

A Primal-Dual Active-Set Algorithm for Bilaterally Constrained Total Variation Deblurring and Piecewise Constant Mumford-Shah Segmentation Problems

D. Krishnan* Quang Vinh Pham* Andy M. Yip*[†]

October 24, 2007

Abstract

In this paper, we propose a fast primal-dual algorithm for solving bilaterally constrained total variation minimization problems which subsume the bilaterally constrained total variation image deblurring model and the two-phase piecewise constant Mumford-Shah image segmentation model. The presence of the bilateral constraints makes the optimality conditions of the primal-dual problem semi-smooth which can be solved by a semi-smooth Newton's method superlinearly. But the linear system to solve at each iteration is very large and difficult to precondition. Using a primal-dual active-set strategy, we reduce the linear system to a much smaller and better structured one so that it can be solved efficiently by conjugate gradient with an approximating inverse preconditioner. Locally superlinear convergence results are derived for the proposed algorithm. Numerical experiments are also provided for both deblurring and segmentation problems. In particular, for the deblurring problem, we show that the addition of the bilateral constraints to the total variation model improves the quality of the solutions.

1 Introduction

Image deblurring and image segmentation are two fundamental problems in image processing. The goal of image deblurring is to recover a sharp image from a blurred and noisy observed image. The goal of image segmentation is to partition the image into homogeneous regions so that pixels within a region belong to the same object.

Among the many approaches, variational and partial differential equation (PDE) methods have been very successful [9, 10]. These methods allow us to impose geometric constraints such as regularity of solutions conveniently. Moreover, existing theory, models and computational methods are available for variational problems and PDEs in various contexts. Two of the most well-known and influential examples are the total variation (TV) image deblurring model [22] and the Mumford-Shah image segmentation model [20].

1.1 The Total Variation Deblurring Model

In image deblurring, the degradation model we use is the linear shift-invariant (LSI) model:

$$f = k * u + \eta,$$

*Department of Mathematics, National University of Singapore, 2, Science Drive 2, Singapore 117543, Singapore.

[†]Corresponding author. Email: andyyip@nus.edu.sg.

where $f = (f_{ij})$ is an observed m -by- n blurred and noisy image, k is a known shift-invariant *point spread function* (PSF), $u = (u_{ij})$ is the unknown data to be recovered, η is an unknown noise, and $*$ is the discrete convolution operator. The indices i and j refer to the pixel location in the image. By rearranging f , u and η into vectors using lexicographical ordering, the model can be expressed in the matrix form $f = Ku + \eta$ where K is an mn -by- mn blurring matrix. The matrix K is often very ill-conditioned so that the straightforward inverse filtering ($K^{-1}f$) gives poor results due to severe noise amplification. Hence, regularization is needed.

The discrete total variation deblurring model seeks for the minimizer of the following Tikhonov-type functional:

$$F_0^{\text{TV}}(u) = \frac{1}{2}\|Ku - f\|^2 + \beta\|u\|_{\text{TV}},$$

where $\beta > 0$ is a regularization parameter, $\|\cdot\|$ is the l^2 norm, and $\|\cdot\|_{\text{TV}}$ is the discrete TV norm. The regularization parameter β controls the tradeoff between the goodness-of-fit of the LSI model and the regularity of the solution. The discrete TV norm is defined by

$$\|u\|_{\text{TV}} = \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} |(\nabla u)_{i,j}|, \quad (1)$$

where

$$(\nabla u)_{i,j} = \begin{bmatrix} u_{i+1,j} - u_{i,j} \\ u_{i,j+1} - u_{i,j} \end{bmatrix},$$

and $|\cdot|$ is the Euclidean norm for \mathbb{R}^2 . The matrix ∇ can be expressed as

$$\nabla = \begin{bmatrix} I_n \otimes \nabla_1 \\ \nabla_1 \otimes I_m \end{bmatrix},$$

where ∇_1 is the 1-dimensional forward difference operator, and I_m and I_n are identity matrices of order m and n respectively. The main advantage of the TV model is the ability to preserve edges in the image. This is due to the piecewise smooth regularization property of the TV norm.

Image values which represent physical quantities such as photon count or energies are not only non-negative, but they often have an upper bound as well, owing to finite number of bits being used for image representation. For example, pixel values in digital images are often bounded by $0 \leq u \leq 255$. Several studies have shown that imposing non-negativity constraints on image values can improve the quality of restoration results [18, 23]. We show in our experiments that adding an upper bound can further improve the quality.

In this paper, we consider the bilaterally constrained problem

$$\min_{\substack{u \in \mathbb{R}^{mn} \\ l \leq u \leq h}} F^{\text{TV}}(u), \quad (2)$$

where

$$F^{\text{TV}}(u) = \frac{1}{2}\|Ku - f\|^2 + \beta\|u\|_{\text{TV}} + \frac{\alpha}{2}\|u\|^2.$$

The constraints are understood componentwise. We include the third term to the model so that it can be applied to a larger class of problems. When $\alpha = 0$, this problem is convex for all K and is strictly convex when K is of full rank. When $\alpha > 0$, the problem is strictly convex. We assume that either K is of full rank or $\alpha > 0$ so that the problem has a unique solution. In deblurring problems, even if the observed data fall within the bounds l and h , the deblurred result from the

unconstrained deblurring model may contain values outside these bounds. Therefore, imposing the bounds is non-redundant.

While the addition of the bilateral constraints can improve the solution quality, the resulting optimization problem is harder to solve than its unconstrained counterpart. Bertsekas [1] presents a Projected Newton’s algorithm which can be used to solve the primal problem Eq. (2). This algorithm has been used in [25, 26] to solve TV deblurring problems with non-negativity constraints. We compare the performance of our proposed algorithm to this method. The proposed algorithm here is a generalization of our previous work in [15] which works for non-negativity constrained deblurring problems only. But the generalized method presented here can be applied to segmentation problems as well.

1.2 The Piecewise Constant Mumford-Shah Image Segmentation Model

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 constants $\mathbf{c} = (c_1, c_2, \dots, c_L)$ which minimize an energy functional given by:

$$F^{\text{MS}}(C, \mathbf{c}) = \sum_{l=1}^L \int_{\Omega_l} [f(\mathbf{x}) - c_l]^2 d\mathbf{x} + \beta \cdot \text{Length}(C). \quad (3)$$

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 non-trivial to optimize especially when the curves need to be split and merged. Chan and Vese [11] proposed a level set-based method which can handle topological change 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

$$F^{\text{CV}}(\phi, c_1, c_2) = \int_{\Omega} H(\phi(\mathbf{x})) [f(\mathbf{x}) - c_1]^2 d\mathbf{x} + \int_{\Omega} [1 - H(\phi(\mathbf{x}))] [f(\mathbf{x}) - c_2]^2 d\mathbf{x} + \beta \cdot \text{Length}(\{\phi = 0\}).$$

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 , for example,

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

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 often correspond to undesirable segmentations, see [16].

Interestingly, Chan *et al.* showed in [7] that for fixed c_1 and c_2 , then the above non-convex objective can be reformulated as a convex problem so that a global minimum can be easily computed. The globalized objective is given by

$$F^{\text{CEN}}(u, c_1, c_2) = \int_{\Omega} \{ [f(\mathbf{x}) - c_1]^2 - [f(\mathbf{x}) - c_2]^2 \} u(\mathbf{x}) d\mathbf{x} + \beta \int_{\Omega} |\nabla u(\mathbf{x})| d\mathbf{x}, \quad (4)$$

which is minimized over all u satisfying $0 \leq u \leq 1$, c_1 and c_2 . The last term is the continuous version of the TV norm. 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 [7]. Some other proposals for computing global solutions can be found in [14, 16].

In this paper, we consider the discrete version given by

$$F^{\text{CEN}}(u, c_1, c_2) = \langle s, u \rangle + \beta \|u\|_{\text{TV}} + \frac{\alpha}{2} \|u - \frac{1}{2}\|^2, \quad (5)$$

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. We therefore assume that $\alpha > 0$ so that solution is unique. The additive constant $\frac{1}{2}$ is introduced in the third term so that the minimizer does not bias towards 0 or 1.

To optimize the globalized objective function (4), Chan *et al.* proposed to use an exact penalty method to convert the bilaterally constrained problem to an unconstrained problem. Then the gradient descent method is applied. It is well-known that gradient descent is very slow. This can cause some practical problems. The final segmentation is obtained by thresholding the solution u . In many practical cases, the exact solution u should be close to a binary function (see [6] for some exceptions though). Thus if it is computed with a high accuracy, then it is also very close to binary so that a large range of thresholds in $[0, 1]$ can be used. However, if u is not computed with a high accuracy, then there could be significantly many values of u spread over $[0, 1]$. In this case, choosing different thresholds can lead to significantly different segmentations.

In [3], a dual method which is a variant of Chambolle's dual method for denoising [5] has been proposed. It is a gradient descent-like method applied to the dual objective function. A splitting strategy is also used so that the terms in the objective are optimized individually and alternatively. This method is very fast to obtain a good segmentation. We found that u converges to a close-to-binary function quickly after which it slows down significantly. This is because of the gradient descent nature of the method. But since the final segmentation is obtained by thresholding, the slowness at the later stage causes no problem at all. However, the dual method of [5] is unsuitable for solving deblurring problems since the step size is restricted by the reciprocal of the condition number of the blurring matrix K [15].

In this paper, we propose a primal-dual active-set method to solve a class of bilaterally constrained TV minimization problems which include TV deblurring and two-phase piecewise constant Mumford-Shah segmentation models as special cases. The main idea is to apply the semi-smooth Newton's method to solve the optimality conditions which are semi-smooth (due to the presence of the bilateral constraints) and to use the active-set strategy to simplify the Newton's equation to a much smaller and better structured system to facilitate preconditioning. The convergence of the semi-smooth Newton's method is locally superlinear. For deblurring, our primal-dual method is much faster than the (primal) projected Newton's method. For segmentation, our method is much faster than the (primal) gradient descent method and is competitive with the dual gradient descent method.

The rest of the paper is organized as follows: Section 2 presents the derivation of our proposed primal-dual method (which we call BCGM) for solving a general class of bilaterally constrained TV minimization problems. Local superlinear convergence is established in Section 3. In Section 4, the numerical performance of our algorithm for solving deblurring and segmentation problems is compared to some state-of-the-art algorithms.

2 Main Results

2.1 The Dual Problem

In this subsection, we derive the dual problem for a general class of bilaterally constrained TV minimization problems.

It is easy to see that the TV deblurring problem (2) and the Mumford-Shah segmentation problem (3) are special cases of the following bilaterally constrained problem:

$$\min_{u \in C_u} \frac{1}{2} \|Ku\|^2 + \langle b, u \rangle + \frac{\alpha}{2} \|u\|^2 + \beta \|u\|_{\text{TV}}, \quad (6)$$

where

$$C_u = \{u \in \mathbb{R}^{mn} : l \leq u \leq h\}$$

is the feasible set for u .

A difficulty to optimize objective functions with a TV norm is that the TV norm is non-differentiable. A commonly used remedy is to regularize the TV norm by replacing $|(\nabla u)_{i,j}|$ in (1) with

$$|(\nabla u)_{i,j}|_\epsilon = \sqrt{|(\nabla u)_{i,j}|^2 + \epsilon},$$

which is a smooth function. The trade-off in choosing this smoothing parameter ϵ is the solution accuracy error versus the speed of convergence.

An alternative way studied by Chan *et al.* [8], Carter [4] and Chambolle [5] is to introduce a Fenchel dual variable p of u . This is done using the Fenchel transform to obtain

$$\|u\|_{\text{TV}} = \sup_{p \in C_p} \langle u, -\text{div} p \rangle.$$

Here $p = (p_{i,j})$ is the dual variable with

$$\begin{aligned} p_{i,j} &= \begin{bmatrix} p_{i,j}^x \\ p_{i,j}^y \end{bmatrix} \in \mathbb{R}^2, \\ C_p &= \{p \in \mathbb{R}^{2mn} : |p_{i,j}| \leq 1 \forall i, j\}, \end{aligned}$$

and div is the discrete divergence matrix defined by

$$(\text{div} p)_{i,j} = p_{i,j}^x - p_{i-1,j}^x + p_{i,j}^y - p_{i,j-1}^y.$$

It can be verified that

$$\text{div} = \begin{bmatrix} I_n \otimes \text{div}_1 & \text{div}_1 \otimes I_m \end{bmatrix} = -\nabla^T,$$

where div_1 is the 1-dimensional backward difference operator. An advantage of solving the dual problem is that the extra smoothing parameter ϵ is not needed and it is therefore more faithful to the original TV model.

Using the Fenchel transform, we derive the dual problem in the following proposition. Due to the presence of the bilateral constraints, some Lagrange dual variables are introduced.

Proposition 1 *The dual problem to (6) is given by*

$$\min_{p \in C_p} \min_{\lambda_l \in \mathbb{R}_+^{mn}} \min_{\lambda_h \in \mathbb{R}_+^{mn}} \left\{ \frac{1}{2} \|B^{1/2}(\beta \text{div} p + \lambda_l - \lambda_h - b)\|^2 - \langle l, \lambda_l \rangle + \langle h, \lambda_h \rangle \right\}, \quad (7)$$

where $B := (K^T K + \alpha I)^{-1}$.

Proof: Using the Fenchel transform, the problem (6) becomes

$$\min_{u \in C_u} \max_{p \in C_p} \left\{ \frac{1}{2} \|Ku\|^2 + \frac{\alpha}{2} \|u\|^2 + \langle u, b - \beta \operatorname{div} p \rangle \right\}.$$

Since the objective is convex in u and concave in p and the constraints are convex, the strong max-min property holds [2]. Therefore, we can change the order of the min and max terms to arrive at:

$$\max_{p \in C_p} \min_{u \in C_u} \left\{ \frac{1}{2} \|B^{-1/2}u\|^2 + \langle u, b - \beta \operatorname{div} p \rangle \right\}. \quad (8)$$

Here $B = (K^T K + \alpha I)^{-1}$.

Consider the inner minimization for a fixed p . The Lagrangian for this problem is given by

$$\begin{aligned} L(u, \lambda_l, \lambda_h) &:= \frac{1}{2} \|B^{-1/2}u\|^2 + \langle u, b - \beta \operatorname{div} p \rangle - \langle u - l, \lambda_l \rangle + \langle u - h, \lambda_h \rangle \\ &= \frac{1}{2} \|B^{-1/2}u\|^2 + \langle u, b - \beta \operatorname{div} p - \lambda_l + \lambda_h \rangle + \langle l, \lambda_l \rangle - \langle h, \lambda_h \rangle, \end{aligned} \quad (9)$$

where $\lambda_l \geq 0$ is the Lagrange multiplier for the constraint $u \geq l$ and $\lambda_h \geq 0$ is the Lagrange multiplier for the constraint $u \leq h$. Then,

$$\nabla_u L = B^{-1}u + b - \beta \operatorname{div} p - \lambda_l + \lambda_h.$$

Solving $\nabla_u L(u^*, \lambda_l^*, \lambda_h^*) = 0$ gives

$$u^* = B(\beta \operatorname{div} p - b + \lambda_l^* - \lambda_h^*).$$

To derive the dual problem, we substitute $u = B(\beta \operatorname{div} p - b + \lambda_l - \lambda_h)$ into the Lagrangian (9). Then we have

$$L(\lambda_l, \lambda_h) = -\frac{1}{2} \|B^{1/2}(\beta \operatorname{div} p - b + \lambda_l - \lambda_h)\|^2 + \langle l, \lambda_l \rangle - \langle h, \lambda_h \rangle.$$

This is the Lagrange dual objective function [2] satisfying

$$\max_{\lambda_l \in \mathbb{R}_+^{mn}} \max_{\lambda_h \in \mathbb{R}_+^{mn}} L(\lambda_l, \lambda_h) = \min_{u \in C_u} \left\{ \frac{1}{2} \|B^{-1/2}u\|^2 + \langle u, b - \beta \operatorname{div} p \rangle \right\}.$$

Therefore, the problem in (8) becomes (7) after multiplying the objective by -1 and changing the max functions to min. \square

For TV minimization without constraints, Carter [4], Chambolle [5], Hintermüller and Stadler [13] and Ng *et al.* [21] studied methods for solving the dual problem directly. These methods work well for denoising problems where $K = I$. But when K is ill-conditioned, the matrix $B = (K^T K + \alpha I)^{-1}$ in the quadratic term causes numerical difficulty. This is owing to the fact that values of α that are used are usually small. As a result, the matrix B is ill-conditioned for reasonable values of α . Therefore, we adopt the primal-dual approach proposed by Chan *et al.* in [8] for solving unconstrained problems.

2.2 The Primal-dual Problem

In this subsection, we present the proposed primal-dual system for the problem (6). The optimality conditions (KKT conditions [2]) for the problem (7) are as follows

$$-\beta \nabla B(\beta \operatorname{div} p - b + \lambda_l - \lambda_h) + \mu \odot p = 0, \quad (10)$$

$$B(\beta \operatorname{div} p - b + \lambda_l - \lambda_h) - l - \nu_l = 0, \quad (11)$$

$$-B(\beta \operatorname{div} p - b + \lambda_l - \lambda_h) + h - \nu_h = 0, \quad (12)$$

$$\mu \odot (|p|^2 - 1) = 0, \quad (13)$$

$$\nu_l \odot \lambda_l = 0, \quad (14)$$

$$\nu_h \odot \lambda_h = 0, \quad (15)$$

together with the inequality constraints $p \in C_p$ and $\lambda_l, \lambda_h, \nu_l, \nu_h, \mu \in \mathbb{R}_+^{mn}$. The equations $\nu_l \odot \lambda_l = 0$ and $\nu_h \odot \lambda_h = 0$ are understood as component-wise multiplication. The equation $\mu \odot (|p|^2 - 1) = 0$ is understood as $\mu_{i,j}(|p_{i,j}|^2 - 1) = 0$ for each i, j . The expression $\mu \odot p$ is understood as $\mu_{i,j} p_{i,j}$ for each i, j .

Next, we show that the above KKT system can be reduced to a set of equalities. The reduced system has the advantages of

1. not involving matrix-vector product of the form $B\mathbf{x}$ which can be difficult or inaccurate to compute since $B = (K^T K + \alpha I)^{-1}$;
2. no inequalities are involved.

But it involves max and min functions which are semi-smooth. Compared to the optimality conditions of the primal problem (2), the singularity in $\nabla u/|\nabla u|$ is removed [8].

Proposition 2 *The KKT system (10)–(15) (including the constraints $p \in C_p$ and $\lambda_l, \lambda_h, \nu_l, \nu_h \in \mathbb{R}_+^{mn}$) is equivalent to the following system:*

$$0 = F_1(p, u, \lambda) = |\nabla u| \odot p - \nabla u, \quad (16)$$

$$0 = F_2(p, u, \lambda) = -\beta \operatorname{div} p + b - \lambda + Au, \quad (17)$$

$$0 = F_3(p, u, \lambda) = \lambda - \min\{0, \lambda - c(u - h)\} - \max\{0, \lambda - c(u - l)\}, \quad (18)$$

where $A = B^{-1} = K^T K + \alpha I$.

Proof: By setting $u = B(\beta \operatorname{div} p - b + \lambda_l - \lambda_h)$, the system (10)–(15) becomes

$$-\beta \nabla u + \mu \odot p = 0, \quad (19)$$

$$\mu \odot (|p|^2 - 1) = 0, \quad (20)$$

$$Au - (\beta \operatorname{div} p - b + \lambda_l - \lambda_h) = 0, \quad (21)$$

$$(u - l) \odot \lambda_l = 0, \quad (22)$$

$$(h - u) \odot \lambda_h = 0, \quad (23)$$

together with the constraints $u \in C_u$, $p \in C_p$ and $\lambda_l, \lambda_h, \mu \in \mathbb{R}_+^{mn}$. In the third equation (21), we have multiplied both sides with A to cancel B .

The fourth and fifth equations (22)–(23) together with the constraints $u \in C_u$ and $\lambda_l, \lambda_h \in \mathbb{R}_+^{mn}$ imply the following equality:

$$\lambda_l - \lambda_h - \min\{0, \lambda_l - \lambda_h - c(u - h)\} - \max\{0, \lambda_l - \lambda_h - c(u - l)\} = 0,$$

where $c > 0$ is an arbitrary constant. By letting $\lambda = \lambda_l - \lambda_h$, we have

$$\lambda - \min\{0, \lambda - c(u - h)\} - \max\{0, \lambda - c(u - l)\} = 0.$$

From the second equation (20), if $|p_{i,j}| < 1$, then we must have $\mu_{i,j} = 0$, and therefore, $\mu_{i,j} = (\nabla u)_{i,j} = 0$. If $|p_{i,j}| = 1$, then the first equation (20) and $\mu_{i,j} \in \mathbb{R}_+$ implies that $\mu_{i,j} = \beta|(\nabla u)_{i,j}|$. In any case, we have $\mu = \beta|\nabla u|$. Thus, the first and second equations (19)–(20) together with $p \in C_p$ and $\mu \in \mathbb{R}_+^{mn}$ imply that

$$|\nabla u| \odot p - \nabla u = 0.$$

Therefore, we have the system (19)–(23).

By reversing the above arguments, it can be shown that the system (19)–(23) imply the KKT system (10)–(15) together with the inequality constraints. Here, the variables λ_l and λ_h are recovered from λ via $\lambda_l = \max\{\lambda, 0\}$ and $\lambda_h = \max\{-\lambda, 0\}$. We omit the details. \square

When solving the reduced system, an ϵ -regularization is added to the term $|\nabla u|$ in F_1 owing to the possibility that $\nabla u = 0$. The addition of this ϵ -regularization also ensures a unique solution for p , thereby avoiding numerical difficulties in the solution of the system.

2.3 The Primal-dual Active-set Strategy

To solve the system (16)–(18), it is natural to consider Newton’s method to aim for superlinear (or even quadratic) convergence. The Newton’s update for our system is given by the following:

$$\begin{bmatrix} |\nabla u|_\epsilon & -\left(I - \frac{p(\nabla u)^T}{|\nabla u|_\epsilon}\right) \nabla & 0 \\ -\beta \operatorname{div} & A & -I \\ 0 & \frac{\partial F_3}{\partial u} & \frac{\partial F_3}{\partial \lambda} \end{bmatrix} \begin{bmatrix} \delta p \\ \delta u \\ \delta \lambda \end{bmatrix} = - \begin{bmatrix} F_1 \\ F_2 \\ F_3 \end{bmatrix}. \quad (24)$$

Here, $|\nabla u|_\epsilon$ denotes a diagonal matrix such that

$$(|\nabla u|_\epsilon \delta p)_{i,j} = |(\nabla u)_{i,j}|_\epsilon (\delta p)_{i,j},$$

and $\frac{p(\nabla u)^T}{|\nabla u|_\epsilon} \nabla$ denotes a matrix such that

$$\left(\frac{p(\nabla u)^T}{|\nabla u|_\epsilon} \nabla \delta u \right)_{i,j} = \frac{1}{|(\nabla u)_{i,j}|_\epsilon} p_{i,j} (\nabla u)_{i,j}^T (\nabla \delta u)_{i,j}.$$

The function F_3 is strongly semi-smooth and the derivatives $\partial F_2 / \partial u$ and $\partial F_3 / \partial \lambda$ are defined in the sense of slant differentiability which is a generalized derivative, see [12]. We can solve the above system directly to obtain the updates. However, it is non-symmetric and the blocks have diverse characteristics. Thus, it is difficult to construct an effective preconditioner for it. We use an active-set strategy to reduce the system.

For an optimization problem with primal variables $y = (y_i)$ and constraints $y \geq \psi$ where $\psi = (\psi_i)$ are given constants, the primal-dual active-set (PDAS) method [12] divides the variables into two sets, the “active set” \mathcal{A} and the “inactive set” \mathcal{I} . If a primal variable is predicted to be “active” (but y_i is not necessarily ψ_i), then y_i is set to ψ_i in the next iteration. If a primal variable is predicted to “inactive”, then the corresponding dual variable is set to 0 in the next iteration. The remaining variables are updated according to some optimality conditions, usually $\nabla_y L = 0$ where L is the Lagrangian. The effectiveness of such a method depends on the accuracy

of the prediction at each iteration. The prediction scheme we use is based on [12]. This scheme is presented next.

Motivated by the piecewise linearity of F_3 in Eq. (18) and the analysis in [12], it is natural to consider the following definition of active and inactive based on the primal and dual variables:

$$\begin{aligned}\mathcal{I} &:= \{(i, j) : \lambda_{i,j} - c(u_{i,j} - h) \geq 0, \lambda_{i,j} - c(u_{i,j} - l) \leq 0\}, \\ \mathcal{A}_l &:= \{(i, j) : \lambda_{i,j} - c(u_{i,j} - h) \geq 0, \lambda_{i,j} - c(u_{i,j} - l) > 0\}, \\ \mathcal{A}_h &:= \{(i, j) : \lambda_{i,j} - c(u_{i,j} - h) < 0, \lambda_{i,j} - c(u_{i,j} - l) \leq 0\}.\end{aligned}$$

The fourth situation of $\lambda_{i,j} - c(u_{i,j} - h) < 0$ and $\lambda_{i,j} - c(u_{i,j} - l) > 0$ contradicts to the assumption that $l \leq h$. These sets define regions in which F_3 is linear.

Let $D_{\mathcal{I}}$, $D_{\mathcal{A}_l}$ and $D_{\mathcal{A}_h}$ be the down-sampling matrix with relation to the inactive set \mathcal{I} and active sets \mathcal{A}_l and \mathcal{A}_h respectively. Then the components of u in \mathcal{I} , \mathcal{A}_l and \mathcal{A}_h are given by $u_{\mathcal{I}} = D_{\mathcal{I}}u$, $u_{\mathcal{A}_l} = D_{\mathcal{A}_l}u$ and $u_{\mathcal{A}_h} = D_{\mathcal{A}_h}u$ respectively. The components $\lambda_{\mathcal{I}}$, $\lambda_{\mathcal{A}_l}$ and $\lambda_{\mathcal{A}_h}$ can be obtained similarly. We also let $A_{\mathcal{X}} = D_{\mathcal{X}}AD_{\mathcal{X}}^T$ and $A_{\mathcal{X}\mathcal{Y}} = D_{\mathcal{X}}AD_{\mathcal{Y}}^T$ for $\mathcal{X}, \mathcal{Y} \in \{\mathcal{I}, \mathcal{A}_l, \mathcal{A}_h\}$.

We are now in position to derive a simple system to compute the Newton's update in (24). Only $\delta u_{\mathcal{I}}$, the update for $u_{\mathcal{I}}$ requires solving a linear system. Other variables can be computed directly.

Proposition 3 *The Newton's update are given by*

$$\delta \lambda_{\mathcal{I}} = -\lambda_{\mathcal{I}}, \quad (25)$$

$$\delta u_{\mathcal{A}_l} = l - u_{\mathcal{A}_l}, \quad (26)$$

$$\delta u_{\mathcal{A}_h} = h - u_{\mathcal{A}_h}, \quad (27)$$

$$\delta u_{\mathcal{I}} = \left[D_{\mathcal{I}} \left(-\beta \operatorname{div} \frac{1}{|\nabla u|_{\epsilon}} H \nabla + A \right) D_{\mathcal{I}}^T \right]^{-1} g, \quad (28)$$

$$\delta p = \frac{1}{|\nabla u|_{\epsilon}} \{ H \nabla [D_{\mathcal{I}}^T \delta u_{\mathcal{I}} + D_{\mathcal{A}_l}^T (l - u_{\mathcal{A}_l}) + D_{\mathcal{A}_h}^T (h - u_{\mathcal{A}_h})] - F_1 \}, \quad (29)$$

$$\delta \lambda_{\mathcal{A}_l} = D_{\mathcal{A}_l} F_2 - \beta D_{\mathcal{A}_l} \operatorname{div} \delta p + A_{\mathcal{A}_l \mathcal{I}} \delta u_{\mathcal{I}} + A_{\mathcal{A}_l} (l - u_{\mathcal{A}_l}) + A_{\mathcal{A}_l \mathcal{A}_h} (h - u_{\mathcal{A}_h}), \quad (30)$$

$$\delta \lambda_{\mathcal{A}_h} = D_{\mathcal{A}_h} F_2 - \beta D_{\mathcal{A}_h} \operatorname{div} \delta p + A_{\mathcal{A}_h \mathcal{I}} \delta u_{\mathcal{I}} + A_{\mathcal{A}_h \mathcal{A}_l} (l - u_{\mathcal{A}_l}) + A_{\mathcal{A}_h} (h - u_{\mathcal{A}_h}), \quad (31)$$

where

$$\begin{aligned}g &= \beta D_{\mathcal{I}} \operatorname{div} \frac{1}{|\nabla u|_{\epsilon}} \{ H \nabla [D_{\mathcal{A}_l}^T (l - u_{\mathcal{A}_l}) + D_{\mathcal{A}_h}^T (h - u_{\mathcal{A}_h})] - F_1 \} - D_{\mathcal{I}} F_2 \\ &\quad - A_{\mathcal{I} \mathcal{A}_l} (l - u_{\mathcal{A}_l}) - A_{\mathcal{I} \mathcal{A}_h} (h - u_{\mathcal{A}_h}) - \lambda_{\mathcal{I}},\end{aligned}$$

and

$$H = I - \frac{p(\nabla u)^T}{|\nabla u|_{\epsilon}}.$$

Proof: By splitting the variables u and λ into active and inactive sets, the Newton's system (24) is given by

$$\begin{bmatrix} |\nabla u|_{\epsilon} & -H \nabla D_{\mathcal{I}}^T & -H \nabla D_{\mathcal{A}_l}^T & -H \nabla D_{\mathcal{A}_h}^T & 0 & 0 & 0 \\ -\beta D_{\mathcal{I}} \operatorname{div} & A_{\mathcal{I}} & A_{\mathcal{I} \mathcal{A}_l} & A_{\mathcal{I} \mathcal{A}_h} & -I & 0 & 0 \\ -\beta D_{\mathcal{A}_l} \operatorname{div} & A_{\mathcal{A}_l \mathcal{I}} & A_{\mathcal{A}_l} & A_{\mathcal{A}_l \mathcal{A}_h} & 0 & -I & 0 \\ -\beta D_{\mathcal{A}_h} \operatorname{div} & A_{\mathcal{A}_h \mathcal{I}} & A_{\mathcal{A}_h \mathcal{A}_l} & A_{\mathcal{A}_h} & 0 & 0 & -I \\ 0 & 0 & 0 & 0 & I & 0 & 0 \\ 0 & 0 & cI & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & cI & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \delta p \\ \delta u_{\mathcal{I}} \\ \delta u_{\mathcal{A}_l} \\ \delta u_{\mathcal{A}_h} \\ \delta \lambda_{\mathcal{I}} \\ \delta \lambda_{\mathcal{A}_l} \\ \delta \lambda_{\mathcal{A}_h} \end{bmatrix} = - \begin{bmatrix} F_1 \\ D_{\mathcal{I}} F_2 \\ D_{\mathcal{A}_l} F_2 \\ D_{\mathcal{A}_h} F_2 \\ D_{\mathcal{I}} F_3 \\ D_{\mathcal{A}_l} F_3 \\ D_{\mathcal{A}_h} F_3 \end{bmatrix}. \quad (32)$$

We note from Eq. (18) that

$$\begin{aligned} D_{\mathcal{I}}F_3 &= \lambda_{\mathcal{I}}, \\ D_{\mathcal{A}_l}F_3 &= c(u_{\mathcal{A}_l} - l), \\ D_{\mathcal{A}_h}F_3 &= c(u_{\mathcal{A}_h} - h). \end{aligned}$$

Thus, the fifth row in Eq. (32) reads

$$\delta\lambda_{\mathcal{I}} = -D_{\mathcal{I}}F_3 = -\lambda_{\mathcal{I}}.$$

The sixth row in Eq. (32) reads

$$c\delta u_{\mathcal{A}_l} = -D_{\mathcal{A}_l}F_3 = -c(u_{\mathcal{A}_l} - l),$$

and therefore,

$$\delta u_{\mathcal{A}_l} = l - u_{\mathcal{A}_l}.$$

The seventh row Eq. (32) reads

$$c\delta u_{\mathcal{A}_h} = -D_{\mathcal{A}_h}F_3 = -c(u_{\mathcal{A}_h} - h),$$

so that

$$\delta u_{\mathcal{A}_h} = h - u_{\mathcal{A}_h}.$$

Hence, we have proved Eq. (25)–(27).

The third and fourth rows in Eq. (32) read

$$\begin{aligned} -\beta D_{\mathcal{A}_l} \operatorname{div} \delta p + A_{\mathcal{A}_l \mathcal{I}} \delta u_{\mathcal{I}} + A_{\mathcal{A}_l} \delta u_{\mathcal{A}_l} + A_{\mathcal{A}_l \mathcal{A}_h} \delta u_{\mathcal{A}_h} - \delta \lambda_{\mathcal{A}_l} &= -D_{\mathcal{A}_l} F_2, \\ -\beta D_{\mathcal{A}_h} \operatorname{div} \delta p + A_{\mathcal{A}_h \mathcal{I}} \delta u_{\mathcal{I}} + A_{\mathcal{A}_h \mathcal{A}_l} \delta u_{\mathcal{A}_l} + A_{\mathcal{A}_h} \delta u_{\mathcal{A}_h} - \delta \lambda_{\mathcal{A}_h} &= -D_{\mathcal{A}_h} F_2, \end{aligned}$$

respectively. They give Eq. (30)–(31) immediately after replacing $\delta u_{\mathcal{A}_l}$ with $l - u_{\mathcal{A}_l}$ and $\delta u_{\mathcal{A}_h}$ with $h - u_{\mathcal{A}_h}$.

The first row in Eq. (32) reads

$$|\nabla u|_{\epsilon} \delta p = H \nabla \left[D_{\mathcal{I}}^T \delta u_{\mathcal{I}} + D_{\mathcal{A}_l}^T \delta u_{\mathcal{A}_l} + D_{\mathcal{A}_h}^T \delta u_{\mathcal{A}_h} \right] - F_1.$$

Notice that $|\nabla u|_{\epsilon}$ is a diagonal matrix with inverse $1/|\nabla u|_{\epsilon}$. We obtain Eq. (29).

Finally, the second row in Eq. (32) reads

$$-\beta D_{\mathcal{I}} \operatorname{div} \delta p + A_{\mathcal{I}} \delta u_{\mathcal{I}} + A_{\mathcal{I} \mathcal{A}_l} \delta u_{\mathcal{A}_l} + A_{\mathcal{I} \mathcal{A}_h} \delta u_{\mathcal{A}_h} - \delta \lambda_{\mathcal{I}} = -D_{\mathcal{I}} F_2.$$

Eliminating δp from the above equation using Eq. (29), we obtain (28). \square

We note that this proposition shows that the next value of $\lambda_{\mathcal{I}}$, $u_{\mathcal{A}_l}$ and $u_{\mathcal{A}_h}$ is 0, l and h respectively. This is exactly the same as the PDAS algorithm presented in [12]. Thus our method can be thought of a PDAS algorithm.

The linear system (28) is non-symmetric and so CG cannot be applied. Chan *et al.* [8] proposed to symmetrize the matrix so that CG could be applied. Following this suggestion, we symmetrize the system as:

$$D_{\mathcal{I}} \left[-\beta \operatorname{div} \frac{1}{|\nabla u|_{\epsilon}} \left(I - \frac{p(\nabla u)^T + (\nabla u)p^T}{2|\nabla u|_{\epsilon}} \right) \nabla + A \right] D_{\mathcal{I}}^T \delta u_{\mathcal{I}} = g. \quad (33)$$

In the next proposition, we show that this linear system is symmetric positive definite. In Section 2.5, we will discuss the use preconditioners for solving this system. We may regard the resulting iteration as a quasi-Newton's method with the coefficient matrix of the quasi-Newton's equation given by:

$$\begin{aligned}
& \begin{bmatrix} |\nabla u|_\epsilon & -\frac{1}{2}(H + H^T)\nabla D_{\mathcal{I}}^T & -H\nabla D_{\mathcal{A}_l}^T & -H\nabla D_{\mathcal{A}_h}^T & 0 & 0 & 0 \\ -\beta D_{\mathcal{I}}\text{div} & A_{\mathcal{I}} & A_{\mathcal{I}\mathcal{A}_l} & A_{\mathcal{I}\mathcal{A}_h} & -I & 0 & 0 \\ -\beta D_{\mathcal{A}_l}\text{div} & A_{\mathcal{A}_l\mathcal{I}} & A_{\mathcal{A}_l} & A_{\mathcal{A}_l\mathcal{A}_h} & 0 & -I & 0 \\ -\beta D_{\mathcal{A}_h}\text{div} & A_{\mathcal{A}_h\mathcal{I}} & A_{\mathcal{A}_h\mathcal{A}_l} & A_{\mathcal{A}_h} & 0 & 0 & -I \\ 0 & 0 & 0 & 0 & I & 0 & 0 \\ 0 & 0 & cI & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & cI & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \delta p \\ \delta u_{\mathcal{I}} \\ \delta u_{\mathcal{A}_l} \\ \delta u_{\mathcal{A}_h} \\ \delta \lambda_{\mathcal{I}} \\ \delta \lambda_{\mathcal{A}_l} \\ \delta \lambda_{\mathcal{A}_h} \end{bmatrix} \\
& = - \begin{bmatrix} F_1 \\ D_{\mathcal{I}}F_2 \\ D_{\mathcal{A}_l}F_2 \\ D_{\mathcal{A}_h}F_2 \\ D_{\mathcal{I}}F_3 \\ D_{\mathcal{A}_l}F_3 \\ D_{\mathcal{A}_h}F_3 \end{bmatrix}. \tag{34}
\end{aligned}$$

The convergence of the quasi-Newton's method is studied in Section 3.

Proposition 4 *The system (33) is symmetric positive definite if $|p_{i,j}| \leq 1$ for all i, j .*

Proof: Let

$$M = \frac{1}{|\nabla u|_\epsilon} \left(I - \frac{p(\nabla u)^T + (\nabla u)p^T}{2|\nabla u|_\epsilon} \right),$$

and let $y = (y_{i,j}) \in \mathbb{R}^{2mn}$. Clearly, M is symmetric since it is the symmetrization of the matrix

$$\frac{1}{|(\nabla u)|_\epsilon} \left(I - \frac{p(\nabla u)^T}{|(\nabla u)|_\epsilon} \right).$$

Note that by definition of M

$$\begin{aligned}
y^T M y &= \sum_{i,j} y_{i,j}^T \frac{1}{|(\nabla u)_{i,j}|_\epsilon} \left(I - \frac{p_{i,j}(\nabla u)_{i,j}^T + (\nabla u)_{i,j}p_{i,j}^T}{2|(\nabla u)_{i,j}|_\epsilon} \right) y_{i,j} \\
&= \sum_{i,j} \frac{1}{|(\nabla u)_{i,j}|_\epsilon} \left(|y_{i,j}|^2 - \frac{y_{i,j}^T p_{i,j}(\nabla u)_{i,j}^T y_{i,j} + y_{i,j}^T p_{i,j}(\nabla u)_{i,j}^T y_{i,j}}{2|(\nabla u)_{i,j}|_\epsilon} \right) \\
&= \sum_{i,j} \frac{1}{|(\nabla u)_{i,j}|_\epsilon} \left(|y_{i,j}|^2 - \frac{y_{i,j}^T p_{i,j}(\nabla u)_{i,j}^T y_{i,j}}{|(\nabla u)_{i,j}|_\epsilon} \right) \\
&\geq \sum_{i,j} \frac{1}{|(\nabla u)_{i,j}|_\epsilon} \left(|y_{i,j}|^2 - \frac{|y_{i,j}||p_{i,j}||(\nabla u)_{i,j}||y_{i,j}|}{|(\nabla u)_{i,j}|_\epsilon} \right) \geq 0.
\end{aligned}$$

The last inequality is due to $|p_{i,j}| \leq 1$ and $|(\nabla u)_{i,j}| \leq |(\nabla u)_{i,j}|_\epsilon$. Thus, M is symmetric positive semi-definite. Since $A = K^T K + \alpha I$ is symmetric positive definite, so is $\beta \nabla^T M \nabla + A$. Finally, The principal submatrix $D_{\mathcal{I}}(\nabla^T M \nabla + A)D_{\mathcal{I}}^T$ of a symmetric positive definite matrix is symmetric positive definite. \square

2.4 The BCGM Algorithm

The full BCGM algorithm for image deblurring and segmentation can now be described as follows:

1. Initialize p^0, u^0, λ^0 and set $k = 0$
2. Set the active and inactive sets:

$$\mathcal{I}^k = \{(i, j) : \lambda_{i,j}^k - c(u_{i,j}^k - h) \geq 0, \lambda_{i,j}^k - c(u_{i,j}^k - l) \leq 0\}$$

$$\mathcal{A}_l^k = \{(i, j) : \lambda_{i,j}^k - c(u_{i,j}^k - h) \geq 0, \lambda_{i,j}^k - c(u_{i,j}^k - l) > 0\}$$

$$\mathcal{A}_h^k = \{(i, j) : \lambda_{i,j}^k - c(u_{i,j}^k - h) < 0, \lambda_{i,j}^k - c(u_{i,j}^k - l) \leq 0\}$$
3. Compute $\delta u_{\mathcal{I}}^k$ by solving the linear system (33) using PCG
4. Compute $\delta p^k, \delta \lambda_{\mathcal{A}_l}^k$ and $\delta \lambda_{\mathcal{A}_h}^k$ by applying (29)–(31)
5. Compute the step size τ :

$$\tau = \rho \sup_{\gamma > 0} \{ |p_{i,j}^k + \gamma \delta p_{i,j}^k| \leq 1 \quad \forall i, j \}$$
6. Update the variables:

$$p^{k+1} = p^k + \tau \delta p^k$$

$$u_{\mathcal{I}}^{k+1} = u_{\mathcal{I}}^k + \delta u_{\mathcal{I}}^k$$

$$u_{\mathcal{A}_l}^{k+1} = l$$

$$u_{\mathcal{A}_h}^{k+1} = h$$

$$\lambda_{\mathcal{I}}^{k+1} = 0$$

$$\lambda_{\mathcal{A}_l}^{k+1} = \lambda_{\mathcal{A}_l}^k + \delta \lambda_{\mathcal{A}_l}^k$$

$$\lambda_{\mathcal{A}_h}^{k+1} = \lambda_{\mathcal{A}_h}^k + \delta \lambda_{\mathcal{A}_h}^k$$
7. Check for convergence. Stop if converges; otherwise set $k = k + 1$ and go to Step 2

To check convergence, we use the residual of the optimality system (16)–(18):

$$r^k = \sqrt{\|F_1(p^k, u^k, \lambda^k)\|^2 + \|F_2(p^k, u^k, \lambda^k)\|^2 + \|F_3(p^k, u^k, \lambda^k)\|^2}.$$

Although the system (16)–(18) is not the original KKT system (10)–(15), we shall still call its residual r^k the KKT residual for convenience. We terminate the iterations if the relative KKT residual r^{k+1}/r^0 is less than a predefined tolerance.

A line search is only required in Step 5, for p . This is to ensure positive definiteness of Eq. (33). The algorithm requires the specification of several parameters. The parameter c is used to determine the active and inactive sets at every iteration, see Step 3 above. As shown in [12] the performance of the algorithm is largely independent of its value. Thus, we simply fix c to be 10^4 . The parameter ρ is merely used to make the step size a little conservative. Setting it to 0.99 worked for all our numerical tests. The parameter ϵ is to be selected at a reasonably small value to achieve a trade-off between reconstruction error and time for convergence. In general, the choice of ϵ may be resolution dependent, but we found that setting it to 10^{-2} provided a good trade-off between quality and speed for all the cases that we tested. Reducing it further did not significantly reduce the reconstruction error.

The regularization parameter β decides the trade-off between the reconstruction error and noise amplification. It is a part of the deblurring model, rather than our algorithm. The value of β must be selected carefully for any TV deblurring algorithm. Section 3 provides a (local) convergence analysis of the BCGM algorithm based on the quasi-Newton approach of Eq. (33). The numerical results of Section 4 show a locally superlinear rate of convergence.

2.5 Preconditioners

The most computationally intensive step of the BCGM algorithm is Step 3 which involves solving the linear system in Eq. (33). Though significantly smaller than the original linear system (24) obtained by linearizing Eq. (16)–(18), it is still a large system. We determined that the Factorized Banded Inverse Preconditioner (FBIP) [17] worked well to speed up the solution of the linear system. Using the FBIP preconditioner to solve the linear system requires essentially $O(N \log N)$ operations, where N is the total number of pixels in the image. This is including the use of FFT's for computations involving the matrix $A = K^T K + \alpha I$.

The original system Eq. (24) has different characteristics in each of its blocks. It is therefore harder to construct an effective preconditioner. In contrast, the reduced system Eq. (33) has a simpler structure so that standard preconditioners work well.

3 Convergence Analysis

In this section, we prove the locally superlinear convergence of the quasi-Newton's iteration (34) to the unique solution of the KKT system (16)–(18).

Let k be the iteration number. We denote the system (16)–(18) by $F(p, u, \lambda) = 0$ and denote the residual at iteration k by $F^k = F(p^k, u^k, \lambda^k)$. We first note that the system $F = 0$ has a solution $x^* = (p^*, u^*, \lambda^*)$ because the original primal problem (6) (with the regularized TV norm) has a solution. Moreover, the solution to $F = 0$ is unique because the primal problem is strictly convex and the dual variable p^* is uniquely determined by u^* via $p^* = \nabla u^* / |\nabla u^*|_\epsilon$.

We denote the Newton's system (32) and quasi-Newton's system (34) by

$$W_k \delta^k = -F^k$$

and

$$V_k \delta^k = -F^k$$

respectively. We will use Theorems 4.1 and 4.2 from [24] to prove the superlinear convergence. For convenience, we combine the two theorems and restate them here:

Theorem 1 *Suppose that $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a locally Lipschitzian function in an open convex set $D \subset \mathbb{R}^n$ and $x^* \in D$ is a solution of $F(x) = 0$. Suppose that F is semi-smooth at x^* and all $W \in \partial_b F(x^*)$ are non-singular. Let β be a positive constant such that $\|W^{-1}\| \leq \beta$ for all $W \in \partial_b F(x^*)$. Let $\Delta = 1/(6\beta)$. Then there exists a $\gamma_1 > 0$ such that for any $x_0 \in D \cap B_{\gamma_1}(x^*)$ the sequence generated by quasi-Newton's iteration $x^{k+1} = x^k - V_k^{-1} F(x^k)$ with $V_k \in \{V \in B_\Delta(W_k) : W_k \in \partial_b F(x^k)\}$ is well-defined (i.e. V_k is non-singular for each k) and converges to x^* linearly.*

Furthermore, if there exists a sequence $\{W_k\}$ with $W_k \in \partial_b F(x^k)$ for each k such that

$$\lim_{k \rightarrow \infty} \frac{\|(V_k - W_k)(x^{k+1} - x^k)\|}{\|x^{k+1} - x^k\|} = 0,$$

then the convergence is superlinear.

Here $\partial_b F(x)$ is defined as

$$\partial_b F(x) = \lim_{\substack{x^k \rightarrow x \\ x^k \in D_F}} F'(x^k), \quad (35)$$

D_F is the set where F is differentiable, $F'(x)$ is the Jacobian of F at a point $x \in D_F$, and $B_\gamma(x)$ is the closed ball centered at x with radius γ . Regarding the constant β , it has been

shown in [24, Lemma 2.2] that if all $W \in \partial_b F(x^*)$ are non-singular, then $\|W^{-1}\|_2 \leq \beta$ for all $W \in \partial_b F(x^*)$ for some $\beta > 0$. Moreover, there exists a neighborhood $N(x^*)$ of x^* such that $\|V^{-1}\| \leq (10/9)\beta$ for all $V \in \partial_b F(x)$ for all $x \in N(x^*)$.

In view of the above theorem, to establish the superlinear convergence of the quasi-Newton's iteration, it suffices to verify that the assumptions in the theorem are satisfied. More precisely, we show in the next five lemmas that the followings hold:

1. F is locally Lipschitzian;
2. F is semi-smooth at $x^* = (p^*, u^*, \lambda^*)$;
3. all $W \in \partial_b F(x^*)$ are non-singular;
4. $V_k \in \{V \in B_\Delta(W_k) : W_k \in \partial_b F(x^k)\}$ for all k ;
5. $\|(V_k - W_k)(x^{k+1} - x^k)\|/\|x^{k+1} - x^k\| \rightarrow 0$ as $k \rightarrow \infty$.

Lemma 1 *The function F is locally Lipschitzian.*

Proof: Both F_1 (regularized) and F_2 are differentiable everywhere. Thus they are locally Lipschitzian. The F_3 is a continuous piecewise linear function which is locally Lipschitzian. Thus, $F = (F_1, F_2, F_3)$ is locally Lipschitzian. \square

Lemma 2 *The function F is semi-smooth everywhere.*

Proof: Since the components of F_1 and F_2 are differentiable everywhere, they are semi-smooth everywhere. The pointwise maximum or minimum of a compact family of continuously differentiable functions is semi-smooth [19]. The sum of semi-smooth functions is semi-smooth. Thus, the components of F_3 are semi-smooth everywhere. Finally, a vector-valued function is semi-smooth if and only if all of its components are semi-smooth. Hence, F is semi-smooth everywhere. \square

Lemma 3 *All $W \in \partial_b F(x^*)$ are non-singular.*

Proof: We first compute $\partial_b F(x^*)$. For the sake of simplicity, we assume $l < h$. The case $l = h$ can be handled easily. The functions F_1 and F_2 are differentiable, and therefore, their derivatives with respect to p, u, λ are well-defined and unique. The function F_3 is independent of p and is piecewise linear with respect to u and λ . Consider a typical component of F_3 :

$$f(u, \lambda) = \lambda - \min\{0, \lambda - c(u - h)\} - \max\{0, \lambda - c(u - l)\}$$

where $u \in \mathbb{R}$ and $\lambda \in \mathbb{R}$. The classical derivative exists in the regions $\{\lambda < c(u - h)\}$, $\{\lambda > c(u - l)\}$ and $\{c(u - h) < \lambda < c(u - l)\}$. In fact, we have

$$\left(\frac{\partial f}{\partial u}, \frac{\partial f}{\partial \lambda} \right) = \begin{cases} (c, 0) & \text{if } \lambda < c(u - h), \\ (0, 1) & \text{if } c(u - h) < \lambda < c(u - l), \\ (c, 0) & \text{if } \lambda > c(u - l). \end{cases}$$

When $\lambda = c(u - h)$ or $\lambda = c(u - l)$, by taking one-sided limits, it is seen that the derivative can be chosen to be either $(c, 0)$ or $(0, 1)$.

Coming back to F_3 , if d number of components of (u^*, λ^*) satisfy $\lambda_{i,j}^* = c(u_{i,j}^* - h)$ or $\lambda_{i,j}^* = c(u_{i,j}^* - l)$, then each corresponding pair $(\partial F_3/\partial u_{i,j}, \partial F_3/\partial \lambda_{i,j})$ has two possible values and thus $\partial_b F(x^*)$ contains 2^d elements.

In any case, each $W \in \partial_b F(x^*)$ has exactly the form in Eq. (32) but the definition of the active and inactive sets need to be adjusted according to the chosen value of $(\partial F_3/\partial u_{i,j}, \partial F_3/\partial \lambda_{i,j})$. Indeed, we have

$$\begin{aligned} \mathcal{I} &:= \{(i, j) : \lambda_{i,j} - c(u_{i,j} - h) > 0, \lambda_{i,j} - c(u_{i,j} - l) < 0\} \\ &\cup \left\{ (i, j) : \lambda_{i,j} - c(u_{i,j} - h) = 0, \left(\frac{\partial F_3}{\partial u_{i,j}}, \frac{\partial F_3}{\partial \lambda_{i,j}} \right) = (0, 1) \right\} \\ &\cup \left\{ (i, j) : \lambda_{i,j} - c(u_{i,j} - l) = 0, \left(\frac{\partial F_3}{\partial u_{i,j}}, \frac{\partial F_3}{\partial \lambda_{i,j}} \right) = (0, 1) \right\}, \\ \mathcal{A}_l &:= \{(i, j) : \lambda_{i,j} - c(u_{i,j} - h) \geq 0, \lambda_{i,j} - c(u_{i,j} - l) > 0\} \\ &\cup \left\{ (i, j) : \lambda_{i,j} - c(u_{i,j} - l) = 0, \left(\frac{\partial F_3}{\partial u_{i,j}}, \frac{\partial F_3}{\partial \lambda_{i,j}} \right) = (c, 0) \right\}, \\ \mathcal{A}_h &:= \{(i, j) : \lambda_{i,j} - c(u_{i,j} - h) < 0, \lambda_{i,j} - c(u_{i,j} - l) \leq 0\} \\ &\cup \left\{ (i, j) : \lambda_{i,j} - c(u_{i,j} - h) = 0, \left(\frac{\partial F_3}{\partial u_{i,j}}, \frac{\partial F_3}{\partial \lambda_{i,j}} \right) = (c, 0) \right\}. \end{aligned}$$

Based on this splitting into active and inactive sets, we can use the same proof for Proposition 3 to show that the equation $W\delta = -F$ can be reduced to Eq. (25)–Eq. (31). Thus, W is non-singular if and only if the matrix in Eq. (28) is non-singular. To see this, we let

$$M = -D_{\mathcal{I}} \beta \operatorname{div} \frac{1}{|\nabla u^*|_{\epsilon}} \left(I - \frac{p^*(\nabla u^*)^T}{|\nabla u^*|_{\epsilon}} \right) \nabla D_{\mathcal{I}}^T.$$

The coefficient matrix in Eq. (28) is thus given by $M + D_{\mathcal{I}} A D_{\mathcal{I}}^T$. If \mathcal{I} is empty, then W is non-singular since the equations (25)–(27) and (29)–(31) define the unique solution. We assume that \mathcal{I} is non-empty.

Using the fact that $|p_{i,j}^*| = |\nabla u^*|/|\nabla u^*|_{\epsilon} \leq 1$ for each i, j , it can be shown that $y^T M y \geq 0$ for all y . Since A is symmetric positive definite, the principal submatrix $D_{\mathcal{I}} A D_{\mathcal{I}}^T$ is also symmetric positive definite. Thus, $y^T D_{\mathcal{I}} A D_{\mathcal{I}}^T y > 0$ for all $y \neq \mathbf{0}$. Suppose that $y \neq \mathbf{0}$ is in the null space of the coefficient matrix in (28). Then, we have $M y + D_{\mathcal{I}} A D_{\mathcal{I}}^T y = \mathbf{0}$. Thus, $y^T M y + y^T D_{\mathcal{I}} A D_{\mathcal{I}}^T y = 0$. But this contradicts to $y^T D_{\mathcal{I}} A D_{\mathcal{I}}^T y > 0$ and $y^T M y \geq 0$. Hence the coefficient matrix $M + D_{\mathcal{I}} A D_{\mathcal{I}}^T$ is non-singular. This implies $W \in \partial_b F(x^*)$ is also non-singular. \square

Lemma 4 *For each $x = (p, u, \lambda)$, let $V = V(x)$ be the coefficient matrix in the quasi-Newton's equation (34). Then, there exists a $\gamma_2 > 0$ such that for each $x \in B_{\gamma_2}(x^*)$ we have $V \in \{V \in B_{\Delta}(W) : W \in \partial_b F(x)\}$.*

Proof: For each $x = (p, u, \lambda)$, we let W be the coefficient matrix in the Newton's equation (32). Then we have $W \in \partial_b F(x)$. It remains to show that $\|V - W\| \leq \Delta$ when x is close enough to x^* .

We first note that $p_{i,j}^* = (\nabla u^*)_{i,j}/|(\nabla u^*)_{i,j}|_{\epsilon}$ so that $|p_{i,j}^*| < 1$ for all i, j . Thus, there exists a $\xi > 0$ such that $|p_{i,j}| < 1$ for all i, j for $x \in B_{\xi}(x^*)$. In this rest of the proof, we assume that $x \in B_{\xi}(x^*)$. Consider the continuously differentiable function $f(u) = \nabla u/|\nabla u|_{\epsilon}$. Using the Mean

Value Theorem, we have

$$\begin{aligned}
\left\| \frac{\nabla u}{|\nabla u|_\epsilon} - \frac{\nabla u^*}{|\nabla u^*|_\epsilon} \right\| &= \|f(u) - f(u^*)\| \\
&\leq \sup_v \|f'(v)\| \|u - u^*\| \\
&= \sup_v \left\| \frac{1}{|\nabla v|_\epsilon} \left(I - \frac{(\nabla v)(\nabla v)^T}{|\nabla v|_\epsilon^2} \right) \nabla \right\| \|u - u^*\| \\
&= \sup_v \left\| \frac{1}{|\nabla v|_\epsilon} \right\| \left\| I - \frac{(\nabla v)(\nabla v)^T}{|\nabla v|_\epsilon^2} \right\| \|\nabla\| \|u - u^*\| \\
&\leq \frac{1}{\sqrt{\epsilon}} \cdot 1 \cdot \sqrt{8} \|u - u^*\|.
\end{aligned}$$

Here, $\|\nabla\| \leq \sqrt{\|\nabla\|_1 \|\nabla\|_\infty} = \sqrt{8}$. Thus we have

$$\begin{aligned}
\left\| p - \frac{\nabla u}{|\nabla u|_\epsilon} \right\| &\leq \|p - p^*\| + \left\| \frac{\nabla u}{|\nabla u|_\epsilon} - \frac{\nabla u^*}{|\nabla u^*|_\epsilon} \right\| + \left\| p^* - \frac{\nabla u^*}{|\nabla u^*|_\epsilon} \right\| \\
&\leq \|p - p^*\| + \sqrt{\frac{8}{\epsilon}} \|u - u^*\| \\
&\leq \sqrt{1 + \left(\frac{8}{\epsilon}\right)^2} \|x - x^*\|.
\end{aligned} \tag{36}$$

From (32), (34) and the facts that $\|\nabla\| \leq \sqrt{8}$ and $\|D_{\mathcal{I}}^T\| = 1$, we have

$$\|V - W\| = \left\| \frac{1}{2}(H - H^T)\nabla D_{\mathcal{I}}^T \right\| \leq \sqrt{2} \|H - H^T\|. \tag{37}$$

Here we assumed that \mathcal{I} is non-empty. If \mathcal{I} is empty, then $V = W$ so that assertion of the lemma is trivial. Let $y \neq 0$. Then,

$$\begin{aligned}
\|(H - H^T)y\|^2 &= \sum_{i,j} |(H - H^T)y_{i,j}|^2 \\
&= \sum_{i,j} \left| \frac{p_{i,j}(\nabla u)_{i,j}^T y_{i,j}}{|\nabla u_{i,j}|_\epsilon} - \frac{(\nabla u)_{i,j} p_{i,j}^T y_{i,j}}{|\nabla u_{i,j}|_\epsilon} \right|^2 \\
&= \sum_{i,j} \left| \left[p_{i,j} - \frac{(\nabla u)_{i,j}}{|\nabla u_{i,j}|_\epsilon} \right] \frac{(\nabla u)_{i,j}^T y_{i,j}}{|\nabla u_{i,j}|_\epsilon} + \frac{(\nabla u)_{i,j}}{|\nabla u_{i,j}|_\epsilon} \left[\frac{(\nabla u)_{i,j}}{|\nabla u_{i,j}|_\epsilon} - p_{i,j} \right]^T y_{i,j} \right|^2 \\
&\leq 2 \sum_{i,j} \left| \left[p_{i,j} - \frac{(\nabla u)_{i,j}}{|\nabla u_{i,j}|_\epsilon} \right] \frac{(\nabla u)_{i,j}^T y_{i,j}}{|\nabla u_{i,j}|_\epsilon} \right|^2 + \left| \frac{(\nabla u)_{i,j}}{|\nabla u_{i,j}|_\epsilon} \left[\frac{(\nabla u)_{i,j}}{|\nabla u_{i,j}|_\epsilon} - p_{i,j} \right]^T y_{i,j} \right|^2 \\
&\leq 4 \sum_{i,j} \left| p_{i,j} - \frac{(\nabla u)_{i,j}}{|\nabla u_{i,j}|_\epsilon} \right|^2 |y_{i,j}|^2 \\
&\leq 4(1 + 64\epsilon^{-2}) \|x - x^*\|^2 \|y\|^2.
\end{aligned} \tag{38}$$

The last inequality follows from (36). Therefore, combining (37) and (38), we have

$$\|V - W\| \leq 2\sqrt{2}\sqrt{1 + 64\epsilon^{-2}} \|x - x^*\| \tag{39}$$

for all $x \in B_\xi(x^*)$. Let

$$\gamma_2 = \min \left\{ \xi, \frac{\Delta}{2\sqrt{2}\sqrt{1+64\epsilon^{-2}}} \right\}.$$

Then $\|V - W\| \leq \Delta$ for all $x \in B_{\gamma_2}(x^*)$. \square

Lemma 5 *Let $\{x^k\}$ be a sequence converging to x^* , the solution to $F = 0$. Let W_k and V_k be the coefficient matrix of the Newton's and quasi-Newton's equation in (32) and (34) respectively. Then,*

$$\lim_k \frac{\|(V_k - W_k)(x^{k+1} - x^k)\|}{\|x^{k+1} - x^k\|} = 0.$$

Proof: From (39), we have

$$\|V_k - W_k\| \leq 4 \max\{1, \sqrt{8\epsilon^{-1}}\} \|x^k - x^*\|$$

for all k large enough so that $x^k \in B_{\gamma_2}(x^*)$. Thus $\|V_k - W_k\| \rightarrow 0$ as $k \rightarrow \infty$. \square

We are now in position to state the convergence theorem for the quasi-Newton's iteration (34).

Theorem 2 *The quasi-Newton's iteration (34) converges locally superlinearly.*

Proof: By Lemmas 1–4 and the first part of Theorem 1, the quasi-Newton's iteration converges linearly in the region $B_\gamma(x^*)$ where $\gamma = \min\{\gamma_1, \gamma_2\}$. Then by Lemma 5 and the second part of Theorem 1, the convergence is superlinear. \square

We remark that Theorem 1 only applies to the quasi-Newton's iteration without a line search. However, our BCGM algorithm uses a line search on p to guarantee $|p_{i,j}| \leq 1$ for all i, j . Fortunately, the exact solution satisfies the strict inequality $|p_{i,j}^*| < 1$ so that the line search becomes obsolete as the iterates approach the optimal solution. Therefore, the local convergence theory can be applied.

4 Numerical Results

4.1 Deblurring

In this subsection, we present numerical results to demonstrate the performance of the BCGM algorithm for solving deblurring problems.

We compare the performance of BCGM with a primal-only Projected Newton's (PN) algorithm. The PN algorithm is based on that presented in [1]. At each outer iteration, active and inactive sets are identified based on the primal variable u . Then a Newton step is taken for the inactive variables whereas a projected steepest descent step is taken for the active ones. A line search ensures that the step size taken in the inactive variables is such that they do not violate the bilateral constraints. A few parameters have to be modified to tune the line search. The method is quite slow, for only a few inactive variables (usually as little as one) are updated at each step. Active variables which are already at the boundary of the feasible set, cannot be updated. Theoretically, once all the active variables are identified, the convergence is quadratic. However, it takes many iterations to find all the active variables. In all our experiments, a quadratic convergence has not been observed within the limit of 300 iterations.

Fig. 1 compares unconstrained, non-negatively constrained and bilaterally constrained deblurring for an artificially generated binary checkerboard image. This is an extreme example constructed for the purpose of demonstrating the usefulness of bilateral constraints. The original image consists of values of either 0 or 255. Therefore, the values of l and h were set to 0 and 255 respectively. The unconstrained reconstruction Fig. 1(c) is derived by resetting the solution u to $\max\{\min\{u, h\}, l\}$. The non-negatively constrained result was derived by first solving the problem using the algorithm in [15], followed by resetting u to $\min\{u, h\}$. It is seen that the number of spurious oscillations is least for the bilaterally constrained solution. This is also seen from the significantly higher PSNR (Peak Signal-to-Noise Ratio) of the bilaterally constrained solution. The histograms of the deblurred images also show that the recovered data is closely clustered around 0 and 255 for the bilaterally constrained deblurring, but closely clustered only around 0 for the nonnegatively constrained deblurring. For the unconstrained deblurring, the data is much less closely clustered around 0 or 255.

Fig. 2 shows the performance of BCGM in recovering a blurred and noisy image of text. The bilateral constraints are useful in recovering a readable version of the text.

Fig. 3(a) compares the convergence of the BCGM and PN methods for the Text image of Fig. 2. The image is degraded with a Gaussian blur of size 5×5 and a SNR of 20dB and 30dB respectively. The value of β is fixed at 0.2. It is seen that the progress of the PN method slows down significantly after around 50 iterations. For the BCGM method, the CPU times for convergence are 155 seconds (20dB) and 167 seconds (30dB). For the PN method, the CPU times for 300 iterations are 280 seconds (20dB) and 282 seconds (30dB).

Fig. 3(b) demonstrates the robustness of the BCGM method for small values of the regularization parameter ϵ . The number of outer iterations increases with decreasing ϵ , but not by a large amount.

Next, we numerically demonstrate the locally superlinear convergence of the BCGM algorithm. Define

$$q_k = \frac{\|u_{k+1} - u^*\|}{\|u_k - u^*\|}$$

where u_k refers to the primal variable u at iteration k , and u^* is the pre-computed solution at convergence. Fig. 3(c) shows the values of q_k for the Text image with a Gaussian blur of size 5×5 and noise of SNR 30dB. It is seen $q_k \rightarrow 0$ in the last 16 iterations out of a total of 33 iterations. Thus a superlinear convergence is achieved. This confirms the results of Section 3.

4.2 Segmentation

In this subsection, we demonstrate the usefulness of our algorithm for solving segmentation problems and compare our algorithm with the primal and dual gradient descent (GD) algorithms of [7] and [3] respectively. The constants c_1 and c_2 are updated after every 100 iterations for the primal and dual GD algorithms and are updated after every iteration for the BCGM algorithm. In all cases, we set $\beta = 0.4$.

Fig. 4 presents the segmentation results of the BCGM algorithm. It is observed that the solution u is very close to a binary function. The final segmentation is obtained by thresholding u with $\mu = 0.5$.

Fig. 5 presents the segmentation results of the BCGM algorithm with respect to various thresholds. We observe that the segmentation is very robust to the threshold due to the fast convergence of the proposed algorithm to a close-to-binary function.

Fig. 6 show the convergence profile of various methods in terms of the relative residual of the respective optimality conditions. Fig. 6(a) and (b) show that the convergence of the BCGM

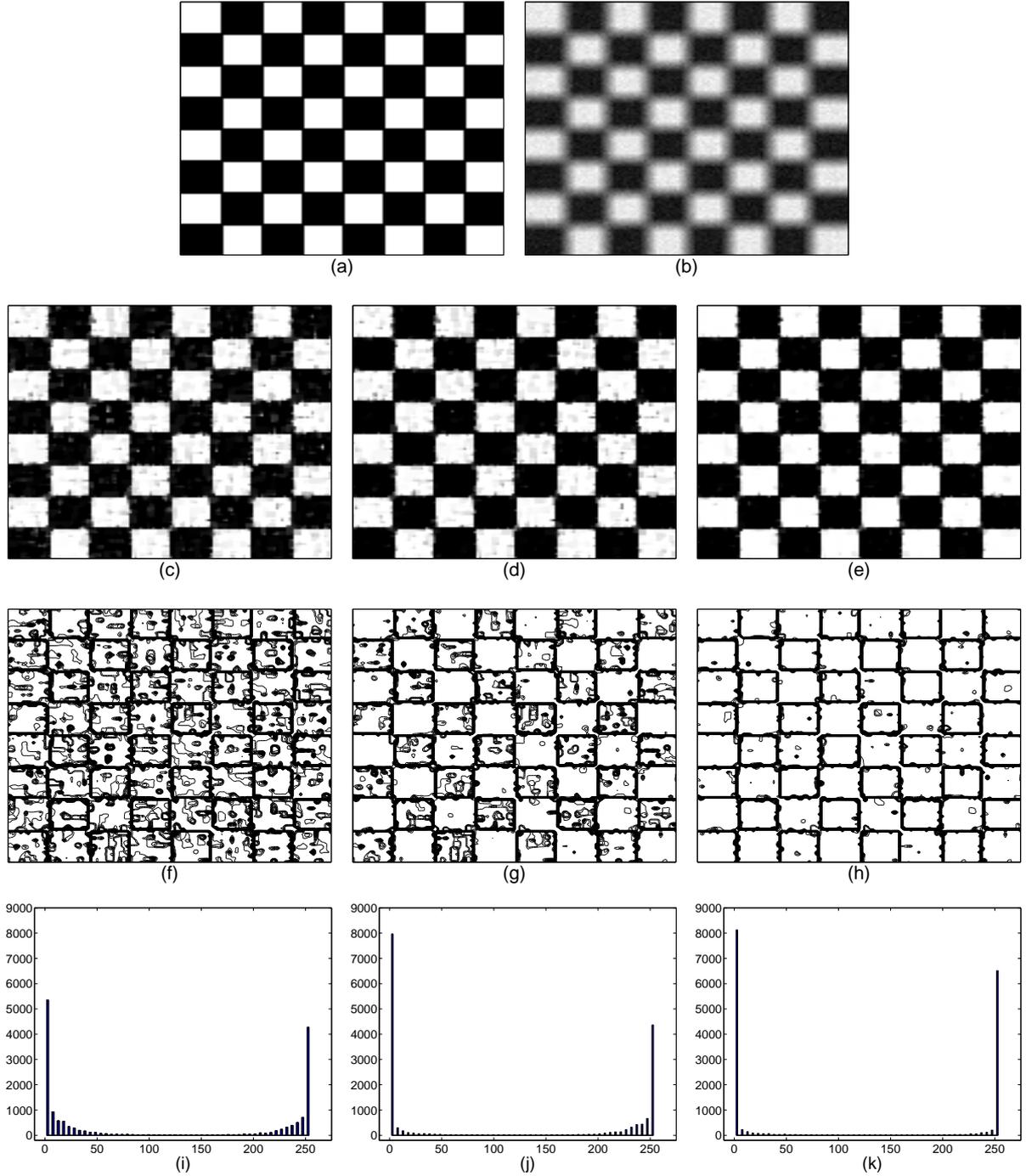


Figure 1: Comparison of bilaterally constrained, non-negatively constrained, and unconstrained deblurring. The value of β is set to 0.2. (a) Original Checker image (128×128). (b) Blurred and noisy data with a Gaussian PSF of size 7×7 and noise of SNR 25dB. (c) Unconstrained deblurring followed by clipping out-of-bound components. PSNR = 19.18dB. (d) Non-negatively constrained deblurring followed by clipping out-of-bound components. PSNR = 20.33dB. (e) Bilaterally constrained deblurring. PSNR = 24.78dB. (f) Contour plot of the result in (c). (g) Contour plot of the result in (d). (h) Contour plot of the result in (e). (i) Histogram of the result in (c). (j) Histogram of the result in (d). (k) Histogram of the result in (e).

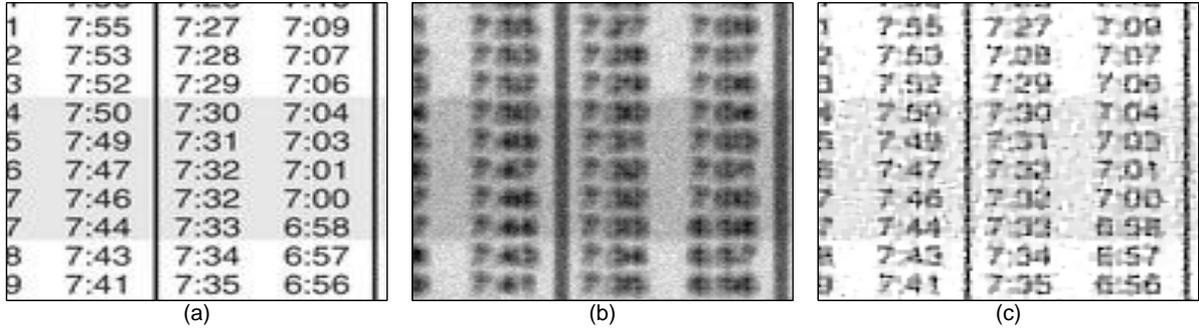


Figure 2: (a) Original Text image (128×128). (b) Degraded with a Gaussian PSF of size 5×5 and noise of SNR 30dB. (c) TV deblurring results with BCGM with $\beta = 0.2$.

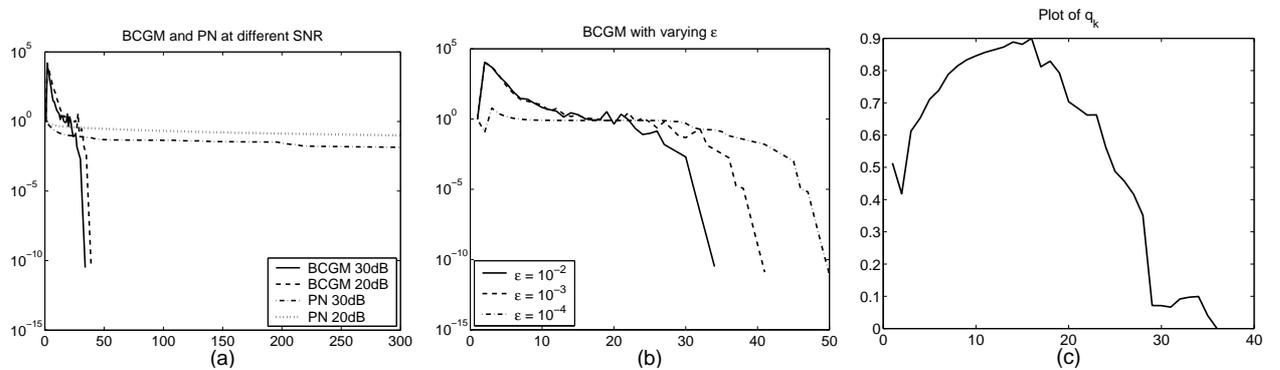


Figure 3: (a) Comparison of convergence of BCGM and PN for different SNRs with a fixed PSF of size 5×5 and $\beta = 0.2$. (b) Convergence of BCGM for varying values of ϵ , with SNR=30dB, PSF= 5×5 . For both (a) and (b), the relative KKT residual versus iteration number is plotted. (c) Plot of q_k for PSF 5×5 , $\beta = 0.2$, SNR=30dB.

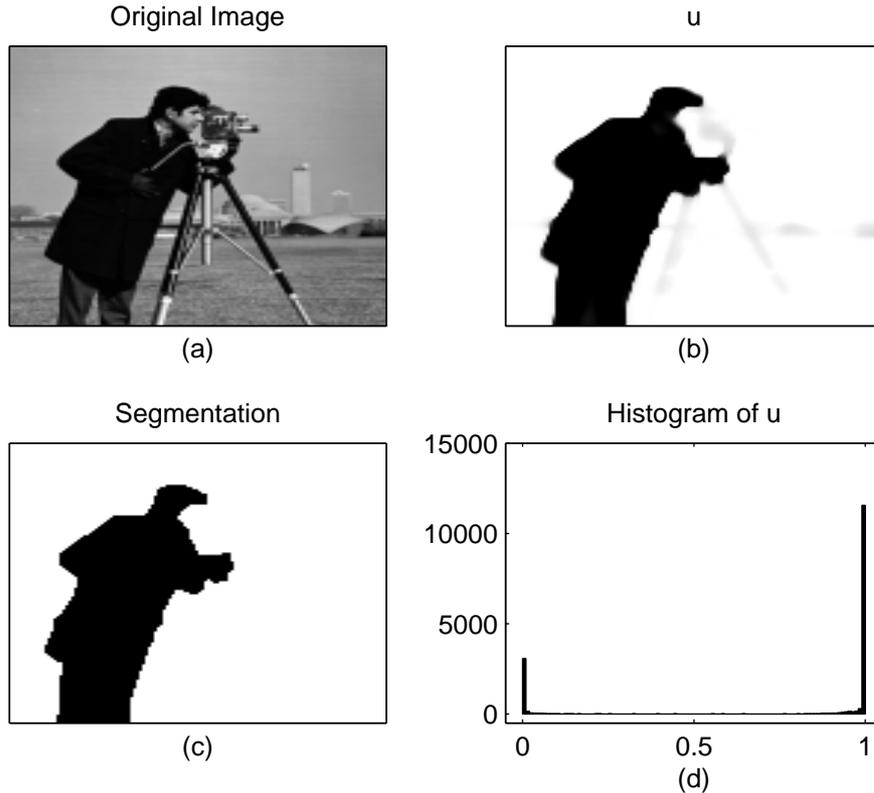


Figure 4: Segmentation results of the BCGM algorithm. (a) Original image. (b) The solution u obtained by minimizing the globalized objective (5). (c) The segmentation obtained by thresholding u with $\mu = 0.5$. (d) The histogram of the values of u . The parameters are set to $\beta = 0.4$, $\alpha = 0.01$, $\epsilon = 10^{-3}$.

algorithm is very robust to various parameters. In all cases, the BCGM algorithm achieves a relative residual of 10^{-10} within 8 iterations. But Fig. 6(c) shows that the primal and dual GD do not achieve such a high accuracy in 3000 iterations. The BCGM algorithm (with $\alpha = 0.01$ and $\epsilon = 10^{-3}$) takes about 16 seconds. The primal and dual GD take about 10 seconds for every 1000 iterations.

For the primal GD, it takes about 5000 iterations to obtain a segmentation similar to the one by BCGM. For the dual GD, it takes about 800 iterations. This shows that the dual GD converges to a close-to-binary function very quickly. Then it slows down and makes small adjustments in fine details which have little effect on the final segmentation. This explains why the residual of the dual GD is still relative large even when the segmentation stabilizes. However, this is not the case for the primal GD. After the progress slows down, the segmentation may still change significantly in a long run. Thus it suggests that the convergence of large scale features is much slower in the primal GD than that in the dual GD. Duality suggests that the convergence of fine scale features is faster in the primal GD than that in the dual GD, but we did not conduct experiments on this aspect. The proposed primal-dual BCGM algorithm appears to have a fast convergence for features in all different scales.

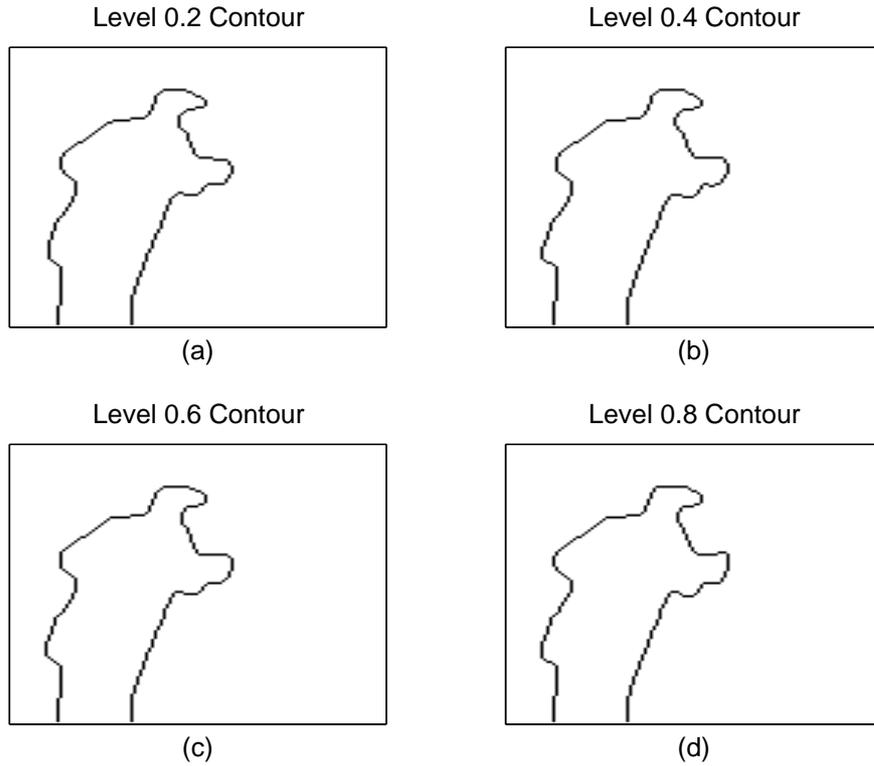


Figure 5: Segmentation results of the BCGM algorithm with respect to various thresholds. (a) $\mu = 0.2$; (b) $\mu = 0.4$; (c) $\mu = 0.6$; (d) $\mu = 0.8$.

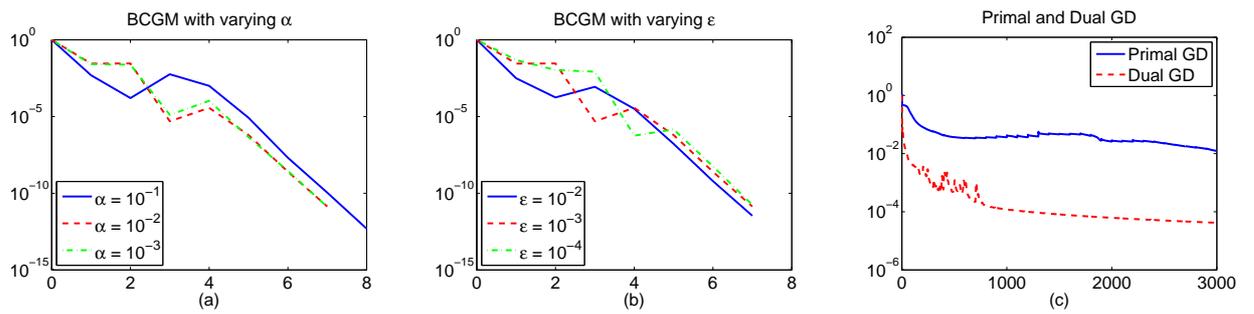


Figure 6: Convergence profile of various methods (the relative residual of the respective optimality conditions versus iteration number). The value of β is fixed at 0.4. (a) Various α 's with $\epsilon = 10^{-3}$. (b) Various ϵ 's with $\alpha = 0.01$. (c) Results of primal gradient descent with $\epsilon = 10^{-3}$ and dual gradient descent with $\theta = 0.04$.

5 Conclusion

In this paper, we propose a unified primal-dual active-set algorithm for solving a class of bilaterally constrained total variation minimization problems including constrained TV deblurring and piecewise constant Mumford-Shah segmentation problems. The convergence of our algorithm is locally superlinear. We also found experimentally that a simple line search in the dual variable p results in global convergence. By splitting the variables into suitable “active” and “inactive” sets, the Newton’s equation is greatly simplified and can be effectively preconditioned by the FBIP preconditioners. Our algorithm inherits the advantages of the primal-dual method by Chan *et al.* [8]. Compared to primal-only methods, the singularity of the term $\nabla u/|\nabla u|$ is overcome by introducing a dual variable. Compared to dual-only methods, our optimality equations do not involve the usually highly ill-conditioned term $(K^T K + \alpha I)^{-1}$ and is therefore more suitable when solving deblurring problems. But for segmentation and denoising problems, both dual and our primal-dual methods converge very fast. Our algorithm is also very robust to various parameters. On the other hand, we also provide empirical evidence to show that the use of the bilateral constraints can greatly improve the quality of restoration results. Due to the improved speed and quality, our algorithm is of great practical interest.

References

- [1] D. P. Bertsekas. Projected Newton methods for optimization problems with simple constraints. *SIAM J. Control and Optim.*, 20(2):221–246, 1982.
- [2] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
- [3] X. Bresson, S. Esedoğlu, P. Vandergheynst, J.-P. Thiran, and S. Osher. Fast global minimization of the active contour/snake model. *Journal of Mathematical Imaging and Vision*, to appear, 2007.
- [4] J. L. Carter. *Dual Methods for Total Variation-Based Image Restoration*. PhD thesis, UCLA, April 2002.
- [5] A. Chambolle. An algorithm for total variation minimization and applications. *J. Math. Imaging Vision*, 20:89–97, 2004.
- [6] T. F. Chan and S. Esedoğlu. Aspects of total variation regularized l^1 function approximation. *SIAM J. Appl. Math.*, 65:1817–1837, 2005.
- [7] T. F. Chan, S. Esedoğlu, and M. Nikolova. Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM J. Appl. Math.*, 66:1632–1648, 2006.
- [8] T. F. Chan, G. H. Golub, and P. Mulet. A nonlinear primal-dual method for total variation-based image restoration. *SIAM J. Sci. Comput.*, 20(6):1964–1977, 1999.
- [9] T. F. Chan and J. Shen. *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods*. SIAM, 2005.
- [10] T. F. Chan, J. Shen, and L. Vese. Variational PDE models in image processing. *Notices Amer. Math. Soc.*, 50:14–26, 2003.
- [11] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Trans. Image Process.*, 10(2):266–277, 2001.

- [12] M. Hintermüller, K. Ito, and K. Kunisch. The primal-dual active set strategy as a semismooth Newton's method. *SIAM J. Optim.*, 13(3):865–888, 2003.
- [13] M. Hintermüller and G. Stadler. A primal-dual algorithm for TV-based inf-convolution-type image restoration. *SIAM J. Sci. Comput.*, 28:1–23, 2006.
- [14] O. Juan, R. Keriven, and G. Postelnicu. Stochastic motion and the level set method in computer vision: Stochastic active contours. *International Journal of Computer Vision*, 69(1):7–25, 2006.
- [15] D. Krishnan, P. Lin, and A. M. Yip. A primal-dual active-set method for non-negativity constrained total variation deblurring problems. *IEEE Trans. Image Process.*, 16(11):2766–2777, 2007.
- [16] Y. N. Law, H. K. Lee, and A. M. Yip. A multi-resolution stochastic level set method for Mumford-Shah image segmentation. *UCLA CAM Report*, #07-43, 2007.
- [17] F. Lin, M. Ng, and W. Ching. Factorized banded inverse preconditioners for matrices with Toeplitz structure. *SIAM J. Sci. Comput.*, 26:1852–1870, 2005.
- [18] R. M. Mersereau and R. W. Schafer. Comparative study of iterative deconvolution algorithms. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process. ICASSP '78*, volume 3, pages 192–195, 1978.
- [19] R. Mifflin. Semismooth and semiconvex functions in constrained optimization. *SIAM J. Control and Optimization*, 15(6):959–972, 1977.
- [20] D. Mumford and J. Shah. Optimal approximation by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42:577–685, 1989.
- [21] M. Ng, L. Qi, Y. Yang, and Y. Huang. On semismooth Newton's methods for total variation minimization. *J. Math. Imaging Vision*, 27(3):265–276, 2007.
- [22] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1999.
- [23] R. W. Schafer, R. M. Mersereau, and M. A. Richards. Constrained iterative restoration algorithms. *Proceedings of the IEEE*, 69(4):432–450, 1981.
- [24] D. Sun and J. Han. Newton and quasi-Newton methods for a class of nonsmooth equations and related problems. *SIAM J. Optim.*, 7:463–480, 1997.
- [25] C. R. Vogel. Solution of linear systems arising in nonlinear image deblurring. In G. Golub, S. Lui, F. Luk, and R. Plemmons, editors, *Scientific Computing: Proceedings of the Workshop*, pages 148–158, Hong Kong, 1997. Springer.
- [26] C. R. Vogel. *Computational Methods for Inverse Problems*. SIAM, Philadelphia, 2002.