Regularizing active set method for nonnegatively constrained ill-posed multichannel image restoration problem

Yanfei Wang,[1,*] Jingjie Cao,[1] Yaxiang Yuan,[2] Changchun Yang,[1] and Naihua Xiu[3]

[1]Key Laboratory of Petroleum Geophysics, Institute of Geology and Geophysics, Chinese Academy of Sciences, P.O. Box 9825, Beijing, 100029, China

[2]Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and System Sciences, Chinese Academy of Sciences, Beijing, 100080, China

[3]Department of Mathematics, Beijing Jiaotong University, Beijing, 100044, China

*Corresponding author: yfwang_ucf@yahoo.com

Received 9 October 2008; revised 15 January 2009; accepted 19 January 2009; posted 21 January 2009 (Doc. ID 101585); published 25 February 2009

In this paper, we consider the nonnegatively constrained multichannel image deblurring problem and propose regularizing active set methods for numerical restoration. For image deblurring problems, it is reasonable to solve a regularizing model with nonnegativity constraints because of the physical meaning of the image. We consider a general regularizing $l_p - l_q$ model with nonnegativity constraints. For p and q equaling 2, the model is in a convex quadratic form, therefore, the active set method is proposed since the nonnegativity constraints are imposed naturally. For p and q not equaling 2, we present an active set method with a feasible Newton-conjugate gradient solution technique. Numerical experiments are presented for ill-posed three-channel blurred image restoration problems. © 2009 Optical Society of America

OCIS codes: 100.1830, 100.3020, 100.3190, 000.4430.

1. Introduction

In applied optics, remote sensing science, and also geophysical science, a major problem is image restoration. It is because the recorded images are usually degraded due to various reasons. For example, the image may be degraded by sensor noise, misfocus of the CCD camera, nonuniform motion, atmospheric aerosols, and random atmospheric turbulence. A key problem in image restoration is to restore the image by solving a blurring model and removing noise. With the advance of observing systems, image data acquisition, nowadays, is moving more and more from single channel/band to multichannel/band, from single sensor to multisensor, from single spectrum to multiple spectra. This leads to more advanced and complex imaging processing as well as image restoration. Multichannel image restoration, like single channel image restoration, is a basic problem. It has received more and more attention recently [1–18]. This is due to the fact that there may be degradation coming from within-channel and between-channel degradation as well as atmospheric aerosols and random atmospheric turbulence for astronomical images. Multichannel images are encountered, with representative examples being color images and sequences of images, in which case each channel is represented by a frame. Such images may be degraded due to within-channel but also between-channel blurs. They can be restored by applying a single channel restoration algorithm to each channel independently. Better results are expected, however, if a multichannel restoration

0003-6935/09/071389-13$15.00/0
© 2009 Optical Society of America

approach is adopted, even when only within-channel degradations exist. This is because with such an approach the within-channel correlations are utilized [9].

A multichannel linear degradation digital model can be formulated as follows:

$$\mathbf{h_n} = \mathbf{h} + \mathbf{n} := \mathcal{K}\mathbf{f} + \mathbf{n}, \tag{1.1}$$

where $\mathcal{K}$ is the degradation matrix, $\mathbf{n}$ is the unknown noise or measurement errors, $\mathbf{h_n}$ is the observed multichannel image with noise/error, $\mathbf{h}$ is assumed to be the noise-free ideal image, and $\mathbf{f}$ is the original multichannel image. For an N-channels linear model with $M \times M$ pixels each, the observed, noise-free, and original images and the noise can be expressed, respectively, as

$$\mathbf{h_n} = \begin{bmatrix} \mathbf{h_{n,1}} \\ \mathbf{h_{n,2}} \\ \vdots \\ \mathbf{h_{n,N}} \end{bmatrix}, \quad \mathbf{h} = \begin{bmatrix} \mathbf{h_1} \\ \mathbf{h_2} \\ \vdots \\ \mathbf{h_N} \end{bmatrix},$$

$$\mathbf{f} = \begin{bmatrix} \mathbf{f_1} \\ \mathbf{f_2} \\ \vdots \\ \mathbf{f_N} \end{bmatrix}, \quad \mathbf{n} = \begin{bmatrix} \mathbf{n_1} \\ \mathbf{n_2} \\ \vdots \\ \mathbf{n_N} \end{bmatrix}, \tag{1.2}$$

where each of the M^2 vectors $\mathbf{h_{n,i}}$, $\mathbf{h}_i$, $\mathbf{f}_i$, and $\mathbf{n}_i$ results from the lexicographic ordering of the two-dimensional signals in each channel. The $NM^2 \times NM^2$ multichannel degradation matrix is given by

$$\mathcal{K} = \begin{bmatrix} \mathcal{K}_{11} & \mathcal{K}_{12} & \cdots & \mathcal{K}_{1N} \\ \mathcal{K}_{21} & \mathcal{K}_{22} & \cdots & \mathcal{K}_{2N} \\ \vdots & \vdots & \cdots & \vdots \\ \mathcal{K}_{N1} & \mathcal{K}_{N2} & \cdots & \mathcal{K}_{NN} \end{bmatrix}, \tag{1.3}$$

where the operators $\mathcal{K}_{ii}$ and $\mathcal{K}_{ij}$ $(i \neq j)$ are of dimensions $M^2 \times M^2$ and represent the within-channel and the between-channel degradation operators, respectively.

Many methods have been developed for solving the linear multichannel system, e.g., minimum mean-squared error multispectral image restoration [12], least-squares filtering for color image restoration based on spectral and spatial correlations [14], multichannel Wiener filtering in color image restoration based on multichannel autoregressive image modeling [1], constrained least-squares restoration of multichannel images using both within- and between-channel deterministic information [9], iterative regularized least-mean mixed-norm image and multichannel restoration [10,11], regularization based on the multichannel cross-validation function [18], double-regularization with conjugate gradient (CG) type alternative-minimization approach for the blind restoration of the multichannel imagery [5], and the related methods contained in these references.

But these papers do not consider the nonnegativity constraints of the solution, and in many cases the negative solutions are physically meaningless. For ill-posed inverse problems, applying *a priori* knowledge to the solution is necessary for successful inversion. For our problems, we consider nonnegativity constraints of the solution as our *a priori* knowledge and apply it for regularizing ill-posedness. Specifically, in this paper, we consider a general regularizing $l_p - l_q$ model with nonnegativity constraints. Due to the fact that the intensity of an image is always nonnegative, imposing a nonnegativity constraint, i.e., $\mathbf{f} \geq 0$, is natural. For p and q equaling 2, the model is a convex quadratic form. Therefore, a regularizing active set method is proposed since the nonnegativity constraints are imposed naturally. For p and q not equaling 2, we propose a regularizing active set method with a feasible Newton-CG technique. We emphasize the importance of the model in the signal/imaging area where the data are observed by sensors. In our model, p, q are any values greater than 0, which is quite important for many signal/image restoration problems. For example, the cases of $p = 2$ and $q = 1$ or $p = 2$ and $q \to 0$ represent "complete" sparse reconstruction of band-limited signals. In our model, the image data are assumed to be observed by sensor (which is always band limited), therefore the model is new to the literature for some choices of values of p and q, and the model does not need to be convex.

The paper is organized as follows. In Section 2, the regularization model on the l_p and l_q scale is formulated, and the solution methods when $p = 2$ are reviewed. A regularizing active set method for $p = q = 2$ is proposed in Subsection 3.B. For general values of p and q, we propose in Subsection 3.C, a regularizing active set method with a feasible Newton-CG technique. Other issues about computing an *a priori* trial solution and choosing regularization parameters and scale matrices are presented in Subsections 3.D and 3.E. Experimental results and conclusions are presented in Sections 4 and 5. Finally, Appendix A introduces the active set method for nonnegatively constrained quadratic programming problems and provides a feasible CG method.

2. Unconstrained Regularization Model on the l_p and l_q Scale

For the linear problem in Eq. (1.1), it is natural to require that the error between the observed and the noise-free images be as small as possible, i.e., that the energy of the noise be minimal,

$$\|\mathbf{h_n} - \mathbf{h}\| = \|\mathbf{n}\| \to \text{minimization}.$$

Here $\|\cdot\|$ is the norm in any form. However, due to the fact that the image restoration problem is

ill-posed, the above problem is unstable. Therefore introducing a regularization technique is necessary.

We consider a regularization model in general form:

$$\min J[\mathbf{f}] := \frac{1}{2}\|\mathcal{K}\mathbf{f} - \mathbf{h}_\mathbf{n}\|_{l_p}^p + \frac{\nu}{2}\|L(\mathbf{f} - \mathbf{f}^0)\|_{l_q}^q, \quad (2.1)$$

where $p, q > 0$, which are specified by users; $\nu > 0$ is the regularization parameter; L is the scale operator; and $\mathbf{f}^0$ is an *a priori* solution of the original model. This formulation includes most of the developed methods. The l_p norm of a vector $\mathbf{f}$ refers to $\|\mathbf{f}\|_{l_p} = (\sum_{i=1}^{NM^2} |\mathbf{f}_i|^p)^{1/p}$ and $\|\mathbf{f}\|_{l_p}^p = \sum_{i=1}^{NM^2} |\mathbf{f}_i|^p$. Norms with different values of p mean a different scale for the vector $\mathbf{f}$. Particularly, for $p = 2$ and $q = 1$, the model represents nonsmooth and sparse regularization, which is important for model parameter retrieval problems [19,20].

A canonical regularization method for solving Eq. (1.1) is Tikhonov regularization, whose standard form is given by setting $p = 2$ in Eq. (2.1) with different choices of values of q, i.e.,

$$\min \|\mathcal{K}\mathbf{f} - \mathbf{h}_\mathbf{n}\|_{l_2}^2 + \nu\Gamma(\mathbf{f}), \quad (2.2)$$

where $\Gamma(\cdot)$ is a function whose role is to give some penalization to the unknown $\mathbf{f}$. There are a lot of tricks for choosing $\Gamma(\cdot)$ (see [21,22] for smooth regularization and [23,24] for nonsmooth regularization). For example, $\Gamma(\mathbf{f})$ can be defined as $(L\mathbf{f}, \mathbf{f})$, where L is a scale operator that can be chosen as a positive definite or positive semi-definite matrix. This choice is equivalent to setting $q = 2$ and $L := L^{1/2}$ in Eq. (2.1). Usually L is sparse for improving the condition of the original kernel. In Eq. (2.2), $\nu > 0$ is the so-called regularization parameter, which plays a major role in regularizing the ill-posedness. The value of ν is positive and will be typically small. But the choice of an appropriate ν is a difficult thing, which is usually related to the spectrum of the discrete kernel $\mathcal{K}$ and the unpredictable noise level in $\mathbf{h}$ [21]. The solution methods of Eq. (2.2) include a singular value decomposition based direct method [25], Newton and quasi-Newton methods [26], gradient methods with various preconditioning techniques, and nonmonotone gradient methods [27,28]. The CG method has been proved to be an efficient iterative regularization method for recovering the correct image from its degradation [29]. Trust region methods have been recently proved to be another useful regularization tool for image restoration [30,31] and have been proved to be a kind of regularization method [30,32]. These methods involve solving a trust region subproblem in each inner iteration and accepting a new trial step within its trust region. In addition, considerable efforts were made to impose nonnegativity constraints on the minimization problem in Eq. (2.2) [28,33–36]. But all of the research is based on single channel imaging problems. We consider enforcing nonnegativity constraints on the regularization model in Eq.(2.1) on the l_p and l_q scale and study using the regularization model (2.1) with new solution methods for solving the multichannel image deblurring problem. And particularly, we concern ourselves with the nonconvex case, i.e., $p, q \neq 2$. The convexity assumption is too strict and may be invalid in many situations, say when the model parameters are sparse and nonsmooth. As is well known that the pixel of the land surface area represents a complex structure, full of nonconvexity. The observed part may contain nondifferentiable structure and may be sparse. Therefore, convexity assumption may not be an "effective model assumption."

3. Regularizing Active Set (RAS) Method for the Constrained Minimization Model

Originally, the active set method was designed as a well-posed quadratic programming problem [37–39]. We consider applying the active set method to an ill-posed multichannel image recovering problem and solve a regularizing problem. Recently, this method was applied for solving an aerosol particle size distribution function retrieval problem [40]. And it was reported in Ref. [41] that the active method is also applicable for large scale computational problems.

A. Active Set Method for Nonnegatively Constrained Minimization

The active set method is originally designed for well-posed quadratic programming problems [37–39]. In particular, it is used to solve constrained quadratic programming problems of the form

$$\min \phi(\mathbf{x}) = c + \mathbf{b}^T\mathbf{x} + \frac{1}{2}\mathbf{x}^T A\mathbf{x}, \qquad \text{s.t. } D\mathbf{x} \geq \mathbf{l}, \quad (3.1)$$

where $\phi(\mathbf{x})$ is a quadratic function, $A \in \mathbb{R}^{N\times N}$, and is symmetric; $\mathbf{b} \in \mathbb{R}^N$; c is a constant; $D \in \mathbb{R}^{M\times N}$; and $\mathbf{l} \in \mathbb{R}^M$. An active method is an implicit Newton-type method for solving the constrained quadratic programming problem (3.1), which describes a method for identifying a correct set of active inequality constraints and temporarily giving up the remaining inequality constraints.

Nonnegatively constrained minimization refers to setting D as an identity and $\mathbf{l} = 0$. Given an iteration point $\mathbf{x}_k$ and the working set W_k, we first need to check whether $\mathbf{x}_k$ minimizes the quadratic functional $\phi(\mathbf{x})$ in the subspace defined by the working set. If not, we compute a step $\mathbf{s}$ by solving an equality-constrained quadratic programming subproblem in which the constraints corresponding to the working set W_k are treated as equalities and all other constraints are temporarily ignored. So, given the iteration point $\mathbf{x}_k$, the gradient g_k, and the working set $W_k := \{j \in S : \mathbf{x}_k^j = 0\}$, the subproblem in terms of the step $\mathbf{s}_k = \mathbf{x} - \mathbf{x}_k$ can be expressed as

$$\min Q[\mathbf{s}_k] = \frac{1}{2}(A\mathbf{s}_k, \mathbf{s}_k) + (g_k, \mathbf{s}_k), \quad \text{s.t. } \mathbf{s}_k^j = 0, j \in W_k. \tag{3.2}$$

We denote the solution of Eq. (3.2) by $\mathbf{s}_k^*$. Note that the constraints in W_k were satisfied at $\mathbf{x}_k$; they are also satisfied at $\mathbf{x}_k + \alpha\mathbf{s}_k^*$, for any value α. It is clear that there is a trivial solution $\mathbf{s}_k^* = 0$. Therefore, we suppose for the moment that the optimal $\mathbf{s}_k^*$ is nonzero. We need to decide how far to move along the direction $\mathbf{s}_k^*$. The strategy is if $\mathbf{x}_k + \mathbf{s}_k^*$ is feasible with regard to all constraints, we set $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{s}_k^*$; otherwise, a line search is made in the direction $\mathbf{s}_k^*$ to find the best feasible point, i.e., we set $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k\mathbf{s}_k^*$, where α_k is the step size satisfies

$$\alpha_k := \min\left\{1, \min_{j \notin W_k, \mathbf{s}_k^j < 0} \frac{-\mathbf{x}_k^j}{\mathbf{s}_k^j}\right\}. \tag{3.3}$$

If $\alpha_k < 1$ in Eq. (3.3), then a new working set W_{k+1} is constructed by adding one active constraint. This constraint is defined by the index, say l, which achieves the minimum in Eq. (3.3), and this index is added to the active set W_k. The procedure for adding constraints on W_k is continued until a point $\mathbf{x}_k^*$ is reached that minimizes the quadratic functional over its current working set W_k^*. The first-order necessary condition for Eq. (3.10) at W_k^* yields

$$G\mathbf{s}_k^* + g_k - \sum_{j \in W_k^*} \lambda_j^* = 0, \tag{3.4}$$

$$\mathbf{s}_k^{*j} = 0, \qquad j \in W_k^*, \tag{3.5}$$

$$\lambda_j^* \geq 0, \qquad j \in W_k^*. \tag{3.6}$$

Details about implementing the algorithm are described in [37–39].

B. RAS for $p = q = 2$

Referring to our problem, we consider the nonnegatively constrained regularizing problem

$$\min J[\mathbf{f}], \qquad \text{s.t. } \mathbf{f} \geq 0, \tag{3.7}$$

where $J[\mathbf{f}]$ is given in Eq. (2.1) with $p = q = 2$, $\nu > 0$ and L a positive (semi-)definite operator. We consider using regularizing active set method for solving this problem. It is clear that Eq. (3.7) is equivalent to

$$\min \phi[\mathbf{f}] := \tfrac{1}{2}\mathbf{f}^T A\mathbf{f} - \mathbf{b}^T\mathbf{f}, \qquad \text{s.t. } \mathbf{f} \geq 0, \tag{3.8}$$

where $A = \mathcal{K}^T\mathcal{K} + \nu L^T L$, $\mathbf{b} = \mathcal{K}^T\mathbf{h}_\mathbf{n} + \nu L^T L\mathbf{f}^0$. This is clearly a special case of Eq. (3.1), therefore the active set method can be performed on the regularization form directly, except that an update of the regularization parameter ν and the matrix L is required in each iteration.

So, given the iteration point $\mathbf{f}_k$ and the working set $W_k := \{j \in S : \mathbf{f}_k^j = 0\}$, the subproblem in terms of the step $\mathbf{s}_k = \mathbf{f} - \mathbf{f}_k$ can be expressed as

$$\min \phi_{\mathbf{s}_k}[\mathbf{s}_k + \mathbf{f}_k] = \tfrac{1}{2}(A\mathbf{s}_k, \mathbf{s}_k) + (g_k, \mathbf{s}_k) + c, \quad \text{s.t. } \mathbf{s}_k^j = 0, j \in W_k, \tag{3.9}$$

with $A = \mathcal{K}^T\mathcal{K} + \nu L^T L$, $g_k = A\mathbf{f}_k - \mathcal{K}^T\mathbf{h}_\mathbf{n}$, and $c = \frac{1}{2}(A\mathbf{f}_k, \mathbf{f}_k) - (\mathcal{K}\mathbf{f}_k, \mathbf{h}_\mathbf{n})$. Since c is a constant, therefore at kth iterative step, we actually solve the equation

$$\min Q[\mathbf{s}_k] = \tfrac{1}{2}(A\mathbf{s}_k, \mathbf{s}_k) + (g_k, \mathbf{s}_k), \quad \text{s.t. } \mathbf{s}_k^j = 0, j \in W_k. \tag{3.10}$$

Based on the above preparation, the regularizing active set algorithm for multichannel image restoration problem is given as follows:

Algorithm 3.1. *(A regularizing active set (RAS) algorithm)*

Step 1. Compute a feasible starting point $\mathbf{f}_0$*; set* W_0 *to be a subset of the active constraints at* $\mathbf{f}_0$*; give the initial regularization parameter* $\nu_0 > 0$ *and the positive (semi-)definite matrix* D*; set* $k := 0$*; and compute* $A = \mathcal{K}^*\mathcal{K} + \nu L^T L$.

Step 2. Solve (3.10) *to find* $\mathbf{s}_k$*; If* $\mathbf{s}_k \neq 0$*, GOTO Step 3; otherwise, GOTO Step 4.*

Step 3. Compute α_k from (3.3)*; set* $\mathbf{f}_{k+1} = \mathbf{f}_k + \alpha_k\mathbf{s}_k$*; if* $\alpha_k = 1$*, GOTO Step 5; otherwise, find* $l \notin W_k$ *such that* $\mathbf{f}_k^l + \alpha_k\mathbf{s}_k^l = 0$ *and set* $W_k := W_k \cup \{l\}$.

Step 4. Compute the Lagrangian multipliers λ_k^j*, set* $W_k^* = W_k$*; If* $\lambda_k^j \geq 0$ *for all* $j \in W_k$*, STOP, output the solution* $\mathbf{f}^* = \mathbf{f}_k$*; Otherwise, set* $j = \text{argmin}_{j \in W_k}\lambda_k^j$*;* $\mathbf{f}_{k+1} = \mathbf{f}_k$*; set* $W_k := W_k \setminus \{j\}$*; GOTO Step 5.*

Step 5. Set $W_{k+1} := W_k$*,* $k := k + 1$ *and update regularization parameter* ν_k*, GOTO Step 2.*

Remark 1. In Step 1, the computation of $A = \mathcal{K}^*\mathcal{K} + \nu L^T L$ is not necessary if the solution method of Eq. (3.10) is given by iterative methods, say gradient type methods, since only the matrix–vector multiplication is performed. In our calculation, we adopt the feasible CG method for finding the approximate solution of Eq. (3.10), details of which are given in Appendix A and Algorithm A.1. Note that for multichannel image deblurring problems, the scale of the problems is large. Therefore, it is reasonable to consider other stopping criteria. For our problem, the stopping criterion is based on the value of the relative decrease of the residual, i.e., the ratio

$$\text{tol} = \frac{\|\mathbf{r}_k\|_{l_2}^2 - \|\mathbf{r}_{k+1}\|_{l_2}^2}{\|\mathbf{r}_{k+1}\|_{l_2}^2} \tag{3.11}$$

is less than ϵ or if a maximum number of iterations has been performed, where $\mathbf{r}_k$ is the residual at the kth iterative step, i.e., $\mathbf{r}_k = \mathcal{K}\mathbf{f}_k - \mathbf{h}_\mathbf{n}$. In our numerical examples, we choose $\epsilon = 1.0 \times 10^{-3}$.

Remark 2. It is generally agreed that a projected gradient method, which allows for the active set to change by many elements per iteration, is more efficient, particularly for large scale problems with a reasonably large active set at the solutions. However, for our problems, the nonnegativity constraints are our "*a priori* knowledge." We do not expect too many zero constrained elements in our model. Actually, zero values in an image usually mean no reflectance on the surface, which seldom occurs. Therefore, too many zero element active constraints are not common in our problem, and not many constraints will be active at the solution. This is why our one-time one-element changing active set techniques work well.

It deserves attention that for the convex quadratic programming problem, methods from [42–44] can also be applied. However, the case $p = q = 2$ is not the main concern of our paper.

C. RAS for $p, q > 0$

We consider the constrained regularizing problem

$$\min J[\mathbf{f}], \qquad \text{s.t.} \mathbf{f} \geq 0, \tag{3.12}$$

where $J[\mathbf{f}]$ is given in Eq. (2.1) with general values of $p, q > 0, \nu > 0$, and L is specified as a (semi-)positive operator. It can be the identity, for simplicity.

Recalling that the components of $\mathcal{K}$, $\mathbf{f}$, and $\mathbf{h}$ are k_{ij}, $\mathbf{f}_i$, and $\mathbf{h}_i$, $i = 1, 2, \cdots, M, j = 1, 2, \cdots, N$, respectively, and setting $k_{i1}\mathbf{f}_1 + k_{i2}\mathbf{f}_2 + \cdots + k_{in}\mathbf{f}_n - \mathbf{h}_i = r_i$, $i = 1, 2, \cdots, M$, then a straightforward calculation yields the gradient and Hessian (the matrix of the second-order partial derivatives) of $J[\mathbf{f}]$ as

$$g(\mathbf{f}) = \frac{1}{2}p\mathcal{K}^T \begin{bmatrix} |r_1|^{p-1}\text{sign}(r_1) \\ |r_2|^{p-1}\text{sign}(r_2) \\ \vdots \\ |r_m|^{p-1}\text{sign}(r_m) \end{bmatrix} + \tfrac{1}{2}\nu q L^T \begin{bmatrix} |\mathbf{f}_1 - \mathbf{f}_1^0|^{q-1}\text{sign}(\mathbf{f}_1 - \mathbf{f}_1^0) \\ |\mathbf{f}_2 - \mathbf{f}_2^0|^{q-1}\text{sign}(\mathbf{f}_2 - \mathbf{f}_2^0) \\ \vdots \\ |\mathbf{f}_n - \mathbf{f}_n^0|^{q-1}\text{sign}(\mathbf{f}_n - \mathbf{f}_n^0) \end{bmatrix} \tag{3.13}$$

and

$$\begin{aligned} H(\mathbf{f}) = {} & \frac{1}{2}p(p-1)\mathcal{K}^T \text{diag}(|r_1|^{p-2}, |r_2|^{p-2}, \cdots |r_m|^{p-2})\mathcal{K} \\ & + \frac{1}{2}\nu q(q-1)L^T \text{diag}(|\mathbf{f}_1 - \mathbf{f}_1^0|^{q-2}, \\ & \times |\mathbf{f}_2 - \mathbf{f}_2^0|^{q-2}, \cdots, |\mathbf{f}_n - \mathbf{f}_n^0|^{q-2})L, \end{aligned} \tag{3.14}$$

respectively, where sign$(\cdot)$ denotes a function that returns -1, 0, or $+1$ when the numeric expression value is negative, zero, or positive, respectively; diag$(\cdot)$ denotes a diagonal matrix whose only nonzero components are the main diagonal line. At the kth step, we define $g_k = g(\mathbf{f}_k)$ and $H_k = H(\mathbf{f}_k)$. Therefore, the minimization of $J[\mathbf{f}]$ can be approximated by solving a subproblem at the kth step,

$$\min \Psi[s] := (s, g_k) + \frac{1}{2}(H_k s, s), \tag{3.15}$$

$$\text{s.t.} \, s + \mathbf{f}_k \geq 0, \tag{3.16}$$

and the solution of the constrained problem (3.12) can be obtained iteratively by setting $\mathbf{f}_{k+1} := \mathbf{f}_k + \mathbf{s}_k$ at the kth iteration, where $\mathbf{s}_k$ is the solution of Eqs. (3.15) and (3.16). The solution of the constrained problem (3.15) and (3.16) can be solved by the Newton method within the constraint set. This process is called the feasible Newton-CG iteration. For the numerical procedure, we refer to Algorithm A.1 for details.

Now, we propose an algorithm for RAS for general values of p, $q > 0$ as follows:

Algorithm 3.2. *(A regularizing active set algorithm with feasible Newton-CG technique)*

Step 1. Choose initial values $\mathbf{f}_0$*; compute* g_0 *and* H_0*; set* $k := 0$.

Step 2. Until convergence, iteratively solve Eqs. (3.15) and (3.16) *by RAS with feasible CG iteration to give* $\mathbf{s}_k$ *and compute* $\mathbf{f}_{k+1} = \mathbf{f}_k + \mathbf{s}_k$.

Step 3. Evaluate g_{k+1} *and* H_{k+1}*; set* $k := k + 1$*; GOTO Step 2*.

In STEP 2, the stopping criterion is based on the value of the relative decrease of the residual, i.e., the ratio

$$tol = \frac{\|r_k\|_{l_p}^p - \|r_{k+1}\|_{l_p}^p}{\|r_{k+1}\|_{l_p}^p} \tag{3.17}$$

is less than ϵ or if a maximum number of iterations has been performed, where r_k is the residual at the kth iterative step with the same definition as before. In our numerical examples, we still choose $\epsilon = 1.0 \times 10^{-3}$. This is, as mentioned before, necessary for large scale scientific computing problems.

D. *A Priori* Solution

Employ an *a priori* trial solution can accelerate the convergence of an iterative regularizing algorithm and is quite useful for applications [21,45,46]. It is well known that the steepest descent gradient method is simple but slowly convergent as iteration proceeds. However, the first 1 or 2 iterations are quite fast, we should employ the information. Therefore, we use the first 2 iteration points as an *a priori* solution $\mathbf{f}^0$ and incorporate it into the regularizing model. The method uses the following iteration formula:

$$\mathbf{f}_{k+1} = \mathbf{f}_k + \lambda_k d_k, \tag{3.18}$$

where $d_k = -g_k$, $\lambda_k = -g_k^T d_k / d_k^T A d_k$ for the quadratic programming form and $\lambda_k = \mathrm{argmin}_\lambda J(\mathbf{f}_k + \lambda d_k)$ by an inexact line search for the general form, and argmin denotes the argument that minimizes the objective function J. Since the *a priori* solution is just a trial step, there is no need to solve for it exactly. In particular, we can just choose the step size λ_k as 1, i.e., we use the full length of the negative gradient step.

E. Choosing the Regularization Parameter and the Scale Matrix L

To ensure the convexity of the quadratic programming problem (3.8), as in Eqs. (3.10) and (3.2), it is necessary to choose the appropriate regularization parameter ν and the scale matrix L [21,26,46,47]. There are several ways to choose the matrix L, however to ensure fast matrix–vector multiplication, we simply choose L as the identity, i.e., the weight imposed to each element is identical.

Choosing the regularization parameter ν is also an important issue. It can be chosen by either an *a priori* technique or in an *a posteriori* way. To simplify the computation, we adopt an *a posteriori* technique in a geometric way, i.e., $\nu_{k+1} = \nu_k \xi^{k-1}$, where $\nu_0, \xi \in (0,1)$ and ξ is a proportional factor. It is clear that ν_k approaches zero as k tends to infinity. A more advanced technique for *a posteriori* choice of the regularization parameter may be used, however complex computation will weaken the efficiency of the algorithm for large scale problems.

4. Numerical Experiments

For image degradation problems, the reason for causing blur is various [25,48]. To show the efficiency of our method, we consider that the point spread function (PSF) is modeled by Gaussian. Actually, the Gaussian function simulates the convolution process of the true signal with the PSF operators well. Both the blurring due to sensors and that due to aerosols and turbulence can be taken as Gaussian. The Gaussian PSF is assumed to be spatially invariant and in the form

$$k(x-\xi, y-\eta) = \frac{1}{2\pi\rho\bar{\rho}} \exp\left(-\frac{1}{2}\left(\frac{x-\xi}{\rho}\right)^2 - \frac{1}{2}\left(\frac{y-\eta}{\bar{\rho}}\right)^2\right), \tag{4.1}$$

where ρ and $\bar{\rho}$ are positive constants. The larger we choose ρ and $\bar{\rho}$, the more f gets smoothed. So by the same argument, the smaller the value of ρ and $\bar{\rho}$ we choose, the more the convolution result resembles f. And the noise level is denoted by level, i.e.,

$$\mathbf{n} = \frac{\mathrm{level}}{N} \|\mathbf{h}\| \times \mathrm{randn}(N^2, 1),$$

where N is the size of the image, $\mathrm{randn}(N^2, 1)$ is a Gaussian normal distributed random vector, and we set *randn* ("*seed*", 0) in our codes to ensure the same random vector is generated every time.

The proposed multichannel regularizing active set method with l_p and l_q norm restoration algorithm was tested with a noisy blurred three channel true color image. The color image is a 256×256 remotely observed image. Two degradation matrices $\mathcal{K}$ were used. The first one introducing only within-channel blurring and is given by

$$\mathcal{K} = \begin{bmatrix} \mathcal{K}_{11} & 0 & \cdots & 0 \\ 0 & \mathcal{K}_{22} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \mathcal{K}_{NN} \end{bmatrix},$$

where $N = 3$, each $\mathcal{K}_{ii}$ is an $M^2 \times M^2$ block Toeplitz submatrix. The second blurring matrix introduces both within-channel and between-channel blurring and is given by

$$\mathcal{K} = \begin{bmatrix} a_{11}\mathcal{K}_{11} & a_{12}\mathcal{K}_{12} & \cdots & a_{13}\mathcal{K}_{1N} \\ a_{21}\mathcal{K}_{21} & a_{22}\mathcal{K}_{22} & \cdots & a_{23}\mathcal{K}_{2N} \\ \vdots & \vdots & \cdots & \vdots \\ a_{N1}\mathcal{K}_{N1} & a_{N2}\mathcal{K}_{N2} & \cdots & a_{N3}\mathcal{K}_{NN} \end{bmatrix},$$

with $N = 3$. Here, each $\mathcal{K}_{ij}$ is again an $M^2 \times M^2$ block Toeplitz submatrix.

The main cost of our algorithm is the matrix–vector multiplication, so an efficient algorithm to compute the matrix–vector multiplication should be investigated. Since the PSF kernel function is spatially invariant, the kernel is separable and can be reformulated as

$$k(x-\xi, y-\eta) = k_x(x-\xi) k_y(y-\eta). \tag{4.2}$$

Numerically, assume that the discretization of k_x and k_y are $\mathcal{K}_{x,ij}$ and $\mathcal{K}_{y,ij}$, respectively, then the matrix $\mathcal{K}_{ij}$ is a tensor of $\mathcal{K}_{x,ij}$ and $\mathcal{K}_{y,ij}$, i.e., the Kronecker product of $\mathcal{K}_{x,ij}$ and $\mathcal{K}_{y,ij}$,

$$\mathcal{K}_{ij} = \mathcal{K}_{x,ij} \otimes \mathcal{K}_{y,ij}. \tag{4.3}$$

$\mathcal{K}_{x,ij}$ and $\mathcal{K}_{y,ij}$ are all Toeplitz matrices, so $\mathcal{K}_{ij}$ is a block Toeplitz matrix with Toeplitz blocks, i.e., the $N^2 \times N^2$ matrix $\mathcal{K}_{ij}$ has the block form

$$\mathcal{K}_{ij} = \begin{pmatrix} \mathcal{K}_0 & \mathcal{K}_{-1} & \cdots & \mathcal{K}_{1-N} \\ \mathcal{K}_1 & \mathcal{K}_0 & \mathcal{K}_{-1} & \vdots \\ \vdots & \ddots & \ddots & \mathcal{K}_{-1} \\ \mathcal{K}_{N-1} & \cdots & \mathcal{K}_1 & \mathcal{K}_0 \end{pmatrix} \tag{4.4}$$

where each block $\mathcal{K}_j$ is an $N \times N$ Toeplitz matrix. For a Toeplitz matrix, it can be determined by its first row and first column elements. By extending the block Toeplitz matrix with Toeplitz blocks into a block circulant with circulant blocks matrix $\tilde{\mathcal{K}}_{ij}$, we can use the two-dimensional discrete Fourier

transform to compute the matrix vector multiplication [26].

The block circulant with circulant blocks matrix can be decomposed as

$$\tilde{\mathcal{K}}_{ij} = \mathcal{F}^{\star}\Lambda\mathcal{F},$$

where $\mathcal{F}$ is the two-dimensional discrete Fourier transform matrix and Λ is a diagonal matrix containing the eigenvalues of $\tilde{\mathcal{K}}_{ij}$. And the eigenvalues of $\tilde{\mathcal{K}}_{ij}$ can be obtained by computing a two-dimensional discrete Fourier transform of the first column of $\tilde{\mathcal{K}}_{ij}$, so we can compute $\tilde{\mathcal{K}}_{ij}x$ by $\mathcal{F}^{\star}\Lambda\mathcal{F}x$. Note that discrete Fourier transforms can be computed at a low computational cost by utilizing the fast Fourier transform. Therefore, the Fourier transform of an m-vector (signal) can be computed in $O(m \times \log_2 m)$ operations.

The precision of the approximation is characterized by the root-mean-square error (rmse)

$$\text{rmse} = \sqrt{\frac{1}{NM^2}\sum_{i=1}^{NM^2}\frac{((\mathcal{K}f^*)_i - \mathbf{h}_i)^2}{((\mathcal{K}f^*)_i)^2}},$$

which describes the average relative deviation of the retrieved signals from the true signals.

A. Experiments for the Form of Quadratic Programming

First, we set $p = q = 2$. This leads to the classical regularization, which corresponds to a quadratic programming problem (3.7) with nonnegativity constraints.

In our numerical experiments, the Gaussian PSF is the integral kernel k, so $\mathcal{K}$ can be represented by a Kronecker product of two low order matrices as $\mathcal{K} = A \otimes B$ with $A \in \mathbb{R}^{m\times m}, B \in \mathbb{R}^{n\times n}$. For the blurring process, A and B are taken to be sparse-banded matrices [25], meaning only pixels within a distance band -1 contribute to the blurring.

Experiment 1. We suppose the half-bands of R, G, and B channels are 5, 4, and 4, respectively, and the values of $\rho = \bar{\rho}$ of R, G, and B channels are 0.6, 0.7, and 0.6, respectively. The blurring matrix is given by

$$\begin{bmatrix} a_{11}\mathcal{K}_{11} & a_{12}\mathcal{K}_{12} & a_{13}\mathcal{K}_{13} \\ a_{21}\mathcal{K}_{21} & a_{22}\mathcal{K}_{22} & a_{23}\mathcal{K}_{22} \\ a_{31}\mathcal{K}_{31} & a_{32}\mathcal{K}_{33} & a_{33}\mathcal{K}_{33} \end{bmatrix}. \tag{4.5}$$

The coefficients a_{ij} are chosen as $a_{11} = 1$, $a_{12} = 0$, $a_{13} = 0, a_{21} = 0, a_{22} = 1, a_{23} = 0$ and $a_{31} = 0, a_{32} = 0$, $a_{33} = 1$. The input original three channel image, blurred image, and restoration are illustrated in Figs. 1–3, respectively.

Experiment 2. We suppose the half-bands of R, G, and B channels are 5, 4, and 4, respectively, and the values of $\rho = \bar{\rho}$ of R, G, and B channels are 0.6, 0.7, and 0.6, respectively. The blurring matrix is given by Eq. (4.5), but the coefficients a_{ij} are chosen as $a_{11} = 0.8$, $a_{12} = 0.1$, $a_{13} = 0.1$, $a_{21} = 0.1$, $a_{22} = 0.8$, $a_{23} = 0.1$ and $a_{31} = 0.1$, $a_{32} = 0.1$, $a_{33} = 0.8$. The blurred three channel image and restoration are illustrated in Figs. 4 and 5.

The within- and between-channel restoration results indicate that their blurring effects are different. And it is more difficult to recover the actual input from between-channel blurs. To characterize the degree of approximation and the cost of computation, we list in Table 1 the rmses and the CPU time. The CPU time consumption and small rmses in within-channel indicate that the method is applicable for within-channel deblurring. But it is not applicable for between-channel deblurring since the values of rmses are greater than 0.1 in some channels.

B. Experiments for $p, q \in [1, 2]$

For general values of p and q, the object functional $J(\mathbf{f})$ of problem (2.1) may be nonsmooth. In this case, the model corresponds to the nonsmooth regularization and optimization model.

Experiment 1. We consider the case $p = 2$ and $q = 1$ in this experiment. We suppose the half-bands of the R, G, and B channels are 5, 4, and 4, respectively, and the values of $\rho = \bar{\rho}$ of the R, G, and B channels are 0.6, 0.7, and 0.6, respectively. The form of blurring matrix is given by Eq. (4.5). The coefficients a_{ij} are chosen as $a_{11} = 1$, $a_{12} = 0$, $a_{13} = 0$, $a_{21} = 0$, and $a_{31} = 0$, $a_{32} = 0$, $a_{33} = 1$. The recovered image is illustrated in Fig. 6.

Experiment 2. We consider the case $p = 2$ and $q = 1$ in this experiment. Again we suppose the half-bands of the R, G, and B channels are 5, 4 and 4 respectively, and the values of $\rho = \bar{\rho}$ of the R, G, and B channels are 0.6, 0.7, and 0.6, respectively. The blurring matrix is given by Eq. (4.5). The coefficients a_{ij} are chosen as $a_{11} = 0.8$, $a_{12} = 0.1$, $a_{13} = 0.1$, $a_{21} = 0.1$, $a_{22} = 0.8$, $a_{23} = 0.1$ and $a_{31} = 0.1$, $a_{32} = 0.1$,

Fig. 1. (Color online) Original input three channel image.

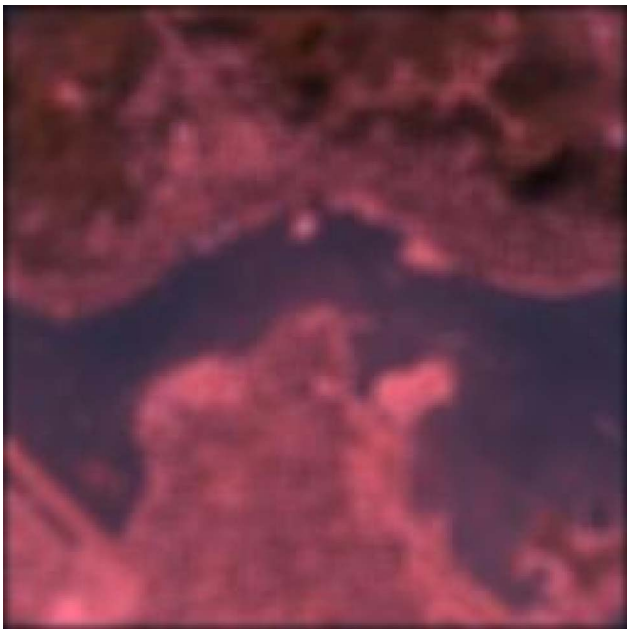

Fig. 2. (Color online) Blurred noisy image with noise level equaling 0.01.

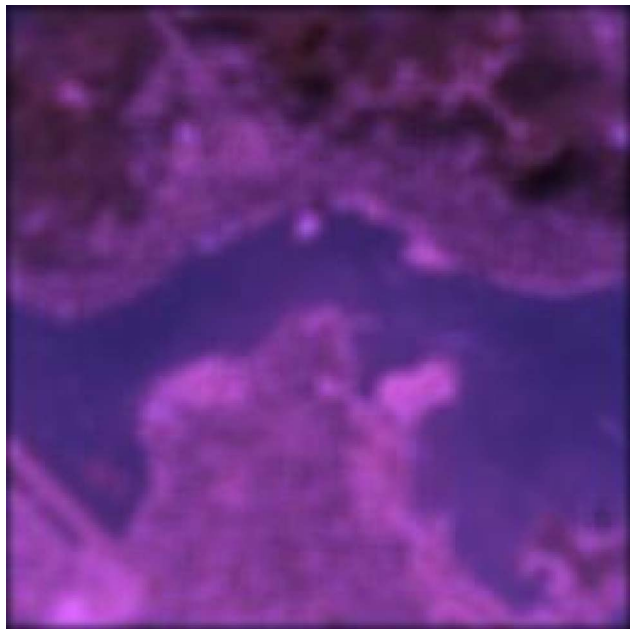

Fig. 4. (Color online) Blurred noisy image with noise level equaling 0.01.

$a_{33} = 0.8$. The recovered image is illustrated in Fig. 7.

Experiment 3. We consider the case $p = 1.5$ and $q = 1$ in this experiment. Again we suppose the half-bands of the R, G, and B channels are 5, 4, and 4, respectively, and the values of $\rho = \bar{\rho}$ of the R, G, and B channels are 0.6, 0.7, and 0.6, respectively. The blurring matrix is given by Eq.(4.5). The coefficients a_{ij} are chosen as $a_{11} = 1$, $a_{12} = 0$, $a_{13} = 0$, $a_{21} = 0$, $a_{22} = 1$, $a_{23} = 0$ and $a_{31} = 0$, $a_{32} = 0$, $a_{33} = 1$. The recovered image is illustrated in Fig. 8.

Experiment 4. We consider the case $p = 1.5$ and $q = 1$ in this experiment. Again we suppose the half-bands of the R, G, and B channels are 5, 4, and 4, respectively, and the values of $\rho = \bar{\rho}$ of the R, G, and B channels are 0.6, 0.7, and 0.6, respectively. The blurring matrix is given by Eq. (4.5). The coefficients a_{ij} are chosen as $a_{11} = 0.8, a_{12} = 0.1, a_{13} = 0.1$, $a_{21} = 0.1$, $a_{22} = 0.8$, $a_{23} = 0.1$ and $a_{31} = 0.1$, $a_{32} = 0.1$, $a_{33} = 0.8$. The recovered image is illustrated in Fig. 9.

Experiment 5. We consider the case $p = 1.6$ and $q = 1.1$ in this experiment. Again we suppose the

Fig. 3. (Color online) Within-channel restoration: the restored image for $\nu = 0.005$ and noise level = 0.01.

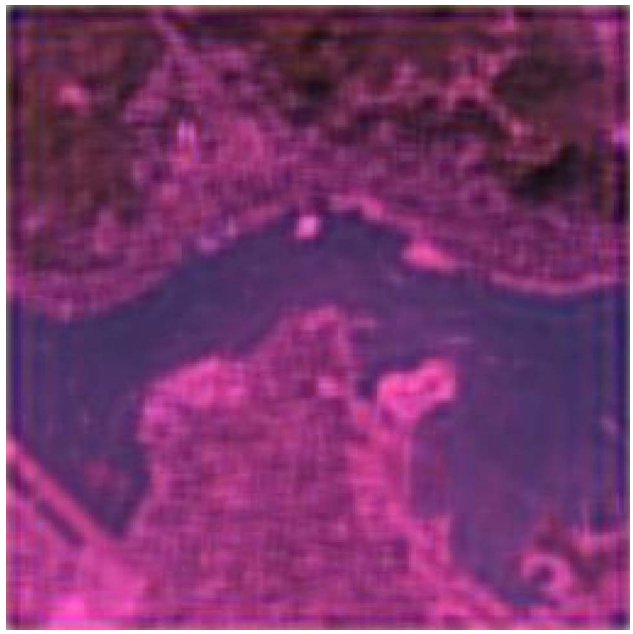

Fig. 5. (Color online) Between-channel restoration: the restored image for $\nu = 0.005$ and noise level = 0.01.

Table 1. RMSEs and Variance Differences for R, G, and B Channels and CPU Time (seconds) of Our Regularizing Algorithm for the Quadratic Inversion Model

	Within-Channel	Between-Channel
CPU	323.4219	72.3281
RMSE (R)	0.0133	0.1356
RMSE (G)	0.0196	0.7195
RMSE (B)	0.0170	0.0513

half-bands of the R, G, and B channels are 5, 4, and 4, respectively, and the values of $\rho = \bar{\rho}$ of R, G, and B channels are 0.6, 0.7, and 0.6, respectively. The blurring matrix is given by Eq. (4.5). The coefficients a_{ij} are chosen as $a_{11} = 0.8, a_{12} = 0, a_{13} = 0$, and $a_{31} = 0$, $a_{32} = 0$, v. The recovered image is illustrated in Fig. 10.

Experiment 6. We consider the case $p = 1.6$ and $q = 1.1$ in this experiment. Again we suppose the half-bands of the R, G, and B channels are 5, 4, and 4, respectively, and the values of $\rho = \bar{\rho}$ of the R, G, and B channels are 0.6, 0.7, and 0.6, respectively. The blurring matrix is given by Eq. (4.5). The coefficients a_{ij} are chosen as $a_{11} = 0.8, a_{12} = 0.1, a_{13} = 0.1$, $a_{21} = 0.1$, $a_{22} = 0.8$, $a_{23} = 0.1$ and $a_{31} = 0.1$, $a_{32} = 0.1$, $a_{33} = 0.8$. The recovered image is illustrated in Fig. 11.

As mentioned before, the blurring effects for within- and between-channel are different. But with the nonquadratic $l_p - l_q$ model and using the regularizing active set method, stable recoveries are obtained. Again, to characterize the degree of approximation and the cost of computation, we list in Table 2 the rmses and the CPU time. The values of rmses reveal that the nonquadratic $l_p - l_q$ model is more suitable for recovering ill-posed nonregular multichannel images.

Fig. 7. (Color online) Between-channel restoration: the restored image for $\nu = 0.005$ and noise level $= 0.01$ in the case $p = 2$ and $q = 1$.

C. Discussion

Large values of p and q for RAS can be also implemented easily with the description of our algorithm. However we find that results are not so satisfactory. For example, for $p = 2$ and $q = 4$, as is used in [10,11], the restorations cannot be accepted. We think the techniques for choosing proper weight

Fig. 6. (Color online) Within-channel restoration: the restored image for $\nu = 0.005$ and noise level $= 0.01$ in the case $p = 2$ and $q = 1$.

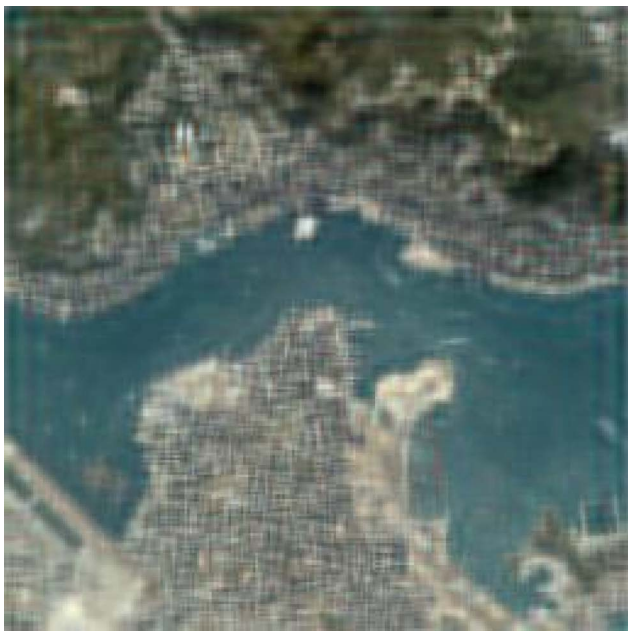

Fig. 8. (Color online) Within-channel restoration: the restored image for $\nu = 0.005$ and noise level $= 0.01$ in the case $p = 1.5$ and $q = 1$.

Fig. 9. (Color online) Between-channel restoration: the restored image for $\nu = 0.005$ and noise level level $= 0.01$ in the case $p = 1.5$ and $q = 1$.

Fig. 11. (Color online) Between-channel restoration: the restored image for $\nu = 0.005$ and noise level $= 0.01$ in the case $p = 1.6$ and $q = 1.1$.

factors for the residual part and the regularized part developed in [10,11] should be employed. However this is beyond the scope of this paper.

The pivot point for the success of the RAS is that it incorporates solving nonnegatively constrained minimization problem naturally, and this is particularly useful for physical problems, say aerosol particle distribution and tomography [40,46].

The numerical results indicate that the active set method for $l_p - l_q$ regularizing model with $1 \le p, q \le 2$ performs better than the active set method for the quadratic model. The use of $p = q = 2$ is best suited for problems with a smooth solution and normally distributed noise in observation. However, irregular observation data blurred and contaminated by turbulence or aerosols lose their smooth property. As is often the case, there are many different types of outliers. This explains why the general $l_p - l_q$ regularizing model gets smaller rmse values than that from quadratic model.

Fig. 10. (Color online) Within-channel restoration: the restored image for $\nu = 0.005$ and noise level $= 0.01$ in the case $p = 1.6$ and $q = 1.1$.

5. Concluding Remarks

This research focused on the development of regularizing active set methods for simulating the nonnegatively constrained multichannel image restoration

Table 2. RMSEs for R, G, and B Channels and CPU Time (seconds) of Our Regularizing Algorithm for the General $I_p - I_q$ Inversion Model

	Within-Channel	Between-Channel
	$p = 2, q = 1.0$	
CPU	342.2500	469.8750
RMSE (R)	0.0133	0.0143
RMSE (G)	0.0196	0.0249
RMSE (B)	0.0170	0.0112
	$p = 1.5, q = 1.0$	
CPU	283.9375	272.1719
RMSE (R)	0.0146	0.0255
RMSE (G)	0.0153	0.0236
RMSE (B)	0.0164	0.0262
	$p = 1.6, q = 1.1$	
CPU	1120.0781	223.5625
RMSE (R)	0.0089	0.0253
RMSE (G)	0.0124	0.0262
RMSE (B)	0.0113	0.0262

problems. The scope of this research is to develop a general $l_p - l_q$ regularizing model and its quadratic approximation, which could characterize regularizing properties in different purpose of numerical inversion.

In optimization research fields, for a general l^p and l^q combined optimization problem, it is convenient to solve it by quasi-Newton methods instead of the exact Newton method. For example, the DFP method, BFGS, L-BFGS, and trust region method [32,37,39,49]. However, for image degradation problems, these methods may lose the structure of the original discrete matrix operator, hence fast matrix–vector multiplication is lost. Therefore, how to use these methods while keeping the structure of the kernel matrix is still an interesting topic. In addition, for $p = 2$ and $q \to 1$ or $p = 2$ and $q \to 0$, the minimization problem corresponding to finding sparse solutions (due to a band-limited system), linear programming methods may be applied. Therefore our model is also useful for problems of recovering sparse model parameters in geophysics [19,20] and for compressive sampling and decoding of signals [50].

Appendix A: Feasible CG Method

We consider the minimization problem

$$\min J[\mathbf{f}], \qquad \text{s.t.}\, \mathbf{f} \geq 0, \tag{A.1}$$

where $J[\mathbf{f}]$ is a nonlinear functional, given in Eq. (3.12). At the kth iteration, the search direction $\mathbf{s}_k$ is computed from

$$\min \Psi[\mathbf{s}_k], \qquad \text{s.t.}\, \tilde{\mathbf{s}}_k^j = 0, j \in W_k, \tag{A.2}$$

where $\tilde{\mathbf{s}}_k^j := \mathbf{s}_k^j + \mathbf{f}_k$. $\Psi[\mathbf{s}_k]$ is a nonlinear function, given in Eq. (3.15). Then $\mathbf{f}_{k+1} = \mathbf{f}_k + \alpha_k \mathbf{s}_k$. The gradient of $\Psi[\mathbf{s}_k]$ is denoted by $\text{grad}_k[\Psi] = H_k \mathbf{s}_k + g_k$.

Initialization for choosing W_0: The initial working set W_0 is related to the initial point $\tilde{\mathbf{s}}_0$. Since the constraint is nonnegative, the components of W_0 can be chosen as i if the ith component of $\tilde{\mathbf{s}}_0$ equals zero, and zero, otherwise. This means that the ith constraint of W_0 is active. The index i is from 1 to N.

Solving the quadratic problem (3.15): Since we are interested in finding the feasible direction $\mathbf{s}_k$, it is unnecessary to solve Eq. (3.15) accurately. We apply a feasible direction of descent method with a CG solution. First, we address the basic concept and procedures of feasible direction of descent methods.

The fundamental concept of feasible direction methods is that of the feasible direction of descent. Denote by $S = \{\mathbf{f} : \mathbf{f} \geq 0\}$. If $\mathbf{f} \in S$, then $\mathbf{s} \neq 0$ is called a feasible direction of descent for $\mathbf{f}$ if there exists α_{upper} such that for all $\alpha \in (0, \alpha_{\text{upper}})$ the following two properties hold: (1) $\mathbf{f} + \alpha\mathbf{s} \in S$; (2) $J[\mathbf{f} + \alpha\mathbf{s}] < J[\mathbf{f}]$. Note that condition (2) is equivalent to requiring that $\text{grad}[J]^T\mathbf{s} < 0$.

The basic steps in feasible direction methods involve solving a nonlinear programming subproblem to find the direction vector and then finding the step size along this direction by performing a constrained one-dimensional line search. After updating the current point, the above steps are repeated until the termination criterion is satisfied.

Based on above comments, the feasible direction of Eq. (3.15) is the vector in null space. There are several ways for solving the nonlinear programming subproblem, such as the steepest descent method, CG method, and Newton and quasi-Newton methods [39]. Due to the fact that the model is quadratic, we apply the CG method, which is fast and efficient. To describe the algorithm, we use the following notations: $G := H_k$ and $g := g_k$.

Algorithm A.1. *(Feasible CG algorithm)*

Step 1. *Input* $\mathbf{s}_0$ *(such that* $n_0 \in S$*)*; compute $\text{grad}_0[\Psi] := G\mathbf{s}_0 + g$ *and such that* $\text{grad}_0[\Psi]$ *is a feasible direction; set* $k := 1$.

Step 2. *If the stopping criterion is satisfied, output* $\mathbf{s}^* = \mathbf{s}_{k-1}$, *STOP; otherwise, set* $\mathbf{z}_{k-1} := -\text{grad}_{k-1}[\Psi]$, $\rho_{k-1} := \mathbf{z}_{k-1}^T \mathbf{z}_{k-1}$.

Step 3. *Compute next iteration points:*
$\alpha_k := \rho_{k-1}/(\mathbf{s}_{k-1}^T(G\mathbf{s}_{k-1}))$,
$\mathbf{s}_k := \mathbf{s}_{k-1} + \alpha_k \mathbf{z}_{k-1}$,
$\text{grad}_k[\Psi] := \text{grad}_{k-1}[\Psi] +$
$\alpha_k G\mathbf{s}_k$*(such that* $\text{grad}_k[\Psi]$*is a feasible direction*),
$\rho_k := \text{grad}_k[\Psi]^T \text{grad}_k[\Psi]$,
$\beta_k := \rho_k/\rho_{k-1}$,
$\mathbf{z}_k := -\text{grad}_{k-1}[\Psi] + \beta_k \mathbf{z}_{k-1}$.

Step 4. *If* k *exceeds the maximum iterative steps, output* $\mathbf{s}^* = s_k$ *STOP; otherwise, set* $k = k + 1$, *GOTO Step 2.*

In our calculation, we choose the initial values of $s_0$as a vector with components all equaling 0.5. We want to mention that for ill-posed problems the stopping criterion must be carefully chosen. Here we choose the stopping criterion in Step 2 as follows: we define

$$\|\text{grad}_k[\Psi]\| \leq \varrho \|\text{grad}_0[\Psi]\|_2, \tag{A.3}$$

where ϱ is a preassigned tolerance or dominant parameter and $\text{grad}_k[\Psi]$ is defined as

$$(\text{grad}_k[\Psi])_i \begin{cases} (\text{grad}_k[\Psi])_i & \text{if}(s_k)_i > 0 \\ \min\{(\text{grad}_k[\Psi])_i, 0\} & \text{if}(s_k)_i > 0 \end{cases}.$$

The role of ϱ is controlling further iteration at preassigned precision. The choice of ϱ is dependent on the degree of ill-posedness of problems. Less value of ϱ yields better approximation but induces more iterative steps and results in more CPU time; larger value of ϱ require fewer iterative steps and less CPU time but yield insufficient approximation. In our numerical examples, we choose $\varrho = 2.0 \times 10^{-3}$. Empirically, we recommend choosing ϱ between 1.0×10^{-4} and 1.0×10^{-2}.

We thank the referees for their valuable suggestions in revising the paper. This work is supported by the National Science Foundation of China (NSFC) under grants 10871191 and 10831006 and Chinese Academy of Sciences grant kjcx-yw-s7-03. It is also partly supported by the National "973" Key Basic Research Developments Program of China under grant no. 2005CB422104 and no. 2007CB714400.

References

1. G. Angelopoulos and I. Pitas, "Multichannel Wiener filters in color image restoration," IEEE Trans. Circuits Syst. Video Technol. **4**, 83–87 (1994).
2. D. L. Angwin and H. Haufman, "Effects of modeling domains on recursive color image restoration," in IEEE Trans. Acoust., Speech, Signal Process. **12**, 1229–1231 (1987).
3. A. Beck, A. Ben, and Y. C. Eldar, "Robust mean-squared error estimation of multiple signals in linear systems affected by model and noise uncertainties," Math. Program. **107**, 155–187 (2006).
4. J. Bescos, J. H. Altamirano, A. Santisteban, and J. Santamaria, "Digital restoration models for color imaging," Appl. Opt. **27**, 419–425 (1988).
5. T. W. S. Chow, X. D. Li, and K.-T. Ng, "Double-regularization approach for blind restoration of multichannel imagery," IEEE Trans. Circuits Syst. I Fundam. Theory Appl. **48**, 1075–1085 (2001).
6. Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Robust mean-squared error estimation in the presence of model uncertainties," IEEE Trans. Signal Process. **53**, 168–181 (ISPRS Working Groups WG VII/1, 2005).
7. N. P. Galatsanos and R. T. Chin, "Digital restoration of multichannel images," IEEE Trans. Acoust., Speech Signal Process. **37**, 415–421 (1989).
8. N. P. Galatsanos and R. T. Chin, "Restoration of color images by multichannel Kalman filtering," IEEE Trans. Signal Process. **39**, 2237–2252 (1991).
9. N. P. Galatsanos, A. K. Katsaggelos, R. T. Chin, and A. D. Hillery, "Least squares restoration of multichannel images," IEEE Trans. Signal Process. **39**, 2222–2236 (1991).
10. M.-C. Hong, T. Stathaki, and A. K. Katsasggelos, "Iterative regularized least-mean mixed-norm image restoration," Opt. Eng. **41**, 2525–2524 (2002).
11. M.-C. Hong, T. Stathaki, and A. K. Katsasggelos, "Iterative regularized mixed norm multichannel image restoration," J. Electron. Imaging **14**, 013004 (2005).
12. B. R. Hunt and O. Kubler, "Karhunen–Loeve multispectral image restoration, part I: theory," IEEE Trans. Acoust., Speech Signal Process. **ASSP-32**, 592–600 (1984).
13. A. K. Katsaggelos and J. K. Paik, "Iterative color image restoration algorithms," IEEE Trans. Acoust., Speech Signal Process. **2**, 1028–1031 (1988).
14. N. Ohyama, M. Yachida, E. Badique, J. Tsujiuchi, and T. Honda, "Least-squares filter for color-image restoration," J. Opt. Soc. Am. A **5**, 19–24 (1988).
15. R. R. Schultz and R. L. Stevenson, "Stochastic modeling and estimation of multispectral image data," IEEE Trans. Image Process. **4**, 1109–1119 (1995).
16. M. A. Sezan and H. J. Trussell, "Use of *a priori* knowledge in multispectral image restoration," in IEEE Trans. Acoust., Speech Signal Process. **3**, 1429–1432 (1989).
17. A. M. Tekalp and G. Pavlovic, "Space-variant and color image restoration using Kalman filtering," in *Multidimensional Signal Processing Workshop*, Vol. 6 (IEEE, 1989), pp. 186–187
18. W. W. Zhu, Nikolas P. Galatsanos, and A. K. Katsaggelos, "Regularized multichannel image restoration using cross-validation," Graphical Models **57**, 345–356 (1995).
19. Y. F. Wang, X. W. Li, S. Q. Ma, H. Yang, Z. Nashed, and Y. N. Guan, "BRDF model inversion of multiangular remote sensing: ill-posedness and interior point solution method," in *Proceedings of the 9th International Symposium on Physical Measurements and Signature in Remote Sensing (ISPMSRS)* (ISPRS Working Groups WG VII/1, 2005), Vol. XXXVI, pp. 328–330.
20. Y. F. Wang, S. Q. Ma, H. Yang, J. D. Wang, and X. W. Li, "On the effective inversion by imposing *a priori* information for retrieval of land surface parameters," Sci. China (Ser. D) **52**, 1–10 (2009).
21. H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems* (Kluwer, 1996).
22. T. Y. Xiao, Sh. G. Yu, and Y. F. Wang, *Numerical Methods for the Solution of Inverse Problems* (Beijing: Science Press, 2003).
23. L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," Physica D (Amsterdam) **60**, 259–268 (1992).
24. C. R. Vogel and M. E. Oman, "A fast, robust algorithm for total variation based reconstruction of noisy, blurred images," IEEE Trans. Image Process. **7**, 813–824 (1998).
25. P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion* (Society for Industrial and Applied Mathematics, 1998).
26. C. R. Vogel, *Computational Methods for Inverse Problems* (Society for Applied and Industrial Mathematics, 2002).
27. R. Barrett, M. Berry, T. F. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and Vorst H. Van der, *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods* (Society for Applied and Industrial Mathematics, 1994).
28. Y. F. Wang and S. Q. Ma, "Projected Barzilai–Borwein method for large scale nonnegative image restoration," Inverse Probl. Sci. Eng. **15**, 559–583 (2007).
29. M. Hanke and J. Nagy, "Restoration of atmospherically blurred images by symmetric indefinite conjugate gradient techniques," Inverse Probl. **12**, 157–173 (1996).
30. Y. F. Wang, "On the regularity of trust region-CG algorithm: with application to image deconvolution problem," Sci. China Ser. A Math., Phys., Astron. **46**, 312–325 (2003).
31. Z. W. Wen and Y. F. Wang, " A new trust region algorithm for image restoration," Sci. China Ser. A Math., Phys., Astron. **48**, 169–184 (2005).
32. Y. F. Wang and Y. Yuan, "Convergence and regularity of trust region methods for nonlinear ill-posed inverse problems," Inverse Probl. **21**, 821–838 (2005).
33. J. Bardsley and C. R. Vogel, "A nonnegatively constrained convex programming method for image reconstruction," SIAM J. Sci. Comput. **25**, 1326–1343 (2004).
34. M. Hanke, J. Nagy, and C. R. Vogel, "Quasi-Newton approach to nonnegative image restorations," Numer. Linear Algebra Appl. **316**, 223–236 (2000).
35. J. Nagy and Z. Strakos, "Enforcing nonnegativity in image reconstruction algorithms," Proc. SPIE **4121**, 182–190 (2000).
36. M. Rojas and T. Steihaug, "An interior-point trust-region-based method for large-scale nonnegative regularization," CERFACS Tech. Rep. TR/PA/01/11 (Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique, 2002).
37. R. Fletcher, *Practical Methods of Optimization*, 2nd ed. (Wiley,, 1987).
38. J. Nocedal and S. J. Wright, *Numerical Optimization* (Springer-Verlag, 1999).

39. Y. X. Yuan, *Numerical Methods for Nonlinear Programming* (Shanghai Science and Technology Publication, 1993).
40. Y. F. Wang and C. C. Yang, "A regularizing active set method for retrieval of atmospheric aerosol particle size distribution function," J. Opt. Soc. Am. A **25**, 348–356 (2008).
41. Y. X. Yuan, "Subspace techniques for nonlinear optimization," in *Some Topics in Industrial and Applied Mathematics*, R. Jeltsch, D. Q. Li, and I. H. Sloan, eds., Series in Contemporary Applied Mathematics CAM 8 (Higher Education Press, (2007), pp. 206–218.
42. J. J. More and G. Toraldo, "On the solution of large quadratic programming problems with bound constraints," SIAM J. Control Optim. **1**, 93–113 (1991).
43. D. P. O'Leary, "A generalized conjugate gradient algorithm for solving a class of quadratic programming problems," Numer. Linear Algebra Appl. **34**, 371–399 (1980).
44. D. P. Bertsekas, "Projected Newton methods for optimization problems with simple constraints," SIAM J. Control Optim. **20**, 221–246 (1982).
45. A. Bakushinsky and A. Goncharsky, *Ill-Posed Problems: Theory and Applications* (Kluwer Academic, 1994).
46. Y. F. Wang, *Computational Methods for Inverse Problems and Their Applications* (Higher Education Press, 2007).
47. Y. F. Wang, S. F. Fan, and X. Feng, "Retrieval of the aerosol particle size distribution function by incorporating *a priori* information," J. Aerosol Sci. **38** 885–901 (2007).
48. M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging* (Institute of Physics, 1998).
49. Y. X. Yuan and W. Y. Sun, *Theory and Methods for Optimization* (Science Press, 1997).
50. D. Donoho, "Compressed sensing," IEEE Trans. Inf. Theory **52**, 1289–1306 (2006).