxford (2000)
3. Arnold, B.C., Beaver, R.J.: The skew-Cauchy distribution. Stat. Prob. Lett. **49**, 285–290 (2000)
4. Attouch, H., Bolte, J., Redont, P., Soubeyran, A.: Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Lojasiewicz inequality. Math. Oper. Res. **35**, 438–457 (2010)
5. Attouch, H., Bolte, J., Svaiter, B.F.: Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized gauss-seidel methods. Math. Program. **137**, 91–129 (2013)
6. Balakrishnan, N., Nevzorov, V.B.: A Primer on Statistical Distributions. Wiley, New York (2004)
7. Bertsekas, D.P., Tsitsiklis, J.N.: Parallel and Distributed Computation: Numerical Methods. Prentice-Hall, Englewood Cliffs (1997)
8. Bolte, J., Daniilidis, A., Lewis, A.: The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. SIAM J. Opt. **17**, 1205–1223 (2007)
9. Cai, J.-F., Osher, S., Shen, Z.: Split Bregman methods and frame based image restoration. Multiscale Model. Simul. **8**, 337–369 (2009)
10. Chambolle, A.: An algorithm for total variation minimization and applications. J. Math. Imaging Vis. **20**, 89–97 (2004)
11. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. J. Math. Imaging Vis. **40**, 120–145 (2011)
12. Chan, R.H., Dong, Y., Hintermüller, M.: An efficient two-phase L^1-TV method for restoring blurred images with impulse noise. IEEE Trans. Image Process. **19**, 1731–1739 (2010)
13. Chang, Y., Kadaba, S., Doerschuk, P., Gelfand, S.: Image restoration using recursive Markov random field models driven by Cauchy distributed noise. IEEE Signal Process. Lett. **8**, 65–66 (2001)
14. Chen, C., Ng, M.K., Zhao, X.-L.: Alternating direction method of multipliers for nonlinear image restoration problems. IEEE Trans. Image Process. **24**, 33–43 (2015)
15. Copas, J.B.: On the unimodality of the likelihood for the Cauchy distribution. Biometrika **62**, 701–704 (1975)
16. Dong, Y., Hintermüller, M., Neri, M.: An efficient primal-dual method for L^1TV image restoration. SIAM J. Imag. Sci. **2**, 1168–1189 (2009)
17. Dong, Y., Zeng, T.: A convex variational model for restoring blurred images with multiplicative noise. SIAM J. Imag. Sci. **6**, 1598–1625 (2013)

18. Esser, E., Zhang, X., Chan, T.F.: A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. SIAM J. Imag. Sci. **3**, 1015–1046 (2010)
19. Feller, W.: An Introduction to Probability Theory and Its Applications, vol. 2. Wiley, New York (2008)
20. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Comput. Math. Appl. **2**, 17–40 (1976)
21. Giusti, E.: Minimal Surfaces and Functions of Bounded Variation. Springer Science & Business Media, Berlin (1984). no. 80
22. Goldfarb, D., Yin, W.: Parametric maximum flow algorithms for fast total variation minimization. SIAM J. Sci. Comput. **31**, 3712–3743 (2009)
23. Goldstein, T., Osher, S.: The split Bregman method for L1-regularized problems. SIAM J. Imag. Sci. **2**, 323–343 (2009)
24. Goncalves, M.L., Melo, J.G., Monteiro, R.D.: Improved pointwise iteration-complexity of a regularized ADMM and of a regularized non-Euclidean HPE framework. arXiv preprint arXiv:1601.01140, (2016)
25. Hintermüller, M., Stadler, G.: An infeasible primal-dual algorithm for total bounded variation-based inf-convolution-type image restoration. SIAM J. Sci. Comput. **28**, 1–23 (2006)
26. Idan, M., Speyer, J.L.: Cauchy estimation for linear scalar systems. IEEE Trans. Automat. Contr. **55**, 1329–1342 (2010)
27. Jeong, T., Woo, H., Yun, S.: Frame-based Poisson image restoration using a proximal linearized alternating direction method. Inverse Probl. **29**, 075007 (2013)
28. Johnson, N.L., Kotz, S., Balakrishnan, N.: Continuous Unltivariate Distributions, vol. 1. Wiley, New York (1994)
29. Kent, J.T., Tyler, D.E.: Maximum likelihood estimation for the wrapped Cauchy distribution. J. Appl. Stat. **15**, 247–254 (1988)
30. Knight, F.B.: A characterization of the Cauchy type. Proc. Am. Math. Soc. **55**, 130–135 (1976)
31. Kuruoğlu, E.E., Fitzgerald, W.J., Rayner, P.J.: Near optimal detection of signals in impulsive noise modeled with a symmetric alpha-stable distribution. IEEE Commun. Lett. **2**, 282–284 (1998)
32. Łojasiewicz, S.: Une propriété topologique des sous-ensembles analytiques réels. Equ. Derivees partielles, Paris 1962, Colloques internat. Centre nat. Rech. sci. 117, 87–89 (1963)
33. Łojasiewicz, S.: Sur la géométrie semi- et sous- analytique. Annales de l'institut Fourier **43**, 1575–1595 (1993)
34. Loza, A., Bull, D., Canagarajah, N., Achim, A.: Non-Gaussian model-based fusion of noisy images in the wavelet domain. Comput. Vis. Image Underst. **114**, 54–65 (2010)
35. Ma, L., Moisan, L., Yu, J., Zeng, T.: A dictionary learning approach for Poisson image deblurring. IEEE Trans. Med. Imaging **32**, 1277–1289 (2013)
36. Nikolova, M.: A variational approach to remove outliers and impulse noise. J. Math. Imaging Vis. **20**, 99–120 (2004)
37. Nolan, J.: Stable Distributions–Models for Heavy Tailed Data. Birkhäuser Boston, Cambridge, MA, To appear (Chapter 1 available online from http://academic2.american.edu/~jpnolan)
38. Nolan, J.P.: Numerical calculation of stable densities and distribution functions. Commun. Stat Stoch. Models **13**, 759–774 (1997)
39. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis, vol. 317. Springer Science & Business Media, Berlin (2009)
40. Samoradnitsky, G., Taqqu, M.S.: Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance, vol. 1. CRC Press, Boca Raton (1994)
41. Sawatzky, A., Brune, C., Kösters, T., Wübbeling, F., Burger, M.: EM-TV methods for inverse problems with Poisson noise. In: Level Set and PDE Based Reconstruction Methods in Imaging, Springer, pp. 71–142 (2013)
42. Sciacchitano, F., Dong, Y., Zeng, T.: Variational approach for restoring blurred images with Cauchy noise. SIAM J. Imaging Sci. **8**, 1894–1922 (2015)
43. Setzer, S.: Split Bregman algorithm, Douglas-Rachford splitting and frame shrinkage. In: Scale space and variational methods in computer vision, Springer, Berlin pp. 464–476 (2009)
44. Tsakalides, P., Nikias, C.L.: Deviation from normality in statistical signal processing: parameter estimation with alpha-stable distributions, A Practical Guide to Heavy Tails: Statistical Techniques and Applications, pp. 379–404 (1998)
45. Tsihrintzis, G.A.: Statistical modeling and receiver design for multi-user communication networks. A Practical Guide to Heavy Tails: Statistical Techniques and Applications (1998)
46. Wan, T., Canagarajah, N., Achim, A.: Segmentation of noisy colour images using Cauchy distribution in the complex wavelet domain. IET Image Process. **5**, 159–170 (2011)
47. Wang, F., Cao, W., Xu, Z.: Convergence of multi-block Bregman ADMM for nonconvex composite problems. arXiv preprint arXiv:1505.03063, (2015)

48. Wang, Y., Yin, W., Zeng, J.: Global convergence of ADMM in nonconvex nonsmooth optimization. arXiv preprint arXiv:1511.06324, (2015)
49. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**, 600–612 (2004)
50. Xu, Y., Huang, T.-Z., Liu, J., Lv, X.-G.: Split Bregman iteration algorithm for image deblurring using fourth-order total bounded variation regularization model. J. Appl. Math. 2013 (2013)
51. Xu, Y., Yin, W.: A block coordinate descent method for multi-convex optimization with applications to nonnegative tensor factorization and completion. tech. report, (2012)
52. Yang, J., Zhang, Y., Yin, W.: A fast alternating direction method for TVL1-L2 signal reconstruction from partial Fourier data. IEEE J. Sel. Top Signal Process. **4**, 288–297 (2010)
53. Yang, L., Pong, T.K., Chen, X.: Alternating direction method of multipliers for nonconvex background/foreground extraction. arXiv preprint arXiv:1506.07029, (2015)
54. Zhao, X.-L., Wang, F., Ng, M.K.: A new convex optimization model for multiplicative noise and blur removal. SIAM J. Imaging Sci. **7**, 456–475 (2014)