Open access • Journal Article • DOI:10.1137/040616048

Preconditioned Iterative Methods for Weighted Toeplitz Least Squares Problems

Michele Benzi, Michael K. Ng

Institutions: Emory University, University of Hong Kong

Published on: 31 Dec 2005 - SIAM Journal on Matrix Analysis and Applications (Society for Industrial and Applied Mathematics)

Topics: Toeplitz matrix, Preconditioner, Iterative method, Least squares and Hermitian matrix

SIAM J. MATRIX ANAL. APPL.
Vol. 27, No. 4, pp. 1106–1124

PRECONDITIONED ITERATIVE METHODS FOR WEIGHTED TOEPLITZ LEAST SQUARES PROBLEMS*

MICHELE BENZI† AND MICHAEL K. NG‡

Abstract. We consider the iterative solution of weighted Toeplitz least squares problems. Our approach is based on an augmented system formulation. We focus our attention on two types of preconditioners: a variant of constraint preconditioning, and the Hermitian/skew-Hermitian splitting (HSS) preconditioner. Bounds on the eigenvalues of the preconditioned matrices are given in terms of problem and algorithmic parameters, and numerical experiments are used to illustrate the performance of the preconditioners.

Key words. augmented matrix, iterative methods, preconditioning, splittings, constraint preconditioning, eigenvalue bounds, image restoration

AMS subject classifications. Primary, 65F10, 65N22, 65F50; Secondary, 15A06

DOI. 10.1137/040616048

1. Introduction.

Linear systems with Toeplitz and Toeplitz-related coefficient matrices arise in many different applications; see [28]. While many efficient algorithms have been developed for solving problems with Toeplitz structure, a few emerging applications lead to Toeplitz-related problems for which the available algorithms are not directly applicable.

In this paper, we consider preconditioners for weighted Toeplitz least squares problems

$$\min_x \|Ax - b\|_2^2, \tag{1.1}$$

where the rectangular coefficient matrix A and the right-hand side b are of the form

$$A = \begin{bmatrix} DK \\ \mu I \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} Df \\ 0 \end{bmatrix}. \tag{1.2}$$

Here K is a Toeplitz matrix, D a nonconstant diagonal matrix with real positive entries, f a given right-hand side, $\mu > 0$ a regularization parameter [38], and I the identity matrix. Applications leading to such least squares problems include image reconstruction [17] and nonlinear image restoration [2]. For these applications the size of the problems can be very large, easily over a million unknowns. Moreover, in some applications, the problems have to be solved in real time. Because of the local nature and spatially variant property of weighted Toeplitz matrices, the displacement rank [22] of A can be very large. Efficient and effective preconditioners need to be

*Received by the editors September 30, 2004; accepted for publication (in revised form) by E. Chow May 6, 2005; published electronically March 17, 2006.
http://www.siam.org/journals/simax/27-4/61604.html

†Department of Mathematics and Computer Science, Emory University, Atlanta, GA 30322 (benzi@mathcs.emory.edu). The work of this author was supported in part by National Science Foundation grant DMS-0207599.

‡Department of Mathematics, The University of Hong Kong, Pokfulam Road, Hong Kong, and Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong (mng@maths.hku.hk). The research of this author was supported in part by Hong Kong Research Grants Council grants HKU 7130/02P, HKU 7046/03P, and HKU 7035/04P, and HKU CRCG grants 10205964, 10205775, and 10204437.

investigated to develop fast iterative methods for solving such weighted Toeplitz least squares problems.

Around 1947, Levinson [25], in studying Wiener filters, developed the first fast direct method for solving symmetric positive definite Toeplitz systems $Kx = b$. The method begins with the solution of a 1-by-1 system and then increases the order, using the lower-order solutions to obtain higher-order solutions recursively. In effect, it obtains the inverse Cholesky factorization of K. The fast Cholesky factorization, or Schur's method, was invented by Kailath, Kung, and Morf [21] in 1979 by exploiting the displacement structure of K and its Schur complements. Both methods have a complexity of $O(n^2)$. Superfast direct methods of complexity $O(n \log^2 n)$ were developed in the 1980s; see, for instance, Ammar and Gragg [1]. All these methods make use of the Toeplitz structure (small displacement rank) to derive efficient solvers; see [24, 31]. They are not effective or efficient for Toeplitz-related systems with large displacement rank.[1]

An alternative is to make use of iterative methods instead of direct ones. One can use the conjugate gradient method to find the solution to the least squares problem without explicitly forming the normal equations; see [9]. The convergence of the conjugate gradient algorithm and its variants depends on the singular values of the matrix A. If the singular values cluster around a point (away from zero), convergence will be rapid. Thus, to make the algorithm a useful iterative method, one usually preconditions the system. More precisely, one can use the conjugate gradient method to solve

$$\min_y \|b - AP^{-1}y\|_2^2,$$

and then set $x = P^{-1}y$.

Circulant matrices C are often employed to precondition Toeplitz matrices. Since circulant matrices can always be diagonalized by the discrete Fourier matrix [14], the product of C with any n-vector v can be computed easily by fast Fourier transforms (FFTs) in $O(n \log n)$ operations. The matrix-vector multiplication Kv can also be computed by using FFTs in $O(n \log n)$ operations by first embedding K into a $2n$-by-$2n$ circulant matrix. It follows that the number of operations per iteration is $O(n \log n)$. The convergence rates of the method for Toeplitz systems have been analyzed by several authors; see [28]. The main result is that if the diagonals of the Toeplitz matrix K are the Fourier coefficients of a positive function, then the spectrum of the preconditioned system $C^{-1}K$ will be clustered around 1 and the method will converge superlinearly. Hence the complexity of solving a large class of Toeplitz systems is $O(n \log n)$ operations. Similar convergence results for Toeplitz least squares problems with small displacement rank can be found in [10, 11, 12, 13]. For example, when D is the identity matrix, $(I + C^T C)^{1/2}$ will be a good preconditioner.

However, such a preconditioner is not well suited for least squares problems with large displacement rank in (1.1). Since C is a good preconditioner for K, one might expect the matrix $(I + C^T D \cdot DC)^{1/2}$ to be a good preconditioner for (1.1). However, such a preconditioner is not easy to obtain, and it is too expensive. One possible choice is to approximate D by a circulant matrix C_D; see [20]. It is clear that the

[1] For any square matrix A, consider the matrix operator

$$\triangle(A) = AZ - ZA, \quad Z = (z_{ij}),\ z_{i+1,i} = 1,\ z_{i,j} = 0,\ \text{otherwise.}$$

The displacement rank of A associated with $\triangle$ is defined as the rank of $\triangle(A)$.

eigendecomposition of $(I + C^T C_D^T C_D C)^{1/2}$ can be obtained easily by using FFTs. Unfortunately, if the entries in D exhibit large variations, the performance of this kind of preconditioner is disappointing; i.e., the number of iterations can be very large; see the numerical results in [26].

In this paper, we propose a different approach to solving (1.1), based on an augmented system formulation. If K is m-by-n (with $m \geq n$), it is straightforward to see that problem (1.1) with K given by (1.2) is equivalent to the $(m+n)$-by-$(m+n)$ augmented system

$$\begin{bmatrix} D^{-2} & K \\ K^T & -\mu I \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} = \begin{bmatrix} f \\ 0 \end{bmatrix}, \tag{1.3}$$

where the auxiliary variable $y = D(f - K\,x)$ represents a weighted residual. Furthermore, both problem (1.1)–(1.2) and problem (1.3) are equivalent to the normal equations

$$(K^T D^2 K + \mu I)\, x = K^T D^2 f\,, \tag{1.4}$$

an n-by-n linear system. As is well known, problem (1.4) is a symmetric positive definite system which may be rather ill-conditioned, even in the presence of regularization. On the other hand, system (1.3) is symmetric indefinite. In recent years, much effort has been put into developing effective preconditioners for linear systems of this type; see [8] for a survey. We identify two preconditioning techniques that appear to be potentially well suited for augmented systems corresponding to weighted Toeplitz least-squares problems: a variant of constraint preconditioning, and the Hermitian/skew-Hermitian splitting (HSS) preconditioner.

The remainder of the paper is organized as follows. In section 2 we recall the idea of constraint preconditioning and discuss how it applies to the regularized weighted least squares problems. In section 3 we recall the Hermitian and skew-Hermitian preconditioner and its application to saddle point (augmented) systems. In section 4 we discuss some theoretical properties of the HSS preconditioner, including eigenvalue bounds for the preconditioned matrices. In section 5 we discuss nonlinear image restoration problems, an important class of applications leading to weighted Toeplitz least squares problems. Numerical experiments are presented in section 6, and in section 7 we briefly sum up with conclusions.

2. Constraint preconditioning. Constraint preconditioning has been extensively used in the solution of saddle point systems arising from mixed finite element formulations of elliptic partial differential equations [3, 16, 32, 33]. The method has also been used for the solution of saddle point (KKT) systems in optimization [15, 18, 23, 27].

We begin by describing constraint preconditioning in the case where K has full column rank ($= n$) and $\mu = 0$ (no regularization). Letting $W = D^{-2} = \mathrm{diag}\,(w_1, w_2, \ldots, w_m)$, the augmented system is

$$\begin{bmatrix} W & K \\ K^T & O \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} = \begin{bmatrix} f \\ 0 \end{bmatrix}.$$

In this paper the constraint preconditioner is the matrix

$$P_c = \begin{bmatrix} \gamma I & K \\ K^T & O \end{bmatrix}, \tag{2.1}$$

where γ is the arithmetic average of the diagonal entries of W:

$$\gamma = \frac{w_1 + w_2 + \cdots + w_m}{m}\,.$$

The application of the preconditioner requires solving a linear system with coefficient matrix P_C in each iteration. Since P_c has a block Toeplitz structure, it is clear that fast solvers can be used to efficiently apply the preconditioner. Note that γI is the nearest Toeplitz (indeed, circulant) approximation to W. Some authors reserve the name "constraint preconditioning" to a preconditioner of the form (2.1) with $\gamma = 1$. We found that such a preconditioner is inferior to the one where γ is given by the arithmetic average of the diagonal entries of W.

Letting $\widehat{W} = \gamma^{-1}W$, the preconditioned matrix is

$$\mathcal{M} = \begin{bmatrix} \gamma I & K \\ K^T & O \end{bmatrix}^{-1} \begin{bmatrix} W & K \\ K^T & O \end{bmatrix} = \begin{bmatrix} (I-\Pi)\,\widehat{W} + \Pi & O \\ X & I \end{bmatrix},$$

where Π is the orthogonal projector onto the range of K and $X = (K^TK)^{-1}K^T(W - \gamma I)$. Hence, $\lambda = 1$ is an eigenvalue of $\mathcal{M}$ of multiplicity at least $2n$. It can be shown that the remaining eigenvalues are eigenvalues of the symmetric matrix $(I-\Pi)\widehat{W}(I-\Pi)$. Note that in the special case $m = n$, we have $\Pi = I$, and all the eigenvalues of the preconditioned matrix are equal to 1; furthermore,

$$\mathcal{M} - I = \begin{bmatrix} O & O \\ X & O \end{bmatrix} \Rightarrow (\mathcal{M} - I)^2 = O\,.$$

Therefore the minimum polynomial of the preconditioned matrix has degree 2, and GMRES [34] applied to the preconditioned system delivers the solution in at most two steps, independently of W. Of course, if $m = n$ and K is nonsingular, then the solution is simply given by $x = K^{-1}f$ and $y = 0$, and the augmented system formulation is neither necessary nor recommended.

In the more general case where $m \neq n$, it can be shown that GMRES with constraint preconditioning terminates after at most $m - n + 2$ steps if K is m-by-n; see [23]. Therefore, constraint preconditioning is an excellent strategy when K has full rank, $\mu = 0$, and $m - n$ is small, no matter what W is.

Things, however, can be quite different when K is rank deficient or highly ill-conditioned and regularization must be included ($\mu \neq 0$). In this case the constraint preconditioner needs to be regularized as well, and the preconditioned matrix becomes

$$\mathcal{M}_\mu = \begin{bmatrix} \gamma I & K \\ K^T & -\mu I \end{bmatrix}^{-1} \begin{bmatrix} W & K \\ K^T & -\mu I \end{bmatrix}.$$

This modified constraint preconditioner has been studied by various authors, most recently in [3]. It can be shown that when $\mu > 0$, the preconditioned matrix $\mathcal{M}_\mu$ has the eigenvalue 1 with multiplicity n, and that all the remaining eigenvalues are real.

If K and W are well conditioned or moderately ill conditioned, the spectrum is clustered around 1 and convergence is fast. If, however, K is very ill conditioned (as it will be if regularization is needed), many of the eigenvalues of $\mathcal{M}_\mu$ will be close to zero and the preconditioner quality will deteriorate; see the numerical experiments in section 6.

3. Hermitian and skew-Hermitian preconditioning. The augmented system (1.3) can be rewritten in the equivalent nonsymmetric form

$$\begin{bmatrix} W & K \\ -K^T & \mu I \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} = \begin{bmatrix} f \\ 0 \end{bmatrix} \quad \text{or} \quad Mu = c\,. \tag{3.1}$$

Let now

$$M = \begin{bmatrix} W & K \\ -K^T & \mu I \end{bmatrix} = \begin{bmatrix} W & O \\ O & \mu I \end{bmatrix} + \begin{bmatrix} O & K \\ -K^T & O \end{bmatrix} = H + S \tag{3.2}$$

be the splitting of M into its symmetric and skew-symmetric parts. Note that H, the symmetric part of M, is a positive diagonal matrix and hence positive definite. This means that M itself is positive definite, in the sense that

$$z^T M z = z^T H z > 0 \quad \text{for all } z \in \mathbb{R}^{m+n}\,,\ z \neq 0\,,$$

and therefore all the eigenvalues of M have positive real part.

Consider the preconditioning matrix

$$P := \frac{1}{2\alpha}(H + \alpha I)(S + \alpha I), \tag{3.3}$$

where I denotes the $(m+n)$-by-$(m+n)$ identity and $\alpha > 0$. Note that $H + \alpha I$ is diagonal and positive definite, and $S + \alpha I$ is nonsingular. Thus, this preconditioner is well defined even when some of the entries on the diagonal of W are zero. We note that this case does indeed arise in many situations of practical interest [9, 20]. The scalar $1/(2\alpha)$ in (3.3) has no impact on the preconditioned system $P^{-1}Mu = P^{-1}c$ (or $MP^{-1}w = c$, $u = P^{-1}w$ for right preconditioning), but we retain it as a normalization factor (see below). The stationary iterative process

$$u^{(k+1)} = u^{(k)} + P^{-1}(c - Mu^{(k)}), \quad k = 0, 1, \ldots, \tag{3.4}$$

where $u^{(0)}$ is an arbitrary initial guess, has been introduced in [4]. It is known as the HSS iteration.[2] It was shown in [4] that if H is positive definite, the iteration (3.4) converges to the unique solution of $Mu = c$ for any initial guess and any $\alpha > 0$. This result was extended in [7] to generalized saddle point problems, in which the symmetric part of M may be singular. When H is positive definite, it is possible to give an explicit formula for the value of α that minimizes an upper bound for the spectral radius of the iteration matrix $I - P^{-1}M$ in terms of the extreme eigenvalues of H; see [4].

Note that each iteration (3.4) requires solving two linear systems:

$$(H + \alpha I)\, u^{(k+\frac{1}{2})} = r^{(k)}, \quad (S + \alpha I)\, u^{(k+1)} = 2\alpha u^{(k+\frac{1}{2})}\,, \tag{3.5}$$

where $r^{(k)} = c - Mu^{(k)}$ is the residual at the kth step. In [4, 7] one can find a discussion of the effect of inexact solves in (3.5). Inexact solves may result in considerable savings in terms of overall solution time. In this paper, however, we do not consider the effect of inexact solves.

[2] In [4] the coefficient matrix M was allowed to have complex entries, and the corresponding splitting was $M = H + S$, where H was the Hermitian part of M and S the skew-Hermitian part. In this paper we consider only matrices with real entries, but we retain the name HSS.

The convergence of (3.4), even with the "optimal" value of α (which may be difficult to estimate in practice) can be quite slow. In [7], it was proposed to use HSS as a preconditioner for Krylov subspace methods (such as GMRES or Bi-CGSTAB [40]); see also [6]. The preconditioner was applied to a variety of saddle point systems arising from the discretization of PDEs, with good results. In this paper, we are interested in using HSS exclusively as a preconditioner. It was observed in [7] that while it is in many cases possible to find efficient solvers for the first of the systems (3.5), solving the second system is usually less straightforward. For the problems studied in this paper, the matrix $H + \alpha I$ is diagonal; therefore the first system in (3.5) can be solved at a cost of just $O(n)$ operations. The second system is of the form

$$\begin{bmatrix} \alpha I & K \\ -K^T & \alpha I \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} = \begin{bmatrix} g \\ h \end{bmatrix}. \tag{3.6}$$

System (3.6) is equivalent to the reduced system

$$(K^T K + \alpha^2 I)\, w = \alpha h + K^T g\,. \tag{3.7}$$

Once (3.7) has been solved for w, v can be obtained from $\alpha v = g - Kw$. If K is a Toeplitz matrix, computing v from w can be done in $O(n \log n)$ operations using FFTs. Hence, the problem is reduced to solving linear systems of the type (3.7). But this is precisely the normal equation formulation of a standard regularized least squares problem for a Toeplitz (or Toeplitz-related) matrix K, and as we mentioned in the Introduction, there are efficient methods for solving (3.7). In particular, for certain types of problems that are important in image processing applications there are $O(n \log n)$ methods for solving (3.7). In this case, the cost of applying the HSS preconditioner is $O(n \log n)$ per iteration. See also the discussion at the end of section 6.

We further note that the scalar shift α^2 plays the role of a regularization parameter in the reduced system (3.7). This suggests that perhaps a sensible choice for α could be $\alpha \approx \sqrt{\mu}$, where μ is the regularization parameter for the weighted least squares problem. We will see in the section on numerical experiments that this is a good choice in some cases, but not always.

4. Spectral properties of the HSS preconditioner. The purpose of this section is to derive some bounds on the eigenvalues of the preconditioned matrix $P^{-1}M$ (or MP^{-1}, which has the same spectrum), where M is the nonsymmetric augmented matrix (3.1) and P is the HSS preconditioner (3.3). Although one should be careful not to infer too much from the eigenvalue distribution alone when the preconditioned matrix is nonnormal (as is the case here), our experience is that for the problems considered in this paper the distribution of the eigenvalues is strongly correlated with the rate of convergence of GMRES. In particular, convergence will be fast if most of the eigenvalues are enclosed in a small rectangular region of the complex plane well separated from the origin.

Let $T = I - P^{-1}M$ be the iteration matrix in the stationary iteration (3.4) associated with the HSS splitting. Then it can be shown [4, 7] that $\rho(T) < 1$ (where $\rho(T)$ denotes the spectral radius of T) for all $\alpha > 0$, and therefore the iteration (3.4) is unconditionally convergent. This implies that the eigenvalues of $P^{-1}M$ are contained in the disk of radius 1 centered at $(1, 0)$. In particular, the preconditioned matrix $P^{-1}M$ is *positive stable*: all its eigenvalues have positive real part, a desirable

property for Krylov subspace methods. In addition, if H is positive definite with extreme eigenvalues $\lambda_1 = \lambda_{\max}(H)$, $\lambda_{m+n} = \lambda_{\min}(H)$, then using $\alpha = \sqrt{\lambda_1 \lambda_{m+n}}$ minimizes an upper bound on $\rho(T)$; see [4]. When H is a diagonal matrix, as in (3.1), the "optimal" α is of course explicitly available. However, this is usually not the best choice when P is used as a preconditioner for a Krylov subspace method [6].

More refined bounds on the eigenvalues of $P^{-1}M$ can be found in [36] under the assumption that W is a symmetric positive definite matrix (not necessarily diagonal) and $\mu = 0$. In our setting, this would correspond to a weighted least squares problem without Tikhonov regularization. The results in [36] show that for small α the spectrum of $P^{-1}M$ is real and is entirely contained in two small intervals, one to the right of 0 and the other to the left of 2. The first interval contains n eigenvalues, which tend to 0 as $\alpha \to 0$, and the second interval contains the remaining m eigenvalues, which tend to 2 as $\alpha \to 0$. Here $m \geq n$. In practice this means that α should be taken small but not too small. In the case $\mu \neq 0$, however, things appear to be more complicated, and a thorough analysis has so far eluded our efforts.

It is straightforward to see that the iteration matrix $T = I - P^{-1}M$ is similar to

$$\hat{T} = (\alpha I - H)(\alpha I + H)^{-1}(\alpha I - S)(\alpha I + S)^{-1},$$

and therefore T and $\hat{T}$ have the same eigenvalues. Note that $(\alpha I - H)(\alpha I + H)^{-1}$ is symmetric and $(\alpha I - S)(\alpha I + S)^{-1}$, being the Cayley transform of a skew-symmetric matrix, is orthogonal [19, p. 440]. When $W = \mathrm{diag}\,(w_1, w_2, \ldots, w_m)$ we find

$$(\alpha I - H)(\alpha I + H)^{-1} = \begin{bmatrix} E & O \\ O & \omega I \end{bmatrix}, \quad E = (\alpha I - W)(\alpha I + W)^{-1}, \quad \omega = \frac{\alpha - \mu}{\alpha + \mu},$$

and therefore, letting $\Sigma = \alpha I + \frac{1}{\alpha} K^T K$,

$$\hat{T} = \begin{bmatrix} E & O \\ O & \omega I \end{bmatrix} \begin{bmatrix} \alpha I & -K \\ K^T & \alpha I \end{bmatrix} \begin{bmatrix} \frac{1}{\alpha} I - \frac{1}{\alpha^2} K \Sigma^{-1} K^T & -\frac{1}{\alpha} K \Sigma^{-1} \\ \frac{1}{\alpha} \Sigma^{-1} K^T & \Sigma^{-1} \end{bmatrix}.$$

We already know that the eigenvalues of $\hat{T}$ are bounded above by 2. Lower bounds, on the other hand, are more difficult to come by. However, the particular choice $\alpha = \mu$ (which implies $\omega = 0$) leads to a significant simplification:

$$\hat{T} = \begin{bmatrix} E - \frac{2}{\mu} E K \Sigma^{-1} K^T & -2 E K \Sigma^{-1} \\ O & O \end{bmatrix}.$$

In this case, the spectrum of $\hat{T}$ consists of the spectrum of the m-by-m matrix $E - \frac{2}{\mu} E K \Sigma^{-1} K^T$, plus the eigenvalue 0 with multiplicity n. It follows that the preconditioned matrix $P^{-1}M$, which is similar to $I - \hat{T}$, has m eigenvalues equal to 1, while the remaining n eigenvalues are of the form $1 - \nu_i$ where $\nu_1, \ldots \nu_m$ are the eigenvalues of the matrix

$$E - \frac{2}{\mu} E K \Sigma^{-1} K^T = E \left(I - \frac{2}{\mu} K \Sigma^{-1} K^T \right) = EF,$$

where we have let $F = I - \frac{2}{\mu} K \Sigma^{-1} K^T$. Note that both E and F are symmetric. The diagonal matrix E is negative definite, provided that $0 < \mu < w_{\min} \equiv \min_i \{w_i\}$, and it has spectral norm $||E||_2 < 1$. The eigenvalues of F can be easily expressed in terms

of the singular values σ_i of K as follows:

$$\lambda_i(F) = 1 - 2\left(\frac{\sigma_i^2}{\mu^2+\sigma_i^2}\right) = \frac{\mu^2-\sigma_i^2}{\mu^2+\sigma_i^2}\,,$$

and therefore

$$||F||_2 = \max_i \frac{|\mu^2-\sigma_i^2|}{|\mu^2+\sigma_i^2|} \le 1\,.$$

If K has full column rank and $0 < \mu \le \sigma_{\min}(K)$, then F is negative semidefinite. Therefore if $\mu < \min\{\min_i\{w_i\}, \sigma_{\min}(K)\}$, then the eigenvalues of EF are all real and between 0 and 1, and so are those of $P^{-1}M$. Note, however, that the condition $0 < \mu \le \sigma_{\min}(K)$ is unrealistic in the case of discrete ill-posed problems, where typically $\sigma_{\max}(K) \approx 1$ and $\sigma_{\min}(K) \approx 0$. Therefore, F is typically indefinite. In this case, provided that E is negative definite, the spectrum of the matrix EF is real and contained in the interval $(-1, 1)$. Clearly, the eigenvalues of $P^{-1}M$ are bounded from below by $1 - ||E||_2$. To get a lower bound for this expression, recall that

$$||E||_2 = \max_{1\le i\le m} \frac{|\mu - w_i|}{|\mu + w_i|} = \max_{1\le i\le m} \frac{1-\mu x_i}{1+\mu x_i}\,,$$

where we have set $x_i = 1/w_i$. The function

$$\phi(x) = \frac{1-\mu x}{1+\mu x}$$

is monotonically decreasing; hence it is maximized when $x = x_* = 1/w_{\max}$, where $w_{\max}$ denotes the largest entry in W. Moreover, $\phi(x_*) = \frac{2\mu}{\mu+w_{\max}}$. Hence, we have established the following lower bound:

$$\lambda(P^{-1}M) \ge \frac{2\mu}{\mu + w_{\max}}\,.$$

On the other hand, if $w_{\min} < \mu \le w_{\max}$, then both E and F are indefinite and $P^{-1}M$ will have eigenvalues with nonzero imaginary part, in general. In this case, the imaginary part of the eigenvalues is between -1 and 1, and the real part is between $\frac{2\mu}{\mu+w_{\max}}$ and 2. Hence, we have established the following result.

THEOREM 4.1. *Let $w_{\min}$, $w_{\max}$ denote the smallest and largest entries, respectively, of the diagonal matrix W, with $w_{\min} > 0$. Let M be as in* (3.2)*, and let P denote the corresponding HSS preconditioner with $\alpha = \mu$. Also, let $a := 2\mu/(\mu+w_{\max})$. Then $P^{-1}M$ has m eigenvalues equal to* 1*, and the remaining $n-m$ eigenvalues are contained in the region*

$$\mathcal{R} = \{x+iy \in \mathbb{C}\,;\, a \le x < 2,\ \ -1 < y < 1\} \cap D(1,1),$$

where $D(1,1) = \{z \in \mathbb{C}\,;\, |z-1| < 1\}$. If, moreover, the regularization parameter μ satisfies $\mu < w_{\min}$, then the eigenvalues of $P^{-1}M$ are all real and contained in $\{1\} \cup [a, 2)$.

Note that the inclusion region, $\mathcal{R}$, is independent of K. The theorem implies that no matter how ill-conditioned K is, the eigenvalues of $P^{-1}M$ remain bounded away from zero, and they will all be real as long as μ is sufficiently small. Furthermore, for a fixed value of $w_{\max}$ the real part of the eigenvalues of $P^{-1}M$ not equal to 1

increases as μ grows, and all the eigenvalues not equal to 1 tend to the real value $\lambda = 2$ as $\mu \to \infty$. Note, however, that the preconditioned matrix $P^{-1}M$ is generally nonnormal. Hence, the condition number of the eigenvector matrix may also play a role in the convergence of a Krylov subspace solver like GMRES; see [35, pp. 206–207]. Since the eigenvectors generally depend on K, so does the rate of convergence. Bounds on the location of the eigenvalues like the one given in the foregoing theorem may be useless if the eigenvector matrix of $P^{-1}M$ is highly ill-conditioned. Unfortunately, bounds on the condition number of the eigenvector matrix are even harder to come by than eigenvalue bounds. We have numerically estimated the condition number of the eigenvector matrix in a few cases of interest; see the results in section 6.

In practice, μ is typically small and the largest entry in W can be rather large; hence the lower bound $a := 2\mu/(\mu + w_{\max})$ can be tiny. Numerical experiments indicate that this lower bound can be orders of magnitude smaller than the actual minimum of the real part of the eigenvalues of $P^{-1}M$. Furthermore, as we will see, taking $\alpha = \mu$ does not lead to very good performance: better results are obtained with a different value of α. Unfortunately, it is difficult to give meaningful bounds on the eigenvalues of $P^{-1}M$ when $\alpha \neq \mu$.

5. Nonlinear image restoration. In this section we discuss nonlinear image restoration problems, an important class of applications leading to weighted Toeplitz least squares problems.

In the literature on image restoration, a blurred image is often modeled as the linear convolution of an original image with the *point spread function* of the blur. However, in practice, image formation systems or image sensors usually incorporate a built-in nonlinearity. For instance, the nonlinearity is introduced in the transformation of light intensity to the output units of the imaging system such as current intensity in photo-electric systems and photographic films. The modeling of sensor nonlinearities was first studied by Andrews and Hunt [2]. In matrix-vector notation, the general space-invariant imaging system with additive noise can be represented by the following nonlinear equation:

$$f = s(Kx) + \eta, \tag{5.1}$$

where f, x, and η represent the observed, the original image, and the noise vectors, respectively. Here $s(\cdot)$ denotes a point nonlinearity, and the matrix K is a blurring matrix. Because of the blurring process, the boundary values of f are not completely determined by the original image x inside the scene. They are also affected by the values of x outside the scene. We remark that K is block Toeplitz with Toeplitz blocks (BTTB) when zero boundary conditions are applied, and block Toeplitz-plus-Hankel with Toeplitz-plus-Hankel blocks (BTHTHB) when reflective boundary conditions are used [29]. Both theoretical and experimental results in [29, 30] show that the restoration results using reflective boundary conditions are better than those using zero boundary conditions.

Different nonlinear image restoration algorithms have been proposed and analyzed. For instance, Andrews and Hunt [2] proposed using Taylor series expansion about the mean value of the observed image to approximate (5.1) by a linear equation. An approximate filter for linear image restoration can then be derived. Trussell and Hunt [39] applied the maximum a posteriori probability (MAP) estimation scheme in nonlinear image restoration algorithms. This approach results in an iterative solution algorithm the computational complexity of which is very large. Tekalp and

Pavlović [37] also proposed to transform the noisy and blurred image into "the exposure domain" using the inverse of the nonlinear sensor characteristics. A linear minimum mean square error deconvolution filter was derived by using the linear convolutional model in the presence of multiplicative noise in the exposure domain. In this paper, we consider solving nonlinear least squares problems with regularization,

$$\min_x \|f - s(Kx)\|_2^2 + \mu\|x\|_2^2, \tag{5.2}$$

to restore the original image. In [41], Zervakis and Venetsanopoulos have considered using the Gauss–Newton method to solve the nonlinear least squares problem (5.2). Given an initial guess $x^{(0)}$, for $j = 0, 1, \ldots$, we solve the linear least squares problem

$$x^{(j+1)} = \arg\min_x \left\{ \|f - s(Kx^{(j)}) - D_s^{(j)} K(x - x^{(j)})\|_2^2 + \mu\|x\|_2^2 \right\} \tag{5.3}$$

until $\|x^{(j+1)} - x(\mu)\|_2$ is small enough, where $x(\mu)$ is the solution of (5.2) with regularization parameter μ. Here $D_s^{(j)}$ is a diagonal matrix with diagonal entries

$$[D_s^{(j)}]_{ii} = \frac{\partial s}{\partial \xi}\Big|_{\xi = \sum_l K_{il} x_l^{(j)}}.$$

We remark that under the practical assumption on the nonlinear function $s(\cdot)$, the diagonal entries of $D_s^{(j)}$ are always positive values; see Andrew and Hunt [2]. The least squares problem (5.3) is equivalent to

$$[\mu I + K^T (D_s^{(j)})^2 K](x^{(j+1)} - x^{(j)}) = K^T D_s^{(j)} [f - s(Kx^{(j)})] - \mu x\,,$$

i.e.,

$$[\mu I + K^T (D_s^{(j)})^2 K] x^{(j+1)} = K^T D_s^{(j)} [f - s(Kx^{(j)}) + D_s^{(j)} K x^{(j)}]. \tag{5.4}$$

In the Gauss–Newton method, it is important to choose a good initial guess. We propose the following algorithm to compute $x^{(0)}$:

(1) Solve $f = s(\hat{f})$ for $\hat{f}$ (in many practical applications, s^{-1} can be easily obtained).

(2) Choose a suitable parameter μ_0 and solve

$$(\mu_0 I + K^T K) x^{(0)} = K^T \hat{f}. \tag{5.5}$$

Numerical results in the next section show that the use of image $x^{(0)}$ as initial guess is quite effective for the nonlinear least squares problem in (5.2). We also see that in computing the initial guess and in each Gauss–Newton iteration, we are required to solve BTTB-related systems with diagonal matrix $D = I$ (cf. (5.5)) and $D = (D_s^{(j)})^2$ (cf. (5.4)), respectively.

6. Numerical examples. We first test the two preconditioners on a sequence of weighted Toeplitz least squares problems of the form (3.1), where $K = (k_{ij})$ is an n-by-n full Toeplitz matrix defined by $k_{ij} = 1/(\sqrt{|i-j|}+1)$, D is an n-by-n positive diagonal random matrix with $\kappa_2(D) \approx 10^3$ (so that $\kappa_2(W) \approx 10^6$), and $\mu = 10^{-3}$. Note that because D is randomly generated, this set of test problems should not

TABLE 6.1
Comparison of preconditioners, 1D problem with well-conditioned K, $\mu = 0.001$.

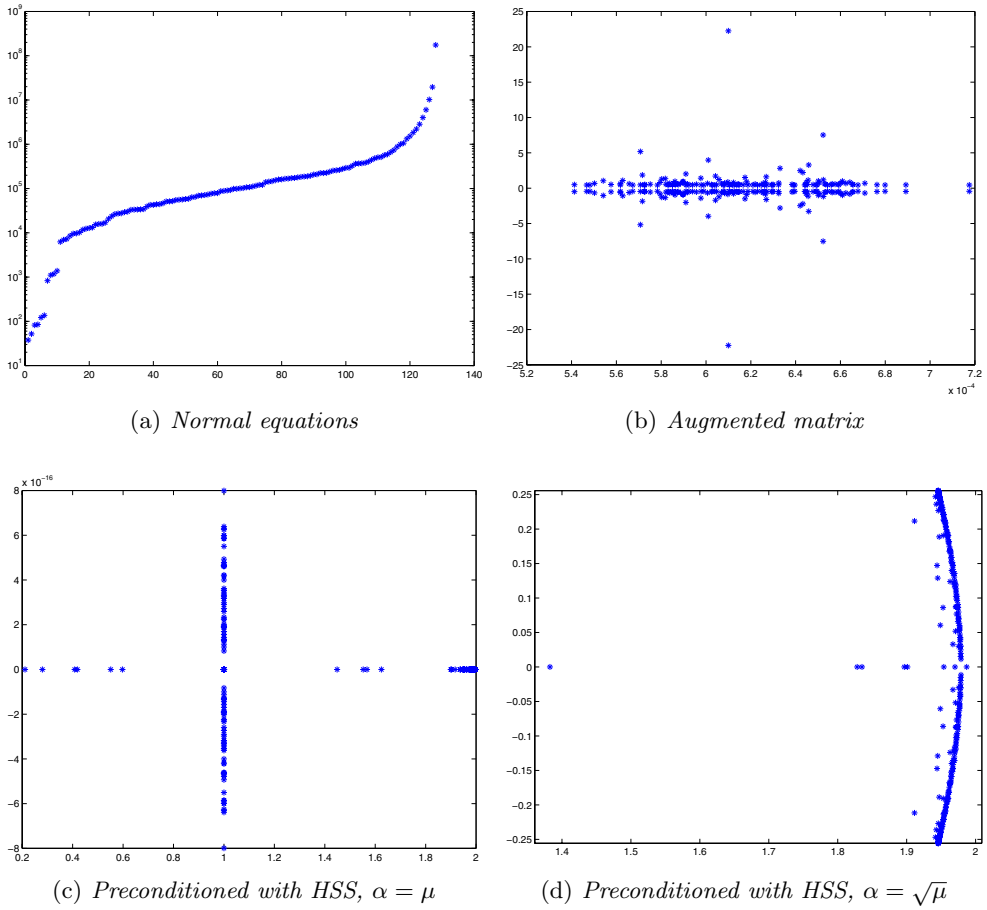
n	CG	GMRES	HSS($\alpha = \mu$)	HSS($\alpha = 0.05$)	HSS($\alpha = \sqrt{\mu}$)	CP
64	159	48	13	7	6	3
128	424	66	13	7	7	3
256	> 1000	90	18	7	7	3
512	> 1000	132	57	16	17	3
1024	> 1000	168	72	14	16	3

be thought of as a sequence of increasingly finer discretizations of one and the same underlying continuous problem. Nevertheless, all the generated problems share similar characteristics. Because of the element of randomness in D, we repeated this set of test runs several times. The iteration counts were found to be fairly stable. Here we report average iteration numbers over five test runs. The initial guess was always the zero vector, and the stopping criterion was a reduction of the initial residual norm by seven orders of magnitude.

Iteration counts for various methods are shown in Table 6.1. Here we have denoted by "CG" the conjugate gradient method on the normal equations (1.4) without preconditioning, by "GMRES" the GMRES method applied to the augmented system (3.1) without preconditioning, and by "HSS(α)" the GMRES method with HSS preconditioning with parameter α applied to the augmented system. Finally, under "CP" we report results for the (modified) constraint preconditioner applied to the augmented system (1.3). As the results show, the number of CG iterations (without preconditioning) grows very fast as the size of the least squares problem goes from $n = 64$ to $n = 1024$. The number of GMRES iterations grows much more slowly, but it is still prohibitive (note that this is full GMRES; no restart was used). When HSS preconditioning is used, the convergence is generally faster. Using $\alpha = \mu$, however, is not satisfactory, since the number of iterations still grows fairly quickly with n, and of course each iteration is now more expensive due to the additional $O(n \log n)$ operations per iteration required by the preconditioner solve. Much better results are obtained with $\alpha = 0.05$. In the last column we show the iteration counts for $\alpha = \sqrt{\mu} \approx 0.0316$, a value not too different from 0.05. The results are very similar to the case with $\alpha = 0.05$, showing that the performance of the preconditioner is not overly sensitive to the value of α. A nice property of the HSS preconditioner is that for the problems considered in this paper, the same value of α works well for all values of n. Note also that the number of iterations appears to be leveling off as n grows. Nevertheless, the results in the last column show that the modified constraint preconditioner works best, with convergence taking place in three iterations regardless of problem size.

In Figure 6.1 we show plots of the eigenvalues of the following matrices: the normal equation matrix $K^T D^2 K + \mu I$, the augmented matrix M in (3.1), the matrix M preconditioned by the HSS preconditioner with $\alpha = \mu$, and M preconditioned by the HSS preconditioner with $\alpha = \sqrt{\mu}$. The eigenvalue plot for the case $\alpha = 0.05$ is qualitatively similar to that for $\alpha = \sqrt{\mu}$. Here $n = m = 128$.

Note the logarithmic scale on the y-axis in Figure 6.1(a), and the 10^{-4} factor multiplying the x-axis in Figure 6.1(b). Also note that the tiny nonzero imaginary part of the eigenvalues in Figure 6.1(c) is due to round-off: the eigenvalues are actually all real, and in fact in this example the conditions for a real spectrum given in Theorem 4.1 are satisfied. In this problem the matrix K is fairly well conditioned, and most of the ill-conditioning is due to the weighting matrix D. We have also observed



(a) *Normal equations* (b) *Augmented matrix*

(c) *Preconditioned with HSS,* $\alpha = \mu$ (d) *Preconditioned with HSS,* $\alpha = \sqrt{\mu}$

FIG. 6.1. *Eigenvalue plots,* 1*D problem with well-conditioned* K.

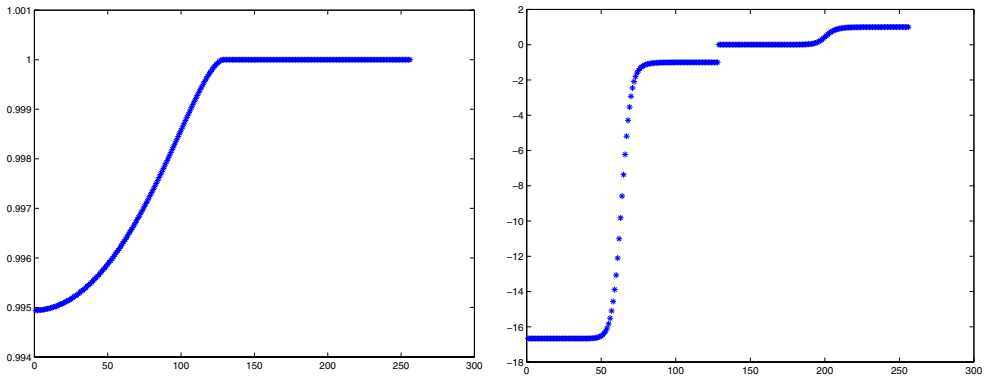
that the preconditioned matrix appears to be diagonalizable with a well-conditioned matrix of eigenvectors, a typical value of the condition number being around 60.

In Figure 6.2(a) we show the spectrum of the augmented matrix preconditioned with the modified constraint preconditioner: note the tight clustering of the eigenvalues near 1.

The second set of experiments is similar to the first one, but now K has entries

$$k_{ij} = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{-|i-j|^2}{(2\sigma^2)}\right].$$

Choosing $\sigma = 2$ results in a highly ill-conditioned Toeplitz matrix with rapidly decaying singular values. The results are reported in Table 6.2. The value $\alpha = 6 \cdot 10^{-5}$ was found to be the optimal value of α for $n = 64$, and we used the same value of α for all values of n. Under "HSS($\alpha = \alpha_{best}$)" we report the best result we could obtain by fine-tuning α. In all cases the value of α_{best} was quite close to $6 \cdot 10^{-5}$, and the corresponding iteration counts not very far apart. This shows, again, that a good value of α can be found on a small problem and used for larger problems of the same type.



(a) *Problem with well-conditioned K* (b) *Problem with ill-conditioned K*

FIG. 6.2. *Eigenvalue plots, modified constraint preconditioner.*

TABLE 6.2
Comparison of preconditioners, 1D problem with ill-conditioned K, $\mu = 0.001$.

n	CG	GMRES	HSS($\alpha = \mu$)	HSS($\alpha = 6 \cdot 10^{-5}$)	HSS($\alpha = \alpha_{best}$)	CP
64	761	117	55	43	43	37
128	> 1000	224	106	74	74	67
256	> 1000	410	159	95	84	125
512	> 1000	770	236	127	117	271
1024	> 1000	> 1000	250	129	117	553

We further observe that this problem is much harder than the previous one, and that using HSS with a good value of α results in much faster convergence than using the modified constraint preconditioner. Also note that, again, the number of iterations with the HSS preconditioners appears to level off as n increases. Eigenvalue plots for this problem are given in Figure 6.2(b) for the modified constraint preconditioner and in Figure 6.3 for the remaining cases. Note the presence of tight clusters of tiny eigenvalues in all cases, except for HSS preconditioning with the optimal α (Figure 6.3(d)), for which there is a visible (albeit small) gap separating the smallest eigenvalues from zero. The condition number of the eigenvectors of the preconditioned matrix is now $O(10^4)$ or less. While larger than in the previous example, this value can still be considered moderate.

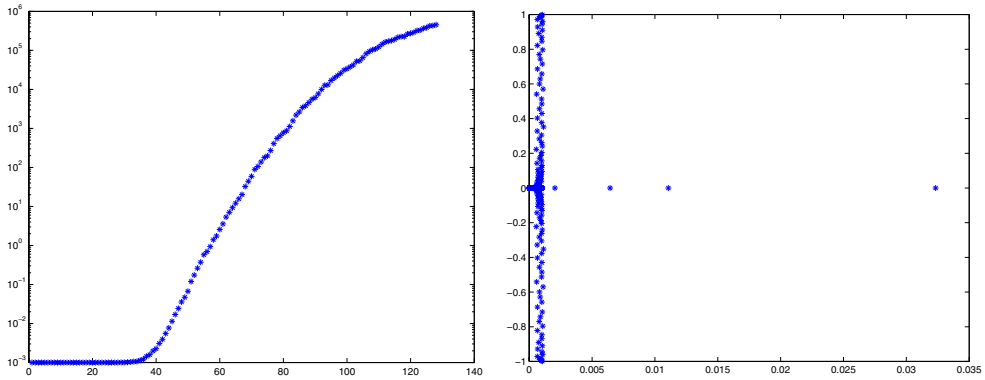
Next, we consider the solution of nonlinear image restoration problems. We test two 128×128 images: Bridge (Figure 6.4(a)) and Cameraman (Figure 6.4(b)). The pointwise nonlinearity employed is of the logarithmic form,

$$s(x) = 30 \log(x)$$

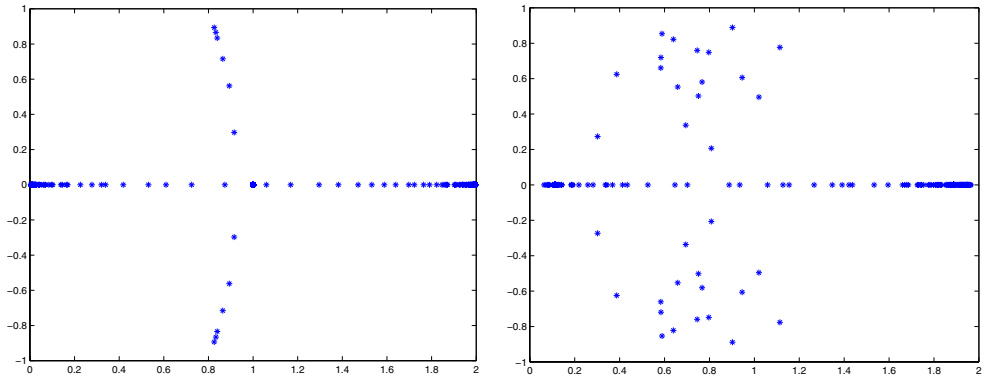
(tested in [41]), and the discrete point spread function is given by

$$k(x, y) = \exp\left[\frac{-(x^2 + y^2)}{2}\right].$$

We construct the observed image by forming the vector $f = s(Kx) + \eta$, where x is a vector formed by row ordering the original image. Here the reflective boundary condition is employed, and therefore the matrix K is BTHTHB. In our tests, the



(a) *Normal equations* (b) *Augmented matrix*



(c) *Preconditioned with HSS,* $\alpha = \mu$ (d) *Preconditioned with HSS,* $\alpha = \alpha_{best}$

FIG. 6.3. *Eigenvalue plots,* $1D$ *problem with ill-conditioned* K.

noise η is set to Gaussian white noise with noise-to-signal ratios of 40dB and 30dB. Observed images for noise-to-signal ratio of 40dB are shown in Figures 6.4(c) and (d).

In the preconditioned method for solving the system (5.5), we use the zero vector as the initial guess, and the stopping criterion is $\|r^{(i)}\|_2/\|K^T\hat{g}\|_2 < 10^{-7}$, where $r^{(i)}$ is the residual after i iterations.

In the preconditioned GMRES(30) methods for solving the augmented system corresponding to (5.4), we use the solution of (5.5) as the initial guess, and the stopping criterion is

$$\frac{\|r^{(i)}\|_2}{\|K^T D_s^{(j)}[f - s(Kx^{(j)}) + D_s^{(j)} Kx^{(j)}]\|_2} < 10^{-7},$$

where $r^{(i)}$ is the residual after the ith iteration.

In Figures 6.5(a) and (b), we present our initial guesses for the restored images, i.e., the solutions of (5.5). The optimal regularization parameter μ_0 is chosen such that it minimizes the relative error of $x^{(0)}(\mu_0)$; i.e., it minimizes

$$R_0 = \frac{\|x - x^{(0)}(\mu_0)\|_2}{\|x\|_2}.$$

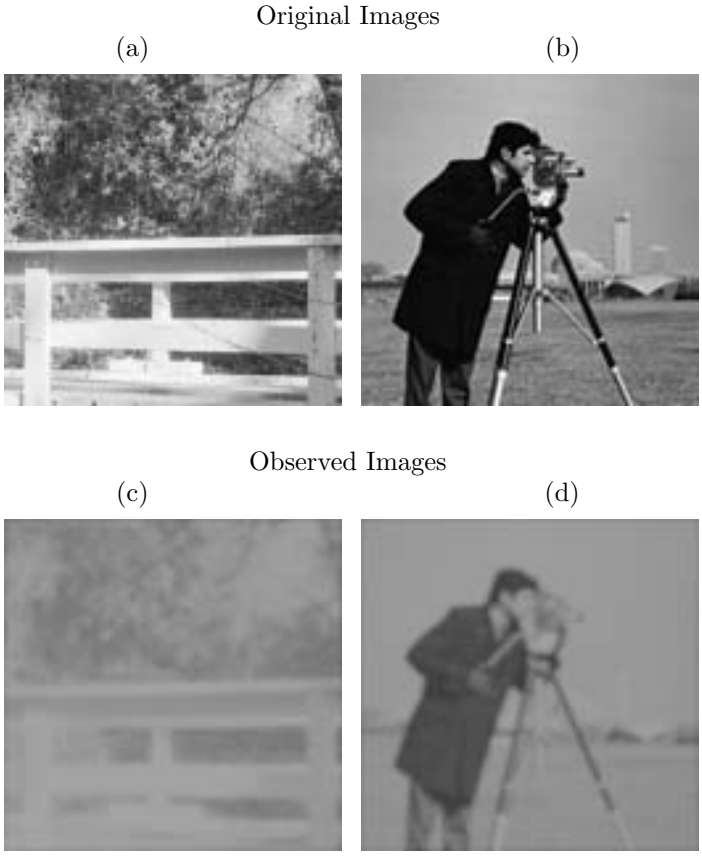


FIG. 6.4. *Original and observed images of Bridge (left) and Cameraman (right).*

TABLE 6.3
Relative errors in the restored images (the number in the bracket refers to the corresponding optimal regularization parameters).

	Bridge		Cameraman	
SNR	R_0	R_1	R_0	R_1
40dB	0.0532 (2.0×10^{-3})	0.0512 (7.5×10^{-3})	0.0548 (5.0×10^{-3})	0.0504 (3.0×10^{-4})
30dB	0.0622 (1.0×10^{-2})	0.0612 (4.0×10^{-3})	0.0824 (2.0×10^{-2})	0.0801 (1.2×10^{-3})

The restored images are shown in Figures 6.5(c) and (d). Again, the optimal regularization parameter μ is chosen such that

$$R_1 = \frac{\|x - x(\mu)\|_2}{\|x\|_2}$$

is minimized, where $x(\mu)$ is the solution of (5.4). It can be seen from the figures that the quality of restored images is visually better than that of initial guess images. The errors R_0 and R_1 are reported in Table 6.3, and their corresponding optimal regularization parameters are also given. We also found that the relative errors of the restored images computed by several iterations of the Gauss–Newton method do

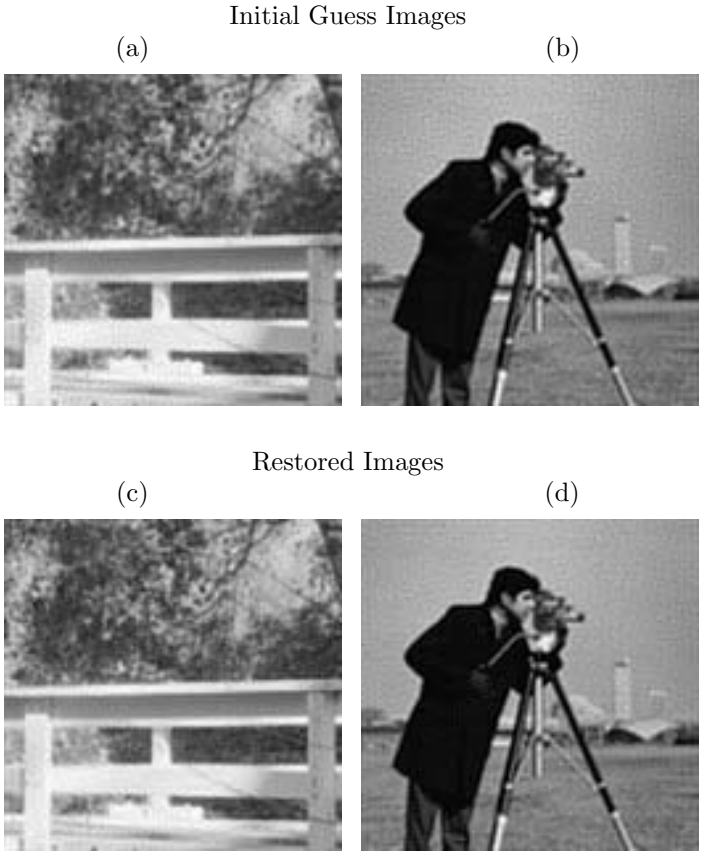


FIG. 6.5. *Initial guess and restored images of Bridge (left) and Cameraman (right).*

TABLE 6.4
Comparison of preconditioners for the nonlinear image restoration problem.

Method	Bridge		Cameraman	
	30dB	40dB	30dB	40dB
CG	289	273	334	768
Circulant-PCG	183	167	189	195
GMRES	> 1000	> 1000	> 1000	> 1000
HSS ($\alpha = 0.01$)	82	76	98	69
HSS ($\alpha = 0.05$)	48	46	51	45
HSS ($\alpha = 0.10$)	39	37	38	49
HSS ($\alpha = 0.50$)	95	87	105	270
CP	312	295	378	> 1000

not improve, and the visual quality of these restored images are about the same as in Figures 6.5(c) and (d).

The number of iterations for CG and GMRES(30) with various preconditioners for solving the augmented system corresponding to (5.4) is listed in Table 6.4. We see that HSS-preconditioned GMRES(30) converges much faster than the other methods tested. These include the unpreconditioned conjugate gradient method applied to (5.4) (under "CG"), unpreconditioned GMRES(30) applied to the corresponding

augmented system (under "GMRES"), the preconditioned conjugate gradient method with the circulant preconditioner (under "Circulant-PCG"), and GMRES(30) with the modified constraint preconditioner (under "CP"). We note that the cost per iteration of the conjugate gradient and the GMRES methods depends mainly on the Toeplitz matrix-vector multiplications. This product can be formed with two two-dimensional FFTs of size $2n$-by-$2n$, i.e., at about the same cost as four two-dimensional FFTs of size n-by-n. With circulant, HSS, and CP preconditioners, we have to solve their corresponding systems in each preconditioned CG/GMRES iteration, and thanks to their special structure, they can be solved by using two two-dimensional FFTs of size n-by-n. Thus the overhead of using different preconditioners is about 50%. Therefore the costs per iteration using different preconditioners are comparable.

7. Conclusions and future work. We have investigated the application of two types of preconditioners to the solution of weighted Toeplitz least squares problems written in augmented system form. Theoretical considerations and numerical experiments indicate that the modified constraint preconditioner is extremely effective when the problem is well conditioned or moderately ill conditioned and $m = n$. On the other hand, this preconditioner is not effective for the highly ill-conditioned cases that arise from the solution of ill-posed problems.

The HSS preconditioner was found to perform quite well (though not as well as constraint preconditioning) for moderately ill-conditioned problems. For harder problems the convergence rate, while not spectacular, was much better than with constraint preconditioning; moreover, for large enough n the rate of convergence appears to be constant. A positive observation is that the iteration parameter α in the HSS preconditioner can be fine-tuned on a small problem instance and then used, with good results, on larger problems.

Although we limited ourselves to the simplest possible form of the regularization term, the techniques used here can easily be applied to the case where the regularization term is of the form $\mu||Lx||_2^2$, where L is a smoothing operator. In particular, HSS preconditioning can be efficiently applied as long as L has a nice structure, such as Toeplitz or Toeplitz-related.

We conclude by mentioning another possible use of HSS preconditioning. In some applications (see, e.g., [5]) it is necessary to solve problems of the form

$$(K^T K + \beta D)x = K^T f\,, \tag{7.1}$$

where K is a Toeplitz-related matrix, $\beta > 0$, and D is a diagonal matrix with positive entries on the main diagonal. Clearly, problem (7.1) is equivalent to the augmented system

$$\begin{bmatrix} I & K \\ -K^T & \beta D \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} = \begin{bmatrix} f \\ 0 \end{bmatrix}, \tag{7.2}$$

for which the HSS preconditioner is well suited.

Acknowledgment. The first author would like to thank the University of Hong Kong for its kind support and warm hospitality during his stay there in January, 2004.

REFERENCES

[1] G. Ammar and W. B. Gragg, *Superfast solution of real positive definite Toeplitz systems*, SIAM J. Matrix Anal. Appl., 9 (1988), pp. 61–76.

[2] H. Andrews and B. Hunt, *Digital Image Restoration*, Prentice-Hall, Englewood Cliffs, NJ, 1977.
[3] O. Axelsson and M. Neytcheva, *Preconditioning methods for linear systems arising in constrained optimization problems*, Numer. Linear Algebra Appl., 10 (2003), pp. 3–31.
[4] Z.-Z. Bai, G. H. Golub, and M. K. Ng, *Hermitian and skew-Hermitian splitting methods for non-Hermitian positive definite linear systems*, SIAM J. Matrix Anal. Appl., 24 (2003), pp. 603–626.
[5] M. Belge, M. E. Kilmer, and E. L. Miller, *Wavelet domain image restoration with adaptive edge-preserving regularization*, IEEE Trans. Image Process., 9 (2000), pp. 597–608.
[6] M. Benzi, M. J. Gander, and G. H. Golub, *Optimization of the Hermitian and skew-Hermitian splitting iteration for saddle-point problems*, BIT, 43 (2003), pp. 881–900.
[7] M. Benzi and G. H. Golub, *A preconditioner for generalized saddle point problems*, SIAM J. Matrix Anal. Appl., 26 (2004), pp. 20–41.
[8] M. Benzi, G. H. Golub, and J. Liesen, *Numerical solution of saddle point problems*, Acta Numer., 14 (2005), pp. 1–137.
[9] A. Björck, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, 1996.
[10] R. H. Chan, J. G. Nagy, and R. J. Plemmons, *FFT-based preconditioners for Toeplitz-block least squares problems*, SIAM J. Numer. Anal., 30 (1993), pp. 1740–1768.
[11] R. H. Chan, J. G. Nagy, and R. J. Plemmons, *Circulant preconditioned Toeplitz least squares iterations*, SIAM J. Matrix Anal. Appl., 15 (1994), pp. 80–97.
[12] R. H. Chan, J. G. Nagy, and R. J. Plemmons, *Displacement preconditioner for Toeplitz least squares iterations*, Electron. Trans. Numer. Anal., 2 (1994), pp. 44–56.
[13] R. H. Chan, J. G. Nagy, and R. J. Plemmons, *Generalization of Strang's preconditioner with applications to Toeplitz least squares problems*, Numer. Linear Algebra Appl., 3 (1996), pp. 45–64.
[14] P. J. Davis, *Circulant Matrices*, John Wiley and Sons, New York, 1979.
[15] N. Dyn and W. E. Ferguson, Jr., *The numerical solution of equality-constrained quadratic programming problems*, Math. Comp., 41 (1983), pp. 165–170.
[16] R. E. Ewing, R. Lazarov, P. Lu, and P. Vassilevski, *Preconditioning indefinite systems arising from mixed finite element discretization of second order elliptic problems*, in Lectures Notes in Math. 1457, O. Axelsson and L. Yu. Kolotilina, eds., Springer-Verlag, Berlin, 1990, pp. 28–43.
[17] J. A. Fessler and S. D. Booth, *Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction*, IEEE Trans. Image Process., 8 (1999), pp. 688–699.
[18] N. I. M. Gould, M. E. Hribar, and J. Nocedal, *On the solution of equality constrained quadratic programming problems arising in optimization*, SIAM J. Sci. Comput., 23 (2001), pp. 1376–1395.
[19] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1994.
[20] X.-Q. Jin, *A preconditioner for constrained and weighted least squares problems with Toeplitz structure*, BIT, 36 (1996), pp. 101–109.
[21] T. Kailath, S. Y. Kung, and M. Morf, *Displacement ranks of a matrix*, Bull. Amer. Math. Soc., 1 (1979), pp. 769–773.
[22] T. Kailath and A. H. Sayed, *Displacement structure: Theory and applications*, SIAM Rev., 37 (1995), pp. 297–386.
[23] C. Keller, N. I. M. Gould, and A. J. Wathen, *Constraint preconditioning for indefinite linear systems*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1300–1317.
[24] M. Kim, H. Park, and L. Eldén, *Comparison of algorithms for Toeplitz least squares and symmetric positive definite linear systems*, in Structured Matrices in Mathematics, Computer Science, and Engineering, Vol. II, Contemporary Mathematics, V. Olshevsky, ed., American Mathematical Society, Providence, RI, 2001, pp. 73–99.
[25] N. Levinson, *The Wiener r.m.s. (root-mean-square) error criterion in filter design and prediction*, J. Math. Phys., 25 (1947), pp. 261–278.
[26] F.-R. Lin, M. K. Ng, and W.-K. Ching, *Factorized banded inverse preconditioners for matrices with Toeplitz structure*, SIAM J. Sci. Comput., 26 (2005), pp. 1852–1870.
[27] L. Lukšan and J. Vlček, *Indefinitely preconditioned inexact Newton method for large sparse equality constrained non-linear programming problems*, Numer. Linear Algebra Appl., 5 (1998), pp. 219–247.
[28] M. K. Ng, *Iterative Methods for Toeplitz Systems*, Numerical Mathematics and Scientific Computation, Oxford University Press, New York, 2004.
[29] 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.

[30] M. Ng and A. Yip, *A fast MAP algorithm for high-resolution image reconstruction with multisensors*, Multidimens. Systems Signal Process., 12 (2001), pp. 143–164.
[31] H. Park and L. Eldén, *Stability analysis and fast algorithms for triangularization of Toeplitz matrices*, Numer. Math., 76 (1997), pp. 383–400.
[32] I. Perugia and V. Simoncini, *Block-diagonal and indefinite symmetric preconditioners for mixed finite element formulations*, Numer. Linear Algebra Appl., 7 (2000), pp. 585–616.
[33] M. Rozložník and V. Simoncini, *Krylov subspace methods for saddle point problems with indefinite preconditioning*, SIAM J. Matrix Anal. Appl., 24 (2002), pp. 368–391.
[34] Y. Saad and M. H. Schultz, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 856–869.
[35] Y. Saad, *Iterative Methods for Sparse Linear Systems*, 2nd ed., SIAM, Philadelphia, PA, 2003.
[36] V. Simoncini and M. Benzi, *Spectral properties of the Hermitian and skew-Hermitian splitting preconditioner for saddle point problems*, SIAM J. Matrix Anal. Appl., 26 (2004), pp. 377–389.
[37] A. Tekalp and G. Pavlović, *Restoration of scanned photographic images*, in Digital Image Restoration, A. Katsaggelos, ed., Springer-Verlag, Berlin, 1991, pp. 209–240.
[38] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*, John Wiley and Sons, New York, 1977.
[39] H. Trussell and B. Hunt, *Improved methods of maximum a posteriori restoration*, IEEE Trans. Comput., 27 (1979), pp. 57–62.
[40] H. A. van der Vorst, *Bi-CGSTAB: A fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 13 (1992), pp. 631–644.
[41] M. Zervakis and A. Venetsanopoulos, *Iterative algorithms with fast convergence rates in nonlinear image restoration*, in SPIE 1452, Image Processing Algorithms and Techniques, SPIE, Bellingham, WA, 1991, pp. 90–103.