A DUAL EM ALGORITHM FOR TV REGULARIZED GAUSSIAN MIXTURE MODEL IN IMAGE SEGMENTATION

SHI YAN, JUN LIU* AND HAIYANG HUANG

Laboratory of Mathematics and Complex Systems (Ministry of Education of China)
School of Mathematical Sciences, Beijing Normal University, Beijing 100875, China

XUE-CHENG TAI

Department of Mathematics, Hong Kong Baptist University
Kowloon Tong, Hong Kong, China

and

Depart of Mathematics, University of Bergen
P. O. Box 7800, N-5020, Bergen, Norway

(Communicated by Yunmei Chen)

Abstract. A dual expectation-maximization (EM) algorithm for total variation (TV) regularized Gaussian mixture model (GMM) is proposed in this paper. The algorithm is built upon the EM algorithm with TV regularization (EM-TV) model which combines the statistical and variational methods together for image segmentation. Inspired by the projection algorithm proposed by Chambolle, we give a dual algorithm for the EM-TV model. The related dual problem is smooth and can be easily solved by a projection gradient method, which is stable and fast. Given the parameters of GMM, the proposed algorithm can be seen as a forward-backward splitting method which converges. This method can be easily extended to many other applications. Numerical results show that our algorithm can provide high quality segmentation results with fast computation speed. Compared with the well-known statistics based methods such as hidden Markov random field with EM method (HMRF-EM), the proposed algorithm has a better performance. The proposed method could also be applied to MRI segmentation such as SPM8 software and improve the segmentation results.

1. Introduction. Image segmentation is a basic part of the image processing with a long history [22]. Images may come from every corner of the world, and some of the natural images may contain noise and are not good enough to segment directly. There are many methods for image segmentation after several decades of development, one of the most important methods is the statistical method. In such a method, a region with its noise can be approximately presented by a Gaussian mixture model (GMM) [26, 17, 23, 39, 41]. A common method of solving GMM is the well-known expectation-maximization (EM) algorithm, which is firstly introduced in [16]. GMM-EM method has been studied a lot in many references such as [9, 30, 38, 41]. However, when using GMM-EM method in image segmentation, one drawback is its sensitive to noise due to the lack of regularization. To fill this flaw, in [41], the authors proposed a hidden Markov random field method (HMRF-EM) to GMM.

2010 Mathematics Subject Classification. Primary: 68U10.
Key words and phrases. Dual algorithm, expectation-maximization, Gaussian mixture model, total variation, projection gradient, image segmentation.
* Corresponding author: Jun Liu.
and solved it by EM algorithm (HMRF-EM), which can achieve stable segmentation results under noise. In fact, the local spatial positions of the neighborhood pixels are taken into consideration in HMRF, which makes the segmentation more robust to noise. However, the classification provided by HMRF may produce some zigzags near the boundaries due to the improper regularization [25], especially when the levels of noise are high. One may see this phenomenon in the numerical experiments section.

As for regularization, total variation (TV) is one of the most successful methods in inverse problems. TV was introduced by Rudin-Osher-Fatemi in [33], known as ROF model. In fact, Mumford and Shah introduced a multi-region segmentation model in [31], where we can also treat the regularization term as TV, as well in [20, 35]. In [12], Chan and Vese proposed the CV model which combined TV regularization and level set method [32] for two-region image segmentation. Some extensions can be found in [1, 11]. It is well-known that there are many fast algorithms to solve TV based minimization problems, such as the dual method [10], split Bregman method [21], augmented Lagrange method [37], graph-cut (max-flow) method [7, 8], continuous max-flow method [40] and so on. Besides, there is a natural connection between TV and the length of the curves theoretically [19]. From the denoising experimental results, we can see that, as a regularization term, TV can keep the jump between classification and suppress noise, which means that TV is good at preserving edges and removing noise.

In [26], Liu et al. proposed an image segmentation model called EM-TV model by integrating EM and TV. They proposed a unified variational functional which brings EM algorithm and TV regularization together, and this model holds both advantages from the statistical and PDE methods. In their method, the minimization problem was solved by a splitting scheme where a $L^1$ penalty term was used to keep the structure of EM. To be different from the traditional $L^2$ penalty problem, the convergence of the algorithm is unknown due to lack of strict convexity of $L^1$. Besides, they introduced some extra auxiliary variables and many penalty parameters in the algorithm. For real implementations, it is very important and difficult to find some good values of these parameters for such a splitting scheme. Bad parameters may cause the algorithm fail to converge or produce undesirable segmentation results.

In this paper, we introduce a new method to solve this EM-TV model. The idea comes from the Chambolle projection method [10]. With the help of Fenchel duality method, we get a smooth dual problem for the EM-TV model. This dual problem can be easily solved by a projection gradient method, which is fast and stable. Moreover, we can get the convergence of this algorithm. Compared with introducing two auxiliary variables in [26], we only introduce one dual variable during the iteration, and the solution can be given by an explicit formulation. The proposed algorithm does not have many control parameters, which tends to be more stable.

Our method is related to the global minimization method described in [5]. Let us point out the main differences of these two methods. Firstly, the theoretical fundaments are totally different. Our method is built upon the statistics method while theirs is from the viewpoint of deterministic approach. In this paper, we try to solve EM-TV model, but [5] studied the Potts model with approximation method. Thus, our model can segment data with the same mean but with different variances due to the property of GMM. In this case, [5] cannot work since the mean is the
only factor for segmenting images. Secondly, in [5], the authors did not provide the convergence of their algorithm. In this paper, we will mathematically show that our algorithm converges.

The rest of the paper is organized as follows. In section 2, we introduce the basic idea of GMM, EM and the EM-TV model. In section 3, we will show the dual algorithm for solving the EM-TV model. In section 4 describes the discrete method of our algorithm. In section 5, some numerical results and comparisons are presented. Finally, we summarize our algorithm.

2. GMM, EM and EM-TV model.

2.1. GMM and EM. GMM and EM algorithm were first introduced in statistics, which can be used in image segmentation. GMM is a probabilistic model, representing the distribution of multi-class. EM algorithm is a 2-step iterations algorithm, the first step is usually called E-step, which uses current estimate of parameters to calculate the maximum a posteriori (MAP) function, the second step is usually called M-step, which uses the function in E-step to update parameters.

Here we take gray image as example, let \( I(x) \) be a gray image where \( x \in \Omega \subset \mathbb{R}^2 \).

A GMM’s distribution for image pixels can be represented as follows:

\[
p_{\text{GMM}}(I(x)) = \sum_{k=1}^{K} \alpha_k p(I(x); \mu_k, \Sigma_k),
\]

where \( K \) is the number of mixtures, \( \alpha_k \) is the weight of \( k \)-th mixture class and it satisfies \( \sum_{k=1}^{K} \alpha_k = 1 \). \( p \) is the Gaussian density function, with \( p(I(x); \mu_k, \Sigma_k) = \frac{1}{\sqrt{2\pi} \Sigma_k} \exp\left(\frac{-(I(x) - \mu_k)^2}{2\Sigma_k}\right) \) and \( \mu_k, \Sigma_k \) are the mean and variance of the \( k \)-th mixture, respectively. Usually, \( K \) is an integer given by the user as a prior.

The related negative log-likelihood function of GMM can be written as

\[
L(\Theta) = -\int_{\Omega} \log \sum_{k=1}^{K} \alpha_k p(I(x); \mu_k, \Sigma_k) \, dx,
\]

where \( \Theta = \{\alpha^1, \ldots, \alpha^K, \mu^1, \ldots, \mu^K, \Sigma^1, \ldots, \Sigma^K\} \) is the parameter set.

Note that optimizing \( \log \sum_{k=1}^{K} \) type functional directly is not efficient, but it can be addressed by a log-sum operation commutativity property [27, 34]:

**Lemma 2.1.** (Commutativity of log-sum operations [27, 34].) Given functions \( f^k(x) > 0 \), then one can get

\[
-\log \sum_{k=1}^{K} f^k(x) = \min_{\phi \in \mathcal{A}} \left\{ -\sum_{k=1}^{K} \phi_k(x) \log f^k(x) + \sum_{k=1}^{K} \phi_k(x) \log \phi_k(x) \right\},
\]

where \( \phi(x) = (\phi^1(x), \phi^2(x), \ldots, \phi^K(x)) \) is a vector-valued function, and \( \phi \in \mathcal{A} \), with

\[
\mathcal{A} = \{ \phi \mid \sum_{k=1}^{K} \phi_k(x) = 1, \phi_k(x) \in (0, 1), \forall x \in \Omega \}.
\]

Let

\[
H(\phi, \Theta) = -\int_{\Omega} \log[\alpha_k p(I(x); \mu_k, \Sigma_k)] \phi_k(x) \, dx
\]
\[ + \int_{\Omega} \sum_{k=1}^{K} \phi^{k}(x) \log \phi^{k}(x) dx, \phi \in \mathcal{A}, \]

then \(H(\phi, \Theta)\) and \(L(\Theta)\) have the same minimizer \(\Theta^{*}\) with respect to variable \(\Theta\), see more in [27].

It is easy to get \(\phi^{*}\) and \(\Theta^{*}\) from \(H(\phi, \Theta)\) by alternating scheme
\[
\begin{align*}
\phi^{m+1} &= \arg \min_{\phi \in \mathcal{A}} H(\phi, \Theta^{m}), \\
\Theta^{m+1} &= \arg \min_{\Theta} H(\phi^{m+1}, \Theta).
\end{align*}
\]

The above iteration strictly corresponds to the E-step and M-step of EM algorithm, respectively.

By applying the first-order optimization condition and Lagrange multiplier technique, we can easily get the solution of E-step and M-step as follows:

**E-step:** for \(x \in \Omega, k = 1, 2, \cdots, K,\)
\[
(\phi^{k})^{m+1}(x) = \frac{(\alpha^{k})^{m}p(I(x); (\mu^{k})^{m}, (\Sigma^{k})^{m})}{\sum_{l=1}^{K} (\alpha^{l})^{m}p(I(x); (\mu^{l})^{m}, (\Sigma^{l})^{m})}.
\]

**M-step:** for \(k = 1, 2, \cdots, K,\)
\[
\begin{align*}
(\alpha^{k})^{m+1} &= \frac{\sum_{x \in \Omega} (\phi^{k})^{m+1}(x)}{\vert \Omega \vert}, \\
(\mu^{k})^{m+1} &= \frac{\sum_{x \in \Omega} (\phi^{k})^{m+1}(x)I(x)}{\sum_{x \in \Omega} (\phi^{k})^{m+1}(x)}, \\
(\Sigma^{k})^{m+1} &= \frac{\sum_{x \in \Omega} (\phi^{k})^{m+1}(x)I(x) - (\mu^{k})^{m+1})^{2}}{\sum_{x \in \Omega} (\phi^{k})^{m+1}(x)}.
\end{align*}
\]

2.2. **TV regularization.** TV regularization was first introduced in ROF model in [33], which aims at noise removal.

In this paper, we use the rotation invariant isotropic TV as
\[
TV(\phi) = \int_{\Omega} \sqrt{\phi x_1(x)^2 + \phi x_2(x)^2} dx,
\]
due to its better performance than anisotropic TV [28], since it can provide more precise boundaries.

2.3. **EM-TV model.** In [26], Liu et al. proposed EM-TV model, which combines GMM-EM method with TV regularization, and got some good results.

GMM-EM algorithm clusters pixels only by the intensity of pixels, which does not take any spatial or geometric information into the segmentation. In [26], the authors added the TV regularization into the energy (2) to get the EM-TV model:
\[
\begin{align*}
\min_{\phi \in \mathcal{A}, \Theta} E(\phi, \Theta) &= \min_{\phi \in \mathcal{A}, \Theta} \{H(\phi, \Theta) + \gamma \sum_{k=1}^{K} TV(\phi^{k})\} \\
&= \min_{\phi \in \mathcal{A}, \Theta} \left\{ - \int_{\Omega} \sum_{k=1}^{K} \log(\alpha^{k}p^{k}(x; \mu^{k}, \Sigma^{k}))\phi^{k}(x) dx \\
& \quad + \int_{\Omega} \sum_{k=1}^{K} \phi^{k}(x) \log \phi^{k}(x) dx + \gamma \sum_{k=1}^{K} TV(\phi^{k}) \right\}.
\end{align*}
\]

Here \(\gamma\) is a regularization parameter.
3. The proposed method. In this section, we discuss how to solve (5) efficiently. Alternating method is applied to solve \( \phi \) and \( \Theta \), which can be written as,

\[
\begin{aligned}
\phi^{m+1} &= \arg \min_{\phi \in \Lambda} \{ E(\phi, \Theta^m) \}, \\
\Theta^{m+1} &= \arg \min_{\Theta} \{ E(\phi^{m+1}, \Theta) \}.
\end{aligned}
\]

(6)

For the minimization of \( \phi \), we shall propose a dual method due to existence of TV norm. This dual algorithm is built upon [10]. Compared with the original dual algorithm [10], there are some segmentation constraints and nonlinear entropy regularization in EM-TV model, which make this problem more difficult. For simplification, we denote our method as Dual EM-TV.

3.1. Minimization of \( \phi \) (Dual EM-TV). Fix \( \Theta^m \), the sub-problem for \( \phi \) is,

\[
\min_{\phi \in \Lambda} \{ E(\phi, \Theta^m) \} = \int_\Omega \sum_{k=1}^K f^k(x) \phi^k(x) dx + \int_\Omega \sum_{k=1}^K \phi^k(x) \log \phi^k(x) dx + \gamma \sum_{k=1}^K TV(\phi^k),
\]

where \((f^k)^m(x) = -\log((\alpha^k)^m p(I(x)); (\mu^k)^m, (\Sigma^k)^m)\). For shortness, in the next, we omit the superscript \( m \) of \((f^k)^m(x)\), and write it as \( f^k(x) \) in this section.

Note that there is a constraint for \( \phi \), we use Lagrange multiplier method and get the following unconstrained energy minimization problem,

\[
\min_{\phi} \{ E_L(\phi, \Theta^m) \} = \int_\Omega \sum_{k=1}^K f^k(x) \phi^k(x) dx + \int_\Omega \sum_{k=1}^K \phi^k(x) \log \phi^k(x) dx \\
+ \gamma \sum_{k=1}^K TV(\phi^k) + \int_\Omega \lambda(x) \left( \sum_{k=1}^K \phi^k(x) - 1 \right) dx,
\]

(8)

It is easy to see that \( E_L(\phi, \Theta^m) \) is convex with respect to \( \phi \), when \( \phi^k(x) > 0, \forall k \in \{1, 2, \cdots, K\}, x \in \Omega \). Moreover, this energy can be separated and thus we can deal with each \( k \) separately as follows,

\[
E_L(\phi, \Theta^m) = \sum_{k=1}^K \left[ \int_\Omega f^k(x) \phi^k(x) dx + \int_\Omega \phi^k(x) \log \phi^k(x) dx \\
+ \gamma TV(\phi^k) + \int_\Omega \lambda(x) \phi^k(x) dx \right] - \int_\Omega \lambda(x) dx.
\]

We can see that the intersection of domain of each term is nonempty, which means the subgradient of \( E_L(\phi, \Theta^m) \) is equal to the sum of subgradient of each term in \( E_L(\phi, \Theta^m) \). Now the necessary and sufficient condition for problem (8) with its minimizer \( \phi^* = ((\phi^*)^1, (\phi^*)^2, \cdots, (\phi^*)^K) \) satisfies

\[
0 \in f^k + \log((\phi^*)^k) + 1 + \lambda + \gamma \partial TV((\phi^*)^k).
\]

(9)

Here \( \partial TV(u) \) is the subdifferential of \( TV(u) \) at \( u \), defined as

\[
\partial TV(u) = \{ \omega | TV(v) \geq TV(u) + \langle \omega, v - u \rangle, \forall v \in \Lambda \},
\]

where \( \langle \cdot, \cdot \rangle \) denotes the corresponding \( L^2 \) inner product.

Then we have:

\[
-(f^k + \log((\phi^*)^k) + 1 + \lambda) \in \partial TV((\phi^*)^k),
\]

\[
\int_\Omega \sum_{k=1}^K f^k(x) \phi^k(x) dx + \int_\Omega \sum_{k=1}^K \phi^k(x) \log \phi^k(x) dx + \gamma \sum_{k=1}^K TV(\phi^k) + \int_\Omega \lambda(x) \left( \sum_{k=1}^K \phi^k(x) - 1 \right) dx,
\]

\[
\int_\Omega \sum_{k=1}^K f^k(x) \phi^k(x) dx + \int_\Omega \sum_{k=1}^K \phi^k(x) \log \phi^k(x) dx + \gamma \sum_{k=1}^K TV(\phi^k) + \int_\Omega \lambda(x) \left( \sum_{k=1}^K \phi^k(x) - 1 \right) dx,
\]
and we get
\[ (\phi^*)^k \in \partial TV^* \left( \frac{-(f^k + \log(\phi^*)^k + 1 + \lambda)}{\gamma} \right), \]
where \( TV^* \) is the Legendre-Fenchel transform of \( TV \), defined by
\[ TV^*(v) = \sup_{u \in \mathbb{A}} \{ (u, v) - TV(u) \}. \]

It is well-known that \( TV^* \) is a "characteristic function" of set \( \mathbb{B} \):
\[ TV^*(v) = \begin{cases} 0, & \text{if } v \in \mathbb{B}, \\ +\infty, & \text{otherwise}, \end{cases} \]
\[ \mathbb{B} = \{ v = \text{div}\mathbf{\eta} = (\eta^1, \eta^2) \in C^1_c(\Omega, \mathbb{R}^2), ||\mathbf{\eta}||_\infty \leq 1 \}, \]
where \( ||\mathbf{\eta}||_\infty = \max_{x \in \Omega} \sqrt{\eta^1(x)^2 + \eta^2(x)^2} \).

More details about the dual presentation of TV norm can be found in Chambolle projection method [10], as well as in [18, 24].

Now, let
\[ (g^*)^k(x) = \frac{-(f^k(x) + \log(\phi^*)^k(x) + 1 + \lambda(x))}{\gamma}, x \in \Omega, \]
and we get the equivalent form
\[ (\phi^*)^k(x) = \exp(-(f^k(x) + 1 + \lambda(x)) - \gamma(g^*)^k(x)), x \in \Omega. \]

Put (12)(13) into (10), we have
\[ 0 \leq -\exp(-(f^k + 1 + \lambda) - \gamma(g^*)^k) + \partial TV^*((g^*)^k), \]
which indicates that \( (g^*)^k \) is the minimizer of energy,
\[ F(g^k) = \int_\Omega \frac{1}{\gamma} \exp(-(f^k(x) + 1 + \lambda(x)) - \gamma(g^*)^k(x)) \, dx + TV^*(g^k). \]

By (11), if \( g^k \notin \mathbb{B}, F(g^k) = +\infty \), thus (14) can be written as a constraint minimization problem
\[ \min_{g^k} F(g^k) = \min_{g^k \in \mathbb{B}} \frac{1}{\gamma} \int_\Omega \exp(-(f^k(x) + 1 + \lambda(x)) - \gamma(g^*)^k(x)) \, dx. \]

According to the definition of \( \mathbb{B} \), computing \( (g^*)^k \) can be transferred to solving the following problem:
\[ (\eta^*)^k = \arg \min_{\eta^k \in \tilde{\mathbb{B}}} \frac{1}{\gamma} \int_\Omega \exp(-(f^k(x) + 1 + \lambda(x)) - \gamma \text{div}\eta^k(x)) \, dx, \]
in which, \( g^k \) and \( \eta^k \) are related as \( g^k = \text{div}\mathbf{\eta}^k \) and \( \tilde{\mathbb{B}} = \{ \eta^k \mid ||\eta^k||_\infty \leq 1, \eta \in C^1_c(\Omega, \mathbb{R}^2) \}. \)

We can use the projection gradient method to solve (16), then one can get the iteration scheme
\[ (\eta^*)^{k+1} = \text{Proj}_\tilde{\mathbb{B}}((\eta^*)^k - \Delta t \nabla \exp(-(f^k(x) + 1 + \lambda(x)) - \gamma \text{div}\eta^k(x)))) \]
Where the operator \( \text{Proj}_\tilde{\mathbb{B}}(\xi) \) means we project \( \xi \) onto a convex set \( \tilde{\mathbb{B}} \), and \( \Delta t \) is a proper time step.

As for the Lagrange multiplier, from (13), we can give out the closed-form solution of \( \lambda(x) \) by
\[ \lambda(x) = \ln \left( \sum_{k=1}^K \exp(-(f^k(x) - \gamma(g^*)^k(x))) - 1, \right) \]
if \((g^*)^k(x)\) is known. Thus, we set the Lagrange multiplier as

\[
\lambda^m(x) = \ln \left( \sum_{k=1}^{K} \exp(- (f^k)^m(x) - \gamma \text{div}(\eta^k)^m(x)) \right) - 1.
\]

And fix \(\lambda = \lambda^m(x)\) when solving \(\eta\).

Now put \(\lambda\) back into (17), we can get the iteration of \((\eta^k)^n\) as

\[
(\eta^k)^{n+1} = \text{Proj}_{\mathbb{B}}((\eta^k)^n) - \Delta t \nabla \left( \sum_{i=1}^{K} (\alpha^k)^m p(I(x); (\mu^k)^m, (\Sigma^k)^m) \exp(- \gamma \text{div}(\eta^k)^m(x)) \right).
\]

3.2. Minimization of \(\Theta\). At the \(m\)-iteration, we fix \((\phi^k)^{m+1}\). Then we have to minimize such an energy respect to \(\Theta\):

\[
E(\phi^{m+1}, \Theta) = - \int \sum_{k=1}^{K} \log[\alpha^k p(I(x); \mu^k, \Sigma^k)](\phi^k)^{m+1}(x)dx.
\]

This optimization problem is the same as the classical EM model, so it can be updated by (4).

4. Implementation. Here we introduce the discrete scheme of our algorithm. First, let \(\mathbb{X} \subset \mathbb{R}^2\) be a \(M \times N\) rectangle area, and \(\mathbb{F}_1 = \{f : \mathbb{X} \to \mathbb{R}\}, \mathbb{F}_2 = \{\eta : \mathbb{X} \to \mathbb{R}^2\}\) are two function spaces defined on \(\mathbb{X}\). \(\forall f \in \mathbb{F}_1, \forall \eta = (\eta^1, \eta^2) \in \mathbb{F}_2, f_{i,j}, \eta_{i,j} = (\eta_{i,j}^1, \eta_{i,j}^2)\) stands for \(f(x), \eta = (\eta^1, \eta^2)\) at location \((i, j)\), respectively, here \(i = 1, 2, \ldots, M, j = 1, 2, \ldots, N\).

The discrete operator \(\nabla_d : \mathbb{F}_1 \to \mathbb{F}_2\) with Neumann boundary condition can be defined as

\[
(\nabla_d f)_{i,j} = \begin{cases} f_{i+1,j} - f_{i,j}, & \text{if } i < M, \\ 0, & \text{if } i = M, \\ f_{i,j-1} - f_{i,j}, & \text{if } j < N, \\ 0, & \text{if } j = N. 
\end{cases}
\]

It is not difficult to calculate that the discrete adjoint operator \(\text{div}_d : \mathbb{F}_2 \to \mathbb{F}_1\) has the formulation

\[
(\text{div}_d(\eta))_{i,j} = \text{div}_d((\eta_{i,j}^1, \eta_{i,j}^2)) = \eta_{i,j+1}^1 - \eta_{i,j-1}^1, \quad \text{if } i = 1, \quad \eta_{i,j+1}^2 - \eta_{i,j-1}^2, \quad \text{if } j = 1, \quad -\eta_{i-1,j}^1, \quad \text{if } j = N.
\]

The projection operator \(\text{Proj}_{\mathbb{B}_d} : \mathbb{F}_2 \to \mathbb{B}_d\) has the discrete formulation

\[
(\text{Proj}_{\mathbb{B}_d}(\eta))_{i,j} = \begin{cases} \eta_{i,j}^1 - \eta_{i,j}^2, & \text{if } 1 < j < N, \\ \eta_{i,j}^1, \eta_{i,j}^2, & \text{if } j = 1, \quad \eta_{i,j}^1, \eta_{i,j}^2, & \text{if } j = N. 
\end{cases}
\]

Note that \(\mathbb{B}_d = \{\eta^k \mid ||\eta^k||_\infty \leq 1, \forall \eta \in C^0_c(\mathbb{X}, \mathbb{R}^2)\}\) is also discrete.

The corresponding discrete scheme of the main iteration becomes

\[
(\eta^k)^{n+1} = \text{Proj}_{\mathbb{B}_d}((\eta^k)^n) - \Delta t \nabla_d \left( \sum_{i=1}^{K} (\alpha^k)^m p(I(x); (\mu^k)^m, (\Sigma^k)^m) \exp(- \gamma \text{div}_d(\eta^k)^m(x)) \right).
\]
Remark 1. In practice, we might slightly modify the second term: \( \int_{\Omega} \sum_{k=1}^{K} f^k(x) \log f^k(x) dx \) as \( \delta \int_{\Omega} \sum_{k=1}^{K} f^k(x) \log f^k(x) dx \). This is actually equivalent to the continuous smooth method in [2, 13]. Then our iteration of \( \eta \) becomes:

\[
(\eta^k)^{n+1} = \text{Proj}_{\mathcal{G}}(\eta^k)^n - \Delta t \nabla \phi \left( \frac{1}{\gamma} \sum_{i=1}^{K} (\alpha^i)^m p(I(x); \mu^i)^m, (\Sigma^i)^m) \right) \exp\left( -\frac{\gamma \text{div}_d(\eta^k)^n(x)}{\delta} \right).
\]

### 4.1. Summary of Algorithm

Now, we can summarize our algorithm in Algorithm 1:

Algorithm 1. Dual EM-TV Algorithm

Set \( \phi^0 = (1, 0, \ldots, 0) \), and \( \Theta^0 \) by several EM iteration (3) (4). Let \( m = 0 \).

1. Dual step: Get \( \eta^{m+1} \) by (21) until convergence, i.e. \( \eta^{m+1} = \lim_{n \to \infty} \eta^{n+1} \).
2. Prime step: Get the smooth segmentation \( \phi^{m+1} \) by (13).
3. Convergence checking. If \( \frac{||\phi^{m+1} - \phi^m||}{||\phi^m||} > \epsilon \), then go to the next step. Else, end the algorithm.
4. Update the parameter \( \Theta^{m+1} \) using (4).
5. let \( m = m + 1 \).

### 4.2. Convergence analysis

For the \( \eta \)-sub problem, we have the following convergence result.

**Proposition 1.** Assume \( f \) is bounded by \( M \), i.e. \( |f_{ij}| \leq M \). Then for all \( 0 < \Delta t < \frac{K}{\|\phi^0\|} \exp\left( -(2M + 8\gamma) \right) \), the sequence generated by (21) converges to one \( \eta^* \in \mathcal{G} \) if the fixed point set \( \mathcal{G} \) of iteration scheme (21) is nonempty. Here \( K \) is the number of classification and \( \gamma \) is the regularization parameter.

We put the proof of Proposition 1 in the appendix section.

5. **Experiment results.** In this section, we demonstrate the segmentation results. The computing platform is a laptop equipped with Processor Intel Core i7-8550U CPU @ 1.80Hz and Matlab R2010b. We take HMRF-EM method [36, 41] as counterparty. In HMRF method, the Gibbs prior is applied as a regularization, while the Gibbs distribution is given by

\[
P(\mathbf{x}) = Z^{-1} \exp(-U(\mathbf{x})),
\]

where \( Z \) is a normalization constant. The prior energy function is set to \( U(x) = \gamma \sum_{y \in N_x} (1 - \text{Class}(x, y)) \), where \( N_x \) denotes the neighborhood of point \( x \).

\[
\text{Class}(x, y) = \begin{cases} 
0 & \text{label}(x) \neq \text{label}(y), \\
1 & \text{label}(x) = \text{label}(y).
\end{cases}
\]

In our numerical experiments, the size of \( N_x \) in regularization term of HMRF-EM is chosen as \( |N_x| = 4, 8, 12, 20 \) neighborhoods, respectively. For short, we call them as HMRF-EM 4n, HMRF-EM 8n, HMRF-EM 12n, HMRF-EM 20n, respectively. The codes for running HMRF-EM are got from [https://se.mathworks.com/matlabcentral/fileexchange/37530-hmrf-em-image](https://se.mathworks.com/matlabcentral/fileexchange/37530-hmrf-em-image) in [36].
5.1. **Parameter selection.** A possible way to speed up the algorithm is to use the Armijo rule [42] to choose the step size $\Delta t$ in (21) by line search. One can set the theoretical upper bound of $\Delta t$ appeared in the proposition 1 as the Armijo rule’s low bound ($\alpha_{\min}$ in [42]). We can also set the initial $\Delta t$ as some proper constants if fast convergence is preferred. In this paper, we do not further study the optimal step size and just use some constants as step size.

There are two parameters $\delta$ and $\gamma$ in our model. $\delta$ controls the smoothness of the classification function. In the experiment, $\delta$ is not very sensitive, so it can be chosen in a wide range. Generally speaking, the heavier noise, the larger $\delta$ should be taken. In our experiments, we choose $\delta$ from 1 to 1000. $\gamma$ is the regularization parameter, it would affect the smoothness of the boundary.

The initialization of parameters for Dual EM-TV method is as follows: $\alpha, \mu, \Sigma$ are got from K-means method. $\eta_k = 0, k = 1, 2, \ldots, K$. $\gamma$ and $\delta$ are set as different values for each image: in heart image, $\delta = 20$, $\gamma = 18$; in 4-color image: $\delta = 20$, $\gamma = 12$. The time step for Dual EM-TV method is $\Delta t = 0.5$ for heart image, and $\Delta t = 1$ for 4 color image.

The initialization of parameters for HMRF-EM method is as follows: $\alpha, \mu, \Sigma$ are got from K-means method as well. For 4-neighborhood, $\gamma = 320$, for 8-neighborhood: $\gamma = 160$, for 20-neighborhood: $\gamma = 40$.

5.2. **Comparison of dual EM-TV, Original EM-TV [26] and HMRF-EM model.** Here, we compare the segmentation results of these three methods on two synthetic images. These two gray images are both normalized to $[0,1]$. We add additive white Gaussian noise with standard deviation $150/255$ and $50/255$ into heart image (size $615 \times 615$) and 4-color image (size $600 \times 600$), respectively.

To compare the segmentation results, we use the segmentation accuracy index (SAI)

$$SAI = \frac{N_c}{N_t},$$



to measure the classification quality. Here $N_c$ is the correctly labeled points, while $N_t$ is the total number of points.

The CPU time and accuracy results are listed in Table 1 and Table 2. The segmentation results comparisons are displayed in Figures 1, 2, 3, 4. In Figures 1 and 3, the first row contains original, noisy images and the energy evaluation of the Dual EM-TV method. While the second and third rows are the ground truth, segmentation results of Dual EM-TV, Original EM-TV [26], HMRF-EM 4n, HMRF-EM 8n, HMRF-EM 20n, respectively. In Figure 2 and 4, the enlarged images of the related red square areas in Figure 1 and 3 are shown.

One can find that our algorithm is faster than HMRF-EM model and provides better segmentation results. Compared to the original EM-TV method, our proposed method produced slightly better results and costed much less CPU time.

5.3. **Comparison of dual EM-TV and CV method.** In this subsection, we show an example that our model can segment an image with two regions which have the same means but different variances.

We first synthetic two regions with the same mean but different variances in Figure 5(a). The size of image is $100 \times 100$, the mean of the $70 \times 70$ inside region is 128, and the variance is 1000. While the mean and variance of outside region are set as 128, 10, respectively. Due to the normalization, the middle of the histogram is
Table 1. Comparison of 2 classes segmentation for Dual EM-TV, Original EM-TV [26] and HMRF-EM method (Heart image with size 615 × 615).

| Methods         | CPU time (seconds) | Accuracy (SAI) |
|-----------------|--------------------|----------------|
| Dual EM-TV      | 1.233              | 99.70%         |
| Original EM-TV  | 9.482              | 98.02%         |
| HMRF-EM 4n      | 4.052              | 91.84%         |
| HMRF-EM 8n      | 3.052              | 97.38%         |
| HMRF-EM 20n     | 1.528              | 99.24%         |

Table 2. Comparison of 4 classes segmentation for Dual EM-TV, Original EM-TV [26] and HMRF-EM methods (4-color image with size 600 × 600).

| Methods         | CPU time (seconds) | Accuracy (SAI) |
|-----------------|--------------------|----------------|
| Dual EM-TV      | 1.699              | 99.90%         |
| Original EM-TV  | 18.502             | 99.02%         |
| HMRF-EM 4n      | 3.480              | 97.39%         |
| HMRF-EM 8n      | 3.202              | 99.81%         |
| HMRF-EM 20n     | 2.388              | 99.83%         |

slightly apart from 128. The histogram in Figure 5(b-d) clearly shows the difference of variance. Please see Figure 5 for details.

Then we apply GMM-EM method on Figure 5(a) and gets the segmentation result on Figure 6(a). And then we use the parameters got from this GMM-EM method and apply out Dual EM-TV method with the segmentation result in Figure 6(b). Figure 6(c) is the segmentation result of CV model. As you can see that GMM-EM based method successfully segmented the target image while CV cannot.

5.4. Natural image segmentation. Here, we show more comparisons for Dual EM-TV and HMRF-EM method in Figure 7. We add Gaussian noise with mean 0, standard deviation of 50/255. We can see that our algorithm can get a very good segmentation. In Figure 7, the first and second columns contain the original and noisy images. All the segmentation results provided by Dual EM-TV, HMRF-EM 4n, HMRF-EM 8n, HMRF-EM 12n are displayed in the right four columns, respectively.

We list the parameters appeared in this experiment: row 1 (the cameraman), \( \Delta t = 1, \delta = 1000, \gamma = 400 \); row 2 (synthetic image objects), \( \Delta t = 1, \delta = 1000, \gamma = 400 \); row 3 (synthetic image 2), \( \Delta t = 1, \delta = 1000, \gamma = 480 \). row 4 (starfish), \( \Delta t = 0.5, \delta = 40, \gamma = 39 \).

5.5. Color images. It is easy to extend the GMM based model to color images. We take some real-life images from database [3, 29] to test our algorithm. In Figure 8 and 9, we apply our model to images with two-region and multi-region, respectively. In these figures, the original images are shown on the left column and the
Figure 1. Comparison of Dual EM-TV, Original EM-TV [26] and HMRF-EM methods for 2 classes segmentation. The segmentation results are displayed in the right column. These results show that our algorithm can produce a good segmentation results.

The parameters we used are as follows. In Figure 8, row 1 (bear image), we let $\Delta t = 0.5, \gamma = \delta = 10$; row 2 (eagle image), $\Delta t = 0.5, \gamma = 7, \delta = 10$; row 3 (horse image), $\Delta t = 0.1, \gamma = 50, \delta = 10$; row 4 (desert image), $\Delta t = 0.5, \gamma = \delta = 10$. In Figure 9: row 1 (moon image), we let $\Delta t = 0.5, \gamma = 2, \delta = 10$; row 2 (stone image) $\Delta t = 0.005, \gamma = 80, \delta = 1$; row 3 (temple image) $\Delta t = 0.005, \gamma = 80, \delta = 1$.

Remark 2. The color images are all segmented in both RGB and HSI format and we choose the better result here.

5.6. Under different levels of noise. Here, we discuss how the noise levels affect our model and the segmentation results. Figure 10 is a two class segmentation.
results. The noise is additive white Gaussian noise, with standard deviation of 120/255, 240/255, 300/255, 390/255, respectively. In Figure 11, We add additive white Gaussian noise, with standard deviation of 50/255, 60/255, 70/255, 80/255, respectively, and test the outcomes. As you can see, we can get better results than HMRF-EM model under different levels of noise.

Parameters are as follows. For heart image with noise 120/255, 240/255, 300/255, $\Delta t = 0.1, \gamma = 60, \delta = 20$, for noise 390/255, $\Delta t = 0.1, \gamma = 80, \delta = 20$. For 4-color image with noise 50/255, $\Delta t = 0.1, \gamma = 12, \delta = 20$, for noise 60/255, $\Delta t = 0.01, \gamma = 12, \delta = 2$, for noise 70/255, $\Delta t = 0.01, \gamma = 18, \delta = 2$, for noise 80/255, $\Delta t = 0.01, \gamma = 18, \delta = 2$.

5.7. Applied to 3-D MR images. In Figure 12, we use the proposed method instead of the segmentation method in [4] on 3-D MR images, based on SPM8 (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/), which is a MATLAB toolbox for MRI analysis. To compare, we based on the newsegment module in [4] to test our method, and apply to the MR images from BrainWeb (http://brainweb.bic.mni.mcgill.ca/brainweb/). The image dimension is $181 \times 217 \times 181$, with voxels of $1 \times 1 \times 1$mm. The data with image intensity non-uniformity 40% RF (a parameter specifies the intensity non-uniformity level, larger value responding to higher level) and level of noise 9% is tested.

The parameters are as follows: $\Delta t = 0.0001, \gamma = \delta = 1$.

We use Dice metric (DM) to compare the segmentation accuracy, as

$$DM = \frac{2N_{ag}}{N_a + N_g},$$
Figure 3. Comparison of Dual EM-TV, Original EM-TV [26] and HMRF-EM methods for 4 classes segmentation.

where \( N_{ag} \) denotes the number of the voxels that are correctly assigned to the \( k \)-th class by both the ground truth and the segmentation algorithm, \( N_a \) and \( N_g \) are the numbers of the voxels assigned to the \( k \)-th class by the algorithm and ground truth, respectively.

We compare our results with new segment in SPM 8, the DM is listed in Table 3, WM denotes white matter, GM denotes gray matter. As you can see that our method is slightly better.

6. Discussion and conclusion. In this paper, we proposed a dual algorithm for solving the EM-TV model. To combine the dual, project gradient and Lagrange multiplier methods, we get an elegant dual EM algorithm which is fast and stable. Compared with the original EM-TV model, the optimization problem is smooth and
Figure 4. Enlarged red square regions in Figure 3

Table 3. Comparison of DM in SPM8 on brain images

| noise level | 5% | 5% | 7% | 7% | 9% | 9% |
|-------------|----|----|----|----|----|----|
| brain part  | WM | GM | WM | GM | WM | GM |
| Dual EM-TV  | 0.9380 | 0.9122 | 0.9218 | 0.8970 | 0.9035 | 0.8818 |
| New Segment | 0.9317 | 0.9099 | 0.9035 | 0.8822 | 0.8759 | 0.8536 |

easier to solve. Compared with HMRF-EM model, our algorithm is faster and has better segmentation performance since the isotropic TV cannot be solved by HMRF type algorithm. Based on the experiment results, we can see that our algorithm can produce good segmentation results under high levels of noise, and can also achieve some good segmentations on both color and gray images. The proposed method could also be applied to SPMS and gets good results. Our method can be extended to non-parameter segmentation mixture model such as kernel method which is parameter free. We will leave these as our future research.

Acknowledgments. Jun Liu and Haiyang Huang were partially supported by The National Key Research and Development Program of China (2017YFA0604903). Jun Liu was also supported by the National Natural Science Foundation of China (No. 11871035). Shi Yan was partially supported by China Scholarship Council. The research of Tai is supported by HKBU startup grant, RG(R)- RC/17-18/02-MATH, and FRG2/17-18/033.
Figure 5. An image with two regions, which are the same mean but different variance, and the histogram.

Figure 6. Comparison of segmentation results of GMM-EM, Dual EM-TV and CV model on region with same mean but different variances image.
7. Appendix A. For self-contained in this paper, we give out the proof of Proposition 1.

Let $T : \mathbb{F}_1 \to \mathbb{F}_2$ be an operator.

Introduce notations,

$$T_1(\eta) = \text{Proj}_{\mathbb{F}_2}(\eta),$$

$$T_2(\eta) = (I - \Delta t T_{20})(\eta),$$

$$T_{20}(\eta) = \nabla_d (\exp(-(f^k + 1 + \lambda^m) - \gamma \text{div}_d \eta)).$$

**Definition 7.1** (Lipschitz continuous). $T$ is $L$-Lipschitz continuous, $L \in \mathbb{R}$, if

$$||Tx - Ty|| \leq L||x - y||, \forall x, y \in \mathbb{F}_1.$$

**Definition 7.2** (Non-expansive operator). $T$ is a non-expansive operator if $T$ is 1-Lipschitz continuous, i.e.,

$$||T(x) - T(y)|| \leq ||x - y||, \forall x, y \in \mathbb{F}_1.$$

**Definition 7.3** ($\alpha$-average non-expansive operator). $T$ is an $\alpha$-average non-expansive operator if there exists a non-expansive operator $R$ and $\alpha \in (0, 1)$, s.t.

$$T = (1 - \alpha)I + \alpha R,$$

here $I$ is the identity operator, i.e., $I(x) = x, \forall x$.

An equivalent definition of $\alpha$-average non-expansive operator is, if and only if $T$ satisfies:

$$||T(x) - T(y)||^2 \leq ||x - y||^2 - \frac{1 - \alpha}{\alpha} ||(I - T)(x) - (I - T)(y)||^2.$$
Definition 7.4 (Firmly non-expansive operator). $T$ is firmly non-expansive if and only if $T$ is a $\frac{1}{2}$-average non-expansive operator. Another equivalent definition of firmly non-expansive is that the following inequality holds,

$$||T(x) - T(y)||^2 \leq \langle T(x) - T(y), x - y \rangle.$$  

In our paper, the spaces $\mathbb{F}_1, \mathbb{F}_2$ are often taken as $\mathbb{R}$ or $\mathbb{R}^n$. In the next, we give some properties of $T_1, T_2, T_20$ appeared in our algorithm.
Figure 9. Segmentation results by Dual EM-TV in color images for multi-region images.

Proposition 2. [15] $T_1$ is firmly non-expansive, i.e. $T_1$ is a $\frac{1}{2}$-average non-expansive operator.

Proof. $T_1$ is a projection operator, which is a special proximate operator, and thus it is firmly non-expansive (more details please find in [15]).

Proposition 3. If $f$ is bounded by $M$, then $T_{20}$ is $\beta$-Lipschitz continuous with $\beta = \frac{8\gamma K}{M} \exp (2M + 8\gamma)$.

To prove Proposition 3, we need the following lemmas, and where $K$ is the classification number and $\gamma$ is the regularization parameter.

Lemma 7.5. Under the assumption of Proposition 3, $\exp(-(f^k + 1 + \lambda m))$ is bounded by

$$\left[ \frac{\exp(-2M - 4\gamma)}{K}, \frac{\exp(2M + 4\gamma)}{K} \right].$$

Proof.

$$\exp(-(f^k(x) + \lambda m(x) + 1) = \frac{\exp(-f^k(x))}{\sum_{i=1}^{K} \exp(-f^i(x) - \gamma \div_d(\eta^i)m(x))}.$$
Figure 10. Segmentation comparison under different noise levels. Noise standard deviation: A: 120/255, B: 240/255, C: 300/255, D: 390/255.

Since \((\eta^k)^m \in [-1, 1]\), then \(\text{div}_d(\eta^k)^m \in [-4, 4]\), due to the definition of \(\text{div}_d\). Then we know that
\[
\exp\left(-f^k(x) + \lambda^m(x) + 1\right) \in \left[\frac{\exp(-2M - 4\gamma)}{K}, \frac{\exp(2M + 4\gamma)}{K}\right].
\]

Here we denote \(K_0 = \frac{\exp(2M + 4\gamma)}{K}\).

**Lemma 7.6.** \(\nabla_d\) is a \(K_1\)-Lipschitz linear operator, where \(K_1 = \sqrt{8}\).

**Proof.** It is easy to check \(\nabla_d\) is linear. Moreover
\[
\|\nabla_d g\|^2 = \sum_{i,j} (g_{i+1,j} - g_{i,j})^2 + (g_{i,j+1} - g_{i,j})^2 \\
\leq \sum_{i,j} 2(g_{i+1,j})^2 + 2(g_{i,j})^2 + 2(g_{i,j+1})^2 + 2(g_{i,j})^2 \\
= 8\|g\|^2,
\]
which means \(\|\nabla_d g\| \leq K_1\|g\|\) with \(K_1 = \sqrt{8}\). □

**Lemma 7.7.** \(\exp(z)\) satisfies \(K_2\)-Lipschitz condition when \(z \in [-4\gamma, 4\gamma]\), \(K_2 = \exp(4\gamma)\).

**Proof.** We can see that \(\exp(z)\) is only affected by the \((i,j)\) position, so we can consider it in each point as in the one dimensional case:
\[
|\exp(z) - \exp(y)| = |\exp(\delta)||z - y|,
\]
where \(\delta\) is between \(z\) and \(y\) by the differential mean value theorem. And since \(\exp(z)\) is monotone, we can say
\[
|\exp(z) - \exp(y)| \leq |\exp(\max\{z,y\})||z - y|.
\]
While in our algorithm, $z = -\gamma \text{div}_d \eta^n$, then $z \in [-4\gamma, 4\gamma]$, and let $K_2 = \exp(4\gamma)$.

Now $\forall g_1, g_2 \in F_1$

$$||\exp(g_1) - \exp(g_2)|| \leq K_2 ||g_1 - g_2||.$$

**Lemma 7.8.** $\text{div}_d$ is a $K_3$-Lipschitz linear operator, $K_3 = \sqrt{8}$.

**Proof.** Same as $\nabla_d$, we can get $||\text{div}_d \eta|| \leq K_3 ||\eta||$, where $K_3 = \sqrt{8}$. \hfill \Box

Now we show the proof of Proposition 3.

**Proof.** By the Lemmas 7.5-7.8, we have

$$||T_{20}(\eta) - T_{20}(\theta)|| = ||\nabla_d(\exp(-f^k + 1 + \lambda^m) - \gamma \text{div}_d \eta) - \nabla_d(\exp(-f^k + 1 + \lambda^m) - \gamma \text{div}_d \theta)||$$

$$||\nabla_d(\exp(-f^k + 1 + \lambda^m))(\exp(-\gamma \text{div}_d \eta) - \exp(-\gamma \text{div}_d \theta))||$$

$$\leq ||\nabla_d|| ||(\exp(-f^k + 1 + \lambda^m))(\exp(-\gamma \text{div}_d \eta) - \exp(-\gamma \text{div}_d \theta))||$$

$$\leq K_1 K_0 ||(\exp(-\gamma \text{div}_d \eta) - \exp(-\gamma \text{div}_d \theta))||$$

$$\leq K_1 K_0 K_2 ||(-\gamma \text{div}_d \eta) - (-\gamma \text{div}_d \theta)||$$

$$\leq K_1 K_0 K_2 \gamma ||(\text{div}_d \eta) - (\text{div}_d \theta)||$$

$$\leq \gamma K_1 K_0 K_2 K_3 ||\eta - \theta||.$$

Let

$$\beta = \gamma K_1 K_0 K_2 K_3 = \frac{8\gamma}{K} \exp(2M + 8\gamma),$$

**Figure 11.** Segmentation comparison of the same multi-region image by different noise levels. Noise standard deviation: A: 50/255, B: 60/255, C: 70/255, D: 80/255.
we complete the proof.

**Proposition 4.** $T_2$ is $\alpha_2$-average non-expansive operators if

$$0 < \Delta t < \frac{K}{4\gamma} \exp(-(2M + 8\gamma)).$$

**Proof.** According to Proposition 3, $T_{20}$ is $\beta$-Lipschitz. Moreover, $T_{20}$ is a form of $\nabla H$ with $H$ is a convex function with respect to $\eta$, so $\frac{1}{\beta}T_{20}$ is firmly non-expansive by [6], then there exist a non-expansive operator $R$, s.t. $\frac{1}{\beta}T_{20} = \frac{1}{2}(I + R)$. Back to $T_2$, we have $T_2 = I - \Delta t T_{20} = (1 - \Delta t \frac{\beta}{2})I + \frac{\Delta t \beta}{2}(-R)$ (Note that if $R$ is non-expansive, $-R$ is non-expansive too). Now if we let $(1 - \Delta t \frac{\beta}{2}) \in (0, 1)$, then $T_2$ is $\alpha_2$-average non-expansive, with $\alpha_2 = \Delta t \frac{\beta}{2} (1 - \Delta t \frac{\beta}{2}) \in (0, 1)$ means that we need $0 < 1 - \Delta t \frac{\beta}{2} < 1$, i.e., $0 < \Delta t < \frac{K}{4\gamma} \exp(-(2M + 8\gamma))$. □

**Proposition 5.** [14]. $T_1 \circ T_2$ is $\alpha$-average non-expansive, with $\alpha = \frac{2}{1 + \max(\alpha_1, \alpha_2)}$.

**Proof.** This proposition is the same as Lemma 2.2(iii) in [14]. But the proof is slightly different.

Let $\alpha_1 = \frac{1}{2}$, then according to the Proposition 2 and Proposition 4, $T_1, T_2$ are $\alpha_1, \alpha_2$-average non-expansive respectively, and we have

$$||T_1(\eta) - T_1(\delta)||^2 \leq ||\eta - \delta||^2 - \frac{1 - \alpha_1}{\alpha_1}||(I - T_1)(\eta) - (I - T_1)(\delta)||^2,$$

$$||T_2(\eta) - T_2(\delta)||^2 \leq ||\eta - \delta||^2 - \frac{1 - \alpha_2}{\alpha_2}||(I - T_2)(\eta) - (I - T_2)(\delta)||^2.$$
Let $\alpha_0 = \max\{\alpha_1, \alpha_2\}$, we have
\[
\frac{1 - \alpha_0}{\alpha_0} \geq \frac{1 - \alpha_2}{\alpha_2},
\]
\[
\frac{1 - \alpha_0}{\alpha_0} \geq \frac{1 - \alpha_1}{\alpha_1}.
\]

Now we back to $T_1 \circ T_2$,
\[
||T_1 \circ T_2(\eta) - T_1 \circ T_2(\delta)||^2
\]
\[
= ||T_1(T_2(\eta)) - T_1(T_2(\delta))||^2
\]
\[
\leq ||T_2(\eta) - T_2(\delta)||^2 - \frac{1 - \alpha_0}{\alpha_1}||(I - T_1)(T_2(\eta)) - (I - T_1)(T_2(\delta))||^2
\]
\[
\leq ||\eta - \delta||^2 - \frac{1 - \alpha_0}{\alpha_1}||(I - T_1)(T_2(\eta)) - (I - T_1)(T_2(\delta))||^2
\]
\[
\leq ||\eta - \delta||^2 - \frac{1 - \alpha_0}{\alpha_1}||T_2(\eta) - (I - T_2)(\delta)||^2
\]
\[
\leq ||\eta - \delta||^2 - \frac{1 - \alpha_0}{\alpha_1}||T_2(\eta) - (I - T_2)(\delta)||^2
\]
\[
\leq ||\eta - \delta||^2 - \frac{1 - \alpha_0}{\alpha_1}||T_2(\eta) - (I - T_2)(\delta)||^2.
\]

Here we use the Cauchy-Schwarz inequality, which is $\langle a^2 + b^2 \rangle \geq \frac{1}{2}(a^2 + b^2)$.

Let $\frac{1 - \alpha}{\alpha} = \frac{1 - \alpha_0}{\alpha_1}$, we get that $\alpha = \frac{1}{\max\{\alpha_1, \alpha_2\}}$. Which means that $T_1 \circ T_2$ is $\alpha$-average non-expansive.

\[\Box\]

**Proposition 6.** If $G = \{\eta | \eta = T_1(T_2(\eta))\}$ is not empty, $f$ is bounded by $M$, and $0 < \Delta t < \frac{\delta}{2} \exp(-(2M + 8\gamma))$ then (21) convergence to one $\eta^* \in G$.

**Proof.** The idea of this proof is from [14].

From Proposition 5, $T = T_1 \circ T_2$ is $\alpha$-average non-expansive. Then we will prove that $\eta^{n+1} = T(\eta^n)$ converges.

\[
||\eta^n - \tilde{\eta}|| = ||T(\eta^n - 1) - T(\tilde{\eta})||
\]
\[
\leq ||\eta^n - 1 - \tilde{\eta}|| - \frac{1 - \alpha}{\alpha}||T(\eta^n - 1) - (I - T)(\tilde{\eta})||
\]
\[
\leq ||\eta^n - 1 - \tilde{\eta}|| - \frac{1 - \alpha}{\alpha}||T(\eta^n - 1)||
\]
\[
\leq ||\eta^n - 1 - \tilde{\eta}||
\]
\[
\leq ||\eta^n - \tilde{\eta}||,
\]

which means that $||\eta^n - \tilde{\eta}||$ is bounded and monotone, then the limit

\[
\lim_{n \to +\infty} ||\eta^n - \tilde{\eta}||
\]

exists. While, $\eta^n$ is bounded due to

\[
||\eta^n|| \leq ||\eta^n - \tilde{\eta}|| + ||\tilde{\eta}||.
\]

So there exists a subsequence of $\eta^n$ denoted as $\eta^{n_j}$, that converges to $\eta^*$.

Back to (25), we rewrite it as

\[
\frac{1 - \alpha}{\alpha}||T(\eta^n - 1)|| \leq ||\eta^n - 1 - \tilde{\eta}|| - ||\eta^n - \tilde{\eta}||.
\]

Let $n = 1, 2, \cdots, N$, and sum them together, we have

\[
\sum_{n=1}^{N} \frac{1 - \alpha}{\alpha}||T(\eta^n - 1)|| \leq \sum_{n=1}^{N} (||\eta^n - 1 - \tilde{\eta}|| - ||\eta^n - \tilde{\eta}||)
\]
\[
= ||\eta^n - \tilde{\eta}|| - ||\eta^n - \tilde{\eta}||.
\]

**Inverse Problems and Imaging**

Volume 13, No. 3 (2019), 653–677
Now let $N \to +\infty$, we have
\[
\lim_{N \to +\infty} \sum_{n=1}^{+\infty} \frac{1-\alpha}{\alpha} \| |(I - T)(\eta^{n-1})| \| \leq \| \eta^0 - \hat{\eta} \| - \lim_{N \to +\infty} \| \eta^N - \hat{\eta} \|,
\]
which means
\[
\lim_{n \to +\infty} \frac{1-\alpha}{\alpha} \| |(I - T)(\eta^{n-1})| \| = 0.
\]
Because $\alpha \in (0, 1)$, this means
\[
\lim_{n \to +\infty} \| \eta^{n-1} - \eta^n \| = 0.
\]
Then
\[
\| T(\eta^*) - \eta^* \| = \lim_{j \to +\infty} \| T(\eta^{n_j}) - \eta^{n_j} \| = \lim_{j \to +\infty} \| \eta^{n_j+1} - \eta^{n_j} \| = 0.
\]
Which means $\eta^* \in \mathbb{G}$, now we can replace $\eta^*$ with $\hat{\eta}$ in all of our results.

For $\forall n$, we take $j = \max \{ i | n_i \leq n \}$, then $n_j \leq n < n_{j+1}$, and
\[
\| \eta^{n_j+1} - \eta^* \| < \| \eta^n - \eta^* \| \leq \| \eta^{n_j} - \eta^* \|,
\]
let $n \to +\infty$, which also implies $j \to +\infty$, we have
\[
\lim_{n \to +\infty} \| \eta^n - \eta^* \| = 0.
\]
i.e.
\[
\lim_{n \to +\infty} \eta^n = \eta^*.
\]

REFERENCES

[1] W. Allard, Total variation regularization for image denoising, i. geometric theory, SIAM Journal on Mathematical Analysis, 39 (2007), 1150–1190.
[2] U. Amato and W. Hughes, Maximum entropy regularization of fredholm integral equations of the first kind, Inverse Problems, 7 (1991), 793–808.
[3] P. Arbelaez, M. Maire, C. Fowlkes and J. Malik, Contour detection and hierarchical image segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 33 (2011), 898–916.
[4] J. Ashburner and K. J. Friston, Unified segmentation, NeuroImage, 26 (2005), 839–851, URL http://www.sciencedirect.com/science/article/pii/S1053811905001102.
[5] E. Bae, J. Yuan and X. Tai, Global minimization for continuous multiphase partitioning problems using a dual approach, International Journal of Computer Vision, 92 (2011), 112–129.
[6] J. Baillon and G. Haddad, Quelques propriétés des opérateurs angle-bornés etn-cycliquement monotones, Israel Journal of Mathematics, 26 (1977), 137–150.
[7] Y. Boykov and V. Kolmogorov, An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision, Energy Minimization Methods in Computer Vision and Pattern Recognition, (2001), 359–374.
[8] Y. Boykov, O. Veksler and R. Zabih, Fast approximate energy minimization via graph cuts, IEEE Transactions on Pattern Analysis and Machine Intelligence, 23 (2001), 1222–1239.
[9] C. Carson, S. Belongie, H. Greenspan and J. Malik, Blobworld: Image segmentation using expectation-maximization and its application to image querying, IEEE Transactions on Pattern Analysis and Machine Intelligence, 24 (2002), 1026–1038.
[10] A. Chambolle, An algorithm for total variation minimization and applications, Journal of Mathematical Imaging and Vision, 20 (2004), 89–97.
[11] T. Chan, S. Esedoglu and A. Yip, Recent developments in total variation image restoration, Mathematical Models of Computer Vision, 17 (2005).
[12] T. Chan and L. Vese, Active contours without edges, IEEE Transactions on Image Processing, 10 (2001), 266–277.
[13] Y. Chiang, P. Borbat and J. Freed, Maximum entropy: A complement to tikhonov regularization for determination of pair distance distributions by pulsed esr, *Journal of Magnetic Resonance*, 177 (2005), 184–196.

[14] P. Combettes, Solving monotone inclusions via compositions of nonexpansive averaged operators, *Optimization*, 53 (2004), 475–504.

[15] P. Combettes and V. Wajs, Signal recovery by proximal forward-backward splitting, *Multiscale Modeling and Simulation*, 4 (2005), 1168–1200.

[16] A. Dempster, N. Laird and D. Rubin, Maximum likelihood from incomplete data via the em algorithm, *Journal of the Royal Statistical Society. Series B (methodological)*, 39 (1977), 1–38.

[17] J. Duarte-Carvajalino, G. Sapiro, G. Yu and L. Carin, Online adaptive statistical compressed sensing of gaussian mixture models, arXiv preprint, arXiv:1112.5895.

[18] I. Ekeland and R. Temam, *Convex Analysis and Variational Problems*, SIAM, 1999.

[19] L. Evans and R. Gariepy, *Measure Theory and Fine Properties of Functions*, CRC Press, 2015.

[20] S. Gao and T. Bui, Image segmentation and selective smoothing by using mumford-shah model, *IEEE Transactions on Image Processing*, 14 (2005), 1537–1549.

[21] T. Goldstein and S. Osher, The split bregman method for l1-regularized problems, *SIAM Journal on Imaging Sciences*, 2 (2009), 323–343.

[22] R. Gonzalez, R. Woods and S. Eddins, *Digital Image Processing Using MATLAB*, Prentice Hall, 2004.

[23] L. Gupta and T. Sortrakul, A gaussian-mixture-based image segmentation algorithm, *Pattern Recognition*, 31 (1998), 315–325.

[24] J. Hiriart-Urruty and C. Lemar chal, *Convex Analysis and Minimization Algorithms I*, Springer, 1993.

[25] H. Ishikawa, Exact optimization for markov random fields with convex priors, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25 (2003), 1333–1336.

[26] J. Liu, Y. Ku and S. Leung, Expectation–maximization algorithm with total variation regularization for vector-valued image segmentation, *Journal of Visual Communication and Image Representation*, 23 (2012), 1234–1244.

[27] J. Liu, X. Tai, H. Huang and Z. Huan, A weighted dictionary learning model for denoising images corrupted by mixed noise, *IEEE Transactions on Image Processing*, 22 (2013), 1108–1120.

[28] J. Liu, X. Tai, S. Leung and H. Huang, A new continuous max-flow algorithm for multiphase image segmentation using super-level set functions, *Journal of Visual Communication & Image Representation*, 25 (2014), 1472–1488.

[29] D. Martin, C. Fowlkes, D. Tal and J. Malik, A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, vol. 2, IEEE, 2001, 416–423.

[30] G. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, A Wiley-Interscience Publication, J ohn Wiley & Sons, Inc., New York, 1997.

[31] D. Mumford and J. Shah, Optimal approximations by piecewise smooth functions and associated variational problems, *Communications on Pure and Applied Mathematics*, 42 (1989), 577–685.

[32] S. Osher and J. Sethian, Fronts propagating with curvature-dependent speed: Algorithms based on hamilton-jacobi formulations, *Journal of Computational Physics*, 79 (1988), 12–49.

[33] L. Rudin, S. Osher and E. Fatemi, Total variation based noise removal algorithms, *Physica D: Nonlinear Phenomena*, 60 (1992), 259–268.

[34] M. Teboulle, A unified continuous optimization framework for center-based clustering methods, *Journal of Machine Learning Research*, 8 (2007), 65–102.

[35] L. Vese and T. Chan, A multiphase level set framework for image segmentation using the mumford and shah model, *International Journal of Computer Vision*, 50 (2002), 271–293.

[36] Q. Wang, Hmrf-em-image: Implementation of the hidden markov random field model and its expectation-maximization algorithm, arXiv preprint arXiv:1207.3510.

[37] C. Wu and X. Tai, Augmented lagrangian method, dual methods, and split bregman iteration for rof, vectorial tv, and high order models, *SIAM Journal on Imaging Sciences*, 3 (2010), 300–339.
[38] C. Wu, On the convergence properties of the em algorithm, The Annals of Statistics, 11 (1983), 95–103.
[39] G. Yu and G. Sapiro, Statistical compressed sensing of gaussian mixture models, IEEE Transactions on Signal Processing, 59 (2011), 5842–5858.
[40] J. Yuan, E. Bae, X. Tai and Y. Boykov, A continuous max-flow approach to potts model, Computer Vision (ECCV 2010), European Conference on, 2010, 379–392.
[41] Y. Zhang, M. Brady and S. Smith, Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm, IEEE Transactions on Medical Imaging, 20 (2001), 45–57.
[42] M. Zhu, S. J. Wright and T. Chan, Duality-based algorithms for total-variation-regularized image restoration, Computational Optimization and Applications, 47 (2010), 377–400.

Received May 2018; 1st revision September 2018; 2nd revision October 2018.

E-mail address: ysicesword@mail.bnu.edu.cn
E-mail address: jliu@bnu.edu.cn
E-mail address: hhywsg@bnu.edu.cn
E-mail address: xuechengtai@hkbu.edu.hk