Geometry of total variation regularized $L^p$-model

Yuying Shi $^{a,*}$, Li-Lian Wang $^b$, Xue-Cheng Tai $^{b,c}$

$^a$ Department of Mathematics and Physics, North China Electric Power University, Beijing, 102206, China
$^b$ Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, 637371, Singapore
$^c$ 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

$$\inf_{u \in \mathcal{X}} \{ J(u) + \lambda H(u - f) \},$$

where $J(\cdot)$ and $H(\cdot)$ are two non-negative functionals defined over a suitable functional space $\mathcal{X}$, and $\lambda > 0$ is a parameter to balance the two terms. Various models with different combinations of $\{\mathcal{X}; H, J\}$ 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 TV-$L^2$ model, takes the form

$$\inf_{u \in \mathcal{X}} \left\{ \text{TV}(u) + \frac{\lambda}{2} \| u - f \|^2_{L^2(\Omega)} \right\}.$$  

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

$$\text{TV}(u) = \int_{\Omega} |\nabla u| d\mathbf{x} := \sup_{p \in \mathcal{S}} \int_{\Omega} u(\mathbf{x}) \text{div} p(\mathbf{x}) d\mathbf{x},$$

where $\mathcal{S} = \{ \mathbf{p} \in C^1_c(\Omega; \mathbb{R}^2) : |\mathbf{p}(\mathbf{x})| \leq 1, \forall \mathbf{x} \in \mathbb{R}^2 \}$, and the space $\mathcal{X} = \{ u \in L^1(\Omega) : \| u \|_{BV} < \infty \}$ with the norm $\| u \|_{BV} := \| u \|_{L^1(\Omega)} + \text{TV}(u)$. As an important variant, the TV-$L^1$ 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]),
Theorem 2.1. Define
\[ C(\Sigma, \Gamma_f) := \text{Per}(\Sigma) + \lambda \left( \int_{\Sigma \setminus \Gamma_f} (\xi - f)^{p-1} \, dx + \int_{\Gamma_f \setminus \Sigma_f} (f - \xi)^{p-1} \, dx \right). \]

Motivated by [6,19], we consider in this paper the geometric properties and scale separation of the TV-L^p model with 0 < p < +∞:
\[
\min_{u \in BV(\Omega)} \left\{ TV^p_L(u) := TV(u) + \frac{\lambda}{p} \int_\Omega |f - u|^p \, dx \right\}, \quad \lambda > 0, \tag{4}
\]
and discuss the numerical algorithms for this general model. We first formulate the cost functional TVL^p (·) 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 λ that allows to extract geometric features of a given scale. Different from the G-value for the TV-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 TV-L^1 model fails to distinguish them, but the TV-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 TV-L^p (0 < p < +∞) 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 TV-L^p (1 < p < +∞) 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 TV-L^p (1 ≤ p < +∞) 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 TV-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 Ω = \( \mathbb{R}^2 \) for the moment, and recall that the perimeter of a Borel measurable set \( \Sigma \subset \Omega \) is defined by
\[
\text{Per}(\Sigma) := TV(1_\Sigma) = \int_\Omega |\nabla 1_\Sigma| \, dx, \tag{5}
\]
where 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
\[
\Gamma_f := \{ \xi \in \mathbb{R} : f(\mathbf{x}) > \xi \}. \tag{6}
\]
For clarity, we particularly denote \( \Sigma_f(\xi) := \Gamma_f(\xi) \), where \( u \) is the restored image. We shall also use the co-area formula (see, e.g., [20,3]):
\[
TV(u) = \int_{-\infty}^{+\infty} \text{Per}(\Sigma_u(\xi)) \, d\xi = \int_{-\infty}^{+\infty} TV(1_{\Sigma_u(\xi)}) \, d\xi. \tag{7}
\]
The following important result shows that the TV-L^p functional (0 < p < ∞) in (4) can be expressed in an equivalent form in terms of the level sets.

Theorem 2.1. Define
\[
C(\Sigma_u, \Gamma_f) := \text{Per}(\Sigma) + \lambda \left( \int_{\Sigma_u \setminus \Gamma_f} (\xi - f)^{p-1} \, dx + \int_{\Gamma_f \setminus \Sigma_u} (f - \xi)^{p-1} \, dx \right). \tag{8}
\]
Then for all $0 < p < \infty$,
\[
TV(u) + \frac{\lambda}{p} \|u - f\|_p^p = \int_{-\infty}^{+\infty} C(\Sigma_u, \Gamma_f) d\xi.
\] (9)

**Proof.** We have that
\[
\|u - f\|_p^p = \int_{\{u \neq f\}} (u - f)^p dx + \int_{\{f > u\}} (f - u)^p dx
\]
\[
= \int_{\{u \neq f\}} (u - f)^p dx + \int_{\{f > u\}} (f - u)^p + \int_{\{f \leq u\}} (f - u)^p dx
\]
\[
= \int_{\Omega} \int_{\|u\|} (\xi - f)^p d\xi dx + \int_{\Omega} \int_{\|u\|} (f - \xi)^p d\xi dx
\]
\[
= \int_{\Omega} \int_{\|u\|} (\xi - f)^p d\xi dx + \int_{\Omega} \int_{\|u\|} (f - \xi)^p d\xi dx
\]
\[
= \int_{\Omega} \int_{\|u\|} (\xi - f)^p d\xi dx + \int_{\Omega} \int_{\|u\|} (f - \xi)^p d\xi dx
\]
\[
= p \left[ \int_{\{u \neq f\}} (\xi - f)^p d\xi dx + \int_{\{f > u\}} (f - \xi)^p d\xi dx \right].
\]
\[
= p \left[ \int_{\{u \neq f\}} (\xi - f)^p d\xi dx + \int_{\{f > u\}} (f - \xi)^p d\xi dx \right]
\]
\[
\text{Here, } 1_{[a,b]} \text{ is the characteristic function of the interval } [a, b] \subset \mathbb{R}. \text{ Observe that}
\]
\[
\int_{\Omega} 1_{\{u \neq f\}} 1_{\Sigma_u \setminus \Gamma_f} dx = \int_{\Sigma_u \setminus \Gamma_f} dx,
\]
and likewise for $1_{\{u \neq f\}} 1_{\Gamma_f \setminus \Sigma_u}$. Hence, we obtain
\[
\|u - f\|_p^p = \int_{\|u\|} (\xi - f)^p d\xi dx + \int_{\Gamma_f \setminus \Sigma_u} (f - \xi)^p d\xi dx.
\] (10)
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 $C(\Sigma, \Gamma_f)$ 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:
\[
\min_{\Sigma} C(\Sigma, \Gamma_f(\xi)) \text{ for given } \xi \in \mathbb{R} \text{ and } f.
\] (11)

Next, we present two fundamental properties of the functional $C$. Hereafter, for $\xi_1 < \xi_2$, we denote $\Gamma_f^i := \Gamma_f(\xi_i) \ (i = 1, 2)$ 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
\[
C(\Sigma^2, \Gamma_f^1) - C(\Sigma^1 \cap \Sigma^2, \Gamma_f^1) \geq C(\Sigma^1 \cup \Sigma^2, \Gamma_f^1) - C(\Sigma^1, \Gamma_f^1) \geq 0.
\] (12)
and
\[
0 \geq C(\Sigma^2, \Gamma_f^2) - C(\Sigma^1 \cup \Sigma^2, \Gamma_f^2) \geq C(\Sigma^2, \Gamma_f^1) - C(\Sigma^1 \cap \Sigma^2, \Gamma_f^1).
\] (13)

**Proof.** We first prove (12). By the definition (8),
\[
C(\Sigma^2, \Gamma_f^1) - C(\Sigma^1 \cap \Sigma^2, \Gamma_f^1) = C(\Sigma^1 \cup \Sigma^2, \Gamma_f^1) + C(\Sigma^1, \Gamma_f^1)
\]
\[
= \text{Per}(\Sigma^2) - \text{Per}(\Sigma^1 \cap \Sigma^2) - \text{Per}(\Sigma^1 \cup \Sigma^2) + \text{Per}(\Sigma^1)
\]
\[
+ \lambda \left[ \int_{\Sigma^1 \setminus \Gamma_f^1} (\xi - f)^p d\xi dx - \int_{(\Sigma^1 \setminus \Sigma^2) \setminus \Gamma_f^1} (\xi - f)^p d\xi dx \right]
\]
\[
- \left[ \int_{(\Sigma^1 \cup \Sigma^2) \setminus \Gamma_f^1} (\xi - f)^p d\xi dx - \int_{\Sigma^1 \setminus \Gamma_f^1} (\xi - f)^p d\xi dx \right].
\] (16)
One verifies readily that
\[
\left(\Sigma \setminus \Gamma_j \right) \setminus \left((\Sigma^1 \setminus \Sigma^2) \setminus \Gamma_j \right) = (\Sigma^2 \setminus \Sigma^1) \setminus \Gamma_j
\]
and
\[
\left((\Sigma^1 \cup \Sigma^2) \setminus \Gamma_j \right) \setminus \left(\Sigma^1 \setminus \Gamma_j \right) = (\Sigma^2 \setminus \Sigma^1) \setminus \Gamma_j.
\]
Thus the summation in the square brackets in (16) and (17) vanishes. Similarly, as \(\Sigma^2 \supseteq (\Sigma^1 \cap \Sigma^2)\), we have
\[
\left(\Gamma_i \setminus (\Sigma^1 \cap \Sigma^2) \right) \setminus \left((\Gamma_i \setminus \Sigma^2) \cap \Sigma^1 \right) = (\Gamma_i \setminus \Sigma^1) \setminus \Sigma^1,
\]
and
\[
\left(\Gamma_i \setminus \Sigma^1 \right) \setminus \left((\Gamma_i \setminus \Sigma^1) \setminus \Sigma^2 \right) = (\Gamma_i \setminus \Sigma^1) \setminus \Sigma^1,
\]
which imply that the summation in (18) and (19) is zero. Hence, a combination of the above facts leads to
\[
C(\Sigma^2, \Gamma_i) - C(\Sigma^1 \cap \Sigma^2, \Gamma_i) - C(\Sigma^1 \cup \Sigma^2, \Gamma_i) + C(\Sigma^1, \Gamma_i) = \text{Per}(\Sigma^2) - \text{Per}(\Sigma^1 \cap \Sigma^2) - \text{Per}(\Sigma^1 \cup \Sigma^2) + \text{Per}(\Sigma^1).
\]
Therefore, the formula (12) follows from the property of the Per-functional (see, e.g., [21]):
\[
\text{Per}(\Sigma^2) + \text{Per}(\Sigma^1) \geq \text{Per}(\Sigma^1 \cap \Sigma^2) + \text{Per}(\Sigma^1 \cup \Sigma^2),
\]
and the assumption that \(\Sigma^1\) is the minimizer with \(\Gamma_j = \Gamma_j^1\).

Now, we turn to the derivation of (13). Similarly, by (12),
\[
C(\Sigma^2, \Gamma_i) - C(\Sigma^1 \cap \Sigma^2, \Gamma_i) - C(\Sigma^2, \Gamma_i) + C(\Sigma^1 \cap \Sigma^2, \Gamma_i) = \left[\text{Per}(\Sigma^2) - \text{Per}(\Sigma^1 \cap \Sigma^2)\right] - \left[\text{Per}(\Sigma^2) - \text{Per}(\Sigma^1 \cap \Sigma^2)\right]
\]
\[
+ \lambda \left[\int_{\Sigma^2 \setminus \Gamma_i} (\xi - f)^{p-1}dx - \int_{(\Sigma^1 \cap \Sigma^2) \setminus \Gamma_i} (\xi - f)^{p-1}dx\right]
\]
\[
- \left(\int_{\Sigma^2 \setminus \Gamma_i} (\xi - f)^{p-1}dx - \int_{(\Sigma^1 \setminus \Sigma^2) \setminus \Gamma_i} (\xi - f)^{p-1}dx\right)
\]
\[
+ \lambda \left[\int_{\Gamma_i \setminus \Sigma^2} (f - \xi)^{p-1}dx - \int_{\Gamma_i \setminus \Sigma^1} (f - \xi)^{p-1}dx\right]
\]
\[
- \left(\int_{\Gamma_i \setminus \Sigma^1} (f - \xi)^{p-1}dx - \int_{\Gamma_i \setminus \Sigma^2} (f - \xi)^{p-1}dx\right).
\]

Obviously, (22) is zero, and for \(i = 1, 2\),
\[
\left(\Sigma^2 \setminus \Gamma_i \right) \setminus \left((\Sigma^1 \setminus \Sigma^2) \setminus \Gamma_i \right) = (\Sigma^2 \setminus \Sigma^1) \setminus \Gamma_i,
\]
and
\[
\left(\Gamma_i \setminus (\Sigma^1 \cap \Sigma^2) \right) \setminus \left((\Gamma_i \setminus \Sigma^2) \cap \Sigma^1 \right) = (\Gamma_i \setminus \Sigma^1) \setminus \Sigma^1.
\]
By (27), the summation in (23) and (24) becomes
\[
\lambda \left[\int_{(\Sigma^2 \setminus \Sigma^1) \setminus \Gamma_i} (\xi - f)^{p-1}dx - \int_{(\Sigma^2 \setminus \Sigma^1) \setminus \Gamma_i} (\xi - f)^{p-1}dx\right] \geq 0,
\]
where we have used the fact: \(\Gamma_i \supseteq \Gamma_j \) and \(\Sigma^2 \setminus \Sigma^1 \supseteq \Gamma_j \supseteq \Sigma^2 \setminus \Sigma^1 \). 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 \(\Gamma_j = \Gamma_j^2\). □
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 $\Gamma^1 = \Gamma^1_\lambda$ and $\Gamma^2 = \Gamma^2_\lambda$, respectively. If $\Gamma^1_\lambda \supseteq \Gamma^2_\lambda$, 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
\[
0 \geq C(\Sigma^2, \Gamma^1_\lambda) - C(\Sigma^1 \cap \Sigma^2, \Gamma^2_\lambda) \geq C(\Sigma^1 \cap \Sigma^2, \Gamma^1_\lambda) - C(\Sigma^1, \Gamma^1_\lambda) \geq 0.
\]
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 $\Gamma^1 = \Gamma^1_\lambda$ and $\Gamma^2 = \Gamma^2_\lambda$, respectively. It is clear that the former is included in the latter. \qed

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

### 3. Properties of the minimizers

In this section, we study properties of the minimizers of the TV-$L^p$ 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 TV-$L^p$ model (4), if and only if $\lambda$ satisfies
\[
\lambda \int_\Omega |u_\lambda - f|^{p-2}(u_\lambda - f)h dx \geq TV(u_\lambda) - TV(u_\lambda + h), \quad \forall h \in BV(\Omega). \tag{29}
\]

**Proof.** If $u_\lambda$ is a minimizer of (4), we have that for any $h \in BV(\Omega)$ and $\epsilon > 0$,
\[
TV(u_\lambda) + \frac{\lambda}{p} \int_\Omega |u_\lambda - f|^p dx \leq TV(u_\lambda + \epsilon h) + \frac{\lambda}{p} \int_\Omega |u_\lambda + \epsilon h - f|^p dx
\]
\[
\quad \leq TV(u_\lambda + \epsilon h) + \frac{\lambda}{p} \int_\Omega |u_\lambda - f|^p dx
\]
\[
+ \lambda \epsilon \int_\Omega |u_\lambda - f|^{p-2}(u_\lambda - f) dx + \lambda \epsilon \eta(\epsilon; p, u_\lambda, h), \tag{30}
\]
where $\eta(\epsilon; p, u_\lambda, h) = o(\epsilon)$. This implies
\[
- \lambda \int_\Omega |u_\lambda - f|^{p-2}(u_\lambda - f) dx \leq \frac{TV(u_\lambda + \epsilon h) - TV(u_\lambda)}{\epsilon} + \lambda \eta(\epsilon; p, u_\lambda, h). \tag{31}
\]
The convexity of $TV(u)$ implies
\[
\frac{TV(u_\lambda + \epsilon h) - TV(u_\lambda)}{\epsilon} \leq TV(u_\lambda + h) - TV(u_\lambda), \quad 0 < \epsilon < 1.
\]
Hence, letting $\epsilon \to 0$ in (31) leads to \[
- \lambda \int_\Omega |u_\lambda - f|^{p-2}(u_\lambda - f) dx \leq TV(u_\lambda + h) - TV(u_\lambda).
\]

Therefore, $\lambda$ satisfies (29).

Conversely, by (29) and the convexity of $| \cdot |^p (1 < p)$, we obtain that
\[
TV(u_\lambda + h) + \frac{\lambda}{p} \int_\Omega |u_\lambda + h - f|^p dx \geq TV(u_\lambda) + \frac{\lambda}{p} \int_\Omega |u_\lambda - f|^p dx
\]
\[
+ \lambda \int_\Omega |u_\lambda - f|^{p-2}(u_\lambda - f) dx + TV(u_\lambda + h) - TV(u_\lambda)
\]
\[
\geq TV(u_\lambda) + \frac{\lambda}{p} \int_\Omega |u_\lambda - f|^p dx. \tag{32}
\]
This ends the proof. \qed
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 $TV-L^1$ model, if and only if $\lambda$ satisfies that for any $h \in BV(\Omega)$,

$$
\lambda \left( - \int_{u_\lambda \neq f} \text{sign}(u_\lambda - f) h dx - \int_{u_\lambda = f} |h| dx \right) \leq TV(u_\lambda + h) - TV(h). \quad (33)
$$

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

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

$$
J_\gamma(u) := \int_\Omega |\nabla u|^\gamma \, dx, \quad \gamma > 1.
$$

(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 $TV(\cdot)$.

For a fixed $\lambda$, the $TV-L^p$ 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 $v = \text{div} \, g$ with $g = (g_1, g_2)$ and $g_i \in L^\infty(\Omega)$, $i = 1, 2$, the $G$-norm of $v$ is defined as

$$
G(v) := \|v\|_G = \inf_g \left\{ \|g_1^2 + g_2^2\|_{L^\infty} \right\}. \quad (34)
$$

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

$$
G(v) = \sup_{h \in BV(\Omega), h \neq 0} \frac{\int_\Omega v(x) h(x) dx}{TV(h)}. \quad (35)
$$

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

$$
0 < \lambda \leq \frac{1}{G(|f|^{p-1})}. \quad (36)
$$

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

$$
-\lambda \int_\Omega |f|^{p-2} h dx \geq -TV(h) \Leftrightarrow \lambda \int_\Omega |f|^{p-1} h dx \leq TV(h), \quad \forall h \in BV(\Omega),
$$

where we used the fact that the (pixel) intensity of $f \geq 0$. In view of (35), we have $\lambda G(|f|^{p-1}) \leq 1$, so (36) follows. \hfill \Box

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.,

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

Then, $u_\lambda = 0$ is the minimizer of the $TV-L^1$ 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 $f = 1_{B_r(0)}$, where $B_r(0) \subset \Omega$ is a disk centered at the origin with radius $r$. Notice that

$$
G(1_{B_r(0)}^{p-1}) = G(|1_{B_r(0)}|) = \frac{r}{2}.
$$

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

The next result shows the feature extraction capability of the $TV-L^p$ 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, $f = f_{\Omega_1} 1_{\Omega_1} + f_{\Omega_2} 1_{\Omega_2}$, where $f_{\Omega_i}$ is the restriction
of $f$ in $\Omega_i$ ($i = 1, 2$), and likewise for $u_{\Omega_2}$ below. Suppose that $G(|f_{\Omega_2}|^{p-1}) > G(|f_{\Omega_1}|^{p-1})$ and $u$ is a minimizer of $TV$-$L^p$ model (4) with $1 < p < +\infty$. If $\lambda$ is chosen such that
\[
\frac{1}{G(|f_{\Omega_2}|^{p-1})} < \lambda < \frac{1}{G(|f_{\Omega_1}|^{p-1})},
\]
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 $BV(\Omega)$.

We first prove $u_{\Omega_2} \neq 0$ by contradiction. The convexity of $TV(u)$ and $| \cdot |^p (1 < p < +\infty)$ implies
\[
\epsilon TV(h) \geq TV(u_{\Omega_1 \setminus \Omega_2} + \epsilon h) - TV(u_{\Omega_2 \setminus \Omega_2}).
\]
and
\[
\frac{\lambda}{p} \int_{\Omega_2} |f_{\Omega_2}|^p dx - \frac{\lambda}{p} \int_{\Omega_2} |f_{\Omega_2} - \epsilon h|^p dx \geq \epsilon \lambda \int_{\Omega_2} |f_{\Omega_2} - \epsilon h|^p - (f_{\Omega_2} - \epsilon h) h dx.
\]
Since $u$ is a minimizer of (4), assuming $u_{\Omega_2} = 0$ leads to
\[
TV(u_{\Omega_1 \setminus \Omega_2} + \epsilon h) \geq TV(u_{\Omega_1 \setminus \Omega_2}) + \frac{\lambda}{p} \int_{\Omega_2} |f_{\Omega_2} - \epsilon h|^p dx + \frac{\lambda}{p} \int_{\Omega_2} |u_{\Omega_1 \setminus \Omega_2} - f_{\Omega_1 \setminus \Omega_2}|^p dx.
\]
Hence, a direct consequence of (39)–(41) is
\[
\epsilon TV(h) \geq \lambda \int_{\Omega_2} |f_{\Omega_2} - \epsilon h|^p - (f_{\Omega_2} - \epsilon h) h dx.
\]
That is,
\[
TV(h) \geq \lambda \int_{\Omega_2} |f_{\Omega_2} - \epsilon h|^p - (f_{\Omega_2} - \epsilon h) h dx.
\]
Letting $\epsilon \to 0$ in (42) yields
\[
TV(h) \geq \lambda \int_{\Omega_2} |f_{\Omega_2}|^{p-1} h dx, \quad \forall h \in BV(\Omega),
\]
where we used the fact $f_{\Omega_2} \geq 0$. In view of the definition (35) and the condition (38), there exists $0 \neq h_0 \in BV(\Omega)$ such that
\[
\lambda \int_{\Omega_2} |f_{\Omega_2}|^{p-1} h_0 dx > TV(h_0),
\]
which contradicts to (43). Therefore, $u_{\Omega_2} \neq 0$.

Now, we turn to proving $u_{\Omega_2} = 0$ by contradiction. As a preparation, we first show that the intensity of the minimizer $u$ in the region of the background, denoted by $u_{\Omega_2}$, is zero, where we denote $\Omega_2 = \Omega \setminus (\Omega_1 \cup \Omega_2)$ and $\Omega_1' = \Omega_1 \cup \Omega_2$. Indeed, by the definition of the total variation (3),
\[
TV(u_{\Omega_1} + u_{\Omega_2}) = \sup_{p \in s} \left\{ \int_{\Omega} u_{\Omega_2} \text{div} \ p \ dx + \int_{\Omega} u_{\Omega_2} \text{div} \ p \ dx \right\}
\]
\[
\geq \sup_{p \in s} \int_{\Omega} u_{\Omega_2} \text{div} \ p \ dx = TV(u_{\Omega_2}).
\]
In view of \( f_{Ω_b} = 0 \), (46) implies

\[
TV(u_{Ω_b} + u_Ω^2) + \frac{λ}{p} \int_Ω |u_{Ω_b}|^p dx ≤ TV(u_Ω^2),
\]

which contradicts to (45). Thus the minimizer \( u \) of (4) must satisfy \( u_{Ω_b} = 0 \).

Now, we are ready to prove \( u_{Ω_1} = 0 \) by contradiction again. By the convexity of \( |·|^p \) (1 < \( p < +∞ \)),

\[
\frac{λ}{p} \int_Ω |f_{Ω_1}|^p dx - \frac{λ}{p} \int_Ω |f_{Ω_1} - h|^p dx ≤ \lambda \int_Ω |f_{Ω_1}|^{p-1} h dx.
\]

Assume that \( u \) is a minimizer of (4) with \( u_{Ω_1} \neq 0 \). Using the facts that \( u_{Ω_b} = 0 \) and \( Ω_1, Ω_2 \) are separated with some distance, we obtain

\[
TV^p(u_{Ω_b \setminus Ω_1}) ≥ TV^p(u_{Ω_b \setminus Ω_1} + u_{Ω_2})
\]

\[
= \int_{Ω_1 \setminus Ω_1} |∇u_{Ω_b \setminus Ω_1}| dx + \int_{Ω_1} |∇u_{Ω_1}| dx
\]

\[
+ \frac{λ}{p} \int_{Ω_2} |u_{Ω_b \setminus Ω_1} - f_{Ω_2}|^p dx + \frac{λ}{p} \int_{Ω_1} |f_{Ω_1} - u_{Ω_1}|^p dx
\]

\[
= TV^p(u_{Ω_b \setminus Ω_1}) + \int_{Ω_1} |∇u_{Ω_1}| dx + \frac{λ}{p} \int_{Ω_1} |f_{Ω_1} - u_{Ω_1}|^p dx - \frac{λ}{p} \int_{Ω_1} |f_{Ω_1}|^p dx,
\]

which implies

\[
\int_{Ω_1} |∇u_{Ω_1}| dx ≤ \frac{λ}{p} \int_{Ω_1} |f_{Ω_1}|^p dx - \frac{λ}{p} \int_{Ω_1} |f_{Ω_1} - u_{Ω_1}|^p dx.
\]

Thus, by (47) and (49),

\[
TV(u_{Ω_1}) = \int_{Ω_1} |∇u_{Ω_1}| dx ≤ \frac{λ}{p} \int_{Ω_1} |f_{Ω_1}|^{p-1} u_{Ω_1} dx.
\]

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

\[
TV(h) ≥ λ \int_{Ω_1} |f_{Ω_1}|^{p-1} h dx, \quad ∀ h \in BV(Ω).
\]

Taking \( h = u_{Ω_1} \) in the above inequality leads to a contradiction. Therefore, \( u_{Ω_1} = 0 \). □

4. Algorithms and numerical results

In this section, we discuss numerical methods for solving the TV-\( L^p \) 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 TV-\( L^1 \) 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 TV-\( L^1 \) 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 TV-\( L^p \) model

The augmented Lagrangian functional for the TV-\( L^p \) model is as follows:

\[
\min_{u,q,z} \max_{λ_q,λ_z} \left\{ g(u, q, z, λ_q, λ_z) = \int_Ω |q| dx + \frac{λ}{p} \| z - f \|_p^p + (λ_q, q - ∇ u) + (λ_z, z - u) + \frac{r}{2} \| q - ∇ u \|_2^2 + \frac{r_z}{2} \| z - u \|_2^2 \right\},
\]

where

\[
g(u, q, z, λ_q, λ_z) = \int_Ω |q| dx + \frac{λ}{p} \| z - f \|_p^p + (λ_q, q - ∇ u) + (λ_z, z - u) + \frac{r}{2} \| q - ∇ u \|_2^2 + \frac{r_z}{2} \| z - u \|_2^2.
\]
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 \( g(u, q, z, \lambda_q, \lambda_z) \), and the algorithm can be summarized as follows.

**Algorithm.** Augmented Lagrangian method for \(TV-L^p\) model.

1. Initialization: \( \lambda^0_q = 0, \lambda^0_z = 0 \);
2. For \( k = 1, 2, \ldots \):
   - Compute the minimizer \( (u^{k+1}, q^{k+1}, z^{k+1}) \):
     \[
     (u^{k+1}, q^{k+1}, z^{k+1}) = \arg \min_{u, q, z} g(u, q, z, \lambda^k_q, \lambda^k_z). \tag{53}
     \]
   - Update the Lagrangian multipliers:
     \[
     \lambda^{k+1}_q = \lambda^k_q + r(q^{k+1} - \nabla u^{k+1}), \quad \lambda^{k+1}_z = \lambda^k_z + r_z(z^{k+1} - u^{k+1}). \tag{54}
     \]
3. Go to 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 \( \lambda^k_q \) and \( \lambda^k_z \), which can be split into three sub-problems.

- **u-subproblem** for given \( q \) and \( z \):
  \[
  \min_u \left\{ - (\lambda^k_q, \nabla u) + (\lambda^k_z, -u) + \frac{r}{2} \|q - \nabla u\|_{L^2}^2 + \frac{r_z}{2} \|z - u\|_{L^2}^2 \right\}. \tag{55}
  \]
- **q-subproblem** for given \( u \) and \( z \):
  \[
  \min_q \left\{ \int_{\Omega} |q| dx + (\lambda^k_q, q) + \frac{r_z}{2} \|q - \nabla u\|_{L^2}^2 \right\}. \tag{56}
  \]
- **z-subproblem** for given \( u \) and \( q \):
  \[
  \min_z \left\{ \frac{\lambda}{p} \|z - f\|_{L^p}^p + (\lambda^k_z, z) + \frac{r_z}{2} \|z - u\|_{L^2}^2 \right\}. \tag{57}
  \]

In contrast with the \(TV-L^2\) 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

\[
\min_z \left\{ \frac{\lambda}{p} \|z - f\|_{L^p}^p + \frac{r_z}{2} \|z - (u - \lambda^k_z/r_z)\|_{L^2}^2 \right\}. \tag{58}
\]

This motivates us to seek \( z = (1 - \beta)f + \beta(u - \lambda^k_z/r_z), \quad 0 \leq \beta \leq 1 \), as in [34] and the problem (58) can be simplified to the one-dimensional problem:

\[
\min_{0 \leq \beta \leq 1} \left\{ E(\beta) := \frac{\lambda^k_z \beta^p}{p} \left\| u - \frac{\lambda^k_z}{r_z} f \right\|_{L^p}^p + \frac{r_z}{2} (\beta - 1)^2 \left\| u - \frac{\lambda^k_z}{r_z} f \right\|_{L^2}^2 \right\}. \tag{59}
\]

Denote \( s = |u - \lambda^k_z/r_z - f| \). The minimization problem (59) is converted to

\[
\min_{0 \leq \beta \leq 1} \left\{ \frac{\lambda}{p} \beta^p s^p + \frac{r_z}{2} (\beta - 1)^2 s^2 \right\}, \tag{60}
\]

which turns out to be a simple calculus problem.

In the computations, the stopping rule is based on the discrete \( L^2 \)-error:

\[
\| \nabla u^k - q^k \|_{L^2} + \| z^k - u^k \|_{L^2} \leq \varepsilon, \tag{61}
\]

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 \( \{s_i\}_{i=1}^3 \) embedded in a black background image, as depicted in Fig. 1 (left). Purposefully, 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 \(TV-L^p\) model.
Accordingly, if we choose \( \lambda = \lambda_{p,i}^\text{min} \) or \( \lambda_{p,i}^\text{max} \) in the model, where

\[
\lambda_{p,i}^\text{min} G(|s_i|^{p-1}) = 0.9, \quad \lambda_{p,i}^\text{max} G(|s_i|^{p-1}) = 1.1, \quad i = 1, 2, 3, \quad 1 < p < \infty,
\]

and if \( p = 1 \), the G-value is in place of the G-norm. We take \( r = r_2 = 80 \) in (52), and the error tolerance in (61) to be \( \varepsilon = 4 \times 10^{-8} \).

Observe from Table 1 that for \( p = 3 \),

\[
G(|s_1|^{p-1}) < G(|s_3|^{p-1}) < G(|s_2|^{p-1}).
\]

Accordingly, if we choose \( \lambda \) sequentially to be

\[
\lambda_{p,2}^\text{min} < \lambda_{p,2}^\text{max} < \lambda_{p,3}^\text{min} < \lambda_{p,3}^\text{max} < \lambda_{p,1}^\text{min} < \lambda_{p,1}^\text{max},
\]

then by Theorem 3.2, the objects will be “switched on” in turn. More precisely, if \( \lambda = \lambda_{p,2}^\text{min} \) 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 \( \lambda = \lambda_{p,3}^\text{max} \) (resp. \( \lambda_{p,3}^\text{max} \)) turns on \( s_3 \) (resp. \( s_2 \) and \( s_1 \)). 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 TV-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 TV-L^1 model cannot do this (cf. Fig. 4).

We plot in Fig. 3 the output images of the TV-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 TV-L^1 model. Noting that the G-values of \( s_1 \) and \( s_2 \) are the same, we sequentially choose \( \lambda = \lambda_{1,3}^\text{min} \), \( \lambda_{1,3}^\text{max} \), \( \lambda_{1,2}^\text{min} \), \( \lambda_{1,2}^\text{max} \), 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 well-preserved, which is an important feature of TV-L^1 model. It also shows that the algorithm produces very good numerical results.
Fig. 2. Restored images by TV-$L^p$ ($p = 3$) model and the profiles of the intensity of $\{s_i\}_{i=1}^3$ with different $\lambda$. Row 1: $\lambda = \lambda_{\text{min}}^{\text{p},1}, \lambda_{\text{min}}^{\text{p},2}, \lambda_{\text{min}}^{\text{p},3}$, and Row 2: $\lambda = \lambda_{\text{max}}^{\text{p},1}, \lambda_{\text{max}}^{\text{p},2}, \lambda_{\text{max}}^{\text{p},3}$.

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

Fig. 4. Restored images by TV-$L^1$ model and the profiles of the intensity of $\{s_i\}_{i=1}^3$ with different $\lambda$. Row 1: $\lambda = \lambda_{\text{min}}^{\text{1},1}, \lambda_{\text{min}}^{\text{1},2}, \lambda_{\text{min}}^{\text{1},3}$, and Row 2: $\lambda = \lambda_{\text{max}}^{\text{1},1}, \lambda_{\text{max}}^{\text{1},2}, \lambda_{\text{max}}^{\text{1},3}$.

5. Concluding remarks

In this paper, we investigated various properties of the TV-$L^p$ model with general real $p$. As with the TV-$L^1$ 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 TV-$L^p$ (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


