A Primal-Dual Active-Set Method for Non-Negativity Constrained Total Variation Deblurring Problems

D. Krishnan, *Member, IEEE*, Ping Lin, and Andy M. Yip

***Abstract*—This paper studies image deblurring problems using a total variation-based model, with a non-negativity constraint. The addition of the non-negativity constraint improves the quality of the solutions, but makes the solution process a difficult one. The contribution of our work is a fast and robust numerical algorithm to solve the non-negatively constrained problem. To overcome the nondifferentiability of the total variation norm, we formulate the constrained deblurring problem as a primal-dual program which is a variant of the formulation proposed by Chan, Golub, and Mulet for unconstrained problems. Here, dual refers to a combination of the Lagrangian and Fenchel duals. To solve the constrained primal-dual program, we use a semi-smooth Newton's method. We exploit the relationship between the semi-smooth Newton's method and the primal-dual active set method to achieve considerable simplification of the computations. The main advantages of our proposed scheme are: no parameters need significant adjustment, a standard inverse preconditioner works very well, quadratic rate of local convergence (theoretical and numerical), numerical evidence of global convergence, and high accuracy of solving the optimality system. The scheme shows robustness of performance over a wide range of parameters. A comprehensive set of numerical comparisons are provided against other methods to solve the same problem which show the speed and accuracy advantages of our scheme.[1]**

***Index Terms*—Image deblurring, non-negativity, primal-dual active-set, semismooth Newton's method, total variation.**

I. Introduction

TOTAL VARIATION (TV) minimization problems were first introduced into the context of image denoising in the seminal paper [3] by Rudin *et al.*. They have proven to be successful in dealing with image denoising and deblurring problems [1], [4]–[6], image inpainting problems [7], and image decomposition [8]. Recently, they have also been applied in various areas such as CT imaging [9], [10] and confocal microscopy [11]. The main advantage of the TV formulation is the ability to preserve edges in the image. This is due to the piecewise smooth regularization property of the TV norm.

A discrete version of the unconstrained TV deblurring problem proposed by Rudin *et al.* in [3] is given by

$$\min_u \frac{1}{2}||Ku - f||^2 + \beta||u||_{\mathrm{TV}} \tag{1}$$

where $\|\cdot\|$ is the l^2 norm, f is the observed (blurred and noisy) data, K is the blurring operator corresponding to a *point spread function* (PSF), u is the unknown data to be recovered, and $\|\cdot\|_{\mathrm{TV}}$ is the discrete TV regularization term. We assume that the $m \times n$ images $u = (u_{i,j})$ and $f = (f_{i,j})$ have been rearranged into a vector form using the lexicographical ordering. Thus, K is an $mn \times mn$ matrix. The discrete TV norm is defined as

$$\|u\|_{\mathrm{TV}} := \sum_{i=1}^{m-1}\sum_{j=1}^{n-1} |(\nabla u)_{ij}| \tag{2}$$

where $(\nabla u)_{ij} = [u_{i+1,j} - u_{i,j}, u_{i,j+1} - u_{i,j}]^T$. Here, i and j refer to the pixel indices in the image and $|\cdot|$ is the Euclidean norm for $\mathbb{R}^2$. Regularization is necessary due to the presence of noise, see [12]. Without regularization, noise amplification would be so severe that the resulting output data is useless, especially when K is very ill conditioned. Even when K is well conditioned, regularization is still needed to remove noise. The regularization parameter β needs to be selected as a tradeoff between oversmoothing and noise amplification. When $K = I$, the deblurring problem becomes a pure denoising problem. When K is derived from a nontrivial PSF (i.e., apart from the Dirac Delta function), the problem is harder to solve since the pixels of f are coupled together to a greater degree. When K is unknown, the problem becomes a blind deblurring problem [4], [13]. In this paper, we assume that the PSF, and, hence, K is known.

One of the major reasons for ongoing research into TV deblurring problems is that the nondifferentiability of the TV norm makes it a difficult task to find a fast numerical method. The (formal) first-order derivative of the TV norm involves the term $(\nabla u/|\nabla u|)$ which is degenerate when $|\nabla u| = 0$. This could happen in flat areas of the image. Methods that can effectively deal with such singularities are still actively sought.

A number of numerical methods have been proposed for unconstrained TV denoising and/or deblurring models. These include partial differential equation-based methods such as explicit [3], semi-implicit [14] or operator splitting schemes [15] and fixed point iterations [5]. Optimization oriented techniques include Newton-like methods [1], [6], [16], multilevel [17], second-order cone programming [18], and interior-point methods [19]. Recently, graph-based approaches have also been studied [20]–[23].

Carter [24] presents a (Fenchel) dual formulation[2] of the TV denoising problem and studies some primal-dual interior-point

Manuscript received February 21, 2007; revised August 7, 2007. A. Yip was supported in part by the National University of Singapore under Grants R-146-050-079-101 and R-146-050-079-133. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Michael Elad.

The authors are with the Department of Mathematics, National University of Singapore, Singapore (e-mail: andyyip@nus.edu.sg).

Digital Object Identifier 10.1109/TIP.2007.908079

[1]Our code is available at http://www.math.nus.edu.sg/~mhyip/nncgm/.

[2]Simply put, a Fenchel dual formulation uses the change of variable $p = (\nabla u/|\nabla u|)$ to express the original primal objective in terms of the dual variable p. See [25] for a rigorous treatment of duality.

and primal-dual relaxation methods.[3] Chambolle [26] presents a semi-implicit scheme and Ng *et al.* [27] present a semi-smooth Newton's method[4] for solving the same dual problem. These algorithms have the advantage of not requiring an extra regularization of the TV norm. Being faithful to the original TV norm without any regularization, these methods often require many iterations to converge to a moderate level of accuracy for the underlying optimization problem is not strictly convex.

Hintermüller and Kunisch [29] have derived a (Fenchel) dual version of an anisotropic TV deblurring problem. In the anisotropic formulation, the TV norm $||u||_{\mathrm{TV}}$ in (2) is replaced with $\sum_{i,j}|(\nabla u)_{ij}|_1$ where $|\cdot|_1$ is the l^1 norm for $\mathbb{R}^2$. This makes the dual problem a quadratic one with linear bilateral constraints. In contrast, the isotropic formulation is based on the l^2 norm and has the advantage of being rotation invariant. However, the dual problem corresponding to the isotropic TV norm has quadratic constraints which are harder to deal with. Hintermüller and Kunisch have solved the anisotropic formulation using a primal-dual active-set (PDAS) method, but the algorithm requires an additional l^2 regularization term when a deblurring problem is solved.

Chan, Golub, and Mulet (CGM) present a primal-dual formulation [1]. This method (which we henceforth call the CGM method) simultaneously solves both the primal and (Fenchel) dual problems. In this paper, we propose a variant of their method to handle the non-negativity constraint. Recently, a study of a primal-dual algorithm for other regularization functionals has been given in [30].

It should be noted that many of the aforementioned numerical methods are specific to denoising and cannot be readily extended to a general deblurring problem. Fewer papers focus on TV deblurring problems. Still, fewer focus on constrained TV deblurring problems. However, our method works for the more difficult non-negativity constrained isotropic TV deblurring problem and is faster than other existing methods for solving the same problem.

Image values which represent physical quantities such as photon count or energies are often non-negative. For example, in applications such as gamma ray spectral analysis [31], astronomical imaging and spectroscopy [32], the physical characteristics of the problem require the recovered data to be non-negative. An intuitive approach to ensuring non-negativity is to solve the unconstrained problem first, followed by setting the negative components of the resulting output to zero. However, this approach may result in the presence of spurious ripples in the reconstructed image. Chopping off the negative values may also introduce patches of black color which could be visually unpleasant. In biomedical imaging, the authors in [10], [11] also stressed the importance of non-negativity in their TV-based models. However, they obtain non-negative results by tuning a regularization parameter similar to the β in (1). This may cause the results to be under- or over-regularized. Moreover, there is no guarantee that such a choice of the parameter exists. Therefore, a non-negativity constraint on the deblurring problem is a natural requirement.

The non-negatively constrained TV deblurring problem is given by

$$\min_{u,u\geq 0}\frac{1}{2}||Ku-f||^2+\beta||u||_{\mathrm{TV}}. \tag{3}$$

This problem is convex for all K and is strictly convex when K is of full rank. Since most practical PSFs lead to a K of full rank, we shall assume that K is of full rank. This assumption means that the problem has a unique solution and that K^TK is invertible. In deblurring problems, even if the observed data are all positive, the deblurred result may contain negative values if non-negativity is not enforced. Therefore, imposing non-negativity is nonredundant.

Schafer *et al.* [31], [33], have studied the non-negativity constraint on gamma-ray spectral data and synthetic data. They have demonstrated that such a constraint helps not only in the interpretability of the results, but also helps in the reconstruction of high-frequency information beyond the Nyquist frequency (in the case of bandlimited signals). Reconstruction of the high-frequency information is an important requirement for image processing since the details of the image are usually the edges. Some examples showing the increased reconstruction quality of imposing non-negativity can be found in [34]. Studies on other non-negativity constrained deblurring problems such as Poisson noise model, linear regularization and entropy-type penalty can be found in [35] and [36].

Fig. 1 gives another example. The reconstruction [Fig. 1(d), (f)] based on the unconstrained primal-dual method presented in [1] shows a larger number of spurious oscillations. In contrast, the constrained solution [Fig. 1(c), (e)] based on the same regularization parameter β, has much fewer spurious oscillations. The unconstrained result has a higher reconstruction error (designated by a lower PSNR value) compared to the constrained reconstruction. See (14) for the definition of PSNR.

Very few numerical approaches have been studied for non-negatively constrained total variation deblurring problems. In [34] and [37], a projected Newton's method based on the algorithm of [38] is presented to solve the non-negatively constrained problem. We study the performance of this algorithm in this paper. Fu *et al.* [19] present an algorithm based on interior-point methods, along with very effective preconditioners. The total number of outer iterations is small. However, the inner iterations, corresponding to Newton steps in the interior-point method, take long to converge. Moreover, Fu *et al.* study the anisotropic TV formulation, which can be reduced to a linear programming problem, whereas the isotropic formulation is more difficult to solve. We have studied the interior-point method for the isotropic TV norm and observed significant slow down in the inner iterations as the outer iterations proceed. This is because of the increased ill conditioning of the linear systems that are to be solved in the inner iterations. In contrast, the primal-dual method presented in this paper does not suffer from this drawback—the number of inner *conjugate gradient*

[3]The primal-dual methods here are for general *constrained* optimization problems. They solve the optimality conditions of both the primal and Lagrange dual problems simultaneously. The Lagrange dual variables are the Lagrange multipliers associated with the constraints. Here, the constrained optimization problem to solve is the Fenchel dual formulation of the original TV denoising problem.

[4]Semi-smooth Newton's method is a generalization of the standard Newton's method for handling some nonsmooth functions [28].

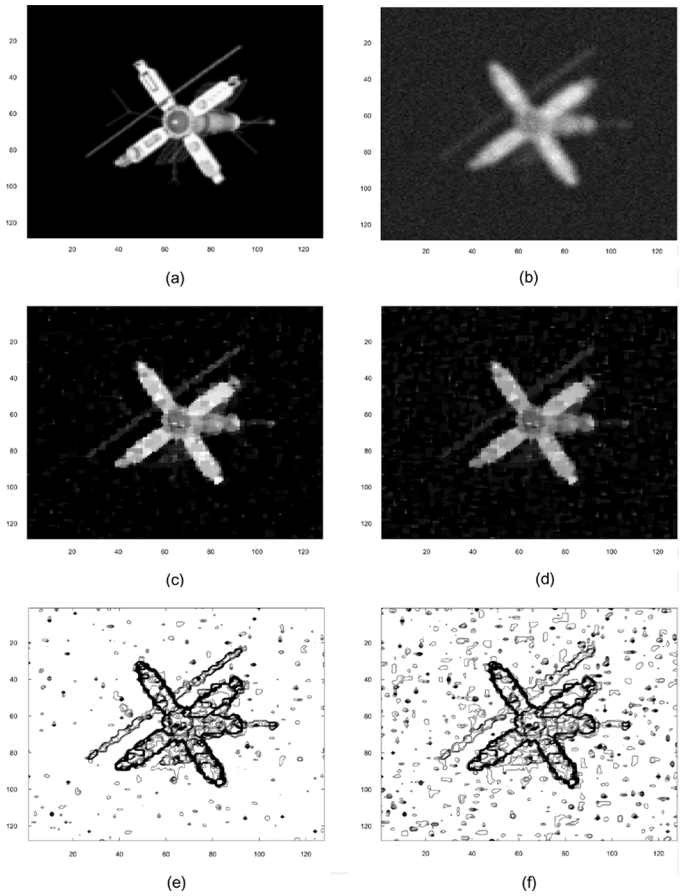


Fig. 1. Comparison of constrained and unconstrained deblurring. (a) Original Satellite (128 × 128). (b) Blurred/noisy data, Gaussian PSF (5 × 5), noise (SNR 15 dB); (c) Non-negatively constrained deblurring result. $\beta = 0.4$, PSNR = 27.61 dB. (d) Unconstrained CGM; $\beta = 0.4$ and negative components set to 0, PSNR = 25.96 dB. (e) Contour plot of (c). (f) Contour plot of (d).

iterations [39] shows no significant increase as the system approaches convergence.

The rest of the paper is organized as follows. Section II presents our proposed primal-dual method (which we call NNCGM) for non-negatively constrained TV deblurring, along with two other algorithms to which we compare the performance of the NNCGM algorithm. These two algorithms are a dual-only Alternating Minimization method and a primal-only Projected Newton's method. Section III provides numerical results to compare NNCGM with these two methods and also shows the robustness of NNCGM. Section IV gives conclusions. Appendix I gives the technical details of the derivation of the primal-dual formulation and the NNCGM algorithm. Appendix II gives the default parameters that were used for all the numerical results given in this paper.

II. Non-Negatively Constrained Primal-Dual Program

A. Dual and Primal-Dual Approaches

Solving the primal TV deblurring problem, whether unconstrained or constrained, poses numerical difficulties due to the nondifferentiability of the TV norm. This difficulty is usually overcome by the addition of a perturbation ϵ. That is, to replace $|\nabla u|$ with $|\nabla u|_\epsilon = \sqrt{|\nabla u|^2 + \epsilon}$ which is a differentiable function. The tradeoff in choosing this smoothing parameter ϵ is the reconstruction error versus the speed of convergence. The smaller the perturbation term, the more accurate is the final reconstruction. However, convergence takes longer since the corresponding objective function becomes increasingly closer to the original nondifferentiable objective function. See [40] for more details on convergence in relation to the value of ϵ.

Owing to the numerical difficulties, some researchers have studied a dual approach to the TV deblurring problem. Carter [24] and Chambolle [26] present a dual problem based on the Fenchel dual formulation for the TV denoising problem. Chambolle's scheme is based on the minimization problem

$$\min_{p,|p_{i,j}|\leq 1} \|f + \beta \mathrm{div} p\|^2. \tag{4}$$

Here

$$p_{i,j} = \begin{bmatrix} p^x_{i,j} \\ p^y_{i,j} \end{bmatrix} \tag{5}$$

is the dual variable at each pixel location with homogeneous Dirichlet boundary conditions $p^x_{0,j} = p^x_{m,j} = 0$ for all j and $p^y_{i,0} = p^y_{i,n} = 0$ for all i. The vector p is a concatenation of all $p_{i,j}$. The discrete divergence operator div is defined such that the vector $\mathrm{div} p$ is given by

$$(\mathrm{div} p)_{i,j} = p^x_{i,j} - p^x_{i-1,j} + p^y_{i,j} - p^y_{i,j-1}. \tag{6}$$

It can be shown (see [26]) that $(-\mathrm{div})^T = \nabla$ defined in (2). The constraints $|p_{i,j}| \leq 1$ are quadratic after squaring both sides. The update is given by

$$p^{n+1}_{i,j} = \frac{p^n_{i,j} + \tau\beta(\nabla(\beta \mathrm{div} p^n + f))_{i,j}}{1 + \tau\beta|(\nabla(\beta \mathrm{div} p^n + f))_{i,j}|}, \quad \forall\, i,j$$

where τ is the step size (which, as shown in [26], needs to be less than 1/8). Once the optimal solution, denoted by p^*, is obtained, the denoised image u^* can be reconstructed by $u^* = \beta \mathrm{div} p^* + f$. An interesting aspect of the algorithm is that, even without the ϵ-perturbation of the TV norm, the objective function becomes a quadratic function which is infinitely differentiable. However, the dual variable p becomes constrained. Unfortunately, being based on a steepest-descent technique, the algorithm slows down towards convergence and requires a large number of iterations for even a moderate accuracy.

Hintermüller and Kunisch [29] have also used the Fenchel dual approach to formulate a constrained quadratic dual problem and to derive a very effective method. They consider the case of anisotropic TV norm so that the dual variable is bilaterally constrained, i.e., $-1 \leq p_{i,j} \leq 1$, whereas the constraints in (4) are quadratic. The smooth (quadratic) nature of the dual problem makes it much more amenable to solution by a Newton-like method. To deal with the bilateral constraints on p, the authors propose to use the 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 $\mathbb{R}^n$. The optimality conditions, known as the Karush–Kuhn–Tucker (KKT) conditions [41], are given by

$$Ay + \lambda = f$$
$$C(y, \lambda) = 0$$

where $C(y, \lambda) = \lambda - \max(0, \lambda + c(y - \psi))$ for an arbitrary positive constant c, and λ is the Lagrange multiplier. The $\max$ operation is understood to be component-wise. Then the PDAS algorithm is given by the following.

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

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

4) Stop, or set $k = k + 1$ and return to Step 2).

In their work in [2], the authors show that the PDAS algorithm is equivalent to a semi-smooth Newton's method for a class of optimization problems that include the dual anisotropic TV deblurring problem. Local superlinear convergence results are derived. Conditional global convergence results based on the properties of the matrix K are also derived. However, their formulation in [29] is based on the anisotropic TV norm and the dual problem requires an extra l^2 regularization term when a deblurring problem is solved.

Chan *et al.* [1] present a primal-dual numerical method which has a much better global convergence behaviour than a primal-only method for the unconstrained problem. As the name suggests, this algorithm simultaneously solves both the primal and dual problems. The algorithm is derived as a Newton step for the following equations:

$$p|\nabla u|_\epsilon - \nabla u = 0$$
$$-\beta \mathrm{div} p - K^T f + K^T K u = 0.$$

At each Newton step, both the primal variable u and the dual variable p are updated. The dual variable can be thought of as helping to overcome the singularity in term $\mathrm{div}(\nabla u/|\nabla u|)$. An advantage of this method is that a line search is required only for the dual variable p to maintain the feasibility $|p_{i,j}| \leq 1$, whereas a line search for u is unnecessary. Furthermore, while still requiring an ϵ-regularization, it converges fast even when the perturbation ϵ is small. Our algorithm is inspired by the CGM method. However, the CGM method does not handle the non-negativity constraint on u.

Note that dual-only methods for TV deblurring need an extra l^2 regularization which is a disadvantage for these methods. This is because the matrix $(K^T K)^{-1}$ is involved and one needs to replace it by $(K^T K + \alpha I)^{-1}$ to make it well conditioned. In denoising problems, we have $K = I$ so that the ill conditioning problem of $(K^T K)^{-1}$ in dual methods is absent. However, in deblurring problems, some extra care needs to be taken. The modified TV deblurring problem is then given by

$$\min_{u, u \geq 0} \frac{1}{2}||Ku - f||^2 + \beta ||u||_{\mathrm{TV}} + \frac{\alpha}{2}||u||^2. \tag{7}$$

Primal-dual methods such as CGM and our NNCGM do not require the extra l^2 regularization term, i.e., $\alpha = 0$. This is because $K^T K$ instead of $(K^T K)^{-1}$ is involved in the primal-dual formulation. See Appendix I.

B. NNCGM Algorithm

As already stated, our algorithm is inspired by the CGM method. Hence, we call it the non-negatively constrained Chan–Golub–Mulet (NNCGM) algorithm. We derived the dual of the constrained problem (7) as follows:

$$\min_{p, |p_i, j| \leq 1} \min_{\lambda, \lambda \geq 0} \left\{ \frac{1}{2} ||B^{1/2}(K^T f + \beta \mathrm{div} p + \lambda)||^2 \right\} \tag{8}$$

where $B := (K^T K + \alpha I)^{-1}$. The dual variable p has constraints on it arising from the Fenchel transform of the TV norm in the primal objective function (7). The variable λ has a non-negativity constraint since it arises as a Lagrange multiplier for the non-negativity constraint on u. See Appendix I for the detailed derivation. We remark that the parameter α can be set to 0 in our method, see Section II-A.

The ϵ-regularized primal-dual program associated with the problem (8) is given by

$$|\nabla u|_\epsilon p - \nabla u = 0 \tag{9}$$
$$-\beta \mathrm{div} p - K^T f - \lambda + Au = 0 \tag{10}$$
$$\lambda - \max\{0, \lambda - cu\} = 0 \tag{11}$$

where $A := K^T K + \alpha I$ and c is an arbitrary positive constant. This system is derived by reformulating the KKT system for (8). See (28)–(30) and Appendix I-B for the derivation. For convenience, we still call (9)–(11) the KKT system. We have identified the Lagrange multiplier for $\lambda \geq 0$ with the primal variable u. This leads to the presence of u in the system. Note that we have transformed all inequality constraints and complementarity conditions on u and λ into the single equality constraint in (11).

The NNCGM algorithm is essentially a semi-smooth Newton's method for the system in (9)–(11). It has been shown by Hintermüller *et al.* [29] that the semi-smooth Newton's method is equivalent to the PDAS algorithm for a certain class of optimization problems. Although the equivalency does not hold in our problem due to the nonlinearity of (9), the two methods are still highly related in the sense that the equivalency holds to the subsystem (10)–(11) with unknowns u, λ. We exploit this relationship, and use some ideas of the PDAS algorithm to significantly simplify the computations involved in solving (9)–(11), see Appendix I. Specifically, see the Sections I-B and C for a full description of the notations and equations used behind the following algorithm. The full NNCGM algorithm is as follows.

1) Select parameters based on Table V.
2) Initialize p^0, u^0, λ^0. Set $k = 0$.
3) Set the inactive set $\mathcal{I}^k = \{i : \lambda_i^k - cu_i^k \leq 0\}$ and the active set $\mathcal{A}^k = \{i : \lambda_i^k - cu_i^k > 0\}$. In the rest of the following algorithm, these two quantities are represented as $\mathcal{I}$ and $\mathcal{A}$ respectively.
4) Compute $\delta u_{\mathcal{I}}^k$, the change of u on $\mathcal{I}$, by solving (36).
5) Compute δp^k, the change of p, by applying (34).
6) Compute $\delta \lambda_{\mathcal{A}}^k$, the change of λ on $\mathcal{A}$, by applying (33).

7) Compute the step size s by $\rho \sup_{\gamma>0}\{|p_{i,j}^k + \gamma\delta p_{i,j}^k| \leq 1 \forall i,j\}$.
8) Update
$$\begin{aligned} p^{k+1} &= p^k + s\delta p^k \\ u_{\mathcal{I}}^{k+1} &= u_{\mathcal{I}}^k + \delta u_{\mathcal{I}}^k \\ u_{\mathcal{A}}^{k+1} &= 0 \\ \lambda_{\mathcal{I}}^{k+1} &= 0 \\ \lambda_{\mathcal{A}}^{k+1} &= \lambda_{\mathcal{A}}^k + \delta\lambda_{\mathcal{A}}^k. \end{aligned}$$
9) Either stop if the residual of the KKT system (9)–(11) has reached the desired accuracy, or set $k = k+1$ and go back to Step 3) otherwise. The residual is given by $(||F_1||^2 + ||F_2||^2 + ||F_3||^2)^{1/2}$ where F_1, F_2, F_3 are defined by the left hand side of (9)–(11).

At every iteration, the current iterates for λ and u are used to predict the active $(\mathcal{A}^k)$ and inactive $(\mathcal{I}^k)$ sets for the next iteration. This is the fundamental mechanism of the PDAS method as presented in [2]. A line search is only required in Step 7), for p. We found numerically that there was no need to have a line search in the u and λ variables. Occasional infeasibility (i.e., violation of non-negativity) of these variables during the iterations did not prevent convergence. The algorithm requires the specification of the following parameters.

1) c: c is a positive value used to determine the active and inactive sets at every iteration; see Step 3). In our tests, we found that the performance of the algorithm is independent of the value of c, as long as c is a positive value. This is consistent with the results obtained in [2]. Hence, using a fixed value of c was sufficient for all our numerical tests.
2) ρ: Setting ρ to 0.99 worked for all our numerical tests. This parameter is used only to make the step size a little conservative.
3) ϵ: The perturbation constant is to be selected at a reasonably small value to achieve a tradeoff 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 tradeoff between quality and speed for all the cases that we tested. Reducing it further did not significantly reduce the reconstruction error. See Section III for results.

The regularization parameter β decides the tradeoff 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.

Our NNCGM algorithm was inspired by using the CGM algorithm [1] to handle the TV deblurring, and the PDAS algorithm [2] to handle the non-negativity constraint. The CGM algorithm was shown to be very fast in solving the unconstrained TV deblurring problem, and involved a minimal number of parameters. It also handles the inequality constraint on the dual variable p by a simple line search. Furthermore, the numerical results in [1] show a locally quadratic rate of convergence. The PDAS algorithm handles unilateral constraints effectively. While Hintermüller and Kunisch [29] apply PDAS to handle the constraints $-1 \leq p \leq 1$, we apply it to handle the non-negativity constraints $u, \lambda \geq 0$. Under our formulation, the quadratic constraints $|p_{i,j}| \leq 1$ for all i,j are implied (9). However, we found it more convenient to maintain the feasibility of these quadratic constraints by a line search as in the CGM method. This can make sure that the linear system (36) to solve in Step 4) is positive definite. Since the NNCGM method is basically a semi-smooth Newton's method and the system of equations to solve in our formulation is strongly semi-smooth, it can, therefore, be expected that the NNCGM algorithm should exhibit a quadratic rate of local convergence. The numerical results of Section III show a locally quadratic rate of convergence.

C. Preconditioners

The most computationally intensive step of the NNCGM algorithm is Step 4) which involves solving the linear system in (36). Though significantly smaller than the original linear system (32) obtained by linearizing (9)–(11), it is still a large system. We, therefore, explored the use of preconditioners, and discovered that the standard ILU preconditioner [39] and the factorized banded inverse preconditioner (FBIP) [42] worked well to speed up the solution of the linear system. The FBIP preconditioner, in particular, worked extremely well. 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 FFTs for computations involving the matrix $A = K^T K + \alpha I$.

The original system (32) has different characteristics in each of its blocks. It is, therefore, harder to construct an effective preconditioner. In contrast, the reduced system (36) has a simpler structure so that standard preconditioners work well.

D. Comparative Algorithms

We compare the performance of the NNCGM with two other algorithms: a primal-only projected Newton's (PN) algorithm, and a dual-only alternating minimization (AM) algorithm. To the best of our knowledge, the PN algorithm is the only algorithm proposed for solving non-negativity constrained isotropic TV deblurring problems that is designed for speed. The AM algorithm, derived by us is a straightforward and natural way to reduce the problem into subproblems that are solvable by existing solvers. A common way used in application-oriented literature is to cast the TV minimization problem as a maximum *a posteriori* estimation problem and then apply the expectation maximization (EM) algorithm with multiplicative updates to ensure non-negativity [43]. The algorithm is usually straightforward to implement. However, it is well-known that the convergence of EM-type algorithms is slow. We experimented with a number of other algorithms, as well, but the performance of those algorithms was quite poor. For example, we tried using a barrier method to maintain feasibility, but the method was very slow. The method given in Fu *et al.*'s paper, [19] uses linear programming. Since they use the anisotropic model for the TV, it is not a problem to maintain feasibility in their approach. We tried to adopt this approach for our isotropic formulation, but it is difficult to maintain the feasibility of the problem. Other interior-point methods require a feasible initial point, which is very difficult to obtain for this problem owing to the nonlinearity.

The PN algorithm is based on that presented in [34] and [37]. 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

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 non-negativity constraint. A few parameters have to be modified to tune the line search. The method is quite slow, for only a few inactive variables 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 of our experiments, a quadratic convergence has not been observed within the limit of 300 iterations. More importantly, the Newton iterations diverge for many initial data since the nondifferentiability of the TV norm has not been dealt with properly [40].

A natural way to solve the dual problem (8) is by alternating minimization. The AM algorithm is based on the convexity of the dual problem. The problem is solved by the alternating solution of two subproblems: the λ subproblem for a fixed p and the p subproblem for a fixed λ. The λ subproblem is given by

$$\min_{\lambda,\lambda\geq 0} \frac{1}{2}\|B^{1/2}(b+\lambda)\|^2 \tag{12}$$

where $b = K^T f + \beta \mathrm{div} p^c$ and p^c is the latest value of p from the previous iteration. The p subproblem is given by

$$\min_{p,|p_{i,j}|\leq 1} \frac{1}{2}\|B^{1/2}(h+\beta \mathrm{div} p)\|^2 \tag{13}$$

where $h = K^T f + \lambda^c$ and λ^c is the latest value of λ from the previous iteration. The solution of the λ subproblem (12) uses the PDAS algorithm presented in [2]. The solution of the p subproblem (13) is based on an extension of Chambolle's steepest descent technique presented in [26], modified for deblurring problems. The Euler–Lagrange equation corresponding to the problem is

$$-\beta\left[\nabla(B(\beta \mathrm{div} p + h))\right]_{i,j} + \mu_{i,j} p_{i,j} = 0, \quad \forall i,j$$

where, as before, i, j refer to individual pixels in the image. Note that, as discussed earlier, $\alpha > 0$ in the case of AM, since this is a dual-only method. The preceding equation is used to derive the following steepest-descent algorithm to solve the p subproblem

$$p_{i,j}^{n+1} = \frac{p_{i,j}^n + \tau\beta[\nabla(B(\beta \mathrm{div} p^n + h))]_{i,j}}{1 + \tau\beta|[\nabla(B(\beta \mathrm{div} p^n + h))]_{i,j}|}, \quad \forall i,j.$$

Here, the step size τ is inversely proportional to the square root of the condition number of $K^T K + \alpha I$ which is very small for most reasonable choice of α (which is set to 0.008 in our experiments). Thus, a large number of steps is expected. Once the dual problem is solved, the solution u^* to the original problem is recovered as

$$u^* = B(K^T f + \beta \mathrm{div} p^* + \lambda^*)$$

where p^* and λ^* are the optimal solution to the dual problem. Duality arguments can be used to show that the recovered optimal u^* satisfies the non-negativity constraint, cf. (23) and (26).

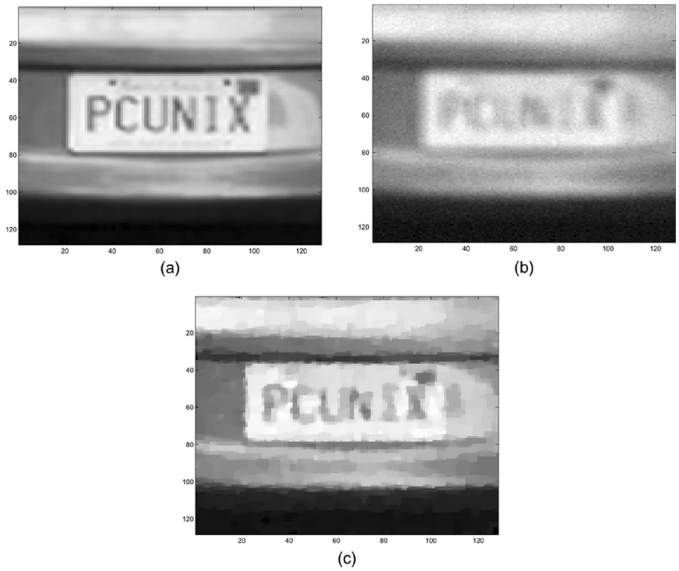


Fig. 2. (a) Original License Plate image (128×128). (b) Blurred with Gaussian PSF 7×7, SNR 30 dB. (c) TV deblurring results with NNCGM, $\beta = 0.4$.

III. Numerical Results

A. Introduction

In this section, we present extensive numerical results to demonstrate the performance of the NNCGM algorithm. We consider various conditions: different *signal-to-noise ratios* (SNRs), different types and sizes of PSFs and different values of the smoothing parameter ϵ. We also show the robustness of NNCGM with respect to various parameters, and the performance of the FBIP preconditioner. The two images that will be used for the comparison purposes are the License Plate and Satellite images. The original images and typical results of TV deblurring for NNCGM are shown in Figs. 1 and 2.

B. Numerical Comparison With the PN and AM Algorithms

In the tests that follow, we vary one condition at a time, leaving the others to their moderately chosen values. In each test, we run each algorithm for a few different values of β and choose the optimal β that minimizes the l^2 reconstruction error. Unless otherwise mentioned, all results for NNCGM are with the use of the FBIP preconditioner, which significantly speeds up the processing. For PN, we tested both the ILU and FBIP preconditioners, and they caused the processing to be slower. Therefore, the results reported are without the use of any preconditioner. Our primary interest in Figs. 3–6 is the outer iterations which are largely independent of the inner iterations.

Fig. 3 compares the convergence of NNCGM, PN and AM for different SNRs of -10, 20, and 50 dB corresponding to high-, medium-, and low-level of noise, respectively. Both the License Plate and Satellite images are compared. A fixed Gaussian PSF of size 9×9 and a fixed ϵ of 10^{-2} were used. It is seen that the NNCGM method reaches a very high accuracy of KKT residual of the order of 10^{-6} and the convergence is quadratic eventually. An even higher accuracy can be achieved with only a few more

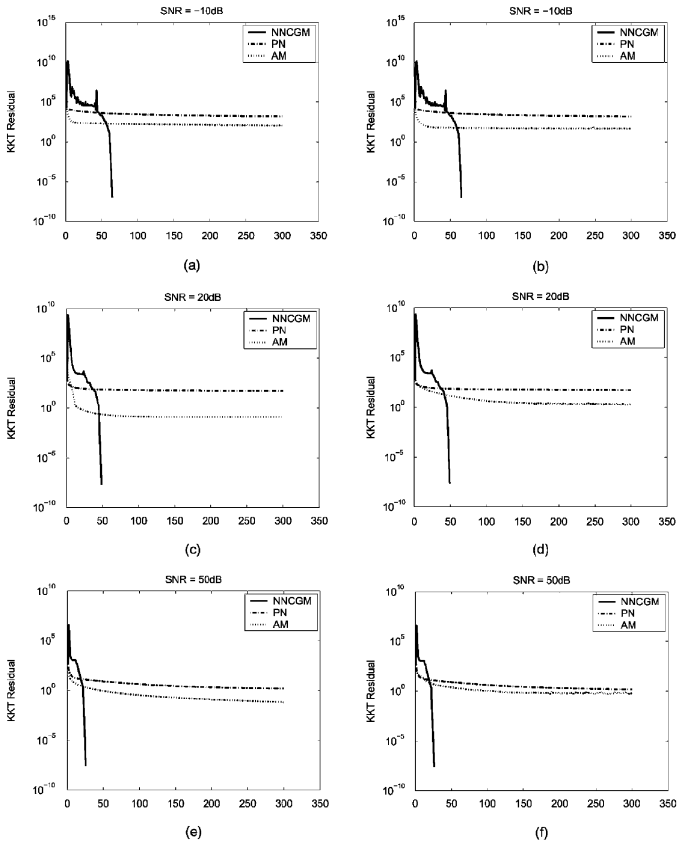


Fig. 3. Convergence profiles (KKT residual versus outer iterations) of varying SNR. (a), (b) -10 dB. (c), (d) 20 dB. (e), (f) 50 dB. Left column corresponds to License Plate, right column to Satellite.

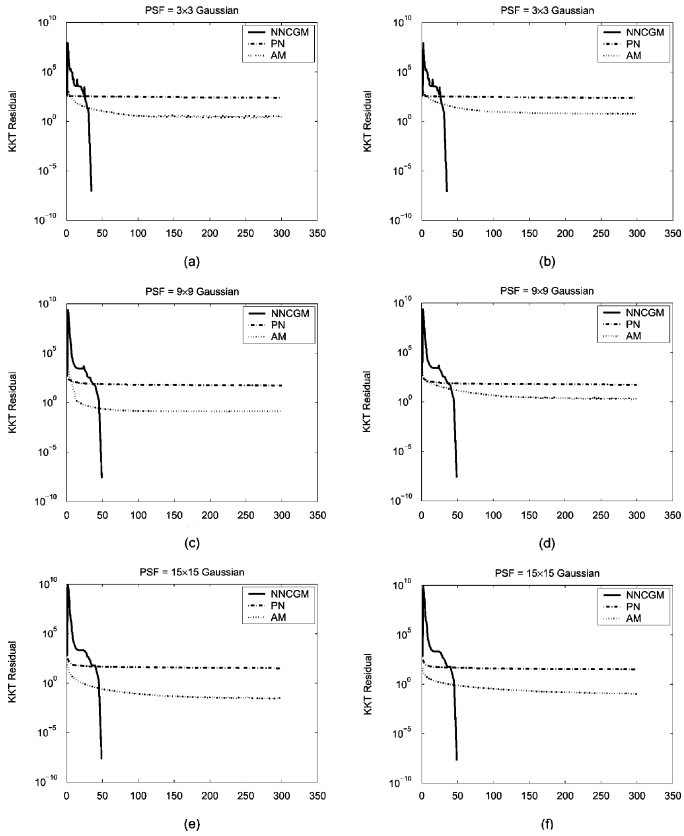


Fig. 4. Convergence profiles (KKT residual versus outer iterations) of varying PSF. (a), (b) Gaussian 3×3. (c), (d) Gaussian 9×9. (e), (f) Gaussian 15×15. Left column corresponds to License Plate, right column to Satellite.

iterations. The PN and AM methods become very slow in their progression after about 50 iterations. The total number of outer iterations for the NNCGM method stays below 70 even for a high noise level of -10 dB.

Fig. 4 compares the convergence of NNCGM, PN, and AM for varying Gaussian PSF size with a fixed SNR of 20 dB and a fixed ϵ of 10^{-2}, for both License Plate and Satellite images. Fig. 5 compares the convergence of NNCGM, PN and AM for varying ϵ with fixed SNR of 20 dB and Gaussian PSF of size 9×9, for both License Plate and Satellite images. Fig. 6 compares the convergence of NNCGM, PN and AM for Gaussian blur and out-of-focus blur with a fixed SNR of 20 dB, a fixed PSF size of 9×9 and a fixed ϵ of 10^{-2}, for both License Plate and Satellite images.

Tables I–IV show the CPU timings in seconds and the *peak signal-to-noise ratio* (PSNR) in decibels for the plots from Figs. 3–6. The PSNR, defined by

$$10 \log_{10} \left(\frac{255^2}{\frac{1}{mn} \|\text{original} - \text{reconstructed}\|^2} \right) \qquad (14)$$

is a measure of image reconstruction error. Here, $m \times n$ are dimensions of the image. The larger the PSNR, the smaller the error will be. The figure in parentheses after each CPU timing for NNCGM refers to the total number of outer iterations required for each method. In all cases, we set the maximum number of iterations to 300, for both the PN and AM algorithms essentially stagnate after 300 iterations. The first subrow in each row are License Plate data and the second subrow are Satellite data. In each case, bold letters highlight the lowest CPU timings and the lowest reconstruction error (highest PSNR) among the three algorithms. All the algorithms were implemented in Matlab 7.2. CPU timings were measured on a Pentium D 3.2-GHz processor with 4 GB of RAM. In the tables, "LP" refers to the License Plate image and "S" refers to the Satellite image, respectively. "G" refers to Gaussian and "OOF" refers to out-of-focus, respectively.

In most cases, the PN algorithm iterated for the maximum 300 iterations but did not reach even a moderate level of accuracy for the stopping criterion (the KKT residual). So, the CPU timings for the PN algorithm would have been much higher if we iterated until convergence. Moreover, it is well known that primal-only methods such as PN have a small domain of convergence [5], [40]. We would have had to use a large value of ϵ to enlarge the domain of convergence to make PN converge. The NNCGM algorithm shows considerable robustness to the initial guess even for a small value of ϵ [see Fig. 7(a)], therefore allowing for a greater accuracy. It is seen that the NNCGM algorithm outperforms the other two both in terms of CPU time and PSNR in most of the tested cases. NNCGM is also the only method that is able to achieve very high levels of accuracy within a reasonable time, and so we do not need the usual tradeoff between time taken and accuracy. The AM algorithm, while having a larger domain of convergence, is too slow for practical purposes.

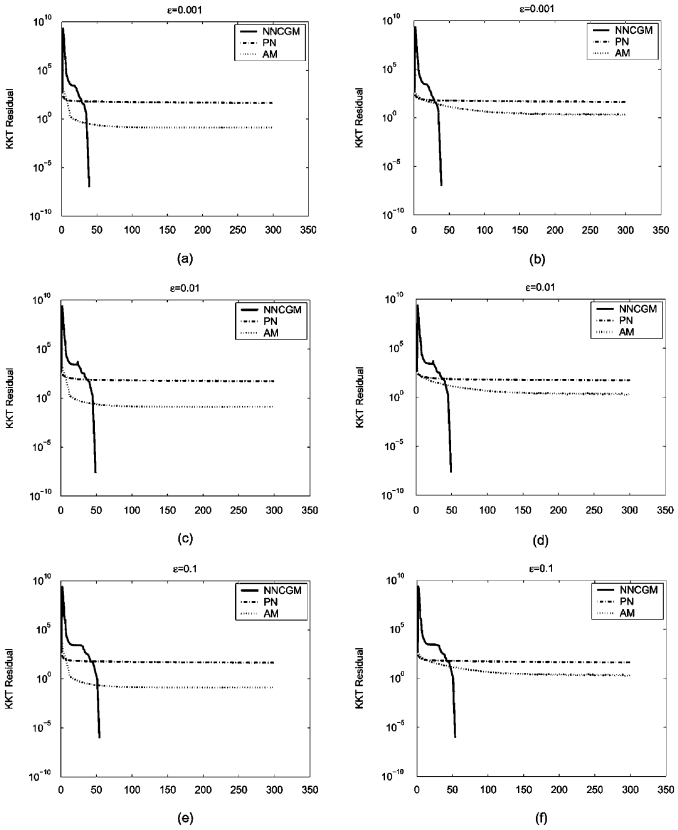


Fig. 5. Convergence profiles (KKT residual versus outer iterations) of varying smoothing parameter ϵ. (a), (b) $\epsilon = 0.001$. (c), (d) $\epsilon = 0.01$. (e), (f) $\epsilon = 0.1$. Left column corresponds to License Plate, right column to Satellite.

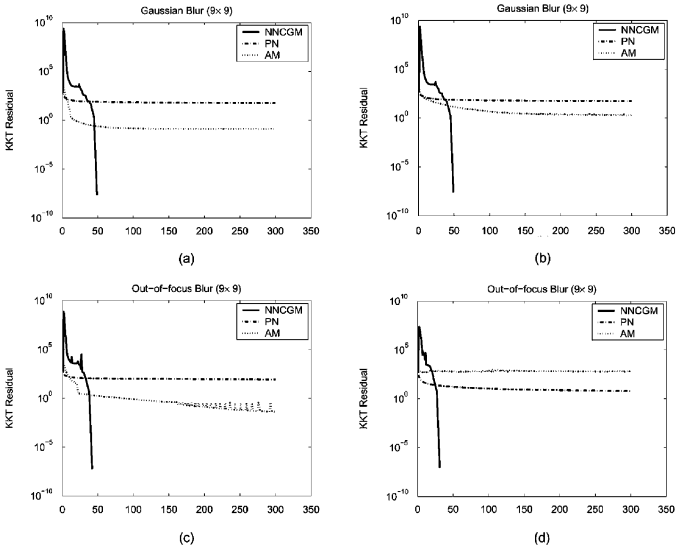


Fig. 6. Convergence profiles (KKT residual versus outer iterations) for different types of blur. (a), (b) Gaussian. (c), (d) Out-of-focus. The left column corresponds to License Plate, right column to Satellite.

C. Robustness of NNCGM

In this section, we test the robustness of the NNCGM algorithm with respect to various parameters. Fig. 7(a) shows the convergence of NNCGM for extremely small values of ϵ, a fixed SNR of 20 dB and a Gaussian PSF of size 9×9 for the License Plate image. It is seen that even for small values of ϵ, the rate of convergence is eventually quadratic. In contrast, PN diverges for such small values of ϵ.

TABLE I
CPU TIME IN SECONDS AND PSNR FOR VARYING SNRS

SNR	Image	NNCGM		PN		AM	
		CPU	PSNR	CPU	PSNR	CPU	PSNR
−10	LP	449.71 (65)	11.88	**369.4**	**12.32**	6630	8.16
	S	**124.92** (62)	**29.64**	150.64	23.42	6360	20.94
20	LP	**192.28** (49)	**31.79**	497.01	29.96	3720	30.82
	S	**49.46** (31)	**37.50**	77.57	36.31	5340	32.22
50	LP	**103.64** (26)	**37.79**	309.95	34.29	3570	32.22
	S	**52.76** (27)	**38.47**	96.17	36.16	6040	36.09

TABLE II
CPU TIME IN SECONDS AND PSNR FOR DIFFERENT PSFS. PSF SIZE IS SYMMETRIC, I.E., 3 IN THE PSF COLUMN MEANS PSF OF SIZE 3×3

PSF	Image	NNCGM		PN		AM	
		CPU	PSNR	CPU	PSNR	CPU	PSNR
3	LP	**49.04** (35)	37.29	382.22	31.91	5280	**37.8**
	S	**28.37** (42)	45.53	148.15	45.78	6050	**45.86**
9	LP	**205.26** (49)	**31.79**	483.29	29.91	3710	30.82
	S	**48.82** (42)	**37.50**	78.92	36.31	5920	36.06
15	LP	**371.74** (49)	**28.25**	614.15	28.16	4200	28.23
	S	107.09 (42)	**35.54**	**84.31**	33.04	5210	32.87

TABLE III
CPU TIME IN SECONDS AND PSNR FOR DIFFERENT ϵ

ϵ	Image	NNCGM		PN		AM	
		CPU	PSNR	CPU	PSNR	CPU	PSNR
0.001	LP	**309.09** (54)	**31.79**	356.84	29.8	3740	30.82
	S	**59.70** (42)	**37.5**	128.87	34.79	5876	36.06
0.01	LP	**206.17** (49)	**31.79**	481.07	29.96	3749	30.82
	S	**48.81** (42)	**37.50**	77.25	36.31	5280	36.06
0.1	LP	**138.06** (39)	**31.8**	352.42	29.8	3740	30.82
	S	41.32 (42)	**37.51**	**30.85**	35.31	5350	36.06

TABLE IV
CPU TIME IN SECONDS AND PSNR FOR DIFFERENT BLUR TYPES. ALL PSFS HAVE SIZE 9×9

Blur Type	Image	NNCGM		PN		AM	
		CPU	PSNR	CPU	PSNR	CPU	PSNR
G	LP	**210.31** (49)	**31.79**	478.35	29.96	3650	30.82
	S	**48** (42)	**37.50**	75.67	36.31	3670	31.02
OOF	LP	**162.70** (42)	**31.79**	478.35	29.96	6096	36.06
	S	**41** (42)	37.89	85.98	**38.62**	6470	37.61

Fig. 7(b) shows the convergence of NNCGM for different values of c, a fixed SNR of 20 dB and a Gaussian PSF of size 9×9 for the License Plate image. The convergence plots are almost identical. This is in keeping with the theoretical results obtained in [2] for the PDAS algorithm.

NNCGM can be made significantly faster with the use of standard preconditioners. We tested the NNCGM method with an ILU preconditioner and a FBIP preconditioner. The best performance was seen with the FBIP preconditioner. Fig. 8 shows the number of inner conjugate gradient (CG) iterations for each outer iteration for the FBIP preconditioner with varying PSF size, a fixed SNR of 20 dB and a Gaussian PSF of size 9×9. These results are for the License Plate image. It is seen that the number of inner CG iterations shows a slight tendency to increase. However, the number of CG iterations still remains low. Moreover, a very high accuracy is obtained so that there is no need to carry out further outer iterations. In this sense the slight increase in the number of CG iterations does not pose any serious problem.

There are three possible reasons for the increase in the number of inner CG iterations and the overall time taken to process as the PSF size increases. First, as the system approaches convergence, the linear system (36) to solve in Step 4) of the NNCGM algorithm is increasingly ill conditioned, and, thus, the CG requires more iterations to solve the problem. Second, as the PSF size increases, the matrix K has more

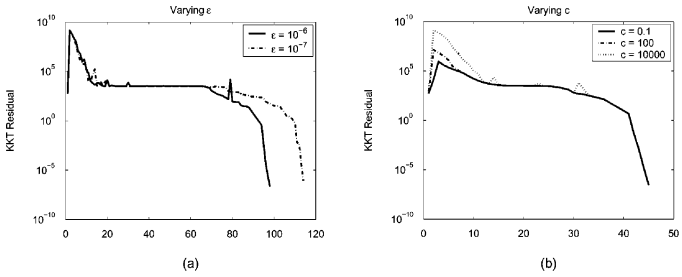


Fig. 7. Convergence plots for NNCGM for (a) very small values of ϵ; (b) varying c. Both plots are for License Plate.

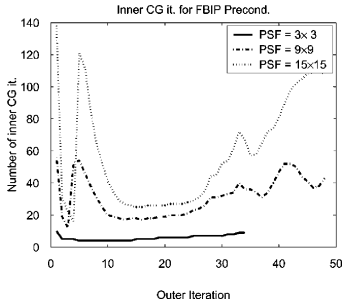


Fig. 8. Inner CG iterations for FBIP preconditioner varying Gaussian PSF sizes.

bands. This means that the construction of the preconditioner takes more time. Third, since we keep the bandwidth of the preconditioner fixed to 2 in most of the tests, the approximation error of the preconditioner becomes larger as the PSF size increases. This can be remedied by increasing the bandwidth of the preconditioner. However, it is not our purpose here to fine-tune the preconditioner performance.

IV. Discussion

Our experiments indicate that the NNCGM method is very robust in performance over a wide range of parameters. The rate of convergence is locally quadratic in all cases. The method is able to solve the non-negatively constrained TV deblurring problem to a high level of accuracy. Both Gaussian and out-of-focus blurs can be solved equally fast. The reduced system (36) has a simple structure and the standard FBIP preconditioner helps in significantly speeding up the processing time. We have also found in our tests that the method is highly robust to the initial guess of the image. In all the tests, the initial guess was simply taken to be the observed image with negative components set to zero. Using the observed image itself gave very similar convergence results. Maintaining feasibility of the variables can be effectively done in a straightforward way whereas we found it difficult to do so with standard interior-point methods. All parameters can be left at their default values, except in the case of extremely small values of the regularization parameter ϵ (values below 10^{-6}). In this case, the bandwidth of the FBIP preconditioner needs to be increased from the default value of 2 to overcome the highly ill conditioning of the linear system (36). However, we found that setting $\epsilon = 10^{-2}$ gives a good balance between quality and speed for all the cases that we tested.

Appendix I
Derivation of the Primal-Dual Program and the NNCGM Algorithm

A. Primal-Dual Program

The general primal problem under consideration (with the l^2 regularization parameter α) is

$$\min_{u,u\geq 0} \frac{1}{2}||Ku-f||^2+\frac{\alpha}{2}||u||^2+\beta||u||_{\mathrm{TV}}$$

which can be written as

$$\min_{u}\left\{\frac{1}{2}||Ku-f||^2+\frac{\alpha}{2}||u||^2+\beta||u||_{\mathrm{TV}}+I(u\geq 0)\right\} \quad (15)$$

where $I(u\geq 0)=0$ if $u\geq 0$ for all components of u, $I(u\geq 0)=\infty$ if any of the components of u are <0. Using the Fenchel transform of the TV norm in (2), we have

$$||u||_{\mathrm{TV}}=\sup_{p,|p_{i,j}|\leq 1}\langle u,-\mathrm{div}p\rangle \quad (16)$$

where p is the dual variable and div is the discrete divergence operator defined in (6). Substituting (16) into (15), we have

$$\min_{u}\max_{p}\left\{\frac{1}{2}||Ku-f||^2+\frac{\alpha}{2}||u||^2 + \beta\langle u,-\mathrm{div}p\rangle+I(u\geq 0)-I(|p_{i,j}|\leq 1)\right\}$$

where it is understood that $I(|p_{i,j}|\leq 1)=\infty$ if the Euclidean norm of any of the components of p is greater than 1. Since the objective is convex in u and concave in p, the strong max-min property holds [41]. Therefore, we can switch the order of the min and max terms to arrive at

$$\max_{p}\min_{u}\left\{\frac{1}{2}||Ku-f||^2+\frac{\alpha}{2}||u||^2+\beta\langle u,-\mathrm{div}p\rangle + I(u\geq 0)-I(|p_{i,j}|\leq 1)\right\}. \quad (17)$$

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

$$L(u,\lambda) := \frac{1}{2}||Ku-f||^2+\frac{\alpha}{2}||u||^2+\langle u,-\beta\mathrm{div}p-\lambda\rangle \quad (18)$$

where λ is the Lagrange multiplier for $u\geq 0$. Then

$$\nabla_u L=(K^TK+\alpha I)u-K^Tf-\beta\mathrm{div}p-\lambda.$$

Solving $\nabla_u L=0$ gives

$$u=B(K^Tf+\beta\mathrm{div}p+\lambda) \quad (19)$$

where $B:=(K^TK+\alpha I)^{-1}$. Next, we plug (19) into (18) to obtain

$$L(\lambda)=-\frac{1}{2}||B^{1/2}(K^Tf+\beta\mathrm{div}p+\lambda)||^2+\frac{1}{2}||f||^2. \quad (20)$$

Now, we plug (20) into (17) and change $\min_u$ to $\max_\lambda$. After simplification, we arrive at the dual problem

$$\min_p \min_\lambda \left\{ \frac{1}{2}||B^{1/2}(K^T f + \beta \text{div} p + \lambda)||^2 - \frac{1}{2}||f||^2 + I(\lambda \geq 0) + I(|p_{i,j}| \leq 1) \right\}. \quad (21)$$

B. Optimality Conditions

The optimality conditions (KKT conditions [41]) for (21) are as follows:

$$-\beta \nabla B(\beta \text{div} p + K^T f + \lambda) + \mu \odot p = 0 \quad (22)$$
$$-B(\beta \text{div} p + K^T f + \lambda) + u = 0 \quad (23)$$
$$\mu \odot (|p|^2 - 1) = 0 \quad (24)$$
$$u \odot \lambda = 0 \quad (25)$$
$$\lambda, u \geq 0 \quad (26)$$
$$|p_{i,j}| \leq 1, \forall i, j. \quad (27)$$

Here, we identify the Lagrange multipliers for $\lambda \geq 0$ with u. This is because of (23), which is the same as (19). The equation $u \odot \lambda = 0$ is 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 all i, j. The expression $\mu \odot p$ is understood as $\mu_{i,j} p_{i,j}$ for all i, j. Instead of solving the preceding KKT system directly, we follow a technique developed in [1]. We solve the following equivalent system of equations:

$$|\nabla u|_\epsilon p - \nabla u = 0 \quad (28)$$
$$-\beta \text{div} p - K^T f - \lambda + Au = 0 \quad (29)$$
$$\lambda - \max\{0, \lambda - cu\} = 0. \quad (30)$$

Here, $A := K^T K + \alpha I = B^{-1}$ and c is an arbitrary positive constant. Note that since K is assumed to be of full rank, A is invertible even if $\alpha = 0$. For notational convenience, we denote the left hand side of (28)–(30) by $F_1(p, u, \lambda)$, $F_2(p, u, \lambda)$ and $F_3(p, u, \lambda)$, respectively. To see the equivalency of the two systems, we note the following.

1) Equation (23) implies that $u = B(\beta \text{div} p + K^T f + \lambda)$. Therefore, (22) can be reduced to $\mu \odot p = \beta \nabla u$. Taking the norm on both sides, along with the complementarity conditions (24), give us $\mu = \beta |\nabla u|$. Hence, we have $p|\nabla u| = \nabla u$ which is the same as $F_1 = 0$ [cf. (28)] when $\epsilon = 0$.
2) The equation $F_2 = 0$ [cf. (29)] is simply the (23) restated.
3) The equation $F_3 = 0$ [cf. (30)] is a standard technique to express (25) and (26) as a single equality constraint.

The backward construction of (22)–(26) from (28)–(30) can also be easily done. In addition, it can be seen that (27) is implied by $F_1 = 0$. Therefore, the two systems are equivalent (when $\epsilon = 0$). Notice that an ϵ-regularization is added to the left-hand side of $F_1 = 0$ owing to the possibility that $\nabla u = 0$.

C. NNCGM Algorithm

In this section, for brevity, we define H as

$$H = -\left(I - \frac{p(\nabla u)^T}{|\nabla u|_\epsilon} \right). \quad (31)$$

The semi-smooth Newton's update for (28)–(30) is given by the following system:

$$\begin{bmatrix} |\nabla u|_\epsilon & H\nabla & 0 \\ -\beta \text{div} & A & -I \\ 0 & \frac{\partial F_3}{\partial u} & \frac{\partial F_3}{\partial \lambda} \end{bmatrix} \begin{bmatrix} \delta p \\ \delta u \\ \delta \lambda \end{bmatrix} = - \begin{bmatrix} F_1 \\ F_2 \\ F_3 \end{bmatrix}. \quad (32)$$

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 $(p(\nabla u)^T / |\nabla u|_\epsilon)\nabla$ 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 $(\partial F_3 / \partial u)$ and $(\partial F_3 / \partial \lambda)$ are defined in the sense of slant differentiability which is a generalized derivative, see [2]. We can solve the preceding system directly to obtain the updates. However, motivated by the PDAS algorithm in [2] and the relationship of the semi-smooth Newton's method to the PDAS method given there, we let

$$\mathcal{I}^k := \left\{ i : \lambda_i^k - cu_i^k \leq 0 \right\}$$
$$\mathcal{A}^k := \left\{ i : \lambda_i^k - cu_i^k > 0 \right\}.$$

These define the (predicted) complementary inactive and active sets respectively after the kth Newton step. Let $D_\mathcal{I}$ and $D_\mathcal{A}$ be the down-sampling matrix with relation to the inactive set $\mathcal{I}$ and active set $\mathcal{A}$ respectively. To extract the components of u in $\mathcal{I}$ and $\mathcal{A}$, we compute $u_\mathcal{I} = D_\mathcal{I} u$ and $u_\mathcal{A} = D_\mathcal{A} u$ respectively. The components $\lambda_\mathcal{I}$ and $\lambda_\mathcal{A}$ can be obtained similarly. Here, all the superscripts ^{k} are dropped for convenience. Let

$$A_\mathcal{I} := D_\mathcal{I} A D_\mathcal{I}^T$$
$$A_\mathcal{A} := D_\mathcal{A} A D_\mathcal{A}^T$$
$$A_{\mathcal{I}\mathcal{A}} := D_\mathcal{I} A D_\mathcal{A}^T$$
$$A_{\mathcal{A}\mathcal{I}} := D_\mathcal{A} A D_\mathcal{I}^T.$$

Using the active and inactive sets, the semi-smooth Newton step is given by

$$\begin{bmatrix} |\nabla u|_\epsilon & H\nabla D_\mathcal{I}^T & H\nabla D_\mathcal{A}^T & 0 & 0 \\ -\beta D_\mathcal{I} \text{div} & A_\mathcal{I} & A_{\mathcal{I}\mathcal{A}} & -I & 0 \\ -\beta D_\mathcal{A} \text{div} & A_{\mathcal{A}\mathcal{I}} & A_\mathcal{A} & 0 & -I \\ 0 & 0 & 0 & I & 0 \\ 0 & 0 & cI & 0 & 0 \end{bmatrix} \times \begin{bmatrix} \delta p \\ \delta u_\mathcal{I} \\ \delta u_\mathcal{A} \\ \delta \lambda_\mathcal{I} \\ \delta \lambda_\mathcal{A} \end{bmatrix} = - \begin{bmatrix} F_1 \\ D_\mathcal{I} F_2 \\ D_\mathcal{A} F_2 \\ D_\mathcal{I} F_3 \\ D_\mathcal{A} F_3 \end{bmatrix}.$$

Here, $\delta u_\mathcal{I} = D_\mathcal{I} \delta u$, etc. The fourth equation reads

$$\delta \lambda_\mathcal{I} = -D_\mathcal{I}(\lambda - 0) = -\lambda_\mathcal{I}$$

which implies $\lambda_\mathcal{I}^{k+1} = \lambda_\mathcal{I}^k + \delta \lambda_\mathcal{I} = 0$. The fifth equation of the system reads

$$c \delta u_\mathcal{A} = -D_\mathcal{A}(\lambda^k - \lambda^k + cu^k) = -cu_\mathcal{A}^k$$

TABLE V
DEFAULT PARAMETER VALUES

Parameter	Value
ϵ	10^{-2}
α (NNCGM/PN)	0
α (AM)	0.008
ρ (NNCGM)	0.99
c (NNCGM)	10^4
FBIP Bandwidth (symmetric, NNCGM)	2
Tol. for outer iter. (KKT residual)	10^{-6}
Max. ♯ outer iter.	300
Tol. for CG iter.	10^{-1}
Max. ♯ CG iter. per outer iter.	200
Max. ♯ line search (p) (NNCGM)	50
Initial guess for u	$u_{i,j} = \max\{f_{i,j}, 0\}$
Initial guess for p (NNCGM/AM)	$p_{i,j} = 0$
Initial guess for λ (NNCGM/AM)	$\lambda_{i,j} = 0$
Tol. for PDAS iter. (AM)	10^{-6}
Max. ♯ of PDAS iter. per outer iter. (AM)	10
Max. ♯ CG iterations per PDAS iter. (AM)	200
Tol. for Chambolle iter. (AM)	10^{-5}
Max. ♯ of Chambolle iter. per outer iter. (AM)	500
Tol. for Newton step (PN)	10^{-6}
Max. ♯ line search (PN)	30
Line search reduction fraction (PN)	0.25

so that $u_{\mathcal{A}}^{k+1} = u_{\mathcal{A}}^k + \delta u_{\mathcal{A}} = 0$. This implies that elements in $\mathcal{A}^k = \{i : \lambda_i^k - cu_i^k > 0\}$ are predicted as active variables in the $(k+1)$st iteration. By setting $\delta u_{\mathcal{A}} = -u_{\mathcal{A}}^k$ and $\delta\lambda_{\mathcal{I}} = -\lambda_{\mathcal{I}}^k$, the Newton's update can be reduced to

$$\begin{bmatrix} |\nabla u|_\epsilon & H\nabla D_{\mathcal{I}}^T & 0 \\ -\beta D_{\mathcal{I}}\mathrm{div} & A_{\mathcal{I}} & 0 \\ -\beta D_{\mathcal{A}}\mathrm{div} & A_{\mathcal{A}\mathcal{I}} & -I \end{bmatrix} \begin{bmatrix} \delta p \\ \delta u_{\mathcal{I}} \\ \delta\lambda_{\mathcal{A}} \end{bmatrix} = -\begin{bmatrix} F_1 \\ D_{\mathcal{I}}F_2 \\ D_{\mathcal{A}}F_2 \end{bmatrix} + \begin{bmatrix} H\nabla D_{\mathcal{A}}^T u_{\mathcal{A}} \\ A_{\mathcal{I}\mathcal{A}}u_{\mathcal{A}} - \lambda_{\mathcal{I}} \\ A_{\mathcal{A}}u_{\mathcal{A}} \end{bmatrix}.$$

It is obvious that $\delta\lambda_{\mathcal{A}}$ can be expressed in terms of δp and $\delta u_{\mathcal{I}}$

$$\delta\lambda_{\mathcal{A}} = -\beta D_{\mathcal{A}}\mathrm{div}\delta p + A_{\mathcal{A}\mathcal{I}}\delta u_{\mathcal{I}} + D_{\mathcal{A}}F_2 - A_{\mathcal{A}}u_{\mathcal{A}}. \quad (33)$$

Therefore, the Newton's update can be further simplified to

$$\begin{bmatrix} |\nabla u|_\epsilon & H\nabla D_{\mathcal{I}}^T \\ -\beta D_{\mathcal{I}}\mathrm{div} & A_{\mathcal{I}} \end{bmatrix} \begin{bmatrix} \delta p \\ \delta u_{\mathcal{I}} \end{bmatrix} = -\begin{bmatrix} F_1 \\ D_{\mathcal{I}}F_2 \end{bmatrix} + \begin{bmatrix} H\nabla D_{\mathcal{A}}^T u_{\mathcal{A}} \\ A_{\mathcal{I}\mathcal{A}}u_{\mathcal{A}} - \lambda_{\mathcal{I}} \end{bmatrix}.$$

Note that the $(1,1)$-block $|\nabla u|_\epsilon$ is a diagonal matrix. Hence, we can express δp in terms of $\delta u_{\mathcal{I}}$

$$\delta p = \frac{1}{|\nabla u|_\epsilon}\left[-H\nabla\left(D_{\mathcal{I}}^T\delta u_{\mathcal{I}} - D_{\mathcal{A}}^T u_{\mathcal{A}}\right) - F_1\right]. \quad (34)$$

Eliminating δp from the Newton's update, we get

$$\left[-\beta D_{\mathcal{I}}\mathrm{div}\frac{1}{|\nabla u|_\epsilon} - H\nabla D_{\mathcal{I}}^T + A_{\mathcal{I}}\right]\delta u_{\mathcal{I}} = g \quad (35)$$

where

$$g = -D_{\mathcal{I}}F_2 + A_{\mathcal{I}\mathcal{A}}u_{\mathcal{A}} - \lambda_{\mathcal{I}} - \beta D_{\mathcal{I}}\mathrm{div}\frac{1}{|\nabla u|_\epsilon}\left[-H\nabla D_{\mathcal{A}}^T u_{\mathcal{A}} + F_1\right].$$

The linear system (35) is nonsymmetric and so CG cannot be applied. Chan *et al.* [1] proposed in their paper to symmetrize the matrix so that CG could be applied. Following this suggestion, we symmetrize the system as

$$D_{\mathcal{I}}\left[-\beta\mathrm{div}\frac{1}{|\nabla u|_\epsilon}\left(I - \frac{p(\nabla u)^T + (\nabla u)p^T}{2|\nabla u|_\epsilon}\right)\nabla + A\right] \times D_{\mathcal{I}}^T\delta u_{\mathcal{I}} = g. \quad (36)$$

It can be shown that the system (36) is symmetric positive definite when $|p_{i,j}| \le 1$ for all i, j. We remark that using the relationship between the PDAS and semi-smooth Newton's method allows us to simplify the original linear system (32) into the much smaller linear system (36). The simpler structure of the reduced system also facilitates the construction of effective preconditioners.

APPENDIX II
DEFAULT PARAMETERS FOR GIVEN DATA

For the three algorithms whose performance we have compared, i.e., NNCGM, PN, and AM, we used the parameters shown in Table V. We used the same values for all results shown in this paper, except for the results in Fig. 7(a), where the bandwidth of the FBIP preconditioner was increased from the default value of 2 to 3 and 4, respectively to overcome the highly ill conditioning of the linear system (36). The parameters that are specific to an algorithm are indicated by parantheses after the name of the algorithm. Other parameters such as ϵ are common to all three algorithms. The subscripts i, j in the initial guess for u, p and λ, refer to all image pixels.

ACKNOWLEDGMENT

The authors would like to thank S. Defeng for useful discussions on semi-smooth Newton's methods.

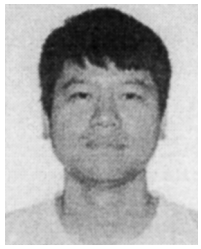
REFERENCES

[1] T. F. Chan, G. H. Golub, and P. Mulet, "A nonlinear primal-dual method for total variation-based image restoration," *SIAM J. Sci. Comput.*, vol. 20, no. 6, pp. 1964–1977, 1999.
[2] M. Hintermüller, K. Ito, and K. Kunisch, "The primal-dual active set strategy as a semismooth Newton's method," *SIAM J. Optim.*, vol. 13, no. 3, pp. 865–888, 2003.
[3] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, pp. 259–268, 1999.
[4] Y. You and M. Kaveh, "Blind image restoration by anisotropic regularization," *IEEE Trans. Image Process.*, vol. 8, no. 3, pp. 396–407, Jun. 1999.
[5] C. R. Vogel and M. E. Oman, "Iterative method for total variation denoising," *SIAM J. Sci. Comput.*, vol. 17, pp. 227–238, 1996.
[6] Y. Li and F. Santosa, "A computational algorithm for minimizing total variation in image reconstruction," *IEEE Trans. Image Process*, vol. 5, no. 6, pp. 987–995, Jun. 1996.
[7] T. F. Chan and J. Shen, "Mathematical models for local non-texture inpainting," *STAM. J. Appl. Math.*, vol. 62, pp. 1019–1043, 2001.
[8] S. Osher, A. Sole, and L. Vese, "Image decomposition and restoration using total variation minimization and the H^{-1} norm," *Multiscale Model. Simul.*, vol. 1, no. 3, pp. 349–370, 2003.
[9] V. Kolehmaincn, A. Vanne, S. Siltanen, S. Jarvenpaa, J. P. Kaipio, M. Lassas, and M. Kalke, "Parallelized Bayesian inversion for three-dimensional dental X-ray imaging," *IEEE Trans. Med. Imag.*, vol. 25, no. 2, pp. 218–228, Feb. 2006.
[10] M. Persson, D. Bone, and H. Elmqvist, "Total variation norm for three-dimensional iterative reconstruction in limited view angle tomography," *Phys. Med. Biol.*, vol. 46, pp. 853–866, 2001.
[11] N. Dey, L. Blanc-Féraud, C. Zimmer, C. Kam, J. Olivo-Marin, and J. Zerubia, "A deconvolution method for confocal microscopy with total variation regularization," in *in Proc. IEEE Int. Symp. Biomcdical Imaging: Macro to Nano*, 2004, vol. 2, pp. 1223–1226.
[12] A. N. Tikhonov and V. Y. Arsenin, *Solutions to Ill-Posed Problems*. Washington, DC: Winston, 1977.
[13] T. F. Chan and C. K. Wong, "Total variation blind deconvolution," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 370–375, Mar. 1998.

[14] D. Krishnan, P. Lin, and X.-C. Tai, "An efficient operator splitting method for noise removal in images," *Con. Comput. Phys.*, vol. 1, no. 5, pp. 847–858, 2006.
[15] M. Lysaker, O. S , and X.-C. Tai, "Noise removal using smoothed normals and surface fitting," *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1345–1357, Oct. 2004.
[16] K. Ito and K. Kunisch, "An active set strategy based on the augumented Lagrangian formulation for Image restoration," *ESAIM: Math. Model. Numer. Anal.*, vol. 33, pp. 1–21, 1999.
[17] T. Chan and K. Chen, "An optimization based, total variation image denoising," *Multiscale Model Simul.*, vol. 5, no. 2, pp. 615–645, 2006.
[18] D. Goldfarb and W. Yin, "Second-order cone programming methods for total variation based image restoration," *SIAM. J. Sci. Comput.*, vol. 27, no. 2, pp. 622–645, 2005.
[19] H. Fu, M. K. Ng, M. Nikolova, and J. L. Barlow, "Efficient minimization methods of mixed l^2-l^1 and l^1-l^1 norms for image restoration," *SIAM J. Sci. Comput.*, vol. 27, no. 6, pp. 1881–1902, 2006.
[20] A. Chambollc, *Total Variation Minimization and a Class of Binary MRF Models*. New York: Springer Verlag, 2005, vol. 3522, pp. 351–359, Lecture Notes Comput. Sci.
[21] D. S. Hochbaum, "An efficient algorithm for image segmentation, markov random fields and related problems," *J. ACM*, vol. 48, pp. 686–701, 2001.
[22] B. A. Zalesky, "Network flow optimization for restoration of images," *J. Appl. Math.*, vol. 2, pp. 199–218, 2002.
[23] J. Darbon and M. Sigelle, "Image restoration with discrete constrained total variation Part 1: Fast and exact optimization," *J. Math. Imag. Vis.*, vol. 26, pp. 261–276, 2006.
[24] J. L. Carter, "Dual methods for total variation-based image restoration," Ph.D. dissertation, Univ. California, Los Angeles, Apr. 4, 2002.
[25] I. Ekehmd and R. Temam, *Convex Analysis and Variational Problems*, ser. Classics in Appl. Math. Philadelphia, PA: SIAM, 1999, no. 28.
[26] A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Imag. Vis.*, vol. 20, pp. 89–97, 2004.
[27] M. Ng, L. Qi, Y. Yang, and Y. Huang, "On semisrnooth Newton's methods for total variation minimization," *J. Math. Imag. Vis.*, vol. 27, no. 3, pp. 265–276, 2007.
[28] L. Qi and J. Sun, "A nonsmooth version of Newton's method," *Math. Programm.*, vol. 58, pp. 353–367, 1993.
[29] M. Hintermüller and K. Kunisch, "Total-bounded variation regularization as a bilaterally constrained optimisation problem," *SIAMJ. Appl. Math*, vol. 64, pp. 1311–1333, 2004.
[30] M. Hintermüller and G. Stadler, "A primal-dual algorithm for TV-based inf-convolution-type image restoration," *SIAM J. Sci. Comput.*, vol. 28, pp. 1–23, 2006.
[31] R. W. Schafer, R. M. Mersereau, and M. A. Richards, "Constrained iterative restoration algorithms," *Proc. IEEE*, vol. 69, no. 4, pp. 432–450, Apr. 1981.
[32] K. Ho, C. Beling, S. Fung, K. Chong, M. Ng, and A. Yip, "Deconvolution of coincidence Doppler broadening spectra using iterative projected Newton method with non-negativity constraints," *Rev. Sci. Instrum.*, vol. 74, pp. 4779–4787, 2003.
[33] R. M. Mersereau and R. W. Schafer, "Comparative study of iterative deconvolution algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1978, vol. 3, pp. 192–195.
[34] C. R. Vogel, *Computational Methods for Inverse Problems*. Philadelphia, PA: SIAM, 2002.
[35] J. M. Bardsley and C. R. Vogel, "A nonnnegatively constrained convex programming method for image reconstruction," *SIAM J. Sci. Comput.*, vol. 25, pp. 1326–1343, 2004.
[36] M. Hanke, J. Nagy, and C. R. Vogel, "Quasi-Newton approach to nonnegative image restoration," *Linear Algebra Appl.*, vol. 316, pp. 223–236, 2000.
[37] C. R. Vogel, "Solution of linear systems arising in nonlinear image deblurring," in *Proc. Sci. Comput. Workshop*, G. Golub, S. Lui, F. Luk, and R. Plemmons, Eds., Hong Kong, 1997, pp. 148–158.
[38] D. P. Bertsekas, "Projected Newton methods for optimization problems with simple constraints," *SIAM J. Control Optim.*, vol. 20, no. 2, pp. 221–246, 1982.
[39] Y. Saad, *Iterative Methods for Sparse Linear Systems*, 2nd ed. Philadelphia, PA: SIAM, 2003.
[40] T. Chan, H. M. Zhou, and R. H. Chan, "Continuation method for total variation denoising problems," in *Proc. SPIE Symp. Advanced Signal Processing*, F. Luk, Ed., 1995, vol. 2563, pp. 314–325, ser. Algorithms, Architectures, and Implementations.
[41] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[42] F. Lin, M. Ng, and W. China, "Factorized banded-inverse, preconditioners for matrices with Toeplitz structure," *SIAM J. Sci. Comput.*, vol. 26, pp. 1852–1870, 2005.
[43] L. K. Saul, F. Sha, and D. D. Lee, "Statistical signal processing with nonnegativity constraints," in *Proc. EuroSpeech*, Geneva, Switzerland, 2003, vol. 2, pp. 1001–1004.

D. Krishnan (M'96) received the B.S. degree in computer engineering from the Nayang Technological University, Singapore, in 1998, and the graduate diploma in mathematics from the National University of Singapore in 2005. He is currently pursuing the M.S. degree in mathematics at the National University of Singapore.

His research interests are in the areas of image/video processing and signal processing.



Ping Lin received the B.Sc. degree in applied mathematics from Nanjing University, China, in 1984 and the Ph.D. degree in applied mathematics from the University of British Columbia, BC, Canada, in 1995.

Since Winter 1998, he has been with the Department of Mathematics, National University of Singapore, where he is now an Associate Professor. Since Fall 2007, he has also been a Professor with the Division of Mathematics, University of Dundee, U.K. His research interests include numerical ordinary and partial differential equations with or without constraints, analysis of multiscale material models, computational incompressible flows and complex fluids, and PDE-based image processing.

Andy M. Yip received the B.S. degree in mathematics from the Chinese University of Hong Kong in 1998, the M.S. degree in mathematics from the University of Hong Kong in 2000, and the Ph.D. degree in mathematics from the University of California, Los Angeles, in 2005.

He joined the Department of Mathematics at the National University of Singapore in July 2005 as an Assistant Professor. His research interests include variational and PDE methods in image processing and data mining algorithms.