Structured Total Least Squares for Color Image Restoration

Fu, Haoying; Ng, Michael K.; Barlow, Jesse L.

Published in:
SIAM Journal on Scientific Computing

DOI:
10.1137/040605436

Published: 04/08/2006

Document Version:
Publisher's PDF, also known as Version of record

SIAM J. SCI. COMPUT.
Vol. 28, No. 3, pp. 1100–1119

STRUCTURED TOTAL LEAST SQUARES FOR COLOR IMAGE RESTORATION*

HAOYING FU†, MICHAEL K. NG‡, AND JESSE L. BARLOW†

Abstract. The problem of 3×3 color mixing image restoration is considered. The blurring matrices, as well as the observed image, are contaminated by noise; therefore the total least squares (TLS) method is employed to restore the original image. Since the blurring matrices are also structured, we apply the structured total least squares (STLS) method [J. B. Rosen, H. Park, and J. Glick, *SIAM J. Matrix Anal. Appl.*, 17 (1996), pp. 110–126]. The blurring matrices are generally ill conditioned; thus Tikhonov's regularization is used to stabilize the solution. Since Neumann boundary conditions are used in the restoration process, the discrete cosine transform (DCT) based preconditioner is effective for the linear systems encountered in our STLS algorithm.

Key words. structured total least squares, color image, image restoration, Toeplitz-like matrices, preconditioners

AMS subject classifications. 65F10, 65F15

DOI. 10.1137/040605436

1. Introduction.

The color image restoration problem has attracted much research effort recently [2, 8, 9, 12, 15, 17, 22]. A color image may be regarded as a set of three images in the three color components, red, green, and blue. A more general term, multichannel images, is often used to denote color images where each color component is considered as a channel.

In most applications [2], the observed image is distorted by blur and additive noise. Let $f_r(x,y)$, $f_g(x,y)$, and $f_b(x,y)$ be the original red, green, and blue scenes. Let $u_r(x,y)$, $u_g(x,y)$, and $u_b(x,y)$ be the observed red, green, and blue scenes. Let $h_{ij}(x,y)$ (for $i,j \in \{r,g,b\}$) be the within-channel and cross-channel point spread functions, normalized such that each function integrates to 1. Let $n_r(x,y)$, $n_g(x,y)$, and $n_b(x,y)$ be the noise present in the observed image. The image formation process can be modeled as follows [8, 9]:

$$\begin{aligned} u_r &= w_{rr}h_{rr} \star f_r + w_{gr}h_{gr} \star f_g + w_{br}h_{br} \star f_b + n_r, \\ u_g &= w_{rg}h_{rg} \star f_r + w_{gg}h_{gg} \star f_g + w_{bg}h_{bg} \star f_b + n_g, \\ u_b &= w_{rb}h_{rb} \star f_r + w_{gb}h_{gb} \star f_g + w_{bb}h_{bb} \star f_b + n_b. \end{aligned} \tag{1.1}$$

Here w_{ij} (for $i,j \in \{r,g,b\}$) are nonnegative weights that measure the contribution of the original channels to each observed channel, and $\star$ denotes two-dimensional (2D) convolution. We note that at least one of w_{ri}, w_{gi}, and w_{bi} ($i \in \{r,g,b\}$) is nonzero.

Let $\mathbf{f}_i$, $\mathbf{u}_i$, and $\mathbf{n}_i$ (for $i \in \{r,g,b\}$) be the row by row ordering of the discretized original scene, observed scene, and additive noise, respectively. Let H_{ij} (for $i,j \in \{r,g,b\}$) be the blurring matrices of appropriate size built according to the discretized

*Received by the editors March 21, 2004; accepted for publication (in revised form) March 3, 2006; published electronically July 7, 2006.
http://www.siam.org/journals/sisc/28-3/60543.html

†Department of Computer Science and Engineering, Pennsylvania State University, University Park, PA 16802 (hfu@cse.psu.edu, barlow@cse.psu.edu). Part of this work was done while the first author was visiting the University of Hong Kong Mathematics Department.

‡Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong (mng@math.hkbu.edu.hk). The research of this author was supported in part by Hong Kong Research Grants Council grants 7046/03P, 7035/04P, and 7035/05P and HKBU FRGs.

point spread functions $\mathbf{h}_{ij}$. Then the discretized image formation process can be put into a matrix-vector form:

$$\begin{aligned} \mathbf{u}_r &= w_{rr}H_{rr}\mathbf{f}_r + w_{gr}H_{gr}\mathbf{f}_g + w_{br}H_{br}\mathbf{f}_b + \mathbf{n}_r, \\ \mathbf{u}_g &= w_{rg}H_{rg}\mathbf{f}_r + w_{gg}H_{gg}\mathbf{f}_g + w_{bg}H_{bg}\mathbf{f}_b + \mathbf{n}_g, \\ \mathbf{u}_b &= w_{rb}H_{rb}\mathbf{f}_r + w_{gb}H_{gb}\mathbf{f}_g + w_{bb}H_{bb}\mathbf{f}_b + \mathbf{n}_b. \end{aligned} \tag{1.2}$$

Let

$$\mathbf{f} = \begin{bmatrix} \mathbf{f}_r \\ \mathbf{f}_g \\ \mathbf{f}_b \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} \mathbf{u}_r \\ \mathbf{u}_g \\ \mathbf{u}_b \end{bmatrix}, \quad \text{and} \quad \mathbf{n} = \begin{bmatrix} \mathbf{n}_r \\ \mathbf{n}_g \\ \mathbf{n}_b \end{bmatrix},$$

let

$$\mathbf{w} = [w_{rr} \quad w_{gr} \quad w_{br} \quad w_{rg} \quad w_{gg} \quad w_{bg} \quad w_{rb} \quad w_{gb} \quad w_{bb}]^T,$$

and define

$$H(\mathbf{w}) = \begin{bmatrix} w_{rr}H_{rr} & w_{gr}H_{gr} & w_{br}H_{br} \\ w_{rg}H_{rg} & w_{gg}H_{gg} & w_{bg}H_{bg} \\ w_{rb}H_{rb} & w_{gb}H_{gb} & w_{bb}H_{bb} \end{bmatrix}.$$

Equation (1.2) can be written as

$$\mathbf{u} = H(\mathbf{w})\mathbf{f} + \mathbf{n}. \tag{1.3}$$

Very often people use the observed image to recover the original image of the same domain. The observed image is not completely determined by the original image in the region of interest (ROI) because of the convolution. To determine the image, we make assumptions about the pixel values of the original image outside the ROI. The zero boundary condition assumes that the scene outside the ROI is totally dark. When the real scene outside the ROI does not satisfy this assumption well, ringing effects will appear at the boundaries of the restored image. Zero boundary conditions cause the blurring matrices H_{ij} to be block Toeplitz with Toeplitz blocks (BTTB). Neumann boundary conditions assume that the scene outside the ROI is a reflection of the scene inside. This reduces the ringing effects when the image is close to stationary at the boundaries. Using Neumann boundary conditions causes the blurring matrices H_{ij} to be block Toeplitz plus Hankel with Toeplitz plus Hankel blocks (BTHTHB), which in general gives better restoration results than those using zero boundary conditions [16].

Image restoration problems tend to be very ill conditioned. Directly solving (1.2) will yield a solution that is extremely sensitive to noise; therefore regularization methods are needed to stabilize the solution.

There is considerable work in the literature on least squares restoration of color images. Based on the assumption that the within-channel and cross-channel point spread functions are separable, Hunt and Kubler [12] derived a minimum mean squares error (MSE) multichannel restoration filter. Filters that utilize the cross-channel correlation without using the separability assumption can be found in [8] and [9]. More results can be found in [15, 17, 22].

The least squares solution assumes that exact information for the point spread functions, $\mathbf{h}_{ij}$, and the weights, w_{ij}, is available. For single channel image restoration problems, it has been demonstrated [14, 18, 19] that total least squares (TLS)

methods, in particular structured total least squares (STLS) [21] methods, yield better results than the ordinary least squares method when error exists in the blurring matrices. In this paper, we consider applying STLS for the multichannel image restoration problem. Our algorithm extends the basic algorithm presented in [19] and incorporates an efficient preconditioning technique.

The rest of this paper is organized as follows. In section 2, we give a description of the regularized STLS problem. In section 3, we derive our algorithm for solving the regularized STLS problem. In section 4, we propose an efficient way to precondition the linear systems encountered in section 3. Experimental results are shown in section 5, and section 6 concludes this paper.

2. Tikhonov's regularization of the STLS problem. In general the matrix H in (1.3) is ill conditioned. The method of regularization can be used to stabilize the solution. In the classical Tikhonov regularization [11], stability is attained by introducing a penalty term that measures the "irregularity" of the solution. One of the widely used regularization techniques for color image restoration is the one proposed by Galatsanos et al. in [9].

An important concept in the regularization technique of Galatsanos et al. is the "weighted three-dimensional (3D) Laplacian." Let

$$\boldsymbol{\beta} = [\beta_{rg} \quad \beta_{rb} \quad \beta_{gb}]^T = \left[\frac{\|\mathbf{f}_r\|_2}{\|\mathbf{f}_g\|_2} \quad \frac{\|\mathbf{f}_r\|_2}{\|\mathbf{f}_b\|_2} \quad \frac{\|\mathbf{f}_g\|_2}{\|\mathbf{f}_b\|_2} \right]^T, \tag{2.1}$$

and let L be the 2D Laplacian matrix of appropriate size. The "weighted 3D Laplacian" matrix is defined as follows:

$$R(\boldsymbol{\beta}) = \begin{bmatrix} 2I + L & -\beta_{rg} I & -\beta_{rb} I \\ -\frac{1}{\beta_{rg}} I & 2I + L & -\beta_{gb} I \\ -\frac{1}{\beta_{rb}} I & -\frac{1}{\beta_{gb}} I & 2I + L \end{bmatrix}.$$

Note that

$$R(\boldsymbol{\beta}) = D \hat{L} D^{-1},$$

where

$$D = \begin{bmatrix} \|\mathbf{f}_r\|_2 I & 0 & 0 \\ 0 & \|\mathbf{f}_g\|_2 I & 0 \\ 0 & 0 & \|\mathbf{f}_b\|_2 I \end{bmatrix}, \quad \hat{L} = \begin{bmatrix} 2I + L & -I & -I \\ -I & 2I + L & -I \\ -I & -I & 2I + L \end{bmatrix}.$$

Therefore, the operation $R(\boldsymbol{\beta})\mathbf{f}$ can be interpreted as applying the standard 3D Laplacian (with circulant boundary conditions imposed on the third dimension) on a scaled version of $\mathbf{f}$ with the results scaled back afterwards.

Let

$$\Omega = \begin{bmatrix} \alpha_r I & 0 & 0 \\ 0 & \alpha_g I & 0 \\ 0 & 0 & \alpha_b I \end{bmatrix},$$

where α_r, α_g, and α_b are positive constants, and I is the identity matrix of appropriate size.

Denote the elementwise square root of the matrix Ω by $\sqrt{\Omega}$ (note that Ω is a diagonal matrix with positive diagonal elements). The regularization technique of Galatsanos et al. takes the following form:

$$\min_{\mathbf{f}} \|\mathbf{u} - H\mathbf{f}\|_2^2 + \|\sqrt{\Omega}R(\beta)\mathbf{f}\|_2^2. \tag{2.2}$$

In this equation, Ω is used to control the degree of regularization. The constants α_r, α_g, and α_b are called the regularization parameters, and R is called the regularization operator. Notice that R depends on the vector $\boldsymbol{\beta}$, which further depends on the solution. In the literature, it is common practice to use an estimate of $\boldsymbol{\beta}$ rather than the exact $\boldsymbol{\beta}$. However, our algorithm will solve for $\boldsymbol{\beta}$ as part of the problem.

From a statistical point of view, if H is known exactly, the solution to (2.2) can be interpreted as the "maximum a posteriori estimator" of the original image. In this interpretation, the regularization term represents prior information on the original image, the regularization operator reflects the type of prior information, and the parameters α_r, α_g, and α_b indicate our level of confidence in the prior information.

As mentioned earlier, the least squares method assumes that exact information of the blurring matrix is available. In practice, the blurring matrix may also be contaminated by noise. The TLS method allows for the possibility of error in the elements of the blurring matrix H, so that the modified blurring matrix is given by $\tilde{H} = H + E$, where E is the error matrix to be determined. For image restoration problems, because H exhibits certain structures, it is more appropriate to apply STLS [21]. The STLS method requires that $\tilde{H}$ has the same structure as H.

In this paper, our focus is on cases where our knowledge of H_{ij}, for $i, j \in \{r, g, b\}$, is precise but that of $\mathbf{w}$ is inexact. However, the basic algorithms of this paper can be easily modified to handle the case in which our knowledge of H_{ij} is also inexact; see, for instance, [7, 19].

Let

$$\tilde{\mathbf{w}} = [\tilde{w}_{rr} \quad \tilde{w}_{gr} \quad \tilde{w}_{br} \quad \tilde{w}_{rg} \quad \tilde{w}_{gg} \quad \tilde{w}_{bg} \quad \tilde{w}_{rb} \quad \tilde{w}_{gb} \quad \tilde{w}_{bb} \quad]^T$$

be the modified $\mathbf{w}$. From our assumption, it follows that the modified blurring matrix is parameterized by $\tilde{\mathbf{w}}$. Therefore, the regularized STLS problem for color image restoration takes the following form:

$$\min_{\mathbf{f},\tilde{\mathbf{w}}} \|H(\tilde{\mathbf{w}})\mathbf{f} - \mathbf{u}\|_2^2 + \|\sqrt{\Omega}R(\boldsymbol{\beta})\mathbf{f}\|_2^2 + \alpha_h\|\mathbf{w} - \tilde{\mathbf{w}}\|_2^2. \tag{2.3}$$

The last term in this formula can be interpreted as a penalty term for the "irregularity" of the error in $\mathbf{w}$, and α_h can be interpreted as the regularization parameter used to control the degree of regularization on that error.

For simplicity, let

$$\begin{aligned}\mathbf{z} &= \tilde{\mathbf{w}} - \mathbf{w} \\ &= [z_{rr} \quad z_{gr} \quad z_{br} \quad z_{rg} \quad z_{gg} \quad z_{bg} \quad z_{rb} \quad z_{gb} \quad z_{bb}]^T;\end{aligned}$$

then (2.3) becomes

$$\min_{\mathbf{f},\mathbf{z}} \|H(\mathbf{w} + \mathbf{z})\mathbf{f} - \mathbf{u}\|_2^2 + \|\sqrt{\Omega}R(\boldsymbol{\beta})\mathbf{f}\|_2^2 + \alpha_h\|\mathbf{z}\|_2^2. \tag{2.4}$$

Experimental results showed that this formula tends to yield a darker version of the original image and a $\tilde{\mathbf{w}}$ whose elements are systematically larger than those of $\mathbf{w}$

(see Figure 5.8). This way it keeps the residual norm almost unchanged but reduces the "irregularity" of the solution.

To overcome this problem, we add the following constraints to (2.3):

$$\begin{aligned} z_{rr} + z_{gr} + z_{br} &= 0, \\ z_{rg} + z_{gg} + z_{bg} &= 0, \\ z_{rb} + z_{gb} + z_{bb} &= 0. \end{aligned} \tag{2.5}$$

That is, we require the sum of the modified weights for each observed channel to be the same as that of the original ones. Equation (2.5) is an expression of the conservation of intensity values in the color image.

This looks as if an unconstrained optimization problem has been turned into a constrained one. However, since

$$\begin{aligned} z_{br} &= -z_{rr} - z_{gr}, \\ z_{bg} &= -z_{rg} - z_{gg}, \\ z_{bb} &= -z_{rb} - z_{gb}, \end{aligned}$$

now the modified blurring matrix, $\tilde{H}$, is parameterized by only six variables, z_{rr}, z_{gr}, z_{rg}, z_{gg}, z_{rb}, and z_{gb}. Let

$$\hat{\mathbf{z}} = [z_{rr} \quad z_{gr} \quad z_{rg} \quad z_{gg} \quad z_{rb} \quad z_{gb}]^T;$$

then $\hat{\mathbf{z}}$ contains all the information in $\mathbf{z}$.

Note that if the constraints stated in (2.5) are satisfied, then $\mathbf{z} = \Gamma\hat{\mathbf{z}}$ with

$$\Gamma = \operatorname{diag}(\hat{\Gamma}, \hat{\Gamma}, \hat{\Gamma}), \quad \hat{\Gamma} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix}.$$

Therefore,

$$\|\mathbf{z}\|_2^2 = \hat{\mathbf{z}}^T \Gamma^T \Gamma \hat{\mathbf{z}}.$$

It is easy to check that the matrix S defined as follows is symmetric positive definite and satisfies $S^2 = \Gamma^T\Gamma$:

$$S = \operatorname{diag}(\hat{S}, \hat{S}, \hat{S}), \quad \text{where } \hat{S} = \begin{bmatrix} \frac{\sqrt{3}+1}{2} & \frac{\sqrt{3}-1}{2} \\ \frac{\sqrt{3}-1}{2} & \frac{\sqrt{3}+1}{2} \end{bmatrix}. \tag{2.6}$$

Therefore, $\|\mathbf{z}\|_2 = \|S\hat{\mathbf{z}}\|_2$.

It follows that our problem can be stated as

$$\min_{\mathbf{f},\hat{\mathbf{z}}} \|H(\mathbf{w} + \Gamma\hat{\mathbf{z}})\mathbf{f} - \mathbf{u}\|_2^2 + \|\sqrt{\Omega}R(\boldsymbol{\beta})\mathbf{f}\|_2^2 + \alpha_h\|S\hat{\mathbf{z}}\|_2^2. \tag{2.7}$$

An important issue in the regularization is the choice of the regularization parameters, α_r, α_g, α_b, and α_h. The optimal value for these parameters depends on the noise level of the observed image and estimated point spread functions. For the least squares solution, Ng and Bose proposed an efficient way to find the optimal $[\alpha_r \quad \alpha_g \quad \alpha_b]$ in [15]. Based on the statistical interpretation of the regularization term, $[\alpha_r \quad \alpha_g \quad \alpha_b]$ reflects our level of confidence in the prior information, which is

largely determined by the noise level on the right-hand side. Considering the TLS solution is also the least squares solution of a problem that has the same right-hand side but a different blurring matrix, we expect the optimal values for $[\alpha_r \quad \alpha_g \quad \alpha_b]$ in the TLS solution to be close to those in the least squares solution. Therefore, we simply use the optimal $[\alpha_r \quad \alpha_g \quad \alpha_b]$ computed by the least squares method for our TLS solution. In order to find the optimal α_h, we use an approach that is similar to the approach employed by Chan and Wong in [4]. That is, we start with a small value for α_h and increase it until a satisfactory solution is found by visually checking the recovered image. One can also use a more systematic approach such as the L-curve method [11], which tries to find a good balance between yielding small TLS residual and "regularity" of the recovered $\hat{\mathbf{z}}$ in (2.7).

3. Solving the regularized STLS problem. In [21], Rosen, Park, and Glick presented an iterative algorithm for structured total least norm in 1-, 2-, and infinity-norms. Pruessner and O'Leary extended the algorithm of Rosen, Park, and Glick to include regularization in all of these norms and demonstrated its use on single channel image deblurring in [19]. For the 2-norm case, the basic idea of the technique of Rosen, Park, and Glick is to approximate the cost function by a quadratic function and minimize that quadratic function at each iteration. Here we show that this technique can be extended to solve (2.7).

For convenience, let

$$\Psi(\mathbf{f}, \hat{\mathbf{z}}) = \|H(\mathbf{w} + \Gamma\hat{\mathbf{z}})\mathbf{f} - \mathbf{u}\|_2^2 + \|\sqrt{\Omega}R(\boldsymbol{\beta})\mathbf{f}\|_2^2 + \alpha_h\|S\hat{\mathbf{z}}\|_2^2;$$

then (2.7) becomes

$$\min_{\mathbf{f},\hat{\mathbf{z}}} \Psi(\mathbf{f}, \hat{\mathbf{z}}). \tag{3.1}$$

If the elements of $\boldsymbol{\beta}$ are constants, then this problem can be solved in a way similar to how Pruessner and O'Leary solved the single channel image deblurring problem in [19]. Taking into account that elements of $\boldsymbol{\beta}$ are not constants, we propose solving (2.7) as follows. At each iteration, we fix $\boldsymbol{\beta}$, approximate the cost function by a quadratic function, and minimize that function. At the end of each iteration, we use the new guess for $\mathbf{f}$ to update $\boldsymbol{\beta}$.

If $\boldsymbol{\beta}$ is fixed, then

$$\begin{aligned}\Psi(\mathbf{f} + \Delta\mathbf{f}, \hat{\mathbf{z}} + \Delta\hat{\mathbf{z}}) = {} & \|H(\mathbf{w} + \Gamma\hat{\mathbf{z}} + \Gamma\Delta\hat{\mathbf{z}})(\mathbf{f} + \Delta\mathbf{f}) - \mathbf{u}\|_2^2 \\ & + \|\sqrt{\Omega}R(\boldsymbol{\beta})(\mathbf{f} + \Delta\mathbf{f})\|_2^2 + \alpha_h\|S(\hat{\mathbf{z}} + \Delta\hat{\mathbf{z}})\|_2^2.\end{aligned} \tag{3.2}$$

Ignoring higher order terms, we get

$$\begin{aligned}\Psi(\mathbf{f} + \Delta\mathbf{f}, \hat{\mathbf{z}} + \Delta\hat{\mathbf{z}}) \approx {} & \|H(\mathbf{w} + \Gamma\hat{\mathbf{z}})(\mathbf{f} + \Delta\mathbf{f}) + H(\Gamma\Delta\hat{\mathbf{z}})\mathbf{f} - \mathbf{u}\|_2^2 \\ & + \|\sqrt{\Omega}R(\boldsymbol{\beta})(\mathbf{f} + \Delta\mathbf{f})\|_2^2 + \alpha_h\|S(\hat{\mathbf{z}} + \Delta\hat{\mathbf{z}})\|_2^2.\end{aligned} \tag{3.3}$$

Let

$$\begin{aligned}\hat{\Psi}(\Delta\mathbf{f}, \Delta\hat{\mathbf{z}}) = {} & \|H(\mathbf{w} + \Gamma\hat{\mathbf{z}})(\mathbf{f} + \Delta\mathbf{f}) + H(\Gamma\Delta\hat{\mathbf{z}})\mathbf{f} - \mathbf{u}\|_2^2 \\ & + \|\sqrt{\Omega}R(\boldsymbol{\beta})(\mathbf{f} + \Delta\mathbf{f})\|_2^2 + \alpha_h\|S(\hat{\mathbf{z}} + \Delta\hat{\mathbf{z}})\|_2^2.\end{aligned} \tag{3.4}$$

When $\boldsymbol{\beta}$ is fixed, minimizing $\hat{\Psi}$ is a Gauss–Newton step for minimizing Ψ.

Define the matrix Z as follows:

$$Z = \begin{bmatrix} Z_1 & 0 & 0 \\ 0 & Z_2 & 0 \\ 0 & 0 & Z_3 \end{bmatrix}, \tag{3.5}$$

where

$$\begin{aligned} Z_1 &= [H_{rr}\mathbf{f}_r - H_{br}\mathbf{f}_b \quad H_{gr}\mathbf{f}_g - H_{br}\mathbf{f}_b], \\ Z_2 &= [H_{rg}\mathbf{f}_r - H_{bg}\mathbf{f}_b \quad H_{gg}\mathbf{f}_g - H_{bg}\mathbf{f}_b], \\ Z_3 &= [H_{rb}\mathbf{f}_r - H_{bb}\mathbf{f}_b \quad H_{gb}\mathbf{f}_g - H_{bb}\mathbf{f}_b]. \end{aligned}$$

Then Z satisfies

$$H(\Gamma\Delta\hat{\mathbf{z}})\mathbf{f} = Z\Delta\hat{\mathbf{z}}. \tag{3.6}$$

Therefore, $\hat{\Psi}$ becomes

$$\begin{aligned} \hat{\Psi}(\Delta\mathbf{f}, \Delta\hat{\mathbf{z}}) &= \|H(\mathbf{w} + \Gamma\hat{\mathbf{z}})(\mathbf{f} + \Delta\mathbf{f}) + Z\Delta\hat{\mathbf{z}} - \mathbf{u}\|_2^2 \\ &\quad + \|\sqrt{\Omega}R(\boldsymbol{\beta})(\mathbf{f} + \Delta\mathbf{f})\|_2^2 + \alpha_h\|S(\hat{\mathbf{z}} + \Delta\hat{\mathbf{z}})\|_2^2. \end{aligned} \tag{3.7}$$

It follows that minimizing $\hat{\Psi}$ is equivalent to solving the following linear least squares system:

$$\min_{\Delta\mathbf{f}, \Delta\hat{\mathbf{z}}} \left\| \begin{bmatrix} H(\mathbf{w} + \Gamma\hat{\mathbf{z}}) & Z \\ \sqrt{\Omega}R(\boldsymbol{\beta}) & 0 \\ 0 & \sqrt{\alpha_h}S \end{bmatrix} \begin{bmatrix} \Delta\mathbf{f} \\ \Delta\hat{\mathbf{z}} \end{bmatrix} - \begin{bmatrix} \mathbf{u} - H(\mathbf{w} + \Gamma\hat{\mathbf{z}})\mathbf{f} \\ -\sqrt{\Omega}R(\boldsymbol{\beta})\mathbf{f} \\ -\sqrt{\alpha_h}S\hat{\mathbf{z}} \end{bmatrix} \right\|_2. \tag{3.8}$$

The STLS algorithm for color image restoration is summarized as follows.

ALGORITHM 3.1.
1. Set $\mathbf{f}$ to be the least squares solution, $\hat{\mathbf{z}} = 0$, construct Z according to (3.5)
2. Repeat
   (a) Solve (3.8) for $\Delta\mathbf{f}$, $\Delta\hat{\mathbf{z}}$.
   (b) Set $\mathbf{f} = \mathbf{f} + \Delta\mathbf{f}$, $\hat{\mathbf{z}} = \hat{\mathbf{z}} + \Delta\hat{\mathbf{z}}$,
   (c) construct Z according to (3.5),
   (d) update $\boldsymbol{\beta}$ according to (2.1).
   until $\|\Delta\mathbf{f}\|$, $\|\Delta\hat{\mathbf{z}}\| \leq \epsilon$.

We have no proof that this algorithm converges; however, experimental results indicate that this algorithm converges very quickly, provided that the regularization parameters are not too small. It is recommended that the least squares solution be used as the initial guess.

4. Solving the linear least squares systems efficiently. The main computational task of one iteration of Algorithm 3.1 is to solve the linear least squares system (3.8). The conjugte gradient least squares (CGLS) method [1] is often used to solve large sparse linear least squares systems in the form of

$$\min_{\mathbf{x}} \|\mathbf{b} - A\mathbf{x}\|_2.$$

The convergence of the CGLS method depends on the condition number of the coefficient matrix, A. When the condition number of A is large, the CGLS method may converge slowly; in that case, preconditioners [10, sect. 10.3] can be used to accelerate the convergence. The idea of preconditioning is to apply the regular CGLS to the transformed system

$$\min_{\tilde{\mathbf{x}}} \|\mathbf{b} - \tilde{A}\tilde{\mathbf{x}}\|_2,$$

where $\tilde{A} = AG^{-1}$ and $\tilde{\mathbf{x}} = G\mathbf{x}$. We need G to be such that $M = G^TG$ has a spectrum similar to A^TA and that the linear system $M\mathbf{x} = \mathbf{b}$ is easy to solve. We call G a preconditioner for A, or M a preconditioner for A^TA.

In [7], Fu and Barlow proposed a preconditioner for a linear system that is similar to (3.8). Unfortunately, the system here requires a different preconditioning strategy, because even though the submatrices of H are BTHTHB, the matrix H itself is not BTHTHB.

4.1. Symmetric blurring matrices. For (3.8), since

$$A = \begin{bmatrix} H(\mathbf{w}+\mathbf{z}) & Z \\ \sqrt{\Omega}R(\boldsymbol{\beta}) & 0 \\ 0 & \sqrt{\alpha_h}S \end{bmatrix}, \tag{4.1}$$

we have

$$A^TA = \begin{bmatrix} H(\mathbf{w}+\mathbf{z})^TH(\mathbf{w}+\mathbf{z}) + R(\boldsymbol{\beta})^T\Omega R(\boldsymbol{\beta}) & H(\mathbf{w}+\mathbf{z})^TZ \\ Z^TH(\mathbf{w}+\mathbf{z}) & Z^TZ + \alpha_h S^TS \end{bmatrix}.$$

Let $\hat{M}$ be the upper left block of A^TA, that is,

$$\hat{M} = H(\mathbf{w}+\mathbf{z})^TH(\mathbf{w}+\mathbf{z}) + R(\boldsymbol{\beta})^T\Omega R(\boldsymbol{\beta}).$$

We remark that the inverse of $\hat{M}$ can be easily computed if all the cross- and within-channel point spread functions are symmetric. By "symmetric" we mean that $h(x,y) = h(-x,y) = h(x,-y) = h(-x,-y)$. In fact, Ng and Bose efficiently computed the inverse of a matrix that is quite similar to $\hat{M}$ in [15].

Ng, Chan, and Tang [16] showed that if the point spread function h is symmetric, then the corresponding blurring matrix H with Neumann boundary conditions can be diagonalized by the 2D discrete cosine transform (DCT) [20] matrix C. Therefore,

$$H = \begin{bmatrix} C^T\Lambda_{rr}C & C^T\Lambda_{gr}C & C^T\Lambda_{br}C \\ C^T\Lambda_{rg}C & C^T\Lambda_{gg}C & C^T\Lambda_{bg}C \\ C^T\Lambda_{rb}C & C^T\Lambda_{gb}C & C^T\Lambda_{bb}C \end{bmatrix},$$

where Λ_{ij} (for $i,j \in \{r,g,b\}$) are real diagonal matrices.

The 2D Laplacian matrix with Neumann boundary conditions can also be diagonalized by the 2D DCT matrix. Let

$$L = C^T\Lambda_L C;$$

then

$$R = \begin{bmatrix} C^T(\Lambda_L + 2I)C & C^T(-\beta_{rg}I)C & C^T(-\beta_{rb}I)C \\ C^T(-\frac{1}{\beta_{rg}}I)C & C^T(\Lambda_L + 2I)C & C^T(-\beta_{gb}I)C \\ C^T(-\frac{1}{\beta_{br}}I)C & C^T(-\frac{1}{\beta_{gb}}I)C & C^T(\Lambda_L + 2I)C \end{bmatrix}.$$

It follows that the matrix $\hat{M}$ can be put into the following form:

$$\hat{M} = \begin{bmatrix} C^T\Lambda_{11}C & C^T\Lambda_{12}C & C^T\Lambda_{13}C \\ C^T\Lambda_{12}C & C^T\Lambda_{22}C & C^T\Lambda_{23}C \\ C^T\Lambda_{13}C & C^T\Lambda_{23}C & C^T\Lambda_{33}C \end{bmatrix} = (I \otimes C^T) \begin{bmatrix} \Lambda_{11} & \Lambda_{12} & \Lambda_{13} \\ \Lambda_{12} & \Lambda_{22} & \Lambda_{23} \\ \Lambda_{13} & \Lambda_{23} & \Lambda_{33} \end{bmatrix} (I \otimes C).$$

Here Λ_{ij} are real diagonal matrices. The matrix with diagonal blocks can be permuted into a block diagonal matrix so that $\hat{M}$ is expressed as

$$\hat{M} = \tilde{C}^T \operatorname{diag}(B_{1,1}, \ldots, B_{1,n}, B_{2,1}, \ldots, B_{2,n}, \ldots, B_{n,1}, \ldots, B_{n,n})\tilde{C}, \tag{4.2}$$

where

$$\tilde{C} = P(I \otimes C)$$

for the appropriate permutation matrix P. Here each $B_{k,\ell}$ $(1 \le k, \ell \le n)$ is a 3×3 symmetric matrix, and we consider an $n \times n$ image in each color channel. Next we show that $\hat{M}$ is symmetric and positive definite.

THEOREM 4.1. *Assume that* $(\mathbf{w}+\mathbf{z})_{ij} \ge 0$ *for* $i, j \in \{r, g, b\}$ *and at least one of* $(\mathbf{w}+\mathbf{z})_{ri}$, $(\mathbf{w}+\mathbf{z})_{gi}$, *and* $(\mathbf{w}+\mathbf{z})_{bi}$ *(*$i \in \{r, g, b\}$*) is nonzero. Then the matrix* $\hat{M}$ *is symmetric and positive definite.*

Proof. It is clear that $\hat{M}$ is symmetric. According to (4.2), it suffices to show that $B_{k,\ell}$ are positive definite for $1 \le k, \ell \le n$. In fact, the matrices $B_{k,\ell}$ are given by

$$B_{k,\ell} = Q_{k,\ell}(\mathbf{w}+\mathbf{z})^T Q_{k,\ell}(\mathbf{w}+\mathbf{z}) + (U + [\Lambda_L]_{k\ell} I)^T \Phi^2 (U + [\Lambda_L]_{k\ell} I),$$

where

$$Q_{k,\ell}(\mathbf{w}+\mathbf{z}) = \begin{bmatrix} (\mathbf{w}+\mathbf{z})_{rr}[\Lambda_{rr}]_{k\ell} & (\mathbf{w}+\mathbf{z})_{gr}[\Lambda_{gr}]_{k\ell} & (\mathbf{w}+\mathbf{z})_{br}[\Lambda_{br}]_{k\ell} \\ (\mathbf{w}+\mathbf{z})_{rg}[\Lambda_{rg}]_{k\ell} & (\mathbf{w}+\mathbf{z})_{gg}[\Lambda_{gg}]_{k\ell} & (\mathbf{w}+\mathbf{z})_{bg}[\Lambda_{bg}]_{k\ell} \\ (\mathbf{w}+\mathbf{z})_{rb}[\Lambda_{rb}]_{k\ell} & (\mathbf{w}+\mathbf{z})_{gb}[\Lambda_{gb}]_{k\ell} & (\mathbf{w}+\mathbf{z})_{bb}[\Lambda_{bb}]_{k\ell} \end{bmatrix},$$

$$U = \begin{bmatrix} 2 & -\beta_{rg} & -\beta_{rb} \\ -\frac{1}{\beta_{rg}} & 2 & -\beta_{gb} \\ -\frac{1}{\beta_{rb}} & -\frac{1}{\beta_{gb}} & 2 \end{bmatrix}, \quad \text{and} \quad \Phi = \begin{bmatrix} \sqrt{\alpha_r} & 0 & 0 \\ 0 & \sqrt{\alpha_g} & 0 \\ 0 & 0 & \sqrt{\alpha_b} \end{bmatrix}.$$

Let

$$F_{k,\ell} = \begin{bmatrix} Q_{k,\ell}(\mathbf{w}+\mathbf{z}) \\ \Phi(U + [\Lambda_L]_{k,\ell} I) \end{bmatrix}. \tag{4.3}$$

It is clear that $B_{k,\ell}$ is positive definite if and only if $F_{k,\ell}$ has full column rank. We note that the eigenvalues of the discrete Laplacian with the Neumann boundary condition are given by

$$[\Lambda_L]_{k\ell} = 4\sin^2\left(\frac{(k-1)\pi}{2n}\right) + 4\sin^2\left(\frac{(\ell-1)\pi}{2n}\right), \quad 1 \le k, \ell \le n.$$

Therefore, we have, $[\Lambda_L]_{1,1} = 0$. It implies that $F_{k,\ell}$ has full column rank for k and ℓ, except when $k = \ell = 1$.

Next we show that $F_{1,1}$ also has full column rank. It is easy to check that the eigenvalues of U are 0, 3, and 3 and that their corresponding eigenvectors are

$$[\beta_{rb}, \beta_{gb}, 1]^T, \quad [-\beta_{rb}, 0, 1]^T, \quad \text{and } [-\beta_{rg}, 1, 0]^T,$$

respectively. Therefore $\Phi(U + [\Lambda_L]_{k,\ell} I)\mathbf{x} = \mathbf{0}$ if and only if $k = \ell = 1$ and $\mathbf{x}$ is proportional to the nonnegative vector $[\beta_{rb}, \beta_{gb}, 1]^T$. Since discrete point spread functions are normalized such that their sums are equal to 1, it is easy to show that $[\Lambda_{ij}]_{1,1} = 1$. By the assumption, it is clear that $Q_{1,1}(\mathbf{w} + \mathbf{z})$ is a positive matrix, and thus $Q_{1,1}(\mathbf{w} + \mathbf{z})\mathbf{x} = 0$. Thus there is no vector $\mathbf{x}$ such that $F_{1,1}\mathbf{x} = \mathbf{0}$. It follows that $F_{1,1}$ must have full column rank. The result follows. □

It follows that

$$\hat{M}^{-1} = (I \otimes C^T)P^T \operatorname{diag}(B_{1,1}^{-1}, \ldots, B_{1,n}^{-1}, B_{2,1}^{-1}, \ldots, B_{2,n}^{-1}, \ldots, B_{n,1}^{-1}, \ldots, B_{n,n}^{-1})P(I \otimes C).$$

Therefore, the inverse of $\hat{M}$ can be computed by performing several 2D DCTs and computing the inverses of n^2 3×3 matrices. The preconditioner $\hat{M}$ can also be factored into $\hat{M} = \hat{G}^T\hat{G}$, where

$$G = \operatorname{diag}(G_{1,1}, \ldots, G_{1,n}, G_{2,1}, \ldots, G_{2,n}, \ldots, G_{n,1}, \ldots, G_{n,n})P(I \otimes C),$$

and $B_{k,\ell} = G_{k,\ell}^T G_{k,\ell}, 1 \le k, \ell \le n$, are the result of either the Cholesky factorization of $B_{k,\ell}$ or the QR factorization of $F_{k,\ell}$ in (4.3).

When all the cross- and within-channel point spread functions are symmetric, we propose to precondition A^TA by the matrix M defined as follows:

$$M = \begin{bmatrix} \hat{M} & 0 \\ 0 & Z^TZ + \alpha_h S^T S \end{bmatrix}.$$

Since the eigenvalues of S are 1 and $\sqrt{3}$ and $\alpha_h > 0$, the matrix $Z^TZ + \alpha_h S^TS$ is positive definite, and the preconditioner M is also positive definite. It is clear that A^TA, when preconditioned by M, is different from the identity matrix only at the last six rows and last six columns, which means it is the identity matrix plus a rank-12 change. Thus preconditioned CGLS is guaranteed to converge to the exact solution within 13 iterations.

4.2. Nonsymmetric blurring matrices. If not all the cross- and within-channel point spread functions are symmetric, we suggest constructing the preconditioner according to the symmetric part of the point spread functions. In [13], Kwan and Ng proved the following theorem and lemma.

THEOREM 4.2. *Let* $\mathbf{h}$ *be an arbitrary point spread function and* H_n *be the blurring matrix of* $\mathbf{h}$ *with the Neumann boundary conditions imposed. Then the optimal cosine transform preconditioner* $c(H_n)$ *of* H_n *is the blurring matrix corresponding to the symmetric point spread function* $\mathbf{h}_s$ *given by*

$$\mathbf{h}_s(i,j) = (\mathbf{h}(i,j) + \mathbf{h}(i,-j) + \mathbf{h}(-i,j) + \mathbf{h}(-i,-j))/4$$

with the Neumann boundary conditions imposed.

LEMMA 4.3. *Let* $\mathbf{h}$ *be an arbitrary point spread function and* H_n *be the blurring matrix of* $\mathbf{h}$ *with the Neumann boundary conditions imposed. Define the distance between the point spread functions* $\mathbf{h}$ *and* $\mathbf{h}_s$ *as*

$$\delta = \sum_{i,j} |\mathbf{h}(i,j) - \mathbf{h}_s(i,j)|.$$

Then

$$\|H_n - c(H_n)\|_1, \|H_n - c(H_n)\|_\infty \leq 4\delta,$$

where $c(H_n)$ is the optimal cosine transform preconditioner for H_n.

Here we consider the preconditioner for A^TA to be given by

$$M_c = \begin{bmatrix} c(H(\mathbf{w}+\mathbf{z}))^T c(H(\mathbf{w}+\mathbf{z})) + R(\boldsymbol{\beta})^T \Omega R(\boldsymbol{\beta}) & 0 \\ 0 & Z^TZ + \alpha_h S^T S \end{bmatrix}, \tag{4.4}$$

where

$$c(H(\mathbf{w}+\mathbf{z})) = \begin{bmatrix} (\mathbf{w}+\mathbf{z})_{rr} c(H_{rr}) & (\mathbf{w}+\mathbf{z})_{gr} c(H_{gr}) & (\mathbf{w}+\mathbf{z})_{br} c(H_{br}) \\ (\mathbf{w}+\mathbf{z})_{rg} c(H_{rg}) & (\mathbf{w}+\mathbf{z})_{gg} c(H_{gg}) & (\mathbf{w}+\mathbf{z})_{bg} c(H_{bg}) \\ (\mathbf{w}+\mathbf{z})_{rb} c(H_{rb}) & (\mathbf{w}+\mathbf{z})_{gb} c(H_{gb}) & (\mathbf{w}+\mathbf{z})_{bb} c(H_{bb}) \end{bmatrix}.$$

Using the above lemma, we have the following result. First, recall that

$$\|\mathbf{w}+\mathbf{z}\|_\infty = \max_{i,j\in\{r,g,b\}} (\mathbf{w}+\mathbf{z})_{ij}. \tag{4.5}$$

LEMMA 4.4. *Let A be the matrix given in* (4.1) *and M_c be the preconditioner given in* (4.4). *If*

$$\max_{i,j\in\{r,g,b\}} \|H_{ij} - c(H_{ij})\|_{1,\infty} \leq 4\delta,$$

then

$$\|H(\mathbf{w}+\mathbf{z})^T H(\mathbf{w}+\mathbf{z}) - c(H(\mathbf{w}+\mathbf{z}))^T c(H(\mathbf{w}+\mathbf{z}))\|_2 \leq 8\delta \|\mathbf{w}+\mathbf{z}\|_\infty.$$

Proof. We note that

$$\begin{aligned} \|H_{ij}^T H_{k\ell} - c(H_{ij})^T c(H_{k\ell})\| &\leq \|H_{ij}^T [H_{k\ell} - c(H_{k\ell})]\| + \|[H_{ij}^T - c(H_{ij})^T] c(H_{k\ell})\| \\ &\leq \|H_{ij}^T\| \|[H_{k\ell} - c(H_{k\ell})]\| + \|[H_{ij}^T - c(H_{ij})^T]\| \|c(H_{k\ell})\| \end{aligned}$$

with $i,j,k,\ell \in \{r,g,b\}$. Since the discrete point spread function is normalized such that its sum is equal to 1,

$$\|H_{ij}^T\|_1 = \|H_{ij}^T\|_\infty = \|c(H_{ij})\|_1 = \|c(H_{ij})\|_\infty = 1.$$

Using the above results and the statement (4.5), we can estimate

$$\|H(\mathbf{w}+\mathbf{z})^T H(\mathbf{w}+\mathbf{z}) - c(H(\mathbf{w}+\mathbf{z}))^T c(H(\mathbf{w}+\mathbf{z}))\|_{1,\infty} \leq 8\delta \|\mathbf{w}+\mathbf{z}\|_\infty.$$

Hence the result follows. ∎

According to Lemma 4.4, we see that if the point spread function is close to symmetric, then M_c will be a good approximation for A^TA. It follows that if we construct the matrix $\hat{M}$ according to the symmetric part of the point spread functions, then the matrix M_c defined as in (4.4) will be a good preconditioner for A^TA.

To show that the spectra of the preconditioned matrices are clustered around 1, we consider a sequence of blurring matrices and investigate the spectra of the sequence of the preconditioned blurring matrices. Therefore, we assume that the given blurring matrix H_{ij} comes from a fixed singly infinite matrix H_{ij}^∞. We associate with H_{ij}^∞ the function ϕ_{ij} where the entries of H_{ij} and H_{ij}^∞ are the Fourier coefficients of ϕ_{ij}. The

function ϕ_{ij} is called the *generating function* of the sequence of blurring matrices; see [3] for details. With the generating function ϕ_{ij} for the blurring matrix H_{ij}, it can be shown that the eigenvalues of the matrix $c(H_{ij})$ are given by

$$\phi_{ij}\left(\frac{(k-1)\pi}{2n}, \frac{(\ell-1)\pi}{2n}\right), \quad 1 \le k, \ell \le n, \quad i, j \in \{r, g, b\},$$

and the largest eigenvalue of $c(H_{ij})$ is given by $\phi_{ij}(0,0)$ when $\ell = k = 1$; see, for instance, [3, 13]. Since discrete point spread functions are normalized such that their sums are equal to 1, it is easy to show that $\phi_{ij}(0,0) = 1$. Next we show that the smallest eigenvalue of M_c is uniformly bounded below away from 0 by a positive constant independent of the size of M_c.

THEOREM 4.5. *Assume that* $(\mathbf{w}+\mathbf{z})_{ij} \ge 0$ *for* $i, j \in \{r, g, b\}$ *and at least one of* $(\mathbf{w}+\mathbf{z})_{ri}$, $(\mathbf{w}+\mathbf{z})_{gi}$, *and* $(\mathbf{w}+\mathbf{z})_{bi}$ $(i \in \{r, g, b\})$ *is nonzero. Then there exists a positive scalar* γ_{min}, *independent of the size of* M_c, *such that the smallest eigenvalue of* M_c *is uniformly bounded away from zero by* γ_{min}.

Proof. We define for simplicity that

$$\xi(x, y) = \lambda_{\min}(F_{x,y}^T F_{x,y}),$$

where

$$F_{x,y} = \begin{bmatrix} Q_{x,y}(\mathbf{w}+\mathbf{z}) \\ \Phi(U + \psi(x,y)I) \end{bmatrix},$$

$$Q_{x,y}(\mathbf{w}+\mathbf{z}) = \begin{bmatrix} (\mathbf{w}+\mathbf{z})_{rr}\phi_{rr}(x,y) & (\mathbf{w}+\mathbf{z})_{gr}\phi_{gr}(x,y) & (\mathbf{w}+\mathbf{z})_{br}\phi_{br}(x,y) \\ (\mathbf{w}+\mathbf{z})_{rg}\phi_{rg}(x,y) & (\mathbf{w}+\mathbf{z})_{gg}\phi_{gg}(x,y) & (\mathbf{w}+\mathbf{z})_{bg}\phi_{bg}(x,y) \\ (\mathbf{w}+\mathbf{z})_{rb}\phi_{rb}(x,y) & (\mathbf{w}+\mathbf{z})_{gb}\phi_{gb}(x,y) & (\mathbf{w}+\mathbf{z})_{bb}\phi_{bb}(x,y) \end{bmatrix},$$

$$\psi(x, y) = 4\sin^2 x + 4\sin^2 y.$$

With this notation, we have shown in Theorem 4.1 that $F_{x,y}$ has full column rank, and hence the smallest eigenvalue of the matrix $F_{x,y}^T F_{x,y}$ is positive. Therefore, $\xi(x, y) > 0$ for all $(x, y) \in [0, \pi/2]^2$. Since $\xi(x, y)$ is a continuous function on $[0, \pi/2]^2$, there exists a constant γ_{min}, independent of the size of the matrix, such that $\xi(x, y) \ge \gamma_{min} > 0$ for all $(x, y) \in [0, \pi/2]^2$. Hence the result follows. ∎

By using Lemma 4.4 and Theorem 4.5, we can prove the following theorem.

THEOREM 4.6. *Assume that* $(\mathbf{w}+\mathbf{z})_{ij} \ge 0$ *for* $i, j \in \{r, g, b\}$, *at least one of* $(\mathbf{w}+\mathbf{z})_{ri}$, $(\mathbf{w}+\mathbf{z})_{gi}$, *and* $(\mathbf{w}+\mathbf{z})_{bi}$ *(*$i \in \{r, g, b\}$*) is nonzero, and*

$$\max_{i,j \in \{r,g,b\}} \|H_{ij} - c(H_{ij})\|_{1,\infty} \le 4\delta.$$

Then

$$M_c^{-1} A^T A = I + E_1 + E_2,$$

where E_1 *is a matrix of rank* 12, *and* $\|E_2\|_2 \le 8\delta \|\mathbf{w}+\mathbf{z}\|_\infty / \gamma_{min}$, *where* γ_{min} *is the positive constant given in Theorem* 4.5.

Proof. We note that

$$\begin{aligned} & M_c^{-1} A^T A \\ &= I + M_c^{-1}(A^T A - M_c) \\ &= I + M_c^{-1} \begin{bmatrix} H(\mathbf{w}+\mathbf{z})^T H(\mathbf{w}+\mathbf{z}) - c(H(\mathbf{w}+\mathbf{z}))^T c(H(\mathbf{w}+\mathbf{z})) & H(\mathbf{w}+\mathbf{z})^T Z \\ Z^T H(\mathbf{w}+\mathbf{z}) & 0 \end{bmatrix}. \end{aligned}$$

The result follows by noting that the matrix

$$M_c^{-1}\begin{bmatrix} 0 & H(\mathbf{w}+\mathbf{z})^T Z \\ Z^T H(\mathbf{w}+\mathbf{z}) & 0 \end{bmatrix}$$

is a matrix of rank 12 and

$$\begin{aligned}
&\left\| M_c^{-1}\begin{bmatrix} (H(\mathbf{w}+\mathbf{z})^T H(\mathbf{w}+\mathbf{z}) - c(H(\mathbf{w}+\mathbf{z}))^T c(H(\mathbf{w}+\mathbf{z}))) & 0 \\ 0 & 0 \end{bmatrix}\right\|_2 \\
&\le \|M_c^{-1}\|_2 \left\|\begin{bmatrix} H(\mathbf{w}+\mathbf{z})^T H(\mathbf{w}+\mathbf{z}) - c(H(\mathbf{w}+\mathbf{z}))^T c(H(\mathbf{w}+\mathbf{z})) & 0 \\ 0 & 0 \end{bmatrix}\right\|_2 \\
&\le 8\delta \|M_c^{-1}\|_2 \|\mathbf{w}+\mathbf{z}\|_\infty \le 8\delta \|\mathbf{w}+\mathbf{z}\|_\infty / \gamma_{min}. \qquad \square
\end{aligned}$$

According to Theorem 4.6, if δ is sufficiently small (i.e., if the point spread function is quite symmetric), then the spectra of the preconditioned matrices $M_c^{-1}A^TA$ are clustered around 1, and therefore the preconditioned conjugate gradient method will converge very quickly. In the next section, this result is confirmed by our experimental results.

5. Experimental results. The image shown in Figure 5.1 is used for computer simulation.

5.1. The first test. The discretized point spread functions are as follows:

$$\mathbf{h}_{rr} = \mathbf{h}_{gr} = \mathbf{h}_{br} = \begin{bmatrix} \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \end{bmatrix},$$

$$\mathbf{h}_{rg} = \mathbf{h}_{gg} = \mathbf{h}_{bg} = \begin{bmatrix} 0 & 0.125 & 0 \\ 0.125 & 0.5 & 0.125 \\ 0 & 0.125 & 0 \end{bmatrix},$$

$$\mathbf{h}_{rb} = \mathbf{h}_{gb} = \mathbf{h}_{bb} = \begin{bmatrix} 0.0625 & 0.125 & 0.0625 \\ 0.125 & 0.25 & 0.125 \\ 0.0625 & 0.125 & 0.0625 \end{bmatrix}.$$

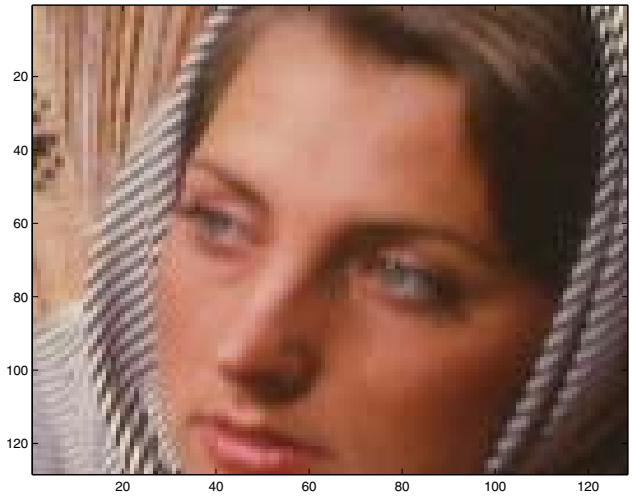

FIG. 5.1. *The original image.*

FIG. 5.2. *The observed image.*

The weights used to generate the observed channels are as follows:

$$\mathbf{w} = [0.5 \quad 0.3 \quad 0.2 \quad 0.25 \quad 0.5 \quad 0.25 \quad 0.2 \quad 0.3 \quad 0.5]^T.$$

A noise returned by the MATLAB command "0.01*(rand-0.5)" [6] is added to the observed channels. Figure 5.2 shows the observed image.

The estimated weights are set to

$$\mathbf{w} = [0.45 \quad 0.33 \quad 0.22 \quad 0.275 \quad 0.45 \quad 0.275 \quad 0.22 \quad 0.33 \quad 0.55]^T.$$

The relative error on the estimated weights is 10%.

We use the peak signal to noise ratio (PSNR) to measure the quality of the recovered images. The PSNR of a discretized color image is defined as

$$PSNR(\mathbf{f}) = 20log_{10}\sqrt{\frac{1}{\|\mathbf{f} - \mathbf{f}_{true}\|_2^2/(3mn)}},$$

where $\mathbf{f}_{true}$ is the true image, and m and n are the size of the image.

Two least squares solutions are computed. One is obtained by using the $\boldsymbol{\beta}$ calculated from the observed image; the other is obtained by using the $\boldsymbol{\beta}$ calculated from the original image. Figures 5.3 and 5.4 show these two images. For both solutions, $[\alpha_r \quad \alpha_g \quad \alpha_b] = [0.0001 \quad 0.0001 \quad 0.0001]$ is used.

For the TLS solution, $\alpha_h = 9$ yields satisfactory results. The recovered weights are as follows:

$$\mathbf{w} = [0.4943 \quad 0.3137 \quad 0.1920 \quad 0.2555 \quad 0.4688 \quad 0.2757 \quad 0.1902 \quad 0.3292 \quad 0.4807]^T.$$

The relative error of the recovered weights is 5.4%, which is much smaller than that of the original estimates. Figure 5.5 shows the image recovered by the STLS method. The PSNRs of the recovered images indicate that the TLS solution is much better.

Figure 5.6 shows the error norm (measured as $\|[\Delta\mathbf{f}^T \Delta\hat{\mathbf{z}}^T]\|$) at each iteration of Algorithm 3.1. As can be seen, Algorithm 3.1 converged to an error level of 10^{-2} in

FIG. 5.3. *The image recovered by the least squares method. The observed image is used to calculate* $\boldsymbol{\beta}$. $PSNR = 27.79db$.

FIG. 5.4. *The image recovered by the least squares method. The original image is used to calculate* $\boldsymbol{\beta}$. $PSNR = 28.01db$.

12 iterations. More iterations will further reduce the error norm but will have very little effect on the recovered image.

Two variants of Algorithm 3.1 are also used to compute TLS solutions. One uses a fixed $\boldsymbol{\beta}$, which is calculated from the observed image. The image recovered by this variation is shown in Figure 5.7. The difference between the PSNRs of this image and the one shown in Figure 5.5 is 0.82db, which is not a small value. The difference can be larger if the $\boldsymbol{\beta}$ values based on the observed image are quite different from those based on the recovered image, which will occur if there is significant cross-channel blur and the original image has significantly different intensity levels in the three channels (for example, when the original image is dominated by the red channel). The other variant does not apply the constraints stated in (2.5). That is, the solution

FIG. 5.5. *The image recovered by Algorithm* 3.1. $PSNR = 34.22db$.

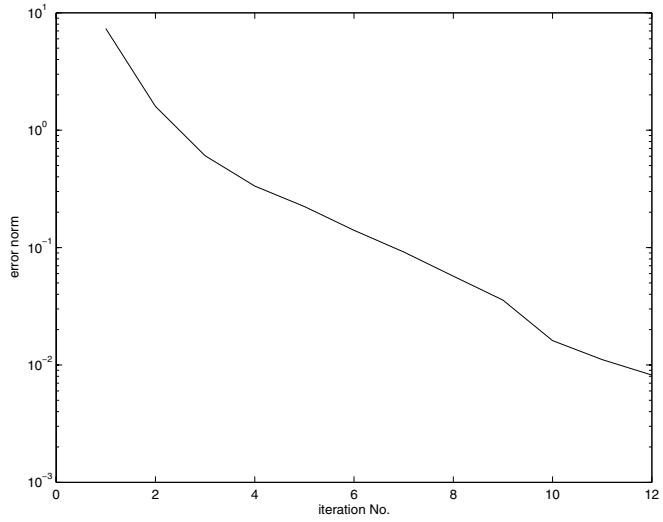

FIG. 5.6. *Error norm at each iteration of Algorithm* 3.1.

of (2.4) is computed. The image recovered by this variant, which has a much lower PSNR, is shown in Figure 5.8. Table 5.1 shows the relative error of each channel when different restoration methods are used, which also indicates that the solution returned by Algorithm 3.1 is the best.

5.2. The second test. We remark that in practice, it is quite possible that no information for the weights is available at all. In that situation, one can use an estimate of 1/3 for all the cross- and within-channel weights. Using the same observed image in Figure 5.2, the recovered weights are as follows:

$$\mathbf{w} = [0.4809 \quad 0.3054 \quad 0.2137 \quad 0.3018 \quad 0.3488 \quad 0.3494 \quad 0.2125 \quad 0.3465 \quad 0.4411]^T.$$

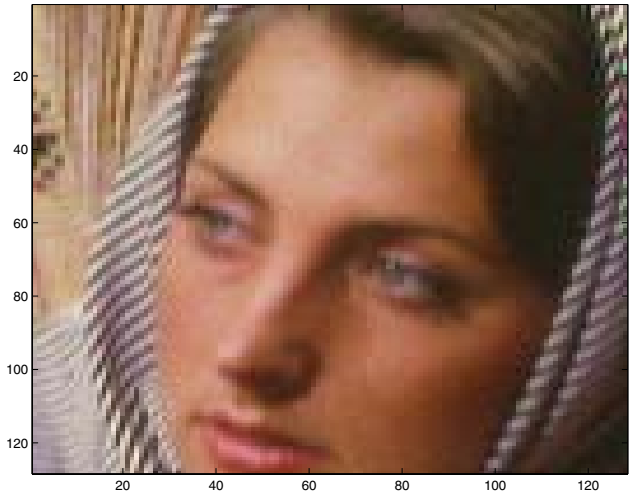

FIG. 5.7. *The image recovered by a variant of Algorithm* 3.1 *(a fixed* $\boldsymbol{\beta}$ *is used). PSNR* $= 33.40db$.

FIG. 5.8. *The image recovered by a variant of Algorithm* 3.1 *(without the constraints stated in* (2.5)). *PSNR* $= 29.47db$.

TABLE 5.1

Relative error of each channel when different restoration methods are used. TLS: total least squares restoration; Algorithm 3.1 *is used. TLS-*1*: total least squares restoration; a fixed* $\boldsymbol{\beta}$ *is used. TLS-*2*: total least squares restoration; the solution of* (2.4) *is computed. LS-*1*: least squares restoration; the observed image is used to calculate* $\boldsymbol{\beta}$*. LS-*2*: least squares restoration; the original image is used to calculate* $\boldsymbol{\beta}$*.*

	TLS	TLS-1	TLS-2	LS-1	LS-2
Red channel	0.0219	0.0224	0.0222	0.0857	0.0826
Green channel	0.0166	0.0251	0.0918	0.0522	0.0492
Blue channel	0.0238	0.0239	0.0674	0.0691	0.0684

The relative error on the estimated weights is 19.19%. The PSNRs of the images recovered by the least squares method are 24.55db and 21.00db when the β are calculated from the original image and the observed image, respectively. The PSNR of the

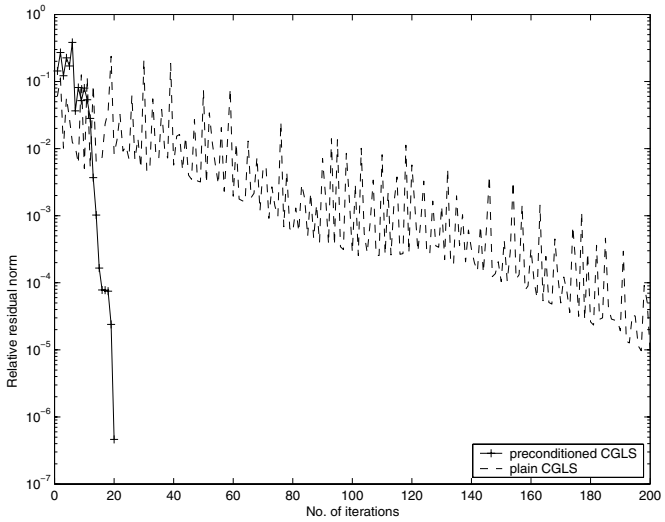

FIG. 5.9. *Relative residual norm at each CGLS iteration.*

image recovered by our proposed method is 29.74db, where the β is calculated from the observed image. Without a doubt, a less accurate estimate of the weights will worsen both the least squares and TLS solutions, but it also makes the TLS solution more likely to recover a better estimate of the weights and therefore outperform the least squares solution.

5.3. The third test. Experiments also showed that the preconditioning technique proposed in section 4 is very effective if the point spread functions are symmetric or close to symmetric. As predicted by theoretical analysis, preconditioned CGLS converges to the exact solution in 13 iterations when the point spread functions are symmetric. When the point spread functions are not symmetric, there is no guarantee that preconditioned CGLS will converge to the exact solution in 13 iterations, but the convergence rate is considerably accelerated if the functions are close to symmetric. This is shown in Figure 5.9. In this example, a noise returned by the MATLAB command "0.01*rand" is added to the elements of the original symmetric point spread functions. The point spread functions are then normalized such that they integrate to 1. As can be seen, preconditioned CGLS converges much faster.

Next we test the performance of the preconditioning techniques for nonsymmetric point spread functions. In the test, we set $H_{rr} = H_{rg} = H_{rb}$, $H_{gr} = H_{gg} = H_{gb}$, and $H_{br} = H_{bg} = H_{bb}$. Their corresponding point spread functions are examples of blurring that occurs in wavefront coding where a cubic phase filter is used to improve depth of field restoration in light efficient wide aperture optical systems [5]. An example of the point spread function for H_{rr} can be found in Figure 5.10. We find in Figure 5.11 that the performance of the preconditioner is poorer than in the previous case, but the convergence rate is considerably accelerated and is much better than the convergence rate of the method without using the preconditioner.

6. Conclusion. In conclusion, we have proposed a regularized STLS algorithm for color image restoration. In our problem setup, we assume that the blurring matrix, as well as the right-hand side, is contaminated by noise. The regularized STLS problem is stated as an unconstrained optimization problem. Our algorithm solves

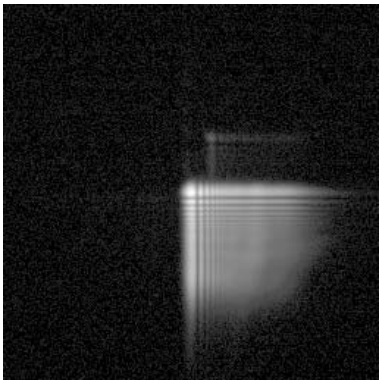

FIG. 5.10. *Image of the nonsymmetric point spread function.*

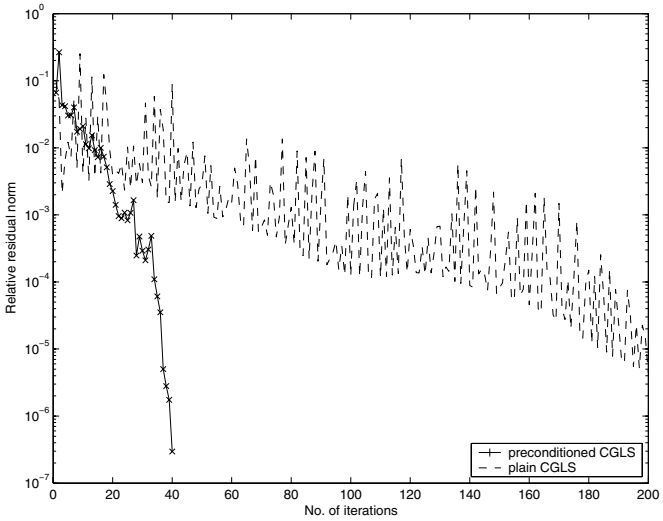

FIG. 5.11. *Relative residual norm at each CGLS iteration.*

the problem iteratively. At each iteration, the objective function is simplified by fixing certain parameters and ignoring high order terms. The simplified version of the objective function can be minimized by solving a linear least squares system. This system is solved by the preconditioned CGLS method. The main contribution of this paper includes solving for the weights of the weighted 3D Laplacian as part of the problem and proposing an efficient preconditioner for the linear systems encountered in the STLS algorithm. Experimental results showed that the STLS method recovers the error on the blurring matrix well, and it yields an image that is superior to the least squares solution.

REFERENCES

[1] Å. BJÖRCK AND T. ELFVING, *Accelerated projection methods for computing pseudo-inverse solutions of linear equations*, BIT, 19 (1979), pp. 145–163.

[2] K. BOO AND N. K. BOSE, *Multispectral image restoration with multisensors*, IEEE Trans. Geosci. Remote Sensing, 35 (1997), pp. 1160–1170.

[3] R. H. Chan and M. K. Ng, *Conjugate gradient methods for Toeplitz systems*, SIAM Rev., 38 (1996), pp. 427–482.
[4] T. F. Chan and C. Wong, *Total variation blind deconvolution*, IEEE Trans. Image Process., 7 (1998), pp. 370–375.
[5] E. R. Dowski and W. T. Cathey, *Extended depth of field through wavefront coding*, Applied Optics, 34 (1995), pp. 1859–1866.
[6] D. M. Etter, D. C. Kuncicky, and D. W. Hull, *Introduction to Matlab* 6, Prentice–Hall, Upper Saddle River, NJ, 2002.
[7] H. Fu and J. Barlow, *A regularized structured total least squares algorithm for high resolution image reconstruction*, Linear Algebra Appl., 391 (2004), pp. 75–98.
[8] N. P. Galatsanos and R. T. Chin, *Digital restoration of multichannel images*, IEEE Trans. Acoust. Speech Signal Process., 37 (1989), pp. 415–421.
[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 (1991), pp. 2222–2236.
[10] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, MD, 1996.
[11] P. C. Hansen and D. P. O'Leary, *The use of the L-curve in the regularization of discrete ill-posed problems*, SIAM J. Sci. Comput., 14 (1993), pp. 1487–1503.
[12] B. R. Hunt and O. Kubler, *Karhunen-Loeve multispectral image restoration, Part* I: *Theory*, IEEE Trans. Acoust. Speech, Signal Process., 32 (1984), pp. 592–600.
[13] W. Kwan and M. K. Ng, *Iterative Regularization for Ill-posed Imaging Problems using Neumann Boundary Conditions*, in Structured Matrices: Recent Developments in Theory and Computation, Nova Science Publishers, Hauppauge, NY, 2001, pp. 177–190.
[14] V. Z. Mesarović, N. P. Galatsanos, and A. K. Katsaggelos, *Regularized constrained total least squares image restoration*, IEEE Trans. Image Process., 4 (1995), pp. 1096–1108.
[15] M. K. Ng and N. K. Bose, *Fast color image restoration with multisensors*, International Journal of Imaging Systems and Technology, 12 (2002), pp. 189–197.
[16] M. K. Ng, R. H. Chan, and W.-C. Tang, *A fast algorithm for deblurring models with Neumann boundary conditions*, SIAM J. Sci. Comput., 21 (1999), pp. 851–866.
[17] M. K. Ng and W. C. Kwan, *Comments on least squares restoration of multichannel images*, IEEE Trans. Signal Process., 49 (2001), p. 2885.
[18] M. K. Ng, R. J. Plemmons, and F. Pimentel, *A new approach to constrained total least squares image restoration*, Linear Algebra Appl., 316 (2000), pp. 237–258.
[19] A. Pruessner and D. P. O'Leary, *Blind deconvolution using a regularized structured total least norm algorithm*, SIAM J. Matrix Anal. Appl., 24 (2003), pp. 1018–1037.
[20] K. Rao and P. Yip, *Discrete Cosine Transform: Algorithms, Advantages, Applications*, Academic Press, Boston, MA, 1990.
[21] J. B. Rosen, H. Park, and J. Glick, *Total least norm formulation and solution for structured problems*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 110–126.
[22] A. Tekalp and G. Pavlovic, *Multichannel image modeling and kalman filtering for multispectral image restoration*, Signal Process., 19 (1990), pp. 221–232.