Electronic Transactions on Numerical Analysis.
Volume 29, pp. 163-177, 2008.

ISSN 1068-9613.

ETNA
Kent State University
etna@mcs.kent.edu

FILTER FACTOR ANALYSIS OF AN ITERATIVE MULTILEVEL REGULARIZING METHOD*

MARCO DONATELLI† AND STEFANO SERRA-CAPIZZANO†

Abstract. Recent results have shown that iterative methods of multigrid type are very precise and efficient for regularizing purposes: the reconstruction quality is of the same level or slightly better than that related to most effective regularizing procedures such as Landweber or conjugate gradients for normal equations, but the associated computational cost is highly reduced. Here we analyze the filter features of one of these multigrid techniques in order to provide a theoretical motivation of the excellent regularizing characteristics experimentally observed in the discussed methods.

Key words. regularization, early termination, filter analysis, boundary conditions, structured matrices

AMS subject classifications. 65Y20, 65F10, 15A12

1. Introduction. We consider the classical de-blurring problem of noisy and blurred signals or images. It is usually modeled as a first kind integral equation which, after discretization and with the imposition of proper boundary conditions (BCs), results in a linear system of the form

$$A\mathbf{x} + \nu = \mathbf{b}. \tag{1.1}$$

The vector $\mathbf{x}$ represents the unknown true object, ν the noise, $\mathbf{b}$ the observed object (the blurred noisy version of $\mathbf{x}$) and A is modelling the blurring phenomena that we assume to be spatially invariant. Moreover, for the sake of notational simplicity, we assume that every involved object has the same size in each direction, and hence $\mathbf{x},\ \mathbf{b},\ \nu \in \mathbb{R}^{n^d}$, while $A \in \mathbb{R}^{n^d \times n^d}$ for a d-dimensional problem. The matrix A has a special d-level structure depending on the imposed BCs (see [9, 12] and references therein), e.g., for zero Dirichlet BCs it is a d-level Toeplitz matrix, see [7].

Since it is not ensured that A is nonsingular, instead of solving (1.1), usually we look for the minimal norm argument of the least square problem

$$\min_{\mathbf{y}\in\mathbb{R}^{n^d}} \|A\mathbf{y} - \mathbf{b}\|_2, \tag{1.2}$$

where $\|\cdot\|_2$ is the vector l_2 Euclidean norm (i.e. $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2}$ for any n-sized vector $\mathbf{x}$). Since we assume the presence of noise and since equation (1.1) arises from the approximation of an ill-posed problem, some regularization is needed also in the formulation (1.2) (see [3, 6]) in order to avoid the magnification of noise-dependent high frequency errors. In several applications the global size n^d is large, especially for $d \geq 2$, and consequently the use of direct methods is not recommended, unless the matrix A belongs to a matrix algebra related to a fast transform (see e.g. [12]). On the other hand, the matrix-vector product can be computed rapidly via the fast Fourier transform for each kind of BCs [8, 4]. Therefore it is attractive to employ iterative methods, enforcing regularization by early termination of the iterations.

In a recent paper [5], we have shown that iterative methods of multigrid type are very effective for regularizing purposes; for a different multigrid regularization of cascadic type

*Received October 23, 2006. Accepted for publication December 17, 2007. Published online on August 9, 2008. Recommended by A. Frommer.

†Dipartimento di Fisica e Matematica, Università dell'Insubria - Sede di Como, Via Valleggio 11, 22100 Como, Italy (marco.donatelli@uninsubria.it, stefano.serrac@uninsubria.it).

ETNA
Kent State University
etna@mcs.kent.edu

see [10]. In [5], under a significant quantity of noise, we observed that the reconstruction quality is of the same level or slightly better than that related to the most established regularizing procedures such as Landweber or the conjugate gradient for normal equations (CGNR) and we proved that the associated cost is highly reduced, when compared with these more classical techniques.

In this paper, we analyze the filter features, in the spirit of [7], of the simplest of the multigrid techniques proposed in [5] in order to provide a theoretical motivation for the excellent regularizing characteristics experimentally observed in that paper. For the sake of notational simplicity, we consider in detail the 1D case (i.e., $d = 1$) with periodic BCs so that the matrix A is circulant (see, e.g., [3, 7]). Furthermore, we do not assume symmetry in the PSF (i.e., the generating function or symbol is complex valued).

The extension to the d-dimensional case is a simple application of the same tools. On the other hand, the zero Dirichlet BCs case is a bit more involved since we cannot diagonalize the related matrix in general: moreover, under the assumption of Dirichlet BCs, a pure Toeplitz structure arises and indeed a similar analysis can be carried out by using the symbol (see [16]). Furthermore, it should be observed that the simplified periodic assumption is classical when analyzing multigrid methods [15], especially for partial differential equations, and leads to the celebrated local Fourier analysis [17]. In order to maintain simple notation, we consider in detail the Two-Level (TL) method (see Section 3 for the algorithm) which is the simplest multigrid-type technique considered in [5]. This is equivalent to considering the two-grid case in the usual multigrid approach. We recall that any of these multigrid-type methods can be employed with any auxiliary iterative regularizing method, used as smoother. For the sake of clarity and owing to our image restoration context, we restrict our attention to the Landweber method.

The paper is organized as follows. In Section 2 we define the filter factor of the classical Landweber method, while Section 3 contains a concise description of the TL method (a prototype of the multigrid-type regularizing methods in [5]) and the derivation of its filter factor. In Section 4 we critically discuss the results and, by comparing the associated filter factors, we demonstrate that the TL is a regularizing method under suitable choices of prolongation and restriction operators. A further analysis about the regularization behavior of the TL method is done in Section 5. Finally Section 6 is concerned with some numerical experiments and Section 7 is devoted to a brief discussion on open problems, extensions (different boundary conditions, multidimensional setting, spatially varying PSFs etc.), and conclusions.

2. The filter factor of the Landweber method. The Landweber method, with initial guess $\mathbf{x}_0 = \mathbf{0}$, can be applied to finding the solution of (1.2) and is defined by

$$\mathbf{x}_{j+1} = \mathbf{x}_j + \alpha A^T(\mathbf{b} - A\mathbf{x}_j), \qquad j = 0, 1, \ldots$$

with $0 < \alpha < 2/\|A^T A\|_2$. Thanks to the PSF normalization condition $\|A^T A\|_2 \le 1$ we can choose $0 < \alpha < 2$ (for the method to be convergent) and, in order to keep simple notations, we set $\alpha = 1$.

With the choice $\mathbf{x}_0 = \mathbf{0}$, it is immediate to verify that

$$\mathbf{x}_j = \sum_{i=0}^{j-1} (I - A^T A)^i A^T \mathbf{b}. \tag{2.1}$$

Let $A = U\Sigma V^T$ be the singular value decomposition (SVD) of A with the singular values $\sigma_1 = \|A\|_2 \ge \sigma_2 \ge \cdots \ge \sigma_n > 0$; see [7]. It follows that

$$\mathbf{x}_j = \sum_{i=1}^{n} (1 - (1 - \sigma_i^2)^j) \frac{\mathbf{u}_i^T \mathbf{b}}{\sigma_i} \mathbf{v}_i,$$

ETNA
Kent State University
etna@mcs.kent.edu

where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the columns of U and V, respectively. For fixed j, and by varying $i = 1, \ldots, n$, the factor $1-(1-\sigma_i^2)^j$ is called "the filter factor" (see e.g. [7]). By comparison with the exact solution, it is clear that the filter factor defines the regularizing properties of the method, i.e., the filtering of the noise. Indeed,

$$\mathbf{x}^* = A^\dagger \mathbf{b} = \sum_{i=1}^{n} \frac{\mathbf{u}_i^T \mathbf{b}}{\sigma_i} \mathbf{v_i}$$

is the minimal (Euclidean) norm solution of (1.2): nevertheless, the noise, which affects the vector $\mathbf{b}$ in the high frequency subspace (in our case the vectors $\mathbf{u}_i$ and $\mathbf{v}_i$ related to the small singular values), produces a great amplification of the noise itself thanks to the latter formula for $\mathbf{x}^*$. The truncated SVD (TSVD, see [3, 7]) solves this problem by ignoring the singular values below a certain threshold (related to the norm of the noise which generally is an available information). However, this binary cut is too strong and some information about the edges could be lost. The factor $1-(1-\sigma_i^2)^j$ represents a smoother filtering since it is close to zero if σ_i is close to zero (the related left and right singular vectors are of high frequency type), while it is, as desired, close to one if σ_i is close to one (the related left and right singular vectors are of low frequency type). Moreover, the number of the frequencies used in the reconstruction grows as the iteration parameter j grows. This explains the classical semi-convergence behavior and the role of the iteration count as regularizing parameter (see Figure 4.2).

In the case of periodic BCs we have $A = \mathcal{C}_n(f)$, i.e., A is a circulant matrix of size n generated by the function f whose Fourier coefficients form the PSF. It is known that $A = \mathcal{C}_n(f) = F_n D_n(f) F_n^H$, where $F_n = [\mathrm{e}^{\mathrm{i} j x_k}]_{k,j=0}^{n-1}/\sqrt{n}$ is the discrete Fourier transform (DFT) matrix and $D_n(f) = \mathrm{diag}([f(x_k)]_{k=0}^{n-1})$ with $x_k = \frac{2\pi k}{n}$. By repeating the previous reasoning and taking into account the circulant structure, we deduce that (2.1) is equivalent to

$$\mathbf{x}_j = F_n \sum_{i=0}^{j-1} (I - D_n(|f|^2))^i D_n(\bar{f}) F_n^H \mathbf{b} \;=\; C_n(g)\mathbf{b} \;=\; C_n(\phi_j) C_n^{-1}(f)\mathbf{b}, \tag{2.2}$$

where $g = \phi_j / f$ and $\phi_j(x) = 1-(1-|f(x)|^2)^j$, $x \in (0, 2\pi]$. In other words $\phi_j(x)$ compactly defines the Landweber filter factor in functional terms. It should be acknowledged that the invertibility of A is not necessary: in the case of a singular A, which implies a vanishing symbol f, we have $g(x) = f^+(1-(1-|f|^2)^j)$ where $f^+ = 1/f$ if $|f| \neq 0$ and zero otherwise. The latter represents a nice correspondence between the matrix pseudo-inverse (see [3, 6]) operation $A^\dagger$ and the functional pseudo-inverse operation f^+.

3. Filter factor estimation of the TL. The TL method [5] is based on the following scheme:

1. Projection of the deblurring problem in a half-sized subspace mainly formed by low and middle frequencies in order to filter the noise (and, by the way, in order to achieve the (optimal) linear complexity in the arising V-cycle [15], when applied to band-structures). The main tool for the projection is the use of a low-pass filter obtained by a centered averaging.
2. Application of an iterative regularizing method (Landweber in our case) at the lower level, i.e., to the problem of size $n/2$.
3. Interpolation of the lower level solution in order to define a full size approximation (i.e., of dimension n) and to apply an early stopping criterium such as the discrepancy principle [3, 6].

ETNA
Kent State University
etna@mcs.kent.edu

More specifically, starting from an initial approximation $\mathbf{x}_n \in \mathbb{R}^n$, one TL iteration provides a new approximation $\mathbf{y}_n \in \mathbb{R}^n$ according to:

$$\begin{array}{l} \mathbf{y}_n := \mathsf{TL}(\mathbf{x}_n, \mathbf{b}_n, S_n, \beta) \\ \hline \mathbf{b}_{n/2} := R_n(\mathbf{b}_n - A_n\mathbf{x}_n) \\ A_{n/2} := P_n A_n R_n \\ \mathbf{y}_{n/2} := \mathsf{Smooth}\big(\mathbf{0}_{n/2}, \mathbf{b}_{n/2}, S_{n/2}, \beta\big) \\ \mathbf{y}_n \quad := \mathbf{x}_n + P_n \mathbf{y}_{n/2} \end{array} \tag{3.1}$$

where S_k defines the regularizing iterative procedure at the dimension k, which is used as smoother (Landweber in our setting). By $\mathsf{Smooth}\big(\mathbf{0}_{n/2}, \mathbf{b}_{n/2}, S_{n/2}, \beta\big)$ we denote the application of β steps of the smoother $S_{n/2}$ with initial guess $\mathbf{0}_{n/2}$ and right-hand side $\mathbf{b}_{n/2}$. Finally, the matrices $P_n \in \mathbb{R}^{n\times \frac{n}{2}}$ and $R_n \in \mathbb{R}^{\frac{n}{2}\times n}$ represent the operators (see [15]) of prolongation and restriction, respectively.

In [5] we proved that applying j TL iterations with $\beta = 1$ is equivalent to applying a single TL iteration with $\beta = j$. By setting $\mathbf{x}_0 = \mathbf{0}$, we have

$$\mathbf{x}_j = P_n G_{\frac{n}{2}} R_n \mathbf{b}, \tag{3.2}$$

with $G_{\frac{n}{2}}$ being the smoother iteration matrix (Landweber iteration matrix in our specific case) at the lower level. More precisely $G_{\frac{n}{2}} = \sum_{i=0}^{j-1} (I - A_{\frac{n}{2}}^T A_{\frac{n}{2}})^i A_{\frac{n}{2}}^T$. For the definition and the associated computational issues of the different matrix structures involved in the algebraic multigrid with circulant pattern; see [14, 1, 2]. In the following, we will recall explicitly only the properties that are necessary for handling our problem. We define the "cutting matrix" $K_n = I_{n/2} \otimes [1,\ 0] \in \mathbb{R}^{\frac{n}{2}\times n}$, where $\otimes$ is the tensor product, i.e., for $\mathbf{s} \in \mathbb{R}^{\frac{n}{2}}$ and $\mathbf{t} \in \mathbb{R}^n$, $\mathbf{s} = K_n\mathbf{t}$ implies that $\mathbf{s}_i = \mathbf{t}_{2i}$, $i = 1, \dots, n/2$. The matrices P_n and R_n are defined as $P_n = \mathcal{C}_n(p)K_n^T$ and $R_n = K_n\mathcal{C}_n(r)$, where p and r are trigonometric polynomials. Since P_n and R_n are chosen to be the operators of linear interpolation and weighted averaging, respectively, the functions p and r are

$$r(x) = (1 + \cos(x))/2, \qquad p(x) = 2r(x). \tag{3.3}$$

REMARK 3.1. It should be acknowledged that the functional interpretation is more general than the classical interpretation provided by geometric multigrid; see also Remark 3.2. Indeed, all convergence and filter properties of these multigrid-type methods are related to the analytical behavior of the functions p and r: in this specific setting it is important that p and r are decreasing in $(0, \pi]$ and have a unique zero at π. In the case of the choices in (3.3), we have also an interpretation in terms of interpolation and averaging, but this is no longer true when considering more general (and often more effective) choices such as, e.g., $r(x) = \gamma_1(x)(1 + \cos(x))^{\beta_1}$, $p(x) = \gamma_2(x)(1 + \cos(x))^{\beta_2}$, β_1, β_2 positive integers, γ_1, γ_2 strictly positive trigonometric polynomials.

Now we show that $G_{\frac{n}{2}} = \mathcal{C}_{\frac{n}{2}}(\hat g)$ for some explicitly computable trigonometric polynomial $\hat g$. We notice that $A_{\frac{n}{2}} = \mathcal{C}_{\frac{n}{2}}(\hat f)$ (see [11, 14]), where

$$\hat f(x) = \frac{1}{2}\left[(prf)\left(\frac{x}{2}\right) + (prf)\left(\pi + \frac{x}{2}\right)\right], \qquad x \in (0, 2\pi].$$

Hence, from (2.2), we infer $\hat g(x) = (1 - (1 - |\hat f(x)|^2)^j)/\hat f(x)$, $x \in (0, 2\pi]$. Taking into account (3.2), we finally have $\mathbf{x}_j = \mathcal{C}_n(p)K_n^T\mathcal{C}_{\frac{n}{2}}(\hat g)K_n\mathcal{C}_n(r)\mathbf{b} = B_n\mathbf{b}$ with

$$B_n = \mathcal{C}_n(p)K_n^T\mathcal{C}_{\frac{n}{2}}(\hat g)K_n\mathcal{C}_n(r). \tag{3.4}$$

The purpose now is to analyze how the TL iteration filters every single frequency and, therefore, we have to study the spectral behavior of B_n with regard to the basic frequencies $F_n^{(k)} = [\mathrm{e}^{\mathrm{i}jx_k}]_{j=0}^{n-1}/\sqrt{n}$, for $k = 0, \dots, n-1$ (which are just the columns of F_n). Since $K_n F_n = [1,\ 1] \otimes F_{\frac{n}{2}}/\sqrt{2}$ (see [14]), from (3.4) we find

$$
\begin{aligned}
B_n &= \frac{1}{2} F_n D_n(p) \begin{bmatrix} D_{\frac{n}{2}}(\hat g) & D_{\frac{n}{2}}(\hat g) \\ D_{\frac{n}{2}}(\hat g) & D_{\frac{n}{2}}(\hat g) \end{bmatrix} D_n(r) F_n^H \\
&= \frac{1}{2} F_n \begin{bmatrix} D_n^{(1)}(p) D_{\frac{n}{2}}(\hat g) D_n^{(1)}(r) & D_n^{(1)}(p) D_{\frac{n}{2}}(\hat g) D_n^{(2)}(r) \\ D_n^{(2)}(p) D_{\frac{n}{2}}(\hat g) D_n^{(1)}(r) & D_n^{(2)}(p) D_{\frac{n}{2}}(\hat g) D_n^{(2)}(r) \end{bmatrix} F_n^H, \qquad (3.5)
\end{aligned}
$$

where $D_n(h) = \mathrm{diag}([D_n^{(1)}(h),\ D_n^{(2)}(h)])$, with $D_n^{(j)}(h) \in \mathbb{R}^{\frac{n}{2} \times \frac{n}{2}}$ and $j = 1, 2$. By defining the permutation matrix $\Pi_n = [K_n^T,\ T_n^T]$ with $T_n = I_{n/2} \otimes [0,\ 1] \in \mathbb{R}^{\frac{n}{2} \times n}$, relation (3.5) transforms into $B_n = F_n \Pi_n^T W_n \Pi_n F_n^H$. Here W_n is the diagonal block matrix of size $(n/2) \times (n/2)$ with blocks of dimension 2×2. For $k = 0, \dots, n/2 - 1$, the k-th diagonal block is given by

$$
W_n^{(k)} = \frac{1}{2} \hat g(x_{2k}) \begin{bmatrix} p(x_k) \\ p(x_{(k+n/2)}) \end{bmatrix} \begin{bmatrix} r(x_k) & r(x_{(k+n/2)}) \end{bmatrix}.
$$

The block $W_n^{(k)}$ has structurally at most rank 1 and therefore we find the trivial null eigenvalue and the nontrivial one λ_k, which coincides with the trace:

$$
\lambda_k = \frac{1}{2} \hat g(x_{2k}) \left((pr)(x_k) + (pr)(x_{(k+n/2)}) \right).
$$

We now determine the corresponding eigenvectors. By definition of p and r in (3.3), $W_n^{(k)}$ is real symmetric and consequently the corresponding two eigenvectors are orthogonal. The two frequencies associated with $W_n^{(k)}$ are the $2k$-th and $(2k+1)$-th columns of $F_n \Pi^T$, $k = 0, \dots, n/2 - 1$, i.e., $F_n^{(k)}$ and $F_n^{(k+n/2)}$. We first derive the eigenvector of $W_n^{(k)}$, which corresponds to the zero eigenvalue. From trivial algebraic manipulations, we deduce $\mu_0^{(k)} = [-1,\ r(x_k)/r(x_{(k+n/2)})]^T$. Such an eigenvector is determined by the following linear combination of frequencies:

$$
\frac{r(x_k)}{r(x_{(k+n/2)})} F_n^{(k+n/2)} - F_n^{(k)}. \qquad (3.6)
$$

Therefore, for a choice of small k ($k = 0, \dots, n/2 - 1$, k is small if and only if $k/n \ll 1$), the dominating frequency is $F_n^{(k+n/2)}$ (high frequency component), while when k grows and is comparable with $n/2$, the component $F_n^{(k)}$ becomes dominating (again high frequency component). The latter discussion shows that the null eigenvalue is always connected to high frequencies (middle frequencies when k is in a neighborhood of $n/4$) and therefore it is a first theoretical evidence of the filtering features of the TL iteration. To complete the analysis, the eigenvector associated to the nontrivial eigenvalue λ_k has to be studied. Its precise expression is given by $\mu_{\lambda_k}^{(k)} = [1,\ p(x_{(k+n/2)})/p(x_k)]^T$ and it can be expressed in terms of the subsequent linear combination of basic frequencies:

$$
F_n^{(k)} + F_n^{(k+n/2)} \frac{p(x_{(k+n/2)})}{p(x_k)}. \qquad (3.7)
$$

By varying k in the range $\{0, \dots, n/2-1\}$, we observe a complementary behavior. For small values of k is $F_n^{(k)}$, while for large values of k the dominating frequency is $F_n^{(k+n/2)}$. This shows that the nontrivial eigenvalue is always associated to low and middle frequencies and, again, this is no surprise given the orthogonality of the two eigenvectors.

REMARK 3.2. Equations (3.6) and (3.7) implicitly define a condition for the choice of r and p, independently (and more generally) of geometric considerations. In fact, in order to force the null eigenvalue to be associated to high and middle frequencies, the functions p and r should have maximal value at the origin (or close to it) and should decrease to zero at π (this implies the classical shape of a function with Fourier coefficients defining a low-pass filter). In addition, their shape (see also Remark 3.1) allows to reinforce or to diminish the filtering characteristics of the TL iteration. Finally, the fact that the functions p and r are chosen as polynomials has nothing to do with the filtering design, but it is only due to computational requirements in order to obtain the optimal cost that characterizes multigrid-type procedures.

Now we come back to a precise analysis of the iterate $\mathbf{x}_j$ and in fact, by exploiting the previous factorization, we obtain

$$\mathbf{x}_j = \sum_{k=0}^{n/2-1} \left[F_n^{(k)},\ F_n^{(k+\frac{n}{2})} \right] Q \begin{bmatrix} 0 & \\ & \lambda_k \end{bmatrix} Q^H \begin{bmatrix} (F_n^{(k)})^H \\ (F_n^{(k+\frac{n}{2})})^H \end{bmatrix} \mathbf{b},$$

where $Q = \left[\frac{\mu_0^{(k)}}{\|\mu_0^{(k)}\|_2},\ \frac{\mu_{\lambda_k}^{(k)}}{\|\mu_{\lambda_k}^{(k)}\|_2} \right]$ is the 2×2 matrix containing the eigenvectors of $W_n^{(k)}$. Therefore, we deduce

$$\begin{aligned}
\mathbf{x}_j &= \sum_{k=0}^{n/2-1} \lambda_k \left(Q_{1,2} F_n^{(k)} + Q_{2,2} F_n^{(k+\frac{n}{2})} \right)^H \mathbf{b} \left(Q_{1,2} F_n^{(k)} + Q_{2,2} F_n^{(k+\frac{n}{2})} \right) \\
&= \sum_{k=0}^{n/2-1} \lambda_k Q_{1,2}^2 f(x_k) \frac{(F_n^{(k)})^H \mathbf{b}}{f(x_k)} F_n^{(k)} + \sum_{k=n/2}^{n-1} \lambda_{k-\frac{n}{2}} Q_{2,2}^2 f(x_k) \frac{(F_n^{(k)})^H \mathbf{b}}{f(x_k)} F_n^{(k)} \\
&\quad + \sum_{k=0}^{n/2-1} \lambda_k Q_{1,2} Q_{2,2} ((F_n^{(k)})^H \mathbf{b} F_n^{(k+n/2)} + (F_n^{(k+n/2)})^H \mathbf{b} F_n^{(k)}) \\
&\approx \sum_{k=0}^{n-1} \theta_j(x_k) \frac{(F_n^{(k)})^H \mathbf{b}}{f(x_k)} F_n^{(k)},
\end{aligned}$$

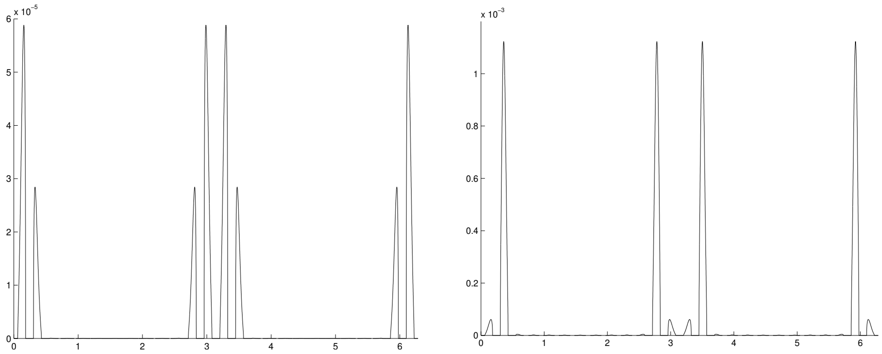
where

$$\theta_j(x_k) = \begin{cases} \lambda_k Q_{1,2}^2 f(x_k), & k = 0, \ldots, \frac{n}{2} - 1, \\ \lambda_{k-\frac{n}{2}} Q_{2,2}^2 f(x_k), & k = \frac{n}{2}, \ldots, n-1. \end{cases}$$

The remainder of the approximation contained in the last step is

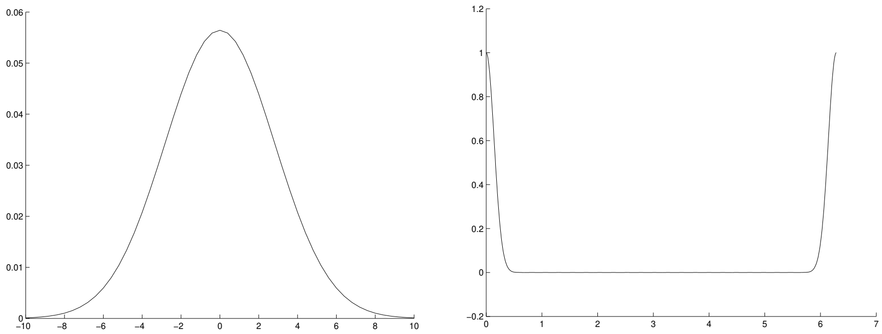
$$\sum_{k=0}^{n-1} \lambda_k Q_{1,2} Q_{2,2} ((F_n^{(k)})^H \mathbf{b} F_n^{(k+n/2)}), \tag{3.8}$$

where we have formally set $\lambda_{k-n/2} := \lambda_k$ for $k = n/2, \ldots, 1$. Since the latter expression is the combination of the k-th and $(k+n/2)$-th frequencies, the quantity in (3.8) can be neglected, because all the functions in the filter have a complementary behavior at x_k and $x_k + \pi$: when one of these functions is large, the other almost vanishes and vice-versa. More specifically, the filter factor associated with (3.8) shows a behavior analogous to $\lambda_k Q_{1,2} Q_{2,2} \sqrt{f(x_k) f(\pi + x_k)}$ which is negligible for every instance of k. Finally, in Figure 3.1 we report the graph of the approximated remainder $\lambda_k Q_{1,2} Q_{2,2} \sqrt{f(x_k) f(\pi + x_k)}$



$j = 10$ (order 10^{-5}) $j = 1000$ (order 10^{-3})

FIG. 3.1. *Graph of the remainder* $\lambda_k Q_{1,2} Q_{2,2} \sqrt{f(x_k) f(\pi + x_k)}$.



Gaussian PSF Generating function $f(x)$

FIG. 4.1. *Gaussian PSF with 51 nodes and related generating function* $f(x)$ *for* $x \in (0, 2\pi]$.

that we decided to neglect (the function f is associated to a prototypal Gaussian PSF; see the next section). It is clear that Figure 3.1 represents numerical evidence of the goodness of the approximation previously proposed, because not only it has small magnitude but also small support, i.e., it is obtained in correspondence of a small number of frequencies.

4. Comparison of the filter factors. We report the graphs of the filter factors for both Landweber and TL methods. In such a way, it is immediate to compare the filtering properties of the two procedures. As an example, let us consider the PSF reported in Figure 4.1 together with its generating function f. In Figure 4.2, for certain values of j, we display the value of $\phi_j(x)$, i.e., the filter factor of Landweber, and the value of $\theta_j(x)$, i.e., the approximation (according to the previous section) of the filter factor of the TL method, for $x \in (0, 2\pi]$. When varying the index j, we observe that the behavior of the TL filter factor follows the same trend of the method used as smoother: in our case the Landweber method is considered, but the same observation can be repeated verbatim for other regularizing smoothers. Furthermore, as it is evident from Figure 4.3, for increasing values of j and with regard to the TL method, the high frequencies contribution is lower than the analogous contribution for Landweber. Such observations perfectly agree with the numerical tests in Section 6. Indeed, for the initial iterations, the reconstruction error decreases in the same manner for both methods. After this phase, the error with Landweber starts to increase while, for the TL method, it keeps going down for some more iterations: this better behavior is essentially due to the higher filtering properties of the TL method when j is large.

ETNA
Kent State University
etna@mcs.kent.edu

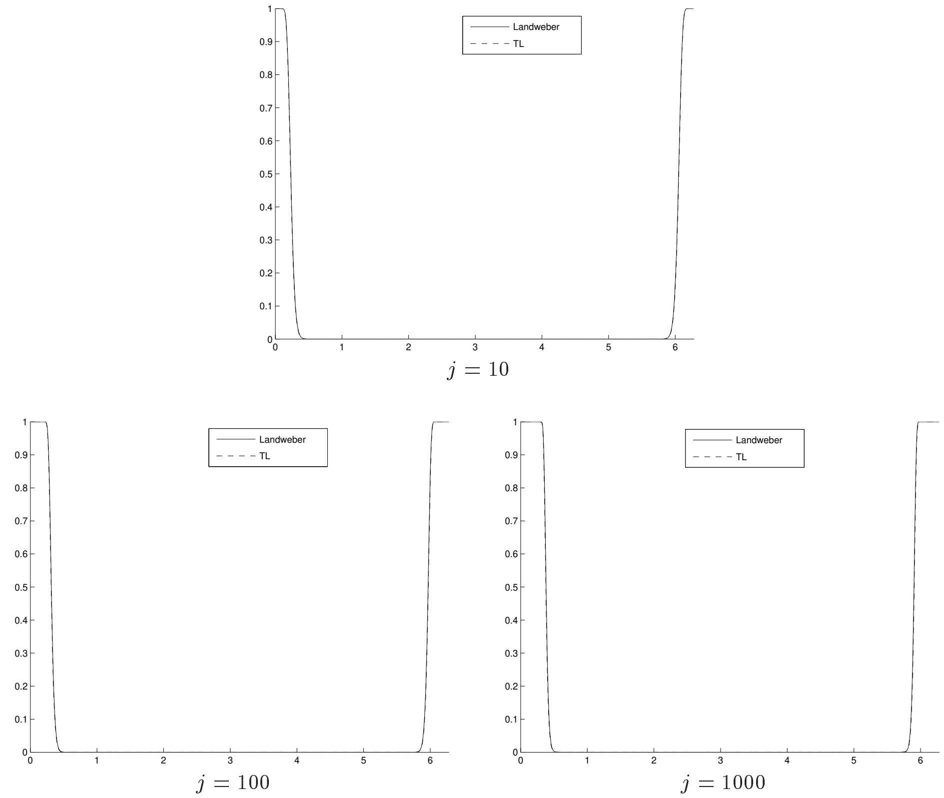


FIG. 4.2. *Filter factor for the Landweber ($\phi_j(x)$) and TL ($\theta_j(x)$) methods for $x \in (0, 2\pi]$ and several instances of j (the two graph are near to be superposed).*

5. A note when the noise level goes to zero. The filter factor analysis in Section 3 shows that in presence of noise, the TL is able to reconstruct the signal filtering the noise, i.e., the TL is a regularizing method. In the literature a regularizing procedure can also studied using the discrepancy principle when the noise level goes to zero [6]. More specifically, let δ be a constant greater than zero such that $\|\nu\|_2 \leq \delta$ and let us assume that the level δ is known. The discrepancy principle is used to stop an iterative method when the current approximation $\mathbf{x}_j$ satisfies the discrepancy principle

$$\|\mathbf{b} - A\mathbf{x}_j\|_2 \leq \alpha\delta$$

for a fixed $\alpha > 1$ independent of δ. Let $\mathbf{x}_{j(\delta)}$ be the first approximation that satisfies this criterium. An iterative method equipped with this stopping rule is said to be a regularization method if the computed approximation $\mathbf{x}_{j(\delta)}$ satisfies

$$\lim_{\delta \to 0^+} \sup_{\nu \leq \delta} \|\mathbf{x} - \mathbf{x}_{j(\delta)}\|_2 = 0. \tag{5.1}$$

This means that in the limit, when there is no noise and under the assumption that the matrix A is invertible, the regularization method reconstructs exactly the true object $\mathbf{x}$; we notice that this more classical approach [6] is the one followed by Reichel and Shyshkov [10] for analyzing a basic multilevel method of multigrid type.

Unfortunately, according to this formal definition, our TL is not a regularization method. Indeed in the noise free case it is not able to reconstruct the true object $\mathbf{x}$. This follows from

ETNA
Kent State University
etna@mcs.kent.edu

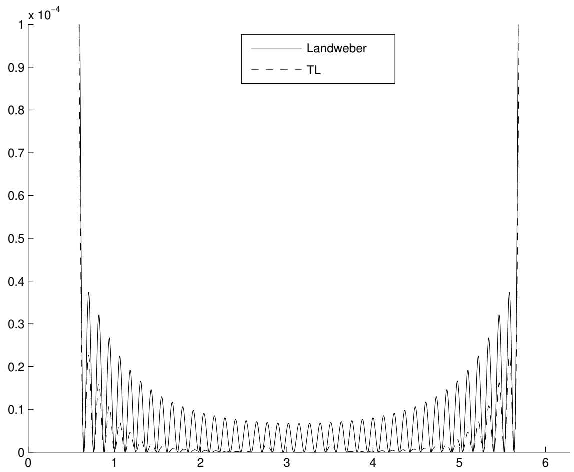


FIG. 4.3. *Filter factor for Landweber ($\phi_j(x)$) and TL ($\theta_j(x)$) method, $x \in (0, 2\pi]$ and $j = 1000$ (note that scale on the y axis equal to 10^{-4}).*

Section 3, where we show that also if A is nonsingular, then the TL operator B_n has a (large) null space of dimension $n/2$. The latter fact means that if the solution $\mathbf{x}$ intersects such a subspace, which is related to high frequencies, then the vector $\mathbf{x}$ cannot be reconstructed exactly. However this limit case ($\delta \to 0^+$) has mainly a purely theoretical impact and it is not usually of interest in the real applications, where we generally encounter nontrivial examples with non-negligible noise levels. We recall that in such case our method is a regularizing method thanks to the analysis in Section 3. Not only this, but the quoted method in (3.1) can be easily modified for low levels of noise in such a way that the resulting algorithm is of regularizing type, according to the framework in [6] as well. For instance, a possibility is to apply the smoother also at the finer level, but with a damping factor depending on the level of noise: in this way we do not lose the good filtering features obtained performing at first a projection in the signal subspace (low frequencies).

In more detail, given an initial approximation $\mathbf{x}_n$, our modified regularizing TL iteration provides a new approximation $\mathbf{y}_n$ according to the following scheme:

$$\mathbf{y}_n := \mathsf{TL}(\mathbf{x}_n, \mathbf{b}_n, S_n, \beta)$$

$$(5.2)\qquad \begin{array}{lll}
\mathbf{b}_{n/2} & := R_n(\mathbf{b}_n - A_n\mathbf{x}_n) & \\
A_{n/2} & := P_n A_n R_n & \\
\mathbf{y}_{n/2} & := \mathsf{Smooth}\big(\mathbf{0}_{n/2}, \mathbf{b}_{n/2}, S_{n/2}, \beta\big) & \\
\mathbf{z}_n & := \mathbf{x}_n + P_n\mathbf{y}_{n/2} & \\
\text{if} & \delta > tol \cdot \|\mathbf{b}_n\|_2 \quad \text{then} & \\
 & \mathbf{y}_n := \mathbf{z}_n & \\
\text{else} & & \\
 & \mathbf{w}_n := \mathsf{Smooth}\big(\mathbf{x}_n, \mathbf{b}_n, S_n, 1\big) & \\
 & \omega := \delta/\|\mathbf{b}_n\|_2 & \\
 & \mathbf{y}_n := \mathbf{w}_n + \omega(\mathbf{z}_n - \mathbf{w}_n) & \\
\text{end if} & &
\end{array}$$

Here the notation is the same as in Algorithm (3.1), tol is a tolerance on the noise level (e.g., $tol = 10^{-3}$) and $\delta/\|\mathbf{b}_n\|_2 \in [0, 1]$ is the percentage of noise. We notice that the smoothing step at the higher dimension n is applied only when the level of the noise is low. In that case,

ETNA
Kent State University
etna@mcs.kent.edu

the whole method behaves as a standard iterative solver used as smoother and this is true especially when $\delta \to 0^+$. The interest of the method relies on the case where the condition $\delta > tol \cdot \|\mathbf{b}_n\|_2$ is satisfied with $tol = 10^{-3}$: in that case, the method filters the noise as we have previously seen since, under this assumption, Algorithm (5.2) reduces to Algorithm (3.1). We notice that the latter case is the only one of interest in real applications. On the other side, the algorithm (5.2) satisfies the condition (5.1) since, when $\delta \to 0^+$, the computed solution tends to the vector computed by the smoother, i.e., to the exact noise-free solution. We recall that, by our assumptions, the smoother is in fact a regularizing method.

Finally, a more precise study, for combining our multigrid technique with methods able to recover the true solution (when the noise goes to zero), will be subject of future researches.

6. Numerical experiments. We present a numerical comparison between the Landweber and the TL methods, where Landweber is used as a smoother. Since we consider real problems with a non-negligible level of noise, we use the TL algorithm (3.1) or, equivalently, algorithm (5.2), where $tol \cdot \|\mathbf{b}_n\|_2$ is lower than the noise percentage used in the experiments. We give two tests, the first one with high noise level and the second with moderate noise, for emphasizing the good regularizing properties of the TL method, independently of the level of the noise corruption.

For the sake of completeness, we report the results also for the multilevel method in the case of one recursive call (V-cycle) or two recursive calls (W-cycle) at every level. For a more complete description of these multilevel methods we refer to [5]. However, it should be stressed that these more sophisticate multilevel procedures derive in a natural way from the TL, by applying recursively the same procedure. In terms of regularization, the above idea means that at the coarser level we apply a smoothing iteration and then we project again into the low frequencies subspace. This can be repeated until we reach a small grid (e.g., 8×8), where the associated linear system can be solved directly, since the problem is computationally negligible and the noise explosion is easily controllable and avoidable.

In order to compare the computational time of the several methods, in [5] we computed the following asymptotic ratios in the 2D case (image deblurring), i.e. for large n, between the cost of one iteration of a multigrid procedure and that of one Landweber iteration: about $1/4$ for the TL iteration, about $1/3$ for the V-cycle, and about 1 for the W-cycle. These ratios can be easily understood by combining the usual computational estimates for multigrid methods and the following remark: at the finest level, the multilevel algorithms perform only a projection (the Landweber method is applied only at the coarser levels) and the problem size after the first projection is $1/2^d$ of the size of the original d-dimensional problem.

In the following experimentation we consider 1D problems (signal deblurring). For a wide experimentation in the 2D case, see [5]. The numerical results are obtained using Matlab 7.0.

6.1. Example 1. We consider the signal test in Fig. 6.1. The observed signal, solid line in Fig. 6.1 (a), is the blurred version of the true signal, dotted line in Fig. 6.1 (a), obtained adding Gaussian blur and 10% of white noise. Since we know the true solution, the methods are stopped at the minimum of the relative restoration errors (RREs) defined as the l_2 norm of the error over the l_2 norm of the true solution. In real applications any classical stopping criterium, like e.g. the discrepancy principle, can be applied both to Landweber and multilevel strategies.

The quality of the restored signals in Fig. 6.1 (b-d) is comparable. In this example the gain in the use of the multilevel methods is mainly in the reduction of the computational cost. Indeed the TL method requires about the same number of iterations of Landweber, but each iteration requires half time. Concerning the V-cycle each iteration costs $2/3$ of one Landweber iteration and moreover it converges faster (within about half iterations).

ETNA
Kent State University
etna@mcs.kent.edu

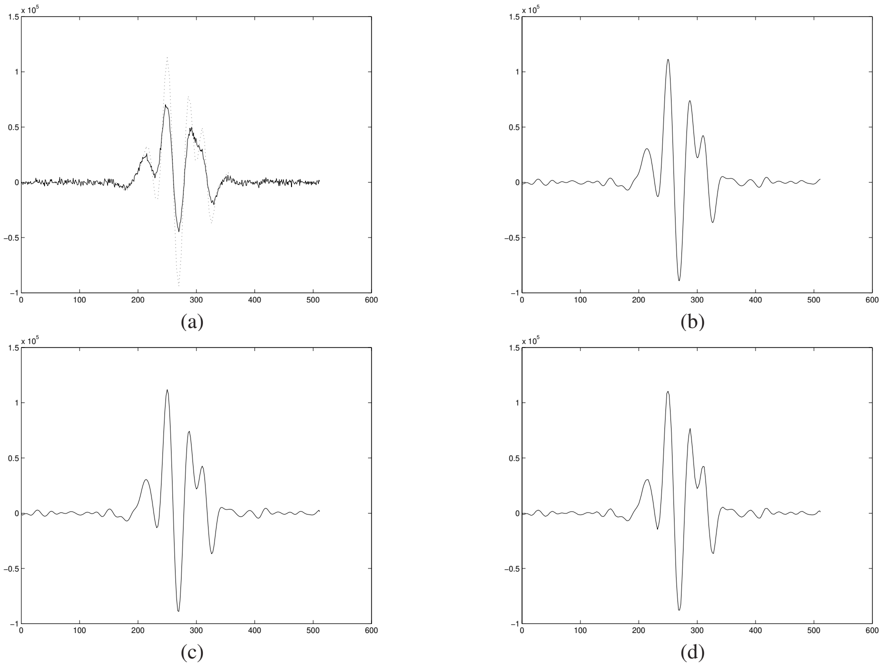


FIG. 6.1. *Test problem: (a) — — — true signal — observed signal with* 10% *of noise, (b) restored signal with Landweber in 39 iterations (RRE = 0.077), (c) restored signal with TL in 42 iterations (RRE = 0.076), (d) restored signal with* V*-cycle in 20 iterations (RRE = 0.078).*

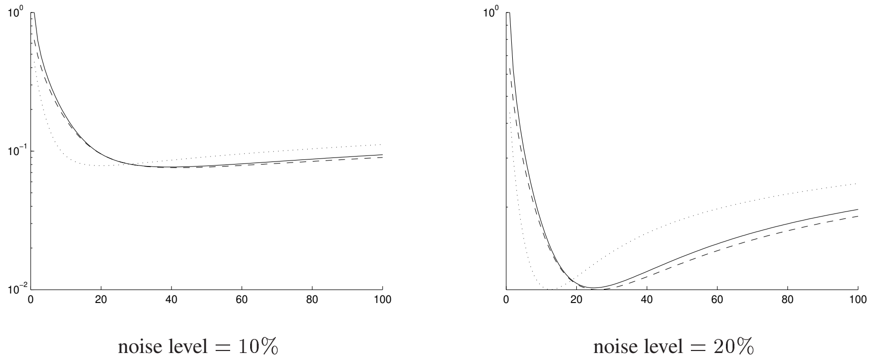


FIG. 6.2. *Relative restorations errors for each iteration (logarithmic scale) for test in Fig. 6.1: — Landweber, — — — TL,* $\cdots$ V*-cycle.*

In Fig. 6.2 the graphs of the RREs are displayed for each iteration of the considered methods. We test also the same signal with a 20% of noise. Similarly to the previous case, we have a considerable gain in time for every multilevel strategy with respect to the Landweber method alone (about half time). Moreover, the graphs in Fig. 6.2 confirm the analysis in Section 4. It is shown that the filter factors of the TL method and the iterative regularizing method used as smoother (Landweber in our case) are essentially the same, even if the TL iteration filters slightly more the high frequencies and indeed it keeps in reducing the noise for more iterations with respect to the smoother. This behavior is evident in the two graphs of Fig. 6.2.

ETNA
Kent State University
etna@mcs.kent.edu

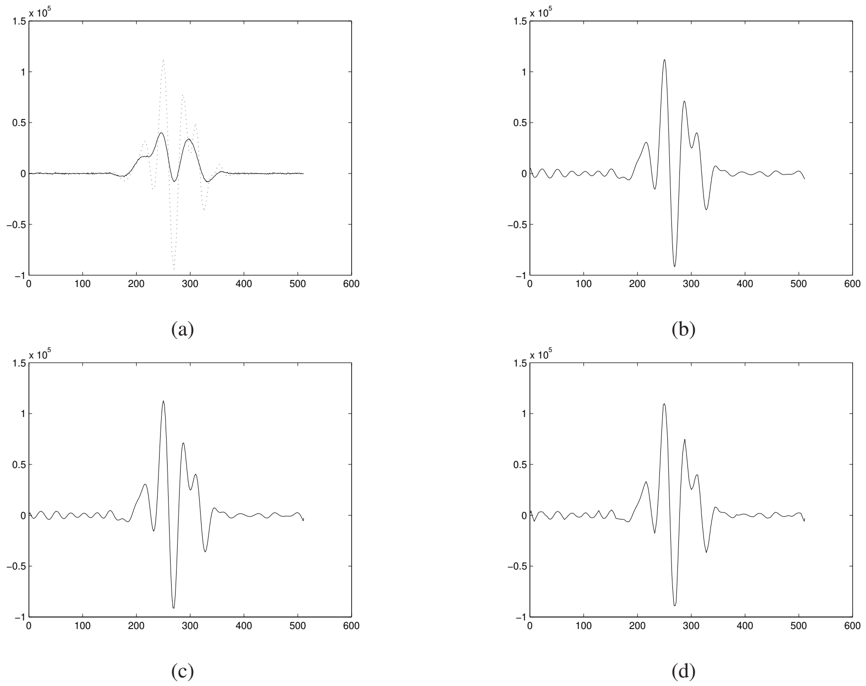


FIG. 6.3. *Test problem: (a) — — — true signal — observed signal with 3% of noise, (b) restored signal with Landweber in 4090 iterations (RRE = 0.113), (c) restored signal with TL in 4587 iterations (RRE = 0.110), (d) restored signal with V-cycle in 1481 iterations (RRE = 0.115).*

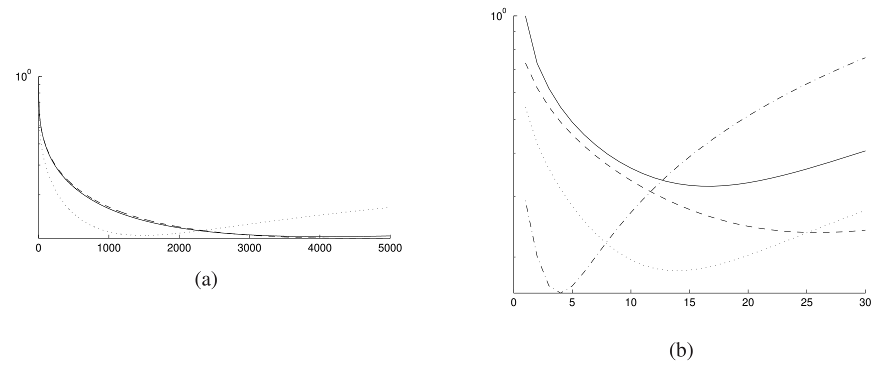


FIG. 6.4. *Relative restorations errors for each iteration (logarithmic scale) for test in Fig.6.3: (a) — Landweber, — — — TL, · · · V-cycle. (b) — Van Cittert, — — — TL, · · · V-cycle, —·— W-cycle.*

6.2. Example 2. The last test problem has a larger PSF and a low level of noise, only 3%. This choice is taken into account in order to show the robustness and the good regularization properties of the multilevel strategies, also for problems with moderate noise level. The true and the observed signals are shown in Fig. 6.3 (a). The numerical results, restored signals in Fig. 6.3 (b-d) and RREs in Fig. 6.4 (a), confirm the previous analysis: the quality of the restored signals are comparable and the TL method improves the regularization of Landweber, while the V-cycle has a faster convergence.

In Fig. 6.3 we can see that Landweber is slowly convergent. Moreover, since the PSF

ETNA
Kent State University
etna@mcs.kent.edu

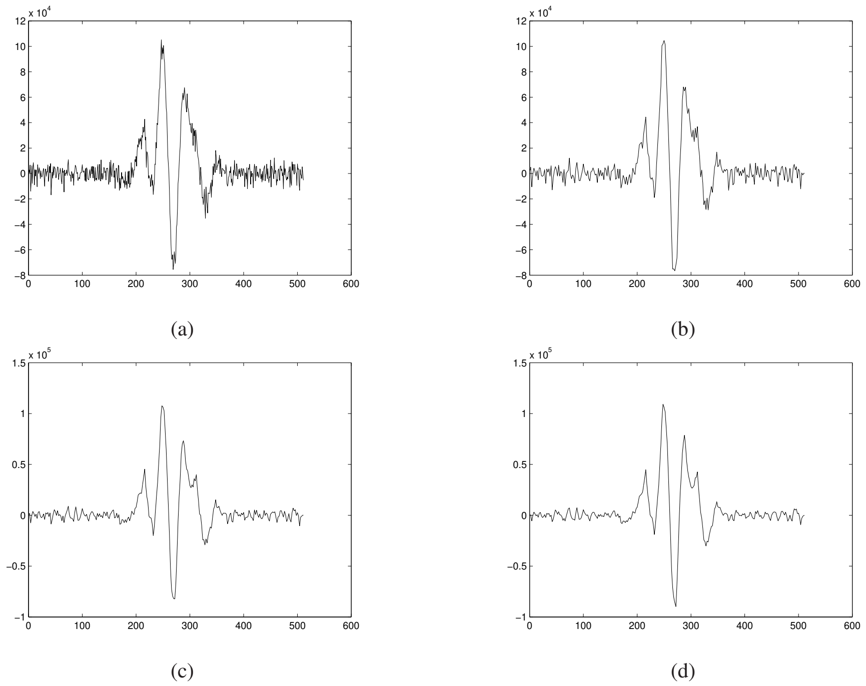


FIG. 6.5. *Restored signals for test in Fig. 6.3: (a) Van Cittert in 17 iterations (RRE = 0.321), (b) TL in 26 iterations (RRE = 0.235), (c) V-cycle in 14 iterations (RRE = 0.182), (d) W-cycle in 4 iterations (RRE = 0.157).*

is a Gaussian blur, the matrix A is positive definite. In such case, instead of Landweber the Van Cittert method [3] can be used. The Van Cittert iteration is defined at the jth step as $\mathbf{x}_{j+1} = \mathbf{x}_j + \tau(\mathbf{b} - A\mathbf{x}_j)$, with $0 < \tau < 2/\|A\|_2$. Notice that no matrix-vector product with A^T is required and, moreover, the convergence is faster than Landweber, but usually the restored quality is lower. In this case, using Van Cittert as smoother, the multilevel strategies can be mainly used for improving the quality of the restoration, as well as the convergence speed. In other words we improve the computational cost of the Van Cittert method and, at the same time, we obtain a substantially more precise reconstruction, comparable with that of the Landweber method.

In Fig. 6.4 (b) and Fig. 6.5 we note the high improvement in the restoration using multilevel strategies. Indeed, the V-cycle in Fig. 6.5 (c) and the W-cycle in Fig. 6.5 (d) are able to recover the rightmost up-peak (around 300) that is not visible in the Van Cittert restoration in Fig. 6.5 (a). Moreover, W-cycle requires only 4 iteration for computing a good enough restoration. This means that the W-cycle with Van Cittert as smoother could be a good method for systems requiring a real-time preview.

Concluding, the proposed tests show that if the smoother is a slowly convergent regularizing method that computes a high quality approximation, then multilevel strategies can be useful for accelerating the convergence, and preserving the quality of the restoration (see Fig. 6.1 and Fig. 6.3). On the other hand, if the smoother is a fast regularizing method that computes a solution of low quality, then the multilevel strategies are able to improve the quality of the restoration, as well as to speed up the convergence (see Fig. 6.5).

7. Discussion, extensions, and conclusions. In this section we briefly discuss the potential of the approach presented so far. In particular we would like to emphasize possible

ETNA
Kent State University
etna@mcs.kent.edu

generalizations of the theoretical analysis.

- The extension to the d-dimensional setting with $d > 1$ does not pose any problem, because of the known block diagonalization of the TL iteration (see e.g. [11, 14, 1, 2] and references there reported): instead of having 2×2 blocks, we obtain blocks of size $2^d \times 2^d$.
- The extension to other BCs is more critical, because we lose the structure of algebra. In other words, the transform that diagonalizes the matrix A is known only in special cases, such as in the case of strongly symmetric PSFs and reflective and anti-reflective BCs (see [9, 12] and references therein).
- The extension to the space varying case poses the same problem as the change of BCs, since the diagonalization of A and therefore the block diagonalization of the corresponding TL iteration cannot be performed in general.

However, recent results in asymptotic linear algebra show that there exists a wide class of matrix sequences (Generalized Locally Toeplitz, [13]) for which the asymptotic spectral behavior can be identified explicitly in terms of generating functions in the sense of spectral distributions (for the notion of spectral distribution see e.g. [16]). For multigrid methods, the symbol is a $2^d \times 2^d$ matrix valued function as shown in Section 3.7 of [13]. Therefore, we believe that the generalizations mentioned in the last two items can be studied properly through these tools, which, as stressed in [13], can be viewed as a generalization of the classical Fourier Analysis to nonconstant coefficient operators.

Acknowledgment. The work of the authors was partially supported by MUR, grant number 2006017542.

REFERENCES

[1] A. ARICÒ AND M. DONATELLI, *A V-cycle Multigrid for multilevel matrix algebras: proof of optimality*, Numer. Math., 105 (2007), pp. 511–547.

[2] A. ARICÒ, M. DONATELLI, AND S. SERRA-CAPIZZANO, *V-cycle optimal convergence for certain (multilevel) structured linear system*, SIAM J. Matrix Anal. Appl., 26 (2004), pp. 186–214.

[3] M. BERTERO AND P. BOCCACCI, *Introduction to Inverse Problems in Imaging*, Institute of Physics Publishing, London, UK, 1998.

[4] M. DONATELLI, C. ESTATICO, A. MARTINELLI, AND S. SERRA-CAPIZZANO, *Improved image deblurring with anti-reflective boundary conditions and re-blurring*, Inverse Problems, 22 (2006), pp. 2035–2053.

[5] M. DONATELLI AND S. SERRA-CAPIZZANO, *On the regularizing power of multigrid-type algorithms*, SIAM J. Sci. Comput., 27 (2006), pp. 2053–2076.

[6] H. ENGL, M. HANKE, AND A. NEUBAUER, *Regularization of Inverse Problems*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.

[7] P. C. HANSEN, J. NAGY, AND D. P. O'LEARY, *Deblurring Images: Matrices, Spectra, and Filtering*, SIAM, Philadelphia, PA, 2006.

[8] J. NAGY, K. PALMER, AND L. PERRONE, *Iterative methods for image deblurring: a Matlab object-oriented approach*, Numer. Algorithms, 36 (2004), pp. 73–93.

[9] M. 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.

[10] L. REICHEL AND A. SHYSHKOV, *Cascadic multilevel methods for ill-posed problems*, J. Comput. Appl. Math., to appear.

[11] S. SERRA-CAPIZZANO, *Convergence analysis of two-grid methods for elliptic Toeplitz and PDEs matrix-sequences*, Numer. Math., 92 (2002), pp. 433–465.

[12] S. SERRA-CAPIZZANO, *A note on anti-reflective boundary conditions and fast deblurring models*, SIAM J. Sci. Comput., 25 (2003), pp. 1307–1325.

[13] S. SERRA-CAPIZZANO, *GLT sequences as a generalized Fourier analysis and applications*, Linear Algebra Appl., 419 (2006), pp. 180–233.

[14] S. SERRA-CAPIZZANO AND C. TABLINO POSSIO, *Multigrid methods for multilevel circulant matrices*, SIAM J. Sci. Comput., 26 (2005), pp. 55–85.

[15] U. TROTTENBERG, C.W. OOSTERLEE, AND A. SCHÜLLER, *Multigrid*, Academic Press, London, 2001.

ETNA
Kent State University
etna@mcs.kent.edu

[16] E. TYRTYSHNIKOV, *A unifying approach to some old and new theorems on distribution and clustering*, Linear. Algebra Appl., 232 (1996), pp. 1–43.

[17] R. WIENANDS AND W. JOPPICH, *Practical Fourier Analysis for Multigrid Methods*, CRC Press, 2004.