# Geometry of total variation regularized $L^{p}$-model 

Yuying Shi ${ }^{\text {a,* }}$, Li-Lian Wang ${ }^{\text {b }}$, Xue-Cheng Tai ${ }^{\text {b,c }}$<br>${ }^{\text {a }}$ Department of Mathematics and Physics, North China Electric Power University, Beijing, 102206, China<br>${ }^{\mathrm{b}}$ Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, 637371, Singapore<br>${ }^{\text {c }}$ Department of Mathematics, University of Bergen, Norway

## ARTICLE INFO

## Article history:

Received 17 October 2010

## Keywords:

TV-L ${ }^{p}$ model
G-norm
Feature selection
Feature separation
Augmented Lagrangian methods


#### Abstract

In this paper, the geometry and scale selection properties of the total variation (TV) regularized $L^{p}$-model are rigorously analyzed. Some intrinsic features different from the TV-L ${ }^{1}$ model are derived and demonstrated. Numerical algorithms based on recent developed augmented Lagrangian methods are implemented and numerical results consistent with the theoretical results are provided.


© 2011 Elsevier B.V. All rights reserved.

## 1. Introduction

The fundamental task of image denoising is to restore a noise-free image $u$ from an observed, noise polluted image $f$, that is, to remove noisy component from $f$. This problem can be modeled by $f=u+n$, where $n$ is the unknown noise. A general way to obtain this decomposition is to solve the problem

$$
\begin{equation*}
\inf _{u \in X}\{J(u)+\lambda H(u-f)\} \tag{1}
\end{equation*}
$$

where $J(\cdot)$ and $H(\cdot)$ are two non-negative functionals defined over a suitable functional space $X$, and $\lambda>0$ is a parameter to balance the two terms. Various models with different combinations of $\{X ; H, J\}$ have been proposed for image restoration and/or decomposition. The seminal ROF model [1], referred to as the total-variation regularized $L^{2}$-model or $T V-L^{2}$ model, takes the form

$$
\begin{equation*}
\inf _{u \in B V(\Omega)}\left\{\operatorname{TV}(u)+\frac{\lambda}{2}\|u-f\|_{L^{2}(\Omega)}^{2}\right\} . \tag{2}
\end{equation*}
$$

Here, the total variation of the function $u(\boldsymbol{x}) \in L_{\text {loc }}^{1}(\Omega)$ is defined to be [2,3]:

$$
\begin{equation*}
\operatorname{TV}(u)=\int_{\Omega}|\nabla u| d \boldsymbol{x}:=\sup _{\boldsymbol{p} \in \delta} \int_{\Omega} u(\boldsymbol{x}) \operatorname{div} \boldsymbol{p}(\boldsymbol{x}) d \boldsymbol{x} \tag{3}
\end{equation*}
$$

where $s=\left\{\boldsymbol{p} \in C_{c}^{1}\left(\Omega ; R^{2}\right):|\boldsymbol{p}(\boldsymbol{x})| \leq 1, \forall \boldsymbol{x} \in R^{2}\right\}$, and the space $B V(\Omega)=\left\{u \in L^{1}(\Omega):\|u\|_{B V}<\infty\right\}$ with the norm $\|u\|_{B V}:=\|u\|_{L^{1}(\Omega)}+\operatorname{TV}(u)$. As an important variant, the $T V-L^{1}$ model (see, e.g., [4-7]) has also been used for denoising and cartoon-texture decomposition. In contrast to the ROF model, it has some interesting properties like morphological invariance and feature extraction by scale. Interestingly, some finer TV-regularized models with $G$ (cf. [8]), $H^{-1}$ (cf. [9]),

[^0]$\operatorname{div}(B M O)$ (cf. [10]) and Besov-type (cf. [11]) fidelity term have also been proposed for image decomposition. We also point out that $[12,13]$ considered some generalized fidelity function for a better denoising effect. On the other hand, some models using regularization term other than TV, e.g., the LLT model [14] and the model with $J(u):=\int_{\Omega}|\nabla u|^{\gamma} d \boldsymbol{x}$ with $\gamma>1$ (cf. [15]), have been adopted to overcome the stair-case effect of the ROF model.

In the last two decades, extensive numerical study has been devoted to these models, but it is also worthwhile to provide some theoretical insights for a better understanding of the models. Using the notion of $G$-norm $\|\cdot\|_{*}$, Meyer [8] showed that for a given image $f$, if $\lambda>\|f\|_{*}^{-1}$, then the minimizer $u_{\lambda}$ of the ROF model meets $\left\|f-u_{\lambda}\right\|_{*}=\lambda^{-1}$. This observation inspired several subsequential works. For instance, Osher and Scherzer [16] extended this analysis to a more general class of regression models. Scherzer et al. [17] generalized the concept of $G$-norms, i.e., $G$-sets, which is used to characterize minimizers of non-differentiable regularization functionals. Tadmor et al. [18] proposed a multi-scale image decomposition based on a hierarchical representation for image features. Moreover, a more subtler analysis has been conducted for the $T V-L^{1}$ model in $[6,19]$. Chan and Esedoglu [6] showed some geometric properties of the $T V-L^{1}$ model with implications for scale selection and multi-scale decomposition. Yin et al. [19] provided an analytic expression of scale based on the $G$-value [17].

Motivated by [6,19], we consider in this paper the geometric properties and scale separation of the TV-L ${ }^{p}$ model with $0<p<+\infty$ :

$$
\begin{equation*}
\min _{u \in B V(\Omega)}\left\{T V L_{\lambda}^{p}(u):=\operatorname{TV}(u)+\frac{\lambda}{p} \int_{\Omega}|f-u|^{p} d \boldsymbol{x}\right\}, \quad \lambda>0 \tag{4}
\end{equation*}
$$

and discuss the numerical algorithms for this general model. We first formulate the cost functional $T V L_{\lambda}^{p}(\cdot)$ into an equivalent geometry energy in terms of upper level sets, which provides some insights for a better understanding of the original model. More importantly, by using the notion of $G$-norm (for $p>1$ ), we characterize the range of $\lambda$ that allows to extract geometric features of a given scale. Different from the $G$-value for the $T V-L^{1}$ setting, the $G$-norm depends on both the scale and intensity values of the features. As a result, given two features with the same scale and different intensities, the $T V-L^{1}$ model fails to distinguish them, but the $T V-L^{p}(p>1)$ model is capable of extracting them. We provide a rigorous theoretical proof and numerical evidences of such a property. Accordingly, this study could be useful for object recognition and image segmentation. Indeed, the ideas and techniques used in this paper can be applied to a much wider class of minimization problems with a convex regularization term.

The rest of the paper is organized as follows. In Section 2, the $T V-L^{p}(0<p<+\infty)$ model is formulated as an equivalent geometry problem in terms of upper level sets, and some properties of the geometry problem are derived. In Section 3, the properties of minimizers of the $T V-L^{p}(1<p<+\infty)$ model are provided. In particular, based on the $G$-norm, the scale selection of features is well studied. In Section 4, numerical algorithms based on augmented Lagrangian algorithm for the $T V-L^{p}(1 \leq p<+\infty)$ model are introduced and some numerical results consistent with the analysis are given, followed by concluding remarks in Section 5.

## 2. Geometric properties of the TV-L ${ }^{p}$ model

This section aims to study the behavior of the $T V-L^{p}$ model. Motivated by [6,19,17], we start with reformulating the model as a geometry problem, and then present some properties of the underlying geometry problem. This provides some new insights of this model with general $p$.

For simplicity of presentation, we assume $\Omega=\mathbb{R}^{2}$ for the moment, and recall that the perimeter of a Borel measurable set $\Sigma \subset \Omega$ is defined by

$$
\begin{equation*}
\operatorname{Per}(\Sigma):=\operatorname{TV}\left(\mathbf{1}_{\Sigma}\right)=\int_{\Omega}\left|\nabla \mathbf{1}_{\Sigma}\right| d \boldsymbol{x} \tag{5}
\end{equation*}
$$

where $\mathbf{1}_{\Sigma}$ is the characteristic function of $\Sigma$. For a given image $f$ and any $\xi \in \mathbb{R}$, we define the upper level set

$$
\begin{equation*}
\Gamma_{f}:=\Gamma_{f}(\xi)=\{\boldsymbol{x} \in \Omega: f(\boldsymbol{x})>\xi\} \tag{6}
\end{equation*}
$$

For clarity, we particularly denote $\Sigma_{u}(\xi):=\Gamma_{u}(\xi)$, where $u$ is the restored image. We shall also use the co-area formula (see, e.g., $[20,3]$ ):

$$
\begin{equation*}
\operatorname{TV}(u)=\int_{-\infty}^{+\infty} \operatorname{Per}\left(\Sigma_{u}(\xi)\right) d \xi=\int_{-\infty}^{+\infty} \operatorname{TV}\left(\mathbf{1}_{\Sigma_{u}(\xi)}\right) d \xi \tag{7}
\end{equation*}
$$

The following important result shows that the $T V-L^{p}$ functional $(0<p<\infty)$ in (4) can be expressed in an equivalent form in terms of the level sets.

Theorem 2.1. Define

$$
\begin{equation*}
C\left(\Sigma_{u}, \Gamma_{f}\right):=\operatorname{Per}\left(\Sigma_{u}\right)+\lambda\left(\int_{\Sigma_{u} \backslash \Gamma_{f}}(\xi-f)^{p-1} d \boldsymbol{x}+\int_{\Gamma_{f} \backslash \Sigma_{u}}(f-\xi)^{p-1} d \boldsymbol{x}\right) \tag{8}
\end{equation*}
$$

Then for all $0<p<\infty$,

$$
\begin{equation*}
\operatorname{TV}(u)+\frac{\lambda}{p}\|u-f\|_{L^{p}}^{p}=\int_{-\infty}^{+\infty} C\left(\Sigma_{u}, \Gamma_{f}\right) d \xi . \tag{9}
\end{equation*}
$$

Proof. We have that

$$
\begin{aligned}
\|u-f\|_{L^{p}}^{p} & =\int_{\{u>f\}}(u-f)^{p} d \boldsymbol{x}+\int_{\{f>u\}}(f-u)^{p} d \boldsymbol{x} \\
& =p\left[\int_{\{u>f\}} \int_{f}^{u}(\xi-f)^{p-1} d \xi d \boldsymbol{x}+\int_{\{f>u\}} \int_{u}^{f}(f-\xi)^{p-1} d \xi d \boldsymbol{x}\right] . \\
& =p\left[\int_{\{u>f\}} \int_{-\infty}^{+\infty}(\xi-f)^{p-1} \mathbf{1}_{[f, u]} d \xi d \boldsymbol{x}+\int_{\{f>u\}} \int_{-\infty}^{+\infty}(f-\xi)^{p-1} \mathbf{1}_{[u, f]} d \xi d \boldsymbol{x}\right] \\
& =p\left[\int_{\Omega} \int_{-\infty}^{+\infty}(\xi-f)^{p-1} \mathbf{1}_{[f, u]} \mathbf{1}_{\Sigma_{u} \backslash \Gamma_{f}} d \xi d \mathbf{x}+\int_{\Omega} \int_{-\infty}^{+\infty}(f-\xi)^{p-1} \mathbf{1}_{[u, f]} \mathbf{1}_{\Gamma_{f} \backslash \Sigma_{u}} d \xi d \boldsymbol{x}\right] \\
& =p\left[\int_{-\infty}^{+\infty} \int_{\Omega}(\xi-f)^{p-1} \mathbf{1}_{[f, u]} \mathbf{1}_{\Sigma_{u} \backslash \Gamma_{f}} d \boldsymbol{x} d \xi+\int_{-\infty}^{+\infty} \int_{\Omega}(f-\xi)^{p-1} \mathbf{1}_{[u, f]} \mathbf{1}_{\Gamma_{f} \backslash \Sigma_{u}} d \boldsymbol{x} d \xi\right] .
\end{aligned}
$$

Here, $\mathbf{1}_{[a, b]}$ is the characteristic function of the interval $[a, b] \subset \mathbb{R}$. Observe that

$$
\int_{\Omega} \mathbf{1}_{[u, f]} \mathbf{1}_{\Sigma_{u} \backslash \Gamma_{f}} d \boldsymbol{x}=\int_{\Sigma_{u} \backslash \Gamma_{f}} d \boldsymbol{x}
$$

and likewise for $\mathbf{1}_{[u, f]} \mathbf{1}_{\Gamma_{f} \backslash \Sigma_{u}}$. Hence, we obtain

$$
\begin{equation*}
\|u-f\|_{L^{p}}^{p}=p \int_{-\infty}^{+\infty}\left[\int_{\Sigma_{u} \backslash \Gamma_{f}}(\xi-f)^{p-1} d \boldsymbol{x}+\int_{\Gamma_{f} \backslash \Sigma_{u}}(f-\xi)^{p-1} d \boldsymbol{x}\right] d \xi . \tag{10}
\end{equation*}
$$

A combination of (7) and (10) leads to the desired result.
It is important to point out that for a given image $f$, the functional $C\left(\Sigma, \Gamma_{f}\right)$ in (8) is well-defined for any Lebesgue measurable subset $\Sigma \subseteq \Omega$. In particular, when $\Sigma=\Sigma_{u}$, the equivalence (9) holds. This suggests that we consider the geometry minimization problem for each level set:

$$
\begin{equation*}
\min _{\Sigma} C\left(\Sigma, \Gamma_{f}(\xi)\right) \text { for given } \xi \in \mathbb{R} \text { and } f . \tag{11}
\end{equation*}
$$

Next, we present two fundamental properties of the functional C. Hereafter, for $\xi_{1}<\xi_{2}$, we denote $\Gamma_{f}^{i}:=\Gamma_{f}\left(\xi_{i}\right)(i=1,2)$ as defined in (6). It is clear that $\Gamma_{f}^{1} \supseteq \Gamma_{f}^{2}$. The functional $C$ enjoys the following "monotone" property.

Theorem 2.2. Assume that $\Sigma^{1}$ and $\Sigma^{2}$ are the minimizers of (11) corresponding to $\Gamma_{f}=\Gamma_{f}^{1}$ and $\Gamma_{f}=\Gamma_{f}^{2}$, respectively, and have finite perimeters. Then we have

$$
\begin{equation*}
C\left(\Sigma^{2}, \Gamma_{f}^{1}\right)-C\left(\Sigma^{1} \cap \Sigma^{2}, \Gamma_{f}^{1}\right) \geq C\left(\Sigma^{1} \cup \Sigma^{2}, \Gamma_{f}^{1}\right)-C\left(\Sigma^{1}, \Gamma_{f}^{1}\right) \geq 0 \tag{12}
\end{equation*}
$$

and

$$
\begin{equation*}
0 \geq C\left(\Sigma^{2}, \Gamma_{f}^{2}\right)-C\left(\Sigma^{1} \cap \Sigma^{2}, \Gamma_{f}^{2}\right) \geq C\left(\Sigma^{2}, \Gamma_{f}^{1}\right)-C\left(\Sigma^{1} \cap \Sigma^{2}, \Gamma_{f}^{1}\right) . \tag{13}
\end{equation*}
$$

Proof. We first prove (12). By the definition (8),

$$
\begin{align*}
& C\left(\Sigma^{2}, \Gamma_{f}^{1}\right)-C\left(\Sigma^{1} \cap \Sigma^{2}, \Gamma_{f}^{1}\right)-C\left(\Sigma^{1} \cup \Sigma^{2}, \Gamma_{f}^{1}\right)+C\left(\Sigma^{1}, \Gamma_{f}^{1}\right)  \tag{14}\\
& \quad=\operatorname{Per}\left(\Sigma^{2}\right)-\operatorname{Per}\left(\Sigma^{1} \cap \Sigma^{2}\right)-\operatorname{Per}\left(\Sigma^{1} \cup \Sigma^{2}\right)+\operatorname{Per}\left(\Sigma^{1}\right)  \tag{15}\\
& \quad+\lambda\left[\left(\int_{\Sigma^{2} \backslash \Gamma_{f}^{1}}(\xi-f)^{p-1} d \boldsymbol{x}-\int_{\left(\Sigma^{1} \cap \Sigma^{2}\right) \backslash \Gamma_{f}^{1}}(\xi-f)^{p-1} d \boldsymbol{x}\right)\right.  \tag{16}\\
& \left.\quad-\left(\int_{\left(\Sigma^{1} \cup \Sigma^{2}\right) \backslash \Gamma_{f}^{1}}(\xi-f)^{p-1} d \boldsymbol{x}-\int_{\Sigma^{1} \backslash \Gamma_{f}^{1}}(\xi-f)^{p-1} d \boldsymbol{x}\right)\right] \tag{17}
\end{align*}
$$

$$
\begin{align*}
& +\lambda\left[\left(\int_{\Gamma_{f}^{1} \backslash \Sigma^{2}}(f-\xi)^{p-1} d \boldsymbol{x}-\int_{\Gamma_{f}^{1} \backslash\left(\Sigma^{1} \cap \Sigma^{2}\right)}(f-\xi)^{p-1} d \boldsymbol{x}\right)\right.  \tag{18}\\
& \left.-\left(\int_{\Gamma_{f}^{1} \backslash\left(\Sigma^{1} \cup \Sigma^{2}\right)}(f-\xi)^{p-1} d \boldsymbol{x}-\int_{\Gamma_{f}^{1} \backslash \Sigma^{1}}(f-\xi)^{p-1} d \boldsymbol{x}\right)\right] . \tag{19}
\end{align*}
$$

One verifies readily that

$$
\left(\Sigma^{2} \backslash \Gamma_{f}^{1}\right) \backslash\left(\left(\Sigma^{1} \cap \Sigma^{2}\right) \backslash \Gamma_{f}^{1}\right)=\left(\Sigma^{2} \backslash \Sigma^{1}\right) \backslash \Gamma_{f}^{1}
$$

and

$$
\left(\left(\Sigma^{1} \cup \Sigma^{2}\right) \backslash \Gamma_{f}^{1}\right) \backslash\left(\Sigma^{1} \backslash \Gamma_{f}^{1}\right)=\left(\Sigma^{2} \backslash \Sigma^{1}\right) \backslash \Gamma_{f}^{1}
$$

Thus the summation in the square brackets in (16) and (17) vanishes. Similarly, as $\Sigma^{2} \supseteq\left(\Sigma^{1} \cap \Sigma^{2}\right)$, we have

$$
\left(\Gamma_{f}^{1} \backslash\left(\Sigma^{1} \cap \Sigma^{2}\right)\right) \backslash\left(\Gamma_{f}^{1} \backslash \Sigma^{2}\right)=\left(\Gamma_{f}^{1} \cap \Sigma^{2}\right) \backslash \Sigma^{1}
$$

and

$$
\left(\Gamma_{f}^{1} \backslash \Sigma^{1}\right) \backslash\left(\Gamma_{f}^{1} \backslash\left(\Sigma^{1} \cup \Sigma^{2}\right)\right)=\left(\Gamma_{f}^{1} \cap \Sigma^{2}\right) \backslash \Sigma^{1}
$$

which imply that the summation in (18) and (19) is zero. Hence, a combination of the above facts leads to

$$
\begin{align*}
& C\left(\Sigma^{2}, \Gamma_{f}^{1}\right)-C\left(\Sigma^{1} \cap \Sigma^{2}, \Gamma_{f}^{1}\right)-C\left(\Sigma^{1} \cup \Sigma^{2}, \Gamma_{f}^{1}\right)+C\left(\Sigma^{1}, \Gamma_{f}^{1}\right) \\
& \quad=\operatorname{Per}\left(\Sigma^{2}\right)-\operatorname{Per}\left(\Sigma^{1} \cap \Sigma^{2}\right)-\operatorname{Per}\left(\Sigma^{1} \cup \Sigma^{2}\right)+\operatorname{Per}\left(\Sigma^{1}\right) \tag{20}
\end{align*}
$$

Therefore, the formula (12) follows from the property of the Per-functional (see, e.g., [21]):

$$
\operatorname{Per}\left(\Sigma^{2}\right)+\operatorname{Per}\left(\Sigma^{1}\right) \geq \operatorname{Per}\left(\Sigma^{1} \cap \Sigma^{2}\right)+\operatorname{Per}\left(\Sigma^{1} \cup \Sigma^{2}\right)
$$

and the assumption that $\Sigma^{1}$ is the minimizer with $\Gamma_{f}=\Gamma_{f}^{1}$.
Now, we turn to the derivation of (13). Similarly, by (12),

$$
\begin{align*}
& C\left(\Sigma^{2}, \Gamma_{f}^{2}\right)-C\left(\Sigma^{1} \cap \Sigma^{2}, \Gamma_{f}^{2}\right)-C\left(\Sigma^{2}, \Gamma_{f}^{1}\right)+C\left(\Sigma^{1} \cap \Sigma^{2}, \Gamma_{f}^{1}\right)  \tag{21}\\
&= {\left[\operatorname{Per}\left(\Sigma^{2}\right)-\operatorname{Per}\left(\Sigma^{1} \cap \Sigma^{2}\right)\right]-\left[\operatorname{Per}\left(\Sigma^{2}\right)-\operatorname{Per}\left(\Sigma^{1} \cap \Sigma^{2}\right)\right] }  \tag{22}\\
&+\lambda\left[\left(\int_{\Sigma^{2} \backslash \Gamma_{f}^{2}}(\xi-f)^{p-1} d \boldsymbol{x}-\int_{\left(\Sigma^{1} \cap \Sigma^{2}\right) \backslash \Gamma_{f}^{2}}(\xi-f)^{p-1} d \boldsymbol{x}\right)\right.  \tag{23}\\
&\left.-\left(\int_{\Sigma^{2} \backslash \Gamma_{f}^{1}}(\xi-f)^{p-1} d \boldsymbol{x}-\int_{\left(\Sigma^{1} \cap \Sigma^{2}\right) \backslash \Gamma_{f}^{1}}(\xi-f)^{p-1} d \boldsymbol{x}\right)\right]  \tag{24}\\
&+\lambda\left[\left(\int_{\Gamma_{f}^{2} \backslash \Sigma^{2}}(f-\xi)^{p-1} d \boldsymbol{x}-\int_{\Gamma_{f}^{2} \backslash\left(\Sigma^{1} \cap \Sigma^{2}\right)}(f-\xi)^{p-1} d \boldsymbol{x}\right)\right.  \tag{25}\\
&\left.-\left(\int_{\Gamma_{f}^{1} \backslash \Sigma^{2}}(f-\xi)^{p-1} d \boldsymbol{x}-\int_{\Gamma_{f}^{1} \backslash\left(\Sigma^{1} \cap \Sigma^{2}\right)}(f-\xi)^{p-1} d \boldsymbol{x}\right)\right] . \tag{26}
\end{align*}
$$

Obviously, (22) is zero, and for $i=1,2$,

$$
\begin{equation*}
\left(\Sigma^{2} \backslash \Gamma_{f}^{i}\right) \backslash\left(\left(\Sigma^{1} \cap \Sigma^{2}\right) \backslash \Gamma_{f}^{i}\right)=\left(\Sigma^{2} \backslash \Sigma^{1}\right) \backslash \Gamma_{f}^{i} \tag{27}
\end{equation*}
$$

and

$$
\begin{equation*}
\left(\Gamma_{f}^{i} \backslash\left(\Sigma^{1} \cap \Sigma^{2}\right)\right) \backslash\left(\Gamma_{f}^{i} \backslash \Sigma^{2}\right)=\left(\Gamma_{f}^{i} \cap \Sigma^{2}\right) \backslash \Sigma^{1} \tag{28}
\end{equation*}
$$

By (27), the summation in (23) and (24) becomes

$$
\lambda\left[\int_{\left(\Sigma^{2} \backslash \Sigma^{1}\right) \backslash \Gamma_{f}^{2}}(\xi-f)^{p-1} d \boldsymbol{x}-\int_{\left(\Sigma^{2} \backslash \Sigma^{1}\right) \backslash \Gamma_{f}^{1}}(\xi-f)^{p-1} d \boldsymbol{x}\right] \geq 0
$$

where we have used the fact: $\Gamma_{f}^{1} \supseteq \Gamma_{f}^{2}$ and $\left(\Sigma^{2} \backslash \Sigma^{1}\right) \backslash \Gamma_{f}^{2} \supseteq\left(\Sigma^{2} \backslash \Sigma^{1}\right) \backslash \Gamma_{f}^{1}$. Similarly, we can show that the summation in (25) and (26) is nonnegative. Hence, (13) follows from the above facts and the assumption that $\Sigma^{2}$ is the minimizer with $\Gamma_{f}=\Gamma_{f}^{2}$.

An important consequence of the above property is as follows.
Corollary 2.1. Assuming that $\Sigma^{1}$ and $\Sigma^{2}$ are the minimizers of (11) corresponding to $\Gamma_{f}=\Gamma_{f}^{1}$ and $\Gamma_{f}=\Gamma_{f}^{2}$, respectively. If $\Gamma_{f}^{1} \supseteq \Gamma_{f}^{2}$,i.e., $\xi_{1}<\xi_{2}$, then the same inclusion holds for a pair of minimizers.

Proof. As a direct consequence of Theorem 2.2, we have

$$
\begin{aligned}
0 & \geq C\left(\Sigma^{2}, \Gamma_{f}^{2}\right)-C\left(\Sigma^{1} \cap \Sigma^{2}, \Gamma_{f}^{2}\right) \geq C\left(\Sigma^{2}, \Gamma_{f}^{1}\right)-C\left(\Sigma^{1} \cap \Sigma^{2}, \Gamma_{f}^{1}\right) \\
& \geq C\left(\Sigma^{1} \cup \Sigma^{2}, \Gamma_{f}^{1}\right)-C\left(\Sigma^{1}, \Gamma_{f}^{1}\right) \geq 0
\end{aligned}
$$

Thus, all inequalities above hold as equalities, that is, $\Sigma^{1} \cap \Sigma^{2}$ and $\Sigma^{1} \cup \Sigma^{2}$ are minimizers of (11) with $\Gamma=\Gamma_{f}^{2}$ and $\Gamma=\Gamma_{f}^{1}$, respectively. It is clear that the former is included in the latter.

It is seen that for $p=1$, the above formulas have simpler representations, and similar analysis for the $T V-L^{1}$ model can be found in [6,19]. Moreover, we may apply an analogous argument in [19] to construct a minimizer $u$ of the original $T V-L^{p}$ model from a series of minimizers of $C\left(\Sigma, \Gamma_{f}\right)$.

## 3. Properties of the minimizers

In this section, we study properties of the minimizers of the $T V-L^{p}$ model. The purpose is to provide some quantitative guidelines for the selection of the parameter $\lambda$, which allows for extracting and separating different scales and intensities.

The first result indicates a close relation between the minimizer and the parameter.
Theorem 3.1. For $1<p<+\infty$ and $\lambda>0, u_{\lambda}$ is a minimizer of the $T V-L^{p}$ model (4), if and only if $\lambda$ satisfies

$$
\begin{equation*}
\lambda \int_{\Omega}\left|u_{\lambda}-f\right|^{p-2}\left(u_{\lambda}-f\right) h d x \geq T V\left(u_{\lambda}\right)-T V\left(u_{\lambda}+h\right), \quad \forall h \in B V(\Omega) \tag{29}
\end{equation*}
$$

Proof. If $u_{\lambda}$ is a minimizer of (4), we have that for any $h \in B V(\Omega)$ and $\epsilon>0$,

$$
\begin{align*}
T V\left(u_{\lambda}\right)+\frac{\lambda}{p} \int_{\Omega}\left|u_{\lambda}-f\right|^{p} d \boldsymbol{x} \leq & T V\left(u_{\lambda}+\epsilon h\right)+\frac{\lambda}{p} \int_{\Omega}\left|u_{\lambda}+\epsilon h-f\right|^{p} d \boldsymbol{x} \\
\leq & T V\left(u_{\lambda}+\epsilon h\right)+\frac{\lambda}{p} \int_{\Omega}\left|u_{\lambda}-f\right|^{p} d \boldsymbol{x} \\
& +\lambda \epsilon \int_{\Omega}\left|u_{\lambda}-f\right|^{p-2}\left(u_{\lambda}-f\right) h d \boldsymbol{x}+\lambda \epsilon \eta\left(\epsilon ; p, u_{\lambda}, h\right) \tag{30}
\end{align*}
$$

where $\eta\left(\epsilon ; p, u_{\lambda}, h\right)=o(\epsilon)$. This implies

$$
\begin{equation*}
-\lambda \int_{\Omega}\left|u_{\lambda}-f\right|^{p-2}\left(u_{\lambda}-f\right) h d \boldsymbol{x} \leq \frac{T V\left(u_{\lambda}+\epsilon h\right)-T V\left(u_{\lambda}\right)}{\epsilon}+\lambda \eta\left(\epsilon ; p, u_{\lambda}, h\right) \tag{31}
\end{equation*}
$$

The convexity of $T V(u)$ implies

$$
\frac{T V\left(u_{\lambda}+\epsilon h\right)-T V\left(u_{\lambda}\right)}{\epsilon} \leq T V\left(u_{\lambda}+h\right)-T V\left(u_{\lambda}\right), \quad 0<\epsilon<1
$$

Hence, letting $\epsilon \rightarrow 0$ in (31) leads to

$$
-\lambda \int_{\Omega}\left|u_{\lambda}-f\right|^{p-2}\left(u_{\lambda}-f\right) h d x \leq T V\left(u_{\lambda}+h\right)-T V\left(u_{\lambda}\right) .
$$

Therefore, $\lambda$ satisfies (29).
Conversely, by (29) and the convexity of $|\cdot|^{p}(1<p)$, we obtain that

$$
\begin{align*}
T V\left(u_{\lambda}+h\right)+\frac{\lambda}{p} \int_{\Omega}\left|u_{\lambda}+h-f\right|^{p} d \boldsymbol{x} \geq & T V\left(u_{\lambda}\right)+\frac{\lambda}{p} \int_{\Omega}\left|u_{\lambda}-f\right|^{p} d \boldsymbol{x} \\
& +\lambda \int_{\Omega}\left|u_{\lambda}-f\right|^{p-2}\left(u_{\lambda}-f\right) h d \boldsymbol{x}+T V\left(u_{\lambda}+h\right)-T V\left(u_{\lambda}\right) \\
\geq & T V\left(u_{\lambda}\right)+\frac{\lambda}{p} \int_{\Omega}\left|u_{\lambda}-f\right|^{p} d \boldsymbol{x} . \tag{32}
\end{align*}
$$

This ends the proof.

Remark 1. The property (29) is valid for $1<p<\infty$. For $p=1$, we refer to [17] the following result.
For $f \in L^{1}(\Omega), u_{\lambda}$ is the minimizer of the $T V-L^{1}$ model, if and only if $\lambda$ satisfies that for any $h \in B V(\Omega)$,

$$
\begin{equation*}
\lambda\left(-\int_{u_{\lambda} \neq f} \operatorname{sign}\left(u_{\lambda}-f\right) h d \boldsymbol{x}-\int_{u_{\lambda}=f}|h| d \boldsymbol{x}\right) \leq T V\left(u_{\lambda}+h\right)-T V(h) \tag{33}
\end{equation*}
$$

Indeed, we can view (33) as a limiting case of (29) (i.e., $p \rightarrow 1^{+}$). We find that the derivation of (33) in [17] is different from that of (29).

Remark 2. If the regularization term $T V(u)$ in (1) is replaced by

$$
J_{\gamma}(u):=\int_{\Omega}|\nabla u|^{\gamma} d \boldsymbol{x}, \quad \gamma>1
$$

(refer to, e.g., $[15,22,23]$ for the applications of such models), then Theorem 3.1 still holds with $J_{\gamma}(\cdot)$ in place of $T V(\cdot)$.
For a fixed $\lambda$, the $T V-L^{p}$ model returns an image $u_{\lambda}$ with certain features. In some applications, it is interesting to choose some $\lambda$, so as to extract desirable feature. Such a critical value is characterized by the so-called $G$-norm of the underlying feature.

We first recall the definition of the $G$-norm (see, e.g., [8,24]).
For $v=\operatorname{div} \boldsymbol{g}$ with $\boldsymbol{g}=\left(g_{1}, g_{2}\right)$ and $g_{i} \in L^{\infty}(\Omega), i=1,2$, the $G$-norm of $v$ is defined as

$$
\begin{equation*}
G(v):=\|v\|_{G}=\inf _{g}\left\|\sqrt{g_{1}^{2}+g_{2}^{2}}\right\|_{L^{\infty}} \tag{34}
\end{equation*}
$$

If $\Omega$ is a bounded connected open domain of $\mathbb{R}^{2}$ with a Lipschitz boundary, $\boldsymbol{g} \cdot \boldsymbol{n}=0$ should be imposed on $\partial \Omega$, where $\boldsymbol{n}$ is the unit outer normal. As shown in [25], the $G$-norm is equivalent to

$$
\begin{equation*}
G(v)=\sup _{h \in B V(\Omega), h \neq 0} \frac{\int_{\Omega} v(\boldsymbol{x}) h(\boldsymbol{x}) d \boldsymbol{x}}{T V(h)} \tag{35}
\end{equation*}
$$

We have the following important result.
Theorem 3.2. $u_{\lambda}=0$ is a minimizer of the TV-L ${ }^{p}$ model (4) with $1<p<+\infty$, if and only if

$$
\begin{equation*}
0<\lambda \leq \frac{1}{G\left(|f|^{p-1}\right)} \tag{36}
\end{equation*}
$$

Proof. It follows from Theorem 3.1 that $u_{\lambda}=0$ is a minimizer if and only if

$$
-\lambda \int_{\Omega}|f|^{p-2} f h d x \geq-T V(h) \Leftrightarrow \lambda \int_{\Omega}|f|^{p-1} h d x \leq T V(h), \quad \forall h \in B V(\Omega)
$$

where we used the fact that the (pixel) intensity of $f \geq 0$. In view of (35), we have $\lambda G\left(|f|^{p-1}\right) \leq 1$, so (36) follows.
Remark 3. For $p=1$, we refer to $[17,19]$ the following result.
Let $\partial|f|$ be the set-valued sub-derivative of $|f|$, i.e.,

$$
\partial|f|(x)= \begin{cases}\operatorname{sign}(f), & \text { if } f(x) \neq 0  \tag{37}\\ {[-1,1],} & \text { if } f(x)=0\end{cases}
$$

Then, $u_{\lambda}=0$ is the minimizer of the $T V-L^{1}$ model if and only if $\lambda \leq \frac{1}{G(\partial|f|)}$, where $(G(\partial|f|))$ is called as the $G$-value of $f$.
Remark 4. As a simple illustration, we assume that the observed image $f=\mathbf{1}_{B_{r}(0)}$, where $B_{r}(0) \subset \Omega$ is a disk centered at the origin with radius $r$. Notice that

$$
G\left(\left|\mathbf{1}_{B_{r}(0)}\right|^{p-1}\right)=G\left(\left|\mathbf{1}_{B_{r}(0)}\right|\right)=\frac{r}{2}
$$

Theorem 3.2 indicates that $B_{r}(0)$ vanishes when $\lambda \leq \frac{2}{r}$. In the forthcoming section, we shall provide numerical results to verify this theoretical result.

The next result shows the feature separation capability of the $T V-L^{p}$ model, which follows from Theorems 3.1 and 3.2.
Corollary 3.1. Let $f$ be an observed image defined in $\Omega$, which consists of two separated (with some distance) piecewise constant features in the disjoint subregions $\Omega_{1}$ and $\Omega_{2}$ over a black background, that is, $f=f_{\Omega_{1}} \mathbf{1}_{\Omega_{1}}+f_{\Omega_{2}} \mathbf{1}_{\Omega_{2}}$, where $f_{\Omega_{i}}$ is the restriction
of $f$ in $\Omega_{i}(i=1,2)$, and likewise for $u_{\Omega_{i}}$ below. Suppose that $G\left(\left|f_{\Omega_{2}}\right|^{p-1}\right)>G\left(\left|f_{\Omega_{1}}\right|^{p-1}\right)$ and $u$ is a minimizer of TV-L model (4) with $1<p<+\infty$. If $\lambda$ is chosen such that

$$
\begin{equation*}
\frac{1}{G\left(\left|f_{\Omega_{2}}\right|^{p-1}\right)}<\lambda<\frac{1}{G\left(\left|f_{\Omega_{1}}\right|^{p-1}\right)} \tag{38}
\end{equation*}
$$

then we have $u_{\Omega_{1}}=0$ and $u_{\Omega_{2}} \neq 0$.
Proof. In this proof, let $0<\epsilon<1$ and $h$ be an arbitrary function in $B V(\Omega)$.
We first prove $u_{\Omega_{2}} \neq 0$ by contradiction. The convexity of $T V(u)$ and $|\cdot|^{p}(1<p<+\infty)$ implies

$$
\begin{equation*}
\epsilon T V(h) \geq T V\left(u_{\Omega \backslash \Omega_{2}}+\epsilon h\right)-T V\left(u_{\Omega \backslash \Omega_{2}}\right) \tag{39}
\end{equation*}
$$

and

$$
\begin{equation*}
\frac{\lambda}{p} \int_{\Omega_{2}}\left|f_{\Omega_{2}}\right|^{p} d \boldsymbol{x}-\frac{\lambda}{p} \int_{\Omega_{2}}\left|f_{\Omega_{2}}-\epsilon h\right|^{p} d \boldsymbol{x} \geq \epsilon \lambda \int_{\Omega_{2}}\left|f_{\Omega_{2}}-\epsilon h\right|^{p-2}\left(f_{\Omega_{2}}-\epsilon h\right) h d \boldsymbol{x} \tag{40}
\end{equation*}
$$

Since $u$ is a minimizer of (4), assuming $u_{\Omega_{2}}=0$ leads to

$$
\begin{align*}
& T V\left(u_{\Omega \backslash \Omega_{2}}\right)+\frac{\lambda}{p} \int_{\Omega_{2}}\left|f_{\Omega_{2}}\right|^{p} d \boldsymbol{x}+\frac{\lambda}{p} \int_{\Omega \backslash \Omega_{2}}\left|u_{\Omega \backslash \Omega_{2}}-f_{\Omega \backslash \Omega_{2}}\right|^{p} d \boldsymbol{x} \\
& \quad \leq T V\left(u_{\Omega \backslash \Omega_{2}}+\epsilon h\right)+\frac{\lambda}{p} \int_{\Omega_{2}}\left|f_{\Omega_{2}}-\epsilon h\right|^{p} d \boldsymbol{x}+\frac{\lambda}{p} \int_{\Omega \backslash \Omega_{2}}\left|u_{\Omega \backslash \Omega_{2}}-f_{\Omega \backslash \Omega_{2}}\right|^{p} d \boldsymbol{x} . \tag{41}
\end{align*}
$$

Hence, a direct consequence of (39)-(41) is

$$
\begin{aligned}
\epsilon T V(h) & \geq T V\left(u_{\Omega \backslash \Omega_{2}}+\epsilon h\right)-T V\left(u_{\Omega \backslash \Omega_{2}}\right) \\
& \geq \frac{\lambda}{p} \int_{\Omega_{2}}\left|f_{\Omega_{2}}\right|^{p} d \boldsymbol{x}-\frac{\lambda}{p} \int_{\Omega_{2}}\left|f_{\Omega_{2}}-\epsilon h\right|^{p} d \boldsymbol{x} \\
& \geq \epsilon \lambda \int_{\Omega_{2}}\left|f_{\Omega_{2}}-\epsilon h\right|^{p-2}\left(f_{\Omega_{2}}-\epsilon h\right) h d \boldsymbol{x}
\end{aligned}
$$

That is,

$$
\begin{equation*}
T V(h) \geq \lambda \int_{\Omega_{2}}\left|f_{\Omega_{2}}-\epsilon h\right|^{p-2}\left(f_{\Omega_{2}}-\epsilon h\right) h d \boldsymbol{x} \tag{42}
\end{equation*}
$$

Letting $\epsilon \rightarrow 0$ in (42) yields

$$
\begin{equation*}
T V(h) \geq \lambda \int_{\Omega_{2}}\left|f_{\Omega_{2}}\right|^{p-1} h d \boldsymbol{x}, \quad \forall h \in B V(\Omega) \tag{43}
\end{equation*}
$$

where we used the fact $f_{\Omega_{2}} \geq 0$. In view of the definition (35) and the condition (38), there exists $0 \neq h_{0} \in B V(\Omega)$ such that

$$
\begin{equation*}
\lambda \int_{\Omega_{2}}\left|f_{\Omega_{2}}\right|^{p-1} h_{0} d \boldsymbol{x}>T V\left(h_{0}\right), \tag{44}
\end{equation*}
$$

which contradicts to (43). Therefore, $u_{\Omega_{2}} \neq 0$.
Now, we turn to proving $u_{\Omega_{1}}=0$ by contradiction. As a preparation, we first show that the intensity of the minimizer $u$ in the region of the background, denoted by $u_{\Omega_{b}}$, is zero, where we denote $\Omega_{b}=\Omega \backslash\left(\Omega_{1} \cup \Omega_{2}\right)$ and $\Omega_{b}^{c}=\Omega_{1} \cup \Omega_{2}$. Indeed, by the definition of the total variation (3),

$$
\begin{align*}
T V\left(u_{\Omega_{b}}+u_{\Omega_{b}^{c}}\right) & =\sup _{p \in \delta}\left\{\int_{\Omega} u_{\Omega_{b}} \operatorname{div} \boldsymbol{p} d \boldsymbol{x}+\int_{\Omega} u_{\Omega_{b}^{c}} \operatorname{div} \boldsymbol{p} d \boldsymbol{x}\right\} \\
& \geq \sup _{\boldsymbol{p} \in \delta} \int_{\Omega} u_{\Omega_{b}^{c}} \operatorname{div} \boldsymbol{p} d \boldsymbol{x}=T V\left(u_{\Omega_{b}^{c}}\right) \tag{45}
\end{align*}
$$

If $u_{\Omega_{b}} \neq 0$, then we have from the fact $u$ is a minimizer that

$$
\begin{align*}
& T V\left(u_{\Omega_{b}}+u_{\Omega_{b}^{c}}\right)+\frac{\lambda}{p} \int_{\Omega}\left|u_{\Omega_{b}}-f_{\Omega_{b}}\right|^{p} d \boldsymbol{x}+\frac{\lambda}{p} \int_{\Omega}\left|u_{\Omega_{b}^{c}}-f_{\Omega_{b}^{c}}\right|^{p} d \boldsymbol{x} \\
& \quad \leq T V\left(u_{\Omega_{b}^{c}}\right)+\frac{\lambda}{p} \int_{\Omega}\left|f_{\Omega_{b}}\right|^{p} d \boldsymbol{x}+\frac{\lambda}{p} \int_{\Omega}\left|u_{\Omega_{b}^{c}}-f_{\Omega_{b}^{c}}\right|^{p} d \boldsymbol{x} \tag{46}
\end{align*}
$$

In view of $f_{\Omega_{b}}=0$, (46) implies

$$
T V\left(u_{\Omega_{b}}+u_{\Omega_{b}^{c}}\right)+\frac{\lambda}{p} \int_{\Omega}\left|u_{\Omega_{b}^{c}}\right|^{p} d \boldsymbol{x} \leq T V\left(u_{\Omega_{b}^{c}}\right)
$$

which contradicts to (45). Thus the minimizer $u$ of (4) must satisfy $u_{\Omega_{b}}=0$.
Now, we are ready to prove $u_{\Omega_{1}}=0$ by contradiction again. By the convexity of $|\cdot|^{p}(1<p<+\infty)$,

$$
\begin{equation*}
\frac{\lambda}{p} \int_{\Omega_{1}}\left|f_{\Omega_{1}}\right|^{p} d \boldsymbol{x}-\frac{\lambda}{p} \int_{\Omega_{1}}\left|f_{\Omega_{1}}-h\right|^{p} d \boldsymbol{x} \leq \lambda \int_{\Omega_{1}}\left|f_{\Omega_{1}}\right|^{p-1} h d \boldsymbol{x} \tag{47}
\end{equation*}
$$

Assume that $u$ is a minimizer of (4) with $u_{\Omega_{1}} \neq 0$. Using the facts that $u_{\Omega_{b}}=0$ and $\Omega_{1}, \Omega_{2}$ are separated with some distance, we obtain

$$
\begin{align*}
T V L_{\lambda}^{p}\left(u_{\Omega \backslash \Omega_{1}}\right) \geq & T V L_{\lambda}^{p}\left(u_{\Omega \backslash \Omega_{1}}+u_{\Omega_{1}}\right) \\
= & \int_{\Omega \backslash \Omega_{1}}\left|\nabla u_{\Omega_{\Omega} \backslash \Omega_{1}}\right| d \boldsymbol{x}+\int_{\Omega_{1}}\left|\nabla u_{\Omega_{1}}\right| d \boldsymbol{x} \\
& +\frac{\lambda}{p} \int_{\Omega \backslash \Omega_{1}}\left|u_{\Omega \backslash \Omega_{1}}-f_{\Omega_{\backslash \Omega_{1}}}\right|^{p} d \boldsymbol{x}+\frac{\lambda}{p} \int_{\Omega_{1}}\left|f_{\Omega_{1}}-u_{\Omega_{1}}\right|^{p} d \boldsymbol{x} \\
= & T V L_{\lambda}^{p}\left(u_{\Omega \backslash \Omega_{1}}\right)+\int_{\Omega_{1}}\left|\nabla u_{\Omega_{1}}\right| d \boldsymbol{x}+\frac{\lambda}{p} \int_{\Omega_{1}}\left|f_{\Omega_{1}}-u_{\Omega_{1}}\right|^{p} d \boldsymbol{x}-\frac{\lambda}{p} \int_{\Omega_{1}}\left|f_{\Omega_{1}}\right|^{p} d \boldsymbol{x}, \tag{48}
\end{align*}
$$

which implies

$$
\begin{equation*}
\int_{\Omega_{1}}\left|\nabla u_{\Omega_{1}}\right| d x \leq \frac{\lambda}{p} \int_{\Omega_{1}}\left|f_{\Omega_{1}}\right|^{p} d \boldsymbol{x}-\frac{\lambda}{p} \int_{\Omega_{1}}\left|f_{\Omega_{1}}-u_{\Omega_{1}}\right|^{p} d \boldsymbol{x} . \tag{49}
\end{equation*}
$$

Thus, by (47) and (49),

$$
\begin{equation*}
T V\left(u_{\Omega_{1}}\right)=\int_{\Omega_{1}}\left|\nabla u_{\Omega_{1}}\right| d \boldsymbol{x} \leq \frac{\lambda}{p} \int_{\Omega_{1}}\left|f_{\Omega_{1}}\right|^{p-1} u_{\Omega_{1}} d \boldsymbol{x} \tag{50}
\end{equation*}
$$

On the other hand, we obtain from the definition (35) and the condition (38) that

$$
\begin{equation*}
T V(h)>\lambda \int_{\Omega_{1}}\left|f_{\Omega_{1}}\right|^{p-1} h d \boldsymbol{x}, \quad \forall h \in B V(\Omega) \tag{51}
\end{equation*}
$$

Taking $h=u_{\Omega_{1}}$ in the above inequality leads to a contradiction. Therefore, $u_{\Omega_{1}}=0$.

## 4. Algorithms and numerical results

In this section, we discuss numerical methods for solving the $T V-L^{p}$ model so as to provide some numerical justifications of the theoretical results.

There have been quite a few types of numerical methods proposed for the TV-L ${ }^{2}$ (i.e., ROF) model, which roughly can be classified into the categories: gradient descent method [1], the lagged diffusivity fixed-point iteration [26], the dual approach [27], the graph-cuts method [28], the Bregman iteration [29], the FTVd method [30] and the multigrid method [31]. It is interesting to point out that as shown in [32-34], the dual approach, the operator splitting, and the Bregman iteration can be derived from the augmented Lagrangian method. However, there is much less work for the $T V-L^{1}$ model as the $L^{1}$-fidelity term induces nonlinearity and non-differentiability. [35] proposed an operator splitting method, and [36,19,37] showed the numerical results using the commercial optimization package Mosek (cf. http://www.mosek.com) formulating the $T V-L^{1}$ model as second-order cone program (SOCP) [38,39]. The G-norm can be computed in an SOCP as in [37].

In the sequel, we adopt the augmented Lagrangian method [34], and describe the algorithm below.

### 4.1. Augmented Lagrangian method for the $T V-L^{p}$ model

The augmented Lagrangian functional for the $T V-L^{p}$ model is as follows:

$$
\begin{align*}
\min _{u, \boldsymbol{q}, z} \max _{\boldsymbol{q}, \lambda_{z}}\left\{g\left(u, \boldsymbol{q}, z, \lambda_{\boldsymbol{q}}, \lambda_{z}\right)=\right. & \int_{\Omega}|\boldsymbol{q}| d \boldsymbol{x}+\frac{\lambda}{p}\|z-f\|_{L^{p}}^{p}+\left(\lambda_{\boldsymbol{q}}, \boldsymbol{q}-\nabla u\right)+\left(\lambda_{z}, z-u\right) \\
& \left.+\frac{r}{2}\|\boldsymbol{q}-\nabla u\|_{L^{2}}^{2}+\frac{r_{z}}{2}\|z-u\|_{L^{2}}^{2}\right\} \tag{52}
\end{align*}
$$

where the bold face letters denote the vectors, and $r$ and $r_{z}$ are positive constants. Here, $(\cdot, \cdot)$ is the $L^{2}$-inner product. This method is to seek a saddle point of the augmented Lagrangian functional $g\left(u, \boldsymbol{q}, z, \lambda_{\boldsymbol{q}}, \lambda_{z}\right)$, and the algorithm can be summarized as follows.

Algorithm. Augmented Lagrangian method for $T V-L^{p}$ model.

1. Initialization: $\lambda_{z}^{0}=0, \lambda_{\boldsymbol{q}}^{0}=\mathbf{0}$;
2. For $k=1,2, \ldots$,
(i) Compute the minimizer $\left(u^{k+1}, \boldsymbol{q}^{k+1}, z^{k+1}\right)$ :

$$
\begin{equation*}
\left(u^{k+1}, \boldsymbol{q}^{k+1}, z^{k+1}\right)=\arg \min _{u, \boldsymbol{q}, z} \mathcal{g}\left(u, \boldsymbol{q}, z, \lambda_{\boldsymbol{q}}^{k}, \lambda_{z}^{k}\right) \tag{53}
\end{equation*}
$$

(ii) Update the Lagrangian multipliers:

$$
\begin{align*}
& \lambda_{\boldsymbol{q}}^{k+1}=\lambda_{\boldsymbol{q}}^{k}+r\left(\boldsymbol{q}^{k+1}-\nabla u^{k+1}\right)  \tag{54}\\
& \lambda_{z}^{k+1}=\lambda_{z}^{k}+r_{z}\left(z^{k+1}-u^{k+1}\right)
\end{align*}
$$

3. Goto step 2 until some stopping rule is met.

We see that it is essential to solve the minimization problem (53), i.e., (52) with fixed $\lambda_{\boldsymbol{q}}^{k}$ and $\lambda_{z}^{k}$, which can be split into three sub-problems.

- $u$-subproblem for given $\boldsymbol{q}$ and $z$ :

$$
\begin{equation*}
\min _{u}\left\{-\left(\lambda_{\boldsymbol{q}}^{k}, \nabla u\right)+\left(\lambda_{z}^{k},-u\right)+\frac{r}{2}\|\boldsymbol{q}-\nabla u\|_{L^{2}}^{2}+\frac{r_{z}}{2}\|z-u\|_{L^{2}}^{2}\right\} . \tag{55}
\end{equation*}
$$

- q-subproblem for given $u$ and $z$ :

$$
\begin{equation*}
\min _{\boldsymbol{q}}\left\{\int_{\Omega}|\boldsymbol{q}| d \boldsymbol{x}+\left(\lambda_{\boldsymbol{q}}^{k}, \boldsymbol{q}\right)+\frac{r}{2}\|\boldsymbol{q}-\nabla u\|_{L^{2}}^{2}\right\} . \tag{56}
\end{equation*}
$$

- $z$-subproblem for given $u$ and $\boldsymbol{q}$ :

$$
\begin{equation*}
\min _{z}\left\{\frac{\lambda}{p}\|z-f\|_{L^{p}}^{p}+\left(\lambda_{z}^{k}, z\right)+\frac{r_{z}}{2}\|z-u\|_{L^{2}}^{2}\right\} . \tag{57}
\end{equation*}
$$

In contrast with the $T V-L^{2}$ case, an additional auxiliary variable $z$ is introduced to circumvent the nonlinearity induced by the $L^{p}$-fidelity term with $p \neq 2$. This does not add any complexity to the algorithm, since the problem (57) can be solved analytically. To show the main idea, we rewrite (57) as

$$
\begin{equation*}
\min _{z}\left\{\frac{\lambda}{p}\|z-f\|_{L^{p}}^{p}+\frac{r_{z}}{2}\left\|z-\left(u-\lambda_{z}^{k} / r_{z}\right)\right\|_{L^{2}}^{2}\right\} \tag{58}
\end{equation*}
$$

This motivates us to seek

$$
z=(1-\beta) f+\beta\left(u-\lambda_{z}^{k} / r_{z}\right), \quad 0 \leq \beta \leq 1
$$

as in [34] and the problem (58) can be simplified to the one-dimensional problem:

$$
\begin{equation*}
\min _{0 \leq \beta \leq 1}\left\{E(\beta):=\frac{\lambda_{z}^{k} \beta^{p}}{p}\left\|u-\frac{\lambda_{z}^{k}}{r_{z}}-f\right\|_{L^{p}}^{p}+\frac{r_{z}}{2}(\beta-1)^{2}\left\|u-\frac{\lambda_{z}^{k}}{r_{z}}-f\right\|_{L^{2}}^{2}\right\} \tag{59}
\end{equation*}
$$

Denote $s=\left|u-\lambda_{z}^{k} / r_{z}-f\right|$. The minimization problem (59) is converted to

$$
\begin{equation*}
\min _{0 \leq \beta \leq 1}\left\{\frac{\lambda}{p} \beta^{p} s^{p}+\frac{r_{z}(\beta-1)^{2}}{2} s^{2}\right\}, \tag{60}
\end{equation*}
$$

which turns out to be a simple calculus problem.
In the computations, the stopping rule is based on the discrete $l^{2}$-error:

$$
\begin{equation*}
\left\|\nabla u^{k}-\boldsymbol{q}^{k}\right\|_{1^{2}}+\left\|z^{k}-u^{k}\right\|_{1^{2}} \leq \varepsilon \tag{61}
\end{equation*}
$$

for some prescribed error tolerance $\varepsilon>0$.

### 4.2. Numerical results

In the sequel, we present some numerical results computed by using the foregoing augmented Lagrangian method to justify the main theoretical results.

Here, we are interested in extracting the objects $\left\{s_{i}\right\}_{i=1}^{3}$ embedded in a black background image, as depicted in Fig. 1 (left). Purposely, we put their centers on a horizontal line so that we can plot the profile of the intensity along this line (see Fig. 1 (right)), and examine the change of the intensity with different choices of $\lambda$ and $p$ in the $T V-L^{p}$ model.


Fig. 1. Left: the given image $f$ with three embedded objects $\left\{s_{i}\right\}_{i=1}^{\}}$. Right: the profile of the intensity along the line of the centers.

Table 1
The $G$-norm or $G$-value of the objects, and $\lambda=\lambda_{p, i}^{\min }$ or $\lambda_{p, i}^{\max }$ for $\left\{s_{i}\right\}_{i=1}^{3}$ and $p=3,1.5,1$.

| $p$ |  | $s_{1}$ | $s_{2}$ | $s_{3}$ |
| :--- | :--- | :--- | :--- | :--- |
| 3 | $G\left(\left\|s_{i}\right\|^{p-1}\right)$ | $2.6077 \mathrm{e}+4$ | $1.0431 \mathrm{e}+5$ | $5.7513 \mathrm{e}+4$ |
|  | $\lambda_{p, i}^{\min }$ | $1.1504 \mathrm{e}-5$ | $2.8760 \mathrm{e}-6$ | $5.2162 \mathrm{e}-6$ |
|  | $\lambda_{p, i}^{\max }$ | $1.4061 \mathrm{e}-5$ | $3.5152 \mathrm{e}-6$ | $6.3754 \mathrm{e}-6$ |
| 1.5 | $G\left(\left\|s_{i}\right\|^{p-1}\right)$ | 26.0767 | 36.8780 | 43.7492 |
|  | $\lambda_{p, i}^{\min }$ | 0.0345 | 0.0244 | 0.0206 |
|  | $\lambda_{p, i}^{\max }$ | 0.0422 | 0.0298 | 0.0251 |
|  | $G\left(\partial\left\|s_{i}\right\|\right)$ | 2.6076 | 2.6076 | 3.7869 |
|  | $\lambda_{p, i}^{\min }$ | 0.3451 | 0.3451 | 0.2377 |
|  | $\lambda_{p, i}^{\max }$ | 0.4218 | 0.4218 | 0.2905 |

We follow [37] to compute the $G$-norm, denoted by $G\left(\left|s_{i}\right|^{p-1}\right)$ for $p>1$, and these values are listed in Table 1 . We find from the definition (37) that the $G$-value is independent of the intensity, but the $G$-norm depends on both the size and intensity. Hence, the objects $s_{1}$ and $s_{2}$ have the same size and $G$-value, but possess different $G$-norms. In the numerical experiments, we choose $\lambda=\lambda_{p, i}^{\min }$ or $\lambda_{p, i}^{\max }$ in the model, where

$$
\begin{equation*}
\lambda_{p, i}^{\min } G\left(\left|s_{i}\right|^{p-1}\right)=0.9, \quad \lambda_{p, i}^{\max } G\left(\left|s_{i}\right|^{p-1}\right)=1.1, \quad i=1,2,3,1<p<\infty, \tag{62}
\end{equation*}
$$

and if $p=1$, the $G$-value is in place of the $G$-norm. We take $r=r_{z}=80$ in (52), and the error tolerance in (61) to be $\varepsilon=4 \times 10^{-8}$.

Observe from Table 1 that for $p=3$,

$$
G\left(\left|s_{1}\right|^{p-1}\right)<G\left(\left|s_{3}\right|^{p-1}\right)<G\left(\left|s_{2}\right|^{p-1}\right) .
$$

Accordingly, if we choose $\lambda$ sequentially to be

$$
\begin{equation*}
\lambda_{p, 2}^{\min }<\lambda_{p, 2}^{\max }<\lambda_{p, 3}^{\min }<\lambda_{p, 3}^{\max }<\lambda_{p, 1}^{\min }<\lambda_{p, 1}^{\max }, \tag{63}
\end{equation*}
$$

then by Theorem 3.2, the objects will be "switched on" in turn. More precisely, if $\lambda=\lambda_{p, 2}^{\min }$, then all three objects will merge into the background as shown by the sub-figs of row 1 and columns 1 and 2 in Fig. 2 . On the other hand, taking $\lambda=\lambda_{p, 2}^{\max }$ (resp. $\lambda_{p, 3}^{\max }$ ) turns on $s_{2}$ (resp. $s_{2}$ and $s_{3}$ ). Such a pattern can be visualized from Fig. 2, which verifies the prediction by Theorem 3.2. We also find from profiles of the intensity that the model with $p=3$ cannot preserve the contrast, which differs from the $T V-L^{1}$ model. And, the $G$-norm depends on both the intensity and size, therefore the model can distinguish them, while based on the $G$-value, the $T V-L^{1}$ model cannot do this (cf. Fig. 4).

We plot in Fig. 3 the output images of the $T V-L^{p}$ model with $p=1.5$. Once again, the numerical evidences verify the theoretical prediction in Theorem 3.2.

To further test the numerical algorithm, we conduct the same experiment on the $T V-L^{1}$ model. Noting that the $G$-values of $s_{1}$ and $s_{2}$ are the same, we sequentially choose $\lambda=\lambda_{1,3}^{\min }, \lambda_{1,3}^{\max }, \lambda_{1,2}^{\min }, \lambda_{1,2}^{\max }$, and by Remark $3, s_{3}$ and $s_{1} / s_{2}$ are "switched on" in turn, which can be visualized from Fig. 4. We also observe from the profiles of the intensity that the contrast is wellpreserved, which is an important feature of $T V-L^{1}$ model. It also shows that the algorithm produces very good numerical results.


Fig. 2. Restored images by $T V-L^{p}(p=3)$ model and the profiles of the intensity of $\left\{s_{i}\right\}_{i=1}^{3}$ with different $\lambda$. Row $1: \lambda=\lambda_{p, 2}^{\min }, \lambda_{p, 3}^{\min }, \lambda_{p, 1}^{\min }$, and Row 2 : $\lambda=\lambda_{p, 2}^{\max }, \lambda_{p, 3}^{\max }, \lambda_{p, 1}^{\max }$.


Fig. 3. Restored images by $T V-L^{p}(p=1.5)$ model and the profiles of the intensity of $\left\{s_{i}\right\}_{i=1}^{3}$ with different $\lambda$. Row $1: \lambda=\lambda_{p, 3}^{\min }, \lambda_{p, 2}^{\mathrm{min}}, \lambda_{p, 1}^{\mathrm{min}}$, and Row 2 : $\lambda=\lambda_{p, 3}^{\max }, \lambda_{p, 2}^{\max }, \lambda_{p, 1}^{\max }$.


Fig. 4. Restored images by $T V-L^{1}$ model and the profiles of the intensity of $\left\{s_{i}\right\}_{i=1}^{3}$ with different $\lambda$. Row $1: \lambda=\lambda_{1,3}^{\min }, \lambda_{1,1}^{\min }\left(=\lambda_{1,2}^{\min }\right)$, and Row 2 : $\lambda=\lambda_{1,3}^{\max }, \lambda_{1,1}^{\max }\left(=\lambda_{1,2}^{\max }\right)$.

## 5. Concluding remarks

In this paper, we investigated various properties of the $T V-L^{p}$ model with general real $p$. As with the $T V-L^{1}$ model, we could formulate the model as a geometric problem, which provides some insights for the original model. By using the notion of $G$-norm, we characterized some important properties of the minimizer of the model, and gave a quantitative criteria for the choice of the parameter $\lambda$ for extracting image features and separating scales. We implemented the augmented Lagrangian method for the minimization problem, and provided illustrative numerical results to verify the analysis. Indeed, the $T V-L^{p}$ (with $1<p<\infty$ ) model exhibited some different properties when compared with the TV-L ${ }^{1}$ model.

## Acknowledgments

The authors want to thank Prof. Wotao Yin and Dr. Chunlin Wu for thoughtful discussions. This research is supported by Singapore MOE Grant T207B2202, and Singapore NRF2007IDM-IDM002-010. The research of the first author is also partially supported by NSFC (No. 10801049) and Foundation of North China Electric Power University. This author would like to thank Mathematical Imaging and Vision Group at Division of Mathematical Sciences in School of Physical and Mathematical Sciences of Nanyang Technological University, Singapore, for the hospitality during the visit.

## References

[1] L. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, Physica D 60 (1992) 259-268.
[2] L. Evans, R. Gariepy, Measure theory and fine properties of functions, in: Studies in Advanced Mathematics, CRC Press, Roca Baton, FL, 1992.
[3] E. Giusti, Minimal Surfaces and Functions of Bounded Variation, in: Monographs in Mathematics, vol. 80, Birkhauser Verlag, Basel, 1984.
[4] S. Alliney, Digital filters as absolute norm regularizers, IEEE Trans. Signal Process. 40 (1992) 1548-1562.
[5] M. Nikolova, Minimizers of cost-functions involving nonsmooth data-fidelity terms, SIAM J. Numer. Anal. 40 (2002) 965-994.
[6] T. Chan, S. Esedoglu, Aspects of total regularized $L^{1}$ function approximation, SIAM J. Appl. Math. 65 (2005) 1817-1837.
[7] J. Darbon, Total variation minimization with L1 data fidelity as a contrast invariant filter, in: Image and Signal Processing and Analysis, 2005. ISPA 2005. Proceedings of the 4th International Symposium on, 15-17, 2005, pp. 221-226.
[8] Y. Meyer, Oscillating Patterns in Image Processing and Nonlinear Evolution Equations, AMS, Providence, 2001.
[9] S. Osher, A. Sole, L. Vese, Image decomposition and restoration using total variation minimization and the $H^{-1}$ norm, Multiscale Model. Simul. 1 (2003) 349-370.
[10] T. Le, L. Vese, Image decomposition using total variation and $\operatorname{div}(B M O)$, Multiscale Model. Simul. 4 (2005) 390-423.
[11] J. Garnett, P. Jones, T. Le, L. Vese, Image decompositions using bounded variation and generalized homogeneous Besov spaces, Appl. Comput. Harmon. Anal. 23 (2007) 25-56.
[12] S. Lee, J. Seo, Noise removal with Gauss curvature-driven diffusion, IEEE Trans. Image Process. 14 (2005) 904-909.
[13] S. Lee, M. Kang, Total variation-based image noise reduction with generalized fidelity function, IEEE Signal Process. Lett. 14 (2007) $832-835$.
[14] M. Lysaker, A. Lundervold, X.C. Tai, Noise removal using fourth-order partial differential equations with applications to medical magnetic resonance images in space and time, IEEE Trans. Image Process. 12 (2003) 1579-1590.
[15] J. Darbon, M. Sigelle, Image restoration with discrete constrained total variation part II: levelable functions, convex priors and non-convex cases, J. Math. Imaging Vision 26 (2006) 277-291.
[16] S. Osher, O. Scherzer, $G$-norm properties of bounded variation regularization, Comm. Math. Sci. 2 (2004) 237-254.
[17] O. Scherzer, W. Yin, S. Osher, Slope and G-set characterization of set-valued functions and applications to non-differentiable optimization problems, Comm. Math. Sci. 3 (2005) 479-492.
[18] E. Tadmor, S. Nezzar, L. Vese, A multiscale image representation using hierarchical ( $B V, L^{2}$ ) decompositions, Multiscale Model. Simul. 2 (2004) 554-579.
[19] W. Yin, D. Goldfarb, S. Osher, The total variation regularized $L^{1}$ model for multiscale decomposition, Multiscale Model. Simul. 6 (2006) 190-211.
[20] H. Federer, Geometric Measure Theory, Springer, New York, 1969.
[21] L. Ambrosio, N. Fusco, D. Pallara, Functions of Bounder Variation and Free Discountinuity Problems, Oxford University Press, New York, 2000.
[22] M. Nikolova, A variational approach to remove outliers and impulse noise, J. Math. Imaging Vision 20 (2004) 99-120.
[23] S. Keeling, Total variation based convex filters for medical imaging, Appl. Math. Comput. 139 (2003) 101-119.
[24] A. Haddad, Y. Meyer, An improvement of Rudin-Osher-Fatemi model, Appl. Comput. Harmon. Anal. 22 (2007) 319-334.
[25] S. Kindermann, S. Osher, J. Xu, Denoising by BV-duality, J. Sci. Comput. 28 (2006) 414-444.
[26] C. Vogel, M. Oman, Iterative methods for total variation denoising, SIAM J. Sci. Comput. 17 (1998) 813-824.
[27] T. Chan, G. Golub, P. Mulet, A nonlinear primal-dual method for total variation-based image restoration, SIAM J. Sci. Comput. 20 (1999) $1964-1977$.
[28] E. Bae, X. Tai, Graph cut optimization for the piecewise constant level set method applied to multiphase image segmentation, in: Sapce and Variational Methods in Computer Vision, SSVM 2009, 2009, pp. 1-13.
[29] S. Osher, M. Burger, D. Goldfarb, J. Xu, W. Yin, An iterative regularization method for total variation based image restoration, Multiscale Model. Simul. 4 (2005) 460-489.
[30] Y. Wang, J. Yang, W. Yin, Y. Zhang, A new alternating minimization algorithm for total variation image reconstruction, SIAM J. Imaging Sci. 1 (2008) 248-272.
[31] Q. Chang, I. Chen, Acceleration methods for total variation-based image denoising, SIAM J. Sci. Comput. 25 (2003) 982-994.
[32] X. Tai, C. Wu, Augmented Lagrangian method, dual methods and split Bregman iteration for ROF model, scale space and variational methods in computer vision, in: Second International Conference, SSVM 2009, Voss, Norway, June 1-5, 2009. Proceedings, in: Lecture Notes in Comput. Sci., vol. 5567, Springer, 2009, pp. 502-513.
[33] C. Wu, X. Tai, Augmented Lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and high order models, SIAM J. Imaging Sci. 3 (2010) 300-339.
[34] C. Wu, J. Zhang, X. Tai, Augmented Lagrangian method for total variation restoration with non-quadratic fidelity, Inverse Probl. Imaging 5 (2011) 237-261.
[35] X. Bresson, S. Esedoglu, P. Vandergheynst, J.P. Thiran, S. Osher, Fast global minimization of the active contour/snake model, J. Math. Imaging Vision 28 (2007) 151-167.
[36] W. Yin, D. Goldfarb, S. Osher, Image cartoon-texture decomposition and feature selection using the total variation regularized $L^{1}$ functional, in: Variational, Geometric, and Level Set Methods in Computer Vision, vol. 3752, 2005, pp. 73-84.
[37] W. Yin, D. Goldfarb, S. Osher, A comparision of three total variation based texture extraction models, J. Vis. Commun. Image Represent. 18 (2007) 240-252.
[38] F. Alizadeh, D. Goldfarb, Second-order cone programming, Math. Program. 95 (2003) 3-51.
[39] D. Goldfarb, W. Yin, Second-order cone programming methods for total variation based image restoration, SIAM J. Sci. Comput. 27 (2005) $622-645$.


[^0]:    * Corresponding author.

    E-mail addresses: yyshi@amss.ac.cn (Y. Shi), lilian@ntu.edu.sg (L.-L. Wang), tai@cma.uio.no (X.-C. Tai).

