

## A MULTIGRID FRAME BASED METHOD FOR IMAGE DEBLURRING\*

ALESSANDRO BUCCINI<sup>†</sup> AND MARCO DONATELLI<sup>‡</sup>

**Abstract.** Iterative soft thresholding algorithms combine one step of Landweber method (or accelerated variants), with one step of thresholding of the wavelet (framelet) coefficients. In this paper, we improve these methods by using the framelet multilevel decomposition for defining a multigrid deconvolution with grid transfer operators given by the low-pass filter of the frame.

Assuming that an estimate of the noise level is available, we combine a recently proposed iterative method for  $\ell_2$ -regularization with linear framelet denoising by soft-thresholding. This combination allows a fast frequency filtering in the Fourier domain and produces a sparse reconstruction in the wavelet domain. Moreover, its employment in a multigrid scheme ensures a stable convergence and a reduced noise amplification. The proposed multigrid method is independent of the imposed boundary conditions and the iterations can be easily projected in a closed and convex set, e.g., the nonnegative cone.

We study the convergence of the proposed algorithm and prove that it is a regularization method. Several numerical results prove that this approach is able to provide highly accurate reconstructions in several different scenarios without requiring the setting of any parameter.

**Key words.** image deblurring, multigrid methods, iterative regularization methods

**AMS subject classifications.** 65F22, 65N55, 65F10, 15B05

**1. Introduction.** Image deblurring is an inverse problem that consists in recovering an unknown image from blurred and noisy measurements. These are obtained as a result of a Fredholm integral of the I kind. Assuming that the blur is spatially invariant, i.e., that it does not depend on the location, the measured data  $g$  can be computed by

$$(1.1) \quad g(x, y) = \int K(s - x, t - y) f(s, t) ds dt = K * f,$$

where  $f$  denotes the true signal and  $K$  is a possibly smooth integral kernel with compact support.  $K$  is usually referred to as point spread function (PSF) since it represents how a single point is spread across its neighborhood. Because  $K$  has compact support, (1.1) is ill-posed [27].

After the discretization of (1.1), using a quadrature formula on a uniform grid, we have to consider that images are available only on finite regions, the field of view (FOV). The values assumed by the blurred image near the boundaries are also determined by data outside the FOV. For simplicity let us assume that the observed image  $\mathbf{B}$  is square with dimensions  $n \times n$  and that the PSF  $\mathbf{H}$  is  $m \times m$  with  $m \leq n$ , thus the true image  $\tilde{\mathbf{X}}$ , which includes all the pixels involved in the definition of  $\mathbf{B}$ , is  $q \times q$  with  $q = n + m - 1$ . By column stacking  $\tilde{\mathbf{X}}$  and  $\mathbf{B}$  the discretized version of (1.1) can be expressed as the linear system of equations

$$(1.2) \quad \tilde{A}\tilde{\mathbf{x}} = \mathbf{b},$$

where  $\tilde{\mathbf{x}}$  and  $\mathbf{b}$  are vectors of dimensions  $q^2 \times 1$  and  $n^2 \times 1$ , respectively, and  $\tilde{A}$  is an underdetermined  $n^2 \times q^2$  rectangular matrix. By imposing boundary conditions (bc's), i.e.,

---

\*Received... Accepted... Published online on... Recommended by....

<sup>†</sup>Dipartimento di Matematica e Informatica  
Università degli Studi di Cagliari  
Via Ospedale 72, 09124 Cagliari (Italy)  
alessandro.buccini@unica.it

<sup>‡</sup>Dipartimento di Scienza e Alta Tecnologia  
Università degli Studi dell'Insubria  
Via Valleggio 11, 22100 Como (Italy)  
marco.donatelli@uninsubria.it

by making assumptions on the behavior of the true image outside the FOV, we can consider only the  $n \times n$  internal points of  $\tilde{\mathbf{X}}$ , denoted by  $\mathbf{X}$ , obtaining a square matrix  $A$  such that

$$(1.3) \quad A\mathbf{x} = \mathbf{b},$$

where  $A \in \mathbb{R}^{N \times N}$  and  $\mathbf{x}, \mathbf{b} \in \mathbb{R}^N$ , with  $N = n^2$ . The choice of the bc's directly influences the structure of  $A$  and some choices let us to use fast transformations for its diagonalization [35].

The discretization process, along with measurement errors, introduces some errors, usually referred to as noise. Thus, the error-free  $\mathbf{b}$  is not available and the real measured signal, denoted by  $\mathbf{b}^\delta$ , is such that

$$(1.4) \quad \|\mathbf{b} - \mathbf{b}^\delta\| \leq \delta,$$

where  $\|\cdot\|$  denotes the Euclidean norm and  $\delta > 0$  is the so-called noise level. Therefore, in practice, instead of solving (1.2), we deal with

$$(1.5) \quad A\mathbf{x} = \mathbf{b}^\delta.$$

Throughout this paper we will assume that a fairly sharp bound for  $\delta$  is known.

Because (1.1) is ill-posed,  $A$  is severely ill-conditioned and may be rank deficient. This, in addition to the unavoidable presence of noise, results in the impossibility of directly solving (1.5); see, e.g., [33]. Thus, regularization methods are essential to compute a good approximation of the true image  $\mathbf{x}^\dagger = A^\dagger \mathbf{b}$ , where by  $A^\dagger$  we denote the Moore-Penrose pseudo-inverse.

One of the most popular regularization method is Tikhonov regularization

$$(1.6) \quad \min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}^\delta\|^2 + \alpha \|\mathbf{x}\|^2,$$

where  $\alpha > 0$  is the regularization parameter. Tikhonov regularization bounds the norm of the solution so that the oscillations, due to the inversion of the noise, are kept under control while ensuring that the reconstructed image fits the measured data. It is well known that Tikhonov regularization leads to over-smoothed solutions and that a poor estimation of the parameter  $\alpha$  may lead to unsatisfactory reconstructions.

The iterative regularization method known as *Iterated Tikhonov (IT)* can be constructed by applying an iterative refinement technique to (1.6); see [27, 33]. Given an initial guess  $\mathbf{x}_0$ , the IT method can be formulated by the following iteration

$$(1.7) \quad \begin{aligned} \mathbf{x}_{k+1} &= \mathbf{x}_k + (A^t A + \alpha_k I)^{-1} A^t (\mathbf{b}^\delta - A\mathbf{x}_k), \\ &= \mathbf{x}_k + A^t (A A^t + \alpha_k I)^{-1} (\mathbf{b}^\delta - A\mathbf{x}_k), \quad k = 0, 1, \dots, \end{aligned}$$

where  $A^t$  denotes the transpose of the matrix  $A$ . IT is an iterative regularization method and hence it displays semi-convergence. During the first iterations the iterates approach the true solution  $\mathbf{x}^\dagger$  and the restoration error decreases until a certain “optimal” iteration is reached. Then, the noise present in  $\mathbf{b}^\delta$  is included in the reconstruction and gets amplified deteriorating, and eventually destroying, the quality of the reconstruction. Therefore, a regularization effect is obtained by early stopping the iteration. It is clear that it is crucial to stop the iterations close to the optimal iteration, whose number is unknown.

Unfortunately, the iteration (1.7) requires at each step the inversion of either  $A^t A + \alpha_k I$  or  $A A^t + \alpha_k I$ , which can be computationally demanding when  $A$  is of large dimensions and cannot be diagonalized by fast transforms. For this reason in [22] the authors developed an iterative method with a nonstationary preconditioner, that can be seen as an approximated

version of IT. In particular, they considered an operator  $C$  which is spectrally equivalent to  $A$  (see Assumption 1) and form the preconditioner at step  $k$  as

$$C^t(CC^t + \alpha_k I)^{-1} \approx A^t(AA^t + \alpha_k I)^{-1},$$

where  $\alpha_k$  is determined by a damped version of the discrepancy principle. In this way, by wisely choosing the structure of  $C$ , the resulting iterative method is fast and computationally cheap without requiring the setting of any parameter. In the following we will refer to this method as *AIT* (*Approximated Iterated Tikhonov*). Similarly, the method proposed in [36] uses an approximation of  $A$  in a small Krylov subspace.

In many applications it is known that  $\mathbf{x}^\dagger \in \Omega$ , where  $\Omega$  is a closed and convex set, e.g., for image deblurring a possible choice for  $\Omega$  is the nonnegative cone. The APIT (Approximated Projected Iterated Tikhonov) algorithm is obtained by projecting at each step the computed approximation in a convex set; see [7].

Multigrid methods are powerful algorithms that are able to achieve fast computations and high accuracy when the main ingredients (smoother and grid transfer operators) are properly combined; see e.g. [47]. They have been initially developed for solving linear systems of equations derived from partial differential equations (PDEs) and later have been successfully applied to more general linear systems [45]. Multigrid methods have already been considered to solve ill-posed problems [14, 15, 24, 34, 38, 39, 44], but usually as solvers for Tikhonov like regularized models. The first attempt of using multigrid methods as iterative regularization methods for image deblurring has been done, to the best of our knowledge, in [23], where the authors combined an iterative regularization method used as pre-smoother with a low-pass filter coarsening. A different multilevel strategy based on the cascadic approach was proposed in [43]. Nonlinear “corrections” to the previous multigrid methods were introduced in [40] using a total variation-type regularization and in [19, 28] combining multigrid and wavelets. More recently, also blind deconvolution has been successfully approached in [29]. Note that, these multigrid methods have been defined to preserve the block Toeplitz with Toeplitz blocks (BTTB) structure of the blurring matrix at each coarser level. This is crucial for the definition of the algorithm and for preserving a fast and simple matrix-vector product.

The main novelty in [19], with respect to [23], was the addition, as post-smoother, of a soft-thresholding denoising. We are going to define our method, starting from the idea presented in [19], by combining framelet denoising and multigrid. Firstly, differently from the previous works, we define a general coarsening strategy independent of the bc’s. This strategy consists in applying the Galerkin projection to the PSF, instead of applying it to the coefficient matrix, and then employing, at each coarser level, the desired bc’s so that fast computations are preserved across the levels. We will prove that it is equivalent to construct a sequence of continuous operators by the Galerkin approach and then discretize each continuous problem on the corresponding grid with appropriate bc’s. In particular, at the coarser levels, exploiting the properties of the error equation, we change the structure of the coefficient matrix resorting to block circulant with circulant blocks (BCCB) matrices, such that all the involved computations can be done by fast Fourier transform (FFT). Furthermore, our proposal differs from the one in [19] also for the usage of framelet denoising as a pre-smoother instead of as post-smoother and for the usage of the described above APIT algorithm instead of CGLS as inner iterative regularization method. Because APIT projects at each iteration into a closed and convex set  $\Omega$  this choice let us ensure that the provided approximation lies in  $\Omega$  as well. Thus, using APIT as post-smoother is equivalent to projecting each multigrid iteration into  $\Omega$ .

Finally, we also provide a proof of convergence of the algorithm in the two grid case (as usual for multigrid methods [47]). We show that proposed method is an iterative regu-

larization method, i.e., that regularization is achieved by early stop of the iteration and that, denoting by  $\mathbf{x}^\delta$  the reconstruction obtained with noise level  $\delta$ , it holds that  $\mathbf{x}^\delta \rightarrow \mathbf{x}^*$  as  $\delta \rightarrow 0$ , where  $\mathbf{x}^*$  denotes a solution of the noise free system (1.3). This analysis was not present in [20]. We would like to stress that, differently from variational methods, iterative regularization methods do not aim at minimizing a known objective function, but achieve regularization by an appropriate choice of the stopping criterion in a semi-convergent method. In other words, the proposed approach will solve the exact system only in the unrealistic scenario where the right-hand side  $\mathbf{b}$  is not affected by noise. When the right-hand side is corrupted by noise the proposed approach will only be able to provide an accurate approximation of the exact solution  $\mathbf{x}^\dagger$ . We emphasize that the proposed algorithm is very robust and that, using only the knowledge of  $\delta$ , is able to achieve high accuracy without tuning any parameter. This is confirmed by numerical comparison with several state of the art algorithms.

The main idea exploited in our multigrid regularization method, both in the post-smoother and in the coarsening strategy, is the refinement of the previous approximations by means of the error equation. The potentiality of the error equation, fully exploited in multigrid methods, is that in some applications there is more information available on the error than on the true solution. In the case of image deblurring the bc's model tries to extrapolate the unavailable information on the true image outside the FOV. Clearly, such model does not always provide a reliable extrapolation and accurate bc's usually lead to coefficient matrices that are not diagonalizable by fast trigonometric transforms. On the other hand, when considering an iterative method, if the chosen bc's well describe the current image and the method itself is near convergence, the error can be accurately modeled as a random image. Hence, the choice of the bc's is not crucial for the error image and classical periodic bc's are a computationally attractive choice since all computations, even (pseudo-)inversions, can be efficiently carried out using the FFT. This is one of the several interpretations behind the robustness of the iterative algorithm proposed in [22], which is the foundation of the APIT method used as post-smoother.

This paper is structured as follows. In Section 2 we briefly describe the tools which are needed for the formulation of our algorithm. In Section 3 we describe our multigrid method, while its convergence and regularization property are proven in Section 4. Finally, Section 5 is devoted to the numerical examples.

**2. Preliminaries.** In this section we discuss the tools, that are already present in the literature, used for constructing our algorithm.

**2.1. Structure of the blurring matrix.** As already mentioned in the Introduction, a popular method to deal with the boundary artifacts is the imposition of appropriate bc's. The boundary artifacts are ringing effects in the restored image due to the fact that the pixels of the observed image close to the boundary highly depend on pixels outside to the FOV. Other strategies could be employed, for instance it is possible to deal directly with the undetermined linear system (1.2) as done in [1, 48] or to enlarge the observed image like in [4, 42]. Nevertheless, if the bc's are well chosen for the problem at hand, they usually provide restored images without ringing artifacts and with a quality comparable to the other strategies; see for instance the numerical results in [13] where some of the numerical methods cited above are compared. In this paper, we consider the bc's approach because the iterative method used as smoother in our multigrid method requires squares matrices. Our proposal could be adapted also to the other approaches for dealing with the boundary artifacts whenever the smoother is properly adopted.

We assume that the position of the PSF center is known and it is denoted by the index  $(0, 0)$ . Thus,  $\mathbf{H} \in \mathbb{R}^{m \times m}$  can be depicted as  $\mathbf{H} = [[h_{j_1, j_2}]_{j_1=-m_{1,1}}^{m_{2,1}}]_{j_2=-m_{1,2}}^{m_{2,2}}$ , where it holds that  $m_{1,i} + m_{2,i} + 1 = m$ , for  $i = 1, 2$ , and the indices are shifted according to the center of

TABLE 2.1

Pad of the original image  $\mathbf{X}$  obtained by imposing the classical bc's considered in [35], with  $\mathbf{X}_c = \text{fliplr}(\mathbf{X})$ ,  $\mathbf{X}_r = \text{flipud}(\mathbf{X})$ , and  $\mathbf{X}_{rc} = \text{flipud}(\text{fliplr}(\mathbf{X}))$ , where  $\text{fliplr}(\cdot)$  and  $\text{flipud}(\cdot)$  are the MATLAB functions that perform the left-right and up-down flip, respectively.

Zero			Periodic			Reflective		
0	0	0	X	X	X	$\mathbf{X}_{rc}$	$\mathbf{X}_r$	$\mathbf{X}_{rc}$
0	X	0	X	X	X	$\mathbf{X}_c$	X	$\mathbf{X}_c$
0	0	0	X	X	X	$\mathbf{X}_{rc}$	$\mathbf{X}_r$	$\mathbf{X}_{rc}$

the PSF. Given the pixels  $h_{j_1, j_2}$  of the PSF, it is possible to associate the so-called generating function  $f(x_1, x_2) = \sum_{j_1=-m_{1,1}}^{m_{2,1}} \sum_{j_2=-m_{1,2}}^{m_{2,2}} h_{j_1, j_2} e^{i(j_1 x_1 + j_2 x_2)}$ , where  $i^2 = -1$ . instance, if  $n$  is odd and the PSF is obtained observing a white pixel on black background in the middle of the  $n \times n$  image, then  $m = n$  and the associated symbol is

$$f(x_1, x_2) = \sum_{j_1, j_2=-(n-1)/2}^{(n-1)/2} h_{j_1, j_2} e^{i(j_1 x_1 + j_2 x_2)},$$

where the coefficients  $h_{j_1, j_2}$  far from the center  $h_{0,0}$  are almost zero due to the compact support of the PSF.

We now discuss some classical bc's and the structure of the resulting matrix  $A$  which can be exploited to achieve fast computations. The matrix-vector product can always be performed by FFT resorting to a proper padding of the vector depending on the imposed bc's. Table 2.1 summarizes the definition of zero, periodic, and reflective bc's; for a detailed description refer to [35]. For antireflective bc's see the review paper [25] and the original proposal in [46]. More sophisticated bc's conditions could be applied without changing our regularizing multigrid method. For instance, the synthetic bc's proposed in [1] or the higher order bc's in [16, 18] could be applied as well.

The *Zero* bc's are obtained by assuming that outside the FOV the image is zero everywhere. This is a good choice when we deal with images having a black background, occurring for instance in astronomical applications. The resulting blurring matrix  $A$  is a block Toeplitz with Toeplitz block (BTTB) matrix defined as

$$A = \mathcal{T}_n(f) = \begin{pmatrix} T_0 & T_{-1} & \cdots & T_{-n+1} \\ T_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & T_{-1} \\ T_{n-1} & \cdots & T_1 & T_0 \end{pmatrix}_{n^2 \times n^2}, \quad T_k = \begin{pmatrix} h_{k,0} & h_{k,-1} & \cdots & h_{k,-n+1} \\ h_{k,1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & h_{k,-1} \\ h_{k,n-1} & \cdots & h_{k,1} & h_{k,0} \end{pmatrix}_{n \times n},$$

for  $k = -n + 1, \dots, n - 1$ . Unfortunately, there are no fast transformation to diagonalize a general matrix of this form. This represents an important drawback in using *Zero* bc's with filtering methods like classical Tikhonov. On the contrary, for other bc's, fast transforms are available; see [35] and below.

*Periodic* bc's assume that outside the FOV the image repeats itself in all directions. In this case the blurring matrix  $A$  is block circulant with circulant block (BCCB) and it is diagonalized by the two dimensional Fourier matrix  $F$  defined by  $F = F_1 \otimes F_1$ , where  $F_1 = [e^{-2jk\pi/n}]_{j,k=0}^{n-1} \in \mathbb{C}^{n \times n}$  and  $\otimes$  denotes the Kronecker product. This factorization allows extremely fast computations with  $A$ , including the (pseudo)-inversion, through the FFT algorithm; see, e.g., [35].

A kind of bc's suitable for generic images are *reflective* bc's proposed in [41]. They assume that the image is reflected (like in a mirror) outside the FOV so that the image is

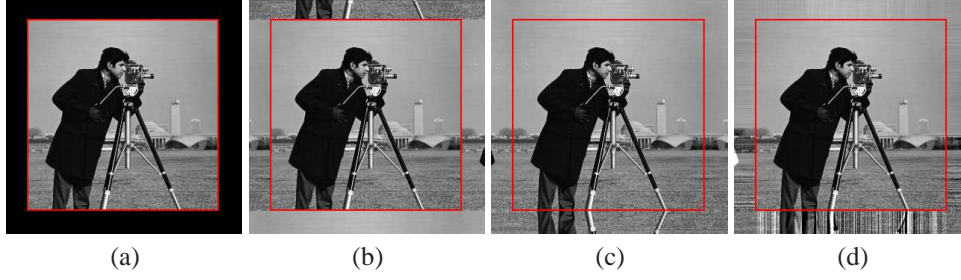


FIGURE 2.1. Examples of bc's, the red box delimits the FOV: (a) zero, (b) periodic, (c) reflective, (d) antireflective.

continuous on the boundaries. The resulting matrix  $A$  has a more involved structure than the ones obtained with periodic or zero bc's. Nevertheless, it can be diagonalized by the discrete cosine transform if the PSF is quadrantly symmetric, i.e., if it is symmetric with regard to both the horizontal and vertical axes.

With a similar spirit Serra-Capizzano proposed the antireflective bc's in [46]. The *antireflective* bc's ensure that on the boundary the image is not only continuous but is continuous also its normal derivative. In this case we antireflect the image outside the FOV; see [25] for details. Even if the structure of the resulting matrix  $A$  is quite complicated, like for reflective bc's,  $A$  can be diagonalized by a modification of the discrete sine transform whenever the PSF is quadrantly symmetric [3, 21, 46].

Figure 2.1 shows an example of the bc's described above.

**2.2. Approximated Iterated Tikhonov.** As discussed in the Introduction, IT can be interpreted as an iterative refinement based on the Tikhonov method. Let  $\mathbf{x}_0$  be an approximation of  $\mathbf{x}^\dagger$ , we can include it in the Tikhonov method (1.6) by computing

$$\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}^\delta\|^2 + \alpha \|\mathbf{x} - \mathbf{x}_0\|^2 \quad \Longleftrightarrow \quad \min_{\mathbf{h}} \|\mathbf{Ah} - \mathbf{r}_0\|^2 + \alpha \|\mathbf{h}\|^2,$$

where  $\mathbf{r}_0 = \mathbf{b}^\delta - \mathbf{Ax}_0$  is the residual and  $\mathbf{h} = \mathbf{x} - \mathbf{x}_0$ . Therefore  $\mathbf{h}$  provides an approximation of the error  $\mathbf{e}_0 = \mathbf{x}^\dagger - \mathbf{x}_0$ , up to the noise in  $\mathbf{b}^\delta$ , and an improved restoration can be computed as  $\mathbf{x}_1 = \mathbf{x}_0 + \mathbf{h}$ . Applying iteratively the same refinement strategy, we obtain the IT method: given  $\mathbf{x}_0$ , for  $k = 0, 1, \dots$

1. compute  $\mathbf{r}_k = \mathbf{b}^\delta - \mathbf{Ax}_k$ ,
2. solve  $\mathbf{h}_k = \arg \min_{\mathbf{h}} \|\mathbf{Ah} - \mathbf{r}_k\|^2 + \alpha_k \|\mathbf{h}\|^2$ ,
3. update  $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{h}_k$ .

The previous algorithm can be formulated as the iteration (1.7).

The choice of  $\alpha_k$  in the IT method has a crucial role and many strategies have been proposed. If the same  $\alpha$  is used for each iteration we call the method stationary, whereas if  $\alpha$  depends on  $k$  we call the method nonstationary. In many applications the latter type has proven to obtain better results, in terms of accuracy and stability than the classical Tikhonov method, avoiding an accurate estimation of the parameter  $\alpha$ . A generalization to the case where a different matrix from the identity is used in the IT iterations has been recently proposed in [8].

The *discrepancy principle* is often chosen as a stopping rule for iterative regularization methods like IT. It stops the iterations after  $k = k_\delta \geq 0$  steps, with  $k_\delta$  such that

$$\|\mathbf{r}_{k_\delta}\| \leq \tau\delta < \|\mathbf{r}_k\|, \quad k = 0, 1, \dots, k_\delta - 1,$$

where  $\tau > 1$  is a fixed constant.



Note that the residual  $\mathbf{r}_k$  at point 1. is computed using the observed image  $\mathbf{b}^\delta$  instead of the true blurred image  $\mathbf{b}$ . Therefore, the error equation, used at point 2. for computing the refinement of the previous approximation, is correct up to the noise in the observation. Therefore, we can replace the matrix  $A$  with an approximation  $C$  if the further introduced error is lower than the noise level  $\delta$ . This is a possible interpretation of the preconditioned iteration proposed in [22] which accordingly requires the following

ASSUMPTION 1. *Let  $C$  be such that*

$$(2.1) \quad \|(C - A)\mathbf{z}\| \leq \rho \|A\mathbf{z}\|, \quad \forall \mathbf{z} \in \mathbb{R}^N,$$

for some  $0 < \rho < \frac{1}{2}$ .

Note that the previous assumption implies that  $C$  is spectrally equivalent to  $A$  and it is crucial for the convergence analysis of the following AIT algorithm [22].

ALGORITHM 1 (AIT). *Let  $\mathbf{x}_0 \in \mathbb{R}^N$  be fixed and set  $k = 0$ . Choose  $\tau = \frac{1+2\rho}{1-2\rho}$  with  $\rho$  as in (2.1), and fix  $q \in [2\rho, 1]$ .*

*While  $\|\mathbf{r}_k\| > \tau\delta$ , let  $\tau_k = \|\mathbf{r}_k\|/\delta$  and  $q_k = \max\left\{q, 2\rho + \frac{1+\rho}{\tau_k}\right\}$ , compute*

$$\mathbf{h}_k = C^t(CC^t + \alpha_k I)^{-1}\mathbf{r}_k,$$

where  $\alpha_k$  is such that

$$(2.2) \quad \|\mathbf{r}_k - C\mathbf{h}_k\| = q_k \|\mathbf{r}_k\|,$$

and update

$$(2.3) \quad \begin{aligned} \mathbf{x}_{k+1} &= \mathbf{x}_k + \mathbf{h}_k, \\ \mathbf{r}_{k+1} &= \mathbf{b}^\delta - A\mathbf{x}_{k+1}. \end{aligned}$$

We summarize the main theoretical results on the convergence of AIT proven in [22]. The parameter  $q$  is not necessary for the theoretical analysis but it is useful in the numerical results preventing a too rapid convergence, which could deteriorate the quality of the computed solution. Firstly, we recall the monotonic decreasing of the norm of the restoration error

$$\mathbf{e}_k = \mathbf{x}^\dagger - \mathbf{x}_k.$$

LEMMA 2.1 ([22]). *Assume that (1.4) and Assumption 1 hold, if  $\tau_k = \|\mathbf{r}_k\|/\delta > \tau_*$ , with  $\tau = (1 + 2\rho)/(1 - 2\rho)$ , then it follows that*

$$\|\mathbf{r}_k - C\mathbf{e}_k\| \leq \left(\rho + \frac{1+\rho}{\tau_k}\right) \|\mathbf{r}_k\| < (1 - \rho) \|\mathbf{r}_k\|$$

PROPOSITION 2.2 ([22]). *Assume Assumption 1 holds and let  $\tau = (1 + 2\rho)/(1 - 2\rho)$ . While  $\|\mathbf{r}_k\| > \tau\delta$  the norm of the reconstruction error  $\mathbf{e}_k$  decreases monotonically, namely  $\|\mathbf{e}_{k+1}\| \leq \|\mathbf{e}_k\|$ , for  $k = 0, 1, \dots$*

From this proposition, it follows that AIT terminates after finitely many iterations according to the following

COROLLARY 2.3 ([22]). *With the assumptions and the notation of Proposition 2.2, it holds*

$$c \sum_{k=0}^{k^\delta-1} \|\mathbf{r}_k\|^2 \leq 2\rho \sum_{k=0}^{k^\delta} \|(CC^t + \alpha_k I)^{-1}\mathbf{r}_k\| \|\mathbf{r}_k\| \leq \|\mathbf{e}_0\|$$

for some constant  $c > 0$ , depending only on  $\rho$  of (2.1) and on the parameter  $q$  of Algorithm 1.

On the other hand, if the data are exact ( $\delta = 0$ ) then the iterates of Algorithm 1 converge to an exact solution of (1.3) as  $k \rightarrow \infty$ .

**THEOREM 2.4 ([22]).** *Assume that the data are exact, i.e.,  $\delta = 0$ , and that  $\mathbf{x}_0$  is not a solution of problem (1.3). Then the sequence  $\{\mathbf{x}_k\}_k$  converges as  $k \rightarrow \infty$  to the solution of (1.3) which is nearest to  $\mathbf{x}_0$ .*

Finally, Algorithm 1 is a regularization method according to the definition in [27].

**THEOREM 2.5 ([22]).** *Let  $\delta \mapsto \mathbf{b}^\delta$  be a function from  $\mathbb{R}^+$  to  $\mathbb{R}^N$  such that (1.4) holds true for all  $\delta > 0$ . Under Assumption 1, for two fixed parameters  $\tau$  and  $q$ , denote by  $\mathbf{x}^\delta$  the resulting approximation obtained with AIT. Then for  $\delta \rightarrow 0$  we have that  $\mathbf{x}^\delta \rightarrow \mathbf{x}_0^\dagger$  which is the nearest solution of (1.3) to  $\mathbf{x}_0$ .*

In [7] it has been proposed an extension of the AIT algorithm which constrains the iterations to lie in a closed and convex set  $\Omega \subset \mathbb{R}^N$ . An example of a suitable  $\Omega$  in the case of image deblurring is the nonnegative cone. In fact, since images are measurements of quantity of light, they can not assume negative values. With the assumption that  $\mathbf{x}^\dagger \in \Omega$ , denoting by  $P_\Omega$  the metric projection over  $\Omega$ , the iterations of projected AIT (APIT) can be computed simply replacing the equation (2.3) by

$$\mathbf{x}_{k+1} = P_\Omega(\mathbf{x}_k + \mathbf{h}_k).$$

All the theoretical results from AIT are directly inherited by APIT as proven in [7].

In both [7, 22] the authors choose  $A$  to be the blurring matrix with the desired bc's and  $C$  as the blurring matrix with the same PSF of  $A$ , but periodic bc's. In this way, independently of the structure of the PSF, it is possible to compute the matrix vector product with  $C^t(CC^t + \alpha_k I)^{-1}$  in  $O(n^2 \log n)$  flops by two FFTs and the computation of  $\alpha_k$  in (2.2) is linear in  $N$  since it requires only few Newton steps in the Fourier domain where the variables are decoupled.

**2.3. Tight frames denoising.** We now describe the classical wavelet (framelet) soft-thresholding signal denoising [26]. A very important feature of tight frames is their redundancy, thus the loss of some information can be tolerated. Hence, wavelets have been often replaced by tight frames in many image applications like inpainting and deblurring [10, 11].

Let the matrix  $\mathcal{W}$  be the analysis operator,

$$\mathcal{W} \text{ is a tight frame} \iff \mathcal{W}^t \mathcal{W} = I.$$

Note that in general  $\mathcal{W}\mathcal{W}^t \neq I$ , unless the system is orthogonal. We can identify some of the elements of the tight frame as low-frequency vectors and the others as high frequency vectors. In other words, we can write

$$\mathcal{W} = \begin{pmatrix} W_0 \\ W_1 \end{pmatrix},$$

where the rows of  $W_0$  are the low-frequency vectors and the rows of  $W_1$  are the high-frequency vectors. Applying  $\mathcal{W}$  to a vector  $\mathbf{x}$ , we obtain  $\mathbf{c}_1 = W_0 \mathbf{x}$  and  $\mathbf{d}_1 = W_1 \mathbf{x}$ . We can recursively apply this decomposition by decomposing again  $\mathbf{c}_1$  obtaining  $\mathbf{c}_2 = W_0^{(1)} \mathbf{c}_1$  and  $\mathbf{d}_2 = W_1^{(1)} \mathbf{c}_1$ , where we indicated with  $\cdot^{(1)}$  the various operators on the (possibly) smaller space in which  $\mathbf{c}_1$  lives. For each level  $j = 0, \dots, l-1$ , starting from  $\mathbf{c}_0 = \mathbf{x}$ , we have that

$$\mathbf{c}_{j+1} = W_0^{(j)} \mathbf{c}_j, \quad \mathbf{d}_{j+1} = W_1^{(j)} \mathbf{c}_j.$$



We can apply the soft-thresholding to the high-frequency components  $\mathbf{d}_j$  to eliminate the noise from a signal. Let  $\theta$  be the threshold parameter, the soft-thresholding operator  $\mu_\theta$  applied to the vector  $\mathbf{d}$  is defined as

$$\mu_\theta(\mathbf{d}) = \text{sgn}(\mathbf{d})(|\mathbf{d}| - \theta)_+,$$

where  $\text{sgn}(x)$  denotes the sign of  $x$  and  $x_+$  is the positive part of  $x$ , i.e.,  $x_+ = \max\{x, 0\}$ , and all the operations are computed element-wise. The choice of the parameter  $\theta$  is of crucial importance. According to [26], we use

$$(2.4) \quad \theta = c \sqrt{\frac{2 \log n}{n}},$$

where  $c > 0$  is a constant that for Gaussian noise can be chosen as  $c = \delta / \|\mathbf{b}^\delta\|$ . Recall that we are assuming that the images are square of dimensions  $n \times n$ .

The denoising algorithm is summarized in Algorithm 2 and will be denoted by

$$S_\theta^l(\mathbf{z}) := \text{Denoise}(\mathbf{z}, \theta, 0, l).$$

**ALGORITHM 2 (Denoise).** Let  $\mathbf{b}^\delta$  denote the noisy signal,  $\theta$  the thresholding parameter, and  $l$  the number of levels.

---

$\mathbf{y} = \text{Denoise}(\mathbf{z}, \theta, lev, l)$

---

```

if lev = l
     $\mathbf{y} = \mathbf{z}$ 
else
     $\begin{pmatrix} \mathbf{c} \\ \mathbf{d} \end{pmatrix} = \mathcal{W}\mathbf{z}$ 
     $\mathbf{c}_1 = \text{Denoise}(\mathbf{c}, \theta, lev + 1, l)$ 
     $\mathbf{d}_1 = \mu_\theta(\mathbf{d})$ 
     $\mathbf{y} = \mathcal{W}^t \begin{pmatrix} \mathbf{c}_1 \\ \mathbf{d}_1 \end{pmatrix}$ 
end

```

The tight frame used in the numerical results is the one related to linear B-splines; see, e.g., [11, 12] and references therein. Note that the low-pass filter of linear B-splines will also be used in the multilevel deblurring algorithm as grid transfer operator. This is not necessary since denoising and projector work on the signal and the residual, respectively, but in practice, this combination provides better restorations with respect to using two independent approximation schemes. The system of linear B-splines is formed by one low-pass filter  $W_0$  and two high-pass filters  $W_1$  and  $W_2$ . The corresponding masks are

$$\mathbf{w}^{(0)} = \frac{1}{2} (1, 2, 1), \quad \mathbf{w}^{(1)} = \frac{\sqrt{2}}{4} (1, 0, -1), \quad \mathbf{w}^{(2)} = \frac{1}{4} (-1, 2, -1).$$

Imposing the reflexive bc's, so that  $\mathcal{W}^t \mathcal{W} = I$ , the resulting filters are

$$W_0 = \frac{1}{4} \begin{pmatrix} 3 & 1 & 0 & \dots & 0 \\ 1 & 2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & 2 & 1 \\ 0 & \dots & 0 & 1 & 3 \end{pmatrix}, \quad W_1 = \frac{\sqrt{2}}{4} \begin{pmatrix} -1 & 1 & 0 & \dots & 0 \\ -1 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 0 & 1 \\ 0 & \dots & 0 & -1 & 1 \end{pmatrix},$$

and

$$W_2 = \frac{1}{4} \begin{pmatrix} 1 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ 0 & \dots & 0 & -1 & 1 \end{pmatrix}.$$

The previous  $W_i$ ,  $i = 0, 1, 2$ , operators are designed for one-dimensional signals. The operators for two space-dimension are defined by using the tensor product

$$(2.5) \quad W_{ij} = W_i \otimes W_j, \quad i, j = 0, 1, 2.$$

Thus, the analysis operator is

$$\mathcal{W} = \begin{pmatrix} W_{00} \\ W_{01} \\ \vdots \\ W_{22} \end{pmatrix}.$$

The low-pass filter is the matrix  $W_{00}$ , while all the other matrices contain at least a high-pass filter in one direction.

**2.4. Multigrid Methods.** Multigrid methods create a sequence of linear systems that decrease in size by consecutive projection. The computational effort is thus reduced and the convergence speed is improved if the smaller linear systems are properly chosen. It is well known that iterative methods for solving linear systems show faster convergence on the well-conditioned space and that the convergence can be very slow in the ill-conditioned space.

REMARK 2.6. The definition of well- and ill-conditioned space is not formal. Let  $V \subset \mathbb{R}^N$  be a linear subspace of  $\mathbb{R}^N$ . We define the *conditioning number of  $A$  restricted to  $V$*  by

$$\kappa_V = \left( \sup_{\mathbf{x} \in V} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} \right) \cdot \left( \sup_{\mathbf{x} \in V} \frac{\|A^\dagger \mathbf{x}\|}{\|\mathbf{x}\|} \right).$$

The *well-conditioned space* is the space  $W$  such that  $\kappa_W$  is not too large, whereas the *ill-conditioned space*  $I$  is the one where  $\kappa_I$  is very large.

For matrices deriving from the discretization of compact integral operator, like, e.g., (1.3), we have that  $W$  corresponds to the low-frequency space and  $I$  is the high-frequency space.

Let us first describe the Two Grid Method (TGM), which is an iterative algorithm. Let  $\mathbf{x}_k$  be an approximation of the solution of (1.3) at the  $k$ th step, apply  $\nu_1$  steps of an iterative method to  $\mathbf{x}_k$  obtaining

$$\tilde{\mathbf{x}}_k = \text{Pre-Smooth}(A, \mathbf{b}, \mathbf{x}_k, \nu_1).$$

This step is called pre-smoothing, since it precedes all other computations and, in the context of differential equations, damps the error in the high-frequencies, i.e., it smooths the error. In order to refine  $\tilde{\mathbf{x}}_k$  using the error equation, we compute the new residual  $\mathbf{g}_k = \mathbf{b} - A\tilde{\mathbf{x}}_k$ .

Let  $0 < N_1 < N$ , we call  $R \in \mathbb{R}^{N_1 \times N}$  the restriction operator. This operator projects a vector from a grid of size  $N$  to a grid of size  $N_1$ . Differently, the interpolation operator

$P \in \mathbb{R}^{N \times N_1}$  interpolates a vector from a grid of size  $N_1$  to a grid of size  $N$ . Usually  $R = cP^t$ , with  $c > 0$ . Using the Galerkin approach the restricted operator is defined as  $A_1 = RAP \in \mathbb{R}^{N_1 \times N_1}$ . Assuming that both  $R$  and  $P$  are of full rank, then  $A_1$  is invertible whenever  $A$  is invertible. To refine the approximation  $\tilde{\mathbf{x}}_k$ , coarser problem is defined using error equation. Assume that  $A_1$  is invertible and compute the coarser approximation of the error as  $\mathbf{h}_k = PA_1^{-1}R\mathbf{g}_k$ , then the refined version of  $\tilde{\mathbf{x}}_k$  is

$$\hat{\mathbf{x}}_k = \tilde{\mathbf{x}}_k + \mathbf{h}_k = \tilde{\mathbf{x}}_k + PA_1^{-1}R(\mathbf{b} - A\tilde{\mathbf{x}}_k) = \tilde{\mathbf{x}}_k + P(RAP)^{-1}R(\mathbf{b} - A\tilde{\mathbf{x}}_k).$$

The procedure that computes  $\hat{\mathbf{x}}_k$  from  $\tilde{\mathbf{x}}_k$  is called Coarse Grid Correction (CGC). The iteration matrix of the CGC is

$$\mathcal{C} = I - P(RAP)^{-1}RA.$$

Note that  $\mathcal{C}$  is a projector and hence the TGM algorithm requires a smoothing step to converge to the solution of (1.3); see [6].

Finally, to obtain the  $(k + 1)$ th approximation, we can apply  $\nu_2$  steps of an iterative method, called post-smoothing, which can be different from the pre-smoother.

It is possible to show that, under mild conditions, this method converges to the solution of (1.3) whenever  $A$  is positive definite. Nevertheless, a linear convergence rate can be achieved only if the smoother and the projector are chosen in order to reduce the error in almost orthogonal subspaces.

When  $N_1$  is large, the main computational issue of the TGM algorithm is the computation of  $\mathbf{h}_k$  since it requires the solution of a linear system of size  $N_1 \times N_1$ . Therefore, the multigrid method applies recursively the TGM restricting the grids until a very small and easily solvable problem is obtained.

Let  $N = N_0 > N_1 > \dots > N_L > 0$ , we call  $R_i$  and  $P_i$  the  $i$ th restriction and interpolation operator, respectively, so that it holds  $R_i \in \mathbb{R}^{N_{i+1} \times N_i}$  and  $P_i \in \mathbb{R}^{N_i \times N_{i+1}}$ , for  $i = 0, \dots, L - 1$ ; then

$$(2.6) \quad A_i = \begin{cases} A & \text{if } i = 0 \\ R_{i-1}A_{i-1}P_{i-1} & i = 1, \dots, L. \end{cases}$$

For instance, for images of size  $2^L \times 2^L$ , we have  $N = 2^{2L}$  and defining  $R_i$  as the operator that selects a pixel every two in each direction (down-sampling), it holds  $N_i = 2^{2(L-i)}$ , for  $i = 0, \dots, L$ , thus  $N_L = 1$ .

Summarizing, the single step of the multigrid iteration goes as follows

$$\begin{array}{l} \mathbf{y}_i = \text{V-Step}(\mathbf{x}_i, A_i, \mathbf{b}_i, i, L) \\ \hline \text{if } (i = L) \text{ then } \mathbf{y}_L = \text{Solve}(A_L \mathbf{y}_L = \mathbf{b}_L) \\ \quad \text{else } \tilde{\mathbf{x}}_i = \text{Pre-Smooth}(A_i, \mathbf{b}_i, \mathbf{x}_i, \nu_1) \\ \quad \quad \mathbf{r}_{i+1} = R_i(\mathbf{b}_i - A_i \tilde{\mathbf{x}}_i) \\ \quad \quad \mathbf{e}_{i+1} = \text{V-Step}(\mathbf{0}, A_{i+1}, \mathbf{r}_{i+1}, i + 1, L) \\ \quad \quad \hat{\mathbf{x}}_i = \tilde{\mathbf{x}}_i + P_i \mathbf{e}_{i+1} \\ \quad \quad \mathbf{y}_i = \text{Post-Smooth}(A_i, \mathbf{b}_i, \hat{\mathbf{x}}_i, \nu_2) \\ \text{end} \end{array}$$

which is called V-cycle since when represented graphically it resembles a V; see [47]. The Multigrid Method is the following iterative application of such V-cycle:

$$\mathbf{x}_{k+1} = \text{V-Step}(\mathbf{x}_k, A, \mathbf{b}, 0, L),$$

given an initial approximation  $\mathbf{x}_0$ .

For well-posed Toeplitz-like linear systems, multigrid methods have been extensively investigated starting from the seminal work [30]. The symbol analysis for Toeplitz matrices can be considered a generalization of the local Fourier analysis, as proven in [17], and can be extended to many Toeplitz-like structures and multidimensional problems; see, e.g., [2]. In all cases the grid transfer operator is constructed as the combination of a square matrix generated by the symbol  $p$  and having the same structure of the coefficient matrix, with a proper down sampling operator which preserves the same structure at the coarser level. In such way, for two dimensional problems, the coarser matrix is generated by the symbol

$$f_{\text{next}}(x_1, x_2) = \frac{1}{4} \left( f p^2 \left( \frac{x_1}{2}, \frac{x_2}{2} \right) + f p^2 \left( \frac{x_1}{2} + \pi, \frac{x_2}{2} \right) + f p^2 \left( \frac{x_1}{2}, \frac{x_2}{2} + \pi \right) + f p^2 \left( \frac{x_1}{2} + \pi, \frac{x_2}{2} + \pi \right) \right),$$

where  $f$  is the symbol of the coefficient matrix. According to this observation, in [37] a multigrid method for BTTB has been defined choosing as coarser matrix the Toeplitz generated by  $f_{\text{next}}$  independently of the downsampling used. This approach can be interpreted as a “geometric” multigrid for structured matrices, where at the coarser level it is not used the rediscritization of the same operator associated to  $f$ , but those associated to  $f_{\text{next}}$  which is computed by applying a Galerkin strategy to the continuous operators, i.e., to the Fourier coefficients of the symbols  $f$  and  $p$ . It is well known that the geometric approach is less robust with respect to the Galerkin approach in term of speed of convergence, but it is more flexible since it does not requires that specific relations on the size of the matrices are preserved.

For well-posed problems, the results in [2] show that for certain matrix algebras related to BTTB, under some hypothesis that link the smoother and the coarsening strategy, the multigrid method has a linear convergence rate, i.e., the number of iterations do not depend on the dimension of the problem. Intuitively the restriction operator should map the error equation into the subspace where  $A$  is ill-conditioned, because classical smoothers reduce the error in the well-conditioned subspace.

Conversely, for ill-posed problems, the projection in the ill-conditioned subspace results to be dangerous because this space is mainly formed by the high frequencies where the noise has its most relevant components. Hence solving the problem projected to the ill-conditioned space may lead to a dramatical amplification of the noise. Therefore, the grid transfer operator has to be chosen differently, as proposed in [23] and proven in [25]. On the other hand the projection in the well-conditioned subspace allows also the inversion of the coarser matrix if  $N_L$  is small enough. Note that using a projection into the well-conditioned subspace, the multigrid that we are going to construct does not have a linear convergence rate since it does not fulfills the hypothesis in [2], but it shows a very stable convergence which is a very useful feature for iterative regularization methods. A fast enough convergence could be preserved using a fast smoother defined for instance using some preconditioning techniques.

**3. Our multigrid iterative regularization method.** Here we describe our algorithmic proposal which combines low-pass filters with fast iterative regularization methods. Firstly, we introduce the coarsening strategy necessary to define the coarser problems and then we describe our pre- and post-smoother.

**3.1. Coarsening.** The restriction  $R_i : \mathbb{R}^{n_{i-1}^2} \rightarrow \mathbb{R}^{n_i^2}$  is the *full weighting* operator and the prolongation is

$$(3.1) \quad P_i = \frac{1}{4} R_i^t, \quad i = 0, \dots, L-1,$$

which is a scaled version of the bilinear interpolation. The stencil associated to  $R_i$  is

$$M = \frac{1}{16} \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix},$$

which corresponds to the symbol

$$p(x_1, x_2) = \frac{1}{4}(1 + \cos(x_1))(1 + \cos(x_2)).$$

In order to obtain a coarser problem, the square matrix generated by  $p(x_1, x_2)$  has to be combined with a down-sampling operator. Let  $K_d^{(i)}$  be the downsampling operator at level  $i$ , it is defined as

$$K_d^{(i)} = \tilde{K}_d^{(i)} \otimes \tilde{K}_d^{(i)},$$

where  $\tilde{K}_d^{(i)}$  is the one-dimensional down-sampling operator which keeps a component every two. Namely,  $\tilde{K}_d^{(i)}$  can be written as an  $n_{i+1} \times n_i$  matrix such that

$$\tilde{K}_d^{(i)} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix}$$

or

$$\tilde{K}_d^{(i)} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix},$$

for  $n_i$  even and odd, respectively. Therefore,  $N_i = n_i^2$ , for  $i = 0, \dots, L$ , with  $n_{i+i} = n_i/2$  if  $n_i$  is even and  $n_{i+i} = (n_i - 1)/2$  for  $n_i$  odd. For instance, in the case of zero Dirichlet bc's, the restriction operator is

$$(3.2) \quad R_i = K_d^{(i)} \mathcal{T}_{n_i}(p), \quad i = 0, \dots, L-1.$$

REMARK 3.1. The restriction matrix  $R_i$ , analogous to that in (3.2) but defined by imposing reflective bc's, coincides with the low-pass filter  $W_{00}$  in (2.5).

We provide now some details on the implementation. Working on 2D problems we store the data in bidimensional arrays. Using the Matlab notation, we denote by  $\text{vec}(\cdot)$  the function that returns a vector containing the columns of the matrix in input stacked below each other. Let  $\mathbf{X} \in \mathbb{R}^{n_i \times n_i}$  and  $\mathbf{x} = \text{vec}(\mathbf{X})$ , the restriction of the vector  $\mathbf{x}$  can be computed by

$$R_i \mathbf{x} = K_d^{(i)} \mathcal{T}_{n_i}(p) \mathbf{x} = (\tilde{K}_d^{(i)} \otimes \tilde{K}_d^{(i)}) \mathcal{T}_{n_i}(p) \text{vec}(\mathbf{X}) = (\tilde{K}_d^{(i)} \otimes \tilde{K}_d^{(i)}) \text{vec}(M * \mathbf{X}) = \text{vec}(\tilde{K}_d^{(i)} (M * \mathbf{X}) (\tilde{K}_d^{(i)})^t),$$

where  $*$  denotes the convolution operator. Similarly, the prolongation  $P_i$  can be applied to an image  $\mathbf{X} \in \mathbb{R}^{n_{i-1} \times n_{i-1}}$  as follows

$$P_i \mathbf{x} = \frac{1}{4} \mathcal{T}_{n_i}^t(p) (K_d^{(i)})^t \mathbf{x} = \frac{1}{4} \text{vec} \left( M * \left( (\tilde{K}_d^{(i)})^t \mathbf{X} \tilde{K}_d^{(i)} \right) \right).$$

A crucial choice concerns the construction of the matrices  $A_i$ . For general bc's and arbitrary size of the true image  $\mathbf{X}$ , the Galerkin approach in (2.6) does not ensure that the structure of the coefficient matrix is preserved through the levels. Indeed the proposals [19, 23] require images of size  $(2^\ell - 1) \times (2^\ell - 1)$ ,  $\ell \in \mathbb{N}$ , and zero Dirichlet boundary conditions. In particular, if  $A$  has a structure defined by reflective or antireflective boundary conditions,  $A_1 = R_1 A P_1$  does not have the same structure of  $A$ . Since the structure of the matrices is essential for fast computations and for the spectral analysis at the coarser levels, we use the approach proposed in [37], where the coarser matrices  $A_{i+1}$  are computed imposing a prescribed structure to the matrix generated by the symbol

$$f_{i+1}(x_1, x_2) = \frac{1}{4} \left( q_i \left( \frac{x_1}{2}, \frac{x_2}{2} \right) + q_i \left( \frac{x_1}{2} + \pi, \frac{x_2}{2} \right) + q_i \left( \frac{x_1}{2}, \frac{x_2}{2} + \pi \right) + q_i \left( \frac{x_1}{2} + \pi, \frac{x_2}{2} + \pi \right) \right),$$

with  $q_i = f_i p_i^2$ , for  $i = 0, \dots, L-1$ , and  $A_0 = A$ . A natural choice is to use the same structure of  $A$  like in [37], but it is possible to save some computational cost, without deteriorate the quality of the computed solution, simply by defining  $A_{i+1}$  as BCCB matrices. This follows from the fact that the CGC works on the error equation. Hence, if the multigrid method is computing a good restoration, then the error can be seen as a random image. Thus, it is not necessary to impose accurate bc's on the error equation, but the periodic assumption, which leads to BCCB matrices, is reliable. Note that if  $A$  is defined by periodic bc's and  $n$  is a power of two, then the proposed coarsening strategy coincides with the standard Galerkin approach, i.e.,  $A_{i+1} = R_i A_i P_i$ .

Computationally, the sequence of symbols  $\{f_i\}_{i=0}^L$  is simply represented by the sequence of the corresponding Fourier coefficients  $\{\text{PSF}_i\}_{i=0}^L$ , where  $f_0 = f$  and  $\text{PSF}_0 = \text{PSF}$  since  $A_0 = A$ . Therefore, according to (2.6), the coarser Fourier coefficients are computed by

$$(3.3) \quad \text{PSF}_i = \begin{cases} \text{PSF} & i = 0 \\ K_d^{(i-1)}(M^t * \text{PSF}_{i-1} * \frac{1}{4}M)K_u^{(i-1)} & i = 1, \dots, L. \end{cases}$$

The matrices  $\text{PSF}_i$  are computed in a setup phase executed before the iterations of the multigrid method, while the construction of the matrices  $A_i$  is not necessary since they are never allocated and only computations by FFT are performed.

The coarsening is repeated until the system is reduced to a single equation in only one variable, such that its solution is fast and stable.

**3.2. Smoothing.** The *pre-smoother* is the framelet denoising described in Section 2.3. This pre-smoother is able to keep under control the effects of the noise without smoothing the edges. Differently from [19], here we use the framelet denoising as a pre-smoother instead of a post-smoother. At the finest level there is no substantial difference in applying a method as pre- or post-smoother thanks to the iterative nature of multigrid methods. This choice is mainly due to two facts:

1. We want to project the iteration inside the nonnegative cone. The denoising of a nonnegative signal can, in principle, insert negative values and thus, since the post-smoother is the very last operation performed, using it as a post-smoother may result in that the determined approximation has negative values.
2. Since the pre-smoother acts directly on the initial approximation at each level, we are going to denoise only the finest level. In fact, the initial approximation of the coarser levels is the zero vector and thus its soft-thresholded version is again 0, independently of the parameter  $\theta$ . Moreover, since we do not apply denoising to the lower levels, there is no need to estimate further parameters.

Framelet denoising is preferable to the usage of few iterations of an iterative regularization method like CGLS because framelet denoising has proven to be able to effectively



remove noise while preserving relevant information on the image. Moreover, few iterations of an iterative regularization method may destroy some of the relevant information in the image, especially when the method is close to convergence, due to the semi-convergence phenomenon. Finally, the introduction of the framelet denoising let us develop a complete convergence theory; see Section 4.

The threshold parameter  $\theta$  for the denoising is chosen in a nonstationary way. We use the following sequence of parameters

$$(3.4) \quad \theta_k = p^{k-1} \frac{\delta}{\|\mathbf{b}^\delta\|} \sqrt{\frac{2 \log n}{n}},$$

where  $0 < p < 1$  and  $k$  denotes the iteration. Note that  $\theta_1$  coincides with the parameter choice in (2.4). However, since the post-smoother which we are going to use also has a denoising effect, thus we choose to decrease the parameter throughout the iterations.

For the *post-smoother*, we use one iteration of Algorithm 1 AIT described in Section 2.2. For the computation of the regularization parameter  $\alpha_1$  at each coarser level, we need an estimate of the norm of the noise at each levels. This is obtained by the estimation in [40], where it is proven that, fixing  $\delta_0 = \delta$ , the norm of the noise  $\delta_i$  can be estimated by

$$(3.5) \quad \delta_i = \frac{\delta_{i-1}}{2}, \quad i = 1, \dots, L-1,$$

Note that at the coarser levels, since we impose periodic bc's, AIT simply reduces to IT saving some computational cost. Nevertheless, we use exactly the same algorithmic proposal of AIT for estimating  $\alpha_1$ .

Enforcing the nonnegativity of the solution can help in achieving better reconstructions, so we want to be sure that our method fulfills this constraint. This can be easily added at each iteration as shown in [13], which is equivalent to use APIT as post-smoother at the finest level.

The projection in the nonnegative cone is performed only at the finest level and not on the coarser ones. This is due to the nature of multigrid methods. As we described above, at the coarser levels multigrid methods solve the error equation and not the problem itself. Each element of the error vector can be either positive or negative and the error vector, in general, does not belong to the nonnegative cone. So the projection in the nonnegative cone could deteriorate the quality of the computed solution since it may destroy relevant information. More in general, determining a closed and convex set in which the error lies is not an easy task and, thus, we do not enforce any constraint on the coarser levels.

**3.3. The algorithm.** To determine at which iteration to stop our multigrid regularization, we use an adaptation of the discrepancy principle, which is the same of Algorithm 1; see [22] and [7] for details on AIT and APIT, respectively. Let  $\mathbf{x}_k$  be the approximated solution at step  $k$ , then the stopping iteration  $k^\delta$  is

$$(3.6) \quad k^\delta = \min_k \left\{ k : \|\mathbf{A}\mathbf{x}_k - \mathbf{b}^\delta\| \leq \frac{1 + 2\rho}{1 - 2\rho} \delta \right\},$$

where  $\rho$  is defined in (2.1).

As mentioned above, we treat differently the finest level from the coarser levels. This choice is motivated by the fact that the coarser levels deal with the error equation which has different properties. In particular, at the coarser levels:

1. We do not apply any denoising because the initial guess is the  $\mathbf{0}$  vector;
2. We do not use projection in the post-smoother because it is not suitable for the error;

3. We employ periodic bc's which usually are a good model for a random image as the error. In fact, the error can be seen almost as a random variable and thus the periodic assumption is well approximating the true behavior of the error. Thus, the matrices  $A_i$ , for  $i = 1, \dots, L$ , are BCCB and all the required computations can be done very efficiently by FFT.

ALGORITHM 3 (Multigrid Method (MgM)). Consider the system (1.5). Let  $A_i$  be defined as the blurring matrix with PSF  $PSF_i$  defined as in (3.3), for  $i = 0, \dots, L$ , and choose suitable bc's for  $i = 0$  and periodic bc's for  $i = 1, \dots, L$ . Let  $R_i$  and  $P_i$  be defined as in (3.1) and (3.2), let the noise levels for each level  $\delta_i$  be defined as in (3.5), the parameter  $\theta_k$  be chosen as in (3.4), and choose the number of framelet levels  $l$  to which apply the denoising. Let  $\mathbf{x}_0$  be an initial guess for the solution of (1.5)

---


$$\mathbf{x} = \text{MGM}(\mathbf{x}_0, A, \mathbf{b}^\delta)$$


---

```

k = 1
while  $\|A\mathbf{x}_k - \mathbf{b}^\delta\| > \frac{1+2\rho}{1-2\rho}\delta$ 
   $\mathbf{x}_k = \text{MGM Single Step}(\mathbf{x}_{k-1}, A, \mathbf{b}^\delta, 0, l, L)$ 
  k = k + 1
end
  
```

The single step of the algorithm is defined by

---


$$\mathbf{y}_i = \text{MGM Single Step}(\mathbf{x}_i, A_i, \mathbf{b}_i^{\delta_i}, i, l, L)$$


---

```

if (i = L) then  $\mathbf{y}_L = A_L^\dagger \mathbf{b}_L^{\delta_L}$ 
else  $\tilde{\mathbf{x}}_i = \begin{cases} S_{\theta_k}^l(\mathbf{x}_i) & i = 0 \\ \mathbf{x}_i & \text{otherwise} \end{cases}$ 
   $\mathbf{r}_{i+1} = R_i(\mathbf{b}_i^{\delta_i} - A_i \tilde{\mathbf{x}}_i)$ 
   $\mathbf{e}_{i+1} = \text{MGM Single Step}(\mathbf{0}, A_{i+1}, \mathbf{r}_{i+1}, i+1, l, L)$ 
   $\hat{\mathbf{x}}_i = \tilde{\mathbf{x}}_i + P_i \mathbf{e}_{i+1}$ 
   $\hat{\mathbf{y}}_i = \text{AIT}(\hat{\mathbf{x}}_i, A_i, \mathbf{b}_i^{\delta_i}, 1)$ 
   $\mathbf{y}_i = \begin{cases} P_\Omega(\hat{\mathbf{y}}_i) & i = 0 \\ \hat{\mathbf{y}}_i & \text{otherwise} \end{cases}$ 
end
  
```

Where  $\text{AIT}(\hat{\mathbf{x}}_i, A_i, \mathbf{b}_i^{\delta_i}, 1)$  denotes the application of one step of Algorithm 1 with initial guess  $\hat{\mathbf{x}}_i$ , system matrix  $A_i$ , right-hand side  $\mathbf{b}_i^{\delta_i}$ , and estimated noise level  $\delta_i$ .

REMARK 3.2. Differently from the well-posed case, it might happen that  $A_L$  is not invertible, then at level  $L$  we solve the least square problem

$$\mathbf{y}_L = \arg \min_{\mathbf{y}} \|A_L \mathbf{y} - \mathbf{b}_L^{\delta_L}\|^2 = A_L^\dagger \mathbf{b}_L^{\delta_L}.$$

We now discuss the choice of  $\mathbf{x}_0$ . As we will see in Section 4, the convergence of the method is guaranteed for any initial guess. This choice can affect very marginally the quality of the final restoration. From our experiments, according to the numerical results in [22], we have observed that the choice of  $\mathbf{x}_0 = \mathbf{b}^\delta$  is very natural and leads to very accurate restorations. Nevertheless, different choices for  $\mathbf{x}_0$ , like  $\mathbf{x}_0 = \mathbf{0}$  and  $\mathbf{x}_0 = A^t \mathbf{b}^\delta$ , can be considered as well.

REMARK 3.3. Our multigrid does not fulfill the hypothesis in [2] and hence we cannot expect a linear convergence rate. In particular, the chosen projector does not project into

the ill-conditioned space of  $A_i$ . On the contrary, being a low-pass filter, it projects into the well-conditioned space of  $A_i$  which is the low-frequency space.

This means that, as we will see in the numerical results in Section 5, our algorithm requires a number of iterations slightly higher if compared to the post-smoother APIT. However, this is needed in order to improve the regularizing effect and improve the quality of the restorations.

Concerning the arithmetic cost of one multigrid iteration, this is not much higher than the cost of a single iteration of the post-smoother APIT at the finer level, which is lower than  $cn^2 \log n$  for a fixed constant  $c > 0$ , up to lower order terms, due to four FFTs (two for computing the residual with the chosen bc's and two for applying the preconditioner). Indeed, at the coarser levels APIT reduces to IT with periodic bc's saving two FFTs and the denoising pre-smoother, which is linear in  $n$  and applied only at the finest level. Hence, since the computational cost at each coarser level  $i = 1, \dots, L$  is lower than  $\frac{c}{2}n_i^2 \log n_i$  up to lower order terms, the total arithmetic cost of one iteration of our MgM for image deblurring is  $\frac{7}{6}cn^2 \log n + O(n)$  according to the computational cost of classical V-cycle [47].

**4. Convergence Analysis.** We are now going to study the convergence and regularization properties of our algorithm. In order to do that, we assume that  $\Omega$  is the nonnegative cone, the discretization of (1.1) is computed imposing periodic bc's so that the coarse matrices computed using (3.3) coincide with the standard Galerkin approach as in (2.6). Note that a PSF usually has nonnegative entries because it performs weighted average. Hence, imposing periodic bc's the resulting matrix has nonnegative entries. Finally, as usually done for the theoretical analysis of multigrid methods like in [47], we consider the simplest case of the two grid method such that  $L = 1$ .

Observe that  $P_i = \frac{1}{4}R_i^t$  according to (3.1) then, denoting the interpolation operator  $P_0$  simply by  $P$ , we obtain that

$$A_1 = 4P^tAP.$$

The two-grid version of our MgM algorithm is reported in Algorithm 4.

**ALGORITHM 4 (TGM).** Consider the system (1.5). Let the parameter  $\theta_k$  be chosen as in (3.4), and choose the number of framelet levels  $l$  to which apply the denoising. Let  $\rho$  be the parameter in equation (2.1) and  $q$  be a constant such that  $2\rho \leq q \leq 1$ . Let  $\mathbf{x}_0$  be an initial guess for the solution of (1.5)

---


$$\mathbf{x} = \text{TGM}(\mathbf{x}_0, A, \mathbf{b}^\delta)$$


---

```

k = 0
while  ||bδ - A $\mathbf{x}_k$ || ≥  $\frac{2+\rho}{2-\rho}\delta$ 
     $\tilde{\mathbf{x}}_k$  =  $S_{\theta_k}^l(\mathbf{x}_k)$ 
     $\mathbf{r}_k$  =  $\mathbf{b}^\delta - A\tilde{\mathbf{x}}_k$ 
     $\mathbf{h}_k$  =  $P(P^tAP)^\dagger P^t\mathbf{r}_k$ 
     $\hat{\mathbf{x}}_k$  =  $\tilde{\mathbf{x}}_k + \mathbf{h}_k$ 
     $q_k$  =  $\max\left\{q, 2\rho + \frac{(1+\rho)\delta}{\|\mathbf{b}^\delta - A\hat{\mathbf{x}}_k\|}\right\}$ 
     $\hat{\mathbf{r}}_k$  =  $\mathbf{b}^\delta - A\hat{\mathbf{x}}_k$ 
     $\alpha_k \leftarrow$  determine  $\alpha_k$  that solves the non-linear equation
        || $\hat{\mathbf{r}}_k - CC^t(CC^t + \alpha_k I)^{-1}\hat{\mathbf{r}}_k$ || =  $q_k \|\hat{\mathbf{r}}_k\|$ 
     $\mathbf{x}_{k+1}$  =  $P_\Omega(\hat{\mathbf{x}}_k + C^t(CC^t + \alpha_k I)^{-1}(\mathbf{b}^\delta - A\hat{\mathbf{x}}_k))$ 
    k = k + 1
end
  
```

Note that  $\alpha_k$  in Algorithm 4 is chosen by solving the non linear equation

$$\|\hat{\mathbf{r}}_k - CC^t(CC^t + \alpha_k I)^{-1}\hat{\mathbf{r}}_k\| = q_k \|\hat{\mathbf{r}}_k\|.$$

This can be performed easily when  $C$  is a BCCB matrix using Newton's method; see [22].

We define the following errors

$$(4.1) \quad \mathbf{e}_k = \mathbf{x}^\dagger - \mathbf{x}_k, \quad \tilde{\mathbf{e}}_k = \mathbf{x}^\dagger - \tilde{\mathbf{x}}_k, \quad \hat{\mathbf{e}}_k = \mathbf{x}^\dagger - \hat{\mathbf{x}}_k.$$

For convenience, we also define

$$(4.2) \quad D = P(P^t AP)^\dagger P^t, \quad Q_k = C^t(CC^t + \alpha_k I)^{-1}.$$

In order to prove the convergence, we need the following

ASSUMPTION 2. *The thresholding parameters  $\theta_k$  and the number of levels  $l$  ensure that*

$$\|\mathbf{e}_k\| \geq \|\tilde{\mathbf{e}}_k\|.$$

ASSUMPTION 3. *The noise vector  $\boldsymbol{\xi} = \mathbf{b}^\delta - \mathbf{b}$  does not have any component in  $\mathcal{R}(P^t)$ , i.e.,  $\boldsymbol{\xi} \in \mathcal{N}(P^t) = \mathcal{R}(P)^\perp$ .*

Assumption 2 requires that the threshold parameter is chosen so that it does not deteriorate the error. Obviously it is satisfied if the parameter is small enough and it is reasonable for a converging algorithm. Since the sequence  $(\theta_k)_k$  is strictly decreasing, then the assumption will be satisfied, at least, for  $k$  large enough. Assumption 3 is unlikely to be completely satisfied in a real case scenario, but it is almost true because the noise is a highly oscillating function, thus most of its components lie in the high frequency space which coincides with the null space of  $P^t$ , being  $P$  a low-pass filter. Thus, while it might not be true that  $\boldsymbol{\xi} \in \mathcal{N}(P^t)$ , in practice  $\|P^t \boldsymbol{\xi}\| \approx 0$ .

REMARK 4.1. Note that when the data are exact, i.e.,  $\delta = 0$ , Assumption 3 is trivially satisfied. Moreover,  $\theta_k = 0$  and so also Assumption 2 holds. In other words, in the noise-free case both Assumptions 2 and 3 are satisfied.

First of all we have to prove the following

LEMMA 4.2. *Let  $\hat{\mathbf{e}}_k$  and  $\tilde{\mathbf{e}}_k$  be defined as in (4.1). Assume that  $\mathbf{x}^\dagger \in \Omega$ , under Assumption 3 it holds*

$$\|\hat{\mathbf{e}}_k\| \leq \|\tilde{\mathbf{e}}_k\|.$$

*Proof.* Note that  $\mathcal{C} = I - DA$  is the CGC matrix where  $D$  is defined as in (4.2). Hence  $\mathcal{C}$  is a projector and  $\|\mathcal{C}\| = 1$ . Thus, it holds

$$\begin{aligned} \|\hat{\mathbf{e}}_k\| &= \|\mathbf{x}^\dagger - \hat{\mathbf{x}}_k\| \\ &= \|\mathbf{x}^\dagger - (\tilde{\mathbf{x}}_k + D(\mathbf{b}^\delta - A\tilde{\mathbf{x}}_k))\| \\ &\stackrel{(a)}{=} \|\mathbf{x}^\dagger - (\tilde{\mathbf{x}}_k + D(\mathbf{b} - A\tilde{\mathbf{x}}_k))\| \\ &= \|\mathbf{x}^\dagger - \tilde{\mathbf{x}}_k - DA(\mathbf{x}^\dagger - \tilde{\mathbf{x}}_k)\| \\ &= \|(I - DA)(\mathbf{x}^\dagger - \tilde{\mathbf{x}}_k)\| \\ &\leq \|\tilde{\mathbf{e}}_k\|, \end{aligned}$$

where (a) is justified by Assumption 3; in fact,

$$D\mathbf{b}^\delta = P(P^t AP)^\dagger P^t \mathbf{b}^\delta = P(P^t AP)^\dagger P^t \mathbf{b} = D\mathbf{b}.$$

□

We can now prove the following preliminary result.

PROPOSITION 4.3. *Let  $\hat{\mathbf{e}}_k$ ,  $\tilde{\mathbf{e}}_k$ , and  $\mathbf{e}_k$  be defined as in (4.1). Under Assumptions 1, 2, and 3, it holds*

$$\|\mathbf{e}_k\|^2 - \|\mathbf{e}_{k+1}\|^2 \geq 2\rho \|(CC^t + \alpha_k I)^{-1} \hat{\mathbf{r}}_k\| \|\hat{\mathbf{r}}_k\|,$$

where  $\hat{\mathbf{r}}_k = \mathbf{b}^\delta - A\hat{\mathbf{x}}_k$ .

*Proof.* Denote with  $\hat{\mathbf{h}}_k = C^t(CC^t + \alpha_k I)^{-1} \hat{\mathbf{r}}_k$ . Consider

$$\begin{aligned} \|\mathbf{e}_{k+1}\|^2 &= \|\mathbf{x}^\dagger - \mathbf{x}_{k+1}\|^2 \\ &= \left\| \mathbf{x}^\dagger - P_\Omega \left( \hat{\mathbf{x}}_k + \hat{\mathbf{h}}_k \right) \right\|^2 \\ &\stackrel{(a)}{=} \left\| P_\Omega \left( \mathbf{x}^\dagger \right) - P_\Omega \left( \hat{\mathbf{x}}_k + \hat{\mathbf{h}}_k \right) \right\|^2 \\ &\stackrel{(b)}{\leq} \left\| \mathbf{x}^\dagger - \hat{\mathbf{x}}_k - \hat{\mathbf{h}}_k \right\|^2 \\ &= \|\hat{\mathbf{e}}_k\|^2 - 2 \langle \hat{\mathbf{e}}_k, \hat{\mathbf{h}}_k \rangle + \|\hat{\mathbf{h}}_k\|^2, \end{aligned}$$

where to obtain (a) we have used the fact that, by assumption,  $\mathbf{x}^\dagger \in \Omega$  and for (b) we have used the fact that the metric projection is a contractive mapping; see [50].

Combining Assumption 2 with Lemma 4.2, it holds that

$$\|\mathbf{e}_k\|^2 - \|\mathbf{e}_{k+1}\|^2 \geq \|\tilde{\mathbf{e}}_k\|^2 - \|\mathbf{e}_{k+1}\|^2 \geq \|\hat{\mathbf{e}}_k\|^2 - \|\mathbf{e}_{k+1}\|^2.$$

The remaining part of the proof is inspired by the proof of Proposition 2.2 in [7, 22]. Thus

$$\begin{aligned} \|\mathbf{e}_k\|^2 - \|\mathbf{e}_{k+1}\|^2 &\geq \|\hat{\mathbf{e}}_k\|^2 - \|\mathbf{e}_{k+1}\|^2 \\ &\geq 2 \langle \hat{\mathbf{e}}_k, \hat{\mathbf{h}}_k \rangle - \|\hat{\mathbf{h}}_k\|^2 \\ &= 2 \langle C\hat{\mathbf{e}}_k, (CC^t + \alpha_k I)^{-1} \hat{\mathbf{r}}_k \rangle - \langle \hat{\mathbf{r}}_k, CC^t(CC^t + \alpha_k I)^{-2} \hat{\mathbf{r}}_k \rangle \\ &= 2 \langle \hat{\mathbf{r}}_k, (CC^t + \alpha_k I)^{-1} \hat{\mathbf{r}}_k \rangle - \langle \hat{\mathbf{r}}_k, CC^t(CC^t + \alpha_k I)^{-2} \hat{\mathbf{r}}_k \rangle \\ &\quad - 2 \langle \hat{\mathbf{r}}_k - C\hat{\mathbf{e}}_k, (CC^t + \alpha_k I)^{-1} \hat{\mathbf{r}}_k \rangle \\ &\geq 2 \langle \hat{\mathbf{r}}_k, (CC^t + \alpha_k I)^{-1} \hat{\mathbf{r}}_k \rangle - 2 \langle \hat{\mathbf{r}}_k, CC^t(CC^t + \alpha_k I)^{-2} \hat{\mathbf{r}}_k \rangle \\ &\quad - 2 \langle \hat{\mathbf{r}}_k - C\hat{\mathbf{e}}_k, (CC^t + \alpha_k I)^{-1} \hat{\mathbf{r}}_k \rangle \\ &= 2\alpha_k \langle \hat{\mathbf{r}}_k, (CC^t + \alpha_k I)^{-2} \hat{\mathbf{r}}_k \rangle - \langle \hat{\mathbf{r}}_k - C\hat{\mathbf{e}}_k, (CC^t + \alpha_k I)^{-1} \hat{\mathbf{r}}_k \rangle \\ &\geq 2\alpha_k \langle \hat{\mathbf{r}}_k, (CC^t + \alpha_k I)^{-2} \hat{\mathbf{r}}_k \rangle - \|\hat{\mathbf{r}}_k - C\hat{\mathbf{e}}_k\| \|(CC^t + \alpha_k I)^{-1} \hat{\mathbf{r}}_k\| \\ &\geq 2 \|(CC^t + \alpha_k I)^{-1} \hat{\mathbf{r}}_k\| (\|\alpha_k(CC^t + \alpha_k I)^{-1} \hat{\mathbf{r}}_k\| - \|\hat{\mathbf{r}}_k - C\hat{\mathbf{e}}_k\|) \end{aligned}$$

Since  $\alpha_k(CC^t + \alpha_k I)^{-1} \hat{\mathbf{r}}_k = \hat{\mathbf{r}}_k - C\hat{\mathbf{h}}_k$  and inserting the definition of  $\alpha_k$  we have that

$$\|\alpha_k(CC^t + \alpha_k I)^{-1} \hat{\mathbf{r}}_k\| = q_k \|\hat{\mathbf{r}}_k\|.$$

Thus

$$\|\mathbf{e}_k\|^2 - \|\mathbf{e}_{k+1}\|^2 \geq 2 \|(CC^t + \alpha_k I)^{-1} \hat{\mathbf{r}}_k\| (q_k \|\hat{\mathbf{r}}_k\| - \|\hat{\mathbf{r}}_k - C\hat{\mathbf{e}}_k\|).$$

Using Lemma 2.1 and the definition of  $q_k$ , it holds

$$\begin{aligned} \|\mathbf{e}_k\|^2 - \|\mathbf{e}_{k+1}\|^2 &\geq 2 \|(CC^t + \alpha_k I)^{-1} \hat{\mathbf{r}}_k\| \left( q_k \|\hat{\mathbf{r}}_k\| - \left( \rho + \frac{1+\rho}{\tau_k} \right) \|\hat{\mathbf{r}}_k\| \right) \\ &\geq 2\rho \|(CC^t + \alpha_k I)^{-1} \hat{\mathbf{r}}_k\| \|\hat{\mathbf{r}}_k\|. \end{aligned}$$

which concludes the proof.  $\square$

COROLLARY 4.4. *With the same notation and assumptions of Proposition 4.3, it holds*

$$c \sum_{k=0}^{k^\delta-1} \|\hat{\mathbf{r}}_{k+1}\|^2 \leq 2\rho \sum_{k=0}^{k^\delta-1} \|(CC^t + \alpha_k I)^{-1} \hat{\mathbf{r}}_k\| \|\hat{\mathbf{r}}_k\| \leq \|\mathbf{e}_0\|,$$

for some constant  $c > 0$  depending only on  $\rho$  and  $q$ .

*Proof.* This is direct consequence of the iterated application of Proposition 4.3 combined with Corollary 2.3.  $\square$

Corollary 4.4 proves that, when  $\delta > 0$ , Algorithm 4 stops after finitely many iterations, independently of the choice of  $\mathbf{x}_0$ . In fact, assuming that  $k^\delta = \infty$ , i.e., that the algorithm does not stop after finitely many iterations, then it should be that

$$\sum_{k=0}^{\infty} \|\hat{\mathbf{r}}_k\|^2 \leq c \|\mathbf{e}_0\| < \infty,$$

for some constant  $c > 0$ . Thus the norm of the residual becomes arbitrarily small which is absurd due to the stopping criteria (3.6).

We are now in a position to prove the convergence of TGM Algorithm 4 in the noise free case and that, if  $\mathbf{x}_0$  is not a solution of the system, an infinite number of iterations are needed.

THEOREM 4.5. *Let  $\delta = 0$  and assume that  $\mathbf{x}_0$  is not a solution of the linear system (1.3). Assume that  $\Omega$  is the nonnegative cone and that all entries of  $A$  are nonnegative. Then the iterates generated by Algorithm 4 converge to a solution of (1.3). Moreover, an infinite number of iterations is needed.*

*Proof.* This proof is inspired by the proof of [22, Theorem 4], but some details are different.

We start first by proving that infinitely many iterations are needed. If  $\delta = 0$  then the stopping criterion can be satisfied for a  $k = k^\delta$  only if  $\mathbf{x}_k$  is a solution of the system (1.5). This implies, in order for  $\mathbf{x}_{k+1}$  to be a solution of (1.5), that  $\hat{\mathbf{h}}_{k-1}$  should coincide with  $\hat{\mathbf{e}}_{k-1}$  up to an element in  $\mathcal{N}(A)$ , where

$$\hat{\mathbf{h}}_{k-1} = C^t(CC^t + \alpha_k I)^{-1} \hat{\mathbf{r}}_{k-1}.$$

From Assumption 1 we that  $\mathcal{N}(A) = \mathcal{N}(C)$  and from the definition of  $\alpha_{k-1}$  and Lemma 2.1 it follows that

$$q_{k-1} \|\hat{\mathbf{r}}_{k-1}\| = \|\hat{\mathbf{r}}_{k-1} - C\hat{\mathbf{h}}_{k-1}\| = \|\hat{\mathbf{r}}_{k-1} - C\hat{\mathbf{e}}_{k-1}\| \leq \left( \rho + \frac{1+\rho}{\tau_{k-1}} \right) \|\hat{\mathbf{r}}_{k-1}\|.$$

However, this contradicts the definition of  $q_{k-1}$  which means that the iterations do not stop if  $\hat{\mathbf{x}}_0$  is not a solution of the system.

We now show that the sequence  $\{\mathbf{x}_k\}_k$  is a Cauchy sequence and thus, that it converges.



Let  $m + 1 \geq l$  and consider

$$\begin{aligned}
 \|\mathbf{x}_{m+1} - P_\Omega(\hat{\mathbf{x}}_l)\|^2 &= \|P_\Omega(\hat{\mathbf{x}}_m + \hat{\mathbf{h}}_m) - P_\Omega(\hat{\mathbf{x}}_l)\|^2 \\
 &\leq \|\hat{\mathbf{x}}_m + \hat{\mathbf{h}}_m - \hat{\mathbf{x}}_l\|^2 \\
 &\leq 2\|\hat{\mathbf{x}}_m - \hat{\mathbf{x}}_l\|^2 + 2\|\hat{\mathbf{h}}_m\|^2 \\
 &= 2\|\hat{\mathbf{e}}_m - 2\hat{\mathbf{e}}_l\|^2 + 2\|\hat{\mathbf{h}}_m\|^2 \\
 &= 2\|\hat{\mathbf{e}}_m\|^2 - 2\|\hat{\mathbf{e}}_l\|^2 - 4\langle \hat{\mathbf{e}}_l, \hat{\mathbf{e}}_m - \hat{\mathbf{e}}_l \rangle + 2\|\hat{\mathbf{h}}_m\|^2 \\
 &= 2\|\hat{\mathbf{e}}_m\|^2 - 2\|\hat{\mathbf{e}}_l\|^2 + 4\sum_{j=l}^{m-1} \langle \hat{\mathbf{e}}_l, \hat{\mathbf{e}}_j - \hat{\mathbf{e}}_{j+1} \rangle + 2\|\hat{\mathbf{h}}_m\|^2, \\
 &= 2\|\hat{\mathbf{e}}_m\|^2 - 2\|\hat{\mathbf{e}}_l\|^2 + 4\sum_{j=l}^{m-1} \langle \hat{\mathbf{e}}_l, \hat{\mathbf{x}}_{j+1} - \hat{\mathbf{x}}_j \rangle + 2\|\hat{\mathbf{h}}_m\|^2 \\
 &= 2\|\hat{\mathbf{e}}_m\|^2 - 2\|\hat{\mathbf{e}}_l\|^2 + 4\sum_{j=l}^{m-1} \langle \hat{\mathbf{e}}_l, \hat{\mathbf{h}}_j \rangle + 2\|\hat{\mathbf{h}}_m\|^2 \\
 &= 2\|\hat{\mathbf{e}}_m\|^2 - 2\|\hat{\mathbf{e}}_l\|^2 + 4\sum_{j=l}^{m-1} \langle C\hat{\mathbf{e}}_l, (C^*C + \alpha_j I)^{-1}\hat{\mathbf{r}}_j \rangle + 2\|\hat{\mathbf{h}}_m\|^2 \\
 &\leq 2\|\hat{\mathbf{e}}_m\|^2 - 2\|\hat{\mathbf{e}}_l\|^2 + 4\sum_{j=l}^{m-1} \|C\hat{\mathbf{e}}_l\| \|(C^*C + \alpha_j I)^{-1}\hat{\mathbf{r}}_j\| + 2\|\hat{\mathbf{h}}_m\|^2 \\
 &\leq 2\|\hat{\mathbf{e}}_m\|^2 - 2\|\hat{\mathbf{e}}_l\|^2 + 4(1 + \rho) \sum_{j=l}^{m-1} \|A\hat{\mathbf{e}}_l\| \|(C^*C + \alpha_j I)^{-1}\hat{\mathbf{r}}_j\| + 2\|\hat{\mathbf{h}}_m\|^2 \\
 &= 2\|\hat{\mathbf{e}}_m\|^2 - 2\|\hat{\mathbf{e}}_l\|^2 + 4(1 + \rho) \sum_{j=l}^{m-1} \|\hat{\mathbf{r}}_l\| \|(C^*C + \alpha_j I)^{-1}\hat{\mathbf{r}}_j\| + 2\|\hat{\mathbf{h}}_m\|^2.
 \end{aligned}$$

Let  $n + 1 \leq l$ , similarly as before we obtain

$$\|P_\Omega(\hat{\mathbf{x}}_l) - \mathbf{x}_{n+1}\|^2 \leq 2\|\hat{\mathbf{e}}_n\|^2 - 2\|\hat{\mathbf{e}}_l\|^2 + 4(1 + \rho) \sum_{j=l}^{n-1} \|\hat{\mathbf{r}}_l\| \|(C^*C + \alpha_j I)^{-1}\hat{\mathbf{r}}_j\| + 2\|\hat{\mathbf{h}}_n\|^2.$$

Let  $l$  be such that  $m + 1 \leq l \leq n + 1$  and  $\|\hat{\mathbf{r}}_l\|$  is minimal, then we obtain

$$\begin{aligned}
 \|\mathbf{x}_{m+1} - \mathbf{x}_{n+1}\|^2 &\leq 2\|P_\Omega(\hat{\mathbf{x}}_l) - \mathbf{x}_{n+1}\|^2 + 2\|\mathbf{x}_{m+1} - P_\Omega(\hat{\mathbf{x}}_l)\|^2 \\
 &\leq 4\|\hat{\mathbf{e}}_m\|^2 + 4\|\hat{\mathbf{e}}_n\|^2 - 8\|\hat{\mathbf{e}}_l\|^2 + 4\|\hat{\mathbf{h}}_m\|^2 + 4\|\hat{\mathbf{h}}_n\|^2 \\
 &\quad + 8(1 + \rho) \sum_{j=n}^{m-1} \|\hat{\mathbf{r}}_l\| \|(C^*C + \alpha_j I)^{-1}\hat{\mathbf{r}}_j\|.
 \end{aligned}$$

This shows that the sequence  $\mathbf{x}_j$  is a Cauchy sequence thanks to the fact that the sequence  $\|\hat{\mathbf{e}}_j\|$  is convergent in force of Proposition 4.3, the summation is the tail of a convergent series thanks to Corollary 4.4, and  $\|\hat{\mathbf{h}}_j\| \rightarrow 0$  as  $j \rightarrow \infty$  thanks to Theorem 2.4.

Since all the entries of  $A$  are nonnegative and  $\Omega$  is the nonnegative cone, it follows that

$$(4.3) \quad \|\mathbf{r}_{k+1}\| = \|A\mathbf{e}_{k+1}\| \leq \|A\hat{\mathbf{e}}_k\| = \|\hat{\mathbf{r}}_k\|,$$

where the inequality is obtained thanks to the fact that all the positive entries of  $\mathbf{e}_{k+1}$  are smaller than the ones in  $\hat{\mathbf{e}}_k$  and the entries of  $A$  are all nonnegative.

From Corollary 4.4 we know that  $\|\hat{\mathbf{r}}_k\| \rightarrow 0$  as  $k \rightarrow \infty$ , then from (4.3) we obtain that  $\|\mathbf{r}_k\| \rightarrow 0$  as  $k \rightarrow \infty$  and thus that the limit of the sequence  $\{\mathbf{x}_k\}_k$  is a solution of the noise-free system.  $\square$

The last result that we would like to prove is that Algorithm 4 is a regularization method.

**THEOREM 4.6.** *Assume that Assumption 1 holds for some  $0 < \rho \leq \frac{1}{2}$  and let  $\delta \mapsto \mathbf{b}^\delta$  be a function from  $\mathbb{R}^+$  to  $\mathbb{R}^N$  such that for all  $\delta$  it holds  $\|\mathbf{b} - \mathbf{b}^\delta\| \leq \delta$ . For fixed  $\tau$  and  $q$  denote by  $\mathbf{x}^\delta$  the approximation of  $\mathbf{x}^\dagger$  obtained with Algorithm 4. Then, as  $\delta \rightarrow 0$ ,  $\mathbf{x}^\delta$  goes to a solution of the system (1.3).*

We omit the proof since it can be copied from [32, Theorem 2.3]; for further reference see also [27, Theorem 11.5]. Its essentials ingredients are the monotonicity proven in Proposition 2.2, the convergence to the exact solution in the exact data case proven in Theorem 4.5 and the continuity of the map  $\delta \mapsto \mathbf{b}^\delta$ .

**5. Numerical Examples.** We firstly recall how we construct, in Algorithm 3, the operators  $A_i$  for  $i = 0, \dots, L$ , the approximation  $C_i$  for  $i = 0, \dots, L - 1$ , and the parameters  $\rho_i$  and  $q_i$  for  $i = 0, \dots, L - 1$ .

For the finest level we select  $A_0 = A$  as the blurring matrix with the appropriate bc's, i.e., the one which better approximate the nature of the image. For the coarser levels, since we are working on the error equation, the bc's are irrelevant and so we select  $A_i$  for  $i = 1, \dots, L$  as the blurring matrix with PSF  $\text{PSF}_i$ , defined as in (3.3), and with periodic bc's.

We set  $C_i$  for  $i = 0, \dots, L - 1$ , like in [7, 22], as the blurring matrix associated to  $\text{PSF}_i$  and with periodic bc's. In this way the computation of  $C_i^t(C_i C_i^t + \alpha I)^{-1}$  can be done in  $O(n^2 \log n)$  flops using two FFTs. However, at the coarser levels, since we are imposing periodic bc's, we have that  $A_i = C_i$  and thus the AIT algorithm is simply IT where the regularization parameter is chosen as in AIT. This choice ensures that Assumption 1 is trivially satisfied for any  $\rho$  and, moreover, let us further damp the computational cost.

We set  $\rho_i = 10^{-4}$ ,  $q_0 = 0.7$ ,  $q_i = 1$  for the levels  $i = 1, \dots, L - 1$ , and we fix  $l = 4$ , where  $l$  denotes the number of decompositions in the denoising algorithm (see Algorithm 2). Finally, we set  $L$  so that  $A_L \in \mathbb{R}$ , i.e., so that the problem at the coarsest level is reduced to a single linear equation in one variable.

We would like to stress that the choice of the parameters above is not crucial for the quality of the computed restorations and different choices lead to practically identical reconstructions. In other words the proposed algorithm is very stable with respect to the choice of these parameters. This is confirmed by the following numerical examples where different kind of blur and noise, and images with different features are considered; see also Figure 5.4. Thanks to this stability we can consider the proposed algorithm completely parameter free whenever an estimation of the noise level  $\delta$  is available.

We compare Algorithm 3 to several methods from the literature with respect to both accuracy and efficiency. For the comparison in accuracy we consider three measures, the Relative Reconstruction Error (RRE), the Peak Signal to Noise Ratio (PSNR), and the Structural SIMilarity index (SSIM). The first two quantities are defined as follows

$$\text{RRE}(\mathbf{x}) = \frac{\|\mathbf{x} - \mathbf{x}^\dagger\|}{\|\mathbf{x}^\dagger\|}, \quad \text{PSNR}(\mathbf{x}) = 20 \log_{10} \left( \frac{nM}{\|\mathbf{x} - \mathbf{x}^\dagger\|} \right),$$

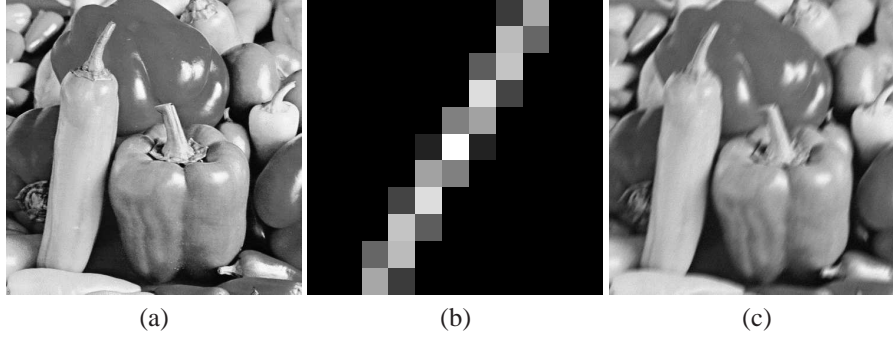


FIGURE 5.1. *Peppers image test problem: (a) True image ( $502 \times 502$  pixels), (b) Gaussian non symmetric PSF ( $11 \times 11$  pixels), (c) Blurred and noisy image with  $\xi = 0.03$  ( $502 \times 502$  pixels).*

where  $n^2$  is the number of elements of  $\mathbf{x}$  and  $M$  denotes the maximum value of  $\mathbf{x}^\dagger$ , the definition of the third is more involved and we refer to [49] for it. Here we just recall that the SSIM measures how well the overall structure of the image is recovered and that the higher the index the better the reconstruction. In particular, the highest value achievable is 1.

For constructing the examples, in order to simulate real test data, we first blur the true image using the periodic bc's, then we cut from the blurred image the boundary pixels affected by the periodic assumption depending on the support of the PSF. Finally, we add white Gaussian noise with noise level  $\xi$  such that

$$\delta = \xi \|\mathbf{b}\|.$$

We compare our MgM Algorithm 3 with other regularization methods recently proposed in the literature

- Approximated Projected Iterated Tikhonov (APIT), see Section 2.2 and [7];
- The Multigrid regularization method (MgM-CGLS) developed in [19];
- Linearized Bregman algorithm (LBA), see [11];
- The Modified Linearized Bregman Algorithm (MLBA-AIT) developed in [13];
- ADMM with unknown boundary conditions (ADMM-UBC) which uses a Total Variation penalty term, see [1];
- The  $\ell_2 - \ell_q$  coupled with the discrepancy principle described in [9];
- The `IRhtv` method in the `IRtools` package [31];
- The `IRell1` method in the `IRtools` package [31].

Some of these methods require the estimation of a parameter, in particular this is true for ADMM-UBC, LBA, and MLBA-AIT. For these methods, we use the parameter which minimizes the RRE (or, equivalently, maximizes the PSNR). ADMM-UBC has a proper treatment of the boundary effects, while for the other algorithms the matrix  $A$  is the same of our MGM, i.e., the same bc's are imposed.

The maximum number of iterations is fixed at 400 for all methods.

*Peppers.* In the first example we consider the peppers image and we blur it with a motion PSF, finally we add white Gaussian noise with  $\xi = 0.03$ . Figure 5.1 shows the true image, the PSF, and the blurred and noisy image. For the deblurring we use the antireflexive bc's.

From the comparison in Table 5.1 of the results obtained with MgM and the other methods considered we can see that our method outperforms all the others in terms of accuracy and SSIM. Moreover, from the visual inspection of the reconstructions in Figure 5.2, we can see that the reconstruction provided by the proposed approach is sharper than the ones obtained with the other methods.

TABLE 5.1

*Peppers test problem: Comparison between the proposed MgM and other methods from the literature. For ADMM-UBC, LBA, and MLBA-AIT the optimal regularization parameter was used. In bold the smallest error and the greatest PSNR and SSIM.*

Method	RRE	PSNR	SSIM	CPU time (sec.)
MgM	<b>0.06649</b>	<b>28.2841</b>	<b>0.74398</b>	65.7461
APIT	0.07966	26.7142	0.64551	3.0685
MgM-CGLS	0.07668	27.0459	0.71774	74.4698
LBA	0.13908	21.8739	0.63246	127.9507
MLBA-AIT	0.07062	27.7611	0.72387	7.9436
ADMM-UBC	0.14474	21.5273	0.61317	8.2949
$\ell^2 - \ell^q$	0.07320	27.4490	0.72364	9.7205
IRhtv	0.15989	20.6631	0.45378	265.1156
IRell1	0.16731	20.2690	0.45302	7.9181

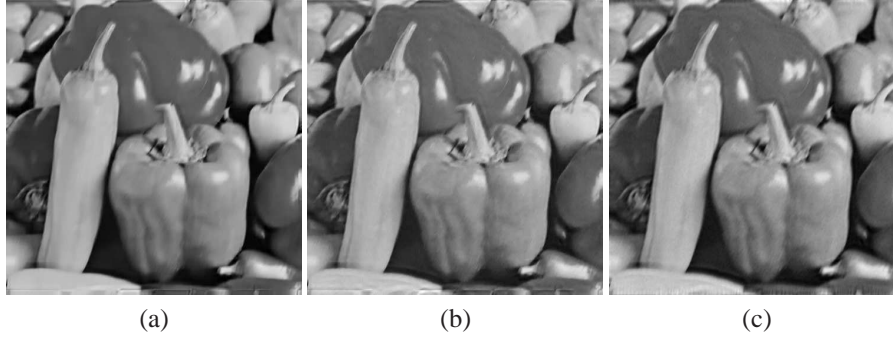


FIGURE 5.2. *Peppers image test problem, reconstructions obtained with different methods: (a) MgM, (b) MLBA-AIT, (c)  $\ell^2 - \ell^q$ .*

*Cameraman.* In the second example we blur the cameraman image with a circular blur PSF and add Gaussian noise so that  $\xi = 0.02$ . Figure 5.3 shows the true image, the PSF, and the blurred and noisy image. We employ the antireflective bc's.

In the considered case the PSF is quadrantly symmetric and thus the operator  $A$  is diagonalizable using the fast antireflective transform [3] and we could chose  $C_0 = A_0 = A$  without introducing any approximation of the operator in the post-smoother. However, we avoid this implementation since there is not a remarkable improvement in the quality of the reconstruction, which is already evident even when the circulant approximation is employed. Moreover, this let us show that AIT, and by extension MgM, is very robust with respect of the choice of the bc's used for the preconditioner.

Table 5.2 reports the results obtained with MgM and the benchmark methods. From this comparison we can see that MgM greatly outperforms most of the other methods in term of accuracy. The error obtained with MLBA-AIT is smaller. However, this error is obtained by hand tuning the regularization parameter inside the method so that the error becomes minimum and small changes in the parameter leads to a larger error. In detail, for the optimum parameter 0.35 the RRE is 0.07911, while when the parameter is set to 0.1 the RRE is 0.083838 and when is set to 1.5 the RRE is 0.083673. Differently, our MgM does not require the tuning of any parameter and has a SSIM larger than the one obtained by all other methods. The only two parameters present in the MgM algorithm are  $\rho$  and  $q$ . This is very important in real applications since it is not possible to rely on the knowledge of the exact

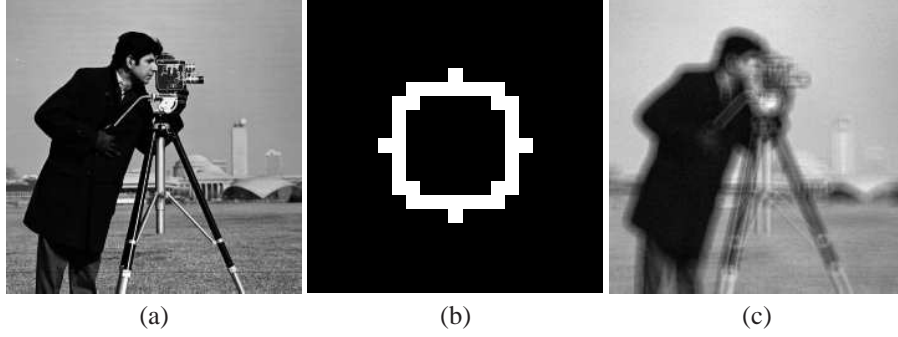


FIGURE 5.3. *Cameraman test problem: (a) True image ( $238 \times 238$  pixels), (b) Circular motion PSF ( $21 \times 21$  pixels), (c) Blurred and noisy image with  $\xi = 0.02$  ( $238 \times 238$  pixels).*

TABLE 5.2

*Cameraman test problem: Comparison between MgM and other methods from the literature. For ADMM-UBC, LBA, and MLBA-AIT the optimal regularization parameter was used. In bold the smallest error and the greatest PSNR and SSIM.*

Method	RRE	PSNR	SSIM	CPU time (sec.)
MgM	0.08259	27.2753	<b>0.83357</b>	17.7704
APIT	0.11637	24.3712	0.62238	2.2912
MgM-CGLS	0.09432	26.2138	0.82228	76.9488
LBA	0.08539	27.0602	0.76072	24.4277
MLBA-AIT	<b>0.07911</b>	<b>27.7236</b>	0.82568	21.2005
ADMM-UBC	0.17498	20.8282	0.69009	2.0235
$\ell^2 - \ell^q$	0.11340	24.5960	0.57334	24.5685
IRhtv	0.11399	24.5510	0.58923	34.6928
IRell1	0.11384	24.5626	0.58173	3.5299

solution for tuning the parameters. Moreover, when the parameters in an algorithm are not set automatically they have to be hand tuned and this process can be sometimes unreliable and extremely time consuming. Finally, we would like to stress that the proposed algorithm is very stable with respect to the choice of these parameters. To show this in Figure 5.4 we show the value of the RRE, PSNR, and SSIM obtained with different choices of  $\rho$  and  $q$ . We can observe that even large changes in these two parameters barely affect the quality of the reconstructed solution. Thus, we can fix these two parameters as described above without compromising the quality of the restoration.

From the visual inspection of the reconstruction in Figure 5.5 we can see how accurate is the approximation given by MgM also in comparison with MLBA-AIT.

*Hubble.* For the last example, we consider an image of the Hubble Telescope blurred with a nonsymmetric PSF and we add white Gaussian noise with  $\xi = 0.05$ . Figure 5.6 shows the true image, the PSF and the blurred and noisy image. Since this is an astronomical image we have employed the zero Dirichlet bc's.

Table 5.3 compares the results obtained with our algorithm against the ones obtained with the other considered methods. MgM gives the highest SSIM while keeping a reasonable computational cost. In Figures 5.7 and 5.8 we can see the different reconstructions obtained with three methods: MgM, ADMM-UBC, and APIT and the blown-ups around the antenna of the telescope. From a visual inspection, note that the reconstruction obtained with ADMM-UBC is not able to reconstruct the black area around the antenna and is affected by heavy

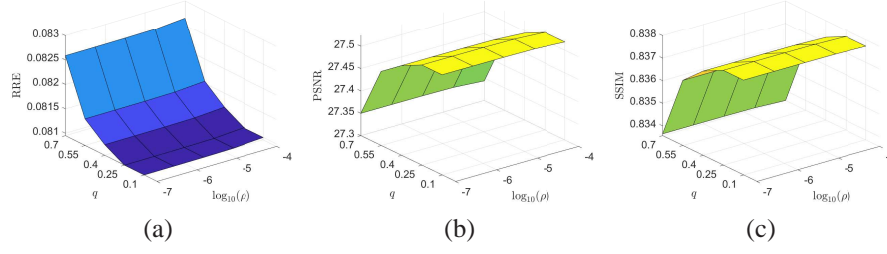


FIGURE 5.4. *Cameraman test problem: Value of the RRE (a), PSNR (b), and SSIM (c) obtained with different parameters  $\rho$  and  $q$ .*

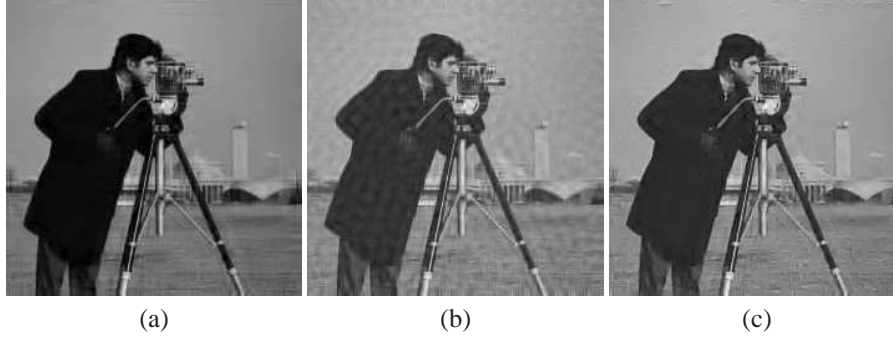


FIGURE 5.5. *Cameraman test problem, reconstructions obtained with different methods: (a) MgM, (b) LBA, (c) MLBA-AIT.*

staircase effect. On the other hand, our MgM is able to correctly recognize as black the background and provides an accurate reconstruction of the antenna of the telescope.

We would like to stress that the most competitive methods among the tested ones, i.e., ADMM-UBC, LBA, and MBLA-AIT, all relies on an estimate of a parameter which has to be hand tuned. All these methods are quite sensible with respect to the selection of the regularization parameter and an imprudent choice can lead to very poor reconstructions. This is confirmed in Figure 5.9 where we plot the RRE obtained with ADMM-UBC, LBA, and MBLA-AIT against their regularization parameters. We can observe that in all three methods the accuracy of the reconstruction deteriorates rapidly if the regularization parameter is chosen far from the optimal one. This is particularly evident for the ADMM-UBC and LBA methods, while the MBLA-AIT method is more stable.

**6. Conclusions.** We have defined a new iterative multigrid method for image deblurring which combines the idea of regularizing multigrid introduced in [23] (and in particular its extension proposed in [19]) with the nonstationary preconditioning developed in [22] and the extensions in [7].

The constructed method can be applied with any boundary condition, preserves the non-negativity of the image pixels, and does not require the estimation of any parameter, provided that an estimation of the norm of the noise is available. This last point is in accordance with the Bakushinskii veto [5] which states that it is not possible to develop a complete convergence analysis of the regularization properties of a method unless some information on the noise is available. A theoretical analysis of the convergence and regularization property of the method is provided.



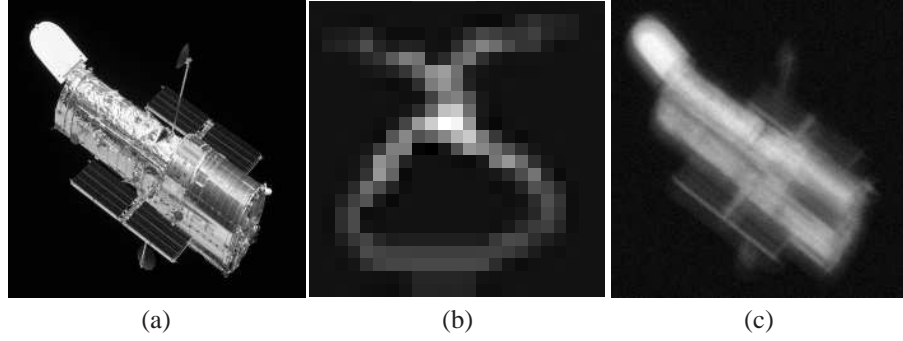


FIGURE 5.6. *Hubble test problem: (a) True image ( $234 \times 234$  pixels), (b) Nonsymmetric PSF ( $17 \times 17$  pixels), (c) Blurred and noisy image with  $\xi = 0.05$  ( $234 \times 234$  pixels).*

TABLE 5.3

*Hubble test problem: Comparison between MgM and other methods from the literature. For ADMM-UBC, LBA, and MLBA-AIT the optimal regularization parameter was used. In bold the smallest error and the greatest PSNR and SSIM.*

Method	RRE	PSNR	SSIM	CPU time (sec.)
MgM	0.14830	26.3100	<b>0.85603</b>	13.2393
APIT	0.16805	25.2242	0.70313	1.3034
MgM-CGLS	0.16140	25.5752	0.71101	0.9462
LBA	0.15976	25.6638	0.58699	21.4703
MLBA-AIT	0.14871	26.2863	0.83654	6.1021
ADMM-UBC	<b>0.14180</b>	<b>26.6995</b>	0.78301	2.5935
$\ell^2 - \ell^q$	0.17360	24.9422	0.52144	21.9041
IRhtv	0.26195	21.3686	0.39063	31.3013
IRell1	0.26270	21.3439	0.38752	1.1514

**Acknowledgments.** The authors would like to thank Silvia Gazzola for the interesting discussion and for providing the codes for some of the benchmark methods. The authors are members of the INdAM Research group GNCS and their work is supported in part by a grant of the group GNCS of INdAM.

#### REFERENCES

- [1] M. S. ALMEIDA AND M. A. FIGUEIREDO, *Deconvolving images with unknown boundaries using the alternating direction method of multipliers*, IEEE Transactions on Image Processing, 22 (2013), pp. 3074–3086.
- [2] A. ARICÒ AND M. DONATELLI, *A V-cycle Multigrid for multilevel matrix algebras: proof of optimality*, Numerische Mathematik, 105 (2007), pp. 511–547.
- [3] A. ARICÒ, M. DONATELLI, J. G. NAGY, AND S. SERRA-CAPIZZANO, *The anti-reflective transform and regularization by filtering*, in Numerical Linear Algebra in Signals, Systems and Control, Springer, 2011, pp. 1–21.
- [4] Z.-J. BAI, D. CASSANI, M. DONATELLI, AND S. SERRA-CAPIZZANO, *A fast alternating minimization algorithm for total variation deblurring without boundary artifacts*, Journal of Mathematical Analysis and Applications, 415 (2014), pp. 373–393.
- [5] A. B. BAKUSHINSKII, *Remarks on the choice of regularization parameter from quasioptimality and relation tests*, Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki, 24 (1984), pp. 1258–1259.
- [6] W. L. BRIGGS, V. E. HENSON, AND S. F. MCCORMICK, *A Multigrid Tutorial, Second Edition*, Society for Industrial and Applied Mathematics, second ed., 2000.
- [7] A. BUCCINI, *Regularizing preconditioners by non-stationary iterated Tikhonov with general penalty term*, Applied Numerical Mathematics, 116 (2017), pp. 64–81.

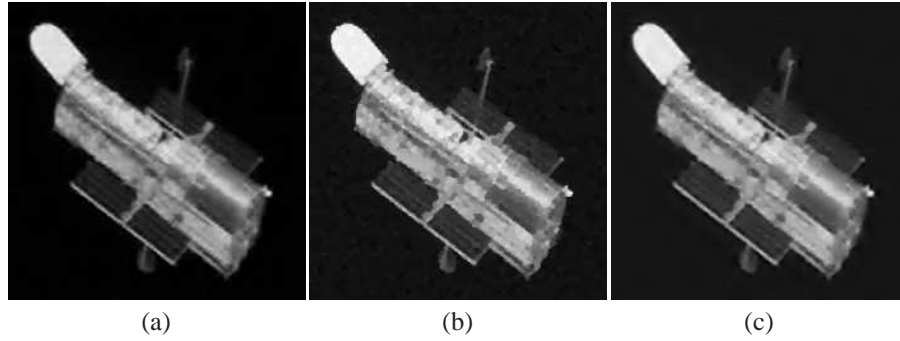


FIGURE 5.7. *Hubble test problem, reconstructions obtained with different methods: (a) MgM, (b) ADMM-UBC, (c) MLBA-AIT.*

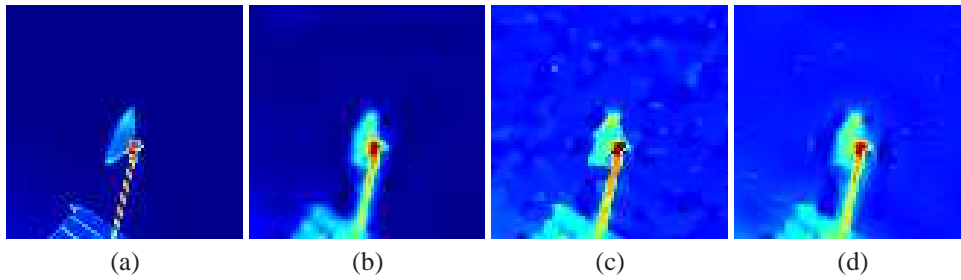


FIGURE 5.8. *Hubble test problem, blown-ups of the reconstruction around the antenna of the Hubble telescope in a different color map: (a) True image, (b) MgM, (c) ADMM-UBC, (d) MLBA-AIT.*

- [8] A. BUCCINI, M. DONATELLI, AND L. REICHEL, *Iterated Tikhonov regularization with a general penalty term*, Numerical Linear Algebra with Applications, 24 (2017), p. e2089.
- [9] A. BUCCINI AND L. REICHEL, *An  $\ell^2 - \ell^q$  regularization method for large discrete ill-posed problems*, Journal of Scientific Computing, 78 (2019), pp. 1526–1549.
- [10] J.-F. CAI, R. H. CHAN, AND Z. SHEN, *A framelet-based image inpainting algorithm*, Applied and Computational Harmonic Analysis, 24 (2008), pp. 131–149.
- [11] J.-F. CAI, S. OSHER, AND Z. SHEN, *Linearized Bregman iterations for frame-based image deblurring*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 226–252.
- [12] J.-F. CAI, S. OSHER, AND Z. SHEN, *Split Bregman methods and frame based image restoration*, Multiscale modeling & simulation, 8 (2009), pp. 337–369.
- [13] Y. CAI, M. DONATELLI, D. BIANCHI, AND T.-Z. HUANG, *Regularization preconditioners for frame-based image deblurring with reduced boundary artifacts*, SIAM Journal on Scientific Computing, 38 (2016), pp. B164–B189.
- [14] R. H. CHAN, T. F. CHAN, , AND W. L. WAN, *Multigrid for differential-convolution problems arising from image processing*, in Processing, in Proceedings of the Workshop on Sci. Comput, Springer-Verlag, 1997, pp. 58–72.
- [15] R. H. CHAN AND K. CHEN, *A multilevel algorithm for simultaneously denoising and deblurring images*, SIAM Journal on Scientific Computing, 32 (2010), pp. 1043–1063.
- [16] P. DELL’ACQUA, *A note on taylor boundary conditions for accurate image restoration*, Advances in Computational Mathematics, (2017).
- [17] M. DONATELLI, *An algebraic generalization of local Fourier analysis for grid transfer operators in multigrid based on Toeplitz matrices*, Numerical Linear Algebra with Applications, 17 (2010), pp. 179–197.
- [18] M. DONATELLI, *Fast transforms for high order boundary conditions in deconvolution problems*, BIT, 50 (2010), pp. 559–576.
- [19] M. DONATELLI, *An iterative multigrid regularization method for Toeplitz discrete ill-posed problems*, Numerical Mathematics: Theory, Methods and Applications, 5 (2012), pp. 43–61.
- [20] M. DONATELLI, *On nondecreasing sequences of regularization parameters for nonstationary iterated Tikhonov*, Numerical Algorithms, 60 (2012), pp. 651–668.
- [21] M. DONATELLI, C. ESTATICO, A. MARTINELLI, AND S. SERRA-CAPIZZANO, *Improved image deblurring*

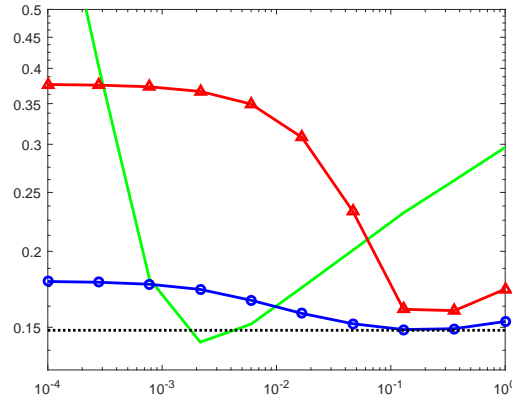


FIGURE 5.9. *Hubble test problem relative reconstruction error against regularization parameter for the ADMM-UBC, LBA, and MLBA-AIT methods. The green line represents the ADMM-UBC method, the red line the LBA method, and the blue line the MLBA-AIT method. The dotted black line is the RRE obtained with MgM. This line is constant since it does not depend on any parameter.*

- with anti-reflective boundary conditions and re-blurring, *Inverse problems*, 22 (2006), p. 2035.
- [22] M. DONATELLI AND M. HANKE, *Fast nonstationary preconditioned iterative methods for ill-posed problems, with application to image deblurring*, *Inverse Problems*, 29 (2013), p. 095008.
- [23] M. DONATELLI AND S. SERRA-CAPIZZANO, *On the regularizing power of multigrid-type algorithms*, *SIAM Journal on Scientific Computing*, 27 (2006), pp. 2053–2076.
- [24] M. DONATELLI AND S. SERRA-CAPIZZANO, *Filter factor analysis of an iterative multilevel regularizing method*, *Electron. Trans. Numer. Anal.*, 29 (2007/2008), pp. 163–177.
- [25] M. DONATELLI AND S. SERRA-CAPIZZANO, *Antireflective boundary conditions for deblurring problems*, *Journal of Electrical and Computer Engineering*, 2010 (2010), p. 2.
- [26] D. L. DONOHO, *De-noising by soft-thresholding*, *IEEE transactions on information theory*, 41 (1995), pp. 613–627.
- [27] H. W. ENGL, M. HANKE, AND A. NEUBAUER, *Regularization of inverse problems*, vol. 375, Springer Science & Business Media, 1996.
- [28] M. I. ESPANOL AND M. E. KILMER, *Multilevel approach for signal restoration problems with toeplitz matrices*, *SIAM Journal on Scientific Computing*, 32 (2010), pp. 299–319.
- [29] M. I. ESPANOL AND M. E. KILMER, *A wavelet-based multilevel approach for blind deconvolution problems*, *SIAM Journal on Scientific Computing*, 36 (2014), pp. A1432–A1450.
- [30] G. FIORENTINO AND S. SERRA, *Multigrid methods for Toeplitz matrices*, *Calcolo*, 28 (1991), pp. 283–305.
- [31] S. GAZZOLA, P. C. HANSEN, AND J. G. NAGY, *Ir tools: a matlab package of iterative regularization methods and large-scale test problems*, *Numerical Algorithms*, (2018), <https://doi.org/10.1007/s11075-018-0570-7>, <https://doi.org/10.1007/s11075-018-0570-7>.
- [32] M. HANKE, *A regularizing Levenberg-Marquardt scheme, with applications to inverse groundwater filtration problems*, *Inverse problems*, 13 (1997), p. 79.
- [33] M. HANKE AND P. C. HANSEN, *Regularization methods for large-scale problems*, *Surv. Math. Ind.*, 3 (1993), pp. 253–315.
- [34] M. HANKE AND R. C. VOGEL, *Two-level preconditioners for regularized inverse problems i: Theory*, *Numerische Mathematik*, 83 (1999), pp. 385–402.
- [35] P. C. HANSEN, J. G. NAGY, AND D. P. O’LEARY, *Deblurring images: matrices, spectra, and filtering*, vol. 3, Siam, 2006.
- [36] G. HUANG, L. REICHEL, AND F. YIN, *Projected nonstationary iterated Tikhonov regularization*, *BIT Numerical Mathematics*, 56 (2016), pp. 467–487.
- [37] T. HUCKLE AND J. STAUDACHER, *Multigrid preconditioning and Toeplitz matrices*, *Electronic Transactions on Numerical Analysis*, 13 (2002), pp. 81–105.
- [38] B. KALTENBACHER, *On the regularizing properties of a full multigrid method for ill-posed problems*, *Inverse Problems*, 17 (2001), p. 767.
- [39] T. J. KING, *Multilevel algorithms for ill-posed problems*, *Numerische Mathematik*, 61 (1992), pp. 311–334.
- [40] S. MORIGI, L. REICHEL, F. SGALLARI, AND A. SHYSHKOV, *Cascadic multiresolution methods for image*

- deblurring*, SIAM Journal on Imaging Sciences, 1 (2008), pp. 51–74.
- [41] M. K. NG, R. H. CHAN, AND W.-C. TANG, *A fast algorithm for deblurring models with neumann boundary conditions*, SIAM Journal on Scientific Computing, 21 (1999), pp. 851–866.
  - [42] S. REEVES, *Fast image restoration without boundary artifacts*, IEEE Transactions on Image Processing, 14 (2005), pp. 1448–1453.
  - [43] L. REICHEL AND A. SHYSHKOV, *Cascadic multilevel methods for ill-posed problems*, Journal of Computational and Applied Mathematics, 233 (2010), pp. 1314 – 1325.
  - [44] A. RIEDER, *A wavelet multilevel method for ill-posed problems stabilized by Tikhonov regularization*, Numerische Mathematik, 75 (1997), pp. 501–522.
  - [45] J. W. RUGE AND K. STUBEN, *Algebraic multigrid*, in Multigrid Methods, S. F. McCormick, ed., SIAM, Philadelphia, Pennsylvania, 1987, pp. 73–130.
  - [46] S. SERRA-CAPIZZANO, *A note on antireflective boundary conditions and fast deblurring models*, SIAM Journal on Scientific Computing, 25 (2004), pp. 1307–1325.
  - [47] U. TROTTEBERG, C. W. OOSTERLEE, AND A. SCHULLER, *Multigrid*, Academic press, 2000.
  - [48] R. VIO, J. BARDSLEY, M. DONATELLI, AND W. WAMSTEKER, *Dealing with edge effects in least-squares image deconvolution problems*, Astronomy and Astrophysics, 442 (2005), pp. 397–403.
  - [49] Z. WANG, A. C. BOVIK, H. R. SHEIKH, AND E. P. SIMONCELLI, *Image quality assessment: from error visibility to structural similarity*, IEEE Transactions on Image Processing, 13 (2004), pp. 600–612.
  - [50] E. H. ZARANTONELLO, *Projections on convex sets in Hilbert space and spectral theory*, University of Wisconsin, 1971.