Electronic Transactions on Numerical Analysis.
Volume 53, pp. 283–312, 2020.

ISSN 1068–9613.
DOI: 10.1553/etna_vol53s283

ETNA
Kent State University and
Johann Radon Institute (RICAM)

A MULTIGRID FRAME BASED METHOD FOR IMAGE DEBLURRING*

ALESSANDRO BUCCINI† AND MARCO DONATELLI‡

Abstract. Iterative soft thresholding algorithms combine one step of a 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 ℓ_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 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 onto 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 operator of the first 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

$$g(x,y) = \int K(s-x, t-y) f(s,t) \mathrm{d}s\mathrm{d}t = K * f, \tag{1.1}$$

where f denotes the true signal and K is a possibly smooth integral kernel with compact support. K is usually referred to as a point spread function (PSF) since it represents how a single point is spread across its neighborhood. Because K has compact support, equation (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 matrix 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 of size $q \times q$ 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

$$\tilde{A}\tilde{\mathbf{x}} = \mathbf{b}, \tag{1.2}$$

where $\tilde{\mathbf{x}}$ and $\mathbf{b}$ are vectors of size $q^2 \times 1$ 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., by making assumptions on the behavior of the true image outside the FOV, we may consider only the

*Received October 23, 2019. Accepted January 25, 2020. Published online on February 25, 2020. Recommended by Giuseppe Rodriguez.
†Dipartimento di Matematica e Informatica, Università degli Studi di Cagliari, Via Ospedale 72, 09124 Cagliari, Italy (alessandro.buccini@unica.it).
‡Dipartimento di Scienza e Alta Tecnologia, Università degli Studi dell'Insubria, Via Valleggio 11, 22100 Como, Italy (marco.donatelli@uninsubria.it).

ETNA
Kent State University and
Johann Radon Institute (RICAM)

$n \times n$ internal points of $\tilde{\mathbf{X}}$, denoted by $\mathbf{X}$, and obtain a square matrix A such that

$$A\mathbf{x} = \mathbf{b}, \tag{1.3}$$

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 allow 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

$$\left\|\mathbf{b} - \mathbf{b}^\delta\right\| \leq \delta, \tag{1.4}$$

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

$$A\mathbf{x} = \mathbf{b}^\delta. \tag{1.5}$$

Throughout this paper we will assume that a fairly sharp bound for δ 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 methods is Tikhonov regularization,

$$\min_{\mathbf{x}} \left\|A\mathbf{x} - \mathbf{b}^\delta\right\|^2 + \alpha \left\|\mathbf{x}\right\|^2, \tag{1.6}$$

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 α 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 is formulated by the following iteration:

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

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 number is reached. After that, the noise present in $\mathbf{b}^\delta$ is included in the reconstruction and gets amplified, thus deteriorating and eventually destroying the quality of the reconstruction. Therefore, a regularization effect is obtained by an early stopping of the iteration. It is clear that it is crucial to stop the iterations close to the optimal iteration number, which is unknown.

Unfortunately, the iteration (1.7) requires at each step the inversion of either $A^tA + \alpha_k I$ or $AA^t + \alpha_k I$, 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

ETNA
Kent State University and
Johann Radon Institute (RICAM)

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

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

where α_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 Ω is a closed and convex set, e.g., for image deblurring a possible choice for Ω is the nonnegative cone. The APIT (Approximated Projected Iterated Tikhonov) algorithm is obtained by projecting at each step the computed approximation into 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 for solving 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 methods. 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 a 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 a multigrid method. Firstly, different from previous works, we define a general coarsening strategy independent of the bc's. This strategy consists in applying a 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 constructing a sequence of continuous operators by the Galerkin approach and then discretizing 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 the fast Fourier transform (FFT). Furthermore, our proposal differs from the one in [19] also in the usage of framelet denoising and in the usage of the above described APIT algorithm instead of CGLS as inner iterative regularization method. Because APIT projects at each iteration into a closed and convex set Ω, this choice let us ensure that the provided approximation lies in Ω as well. Thus, using APIT as post-smoother is equivalent to projecting each multigrid iteration into Ω.

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 the proposed method is an iterative regularization method, i.e., that regularization is achieved by an early stopping of the iteration

ETNA
Kent State University and
Johann Radon Institute (RICAM)

and that, denoting by $\mathbf{x}^\delta$ the reconstruction obtained with the noise level δ, it holds that $\mathbf{x}^\delta \to \mathbf{x}^*$ as $\delta \to 0$, 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 δ, it 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 potential 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 a 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, then 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, which are already present in the literature, used for constructing our algorithm.

2.1. The structure of the blurring matrix. As already mentioned in the introduction, a popular method to deal with the boundary artifacts is to impose 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 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, then 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 square matrices. Our proposal could be adapted also to the other approaches for dealing with the boundary artifacts whenever the smoother is properly adjusted.

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 the PSF. Given the pixels h_{j_1,j_2} of the PSF, it is possible to associate the so-called

ETNA
Kent State University and
Johann Radon Institute (RICAM)

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
$\mathbf{0}\ \mathbf{0}\ \mathbf{0}$	$\mathbf{X}\ \mathbf{X}\ \mathbf{X}$	$\mathbf{X}_{rc}\ \mathbf{X}_r\ \mathbf{X}_{rc}$
$\mathbf{0}\ \mathbf{X}\ \mathbf{0}$	$\mathbf{X}\ \mathbf{X}\ \mathbf{X}$	$\mathbf{X}_c\ \mathbf{X}\ \mathbf{X}_c$
$\mathbf{0}\ \mathbf{0}\ \mathbf{0}$	$\mathbf{X}\ \mathbf{X}\ \mathbf{X}$	$\mathbf{X}_{rc}\ \mathbf{X}_r\ \mathbf{X}_{rc}$

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

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

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 that can be exploited to achieve fast computations. The matrix-vector product can always be performed by an 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 we refer to [35]; for antireflective bc's, see the review paper [25] and the original proposal in [46]. More sophisticated bc's 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

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

for $k = -n+1, \ldots, n-1$. Unfortunately, there are no fast transformations to diagonalize a general matrix of this form. This represents an important drawback of the use of 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 periodically 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 $F_1 = [\mathrm{e}^{-2jk\mathrm{i}\pi/n}]_{j,k=0}^{n-1} \in \mathbb{C}^{n \times n}$ 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 continuous on the boundaries. The resulting matrix A has a more involved structure than the

ETNA
Kent State University and
Johann Radon Institute (RICAM)

(a) (b) (c) (d)

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

ones obtained with periodic or zero bc's. Nevertheless, it can be diagonalized by the discrete cosine transform if the PSF is quadrantally 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 also its normal derivative is continuous. 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 quadrantally symmetric [3, 21, 46]. Figure 2.1 shows an example of the bc's described above.

2.2. Approximated Iterated Tikhonov method. 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 included it into the Tikhonov method (1.6) by computing

$$\min_{\mathbf{x}} \left\| A\mathbf{x} - \mathbf{b}^\delta \right\|^2 + \alpha \left\| \mathbf{x} - \mathbf{x}_0 \right\|^2 \qquad \Longleftrightarrow \qquad \min_{\mathbf{h}} \left\| A\mathbf{h} - \mathbf{r}_0 \right\|^2 + \alpha \left\| \mathbf{h} \right\|^2,$$

where $\mathbf{r}_0 = \mathbf{b}^\delta - A\mathbf{x}_0$ 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 $k = 0, 1, \ldots$,

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

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

The choice of α_k in the IT method plays a crucial role, and many strategies have been proposed. If the same α is used for each iteration, then we call the method stationary, whereas if α 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 α. A generalization to the case where a matrix different from the identity is used in the IT iterations has recently been 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_δ such that

$$\left\| \mathbf{r}_{k_\delta} \right\| \leq \tau\delta < \left\| \mathbf{r}_k \right\|, \qquad k = 0, 1, \ldots, k_\delta - 1,$$

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

Note that the residual $\mathbf{r}_k$ at item 1. is computed using the observed image $\mathbf{b}^\delta$ instead of the true blurred image $\mathbf{b}$. Therefore, the error equation, used at item 2. for computing the

ETNA
Kent State University and
Johann Radon Institute (RICAM)

refinement of the previous approximation, is correct up to the noise in the observation. Hence, we can replace the matrix A with an approximation C if the further introduced error is lower than the noise level δ. This is a possible interpretation of the preconditioned iteration proposed in [22], which accordingly requires the following assumption:

ASSUMPTION 1. *Let C be such that*

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

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

Note that the above 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 ρ 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 α_k is such that

$$\|\mathbf{r}_k - C\mathbf{h}_k\| = q_k \|r_k\|, \tag{2.2}$$

and

- *update*

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

We summarize the main theoretical results for 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 reduce the quality of the computed solution. Firstly, we recall the monotone decrease 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 $\tau = (1+2\rho)/(1-2\rho)$. As long as $\|\mathbf{r}_k\| > \tau\delta$, the norm of the reconstruction error $\mathbf{e}_k$ decreases monotonically, namely $\|\mathbf{e}_{k+1}\| \leq \|\mathbf{e}_k\|$, for $k = 0, 1, \ldots$*

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

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

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

ETNA
Kent State University and
Johann Radon Institute (RICAM)

for some constant $c > 0$, depending only on ρ 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 $k \to \infty$.

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 $k \to \infty$ 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 τ and q, denote by $\mathbf{x}^\delta$ the resulting approximation obtained with AIT. Then for $\delta \to 0$ we have that $\mathbf{x}^\delta \to \mathbf{x}_0^\dagger$ which is the nearest solution of (1.3) to $\mathbf{x}_0$.*

In [7] an extension of the AIT algorithm has been proposed which constrains the iterations to lie in a closed and convex set $\Omega \subset \mathbb{R}^N$. An example of a suitable Ω in the case of image deblurring is the nonnegative cone. In fact, since images are measurements of a quantity of light, they can not assume negative values. With the assumption that $\mathbf{x}^\dagger \in \Omega$, denoting by P_Ω the metric projection onto Ω, the iterations of the projected AIT (APIT) can be computed simply by replacing 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 with 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 α_k in (2.2) is linear in N since it requires only a 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. Then

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

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

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

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 $\mathbf{c}_1$ again, thus obtaining $\mathbf{c}_2 = W_0^{(1)}\mathbf{c}_1$ and $\mathbf{d}_2 = W_1^{(1)}\mathbf{c}_1$, where we indicated with a superscript $^{(1)}$ the various operators on the (possibly) smaller space in which $\mathbf{c}_1$ lives. For each level $j = 0, \ldots, l-1$, starting from $\mathbf{c}_0 = \mathbf{x}$, we have that

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

ETNA
Kent State University and
Johann Radon Institute (RICAM)

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

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

where $\operatorname{sgn}(x)$ 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 θ is of crucial importance. According to [26], we use

$$\theta = c\sqrt{\frac{2\log n}{n}}, \tag{2.4}$$

where $c > 0$ is a constant that for Gaussian noise can be chosen as $c = \delta / \left\|\mathbf{b}^\delta\right\|$. 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

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

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

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

if $lev = l$

$$\mathbf{y} = \mathbf{z}$$

else

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

end

The tight frame used in the numerical results is the one related to linear B-splines; see, e.g., [11, 12] and the 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 projecting work on the signal and the residual, respectively, but in practice, this combination provides better restorations compared 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

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

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

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

ETNA
Kent State University and
Johann Radon Institute (RICAM)

and

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

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

$$W_{ij} = W_i \otimes W_j, \qquad i, j = 0, 1, 2. \tag{2.5}$$

Thus, the analysis operator is

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

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 projections. 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 convergence can be very slow in the ill-conditioned space.

REMARK 2.6. The definition of the well- and ill-conditioned space is informal. 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

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

The *well-conditioned space* is the space W such that κ_W is not too large, whereas the *ill-conditioned space* I is the one where κ_I is very large. For matrices derived from the discretization of a 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 kth step, apply ν_1 steps of an iterative method to $\mathbf{x}_k$ obtaining

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

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 denote by $R \in \mathbb{R}^{N_1 \times N}$ 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

ETNA
Kent State University and
Johann Radon Institute (RICAM)

$R = cP^t$ 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 $\tilde{\mathbf{x}}_k$, a coarser problem is defined using the 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 $(k+1)$st approximation, we can apply ν_2 steps of an iterative method, called post-smoothing, which may 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 $N = N_0 > N_1 > \ldots > N_L > 0$. We denote by R_i and P_i the ith restriction and interpolation operator, respectively, so that it holds that $R_i \in \mathbb{R}^{N_{i+1} \times N_i}$ and $P_i \in \mathbb{R}^{N_i \times N_{i+1}}$, for $i = 0, \ldots, L-1$. Then

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

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 every second pixel in each direction (downsampling), it holds that $N_i = 2^{2(L-i)}$, for $i = 0, \ldots, L$, thus $N_L = 1$.

Summarizing, the single step of the multigrid iteration reads as follows:

$$\mathbf{y}_i = \text{V-Step}(\mathbf{x}_i, A_i, \mathbf{b}_i, i, L)$$

```
if (i = L) then y_L  = Solve(A_L y_L = b_L)
           else x̃_i  = Pre-Smooth(A_i, b_i, x_i, ν_1)
                r_{i+1} = R_i(b_i − A_i x̃_i)
                e_{i+1} = V-Step(0, A_{i+1}, r_{i+1}, i+1, L)
                x̂_i  = x̃_i + P_i e_{i+1}
                y_i  = Post-Smooth(A_i, b_i, x̂_i, ν_2)
end
```

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

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

ETNA
Kent State University and
Johann Radon Institute (RICAM)

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 downsampling operator which preserves the same structure at the coarser level. In such a way, for two dimensional problems, the coarser matrix is generated by the symbol

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

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 matrix generated by f_{next} independently of the used downsampling. This approach can be interpreted as a "geometric" multigrid for structured matrices, where at the coarser level, the rediscretization of the same operator associated to f is not used but those associated to f_{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 between the sizes 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 also allows the inversion of the coarser matrix if N_L is small enough. Note that using a projection into the well-conditioned subspace, the multigrid method 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 that is necessary to define the coarser problems, and then we describe our pre- and post-smoother.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

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

$$P_i = \frac{1}{4} R_i^t, \qquad i = 0, \ldots, L-1, \tag{3.1}$$

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

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

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 downsampling 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 downsampling operator which keeps every second component. Namely, $\tilde{K}_d^{(i)}$ can be written as an $n_{i+1} \times n_i$ matrix such that

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

or

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

for n_i even and odd, respectively. Therefore, $N_i = n_i^2$, for $i = 0, \ldots, L$, 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

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

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 now provide some details on the implementation. Working on 2D problems we store the data in bidimensional arrays. Using Matlab notation, we denote by $\mathsf{vec}(\cdot)$ the function

ETNA
Kent State University and
Johann Radon Institute (RICAM)

that returns a vector containing the columns of the input matrix stacked below each other. Let $\mathbf{X} \in \mathbb{R}^{n_i \times n_i}$ and $\mathbf{x} = \mathsf{vec}(\mathbf{X})$. The restriction of the vector $\mathbf{x}$ can be computed by

$$\begin{aligned} R_i\mathbf{x} = K_d^{(i)}\mathcal{T}_{n_i}(p)\mathbf{x} &= \left(\tilde{K}_d^{(i)} \otimes \tilde{K}_d^{(i)}\right)\mathcal{T}_{n_i}(p)\mathsf{vec}(\mathbf{X}) \\ &= \left(\tilde{K}_d^{(i)} \otimes \tilde{K}_d^{(i)}\right)\mathsf{vec}\left(M * \mathbf{X}\right) = \mathsf{vec}\left(\tilde{K}_d^{(i)}\left(M * \mathbf{X}\right)\left(\tilde{K}_d^{(i)}\right)^t\right), \end{aligned}$$

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:

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

A crucial choice concerns the construction of the matrices A_i. For general bc's and an 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, then $A_1 = R_1AP_1$ does not have the same structure as 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

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

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 deteriorating the quality of the computed solution simply by defining A_{i+1} as BCCB matrices. This follows from the fact that the CGC operates 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 for 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., $A_{i+1} = R_iA_iP_i$.

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

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

The matrices $\mathtt{PSF}_i$ 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 FFTs 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.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

3.2. Smoothing. The *pre-smoother* is the framelet denoising described in Section 2.3. This pre-smoother is able to keep the effects of the noise under control 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 negative values of the determined approximation.
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 $\mathbf{0}$ independently of the parameter θ. 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 of 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 θ for the denoising is chosen in a nonstationary way. We use the following sequence of parameters

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

where $0 < p < 1$ and k denotes the iteration. Note that θ_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, we thus 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 α_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 δ_i can be estimated by

$$\delta_i = \frac{\delta_{i-1}}{2}, \qquad i = 1, \ldots, L-1. \tag{3.5}$$

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 α_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 using APIT as post-smoother at the finest level.

The projection onto 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 onto the nonnegative cone

ETNA
Kent State University and
Johann Radon Institute (RICAM)

may deteriorate the quality of the computed solution since it may destroy relevant information. More 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 as in 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^δ is

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

where ρ is defined in (2.1).

As mentioned above, we treat the finest level differently 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 such 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. Hence, the matrices A_i, for $i = 1, \ldots, L$, are BCCB, and all the required computations can be done very efficiently by FFTs.

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

$$\mathbf{x} = \mathsf{MGM}(\mathbf{x_0}, A, \mathbf{b}^\delta)$$

$k = 1$
While $\left\| A\mathbf{x}_k - \mathbf{b}^\delta \right\| > \frac{1+2\rho}{1-2\rho}\delta$
$\quad \mathbf{x_k} = \mathsf{MGM\ Single\ Step}(\mathbf{x}_{k-1}, A, \mathbf{b}^\delta, 0, l, L)$
$\quad k = k + 1$
end

The single step of the algorithm is defined by

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

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

ETNA
Kent State University and
Johann Radon Institute (RICAM)

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

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

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 the quality of the final restoration only marginally. 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 method 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, is applied only at the finest level. Hence, since the computational cost at each coarser level $i = 1, \ldots, L$ 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 a 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 Ω 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 averages. Hence, by 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

$$A_1 = 4 P^t A P.$$

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

ALGORITHM 4 (TGM). *Consider the system* (1.5)*. Let the parameter* θ_k *be chosen as in* (3.4)*, and choose the number of framelet levels* l *to which we apply the denoising. Let* ρ *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*

ETNA
Kent State University and
Johann Radon Institute (RICAM)

guess for the solution of (1.5).

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

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

Note that α_k in Algorithm 4 is chosen by solving the non linear equation

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

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

We define the following errors

$$\mathbf{e}_k = \mathbf{x}^\dagger - \mathbf{x}_k, \qquad \tilde{\mathbf{e}}_k = \mathbf{x}^\dagger - \tilde{\mathbf{x}}_k, \qquad \hat{\mathbf{e}}_k = \mathbf{x}^\dagger - \hat{\mathbf{x}}_k. \tag{4.1}$$

For convenience, we also define

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

In order to prove convergence, we need the following assumptions:

ASSUMPTION 2. *The thresholding parameters θ_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 this is reasonable for a converging algorithm. Since the sequence $(\theta_k)_k$ is strictly decreasing, 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 since P represents 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 result:

ETNA
Kent State University and
Johann Radon Institute (RICAM)

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 that*

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

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 that

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

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

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

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 that*

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

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

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

where in order 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

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

ETNA
Kent State University and
Johann Radon Institute (RICAM)

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 α_k, we have that

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

Thus,

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

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

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

which concludes the proof. □

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

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

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

Proof. This is direct consequence of the iterated application of Proposition 4.3 combined with Corollary 2.3. □

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 has to be the case that

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

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 convergence of the 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 Ω 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 are 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 $\hat{\mathbf{h}}_{k-1}$ coincides 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 obtain that $\mathcal{N}(A) = \mathcal{N}(C)$, and from the definition of α_{k-1} and Lemma 2.1 it follows that

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

ETNA
Kent State University and
Johann Radon Institute (RICAM)

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 hence that it converges. Let $m+1 \geq l$ and consider the following estimates:

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

Let $n+1 \leq l$. Similarly as before we obtain

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

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

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

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

ETNA
Kent State University and
Johann Radon Institute (RICAM)

Since all the entries of A are nonnegative and Ω is the nonnegative cone, it follows that

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

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 $\|\hat{\mathbf{r}}_k\| \to 0$ as $k \to \infty$. Then from (4.3) we obtain that $\|\mathbf{r}_k\| \to 0$ as $k \to \infty$ and thus that the limit of the sequence $\{\mathbf{x}_k\}_k$ is a solution of the noise-free system. □

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 δ it holds that $\left\|\mathbf{b} - \mathbf{b}^\delta\right\| \leq \delta$. For fixed τ and q, denote by $\mathbf{x}^\delta$ the approximation of $\mathbf{x}^\dagger$ obtained with Algorithm 4. Then, as $\delta \to 0$, $\mathbf{x}^\delta$ tends 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 $i = 0, \ldots, L$, the approximation C_i, for $i = 0, \ldots, L-1$, and the parameters ρ_i and q_i, for $i = 0, \ldots, L-1$.

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 $i = 1, \ldots, L$, as the blurring matrix with the PSF $\texttt{PSF}_i$ defined as in (3.3) and with periodic bc's.

We set C_i, for $i = 0, \ldots, L-1$, as in [7, 22] as the blurring matrix associated to $\texttt{PSF}_i$ and with periodic bc's. In this way the computation of $C_i^t(C_iC_i^t + \alpha I)^{-1}$ 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 the IT method, where the regularization parameter is chosen as in AIT. This choice ensures that Assumption 1 is trivially satisfied for any ρ, and, moreover, let us further dampen the computational cost.

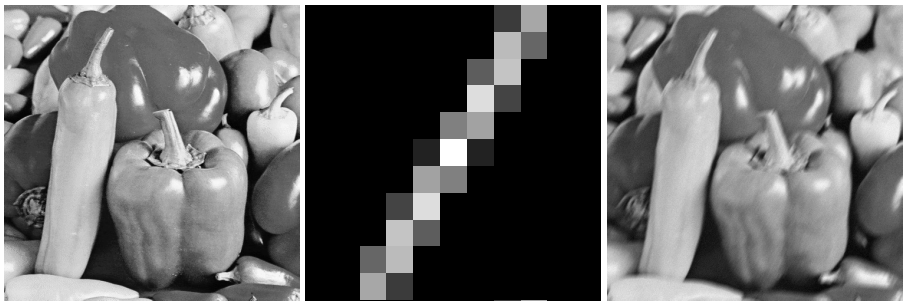
We set $\rho_i = 10^{-4}$, $q_0 = 0.7$, $q_i = 1$ for the levels $i = 1, \ldots, L-1$, and we fix $l = 4$, where l denotes the number of decompositions in the denoising algorithm (see Algorithm 2). Finally, we set L so that $A_L \in \mathbb{R}$, i.e., the problem at the coarsest level is reduced to a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

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

ETNA
Kent State University and
Johann Radon Institute (RICAM)



(a) (b) (c)

FIG. 5.1. *Peppers image test problem: (a) true image* (502×502 *pixels), (b) Gaussian non symmetric PSF* (11×11 *pixels), (c) blurred and noisy image with* $\xi = 0.03$ (502×502 *pixels).*

where n^2 is the the number of elements of $\mathbf{x}$ and M denotes the maximum value of $\mathbf{x}^\dagger$; the definition of the third one 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 periodicity assumption depending on the support of the PSF. Finally, we add white Gaussian noise with a noise level ξ 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 as for 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.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

TABLE 5.1
Peppers test problem: Comparison of 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	**0.06649**	**28.2841**	**0.74398**	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

(a) (b) (c)

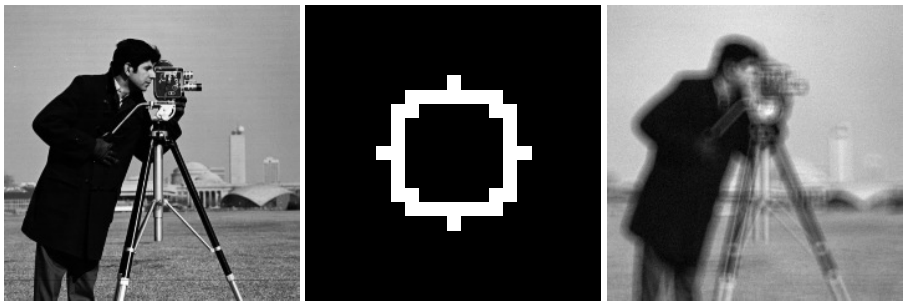
FIG. 5.2. *Peppers image test problem, reconstructions obtained with different methods: (a) MgM, (b) MLBA-AIT, (c)* $\ell^2 - \ell^q$.

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 quadrantally 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 no 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 minimal, and small changes in the parameter leads to a larger error. In detail, for the optimal 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 ρ and q. This is very important in real applications since it is not possible to rely on the knowledge of the exact solution for tuning

ETNA
Kent State University and
Johann Radon Institute (RICAM)



(a) (b) (c)

FIG. 5.3. *Cameraman test problem: (a) true image (*238×238 *pixels), (b) circular motion PSF (*21×21 *pixels), (c) blurred and noisy image with* $\xi = 0.02$ *(*238×238 *pixels).*

TABLE 5.2
Cameraman test problem: Comparison of 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	**0.83357**	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	**0.07911**	**27.7236**	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

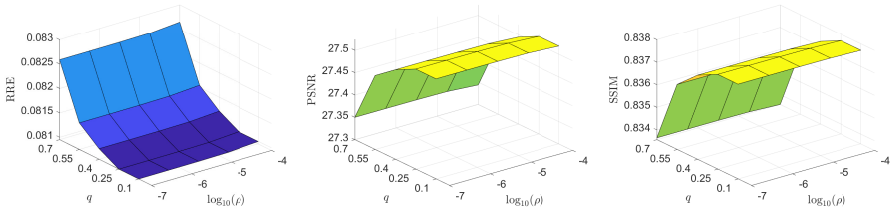
the parameters. Moreover, when the parameters in an algorithm are not set automatically, they have to be hand-tuned, and this process can be sometimes unreliable and extremely time consuming. Finally, we would like to stress that the proposed algorithm is very stable with respect to the choice of these parameters. To illustrate this in Figure 5.4, we present the value of the RRE, PSNR, and SSIM obtained with different choices of ρ 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 the approximation given by MgM is 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 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 observe the different reconstructions obtained with three methods: MgM, ADMM-UBC, and APIT, and blow-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

ETNA
Kent State University and
Johann Radon Institute (RICAM)



(a) (b) (c)

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

(a) (b) (c)

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

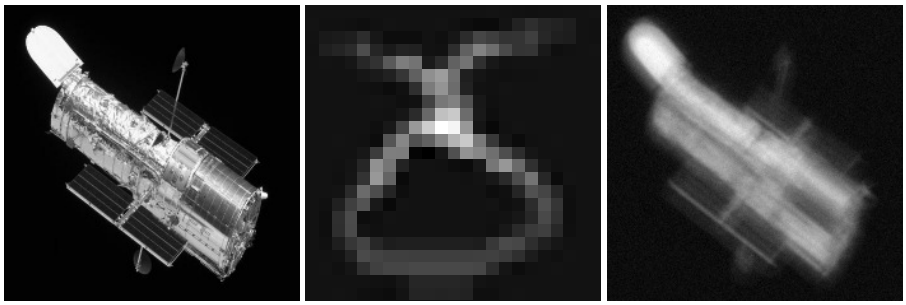
by a heavy staircase effect. On the other hand, our MgM is able to correctly recognize the background as black and provides an accurate reconstruction of the antenna of the telescope.

We would like to stress that the most competitive methods among the tested ones, i.e., ADMM-UBC, LBA, and MBLA-AIT, all rely 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 a regularizing multigrid method 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 nonnegativity 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 the regularization property of the method is provided.

ETNA
Kent State University and
Johann Radon Institute (RICAM)



(a) (b) (c)

FIG. 5.6. *Hubble test problem: (a) true image (*234×234 *pixels), (b) nonsymmetric PSF (*17×17 *pixels), (c) blurred and noisy image with* $\xi = 0.05$ *(*234×234 *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.

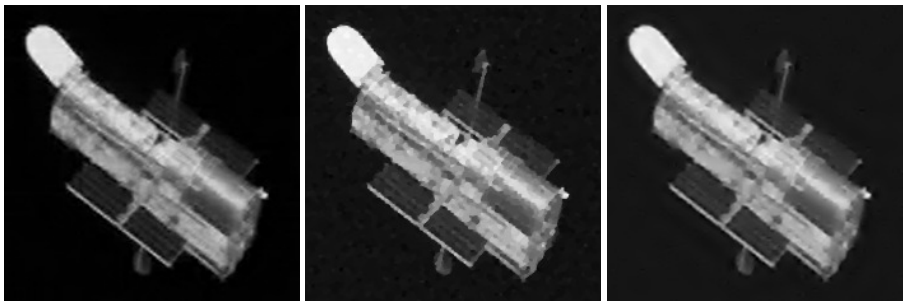
Method	RRE	PSNR	SSIM	CPU time (sec.)
MgM	0.14830	26.3100	**0.85603**	13.2393
APIT	0.16805	25.2242	0.70313	1.3034
MgM-CGLS	0.16140	25.5752	0.71101	0.9462
LBA	0.15976	25.6638	0.58699	21.4703
MLBA-AIT	0.14871	26.2863	0.83654	6.1021
ADMM-UBC	**0.14180**	**26.6995**	0.78301	2.5935
$\ell^2 - \ell^q$	0.17360	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

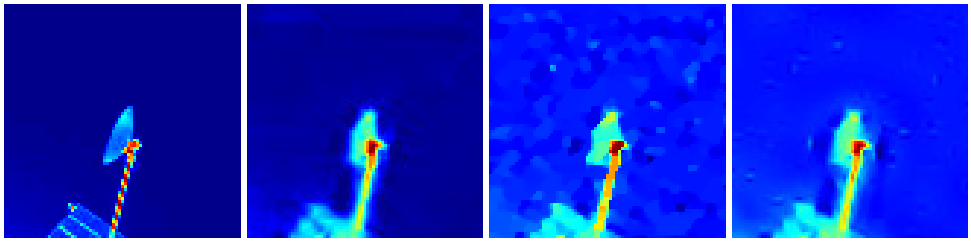
[1] M. S. C. ALMEIDA AND M. A. T. FIGUEIREDO, *Deconvolving images with unknown boundaries using the alternating direction method of multipliers*, IEEE Trans. Image Process., 22 (2013), pp. 3074–3086.
[2] A. ARICÒ AND M. DONATELLI, *A V-cycle multigrid for multilevel matrix algebras: proof of optimality*, Numer. Math., 105 (2007), pp. 511–547.
[3] A. ARICÒ, M. DONATELLI, J. NAGY, AND S. SERRA-CAPIZZANO, *The anti-reflective transform and regularization by filtering*, in Numerical Linear Algebra in Signals, Systems and Control, P. Van Dooren, S. P. Bhattacharyya, R. H. Chan, V. Olshevsky, and A. Routray, eds., vol. 80 of Lect. Notes Electr. Eng., Springer, Dordrecht, 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*, J. Math. Anal. Appl., 415 (2014), pp. 373–393.
[5] A. B. BAKUSHINSKIĬ, *Remarks on the choice of regularization parameter from quasioptimality and relation tests*, Zh. Vychisl. Mat. i Mat. Fiz., 24 (1984), pp. 1258–1259.
[6] W. L. BRIGGS, V. E. HENSON, AND S. F. MCCORMICK, *A Multigrid Tutorial*, 2nd ed., SIAM, Philadelphia, 2000.
[7] A. BUCCINI, *Regularizing preconditioners by non-stationary iterated Tikhonov with general penalty term*, Appl. Numer. Math., 116 (2017), pp. 64–81.

ETNA
Kent State University and
Johann Radon Institute (RICAM)



(a) (b) (c)

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



(a) (b) (c) (d)

FIG. 5.8. *Hubble test problem, blow-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*, Numer. Linear Algebra Appl., 24 (2017), Art. e2089, 12 pages.
[9] A. BUCCINI AND L. REICHEL, *An $\ell^2 - \ell^q$ regularization method for large discrete ill-posed problems*, J. Sci. Comput., 78 (2019), pp. 1526–1549.
[10] J.-F. CAI, R. H. CHAN, AND Z. SHEN, *A framelet-based image inpainting algorithm*, Appl. Comput. Harmon. Anal., 24 (2008), pp. 131–149.
[11] J.-F. CAI, S. OSHER, AND Z. SHEN, *Linearized Bregman iterations for frame-based image deblurring*, SIAM J. Imaging Sci., 2 (2009), pp. 226–252.
[12] ———, *Split Bregman methods and frame based image restoration*, Multiscale Model. Simul., 8 (2009/10), 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 J. Sci. Comput., 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 Scientific Computing, G. H. Golub, S. H. Lui, F. T. Luk, and R. J. Plemmons, eds., Springer, Singapore, 1997, pp. 58–72.
[15] R. H. CHAN AND K. CHEN, *A multilevel algorithm for simultaneously denoising and deblurring images*, SIAM J. Sci. Comput., 32 (2010), pp. 1043–1063.
[16] P. DELL'ACQUA, *A note on Taylor boundary conditions for accurate image restoration*, Adv. Comput. Math., 43 (2017), pp. 1283–1304.
[17] M. DONATELLI, *An algebraic generalization of local Fourier analysis for grid transfer operators in multigrid based on Toeplitz matrices*, Numer. Linear Algebra Appl., 17 (2010), pp. 179–197.
[18] ———, *Fast transforms for high order boundary conditions in deconvolution problems*, BIT, 50 (2010), pp. 559–576.
[19] ———, *An iterative multigrid regularization method for Toeplitz discrete ill-posed problems*, Numer. Math. Theory Methods Appl., 5 (2012), pp. 43–61.
[20] ———, *On nondecreasing sequences of regularization parameters for nonstationary iterated Tikhonov*, Numer. Algorithms, 60 (2012), pp. 651–668.
[21] M. DONATELLI, C. ESTATICO, A. MARTINELLI, AND S. SERRA-CAPIZZANO, *Improved image deblurring with anti-reflective boundary conditions and re-blurring*, Inverse Problems, 22 (2006), pp. 2035–2053.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

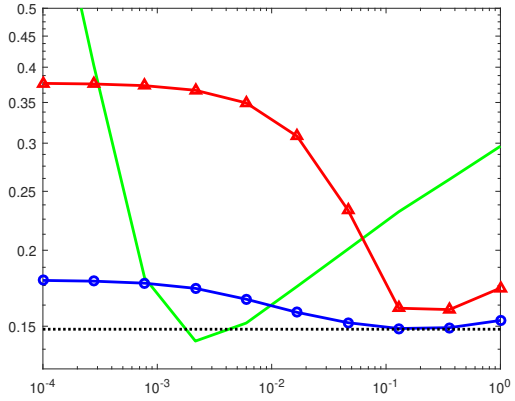


FIG. 5.9. *Hubble test problem relative reconstruction error against regularization parameter for the ADMM-UBC, LBA, and MLBA-AIT methods. The green line represents the ADMM-UBC method, the red line the LBA method, and the blue line the MLBA-AIT method. The dotted black line is the RRE obtained with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

[22] M. DONATELLI AND M. HANKE, *Fast nonstationary preconditioned iterative methods for ill-posed problems, with application to image deblurring*, Inverse Problems, 29 (2013), Art. 095008, 16 pages.
[23] M. DONATELLI AND S. SERRA-CAPIZZANO, *On the regularizing power of multigrid-type algorithms*, SIAM J. Sci. Comput., 27 (2006), pp. 2053–2076.
[24] ———, *Filter factor analysis of an iterative multilevel regularizing method*, Electron. Trans. Numer. Anal., 29 (2007/08), pp. 163–177. http://etna.ricam.oeaw.ac.at/vol.29.2007-2008/pp163-177.dir/pp163-177.pdf
[25] ———, *Antireflective boundary conditions for deblurring problems*, J. Electr. Comput. Eng., (2010), Art. 241467, 18 pages.
[26] D. L. DONOHO, *De-noising by soft-thresholding*, IEEE Trans. Inform. Theory, 41 (1995), pp. 613–627.
[27] H. W. ENGL, M. HANKE, AND A. NEUBAUER, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996.
[28] M. I. ESPAÑOL AND M. E. KILMER, *Multilevel approach for signal restoration problems with Toeplitz matrices*, SIAM J. Sci. Comput., 32 (2010), pp. 299–319.
[29] ———, *A wavelet-based multilevel approach for blind deconvolution problems*, SIAM J. Sci. Comput., 36 (2014), pp. A1432–A1450.
[30] G. FIORENTINO AND S. SERRA, *Multigrid methods for Toeplitz matrices*, Calcolo, 28 (1991), pp. 283–305 (1992).
[31] S. GAZZOLA, P. C. HANSEN, AND J. G. NAGY, *IR Tools: a MATLAB package of iterative regularization methods and large-scale test problems*, Numer. Algorithms, 81 (2019), pp. 773–811.
[32] M. HANKE, *A regularizing Levenberg-Marquardt scheme, with applications to inverse groundwater filtration problems*, Inverse Problems, 13 (1997), pp. 79–95.
[33] M. HANKE AND P. C. HANSEN, *Regularization methods for large-scale problems*, Surveys Math. Indust., 3 (1993), pp. 253–315.
[34] M. HANKE AND C. R. VOGEL, *Two-level preconditioners for regularized inverse problems. I. Theory*, Numer. Math., 83 (1999), pp. 385–402.
[35] P. C. HANSEN, J. G. NAGY, AND D. P. O'LEARY, *Deblurring Images. Matrices, Spectra, and Filtering*, SIAM, Philadelphia, 2006.
[36] G. HUANG, L. REICHEL, AND F. YIN, *Projected nonstationary iterated Tikhonov regularization*, BIT, 56 (2016), pp. 467–487.
[37] T. HUCKLE AND J. STAUDACHER, *Multigrid preconditioning and Toeplitz matrices*, Electron. Trans. Numer. Anal., 13 (2002), pp. 81–105. http://etna.ricam.oeaw.ac.at/vol.13.2002/pp81-105.dir/pp81-105.pdf
[38] B. KALTENBACHER, *On the regularizing properties of a full multigrid method for ill-posed problems*, Inverse Problems, 17 (2001), pp. 767–788.
[39] J. T. KING, *Multilevel algorithms for ill-posed problems*, Numer. Math., 61 (1992), pp. 311–334.
[40] S. MORIGI, L. REICHEL, F. SGALLARI, AND A. SHYSHKOV, *Cascadic multiresolution methods for image deblurring*, SIAM J. Imaging Sci., 1 (2008), pp. 51–74.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

[41] M. K. NG, R. H. CHAN, AND W.-C. TANG, *A fast algorithm for deblurring models with Neumann boundary conditions*, SIAM J. Sci. Comput., 21 (1999), pp. 851–866.
[42] S. REEVES, *Fast image restoration without boundary artifacts*, IEEE Trans. Image Process., 14 (2005), pp. 1448–1453.
[43] L. REICHEL AND A. SHYSHKOV, *Cascadic multilevel methods for ill-posed problems*, J. Comput. Appl. Math., 233 (2010), pp. 1314–1325.
[44] A. RIEDER, *A wavelet multilevel method for ill-posed problems stabilized by Tikhonov regularization*, Numer. Math., 75 (1997), pp. 501–522.
[45] J. W. RUGE AND K. STÜBEN, *Algebraic multigrid*, in Multigrid methods, S. F. McCormick, ed., vol. 3 of Frontiers Appl. Math., SIAM, Philadelphia, 1987, pp. 73–130.
[46] S. SERRA-CAPIZZANO, *A note on antireflective boundary conditions and fast deblurring models*, SIAM J. Sci. Comput., 25 (2003/04), pp. 1307–1325.
[47] U. TROTTENBERG, C. W. OOSTERLEE, AND A. SCHÜLLER, *Multigrid*, Academic Press, Inc., San Diego, 2001.
[48] R. VIO, J. BARDSLEY, M. DONATELLI, AND W. WAMSTEKER, *Dealing with edge effects in least-squares image deconvolution problems*, Astronom. and Astrophys., 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 Trans. Image Process., 13 (2004), pp. 600–612.
[50] E. H. ZARANTONELLO, *Projections on convex sets in Hilbert space and spectral theory*, University of Wisconsin, 1971.