SUPEROPTIMAL PRECONDITIONED CONJUGATE GRADIENT ITERATION FOR IMAGE DEBLURRING[*]

FABIO DI BENEDETTO[†], CLAUDIO ESTATICO[†], AND STEFANO SERRA CAPIZZANO[‡]

Abstract. We study the superoptimal Frobenius operators in the two-level circulant algebra. We consider two specific viewpoints: (1) the regularizing properties in imaging and (2) the computational effort in connection with the preconditioned conjugate gradient method. Some numerical experiments illustrating the effectiveness of the proposed technique are given and discussed.

Key words. two-level Toeplitz and circulant matrices, preconditioning, regularization

AMS subject classifications. 65F10, 15A30, 15A60, 47B48

DOI. 10.1137/S1064827503421653

1. Introduction. Many image restoration problems lead to square linear systems having the Toeplitz structure: this happens when the blurring process is space invariant [1, 19]. In the case of two-dimensional (2-D) images, the discretized problem is well described by means of two-level partitioning of the matrix involved.

The most compact notation makes use of multi-indices: in our case $n = (n_1, n_2)$ represents the image size. By identifying images with vectors of $N(n) = n_1 \cdot n_2$ pixels, the blurring operator is a matrix $A_n \in \mathbb{R}^{N(n) \times N(n)}$; we assume that A_n comes from a sequence of two-level Toeplitz matrices generated [17] by a Lebesgue-integrable bivariate function $f(x^1, x^2)$. This means that the entries of $A_n = T_n(f)$ along the kth diagonal are given by the kth Fourier coefficient a_k of $f : Q^2 \to \mathbb{R}$, $Q = [-\pi, \pi]$ (f is even and real-valued because each A_n is assumed to be symmetric). In the case where f is a trigonometric polynomial, $\{A_n\}$ is a sequence of block-banded matrices with banded blocks: we emphasize that the banded case occurs when dealing with the out-of-focus blurring [24] or with the blurring related to linear motion [1].

In recent years some attention has been paid to the analysis of the superoptimal preconditioner proposed by Tyrtyshnikov in [30]. The main result, quite evident from the experimental point of view, is a negative one. Indeed, in [10] it has been analytically shown that the superoptimal approximation of Toeplitz structures is a poor approximation when the Toeplitz matrix is ill-conditioned and is not competitive with other strategies proposed in the literature, which are easier to define and more efficient to construct. More precisely, in the frequency spaces where the eigenvalues of $T_n(f)$ collapse to zero we observe that the corresponding eigenvalues of the superoptimal approximation are well separated from zero, and therefore there is a consequent slowdown of the PCG method when the linear system has to be solved exactly.

Conversely, this behavior is appreciated in the field of image restoration, and indeed the usual approximations are generally "stabilized" [19] in order to force this property that arises naturally in the superoptimal approximation. More precisely, conjugate gradient–like iterations are widely used to obtain a *regularized* solution

[†]Dipartimento di Matematica, Via Dodecaneso 35, 16146 Genova, Italy (dibenede@dima.unige.it, estatico@dima.unige.it).

[‡]Dipartimento di Chimica, Fisica e Matematica, Via Valleggio 11, 22100 Como, Italy (stefano.serrac@uninsubria.it, serra@mail.dm.unipi.it).

(instead of an exact solution), where the iteration count plays the role of the regularization parameter [25, 2]. The regularizing power of iterations is preserved by introducing a preconditioner, provided that the action of the preconditioned matrix on the frequencies mentioned above is comparable with $T_n(f)$.

Motivated by this observation, we used again in [10] this preconditioning strategy in one-dimensional (1-D) reconstruction problems of a blurred signal corrupted by some noise. The results were very promising particularly with respect to the quality of the reconstructed signal: the obtained results were very close to the one achieved by using the best choice of the Tikhonov regularization parameter. We recall that the determination of this parameter is never an easy task. Having in mind this background, the idea in the present paper is to perform a more systematic analysis in this direction with special attention to and emphasis on the more realistic 2-D setting.

The results of our analysis are controversial: when the amount of noise is within a given level related to the ill-posedness of the problem (e.g., 0.1% with a Gaussian blur), then the superoptimal approximation gives a qualitatively good reconstruction with a small number of iterations when compared with the nonpreconditioned case. We recall that the standard preconditioning strategies (optimal, natural, etc.; see [7]) are not employable in this context since they destroy the image by putting in the solution all the noisy components. On the other hand, if the level of noise is higher, then the quality of the reconstruction deteriorates, and hence other more specialized techniques should be used: higher order filtering [12] or stabilized preconditioning [19]. In this case the superoptimal preconditioner is just the basic element of a wider family, and could be a good alternative to the (nonregularizing) optimal one.

The paper is organized as follows. In section 2 we introduce the formal definition of superoptimal operator and study the eigenvalue distribution in the 2-D case. Section 3 is devoted to the study of the computational effort for computing the superoptimal approximation and for using it in the PCG method. Finally, in section 4 we perform a wide numerical simulation, and in section 5 we draw conclusions.

2. The superoptimal operator. The superoptimal operator, introduced for the (scalar) circulant class by Tyrtyshnikov [30], forces the preconditioned matrix $P^{-1}A$ to approximate the identity in the Frobenius norm among all the elements P of a prescribed matrix algebra $\mathcal{M}(U) = \{X = U\Delta U^* : \Delta \text{ is diagonal}\}$, diagonalized by a unitary transform U. We recall that the Frobenius norm $\|\cdot\|_{\mathrm{F}}$ of a matrix X is defined as $\|X\|_{\mathrm{F}}^2 = \sum_{i,j} |(X)_{i,j}|^2$ and coincides with the Euclidean norm of the vector containing the singular values of X.

DEFINITION 2.1. *The superoptimal operator $\hat{\mathcal{P}}_U$ maps a matrix $A \in \mathbb{C}^{N \times N}$ to a matrix $P \in \mathcal{M}(U) \subset \mathbb{C}^{N \times N}$ such that its Moore–Penrose inverse P^+ solves the minimization problem*

$$(2.1) \qquad \min_{X \in \mathcal{M}(U)} \|AX - I\|_{\mathrm{F}},$$

I being the $N \times N$ identity matrix and $\|\cdot\|$ being the Frobenius norm.

It is possible to give an explicit expression for the construction of $\hat{\mathcal{P}}_U$, involving the well-known *optimal operator* [8] $\mathcal{P}_U(A) = \arg\min_{X \in \mathcal{M}(U)} \|A - X\|_{\mathrm{F}}$ (the former expression is well defined due to the strict convexity of the Frobenius norm that implies the uniqueness of the minimizer from a given convex closed set).

THEOREM 2.2 (see [5, 10]). *Assume that A is nonsingular. Then the solution of problem* (2.1) *has the expression $P^+ = [\mathcal{P}_U(A^*A)]^{-1} (\mathcal{P}_U(A))^*$; if A is also Hermitian,*

it follows that

$$\hat{\mathcal{P}}_U(A) = [\mathcal{P}_U(A)]^{-1} \mathcal{P}_U(A^2) = \frac{\mathcal{P}_U(A^2)}{\mathcal{P}_U(A)},$$

where the fractional notation emphasizes the commutativity between $\mathcal{P}_U(A)^{-1}$ and $\mathcal{P}_U(A^2)$ inside $\mathcal{M}(U)$.

COROLLARY 2.3. *If A is Hermitian positive definite, then $\hat{\mathcal{P}}_U(A)$ is Hermitian and positive definite.*

In this paper we will treat the case where $\mathcal{M}(U)$ is the circulant algebra, associated with the Fourier matrix

$$U = F_m = \left(\frac{1}{\sqrt{m}} e^{2\pi \mathbf{i}jk/m} \right), \quad k, j = 0, \ldots, m-1,$$

in the 1-D case, and $U = F_{n_1} \otimes F_{n_2}$ in the 2-D case. We will use for simplicity the compact notation $\mathcal{P}_F$ and $\hat{\mathcal{P}}_F$ in both instances.

The goal of this section is the theoretical estimation of the smallest eigenvalues of the preconditioned matrix $P^{-1}A$, where $A = T_n(f)$ is assumed to be symmetric and positive definite (hence f is nonnegative) and $P = \hat{\mathcal{P}}_F(A)$.

Our investigation will be carried out through the following steps:

    (i) several explicit expressions of the eigenvalues of P will be given in terms of the matrix A and the function f;

    (ii) the asymptotical behavior of some "critical" eigenvalues will be investigated as n increases;

    (iii) the minimal eigenvalues of $P^{-1}A$ will be estimated.

2.1. The eigenvalues of the preconditioner. We give below some different characterizations according to the specific assumptions made on the function f. Notice that any 1-D result allows for an immediate extension to the 2-D *separable* case, that is, $f(x^1, x^2) = f_1(x^1)f_2(x^2)$. In fact, under such a special assumption both A, P and $P^{-1}A$ can be represented as tensor products, so that their eigenvalues are simply related to those of 1-D matrices.

DEFINITION 2.4. *For $x \in Q = [-\pi, \pi]$, the unitary vector $u_n(x) \in \mathbb{C}^n$ in the 1-D case is*

$$u_n(x) = \left(\frac{1}{\sqrt{n}} \exp(\mathbf{i}kx) \right)_{k=0}^{n-1};$$

in the 2-D case, for $x = (x^1, x^2) \in Q^2$ we introduce the unitary vector $u_n(x) \in \mathbb{C}^{N(n)}$, $N(n) = n_1 n_2$, such that

$$u_n(x) = u_{n_1}(x^1) \otimes u_{n_2}(x^2) = \left(\frac{1}{\sqrt{N(n)}} \exp(\mathbf{i}k \cdot x) \right)_{k=0}^{n-e},$$

where $k \cdot x$ is a scalar product and $e = (1, 1)$.

REMARK 2.1. *The columns of a 1-D Fourier matrix are obtained through suitable evaluations of $u_n(x)$ at the uniform grid $x_j = (2\pi j/n)$ on Q; in fact, $F_n = (u_n(x_0)\ u_n(x_1)\ \ldots\ u_n(x_{n-1}))$.*

The same result holds in two dimensions. There, if the grid is meant component-wise, that is,

$$x_j = (x_{j_1}^1, x_{j_2}^2) = \left(\frac{2\pi j_1}{n_1}, \frac{2\pi j_2}{n_2} \right),$$

and we assume the lexicographical ordering for the multi-indices $j = (j_1, j_2)$, then the 2-D Fourier matrix can be represented as follows:

$$F_{n_1} \otimes F_{n_2} = (u_n(x_j))_{j=0}^{n-e}.$$

Notation. We use the symbol $\#(\cdot)$ to indicate the cardinality of a set. In the following, let $\mathbb{T}_d$ denote the class of trigonometric polynomials in d variables.

Concerning the eigenvalues of the superoptimal preconditioner P, we are able to give a first result.

PROPOSITION 2.5. *The following four relations hold:*

1. *Let $C_n(x) = \|Au_n(x)\|_2^2$, $D_n(x) = u_n(x)^* Au_n(x)$; then the eigenvalues of $P = \hat{\mathcal{P}}_F(A)$ are given by*

$$\lambda_j(P) = \frac{C_n(x_j)}{D_n(x_j)},$$

where $x_j = 2\pi j/n$ (also in the 2-D case).

2. *If $A = T_n(f)$ with $f \in \mathbb{T}_1$ (trigonometric polynomials in one variable), then*

$$C_n(x) = f(x)^2 - \frac{2}{n} s_2(f; x), \quad D_n(x) = f(x) - \frac{2}{n} s_1(f; x),$$

where $s_1, s_2 \in \mathbb{T}_1$ do not depend on n.

3. *$D_n(x)$ is a Césaro sum [32] related to f.*

4. *Let $g_r(x) := \sum_{k \geq 0} a_{r-k} \exp(ik \cdot x)$ for $r \geq 0$, in either the 1-D or the 2-D case. Then*

$$\frac{C_n(x)}{D_n(x)} \geq \frac{\sum_{r \geq 0} |g_r(x)|^2}{\sum_{r \geq 0} |g_r(x)|}.$$

Proof. 1. From the explicit expression provided by Theorem 2.2, we have

$$\lambda_j(P) = \frac{\lambda_j(\mathcal{P}_F(A^2))}{\lambda_j(\mathcal{P}_F(A))},$$

where all the eigenvalues are labeled according to the column ordering of the eigenvector matrix F. Since the corresponding eigenvector is a unitary vector like $u_n(x_j)$ by Remark 2.1, the two terms of the fraction can be represented as Rayleigh quotients:

$$\lambda_j(\mathcal{P}_F(A^2)) = u_n(x_j)^* A^2 u_n(x_j) = \|Au_n(x_j)\|_2^2 = C_n(x_j),$$
$$\lambda_j(\mathcal{P}_F(A)) = u_n(x_j)^* Au_n(x_j) = D_n(x_j).$$

2. This part has been proved in [10, Theorem 5.1], where the explicit expression of s_1 and s_2 can be found. In that reference some additional assumptions are made, but just $f \in \mathbb{T}_1$ is essential.

3. This fact is well known (see, e.g., [31]) and can be proved by direct inspection.

4. $g_r(x)$ represents the rth entry of the vector $Au_n(x)$, up to a scaling factor ($\sqrt{N(n)}$ in two dimensions). Hence the functions $C_n(x)$ and $D_n(x)$ introduced in part 1 can be expressed as

$$C_n(x) = \frac{1}{N(n)} \sum_{r \geq 0} |g_r(x)|^2, \quad D_n(x) = \frac{1}{N(n)} \sum_{r \geq 0} g_r(x) \exp(-irx);$$

since A is positive definite, $D_n(x) = |D_n(x)| \leq \frac{1}{N(n)} \sum_{r \geq 0} |g_r(x)|$ and the thesis follows. The same argument works in the 1-D case. □

2.2. Estimating the eigenvalues of P. The different characterizations given above are useful for bounding from below the eigenvalues of P, and, in fact, the proving technique varies according to the assumptions made on f. In the following theorems, the only general hypothesis we assume on f is that every zero $\bar{x}$ is of order at least p, that is,

$$(2.2) \qquad \exists \delta > 0 : \quad f(x) \leq c_0 \|x - \bar{x}\|_{\infty}^p \quad \forall x \in B(\bar{x}, \delta),$$

where δ is a vector and $B(\bar{x}, \delta)$ denotes the hyperrectangle containing all the x such that $|(x - \bar{x})_i| \leq \delta_i$. Moreover when δ is scalar, we mean that $B(\bar{x}, \delta)$ is the hypercube such that $|(x - \bar{x})_i| \leq \delta$. Further assumptions on f will vary according to the characterization of the eigenvalues of P we will use, given by Proposition 2.5.

THEOREM 2.6. *Let $f \in \mathbb{T}_1$. For any root $\bar{x}$ of f satisfying (2.2), there exists a positive constant ρ such that*

$$\frac{C_n(x)}{D_n(x)} \geq \rho > 0 \quad \forall x \in B(\bar{x}, n^{-1/p})$$

for n sufficiently large, where C_n and D_n are defined as in Proposition 2.5.

Proof. Consider the evaluation of C_n and D_n at the root $\bar{x}$. By definition,

$$C_n(\bar{x}) = u_n(\bar{x})^* A^2 u_n(\bar{x}) \qquad \text{and} \qquad D_n(\bar{x}) = u_n(\bar{x})^* A u_n(\bar{x}).$$

If we construct a unitary matrix $U \in \mathbb{C}^{n \times n}$ having $u_n(\bar{x})$ as its first column (this can be achieved by taking the other columns in the orthogonal complement), we know from the theory of the optimal matrix algebra operators (see [9]) that for each matrix $B \in \mathbb{C}^{n \times n}$ its optimal preconditioner $\mathcal{P}_U(B)$ has the eigenvalue $u_n(\bar{x})^* B u_n(\bar{x})$ associated with the eigenvector $u_n(\bar{x})$. Moreover, if B is symmetric positive definite, then the same holds for $\mathcal{P}_U(B)$.

Applying this result to the cases $B = A^2$ and $B = A$, we obtain the eigenvalues $C_n(\bar{x}) > 0$ and $D_n(\bar{x}) > 0$, respectively.

Now we can use the assumption $f \in \mathbb{T}_1$ and consequently the expressions for C_n and D_n given by part 2 of Proposition 2.5. Since $f(\bar{x}) = 0$, we have

$$C_n(\bar{x}) = -\frac{2}{n} s_2(f; \bar{x}), \quad D_n(\bar{x}) = -\frac{2}{n} s_1(f; \bar{x});$$

from the positivity of $C_n(\bar{x})$ and $D_n(\bar{x})$ it follows that $s_1(f; \bar{x}) < 0$ and $s_2(f; \bar{x}) < 0$. Since $s_1, s_2 \in \mathbb{T}_1$ do not depend on n, suitable constants $\eta_1, \eta_2 > 0$ and $\delta_1 < \delta$ can be found, such that

$$(2.3) \qquad -\eta_1 \leq s_1(f; x) < 0 \text{ and } s_2(f; x) \leq -\eta_2 \qquad \forall x \in B(\bar{x}, \delta_1)$$

by continuity. For those values of x

$$C_n(x) = f(x)^2 - \frac{2}{n} s_2(f; x) \geq \frac{2}{n} \eta_2,$$

$$D_n(x) = f(x) - \frac{2}{n} s_1(f; x) \leq c_0 |x - \bar{x}|^p + \frac{2}{n} \eta_1,$$

where in the last inequality we have applied (2.2). If n is sufficiently large, $n^{-1/p} < \delta_1$, and therefore

$$\frac{C_n(x)}{D_n(x)} \geq \frac{2\eta_2/n}{c_0/n + 2\eta_1/n} = \frac{2\eta_2}{c_0 + 2\eta_1} =: \rho > 0$$

for all $x \in B(\bar{x}, n^{-1/p})$.   □

We remark that if the assumption that f is a polynomial is dropped, there is no easy way to relate the signs of $C_n(\bar{x})$ and $s_2(f; \bar{x})$, $\tilde{k}_1(\bar{x})$, $\tilde{k}_2(\bar{x})$ in order to obtain the bounds (2.3) and their multivariate generalization. Hence a different tool must be used in order to extend the result to the nonpolynomial case.

THEOREM 2.7. *Let $f : \mathbb{R} \to \mathbb{R}$ be an arbitrary univariate 2π-periodic C^1 function. Then for any root $\bar{x}$ satisfying (2.2), the same conclusions as those in Theorem 2.6 hold.*

Proof. By virtue of part 3 of Proposition 2.5, we can study $D_n(x)$ by using classical error bounds involving the Césaro sum of a continuous function (see, e.g., [32]):

$$E(D_n) := \|f - D_n\|_\infty \leq c_1/n \qquad \text{for large } n.$$

It follows that $D_n(x) = f(x) + (D_n(x) - f(x)) \leq c_0|x - \bar{x}|^p + c_1/n$ for all $x \in B(\bar{x}, \delta)$.

Concerning the numerator $C_n(x)$, we need the characterization

$$C_n(x) = \sum_{r \geq 0} |g_r(x)|^2/n$$

given in the proof of Proposition 2.5, part 4. It follows that the sharp bound

$$(2.4) \qquad C_n(x) \geq \frac{1}{n}|g_0(x)|^2 \geq \frac{1}{n}[\mathrm{Re}(g_0(x))]^2$$

holds. From the expression $g_0(x) := \sum_{k=0}^{n-1} a_k \exp(\mathbf{i}kx)$ we obtain

$$\mathrm{Re}(g_0(x)) = \sum_{k=0}^{n-1} a_k \cos kx = \frac{1}{2}(a_0 + f_n(x)),$$

where $f_n(x)$ is the nth partial Fourier sum of $f(x)$. Hence

$$|\mathrm{Re}(g_0(x))| \geq \frac{1}{2}(a_0 - f(x) - |f(x) - f_n(x)|)$$

because $a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\,dx > 0$. Substituting into (2.4) we have

$$C_n(x) \geq \frac{1}{4n}[a_0 - f(x) - E(f_n)]^2 \geq \frac{1}{4n}[a_0^2 - 2a_0(f(x) + E(f_n))],$$

where $E(f_n) = \|f - f_n\|_\infty$. From the bounds on $C_n(x)$ and $D_n(x)$ and from (2.2) we obtain, for large n and $x \in B(\bar{x}, n^{-1/p})$,

$$\frac{C_n(x)}{D_n(x)} \geq \frac{a_0^2 - 2a_0(c_0/n + E(f_n))}{4(c_0 + c_1)} \geq \rho > 0,$$

since $E(f_n)$ goes to zero when f is regular enough.   □

We emphasize that the above proof cannot be immediately extended to the case of two dimensions, where the bound of $C_n(x)$ becomes proportional to $1/N(n)$, while $E(D_n)$ does not improve accordingly.

In order to manage the case of two variables, we need a different argument: for band matrices, we cannot even extend the technique of Theorem 2.6, since the analogue of Proposition 2.5, part 2 in several dimensions is more involved. The right tool is the characterization in terms of $g_r(x)$ (part 4), for which we perform a deeper investigation.

LEMMA 2.8. *Let $f \in \mathbb{T}_d$, and for a fixed x denote*

$$G_x := \{|g_r(x)| \; : \; 0 \le r \le n - e\}$$

as the set of all possible values taken by $|g_r(x)|$ among all the multi-indices. Then $\#(G_x) = O(1)$ with respect to n.

Proof. For brevity we treat the two-level case. Thanks to the band structure of A, there exist integers p_1, p_2 such that

$$a_{l_1, l_2} = 0 \qquad \text{whenever } l_1 > p_1 \text{ or } l_2 > p_2.$$

The expanded formula for $g_r(x)$ is

$$(2.5) \qquad g_{r_1, r_2}(x^1, x^2) = \sum_{k_1=0}^{n_1-1} \sum_{k_2=0}^{n_2-1} a_{r_1-k_1, r_2-k_2} \exp(\mathbf{i}(k_1 x^1 + k_2 x^2));$$

if we consider just the subscripts r_1 between p_1 and $n_1 - p_1 - 1$, the first summation in (2.5) can be restricted to the range from $r_1 - p_1$ to $r_1 + p_1$. It follows that

$$g_{r_1, r_2}(x^1, x^2) = \exp(\mathbf{i} r_1 x^1)\hat{g}_{r_2}(x^1, x^2),$$

where $\hat{g}_{r_2}(x^1, x^2) = \sum_{l_1=-p_1}^{p_1} \sum_{k_2=0}^{n_2-1} a_{l_1, r_2-k_2} \exp(\mathbf{i}(l_1 x^1 + k_2 x^2))$ no longer depends on r_1. Thus

$$|g_{r_1, r_2}(x)| = |\hat{g}_{r_2}(x)| \quad \forall r_1 = p_1, \ldots, n_1 - p_1 - 1;$$

through a similar argument we can prove that

$$|g_{r_1, r_2}(x)| = |\tilde{g}_{r_1}(x)| \quad \forall r_2 = p_2, \ldots, n_2 - p_2 - 1$$

for $\tilde{g}_{r_1}$ suitably defined. Then the set G_x reduces to the values

$$\{|g_r(x)| \; : \quad 0 \le r \le p \text{ or } n - p \le r \le n - e\}, \qquad p = (p_1, p_2),$$

whose cardinality is at most $(2p_1 + 1)(2p_2 + 1)$, independent of n. ☐

THEOREM 2.9. *Let $f \in \mathbb{T}_2$; for any root $\bar{x}$ satisfying (2.2), the same conclusions as those in Theorem 2.6 hold.*

Proof. It suffices to prove the thesis for $x = \bar{x}$ and then to apply a continuity argument. Let $J_+ := \{r : g_r(x) \ne 0\}$, $g_{\min} := \min_{r \in J_+} |g_r(\bar{x})|$, $g_{\max} := \max_{r \in J_+} |g_r(\bar{x})|$. By virtue of Lemma 2.8, the minimum $g_{\min}$ is taken over a set of cardinality $O(1)$ and therefore is strictly positive and constant with respect to n; the same holds for $g_{\max}$ (even though it can be simply derived through a continuity argument without exploiting the band structure of A). By part 4 of Proposition 2.5,

$$\frac{C_n(\bar{x})}{D_n(\bar{x})} \ge \frac{\sum_{r \in J_+} |g_r(\bar{x})|^2}{\sum_{r \in J_+} |g_r(\bar{x})|} \ge \frac{g_{\min}^2 \cdot \#(J_+)}{g_{\max} \cdot \#(J_+)} = \frac{g_{\min}^2}{g_{\max}} =: \rho > 0,$$

and for n large enough and $x \in B(\bar{x}, n^{-1/p})$ we are able to bound $\frac{C_n(x)}{D_n(x)}$ by $\rho/2$, along similar lines to those of the preceding proofs. ☐

We have always found for every root of f a suitable neighborhood where the eigenvalue function is bounded from below. Our next step is to quantify the asymptotical number of eigenvalues corresponding to those neighborhoods.

LEMMA 2.10. *For a given infinitesimal positive sequence $\{\alpha_n\}_{n\in\mathbb{N}}$ and a fixed $\bar{x} \in Q$, define the index sets*

$$J_{\alpha_n} = \{j \in \{0, \ldots, n-1\}: \quad |x_j - \bar{x}| \le \alpha_n\},$$

where $x_j = 2\pi j/n$; then $\#(J_{\alpha_n}) = O(n\alpha_n)$.

Proof. By using the expression of x_j, the extremal values of J_{α_n} are easily obtained as

$$\min J_{\alpha_n} = \lceil n(\bar{x} - \alpha_n)/2\pi \rceil, \quad \max J_{\alpha_n} = \lfloor n(\bar{x} + \alpha_n)/2\pi \rfloor,$$

whence $\#(J_{\alpha_n}) = \max J_{\alpha_n} - \min J_{\alpha_n} + 1 \le n(\bar{x}+\alpha_n)/2\pi - n(\bar{x}-\alpha_n)/2\pi + 1$ and, on the other hand, $\#(J_{\alpha_n}) \ge n(\bar{x}+\alpha_n)/2\pi - 1 - (n(\bar{x}-\alpha_n)/2\pi + 1) + 1$. In summary,

$$\left| \#(J_{\alpha_n}) - \frac{n\alpha_n}{\pi} \right| \le 1,$$

and therefore $\#(J_{\alpha_n}) = O(n\alpha_n)$. □

In several dimensions, the number of grid points x_j falling inside the hypercube $B(\bar{x}, n^{-1/p})$ can be counted as well by means of Lemma 2.10 applied to the individual components, obtaining the bound

$$\#(J_{n^{-1/p}}) = O\left(N(n)^{1-1/p}\right).$$

DEFINITION 2.11. *Let $\bar{x}$ be a zero of f. A sequence $\{\mathcal{I}_n(\bar{x})\}$ of multi-index sets is called a* degenerating family *(related to $\bar{x}$) if for any $j \in \mathcal{I}_n(\bar{x})$ we have $|j - \bar{x}n/(2\pi)| = o(n)$ or, equivalently, $x_j \in B(\bar{x}, o(1))$. For any n, the linear space spanned by the Fourier vectors $\{u_n(x_j) : j \in \mathcal{I}_n(\bar{x})\}$ is called a* degenerating subspace.

It is clear that the idea behind this is to give a formal definition of the subspaces where the sequence $\{T_n(f)\}$ is asymptotically ill-conditioned: in the case of image/signal deblurring problems, since the related eigenvectors are close to highly oscillating Fourier vectors, it follows that a union of *degenerating subspaces*, taken over the set of all the zeros of f, has a very large intersection with the one where the noise tends to concentrate (noisy subspace).

REMARK 2.2. *Let S_n be the Strang circulant approximation [29] of the matrix $A = T_n(f)$. Under one of the assumptions of Theorems 2.6 and 2.7 concerning the generating function f and its zeros $\bar{x}$, the eigenvalues of S_n associated with eigenvectors in any degenerating subspace are infinitesimal with n.*

Proof. Let $\lambda_j(S_n)$ be the eigenvalue of S_n associated with $u_n(x_j)$, where $j \in \mathcal{I}_n(\bar{x})$. It is well known that $\lambda_j(S_n) = f_n(x_j)$, where f_n is the nth partial Fourier sum of f and therefore coincides with $f \in \mathbb{T}_d$ for n large enough. In this polynomial case

$$|\lambda_j(S_n)| = |f(x_j)| = |f(\bar{x}) + f'(\xi_j) \cdot (x_j - \bar{x})| \le \|f'(\xi_j)\|_1 \cdot \|x_j - \bar{x}\|_\infty,$$

where $\bar{x}$ is the zero of f such that $x_j \in B(\bar{x}, o(1))$. Since the Jacobian f' is bounded by a constant on Q^d, then $\lambda_j(S_n) = o(1)$.

If f is a general univariate C^1 function, then

$$|\lambda_j(S_n)| = |f_n(x_j)| \le |f(x_j)| + \|f_n - f\|_\infty$$

and the same conclusions hold as before, since f_n uniformly converges to f. □

Combining the bound of Lemma 2.10, part 1 of Proposition 2.5, and the various results of this section, the proof of the next theorem becomes straightforward and tells us that P has a very different spectral behavior than S_n near the "singularities."

THEOREM 2.12. *Assume that one of the following assumptions on f holds true:*

(i) $f \in \mathbb{T}_d$;

(ii) $f \in C^1(\mathbb{R})$, f 2π-periodic,

and all the zeros $\bar{x}$ of f satisfy (2.2). Then for each $\bar{x}$ there exists a degenerating family $\{\mathcal{I}_n(\bar{x})\}$ and a positive constant ρ such that $\#(\mathcal{I}_n(\bar{x})) = O\left(N(n)^{1-1/p}\right)$ and the eigenvalues related to $\mathcal{I}_n(\bar{x})$ of the superoptimal preconditioner P are not less than ρ.

2.3. The preconditioned eigenvalues. The spectrum of P staying bounded from below in the "bad" frequencies allows us to infer a regularizing behavior for the preconditioned matrix $P^{-1}A$. By this concept we mean that in correspondence to the degenerating subspaces, the eigenvalues of $P^{-1}A$ are infinitesimal, so that the related frequencies are damped out.

We need for the proof a key observation concerning the difference between a Toeplitz matrix and its Strang approximation.

LEMMA 2.13 (see [23]). *Let $A = T_n(f)$, where f belongs to the Wiener class (that is, the series of its Fourier coefficients is absolutely summable). If S_n denotes its Strang approximation, then for any $\epsilon > 0$ there exists $\mu \in \mathbb{N}$ such that for n large enough the symmetric splitting*

$$A = S_n + R_n + E_n$$

holds, where $\|E_n\| < \epsilon$ and R_n is a sparse matrix having the following properties:

(R1) *at most $\mu N(n) \sum_{i=1}^{d} \frac{1}{n_i}$ rows are nonzero;*

(R2) *each row contains at most μ nonzero entries;*

(R3) *any nonzero element is bounded in absolute value by a constant α independent of n.*

In the polynomial case $f \in \mathbb{T}_d$, E_n can be chosen as the null matrix and μ does not depend on ϵ.

THEOREM 2.14. *Assume that one of the following assumptions on f holds true:*

(i) $f \in \mathbb{T}_d$;

(ii) $f \in C^1(\mathbb{R})$, f 2π-periodic,

and all the zeros $\bar{x}$ of f satisfy (2.2). Then for any $\epsilon > 0$ there exist k_n eigenvalues of the preconditioned matrix $P^{-1}A$ less than ϵ, where k_n is a diverging function as n tends to infinity. More precisely, $k_n = O\left(n^{1-1/p}\right)$ if f is univariate, $k_n = o\left(\min_{i=1,2} n_i\right)$ if f is bivariate.

Proof. We analyze the tth eigenvalue of the preconditioned matrix by using the classical minmax characterization

$$\lambda_t \left(P^{-1}A\right) = \min_{\dim \mathcal{V} = t} \max_{v \in \mathcal{V} \setminus \{0\}} \frac{v^* A v}{v^* P v}.$$

By Theorem 2.12, we are able to find a degenerating family $\{\mathcal{I}_n(\bar{x})\}$ corresponding to sufficiently large values of $v^* P v$. Therefore, for any n we can restrict the minmax to the particular degenerating subspace

$$\mathcal{V} = \text{span}\{u_n(x_j): \ j \in \mathcal{H}_n \subset \cup_{\bar{x} \in K} \mathcal{I}_n(\bar{x}), \ \#(\mathcal{H}_n) = t\},$$

where K denotes the set of all the zeros of f, provided that $t \leq O\left(N(n)^{1-1/p}\right) \leq \#(\cup_{\bar{x} \in K} \mathcal{I}_n(\bar{x}))$. Thus

$$\lambda_t\left(P^{-1}A\right) \leq \frac{1}{\rho} \cdot \max_{v \in \mathcal{V} \setminus \{0\}} \frac{v^* A v}{v^* v}.$$

In order to study this Rayleigh quotient, we use the splitting of A related to $\rho\epsilon/3$, given by Lemma 2.13. We obtain the sum of three terms:

$$(2.6) \qquad \lambda_t\left(P^{-1}A\right) \leq \frac{1}{\rho}\left\{\max_{v \in \mathcal{V} \setminus \{0\}} \frac{v^* S_n v}{v^* v} + \max_{v \in \mathcal{V} \setminus \{0\}} \frac{v^* R_n v}{v^* v}\right\} + \frac{\epsilon}{3}.$$

The first term in (2.6) represents the action of the Strang preconditioner on a degenerating subspace and therefore is infinitesimal. More precisely, since $v = \sum_{j \in \mathcal{H}_n} c_j u_n(x_j)$ for some c_j, we have

$$v^* S_n v = \sum_{j,j' \in \mathcal{H}_n} \bar{c}_j c_{j'} u_n(x_j)^* S_n u_n(x_{j'}) = \sum_{j \in \mathcal{H}_n} |c_j|^2 \lambda_j(S_n) \leq \frac{\rho\epsilon}{3} \sum_{j \in \mathcal{H}_n} |c_j|^2 = \frac{\rho\epsilon}{3} v^* v$$

for n large enough, in light of Remark 2.2.

The analysis of the second term in (2.6) is more delicate and makes use of the properties of R_n stated by Lemma 2.13. Consider the vector $w = R_n u_n(x_{j'})$: by property (R1), at most $\mu N(n) \sum_{i=1}^d \frac{1}{n_i}$ entries of w are nonzero, and each of them has an expression like

$$w_l = \sum_{k=1}^{N(n)} (R_n)_{l,k} \frac{1}{\sqrt{N(n)}} \exp(\mathbf{i}k \cdot x_{j'}),$$

where the summation involves at most μ terms by property (R2); hence $|w_l| \leq \mu\alpha/\sqrt{N(n)}$, where α is the constant guaranteed by property (R3).

Now the inner product between w and any Fourier vector can be bounded as follows:

$$|u_n(x_j)^* R_n u_n(x_{j'})| = \left|\sum_{l=1}^{N(n)} w_l \frac{1}{\sqrt{N(n)}} \exp(-\mathbf{i}l \cdot x_j)\right|$$

$$\leq \frac{1}{\sqrt{N(n)}} \cdot \#\{w_l \neq 0\} \cdot \max |w_l| \leq \mu^2 \alpha \sum_{i=1}^d \frac{1}{n_i}.$$

In summary, the Rayleigh quotient of R_n under consideration is

$$v^* R_n v = \sum_{j,j' \in \mathcal{H}_n} \bar{c}_j c_{j'} u_n(x_j)^* R_n u_n(x_{j'}) = c^* \hat{R} c,$$

where $c = (c_j)_{j \in \mathcal{H}_n} \in \mathbb{C}^t$ and $\hat{R} = (u_n(x_j)^* R_n u_n(x_{j'}))_{j,j' \in \mathcal{H}_n}$ is a $t \times t$ matrix whose entries are bounded by $\mu^2 \alpha \sum_{i=1}^d \frac{1}{n_i}$. Thus

$$v^* R_n v \leq \lambda_{\max}(\hat{R}) c^* c \leq \|\hat{R}\|_\infty v^* v \leq t\mu^2 \alpha \sum_{i=1}^d \frac{1}{n_i} v^* v.$$

In the univariate case, the condition $t \leq \#(\mathcal{I}_n) = O\left(n^{1-1/p}\right)$ makes this bound less than $(\rho\epsilon/3)v^* v$ for a sufficiently large n. In two variables, we must force the index t to take values less than $o\left(\min_{i=1,2} n_i\right)$ because $N(n)^{1-1/p} > \min_{i=1,2} n_i$ for $p \geq 2$ (since f is a nonnegative trigonometric polynomial, p cannot be smaller).

In both instances, substituting in (2.6) proves that $\lambda_t\left(P^{-1}A\right) \leq \epsilon$. $\quad\square$

3. The computation of the superoptimal approximation. In this section we consider the algorithmic problem of building the circulant superoptimal preconditioner in the 2-D setting: the computation for the 1-D case has been described in [5], but its multilevel extension is not trivial and deserves a specific analysis.

By virtue of Theorem 2.2, the superoptimal operator $\hat{\mathcal{P}}_F$ can be computed with the help of the simpler optimal one $\mathcal{P}_F$ thanks to the formula

$$(3.1) \qquad \hat{\mathcal{P}}_F(A) = [\mathcal{P}_F(A)]^{-1}\mathcal{P}_F(A^2) = \frac{\mathcal{P}_F(A^2)}{\mathcal{P}_F(A)}.$$

Inside the denominator, the optimal preconditioner $\mathcal{P}_F(A)$ of a Toeplitz matrix A is exactly computed within $O(N \log N)$ operations [6], where $N = n_1 n_2$ and n_1, n_2 denote the 2-D input (i.e., image) sizes. Unfortunately, in the explicit formula (3.1), the matrix A^2 (or A^*A in the nonsymmetric case) appearing in the numerator does not preserve the Toeplitz structure in the most relevant cases; therefore the previously developed $O(N \log N)$ efficient procedures for the construction of the optimal operator cannot be applied directly. Under these considerations, our effort is addressed to the computation of the term $\mathcal{P}_F(A^2)$ (the one hard to construct in $O(N \log N)$ time).

It is well known that a 1-D (scalar) Toeplitz matrix can be decomposed into the sum of a circulant and a skewcirculant matrix [16] by taking simple averages of its elements. We recall that an $m \times m$ skewcirculant 1-D matrix $(b_{i,j})$ is a Toeplitz matrix such that $b_{i,j} = b_{(i-j) \bmod n}$ if $i \geq j$ and $b_{i,j} = -b_{(i-j) \bmod n}$ if $i < j$, where b_r denotes the rth entry of its first column. The linear space of the skewcirculants is diagonalized by the unitary transform $S_m = F_m \Theta^*$, where $\Theta = \mathrm{diag}\left(1, \sigma, \sigma^2, \ldots, \sigma^{m-1}\right)$ and $\sigma^m = -1$; hence it can be viewed as the algebra $\mathcal{M}(S_m)$ according to the notation of section 2. The key point of the algorithm we are going to develop for the 2-D setting is the following: if we perform the same decomposition into circulant and skewcirculant matrices at each block level of the two-level Toeplitz matrix A, then the optimal preconditioner $\mathcal{P}_F(A^2)$ can be computed on the individual terms only by using one-level algebras arguments.

DEFINITION 3.1. *The algebra $\mathcal{C}^{(1)}$ of external-level circulant matrices is the set of block matrices that are diagonalized by the Fourier transform just at the outer block level:*

$$(3.2) \qquad \mathcal{C}^{(1)} = \{A = (F_{n_1} \otimes I_{n_2})\Delta(F_{n_1}^* \otimes I_{n_2}) \mid \Delta \ block \ diagonal\}.$$

Similarly, the algebra $\mathcal{C}^{(2)}$ of internal-level circulant matrices is just diagonalized at the inner level:

$$(3.3) \qquad \mathcal{C}^{(2)} = \{A = (I_{n_1} \otimes F_{n_2})\Delta(I_{n_1} \otimes F_{n_2}^*) \mid \Delta \ has \ diagonal \ blocks\}.$$

In the same way, the one-level skewcirculant algebras $\mathcal{S}^{(1)}$ and $\mathcal{S}^{(2)}$ are defined as in (3.2) and (3.3) with the unitary transform S_{n_i} in place of F_{n_i}, $i = 1, 2$.

One-level algebras have been introduced for the circulant case in [4]. The case of general algebras can be found in [9], where the one-level diagonal form is extended by using the concept of a masking operator.

Now report the two-level generalization of the decomposition of Toeplitz matrices into sums of circulant and skewcirculant ones. Let $A_C \in \mathcal{C}^{(1)}$ and $A_S \in \mathcal{S}^{(1)}$ denote the external-level decomposition of the two-level matrix A, that is, $A = A_C + A_S$, obtained by using the same averaging rule as in the scalar setting; notice that the

inner blocks of A_C and A_S retain the Toeplitz structure since they are obtained by averaging those of A.

Therefore, if we consider the internal level, in a similar way, we can decompose both A_C and A_S into the sum of matrices which belong to the corresponding algebras $\mathcal{C}^{(2)}$ and $\mathcal{S}^{(2)}$. In summary, we obtain a total decomposition of the two-level block Toeplitz matrix A into matrices that belong to matrix algebras endowed with low computational costs, that is,

$$A = A_{C,C} + A_{C,S} + A_{S,C} + A_{S,S},$$

where the first index is related to the external level and the second to the internal level of partitioning; for instance, $A_{S,C} \in \mathcal{M}(S_{n_1} \otimes F_{n_2})$. From such a decomposition, and thanks to the linearity of the optimal approximation [9], we obtain the following constructive formula for the numerator of the superoptimal preconditioner:

$$(3.4) \qquad \mathcal{P}_F(A^2) = \sum_{M_1,M_2,N_1,N_2 \in \{C;S\}} \mathcal{P}_F(A_{M_1,M_2} A_{N_1,N_2}).$$

The computation of $\mathcal{P}_F(A^2)$, where A^2 is generally unstructured (i.e., with a higher displacement rank structure) and hard to manage, is now reduced to the computation of the 16 terms of the latter formula, where all the matrices A_{M_1,M_2} and A_{N_1,N_2} belong to simple matrix algebras related to fast transforms.

The next basic step is to compute the optimal approximation at the external and internal levels separately.

DEFINITION 3.2 (see [4]). *The optimal circulant operator* $\mathcal{P}^{(1)}$ *at the external level in the algebra* $\mathcal{C}^{(1)}$ *and the optimal circulant operator* $\mathcal{P}^{(2)}$ *at the internal level in the algebra* $\mathcal{C}^{(2)}$ *are defined as follows:*

$$(3.5) \qquad \mathcal{P}^{(i)}(A) = \arg \min_{X \in \mathcal{C}^{(i)}} \|A - X\|_F,$$

where $i = 1$ *for the external level case and* $i = 2$ *for the internal level case.*

In order to explain all the necessary tools for developing a fast algorithm, we now summarize some useful properties related to the optimal operators $\mathcal{P}^{(1)}$ and $\mathcal{P}^{(2)}$.

PROPOSITION 3.3. *The following facts hold true:*

1. *The two single-level optimal preconditioners commute with respect to the composition operator:*

$$\mathcal{P}_F(A) = \mathcal{P}^{(1)}(\mathcal{P}^{(2)}(A)) = \mathcal{P}^{(2)}(\mathcal{P}^{(1)}(A)).$$

2. *If* $A \in \mathcal{C}^{(i)}$, *then the single-level optimal operator* $\mathcal{P}^{(i)} \in \mathcal{C}^{(i)}$ *satisfies the two equalities*

$$\mathcal{P}^{(i)}(AB) = A\,\mathcal{P}^{(i)}(B) \qquad and \qquad \mathcal{P}^{(i)}(BA) = \mathcal{P}^{(i)}(B)\,A \qquad (i = 1, 2),$$

with no special assumption on B.

3. *If* A *is a block Toeplitz matrix, then the block entries of* $\mathcal{P}^{(1)}(A)$ *are suitable weighted averages of those of* A; *similarly, if* A *is a block matrix with Toeplitz blocks, then the inner blocks of* $\mathcal{P}^{(2)}(A)$ *are obtained by correcting in* $\mathcal{M}(F_{n_2})$ *the inner blocks of* A.

Proof. The first equality of part 1 is proved in [4, Theorem 3], but the same argument can be applied for the second equality. Part 2 is a block extension of

[4, Lemma 1] and a particular instance of [9, part 6 of Theorem 3.1]. Part 3 is proved in [6]. ☐

If we apply the results of Proposition 3.3, the fast computation of each term in the decomposition (3.4) is completely provided.

COROLLARY 3.4. *Let us consider the decomposition* (3.4) *of the optimal precon-ditioner* $\mathcal{P}_F(A^2)$, *and let* $\mathcal{P}^{(i)}$ *be the optimal preconditioner of level* i *in the algebra* $\mathcal{C}^{(i)}$, *with* $i \in \{1;2\}$ *fixed. The following four constructive relations hold:*

(i) *If* $M_i = N_i = C$, *then* $\mathcal{P}^{(i)}(A_{M_1,M_2}A_{N_1,N_2}) = A_{M_1,M_2}A_{N_1,N_2}$.

(ii) *If* $M_i = C$ *and* $N_i = S$, *then* $\mathcal{P}^{(i)}(A_{M_1,M_2}A_{N_1,N_2}) = A_{M_1,M_2}\mathcal{P}^{(i)}(A_{N_1,N_2})$.

(iii) *If* $M_i = S$ *and* $N_i = C$, *then* $\mathcal{P}^{(i)}(A_{M_1,M_2}A_{N_1,N_2}) = \mathcal{P}^{(i)}(A_{M_1,M_2})A_{N_1,N_2}$.

(iv) *If* $M_i = N_i = S$ *and* $M_j = N_j$ *with* $j \neq i$, *then* $B = A_{M_1,M_2}A_{N_1,N_2}$ *belongs* *to the same matrix algebra of* A_{M_1,M_2} *and* A_{N_1,N_2}, *which is a subspace of* $\mathcal{S}^{(i)}$.

Item (i) of Corollary 3.4 simply follows from Definition 3.2, which implies that if $B \in \mathcal{C}^{(i)}$, then $\mathcal{P}^{(i)}(B) \equiv B$. Items (ii) and (iii) are direct effects of Proposition 3.3, part 2.

On the other hand, a little more detail should be devoted to item (iv). By the assumptions, both A_{M_1,M_2} and A_{N_1,N_2} belong to the same algebra, which is one of either $\mathcal{M}' = \mathcal{S}^{(i)} \cap \mathcal{C}^{(j)}$ or $\mathcal{M}'' = \mathcal{S}^{(i)} \cap \mathcal{S}^{(j)}$, accordingly to the hypotheses. In such algebras $\mathcal{M}'$ and $\mathcal{M}''$, which concern circulant and skewcirculant matrices only, the product of A_{M_1,M_2} and A_{N_1,N_2} costs $O(N \log N)$, $N = n_1 n_2$. Furthermore, it is important to note that all arguments of the i-level preconditioners $\mathcal{P}^{(i)}$ in the right-hand sides of the four items of Corollary 3.4 (that is, A_{M_1,M_2}, A_{N_1,N_2}, and the matrix B) are two-level Toeplitz matrices: the computation of such preconditioners simply reduces to $O(N \log N)$ operations in light of part 3 of Proposition 3.3.

The algorithm for the computation of each term (3.4) of the superoptimal pre-conditioner $\mathcal{P}_F(A)$ is now totally explained: for any term, we have to build the com-position of the external and internal level optimal operators, regardless of the order of composition (according to part 1 of Proposition 3.3). We look for the right optimal operators to use, in order to apply one of the reductions (i), (ii), or (iii) of the above Corollary 3.4, within $O(N \log N)$ operations. If none of the first three equalities holds, then we use item (iv) with the same $O(N \log N)$ complexity.

For a better explanation, we show, for instance, how the term $A_{S,S}A_{C,S}$ must be manipulated.

Since case (iii) of Corollary 3.4 occurs at the outer level, it is convenient to apply the high-level operator $\mathcal{P}^{(1)}$ first. We obtain

$$(3.6) \qquad \mathcal{P}^{(1)}(A_{S,S}A_{C,S}) = \mathcal{P}^{(1)}(A_{S,S})A_{C,S},$$

where the first factor belongs to $\mathcal{M}(F_{n_1} \otimes S_{n_2})$ as (trivially) the second one: in fact, $\mathcal{P}^{(1)}$ is computed by averaging the blocks of $A_{S,S}$, and therefore the inner skewcirculant structure is preserved. Thus the matrix multiplication in (3.6) can be performed by a fast algorithm in the algebra $\mathcal{M}(F_{n_1} \otimes S_{n_2})$. The application of $\mathcal{P}^{(2)}$ can be made just by correcting each inner Toeplitz block separately, yielding $\mathcal{P}_F(A_{S,S}A_{C,S}) = \mathcal{P}^{(2)}(\mathcal{P}^{(1)}(A_{S,S})A_{C,S})$ in $O(N \log N)$ operations.

In conclusion, by coming back to the explicit formula of Theorem 2.2, we can summarize the previous computational analysis in the following proposition.

PROPOSITION 3.5. *The superoptimal preconditioner* $\hat{\mathcal{P}}_U(A)$ *of the two-level Toe-plitz matrix* A *is computable within* $O(N \log N)$ *operations.*

Clearly, the algorithm is more involved than for the optimal preconditioner, and the hidden constant in the $O(\cdot)$ notation is higher (see [14] for details). Anyway,

this is just a data-independent, preprocessing step; once the preconditioner has been constructed, for each input data every PCG iteration has exactly the same complexity in both the cases.

4. A regularizing preconditioner. The present section is devoted to testing how the distribution of the eigenvalues of the superoptimal approximation can be favorable for solving ill-conditioned Toeplitz systems arising in inverse problems of image restoration.

The mathematical model which represents the process of image formation can be described, in first analysis, by the following Fredholm operator of the first kind:

$$(4.1) \qquad g(x,y) \,=\, \int_{\mathbb{R}^2} \tilde{\mathcal{K}}(x,y,\theta,\xi)\, h_o(\theta,\xi)\, d\theta d\xi \,+\, \omega(x,y),$$

where h_o is the (true) input object; $\tilde{\mathcal{K}}$ is the integral kernel of the operator, also called the point spread function (PSF); ω is the noise which arises in the process; and g is the observed image.

In most applications, the kernel $\tilde{\mathcal{K}}$ is assumed to be spatially invariant. This simplifies the expression of the PSF, which can be written as $\tilde{\mathcal{K}}(x,y,\theta,\xi) = \mathcal{K}(x-\theta,y-\xi)$, where $\mathcal{K}$ represents the space-invariant kernel. Hence the discretization of (4.1), with image size $n = (n_1, n_2)$, leads to the equation $g = Ah_o + \omega$, where g, h_o, ω represent the column-ordered vectors of size $N = n_1 n_2$ of the corresponding quantities and A is an $N \times N$ block Toeplitz matrix with Toeplitz blocks (BTTB) that discretizes the kernel $\mathcal{K}$ [1]. When the PSF is space-variant (with smooth variation), it is worth mentioning that both the matrix vector product and the spectral analysis can be reduced in principle to the shift-invariant case (see, e.g., [21, 27]).

The image restoration problem is the inverse of (4.1). Given the blurred and noisy image g, we have to recover an approximation h of the true solution h_o by computing an approximate solution of the linear system $Ah = g$; in this approximation process it is crucial to neglect those components which carry noise information, typically corresponding to the smallest eigenvalues of A [1].

From the theory developed in section 2, if $A = T_n(f)$, the eigenvalues of the superoptimal preconditioner $P = \hat{\mathcal{P}}_F(A)$, related to the vanishing ones of A, stay bounded away from zero. In the recovering process of the true solution by the preconditioned Toeplitz system, the main consequence is that a number of components of the data usually corrupted by noise are suitably filtered out; in other words, the action of $P^{-1}A$ is essentially the same as that of A in the "noise subspace."

We recall that the conjugate gradient method is an iterative regularization algorithm for inverse problems [25]: an early stop of the iterations prevents reconstruction from components related to the noise, thanks to its filtering action on the noise subspace of A, so that the iteration count plays the role of a regularization parameter. The preconditioned method, essentially based on the matrix $P^{-1}A$, should give comparable results in fewer iterations due to the eigenvalue clustering along the "signal subspace"; this observation led Hanke, Nagy, and Plemmons [19] to introduce a modification of the optimal preconditioner, obtained through the "cut" of the bad eigenvalues below a threshold level, which plays here the role of a *second* regularization parameter. Other approaches need a similar filtering, based on estimates of the "size" of the noise subspace [22, 18], which acts as a regularization parameter.

We emphasize that the latter strategies are not our comparison term here, since it is not fair to compare a parameter-dependent family with a "fixed" approach, which the superoptimal preconditioner is. In particular, it should be stressed that the level

of regularization given by the superoptimal preconditioner is related only to the size of A and to the order p defined in (2.2) for the zeros of the generating function f by virtue of the result of Theorem 2.14. Basically, the degenerating subspace, where the filtering action is located, has a dimension depending on these fixed parameters. However, in some severely ill-posed cases, such a level of regularization may be not enough, since the actual noise space is greater than the degenerating one where the preconditioner acts.

In the presence of such difficulties, the level of regularization must be increased by generalizing the superoptimal preconditioner to a parameter-dependent family, developed in [12, 13], where the additional parameter plays a role very similar to the threshold introduced by Hanke, Nagy, and Plemmons. Hence, a fully correct comparison should be made between the filtered optimal family of [19] and the generalized superoptimal one, and this is beyond the scope of this paper. Here we deal only with the raw superoptimal preconditioner (2.1), which is the basic one of that family, and whose (even limited) filtering capabilities are essential for the effectiveness of the whole family.

This section is therefore devoted to showing that the superoptimal preconditioner provides regularization capabilities which allow us to obtain good reconstructions for a certain range of discrete ill-posed problems. We will discuss some 2-D numerical experiments arising in image processing, by comparing the results of the superoptimal preconditioned conjugate gradient method with respect to the nonpreconditioned one, in order to confirm both the expected regularization and acceleration behaviors.

Since all our BTTB matrices are not symmetric and positive definite, in our examples we always use the least squares version of both methods, namely, CGLS and PCGLS [6].

It is important to mention here a result of Kailath and Olshevsky [20] stating that in principle one iteration of PCGLS with a trigonometric transform–based preconditioner can be implemented at the same cost as a CGLS iteration (we did not apply this trick in our experiments); therefore the number of iterations performed is a good parameter for evaluating the performances of the preconditioner.

The numerical tests have been implemented on IDL 5.3 (Interactive Data Language) and performed on an IBM PC with floating-point precision of 10^{-8}. The superoptimal preconditioner P is computed according to the fast, $O(N \log N)$, algorithm of section 3.

For each method we run 100 iterations, and among all we choose the iteration giving the best relative restoration error (RRE) $\|h - h_o\|/\|h_o\|$. This choice is motivated by the need for full control on the reconstruction level as the iterations proceed.

In practice, if the method were applied to real data, careful attention would have to be paid to the selection of a good stopping criterium (which could be equivalent in work to finding a Tikhonov regularization parameter).

In our numerical tests, we consider two different PSFs. Only the generating function of the first one satisfies the assumptions used in all the previous sections. On the other hand, the second one is an experimental PSF, which has a more practical effectiveness and allows us an actual comparison of the results.

4.1. First example: Banded matrix. The BTTB system matrix of the first example is related to a trigonometric polynomial generating function $f \in \mathbb{T}_2$ and hence is banded. Bounded PSFs arise in out-of-focus image formation problems, where the blurring function is uniform on a bounded support [24], or in the case of motion blur [1]. Even though it is not very usual in practice, this simple example will

allow us to study the performances in a case where our theory developed in section 2 fully applies.

In our tests, the Toeplitz matrix is constructed by means of a 1-D convolution kernel proposed by Eldén [11] as a prototype of image restoration problems. In particular, $A = T_{n_1,n_2}(f(x_1, x_2)) = \tilde{T}_{n_1}(\tilde{f}(x_1)) \otimes \tilde{T}_{n_2}(\tilde{f}(x_2))$ for $n_1 = n_2 = 256$, where $\tilde{T}_m(\tilde{f}(x))$ is the $m \times m$ symmetric Toeplitz matrix such that

$$(4.2) \qquad [\tilde{T}_m(\tilde{f})]_{r,s} = a_{r-s} = \begin{cases} \frac{4}{51} k_{0.15}(x_r - x_s) & \text{if } |r - s| < B, \\ 0 & \text{otherwise,} \end{cases}$$

$k_\sigma(t)$ is the Gaussian distribution with zero mean and standard deviation σ, the points x_j for $j = 1, \ldots, n$ are equally distributed in $[-2, 2]$, and B is a preassigned bandwidth.

We provide numerical results for the three following band sizes $B = 9$, $B = 15$, and $B = 30$, which correspond to three different degrees of the trigonometric polynomial generating function $f \in \mathbb{T}_2$. These increasing values of B yield a worse conditioning for A: in Table 4.1 we show the estimates obtained through the condition numbers of the corresponding circulant optimal preconditioners $\mathcal{P}_F(A)$.

TABLE 4.1
Estimates for the conditioning of A.

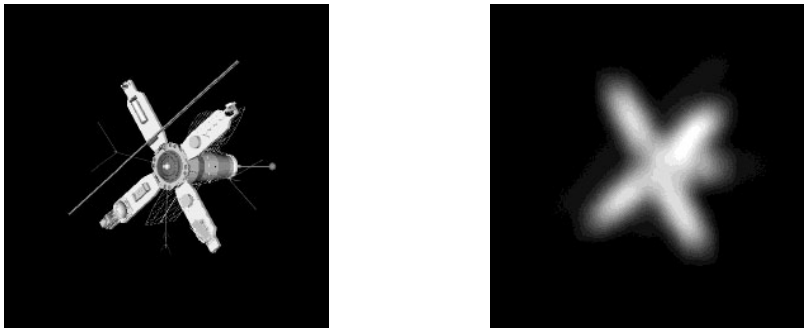| B	$\mathrm{cond}(\mathcal{P}_F(A))$
9	$4.468 \cdot 10^4$
15	$6.464 \cdot 10^6$
30	$4.834 \cdot 10^7$

FIG. 4.1. *Satellite (true) image and "blurred and noisy" data for banded PSF ($B = 15$).*

The input image h_o is of a satellite (see Figure 4.1) in 256×256 pixels; the image was developed at the U.S. Air Force Phillips Laboratory, Lasers and Imaging Directorate, Kirtland Air Force Base, NM, and it is widely used in numerical tests for image restoration problems [26]. The system data g shown in Figure 4.1 are contaminated by a noise ω, coming from a normal distribution with zero mean (white noise). We adopt two different standard deviations for the white noise, which, respectively, give rise to relative noise $E_\omega = \|Ah_o - g\|/\|Ah_o\|$ of about $2 \cdot 10^{-5}$ and $2 \cdot 10^{-4}$.

The numerical results are shown in Table 4.2. Since the condition number of the matrix A increases with the size of the band, the superoptimal approximation seems to give a bad RRE only for the case $B = 30$, in which the ill-conditioning is too large;

TABLE 4.2
Best relative errors and numbers of iterations for banded PSF.

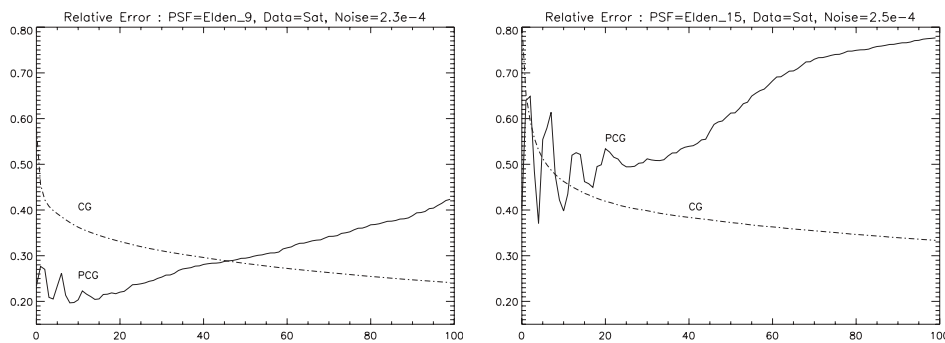| Banded PSF	CGLS		PCGLS	
	Best RRE	Iteration	Best RRE	Iteration
$B = 9$; $E_\omega = 2 \cdot 10^{-5}$	0.2413	100	0.1510	19
$B = 9$; $E_\omega = 2 \cdot 10^{-4}$	0.2413	100	0.1968	9
$B = 15$; $E_\omega = 2 \cdot 10^{-5}$	0.3336	100	0.2518	10
$B = 15$; $E_\omega = 2 \cdot 10^{-4}$	0.3335	100	0.3707	5
$B = 30$; $E_\omega = 2 \cdot 10^{-5}$	0.5205	100	0.9999	9

FIG. 4.2. *Reconstruction errors for the banded PSF versus iterations.*

in the first two cases, for $B = 9$ and $B = 15$, the results are very satisfactory, even though the performance deteriorates when the noise level is higher.

Figure 4.2 shows the convergence history of the RREs in the first 100 iterations, with $E_\omega \approx 2 \cdot 10^{-4}$, for $B = 9$ and $B = 15$. The convergence of the CGLS is slow in both instances, and the minimum values of the RREs decrease for many iterations (more than 100). When the conditioning is high, the PCGLS is less stable, but it starts to give worse results early, within few iterations (about 20 for $B = 9$ and about 10 for $B = 15$); hence the PCGLS has to be stopped first. In such a case, the early stop avoids the reconstruction from the noise space, and simultaneously the preconditioner gives a good reconstruction from the signal space.

Figure 4.3 shows the reconstructed images with minimal RRE, for both the CGLS and the PCGLS methods, with relative noise $E_\omega \approx 2 \cdot 10^{-4}$ for $B = 9$, $B = 15$ and $E_\omega \approx 2 \cdot 10^{-5}$ for $B = 30$. As already mentioned, the convergence of the CGLS method is slow, as shown by the figures in the left column, but it gives better results when the ill-conditioning is very high, that is, for the case $B = 30$ in the third row. On the other hand, in the first two rows, where $B = 9$ and $B = 15$, the superoptimal preconditioner gives a good reconstruction, which is affected only by small periodic artifacts, probably due to the discontinuity of the PSF at the end of the band. The case $B = 30$ shows that for high ill-conditioning, the regularization properties of the superoptimal preconditioner can be insufficient. This is the case where it would be useful to enforce the regularization of the superoptimal preconditioner by means of its parametric extension [12].

As we expect, the raw T. Chan optimal approximation, which we have also tested, is not competitive since it always gives very high restoration errors (RRE > 0.9) by confirming the absence of regularization properties and the need of a *good* choice

CGLS   PCGLS

$It = 100 \ (B = 9)$ $It = 9 \ (B = 9)$

$It = 100 \ (B = 15)$ $It = 5 \ (B = 15)$

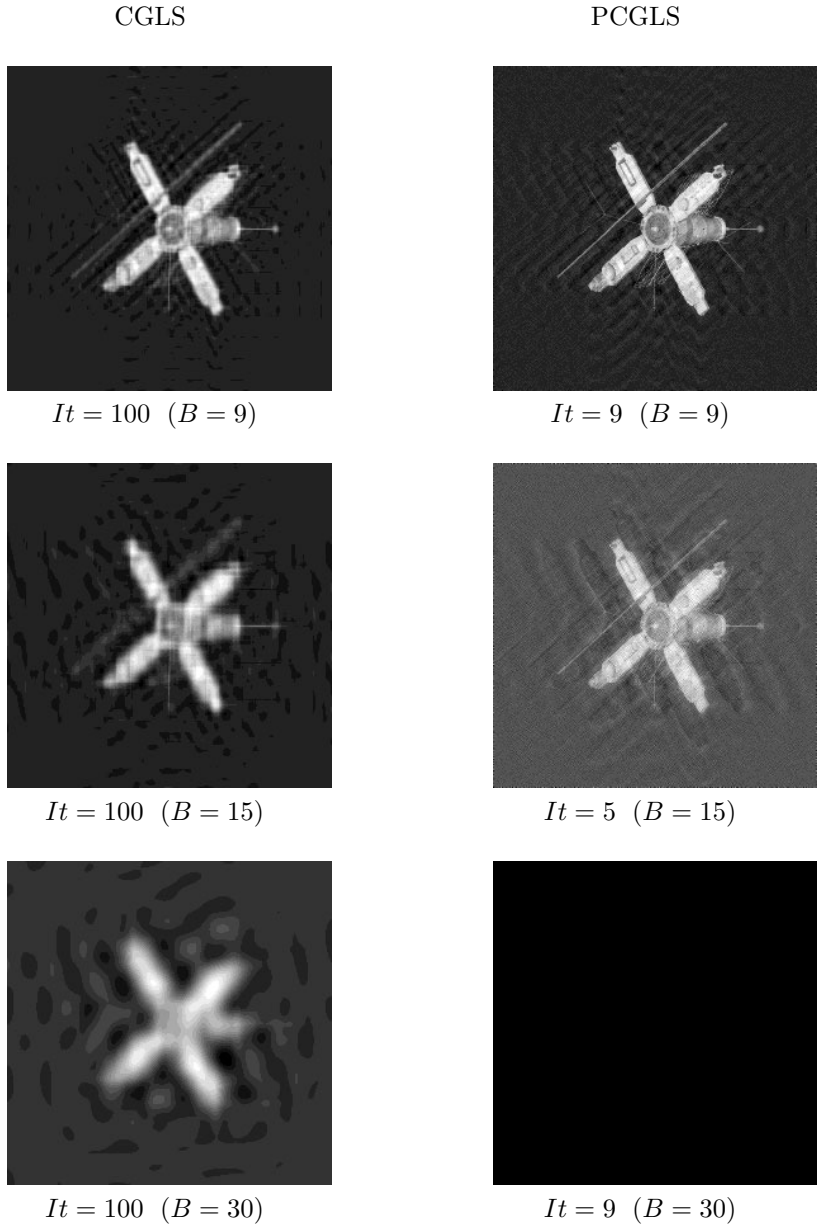$It = 100 \ (B = 30)$ $It = 9 \ (B = 30)$

FIG. 4.3. *Best reconstructed images for the banded PSF.*

of the threshold parameter in the filtered version of [19]. With low levels of error such as those in these experiments, this problem does not occur for the superoptimal preconditioner, which works as it is.

4.2. Second example: Full matrix. In the second example, we test the regularization properties of the superoptimal preconditioning for a full BTTB system matrix. The related 256×256 PSF is a widely used experimental PSF developed by the U.S. Air Force Phillips Laboratory, as is the satellite data of the previous example. We point out that our theoretical results do not cover the present case.

In this more practical case, we must not forget the "original sin" of circulant approximations (either optimal or superoptimal): roughly speaking, they add artificial periodic boundary conditions to the blurring operator A. This implies a lack of information near the boundaries of the image to be recovered (the resulting artifacts are usually called ringing effects), and in that portion we should expect a better behavior of the nonpreconditioned CGLS method.

In the previous example with bounded PSFs, both the image and the PSF have an "empty" boundary, and hence they are easier to process by circulant preconditioning. In that case, both Toeplitz and circulant approximations lead indeed to the same exactly null convolution product close to the boundary. On the contrary, in this second example the PSF is full, which means that it differs from zero everywhere, and artifacts turn up near the boundaries of the reconstructions. This problem was recently investigated by enforcing more appropriate boundary conditions, such as reflective [24] or antireflective [28], that are able to reduce the ringing effects, obtaining a remarkable improvement of the reconstructions. It would be interesting to integrate this approach in our PCG solver, but this is beyond the scope of the paper.

The data error E_ω on the blurred data Ah_o is again white noise, and we test both the noise-free case $E_\omega = 0$ and $E_\omega \approx 3 \cdot 10^{-4}$, $1 \cdot 10^{-3}$. In addition, we consider experimental blurred data which have $E_\omega \approx 3 \cdot 10^{-2}$ [26].

Table 4.3 shows the best RREs and the number of iterations needed; graphs of convergence history of the first 100 iterations are shown in Figure 4.4.

TABLE 4.3
Best relative errors and numbers of iterations for experimental full PSF.

Full PSF	CGLS		PCGLS	
	Best RRE	Iteration	Best RRE	Iteration
$E_\omega = 0$	0.3091	100	0.6744	19
(center)	0.3072	100	0.2092	19
$E_\omega = 3 \cdot 10^{-4}$	0.3092	100 (38)	0.7628	11
(center)	0.3074	100 (38)	0.3777	11
$E_\omega = 1 \cdot 10^{-3}$	0.3092	100 (9)	0.9165	4
(center)	0.3074	100 (9)	0.6916	4
$E_\omega = 3 \cdot 10^{-2}$	0.3734	58	≥ 0.95	-

For any noise level, we compute two different RREs. The former is related to the whole image size and the latter to the internal portion of the image, that is, in the pixels $(33 : 224, \ 33 : 224)$. The reconstruction errors of the PCGLS method decrease a lot in the internal portion of the image, as shown by the fourth column of Table 4.3; the same does not happen for the CGLS method, shown in the left side of the table, which exhibits uniform reconstruction errors. Some of the images are shown in Figure 4.5, where the automatic scaling of gray levels has been used.

The superoptimal preconditioner gives very good results only if the noise level is low. The first part of Table 4.3 concerns the noise-free data. Here the PCGLS reconstruction in the internal portion at the 19th iteration is much better than the best one of the CGLS reconstruction within 100 iterations. The quality of that PCGLS image is quite higher than the CGLS one, as the first two images of Figure 4.5 show. The correspondent convergence history is the upper graph of Figure 4.4. The reconstruction of the PCGLS method is very fast, and it needs about 10 iterations to give the same result as that of the hundredth iteration of the nonpreconditioned method.
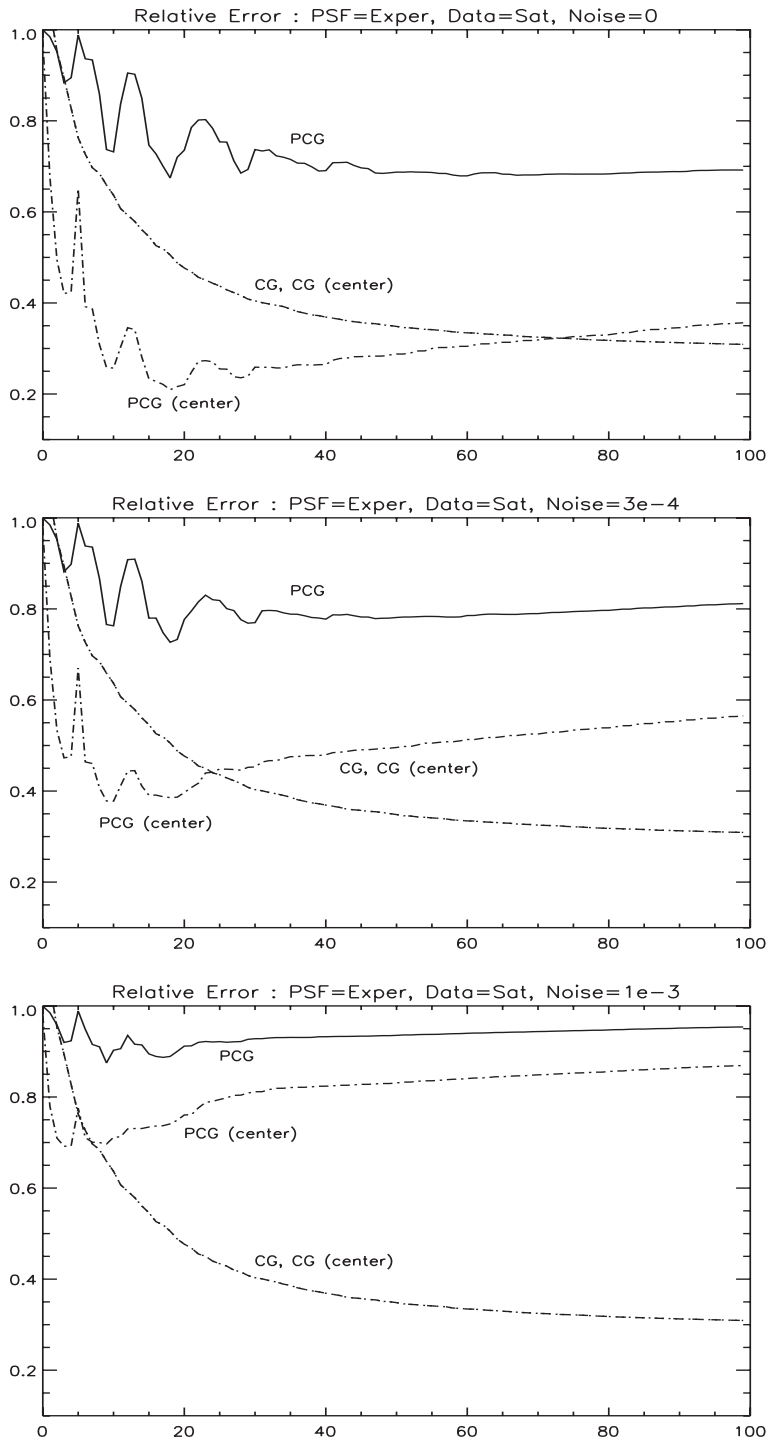
FIG. 4.4. *RREs versus iterations for the experimental full PSF.*

CGLS   PCGLS

$E_\omega = 0$; $It = 100$ $E_\omega = 0$; $It = 19$

$E_\omega = 3 \cdot 10^{-4}$; $It = 38$ $E_\omega = 3 \cdot 10^{-4}$; $It = 11$

$E_\omega = 1 \cdot 10^{-3}$; $It = 9$ $E_\omega = 1 \cdot 10^{-3}$; $It = 4$
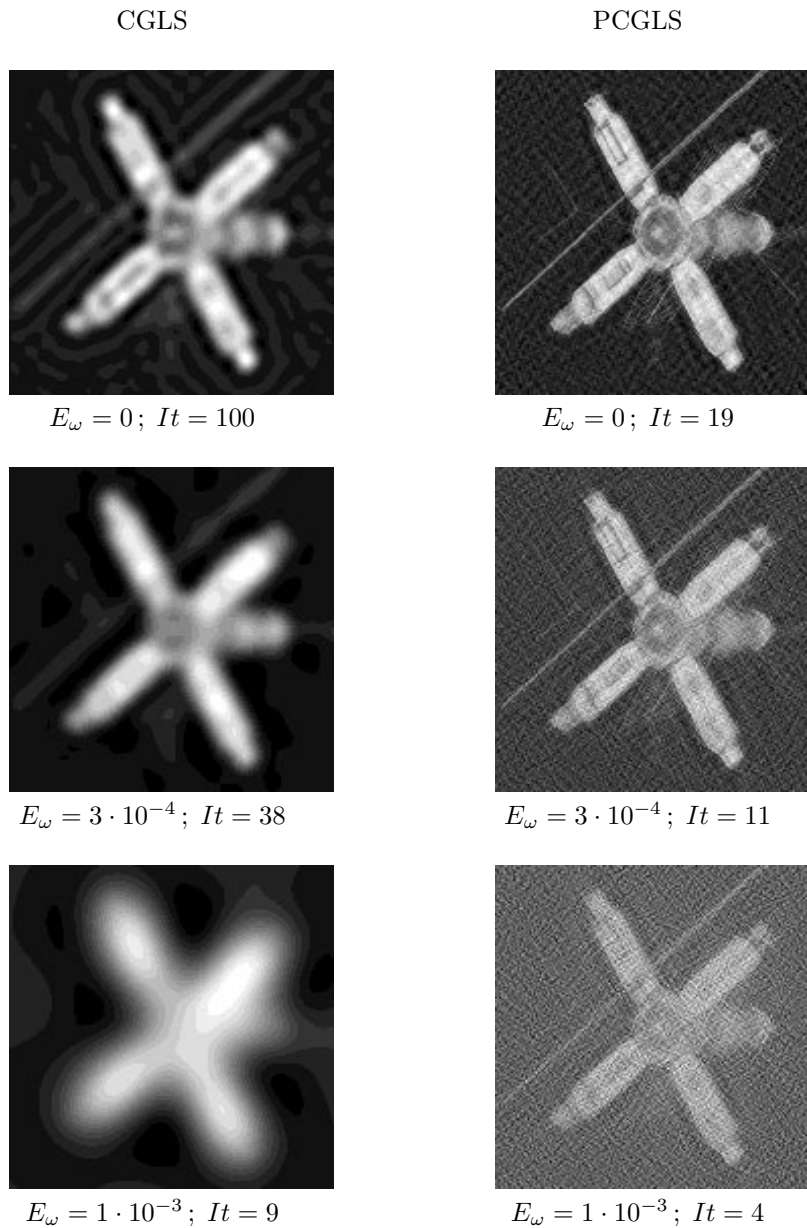
FIG. 4.5. *Reconstructions for the experimental full PSF.*

On the other hand, we have that the reconstruction is good only in the internal part, while it is unsatisfactory at the boundaries, especially close to the corners, where high ringing effects occur.

We remark that the T. Chan optimal preconditioner does not even work at all in this noise-free case (RRE > 0.95), where the reconstruction is completely meaningless.

When the data is corrupted by noise, the PCGLS method is still fast, but the lowest RREs are given by the nonpreconditioned method with a lot more iterations. The second parts of Table 4.3 and Figure 4.4 show the results with low noise

$E_\omega \approx 3 \cdot 10^{-4}$. The best result of the PCGLS method is obtained at the 11th iteration with RRE ≈ 0.37, and a similar RRE needs 38 iterations in the nonpreconditioned CGLS, as shown in Table 4.3 by the numbers in parentheses. In subsequent iterations, the results of the CGLS method keep giving smaller RREs, while the noise effects start to dominate the PCGLS method by the 20th iteration, as Figure 4.4 shows.

However, the quality of the images is quite different, and it is inadequate to compare the performances just on the basis of the values of RRE. The two images in the second row of Figure 4.5 are the 38th iteration of the CGLS method (left side) and the 11th iteration of the PCGLS method (right side). Note that the RRE of both the images is about 0.37; the first image is "cleaner" but exhibits fewer details than the second one. The large RRE for the PCGLS method is due to high errors localized in few pixels. On the other hand, the image of the CGLS method is more uniform, that is, without large errors; hence the RRE of the CGLS method is smaller, even though the quality of its image is worse.

When the noise increases, the reconstructions of the conjugate gradient method are generally better. Here the superoptimal preconditioner begins to amplify far too early the data error components in the recovering process. In the case of noise $E_\omega \approx 1 \cdot 10^{-3}$, the PCGLS method can be used within very few iterations in order to have a first approximation of the solution. Only in those iterations, indeed, the results of the PCGLS method are better than the ones of CGLS, as shown in the third row of Figure 4.4. In this case, the regularization effects of the superoptimal preconditioner are not sufficient to compensate the data errors, and it would be better to adopt a higher superoptimal-based noise filtering [12], as previously explained. Such a fault is confirmed by the experimental data with high noise $E_\omega \approx 3 \cdot 10^{-2}$. In this case the superoptimal preconditioner is useless, while the nonpreconditioned CGLS method gives the best solution at the 58th iteration, with RRE $= 0.3734$ (see Figure 4.6).

CGLS   PCGLS
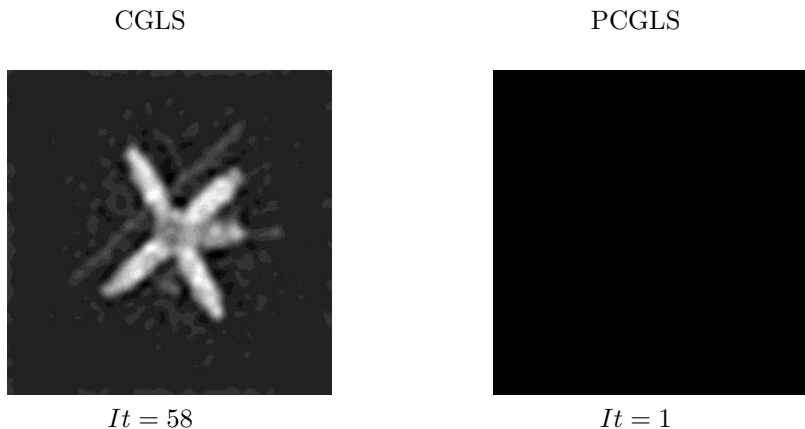

$It = 58$ $It = 1$

FIG. 4.6. *Reconstructions for the experimental PSF with $E_\omega \approx 3 \cdot 10^{-2}$.*

In summary, we can state that the superoptimal preconditioner can work as a basic regularizing tool, which, in the absence of critical conditions (e.g., high noise level and/or excessive ill-conditioning), provides reconstructions of acceptable quality in much fewer iterations when compared to the nonpreconditioned method. More precisely, the following observations can be made:

(i) The superoptimal approximation leads to a basic level of regularization without further operations. This feature is not common in other techniques where the

preconditioner has to be filtered in some way, and this operation is often very difficult to do (see, e.g., [13]).

(ii) The level of regularization achieved by the superoptimal operator can be not sufficient, but it is higher when compared with any other classical preconditioning operator (natural, optimal, etc.; see [7]); therefore the superoptimal approximation can be used as a starting point and can be favorably combined with a Tykhonov-like shift of the spectrum (as in [21]) or with the stabilization idea used in [22, 18].

(iii) The computational cost is $O(N \log N)$, which is the minimal one when dealing with trigonometric algebras.

(iv) In some cases, such as large binocular telescope (LBT) applications [15], the superoptimal preconditioning is enough for obtaining qualitatively good images.

5. Conclusions. In this paper we reported a detailed analysis of the superoptimal preconditioning strategy in the 2-D setting. A theoretical analysis and a wide numerical experimentation have shown that the superoptimal preconditioner, in the presence of moderate levels of noise, can represent a good basis for the fast PCG solution of some image deblurring problems, avoiding the need of finding additional regularization parameters. If the level of noise is higher, then the quality of the reconstruction deteriorates, and thus other, more specialized techniques should be used: higher order filtering [12] or stabilized preconditioning [19]. Many open questions remain:

(i) how to insert the information regarding the specific kind of noise carried in the known data vector, in the cases where the regularization effect is not sufficient (in this respect there is an interesting proposal by Calvetti, Reichel, and Shuibi [3] concerning enriched Krylov subspace methods); future work will concern this issue by also using the extension of the notion of superoptimal approximation to a parameter-dependent regularizing family introduced in [12, 13];

(ii) how to impose nonnegativity constraints and other boundary conditions in order to reduce artifacts caused by ringing effects (see, e.g., [24, 28]);

(iii) to investigate if other preconditioned Krylov subspace methods (GMRES, MINRES, etc.) can improve the quality of the reconstructed images with respect to the PCG techniques considered in this paper.

REFERENCES

[1] M. BERTERO AND P. BOCCACCI, *Introduction to Inverse Problems in Imaging*, Institute of Physics, Bristol, 1998.

[2] D. CALVETTI, B. LEWIS, AND L. REICHEL, *GMRES, L-curves, and discrete ill-posed problems*, BIT, 42 (2002), pp. 44–65.

[3] D. CALVETTI, L. REICHEL, AND A. SHUIBI, *Enriched Krylov subspace methods for ill-posed problems*, Linear Algebra Appl., 362 (2003), pp. 257–273.

[4] R. H. CHAN AND X.-Q. JIN, *A family of block preconditioners for block systems*, SIAM J. Sci. Statist. Comput., 13 (1992), pp. 1218–1235.

[5] R. H. CHAN, X.-Q. JIN, AND M.-C. YEUNG, *The spectra of super-optimal circulant preconditioned Toeplitz systems*, SIAM J. Numer. Anal., 28 (1991), pp. 871–879.

[6] 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.

[7] R. H. CHAN AND M. K. NG, *Conjugate gradient methods for Toeplitz systems*, SIAM Rev., 38 (1996), pp. 427–482.

[8] T. F. CHAN, *An optimal circulant preconditioner for Toeplitz systems*, SIAM J. Sci. Statist. Comput., 9 (1988), pp. 766–771.

[9] F. DI BENEDETTO AND S. SERRA CAPIZZANO, *Optimal multilevel matrix algebra operators*, Linear and Multilinear Algebra, 48 (2000), pp. 35–66.

[10] F. DI BENEDETTO AND S. SERRA CAPIZZANO, *A note on the superoptimal matrix algebra operators*, Linear and Multilinear Algebra, 50 (2002), pp. 343–372.

[11] L. ELDÉN, *An algorithm for the regularization of ill-conditioned, banded least squares problems*, SIAM J. Sci. Statist. Comput., 5 (1984), pp. 237–254.

[12] C. ESTATICO, *A class of filtering superoptimal preconditioners for highly ill-conditioned linear systems*, BIT, 42 (2002), pp. 753–778.

[13] C. ESTATICO, *Preconditioners for Ill-Conditioned Toeplitz Matrices with Differentiable Generating Functions*, Tech. Rep. 483, Dipartimento di Matematica, Università di Genova, 2003.

[14] C. ESTATICO, *A Fast Algorithm for the Computation of Multilevel Filtering Superoptimal Preconditioners*, Tech. Rep. 489, Dipartimento di Matematica, Università di Genova, 2003.

[15] C. ESTATICO, *Regularized Fast Deblurring for the Large Binocular Telescope*, Tech. Rep. 490, Dipartimento di Matematica, Università di Genova, 2003.

[16] B. GRASSO, *Tecniche di Precondizionamento per la Risoluzione Numerica di Sistemi Lineari tramite Proiezione su Particolari Algebre Matriciali*, Graduate Thesis in Mathematics, University of Genova, 1995.

[17] U. GRENANDER AND G. SZEGÖ, *Toeplitz Forms and Their Applications*, 2nd ed., Chelsea, New York, 1984.

[18] M. HANKE AND J. NAGY, *Inverse Toeplitz preconditioners for ill-posed problems*, Linear Algebra Appl., 284 (1998), pp. 137–156.

[19] M. HANKE, J. NAGY, AND R. PLEMMONS, *Preconditioned iterative regularization for ill-posed problems*, in Numerical Linear Algebra (Kent, OH, 1992), de Gruyter, Berlin, 1993, pp. 141–163.

[20] T. KAILATH AND V. OLSHEVSKY, *Displacement structure approach to discrete-trigonometric-transform based preconditioners of G. Strang type and T. Chan type*, Calcolo, 33 (1996), pp. 191–208.

[21] J. KAMM AND J. NAGY, *Kronecker product and SVD approximation in image restoration*, Linear Algebra Appl., 284 (1998), pp. 177–192.

[22] M. E. KILMER, *Cauchy-like preconditioners for two-dimensional ill-posed problems*, SIAM J. Matrix Anal. Appl., 20 (1999), pp. 777–799.

[23] T. KU AND C.-C. J. KUO, *On the spectrum of a family of preconditioned block Toeplitz matrices*, SIAM J. Sci. Statist. Comput., 13 (1992), pp. 948–966.

[24] 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.

[25] R. PLATO, *Optimal algorithms for linear illposed problems yield regularization methods*, Numer. Funct. Anal. Optim., 11 (1990), pp. 111–118.

[26] M. C. ROGGEMANN AND B. WELSH, *Imaging through Turbulence*, CRC Press, Boca Raton, FL, 1996.

[27] S. SERRA CAPIZZANO, *Generalized locally Toeplitz sequences: Spectral analysis and applications to discretized partial differential equations*, Linear Algebra Appl., 366 (2003), pp. 371–402.

[28] S. SERRA CAPIZZANO, *A note on antireflective boundary conditions and fast deblurring methods*, SIAM J. Sci. Comput., 25 (2003), pp. 1307–1325.

[29] G. STRANG, *A proposal for Toeplitz matrix calculations*, Stud. Appl. Math., 74 (1986), pp. 171–176.

[30] E. E. TYRTYSHNIKOV, *Optimal and superoptimal circulant preconditioners*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 459–473.

[31] E. E. TYRTYSHNIKOV, *Circulant preconditioners with unbounded inverses*, Linear Algebra Appl., 216 (1995), pp. 1–23.

[32] A. ZYGMUND, *Trigonometric Series*, Cambridge University Press, Cambridge, UK, 1959.