Acta Mathematica Scientia, 2022, **42B**(5): 1779–1808
https://doi.org/10.1007/s10473-022-0505-1

Acta Mathematica Scientia 数学物理学报
http://actams.apm.ac.cn

# A NON-LOCAL DIFFUSION EQUATION FOR NOISE REMOVAL*

*Jingfeng SHAO* (邵景峰)[1] *Zhichang GUO* (郭志昌)[1] *Wenjuan YAO* (姚文娟)[1†]
*Dong YAN* (严冬)[2] *Boying WU* (吴勃英)[1]
*1. School of Mathematics, Harbin Institute of Technology, Harbin 15000, China*
*2. School of Mathematics, University of California at Irvine, Irvine 92697, U.S.A.*
*E-mail*: *sjfmath@foxmail.com; mathgzc@hit.edu.cn; mathywj@hit.edu.cn; dyan6@uci.edu; mathwby@hit.edu.cn*

**Abstract** In this paper, we propose a new non-local diffusion equation for noise removal, which is derived from the classical Perona-Malik equation (PM equation) and the regularized PM equation. Using the convolution of the image gradient and the gradient, we propose a new diffusion coefficient. Due to the use of the convolution, the diffusion coefficient is non-local. However, the solution of the new diffusion equation may be discontinuous and belong to the bounded variation space (BV space). By virtue of Young measure method, the existence of a BV solution to the new non-local diffusion equation is established. Experimental results illustrate that the new method has some non-local performance and performs better than the original PM and other methods.

**Key words** image denoising; non-local diffusion; BV solutions; Perona-Malik method

**2010 MR Subject Classification** 35K59; 68U10

## 1 Introduction

Images are inevitably affected by noise in the process of acquisition and transmission, so that image denoising is an important task in image processing. The main difficulty in image denoising is to preserve the details of information such as the edges and textures of the original image from the noisy observation. Most of the time, those details of information are of vital importance in subsequent image processing tasks, such as image segmentation and image restoration.

Research on image denoising based on partial differential equations (PDE) has developed extensively over the course of the last thirty years. Osher, in [1], proposed a pioneering denoising model by a total variation (TV) method. Later, Vese, in [2], summarized a family of denoising methods based on a variational PDE, then more in-depth studies on its theoretical background

*Received March 26, 2021; revised May 18, 2022. This work was partially supported by the National Natural Science Foundation of China (11971131, 12171123, 11871133, 11671111, U1637208, 61873071, 51476047), the Guangdong Basic and Applied Basic Research Foundation (2020B1515310006), and the Natural Sciences Foundation of Heilongjiang Province (LH2021A011) and China Postdoctoral Science Foundation (2020M670893).
†Corresponding author: Wenjuan YAO.

were proposed in [3] and [4], and the relevant numerical methods were proposed in [5–8]. After that, denoising methods based on second-order equations were generalized to other types of PDEs, such as fourth-order PDEs [9–11] and fractional PDEs [12–14].

Perona and Malik proposed an anisotropic diffusion equation [15] (the PM method) as follows:

$$u_t - \operatorname{div}\left(g\left(|\nabla u|\right)\nabla u\right) = 0.$$

Here $g(s)$ is a diffusivity function with the form

$$g(s) = \frac{1}{1+s^2/K^2}, \quad K > 0. \tag{1.1}$$

The anisotropic diffusion enables the PM equation to detect the edge area of a noisy image and implement an edge preserved diffusion. Moreover, the diffusion approach of the PM equation was proved to be forward-backward, i.e., the edge area can not only be preserved, but also be enhanced. However, this edge enhancement also leads to artificial phenomena; new features are added to the restored images so that it no longer preserves the original information. As a result, there will appear the "staircase" and "speckle" effects in the restored images. A regularized PM equation (RPM [16]) was proposed to avoid forward-backward diffusion in the original PM equation by refining the diffusivity function from $g\left(|\nabla u|\right)$ to $g\left(|\nabla G_\sigma * u|\right)$, where $G_\sigma$ is a Gaussian function

$$G_\sigma(x) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\exp\left(-\frac{|x|^2}{2\sigma^2}\right), \quad \sigma > 0. \tag{1.2}$$

With Gaussian convolution, the diffusivity becomes totally forward. Furthermore, the RPM method becomes a well-posed problem, and the solution of the RPM equation is smooth, which effectively avoids the "staircase" effect of the PM method. Unfortunately, this smooth solution sometimes will result in blurry edges in the restored images.

Guidotti, in [17–19], regularized the diffusivity $g(|\nabla u|)$ in the PM equation by fractional derivatives $g(\left|\nabla^{1-\epsilon}u\right|)$, and proved in a one dimensional situation that the new equation could be changed from a PM equation to a linear heat equation, as $\epsilon$ goes from 0 to 1. Later, Guidotti, in [20], proposed the "milder" regularized PM method by modifying the diffusivity function of the PM equation into

$$g(|\nabla u|) = \frac{1}{1+|\nabla u|^2} + \delta;$$

the "milder" regularization preserves the forward and backward diffusion property of the PM method but enables the equation to remove the "staircase" and "speckle" effects. Furthermore, in [21], the authors extended this"milder" regularization diffusivity to

$$g(|\nabla u|) = \frac{1}{1+|\nabla u|^2} + \delta|\nabla u|^{p-2}, \tag{1.3}$$

where $p \in [1, +\infty)$.

In this paper, we consider a new model for noise removal:

$$\begin{cases} \dfrac{\partial u}{\partial t} = \operatorname{div}\left(\dfrac{\nabla u}{1+|\nabla G_\sigma * u|\,|\nabla u|}\right), & \text{in } \Omega_T, \quad (1.4)\\ u(x,t) = 0, & \text{on } \Gamma, \quad (1.5)\\ u(x,0) = u_0(x), & \text{in } \Omega. \quad (1.6) \end{cases}$$

Define $\Omega_T := \Omega \times (0,T)$, where $\Omega \subset \mathbb{R}^N$ is a bounded Lipschitz domain with boundary $\partial\Omega$, and $\Gamma := \partial\Omega \times (0,T)$. $u_0(x)$ is the initial noisy image. The diffusivity function contains both the gradient term and the convolutional gradient term, which combines the advantages of PM and RPM simultaneously. It is worth noticing that our new model is not just a trivial combination of two existing models; the properties of the solution of our model are totally different from that of the others and we will discuss this in depth in the ensuing sections. Because of the non-local term $|\nabla G_\sigma * u|$, the new diffusion at every point depends on a neighborhood of this point.

In terms of mathematical theory, we investigate the first initial-boundary value problem of diffusion equations in more general form as follows:

$$\begin{cases} u_t = \operatorname{div}\left(\dfrac{\nabla u}{1+\alpha(u)|\nabla u|}\right) & \text{in } \Omega_T, \quad (1.7)\\ u(x,t) = 0 & \text{on } \Gamma, \quad (1.8)\\ u(x,0) = u_0(x) & \text{in } \Omega; \quad (1.9) \end{cases}$$

this will be denoted by $\mathcal{P}$. Here $u_0(x) \in \mathrm{BV}(\Omega)$ with a zero trace on $\partial\Omega$. We also have that $\alpha(u) = a(|\overrightarrow{f} * \tilde{u}|)$, $\overrightarrow{f} = (f_1, f_2, \cdots, f_n)$, $f_i \in \mathcal{S}(\mathbb{R}^N)$, and by $|\cdot|$ we denote the Euclidean norm and we define $\mathcal{S}(\mathbb{R}^N)$ to be the Schwartz space. Given $u \in L^2(\Omega)$, we extend this by zero to the whole of $\mathbb{R}^N$ and denote $\tilde{u}$ as the extended function. $a(\cdot) \in C^2[0,\infty)$ and $a(x) \geq 0$. Let $f * \tilde{u}$ be the convolution of $f$ and $\tilde{u}$ in $\mathbb{R}^N$. In particular, for the case $\alpha(u) = |\nabla G_\sigma * u|$, (1.7)–(1.9) become (1.4)–(1.6). The loss of growth, however, brings some technical difficulties when $\alpha(u)$ arrives at zero. We prove the existence of the solution to $\mathcal{P}$ for the case in which $\alpha(u) \geq c > 0$, in theory. This means that $\alpha(u) = |\nabla G_\sigma * u| + \varepsilon$ is allowed, for a small $\varepsilon > 0$. The uniqueness is proved in a narrower class of functions than the existence.

The rest of this paper is organized as follows: in Section 2, we propose the non-local PM method. Section 3, Section 4 and Section 5 are the mathematical theory parts of the paper. We first state the mathematical preliminaries in Section 3. The existence of solutions to problem $\mathcal{P}$ is proved in Section 4. In Section 5, we investigate the uniqueness of solutions to problem $\mathcal{P}$. Finally, Section 6 verifies the effectiveness and superiority of our proposed method by comparative numerical experiments.

## 2 Non-local PM Denoising Method

In this section, comparing our new model with the PM and RPM models, we discuss how our new model works and the non-triviality of our modification.

As we know, PM diffusion is forward-backward, and this creates is an ill-posed problem. An image restored by the PM method may have black and white speckles and false edges because of the backward diffusion. RPM diffusion is regularized and well-posed, and produces smoothing solutions, but this regularization property may destroy the edges in the restored image. Combining the advantages of PM and RPM, and the corresponding diffusivities,

$$g_{\mathrm{PM}} = \frac{1}{1+|\nabla u|^2},$$
$$g_{\mathrm{RPM}} = \frac{1}{1+|\nabla G_\sigma * u|^2},$$

we propose the diffusivity function

$$g = \frac{1}{1 + |\nabla G_\sigma * u|\,|\nabla u|}. \tag{2.1}$$

PM diffusivity only involves point-wise operation, and thus the PM equation is a local method. RPM is non-local because of the convolution. Our new model combines the two features so that it is also non-local. We name our new model the "non-local PM" (NLPM) method.

The non-triviality of our model is mainly reflected in the properties of the solutions. The solution of the PM equation may not exist, or exist but not be unique [22]. The solution space of the RPM equation is the $C^\infty$ space [16]. However, some researchers regard the BV space as the right space for image processing [1, 31]. With our new diffusivity (2.1), the NLPM model is proved, in the next section, to have a BV solution, and so is similar to the TV model [1].

Next, a prototypical example is used to illustrate the edge-preserving ability of our NLPM method. Formally, one can regard the edge of an image as a function with a large gradient norm, observing that the diffusivity of the NLPM method consists of $|\nabla u|^{-1}$, so the NLPM method should undertake no diffusion on the edge area of an image, i.e., the edge will be preserved by the NLPM method. Although this is not a strict mathematical analysis, numerical experiments enable to validate this conclusion. Now we consider that

$$f(x) = \begin{cases} 5, \, x > 0, \\ 0, \, \text{otherwise}. \end{cases}$$

This function has a jump discontinuity at $x = 0$, which can be regarded as an edge of the one-dimensional signal $f$. Also, we consider the mollified signals $f_\sigma$ by the Gaussian kernels (1.2) with a different parameter $\sigma$, as shown by the red lines in Figure 1. Then we use the NLPM method to process these signals; the output signals are shown as the black circled lines. One can observe clearly that the NLPM method preserves the edge very well.

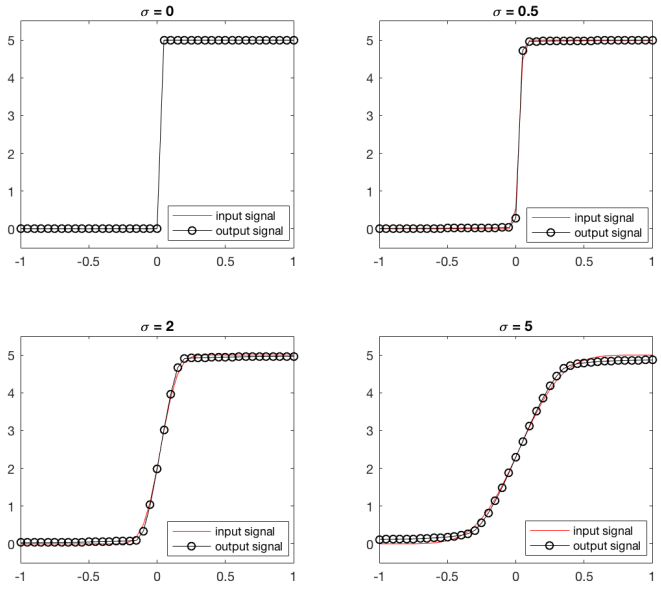



Figure 1 The NLPM method enables one to maintain the jump discontinuity

Another prototypical experiment on one-dimensional signal processing demonstrates, intuitively, the differences between the PM method, the RPM method and the NLPM method. Consider the one-dimensional forms of the three respective methods as follows:

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left(\frac{u_x}{1+|u_x|^2}\right),$$
$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left(\frac{u_x}{1+|(G_\sigma * u)_x|^2}\right),$$
$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left(\frac{u_x}{1+|(G_\sigma * u)_x|\,|u_x|}\right).$$

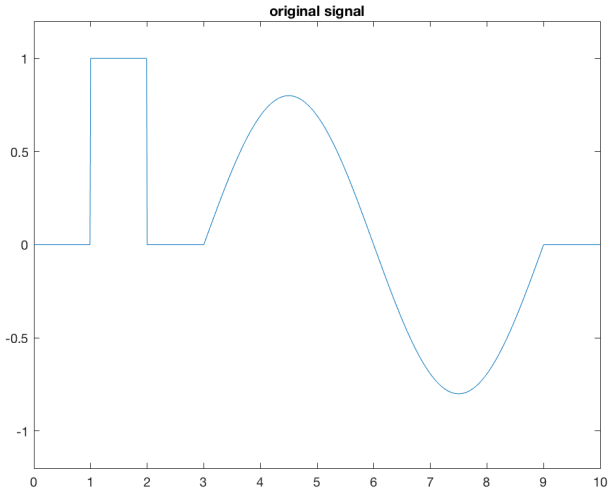



Figure 2 Synthetic one-dimensional signal

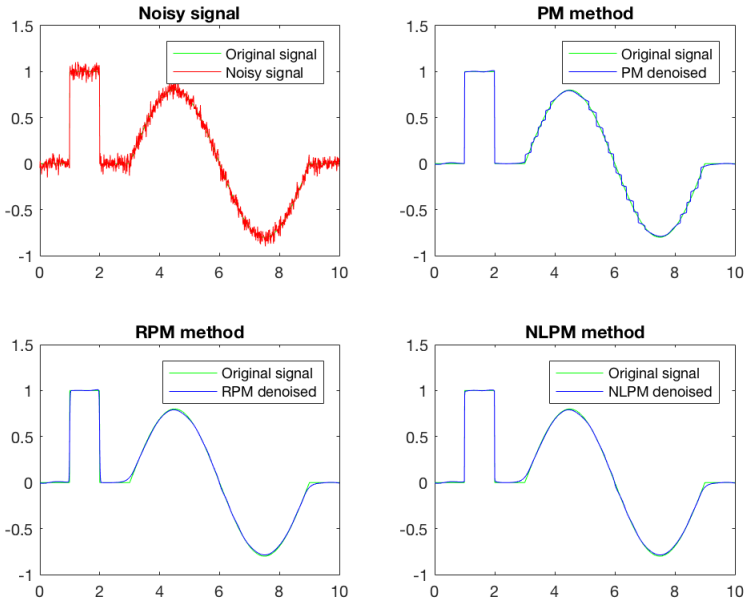



Figure 3 The PM method has a "staircase" effect

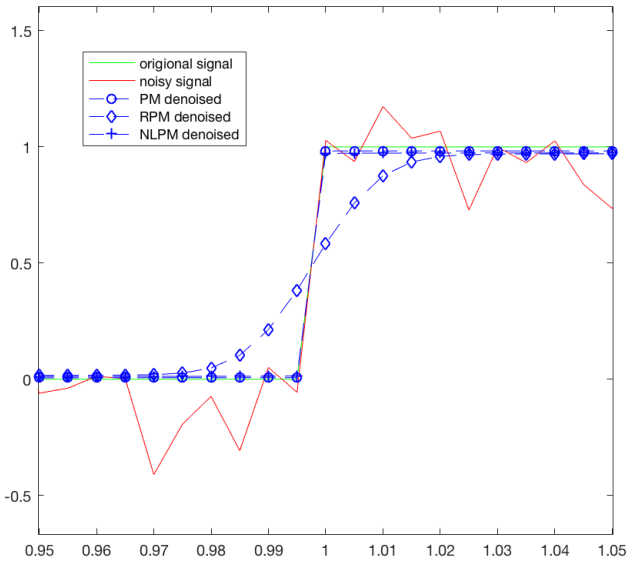



Figure 4 The RPM method mollifies the edge

We use these equations to process a synthetic one-dimensional signal, as shown in Figure 2. Notice that this example consists of various cases such as jump discontinuities and the smooth periodic signal. Figure 3 shows that the PM method has a severe "staircase" effect in smooth intensity transition areas (i.e., the trigonometric function part of the synthetic signal), while the RPM and NLPM methods maintain the smoothness well. Furthermore, we zoom on the discontinuous region of the signal (i.e., the staircase function part) and put the processing results together in Figure 4. One can easily see that the RPM method mollifies the edge while the other two methods preserve the edge well. Noticing that the x-ticks in Figure 4 are close to the jump discontinuity $x = 1$, it is natural that this phenomenon is not quite as obvious in the original figures as in Figure 3.

In addition, the parameter $\sigma$ in the NLPM diffusivity function (2.1) also increases the flexibility of the model. The parameter $\sigma$ is related to the window size of the Gaussian convolution, which shows the non-locality of the model. When $\sigma$ approaches zero, the non-locality of the model is minimized and the equation degenerates to the original PM equation. When $\sigma$ approaches $\infty$, the model is approximate to the Laplacian equation.

## 3 Mathematical Preliminaries

### 3.1 Pairings between Measures and Bounded Functions

Denote $\mathcal{M}(X)$ as the set of finite Radon measures on a Borel set $X$, and denote $M_1^+(X)$ as its subset of probability measures. As usual, $\mathcal{L}^N$ and $\mathcal{H}^{N-1}$ represent the $N$-dimensional Lebesgue measure and $(N-1)$-dimensional Hausdorff measure in $\mathbb{R}^N$, respectively. We denote $\mathcal{M}_+(X)$ as the positive Radon measures. For all $\lambda_1, \lambda_2 \in \mathcal{M}(X)$,

$$\lambda_1 \geqslant \lambda_2 \ \text{ in } \ \mathcal{M}(X) \Leftrightarrow \lambda_1 - \lambda_2 \in \mathcal{M}_+(X).$$

Denote $C_0(\mathbb{R}^N)$ as the closure of continuous functions on $\mathbb{R}^N$ with compact support. The

dual of $C_0(\mathbb{R}^N)$ can be identified with the Radon measure space $\mathcal{M}(\mathbb{R}^N)$ via the pairing

$$\langle \nu, f\rangle = \int_{\mathbb{R}^N} f \mathrm{d}\nu, \ \ \forall \nu \in \mathcal{M}(\mathbb{R}^N), f \in C_0(\mathbb{R}^N).$$

Let the set $D \subset R^n$ be a measurable set with finite measure. A map $\nu : D \to \mathcal{M}(\mathbb{R}^N)$ is said to be weakly* measurable if the functions $x \mapsto \int_{\mathbb{R}^N} f \mathrm{d}\nu_x$ are measurable for all $f \in C_0(\mathbb{R}^N)$, where $\nu_x = \nu(x)$.

Throughout the rest of this paper, $I$ represents the identity operator in $\mathbb{R}^N$.

A function $u \in L^1(\Omega)$ is called a function of bounded variation if its distributional derivative $Du$ is a $\mathbb{R}^N$-valued Radon measure with finite total variation in $\Omega$. The vector space of functions of bounded variation in $\Omega$ is denoted by $\mathrm{BV}(\Omega)$. The space $\mathrm{BV}(\Omega)$ is a non-reflexive Banach space under the norm $\|u\|_{BV} := \|u\|_{L^1} + |Du|(\Omega)$, where $|Du|$ is the total variation measure of $Du$.

For $u \in \mathrm{BV}(\Omega)$, the gradient $Du$ is a Radon measure that can be decomposed into its absolutely continuous and singular parts

$$Du = D^a u + D^s u.$$

Then, $D^a u = \nabla u \mathcal{L}^N$, where $\nabla u$ is the Radon-Nikodým derivative of the measure $Du$ with respect to the Lebesgue measure $\mathcal{L}^N$. There is also the polar decomposition $D^s u = \frac{D^s u}{|D^s u|} |D^s u|$, where $\frac{D^s u}{|D^s u|} \in L^1\left(\Omega, |D^s u|; \partial\mathbb{B}^N\right)$ is the Radon-Nikodým derivative of $D^s u$ with respect to its total variation measure $|D^s u|$.

Set that $\tilde{\Omega}$ is a bounded smooth domain and that is satisfies $\Omega \subset\subset \tilde{\Omega}$. For $v \in \mathrm{BV}(\Omega)$, we denote $\tilde{v}$ by

$$\tilde{v}(x) := \begin{cases} v(x), & x \in \Omega, \\ 0, & x \in \tilde{\Omega} \backslash \overline{\Omega}. \end{cases}$$

**Lemma 3.1** Assume that $u \in \mathrm{BV}(\Omega)$. There exists a sequence of functions $u_i \in W^{1,1}(\Omega) \bigcap C^\infty(\Omega)$ such that

(i) $u_i \to u$ in $L^1(\Omega)$;

(ii) $|Du_i|(\Omega) \to |Du|(\Omega)$;

(iii) $u_i|_{\partial\Omega} = u|_{\partial\Omega}$ for all $i$.

Moreover,

(iv) if $u \in \mathrm{BV}(\Omega) \bigcap L^q(\Omega)$, $q < \infty$, we can find functions $u_i$ such that $u_i \in L^q(\Omega)$ and $u_i \to u$ in $L^q(\Omega)$;

(v) if $u \in \mathrm{BV}(\Omega) \bigcap L^\infty(\Omega)$, we can find $u_i$ such that $\|u_i\|_\infty \le \|u\|_\infty$ and $u_i \to u$ weakly* in $L^\infty(\Omega)$.

It is well known that the summability conditions on the divergence of a vector field $z$ in $\Omega$ yield trace properties for the normal component of $z$ on $\partial\Omega$. As in [23], we define a function $[z, \vec{n}] \in L^\infty(\partial\Omega)$, which is associated to any vector field $z \in L^\infty(\Omega, \mathbb{R}^N)$ such that $\mathrm{div}(z)$ is a bounded measure in $\Omega$.

Assume that $\Omega$ is an open bounded set in $\mathbb{R}^N$ with $\partial\Omega$ Lipschitz, $N \ge 2$, and $1 \le p \le N$, $\frac{N}{N-1} \le q \le \infty$. Since $\partial\Omega$ is Lipschitz, the outer unit normal $\vec{n}$ exists in $\mathcal{H}^N$ a.e. on $\partial\Omega$. We shall consider the following spaces:

$$\mathrm{BV}(\Omega)_q := \mathrm{BV}(\Omega) \bigcap L^q(\Omega),$$

$$\mathrm{BV}(\Omega)_c := \mathrm{BV}(\Omega)\bigcap L^\infty(\Omega)\bigcap C(\Omega),$$
$$X(\Omega)_p := \{z \in L^\infty(\Omega,\mathbb{R}^N) : \mathrm{div}(z) \in L^p(\Omega)\},$$
$$X(\Omega)_\mu := \{z \in L^\infty(\Omega,\mathbb{R}^N) : \mathrm{div}(z) \in \mathcal{M}(\Omega)\}.$$

**Lemma 3.2** Assume that $\Omega \subseteq \mathbb{R}^N$ is a open bounded set with Lipschitz boundary $\partial\Omega$. Then there exists a bilinear map $\langle z,u\rangle_{\partial\Omega} : X(\Omega)_\mu \times \mathrm{BV}(\Omega)_c \to \mathbb{R}$ such that

$$\langle a,u\rangle_{\partial\Omega} = \int_{\partial\Omega} u(x)z(x)\cdot\vec{n}\,\mathrm{d}\mathcal{H}^{\mathcal{N}-\infty} \ \text{ if } z \in C^1(\overline{\Omega},\mathbb{R}^N), \tag{3.1}$$
$$|\langle a,u\rangle_{\partial\Omega}| \le \|z\|_\infty \int_{\partial\Omega} |u(x)|\mathrm{d}\mathcal{H}^{\mathcal{N}-\infty}. \tag{3.2}$$

**Lemma 3.3** Let $\Omega$ be as in Lemma 3.2. Then there exists a linear operator $\gamma : X(\Omega)_\mu \to L^\infty(\partial\Omega)$ such that

$$\|\gamma\|_\infty \le \|z\|_\infty, \tag{3.3}$$
$$\langle a,u\rangle_{\partial\Omega} = \int_{\partial\Omega} u(x)\gamma(z)(x)\cdot\vec{n}\,\mathrm{d}\mathcal{H}^{\mathcal{N}-\infty} \quad \text{for all } u \in \mathrm{BV}(\Omega)_c, \tag{3.4}$$
$$\gamma(z)(x) = z(x)\cdot\vec{n} \quad \text{for all } x \in \partial\Omega \ \text{ if } \ z \in C^1(\overline{\Omega},\mathbb{R}^N). \tag{3.5}$$

The function $\gamma(z)$ is a weakly defined trace on $\partial\Omega$ of the normal component of $z$. We denote $\gamma(z)$ by $[z,\vec{n}]$.

In the sequel, we shall consider pairs $(z,u)$ such that one of the following conditions holds:

$$\begin{cases} \text{a)} \quad z \in X(\Omega)_p, u \in \mathrm{BV}(\Omega)_{p'} \ \text{ and } 1 < p \le N; \\ \text{b)} \quad z \in X(\Omega)_1, u \in \mathrm{BV}(\Omega)_\infty; \\ \text{c)} \quad z \in X(\Omega)_\mu, u \in \mathrm{BV}(\Omega)_c. \end{cases} \tag{3.6}$$

**Definition 3.4** Let $z$ and $u$ be two functions such that one of the conditions of (3.6) holds. Then we define a functional $(z,Du) : \mathcal{D}(\Omega) \to \mathbb{R}$ as

$$\langle (z,Du),\phi\rangle := -\int_\Omega u(x)\phi\,\mathrm{div}(z)\mathrm{d}x - \int_\Omega u(x)z\cdot\nabla\phi\mathrm{d}x, \ \ \phi \in \mathcal{D}(\Omega).$$

**Lemma 3.5** For all open sets $V \subseteq \Omega$ and for all functions $\phi \in \mathcal{D}(V)$, one has that

$$|\langle (z,Du),\phi\rangle| \le \sup\|\phi\|_\infty \|z\|_{L^\infty(V)}|Du|(V), \tag{3.7}$$

and thus, $(z,Du)$ is a Radon measure in $\Omega$.

We give the Green's formula relating the function $[z,\vec{n}]$ and the measure $(z,Du)$.

**Lemma 3.6** Let $\Omega$ be as in Lemma (3.2). If $z$ and $u$ are two functions such that one of the conditions of (3.6) holds, then we have that

$$\int_\Omega u\,\mathrm{div}(z)\mathrm{d}x + \int_\Omega (z,Du) = \int_{\partial\Omega} [z,\vec{n}]\,u\,\mathrm{d}\mathcal{H}^{N-1}. \tag{3.8}$$

More details about the pairings $\langle z,u\rangle_{\partial\Omega}$ and $(z,Du)$ can be found in [23].

### 3.2 The Generalized Young Measure

This subsection gives a brief overview of the basic theory of generalized Young measures and recalls results that will be used later. We refer to [24] for further information about the generalized Young measures.

First, we need a suitable class of integrands. Let $\mathcal{E}(\Omega;\mathbb{R}^N)$ be the set of all $f \in C(\overline{\Omega}\times\mathbb{R}^N)$ such that

$$(Tf)(x,\hat{A}) := (1-|\hat{A}|)f\left(x,\frac{\hat{A}}{1-|\hat{A}|}\right),\ \ x\in\overline{\Omega},\ \ \hat{A}\in\mathbb{B}^N \tag{3.9}$$

extends into a continuous function $Tf \in (\overline{\Omega\times\mathbb{B}^N})$, where $\mathbb{B}^N$ stands for the unit sphere in $\mathbb{R}^N$. In particular, this implies that $f$ has linear growth at infinity, i.e., there exists a constant $M>0$ (in fact, $M := \|f\|_{E(\Omega;\mathbb{R}^N)} := \|Tf\|_{\infty,\overline{\Omega\times\mathbb{B}^N}}$ ) such that

$$|f(x,A)| \leqslant M(1+|A|) \quad \text{for all } (x,A)\in\overline{\Omega}\times\mathbb{R}^N.$$

For all $f\in\mathcal{E}(\Omega;\mathbb{R}^N)$, the recession function

$$f^\infty(x,A) := \lim_{\substack{A'\to A\\ x'\to x\\ t\to\infty}} \frac{f(x',tA')}{t},\ \ x\in\overline{\Omega},\ A\in\mathbb{R}^N \tag{3.10}$$

exists as a continuous function. Sometimes this notion of a recession function is too strong, so for any function $g\in C(\mathbb{R}^N)$ with linear growth at infinity, we define the generalized recession function as

$$g^\sharp(A) := \lim_{\substack{A'\to A\\ t\to\infty}} \frac{g(tA')}{t},\ \ A\in\mathbb{R}^N. \tag{3.11}$$

**Definition 3.7** A generalized Young measure with target space $\mathbb{R}^N$ is a triple $\lambda = (\nu_x,\lambda_\nu,\nu_x^\infty)$ comprising

(i) a parametrized family of probability measures $(\nu_x)_{x\in\Omega}\subseteq\mathcal{M}_1^+(\mathbb{R}^N)$;

(ii) a positive finite measure $\lambda_\nu\in\mathcal{M}^+(\overline{\Omega})$;

(iii) a parametrized family of probability measures $(\nu_x^\infty)_{x\in\overline{\Omega}}\subseteq\mathcal{M}_1^+(\partial\mathbb{B}^N)$ ($\partial\mathbb{B}^N$ denotes the unit sphere in $\mathbb{R}^N$);

(iv) a map $x\mapsto\nu_x$ is weakly* measurable with respect to $\mathcal{L}^N$, i.e., the function $x\mapsto\langle\nu_x,f(x,\cdot)\rangle$ is $\mathcal{L}^N$-measurable for all bounded Borel functions $f:\Omega\times\mathbb{R}^N\to\mathbb{R}$;

(v) a map $x\mapsto\nu_x^\infty$ is weakly* measurable with respect to $\lambda_\nu$;

(vi) $x\mapsto\langle\nu_x,|\cdot|\rangle\in L^1(\Omega)$.

By $Y(\Omega;\mathbb{R}^N)$, we denote the set of all such generalized Young measures. The parametrized measure $(\nu_x)$ is called the oscillation measure, the measure $\lambda_\nu$ is the concentration measure, and $(\nu_x^\infty)$ is the concentration-angle measure.

The duality pairing $\langle\langle\lambda,f\rangle\rangle$ for $\lambda\in Y(\Omega;\mathbb{R}^N)$ and $f\in(\Omega;\mathbb{R}^N)$ is defined by

$$\begin{aligned}\langle\langle\lambda,f\rangle\rangle &:= \int_\Omega\langle\nu_x,f(x,\cdot)\rangle\,\mathrm{d}x + \int_{\overline{\Omega}}\langle\nu_x^\infty,f^\infty(x,\cdot)\rangle\,\mathrm{d}\lambda_\nu(x)\\ &= \int_\Omega\int_{\mathbb{R}^N} f(x,A)\mathrm{d}\nu_x(A)\mathrm{d}x + \int_{\overline{\Omega}}\int_{\partial\mathbb{R}^N} f^\infty(x,\cdot)\mathrm{d}\nu_x^\infty(A)\mathrm{d}\lambda_\nu(x).\end{aligned}$$

The space $Y(\Omega;\mathbb{R}^N)$ of Young measures can be considered as a part of the dual space $\mathcal{E}(\Omega;\mathbb{R}^N)^*$ (the inclusion is strict since, for instance, $f\mapsto s\int_\Omega f(x,\cdot)\mathrm{d}x$ lies in $\mathcal{E}(\Omega;\mathbb{R}^N)^*\setminus Y(\Omega;\mathbb{R}^N)$ whenever $s\neq 1$). This embedding gives rise to a weak* topology on $Y(\Omega;\mathbb{R}^N)$ and so we say that $(\lambda_j)\subseteq Y(\Omega;\mathbb{R}^N)$ (where $\lambda_j := (\nu_x^j,\lambda_{\nu_j},(\nu_x^\infty)_j)$) weakly* converges to $\lambda\in Y(\Omega;\mathbb{R}^N)$, in symbols, $\lambda_j\overset{*}{\rightharpoonup}\lambda$, if $\langle\langle\lambda_j,f\rangle\rangle\to\langle\langle\lambda,f\rangle\rangle$ for all $f\in\mathcal{E}(\Omega;\mathbb{R}^N)$. The set $Y(\Omega;\mathbb{R}^N)$ is topologically weakly*-closed in $\mathcal{E}(\Omega;\mathbb{R}^N)^*$.

The main compactness result in the space $Y(\Omega;\mathbb{R}^N)$ is listed as follows:

**Lemma 3.8** Let $(\lambda_i) \subseteq Y(\Omega;\mathbb{R}^N)$ be a sequence such that

(i) the functions $x \mapsto \langle (\nu_j)_x, |\cdot| \rangle$ are uniformly bounded in $L^1(\Omega)$;

(ii) the sequence $(\lambda_{\nu_j}(\overline{\Omega}))$ is uniformly bounded.

Then, $(\lambda_j)$ is weakly$^*$ sequentially relatively compact in $Y(\Omega;\mathbb{R}^N)$, i.e., there exist $\lambda \in Y(\Omega;\mathbb{R}^N)$ and a subsequence of $(\lambda_j)$ (not relabeled) such that $\lambda_j \overset{*}{\rightharpoonup} \lambda$.

Next, we define the set $GY(\Omega;\mathbb{R}^N)$ of generalized gradient Young measures as the collection of the generalized Young measures $\lambda \in Y(\Omega;\mathbb{R}^N)$ with the property that there exists a norm-bounded sequence $(u_j) \subseteq \mathrm{BV}(\Omega)$ such that the sequence $(Du_j)$ generates $\lambda$, which, in symbols, we will write as $Du_j \overset{Y}{\to} \lambda$, meaning that

$$\int_\Omega f(x, \nabla u_j(x))\mathrm{d}x + \int_{\overline{\Omega}} f^\infty(x, \frac{D^s u_j}{|D^s u_j|})\mathrm{d}\,|D^s u_j| \to \langle\langle \lambda, f \rangle\rangle$$

for all $f \in \mathcal{E}(\Omega;\mathbb{R}^N)$.

**Lemma 3.9** Let $\lambda \in Y(\Omega;\mathbb{R}^N)$ be a generalized Young measure with $\lambda_\nu(\partial\Omega) = 0$. Then $\lambda \in GY(\Omega;\mathbb{R}^N)$, if and only if there exists $u \in \mathrm{BV}(\Omega)$ with

$$Du = \langle \nu, I \rangle\, \mathcal{L}^N \llcorner \Omega + \langle \nu^\infty, I \rangle\, \lambda_\nu \llcorner \Omega \ \text{ in } \ \mathcal{M}(\Omega),$$

and for all quasiconvex $g \in C(\mathbb{R}^N)$ with linear growth at infinity, the following Jensen-type inequalities hold:

(i)

$$g\left(\langle \nu, I \rangle + \langle \nu^\infty, I \rangle \frac{\mathrm{d}\lambda_\nu}{\mathrm{d}\mathcal{L}^\mathcal{N}}\right) \leqslant \langle \nu, g \rangle + \left\langle \nu^\infty, g^\sharp \right\rangle \frac{\mathrm{d}\lambda_\nu}{\mathrm{d}\mathcal{L}^\mathcal{N}} \quad \text{for } \mathcal{L}^N\text{-a.e. } x \in \Omega,$$

where $\frac{\mathrm{d}\lambda_\nu}{\mathrm{d}\mathcal{L}^\mathcal{N}}$ is the Radon-Nikodým derivative of the measure $\lambda$ with respect to the Lebesgue measure $\mathcal{L}^N$;

(ii)

$$g^\sharp\left(\langle \nu^\infty, I \rangle\right) \leqslant \left\langle \nu^\infty, g^\sharp \right\rangle \quad \text{for } \lambda_\nu^s\text{-a.e. } x \in \Omega,$$

where $\lambda_\nu^s$ is the singular part of $\lambda$ with respect to the Lebesgue measure $\mathcal{L}^N$.

We remark that for both flavors of recession function, one can drop the additional sequence $A' \to A$ if the functional is Lipschitz continuous (see [25]).

Given $u \in \mathrm{BV}(\Omega)$, denote that $\sigma_{Du} := (\delta_{\nabla u}, |D^s u|, \delta_p) \in GY(\Omega;\mathbb{R}^N)$ and $p := \frac{D^s u}{|D^s u|} \in L^1\left(\Omega, |D^s u|\,; \partial\mathbb{B}^N\right)$. $\delta_A$ is denoted as the Dirac measure on $\mathbb{R}^N$ giving unit mass to the point $A \in \mathbb{R}^N$.

Throughout the rest of this paper, $u^\Omega$ represents the trace of $u \in \mathrm{BV}(\Omega)$ on $\partial\Omega$.

Referring to [26], Corollary 4, we have the following necessary lemma:

**Lemma 3.10** Assume $X \subset B \subset Y$ with compact imbedding $X \subset B$ ($X, B$ and $Y$ are Banach spaces).

i) Let $F$ be bounded in $L^p(0,T;X)$, where $1 \leq p < \infty$, and let $\frac{\partial F}{\partial t} = \left\{\frac{\partial f}{\partial t}; f \in F\right\}$ be bounded in $L^1(0,T;Y)$. Then $F$ is relatively compact in $L^p(0,T;B)$;

ii) Let $F$ be bounded in $L^\infty(0,T;X)$, where $1 \leq p < \infty$, and let $\frac{\partial F}{\partial t} = \left\{\frac{\partial f}{\partial t}; f \in F\right\}$ be bounded in $L^r(0,T;Y)$ where $r > 1$. Then $F$ is relatively compact in $C(0,T;B)$.

## 4 Existence

This section is divided into four subsections. In Subsection 4.1, we list the main results. In Subsection 4.2, we employ the regularization method and prove the existence of weak solutions of the auxiliary problems. In order to obtain the solution of problem $\mathcal{P}$, some regular estimates for weak solutions of the auxiliary problems are also necessary in Subsection 4.3. Finally, we prove the main two theorems (Theorem 4.4 and Theorem 4.5) in Subsection 4.4.

### 4.1 Main Results

For the sake of simplicity, throughout the rest of this paper we use the notation

$$\overrightarrow{q}_v(A) := \overrightarrow{q}(v, A) = \frac{A}{1+\alpha(v)|A|},$$

where $v \in L^2(\Omega_T)$, $A \in \mathbb{R}^N$.

**Definition 4.1** A measurable function $u : (0,T) \times \Omega \to \mathbb{R}$ is a strong solution of $(P)$ in $\Omega_T$ if $u \in C([0,t], L^2(\Omega))$, $u(0) = u_0$, $u'(t) \in L^2(\Omega)$, $u(t) \in \mathrm{BV}(\Omega) \bigcap L^2(\Omega)$ a.e. $t \in [0,T]$, and there exists $z(t) \in X(\Omega)_1$ with $\|\alpha(u)z\|_\infty \le 1$, satisfying, for almost all $t \in [0,T]$, that

$$u'(t) = \mathrm{div}(z(t)) \ \text{ in } \mathcal{D}(\Omega), \tag{4.1}$$

$$z(t) = \frac{\nabla u}{1+\alpha(u)|\nabla u|} \ \ \mathcal{L}^N\text{-a.e. on } \Omega, \tag{4.2}$$

$$\alpha(u)\,(z(t), D^s u(t)) = |D^s u(t)| \ \text{ in } \mathcal{M}(\Omega), \tag{4.3}$$

$$\alpha(u)[z(t), \vec{n}] \in -\mathrm{sign}(u^\Omega) \ \ \mathcal{H}^{N-1}\text{-a.e. on } \partial\Omega. \tag{4.4}$$

**Definition 4.2** A generalized Young measure valued solution of $\mathcal{P}$ is a pair $(u, \lambda)$, where $u \in L^\infty(0,T; \mathrm{BV}(\Omega) \bigcap L^2(\Omega))$, $\frac{\partial u}{\partial t} \in L^2(\Omega_T)$, and, for almost all $t \in (0,T)$, $\lambda = (\nu, \lambda_\nu, \nu^\infty)_{x,t}$ is a parametrized family of generalized Young measures in $Y(\Omega; \mathbb{R}^N)$ such that

$$u_t = \mathrm{div}\langle \nu, \overrightarrow{q}_u \rangle, \ \text{in } \mathcal{D}'(\Omega). \tag{4.5}$$

$$Du = \langle \nu, I\rangle\, \mathcal{L}^N \llcorner \Omega + \langle \nu^\infty, I\rangle\, \lambda_\nu \llcorner \Omega \ \text{ in } \ \mathcal{M}(\Omega) \tag{4.6}$$

and

$$\langle \nu^\infty, I\rangle\, \lambda_\nu \llcorner \partial\Omega = \left(-u^\Omega \otimes \vec{n}\right) \mathcal{H}^{N-1} \llcorner \partial\Omega \ \text{ in } \ \mathcal{M}(\overline{\Omega}), \tag{4.7}$$

$$u(x,0) = u_0(x), \ x \in \Omega. \tag{4.8}$$

Furthermore,

$$\langle\langle |\cdot|, \lambda\rangle\rangle \in L^\infty(0,T), \tag{4.9}$$

and the (JF) inequality

$$\begin{aligned} &\langle \nu, \overrightarrow{q}_u \rangle \cdot \left(\langle \nu, I\rangle\, \mathcal{L}^N \llcorner \Omega + \langle \nu^\infty, I\rangle\, \lambda_\nu \llcorner \overline{\Omega}\right) \\ &\ge \left(\langle \nu, \overrightarrow{q}_u \cdot I\rangle\, \mathcal{L}^N \llcorner \Omega + \langle \nu^\infty, (\overrightarrow{q}_u \cdot I)^\infty\rangle\, \lambda_\nu \llcorner \overline{\Omega}\right) \end{aligned} \tag{4.10}$$

holds in $\mathcal{M}(\overline{\Omega})$.

**Remark 4.3** If $(u, \lambda)$ is a generalized Young measure solution, and, for almost all $t \in (0,T)$, $\lambda \in GY(\Omega; \mathbb{R}^N)$, we say that $(u, \lambda)$ is a generalized gradient Young measure solution.

**Theorem 4.4** Given $u_0 \in \mathrm{BV}(\Omega) \bigcap L^\infty(\Omega)$ with a trace of zero on $\partial\Omega$, then there exists a generalized Young measure solution of $\mathcal{P}$ for every $T > 0$.

**Theorem 4.5** Suppose that all the assumptions of Theorem 4.4 hold. This admits a strong solution of $\mathcal{P}$ for every $T > 0$.

### 4.2 Weak Solution for Auxiliary Problem

We use $\mathcal{P}_\varepsilon$ to denote the following auxiliary problem:

$$\begin{cases} \dfrac{\partial u}{\partial t} = \operatorname{div}\left(\dfrac{\nabla u}{1+\alpha(u)|\nabla u|}\right) + \varepsilon\triangle u & \text{in } \Omega_T, \quad (4.11)\\ u(x,t) = 0 & \text{on } \Gamma, \quad (4.12)\\ u(x,0) = u_0^\varepsilon(x) & \text{in } \Omega. \quad (4.13) \end{cases}$$

Here $u_0^\varepsilon(x) \in H_0^1(\Omega)$ and $u_0^\varepsilon(x) \in H_0^1(\Omega)$ satisfies, when $\varepsilon \to 0$, that

$$\begin{cases} \|u_0^\varepsilon(x)\|_{\mathrm{BV}(\Omega)} \to \|u_0(x)\|_{\mathrm{BV}(\Omega)}, & (4.14)\\ u_0^\varepsilon(x) \to u_0(x) \text{ in } L^2(\Omega), & (4.15)\\ \|u_0^\varepsilon(x)\|_{L^\infty(\Omega)} \leqslant \|u_0(x)\|_{L^\infty(\Omega)}, & (4.16)\\ \sqrt{\varepsilon_n}\|u_0^{\varepsilon_n}(x)\|_{H^1(\Omega)} \to 0, \text{ for a subsequence of } \{\varepsilon\}. & (4.17) \end{cases}$$

First, we are going to prove that there exists a weak solution (in that usual sense) of $\mathcal{P}_\varepsilon$. For any $v \in L^2(Q_T)$, we denote $u = Tv$ as the solution to the following problem:

$$\begin{cases} \dfrac{\partial u}{\partial t} = \operatorname{div}\left(\dfrac{\nabla u}{1+\alpha(v)|\nabla u|}\right) + \varepsilon\triangle u & \text{in } \Omega_T, \quad (4.18)\\ u(x,t) = 0 & \text{on } \Gamma, \quad (4.19)\\ u(x,0) = u_0^\varepsilon(x) & \text{in } \Omega. \quad (4.20) \end{cases}$$

By the usual theory of monotone operators, there exists a unique $u = Tv$, and $u \in L^2(0,T; H_0^1(\Omega))$, $\frac{\partial u}{\partial t} \in L^2(0,T; H^{-1}(\Omega))$. The following estimate holds:

$$|f * \tilde{v}| = \left|\int_{\mathbb{R}^N} f(x-y)\tilde{v}(y)\mathrm{d}y\right| \leqslant \|f\|_{L^2(\mathbb{R}^N)}\|\tilde{v}\|_{L^2(\mathbb{R}^N)} \leqslant \|f\|_{L^2(\mathbb{R}^N)}\|v\|_{L^2(\Omega)}.$$

Thus,

$$0 < c \leq \alpha(v) \leq C\|v\|_{L^2(\Omega)}. \quad (4.21)$$

According to the definition of weak solutions,

$$\int_0^s \langle u_t, \varphi\rangle \mathrm{d}t + \int_0^s \int_\Omega \left(\frac{\nabla u}{1+\alpha(v)\,|\nabla u|} + \varepsilon\nabla u\right)\cdot\nabla\varphi \mathrm{d}x\mathrm{d}t = 0.$$

Taking $\varphi = u$, we have that

$$\|u(s)\|_{L^2(\Omega)}^2 + 2\int_0^s \int_\Omega \left(\frac{|\nabla u|^2}{1+\alpha(v)\,|\nabla u|} + \varepsilon|\nabla u|^2\right)\mathrm{d}x\mathrm{d}t = \|u_0^\varepsilon\|_{L^2(\Omega)}^2 \leqslant M,$$

where $M$ is a constant independent of $\varepsilon$ and $T$. It follows that

$$\operatorname*{ess\,sup}_{0<t<T} \|u(t)\|_{L^2(\Omega)}^2 + \int_0^T \int_\Omega \left(\frac{|\nabla u|^2}{1+\alpha(v)\,|\nabla u|} + \varepsilon|\nabla u|^2\right)\mathrm{d}x\mathrm{d}t \leqslant M.$$

Then, by

$$\left|\int_0^T \langle u_t, \varphi\rangle \mathrm{d}t\right| = \left|\int_0^T \int_\Omega \left(\frac{\nabla u}{1+\alpha(v)\,|\nabla u|} + \varepsilon\nabla u\right)\cdot\nabla\varphi \mathrm{d}x\mathrm{d}t\right|$$

$$
\begin{aligned}
&\leqslant \left| \iint_{\Omega_T} \frac{\nabla u}{1+\alpha(v)\,|\nabla u|} \cdot \nabla\varphi \mathrm{d}x\mathrm{d}t \right| + \left| \iint_{\Omega_T} \varepsilon \nabla u \cdot \nabla\varphi \mathrm{d}x\mathrm{d}t \right| \\
&\leqslant \left( \iint_{\Omega_T} \frac{|\nabla u|^2}{1+\alpha(v)\,|\nabla u|} \mathrm{d}x\mathrm{d}t \right)^{\frac{1}{2}} \cdot \left( \iint_{\Omega_T} \frac{|\nabla \varphi|^2}{1+\alpha(v)\,|\nabla u|} \mathrm{d}x\mathrm{d}t \right)^{\frac{1}{2}} \\
&\quad + \sqrt{\varepsilon} \left( \iint_{\Omega_T} \varepsilon |\nabla u|^2 \mathrm{d}x\mathrm{d}t \right)^{\frac{1}{2}} \cdot \left( \iint_{\Omega_T} |\nabla \varphi|^2 \mathrm{d}x\mathrm{d}t \right)^{\frac{1}{2}} \\
&\leqslant C \|\nabla \varphi\|_{L^2(\Omega_T)},
\end{aligned}
$$

where $C$ is a constant independent of $\varepsilon$ and $T$, we arrive at

$$
\begin{cases}
\|u\|_{L^2(0,T;H_0^1(\Omega))} \leqslant \dfrac{C}{\sqrt{\varepsilon}}, & (4.22) \\
\|u\|_{L^\infty(0,T;L^2(\Omega))} \leqslant C, & (4.23) \\
\|u_t\|_{L^2(0,T;H^{-1}(\Omega))} \leqslant C. & (4.24)
\end{cases}
$$

Now, we write

$$
W(0,T) = \left\{ w \in L^2(\Omega_T) \;\middle|\; \begin{array}{l} \|w\|_{L^2(0,T;H_0^1(\Omega))} \leqslant \dfrac{C}{\sqrt{\varepsilon}}, \\ \|w\|_{L^\infty(0,T;L^2(\Omega))} \leqslant C, \\ \|w_t\|_{L^2(0,T;H^{-1}(\Omega))} \leqslant C. \end{array} \right\}
$$

Applying the compact embedding theorem of Aubin-Lions (see also Lemma 3.10), we get that $W(0,T)$ is precompact in $L^2(\Omega_T)$. Let us now consider the mapping $T : L^2(\Omega_T) \to L^2(\Omega_T)$. Based on Schauder's fixed point theorem, it is clear that this mapping will have a fixed point provided that it is continuous. To prove this, note that $u_1 = T(v_1)$, $u_2 = T(v_2)$. According to (4.18), taking the texting function $\varphi = u_1 - u_2$, it follows that

$$
\begin{aligned}
&\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u_1 - u_2\|^2 + \int_\Omega \varepsilon |\nabla(u_1 - u_2)|^2 \mathrm{d}x \\
&\quad + \int_\Omega \left( \frac{\nabla u_1}{1+\alpha(v_1)\,|\nabla u_1|} - \frac{\nabla u_2}{1+\alpha(v_1)\,|\nabla u_2|} \right) \cdot \nabla(u_1 - u_2) \mathrm{d}x \\
&= -\int_\Omega \left( \frac{\nabla u_2}{1+\alpha(v_1)\,|\nabla u_2|} - \frac{\nabla u_2}{1+\alpha(v_2)\,|\nabla u_2|} \right) \cdot (\nabla u_1 - \nabla u_2)\,\mathrm{d}x \\
&\leqslant \int_\Omega \frac{|\alpha(v_1) - \alpha(v_2)| \cdot |\nabla u_2|^2}{(1+\alpha(v_1)\,|\nabla u_2|) \cdot (1+\alpha(v_2)\,|\nabla u_2|)} \cdot |\nabla u_1 - \nabla u_2|\,\mathrm{d}x. \qquad (4.25)
\end{aligned}
$$

Then, by

$$
|\alpha(v_1) - \alpha(v_2)| \leqslant C \left| \int_{\mathbb{R}^N} f(x-y)\,(\tilde{v}_1(y) - \tilde{v}_2(y))\,\mathrm{d}y \right|,
$$

we get that

$$
|\alpha(v_1) - \alpha(v_2)|_{L^\infty(\mathbb{R}^N)} \leqslant C \|v_1 - v_2\|_{L^2(\Omega)}. \qquad (4.26)
$$

Hence, by (4.25), (4.26) and Young's inequality, we have that

$$
\frac{\mathrm{d}}{\mathrm{d}t}\,\|u_1 - u_2\|_{L^2(\Omega)}^2 \leqslant \frac{C}{2\varepsilon c^2}\,\|v_1 - v_2\|_{L^2(\Omega)}^2, \qquad (4.27)
$$

which proves the continuity of the mapping $T$, and thus that there exists a weak solution $u^\varepsilon$ to the problem $\mathcal{P}_\varepsilon$. On the other hand, using Gronwall's inequality, (4.27) ensures the uniqueness of $\mathcal{P}_\varepsilon$.

### 4.3 Regularity for Weak Solution of $\mathcal{P}_\varepsilon$

In order to obtain the general Young measure solution of $\mathcal{P}$, some estimates are also necessary.

Since $L^2(0,T;H^{-1}(\Omega))$ is the dual space of $L^2(0,T;H_0^1(\Omega))$, for every $w\in L^2(0,T;H^{-1}(\Omega))$, it follows that

$$w\in L^2(0,T;H^{-1}(\Omega))\Leftrightarrow\begin{cases} w=w_0+\sum\limits_i D_{x_i}w_i,\ \ w_0,w_i\in L^2\left(\Omega_T\right),i=1,2,\cdots,n,\\ \forall\varphi\in L^2(0,T;H_0^1(\Omega)),\ \ \langle w,\varphi\rangle=\displaystyle\int_{\Omega_T}\left(w_0\varphi+\sum_i w_iD_{x_i}\varphi\right)\mathrm{d}x\mathrm{d}t. \end{cases}$$

**Lemma 4.6** Let $v\in W(0,T)$ and $f\in\mathcal{S}(\mathbb{R}^N)$. We note that $v_t=v_0+\sum\limits_i D_{x_i}v_i$, $v_i\in L^2\left(\Omega_T\right),i=0,1,\cdots,n$. Then it follows that

$$(f*\tilde{v})_t=v_0-\sum_i\left(D_{x_i}f\right)*\tilde{v}_i. \tag{4.28}$$

Moreover, for almost all $t\in[0,T]$,

$$\|(f*\tilde{v})_t\|_{L^\infty(\Omega)}\le C\|v_t\|_{H^{-1}(\Omega)}, \tag{4.29}$$

where $C$ is a constant that depends only on $\sum\limits_i\|D_{x_i}f\|_{L^2(\mathbb{R}^N)}$.

**Proof** By the definition of the weak derivative of $f*\tilde{v}$ and Fubini's Theorem, for every $\varphi\in\mathcal{D}(\Omega_T)$, one has that

$$\begin{aligned}\left\langle(f*\tilde{v})_t,\varphi\right\rangle&=-\int_0^T\int_{R^N}f*\tilde{v}\cdot\varphi_t(x,t)\mathrm{d}x\mathrm{d}t\\&=-\int_0^T\int_{R^N}\tilde{v}*f\cdot\varphi_t(x,t)\mathrm{d}x\mathrm{d}t\\&=-\int_0^T\int_{R^N}\left(\int_{R^N}f(x-y)\tilde{v}(y,t)dy\right)\varphi_t(x,t)\mathrm{d}x\mathrm{d}t\\&=-\int_0^T\int_{R^N}\left(\int_{R^N}f(x-y)\varphi_t(x,t)\mathrm{d}x\right)\tilde{v}(y,t)\mathrm{d}y\mathrm{d}t\\&=-\int_0^T\int_{R^N}\left(\int_{R^N}f(x)\varphi_t(x+y,t)\mathrm{d}x\right)\tilde{v}(y,t)\mathrm{d}y\mathrm{d}t\\&=-\int_{R^N}\left(\int_0^T\int_{R^N}\varphi_t(x+y,t)\tilde{v}(y,t)\mathrm{d}y\mathrm{d}t\right)f(x)\mathrm{d}x\\&=\int_{R^N}\langle\tilde{v}_t(y,t),\varphi(x+y,t)\rangle f(x)\mathrm{d}x.\end{aligned}$$

On the other hand,

$$\begin{aligned}&\int_{R^N}\left(\int_0^T\int_{R^N}\tilde{v}_i(y,t)D_{y_i}\varphi(x+y,t)\mathrm{d}y\mathrm{d}t\right)f(x)\mathrm{d}x\\=&\int_0^T\int_{R^N}\tilde{v}_i(y,t)\left(\int_{R^N}f(x)D_{y_i}\varphi(x+y,t)\mathrm{d}x\right)\mathrm{d}y\mathrm{d}t\\=&\int_0^T\int_{R^N}\tilde{v}_i(y,t)\left(\int_{R^N}f(x)D_{x_i}\varphi(x+y,t)\mathrm{d}x\right)\mathrm{d}y\mathrm{d}t\end{aligned}$$

$$
\begin{aligned}
&= -\int_0^T \int_{R^N} \tilde{v}_i(y,t) \left( \int_{R^N} \varphi(x,t) D_{x_i} f(x-y) \mathrm{d}x \right) \mathrm{d}y \mathrm{d}t \\
&= -\int_0^T \int_{R^N} \varphi(x,t) \left( \int_{R^N} \tilde{v}_i(y,t) D_{x_i} f(x-y) \mathrm{d}y \right) \mathrm{d}x \mathrm{d}t \\
&= -\left\langle (D_{x_i} f) * \tilde{v}_i, \varphi \right\rangle .
\end{aligned}
$$

Collecting all of these facts, we obtain (4.28).

Moreover, by Hölder's inequality, it follows that

$$
|(f * \tilde{v})_t| \leqslant \sum_i \|D_{x_i} f\|_{L^2(\mathbb{R}^N)} \cdot \|v_i\|_{L^2(\Omega)} \leqslant \sum_i \|D_{x_i} f\|_{L^2(\mathbb{R}^N)} \cdot \|v_t\|_{H^{-1}(\Omega)},
$$

and this implies (4.29). □

**Lemma 4.7** If $u^\varepsilon$ is a solution of $\mathcal{P}_\varepsilon$, then

$$
\operatorname*{ess\,sup}_{0<t<T} \int_\Omega \left( \frac{|\nabla u^\varepsilon|^2}{1+\alpha(v)\,|\nabla u^\varepsilon|} + \varepsilon |\nabla u^\varepsilon|^2 \right) \mathrm{d}x + \int_0^T \int_\Omega |u_t^\varepsilon|^2 \mathrm{d}x \mathrm{d}t \leqslant C, \tag{4.30}
$$

where $C$ is a constant independent of $\varepsilon$.

**Proof** Let us use the notation

$$
h(t) = \max_{x \in \overline{\Omega}} |\alpha(v)_t| . \tag{4.31}
$$

Multiply relations (4.18) by $u_t$. This gives the equality

$$
\|u_t\|^2_{L^2(\Omega)} + \int_\Omega \frac{\nabla u}{1+\alpha(v)\,|\nabla u|} \cdot \nabla u_t \mathrm{d}x + \int_\Omega \varepsilon \nabla u \cdot \nabla u_t \mathrm{d}x = 0. \tag{4.32}
$$

We use the formulas

$$
\frac{\nabla u}{1+\alpha(v)\,|\nabla u|} \cdot \nabla u_t = \frac{\partial}{\partial t} \int_0^{|\nabla u|} \frac{s}{1+\alpha(v)s} \mathrm{d}s + \alpha(v)_t \int_0^{|\nabla u|} \left( \frac{s}{1+\alpha(v)s} \right)^2 \mathrm{d}s,
$$

$$
\varepsilon \nabla u \cdot \nabla u_t = \frac{\partial}{\partial t} \left( \frac{\varepsilon}{2} |\nabla u|^2 \right).
$$

Substituting these into (4.32), we rewrite it in the form

$$
\|u_t\|^2_{L^2(\Omega)} + Y'(t) = J, \tag{4.33}
$$

where

$$
Y(t) = \int_\Omega \int_0^{|\nabla u|} \frac{s}{1+\alpha(v)s} \mathrm{d}s \mathrm{d}x + \int_\Omega \frac{\varepsilon}{2} |\nabla u|^2 \mathrm{d}x, \tag{4.34}
$$

$$
J = -\int_\Omega \alpha(v)_t \int_0^{|\nabla u|} \left( \frac{s}{1+\alpha(v)s} \right)^2 \mathrm{d}s \mathrm{d}x. \tag{4.35}
$$

Since $c \leq \alpha(v)$, it follows that

$$
|J| \leqslant \frac{h(t)}{c} \int_\Omega \int_0^{|\nabla u|} \frac{s}{1+\alpha(v)s} \mathrm{d}s \mathrm{d}x \leqslant \frac{h(t)}{c} Y(t).
$$

On the other hand, by Lemma 4.6,

$$
h(t) \leq C \|v_t\|_{H^{-1}}(\Omega),
$$

Thus, the function $Y(t)$ satisfies the differential inequality

$$
\|u_t\|^2_{L^2(\Omega)} + Y'(t) \leqslant \frac{h(t)}{c} Y(t). \tag{4.36}
$$

Omitting the nonnegative terms on the left-hand side, we estimate $Y(t)$ by using Gronwall's lemma. The estimate follows from the integration of (4.36) over the interval $(0,T)$. □

### 4.4 Letting $\varepsilon \to 0$

**Proof of Theorem 4.4** We set

$$\begin{cases} \overrightarrow{p} := \overrightarrow{q}(u,A) = \dfrac{A}{1+\alpha(u)|A|}, \\ \overrightarrow{p}_n := \overrightarrow{q}(u^{\varepsilon_n},A) = \dfrac{A}{1+\alpha(u^{\varepsilon_n})|A|}, \\ \lambda^{\varepsilon_n} := (\nu^{\varepsilon_n},0,0) = \sigma_{Du^{\varepsilon_n}}. \end{cases}$$

According to Lemma 3.8 and Lemma 4.7, for almost all $t \in (0,T)$, we can extract from $\{u^{\varepsilon_n}\}$ and $\{\nu^{\varepsilon_n}\}$ a subsequence (still labeled $\{u^{\varepsilon_n}\}$ and $\{\nu^{\varepsilon_n}\}$) such that

$$\begin{cases} u^{\varepsilon_n} \to u \ \text{ in } L^2(\Omega), & (4.37) \\ u^{\varepsilon_n} \overset{*}{\rightharpoonup} u \ \text{ in } L^\infty(\Omega), & (4.38) \\ u_t^{\varepsilon_n} \rightharpoonup u_t \ \text{ in } L^2(\Omega), & (4.39) \\ \nabla \tilde{u}^{\varepsilon_n} \mathcal{L}^N \llcorner \tilde{\Omega} \overset{*}{\rightharpoonup} D\tilde{u} \ \text{ in } \mathcal{M}(\tilde{\Omega}), & (4.40) \\ \varepsilon_n |\nabla u^{\varepsilon_n}| \to 0 \ \text{ in } L^2(\Omega). & (4.41) \end{cases}$$

$$\begin{cases} \langle \nu^{\varepsilon_n}, \overrightarrow{p} \rangle \overset{*}{\rightharpoonup} \langle \nu, \overrightarrow{p} \rangle \ \text{ in } L^\infty(\Omega;\mathbb{R}^N), & (4.42) \\ \langle \nu^{\varepsilon_n}, \overrightarrow{p} \cdot I \rangle \mathcal{L}^N \llcorner \Omega \overset{*}{\rightharpoonup} \langle \nu, \overrightarrow{p} \cdot I \rangle \mathcal{L}^N \llcorner \Omega + \left\langle \nu^\infty, (\overrightarrow{p} \cdot I)^\infty \right\rangle \lambda_\nu \ \text{ in } \mathcal{M}(\overline{\Omega}), & (4.43) \\ \lambda^{\varepsilon_n} \overset{*}{\rightharpoonup} \lambda \text{ in } Y\left(\Omega;\mathbb{R}^N\right), & (4.44) \end{cases}$$

where $\lambda = (\nu, \lambda_\nu, \nu^\infty)$.

**Step 1** According to Lemma 4.7, since $\frac{\partial u^{\varepsilon_n}}{\partial t}$ is bounded in $L^2(\Omega_T)$ and $u^{\varepsilon_n}$ is bounded in $L^\infty(0,T;W_0^{1,1}(\Omega))$, we have that $\frac{\partial \alpha(u^{\varepsilon_n})}{\partial t}$ is bounded in $L^2(0,T;C(\overline{\Omega}))$ and $\alpha(u^{\varepsilon_n})$ is bounded in $L^\infty(0,T;W^{1,\infty}(\Omega))$. Thus, according to Lemma 3.10, there exists a subsequence $\{\alpha(u^{\varepsilon_n})\}$ (not relabeled) satisfying

$$\alpha(u^{\varepsilon_n}) \to \alpha(u) \ \text{ uniformly in } \ C([0,T];C(\overline{\Omega})). \tag{4.45}$$

Moreover,

$$\begin{aligned} |\langle \nu^{\varepsilon_n}, \overrightarrow{p}_n \rangle - \langle \nu^{\varepsilon_n}, \overrightarrow{p} \rangle| &\leqslant \int_{R^N} \frac{|A|^2 \, |\alpha(u^{\varepsilon_n}) - \alpha(u)|}{(1+\alpha(u)|A|)\,(1+\alpha(u^{\varepsilon_n})|A|)} \mathrm{d}\nu^{\varepsilon_n} \\ &\leqslant \frac{1}{c^2} \, |\alpha(u^{\varepsilon_n}) - \alpha(u)|, \end{aligned}$$

which converges to zero uniformly in $\overline{\Omega}$. Together with (4.42), it follows that

$$\langle \nu^{\varepsilon_n}, \overrightarrow{p}_n \rangle \overset{*}{\rightharpoonup} \langle \nu, \overrightarrow{p} \rangle \ \text{ in } L^\infty(\Omega;\mathbb{R}^N). \tag{4.46}$$

Therefore, by (4.39) and (4.41), we get that

$$u_t = \mathrm{div}\langle \nu, \overrightarrow{p} \rangle, \ \text{ in } \mathcal{D}'(\Omega).$$

Since $u^{\varepsilon_n} \in H_0^1(\Omega)$,

$$D\tilde{u}^{\varepsilon_n} = \langle \nu^{\varepsilon_n}, I \rangle \mathcal{L}^N \llcorner \Omega. \tag{4.47}$$

As

$$D\tilde{u}^{\varepsilon_n} \overset{*}{\rightharpoonup} D\tilde{u} \text{ in } \mathcal{M}(\tilde{\Omega})$$

and

$$\langle \nu^{\varepsilon_n}, I\rangle \, \mathcal{L}^N \llcorner \Omega \overset{*}{\rightharpoonup} \langle \nu, I\rangle \, \mathcal{L}^N \llcorner \Omega + \langle \nu^{\infty}, I\rangle \, \lambda_{\nu} \ \text{ in } \mathcal{M}(\overline{\Omega}), \tag{4.48}$$

we obtain that

$$D\tilde{u} = \langle \nu, I\rangle \, \mathcal{L}^N \llcorner \Omega + \langle \nu^{\infty}, I\rangle \, \lambda_{\nu} \ \text{ in } \mathcal{M}(\tilde{\Omega}). \tag{4.49}$$

Therefore,

$$Du = \langle \nu, I\rangle \, \mathcal{L}^N \llcorner \Omega + \langle \nu^{\infty}, I\rangle \, \lambda_{\nu} \llcorner \Omega \ \text{ in } \ \mathcal{M}(\Omega) \tag{4.50}$$

and

$$D\tilde{u} \llcorner \partial\Omega = \langle \nu^{\infty}, I\rangle \, \lambda_{\nu} \llcorner \partial\Omega = \left(-u^{\Omega} \otimes \vec{n}\right) \mathcal{H}^{N-1} \llcorner \partial\Omega \ \text{ in } \ \mathcal{M}(\overline{\Omega}). \tag{4.51}$$

**Step 2** Since $u_t = \operatorname{div}\langle \nu, \overrightarrow{p}\rangle$, in $\mathcal{D}'(\Omega)$, with $u_t \in L^2(\Omega)$, we get that

$$\langle \nu, \overrightarrow{p}\rangle \in X_2(\Omega).$$

Similarly, by (4.18) and $u_t^{\varepsilon_n} \in L^2(\Omega)$, one has that

$$\langle \nu^{\varepsilon_n}, \overrightarrow{p} + \varepsilon_n I\rangle \in X_2(\Omega).$$

For any $0 \leq \varphi \in C_c^{\infty}(\tilde{\Omega})$, we obtain that

$$\begin{aligned} -\int_{\tilde{\Omega}} u_t^{\varepsilon_n} \varphi u^{\varepsilon_n} \mathrm{d}x &= -\int_{\tilde{\Omega}} \operatorname{div}\left(\langle \nu^{\varepsilon_n}, \overrightarrow{p} + \varepsilon_n I\rangle\right) \varphi u^{\varepsilon_n} \mathrm{d}x \\ &= \int_{\tilde{\Omega}} \varphi \, \langle \nu^{\varepsilon_n}, \overrightarrow{p} + \varepsilon_n I\rangle \cdot \nabla u^{\varepsilon_n} \mathrm{d}x + \int_{\tilde{\Omega}} u^{\varepsilon_n} \, \langle \nu^{\varepsilon_n}, \overrightarrow{p} + \varepsilon_n I\rangle \cdot \nabla \varphi \mathrm{d}x. \end{aligned} \tag{4.52}$$

Observe that, by (4.45) and Lemma 4.7, we obtain that

$$\begin{aligned} &\left| \int_{\tilde{\Omega}} \varphi \left( \langle \nu^{\varepsilon_n}, \overrightarrow{p}_n \cdot I\rangle - \langle \nu^{\varepsilon_n}, \overrightarrow{p} \cdot I\rangle \right) \mathrm{d}x \right| \\ &\leqslant \int_{\tilde{\Omega}} \varphi \int_{\mathbb{R}^N} \left| \frac{|A|^2}{1 + \alpha(u^{\varepsilon_n})|A|} - \frac{|A|^2}{1 + \alpha(u)|A|} \right| \mathrm{d}\nu^{\varepsilon_n} \mathrm{d}x \\ &\leqslant \frac{1}{c^2} \int_{\tilde{\Omega}} \left( \varphi \, |\alpha(u^{\varepsilon_n}) - \alpha(u)| \int_{\mathbb{R}^N} |A| \mathrm{d}\nu^{\varepsilon_n} \right) \mathrm{d}x \\ &\leqslant \frac{1}{c^2} \|\varphi\|_{C(\overline{\Omega})} \cdot \|\alpha(u^{\varepsilon_n}) - \alpha(u)\|_{C(\overline{\Omega})} \cdot \|\nabla u^{\varepsilon_n}\|_{L^1(\Omega)} \\ &\leqslant C \|\alpha(u^{\varepsilon_n}) - \alpha(u)\|_{C(\overline{\Omega})} \to 0. \end{aligned}$$

By (4.47), we have that

$$\begin{aligned} \int_{\tilde{\Omega}} \varphi \, \langle \nu^{\varepsilon_n}, \overrightarrow{p} + \varepsilon_n I\rangle \cdot \nabla u^{\varepsilon_n} \mathrm{d}x &= \int_{\tilde{\Omega}} \varphi \, \langle \nu^{\varepsilon_n}, \overrightarrow{p} + \varepsilon_n I\rangle \cdot \langle \nu^{\varepsilon_n}, I\rangle \, \mathrm{d}x \\ &= \int_{\tilde{\Omega}} \varphi \, \langle \nu^{\varepsilon_n}, \overrightarrow{p} \cdot I\rangle \, \mathrm{d}x + \varepsilon_n \int_{\tilde{\Omega}} \varphi |\langle \nu^{\varepsilon_n}, I\rangle|^2 \mathrm{d}x \\ &\geqslant \int_{\tilde{\Omega}} \varphi \, \langle \nu^{\varepsilon_n}, \overrightarrow{p} \cdot I\rangle \, \mathrm{d}x. \end{aligned} \tag{4.53}$$

Combining (4.37), (4.41) and (4.46), we get that

$$\begin{aligned} &\int_{\tilde{\Omega}} u^{\varepsilon_n} \, \langle \nu^{\varepsilon_n}, \overrightarrow{p}_n + \varepsilon_n I\rangle \cdot \nabla \varphi \mathrm{d}x \\ &= \int_{\tilde{\Omega}} u^{\varepsilon_n} \, \langle \nu^{\varepsilon_n}, \overrightarrow{p}_n\rangle \cdot \nabla \varphi \mathrm{d}x + \int_{\tilde{\Omega}} u^{\varepsilon_n} \left( \varepsilon_n \nabla u^{\varepsilon_n} \cdot \nabla \varphi \right) \mathrm{d}x \end{aligned}$$

$$\to \int_{\tilde{\Omega}} u \langle \nu, \overrightarrow{p} \rangle \cdot \nabla \varphi \mathrm{d}x, \quad n \to \infty. \tag{4.54}$$

On the other hand, by Green's formula of Lemma 3.6, it follows that

$$\begin{aligned} -\int_{\tilde{\Omega}} u_t \varphi u \mathrm{d}x &= -\int_{\tilde{\Omega}} \operatorname{div}\left(\langle \nu, \overrightarrow{p} \rangle\right) \varphi u \mathrm{d}x \\ &= \int_{\tilde{\Omega}} \varphi \left(\langle \nu, \overrightarrow{p} \rangle, D\tilde{u}\right) + \int_{\tilde{\Omega}} u \langle \nu, \overrightarrow{p} \rangle \cdot \nabla \varphi \mathrm{d}x \\ &= \int_{\tilde{\Omega}} \varphi \langle \nu, \overrightarrow{p} \rangle \cdot \left(\langle \nu, I \rangle \mathrm{d}x + \langle \nu^{\infty}, I \rangle \mathrm{d}\lambda_{\nu}\right) + \int_{\tilde{\Omega}} u \langle \nu, \overrightarrow{p} \rangle \cdot \nabla \varphi \mathrm{d}x. \end{aligned} \tag{4.55}$$

Hence, by (4.52), (4.53), (4.54) and (4.55), letting $n \to \infty$, we have that

$$\int_{\tilde{\Omega}} \varphi \langle \nu, \overrightarrow{p} \rangle \cdot \left(\langle \nu, I \rangle \mathrm{d}x + \langle \nu^{\infty}, I \rangle \mathrm{d}\lambda_{\nu}\right) \geqslant \int_{\tilde{\Omega}} \varphi \left(\langle \nu, \overrightarrow{p} \cdot I \rangle \mathrm{d}x + \left\langle \nu^{\infty}, \left(\overrightarrow{p} \cdot I\right)^{\infty} \right\rangle \mathrm{d}\lambda_{\nu}\right).$$

Furthermore, by the arbitrariness of $\varphi$ and the definition of $\tilde{u}$,

$$\langle \nu, \overrightarrow{p} \rangle \cdot \left(\langle \nu, I \rangle \mathcal{L}^N \llcorner \Omega + \langle \nu^{\infty}, I \rangle \lambda_{\nu} \llcorner \overline{\Omega}\right) \geq \left(\langle \nu, \overrightarrow{p} \cdot I \rangle \mathcal{L}^N \llcorner \Omega + \langle \nu^{\infty}, \left(\overrightarrow{p} \cdot I\right)^{\infty} \rangle \lambda_{\nu} \llcorner \overline{\Omega}\right) \tag{4.56}$$

holds in $\mathcal{M}(\overline{\Omega})$. Moreover, $\lambda^{\varepsilon_n} \in GY(\Omega)$ yields that $\lambda \in GY(\Omega)$, and $u$ is a generalized gradient Young measure solution. □

**Proof of Theorem 4.5** To prove Theorem 4.5, we only need to prove that the generalized gradient Young measure solution in Theorem 4.4 is also a strong solution to problem $\mathcal{P}$. As in the proof of Theorem 4.4, one has that

$$\begin{cases} u^{\varepsilon_n} \to u \ \text{ in } L^2(\Omega), & (4.57) \\ u_t^{\varepsilon_n} \rightharpoonup u_t \ \text{ in } L^2(\Omega), & (4.58) \\ \varepsilon_n |\nabla u^{\varepsilon_n}| \to 0 \ \text{ in } L^2(\Omega), & (4.59) \\ \langle \nu^{\varepsilon_n}, \overrightarrow{p}_n \rangle \overset{*}{\rightharpoonup} \langle \nu, \overrightarrow{p} \rangle \ \text{ in } L^{\infty}(\Omega; \mathbb{R}^N). & (4.60) \end{cases}$$

Since

$$u_t = \operatorname{div}\langle \nu, \overrightarrow{p} \rangle, \ \text{in } \mathcal{D}'(\Omega), \tag{4.61}$$

we have that $\langle \nu, \overrightarrow{p} \rangle \in X_2(\Omega)$. Letting $0 \leq \theta \in \mathcal{D}(\Omega)$ and $\eta \in C^1(\overline{\Omega})$, by the monotonicity of $\overrightarrow{q}(u^{\varepsilon_n}, \cdot)$, it follows that

$$\int_{\Omega} \theta \left( \left(\overrightarrow{p}_n \left(\nabla u^{\varepsilon_n}\right) + \varepsilon_n \nabla u^{\varepsilon_n}\right) - \left(\overrightarrow{p}_n \left(\nabla \eta\right) + \varepsilon_n \nabla \eta\right)\right) \cdot \left(\nabla u^{\varepsilon_n} - \nabla \eta\right) \mathrm{d}x \geqslant 0;$$

since

$$\lim_{n \to \infty} \int_{\Omega} \theta u_t^{\varepsilon_n} \left(u^{\varepsilon_n} - \eta\right) \mathrm{d}x = \int_{\Omega} \theta u_t \left(u - \eta\right) \mathrm{d}x$$

and

$$\lim_{n \to \infty} \int_{\Omega} \left(u^{\varepsilon_n} - \eta\right) \nabla \theta \cdot \left(\overrightarrow{p}_n \left(\nabla u^{\varepsilon_n}\right) + \varepsilon_n \nabla u^{\varepsilon_n}\right) \mathrm{d}x = \int_{\Omega} \left(u - \eta\right) \nabla \theta \cdot \langle v, \overrightarrow{p} \rangle \, \mathrm{d}x.$$

Hence, by Green's formula of Lemma 3.6, we obtain that

$$\begin{aligned} &\lim_{n \to \infty} \int_{\Omega} \theta \left(\overrightarrow{p}_n \left(\nabla u^{\varepsilon_n}\right) + \varepsilon_n \nabla u^{\varepsilon_n}\right) \cdot \nabla \left(u^{\varepsilon_n} - \eta\right) \mathrm{d}x \\ &= \lim_{n \to \infty} \int_{\Omega} -\theta u_t^{\varepsilon_n} \left(u^{\varepsilon_n} - \eta\right) \mathrm{d}x + \lim_{n \to \infty} \int_{\Omega} -\left(u^{\varepsilon_n} - \eta\right) \nabla \theta \cdot \left(\overrightarrow{p}_n \left(\nabla u^{\varepsilon_n}\right) + \varepsilon_n \nabla u^{\varepsilon_n}\right) \mathrm{d}x \\ &= -\int_{\Omega} \theta u_t \left(u - \eta\right) \mathrm{d}x - \int_{\Omega} \left(u - \eta\right) \nabla \theta \cdot \langle v, \overrightarrow{p} \rangle \, \mathrm{d}x \end{aligned}$$

$$= \int_\Omega \theta\left(\langle \nu, \overrightarrow{p}\rangle, D\left(u-\eta\right)\right).$$

With this, and

$$\lim_{n\to\infty} \int_\Omega \theta\left(\overrightarrow{p}_n\left(\nabla\eta\right) + \varepsilon_n \nabla\eta\right) \cdot \left(\nabla u^{\varepsilon_n} - \nabla\eta\right) \mathrm{d}x = \int_\Omega \theta\left(\overrightarrow{p}\left(\nabla\eta\right), D\left(u-\eta\right)\right),$$

we have that

$$\int_\Omega \theta\left(\langle \nu, \overrightarrow{p}\rangle - \overrightarrow{p}\left(\nabla\eta\right), D\left(u-\eta\right)\right) \geqslant 0.$$

Therefore,

$$\left(\langle \nu, \overrightarrow{p}\rangle - \overrightarrow{p}\left(\nabla\eta\right), D\left(u-\eta\right)\right) \geqslant 0 \ \text{ in } \mathcal{M}(\Omega),$$

so its absolutely continuous part is

$$\left(\langle \nu, \overrightarrow{p}\rangle - \overrightarrow{p}\left(\nabla\eta\right)\right) \cdot \nabla\left(u-\eta\right) \geqslant 0 \quad \text{a.e. in } \Omega.$$

Since $C^1(\overline{\Omega})$ is separable, by taking a countable set as dense in $C^1(\overline{\Omega})$, we have that the above inequality holds for all $x \in \Omega'$, where $\Omega' \subseteq \Omega$ satisfies that $\mathcal{L}^N(\Omega \setminus \Omega') = 0$ and all $\eta \in C^1(\overline{\Omega})$. Now, fixing $x \in \Omega'$, and given $A \in \mathbb{R}^N$, there exists $\eta \in C^1(\overline{\Omega})$ such that $\nabla\eta = A$. Thus,

$$\left(\langle \nu, \overrightarrow{p}\rangle - \overrightarrow{p}\left(A\right)\right) \cdot \left(\nabla u(x) - A\right) \geqslant 0, \ \ \forall\, A \in \mathbb{R}^N.$$

By choosing $A = \nabla u(x) \pm \epsilon\xi$, $\forall \xi \in \mathbb{R}^N$ and letting $\epsilon \to 0^+$, we obtain that

$$\langle \nu, \overrightarrow{p}\rangle = \overrightarrow{p}(\nabla u(x)) = \frac{\nabla u}{1+\alpha(u)|\nabla u|}, \ \text{ a.e. } x \in \Omega, \tag{4.62}$$

which implies (4.2).

According to (4.49) and (4.56), we have that

$$D\tilde{u} = \langle \nu, I\rangle\, \mathcal{L}^N \llcorner \Omega + \langle \nu^\infty, I\rangle\, \lambda_\nu \ \text{ in } \mathcal{M}(\tilde{\Omega}) \tag{4.63}$$

and

$$\begin{aligned} &\langle \nu, \overrightarrow{p}\rangle \cdot \left(\langle \nu, I\rangle\, \mathcal{L}^N \llcorner \Omega + \langle \nu^\infty, I\rangle\, \lambda_\nu \llcorner \overline{\Omega}\right) \\ &\geq \left(\langle \nu, \overrightarrow{p}\cdot I\rangle\, \mathcal{L}^N \llcorner \Omega + \langle \nu^\infty, (\overrightarrow{p}\cdot I)^\infty\rangle\, \lambda_\nu \llcorner \overline{\Omega}\right) \ \text{ in } \mathcal{M}(\overline{\Omega}). \end{aligned} \tag{4.64}$$

Thus,

$$\left(\langle \nu, \overrightarrow{p}\rangle \langle \nu, I\rangle - \langle \nu, \overrightarrow{p}\cdot I\rangle\right) \mathcal{L}^N \llcorner \Omega + \left(\langle \nu, \overrightarrow{p}\rangle \langle \nu^\infty, I\rangle - \left\langle \nu^\infty, (\overrightarrow{p}\cdot I)^\infty\right\rangle\right) \lambda_\nu \llcorner \bar{\Omega} \geqslant 0, \tag{4.65}$$

and its singular parts are

$$\left(\langle \nu, \overrightarrow{p}\rangle \langle \nu^\infty, I\rangle - \left\langle \nu^\infty, (\overrightarrow{p}\cdot I)^\infty\right\rangle\right) \lambda_\nu^s \llcorner \bar{\Omega} \geqslant 0 \text{ in } \mathcal{M}(\overline{\Omega}).$$

Then, it follows that

$$\langle \nu, \overrightarrow{p}\rangle \langle \nu^\infty, I\rangle \geqslant \left\langle \nu^\infty, (\overrightarrow{p}\cdot I)^\infty\right\rangle \quad \lambda_\nu^s - \text{a.e. in } \bar{\Omega}.$$

With this, and by

$$D^s\tilde{u} = \langle \nu^\infty, I\rangle\, \lambda_\nu^s \ \text{ in } \mathcal{M}(\widetilde{\Omega}), \tag{4.66}$$

we have that

$$\left\langle \nu^\infty, (\overrightarrow{p}\cdot I)^\infty\right\rangle \lambda_\nu^s = \frac{1}{\alpha(u)} \left|\langle v^\infty, I\rangle\right| \lambda_\nu^s = \frac{1}{\alpha(u)} \left|D^s\tilde{u}\right| \ \text{ in } \mathcal{M}(\widetilde{\Omega}).$$

On the other hand, by (4.66), we have that

$$\langle \nu, \overrightarrow{p} \rangle \langle \nu^{\infty}, I \rangle \lambda_{\nu}^{s} = \langle \nu, \overrightarrow{p} \rangle D^{s}\tilde{u} = \langle \nu, \overrightarrow{p} \rangle \langle \delta_{p}, I \rangle |D^{s}\tilde{u}|,$$

where $p = \frac{D^{s}\tilde{u}}{|D^{s}\tilde{u}|} \in L^{1}(\overline{\Omega}, |D^{s}\tilde{u}|)$. Then,

$$\begin{aligned}
\langle \nu, \overrightarrow{p} \rangle \langle \delta_{p}, I \rangle &= \int_{\mathbb{R}^{N}} \int_{\partial \mathbb{B}^{N}} \overrightarrow{p}(B) \cdot A \mathrm{d}\delta_{p}(A) \mathrm{d}\nu(B) \\
&= \int_{\mathbb{R}^{N}} \int_{\partial \mathbb{B}^{N}} \frac{A \cdot B}{1 + \alpha(u)|B|} \mathrm{d}\delta_{p}(A) \mathrm{d}\nu(B) \\
&\leqslant \int_{\mathbb{R}^{N}} \int_{\partial \mathbb{B}^{N}} \frac{|A|}{\alpha(u)} \mathrm{d}\delta_{p}(A) \mathrm{d}\nu(B) \\
&= \frac{1}{\alpha(u)},
\end{aligned}$$

with $|D^{s}\tilde{u}|$ − a.e. in $\overline{\Omega}$. Thus,

$$\begin{aligned}
\frac{1}{\alpha(u)} |D^{s}\tilde{u}| &\geqslant \langle \nu, \overrightarrow{p} \rangle \langle \delta_{p}, I \rangle |D^{s}\tilde{u}| = \langle \nu, \overrightarrow{p} \rangle \langle \nu^{\infty}, I \rangle \lambda_{\nu}^{s} \\
&\geqslant \left\langle \nu^{\infty}, (\overrightarrow{p} \cdot I)^{\infty} \right\rangle \lambda_{\nu}^{s} = \frac{1}{\alpha(u)} |D^{s}\tilde{u}| \quad \text{in } \mathcal{M}(\widetilde{\Omega}),
\end{aligned}$$

i.e.,

$$\langle \nu, \overrightarrow{p} \rangle D^{s}\tilde{u} = \frac{1}{\alpha(u)} |D^{s}\tilde{u}| \quad \text{in } \mathcal{M}(\widetilde{\Omega}).$$

This implies that

$$\langle \nu, \overrightarrow{p} \rangle D^{s}\tilde{u} \llcorner \Omega = \langle \nu, \overrightarrow{p} \rangle D^{s}u \llcorner \Omega = \frac{1}{\alpha(u)} |D^{s}u| \quad \text{in} \quad \mathcal{M}(\bar{\Omega}) \tag{4.67}$$

and

$$\langle \nu, \overrightarrow{p} \rangle D^{s}\tilde{u} \llcorner \partial\Omega = \frac{1}{\alpha(u)} |D^{s}\tilde{u}| \llcorner \partial\Omega \text{ in } \mathcal{M}(\bar{\Omega}).$$

Since $u \in \mathrm{BV}(\Omega)$, by the boundary trace theorem (see Theorem 3.87 in [27]), we get that

$$D^{s}\tilde{u} \llcorner \partial\Omega = D^{s}u \llcorner \Omega = \left(-u^{\Omega} \otimes \vec{n}\right) \mathcal{H}^{N-1} \llcorner \partial\Omega \text{ in } \mathcal{M}(\bar{\Omega}).$$

Hence,

$$\langle \nu, \overrightarrow{p} \rangle \cdot \left(-u^{\Omega} \otimes \vec{n}\right) \mathcal{H}^{N-1} \llcorner \partial\Omega = \frac{1}{\alpha(u)} \left|u^{\Omega}\right| \mathcal{H}^{N-1} \llcorner \partial\Omega;$$

i.e.,

$$\left[\langle \nu, \overrightarrow{p} \rangle, \vec{n}\right] \in -\mathrm{sign}\left(u^{\Omega}\right) \cdot \frac{1}{\alpha(u)} \quad \mathcal{H}^{N-1}\text{-a.e. on } \partial\Omega. \tag{4.68}$$

Letting $z = \langle \nu, \overrightarrow{p} \rangle$, and combining (4.61), (4.62), (4.67) and (4.68), we complete the proof. □

## 5 Uniqueness Results

The uniqueness is proved in a narrower class of functions than the existence. However, since the proof of Theorem 5.1 is practically independent of the proof of existence, the regularity on $u_t$ is less restrictive.

**Theorem 5.1** There is at most one solution (in the sense of Definition 4.1 or Definition 4.2) to the problem $\mathcal{P}$ in the class

$$W(\Omega_T) = \{u(x,t) | u \in L^{2}(0,T; H_{0}^{1}(\Omega)), u_t \in L^{2}(0,T; H^{-1}(\Omega))\}.$$

**Proof** Actually, a solution in the class $W(\Omega_T)$ is also a weak solution in usual sense. Let both $u_1$ and $u_2$ be weak solutions of $\mathcal{P}$. Then one has that

$$\begin{aligned}\left\langle \frac{\partial u_1}{\partial t} - \frac{\partial u_2}{\partial t}, u_1 - u_2 \right\rangle &= -\int_\Omega \left( \frac{\nabla u_1}{1+\alpha(u_1)|\nabla u_1|} - \frac{\nabla u_2}{1+\alpha(u_2)|\nabla u_2|} \right) \cdot (\nabla u_1 - \nabla u_2)\,\mathrm{d}x \\ &= -\int_\Omega \left( \frac{\nabla u_1}{1+\alpha(u_1)|\nabla u_1|} - \frac{\nabla u_2}{1+\alpha(u_1)|\nabla u_2|} \right) \cdot (\nabla u_1 - \nabla u_2)\,\mathrm{d}x \\ &\quad + \int_\Omega \left( \frac{\nabla u_2}{1+\alpha(u_2)|\nabla u_2|} - \frac{\nabla u_2}{1+\alpha(u_1)|\nabla u_2|} \right) \cdot (\nabla u_1 - \nabla u_2)\,\mathrm{d}x \\ &:= -I_1 + I_2,\end{aligned}$$

where $\langle \cdot, \cdot \rangle$ denotes the dual product of $H^{-1}(\Omega)$ and $H_0^1(\Omega)$.

By some calculations, we get that

$$\begin{aligned}I_1 &= \int_\Omega \frac{|\nabla u_1 - \nabla u_2|^2}{(1+\alpha(u_1)|\nabla u_1|)\,(1+\alpha(u_1)|\nabla u_2|)}\mathrm{d}x \\ &\quad + \int_\Omega \frac{\alpha(u_1)\,(|\nabla u_1| + |\nabla u_2|)\,(|\nabla u_1||\nabla u_2| - \nabla u_1 \cdot \nabla u_2)}{(1+\alpha(u_1)|\nabla u_1|)\,(1+\alpha(u_1)|\nabla u_2|)}\mathrm{d}x \\ &\geqslant \int_\Omega \frac{|\nabla u_1 - \nabla u_2|^2}{(1+\alpha(u_1)|\nabla u_1|)\,(1+\alpha(u_1)|\nabla u_2|)}\mathrm{d}x,\end{aligned}$$

and

$$\begin{aligned}|I_2| &\leqslant \int_\Omega \frac{|\alpha(u_1) - \alpha(u_2)|\,|\nabla u_2|^2\,|\nabla u_1 - \nabla u_2|}{(1+\alpha(u_1)|\nabla u_2|)\,(1+\alpha(u_2)|\nabla u_2|)}\mathrm{d}x \\ &\leqslant \int_\Omega \frac{|\nabla u_1 - \nabla u_2|^2}{(1+\alpha(u_1)|\nabla u_1|)\,(1+\alpha(u_1)|\nabla u_2|)}\mathrm{d}x \\ &\quad + \frac{1}{4}\int_\Omega \frac{|\alpha(u_1) - \alpha(u_2)|^2 |\nabla u_2|^4\,(1+\alpha(u_1)|\nabla u_1|)}{(1+\alpha(u_2)|\nabla u_2|)^2\,(1+\alpha(u_1)|\nabla u_2|)}\mathrm{d}x.\end{aligned}$$

Collecting all of these facts, by $\alpha \geq c > 0$, we obtain that

$$\begin{aligned}&\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega (u_1(t) - u_2(t))^2\mathrm{d}x \\ \leqslant &\frac{1}{4c^3}\int_\Omega |\alpha(u_1) - \alpha(u_2)|^2\,|\nabla u_2|\,(1+\alpha(u_1)|\nabla u_1|)\,\mathrm{d}x \\ \leqslant &\frac{1}{4c^3}\,|\alpha(u_1) - \alpha(u_2)|^2_{C(\overline{\Omega})} \cdot \|\nabla u_2\|_{L^2(\Omega)} \cdot \|1+\alpha(u_1)|\nabla u_1|\|_{L^2(\Omega)}.\end{aligned}$$

Thus, by (4.26), we have that

$$Y'(t) \leqslant C\lambda(t)Y(t),$$

where $C$ is a constant and

$$\begin{aligned}Y(t) &= \|u_1(t) - u_2(t)\|^2_{L^2(\Omega)}, \\ \lambda(t) &= \|\nabla u_2(t)\|_{L^2(\Omega)} \cdot \|1 + |\nabla u_1(t)|\|_{L^2(\Omega)}.\end{aligned}$$

Observing that $\lambda(t) \in L^1(0,T)$, by Gronwall's inequality, it follows that

$$Y(t) \leqslant \exp(C\int_0^T \lambda(s)\mathrm{d}s) \cdot Y(0).$$

Since $Y(0) = \|u_1(0) - u_2(0)\|^2_{L^2(\Omega)} = 0$, we have that $u_1 = u_2$, and the proof is concluded. □

## 6 Numerical Scheme and Experiments

In this section, the PM scheme in [15] and a fast numerical scheme are used to implement our NLPM method. Furthermore, we compare our NLPM method with other denoising methods through experiments on different types of images. It is worth noting that, starting from this section, we introduce a positive modifier parameter $K$ into our NLPM diffusivity function (2.1), as the original PM diffusivity (1.1) did. More specifically, the modified diffusivity

$$g_{\mathrm{NLPM}} = \frac{1}{1 + |\nabla G_\sigma * u||\nabla u|/K^2}$$

will be used in our experiments. The reason we add the modifier $K$ is to enhance the flexibility of our NLPM method. In order to maintain the fairness of the comparison experiments, we add the modifier $K$ to all the other methods, and adjust the parameters in order for each to reach its best results. The analytical results we proved in the previous sections can be viewed as a special case of when $K = 1$. We will carefully discuss the selection of parameters in Section 6.3.

### 6.1 Numerical Schemes

(1) PM scheme (PMS)

Similarly to the numerical scheme in [15], which was originally introduced to implement the PM method, and thus called the "PM scheme", the scheme that follows is used for our NLPM method.

First, the gradient terms of $u$ and $G_\sigma * u$ are respectively discretized into four orthogonal directions: north, south, east and west. Assume that the width and length of the images are $I$ and $J$, respectively, and also assume that the spatial step sizes are $h_x = h_y = 1$.

For $0 \le i \le I, 0 \le j \le J$,

$$\begin{aligned}
&\nabla_W u_{i,j} = u_{i-1,j} - u_{i,j},\ \nabla_E u_{i,j} = u_{i+1,j} - u_{i,j},\\
&\nabla_N u_{i,j} = u_{i,j+1} - u_{i,j},\ \nabla_S u_{i,j} = u_{i,j-1} - u_{i,j},\\
&\nabla_W [G_\sigma * u]_{i,j} = [G_\sigma * u]_{i-1,j} - [G_\sigma * u]_{i,j},\\
&\nabla_E [G_\sigma * u]_{i,j} = [G_\sigma * u]_{i+1,j} - [G_\sigma * u]_{i,j},\\
&\nabla_N [G_\sigma * u]_{i,j} = [G_\sigma * u]_{i,j+1} - [G_\sigma * u]_{i,j},\\
&\nabla_S [G_\sigma * u]_{i,j} = [G_\sigma * u]_{i,j-1} - [G_\sigma * u]_{i,j}.
\end{aligned}$$

Then, we denote $\Lambda \in \Omega = \{N, S, E, W\}$ as any of the four directions, and define $C^n_{\Lambda i,j}$ as

$$C^n_{\Lambda i,j} = \frac{1}{1 + \left|\nabla_\Lambda [G_\sigma * u]_{i,j}\right| |\nabla_\Lambda u_{i,j}| / K^2}.$$

Finally, the NLPM equation is put into a simple form as follows:

$$u^{n+1}_{i,j} = u^n_{i,j} + \tau \sum_{\Lambda \in \Omega} C^n_{\Lambda i,j} \nabla_\Lambda u^n_{i,j}, \tag{6.1}$$

$$u^0_{i,j} = f_{i,j} = f(i,j),$$

$$u^n_{i,0} = u^n_{i,1},\ \ u^n_{0,j} = u^n_{1,j},\ \ u^n_{I,j} = u^n_{I-1,j},\ \ u^n_{i,J} = u^n_{i,J-1}.$$

Here $\tau$ is unit time step, and $0 \le \tau \le \frac{1}{4}$ is required to ensure that the scheme is stable.

Furthermore, if we denote $U^n \in \mathbb{R}^{IJ}$ as the vector with entries $u^n_{i,j}$, then we can rewrite equation (6.1) as the matrix form

$$U^{n+1} = (I + \tau A(U^n))U^n, \tag{6.2}$$

where $I$ is a $IJ \times IJ$ identity matrix, and $A(U^n)$ with respect to $U^n$ is a negative semidefinite $IJ \times IJ$ matrix derived from the diffusion process at time level $n$.

The PMS is a very simple and intuitive scheme to implement our new model, however since PMS is an explicit scheme with a fixed time step size, it suffers from a strict time step restriction which will cause a severe inefficiency. Therefore, we use another numerical scheme which can realize a rapid implementation for our NLPM method.

(2) Fast explicit diffusion scheme (FED)

According to Grewenig et al. [28], splitting a long fixed time step iteration into several variant time step cycles (i.e., the FED cycles) will speed up the implementation, as well as maintain the stability. More specifically, the corresponding FED scheme for our NLPM model is as follows (given that the default spatial step size $h$ in image processing is 1):

1) Input: total processing time $T$, the number of FED cycles $M$, noisy image $f$.

2) Initialization: let $U^0$ be the vector generalized by $f$.

3) Iteration: for $k = 0, \cdots, M-1$,

a) Find the current diffusion matrix $P$, where $P = A(U^k)$, according to equation (6.2).

b) Find the largest module $\mu_{\mathrm{m}}$ of eigenvalues of $P$ (i.e., $\mu_{\mathrm{m}} = \max_i |\mu_i|$, where $\mu_i$ is the eigenvalue of $P$).

c) Find the smallest $n$ satisfying that

$$\frac{2(n^2+n)}{3\mu_{\mathrm{m}}} \geq \frac{T}{M}.$$

d) Compute

$$\tilde{\tau}_i = \frac{3T}{2M(n^2+n)\cos^2\left(\pi\frac{2i+1}{4n+2}\right)} \quad \text{for } i = 0, \cdots, n-1.$$

e) Reorder $\tilde{\tau}_i$, for $i = 0, \cdots, n-1$, using Leja ordering [29] for the sake of numerical stability. For the simplicity of notation, we still use $\{\tilde{\tau}_i\}_{i=0}^{n-1}$ to denote the reordered time step sizes.

f) Do one FED cycle:

$$U^{k+1} = \prod_{i=0}^{n-1}(I + \tilde{\tau}_i P)U^k.$$

End

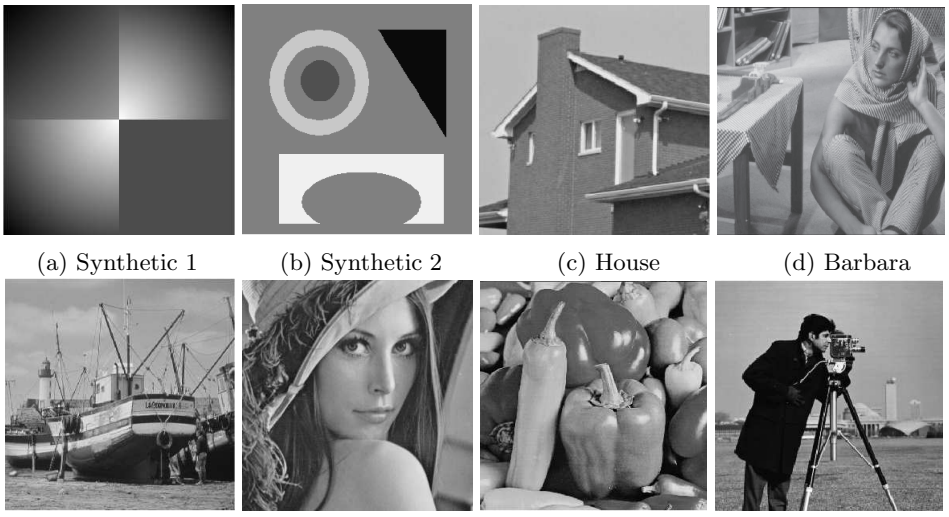
### 6.2 Numerical Experiments

We use peak signal-to-noise ratio (PSNR) to measure the effectiveness of different image denoising methods:

$$\mathrm{PSNR} = 10\log_{10}\frac{MN|\max u_o - \min u_o|^2}{||\hat{u} - u_o||_2^2}(\mathrm{dB}).$$

Here $u_o$ is the noise-free image, $\hat{u}$ is the denoised image, and $M$ and $N$ are the width and length of the images, respectively.

First we compare the results of our NLPM method implemented by two schemes introduced in Section 6.1. We use these two methods (i.e., NLPM implemented by the PMS scheme, and NLPM implemented by the FED scheme) to process the images shown in Figure 5. For each method, the iteration is stopped when the maximal PSNR value is attained. The PSNR results, processing time and corresponding parameters are listed in Table 1. We can find that the FED scheme sharply reduces the processing time of the PMS scheme, however, the PSNR of the FED scheme drops by around $0.2 - 0.5$ dB; this is a compromise of fast implementation.



(a) Synthetic 1 (b) Synthetic 2 (c) House (d) Barbara

(e) Boat (f) Lena (g) Pepper (h) Cameraman

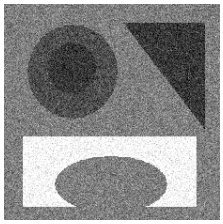
Figure 5 Test images used in comparison experiments of the PMS and FED schemes

**Tabld 1** Test results between the PMS and FED schemes. All of the test images are from Figure 5. $\tau_{\mathrm{PMS}}$ is the time step size of the PMS scheme, $Iter_{\mathrm{PMS}}$ is the total iterations of the PMS scheme, $T_{\mathrm{FED}}$ is the total processing time of the FED scheme, $M_{\mathrm{FED}}$ is the number of FED cycles, $\sigma_{\mathrm{NLPM}}$ is the parameters of the Gaussian in our NLPM method

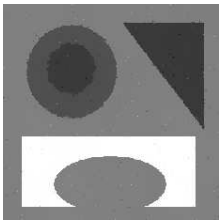
| Test image | PMS time | FED time | PMS PSNR | FED PSNR | $\tau_{\mathrm{PMS}}$ | $Iter_{\mathrm{PMS}}$ | $T_{\mathrm{FED}}$ | $M_{\mathrm{FED}}$ | $\sigma_{\mathrm{NLPM}}$ |
|---|---|---|---|---|---|---|---|---|---|
| Synthetic 1 | 0.4779 | 0.1097 | 39.8452 | 39.9370 | 0.1 | 353 | 35 | 6 | 3 |
| Synthetic 2 | 2.1384 | 0.4702 | 41.1160 | 40.9414 | 0.1 | 550 | 55 | 9 | 3 |
| House | 1.5085 | 0.2790 | 30.3166 | 30.1993 | 0.1 | 450 | 45 | 8 | 3 |
| Barbara | 2.8479 | 0.3147 | 27.8142 | 27.3395 | 0.05 | 350 | 17 | 5 | 1 |
| Boat | 6.6273 | 0.7257 | 31.3769 | 31.2265 | 0.05 | 750 | 75 | 10 | 1 |
| Lena | 2.8780 | 0.3339 | 31.5354 | 31.2055 | 0.05 | 300 | 15 | 3 | 3 |
| Pepper | 8.5388 | 0.6982 | 30.0658 | 30.0089 | 0.05 | 950 | 47 | 8 | 3 |
| Cameraman | 1.8377 | 0.2232 | 30.3965 | 30.2106 | 0.05 | 700 | 35 | 10 | 1 |

Furthermore, to verify the effectiveness of our NLPM method, we perform comparison experiments among several different image denoising methods. In the experiments, we care more about PSNR comparison for the different methods rather than the processing time. In order to make a fair comparison between different methods, we uniformly use the simple explicit finite difference schemes for implementation (i.e., PMS for NLPM). The corresponding parameters are set to reach the maximal PSNR for each method.

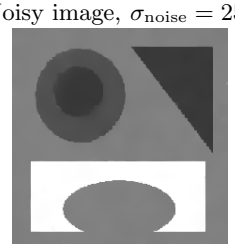
We first add Gaussian noise with a standard deviation 25 to a clean synthetic image, as shown in Figure 6(a). Then Figures 6(b)–6(d) show the denoising results of PM, RPM and our NLPM method, respectively. To make a detailed comparison, we zoom on the restored images to compare the detail performance of our NLPM method with the other two diffusion methods. Figures 7(a) and 7(e) are restored images of the PM and NLPM methods, respectively. We box three pairs of local blocks in these, and display in pairs the rest of the subfigures of Figure 7. Figures 7(b) and 7(c) demonstrate that there exists a "staircase" effect in PM restored images, while this kind of oscillation is well avoided by the NLPM method, as shown in Figures7(f) and 7(e). Figures 7(d) and 7(h) demonstrate that NLPM method can avoid black and white "speckle" effect occurring in PM restored images. Figure 8 compares the performance of the RPM and NLPM methods. From Figures 8(b) and 8(c) we can see that, in the edge areas, the RPM method has serious blurring; the contrast of two adjacent colors becomes much less obvious. Figures 8(e) and 8(f) demonstrate that the NLPM method can preserve the edge area much better.



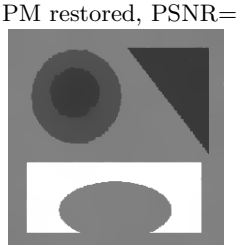
(a) Noisy image, $\sigma_{\text{noise}} = 25$



(b) PM restored, PSNR= 36.96

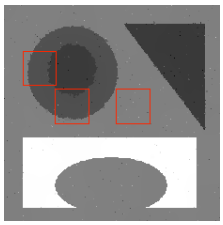


(c) RPM restored, PSNR= 37.78



(d) NLPM restored, PSNR= 39.11

Figure 6 Restored results of three different diffusion methods

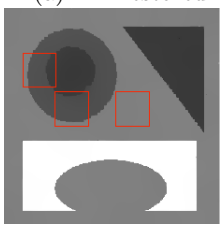


(a) PM restored

(b) PM detail 1

(c) PM detail 2

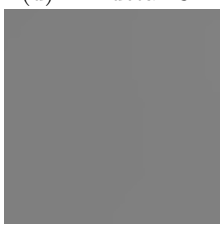
(d) PM detail 3



(e) NLPM restored

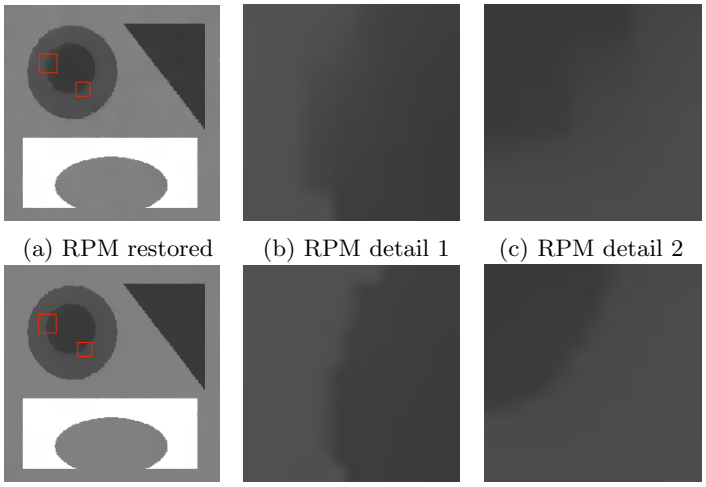
(f) NLPM detail 1

(g) NLPM detail 2



(h) NLPM detail 3

Figure 7 Detailed comparisons of restored images by the PM method and the NLPM method



(a) RPM restored (b) RPM detail 1 (c) RPM detail 2

(d) NLPM restored (e) NLPM detail 1 (f) NLPM detail 2

Figure 8 Detailed comparisons of restored images by the RPM method and the NLPM method

Our next examples compare the performance of the proposed NLPM method with more PDE-based image denoising methods, namely, the total variation method (TV) [1], the PM method, the RPM method, the mild regularized PM method (M-RPM) [21], and the adaptive PM method (we use the same notation "D-$\alpha$-PM" as the original article [30]). All of the parameters are chosen to be optimal based on the corresponding articles, also, the iteration is stopped when maximal PSNR value is reached. The best PSNR results of these six different PDE-based methods are listed in Table 2, and the test images are chosen from Figure 5. The highest PSNR value of each experiment is shown in bold face. For various types of images, our NLPM method demonstrates the best denoising ability among these six PDE-based methods.

**Table 2** The optimal PSNR results of different PDE-based methods

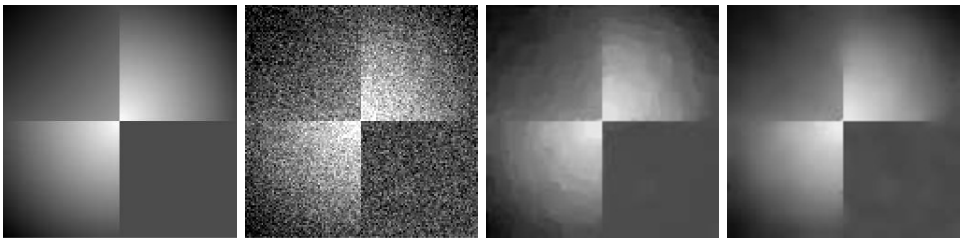
| Image | TV | PM | RPM | M-RPM | D-$\alpha$-PM | NLPM |
|---|---|---|---|---|---|---|
| Synthetic1 | 34.68532959 | 36.93236781 | 35.60422723 | 33.39559309 | 37.2309 | **37.7621703** |
| House | 31.40358253 | 30.87583707 | 31.75864565 | 31.0057527 | 31.60348162 | **31.8226579** |
| Lena | 28.59233379 | 28.4242997 | 28.78708987 | 27.98096444 | 28.55191082 | **28.82056526** |
| Pepper | 30.28443428 | 29.99339522 | 30.57299094 | 29.46266412 | 30.2640025 | **30.61255026** |

The visual results are presented in Figures 9–12, with the original image, the noisy image contaminated by additive Gaussian noise, and the images denoised by different methods shown in the respective subfigures. One notices that there are several random black and white speckles appearing in the images denoised by the PM method; that is because of the forward-backward diffusion. The RPM method loses the preservation of edge lines (such as the junction lines of the four squares in Figure 9(e)), due to the high regularization of its solutions. In the smooth intensity transition areas (such as the first three quadrants of "Synthetic 1" in Figure 9(a), the cheeks of "Lena" in Figure 11(a), and the smooth surface of "Pepper" in Figure 12(a)), the TV method has an obvious "staircase" effect as shown in Figures 9(c), 11(c), and 12(c), while the M-RPM method exposes a more serious "staircase" effect in the corresponding experiments, demonstrated especially in Figure 9(f). This is because the M-RPM method depends too much on the parameter $p$ in terms of its diffusivity (1.3), and the fixed parameter $p$ results in a lack

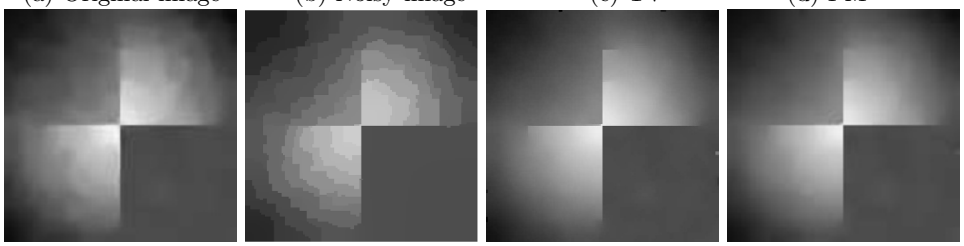
of flexibility regarding its diffusion process. The D-$\alpha$-PM method, specifically,

$$u_t = \operatorname{div}\left(\frac{\nabla u}{1+(|\nabla u|/K)^{2-\frac{2}{1+k|\nabla G_\sigma * u|^2}}}\right) - \lambda(u-f), \tag{6.3}$$

adopts an adaptive diffusivity which can effectively improve the drawbacks of the M-RPM method (as shown in Figures 9(g), 11(g) and 12(g)). However, the D-$\alpha$-PM method has too many parameters (namely, $K$, $k$, $\sigma$, $\lambda$ in (6.3), the time step $\tau$, and the iteration times) to modify, which increases the difficulty of implementation and limits the application of the D-$\alpha$-PM method. Our NLPM method prevents the shortcomings mentioned above. Specifically, the NLPM method preserves the edges well (Figures 9(g) and 10(h)), restores the smooth transitions well, and avoids the "staircase" artifacts (Figures 9(h), 11(h) and (12)(h)). It also presents visual effects as good as the D-$\alpha$-PM method while having less parameters to modify.

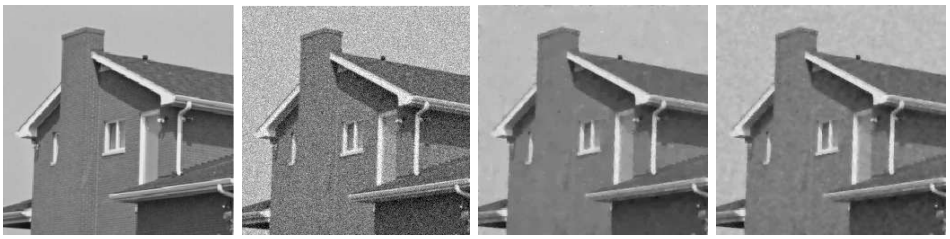


(a) Original image (b) Noisy image (c) TV (d) PM



(e) RPM (f) M-RPM (g) D-$\alpha$-PM (h) NLPM

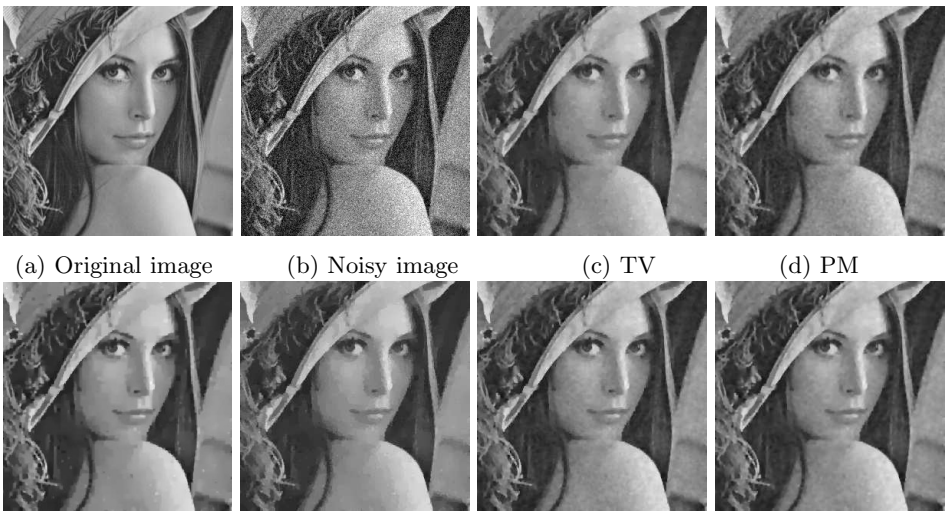
Figure 9 Experiments on "Synthetic 1", of size $128 \times 128$, the noise standard deviation being 30



(a) Original image (b) Noisy image (c) TV (d) PM

(e) RPM (f) M-RPM (g) D-$\alpha$-PM (h) NLPM

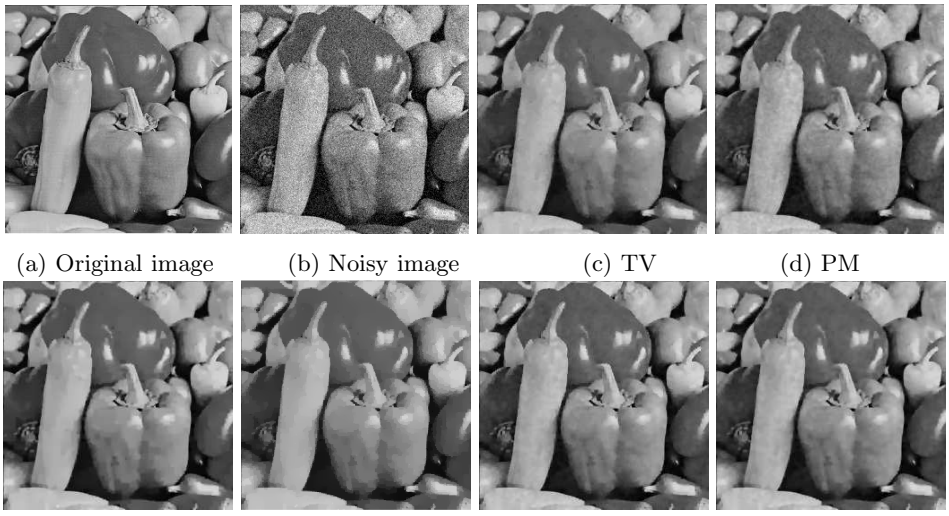
Figure 10 Experiments on "House", of size $256 \times 256$, the noise standard deviation being 20



(a) Original image (b) Noisy image (c) TV (d) PM

(e) RPM (f) M-RPM (g) D-$\alpha$-PM (h) NLPM

Figure 11 Experiments on "Lena", of size $300 \times 300$, the noise standard deviation being 25



(a) Original image (b) Noisy image (c) TV (d) PM

(e) RPM (f) M-RPM (g) D-$\alpha$-PM (h) NLPM

Figure 12 Experiments on "Pepper", of size $256 \times 256$, the noise standard deviation being 20

### 6.3 Selection of Parameters

There are two main parameters, $K$ and $\sigma$, in our NLPM method that need to be modified in the experiments.

The parameter $K$ is a threshold value of diffusion progress; it has more influence on the boundary and edge areas than the smooth areas. Consequently, there are differences of the optimal $K$ among different types of images. As for the images with fewer details, such as the "Synthetic 1" and "House" in Figures 9(a) and 10(a), the optimal $K$ is smaller and lies between 3 and 5. For the images with more details, i.e., the real natural images such as "Lena" and "Pepper" in Figures 11(a) and 12(a), the optimal $K$ is supposed to be larger; around $7 - 9$. On the other hand, the noise level will also affect the optimal value of $K$. Specifically, images

with larger noise levels will have a larger optimal $K$, which is reasonable, since the noise will increase the "details" in the image.

The parameter $\sigma$, as discussed in Section 2, represents the non-locality level of diffusion progress. Throughout the experiments, we found that $\sigma = 1$ is an optimal choice for different types of images. On the other hand, to ensure the numerical stability of the PMS scheme, the time step $\tau$ is required to be less than 0.25. Also, according to our empirical evidence, the PSNR result will not increase after $\tau$ gets small enough. Therefore, we choose $\tau = 0.02$ for all experiments to both satisfy the stability requirements and preserve the time efficiency. Finally, for the number of iteration times, we apply the maximal PSNR criterion, i.e., stop the iteration when the maximal PSNR is reached. Therefore, the number of iteration times is fixed if all of the other parameters are given. In general, the number of iteration times is in proportion to noise level, and is inversely proportional to $K$.

## References


[1] Rudin L I, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena, 1992, **60**(1/4): 259–268

[2] Vese L. A study in the BV space of a denoising–deblurring variational problem. Applied Mathematics and Optimization, 2001, **44**(2): 131–161

[3] Andreu F, Caselles V, Mazón J M. Existence and uniqueness of a solution for a parabolic quasilinear problem for linear growth functionals with $L^1$ data. Mathematische Annalen, 2002, **322**(1): 139–206

[4] Andreu F, Ballester C, Caselles V, et al. Minimizing total variation flow. Differential and Integral Equations, 2001, **14**(3): 321–360

[5] Chan T F, Golub G H, Mulet P. A nonlinear primal-dual method for total variation-based image restoration. SIAM Journal on Scientific Computing, 1999, **20**(6): 1964–1977

[6] Chan T F, Mulet P. On the convergence of the lagged diffusivity fixed point method in total variation image restoration. SIAM Journal on Numerical Analysis, 1999, **36**(2): 354–367

[7] Chambolle A. An algorithm for total variation minimization and applications. Journal of Mathematical Imaging and Vision, 2004, **20**(1/2): 89–97

[8] Goldstein T, Osher S. The split bregman method for $L^1$-regularized problems. SIAM Journal on Imaging Sciences, 2009, **2**(2): 323–343

[9] Chan T, Marquina A, Mulet P. High-order total variation-based image restoration. SIAM Journal on Scientific Computing, 2000, **22**(2): 503–516

[10] Lysaker M, Lundervold A, Tai X C. Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time. IEEE Transactions on Image Processing, 2003, **12**(12): 1579–1590

[11] Chan T F, Esedoglu S, Park F. A fourth order dual method for staircase reduction in texture extraction and image restoration problems//2010 IEEE International Conference on Image Processing. IEEE, 2010: 4137–4140

[12] Pu Y F, Zhou J L, Yuan X. Fractional differential mask: a fractional differential-based approach for multiscale texture enhancement. IEEE Transactions on Image Processing, 2010, **19**(2): 491–511

[13] Pu Y, Wang W, Zhou J, et al. Fractional differential approach to detecting textural features of digital image and its fractional differential filter implementation. Science in China Series F: Information Sciences, 2008, **51**(9): 1319–1339

[14] Bai J, Feng X C. Fractional-order anisotropic diffusion for image denoising. IEEE Transactions on Image Processing, 2007, **16**(10): 2492–2502

[15] Perona P, Malik J. Scale-space and edge detection using anisotropic diffusion. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1990, **12**(7): 629–639

[16] Catté F, Lions P L, Morel J M, et al. Image selective smoothing and edge detection by nonlinear diffusion. SIAM Journal on Numerical Analysis, 1992, **29**(1): 182–193

[17] Guidotti P, Lambers J V. Two new nonlinear nonlocal diffusions for noise reduction. Journal of Mathematical Imaging and Vision, 2009, **33**(1): 25–37

[18] Guidotti P. A new nonlocal nonlinear diffusion of image processing. Journal of Differential Equations, 2009, **246**(12): 4731–4742

[19] Guidotti P. A family of nonlinear diffusions connecting perona-malik to standard diffusion. Discrete & Continuous Dynamical Systems-Series S, 2012, **5**(3): 581–590

[20] Guidotti P. A backward–forward regularization of the Perona–Malik equation. Journal of Differential Equations, 2012, **252**(4): 3226–3244

[21] Guidotti P, Kim Y, Lambers J. Image restoration with a new class of forward-backward-forward diffusion equations of Perona–Malik type with applications to satellite image enhancement. SIAM Journal on Imaging Sciences, 2013, **6**(3): 1416–1444

[22] Chen Y, Zhang K. Young measure solutions of the two-dimensional Perona-Malik equation in image processing. Communications on Pure and Applied Analysis, 2006, **5**(3): 615–635

[23] Anzellotti G. Pairings between measures and bounded functions and compensated compactness. Annali di Matematica Pura ed Applicata, 1983, **135**: 293–318

[24] Kristensen J, Rindler F. Characterization of generalized gradient Young measures generated by sequences in $W^{1,1}$ and BV. Archive for Rational Mechanics & Analysis, 2010, **197**(2): 539–598

[25] Rindler F. A local proof for the characterization of Young measures generated by sequences in BV. Journal of Functional Analysis, 2014, **266**(11): 6335–6371

[26] Simon J. Compact sets in the space $L^p(0,T;B)$. Annali di Matematica Pura ed Applicata, 1986, **146**(1): 65–96

[27] Ambrosio L, Fusco N, Pallara D. Functions of Bounded Variation and Free Discontinuity Problems. New York: The Clarendon Press, Oxford University Press, 2000: 180–183

[28] Grewenig S, Weickert J, Bruhn A. From box filtering to fast explicit diffusion//Joint Pattern Recognition Symposium. Berlin: Springer, 2010: 533–542

[29] Calvetti D, Reichel L. Adaptive Richardson iteration based on Leja points. Journal of Computational and Applied Mathematics, 1996, **71**(2): 267–286

[30] Guo Z, Sun J, Zhang D, et al. Adaptive Perona-Malik model based on the variable exponent for image denoising. IEEE Transactions on Image Processing, 2012, **21**(3): 958–967

[31] Kong L, Huan Z, Guo B. BV solutions to a degenerate parabolic equation for image denoising. Acta Mathematica Scientia, 2007, **27B**(1): 169–179