Received: 22 April 2017 | Revised: 24 October 2018 | Accepted: 11 November 2018

DOI: 10.1002/nla.2224

RESEARCH ARTICLE

WILEY

A primal-dual method for the Meyer model of cartoon and texture decomposition

You-Wei Wen[1] | Hai-Wei Sun[2] | Michael K. Ng[3]

[1]Key Laboratory of High Performance Computing and Stochastic Information Processing, College of Mathematics and Computer Science, Hunan Normal University, Changsha, China

[2]Department of Mathematics, University of Macau, Macao, China

[3]Department of Mathematics, Hong Kong Baptist University, Hong Kong

Correspondence
You-Wei Wen, Key Laboratory of High Performance Computing and Stochastic Information Processing, College of Mathematics and Computer Science, Hunan Normal University, Changsha 410081, China.
Email: wenyouwei@gmail.com

Funding information
NSFC, Grant/Award Number: 11871210; Construct Program of the Key Discipline in Hunan Province; Scientific Research Fund of Hunan Provincial Education Department, Grant/Award Number: 17A128; HKRGC GRF, Grant/Award Number: 1202715, 12306616, 12200317, 12300218, and HKBU RC-ICRS/16-17/03

Summary

In this paper, we study the original Meyer model of cartoon and texture decomposition in image processing. The model, which is a minimization problem, contains an l_1-based TV-norm and an l_∞-based G-norm. The main idea of this paper is to use the dual formulation to represent both TV-norm and G-norm. The resulting minimization problem of the Meyer model can be given as a minimax problem. A first-order primal-dual algorithm can be developed to compute the saddle point of the minimax problem. The convergence of the proposed algorithm is theoretically shown. Numerical results are presented to show that the original Meyer model can decompose better cartoon and texture components than the other testing methods.

KEYWORDS

G-norm, image decomposition, Meyer's model, primal-dual algorithm, total variation

1 | INTRODUCTION

In many problems of image processing and computer vision, we have a target image that may contain different components such as cartoon component and oscillatory component. The cartoon component is the geometrical part or sketchy approximation of an image, which is well structured. It is used to model the homogeneous regions with sharp boundaries. The oscillatory component is a pattern of small-scale details and of zero mean. An important image processing task is to decompose the target image into these two components in order to improve the subsequent processing operations such as image denoising,[1] image compression,[2] image segmentation,[3,4] face recognition,[5] and image inpainting.[6] For example, Meyer et al.[7] proposed an image compression scheme where an image is treated as a superposition of cartoon component and texture component. Each components can be sparsely represented using a suitable dictionary, which is chosen such that each leads to a sparse representation over the content part it is serving, while yielding a nonsparse representation for the other content part. Examples of dictionaries that can sparsely approximate cartoon components in images are curvelets,[8,9] orthonormal wavelets,[10] and wavelet tight frames constructed by the unitary extension principle in the work of Ron et al.[11] Examples of dictionaries that can sparsely approximate textures are the local discrete cosine transform and

the Gabor transform; see other works.[12–16] If the cartoon component and the texture component are encoded separately, higher overall compression gain can be obtained.

The problem of image decomposition can be modeled to decompose a target gray image $\boldsymbol{f} : (x, y) \in \Omega \to \mathbb{R}$ into two components

$$\boldsymbol{f} = \boldsymbol{u} + \boldsymbol{v}, \tag{1}$$

where $\Omega \subset \mathbb{R}^2$ is an open bounded domain with a Lipschitz boundary, $\boldsymbol{u}$ represents a cartoon component, and $\boldsymbol{v}$ is the oscillatory component. Mathematically, the cartoon component can be described by a piecewise smooth function, which can be represented in the space of function of bounded variation

$$\mathrm{BV}(\Omega) = \left\{ \boldsymbol{u} \in L^1(\Omega) : \int_\Omega |\mathcal{D}\boldsymbol{u}| < \infty \right\},$$

where $\mathcal{D}\boldsymbol{u}$ denotes the generalized derivative of $\boldsymbol{u}$ and $\int_\Omega |\mathcal{D}\boldsymbol{u}|$ stands for the total variation (TV) norm of $\boldsymbol{u}$. When $\boldsymbol{u}$ is regular, $\int_\Omega |\mathcal{D}\boldsymbol{u}|$ is often written as $\int_\Omega |\nabla \boldsymbol{u}|$, which can be also reformulated as a dual approach as follows:

$$\int_\Omega |\nabla \boldsymbol{u}| = \sup \left\{ \int_\Omega \boldsymbol{u} \mathrm{div} \boldsymbol{p} : \boldsymbol{p} = (\boldsymbol{p}_1, \boldsymbol{p}_2) \in Y, \|\boldsymbol{p}\|_\infty \le 1 \right\}.$$

Here, $\|\boldsymbol{p}\|_\infty = \max_{x,y}\{|\boldsymbol{p}(x, y)|\}$ with $|\boldsymbol{p}(x, y)| = \sqrt{(\boldsymbol{p}_1(x, y))^2 + (\boldsymbol{p}_2(x, y))^2}$, $Y = L^2(\Omega) \times L^2(\Omega)$, and $\mathrm{div} = -\nabla^*$ with the adjoint operator $*$. We denote the TV-norm of $\boldsymbol{u}$ as $\mathrm{TV}(\boldsymbol{u})$.

The oscillatory part $\boldsymbol{v}$ contains essentially the noise and the texture. The noise is the unwanted component of the image acquired during the image acquisition process or the transmission process. Rudin, Osher, and Fatemi (ROF[1]) proposed a TV minimization model to remove the Gaussian white noise

$$\min_{\boldsymbol{u}+\boldsymbol{v}=\boldsymbol{f}} \frac{\beta}{2} \int_\Omega |\boldsymbol{v}|^2 + \mathrm{TV}(\boldsymbol{u}), \tag{2}$$

where β is a regularization parameter. The L^2-norm in the first term of (2) is replaced by L^1-norm for random noise or impulsive noise[17] and by Kullback–Leibler (KL) divergence (I-divergence) for Poisson noise.[18] Meyer characterized the solution $\boldsymbol{v}$ in (2) in G space,[15] a dual space of $BV(\Omega)$, which is defined by

$$G(\Omega) = \left\{ \boldsymbol{v} \in L^2(\Omega) | \boldsymbol{v} = \mathrm{div} \boldsymbol{w}, \boldsymbol{w} \in Y \right\}.$$

An image $\boldsymbol{v}$ belongs to G-space if and only if there exists $\boldsymbol{w} \in Y$ with $\boldsymbol{v} = \mathrm{div}\boldsymbol{w}$. The G-norm of $\boldsymbol{v}$ is defined as the infimum of all L^∞-norm of the functions $\boldsymbol{w} \in Y$ with $\boldsymbol{v} = \mathrm{div}\boldsymbol{w}$, that is,

$$\|\boldsymbol{v}\|_G = \inf \left\{ \|\boldsymbol{w}\|_\infty | \boldsymbol{v} = \mathrm{div} \boldsymbol{w}, \boldsymbol{w} \in Y \right\}.$$

Meyer observed that the target image $\boldsymbol{f}$ will be treated by the model (2) as texture or noise even though $\boldsymbol{f}$ is regular (e.g., $\boldsymbol{f}$ is a characteristic function) when the G-norm of $\boldsymbol{f}$ is not larger than $1/\beta$. Theoretical analysis shows that a sequence of oscillating functions having bounded L^q-norm with $q \ge 2$ tends to zero in G-space without converging strongly to zero in $L^q(\Omega)$.[19] Thus, if $\boldsymbol{v}$ is the texture or oscillating patterns of the image $\boldsymbol{f}$, it has a smaller value of G-norm than that of L^2-norm. Numerical experiments have shown that G-norm indeed captures oscillatory patterns and texture better than the standard ROF model.[20,21] To better capture textures or oscillating patterns, Meyer proposed to replace the L^2-norm in the model (2) with G-norm[15]; thus, the minimization model became

$$\min_{\boldsymbol{u}+\boldsymbol{v}=\boldsymbol{f}} \beta \|\boldsymbol{v}\|_G + \mathrm{TV}(\boldsymbol{u}). \tag{3}$$

Because G-norm cannot be expressed as integrals, the corresponding Euler–Lagrange equation of (3) cannot be expressed directly. It is a challenge to study numerical methods for solving the minimization problem in (3). In the literature, several numerical methods[20–22] have been proposed to modify Meyer's model (3) and use the other norms, which can be implemented numerically in practice, to approximate G-norm. Vese and Osher (VO[21]) proposed to relax the constraint $\boldsymbol{f} = \boldsymbol{u} + \boldsymbol{v}$ and approximate G-norm by L^p-norm in the original Meyer model (3). According to the definition of

G-norm, if $\boldsymbol{v} \in G(\Omega)$, then there exists $\boldsymbol{w} \in Y$ with $\boldsymbol{v} = \text{div}\boldsymbol{w}$ such that $\|\boldsymbol{v}\|_G = \|\boldsymbol{w}\|_\infty$ (see lemma 5.2.1 in the work of Aubert et al.[19]). As a result, the minimization problem (3) can be rewritten as

$$\min_{\boldsymbol{u},\boldsymbol{v}} \beta\|\boldsymbol{w}\|_\infty + \text{TV}(\boldsymbol{u}), \tag{4}$$

subject to $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{f}$ and $\boldsymbol{v} = \text{div}\boldsymbol{w}$. To numerically compute the solution of (4), Vese et al.[21] proposed to use L^p-norm to replace L^∞-norm in (4); then, the minimization problem considered in the work of Vese et al.[21] is

$$\min_{\boldsymbol{u},\boldsymbol{w}} \frac{\alpha}{2}\int_\Omega |\boldsymbol{f} - \boldsymbol{u} - \text{div}\boldsymbol{w}|^2 + \beta\left[\int_\Omega |\boldsymbol{w}|^p\right]^{\frac{1}{p}} + \text{TV}(\boldsymbol{u}), \tag{5}$$

where α and β are tuning parameters. The first and second terms ensure $\text{div}\boldsymbol{w} \approx \boldsymbol{f} - \boldsymbol{u} \in G(\Omega)$. If $\alpha \to \infty$ and $p \to \infty$, the VO model mathematically coincides with the original Meyer model in (4); see (3). Once the solution $(\widehat{\boldsymbol{u}}, \widehat{\boldsymbol{w}})$ of (5) is obtained, the texture component is simply given by $\widehat{\boldsymbol{v}} = \text{div}\widehat{\boldsymbol{w}}$. Formally, minimizing the objective function in (5) with respect to $\boldsymbol{u}, \boldsymbol{w}$ yields the Euler–Lagrange equations. Later, Osher et al.[20] considered the case $p = 2$, which approximates L^∞-norm by the seminorm in H^{-1}.

Aujol, Aubert, Blanc-Feraud, and Chambolle[22] (AABC) proposed to approximately solve the original Meyer model by replacing the G-norm in (3) with a constraint $\|\boldsymbol{v}\|_G \le \mu$ with a parameter $\mu > 0$ and adding extra regularization terms

$$\min_{\boldsymbol{u},\boldsymbol{v}} \frac{\alpha}{2}\int_\Omega |\boldsymbol{f} - \boldsymbol{u} - \boldsymbol{v}|^2 + \text{TV}(\boldsymbol{u}), \quad \text{subject to } \|\boldsymbol{v}\|_G \le \mu. \tag{6}$$

Here, α is a regularization parameter. When $\alpha \to \infty$, the problem in (6) tends to the limit problem

$$\min_{\boldsymbol{u},\boldsymbol{v}} \text{TV}(\boldsymbol{u}), \text{subject to } \|\boldsymbol{v}\|_G \le \mu, \boldsymbol{f} = \boldsymbol{u} + \boldsymbol{v}. \tag{7}$$

Mathematically, (3) and (7) are equivalent. Assume that $(\widehat{\boldsymbol{u}}, \widehat{\boldsymbol{v}})$ is the solution of (3), it will also be the solution of problem (7) for a particular parameter $\mu = \|\widehat{\boldsymbol{v}}\|_G$. Similarly, if there exists a solution for problem (7), it should be a solution of (3) for a particular choice of β, which is the Lagrange multiplier corresponding to the constraint $\|\boldsymbol{v}\|_G \le \mu$ in (7). Thus, the problem (6) asymptotically converges to the original Meyer model (3) when α tends to infinity. The solution of the AABC model (6) is unique[22] and can be obtained by the alternating minimization method. Fixed the variable $\boldsymbol{v}$, we obtain the $\boldsymbol{u}$-subproblem, which is the classical TV denoising problem. While fixed the variable $\boldsymbol{u}$, we obtain the $\boldsymbol{v}$-subproblem, which is equivalent to compute the projection of $\boldsymbol{f} - \boldsymbol{u}$ to the set $\{\boldsymbol{v}|\|\boldsymbol{v}\|_G \le \mu\}$. The corresponding $\boldsymbol{u}$- and $\boldsymbol{v}$-subproblems can be solved approximately by Chambolle's projection method.[23] Gilles et al. proposed to solve (6) by the splitting Bregman iteration method.[24]

In the literature, other spaces were used to model the oscillating component. Aujol et al.[25] proposed to replace G-norm in (6) by E-norm,[15] which is the norm endowed in the E-space, the dual space of the standard Besov space B_1^1. The E-norm model can well represent noise and, hence, is suited to denoising a textured image. In the work of Aujol et al.,[26] a Hilbert-space norm was proposed to represent the oscillatory component. The norm of the oscillatory component can be easier to compute by convolution with a smoothing kernel in L^2-norm. Elion et al.[27] proposed to model the oscillatory part as the Laplacian of a single-valued function whose gradient belongs to L^∞ space. A splitting Bregman iteration approach was applied to find the solution.[28] In the work of Ng et al.,[29] the authors proposed a cartoon+texture decomposition model to restore blurred images with missing pixels.

The main goal of this paper is to develop a numerical scheme to solve the original Meyer model in (3) rather than an approximation one. To overcome the numerical difficulty caused by the G-norm, we reformulate it using Legendre–Fenchel duality to represent the G-norm

$$\|\boldsymbol{v}\|_G = \sup_{\boldsymbol{z}} \left\{ \int_\Omega \boldsymbol{w}\cdot\boldsymbol{z} \,\middle|\, \boldsymbol{w}, \boldsymbol{z} \in Y, \boldsymbol{v} = \text{div}\boldsymbol{w}, \int_\Omega |\boldsymbol{z}| \le 1 \right\}.$$

Denote $A = \{\boldsymbol{p}|\boldsymbol{p} \in Y, \|\boldsymbol{p}\|_\infty \le 1\}$ and $B = \{\boldsymbol{z}|\boldsymbol{z} \in Y, \int_\Omega |\boldsymbol{z}| \le 1\}$. Notice that if $\boldsymbol{v} \in G(\Omega)$, then there exists $\boldsymbol{w} \in Y$ with $\boldsymbol{v} = \text{div}\boldsymbol{w}$. Once $\boldsymbol{w}$ is given, the cartoon component of $\boldsymbol{f}$ is obtained by $\boldsymbol{u} = \boldsymbol{f} - \text{div}\boldsymbol{w}$. We use the duality to represent both

TV-norm and G-norm in the original Meyer model; then, the problem in (3) becomes

$$\inf_{\boldsymbol{w}} \sup_{\boldsymbol{p}\in A,\boldsymbol{z}\in B} \mathcal{J}(\boldsymbol{w},\boldsymbol{p},\boldsymbol{z}) \equiv \int_{\Omega} \left[(\boldsymbol{f}-\operatorname{div}\boldsymbol{w})\operatorname{div}\boldsymbol{p} + \beta \boldsymbol{w}\cdot\boldsymbol{z}\right]. \tag{8}$$

We will show that the infimum and the supremum in (8) can be swapped and then apply the primal-dual proximal point method[30–37] to compute the solution of the minimax problem (8) alternately with the primal variable $\boldsymbol{w}$ and the dual variables $\boldsymbol{p},\boldsymbol{z}$ fixed alternatively.

We highlight the major differences between our method and the previous methods[20–22,24] to solve the G-norm–based image decomposition problem. First of all, the existing methods[20–22,24] were to solve a related problem, which was to approximate the original one (3). Through asymptotic analysis, it can be shown that the VO model (5) and the AABC model (6) mathematically coincide with the original Meyer model (3). However, for a fixed parameter α, they are of course different from the original model (3). To our best knowledge, no work is proposed in the literature to develop an algorithm that solves the exact Meyer model (3). On the other hand, the VO model (5) and the AABC model (6) introduce an additional parameter α to relax the constraint $\boldsymbol{f}=\boldsymbol{u}+\boldsymbol{v}$; thus, the original Meyer model is simpler than the VO model and the AABC model in the sense that there are fewer parameters to tune. Our numerical results also show that the cartoon component obtained by the original Meyer model has a sharper boundary than the AABC model. Moreover, the advantage of the proposed approach is that there are closed-form solutions for each subproblem in the minimization process.

The rest of this paper is organized as follows. In Section 2, we study a primal-dual algorithm to solve the minimax problem (8) numerically. We also describe the algorithm using a discrete setting. In Section 3, we give the convergence result for the proposed algorithm. In Section 4, we present some numerical examples for image decompositions. In Section 5, a conclusion is finally given.

2 | PRIMAL-DUAL ALGORITHM

In this section, we consider to calculate the solution of the minimax problem (8) numerically. The presentation is given below in the discrete setting, but it is the same as in the continuous case.

2.1 | Notations

Assume that the size of the discrete image is $n\times n$. We denote by X the Euclidean space $\mathbb{R}^{n\times n}$, and $Y=X\times X$. We use the classical notions $\boldsymbol{f}_{i,j}=\boldsymbol{f}(ih,jh)$, $\boldsymbol{u}_{i,j}=\boldsymbol{u}(ih,jh)$, $\boldsymbol{v}_{i,j}=\boldsymbol{v}(ih,jh)$, and $\boldsymbol{w}_{i,j}=(\boldsymbol{w}_{1,i,j},\boldsymbol{w}_{2,i,j})$ with $\boldsymbol{w}_{r,i,j}=\boldsymbol{w}_r(ih,jh)(r=1,2)$, $\boldsymbol{p}_{i,j}=(\boldsymbol{p}_{1,i,j},\boldsymbol{p}_{2,i,j})$ with $\boldsymbol{p}_{r,i,j}=\boldsymbol{p}_r(ih,jh)(r=1,2)$, $\boldsymbol{z}_{i,j}=(\boldsymbol{z}_{1,i,j},\boldsymbol{z}_{2,i,j})$ with $\boldsymbol{z}_{r,i,j}=\boldsymbol{z}_r(ih,jh)(r=1,2)$, where $h>0$ is the step size and (ih,jh) denotes a discrete point for $1\le i,j\le n$.

Define the inner product $\langle\boldsymbol{u},\boldsymbol{f}\rangle_X=\sum_{i,j}\boldsymbol{u}_{i,j}\boldsymbol{f}_{i,j}$, $\langle\boldsymbol{p},\boldsymbol{q}\rangle_Y=\sum_{i,j}\sum_{r=1}^{2}p_{r,i,j}q_{r,i,j}$ and the norm $\|\boldsymbol{f}\|_2=\sqrt{\langle\boldsymbol{f},\boldsymbol{f}\rangle_X}$, $\|\boldsymbol{p}\|_2=\sqrt{\langle\boldsymbol{p},\boldsymbol{p}\rangle_Y}$, $\|\boldsymbol{p}\|_1=\sum_{r,s}|p_{i,j}|$ and $\|\boldsymbol{p}\|_\infty=\max_{i,j}\{|p_{i,j}|\}$ with $|p_{i,j}|=\sqrt{p_{1,i,j}^2+p_{2,i,j}^2}$. For any $\boldsymbol{u}\in X$, the discrete gradient of $\boldsymbol{u}$ is defined as

$$(\nabla\boldsymbol{u})_{i,j}=\left((\nabla_x\boldsymbol{u})_{i,j},(\nabla_y\boldsymbol{u})_{i,j}\right).$$

Here,

$$(\nabla_x\boldsymbol{u})_{i,j}=\boldsymbol{u}_{i+1,j}-\boldsymbol{u}_{i,j}\quad\text{and}\quad(\nabla_y\boldsymbol{u})_{i,j}=\boldsymbol{u}_{i,j+1}-\boldsymbol{u}_{i,j}$$

with $\boldsymbol{u}_{n+1,j}=\boldsymbol{u}_{n,j}$ and $\boldsymbol{u}_{i,n+1}=\boldsymbol{u}_{i,n}$ for $i,j=1,2,\dots,n$. The discrete version of the divergence operator is defined by analogy with the continuous case by $\operatorname{div}=-\nabla^T$. Thus, the discrete divergence of $\boldsymbol{p}$ is defined as

$$(\operatorname{div}\boldsymbol{p})_{i,j}=\boldsymbol{p}_{1,i,j}-\boldsymbol{p}_{1,i-1,j}+\boldsymbol{p}_{2,i,j}-\boldsymbol{p}_{2,i,j-1}$$

with $\boldsymbol{p}_{1,0,j}=\boldsymbol{p}_{2,i,0}=0$ for $i,j=1,2,\dots,n$. Moreover, for every $\boldsymbol{p}\in Y$ and $\boldsymbol{f}\in X$, we have $\langle-\operatorname{div}\boldsymbol{p},\boldsymbol{f}\rangle_X=\langle\boldsymbol{p},\nabla\boldsymbol{f}\rangle_Y$.

2.2 | Numerical method

First, we describe the minimax problem in a discrete setting. The discrete version of (8) can be reformulated as

$$\mathcal{J}(\boldsymbol{w},\boldsymbol{p},\boldsymbol{z})=\langle\boldsymbol{f}-\operatorname{div}\boldsymbol{w},\operatorname{div}\boldsymbol{p}\rangle_X+\beta\langle\boldsymbol{w},\boldsymbol{z}\rangle_Y. \tag{9}$$

The sets A, B are defined as $A = \{\boldsymbol{p} | \boldsymbol{p} \in Y, \|\boldsymbol{p}\|_\infty \leq 1\}$ and $B = \{\boldsymbol{z} | \boldsymbol{z} \in Y, \|\boldsymbol{z}\|_1 \leq 1\}$. It is clear that A, B are convex and closed.

We now recall that a pair $(\boldsymbol{w}^*, \boldsymbol{p}^*, \boldsymbol{z}^*)$ is a saddle point of $\mathcal{J}(\boldsymbol{w}, \boldsymbol{p}, \boldsymbol{z})$ if

$$\mathcal{J}(\boldsymbol{w}^*, \boldsymbol{p}, \boldsymbol{z}) \leq \mathcal{J}(\boldsymbol{w}^*, \boldsymbol{p}^*, \boldsymbol{z}^*) \leq \mathcal{J}(\boldsymbol{w}, \boldsymbol{p}^*, \boldsymbol{z}^*), \quad \forall \boldsymbol{w} \in \Omega, \boldsymbol{p}, \boldsymbol{z} \in Y.$$

We note that $\mathcal{J}(\cdot, \boldsymbol{p}, \boldsymbol{z})$ and $\mathcal{J}(\boldsymbol{w}, \cdot, \cdot)$ are linear functions for each $\boldsymbol{p}, \boldsymbol{z}$ and each $\boldsymbol{w}$, respectively. It is well known that a linear function is both convex and concave.

We will demonstrate that the set of saddle points of $\mathcal{J}$ defined in (9) is nonempty. Let

$$\psi(\boldsymbol{w}) = \begin{cases} \sup_{\boldsymbol{p} \in A, \boldsymbol{z} \in B} \mathcal{J}(\boldsymbol{w}, \boldsymbol{p}, \boldsymbol{z}) = \mathrm{TV}(\boldsymbol{f} - \mathrm{div}\boldsymbol{w}) + \beta \|\boldsymbol{w}\|_\infty, & \text{if} \quad \boldsymbol{w} \in \Omega, \\ +\infty, & \text{if} \quad \boldsymbol{w} \notin \Omega. \end{cases}$$

It is easier to know that $\psi(\boldsymbol{w}) > -\infty$ for all $\boldsymbol{w}$ and $\psi(\boldsymbol{0}) < \infty$, which implies that $\psi(\boldsymbol{w})$ is proper. As $\{\boldsymbol{w} : \psi(\boldsymbol{w}) \leq \lambda\} \subseteq \{\boldsymbol{w} : \|\boldsymbol{w}\|_\infty \leq \lambda\}$, it implies that the level sets $\{\boldsymbol{w} : \psi(\boldsymbol{w}) \leq \lambda\}$ are closed and bounded, and they are compact. Let

$$\phi(\boldsymbol{p}, \boldsymbol{z}) = \begin{cases} -\inf_{\boldsymbol{w} \in \Omega} \mathcal{J}(\boldsymbol{w}, \boldsymbol{p}, \boldsymbol{z}), & \text{if} \quad \boldsymbol{p} \in A, \boldsymbol{z} \in B, \\ +\infty, & \text{if} \quad \boldsymbol{p} \notin A \ \text{ or } \ \boldsymbol{z} \notin B. \end{cases}$$

We know that $\phi(\boldsymbol{p}, \boldsymbol{z}) = 0$ if $\nabla \mathrm{div}\, \boldsymbol{p} + \beta \boldsymbol{z} = 0$ and $\boldsymbol{p} \in A, \boldsymbol{z} \in B$, and $\phi(\boldsymbol{p}, \boldsymbol{z}) = +\infty$ if $\nabla \mathrm{div}\, \boldsymbol{p} + \beta \boldsymbol{z} \neq 0$. Therefore, the level sets $\{(\boldsymbol{p}, \boldsymbol{z}) : \phi(\boldsymbol{p}, \boldsymbol{z}) \leq \lambda\}$ are closed and bounded, and they are compact. Hence, according to the saddle-point theorem (proposition 5.5.6 in the work of Bertsekas[38]), the set of the saddle points of $\mathcal{J}(\boldsymbol{w}, \boldsymbol{p}, \boldsymbol{z})$ is nonempty.

By the existence of the saddle point of $\mathcal{J}(\boldsymbol{w}, \boldsymbol{p}, \boldsymbol{z})$, convex analysis can be applied to show that the infimum and the supremum in (8) can be swapped, that is,

$$\inf_{\boldsymbol{w}} \sup_{\boldsymbol{p} \in A, \boldsymbol{z} \in B} \mathcal{J}(\boldsymbol{w}, \boldsymbol{p}, \boldsymbol{z}) = \sup_{\boldsymbol{p} \in A, \boldsymbol{z} \in B} \inf_{\boldsymbol{w}} \mathcal{J}(\boldsymbol{w}, \boldsymbol{p}, \boldsymbol{z}) = \mathcal{J}(\boldsymbol{w}^*, \boldsymbol{p}^*, \boldsymbol{z}^*).$$

Numerous methods have been proposed to solve saddle-point problems. They include the finite envelope method, the primal-dual gradient algorithm,[39] and the alternating direction method (see the work of Bertsekas et al.[40] for discussion). Here, we apply the (primal-dual) proximal-point method[40] to compute the saddle point of (8). In particular, proximal-point iterations are applied to the subdifferential of the primal-dual function alternately with the primal variable and the dual variable fixed. This can be equivalently interpreted as a first-order primal-dual relaxed Arrow–Hurwicz algorithm, proposed in the work of Chambolle et al.[30] More specifically, starting from an initial guess $(\boldsymbol{w}^{(0)}, \boldsymbol{p}^{(0)}, \boldsymbol{z}^{(0)})$, this method computes a sequence of iterates:

$$\boldsymbol{w}^{(k+1)} = \underset{\boldsymbol{w}}{\mathrm{argmin}}\, \mathcal{J}(\boldsymbol{w}, \boldsymbol{p}^{(k)}, \boldsymbol{z}^{(k)}) + \frac{1}{2s} \left\| \boldsymbol{w} - \boldsymbol{w}^{(k)} \right\|_2^2, \tag{10}$$

$$\overline{\boldsymbol{w}}^{(k+1)} = \boldsymbol{w}^{(k+1)} + \theta(\boldsymbol{w}^{(k+1)} - \boldsymbol{w}^{(k)}), \tag{11}$$

$$\boldsymbol{p}^{(k+1)} = \underset{\boldsymbol{p} \in A}{\mathrm{argmax}}\, \mathcal{J}(\overline{\boldsymbol{w}}^{(k+1)}, \boldsymbol{p}, \boldsymbol{z}) - \frac{1}{2t} \left\| \boldsymbol{p} - \boldsymbol{p}^{(k)} \right\|_2, \tag{12}$$

$$\boldsymbol{z}^{(k+1)} = \underset{\boldsymbol{z} \in B}{\mathrm{argmax}}\, \mathcal{J}(\overline{\boldsymbol{w}}^{(k+1)}, \boldsymbol{p}, \boldsymbol{z}) - \frac{1}{2t} \left\| \boldsymbol{z} - \boldsymbol{z}^{(k)} \right\|_2. \tag{13}$$

Here, the parameters $s, t > 0$ are the step sizes of the primal and dual steps, respectively, and $\theta \in [0, 1]$ is called the combination parameter. We remark that the parameter $\theta = 0$ corresponds to the semi-implicit classical Arrow–Hurwicz algorithm and has been used in the work of Zhu et al.[39] for TV minimization. The proposed method can resemble in some way the one in the work of Chambolle et al.,[30] where the convergence analysis is given for the case $\theta = 1$. The main difference between the proposed method and the one in the work of Chambolle et al.[30] is the order of the iterations among the variables. In the work of Chambolle et al.,[30] the variables $\boldsymbol{p}^{(k+1)}, \boldsymbol{z}^{(k+1)}$ are updated for given $\overline{\boldsymbol{w}}^{(k)}$ rather than $\overline{\boldsymbol{w}}^{(k+1)}$, whereas $\boldsymbol{w}^{(k+1)}$ is updated for given $\boldsymbol{p}^{(k+1)}, \boldsymbol{z}^{(k+1)}$ rather than $\boldsymbol{p}^{(k)}, \boldsymbol{z}^{(k)}$. In Section 3, we give another approach to show the convergence of the proposed method.

Next, we discuss how to solve the subproblems (10)–(13).

2.2.1 | Subproblem for variable $\boldsymbol{w}$

Because $\mathcal{J}(\boldsymbol{w}, \boldsymbol{p}^{(k)}, \boldsymbol{z}^{(k)})$ is linear with respect to $\boldsymbol{w}$, we see that the objective function in (10) is quadratic with respect to $\boldsymbol{w}$, that is, $\boldsymbol{w}^{(k+1)}$ is the minimizer of the quadratic function $\varphi_k(\boldsymbol{w})$, which is defined by

$$\varphi_k(\boldsymbol{w}) \equiv \frac{1}{2s}\left\|\boldsymbol{w}-\boldsymbol{w}^{(k)}\right\|_2^2 + \langle \boldsymbol{w}, \nabla \mathrm{div}\boldsymbol{p}^{(k)} + \beta \boldsymbol{z}^{(k)}\rangle_Y.$$

Hence, $\boldsymbol{w}^{(k+1)}$ can easily be computed by the formula

$$\boldsymbol{w}^{(k+1)} = \boldsymbol{w}^{(k)} - s\nabla \mathrm{div}\boldsymbol{p}^{(k)} - \beta s \boldsymbol{z}^{(k)}.$$

Because $\varphi_k(\boldsymbol{w})$ is strongly convex functions with modulus $1/s$, it follows that (proposition 6 in the work of Rockafellar[36])

$$\varphi_k(\boldsymbol{w}) \geq \varphi_k(\widehat{\boldsymbol{w}}) + \langle \nabla\varphi_k(\widehat{\boldsymbol{w}}), \boldsymbol{w}-\widehat{\boldsymbol{w}}\rangle_Y + \frac{1}{2s}\left\|\boldsymbol{w}-\widehat{\boldsymbol{w}}\right\|_2^2 \tag{14}$$

for any $\boldsymbol{w}$ and $\widehat{\boldsymbol{w}}$. Notice that $\boldsymbol{w}^{(k+1)} = \mathrm{argmin}_{\boldsymbol{w}}\varphi_k(\boldsymbol{w})$, we have $\nabla\varphi_k(\boldsymbol{w}^{(k+1)}) = 0$. Setting $\widehat{\boldsymbol{w}} = \boldsymbol{w}^{(k+1)}$ in the inequality (14), we obtain

$$\left\|\boldsymbol{w}-\boldsymbol{w}^{(k)}\right\|_2^2 + 2s\langle \boldsymbol{w}-\boldsymbol{w}^{(k+1)}, \nabla \mathrm{div}\boldsymbol{p}^{(k)} + \beta\boldsymbol{z}^{(k)}\rangle_Y \geq \left\|\boldsymbol{w}^{(k+1)}-\boldsymbol{w}^{(k)}\right\|_2^2 + \left\|\boldsymbol{w}-\boldsymbol{w}^{(k+1)}\right\|_2^2. \tag{15}$$

2.2.2 | Subproblem for variable $\boldsymbol{p}$

We consider to solve the subproblem (12). Notice that

$$\begin{aligned}\boldsymbol{p}^{(k+1)} &= \underset{\boldsymbol{p}\in A}{\mathrm{argmin}}\ \frac{1}{2}\left\|\boldsymbol{p}-\boldsymbol{p}^{(k)}\right\|_2^2 - t\langle \boldsymbol{f} - \mathrm{div}\overline{\boldsymbol{w}}^{(k+1)}, \mathrm{div}\boldsymbol{p}\rangle_X \\ &= \underset{\boldsymbol{p}\in A}{\mathrm{argmin}}\ \frac{1}{2}\left\|\boldsymbol{p}-\left(\boldsymbol{p}^{(k)} - t\nabla(\boldsymbol{f}-\mathrm{div}\overline{\boldsymbol{w}}^{(k+1)})\right)\right\|_2^2.\end{aligned}$$

Thus, the solution $\boldsymbol{p}^{(k+1)}$ is the projection of $(\boldsymbol{p}^{(k)} - t\nabla(\boldsymbol{f}-\mathrm{div}\overline{\boldsymbol{w}}^{(k+1)}))$ onto the set A. The projection of a vector $\boldsymbol{q}$ onto A can be conveniently expressed as

$$\mathcal{P}_A(\boldsymbol{q}) = \underset{\boldsymbol{p}\in A}{\mathrm{argmin}}\ \|\boldsymbol{p}-\boldsymbol{q}\|_2^2. \tag{16}$$

The projection operator $\mathcal{P}_A$ can be computed by the Lagrangian method. The Lagrangian function associated with (16) is

$$\|\boldsymbol{p}-\boldsymbol{q}\|_2^2 + \sum_{i,j}\beta_{i,j}\left(|p_{i,j}|^2 - 1\right),$$

where $\beta_{i,j} \geq 0$ is the Lagrange multiplier associated with the constraint $|p_{i,j}|^2 \leq 1$. Its complementarity condition implies that, for the optimal $\beta_{i,j}$, either $\beta_{i,j} = 0$ with $|p_{i,j}| \leq 1$ and $|q_{i,j}| \leq 1$, or $\beta_{i,j} > 0$ with $|p_{i,j}| = 1$ and $|q_{i,j}| > 1$. In the former case, we have $p_{i,j} = q_{i,j}$. In the latter case, the Karush–Kuhn–Tucker conditions yield

$$p_{i,j} - q_{i,j} + \beta_{i,j}p_{i,j} = 0, \quad \forall i,j.$$

Therefore, we have $\beta_{i,j} = |q_{i,j}| - 1$, and thus, $p_{i,j} = q_{i,j}/|q_{i,j}|$. Hence, we obtain

$$(\mathcal{P}_A(\boldsymbol{q}))_{i,j} = \frac{q_{i,j}}{\max(1, |q_{i,j}|)}.$$

Therefore, we have

$$\boldsymbol{p}^{(k+1)} = \mathcal{P}_A\left(\boldsymbol{p}^{(k)} - t\nabla(\boldsymbol{f}-\mathrm{div}\overline{\boldsymbol{w}}^{(k+1)})\right).$$

2.2.3 | Subproblem for variable $\boldsymbol{z}$

The subproblem (13) can be reformulated as

$$\boldsymbol{z}^{(k+1)} = \underset{\boldsymbol{z}\in B}{\mathrm{argmin}}\ \frac{1}{2t}\left\|\boldsymbol{z} - (\boldsymbol{z}^{(k)} + \beta t\overline{\boldsymbol{w}}^{(k+1)})\right\|_2^2.$$

Thus, $\boldsymbol{z}^{(k+1)}$ is the projection of $(\boldsymbol{z}^{(k)} + \beta t\overline{\boldsymbol{w}}^{(k+1)})$ onto the set B. The projection of a vector $\boldsymbol{q}$ onto B can be conveniently expressed as

$$\mathcal{P}_B(\boldsymbol{q}) = \underset{\boldsymbol{p}\in B}{\mathrm{argmin}}\ \|\boldsymbol{p}-\boldsymbol{q}\|_2^2. \tag{17}$$

Hence, the subproblem (13)

$$\boldsymbol{z}^{(k+1)} = \mathcal{P}_B(\boldsymbol{z}^{(k)} + \beta t\overline{\boldsymbol{w}}^{(k+1)}).$$

We also apply the Lagrangian method to calculate the projection operator $\mathcal{P}_B$. Instead of solving the problem (17), we consider the following unconstrained problem:

$$\min_{\boldsymbol{p}} \left\{ \frac{1}{2}\|\boldsymbol{p}-\boldsymbol{q}\|_2^2 + \beta\left(\|\boldsymbol{p}\|_1 - 1\right)\right\}. \tag{18}$$

Here, β is the Lagrange multiplier corresponding to the inequality constraint $\|\boldsymbol{p}\|_1 \leq 1$. Its complementarity condition implies that, for the optimal $\boldsymbol{p}$, either (i) $\|\boldsymbol{p}\|_1 \leq 1$ with $\beta = 0$ or (ii) $\|\boldsymbol{p}\|_1 = 1$ with $\beta > 0$.

For (i), we can deduce that $\|\boldsymbol{q}\|_1 \leq 1$ and the solution is given by $\boldsymbol{p} = \boldsymbol{q}$. Now, we consider how to find the solution for the case (ii). For (ii), there exists a suitable parameter β such that the solution of (18) is given by

$$\boldsymbol{p}_i = \begin{cases} \boldsymbol{q}_i - \beta\frac{\boldsymbol{q}_i}{|\boldsymbol{q}_i|}, & \text{if } |\boldsymbol{q}_i| > \beta, \\ 0, & \text{others.} \end{cases} \tag{19}$$

According to the complementary condition, the parameter $\beta > 0$ should satisfy

$$\|\boldsymbol{p}\|_1 = \sum_{i\in\{i:|\boldsymbol{q}_i|\geq\beta\}} \left|\boldsymbol{q}_i - \beta\frac{\boldsymbol{q}_i}{|\boldsymbol{q}_i|}\right| = \sum_{i\in\{i:|\boldsymbol{q}_i|\geq\beta\}} |\boldsymbol{q}_i|\left|1 - \frac{\beta}{|\boldsymbol{q}_i|}\right| = 1. \tag{20}$$

Notice that $0 < 1 - \beta/|\boldsymbol{q}_i| < 1$ for $|\boldsymbol{q}_i| > \beta$, the above inequality implies that

$$1 = \|\boldsymbol{p}\|_1 \leq \sum_{i\in\{i:|\boldsymbol{q}_i|\geq\beta\}} \left|\boldsymbol{q}_i\right| \leq \|\boldsymbol{q}\|_1.$$

If $\|\boldsymbol{q}\|_1 > 1$, we know that there exists $\beta > 0$.

Now, we consider how to find β to satisfy the Equation (20). Let $\boldsymbol{r}$ be the vector obtained by sorting the elements of the vector $|\boldsymbol{q}|$ in ascending order, that is,

$$\boldsymbol{r}_1 \leq \boldsymbol{r}_2 \leq \cdots \leq \boldsymbol{r}_L.$$

Assume that i_0 is the minimum index such that $\boldsymbol{r}_{i_0} \geq \beta$, the equation in (20) can be rewritten as

$$\|\boldsymbol{p}\|_1 = \sum_{j=i_0}^{L}(\boldsymbol{r}_j - \beta).$$

We let $\boldsymbol{r}_0 = 0$ and define the sequence $E_i (i = 0, 1, \ldots, L-1)$ as follows:

$$E_i = \sum_{j=i+1}^{L}(\boldsymbol{r}_j - \boldsymbol{r}_i).$$

Then, E_i is a decreasing sequence with $E_0 = \sum_{j=1}^{L}\boldsymbol{r}_j = \|\boldsymbol{q}\|_1 \geq 1$. We have either $E_{L-1} \geq 1$ or $E_{L-1} < 1$. When $E_{L-1} \geq 1$, we have $\boldsymbol{r}_L \geq \boldsymbol{r}_{L-1} + 1$. We set $\beta = \boldsymbol{r}_L - 1$. When $E_{L-1} < 1$, there must exist an index $i_0 < L-1$ such that $E_{i_0} \geq 1 > E_{i_0+1}$. We consider the function

$$\varphi(x) = \sum_{j=i_0+1}^{L}(\boldsymbol{r}_j - x).$$

We can verify that $\varphi(\boldsymbol{r}_{i_0}) = E_{i_0}$ and $\varphi(\boldsymbol{r}_{i_0+1}) = E_{i_0+1}$. The function $\varphi(x)$ is continuous in $[\boldsymbol{r}_{i_0}, \boldsymbol{r}_{i_0+1}]$ with $\varphi(\boldsymbol{r}_{i_0}) \geq 1$ and $\varphi(\boldsymbol{r}_{i_0+1}) < 1$; there must exist a β such that $\varphi(\beta) = 1$. Hence, we have

$$\beta = \boldsymbol{r}_{i_0} + \frac{E_{i_0} - 1}{L - i_0}. \tag{21}$$

3 | CONVERGENCE ANALYSIS

We show that the sequence $\{(\boldsymbol{w}^{(k)}, \boldsymbol{p}^{(k)}, \boldsymbol{z}^{(k)})\}$ generated by (10)–(13) converges to the saddle point of $\mathcal{J}(\boldsymbol{w}, \boldsymbol{p}, \boldsymbol{z})$ in (9). First, we introduce the inner product of the vector $\boldsymbol{x}_i = \begin{pmatrix} \boldsymbol{w}_i \\ \boldsymbol{p}_i \\ \boldsymbol{z}_i \end{pmatrix} (i = 1, 2)$ with $\boldsymbol{w}_i, \boldsymbol{p}_i, \boldsymbol{z}_i \in Y$. We define the inner product

$$\langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle = \langle \boldsymbol{w}_1, \boldsymbol{w}_2 \rangle_Y + \langle \boldsymbol{p}_1, \boldsymbol{p}_2 \rangle_Y + \langle \boldsymbol{z}_1, \boldsymbol{z}_2 \rangle_Y.$$

We have the following lemma.

Lemma 1. *Assume that s, t, θ are defined by (10)–(13). Let $\lambda = \|\nabla \mathrm{div}\|_2^2$ and*

$$\boldsymbol{W} = \begin{pmatrix} \frac{1}{s}\boldsymbol{I} & -\nabla \mathrm{div} & -\beta \boldsymbol{I} \\ -\nabla \mathrm{div} & \frac{1}{t}\boldsymbol{I} & 0 \\ -\beta \boldsymbol{I} & 0 & \frac{1}{t}\boldsymbol{I} \end{pmatrix}.$$

Then, when $s(\lambda + \beta) < 1$ and $t < min(1, 1/\beta)$, $\boldsymbol{W}$ is a positive definite matrix.

Proof. Denote $\boldsymbol{x} = \begin{pmatrix} \boldsymbol{w} \\ \boldsymbol{p} \\ \boldsymbol{z} \end{pmatrix}$ with $\boldsymbol{w}, \boldsymbol{p}, \boldsymbol{z} \in Y$, we have

$$\begin{aligned} \langle \boldsymbol{x}, \boldsymbol{W}\boldsymbol{x} \rangle &= \left\langle \boldsymbol{w}, \frac{1}{s}\boldsymbol{w} - \nabla \mathrm{div}\boldsymbol{p} - \beta \boldsymbol{z} \right\rangle_Y + \left\langle \boldsymbol{p}, -\nabla \mathrm{div}\boldsymbol{w} + \frac{1}{t}\boldsymbol{p} \right\rangle_Y + \left\langle \boldsymbol{z}, -\beta \boldsymbol{w} + \frac{1}{t}\boldsymbol{z} \right\rangle_Y \\ &= \frac{1}{s}\|\boldsymbol{w}\|_2^2 + \frac{1}{t}\|\boldsymbol{p}\|_2^2 + \frac{1}{t}\|\boldsymbol{z}\|_2^2 - 2\langle \boldsymbol{p}, \nabla \mathrm{div}\boldsymbol{w} \rangle_Y - 2\beta \langle \boldsymbol{w}, \boldsymbol{z} \rangle_Y. \end{aligned}$$

Using the inequality $2ab \le a^2 + b^2$, we obtain

$$\begin{aligned} \langle \boldsymbol{x}, \boldsymbol{W}\boldsymbol{x} \rangle &\ge \frac{1}{s}\|\boldsymbol{w}\|_2^2 + \frac{1}{t}\|\boldsymbol{p}\|_2^2 + \frac{1}{t}\|\boldsymbol{z}\|_2^2 - \|\boldsymbol{p}\|_2^2 - \|\nabla \mathrm{div}\|_2^2 \|\boldsymbol{w}\|_2^2 - \beta \|\boldsymbol{w}\|_2^2 - \beta \|\boldsymbol{z}\|_2^2 \\ &= \left(\frac{1}{s} - \|\nabla \mathrm{div}\|_2^2 - \beta\right)\|\boldsymbol{w}\|_2^2 + \left(\frac{1}{t} - 1\right)\|\boldsymbol{z}\|_2^2 + \left(\frac{1}{t} - \beta\right)\|\boldsymbol{z}\|_2^2. \end{aligned}$$

Therefore, when $s(\lambda + \beta) < 1$ and $t < \min(1, 1/\beta)$, we have $\langle \boldsymbol{x}, \boldsymbol{W}\boldsymbol{x} \rangle_Y > 0$ for any $\boldsymbol{x} \neq 0$. It follows that $\boldsymbol{W}$ is a positive definite matrix. □

The following theorem states that the sequence $(\boldsymbol{w}^{(k)}, \boldsymbol{p}^{(k)}, \boldsymbol{z}^{(k)})$ generated by (10)–(13) with $\theta = 1$ will converge to the saddle point of $\mathcal{J}(\boldsymbol{w}, \boldsymbol{p}, \boldsymbol{z})$ under some assumption.

Theorem 1. *Let $\lambda = \|\nabla \mathrm{div}\|_2^2$ and $(\boldsymbol{w}^{(k)}, \boldsymbol{p}^{(k)}, \boldsymbol{z}^{(k)})$ be the sequence generated by (10)–(13) with $\theta = 1$. Assume that $\mathcal{J}(\boldsymbol{w}, \boldsymbol{p}, \boldsymbol{z})$ in (9) has a saddle point $(\boldsymbol{w}^*, \boldsymbol{p}^*, \boldsymbol{z}^*)$. When the step sizes s, t satisfy the conditions $s(\lambda + \beta) < 1$ and $t < \min(1, 1/\beta)$, respectively, there exists a saddle point $(\boldsymbol{w}^*, \boldsymbol{p}^*, \boldsymbol{z}^*)$ such that*

$$\boldsymbol{w}^{(k)} \to \boldsymbol{w}^*, \boldsymbol{p}^{(k)} \to \boldsymbol{p}^* \quad \text{and} \quad \boldsymbol{z}^{(k)} \to \boldsymbol{z}^*.$$

Proof. A solution of $\mathcal{J}$ is a saddle point if and only if it satisfies the optimal condition

$$\langle \boldsymbol{w} - \boldsymbol{w}^*, \nabla \mathrm{div}\boldsymbol{p}^* + \beta \boldsymbol{z}^* \rangle_Y = 0 \tag{22a}$$

$$\langle \boldsymbol{p} - \boldsymbol{p}^*, \nabla \boldsymbol{f} - \mathrm{div}\boldsymbol{w}^* \rangle_Y \ge 0 \tag{22b}$$

$$\langle \boldsymbol{z} - \boldsymbol{z}^*, -\beta \boldsymbol{w}^* \rangle_Y \ge 0. \tag{22c}$$

On the other hand, $\boldsymbol{w}^{(k)}, \boldsymbol{p}^{(k)}, \boldsymbol{z}^{(k)}$ are the solution of the minimization problems in (10)–(13) if and only if they satisfy

$$\left\langle \boldsymbol{w} - \boldsymbol{w}^{(k+1)}, \frac{1}{s}(\boldsymbol{w}^{(k+1)} - \boldsymbol{w}^{(k)}) + \nabla \mathrm{div} \boldsymbol{p}^{(k)} + \beta \boldsymbol{z}^{(k)} \right\rangle_Y = 0 \tag{23a}$$

$$\left\langle \boldsymbol{p} - \boldsymbol{p}^{(k+1)}, \frac{1}{t}(\boldsymbol{p}^{(k+1)} - \boldsymbol{p}^{(k)}) + \nabla(\boldsymbol{f} - \mathrm{div}\overline{\boldsymbol{w}}^{(k)}) \right\rangle_Y \geq 0 \tag{23b}$$

$$\left\langle \boldsymbol{z} - \boldsymbol{z}^{(k+1)}, \boldsymbol{z}^{(k+1)} - \boldsymbol{z}^{(k)} - \beta t \overline{\boldsymbol{w}}^{(k)} \right\rangle_Y \geq 0 \tag{23c}$$

for any $\boldsymbol{w}, \boldsymbol{p}, \boldsymbol{z} \in Y$.

For simplicity, we define

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{w} \\ \boldsymbol{p} \\ \boldsymbol{z} \end{pmatrix}, \quad \phi(\boldsymbol{x}) = \begin{pmatrix} \nabla \mathrm{div} \boldsymbol{p} + \beta \boldsymbol{z} \\ \nabla (\boldsymbol{f} - \mathrm{div} \boldsymbol{w}) \\ -\beta \boldsymbol{w} \end{pmatrix}.$$

Then, the optimal condition in (22) and (23) can be rewritten as

$$\langle \boldsymbol{x} - \boldsymbol{x}^*, \phi(\boldsymbol{x}^*) \rangle \geq 0 \tag{24}$$

and

$$\left\langle \boldsymbol{x} - \boldsymbol{x}^{(k+1)}, \phi(\boldsymbol{x}^{(k+1)}) + \boldsymbol{W}(\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}) \right\rangle \geq 0, \tag{25}$$

respectively. Here, $\boldsymbol{W}$ is the matrix defined in Lemma 1, which states that $\boldsymbol{W}$ is a positive definition matrix under the assumption of the step size.

Now, we set $\boldsymbol{x}$ to $\boldsymbol{x}^*$ in (25) and $\boldsymbol{x}$ to $\boldsymbol{x}^{k+1}$ in (24), and then, sum the inequalities (24) and (25), we obtain

$$\left\langle \boldsymbol{x}^* - \boldsymbol{x}^{(k+1)}, \phi(\boldsymbol{x}^{(k+1)}) - \phi(\boldsymbol{x}^*) \right\rangle + \left\langle \boldsymbol{x}^* - \boldsymbol{x}^{(k+1)}, \boldsymbol{W}(\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}) \right\rangle \geq 0. \tag{26}$$

Notice that

$$\boldsymbol{x} - \boldsymbol{x}^* = \begin{pmatrix} \boldsymbol{w} - \boldsymbol{w}^* \\ \boldsymbol{p} - \boldsymbol{p}^* \\ \boldsymbol{z} - \boldsymbol{z}^* \end{pmatrix} \quad \text{and} \quad \phi(\boldsymbol{x}) - \phi(\boldsymbol{x}^*) = \begin{pmatrix} \nabla \mathrm{div}(\boldsymbol{p} - \boldsymbol{p}^*) + \beta(\boldsymbol{z} - \boldsymbol{z}^*) \\ -\nabla \mathrm{div}(\boldsymbol{w} - \boldsymbol{w}^*) \\ -\beta(\boldsymbol{w} - \boldsymbol{w}^*) \end{pmatrix}.$$

We have

$$\langle \boldsymbol{x} - \boldsymbol{x}^*, \phi(\boldsymbol{x}) - \phi(\boldsymbol{x}^*) \rangle = 0. \tag{27}$$

On the other hand, similarly to triangle identity of Euclidean geometry, we have

$$\left\| \boldsymbol{x}^* - \boldsymbol{x}^{(k)} \right\|_W^2 = \left\| \boldsymbol{x}^* - \boldsymbol{x}^{(k+1)} \right\|_W^2 + \left\| \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)} \right\|_W^2 + \langle \boldsymbol{x}^* - \boldsymbol{x}^{(k+1)}, \boldsymbol{W}(\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}) \rangle. \tag{28}$$

Setting $\boldsymbol{x} = \boldsymbol{x}^{(k+1)}$ in (27) and then summing (26), (27), and (28), we have

$$\left\| \boldsymbol{x}^* - \boldsymbol{x}^{(k)} \right\|_W^2 \geq \left\| \boldsymbol{x}^* - \boldsymbol{x}^{(k+1)} \right\|_W^2 + \left\| \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)} \right\|_W^2. \tag{29}$$

The inequality (29) implies that the sequence $\{\|\boldsymbol{x}^* - \boldsymbol{x}^{(k)}\|_W^2\}$ is both bounded and nonincreasing. It is known that a bounded nonincreasing sequence is convergent; therefore, the sequence $\{\|\boldsymbol{x}^* - \boldsymbol{x}^{(k)}\|_W^2\}$ is convergent. Hence, the sequence $\{\boldsymbol{x}^{(k)}\}$ contains a subsequence $\boldsymbol{x}^{(k_j)}$ converging to some limit point $\boldsymbol{x}^\dagger$. Notice that the inequality (29) also implies that

$$\lim_{k \to \infty} (\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}) = 0.$$

In particular, $\boldsymbol{x}^{(k_j+1)}$ also converges to $\boldsymbol{x}^\dagger$. Set $\boldsymbol{x}^{(k_j)}$ in (23) and let $j \to \infty$, we obtain

$$\langle \boldsymbol{x} - \boldsymbol{x}^\dagger, \phi(\boldsymbol{x}^\dagger) \rangle \geq 0.$$

It implies that $\boldsymbol{x}^{\dagger}$ satisfies the optimal condition and, hence, is a saddle point of $\mathcal{J}$. Set $\boldsymbol{x}^{*} = \boldsymbol{x}^{\dagger}$ and assume that α is the minimum eigenvalue of the positive definition matrix $\boldsymbol{W}$, we have $\alpha > 0$ and

$$\alpha \left\| \boldsymbol{x}^{\dagger} - \boldsymbol{x}^{(k)} \right\|_2^2 \leq \left\| \boldsymbol{x}^{\dagger} - \boldsymbol{x}^{(k)} \right\|_W^2 .$$

Because the sequence $\{\boldsymbol{x}^{(k)}\}$ contains a subsequence converging to $\boldsymbol{x}^{\dagger}$ and $\|\boldsymbol{x}^{\dagger} - \boldsymbol{x}^{(k)}\|_W^2$ is nonincreasing, the whole sequence $\{\boldsymbol{x}^{(k)}\}$ converges to $\boldsymbol{x}^{\dagger}$. □

4 | NUMERICAL RESULTS

In this section, we illustrate the performance of our proposed algorithm for the image decomposition problem. Though several efficient numerical algorithms have been developed to solve the variants of the Meyer G-norm–based minimization problem, few works have been developed to solve the exact G-norm–based minimization problem due to the definition of the Meyer G-norm. More recently, the alternating direction method of multipliers (ADMM) and the nonlinear projection (NLP) method were applied to solve the exact G-norm–based minimization problem.[25,29] We remark that the algorithms in the works of Aujol et al.[25] and Ng et al.[29] cannot be directly applied to solve the original Meyer image decomposition model. In fact, the ADMM was applied to solve the VO model in (5) with $p = \infty$,[29] whereas Chambolle's projection method was developed to solve the AABC model in (6). We will compare the proposed method with the ADMM and Chambolle's projection method proposed in the works of Aujol et al.[25] and Ng et al.[29] The stopping criterion of all the methods is that the relative difference between the successive iterate of the restored image should be less than 5×10^{-5} or the number of iterations reaches 1,000.

In order to measure the decomposition results, we generate two synthetic images. We generate the original cartoon image $\boldsymbol{u}$ and the original texture image $\boldsymbol{v}$ by the image decomposition algorithm such that $\boldsymbol{u}$ and $\boldsymbol{v}$ are consistent with Meyer's model, where the cartoon component is well structured and the texture is modeled by oscillatory functions taking both positive and negative values, and of zero mean.[15] Then, the synthetic image is superposed by a cartoon image and a texture image (see Figure 1). The quality of the decomposed images is assessed using the signal-to-noise ratio (SNR) and

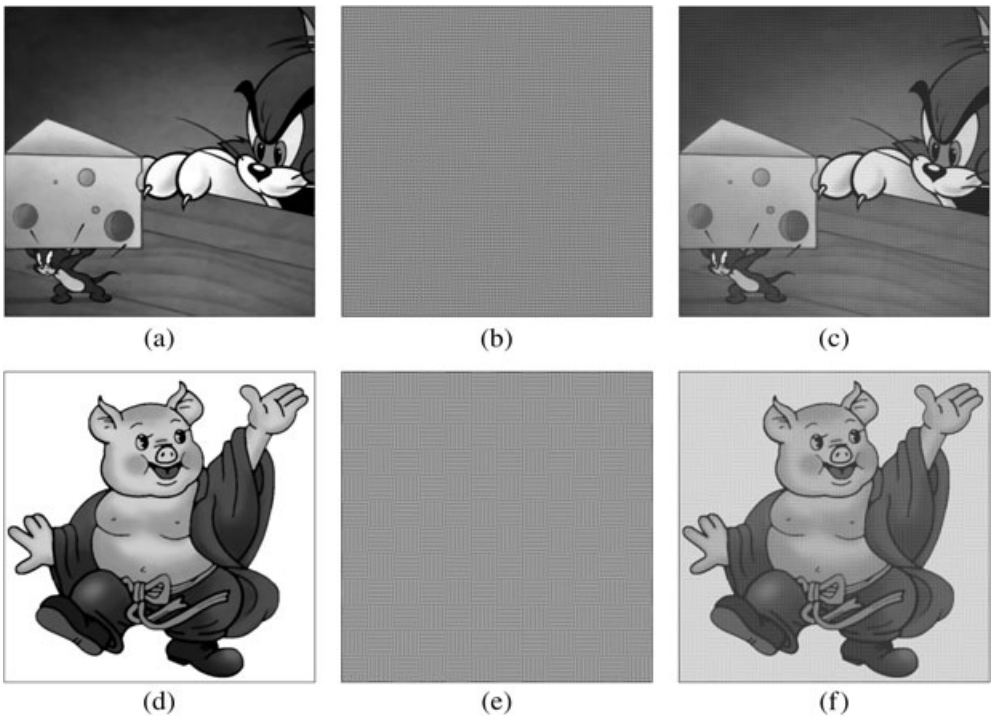

FIGURE 1 The original cartoon image, the original texture image, and the synthetic image. (Top) Tom & Jerry image. (Bottom) pig image. The range of the pixel value of the images is normalized to [0, 255]. (a) Cartoon part $\boldsymbol{u}$. (b) Texture part $\boldsymbol{v}$. (c) Synthetic image $\boldsymbol{f} = \boldsymbol{u} + \boldsymbol{v}$. (d) Cartoon part $\boldsymbol{u}$. (e) Texture part $\boldsymbol{v}$. (f) Synthetic image $\boldsymbol{f} = \boldsymbol{u} + \boldsymbol{v}$

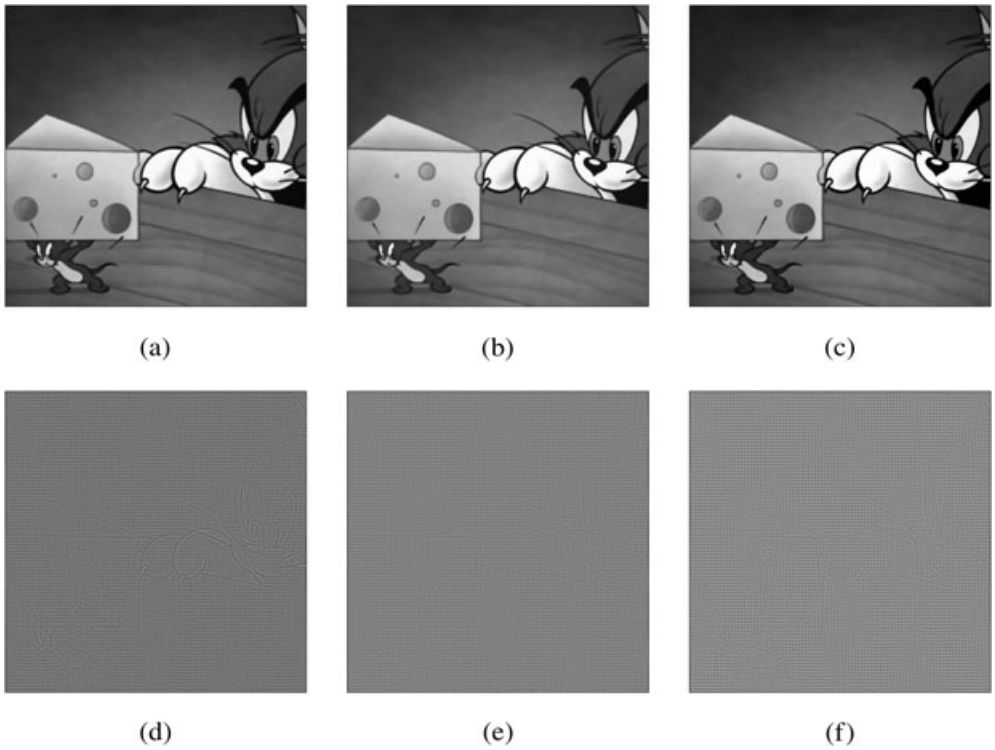

FIGURE 2 The parameters are as follows: $\alpha = 0.19, \beta = 0.12$ for the alternating direction method of multipliers (ADMM) in the work of Ng et al.,[29] $\alpha = 0.4, \mu = 31.1$ for the nonlinear projection (NLP) method in the work of Aujol et al.,[25] and $\beta = 1$ for the proposed method. The range of the pixel value of the images is normalized to $[0, 255]$. (a) $\widehat{\boldsymbol{u}}$ (Ng et al.[29]). (b) $\widehat{\boldsymbol{u}}$ (Aujol et al.[25]). (c) $\widehat{\boldsymbol{u}}$ (proposed). (d) $\widehat{\boldsymbol{v}}$ (Ng et al.[29]). (e) $\widehat{\boldsymbol{v}}$ (Aujol et al.[25]). (f) $\widehat{\boldsymbol{v}}$ (proposed)

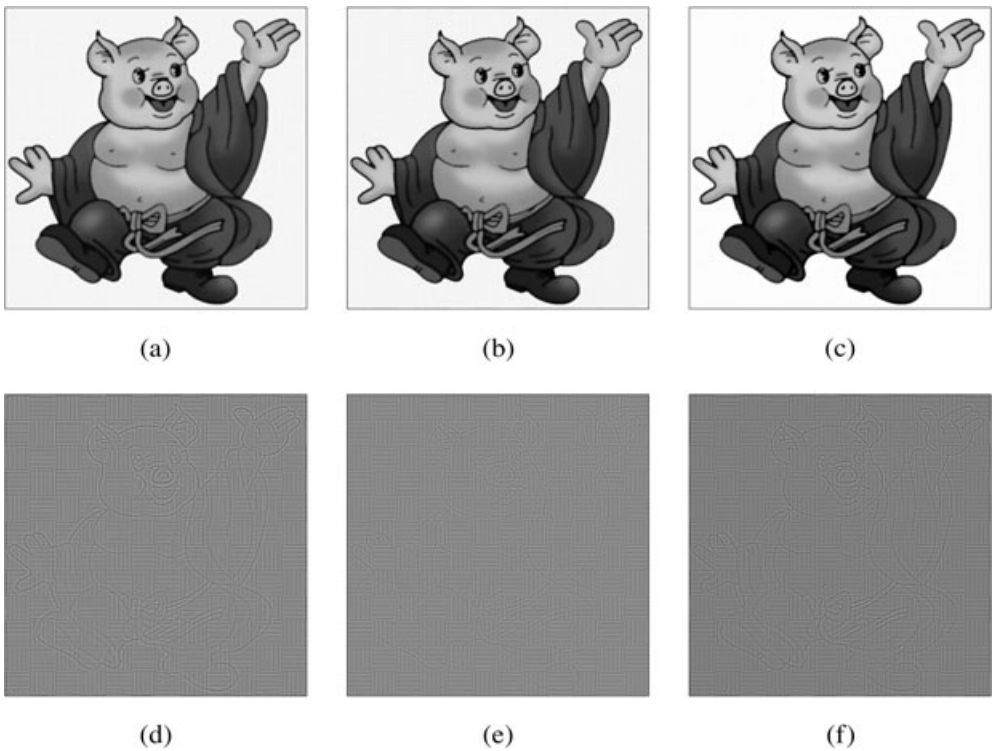

FIGURE 3 The parameters are as follows: $\alpha = 0.32, \beta = 0.20$ for the alternating direction method of multipliers (ADMM) in the work of Ng et al.,[29] $\alpha = 0.5, \mu = 21.38$ for the nonlinear projection (NLP) method in the work of Aujol et al.,[25] and $\beta = 1$ for the proposed method. The range of the pixel value of the images is normalized to $[0, 255]$. (a) $\widehat{\boldsymbol{u}}$ (Ng et al.[29]). (b) $\widehat{\boldsymbol{u}}$ (Aujol et al.[25]). (c) $\widehat{\boldsymbol{u}}$ (proposed). (d) $\widehat{\boldsymbol{v}}$ (Ng et al.[29]). (e) $\widehat{\boldsymbol{v}}$ (Aujol et al.[25]). (f) $\widehat{\boldsymbol{v}}$ (proposed)

absolute error. Mathematically, when the parameter α in (5) and (6) tends to infinity, these two models coincide with the original Meyer model. In theory, α should be chosen as large as possible. However, we find that, when we choose a large α, the SNRs for both decomposed cartoon image and the texture image become worse. Therefore, we apply a trial-and-error method to choose the best value of α from the tested values such that the sum of the SNRs of the decomposed cartoon and texture images is the largest. Once α is determined, we apply the similar rule to choose the parameter β in (5). The parameter β in the exact Meyer model (3) is fixed to 1. The parameter μ in (6) is set the G-norm of the original texture image. The decomposed results are shown in Figures 2 and 3 with their largest SNRs in different image decomposition methods. We list the SNRs for the decomposed cartoon and texture images obtained by different methods in Table 1. The results show that the SNR of the proposed method for both the cartoon image and the texture image are better than those of the methods in the works of Aujol et al.[25] and Ng et al.[29]

Assume that $\boldsymbol{f}$ is the target image, and $\hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{v}}$ are respectively the cartoon part and texture part, the errors for the decomposed images obtained by different methods are shown in Table 2. The residual errors $\|\boldsymbol{f}-\hat{\boldsymbol{u}}-\hat{\boldsymbol{v}}\|_2$ for the method[29] is small; we can observe that $\|\boldsymbol{u}-\hat{\boldsymbol{u}}\|_2 \approx \|\boldsymbol{v}-\hat{\boldsymbol{v}}\|_2$ for the method[29] in Table 2. Because there is $\boldsymbol{f}=\boldsymbol{u}+\boldsymbol{v}=\hat{\boldsymbol{u}}+\hat{\boldsymbol{v}}$ for the proposed method, it is obviously $\|\boldsymbol{u}-\hat{\boldsymbol{u}}\|_2 = \|\boldsymbol{v}-\hat{\boldsymbol{v}}\|_2$. Although the method in the work of Aujol et al.[25] is more efficient than that in the work of Ng et al.[29] with respect to SNR measurement, it is poor in residual errors $\|\boldsymbol{f}-\hat{\boldsymbol{u}}-\hat{\boldsymbol{v}}\|_2$ (see Figure 4.4). A possible explanation is that the VO model (5) is a relaxed version of the AABC model (6), whereas the methods in the works of Ng et al.[29] and Aujol et al.[25] are the numerical algorithm for the VO model and the AABC model, respectively.

Next, we apply the proposed method to two 512×512 images: "Barbara" image and "Dollar" image, as shown in Figure 5. We show the results of the proposed algorithm in Figures 6 and 7. The regularization parameter in (3) is set to $\beta = 0.1$ in the experiments. For the VO model in (5), we set $\alpha = 0.1$ and $\beta = 0.01$. For the AABC model in (6), we set $\alpha = 10$ and $\mu = \|\hat{\boldsymbol{w}}_\beta\|_\infty$, where $\hat{\boldsymbol{w}}_\beta$ is the solution obtained by the proposed method with regularization parameter β. We also test other values of parameter μ; similar decomposed results are obtained. According to decomposed images in Figures 6 and 7, the proposed method provide better visual cartoon and texture components than the other two methods. In particular, we present the zoomed versions of the cartoon and texture components of "Barbara" image in Figure 6(g)–(l) and "Dollar" image in Figure 7(g)–(l). In the "Tablecloth" of "Barbara" image, the texture component generated by the

TABLE 1 Signal-to-noise ratios for the decomposed cartoon and texture images obtained by the alternating direction method of multipliers (ADMM) in the work of Ng et al.,[29] the nonlinear projection (NLP) method in the work of Aujol et al.,[25] and the proposed method

	Method	Cartoon $\hat{\boldsymbol{u}}$	Texture $\hat{\boldsymbol{v}}$	$\hat{\boldsymbol{u}}+\hat{\boldsymbol{v}}$
Tom & Jerry	ADMM ($\alpha = 0.32, \beta = 0.24$)	29.28	12.37	41.65
	NLP($\alpha = 0.25, \mu = 28.5$)	30.78	12.02	42.80
	Proposed ($\beta = 1$)	31.65	14.84	46.49
Pig	ADMM ($\alpha = 0.32, \beta = 0.20$)	30.97	11.51	42.48
	NLP ($\alpha = 0.25, \mu = 21.4$)	31.62	10.00	41.62
	Proposed ($\beta = 1$)	31.87	12.42	44.29

TABLE 2 Relative errors for the decomposed images obtained by the alternating direction method of multipliers (ADMM) in the work of Ng et al.,[29] the nonlinear projection (NLP) method in the work of Aujol et al.,[25] and the proposed method

	Method	ADMM	NLP	Proposed
Tom & Jerry	$\|\boldsymbol{u}-\hat{\boldsymbol{u}}\|_2/\|\boldsymbol{f}\|_2$	0.0339	0.0286	0.0258
	$\|\boldsymbol{v}-\hat{\boldsymbol{v}}\|_2/\|\boldsymbol{f}\|_2$	0.0342	0.0357	0.0258
	$\|\boldsymbol{f}-\hat{\boldsymbol{u}}-\hat{\boldsymbol{v}}\|_2/\|\boldsymbol{f}\|_2$	0.0017	0.0169	**0**
Pig	$\|\boldsymbol{u}-\hat{\boldsymbol{u}}\|_2/\|\boldsymbol{f}\|_2$	0.0281	0.0261	0.0254
	$\|\boldsymbol{v}-\hat{\boldsymbol{v}}\|_2/\|\boldsymbol{f}\|_2$	0.0282	0.0335	0.0254
	$\|\boldsymbol{f}-\hat{\boldsymbol{u}}-\hat{\boldsymbol{v}}\|_2/\|\boldsymbol{f}\|_2$	0.0012	0.0181	**0**

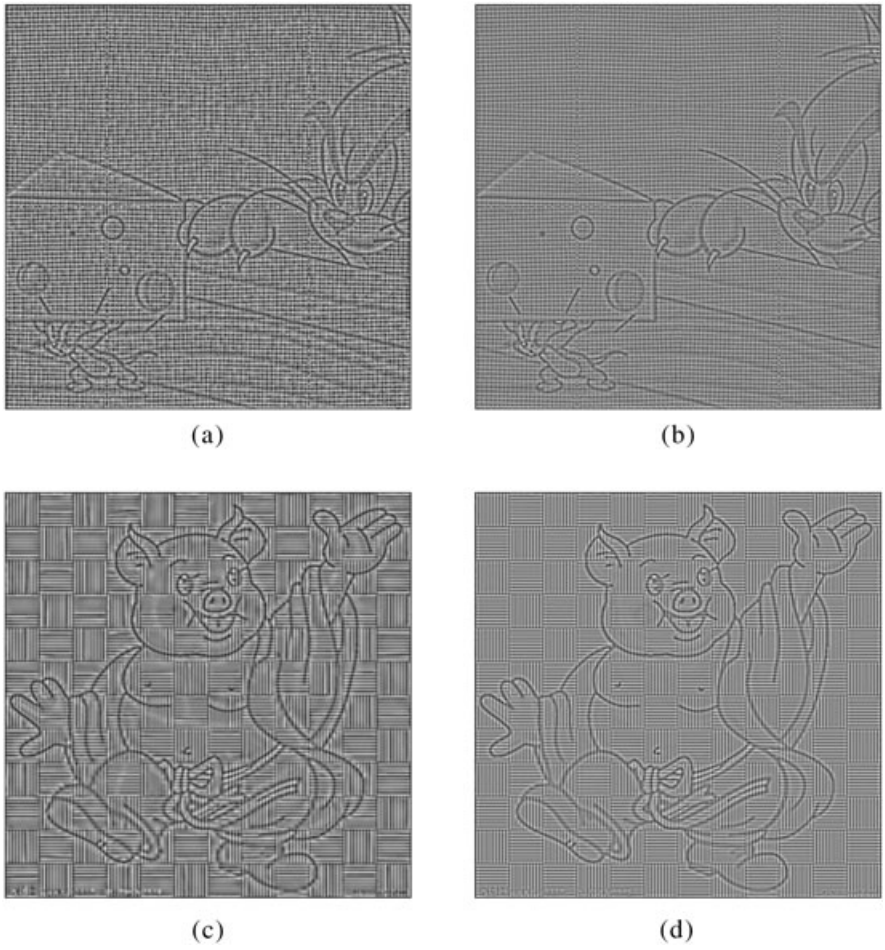

FIGURE 4 The residuals of the decomposed images in Figures 4.1 and 4.2 by the alternating direction method of multipliers (ADMM) in the work of Ng et al.[29] and the nonlinear projection (NLP) method in the work of Aujol et al.[25] (a) $\boldsymbol{f} - \hat{\boldsymbol{u}} - \hat{\boldsymbol{v}}$ (Ng et al.[29]). (b) $\boldsymbol{f} - \hat{\boldsymbol{u}} - \hat{\boldsymbol{v}}$ (Aujol et al.[25]). (c) $\boldsymbol{f} - \hat{\boldsymbol{u}} - \hat{\boldsymbol{v}}$ (Ng et al.[29]). (d) $\boldsymbol{f} - \hat{\boldsymbol{u}} - \hat{\boldsymbol{v}}$ (Aujol et al.[25])

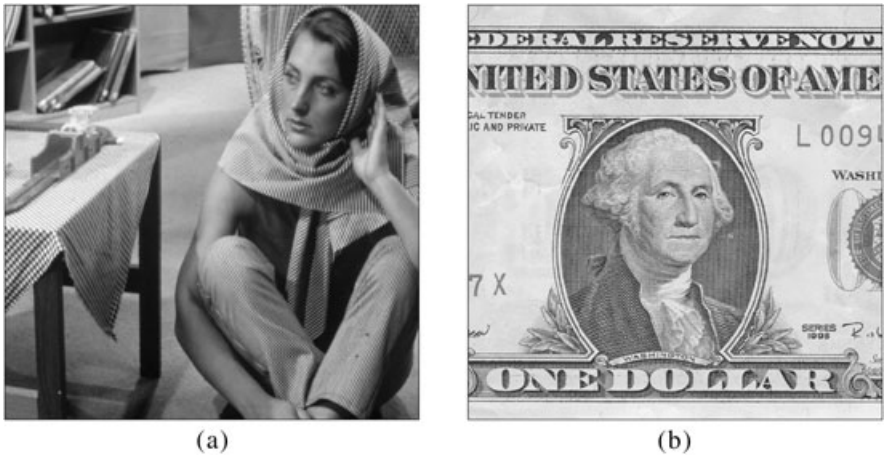

FIGURE 5 The original real images. (a) 512 × 512 Barbara image. (b) 512 × 512 Dollar note image

proposed method is visually decomposed better than those by the other two testing methods. In the "One" symbol of "Dollar" image, the cartoon component generated by the proposed method is visually decomposed better than those by the other two testing methods.

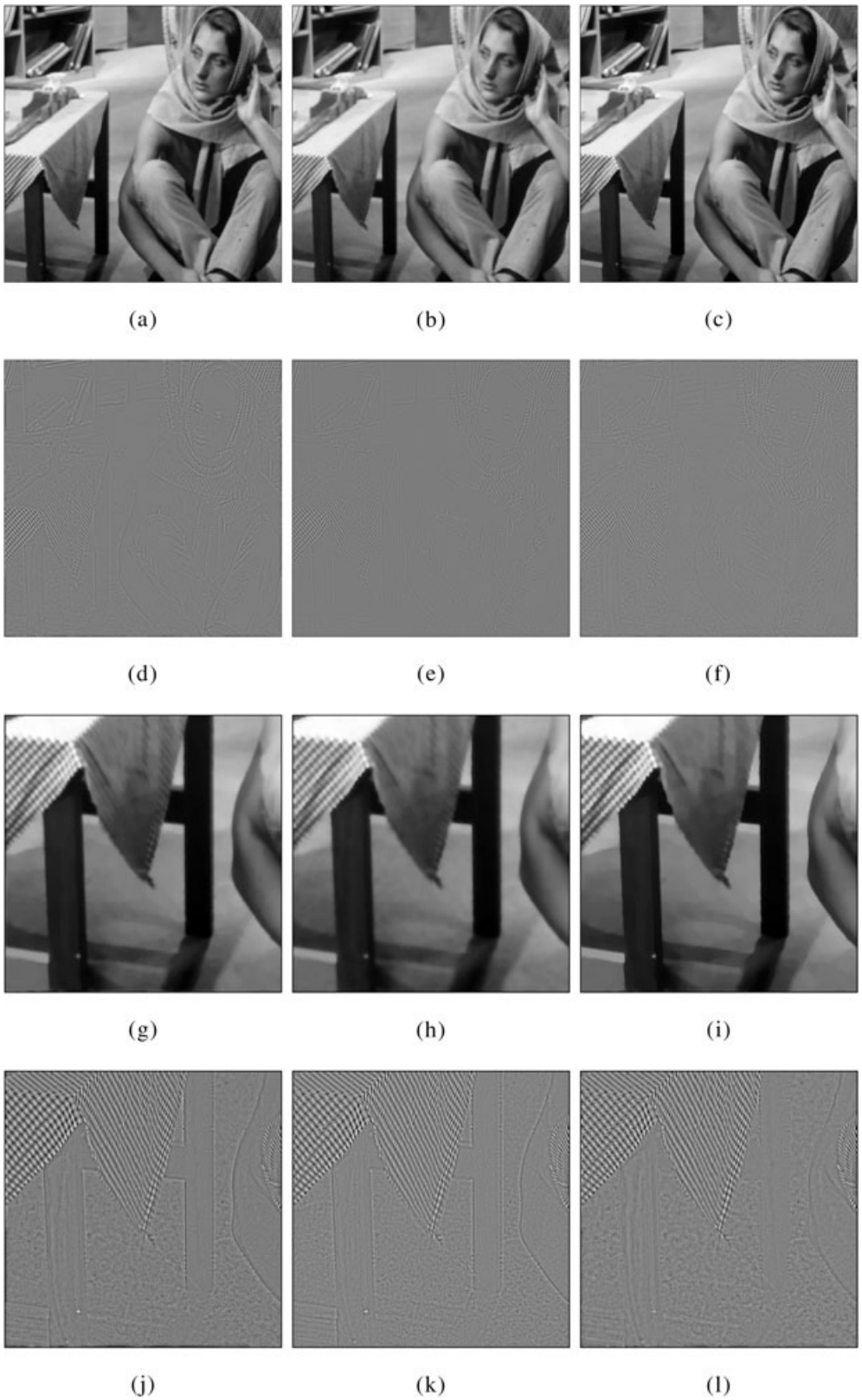

FIGURE 6 The decomposition results for the Dollar image by the alternating direction method of multipliers (ADMM) in the work of Ng et al.,[29] the nonlinear projection (NLP) method in the work of Aujol et al.,[25] and the proposed method. The range of the pixel value of the images is normalized to $[0, 255]$. (a) $\hat{\boldsymbol{u}}$ (Ng et al.[29]). (b) $\hat{\boldsymbol{u}}$ (Aujol et al.[25]). (c) $\hat{\boldsymbol{u}}$ (proposed). (d) $\hat{\boldsymbol{v}}$ (Ng et al.[29]). (e) $\hat{\boldsymbol{v}}$ (Aujol et al.[25]). (f) $\hat{\boldsymbol{v}}$ (proposed). (g) $\hat{\boldsymbol{u}}$ (Ng et al.[29]). (h) $\hat{\boldsymbol{u}}$ (Aujol et al.[25]). (i) $\hat{\boldsymbol{u}}$ (proposed). (j) $\hat{\boldsymbol{v}}$ (Ng et al.[29]). (k) $\hat{\boldsymbol{v}}$ (Aujol et al.[25]). (l) $\hat{\boldsymbol{v}}$ (proposed)

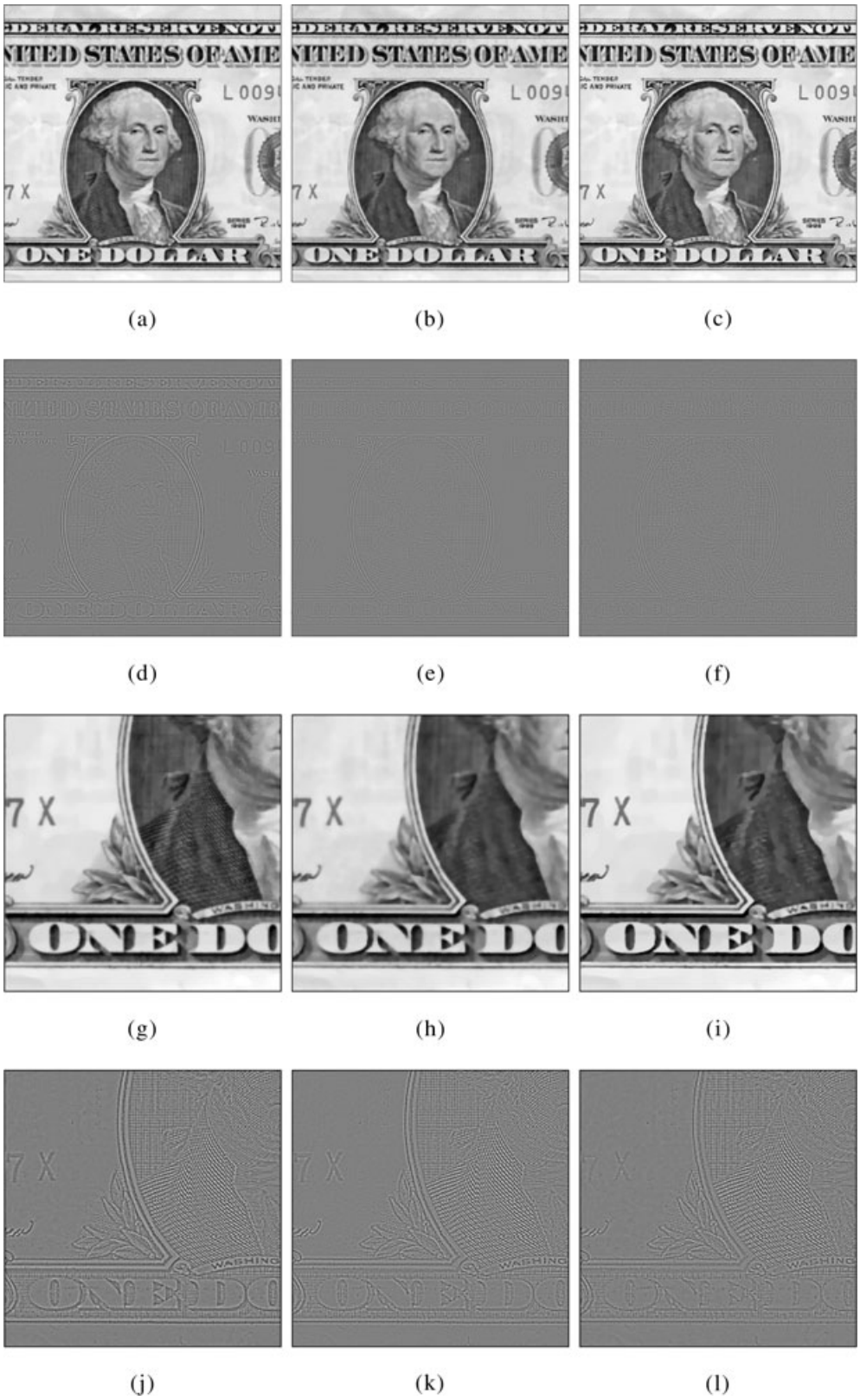

FIGURE 7 The decomposition results for the Barbara image by the alternating direction method of multipliers (ADMM) in the work of Ng et al.,[29] the nonlinear projection (NLP) method in the work of Aujol et al.,[25] and the proposed method. The range of the pixel value of the images is normalized to [0, 255]. (a) $\hat{\boldsymbol{u}}$ (Ng et al.[29]). (b) $\hat{\boldsymbol{u}}$ (Aujol et al.[25]). (c) $\hat{\boldsymbol{u}}$ (proposed). (d) $\hat{\boldsymbol{v}}$ (Ng et al.[29]). (e) $\hat{\boldsymbol{v}}$ (Aujol et al.[25]). (f) $\hat{\boldsymbol{v}}$ (proposed). (g) $\hat{\boldsymbol{u}}$ (Ng et al.[29]). (h) $\hat{\boldsymbol{u}}$ (Aujol et al.[25]). (i) $\hat{\boldsymbol{u}}$ (proposed). (j) $\hat{\boldsymbol{v}}$ (Ng et al.[29]). (k) $\hat{\boldsymbol{v}}$ (Aujol et al.[25]). (l) $\hat{\boldsymbol{v}}$ (proposed)

5 | CONCLUSION

In this paper, we have developed a numerical algorithm to solve the exact Meyer G-norm–based image decomposition problem. To handle the difficulty caused by the infinity norm, we represented it using a dual formulation and then applied a primal-dual algorithm to compute the solution. An advantage of the proposed method is that there exists a closed-form solution for each subproblem, and therefore, no inner iterations are needed in the proposed method. The convergence of the numerical algorithm is also studied and shown. Numerical results have shown that the exact Meyer model can give better image decomposition results than the other testing methods.

ACKNOWLEDGEMENTS

Y. Wen's research is supported in part through NSFC Grant 11871210, the Construct Program of the Key Discipline in Hunan Province, and the Scientific Research Fund of Hunan Provincial Education Department Grant 17A128. M. Ng's research is supported in part through the HKRGC GRF 1202715, 12306616, 12200317, 12300218, and HKBU RC-ICRS/16-17/03.

ORCID

You-Wei Wen https://orcid.org/0000-0002-9892-3668
Hai-Wei Sun https://orcid.org/0000-0002-5507-6083
Michael K. Ng https://orcid.org/0000-0001-6833-5227

REFERENCES

1. Rudin L, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms. Phys D Nonlinear Phenom. 1992;60:259–268.
2. Sprljan N, Mrak M, Izquierdo E. Image compression using a cartoon-texture decomposition technique. In: Proceedings of the 5th International Workshop on Image Analysis for Multimedia Interactive Services; 2004 Apr 21–23; Lisboa, Portugal. p. 91.
3. Casaca W, Paiva A, Gomez-Nieto E, Joia P, Nonato LG. Spectral image segmentation using image decomposition and inner product-based metric. J Math Imaging Vis. 2013;45(3):227–238.
4. Malik J, Belongie S, Leung T, Shi J. Contour and texture analysis for image segmentation. Int J Comput Vis. 2001;43(1):7–27.
5. Chen T, Yin W, Zhou XS, Comaniciu D, Huang TS. Total variation models for variable lighting face recognition. IEEE Trans Pattern Anal Mach Intell. 2006;28(9):1519–1524.
6. Bertalmio M, Vese L, Sapiro G, Osher S. Simultaneous structure and texture image inpainting. IEEE Trans Image Process. 2003;12(8):882–889.
7. Meyer FG, Averbuch AZ, Coifman RR. Multilayered image representation: application to image compression. IEEE Trans Image Process. 2002;11(9):1072–1080.
8. Candès EJ, Donoho DL. Curvelets: a surprisingly effective nonadaptive representation for objects with edges. Stanford, CA: Department of Statistics, Stanford University2000. Technical report.
9. Candes EJ, Donoho DL. New tight frames of curvelets and optimal representations of objects with piecewise $C2$ singularities. Comm Pure Appl Math. 2004;57:219–266.
10. Daubechies I. Ten lectures on wavelets. Philadelphia, PA: SIAM; 1992.
11. Ron A, Shen Z. Affine systems in $L^2(R^d)$: the analysis of the analysis operator. J Func Anal. 1997;148(2):408–447.
12. Elad M, Starck J-L, Querre P, Donoho DL. Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA). Appl Comput Harmon Anal. 2005;19(3):340–358. Available from: https://doi.org/10.1016/j.acha.2005.03.005
13. Fadili MJ, Starck J-L, Murtagh F. Inpainting and zooming using sparse representations. Comput J. 2009;52(1):64–79.
14. Mallat S. A wavelet tour of signal processing. 2nd ed. San Diego, CA: Academic Press; 1999.
15. Meyer Y. Oscillating patterns in image processing and nonlinear evolution equations. Mathematical foundation of turbulent viscous flows. Springer-Verlag Berlin Heidelberg: Berlin, Germany; 2002. p. 101–187. Lecture notes in mathematics, No. 1871.
16. Starck J-L, Elad M, Donoho DL. Image decomposition via the combination of sparse representations and a variational approach. IEEE Trans Image Process. 2005;14(10):1570–1582.
17. Nikolova M. A variational approach to remove outliers and impulse noise. J Math Imaging Vision. 2004;20(1–2):99–120.
18. Chan RH, Chen K. Multilevel algorithm for a Poisson noise removal model with total-variation regularization. Int J Comput Math. 2007;84(8):1183–1198.
19. Aubert G, Kornprobst P. Mathematical problems in image processing: Partial differential equations and the calculus of variations. Vol. 147. New York, NY: Springer-Verlag New York; 2006.
20. Osher S, Solé A, Vese L. Image decomposition and restoration using total variation minimization and the H^{-1}. Multiscale Model Simul. 2003;3:349–370.

21. Vese LA, Osher SJ. Modeling textures with total variation minimization and oscillating patterns in image processing. J Sci Comput. 2003;19(1–3):553–572. Available from: citeseer.ist.psu.edu/vese02modeling.html
22. Aujol J, Aubert G, Blanc-Feraud L, Chambolle A. Image decomposition application to SAR images. Berlin, Germany: Springer; 2003. p. 297–312. *Lecture notes in computer science*, No. 2695.
23. Chambolle A. An algorithm for total variation minimization and applications. J Math Imag Vis. 2004;20(1-2):89–97.
24. Gilles J, Osher S. Bregman implementation of Meyer's *G*-norm for cartoon+textures decomposition. 2011; 11–73. UCLA CAM report.
25. Aujol J-F, Chambolle A. Dual norms and image decomposition models. Int J Comput Vis. 2005;63(1):85–104.
26. Aujol J, Gilboa G, Chan T, Osher S. Structure-texture image decomposition—modeling, algorithms, and parameter selection. Int J Comput Vis. 2006;67(1):111–136.
27. Elion C, Vese LA. An image decomposition model using the total variation and the infinity Laplacian. Electronic Imaging 2007; 2007 Jan 28–Feb 1; San Jose, CA. Bellingham, WA: International Society for Optics and Photonics; 2007.
28. Bonamy C, Le Guyader C. Split Bregman iteration and infinity Laplacian for image decomposition. J Comput Appl Math. 2013;240:99–110.
29. Ng MK, Yuan X, Zhang W. Coupled variational image decomposition and restoration model for blurred cartoon-plus-texture images with missing pixels. IEEE Trans Image Process. 2013;22(6):2233–2246.
30. Chambolle A, Pock T. A first-order primal-dual algorithm for convex problems with applications to imaging. J Math Imaging Vis. 2011;40(1):120–145.
31. Chen G, Teboulle M. A proximal-based decomposition method for convex minimization problems. Math Program Ser A. 1994;64(1–3):81–101.
32. Combettes PL, Pesquet J-L. A proximal decomposition method for solving convex variational inverse problems. Inverse Probl. 2008;24(6). Article No. 065014.
33. Dupe F-X, Fadili JM, Starck J-L. A proximal iteration for deconvolving Poisson noisy images using sparse representations. IEEE Trans Image Process. 2009;18(2):310–321.
34. Eckstein J, Bertsekas DP. On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math Program Ser A. 1992;55(1–3):293–318.
35. Rockafellar RT. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. Math Oper Res. 1976;1(2):97–116.
36. Rockafellar R. Monotone operators and the proximal point algorithm. SIAM J Control Optim. 1976;14(5):877–898.
37. Tseng P. Applications of a splitting algorithm to decomposition in convex programming and variational inequalities. SIAM J Control Optim. 1991;29(1):119–138.
38. Bertsekas D. Convex optimization theory. Belmont, MA: Athena Scientific; 2009.
39. Zhu M, Chan T. An efficient primal-dual hybrid gradient algorithm for total variation image restoration. 2007; 8–34. UCLA CAM report.
40. Bertsekas D, Nedic A, Ozdaglar E. Convex analysis and optimization. Belmont, MA: Athena Scientific; 2003.

How to cite this article: Wen Y-W, Sun H-W, Ng MK. A primal-dual method for the Meyer model of cartoon and texture decomposition. *Numer Linear Algebra Appl*. 2018;e2224. https://doi.org/10.1002/nla.2224