
Numerical Methods and Applications in Total Variation Image Restoration

Raymond Chan, Tony F. Chan, and Andy Yip

Contents

1	Introduction.....	1502
2	Background.....	1502
3	Mathematical Modeling and Analysis.....	1503
	Variants of Total Variation.....	1503
	Further Applications.....	1508
4	Numerical Methods and Case Examples.....	1514
	Dual and Primal-Dual Methods.....	1515
	Bregman Iteration.....	1521
	Graph Cut Methods.....	1524
	Quadratic Programming.....	1527
	Second-Order Cone Programming.....	1528
	Majorization-Minimization.....	1530
	Splitting Methods.....	1532
5	Conclusion.....	1534
	Cross-References.....	1534
	References.....	1534

Abstract

Since their introduction in a classic paper by Rudin, Osher, and Fatemi (Physica D 60:259–268, 1992), total variation minimizing models have become one of

R. Chan (✉)

Department of Mathematics, The Chinese University of Hong Kong, Shatin, Hong Kong
e-mail: rchan@math.cuhk.edu.hk

T.F. Chan

Office of the President, Hong Kong University of Science and Technology, Clear Water Bay,
Hong Kong
e-mail: tonyfchan@ust.hk

A. Yip

Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong
e-mail: mhyipa@hkbu.edu.hk

the most popular and successful methodologies for image restoration. New developments continue to expand the capability of the basic method in various aspects. Many faster numerical algorithms and more sophisticated applications have been proposed. This chapter reviews some of these recent developments.

1 Introduction

Images acquired through an imaging system are inevitably degraded in various ways. The types of degradation include noise corruption, blurring, missing values in the pixel domain or transformed domains, intensity saturation, jittering, etc. Such degradations can have adverse effects on high-level image processing tasks such as object detection and recognition. Image restoration aims at recovering the original image from its degraded version(s) to facilitate subsequent processing tasks. Image data differ from many other kinds of data due to the presence of edges, which are important features in human perception. It is therefore essential to preserve and even reconstruct edges in the processing of images. Variational methods for image restoration have been extensively studied in the past couple of decades. A promise of these methods is that the geometric regularity of the resulting images is explicitly controlled by using well-established descriptors in geometry. For example, smoothness of object boundaries can be easily manipulated by controlling their length. There has also been much research in designing variational methods for preserving other important image features such as textures.

Among the various restoration problems, denoising is perhaps the most fundamental one. Indeed, all algorithms for solving ill-posed restoration problems have to have some denoising capabilities either explicitly or implicitly, for otherwise they cannot cope with any error (noise) introduced during image acquisition or numerical computations. Moreover, the noise removal problem boils down to the fundamental problem of modeling natural images which has great impacts on any image processing tasks. Therefore, research on image denoising has been very active.

2 Background

Total variation (TV)-based image restoration models are introduced by Rudin, Osher, and Fatemi (ROF) in their seminal work [51] on edge-preserving image denoising. It is one of the earliest and best-known examples of variational partial differential equation (PDE)-based edge-preserving denoising models. In this model, the geometric regularity of the resulting image is explicitly imposed by reducing the amount of oscillation while allowing for discontinuities (edges). The unconstrained version introduced in [1] reads:

$$\inf_{u \in L^2(\Omega)} \int_{\Omega} |\nabla u| + \mu \int_{\Omega} (u - f)^2 d\mathbf{x}.$$

Here, Ω is the image domain, $f : \Omega \rightarrow \mathcal{R}$ is the observed noisy image, $u : \Omega \rightarrow \mathcal{R}$ is the denoised image, and $\mu \geq 0$ is a parameter depending on the noise level. The first term is the *total variation* (TV) which is a measure of the amount of oscillation in the resulting image u . Its minimization would reduce the amount of oscillation which presumably reduces noise. The second term is the L^2 distance between u and f , which encourages the denoised image to inherit most features from the observed data. Thus, the model trades off the closeness to f by gaining the regularity of u . The noise is assumed to be additive and Gaussian with zero mean. If the noise variance level σ^2 is known, then the parameter μ can be treated as the Lagrange multiplier, restraining the resulting image to be consistent with the known noise level, i.e., $\int_{\Omega} (u - f)^2 d\mathbf{x} = |\Omega|\sigma^2$ [16].

The ROF model is simple and elegant for edge-preserving denoising. Since its introduction, this model has ignited a great deal of research in constructing more sophisticated variants which can give better reconstructed images, designing faster numerical algorithms for solving the optimization problem numerically, and finding new applications in various domains. In a previous book chapter [21] published in 2005, the authors surveyed some recent progresses in the research of total variation-based models. The present chapter aims at highlighting some exciting latest developments in numerical methods and applications of total variation-based methods since the last survey.

3 Mathematical Modeling and Analysis

In this section, the basic definition of total variation and some of its variants are presented. Then, some recent TV-based mathematical models in imaging are reviewed.

Variants of Total Variation

Basic Definition

The use of TV as a regularizer has been shown to be very effective for processing images because of its ability to preserve edges. Being introduced for different reasons, several variants of TV can be found in the literature. Some variants can handle more sophisticated data such as vector-valued imagery and matrix-valued tensors; some are designed to improve restoration quality, and some are modified versions for the ease of numerical implementation. It is worthwhile to review the basic definition and its variants.

In Rudin, Osher, and Fatemi's work [51], the TV of an image $f : \Omega \rightarrow \mathcal{R}$ is defined as

$$\int_{\Omega} |\nabla f| d\mathbf{x},$$

where $\Omega \subseteq \mathbb{R}^2$ is a bounded open set. Since the image f may contain discontinuities, the gradient ∇f must be interpreted in a generalized sense. It is well known that elements of the Sobolev space $W^{1,1}(\Omega)$ cannot have discontinuities [2]. Therefore, the TV cannot be defined through the completion of the space C^1 of continuously differentiable functions under the Sobolev norm. The ∇f is thus interpreted as a distributional derivative, and its integral is interpreted as a distributional integral [40]. Under this framework, the minimization of TV naturally leads to a PDE with a distribution as a solution.

Besides defining TV as a distributional integral, other perspectives can offer some unique advantages. A set theoretical way is to define TV as a Radon measure of the domain Ω [50]. This has an advantage of allowing Ω to be a more general set. But a more practical and simple alternative is the “dual formulation.” It uses the usual trick in defining weak derivatives – integration by parts – together with the Fenchel transform,

$$\int_{\Omega} |\nabla f| = \sup \left\{ \int_{\Omega} f \operatorname{div} \mathbf{g} \, d\mathbf{x} \mid \mathbf{g} \in C_c^1(\Omega, \mathbb{R}^2), |\mathbf{g}(\mathbf{x})| \leq 1 \forall \mathbf{x} \in \Omega \right\} \quad (1)$$

where $f \in L^1(\Omega)$ and div is the divergence operator. Using this definition, one can bypass the discussion of distributions. It also plays an important role in many recent works in dual and primal-dual methods for solving TV minimization problems. The space BV can now be defined as

$$BV(\Omega) := \left\{ f \in L^1(\Omega) \mid \int_{\Omega} |\nabla f| < \infty \right\}.$$

Equipped with the norm $\|f\|_{BV} = \|f\|_{L^1} + \int_{\Omega} |\nabla f|$, this space is complete and is a proper superset of $W^{1,1}(\Omega)$ [32].

Multichannel TV

Many practical images are acquired in a multichannel way, where each channel emphasizes a specific kind of signal. For example, color images are often acquired through the RGB color components, whereas microscopy images consist of measurements of different fluorescent labels. The signals in the different channels are often correlated (contain redundant information). Therefore, in many practical situations, regularization of multichannel images should not be done independently on each channel.

There are several existing ways to generalize TV to vectorial data. A review of some generalizations can be found in [20]. Many generalizations are very intuitive. But only some of them have a natural dual formulation. Sapiro and Ringach [52] proposed to define

$$\int_{\Omega} |\nabla f| := \int_{\Omega} \sqrt{\sum_{i=1}^M |\nabla f_i|^2} \, d\mathbf{x} = \int_{\Omega} |\nabla f|_F \, d\mathbf{x},$$

where $f = (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_M(\mathbf{x}))$ is the vectorial data with M channels. Thus, it is the integral of the Frobenius norm $|\cdot|_F$ of the Jacobian ∇f . The dual formulation given in [10] is

$$\sup \left\{ \int_{\Omega} \langle f, \operatorname{div} \mathbf{g} \rangle d\mathbf{x} \mid \mathbf{g} \in C_c^1(\Omega, \mathbb{R}^{2 \times M}), |\mathbf{g}(\mathbf{x})|_F \leq 1 \forall \mathbf{x} \in \Omega \right\},$$

where $\langle f, \operatorname{div} \mathbf{g} \rangle = \sum_{i=1}^M f_i \operatorname{div} \mathbf{g}_i$.

Matrix-Valued TV

In applications such as diffusion tensor imaging (DTI), the measurements at each spatial location are represented by a diffusion tensor, which is a 3×3 symmetric positive semi-definite matrix. Recent efforts have been devoted to generalize the TV to matrix-valued images. Some natural generalizations can be obtained by identifying an $M \times N$ matrix with an MN vector, so that a vector-valued total variation can be applied. This was done by Tschumperlé and Deriche [57], who generalized the vectorial TV of [7]. The main challenge is to preserve the positive definiteness of the denoised solution. This will be elaborated in section “Diffusion Tensors Images.”

Another interesting approach proposed by Setzer et al. [54] is the so-called operator-based regularization. Given a matrix-valued function $f = (f_{ij}(\mathbf{x}))$, define a matrix function $A := (a_{ij})$ where $a_{ij} = |\nabla f_{ij}|^2$. Let $\Phi(A)$ be the matrix obtained by replacing each eigenvalue λ of A with $\sqrt{\lambda}$. Then the total variation is defined to be $\int_{\Omega} |\Phi(A)|_F d\mathbf{x}$, where $|\cdot|_F$ is the Frobenius norm. While this formulation seems complicated, its first variation turns out to have a nice simple formula. However, when combined with the ROF model, the preservation of positive definiteness is an issue.

Discrete TV

The ROF model is cast as an infinite-dimensional optimization problem over the BV space. To solve the problem numerically, one must discretize the problem at some stage. The approach proposed by Rudin et al. in [51] is to “optimize then discretize.” The gradient flow equation is discretized with a standard finite difference scheme. This method works very well in the sense that the numerical solution converges to a steady state which is qualitatively consistent with the expected result of the (continuous) ROF model. However, to the best of the authors’ knowledge, a theoretical proof of convergence of the numerical solution to the exact solution of the gradient flow equation as the grid size tends to zero is not yet available. A standard convergence result of finite difference schemes for nonlinear PDE is based on the compactness of TV-bounded sets in L^1 [46]. However, proving TV boundedness in two or more dimensions is often difficult.

An alternative approach is to “discretize then optimize.” In this case, only one has to solve a finite-dimensional optimization problem, whose numerical solution can in many cases be shown to converge. But the convergence of the exact solution

of the finite-dimensional problems to the exact solution of the original infinite-dimensional problem is often hard to obtain too. So, both approaches suffer from the theoretical convergence problem. But the latter method has a precise discrete objective to optimize.

To discretize the ROF objective, the fitting term is often straightforward. But the discretization of the TV term has a strong effect on the numerical schemes. The most commonly used versions of discrete TV are

$$\|f\|_{TV} = \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} \sqrt{(f_{i+1,j} - f_{i,j})^2 + (f_{i,j+1} - f_{i,j})^2} \Delta x, \tag{2}$$

$$\|f\|_{TV} = \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} (|f_{i+1,j} - f_{i,j}| + |f_{i,j+1} - f_{i,j}|) \Delta x, \tag{3}$$

where $f = (f_{i,j})$ is the discrete image and Δx is the grid size. They are sometimes referred as the *isotropic* and *anisotropic* versions, respectively, for they are a formal discretization of the isotropic TV $\int_{\Omega} \sqrt{f_x^2 + f_y^2} d\mathbf{x}$ and the anisotropic TV $\int_{\Omega} (|f_x| + |f_y|) d\mathbf{x}$, respectively. The anisotropic TV is not rotational invariant; an image and its rotation can have a different TV value. Therefore, the discrete TV (3) deviates from the original isotropic TV. But being a piecewise linear function, some numerical techniques for quadratic and linear problems can be applied. Indeed, by introducing some auxiliary variables, the corresponding discrete ROF objective can be converted into a canonical quadratic programming problem [30].

Besides using finite difference approximations, a recent popular way is to represent TV on graphs [27]. To make the problem fully discrete, the range of the image is quantized to a finite set of K integers only, usually 0–255. The image is “leveled,” so that $f_{i,j}^k = 1$ if the intensity of the (i, j) th pixel is at most k , and $f_{i,j}^k = 0$ otherwise. Then, the TV is given by

$$\|f\|_{TV} = \sum_{k=0}^{K-1} \sum_{i,j} \sum_{s,t} w_{i,j,s,t} |f_{i,j}^k - f_{s,t}^k|, \tag{4}$$

where $w_{i,j,s,t}$ is a nonnegative weight. A simple choice is the four-connectivity model, where $w_{i,j,s,t} = 1$ if $|i - s| + |j - t| \leq 1$ and $w_{i,j,s,t} = 0$ otherwise. In this case, it becomes the anisotropic TV (3). Different choices of the weights penalize edges in different orientations.

A related concept introduced by Shen and Kang is the *quantum total variation* [55]. They studied the ROF model when the range of an image is a finite discrete set (preassigned or determined on the fly), but the image domain is a continuous one. The model is suitable for problems such as bar code scanning, image quantization, and image segmentation. An elegant analysis of the model and some stochastic gradient descent algorithms were presented there.

Nonlocal TV

First proposed by Buades et al. [11], the nonlocal means algorithm renounces the use of local smoothness to denoise an image. Patches which are spatially far away but photometrically similar are also utilized in the estimation process – a paradigm which has been used in texture synthesis [28]. The denoising results are surprisingly good. Since then, the use of nonlocal information becomes increasingly popular. In particular, Bresson and Chan [10] and Gilboa and Osher [31] considered the nonlocal TV. The nonlocal gradient $\nabla_{NL} f$ for a pair of points $\mathbf{x} \in \Omega$ and $\mathbf{y} \in \Omega$ is defined by

$$\nabla_{NL} f(\mathbf{x}, \mathbf{y}) = \sqrt{w(\mathbf{x}, \mathbf{y})} (f(\mathbf{x}) - f(\mathbf{y})),$$

where $w(\mathbf{x}, \mathbf{y})$ is a nonnegative weight function which is presumably a similarity measure between a patch around \mathbf{x} and a patch around \mathbf{y} . As an illustration, a simple choice of the weight function is

$$w(\mathbf{x}, \mathbf{y}) = \alpha_1 e^{-|\mathbf{x}-\mathbf{y}|^2/\sigma_1^2} + \alpha_2 e^{-|F(\mathbf{x})-F(\mathbf{y})|^2/\sigma_2^2},$$

where α_i and σ_i are positive constants, and $F(\mathbf{x})$ is a feature vector derived from a patch around \mathbf{x} . The constants α_i may sometimes be defined to depend on \mathbf{x} , so that the total weight over all $\mathbf{y} \in \Omega$ is normalized to 1. In this case, the weight function is nonsymmetric with respect to its arguments. The first term in w is a measure of geometric similarity, so that nearby pixels have a higher weight. The second term is a measure of photometric similarity. The feature vector F can be the color histogram or any texture descriptor over a window around \mathbf{x} . The norm of the nonlocal gradient at \mathbf{x} is defined by

$$|\nabla_{NL} f(\mathbf{x})| = \sqrt{\int_{\Omega} [\nabla_{NL} f(\mathbf{x}, \mathbf{y})]^2 d\mathbf{y}},$$

which adds up all the squared intensity variation relative to $f(\mathbf{x})$, weighted by the similarity between the corresponding pair of patches. The nonlocal TV is then naturally defined by summing up all the norms of the nonlocal gradients over the image domain:

$$\int_{\Omega} |\nabla_{NL} f| d\mathbf{x}.$$

Therefore, the nonlocal TV is small if, for each pair of similar patches, the intensity difference between their centers is small. An advantage of using the nonlocal TV to regularize images is its tendency to preserve highly repetitive patterns better. In practice, the weight function is often truncated to reduce the computation costs spent in handling the many less similar patches.

Further Applications

Inpainting in Transformed Domains

After the release of the image compression standard JPEG2000, images can be formatted and stored in terms of wavelet coefficients. For instance, in Acrobat 6.0 or later, users can opt to use JPEG2000 to compress embedded images in a PDF file. During the process of storing or transmission, some wavelet coefficients may be lost or corrupted. This prompts the need of restoring missing information in wavelet domains. The setup of the problem is as follows. Denote the standard orthogonal wavelet expansion of the images f and u by

$$f(\alpha) = \sum_{j,k} \alpha_{j,k} \psi_{j,k}(x), \quad j \in \mathbb{Z}, k \in \mathbb{Z}^2,$$

and

$$u(\beta) = \sum_{j,k} \beta_{j,k} \psi_{j,k}(x), \quad j \in \mathbb{Z}, k \in \mathbb{Z}^2,$$

where $\{\Psi_{j,k}\}$ is the wavelet basis, and $\{\alpha_{j,k}\}, \{\beta_{j,k}\}$ are the wavelet coefficients of f and u given by

$$\alpha_{j,k} = \langle f, \psi_{j,k} \rangle \quad \text{and} \quad \beta_{j,k} = \langle u, \psi_{j,k} \rangle,$$

respectively, for $j \in \mathbb{Z}, k \in \mathbb{Z}^2$. For convenience, $u(\beta)$ is denoted by u when there is no ambiguity. Assume that the wavelet coefficients in the index set I are known, i.e., the available wavelet coefficients are given by

$$\xi_{j,k} = \begin{cases} \alpha_{j,k}, & (j, k) \in I, \\ 0, & (j, k) \in \Omega \setminus I. \end{cases}$$

The aim of wavelet domain inpainting is to reconstruct the wavelet coefficients of u from the given coefficients ξ . It is well known that the inpainting problem is ill posed, i.e., it admits more than one solution. There are many different ways to fill in the missing coefficients, and therefore many different reconstructions in the pixel domain are possible. Regularization methods can be used to incorporate prior information about the reconstruction. In [23], Chan, Shen, and Zhou used TV to solve the wavelet inpainting problem, so that the missing coefficients are filled while preserving sharp edges in the pixel domain faithfully. More precisely, they considered the minimization of the following objective

$$F(\beta) = \frac{1}{2} \sum_{j,k} \chi_{j,k} (\xi_{j,k} - \beta_{j,k})^2 + \lambda \|u\|_{TV}, \tag{5}$$

with $\chi_{j,k} = 1$ if $(j, k) \in I$ and $\chi_{j,k} = 0$ if $(j, k) \in \Omega \setminus I$, and λ is the regularization parameter. The first term in F is the data-fitting term, and the second is the TV regularization term. The method Chan, Shen, and Zhou used to optimize the objective is the standard gradient descent. The method is very robust, but it often slows down significantly before it converges.

In [18], Chan, Wen, and Yip proposed an efficient *optimization transfer algorithm* to minimize the objective (5). An auxiliary variable ζ is introduced to yield a new objective function:

$$G(\zeta, \beta) = \frac{1 + \tau}{2\tau} \left(\|\chi(\zeta - \xi)\|_2^2 + \tau \|\zeta - \beta\|_2^2 \right) + \lambda \|u(\beta)\|_{TV},$$

where χ denotes a diagonal matrix with diagonal entries $\chi_{j,k}$, and τ is an arbitrary positive parameter. The function G is a quadratic majorizing function [43] of F . The method also has a flavor of the splitting methods introduced in section “Splitting Methods.” But a major difference is that the method here solves the original problem (5) without any alteration. It can be easily shown that

$$F(\beta) = \min_{\zeta} G(\zeta, \beta)$$

for any positive regularization parameter τ . Thus, the minimization of G w.r.t. (ζ, β) is equivalent to the minimization of F w.r.t. β for any $\tau > 0$. Unlike the gradient descent method of [23], the optimization transfer algorithm avoids the use of derivatives of the TV. It also does not require smoothing out the TV to make it differentiable. The experimental results in [18] showed that the algorithm is very efficient and outperforms the gradient descent method.

Superresolution

Image superresolution refers to the process of increasing spatial resolution by fusing information from a sequence of low-resolution images of the same scene. The images are assumed to contain subpixel information (due to subpixel displacements or blurring), so that the superresolution is possible.

In [24], Chan et al. proposed a unified TV model for superresolution imaging problems. They focused on the problem of reconstructing a high-resolution image from several decimated, blurred, and noisy low-resolution versions of the high-resolution image. They derived a low-resolution image formation model which allows multiple-shifted and blurred low-resolution image frames, so that it subsumes several well-known models. The model also allows an arbitrary pattern of missing pixels (in particular an arbitrary pattern of missing frames). The superresolution image reconstruction problem is formulated as an optimization problem which combines the image formation model and the TV inpainting model. In this method, TV minimization is used to suppress noise amplification, repair corrupted pixels in regions without missing pixels, and reconstruct intensity levels in regions with missing pixels.

Image Formation Model The observation model, Chan et al. considered, consists of various degradation processes. Assume that a number of $m \times n$ low-resolution frames are captured by an array of charge-coupled device (CCD) sensors. The goal is to reconstruct an $Lm \times Ln$ high-resolution image. Thus, the resolution is increased by a factor of L in each dimension. Let u be the ideal $Lm \times Ln$ high-resolution clean image.

1. *Formation of low-resolution frames.* A low-resolution frame is given by

$$D_{p,q}Cu,$$

where C is an averaging filter with window size L -by- L , and $D_{p,q}$ is the downsampling matrix which, starting at the (p, q) th pixel, samples every other L pixels in both dimensions to form an $m \times n$ image.

2. *Blurring of frames.* This is modeled by

$$H_{p,q}D_{p,q}Cu,$$

where $H_{p,q}$ is the blurring matrix for the (p, q) th frame.

3. *Concatenation of frames.* The full set of L^2 frames are interlaced to form an $mL \times nL$ image:

$$Au,$$

where

$$A = \sum_{p,q} D_{p,q}^T H_{p,q} D_{p,q} C.$$

4. *Additive Noise.*

$$Au + \eta,$$

where each pixel in η is a Gaussian white noise.

5. *Missing pixels and missing frames.*

$$f = \Lambda_{\mathcal{D}}(Au + \eta),$$

where \mathcal{D} denotes the set of missing pixels, and $\Lambda_{\mathcal{D}}$ is the downsampling matrix from the image domain to \mathcal{D} .

6. *Multiple observations.* Finally, multiple observations of the same scene, but with different noise and blurring, are allowed. This leads to the model

$$f_r = \Lambda_{\mathcal{D}_r}(A_r u + \eta_r) \quad r = 1, \dots, R, \quad (6)$$

where

$$A_r = \sum_{p,q} D_{p,q}^T H_{p,q,r} D_{p,q} C.$$

TV Superresolution Imaging Model To invert the degradation processes in (6), a Tikhonov-type regularization model has been used. It requires minimization of the following energy:

$$F(u) = \frac{1}{2} \sum_{r=1}^R \|\Lambda_{\mathcal{D}_r} A_r u - f_r\|^2 + \lambda \|u\|_{TV}.$$

This model simultaneously performs denoising, deblurring, inpainting, and super-resolution reconstruction. Experimental results show that reasonably good reconstruction can be obtained even if five-sixth of the pixels are missing and the frames are blurred.

Image Segmentation

TV minimization problems also arise from image segmentation. When one seeks for a partition of the image into homogeneous segments, it is often helpful to regularize the shape of the segments. This can increase the robustness of the algorithm against noise and avoid spurious segments. It may also allow the selection of features of different scales. In the classical Mumford-Shah model [47], the regularization is done by minimizing the total length of the boundary of the segments. In this case, if one represents a segment by its characteristic function, then the length of its boundary is exactly the TV of the characteristic function. Therefore, the minimization of length becomes the minimization of TV of characteristic functions.

Given an observed image f on an image domain Ω , the piecewise constant Mumford-Shah model seeks a set of curves C and a set of constant $\mathbf{c} = (c_1, c_2, \dots, c_L)$ which minimize the energy functional given by

$$F^{MS}(C, \mathbf{c}) = \sum_{l=1}^L \int_{\Omega_l} [f(\mathbf{x}) - c_l]^2 d\mathbf{x} + \beta \cdot \text{Length}(C).$$

The curves in C partition the image into L mutually exclusive segments Ω_l for $l = 1, 2, \dots, L$. The idea is to partition the image, so that the intensity of f in each segment Ω_l is well approximated by a constant c_l . The goodness of fit is measured by the L^2 difference between f and c_l . On the other hand, a minimum description length principle is employed which requires the curves C to be as short as possible. This increases the robustness to noise and avoids spurious segments. The parameter $\beta > 0$ controls the trade-off between the goodness of fit and the length of the curves C . The Mumford-Shah objective is nontrivial to optimize especially when the curves need to be split and merged. Chan and Vese [24] proposed a level set-based method which can handle topological changes effectively. In the two-phase

version of this method, the curves are represented by the zero level set of a Lipschitz level set function Φ defined on the image domain. The objective function then becomes

$$F^{CV}(\phi, c_1, c_2) = \int_{\Omega} H(\phi(\mathbf{x})) [f(\mathbf{x}) - c_1]^2 d\mathbf{x} + \int_{\Omega} [1 - H(\phi(\mathbf{x}))] [f(\mathbf{x}) - c_2]^2 d\mathbf{x} + \beta \int_{\Omega} |\nabla H(\phi)|.$$

The function H is the Heaviside function defined by $H(x) = 1$ if $x \geq 0$, $H(x) = 0$ otherwise. In practice, we replace H by a smooth approximation H_{ϵ} , e.g.,

$$H_{\epsilon}(x) = \frac{1}{2} \left[1 + \frac{2}{\pi} \arctan\left(\frac{x}{\epsilon}\right) \right].$$

Although this method makes splitting and merging of curves a simple matter, the energy functional is non-convex which possesses many local minima. These local minima may correspond to undesirable segmentations; see [45].

Interestingly, for fixed c_1 and c_2 , the above non-convex objective can be reformulated as a convex problem, so that a global minimum can be easily computed; see [22, 56]. The globalized objective is given by

$$F^{CEN}(u, c_1, c_2) = \int_{\Omega} \left\{ [f(\mathbf{x}) - c_1]^2 - [f(\mathbf{x}) - c_2]^2 \right\} u(\mathbf{x}) d\mathbf{x} + \beta \int_{\Omega} |\nabla u| \tag{7}$$

which is minimized over all u satisfying the bilateral constraints $0 \leq u \leq 1$ and all scalars c_1 and c_2 . After a solution u is obtained, a global solution to the original two-phase Mumford-Shah objective can be obtained by thresholding u with μ for almost every $\mu \in [0, 1]$, see [22, 56]. Some other proposals for computing global solutions can be found in [45].

To optimize the globalized objective function (7), Chan et al. [22] proposed to use an exact penalty method to convert the bilaterally constrained problem to an unconstrained problem. Then the gradient descent method is applied. This method is very robust and easy to implement. Moreover, the exact penalty method treats the constraints gracefully, as if there is no constraint at all. But of course the gradient descent is not particular fast.

In [42], Krishnan et al. considered the following discrete two-phase Mumford-Shah model:

$$F^{CEN}(u, c_1, c_2) = \langle s, u \rangle + \beta \|u\|_{TV} + \frac{\alpha}{2} \left\| u - \frac{1}{2} \right\|^2,$$

where $\langle \cdot, \cdot \rangle$ is the l^2 inner product, $s = (s_{i,j})$, and

$$s_{i,j} = (f_{i,j} - c_1)^2 - (f_{i,j} - c_2)^2.$$

The variable u is bounded by the bilateral constraints $0 \leq u \leq 1$. When $\alpha = 0$, this problem is convex but not strictly convex. When $\alpha > 0$, this problem is strictly convex. The additive constant $\frac{1}{2}$ is introduced in the third term so that the minimizer does not bias toward $u = 0$ or $u = 1$. This problem is exactly a TV denoising problem with bound constraints. Krishnan et al. proposed to use the primal-dual active-set method to solve the problem. Superlinear convergence has been established.

Diffusion Tensors Images

Recently, diffusion tensor imaging (DTI), a kind of magnetic resonance (MR) modality, becomes increasingly popular. It enables the study of anatomical structures such as nerve fibers in human brains noninvasively. Moreover, the use of direction-sensitive acquisitions results in its lower signal-to-noise ratio compared to convectional MR. At each voxel in the imaging domain, the anisotropy of diffusion water molecules is interested. Such an anisotropy can be described by a diffusion tensor D , which is a 3×3 positive semi-definite matrix. By standard spectral theory results, D can be factorized into

$$D = V\Lambda V^T,$$

where V is an orthogonal matrix whose columns are the eigenvectors of D , and Λ is a diagonal matrix whose diagonal entries are the corresponding eigenvalues. These eigenvalues provide the diffusion rate along the three orthogonal directions defined by the eigenvectors. The goal is to estimate the matrix D (one at each voxel) from the data. Under the Stejskal-Tanner model, the measurement S_k from the imaging device and the diffusion tensor are related by

$$S_k = S_0 e^{-b g_k^T D g_k}, \tag{8}$$

where S_0 is the baseline measurement, g_k is the prescribed direction in which the measurement is done, and $b > 0$ is a scalar depending the strength of the magnetic field applied and the acquisition time. Since D has six degrees of freedom, six measurements at different orientations are needed to reconstruct D . In practice, the measurements are very noisy. Thus, matrix D obtained by directly solving (8) for $k = 1, 2, \dots, 6$ may not be positive semi-definite and is error-prone. It is thus often helpful to take more than six measurements and to use some least squares methods or regularization to obtain a robust estimate while preserving the positive semi-definiteness for physical correctness.

In [60] Wang et al. and in [25] Christiansen et al. proposed an extension of the ROF to denoise tensor-valued data. Two major differences between the two works are that the former regularizes the Cholesky factor of D and uses a channel-by-channel TV regularization, whereas the latter regularizes the tensor D directly and uses a multichannel TV.

The method in [25] is two staged. The first stage is to estimate the diffusion tensors from the raw data based on the Stejskal-Tanner model (8). The obtained

tensors are often noisy and may not be positive semi-definite. The next stage is to use the ROF model to denoise the tensor while restricting the results to be positive semi-definite. The trick they used to ensure positive semi-definiteness is very simple and practical. They observed that a symmetric matrix is positive semi-definite if and only if it has a Cholesky factorization of the form

$$D = LL^T,$$

where L is a lower triangular matrix

$$L = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix}.$$

Then one can easily express D in terms of l_{ij} for $1 \leq j \leq i \leq 3$:

$$D = D(L) = \begin{bmatrix} l_{11}^2 & l_{11}l_{21} & l_{11}l_{31} \\ l_{11}l_{21} & l_{21}^2 + l_{22}^2 & l_{21}l_{31} + l_{22}l_{32} \\ l_{11}l_{31} & l_{21}l_{31} + l_{22}l_{32} & l_{31}^2 + l_{32}^2 + l_{33}^2 \end{bmatrix}.$$

The ROF problem, written in a continuous domain, is then formulated as

$$\min_L \left\{ \frac{1}{2} \sum_{ij} \int_{\Omega} [d_{ij}(L) - \hat{d}_{ij}]^2 d\mathbf{x} + \lambda \sqrt{\sum_{ij} \left[\int_{\Omega} |\nabla d_{ij}(L)| \right]^2} \right\},$$

where $\hat{D} = (\hat{d}_{ij})$ is the observed noisy tensor field, and L is the unknown lower triangular matrix-valued function from Ω to $\mathcal{R}^{3 \times 3}$. Here, the matrix-valued version of TV is used. The objective is then differentiated w.r.t. the lower triangular part of L to obtain a system of six first-order optimality conditions. Once the optimal L is obtained, the tensor D can be formed by taking $D = LL^T$ which is a positive semi-definite.

The original ROF problem is strictly convex so that one can obtain the globally optimal solution. However, in this problem, due to the nonlinear change of variables from D to L , the problem becomes non-convex. But the authors of [25] reported that in their experiments, different initial data often resulted in the same solution, so that the non-convexity does not pose any significant difficulty to the optimization of the objective.

4 Numerical Methods and Case Examples

Fast numerical methods for TV minimization continue to be an active research area. Researchers from different fields have been bringing many fresh ideas to the

problem and led to many exciting results. Some categories of particular mention are dual/primal-dual methods, Bregman iterative methods, and graph cut methods. Many of these methods have a long history with a great deal of general theories developed. But when it comes to their application to the ROF model, many further properties and specialized refinements can be exploited to obtain even faster methods. Having said so, different algorithms may adopt different versions of TV. They have different properties and thus may be used for different purposes. Thus, some caution needs to be taken when one attempts to draw conclusions such as method A is faster than method B. Moreover, different methods have different degree of generality. Some methods can be extended directly to deblurring, while some can only be applied to denoising. (Of course, one can use an outer iteration to solve a deblurring problem by a sequence of denoising problems, so that any denoising algorithm can be used. But the convergence of the outer iteration has little, if not none, to do with the inner denoising algorithm.) This section surveys some recent methods for TV denoising and/or deblurring. The model considered here is a generalized ROF model which simultaneously performs denoising and deblurring. The objective function reads

$$F(u) = \frac{1}{2} \int_{\Omega} (Ku - f)^2 d\mathbf{x} + \lambda \int_{\Omega} |\nabla u|, \quad (9)$$

where K is a blurring operator and $\lambda > 0$ is the regularization parameter. For simplicity, we assume that K is invertible. When K is the identity operator, (9) is the ROF denoising model.

Dual and Primal-Dual Methods

The ROF objective is non-differentiable in flat regions where $|\nabla u| = 0$. This leads to much difficulty in the optimization process since gradient information (hence, Taylor's expansion) becomes unreliable in predicting the function value even locally. Indeed, the staircase effects of TV minimization can introduce some flat regions which make the problem worse. Even if the standard procedure of replacing the TV with a reasonably smoothed version is used so that the objective becomes differentiable, the Euler-Lagrange equation for (9) is still very stiff to solve. Higher-order methods such as Newton's methods often fail to work because higher-order derivatives are even less reliable.

Due to the difficulty in optimizing the ROF objective directly, much recent research has been directed toward solving some reformulated versions. In particular, methods based on dual and primal-dual formulations have been shown to be very fast in practice. Actually, the dual problem (see (12) below) also has its own numerical difficulties to face, e.g., the objective is rank deficient and some extra work is needed to deal with the constraints. But the dual formulation brings many well-developed ideas and techniques from numerical optimization to bear on this problem. Primal-dual methods have also been studied to combine information from

the primal and dual solutions. Several successful dual and primal-dual methods are reviewed.

Chan-Golub-Mulet’s Primal-Dual Method

Some early work in dual and primal-dual methods for the ROF model can be found in [13, 20]. In particular, Chan, Golub, and Mulet (CGM) [20] introduced a primal-dual system involving a primal variable u and a Fenchel dual variable \mathbf{p} . It remains one of the most efficient methods today and is perhaps the most intuitive one. It is worthwhile to review it and see how it relates to the more recent methods. Their idea is to start with the Euler-Lagrange equation of (9):

$$K^T K u - K^T f - \lambda \operatorname{div} \left(\frac{\nabla u}{\sqrt{|\nabla u|^2 + \varepsilon}} \right) = 0. \tag{10}$$

Owing to the singularity of the third term, they introduced an auxiliary variable

$$\mathbf{p} = \frac{\nabla u}{\sqrt{|\nabla u|^2 + \varepsilon}}$$

to form the system

$$\begin{aligned} \mathbf{p} \sqrt{|\nabla u|^2 + \varepsilon} &= \nabla u \\ K^T K u - K^T f - \lambda \operatorname{div} \mathbf{p} &= 0. \end{aligned}$$

Thus, the blowup singularity is canceled. They proposed to solve this system by Newton’s method which is well known to converge quadratically locally if the Jacobian of the system is Lipschitz. Global convergence is observed when coupled with a simple Armijo line search [8]. The variable \mathbf{p} is indeed the same as the Fenchel dual variable \mathbf{g} in (1) when $\nabla u \neq 0$ and $\varepsilon = 0$. Thus, \mathbf{p} is a smoothed version of the dual variable \mathbf{g} . Without the introduction of the dual variable, a direct application of the Newton’s method to the Euler-Lagrange equation (10) often fails to converge because of the small domain of convergence.

Chambolle’s Dual Method

A pure dual method is proposed by Chambolle in [14], where the ROF objective is written solely in terms of the dual variable. By the definition of TV in (1), it can be deduced using duality theory that

$$\begin{aligned} \inf_u \left\{ \frac{1}{2} \int_{\Omega} (K u - f)^2 d\mathbf{x} + \lambda \int_{\Omega} |\nabla u| \right\} \\ \iff \inf_u \sup_{|\mathbf{p}| \leq 1} \left\{ \frac{1}{2} \int_{\Omega} (K u - f)^2 d\mathbf{x} + \lambda \int_{\Omega} u \operatorname{div} \mathbf{p} d\mathbf{x} \right\} \end{aligned} \tag{11}$$

$$\begin{aligned} &\iff \sup_{|\mathbf{p}| \leq 1} \inf_u \left\{ \frac{1}{2} \int_{\Omega} (Ku - f)^2 d\mathbf{x} + \lambda \int_{\Omega} u \operatorname{div} \mathbf{p} d\mathbf{x} \right\} \\ &\iff \sup_{|\mathbf{p}| \leq 1} \left\{ -\frac{\lambda^2}{2} \int_{\Omega} \left| K^{-T} \operatorname{div} \mathbf{p} - \frac{f}{\lambda} \right|^2 d\mathbf{x} \right\}. \end{aligned} \tag{12}$$

The resulting problem has a quadratic objective with quadratic constraints. In contrast, the primal objective is only piecewise smooth which is badly behaved when $\nabla u = 0$. Thus the dual objective function is very simple, but additional efforts are needed to handle the constraints. One can write down the Karush-Kuhn-Tucker (KKT) optimality system [8] of the discretized objective, which amounts to solving a nonlinear system of equations involving complementarity conditions and inequality constraints on the Lagrange multipliers. Interestingly, the Lagrange multipliers have a closed-form solution which greatly simplifies the problem. More precisely, the KKT system consists of the equations

$$\mu p = H(\mathbf{p}) \tag{13}$$

$$\mu (|\mathbf{p}|^2 - 1) = 0 \tag{14}$$

$$\mu \geq 0 \tag{15}$$

$$|\mathbf{p}|^2 \leq 1, \tag{16}$$

where μ is the nonnegative Lagrange multiplier and

$$H(\mathbf{p}) := \nabla \left[(K^T K)^{-1} \operatorname{div} \mathbf{p} - \frac{1}{\lambda} K^{-1} f \right].$$

Since

$$\mu |\mathbf{p}| = |H(\mathbf{p})|,$$

if $|\mathbf{p}| = 1$, then $\mu = |H(\mathbf{p})|$; if $|\mathbf{p}| < 1$, then the complementarity (14) implies $\mu = 0$ and by (13) $H(\mathbf{p}) = 0$, so that μ is also equal to $|H(\mathbf{p})|$. This simplifies the KKT system into a nonlinear system of \mathbf{p} only

$$|H(\mathbf{p})| \mathbf{p} = H(\mathbf{p}).$$

Chambolle proposes a simple semi-implicit scheme to solve the system:

$$\mathbf{p}^{n+1} = \frac{\mathbf{p}^n + \tau H(\mathbf{p}^n)}{\mathbf{p}^n + \tau |H(\mathbf{p}^n)|}.$$

Here, τ is a positive parameter controlling the stepsize. The method is proven to be convergent for any

$$\tau \leq \frac{1}{8} \left\| (K^T K)^{-1} \right\|, \tag{17}$$

where $\|\cdot\|$ is the spectral norm. This method is also faithful to the original ROF problem; it does not require approximating the TV by smoothing.

The convergence rate of this method is at most linear, but for denoising problems, it usually converges fast (measured by the relative residual norm of the optimality condition) in the beginning but stagnates after some iterations (at a level several orders of magnitude higher than the machine epsilon). This is very typical for simple relaxation methods. Fortunately, visually good results (measured by the number of pixels having a gray level different from the optimal one after they are quantized to their 8-bit representation) are often achieved before the method stagnates [64]. However, when applied to deblurring, K is usually ill conditioned, so that the stepsize restriction (17) is too stringent. In this case, another outer iteration is often used in conjunction with the method; see the splitting methods in section “Splitting Methods.”

Chambolle’s method has been successfully adapted to solve a variety of related image processing problems, e.g., the ROF with nonlocal TV [9], multichannel TV [10], and segmentation problems [4]. We remark that many other approaches for solving (12) have been proposed. A discussion of some first-order methods including projected gradient methods and Nesterov methods can be found in [3, 26, 61].

Primal-Dual Hybrid Gradient Method

As mentioned in the beginning of Sect. 4, the primal and dual problems have their own advantages and numerical difficulties to face. It is therefore tempting to combine the best of both. In [64], Zhu and Chan proposed the *primal-dual hybrid gradient* (PDHG) algorithm which alternates between primal and dual formulations.

The method is based on the primal-dual formulation

$$G(u, \mathbf{p}) := \frac{1}{2} \int_{\Omega} (Ku - f)^2 d\mathbf{x} + \lambda \int_{\Omega} u \operatorname{div} \mathbf{p} d\mathbf{x} \rightarrow \inf_u \sup_{|\mathbf{p}| \leq 1};$$

cf. formulation (11). By fixing its two variables one at a time, this saddle point formulation has two subproblems:

$$\sup_{|\mathbf{p}| \leq 1} G(u, \mathbf{p}) \quad \text{and} \quad \inf_u G(u, \mathbf{p}).$$

While one may obtain an optimal solution by solving the two subproblems to a high accuracy alternatively, the PDHG method applies only one step of gradient descent/ascent to each of the two subproblems alternatively. The rationale is that when neither of the two variables are optimal, there is little to gain by iterating each subproblem until convergence. Starting with an initial guess u^0 , the following two steps are repeated:

$$\begin{aligned} \mathbf{p}^{k+1} &= P_{|\mathbf{p}| \leq 1} (\mathbf{p}^k - \tau_k \nabla u^k) \\ u^{k+1} &= u^k - \theta_k (K^T (K u^k - f) + \lambda \operatorname{div} \mathbf{p}^{k+1}) \end{aligned}$$

Here, $P_{|\mathbf{p}| \leq 1}$ is the projector onto the feasible set $\{\mathbf{p} : |\mathbf{p}| \leq 1\}$. The stepsizes τ_k and θ_k can be chosen to optimize the performance. Some stepping strategies were presented in [64]. In [65], Zhu, Wright, and Chan studied a variety of stepping strategies for a related dual method.

Numerical results in [64] show that this simple algorithm is faster than the split Bregman iteration (see section “Split Bregman Iteration”), which is faster than Chambolle’s semi-implicit dual method (see section “Chambolle’s Dual Method”). Some interesting connections between the PDHG algorithm and other algorithms such as *proximal forward-backward splitting*, *alternating minimization*, *alternating direction method of multipliers*, *Douglas-Rachford splitting*, *split inexact Uzawa*, and *averaged gradient* methods applied to different formulations of the ROF model are studied by Esser et al. in [29]. Such connections reveal some convergence theory of the PDHG algorithm in several important cases (special choices of the stepsizes) in a more general setting.

Semi-smooth Newton’s Method

Given the dual problem, it is natural to consider other methods to solve its optimality conditions (13)–(16). A standard technique in optimization to handle complementarity and Lagrange multipliers is to combine them into a single equality constraint. Observe that the constraints $a \geq 0$, $b \geq 0$ and $ab = 0$ can be consolidated into the equality constraint

$$\phi(a, b) := \sqrt{a^2 + b^2} - a - b = 0, \tag{18}$$

where ϕ is known as the Fischer-Burmeister function. Therefore, the KKT system (13)–(16) can be written as

$$\begin{aligned} \mu \mathbf{p} &= H(\mathbf{p}) \\ \phi(\mu, 1 - |\mathbf{p}|^2) &= 0. \end{aligned}$$

Ng et al. [48] observed that this system is semi-smooth and therefore proposed solving this system using a *semi-smooth Newton’s method*. In this method, if the Jacobian of the system is not defined in the classical sense due to the system’s lack of enough smoothness, then the Jacobian is replaced by a generalized Jacobian evaluated at a nearby point. It is proven that this method converges superlinearly if the system to solve is at least semi-smooth and if the generalized Jacobians at convergence satisfy some invertibility conditions. For the dual problem (12), the Newton’s equation may be singular. This problem is fixed by regularizing the Jacobian.

Primal-Dual Active-Set Method

Hintermüller and Kunisch [37] considered the Fenchel dual approach to formulate a constrained quadratic dual problem and derived a very effective active-set method to handle the constraints. The method separates the variables into active and inactive

sets, so that they can be treated differently accordingly to their characteristics. They considered the case of anisotropic discrete TV norm (3), so that the dual variable is bilaterally constrained, i.e., $-1 \leq \mathbf{p} \leq 1$, whereas the constraints in (12) are quadratic. In this setting, superlinear convergence can be established.

To deal with the bilateral constraints on \mathbf{p} , they proposed to use the *primal-dual active-set* (PDAS) algorithm. Consider the general quadratic problem,

$$\min_{y, y \leq \psi} \frac{1}{2} \langle y, Ay \rangle - \langle f, y \rangle,$$

where ψ is a given vector in \mathcal{R}^n . This problem includes (12) as a special instance. The KKT conditions are given by

$$\begin{aligned} Ay + v &= f, \\ v \odot (\psi - y) &= 0, \\ v &\geq 0, \\ \psi - y &\geq 0, \end{aligned}$$

where v is a vector of Lagrange multipliers and \odot denotes the entrywise product. The idea of the PDAS algorithm is to predict the active variables \mathcal{A} and inactive variables \mathcal{I} to speed up the determination of the final active and inactive variables. The prediction is done by comparing the closeness of v and $\psi - y$ to zero. If $\psi - y$ is c times closer to zero than v does, then the variable is predicted as active. The PDAS algorithm is given by

1. Initialize y^0, v^0 . Set $k = 0$.
2. Set $\mathcal{I}^k = \{i : v_i^k - c(\psi - y^k)_i \leq 0\}$ and $\mathcal{A}^k = \{i : v_i^k - c(\psi - y^k)_i > 0\}$.
3. Solve

$$\begin{aligned} Ay^{k+1} + v^{k+1} &= f, \\ y^{k+1} &= \psi \quad \text{on } \mathcal{A}^k, \\ v^{k+1} &= 0 \quad \text{on } \mathcal{I}^k. \end{aligned}$$

4. Stop or set $k = k + 1$ and return to Step 2.
 Notice that the constraints $a \geq 0, b \geq 0$, and $ab = 0$ can be combined as a single equality constraint:

$$\min(a, cb) = 0$$

for any positive constant c . Thus, the KKT system can be written as

$$\begin{aligned} Ay + v &= f, \\ C(y, v) &= 0, \end{aligned}$$

where $C(y, v) = \min(v, c(\psi - y))$ for an arbitrary positive constant c . The function C is piecewise linear, whereas the Fisher-Burmeister formulation (18)

is nonlinear. More importantly, applying Newton’s method (using a generalized derivative) to such a KKT system yields exactly the PDAS algorithm. This allows Hintermüller et al. to explain the local superlinear convergence of the PDAS algorithm for a class of optimization problems that include the dual of the anisotropic TV deblurring problem [36]. In [37], some conditional global convergence results based on the properties of the blurring matrix K have also been derived. Their formulation is based on the anisotropic TV norm, and the dual problem requires an extra l^2 regularization term when a deblurring problem is solved.

The dual problem (12) is rank deficient and does not have a unique solution in general. In [37], Hintermüller and Kunisch proposed to add a regularization term, so that the solution is unique. The regularized objective function is

$$\int_{\Omega} |K^{-1} \operatorname{div} \mathbf{p} - \lambda^{-1} f|^2 d\mathbf{x} + \gamma \int_{\Omega} |P\mathbf{p}|^2 d\mathbf{x},$$

where P is the orthogonal projector onto the null space of the divergence operator div . Later in [38], Hintermüller and Stadler showed that adding such a regularization term to the dual objective is equivalent to smoothing out the singularity of the TV in the primal objective. More precisely, the smoothed TV is given by $\int_{\Omega} \Phi(|\nabla f|) d\mathbf{x}$, where

$$\Phi(s) = \begin{cases} s & \text{if } |s| \geq \gamma, \\ \frac{\gamma}{2} + \frac{1}{2\gamma} s^2 & \text{if } |s| < \gamma. \end{cases}$$

An advantage of using this smoothed TV is that the staircase artifacts are reduced.

In [41, 42], Krishnan et al. considered the TV deblurring problem with bound constraints on the image u . An algorithm, called *nonnegatively constrained CGM*, combining the CGM and the PDAS algorithms has been proposed. The image u and its dual \mathbf{p} are treated as in the CGM method, whereas the bound constraints on u are treated as in the PDAS method. The resulting optimality conditions are shown to be semi-smooth. The scheme can also be interpreted as a *semi-smooth quasi-Newton’s method* and is proven to converge superlinearly. The method is formulated for isotropic TV, but it can also be applied to anisotropic TV after minor changes.

However, Hintermüller and Kunisch’s PDAS method [37] can only be applied to anisotropic TV because they used PDAS that can only handle linear constraints to treat the constraints on \mathbf{p} .

Bregman Iteration

Original Bregman Iteration

The *Bregman iteration* is proposed by Osher et al. in [49] for TV denoising. It has also been generalized to solving many convex inverse problems, e.g., [12]. In each

step, the signal removed in the previous step is added back. This is shown to alleviate the loss of contrast problem presented in the ROF model. Starting with the noisy image $f_0 = f$, the following steps are repeated for $j = 0, 1, 2, \dots$:

1. Set

$$u_{j+1} = \arg \min_u \left\{ \frac{1}{2} \int_{\Omega} (u - f_j)^2 d\mathbf{x} + \lambda \int_{\Omega} |\nabla u| \right\}.$$

2. Set $f_{j+1} = f_j + (f - u_{j+1})$.

In the particular case when f consists of a disk over a constant background, it can be proved that the loss of contrast can be totally recovered. Some theoretical analysis of the method can be found in [49].

For a general regularization functional $J(u)$, the Bregman distance is defined as

$$D_J^p(u, v) = J(u) - J(v) - \langle p, u - v \rangle,$$

where p is an element of the subgradient of J . In case of TV denoising, $J(u) = \lambda \int_{\Omega} |\nabla u|$. Then, starting with $f_0 = f$, the Bregman iteration is given by

1. Set

$$u_{j+1} = \arg \min_u \left\{ \frac{1}{2} \int_{\Omega} (u - f)^2 d\mathbf{x} + D_J^{p_j}(u, u_j) \right\}.$$

2. Set $f_{j+1} = f_j + (f - u_{j+1})$.
3. Set $p_{j+1} = f_{j+1} - f$.

In fact, steps 2 and 3 can be combined to $p_{j+1} = p_j + f - u_{j+1}$ without the need of keeping track of f_j . The above expression is for illustrating how the residual is added back to f_j . In this iteration, it has been shown that the Bregman distance between u_j and the *clean image* is monotonically decreasing as long as the L_2 -distance is larger than the magnitude of the noise component. But if one iterates until convergence, then $u_j \rightarrow f$, i.e., one just gets the noisy image back. This counterintuitive feature is indeed essential to solving other TV minimization problems, e.g., the basis pursuit problem presented next.

The Basis Pursuit Problem

An interesting feature of the Bregman iteration is that, in the discrete setting, if one replaces the term $\|u - f\|^2$ in the objective by $\|Au - f\|^2$, where $Au = f$ is underdetermined, then upon convergence of the Bregman iterations, one obtains the solution of the following *basis pursuit problem* [63]:

$$\min_u \{J(u) | Au = f\}.$$

When $\|Au - f\|^2$ is used in the objective instead of $\|u - f\|^2$, the Bregman iteration is given by:

1. Set

$$u_{j+1} = \arg \min_u \left\{ \frac{1}{2} \int_{\Omega} (Au - f)^2 d\mathbf{x} + D_J^{p_j}(u, u_j) \right\}.$$

2. Set $f_{j+1} = f_j + (f - Au_{j+1})$.

3. Set $p_{j+1} = A^T (f_{j+1} - f)$.

Split Bregman Iteration

Recently, Goldstein and Osher [35] proposed the *split Bregman iteration* which can be applied to solve the ROF problem efficiently. The main idea is to introduce a new variable so that the TV minimization becomes an L^1 minimization problem which can be solved efficiently by the Bregman iteration. This departs from the original Bregman iteration which solves a sequence of ROF problems to improve the quality of the restored image by bringing back the loss signal. The original Bregman iteration is not iterated until convergence. Moreover, it assumes the availability of a basic ROF solver. The split Bregman method, on the other hand, is an iterative method whose iterates converge to the solution of the ROF problem. In this method, a new variable $\mathbf{q} = \nabla u$ is introduced into the objective function:

$$\min_{u, \mathbf{q}} \left\{ \frac{1}{2} \int_{\Omega} (u - f)^2 d\mathbf{x} + \lambda \int_{\Omega} |\mathbf{q}| d\mathbf{x} \right\}. \tag{19}$$

This problem is solved using a penalty method to enforce the constraint $\mathbf{q} = \nabla u$. The objective with an added penalty is given by

$$G(u, \mathbf{q}) = \frac{\alpha}{2} \int_{\Omega} |\mathbf{q} - \nabla u|^2 d\mathbf{x} + \frac{1}{2} \int_{\Omega} (u - f)^2 d\mathbf{x} + \lambda \int_{\Omega} |\mathbf{q}| d\mathbf{x}. \tag{20}$$

Notice that if the variables (u, \mathbf{q}) are denoted by \mathbf{y} , then the above objective can be identified as

$$\min_{\mathbf{y}} \left\{ \frac{\alpha}{2} \int_{\Omega} |A\mathbf{y}|^2 d\mathbf{x} + J(\mathbf{y}) \right\},$$

where

$$\begin{aligned} A\mathbf{y} &= \mathbf{q} - \nabla u, \\ J(\mathbf{y}) &= \frac{1}{2} \int_{\Omega} (u - f)^2 d\mathbf{x} + \lambda \int_{\Omega} |\mathbf{q}| d\mathbf{x}. \end{aligned}$$

This is exactly the basis pursuit problem when $\alpha \rightarrow \infty$. Actually, even with a fixed finite α , as mentioned in section ‘‘The Basis Pursuit Problem,’’ when the Bregman

iteration is used, it converges to the solution of the problem

$$\min_{\mathbf{y}} \{J(\mathbf{y}) \mid A\mathbf{y} = 0\},$$

so that the constraint $\mathbf{q} = \nabla u$ is satisfied at convergence.

It is interesting to note that the split Bregman iteration can be viewed as a forward-backward splitting method [53]. Yet another point of view is provided next.

Augmented Lagrangian Method

In [62, 63], it is recognized that the split Bregman iteration is an *augmented Lagrangian method* [33]. This explains some good convergence behaviour of the split Bregman iteration. To motivate the augmented Lagrangian method, consider a general objective function $J(u)$ with equality constraint $H(u) = 0$. The idea of penalty methods is to solve a sequence of unconstrained problems

$$\min_u \left\{ J(u) + \frac{1}{\beta} \|H(u)\|^2 \right\}$$

with $\beta \rightarrow 0^+$, so that the constraint $H(u) = 0$ is enforced asymptotically. However, one may run into the embarrassing situation where both $H(u(\beta))$ (where $u(\beta)$ is the optimal u for a given β) and β converge to zero in the limit. This could mean that the objective function is stiff when β is very small. The idea of augmented Lagrangian methods is to use a *fixed* parameter. But the penalty term is added to the Lagrangian function, so that the resulting problem is equivalent to the original problem even without letting $\beta \rightarrow 0^+$. The augmented Lagrangian function is

$$L(u, v) = J(u) + v \cdot H(u) + \frac{1}{\beta} \|H(u)\|^2,$$

where v is a vector of Lagrange multipliers. Solving $\frac{\partial L}{\partial u} = \frac{\partial L}{\partial v} = 0$ for a saddle point yields exactly $H(u) = 0$ for any $\beta > 0$. The Bregman iteration applied to the penalized objective (20) is indeed computing a saddle point of the augmented Lagrangian function of (19) rather than optimizing (20) itself. Therefore, the constraint $\nabla u = \mathbf{q}$ accompanied with (19) is exact even with a fixed α .

Graph Cut Methods

Recently, there is a burst of interest in graph cut methods for solving various variational problems. The promises of these methods are that they are fast for many practical problems and they can provide globally optimal solution even for “non-convex problems.” The discussion below is extracted from [15, 27]. Readers are referred to [15, 27] and the references therein for a more thorough discussion of the subject. Since graph cut problems are combinatoric, the objective has to be cast in

a fully discrete way. That is, not only the image domain has to be discretized to a finite set but also the range of the intensity values. Therefore, in this framework, the given m -by- n image f is a function from $\mathbb{Z}_m \times \mathbb{Z}_n$ to \mathbb{Z}_K . The ROF problem thus becomes

$$F(u) = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n (u_{i,j} - f_{i,j})^2 + \lambda \|u\|_{TV} \rightarrow \min_{u: \mathbb{Z}_m \times \mathbb{Z}_n \rightarrow \mathbb{Z}_K},$$

where $\|u\|_{TV}$ is a discrete TV (4). The next question is how to transform this problem to a graph cut problem in such a way that it can be solved efficiently. It turns out that the (fully discretized) ROF problem can be converted to a finite sequence of graph cut problems. This is due to the co-area formula which is unique to TV. Details are described next.

Leveling the Objective

Some notations and basic concepts are in place. For simplicity, the following discrete TV is adopted:

$$\|u\|_{TV} = \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} |u_{i+1,j} - u_{i,j}| + |u_{i,j+1} - u_{i,j}|,$$

which is the anisotropic TV in (3), but with the range of u restricted to \mathbb{Z}_K . Recall that the binary image u^k is defined such that each $u_{i,j}^k$ equals 1 if $u_{i,j} \leq k$ and equals 0 otherwise. Thus, it is the k th lower level set of u . Then the co-area formula states that the discrete TV can be written as

$$\|u\|_{TV} = \sum_{k=0}^{K-2} \|u^k\|_{TV}.$$

Thus, it reduces to the TV of each ‘‘layer’’. Note that the TV of the $(K - 1)$ st level set must be zero, and therefore the above sum is only up to $K - 2$.

The fitting term in the objective can also be treated similarly as follows. Notice that for any function $g_{i,j}(s)$, it holds that

$$\begin{aligned} g_{i,j}(s) &= \sum_{k=0}^{s-1} [g_{i,j}(k+1) - g_{i,j}(k)] + g_{i,j}(0) \\ &= \sum_{k=0}^{K-2} [g_{i,j}(k+1) - g_{i,j}(k)] \chi_{k < s} + g_{i,j}(0), \end{aligned}$$

where $\chi_{k < s} = 1$ if $k < s$ and 0 otherwise. Define $g_{i,j}(k) = \frac{1}{2}(s - f_{i,j})^2$. Then,

$$\begin{aligned} \frac{1}{2}(u_{i,j} - f_{i,j})^2 &= g_{i,j}(u_{i,j}) \\ &= \sum_{k=0}^{K-2} [g_{i,j}(k+1) - g_{i,j}(k)] \chi_{k < u_{i,j}} + g_{i,j}(0) \\ &= \sum_{k=0}^{K-2} [g_{i,j}(k+1) - g_{i,j}(k)] (1 - u_{i,j}^k) + g_{i,j}(0). \end{aligned}$$

As a result, the ROF objective can be expressed as

$$\sum_{k=0}^{K-2} \left\{ \sum_{i,j} [g_{i,j}(k+1) - g_{i,j}(k)] (1 - u_{i,j}^k) + \lambda \|u^k\|_{TV} \right\} + C,$$

where $C = \sum_{i,j} g_{i,j}(0)$.

By defining the objective function

$$F^k(v^k) = \sum_{i,j} [g_{i,j}(k+1) - g_{i,j}(k)] (1 - v_{i,j}^k) + \lambda \|v^k\|_{TV},$$

where v^k is a binary function, the ROF problem is seen to be equivalent to

$$\min_{v^1, v^2, \dots, v^{K-2}} \sum_{k=0}^{K-2} F^k(v^k)$$

subject to the inclusion constraints $v_{i,j}^k \leq v_{i,j}^{k+1}$ for all i, j, k . The constraints make sure the binary functions $\{v^k\}_k$ define the lower level sets of some function v . A very important result is that the minimization can be done independently for each v^k ; amazingly, the solutions $\{v^k\}$ satisfy the inclusion property automatically! See [27] for further details.

Defining a Graph

To minimize each F^k w.r.t. a binary function v^k , a graph cut method is used. First observe that since $g_{i,j}(k) = \frac{1}{2}(k - f_{i,j})^2$, F^k can be simplified to

$$F^k(v^k) = \sum_{i,j} \left[\frac{1}{2} + (k - f_{i,j}) \right] (1 - v_{i,j}^k) + \lambda \|v^k\|_{TV}.$$

By absorbing some constants and dropping the superscript on v^k , the objective takes the following form:

$$F^k(v) = \sum_{i,j} \left(f_{i,j} - k - \frac{1}{2} \right) v_{i,j} + \lambda \|v\|_{TV}. \tag{21}$$

Then, a graph with $mn + 2$ nodes is constructed in the following way:

1. Each of the mn pixels is a node, labeled by (i, j) for $i = 1, 2, \dots, m$ and $j = 1, 2, \dots, n$.
2. Add two additional nodes, called the source S and the sink T .
3. For each (i, j) , connect it to $(i \pm 1, j)$ and $(i, j \pm 1)$ with capacity λ .
4. For each (i, j) , connect S to it with capacity $\frac{1}{2} + k - f_{i,j}$ if $\frac{1}{2} + k - f_{i,j} > 0$ and connect it to T with capacity $f_{i,j} - k - \frac{1}{2}$ if $f_{i,j} - k - \frac{1}{2} > 0$.

A cut (a.k.a. an st -cut) in the graph is a partition $(\mathcal{S}, \mathcal{T})$ such that $S \in \mathcal{S}$ and $T \in \mathcal{T}$. The cost of the cut $\mathcal{C}(\mathcal{S}, \mathcal{T})$ is defined as the sum of the capacities of all edges from \mathcal{S} to \mathcal{T} . For a given cut, let $v_{i,j}$ equals 1 if $(i, j) \in \mathcal{S}$ and equals 0 if $(i, j) \in \mathcal{T}$. Then it can be verified that

$$C(\mathcal{S}, \mathcal{T}) = \sum_{i,j} \max \left\{ \frac{1}{2} + k - f_{i,j}, 0 \right\} v_{i,j} + \max \left\{ f_{i,j} - k - \frac{1}{2}, 0 \right\} (1 - v_{i,j}) + \lambda \|v^k\|_{TV}$$

which is the same as F^k in (21), up to the constant $\sum_{i,j} \max \{f_{i,j} - k, 0\}$. Therefore, computing the minimum cut is equivalent to minimizing (21). It is also well known that the minimum cut problem is equivalent to the maximum flow problem.

Recall that there are $K - 1$ graphs to cut. A simple way is to do them one by one using any classical maximum flow algorithm. But one can exploit the inclusion property to reduce the work; for instance, see the divide-and-conquer algorithm proposed in [27].

In graph cut methods, a fundamental question is what kind of optimization problems can be transformed to a graph cut problem. A particularly relevant question is whether a function is levelable, i.e., its minimization can be done by first solving the simpler problem on each of its level set, followed by assembling the resulting level sets. Interestingly, the only levelable convex regularization function (satisfying some very natural and mild conditions) is TV [27]. This indicates that TV is much more than just an ordinary semi-norm.

Quadratic Programming

The discrete anisotropic TV is a piecewise linear function. Fu et al. [30] showed that by introducing some auxiliary variables, one can transform the TV to a linear function but with some additional linear constraints. Together with the fitting term, the problem to solve has a quadratic objective function with linear constraints.

The objective function considered is by Fu et al.

$$F(u) = \frac{1}{2} \|Ku - f\|^2 + \lambda \sum_{i,j} |u_{i+1,j} - u_{i,j}| + |u_{i,j+1} - u_{i,j}|,$$

which can also be written as

$$F(u) = \frac{1}{2} \|Ku - f\|^2 + \lambda \|Ru\|_1$$

where R is a $2mn$ -by- mn matrix. If the original isotropic TV is used, then it cannot be written in this form.

The trick they used is to let $v = Ru$ and then split it into positive and negative parts: $v^+ = \max(v, 0)$ and $v^- = \max(-v, 0)$. Then, the objective can be written as

$$G(u, v^+, v^-) = \frac{1}{2} \|Ku - f\|^2 + \lambda (1^T v^+ + 1^T v^-),$$

which is a quadratic function. But some linear constraints are added:

$$\begin{aligned} Ru &= v^+ - v^-, \\ v^+, v^- &\geq 0. \end{aligned}$$

Now, this problem can be solved by standard primal-dual interior-point methods. Here, “dual” refers to the Lagrange multipliers for the linear constraints. The major steps can be summarized as follows:

1. Write down the KKT system of optimality conditions, which has a form of $f(x, \mu, s) = 0$, where $x \geq 0$ is the variable of the original problem ($x = (u, v^+, v^-)$ in the present case); μ is the Lagrange multipliers for the equality constraints; and $s \geq 0$ is the Lagrange multipliers for the inequality constraints.
2. Relax the complementarity $xs = 0$ (part of $f(x, \mu, s) = 0$) to $xs = v$, where $v > 0$.
3. Solve the relaxed problem $f_v(x, \mu, s) = 0$ by Newton’s method.
4. After each Newton’s iteration, reduce the value of v so that the solution of $f(x, \mu, s) = 0$ is obtained at convergence.

In this method, the relaxed complementarity $xs = v$ forces the variables x, s to lie in the interior of the feasible region. Once the variables are away from the boundary, the problem becomes a nice unconstrained quadratic problem locally. The main challenge here is that the linear system to solve in each Newton’s iteration becomes increasingly ill conditioned. Under this framework, bound constraints such as $u_{\min} \leq u \leq u_{\max}$ or any linear equality constraints can be easily added.

Second-Order Cone Programming

The trick to “linearize” the TV presented in the last section does not work for isotropic TV. Goldfarb and Yin [34] proposed a *second-order cone programming*

(SOCP) formulation which works for the isotropic version (2). Moreover, its connection to SOCP allows the use of available SOCP solvers to obtain the solutions. The problem they considered is the constrained ROF problem:

$$\min_u \|u\|_{TV}$$

subject to

$$\|u - f\| \leq \sigma,$$

where σ is the standard deviation of the noise which is assumed to be known.

Let $w_{i,j}^x = u_{i+1,j} - u_{i,j}$ and $w_{i,j}^y = u_{i,j+1} - u_{i,j}$. The TV becomes

$$\sum_{i,j} \sqrt{(w_{i,j}^x)^2 + (w_{i,j}^y)^2}.$$

By introducing the variables $v = f - u$ and t and the constraint

$$(w_{i,j}^x)^2 + (w_{i,j}^y)^2 \leq t_{i,j}^2,$$

the TV minimization problem becomes

$$\begin{aligned} & \min \sum_{i,j} t_{i,j} \\ \text{s.t.} \quad & u + v = f \\ & w_{i,j}^x = u_{i+1,j} - u_{i,j} \\ & w_{i,j}^y = u_{i,j+1} - u_{i,j} \\ & (\sigma, v) \in \text{cone}^{mn+1} \\ & (t_{i,j}, w_{i,j}^x, w_{i,j}^y) \in \text{cone}^3. \end{aligned}$$

Here, cone^n is the second-order cone in \mathcal{R}^n :

$$\{x \in \mathbb{R}^n : \|(x_2, x_3, \dots, x_n)\| \leq x_1\}.$$

The optimal solution satisfies

$$t_{i,j}^2 = (w_{i,j}^x)^2 + (w_{i,j}^y)^2,$$

so that

$$\sum_{i,j} t_{i,j} = \sum_{i,j} \sqrt{(w_{i,j}^x)^2 + (w_{i,j}^y)^2} = \|u\|_{TV}.$$

A SOCP formulation of the dual ROF problem is also given in [34].

The SOCP can be solved by interior-point methods. The above formulation can be slightly simplified by eliminating u . But the number of variables (hence, the size of the Newton's equation) is still several times larger than the original problem. Goldfarb and Yin proposed a domain decomposition method to split the large programming problem into smaller ones, so that each subproblem can be solved efficiently. Of course, the convergence rate of the method deteriorates as the domain is further split.

Majorization-Minimization

Majorization-minimization (MM) (or *minorization-maximization*) [43] is a well-studied technique in optimization. The main idea is that at each step of the method, the objective function is replaced by a simple one, called the *surrogate function*, such that its minimization is easy to carry out and the result gives a smaller objective value of the original problem. For a given objective, usually many surrogate functions are possible. In many cases, one can even reduce multidimensional problems into a set of one-dimensional problems. Methods of this class have been heavily used in statistics communities. Indeed expectation-maximization (EM) algorithms are special cases of MM.

The use of MM to solving discrete TV problems can be traced back to the study of emission and transmission tomography reconstruction problems by Lange and Carson in [44]. Recently, some authors have applied the method to solving TV deblurring problems [6]. However, the method is actually the same as the classical lagged diffusivity fixed point iteration proposed by [58] for the particular surrogate function used in [6]. Nevertheless, it is still worthy to present the framework here because other surrogate functions can lead to different schemes.

Denote by u^k the k th iterate. In this method, the surrogate function (majorizer) $Q(u|u^k)$ is defined such that

$$\begin{aligned} F(u^k) &= Q(u^k|u^k) \\ F(u) &\leq Q(u|u^k), \text{ for all } u \end{aligned}$$

Then, the next iterate is defined to be the minimizer of the surrogate function

$$u^{k+1} := \arg \min_u Q(u|u^k).$$

In this way, the following monotonic decreasing property holds

$$F(u^{k+1}) \leq Q(u^{k+1}|u^k) \leq Q(u^k|u^k) = F(u^k).$$

Presumably, the function Q should be chosen so that its minimum is easy to compute. In many applications, it may even be chosen to have a separable form

$$Q(u|u^k) = Q_1(u_1|u^k) + Q_2(u_2|u^k) + \cdots + Q_n(u_n|u^k),$$

so that its minimization reduces to n 's one-dimensional (1D) problems. A promise of this method is that each iteration is very easy to carry out, which compensates its linear-only convergence. To construct a surrogate Q_{TV} for TV, first note that

$$\sqrt{a} = \left(\sqrt[4]{b}\right) \left(\frac{\sqrt{a}}{\sqrt[4]{b}}\right) \leq \frac{\sqrt{b}}{2} + \frac{a}{2\sqrt{b}}$$

for all $a, b \geq 0$. Let D_x and D_y be the forward difference operator in x and in y directions, respectively. Then,

$$\begin{aligned} \|u\|_{TV} &= \sum_{i,j} \sqrt{(D_x u_{i,j})^2 + (D_y u_{i,j})^2} \\ &\leq \frac{1}{2} \sum_{i,j} \sqrt{(D_x u_{i,j}^k)^2 + (D_y u_{i,j}^k)^2} + \frac{1}{2} \sum_{i,j} \frac{(D_x u_{i,j})^2 + (D_y u_{i,j})^2}{\sqrt{(D_x u_{i,j}^k)^2 + (D_y u_{i,j}^k)^2}} \end{aligned}$$

The surrogate is thus defined as

$$Q_{TV}(u|u^k) = \frac{1}{2} \|u^k\|_{TV} + \frac{1}{2} \sum_{i,j} \frac{(D_x u_{i,j})^2 + (D_y u_{i,j})^2}{\sqrt{(D_x u_{i,j}^k)^2 + (D_y u_{i,j}^k)^2}}$$

which is quadratic in u . Notice that the 2D discrete gradient matrix is given by

$$\nabla = \begin{bmatrix} \nabla_n \otimes I_m \\ I_n \otimes \nabla_m \end{bmatrix},$$

where ∇_m is the m -by- m 1D forward difference matrix (under Neumann boundary conditions)

$$\nabla_m = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \\ & & & & 0 \end{bmatrix}$$

Let $\lambda_{i,j}^k = 1/\sqrt{(D_x u_{i,j}^k)^2 + (D_y u_{i,j}^k)^2}$ and let

$$\Lambda^k = \text{diag}(\lambda_{1,1}^k, \dots, \lambda_{m,n}^k, \lambda_{1,1}^k, \dots, \lambda_{m,n}^k).$$

The surrogate becomes

$$Q_{TV}(u|u^k) = \frac{1}{2} \|u^k\|_{TV} + \frac{1}{2} u^T \nabla^T \Lambda^k \nabla u.$$

In this case, the minimization of Q_{TV} cannot be reduced to a set of 1D problems. But it does become quadratic.

Finally, the majorizer for the ROF model is

$$Q(u|u^k) = \frac{1}{2} \|Ku - f\|^2 + \lambda Q_{TV}(u|u^k).$$

While this method completely bypasses the need to optimize the TV term directly, each iteration requires solving the linear system

$$(K^T K + \lambda \nabla^T \Lambda^k \nabla) u^{k+1} = K^T f.$$

This scheme is exactly the lagged diffusivity fixed point iteration. Assume that K is full rank, then the linear system is positive definite. A standard way is to use preconditioned conjugate gradient to solve. Many preconditioners have been proposed for this problem in the 1990s, e.g., cosine transform and multigrid and multiplicative operator splitting; see [17] and the references therein. However, due to the highly varying coefficients in Λ^k , it can be nontrivial to solve efficiently.

Splitting Methods

Recently, there have been several proposals for solving TV deblurring problems based on the idea of separating the deblurring process and the TV regularization process. Many of them are based on the idea that the minimization of an objective of the form

$$F(u) = J_1(u) + J_2(Au),$$

with A a linear operator, can be approximated by the minimization of either of the following two objectives:

$$\begin{aligned} G(u, v) &= J_1(u) + \frac{\alpha}{2} \|u - v\|^2 + J_2(Av), \\ G(u, v) &= J_1(u) + \frac{\alpha}{2} \|Au - v\|^2 + J_2(v), \end{aligned}$$

where α is a large scalar. Then G is minimized w.r.t. u and v alternatively. In this way, at each iteration, the minimization of J_1 and J_2 is done separately. The same idea can be generalized to split an objective with n terms to an objective with n variables.

Consider the discrete ROF model:

$$F(u) = \frac{1}{2} \|Ku - f\|^2 + \lambda \|\nabla u\|_1.$$

Huang et al. [39] and Bresson and Chan [10] considered the splitting

$$G(u, v) = \frac{1}{2} \|Ku - f\|^2 + \frac{\alpha}{2} \|u - v\|^2 + \lambda \|\nabla v\|_1.$$

In this case, the minimization w.r.t. u becomes

$$(K^T K + \alpha I) u = K^T f + \alpha v,$$

which can be solved with the fast Fourier transform (FFT) in $O(N \log N)$ operations when the blurring matrix K can be diagonalized by a fast transform matrix. The minimization w.r.t. v is the ROF denoising problem which can be solved using any of the aforementioned denoising method. Both [39] and [10] employed Chambolle’s dual algorithm. The point is that solving TV denoising is much easier than solving TV deblurring (directly). Moreover, some algorithms such as those based on graph cut cannot be applied to deblurring directly. The reason is that the pixel values in the fitting are no longer separable, which in turn makes the fitting term not “levelable.” However, using the splitting technique, one can now apply graph cut methods to solve each denoising problem.

This method is generally very fast. Moreover, it often works for a large range of α . But when α is too large, the Chambolle’s iteration may slow down. This splitting method has also been applied to other image processing problems such as segmentation [10].

An alternative splitting is proposed by Wang et al. [59]. The bivariate function they used is given by

$$G(u, v) = \frac{1}{2} \|Ku - f\|^2 + \frac{\alpha}{2} \|\nabla u - v\|^2 + \lambda \|v\|_1.$$

The minimization w.r.t. u requires solving

$$(K^T K - \alpha \Delta) u = K^T f + \alpha v,$$

where Δ is the 2D Laplacian. This equation can again be solved with FFT in $O(N \log N)$ operations. The minimization w.r.t. v gs decoupled into N minimization problems (one for each pixel) of two variables. A simple closed-form solution for the 2D minimization problems is available. Therefore, the computation cost per iteration is even less than the approach taken in [39] and [10]. Remark that this objective is indeed the same as the split Bregman method (20). A difference is that when the split Bregman iteration converges, it holds exactly that $\nabla u = v$. But the simple alternating minimization used in most splitting methods does not guarantee $\nabla u = v$ at convergence.

An alternative splitting is introduced by Bect et al. in [5]. It is based on the observation that, for any symmetric positive definite matrix B with $\|B\| < 1$, it holds that

$$\langle Bv, v \rangle = \min_{u \in \mathbb{R}^N} \left\{ \|u - v\|^2 + \langle Cu, u \rangle \right\}$$

for all $v \in \mathcal{R}^N$, where $C = B(I - B)^{-1}$. Then, the ROF model can be formulated as the minimization of the following bivariate function:

$$G(u, v) = \frac{1}{2\mu} \left(\|u - v\|^2 + \langle Cu, u \rangle \right) + \frac{1}{2} \left(\|f\|^2 - 2 \langle Kv, f \rangle \right) + \lambda \|\nabla v\|_1,$$

where $\mu > 0$ such that $\mu \|K^T K\| < 1$ and $B = \mu K^T K$. The minimization of G w.r.t. u has a closed-form solution $u = (I - B)v = (I - \mu K^T K)v$. The minimization of G w.r.t. v is a TV denoising problem. At convergence, the minimizer of F is exactly recovered. An interesting property of this splitting is that it does not involve any matrix inversion in the alternating minimization of G .

5 Conclusion

In this chapter, some recent developments of numerical methods for TV minimization and their applications are reviewed. The chosen topics only reflect the interest of the authors and are by no means comprehensive. It is also hoped that this chapter can serve as a guide to recent literature on some of these recent developments.

Cross-References

- ▶ [Compressive Sensing](#)
- ▶ [Duality and Convex Programming](#)
- ▶ [Energy Minimization Methods](#)
- ▶ [Large-Scale Inverse Problems in Imaging](#)
- ▶ [Mumford and Shah Model and Its Applications to Image Segmentation and Image Restoration](#)
- ▶ [Numerical Methods and Applications in Total Variation Image Restoration](#)
- ▶ [Regularization Methods for Ill-Posed Problems](#)
- ▶ [Total Variation in Imaging](#)
- ▶ [Variational Methods in Shape Analysis](#)

References

1. Acar, A., Vogel, C.: Analysis of bounded variation penalty methods for ill-posed problems. *Inverse Probl.* **10**(6), 1217–1229 (1994)
2. Adams, R., Fournier, J.: *Sobolev Spaces*. Volume 140 of Pure and Applied Mathematics, 2nd edn. Academic, New York (2003)
3. Aujol, J.-F.: Some first-order algorithms for total variation based image restoration. *J. Math. Imaging Vis.* **34**(3), 307–327 (2009)
4. Aujol, J.-F., Gilboa, G., Chan, T., Osher, S.: Structure-texture image decomposition – modeling, algorithms, and parameter selection. *Int. J. Comput. Vis.* **67**(1), 111–136 (2006)

5. Bect, J., Blanc-Féraud, L., Aubert, G., Chambolle, A.: A l^1 -unified variational framework for image restoration. In: Proceedings of ECCV. Volume 3024 of Lecture Notes in Computer Sciences, Prague, Czech Republic, pp. 1–13 (2004)
6. Bioucas-Dias, J., Figueiredo, M., Nowak, R.: Total variation-based image deconvolution: a majorization-minimization approach. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2006), Toulouse, France, vol. 2, pp. 14–19 (2006)
7. Blomgren, P., Chan, T.: Color TV: total variation methods for restoration of vector-valued images. *IEEE Trans. Image Process.* **7**, 304–309 (1998)
8. Boyd, S., Vandenberghe, L.: *Convex Optimization*. Cambridge University Press, Cambridge (2004)
9. Bresson, X., Chan, T.: Non-local unsupervised variational image segmentation models. UCLA CAM Report, 08–67 (2008)
10. Bresson, X., Chan, T.: Fast dual minimization of the vectorial total variation norm and applications to color image processing. *Inverse Probl. Imaging* **2**(4), 455–484 (2008)
11. Buades, A., Coll, B., Morel, J.: A review of image denoising algorithms, with a new one. *Multiscale Model. Simul.* **4**(2), 490–530 (2005)
12. Burger, M., Frick, K., Osher, S., Scherzer, O.: Inverse total variation flow. *Multiscale Model. Simul.* **6**(2), 366–395 (2007)
13. Carter, J.: Dual methods for total variation-based image restoration. Ph.D. thesis, UCLA, Los Angeles (2001)
14. Chambolle, A.: An algorithm for total variation minimization and applications. *J. Math. Imaging Vis.* **20**, 89–97 (2004)
15. Chambolle, A., Darbon, J.: On total variation minimization and surface evolution using parametric maximum flows. *Int. J. Comput. Vis.* **84**(3), 288–307 (1997)
16. Chambolle, A., Lions, P.: Image recovery via total variation minimization and related problems. *Numer. Math.* **76**, 167–188 (1997)
17. Chan, R., Chan, T., Wong, C.: Cosine transform based preconditioners for total variation deblurring. *IEEE Trans. Image Process.* **8**, 1472–1478 (1999)
18. Chan, R., Wen, Y., Yip, A.: A fast optimization transfer algorithm for image inpainting in wavelet domains. *IEEE Trans. Image Process.* **18**(7), 1467–1476 (2009)
19. Chan, T., Vese, L.: Active contours without edges. *IEEE Trans. Image Process.* **10**(2), 266–277 (2001)
20. Chan, T., Golub, G., Mulet, P.: A nonlinear primal-dual method for total variation-based image restoration. *SIAM J. Sci. Comput.* **20**, 1964–1977 (1999)
21. Chan, T., Esedoğlu, S., Park, F., Yip, A.: Recent developments in total variation image restoration. In: Paragios, N., Chen, Y., Faugeras, O. (eds.) *Handbook of Mathematical Models in Computer Vision*. Springer, Berlin, pp. 17–32 (2005)
22. Chan, T., Esedoğlu, S., Nikolova, M.: Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM J. Appl. Math.* **66**(5), 1632–1648 (2006)
23. Chan, T., Shen, J., Zhou, H.: Total variation wavelet inpainting. *J. Math. Imaging Vis.* **25**(1), 107–125 (2006)
24. Chan, T., Ng, M., Yau, C., Yip, A.: Superresolution image reconstruction using fast inpainting algorithms. *Appl. Comput. Harmon. Anal.* **23**(1), 3–24 (2007)
25. Christiansen, O., Lee, T., Lie, J., Sinha, U., Chan, T.: Total variation regularization of matrix-valued images. *Int. J. Biomed. Imaging* **2007**, 27432 (2007)
26. Combettes, P., Wajs, V.: Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.* **4**(4), 1168–1200 (2004)
27. Darbon, J., Sigelle, M.: Image restoration with discrete constrained total variation part I: fast and exact optimization. *J. Math. Imaging Vis.* **26**, 261–276 (2006)
28. Efros, A., Leung, T.: Texture synthesis by non-parametric sampling. In: Proceedings of the IEEE International Conference on Computer Vision, Corfu, vol. 2, pp. 1033–1038 (1999)
29. Esser, E., Zhang, X., Chan, T.: A general framework for a class of first order primal-dual algorithms for TV minimization. *SIAM J. Imaging Sci.* **3**(4), 1015–1046 (2010)

30. Fu, H., Ng, M., Nikolova, M., Barlow, J.: Efficient minimization methods of mixed l_2 - l_1 and l_1 - l_1 norms for image restoration. *SIAM J. Sci. Comput.* **27**(6), 1881–1902 (2006)
31. Gilboa, G., Osher, S.: Nonlocal operators with applications to image processing. *Multiscale Model. Simul.* **7**(3), 1005–1028 (2008)
32. Giusti, E.: *Minimal Surfaces and Functions of Bounded Variation*. Birkhäuser, Boston (1984)
33. Glowinski, R., Le Tallec, P.: *Augmented Lagrangians and Operator-Splitting Methods in Nonlinear Mechanics*. SIAM, Philadelphia (1989)
34. Goldfarb, D., Yin, W.: Second-order cone programming methods for total variation based image restoration. *SIAM J. Sci. Comput.* **27**(2), 622–645 (2005)
35. Goldstein, T., Osher, S.: The split Bregman method for l^1 -regularization problems. *SIAM J. Imaging Sci.* **2**(2), 323–343 (2009)
36. Hintermüller, M., Kunisch, K.: Total bounded variation regularization as a bilaterally constrained optimisation problem. *SIAM J. Appl. Math.* **64**, 1311–1333 (2004)
37. Hintermüller, M., Stadler, G.: A primal-dual algorithm for TV-based inf-convolution-type image restoration. *SIAM J. Sci. Comput.* **28**, 1–23 (2006)
38. Hintermüller, M., Ito, K., Kunisch, K.: The primal-dual active set strategy as a semismooth Newton's method. *SIAM J. Optim.* **13**(3), 865–888 (2003)
39. Huang, Y., Ng, M., Wen, Y.: A fast total variation minimization method for image restoration. *Multiscale Model. Simul.* **7**(2), 774–795 (2008)
40. Kanwal, R.P.: *Generalized Functions: Theory and Applications*. Birkhäuser, Boston (2004)
41. Krishnan, D., Lin, P., Yip, A.: A primal-dual active-set method for non-negativity constrained total variation deblurring problems. *IEEE Trans. Image Process.* **16**(2), 2766–2777 (2007)
42. Krishnan, D., Pham, Q., Yip, A.: A primal dual active set algorithm for bilaterally constrained total variation deblurring and piecewise constant Mumford-Shah segmentation problems. *Adv. Comput. Math.* **31**(1–3), 237–266 (2009)
43. Lange, K.: (2004) *Optimization*. Springer, New York
44. Lange, K., Carson, R.: (1984) EM reconstruction algorithms for emission and transmission tomography. *J. Comput. Assist. Tomogr.* **8**, 306–316
45. Law, Y., Lee, H., Yip, A.: A multi-resolution stochastic level set method for Mumford-Shah image segmentation. *IEEE Trans. Image Process.* **17**(3), 2289–2300 (2008)
46. LeVeque, R.: *Numerical Methods for Conservation Laws*, 2nd edn. Birkhäuser, Basel (2005)
47. Mumford, D., Shah, J.: Optimal approximation by piecewise smooth functions and associated variational problems. *Commun. Pure Appl. Math.* **42**, 577–685 (1989)
48. Ng, M., Qi, L., Tang, Y., Huang, Y.: On semismooth Newton's methods for total variation minimization. *J. Math. Imaging Vis.* **27**(3), 265–276 (2007)
49. Osher, S., Burger, M., Goldfarb, D., Xu, J., Yin, W.: An iterative regularization method for total variation based image restoration. *Multiscale Model. Simul.* **4**, 460–489 (2005)
50. Royden, H.: *Real Analysis*, 3rd edn. Prentice-Hall, Englewood Cliffs (1988)
51. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Physica D* **60**, 259–268 (1992)
52. Sapiro, G., Ringach, D.: Anisotropic diffusion of multivalued images with applications to color filtering. *IEEE Trans. Image Process.* **5**, 1582–1586 (1996)
53. Setzer, S.: Split Bregman algorithm, Douglas-Rachford splitting and frame shrinkage. In: *Proceedings of Scale-Space, Voss, Norway*, pp. 464–476 (2009)
54. Setzer, S., Steidl, G., Popilka, B., Burgeth, B.: Variational methods for denoising matrix fields. In: Laidlaw, D., Weickert, J. (eds.) *Visualization and Processing of Tensor Fields: Advances and Perspectives*, Mathematics and Visualization, pp. 341–360. Springer, Berlin (2009)
55. Shen, J., Kang, S.: Quantum TV and application in image processing. *Inverse Probl. Imaging* **1**(3), 557–575 (2007)
56. Strang, G.: Maximal flow through a domain. *Math. Program.* **26**(2), 123–143 (1983)
57. Tschumperlé, D., Deriche, R.: Diffusion tensor regularization with constraints preservation. In: *Proceedings of 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Kauai*, vol. 1, pp. 948–953. IEEE Computer Science Press (2001)

58. Vogel, C., Oman, M.: Iteration methods for total variation denoising. *SIAM J. Sci. Comput.* **17**, 227–238 (1996)
59. Wang, Y, Yang, J., Yin, W., Zhang, Y.: A new alternating minimization algorithm for total variation image reconstruction. *SIAM J. Imaging Sci.* **1**(3), 248–272 (2008)
60. Wang, Z., Vemuri, B., Chen, Y., Mareci, T.: A constrained variational principle for direct estimation and smoothing of the diffusion tensor field from complex DWI. *IEEE Trans. Med. Imaging* **23**(8), 930–939 (2004)
61. Weiss, P., Aubert, G., Blanc-Fèraud, L.: Efficient schemes for total variation minimization under constraints in image processing. *SIAM J. Sci. Comput.* **31**(3), 2047–2080 (2009)
62. Wu, C., Tai, X.C.: Augmented Lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and high order models. *SIAM J. Imaging Sci.* **3**(3), 300–339 (2010)
63. Yin, W., Osher, S., Goldfarb, D., Darbon, J.: Bregman iterative algorithms for l^1 -minimization with applications to compressed sensing. *SIAM J. Imaging Sci.* **1**(1), 143–168 (2008)
64. Zhu, M., Chan, T.: An efficient primal-dual hybrid gradient algorithm for total variation image restoration. *UCLA CAM Report*, 08–34 (2008)
65. Zhu, M., Wright, S.J., Chan, T.F.: Duality-based algorithms for total-variation-regularized image restoration. *Comput. Optim. Appl.* **47**(3), 377–400 (2010)