

## BLOCK-TRIANGULAR PRECONDITIONERS FOR SYSTEMS ARISING FROM EDGE-PRESERVING IMAGE RESTORATION\*

Zhong-Zhi Bai

*LSEC, ICMSEC, Academy of Mathematics and Systems Science,  
Chinese Academy of Sciences, Beijing 100190, China*

*Email: bzz@lsec.cc.ac.cn*

Yu-Mei Huang

*School of Mathematics and Statistics, Lanzhou University, Lanzhou 730000, China*

*Email: ymhuang08@gmail.com*

Michael K. Ng

*Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong*

*Email: mng@math.hkbu.edu.hk*

### Abstract

Signal and image restoration problems are often solved by minimizing a cost function consisting of an  $\ell_2$  data-fidelity term and a regularization term. We consider a class of convex and edge-preserving regularization functions. In specific, half-quadratic regularization as a fixed-point iteration method is usually employed to solve this problem. The main aim of this paper is to solve the above-described signal and image restoration problems with the half-quadratic regularization technique by making use of the Newton method. At each iteration of the Newton method, the Newton equation is a structured system of linear equations of a symmetric positive definite coefficient matrix, and may be efficiently solved by the preconditioned conjugate gradient method accelerated with the modified block SSOR preconditioner. Our experimental results show that the modified block-SSOR preconditioned conjugate gradient method is feasible and effective for further improving the numerical performance of the half-quadratic regularization approach.

*Mathematics subject classification:* 65C20, 65F10.

*Key words:* Block system of equations, Matrix preconditioner, Edge-preserving, Image restoration, Half-quadratic regularization.

### 1. Introduction

We consider the signal and image restoration problems for which a vector  $\hat{\mathbf{x}} \in \mathbb{R}^p$  (an image or a signal) is estimated based upon a degraded data vector  $\mathbf{b} \in \mathbb{R}^q$  by minimizing a cost function  $J : \mathbb{R}^p \rightarrow \mathbb{R}$ . The function  $J$  is a combination of a data-fidelity term with a regularization term  $\Phi$  that is weighted by a parameter  $\beta > 0$ . More precisely, the problems are of the form

$$\hat{\mathbf{x}} = \min_{\mathbf{x} \in \mathbb{R}^p} J(\mathbf{x}),$$
$$J(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|_2^2 + \beta\Phi(\mathbf{x}),$$

where  $\mathbf{A} \in \mathbb{R}^{q \times p}$  is the known blurring matrix. The data-fidelity term given above assumes that  $\mathbf{b}$  and  $\mathbf{x}$  satisfy an approximate linear relation  $\mathbf{Ax} \approx \mathbf{b}$ , but that  $\mathbf{b}$  is contaminated by noise.

---

\* Received December 18, 2007 / Revised version received November 6, 2009 / Accepted December 7, 2009 /  
Published online August 9, 2010 /

Such data-fidelity terms are popular in numerous inverse problems such as seismic imaging, non-destructive evaluation, and x-ray tomography, see for instance [11]. Here, we consider regularization terms  $\Phi$  of the form

$$\Phi(\mathbf{x}) = \sum_{i=1}^r \phi(\mathbf{g}_i^T \mathbf{x}), \tag{1.1}$$

where  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  is a continuously differentiable function, and  $\mathbf{g}_i : \mathbb{R}^p \rightarrow \mathbb{R}$ , for  $i = 1, \dots, r$ , are linear operators. Typically,  $\{\mathbf{g}_i^T \mathbf{x}\}$  are the first- or the second-order differences between the neighboring samples in  $\mathbf{x}$ . Let  $\mathbf{G}$  denote the  $r \times p$  matrix whose  $i$ th row is  $\mathbf{g}_i^T$ , for  $i = 1, \dots, r$ , and assume that

$$\mathbf{A} \neq 0, \quad \mathbf{G} \neq 0, \quad \phi \neq 0 \quad \text{and} \quad \ker(\mathbf{A}^T \mathbf{A}) \cap \ker(\mathbf{G}^T \mathbf{G}) = \{0\}, \tag{1.2}$$

where  $\ker(\cdot)$  denotes the kernel of the corresponding matrix. Clearly, this assumption guarantees that  $\alpha_1 \mathbf{A}^T \mathbf{A} + \alpha_2 \mathbf{G}^T \mathbf{G}$  is a symmetric positive definite matrix provided both  $\alpha_1$  and  $\alpha_2$  are positive constants.

In this paper, we focus on convex, *edge-preserving* potential functions  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  defined in (1.1), because they can yield image and signal estimates of high quality, involving edges and homogeneous regions. Typical examples of such functions are:

$$\phi_1(t) = |t|/\alpha - \log(1 + |t|/\alpha), \tag{1.3}$$

$$\phi_2(t) = \sqrt{\alpha + t^2}, \tag{1.4}$$

$$\phi_3(t) = \log(\cosh(\alpha t))/\alpha, \tag{1.5}$$

$$\phi_4(t) = \begin{cases} t^2/(2\alpha), & \text{if } |t| \leq \alpha, \\ |t| - \alpha/2, & \text{if } |t| > \alpha, \end{cases} \tag{1.6}$$

where  $\alpha > 0$  is a prescribed parameter. See [15] and the references therein. We will consider the case that  $\phi$  is convex, even, and is  $\mathbb{C}^2$ , and that

$$\mathbf{A}^T \mathbf{A} \text{ is invertible} \quad \text{and/or} \quad \phi''(t) > 0, \quad \forall t \in \mathbb{R}. \tag{1.7}$$

The assumptions in (1.7) and (1.2) guarantee that for every  $\mathbf{x} \in \mathbb{R}^p$ , the function  $J$  has a unique minimum and that this minimum is strict. As now  $\mathbf{A}^T \mathbf{A}$  is symmetric positive definite, we know that  $\ker(\mathbf{A}^T \mathbf{A}) \cap \ker(\mathbf{G}^T \mathbf{G}) = \{0\}$  holds true. Moreover, we easily see that  $\mathbf{A}^T \mathbf{A}$  is symmetric positive definite if and only if  $q \geq p$  and  $\mathbf{A}$  is of full column rank.

However, the minimizers  $\hat{\mathbf{x}}$  of the cost-functions  $J$  involving edge-preserving regularization are nonlinear with respect to  $\mathbf{x}$  and their computations are quite costly. To simplify the computation, a half-quadratic reformulation of  $J$  was pioneered in [12] and [13]. One of the basic idea is to construct an augmented cost function  $\tilde{J} : \mathbb{R}^p \times \mathbb{R}^r \rightarrow \mathbb{R}$  that involves an auxiliary variable  $\mathbf{z} \in \mathbb{R}^r$  in the following form:

$$\tilde{J}(\mathbf{x}, \mathbf{z}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \beta \sum_{i=1}^r \left( \frac{1}{2} (\mathbf{g}_i^T \mathbf{x} - z_i)^2 + \psi(z_i) \right), \tag{1.8}$$

where

$$\psi(t) = \min_{s \in \mathbb{R}} \left\{ -\frac{1}{2} (t - s)^2 + \phi(s) \right\}, \quad \forall t \in \mathbb{R}, \tag{1.9}$$

and  $\psi : \mathbb{R} \rightarrow \mathbb{R}$  is a prescribed dual potential function. We remark that  $\phi(\cdot)$  is a potential function used as the regularization term in (1.1). Usually, such a dual potential function  $\psi$  can be determined by using the theory of convex conjugacy. The condition (1.9) ensures that

$$J(\mathbf{x}) = \min_{\mathbf{z} \in \mathbb{R}^r} \tilde{J}(\mathbf{x}, \mathbf{z}), \quad \forall \mathbf{x} \in \mathbb{R}^p.$$

The regularization term involved in  $\tilde{J}$  is half-quadratic, hence the name of the method is termed. The minimizer  $(\hat{\mathbf{x}}, \hat{\mathbf{z}})$  of  $\tilde{J}$  can be calculated by using alternating minimization. That is to say, if the solution at iteration  $(k - 1)$  reads  $(\mathbf{x}^{(k-1)}, \mathbf{z}^{(k-1)})$ , then at iteration  $k$  one calculates

$$\begin{cases} \mathbf{z}^{(k)} & \text{such that} & \tilde{J}(\mathbf{x}^{(k-1)}, \mathbf{z}^{(k)}) \leq \tilde{J}(\mathbf{x}^{(k-1)}, \mathbf{z}), & \forall \mathbf{z} \in \mathbb{R}^r, \\ \mathbf{x}^{(k)} & \text{such that} & \tilde{J}(\mathbf{x}^{(k)}, \mathbf{z}^{(k)}) \leq \tilde{J}(\mathbf{x}, \mathbf{z}^{(k)}), & \forall \mathbf{x} \in \mathbb{R}^p. \end{cases}$$

The major cost at each iteration of this approach is in computing  $\mathbf{x}^{(k)}$ , which requires to solve a linear system of the form:

$$(2\mathbf{A}^T \mathbf{A} + \beta \mathbf{G}^T \mathbf{G}) \mathbf{x}^{(k)} = 2\mathbf{A}^T \mathbf{b} + \beta \mathbf{G}^T \mathbf{z}^{(k)}. \tag{1.10}$$

We consider spatial-invariant blurring in which case  $\mathbf{A}$  is a Toeplitz-like matrix [16]. In the regularization term,  $\mathbf{G}$  is the discretization matrix of the first-order difference operator. Thus we can solve (1.10) quite efficiently by utilizing the special properties of the involved matrices. Numerical results have shown that the minimization using half-quadratic regularization can speed up computation. Unfortunately, the main drawback of this method is that its convergence rate is only linear [15].

In order to speed up the convergence of the method, we may adopt the Newton-type method to solve (1.8). To this end, we revisit the Hessian of  $\tilde{J}(\mathbf{x}, \mathbf{z})$ , which is given by

$$\mathbf{H}(\mathbf{x}, \mathbf{z}) = \begin{bmatrix} 2\mathbf{A}^T \mathbf{A} + \beta \mathbf{G}^T \mathbf{G} & -\beta \mathbf{G}^T \\ -\beta \mathbf{G} & \beta \mathbf{I} + \beta \text{diag}(\psi''(z_i)) \end{bmatrix} := \begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22}(\mathbf{z}) \end{bmatrix}. \tag{1.11}$$

Here,  $\mathbf{I}$  represents the identity matrix,  $\text{diag}(\psi''(z_i))$  is a diagonal matrix whose diagonal entries are given by  $\{\psi''(z_i)\}$ , and  $\{z_i\}$  are the entries of the vector  $\mathbf{z}$ .

**Theorem 1.1.** *Under the assumptions in (1.7) and (1.2), the Hessian matrix  $\mathbf{H}(\mathbf{x}, \mathbf{z})$  is symmetric positive definite for all  $\mathbf{x}$  and  $\mathbf{z}$ .*

*Proof.* Obviously,  $\mathbf{H}(\mathbf{x}, \mathbf{z})$  is symmetric and its diagonal blocks  $2\mathbf{A}^T \mathbf{A} + \beta \mathbf{G}^T \mathbf{G}$  and  $\beta \mathbf{I} + \beta \text{diag}(\psi''(z_i))$  are symmetric positive definite. Because the Schur complement of  $\mathbf{H}(\mathbf{x}, \mathbf{z})$  with respect to its (2,2) block is given by

$$2\mathbf{A}^T \mathbf{A} + \beta \mathbf{G}^T \mathbf{G} - \beta^2 \mathbf{G}^T (\beta \mathbf{I} + \beta \text{diag}(\psi''(z_i)))^{-1} \mathbf{G},$$

which is symmetric positive definite due to

$$\mathbf{x}^T [2\mathbf{A}^T \mathbf{A} + \beta \mathbf{G}^T \mathbf{G} - \beta^2 \mathbf{G}^T (\beta \mathbf{I} + \beta \text{diag}(\psi''(z_i)))^{-1} \mathbf{G}] \mathbf{x} \geq \mathbf{x}^T [2\mathbf{A}^T \mathbf{A}] \mathbf{x} > 0, \quad \forall \mathbf{x} \in \mathbb{R}^p,$$

we know that  $\mathbf{H}(\mathbf{x}, \mathbf{z})$  is a symmetric positive definite matrix. □

In each Newton's iteration, the Newton equation leads to a structured linear system

$$\mathbf{H}(\mathbf{x}, \mathbf{z}) \mathbf{d} = \mathbf{r}.$$

The main aim of this paper is to construct effective iterative methods such that this kind of structured linear systems can be solved quickly and efficiently. We employ the *preconditioned conjugate gradient (PCG)* method with modified block SSOR (*symmetric successive overrelaxation*) preconditioners [8,9,14,21] for  $\mathbf{H}(\mathbf{x}, \mathbf{z})$  to solve these structured linear systems. We will show that the PCG method with these modified block SSOR preconditioners is very effective for solving the edge-preserving signal and image restoration problems.

The rest of this paper is outlined as follows. In Section 2, we construct modified block SSOR preconditioners and discuss their preconditioning properties. In Section 3, we analyze the modified block SSOR preconditioners when they are specified to the edge-preserving signal and image restoration problems. Experimental results are presented in Section 4.

## 2. Modified Block SSOR Preconditioners

In this section, we present modified block SSOR preconditioners for the block 2-by-2 linear system  $\mathbf{H}(\mathbf{x}, \mathbf{z})\mathbf{d} = \mathbf{r}$ , which was first established and studied in [8,9] for general symmetric positive definite matrices of block structures. By decomposing the symmetric positive definite block 2-by-2 matrix  $\mathbf{H}(\mathbf{x}, \mathbf{z})$  (or  $\mathbf{H}$  for simplicity) into its block lower-triangular part  $\mathbf{L}$ , block diagonal part  $\mathbf{D}$  and block upper-triangular part  $\mathbf{L}^T$ , i.e.,

$$\mathbf{H} = \mathbf{L} + \mathbf{D} + \mathbf{L}^T,$$

with

$$\mathbf{L} = \begin{bmatrix} 0 & 0 \\ \mathbf{H}_{21} & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ -\beta\mathbf{G} & 0 \end{bmatrix},$$

$$\mathbf{D} = \begin{bmatrix} \mathbf{H}_{11} & 0 \\ 0 & \mathbf{H}_{22}(\mathbf{z}) \end{bmatrix} = \begin{bmatrix} 2\mathbf{A}^T\mathbf{A} + \beta\mathbf{G}^T\mathbf{G} & 0 \\ 0 & \beta\mathbf{I} + \beta \text{diag}(\psi^n(z_i)) \end{bmatrix},$$

we can construct the following modified block SSOR preconditioner for the matrix  $\mathbf{H}$ :

$$\mathbf{P}(\omega) = (\tilde{\mathbf{D}} + \omega\mathbf{L})^T \mathbf{Q}(\omega)^{-1} (\tilde{\mathbf{D}} + \omega\mathbf{L}), \quad (2.1)$$

where  $\tilde{\mathbf{D}}$  is an approximation to  $\mathbf{D}$  and

$$\mathbf{Q}(\omega) = \omega(\tilde{\mathbf{D}} + \tilde{\mathbf{D}}^T - \omega\mathbf{D}).$$

From [8,9] we know that the preconditioner  $\mathbf{P}(\omega)$  possesses the following property.

**Lemma 2.1.** ([8,9]) *Let  $\mathbf{Q}(\omega)$  be nonsingular. Then*

(a)  $\mathbf{P}(\omega) = \mathbf{H} + \mathbf{F}(\omega)$ , with

$$\mathbf{F}(\omega) = [\omega(\mathbf{D} + \mathbf{L}) - \tilde{\mathbf{D}}^T]^T \mathbf{Q}(\omega)^{-1} [\omega(\mathbf{D} + \mathbf{L}) - \tilde{\mathbf{D}}^T];$$

(b) if  $\mathbf{H}$  and  $\mathbf{Q}(\omega)$  are symmetric positive definite matrices, it holds that

$$\begin{cases} \max_{\mathbf{x} \neq \mathbf{0}} \frac{\langle \mathbf{x}, \mathbf{H}\mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{P}(\omega)\mathbf{x} \rangle} = 1, & \text{for } (\omega\mathbf{D} - \tilde{\mathbf{D}}^T) \text{ singular,} \\ \max_{\mathbf{x} \neq \mathbf{0}} \frac{\langle \mathbf{x}, \mathbf{H}\mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{P}(\omega)\mathbf{x} \rangle} < 1, & \text{for } (\omega\mathbf{D} - \tilde{\mathbf{D}}^T) \text{ nonsingular.} \end{cases}$$

Define

$$\mu_1 = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\langle \mathbf{x}, \mathbf{P}(\omega)\mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{H}\mathbf{x} \rangle} \quad \text{and} \quad \mu_2 = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\langle \mathbf{x}, \mathbf{H}\mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{P}(\omega)\mathbf{x} \rangle}.$$

Then it is obvious that the condition number of the matrix  $\mathbf{P}(\omega)^{-1}\mathbf{H}$ , denoted by  $\kappa(\mathbf{P}(\omega)^{-1}\mathbf{H})$ , is given by

$$\kappa(\mathbf{P}(\omega)^{-1}\mathbf{H}) = \mu_1\mu_2.$$

It follows from Lemma 2.1 that  $\mu_2$  is always less than or equal to one and, therefore,

$$\kappa(\mathbf{P}(\omega)^{-1}\mathbf{H}) \leq \mu_1.$$

Hence, to estimate  $\kappa(\mathbf{P}(\omega)^{-1}\mathbf{H})$ , we only need to estimate  $\mu_1$ .

To this end, we first symmetrically scale the involved matrices by the block-diagonal matrix  $\mathbf{D}^{-1/2}$  and then estimate the Rayleigh quotients about some matrix pairs among them. Let

$$\mathbf{W} = \frac{1}{2}(\tilde{\mathbf{D}} + \tilde{\mathbf{D}}^T), \quad \mathbf{V} = \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}, \tag{2.2a}$$

$$\mathbf{Z} = \frac{1}{2}(\tilde{\mathbf{D}} - \tilde{\mathbf{D}}^T), \quad \mathbf{Y} = \mathbf{D}^{-1/2}\mathbf{Z}\mathbf{D}^{-1/2}, \tag{2.2b}$$

$$\mathbf{K} = \mathbf{D}^{-1/2}\tilde{\mathbf{D}}\mathbf{D}^{-1/2}, \quad \mathbf{R} = \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}, \tag{2.2c}$$

$$\mathbf{M} = \mathbf{D}^{-1/2}\mathbf{H}\mathbf{D}^{-1/2}, \quad \mathbf{B}(\omega) = \mathbf{D}^{-1/2}\mathbf{P}(\omega)\mathbf{D}^{-1/2}. \tag{2.2d}$$

Then we immediately have

$$\begin{aligned} \mathbf{V} &= \frac{1}{2}(\mathbf{K} + \mathbf{K}^T), \quad \mathbf{Y} = \frac{1}{2}(\mathbf{K} - \mathbf{K}^T), \quad \mathbf{K} = \mathbf{V} + \mathbf{Y}, \quad \mathbf{K}^T = \mathbf{V} - \mathbf{Y}, \\ \mathbf{M} &= \mathbf{I} + \mathbf{R} + \mathbf{R}^T, \quad \mathbf{B}(\omega) = \frac{1}{\omega}(\mathbf{K} + \omega\mathbf{R})^T(2\mathbf{V} - \omega\mathbf{I})^{-1}(\mathbf{K} + \omega\mathbf{R}). \end{aligned}$$

Assume that  $\mathbf{W} \succeq \mathbf{D}$  (i.e.,  $\mathbf{W} - \mathbf{D}$  is symmetric positive semidefinite). Then it is easily seen that

$$\mathbf{V} \succeq \mathbf{I} \quad \text{and} \quad 2\mathbf{V} - \omega\mathbf{I} \succeq (2 - \omega)\mathbf{V}.$$

Hence, when  $0 < \omega < 2$  the matrix  $2\mathbf{V} - \omega\mathbf{I}$  is symmetric positive definite. It follows from the definition of  $\mathbf{B}(\omega)$  that

$$\mathbf{B}(\omega) \preceq \frac{1}{\omega}(\mathbf{K} + \omega\mathbf{R})^T[(2 - \omega)\mathbf{V}]^{-1}(\mathbf{K} + \omega\mathbf{R}),$$

or

$$\begin{aligned} \mathbf{B}(\omega) &\preceq \frac{1}{\omega(2 - \omega)}(\mathbf{K} + \omega\mathbf{R})^T\mathbf{V}^{-1}(\mathbf{K} + \omega\mathbf{R}) \\ &= \frac{1}{\omega(2 - \omega)}(\mathbf{V} + \mathbf{Y} + \omega\mathbf{R})^T\mathbf{V}^{-1}(\mathbf{V} + \mathbf{Y} + \omega\mathbf{R}) \\ &= \frac{1}{\omega(2 - \omega)}\left(\mathbf{V} + \omega(\mathbf{R} + \mathbf{R}^T) + \mathbf{Y}^T\mathbf{V}^{-1}\mathbf{Y} + \omega(\mathbf{R}^T\mathbf{V}^{-1}\mathbf{Y} + \mathbf{Y}^T\mathbf{V}^{-1}\mathbf{R}) \right. \\ &\quad \left. + \omega^2\mathbf{R}^T\mathbf{V}^{-1}\mathbf{R}\right). \end{aligned} \tag{2.3}$$

Note that for all  $\mathbf{x}$ , it holds that

$$\begin{aligned}
\mathbf{x}^T(\mathbf{R}^T\mathbf{V}^{-1}\mathbf{Y} + \mathbf{Y}^T\mathbf{V}^{-1}\mathbf{R})\mathbf{x} &= 2\mathbf{x}^T(\mathbf{R}^T\mathbf{V}^{-1}\mathbf{Y})\mathbf{x} \\
&= 2(\mathbf{V}^{-1/2}\mathbf{R}\mathbf{x})^T(\mathbf{V}^{-1/2}\mathbf{Y}\mathbf{x}) \\
&\leq 2(\mathbf{x}^T\mathbf{R}^T\mathbf{V}^{-1}\mathbf{R}\mathbf{x})^{1/2}(\mathbf{x}^T\mathbf{Y}^T\mathbf{V}^{-1}\mathbf{Y}\mathbf{x})^{1/2} \\
&\leq \mathbf{x}^T\mathbf{R}^T\mathbf{V}^{-1}\mathbf{R}\mathbf{x} + \mathbf{x}^T\mathbf{Y}^T\mathbf{V}^{-1}\mathbf{Y}\mathbf{x} \\
&= \mathbf{x}^T(\mathbf{R}^T\mathbf{V}^{-1}\mathbf{R} + \mathbf{Y}^T\mathbf{V}^{-1}\mathbf{Y})\mathbf{x}.
\end{aligned}$$

Combining (2.2)-(2.3), we obtain

$$\begin{aligned}
\mathbf{B}(\omega) &\preceq \frac{1}{\omega(2-\omega)} \left( \mathbf{V} + \omega\mathbf{M} + \omega^2\mathbf{I} + (1+\omega)\mathbf{Y}^T\mathbf{V}^{-1}\mathbf{Y} + \omega(1+\omega)(\mathbf{R}^T\mathbf{V}^{-1}\mathbf{R} - \mathbf{I}) \right) \\
&\preceq \frac{1}{\omega(2-\omega)} \left( \mathbf{V} + \omega\mathbf{M} + \omega^2\mathbf{I} + (1+\omega)\mathbf{Y}^T\mathbf{V}^{-1}\mathbf{Y} + \omega(1+\omega)(\mathbf{R}^T\mathbf{R} - \mathbf{I}) \right).
\end{aligned}$$

It then follows that

$$\mathbf{P}(\omega) \preceq \frac{1}{\omega(2-\omega)} \left( \mathbf{W} + \omega\mathbf{H} + \omega^2\mathbf{D} + (1+\omega)\mathbf{Z}^T\mathbf{W}^{-1}\mathbf{Z} + \omega(1+\omega)(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L} - \mathbf{D}) \right).$$

By using this property of  $\mathbf{P}(\omega)$ , we can estimate  $\mu_1$  as follows:

$$\begin{aligned}
\mu_1 &= \max_{\mathbf{x} \neq \mathbf{0}} \frac{\langle \mathbf{x}, \mathbf{P}(\omega)\mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{H}\mathbf{x} \rangle} \\
&\leq \max_{\mathbf{x} \neq \mathbf{0}} \frac{\langle \mathbf{x}, \frac{1}{\omega(2-\omega)} [\mathbf{W} + \omega\mathbf{H} + \omega^2\mathbf{D} + (1+\omega)\mathbf{Z}^T\mathbf{W}^{-1}\mathbf{Z} + \omega(1+\omega)(\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L} - \mathbf{D})] \mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{H}\mathbf{x} \rangle} \\
&= \frac{1}{\omega(2-\omega)} \left( \gamma^2\tau^2 + \omega + \gamma^2\omega^2 + (1+\omega)\gamma^2\theta^2 + \omega(1+\omega)\delta \right) \\
&= \frac{1}{\omega(2-\omega)} \left( \gamma^2(\theta^2 + \tau^2) + (1 + \gamma^2\theta^2 + \delta)\omega + (\gamma^2 + \delta)\omega^2 \right) \\
&:= f(\omega),
\end{aligned}$$

where

$$\gamma^2 = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\langle \mathbf{x}, \mathbf{D}\mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{H}\mathbf{x} \rangle}, \quad \tau^2 = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\langle \mathbf{x}, \mathbf{W}\mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{D}\mathbf{x} \rangle}, \quad \theta^2 = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\langle \mathbf{Z}\mathbf{x}, \mathbf{Z}\mathbf{x} \rangle}{\langle \mathbf{D}\mathbf{x}, \mathbf{D}\mathbf{x} \rangle}, \quad (2.4a)$$

$$\delta = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\langle \mathbf{x}, (\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L} - \mathbf{D})\mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{H}\mathbf{x} \rangle}. \quad (2.4b)$$

In addition, because

$$\begin{aligned}
f'(\omega) &= \frac{-2\gamma^2(\theta^2 + \tau^2) + 2\gamma^2(\theta^2 + \tau^2)\omega + 2(\gamma^2 + \delta)\omega^2 - 2(\gamma^2 + \delta)\omega^3}{\omega^2(2-\omega)^2} \\
&= \frac{2(1-\omega)[(\gamma^2 + \delta)\omega^2 - \gamma^2(\theta^2 + \tau^2)]}{\omega^2(2-\omega)^2},
\end{aligned}$$

we know that

$$\omega_1 = 1, \quad \text{and} \quad \omega_2 = \frac{\gamma\sqrt{\theta^2 + \tau^2}}{\sqrt{\gamma^2 + \delta}} \quad \text{for } \gamma^2 + \delta > 0, \quad (2.5)$$

are two stationary points of the function  $f(\omega)$ . After direct computations we have

$$f(\omega_1) = 1 + 2\delta + \gamma^2(1 + \tau^2 + 2\theta^2), \tag{2.6a}$$

$$f(\omega_2) = \frac{2\gamma\sqrt{(\theta^2 + \tau^2)(\gamma^2 + \delta)} + \gamma^2\theta^2 + \delta + 1}{\sqrt{\gamma^2 + \delta}[2\sqrt{\gamma^2 + \delta} - \gamma\sqrt{\theta^2 + \tau^2}]}, \tag{2.6b}$$

for

$$\delta > \gamma^2 \left[ \frac{1}{4}(\theta^2 + \tau^2) - 1 \right]. \tag{2.7}$$

Hence, the optimal relaxation parameter  $\omega^*$  is given by

$$\omega^* = \begin{cases} \omega_1, & \text{if } f(\omega_1) \leq f(\omega_2), \\ \omega_2, & \text{if } f(\omega_2) \leq f(\omega_1), \end{cases} \tag{2.8}$$

and the corresponding optimal estimate for  $\mu_1$  is given by

$$\mu_1 \leq \min\{f(\omega_1), f(\omega_2)\}.$$

In particular, when  $\delta \leq -\gamma^2$ , we have  $\omega^* = \omega_1 = 1$  and  $\mu_1 \leq f(\omega_1)$ .

By summarizing the above analysis, we can immediately obtain the following theorem.

**Theorem 2.1.** *Let  $\mathbf{H} := \mathbf{H}(\mathbf{x}, \mathbf{z})$  defined in (1.11) be the coefficient matrix of the Newton equation, and*

$$\mathbf{H} = \mathbf{L} + \mathbf{D} + \mathbf{L}^T$$

*be a splitting such that  $\mathbf{L}$  is a strictly block-lower triangular matrix and  $\mathbf{D}$  a block-diagonal matrix. Assume that  $\tilde{\mathbf{D}}$  is an approximation of  $\mathbf{D}$  satisfying  $\frac{1}{2}(\tilde{\mathbf{D}} + \tilde{\mathbf{D}}^T) \succeq \mathbf{D}$ . Then for the modified block SSOR preconditioner  $\mathbf{P}(\omega)$  defined in (2.1), it holds that*

$$\kappa(\mathbf{P}(\omega^*)^{-1}\mathbf{H}) \leq \min\{f(\omega_1), f(\omega_2)\},$$

*provided (2.7) is satisfied, where  $\omega^*$  is given by (2.8) and (2.5), and  $f(\omega_1)$  and  $f(\omega_2)$  are given by (2.6), with the constants  $\gamma, \tau$  and  $\theta$  being defined by (2.4a) and the constant  $\delta$  being defined by (2.4b).*

*When  $\delta \leq -\gamma^2$ , it holds that  $\omega^* = 1$  and*

$$\kappa(\mathbf{P}(\omega^*)^{-1}\mathbf{H}) \leq f(\omega_1) = 1 + 2\delta + \gamma^2(1 + \tau^2 + 2\theta^2).$$

We remark that  $\tau \geq 1$  because of  $\mathbf{W} \succeq \mathbf{D}$ . When  $\mathbf{W} = \mathbf{D}$ , it holds that  $\tau = 1$  and  $\theta = 0$ . In general, to estimate the optimal condition number  $\kappa(\mathbf{P}(\omega^*)^{-1}\mathbf{H})$  and the optimal relaxation parameter  $\omega^*$ , we need to compute the constants  $\gamma, \tau, \theta$  and  $\delta$ . This can be done by making use of either the actual structure and properties of the original problem, or some iterative methods for computing the largest eigenvalues of symmetric matrix pairs [14, 19].

### 3. Application and Analysis of Modified Block SSOR Preconditioners

In this section, we apply the modified block SSOR preconditioners discussed in Section 2 to the edge-preserving signal and image restoration problems, see also [8, 9].

Let  $\mathbf{C}$  be a preconditioner for the matrix  $\mathbf{H}_{11} = 2\mathbf{A}^T\mathbf{A} + \beta\mathbf{G}^T\mathbf{G}$  such that

$$\frac{1}{2}(\mathbf{C} + \mathbf{C}^T) \succeq 2\mathbf{A}^T\mathbf{A} + \beta\mathbf{G}^T\mathbf{G},$$

i.e.,

$$\mathbf{C} \approx 2\mathbf{A}^T\mathbf{A} + \beta\mathbf{G}^T\mathbf{G},$$

and define

$$\tilde{\mathbf{D}} = \begin{bmatrix} \mathbf{C} & 0 \\ 0 & \beta\mathbf{I} + \beta \operatorname{diag}(\psi''(z_i)) \end{bmatrix}.$$

Then it obviously holds that  $\frac{1}{2}(\tilde{\mathbf{D}} + \tilde{\mathbf{D}}^T) \succeq \mathbf{D}$ .

According to Theorem 2.1, to obtain an estimate about the condition number  $\kappa(\mathbf{P}(\omega)^{-1}\mathbf{H})$ , we only need to specifically estimate the constants  $\gamma^2$ ,  $\tau^2$ ,  $\theta^2$  and  $\delta$ . To this end, we note that

$$\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L} = \begin{bmatrix} \beta\mathbf{G}^T[\mathbf{I} + \operatorname{diag}(\psi''(z_i))]^{-1}\mathbf{G} & 0 \\ 0 & 0 \end{bmatrix}$$

and, therefore,

$$\begin{bmatrix} \beta\mathbf{G}^T\mathbf{G} & 0 \\ 0 & 0 \end{bmatrix} \succeq \mathbf{L}^T\mathbf{D}^{-1}\mathbf{L}.$$

Since

$$\begin{aligned} \mathbf{L}^T\mathbf{D}^{-1}\mathbf{L} - \mathbf{D} &\preceq \begin{bmatrix} \beta\mathbf{G}^T\mathbf{G} & 0 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} 2\mathbf{A}^T\mathbf{A} + \beta\mathbf{G}^T\mathbf{G} & 0 \\ 0 & \beta\mathbf{I} + \beta \operatorname{diag}(\psi''(z_i)) \end{bmatrix} \\ &= - \begin{bmatrix} 2\mathbf{A}^T\mathbf{A} & 0 \\ 0 & \beta\mathbf{I} + \beta \operatorname{diag}(\psi''(z_i)) \end{bmatrix} \\ &\preceq 0, \end{aligned}$$

or in other words,  $\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L} \preceq \mathbf{D}$ , we have

$$\begin{aligned} \delta &= \max_{\mathbf{x} \neq \mathbf{0}} \frac{\langle \mathbf{x}, (\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L} - \mathbf{D})\mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{H}\mathbf{x} \rangle} \\ &= \max_{\mathbf{x} \neq \mathbf{0}} \left\{ \frac{\langle \mathbf{x}, (\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L} - \mathbf{D})\mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{D}\mathbf{x} \rangle} \cdot \frac{\langle \mathbf{x}, \mathbf{D}\mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{H}\mathbf{x} \rangle} \right\} \\ &\leq \max_{\mathbf{x} \neq \mathbf{0}} \frac{\langle \mathbf{x}, (\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L} - \mathbf{D})\mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{D}\mathbf{x} \rangle} \cdot \min_{\mathbf{x} \neq \mathbf{0}} \frac{\langle \mathbf{x}, \mathbf{D}\mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{H}\mathbf{x} \rangle} \\ &= \max_{\mathbf{x} \neq \mathbf{0}} \frac{\langle \mathbf{x}, (\mathbf{L}^T\mathbf{D}^{-1}\mathbf{L} - \mathbf{D})\mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{D}\mathbf{x} \rangle} \cdot \eta^2, \end{aligned}$$

where

$$\eta^2 = \min_{\mathbf{x} \neq \mathbf{0}} \frac{\langle \mathbf{x}, \mathbf{D}\mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{H}\mathbf{x} \rangle}.$$



Let  $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2]^T$ . As

$$\begin{aligned} \max_{\mathbf{x} \neq \mathbf{0}} \frac{\langle \mathbf{x}, (\mathbf{L}^T \mathbf{D}^{-1} \mathbf{L} - \mathbf{D}) \mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{D} \mathbf{x} \rangle} &= - \min_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}_1^T (2\mathbf{A}^T \mathbf{A}) \mathbf{x}_1 + \mathbf{x}_2^T (\beta \mathbf{I} + \beta \operatorname{diag}(\psi''(z_i))) \mathbf{x}_2}{\mathbf{x}_1^T (2\mathbf{A}^T \mathbf{A} + \beta \mathbf{G}^T \mathbf{G}) \mathbf{x}_1 + \mathbf{x}_2^T (\beta \mathbf{I} + \beta \operatorname{diag}(\psi''(z_i))) \mathbf{x}_2} \\ &\leq - \min_{\mathbf{x}_1 \neq \mathbf{0}} \min \left\{ \frac{\mathbf{x}_1^T (2\mathbf{A}^T \mathbf{A}) \mathbf{x}_1}{\mathbf{x}_1^T (2\mathbf{A}^T \mathbf{A} + \beta \mathbf{G}^T \mathbf{G}) \mathbf{x}_1}, 1 \right\} \\ &= - \min_{\mathbf{x}_1 \neq \mathbf{0}} \frac{\mathbf{x}_1^T (2\mathbf{A}^T \mathbf{A}) \mathbf{x}_1}{\mathbf{x}_1^T (2\mathbf{A}^T \mathbf{A} + \beta \mathbf{G}^T \mathbf{G}) \mathbf{x}_1} \\ &= -1 + \max_{\mathbf{x}_1 \neq \mathbf{0}} \frac{\mathbf{x}_1^T (\beta \mathbf{G}^T \mathbf{G}) \mathbf{x}_1}{\mathbf{x}_1^T (2\mathbf{A}^T \mathbf{A} + \beta \mathbf{G}^T \mathbf{G}) \mathbf{x}_1} \\ &\leq -1 + \nu^2, \end{aligned}$$

where

$$\nu^2 = \max_{\mathbf{x}_1 \neq \mathbf{0}} \frac{\mathbf{x}_1^T (\beta \mathbf{G}^T \mathbf{G}) \mathbf{x}_1}{\mathbf{x}_1^T (2\mathbf{A}^T \mathbf{A} + \beta \mathbf{G}^T \mathbf{G}) \mathbf{x}_1} < 1, \quad (3.1)$$

we obtain

$$\delta \leq -(1 - \nu^2) \eta^2. \quad (3.2)$$

Next, we can demonstrate that the matrix  $\mathbf{H}$  satisfies the Cauchy-Bunyakovski-Schwarz (CBS) inequality [1, 5–7, 10, 20], i.e.,

$$|\mathbf{x}_1^T \mathbf{H}_{12} \mathbf{x}_2| \leq \nu \left( \mathbf{x}_1^T \mathbf{H}_{11} \mathbf{x}_1 \right)^{1/2} \cdot \left( \mathbf{x}_2^T \mathbf{H}_{22} \mathbf{x}_2 \right)^{1/2}.$$

See also [2–4]. In fact,

$$\begin{aligned} |\mathbf{x}_1^T \mathbf{H}_{12} \mathbf{x}_2|^2 &= |\mathbf{x}_1^T (-\beta \mathbf{G}^T) \mathbf{x}_2|^2 \\ &\leq \beta^2 \mathbf{x}_1^T \mathbf{G}^T \mathbf{G} \mathbf{x}_1 \cdot \mathbf{x}_2^T \mathbf{x}_2 \\ &\leq \beta \nu^2 \mathbf{x}_1^T (2\mathbf{A}^T \mathbf{A} + \beta \mathbf{G}^T \mathbf{G}) \mathbf{x}_1 \cdot \mathbf{x}_2^T \mathbf{x}_2 \\ &\leq \nu^2 \mathbf{x}_1^T (2\mathbf{A}^T \mathbf{A} + \beta \mathbf{G}^T \mathbf{G}) \mathbf{x}_1 \cdot \mathbf{x}_2^T (\beta \mathbf{I} + \beta \operatorname{diag}(\psi''(z_i))) \mathbf{x}_2 \\ &= \nu^2 \cdot \mathbf{x}_1^T \mathbf{H}_{11} \mathbf{x}_1 \cdot \mathbf{x}_2^T \mathbf{H}_{22} \mathbf{x}_2. \end{aligned}$$

By using this CBS inequality, we have

$$\begin{aligned} \mathbf{x}^T \mathbf{H} \mathbf{x} &= \mathbf{x}^T \mathbf{D} \mathbf{x} + 2\mathbf{x}_1^T \mathbf{H}_{12} \mathbf{x}_2 \\ &\leq \mathbf{x}^T \mathbf{D} \mathbf{x} + 2|\mathbf{x}_1^T \mathbf{H}_{12} \mathbf{x}_2| \\ &\leq \mathbf{x}^T \mathbf{D} \mathbf{x} + 2\nu \left( \mathbf{x}_1^T \mathbf{H}_{11} \mathbf{x}_1 \right)^{1/2} \cdot \left( \mathbf{x}_2^T \mathbf{H}_{22} \mathbf{x}_2 \right)^{1/2} \\ &\leq \mathbf{x}^T \mathbf{D} \mathbf{x} + \nu \left( \mathbf{x}_1^T \mathbf{H}_{11} \mathbf{x}_1 + \mathbf{x}_2^T \mathbf{H}_{22} \mathbf{x}_2 \right) \\ &= (1 + \nu) \mathbf{x}^T \mathbf{D} \mathbf{x}. \end{aligned}$$

Hence,

$$\eta^2 = \min_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}^T \mathbf{D} \mathbf{x}}{\mathbf{x}^T \mathbf{H} \mathbf{x}} \geq \frac{1}{1 + \nu}$$

and from (3.2),

$$\delta \leq -(1 - \nu^2) \eta^2 \leq -(1 - \nu^2) \cdot \frac{1}{1 + \nu} = -(1 - \nu).$$

Analogously, we can obtain

$$\begin{aligned}
\mathbf{x}^T \mathbf{H} \mathbf{x} &\geq \mathbf{x}^T \mathbf{D} \mathbf{x} - 2|\mathbf{x}_1^T \mathbf{H}_{12} \mathbf{x}_2| \\
&\geq \mathbf{x}^T \mathbf{D} \mathbf{x} - 2\nu \left( \mathbf{x}_1^T \mathbf{H}_{11} \mathbf{x}_1 \right)^{1/2} \cdot \left( \mathbf{x}_2^T \mathbf{H}_{22} \mathbf{x}_2 \right)^{1/2} \\
&\geq \mathbf{x}^T \mathbf{D} \mathbf{x} - \nu \left( \mathbf{x}_1^T \mathbf{H}_{11} \mathbf{x}_1 + \mathbf{x}_2^T \mathbf{H}_{22} \mathbf{x}_2 \right) \\
&= (1 - \nu) \mathbf{x}^T \mathbf{D} \mathbf{x}.
\end{aligned}$$

Therefore,

$$\gamma^2 = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}^T \mathbf{D} \mathbf{x}}{\mathbf{x}^T \mathbf{H} \mathbf{x}} \leq \frac{1}{1 - \nu}.$$

In addition, by noticing the special structure of the matrix  $\tilde{\mathbf{D}}$ , we can straightforwardly obtain the estimates about the constants  $\tau$  and  $\theta$ . In fact, from (2.4a) we have

$$\tau^2 = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\langle \mathbf{x}, \mathbf{W} \mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{D} \mathbf{x} \rangle} \leq \max \left\{ \max_{\mathbf{x}_1 \neq \mathbf{0}} \frac{\langle \mathbf{x}_1, \frac{1}{2}(\mathbf{C} + \mathbf{C}^T) \mathbf{x}_1 \rangle}{\langle \mathbf{x}_1, (2\mathbf{A}^T \mathbf{A} + \beta \mathbf{G}^T \mathbf{G}) \mathbf{x}_1 \rangle}, \quad 1 \right\}, \quad (3.3)$$

and

$$\begin{aligned}
\theta^2 &= \max_{\mathbf{x} \neq \mathbf{0}} \frac{\langle \mathbf{Z} \mathbf{x}, \mathbf{Z} \mathbf{x} \rangle}{\langle \mathbf{D} \mathbf{x}, \mathbf{D} \mathbf{x} \rangle} \\
&\leq \max \left\{ \max_{\mathbf{x}_1 \neq \mathbf{0}} \frac{\langle \frac{1}{2}(\mathbf{C} - \mathbf{C}^T) \mathbf{x}_1, \frac{1}{2}(\mathbf{C} - \mathbf{C}^T) \mathbf{x}_1 \rangle}{\langle (2\mathbf{A}^T \mathbf{A} + \beta \mathbf{G}^T \mathbf{G}) \mathbf{x}_1, (2\mathbf{A}^T \mathbf{A} + \beta \mathbf{G}^T \mathbf{G}) \mathbf{x}_1 \rangle}, \quad 1 \right\}. \quad (3.4)
\end{aligned}$$

By summarizing the above analysis, we can obtain the following theorem.

**Theorem 3.1.** *Let  $\mathbf{H} := \mathbf{H}(\mathbf{x}, \mathbf{z})$  defined in (1.11) be the coefficient matrix of the Newton equation, and*

$$\mathbf{H} = \mathbf{L} + \mathbf{D} + \mathbf{L}^T$$

*be a splitting such that  $\mathbf{L}$  is a strictly block-lower triangular matrix and  $\mathbf{D}$  a block-diagonal matrix. Assume that  $\mathbf{C}$  is an approximation of  $2\mathbf{A}^T \mathbf{A} + \beta \mathbf{G}^T \mathbf{G}$  satisfying*

$$\frac{1}{2}(\mathbf{C} + \mathbf{C}^T) \succeq 2\mathbf{A}^T \mathbf{A} + \beta \mathbf{G}^T \mathbf{G}.$$

*Then for the modified block SSOR preconditioner  $\mathbf{P}(\omega)$  defined in (2.1), it holds that*

$$\kappa(\mathbf{P}(\omega^*)^{-1} \mathbf{H}) \leq \min\{f(\omega_1), f(\omega_2)\},$$

where

$$\begin{aligned}
\omega_1 &= 1, \quad \omega_2 = \sqrt{\frac{\theta^2 + \tau^2}{\nu(2 - \nu)}}, \\
f(\omega_1) &= \frac{\nu(3 - 2\nu) + \tau^2 + 2\theta^2}{1 - \nu}, \\
f(\omega_2) &= \frac{2\sqrt{\nu(2 - \nu)(\theta^2 + \tau^2)} - \theta^2 + \nu(1 - \nu)}{2\nu(2 - \nu) - \sqrt{\nu(2 - \nu)(\theta^2 + \tau^2)}},
\end{aligned}$$

and

$$\omega^* = \begin{cases} \omega_1, & \text{if } f(\omega_1) \leq f(\omega_2), \\ \omega_2, & \text{if } f(\omega_2) \leq f(\omega_1), \end{cases}$$

with  $\nu$  the CBS constant defined in (3.1) and  $\tau$  and  $\theta$  the constants defined in (3.3) and (3.4), respectively.

**Remark 3.1.** When  $\mathbf{C}$  is symmetric, we have  $\theta = 0$ . In particular, when  $\mathbf{C} = 2\mathbf{A}^T\mathbf{A} + \beta\mathbf{G}^T\mathbf{G}$ , we have  $\theta = 0$  and  $\tau = 1$ .

**Remark 3.2.** In actual implementations, the modified block SSOR preconditioner can be applied by adopting the Eisenstat trick [18] to save the computing cost.

#### 4. Numerical Results

In this section, we present experimental results to illustrate the effectiveness of the preconditioning and the corresponding PCG for solving the image restoration problem in which half-quadratic regularization is applied. All codes are written in MATLAB 7.01 and all experiments are implemented on a personal computer with 1.86GHz central processing unit and 512M memory. In our computations, the initial vector  $\mathbf{x}$  is set to be the observed image and  $\mathbf{z}$  is set to be a constant vector; and the outer Newton iteration is stopped once the current residual satisfies the criteria

$$\frac{\|\mathbf{r}^{(k)}\|_2}{\|\mathbf{r}^{(0)}\|_2} \leq 10^{-6}.$$

The image “Cameraman” from the MATLAB toolbox is used in our experiments. The original “Cameraman” image is shown in Figure 4.1. An averaging function [17] is used to blur this image and a Gaussian white noise with the standard deviation 0.001 is added. The observed image is shown in Figure 4.2 (left).



Fig. 4.1. The original “Cameraman” image.

In computations of the image restoration, we take  $\mathbf{C} = 2\mathbf{A}^T\mathbf{A} + \beta\mathbf{G}^T\mathbf{G}$ . With this choice, we obtain the estimates stated in Remark 3.1. In the Newton method, the updating scheme is given by

$$\begin{bmatrix} \mathbf{x}^{(k+1)} \\ \mathbf{z}^{(k+1)} \end{bmatrix} = \begin{bmatrix} \mathbf{x}^{(k)} \\ \mathbf{z}^{(k)} \end{bmatrix} - \gamma_k \mathbf{H}(\mathbf{x}^{(k)}, \mathbf{z}^{(k)})^{-1} \nabla \tilde{J}(\mathbf{x}^{(k)}, \mathbf{z}^{(k)})^T,$$

where  $\gamma_k$  is the step-size determined by a line-search procedure of lower computational cost. The PCG iteration method incorporated with the modified block SSOR preconditioner is applied to solve the sub-system of linear equations  $\mathbf{H}(\mathbf{x}, \mathbf{z})\mathbf{d} = \mathbf{r}$ . In general, the Newton method is more efficient than the alternating iteration method proposed in [15] when an accurate solution is desired. Figure 4.2 (right) shows the restored image using the edge-preserving regularization function  $\phi_1$  in (1.3). We see from this figure that the edge-preserving solution tends to sharpen the edges. Table 4.1 summarizes the computing results corresponding to this image restoration, where the column labelled with “**I**” represents the results without using a preconditioner, while the column labelled with “**P**” represents the results using the modified block SSOR preconditioner **P**. In addition, we use “IT” to denote the number of iteration steps of the Newton method, “PCG” the average number of iteration steps of the PCG method with respect to different values of the relaxation parameter  $\omega$  listed in the first column, and “CPU” the total computing time for the overall iteration process. From Table 4.1 we see that the average number of PCG iteration steps is greater than 200 when no preconditioner is applied, and it is less than 10, however, when the modified block SSOR preconditioner is used. The computing efficiency can be improved up to about 50 times for the preconditioned case. Also, we note that the number of PCG iteration steps is about the same for different values of  $\omega$ , which shows that the modified block SSOR preconditioner is not sensitive to the relaxation parameter  $\omega$ . In particular, when  $\omega = 1$  we get the modified block symmetric Gauss-Seidel preconditioner, and the corresponding computing results are: IT = 3, PCG = 4.7 and CPU = 14.9. It is interesting that these results are the best among all values of  $\omega$  in the table. We remark that the image restoration effect and computing results of using other regularization functions  $\phi_2, \phi_3$  and  $\phi_4$  defined by (1.4)-(1.6), respectively, are about the same.

Next, we present experimental results in which the “Cameraman” image is blurred by a

Table 4.1: The numerical results for the image restoration where “Cameraman” is blurred by the average function.

$\omega$	<b>I</b>			<b>P</b>		
	IT	PCG	CPU	IT	PCG	CPU
0.3				3	7.3	22.6
0.6				3	6.3	19.8
0.9	3	> 200	> 240	3	5.0	16.0
1.2				3	5.0	16.1
1.5				3	6.0	18.5
1.8				3	8.0	24.1

Table 4.2: The numerical results for the image restoration where “Cameraman” is blurred by the Gaussian function.

$\omega$	<b>I</b>			<b>P</b>		
	IT	PCG	CPU	IT	PCG	CPU
0.3				2	6.5	12.2
0.6				2	5.0	10.6
0.9	3	> 200	> 240	2	4.0	8.4
1.2				2	4.0	8.6
1.5				2	5.5	11.7
1.8				2	6.5	12.7



Fig. 4.2. The blurred and noisy image by the average blur (left) and the restored image (right).



Fig. 4.3. The blurred and noisy image by Gaussian blur (left) and the restored image (right).

two-dimensional Gaussian function, say,

$$h(i, j) = e^{-2(i/3)^2 - 2(j/3)^2},$$

that is truncated such that the blurring function has a support of  $7 \times 7$ . A Gaussian white noise with standard deviation 0.001 is added, too. The blurred and noisy image is shown in Figure 4.3 (left). Again, we take  $\mathbf{C} = 2\mathbf{A}^T\mathbf{A} + \beta\mathbf{G}^T\mathbf{G}$ . As an example, the restored image using the edge-preserving regularization function  $\phi_1$  in (1.3) is shown in Figure 4.3 (right). In Table 4.2 we list the computing results corresponding to this restoration. The computing efficiency for this example can be improved up to about 50 times for the preconditioned case. Hence, we can conclude that the modified block SSOR preconditioner shows about the same efficiency for image restoration degraded by different blurs.

Figure 4.4 depicts the spectral distribution of the original coefficient matrix  $\mathbf{H}$  and the preconditioned matrix  $\mathbf{P}^{-1}\mathbf{H}$  when the Newton method is applied to the first restoration. This figure clearly shows that the matrices without preconditioning are very ill-conditioned and, therefore, the corresponding conjugate gradient method may be convergent slowly; the matrices with preconditioning are, however, well-conditioned as they have tightly clustered eigenvalues and, therefore, the corresponding PCG method converges faster.

Numerical implementations also indicate that using the other regularization functions  $\phi_2$ ,  $\phi_3$  and  $\phi_4$  defined by (1.4)-(1.6), respectively, and different parameters  $\alpha$  and  $\beta$ , leads to similar numerical results.

Finally, let us discuss the computational aspects of the Newton method and the proposed

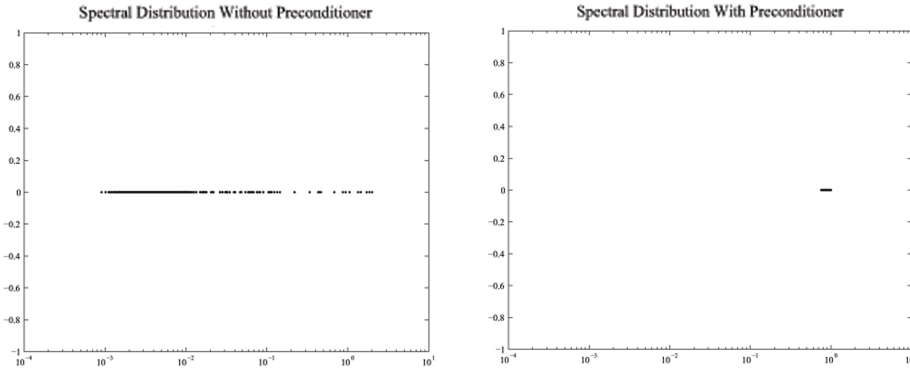


Fig. 4.4. The spectral distributions without preconditioner (left) and with the modified block SSOR preconditioner (right).

modified block SSOR preconditioner. In the method proposed in [15], the calculation of  $\mathbf{z}^{(k)}$  is of linear complexity. The iterate  $\mathbf{x}^{(k)}$  is the solution of the linear system

$$(2\mathbf{A}^T \mathbf{A} + \beta \mathbf{G}^T \mathbf{G})\mathbf{x}^{(k)} = 2\mathbf{A}^T \mathbf{b} + \beta \mathbf{G}^T \mathbf{z}^{(k)}. \tag{4.1}$$

Evidently, the coefficient matrix is constantly fixed, and only the right-hand side is changing with the iteration step. When  $\mathbf{A}$  is a Toeplitz or a block-Toeplitz-Toeplitz-block matrix and  $\mathbf{G}$  is the discretization matrix of the first-order or the second-order difference operator, the transform-based preconditioning techniques have been proved to be very successful [16]. For instance, if  $\mathbf{A}$  is a blurring matrix generated by a symmetric point spread function, then the coefficient matrix of the linear system (4.1) can be diagonalized by a fast transform matrix. It then follows that this linear system can be solved by using three fast transforms in  $\mathcal{O}(p^2 \log p)$  operations [16], where  $p$  is the size of the solution of the linear system (4.1). For the afore-established modified block SSOR preconditioner, the main computational cost is to solve linear systems with the coefficient matrix  $\mathbf{P}(\omega)$  defined by (2.1), with

$$\tilde{\mathbf{D}} = \begin{bmatrix} \mathbf{C} & 0 \\ 0 & \beta \mathbf{I} + \beta \text{diag}(\psi''(z_i^{(k)})) \end{bmatrix}, \quad \mathbf{L} = \begin{bmatrix} 0 & 0 \\ -\beta \mathbf{G} & 0 \end{bmatrix}.$$

It is interesting to see that we have chosen  $\mathbf{C} \approx 2\mathbf{A}^T \mathbf{A} + \beta \mathbf{G}^T \mathbf{G}$  such that

$$\frac{1}{2}(\mathbf{C} + \mathbf{C}^T) \succeq 2\mathbf{A}^T \mathbf{A} + \beta \mathbf{G}^T \mathbf{G}$$

(see Section 3). Note that the matrix  $\beta \mathbf{I} + \beta \text{diag}(\psi''(z_i^{(k)}))$  is diagonal and the matrix  $\mathbf{C}$  may be more easily invertible than the matrix  $2\mathbf{A}^T \mathbf{A} + \beta \mathbf{G}^T \mathbf{G}$ . Both Newton method and the approach proposed in [15] require to solve linear systems of the same type in about the same computational costs. However, after the modified block SSOR preconditioner is used, each inner PCG iteration converges fast, and outer Newton iteration requires only several steps to achieve the prescribed tolerance, the computing time costed in the proposed Newton-PCG iteration process is much less, too.

In summary, we have considered a class of convex and edge-preserving regularization functions for image restoration problems and solved them by the Newton method incorporated with

the preconditioned conjugate gradient method using the modified block SSOR preconditioner. The experimental results have shown that this approach is more feasible and effective than the alternating iteration method in [15].

**Acknowledgements.** The research of Bai is supported by the China NSF Outstanding Young Scientist Foundation (No. 10525102), the National Natural Science Foundation (No. 10471146), and the National Basic Research Program (No. 2005CB321702), P.R. China. The research of Huang is supported in part by the Fundamental Research Fund for Physics and Mathematics of Lanzhou University, P.R. China. The research of Ng is supported in part by Hong Kong Research Grants Council Grant Nos. 7035/04P and 7035/05P, and HKBU FRGs.

## References

- [1] O. Axelsson, *Iterative Solution Methods*, Cambridge University Press, Cambridge, 1994.
- [2] O. Axelsson and P.S. Vassilevski, Algebraic multilevel preconditioning methods. I, *Numer. Math.*, **56** (1989), 157-177.
- [3] O. Axelsson and P.S. Vassilevski, Algebraic multilevel preconditioning methods. II, *SIAM J. Numer. Anal.*, **27** (1990), 1569-1590.
- [4] O. Axelsson and P.S. Vassilevski, A survey of multilevel preconditioned iterative methods, *BIT Numer. Math.*, **29** (1989), 769-793.
- [5] Z.-Z. Bai, *Parallel Iterative Methods for Large-Scale Systems of Algebraic Equations*, *Ph.D. Thesis*, Shanghai University of Science and Technology, Shanghai, June 1993 (in Chinese).
- [6] Z.-Z. Bai, A class of hybrid algebraic multilevel preconditioning methods, *Appl. Numer. Math.*, **19** (1996), 389-399.
- [7] Z.-Z. Bai, Parallel hybrid algebraic multilevel iterative methods, *Linear Algebra Appl.*, **267** (1997), 281-315.
- [8] Z.-Z. Bai, A class of modified block SSOR preconditioners for symmetric positive definite systems of linear equations, *Adv. Comput. Math.*, **10** (1999), 169-186.
- [9] Z.-Z. Bai, Modified block SSOR preconditioners for symmetric positive definite linear systems, *Ann. Oper. Res.*, **103** (2001), 263-282.
- [10] Z.-Z. Bai and D.-R. Wang, A class of new hybrid algebraic multilevel preconditioning methods, *Linear Algebra Appl.*, **260** (1997), 223-255.
- [11] M.R. Banham and A.K. Katsaggelos, Digital image restoration, *IEEE Signal Proc. Mag.*, **March** (1997), 24-41.
- [12] D. Geman and G. Reynolds, Constrained restoration and the recovery of discontinuities, *IEEE Trans. Pattern Anal. Machine Intelligence*, **14** (1992), 367-383.
- [13] D. Geman and C. Yang, Nonlinear image recovery with half-quadratic regularization, *IEEE Trans. Image Process.*, **4** (1995), 932-946.
- [14] G.H. Golub and C.F. Van Loan, *Matrix Computations*, The 3rd Edition, The Johns Hopkins University Press, Baltimore, 1996.
- [15] M.P. Nikolova and M.K. Ng, Analysis of half-quadratic minimization methods for signal and image recovery, *SIAM J. Sci. Comput.*, **27** (2005), 937-966.
- [16] M.K. Ng, *Iterative Methods for Toeplitz Systems*, Oxford University Press, New York, 2004.
- [17] M.K. Ng and N.K. Bose, Mathematical analysis of super-resolution methodology, *IEEE Signal Proc. Mag.*, **20** (2003), 62-74.
- [18] J.M. Ortega, *Introduction to Parallel and Vector Solution of Linear Systems*, Plenum Press, New York, 1989.
- [19] B.N. Parlett, *The Symmetric Eigenvalue Problem*, Prentice-Hall, Englewood Cliffs, N.J., 1980.
- [20] D.-R. Wang and Z.-Z. Bai, Parallel multilevel iterative methods, *Linear Algebra Appl.*, **250** (1997), 317-347.

- [21] D.M. Young, *Iterative Solution of Large Linear Systems*, Academic Press, New York, 1971.