Improvement of space-invariant image deblurring by preconditioned Landweber iterations

Paola Brianzi, Fabio Di Benedetto, Claudio Estatico *

March 16, 2005

Abstract

The Landweber method is a simple and flexible iterative regularization algorithm, whose projected variant provides nonnegative image reconstructions. Since the method is usually very slow, we apply circulant preconditioners, exploiting the shift invariance of many deblurring problems, in order to accelerate the convergence. This way reasonable reconstructions can be obtained within few iterations: the method becomes competitive and more robust than other approaches that, although faster, sometimes lead to lower accuracy. Some theoretical analysis of convergence is given, together with numerical validations.

Keywords: Landweber, two-level Toeplitz and circulant matrices, preconditioning, regularization.
AMS SC: 65F22, 65F10, 45Q05, 15A18.

1 Introduction

Image deblurring is the process of correcting degradations from a detected image. In first analysis [3], the process of image formation is described by a Fredholm operator of first kind; in many applications the blurring system is assumed to be space-invariant, so that the mathematical model is the following:

\[ g(x, y) = \int_{\mathbb{R}^2} K(x - \theta, y - \xi) f_s(\theta, \xi) d\theta d\xi + \omega(x, y) \]  

(1)

where \( f_s \) is the (true) input object, \( K \) is the space-invariant integral kernel of the operator, also called point spread function (PSF), \( \omega \) is the noise which arises in the process, and \( g \) is the observed data.

The image restoration problem is the inversion of (1): given the observed data \( g \), we want to recover (an approximation of) the true data \( f_s \). Its discrete version requires to invert a linear system, typically of very large size and very sensitive to data error, due to the ill-posed nature of the continuous problem [15].

Space invariance leads to strong algebraic structures in the system matrix; depending on the boundary conditions enforced in the discretization, we find circulant, Toeplitz or even more complicated structures related to sine/cosine transforms (see [27] for details). Exploiting structures in the inversion algorithm is necessary to face out computational issues.

The problem of noise sensitivity is usually addressed by using regularization methods, where a suitable (sometimes more than one) parameter controls the degree of bias in the computed solutions. There are several techniques in literature (like Tikhonov [15] or truncated SVD [21]),

* Dipartimento di Matematica, Via Dodecaneso 35, 16146 Genova (ITALY) (E-mail: brianzi@dima.unige.it, dibenede@dima.unige.it, estatico@dima.unige.it),
but in large-scale problems the main choice is given by *iterative regularization algorithms*, where the parameter is represented by the number of iterations: the method works if an early stop prevents from reconstruction of noisy components in the approximated solution \([13, 3]\).

The simplest iterative technique in this class is the Landweber method \([24]\), proposed in 1951 but in agreement to an older work of Cimmino (see the historical overview in \([1]\)); besides its easy implementation, this method presents very good regularization and robustness features (a wide variety of even recent applications can be found in \([2, 5, 4, 25, 31]\)).

In many real problems, the use of a priori information is basic for obtaining a substantial improvement in reconstructions: an important instance in imaging is taking into account nonnegativity constraints. In the last years more attention has been paid to faster Krylov methods like CGLS \([17]\) or GMRES \([7]\), but, unfortunately, these methods do not provide nonnegative iterates, independently of the initial guess.

Although the recent literature has proposed specific approaches to enforce sign constraints (see, e.g., \([20]\)), the Landweber method allows for a straightforward extension in order to do so, leading to the *projected Landweber method* \([11]\). On the other hand, its main disadvantage is that convergence may be very slow in practical applications.

In this paper we aim to overcome this disadvantage of the Landweber method, by proposing an acceleration technique specifically designed for the space-invariant setting. Following a general idea introduced in \([6]\), we study the effect on this method of structure-based preconditioning techniques recently investigated for conjugate gradient iterations \([19, 14]\). We prove that such preconditioners can improve the convergence speed of the Landweber method (preserving its regularization capabilities), making it competitive to other algorithms from a computational point of view. The same considerations could be extended to other preconditioning proposals \([18, 23]\).

We stress that the removal of computational disadvantages allows us to emphasize the great advantages of Landweber in comparison to other iterative methods:

- *simplicity* (we are able to give a formal proof of convergence and regularization behaviour, at least for the nonprojected version);
- *flexibility* (sign or even other constraints are easily incorporated; there are several parameters at our disposal, so that a fine tuning can be performed according to time/accuracy demands);
- *robustness* (little sensitivity to the wrong choice of a parameter).

N.B.: it is worth to mention that the Landweber idea has applications in the context of nonlinear inverse problems too; in this context, it has been successfully applied to many real problems due to its robustness and strong regularization capabilities (see the survey in \([12]\) and the references therein).

The paper is organized as follows. In §2 we introduce the Landweber method, the circulant preconditioning technique and a first convergence analysis in the simplest case of periodic boundary conditions (§2.2). In §3 we use this analysis to discuss the choice of parameters which define the preconditioned method. In §4 we show how a convergence analysis can be performed without simplifying assumptions, by developing the case study of Dirichlet boundary conditions. Numerical results are presented in §5 and final remarks are given in §6.
2 Landweber method and preconditioning

The discretization of (1), with image size \( n = (n_1, n_2) \), reads as \( g = Af_* + \omega \), where \( g, f_*, \omega \) represent the column-ordered vectors of the corresponding quantities and the matrix \( A \) discretizes the kernel \( K \) [3].

In order to enforce the same finite length \( N = n_1 n_2 \) to all the vectors \( g, f_*, \omega \), appropriate “boundary conditions” must be applied: for an exhaustive survey of possible choices, see [27]. This way \( A \) is a square \( N \times N \) matrix having a multilevel structure depending on the specific choice: for instance, \( A \) is a block Toeplitz matrix with Toeplitz blocks in the case of zero (Dirichlet) boundary conditions, and \( A \) is a block circulant matrix with circulant blocks if periodic boundary conditions are assumed.

In the discrete setting, given the blurred and noisy image \( g \), we want to recover a suitable approximation \( f \) of the true image \( f_* \), by computing a regularized solution of the least squares problem \( \min \| Af - g \|_2 \) [3]. Since the continuous problem is known to be ill-posed, the matrix \( A \) has ill-determined rank, since its smallest singular values accumulate to zero as \( N \) increases.

In this paper we deal with the Landweber method, which is the following iterative method for solving the normal equation \( A^* Af = A^* g \). Let \( f_0 \) be an arbitrarily chosen initial guess; as we will see later, a recommended choice is \( f_0 \equiv 0 \). Compute, for \( k = 0, 1, 2, \ldots \), the iterate

\[
f_{k+1} = f_k + \tau (A^* g - A^* A f_k) ,
\]

where \( \tau \) is a fixed value belonging to \((0, 2/\|A^* A\|)\) in order to ensure the convergence.

The Landweber method (2) can be studied in several ways. It corresponds to the method of successive approximations for the computation of a fixed point of the operator \( H(f) = f + \tau (A^* g - A^* A f) \). Moreover, it is the simplest method which returns the minimum point of the convex operator \( H(f) = \frac{1}{2} \| Af - g \|_2^2 \), since \( A^* g - A^* A f_k = -\nabla H(f_k) \) is the steepest descent direction. By induction, it is simple to verify that

\[
f_{k+1} = \tau \sum_{i=0}^k (I - \tau A^* A)^i A^* g + (I - \tau A^* A)^{k+1} f_0 .
\]

The Landweber algorithm belongs to the class of Krylov methods [3]. If we consider the fixed initial guess \( f_0 \equiv 0 \), the iteration can be written as \( f_{k+1} = Q_{k,\tau}(A^* A)A^* g \), where \( Q_{k,\tau} \) is the polynomial of degree \( k \) defined as \( Q_{k,\tau}(t) = \tau P_k(\tau t) \) with

\[
P_k(s) = \sum_{i=0}^k (1 - s)^i = \sum_{i=0}^k \left( \begin{array}{c} k+1 \\ i+1 \end{array} \right) (-s)^i = \frac{1 - (1 - s)^{k+1}}{s} .
\]

The method is linear since the polynomial \( Q_{k,\tau} \) does not depend on \( g \). We remark that, if \( t \in (0, 2/\tau) \supseteq (0, \|A^* A\|] \), then \( Q_{k,\tau}(t) \rightarrow t^{-1} (k \rightarrow +\infty) \), and, if \( t \in [0, 2/\tau) \supseteq [0, \|A^* A\|] \), then \( \|Q_{k,\tau}(t)\| = |1 - (1 - \tau t)^k| \leq 1 \). These two latter properties of \( Q_{k,\tau} \) state that the Landweber method is a continuous regularization algorithm, where the number of iterations \( k \) plays the role of regularization parameter [13, Theorem 6.1]. Basically, the first iterations of the method filter out the components of data mainly corrupted by noise; hence an early stop of the deblurring process improves the stability and gives a good noise filtering. Notice that \( Q_{k,\tau}(t) \geq Q_{1,\tau}(\|A^* A\|) = \tau (2 - \tau \|A^* A\|) > 0 \ \forall t \in [0, \|A^* A\|] \), which can be useful to improve the numerical stability.

In order to study the regularization properties of the algorithm, we analyze the convergence of \( Q_{k,\tau}(t) \) in a (right) neighborhood of 0. The function \( Q_{k,\tau}(t) \) is a polynomial of degree \( k \) such that

\[
Q_{k,\tau}(0) = \tau (k+1) , \quad Q'_{k,\tau}(0) = -(1/2) \tau^2 (k+1) k .
\]
By continuity arguments, the values of $Q_{k,\tau}(t)$ are bounded by $\tau(k+1)$ in a right neighborhood of 0.

The “level” of regularization of the $k$-th iteration is summarized by the values $Q_{k,\tau}(0) = O(k)$ and $Q'_{k,\tau}(0) = O(k^2)$, with $Q'_{k,\tau}(0) < 0$ for $k \geq 1$. At the $k$-th iteration, the approximation to the highest eigenvalues of the Moore-Penrose generalized inverse of $A$, that is, the approximation to the reciprocal of the smallest non-null eigenvalues of $A^*A$, is bounded by $\tau(k+1)$ and highly decreasing at a rate $O(k^2)$. Furthermore, the $k$-th iteration of the Landweber algorithm has basically the same regularization effects of the Tikhonov regularization method with regularization parameter $\alpha = \left(\tau(k + 1)\right)^{-1} > 0$, where $f_\alpha = (A^*A + \alpha I)^{-1}A^*g$ is the Tikhonov’s $\alpha$-regularized solution of the normal equations [13].

The Landweber method is a linear regularization algorithm, which allows a modification denoted as projected Landweber method, very useful to solve inverse problems where some specific constraints on the solution play an important role. For example, the nonnegative constraint $f_s \geq 0$ is very common in image deblurring, whereas many classical regularization methods do not ensure any sign property for computed reconstructions.

The projected variant consists of the following simple modification of (2):

$$f_{k+1} = P_+[f_k + \tau(A^*g - A^*Af_k)],$$

where $P_+$ is the projection onto the nonnegative cone. This leads to a nonlinear algorithm, for which the theoretical understanding is not complete [11]: it is proved that the iterates weakly converge, for exact data, to a minimizer of $\|Af - g\|_2$ among nonnegative vectors.

Other important observed properties are just conjectures at this moment: we have a numerical evidence of semiconvergence for noisy data, and the natural initial guess $f_0 = 0$ seems to provide the convergence of $f_k$ to the least squares nonnegative solution having minimal norm [6].

Because of this absence of effective mathematical tools for investigating the convergence, the projected version of the method will not be considered in the following theoretical analysis. Besides, numerical experiments of §5 will concern both the projected and nonprojected variants.

It is interesting to make a comparison with the widely used conjugate gradient (CGLS) method applied to normal equations [17]. The CGLS method is a nonlinear regularization algorithm and its $k$-th iteration is $f_{k+1} = P_{k,g}(A^*A)A^*g$, provided that $f_0 \equiv 0$ as before. Here $P_{k,g} = P_{k,g}(t)$ is a polynomial of degree $k$ which depends on the input data $g$. The value $P_{k,g}(0)$, which mainly controls the level of regularization at the $k$-th iteration, is usually much greater than $k + 1$ [17]. This implies that the CGLS method is faster than the Landweber one. This is confirmed by recalling that the CGLS method is an optimal Krylov method in the sense that the error at any iteration is minimized among all the Krylov polynomials. On the other hand, the high convergence speed of the CGLS method may be a negative fact in image deblurring, since it gives rise to a fast amplification of the components of the restored image $f_k$ which are related to the noise on input data [16].

We can summarize that the regularization of the Landweber method is high, whereas the convergence speed is low. In the following, our aim is to improve its convergence speed without loosing its very favorable regularization capabilities.

### 2.1 Circulant regularizing preconditioners

As already noticed, the discrete image deblurring system $A^*Af = A^*g$ has ill-determined rank, since the continuous problem is ill-posed. The Landweber scheme is a suitable algorithm for solving that system, since it is a very effective regularization method. The negative fact is that the method is often quite slow: in several application, such as astronomical image deblurring, thousands of iterations could be necessary.
Here we improve the convergence speed by means of preconditioning techniques. Basically, in order to speed up the convergence of any iterative method, a preconditioner has to be an approximation of the, possibly generalized, inverse of the system matrix.

Following [6], the $N \times N$ linear system $A^*Af = A^*g$ is replaced by an algebraic equivalent system

$$DA^*Af = DA^*g,$$

where the $N \times N$ matrix $D$ is the preconditioner which approximates the generalized inverse of $A^*A$. This way the preconditioned version of the method reads as follows:

$$f_{k+1} = P_+ [\tau DA^*g + (I - \tau M) f_k],$$

where $M := DA^*A$ is the preconditioned matrix.

From now on we will consider for the theoretical discussion just the nonprojected variant, for which the operator $P_+$ does not appear on the right hand side of (5). In this case $f_{k+1}$ linearly depends on $f_k$, whence we obtain the closed formula for the case $f_0 = 0$

$$f_k = G_k A^* g, \quad G_k := \tau P_{k-1}(\tau M) D,$$

$P_k(t)$ being the polynomial introduced in (3). The new limitation for $\tau$ becomes $0 < \tau < 2/\|DA^*A\|$. If $B$ denotes an approximation of the matrix $A$, we construct the preconditioner $D$ by computing $D = (B^*B)\dagger$, where the symbol $\dagger$ denotes the Moore-Penrose generalized inverse.

Since for space-invariant problems $A$ has a two-level Toeplitz-like structure, we look for $B$ inside the matrix algebra $C$ of block circulant matrices with circulant blocks (BCCB). The BCCB matrices are very useful in the Toeplitz context since they provide fast diagonalization and matrix-vector multiplication within $O(N \log N)$ operations, via 2-d FFT. From now on, we consider the T. Chan optimal approximation $B = B_{opt}$ of the system matrix $A$, that is, $B_{opt}$ solves the following minimization problem [9]

$$B_{opt} = \arg \min_{X \in C} \|A - X\|_F,$$

where $\| \cdot \|_F$ is the Frobenius norm $\|G\|_F^2 = \sum_{i,j} |(G)_{i,j}|^2$.

Since $B_{opt}$ is the best approximation of $A$ in the space $C$ of the BCCB matrices, with respect to the Frobenius norm, it strongly “inherits” the entire spectral distribution of $A$. This means that, if $A$ has ill-determined rank, the same will hold for $B_{opt}$. The solution of the preconditioned system $DA^*Af = DA^*g$, with $D = (B_{opt}^*B_{opt})\dagger$, leads to worse numerical results due to high amplification of the components related to the noise of $g$. Differing from $B_{opt}$, any useful preconditioner for deblurring linear systems should approximate its system matrix only in the subspace which is less sensitive to data errors.

According to [19], this so-called signal subspace corresponds to the largest singular values of $A$, in the sense that it is spanned by the associated singular vectors. On the other hand, the noise subspace is related to the smallest singular values and represents the components where the direct reconstruction is more contaminated by data errors.

In Toeplitz deblurring context, the problem of locating these two fundamental subspaces was first studied by M. Hanke, J. Nagy and R. Plemmons [19]. Fixed a small real value $\alpha > 0$ called truncation parameter, the signal space can be roughly identified by the eigendirections corresponding to the eigenvalues of $B_{opt}$ greater than $\alpha$. If the parameter is well chosen we can believe that the noise space falls into the directions related to the eigenvalues of $B_{opt}$ with absolute value smaller than $\alpha$. Therefore the authors proposed in [19] to set equal to 1 all these eigenvalues: in that way, the convergence speed increases in the signal space only, without fast amplification of the noisy components.
On these grounds, now we extend the approach of [19], by providing a family of different filtering procedures. Given a BCCB matrix $G$, let $\lambda_1(G), \lambda_2(G), \ldots, \lambda_N(G)$ denote its eigenvalues with respect to the fixed base of eigenvectors collected into the columns of the 2-d unitary Fourier matrix [6]. If $\alpha > 0$ is a truncation parameter, we define the regularizing BCCB preconditioner $D = D_\alpha$ as the matrix whose eigenvalues $\lambda_1(D_\alpha), \lambda_2(D_\alpha), \ldots, \lambda_N(D_\alpha)$ are such that

$$\lambda_i(D_\alpha) = \mathcal{F}_\alpha(\lambda_i(B_{\text{opt}}^* B_{\text{opt}})),$$

where the function $\mathcal{F}_\alpha : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is one of the eight filters of Table 1. Observe that the central column prescribes the eigenvalues along the noise space, and the last one is referred to the signal space.

<table>
<thead>
<tr>
<th>(\mathcal{F}_\alpha(t))</th>
<th>(0 \leq t &lt; \alpha)</th>
<th>(t \geq \alpha)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>((t + \alpha)^{-1})</td>
<td>((t + \alpha)^{-1})</td>
</tr>
<tr>
<td>II</td>
<td>0</td>
<td>(t^{-1})</td>
</tr>
<tr>
<td>III</td>
<td>(\alpha^{-1})</td>
<td>(t^{-1})</td>
</tr>
<tr>
<td>IV</td>
<td>(\alpha^{-(p+1)p})</td>
<td>(t^{-1})</td>
</tr>
<tr>
<td>V</td>
<td>(\alpha^{-\frac{\alpha+1}{\alpha} t^\frac{1}{\alpha}})</td>
<td>(t^{-1})</td>
</tr>
<tr>
<td>VI</td>
<td>(\alpha^{-1} e^{-\alpha t})</td>
<td>(t^{-1})</td>
</tr>
<tr>
<td>VII</td>
<td>(\int_0^{1/\alpha} e^{-t^\alpha} ds)</td>
<td>(\int_0^{1/\alpha} e^{-t^\alpha} ds)</td>
</tr>
<tr>
<td>VIII</td>
<td>(\int_0^{1/\alpha} e^{-t^\alpha} ds)</td>
<td>(\int_0^{1/\alpha} e^{-t^\alpha} ds)</td>
</tr>
</tbody>
</table>

Table 1: Regularizing functions for the optimal-based preconditioners

The filter I comes from the Tikhonov regularization [3] and the filter III is the Hanke, Nagy and Plemmons’ one [19]. The filter IV has been introduced by Tyrtyshnikov, Yeremin and Zamarashkin in [30], and VIII is the Showalter filter for asymptotic regularization [13]. For the polynomial filter V, we consider $p > 0$.

It is worth to notice two important properties common to all of these filters:

- if the truncation parameter $\alpha$ is small enough, then $\lambda_i(D_\alpha)$ approximates the reciprocal of $\lambda_i(B_{\text{opt}}^* B_{\text{opt}})$ on the signal space;
- since $\mathcal{F}_\alpha$ is a bounded function in all the eight cases, the eigenvalues of $D_\alpha$ have a uniform upper bound independent of the dimension.

Further properties of these filters, together with a wide numerical comparison, can be found in [14].

2.2 A simplified convergence analysis: periodic boundary conditions

The theoretical analysis of preconditioned iterations is in general a complicated task, due to the incomplete knowledge on how the singular vectors of the blurring matrix $A$ and of the preconditioner $D$ are related. We will try to give partial results in §4, with special emphasis to the case where $A$ is Toeplitz; by now we want to carry just a preliminary analysis, in order to give useful insight about the role played by the different parameters at our disposal.

For this reason, in the present section we overcome the main difficulty by considering the following simplifying assumption:
It is understood that assuming this is restrictive, but it corresponds to the practical application of *periodic* boundary conditions in the discretization of the continuous space-invariant blurring model, as described in [3, 22, 27]. This is very common in some imaging context where the boundary effect is negligible, e.g. in astronomical observations of isolated sources.

With this choice, $A^*A$ can be regarded as the classical Strang [28] circulant approximation of the convolution between the PSF matrix and its adjoint, whereas $D$ is obtained from the T. Chan approximation. Since the two approximations are spectrally equivalent (see [8, Lemma 1]), we can state that every $\lambda_i(B_{opt}^{*}B_{opt})$ considered in (8) is a good approximation of the corresponding eigenvalue of $A^*A$.

In the light of the circulant structure, $A^*A$ and $D$ are diagonalized by a common eigenvector basis specifically represented by the discrete Fourier transform $F$ (here we intend the two-dimensional transform, but we will use the same symbol independently on the number of dimensions):

$$
A^*A = F \Lambda_A F^* , \quad D = F \Lambda_D F^* ,
$$

$$
\Lambda_A = \mathrm{diag}(\lambda_1^A, \ldots, \lambda_N^A) , \quad \Lambda_D = \mathrm{diag}(\lambda_1^D, \ldots, \lambda_N^D) .
$$

In this case $D$ satisfies all the so-called Strand conditions for regularizing preconditioners and therefore the sequence $f_k$ of preconditioned iterations preserves the semiconvergence property [6, Proposition 3.1].

Under this strong assumption, convergence analysis of the preconditioned Landweber method becomes straightforward, since both the preconditioned matrix $M$ and the iteration matrix $G_k$ such that $f_k = G_k A^* g$, introduced in (6), can be diagonalized in the same way:

$$
M = DA^* A = F \Lambda_M F^* , \quad \Lambda_M = \Lambda_D \Lambda_A = \mathrm{diag}(\lambda_1^M, \ldots, \lambda_N^M) , \quad \lambda_j^M = \lambda_j^D \lambda_j^A ;
$$

$$
G_k = \tau P_{k-1}(\tau M) D = F \Lambda_{G_k} F^* , \quad \Lambda_{G_k} = \tau P_{k-1}(\tau \Lambda_M) \Lambda_D = \mathrm{diag}(\lambda_1^{G_k}, \ldots, \lambda_N^{G_k}) .
$$

Recalling the closed formula (3) for $P_{k-1}(t)$, whenever $\lambda_j^M \neq 0$ we have

$$
\lambda_j^{G_k} = \tau P_{k-1}(\tau \lambda_j^M) \lambda_j^D = \tau \frac{1 - (1 - \tau \lambda_j^M)^k}{\lambda_j^M} \lambda_j^D = \frac{1 - (1 - \tau \lambda_j^D \lambda_j^A)^k}{\lambda_j^A} .
$$

Now we are ready to study the components of the $k$-th Landweber approximation $f_k$ and of the generalized solution $f^\dagger$ along the Fourier directions $v_1, \ldots, v_N$. On one hand,

$$
f_k = G_k A^* g \implies F^* f_k = \Lambda_{G_k} F^* A^* g
$$

and therefore the $j$-th Fourier component of $f_k$ is

$$
v_j^* f_k = \lambda_j^{G_k} \gamma_j , \quad \text{where} \quad \gamma_j = v_j^* A^* g .
$$

On the other hand, we can use for $f^\dagger$ the expression

$$
f^\dagger = (A^* A)^\dagger A^* g \implies F^* f^\dagger = \Lambda_A^\dagger F^* A^* g
$$

whence

$$
v_j^* f^\dagger = \begin{cases} 
\frac{\gamma_j}{\lambda_j^A} & \text{if } \lambda_j^A > 0 \\
0 & \text{if } \lambda_j^A = 0 .
\end{cases}
$$
We obtain the following relation for the relative error on the components lying outside the null space of $A$:

$$\frac{|v_j^* f_k - v_j^* f^\dagger|}{|v_j^* f^\dagger|} = \left| \frac{\lambda_j^G  - 1/\lambda_j^A}{\lambda_j^A} \right| = |1 - \tau \lambda_j^D \lambda_j^A|^k .$$  \hspace{1cm} (10)

It is evident that the rate of decay for this error, as the iterations proceed, heavily depends on the value of the parameter $\tau$ and on the distance of the preconditioned eigenvalues $\lambda_j^D \lambda_j^A$ from 1. As described in the previous subsection, this distance is reduced just in the *signal space*, where $\lambda_j^D$ approximates the reciprocal of the optimal circulant preconditioner eigenvalue (which in turn is a good approximation of $\lambda_j^A$).

It is worth to notice that in the *null space* directions (9) do not apply; instead we may deduce from (4) the relation

$$v_j^* f_k = \tau k \lambda_j^D \gamma_j ,$$  \hspace{1cm} (11)

so that this component is amplified as the iteration count $k$ increases, unless $\lambda_j^D$ is very small.

A similar behaviour can be generally observed along the *noise space*, since under our assumptions these directions correspond to the indices such that $\lambda_j^A < \alpha$, the small threshold used to define the regularizing preconditioner. If this occurs, from the inequalities $1 - i \tau \lambda_j^D \alpha \leq (1 - \tau \lambda_j^D \lambda_j^A)^i \leq 1$ we obtain for $P_{k-1} (\tau \lambda_j^M) = \sum_{i=0}^{k-1} (1 - \tau \lambda_j^D \lambda_j^A)^i$ the bounds

$$k \left(1 - \frac{k}{2} \tau \lambda_j^D \alpha \right) \leq P_{k-1} (\tau \lambda_j^M) \leq k ;$$

substituting into the relation $v_j^* f_k = \tau P_{k-1} (\tau \lambda_j^M) \lambda_j^D \gamma_j$ gives

$$\tau k \lambda_j^D |\gamma_j| \left(1 - \frac{k}{2} \tau \lambda_j^D \alpha \right) \leq |v_j^* f_k| \leq \tau k \lambda_j^D |\gamma_j| .$$  \hspace{1cm} (12)

The last relation also gives a dependence between the chosen threshold $\alpha$ and the iteration count $k$, since both of them affect the lower bound for the amplification of the undesired components.

The different role of the parameters $k$, $\tau$, $\alpha$ and of the filter (which mainly involves the magnitude of $\lambda_j^D$ in the noise space) is now made explicit in the formulae (10), (11), (12): we are ready to present a full discussion in the next section.

### 3 Choice of parameters

The preconditioned Landweber method described so far involves four parameters to choose: the relaxation parameter $\tau$, the regularization parameter $k$ (iteration count), the filter parameter $\alpha$ and finally the type of filtering function $F_{\alpha}$.

As already shown in equation (10), the parameter $\tau$ allows us to control the convergence of the iterations towards the solution of the normal equations. More precisely, the relative reconstruction error is the largest one ($\approx 100\%$) along the components for which the value of $\lambda_j^D \lambda_j^A$ is close to 0 or $2\tau^{-1}$, whereas it is the smallest one ($\approx 0\%$) when $\lambda_j^D \lambda_j^A$ is close to $\tau^{-1}$. This implies that the convergence is always slow in the noise space where $\lambda_j^D \lambda_j^A$ is small, but an appropriate choice of $\tau$ enables us to “select” the most important subspace of components to be first resolved in the reconstruction process.
We recall that, by using a suitable filtering preconditioner $D$ as explained in §2.1, the numbers $\lambda_j D\lambda_j^A$ giving the spectrum of $DA^*A$ can be made clustered at unity in the signal space and clustered at zero in the noise space; in that case, the simplest choice $\tau \equiv 1$ should provide good results in terms of convergence speed.

It is worth noticing that for some choices of the filter (as those labeled by $I$ and $VIII$ in Table 1) the reciprocal function is only approximated and therefore the preconditioned spectrum has a slightly different distribution: in these cases another value of $\tau$ could provide better results. In any case, $\tau$ must be inside $(0,2/\|DA^*A\|)$: it can be proved that, for $\alpha$ not too large, all the filters listed in Table 1 ensure that the choice $\tau = 1$ verifies the constraint above.

Choosing the number of iterations $k$ is of fundamental importance for the performance of the method. If we underestimate $k$, we can provide restorations which are not sufficiently accurate because the “signal” component (10) of the relative error has not yet reached an acceptable value. On the other hand, if we overestimate $k$ we perform too many iterations and this way we do not improve the efficiency of the method; moreover, we can obtain satisfactory restorations, provided that the unwanted components (related to $k$ through the expressions (11) and (12)) have not been amplified too much. By the numerical experiments performed in §5 on simulated data (see, in particular, Figure 2), the restoration error is decreasing first and increasing afterwards preserving the property of semiconvergence [3]. Therefore we propose to apply for general problems the most simple and efficient stopping rule for estimating the optimal value of $k$, that is the discrepancy principle; if we have an estimate of the noise level $\epsilon = \|\omega\| = \|Af^*-g\|$, we stop the iterations when the residual $r_k = \|Af_k - g\|$ becomes less than $\epsilon$.

Concerning the choice of $\alpha$, we recall here the “recipe” of [19]. We consider the discrete Fourier transform (DFT) of the right hand side $g$ and we observe that there is an index $r$ at which the Fourier coefficients begin to stagnate: this corresponds to the components where the random error starts to dominate the data vector $g$. Then we consider the eigenvalues $\lambda_i$ of $B_{opt}^*B_{opt}$, obtained again by means of a DFT as discussed in §2.1: we take as approximation of the filter parameter $\alpha$ the magnitude of the eigenvalue $\lambda_r$.

In many cases this strategy is difficult to apply, due to a great uncertainty in locating the index $p$; in case of doubt, it is better to underestimate its value because this is equivalent to take a greater value of $\alpha$ in order to prevent the reconstruction of noisy components.

The choice of the type of filtering function $F_{\alpha}$ depends on the features of the problem and on the action we want to apply to the high frequency components in the reconstruction. The filters presented in Table 1 can be classified in two categories: “smooth” filters and “sharp” filters. For instance, the “smooth” filters $I$ and $VIII$ have the same expression for all $t$, whereas the “sharp” filtering functions from $II$ to $VII$ have some discontinuity in $t = \alpha$.

The smooth filters give an approximation to the reciprocal function everywhere, so that the error in (10) is somehow reduced in the noise space too; therefore they allow us to modulate the restoration from high frequencies according to the problem. On the other hand, the sharp filters do not try to invert on the noise space and so they do not introduce further improvement in the solution: hence they produce a sequence that at first reduces the restoration error and after becomes stable because for $t \leq \alpha$ the filter functions are slightly varying. This behaviour is particularly desirable in the case of a strongly ill-conditioned problem, since it makes the method more robust with respect to a wrong choice of the “critical” parameters $k$ and $\alpha$.

As we can see in the numerical results of §5, the nature of the problem may suggest to use a type of filter instead of another.
4 Convergence analysis: the Toeplitz case

If the assumption of a common eigenvector basis for $A$ and $D$ is dropped, convergence analysis of the preconditioned Landweber method becomes more involved, if we look at the Fourier components. An alternative way to quantify the acceleration and regularization behaviour of iterations is to perform the analysis with respect to the eigenvector basis of the preconditioned matrix $M = DA^*A$; this choice needs no particular assumption on the structure of $A^*A$.

The argument is the classical one for stationary iterative methods: recalling that the iterates satisfy the recurrence relation

$$f_{k+1} = \tau DA^*g + (I - \tau M)f_k,$$

the generalized solution $f^\dagger$ can be expressed in a fixed-point form:

$$A^*A f^\dagger = A^*g \Rightarrow f^\dagger = \tau DA^*g + (I - \tau M)f^\dagger.$$

Subtracting (14) from (13), we obtain for the $k$-th error a recurrence relation leading to the closed formula

$$f_k - f^\dagger = (I - \tau M)^k(f_0 - f^\dagger) = -(I - \tau M)^k f^\dagger,$$

having assumed the standard choice for the initial guess, that is $f_0 = 0$.

Let $M = V_M\Lambda_M V_M^{-1}$ be the eigenvalue decomposition of the preconditioned matrix, where $\Lambda_M$ is real nonnegative because $M$ is symmetrizable and semidefinite, but $V_M$ may be not unitary. If we define $\phi_k^j$ and $\phi_j^\dagger$ as the $j$-th components of $f_k$ and $f^\dagger$ respectively along the eigenvector basis $V_M$, by (15) we obtain

$$\phi_k^j - \phi_j^\dagger = -(1 - \tau \lambda_j^M)^k \phi_j^\dagger,$$

$\lambda_j^M$ being the generic eigenvalue of $M$. From this characterization we can draw the following general conclusions:

1. Along the directions where $\lambda_j^M \approx 1$, the relative error on the associated component of $f^\dagger$ decreases in magnitude with a linear rate close to $1 - \tau$; in the case $\tau = 1$, the reconstruction of such components is heavily accelerated.

2. Along the directions where $\lambda_j^M$ is small enough, the associated component of $f_k k$ stays bounded as follows:

$$|\phi_k^j| = \left[1 - (1 - \tau \lambda_j^M)^k\right]|\phi_j^\dagger| = \tau \lambda_j^M \sum_{l=0}^{k-1}(1 - \tau \lambda_j^M)^l |\phi_j^\dagger| \leq \tau k \lambda_j^M |\phi_j^\dagger|.$$

Hence such components are poorly reconstructed, provided that the iterations are early stopped; in particular, $f_k$ has no component along the null space of $M$.

On the other hand, the spectral analysis carried out in [19, Theorem 6.1] for the filter III (but extendable to several other choices) proves that the eigenvalues of $M$ have a unit cluster and accumulate to zero, without giving any insight about the related eigenvectors; hence we are sure that statements in items 1 and 2 do not refer to the empty set!

Anyway, our conclusions are of no practical relevance unless we give an answer to some crucial questions.

- Are the directions considered in item 1 related to the signal space? (here we desire a fast reconstruction)
• Are we sure that the noise space (where reconstruction is not wanted) falls into the directions considered in item 2?

The matter is very delicate, and no exact knowledge is at our disposal. In space-invariant deblurring problems, it is known that signal and noise spaces can be described in terms of low and high frequencies (see, e.g., [19]), but no direct relation between the frequency-related Fourier basis and the eigenvectors of \( M \) is known in literature, except for the trivial case of common bases assumed in §2.2.

In the case where the blurring operator \( A \) is a Toeplitz matrix (this occurs when Dirichlet boundary conditions are imposed to the PSF, see [3, 22, 27]), we take some insight from an indirect

**Definition 4.1** [32] The eigenvectors of the sequence of \( n \times n \) matrices \( \{B_n\} \) are distributed like the sequence of unitary vectors \( \{q_k^{(n)}\} \), where \( q_k^{(n)} \in \mathbb{C}^n \) for \( k = 1, \ldots, n \), if the discrepancies

\[
r_k^{(n)} := \|B_nq_k^{(n)} - \langle q_k^{(n)}, B_nq_k^{(n)} \rangle q_k^{(n)}\|_2
\]

are clustered around zero, in the sense that

\[
\forall \epsilon > 0 : \# \left\{ k \in \{1, \ldots, n\} : |r_k^{(n)}| > \epsilon \right\} = o(n) .
\]

Since the discrepancies are a sort of measure of how much the \( \{q_k^{(n)}\} \) behave like the eigenvectors of \( B_n \), our goal is now to show that the eigenvectors of \( M \) are distributed like the Fourier vectors, and therefore are frequency-related; this way we have a partial positive answer to the questions addressed above.

Clearly, in order to apply Definition 4.1 we should think at \( M \) as the element of a sequence of matrices indexed by their total size, which of course is the product of the individual ones along the different directions. Thus we are not sure that the sequence is well defined for any positive integer \( n \), and it is better to adopt the typical multi-index notation of multilevel matrices (see e.g. [29]), with a slight change in the original Definition 4.1.

From now on, we put \( n = (n_1, n_2) \in \mathbb{N}^2 \) as the vector of individual dimensions, in the sense that \( M_n \) is a (block) two-level matrix of total size \( N(n) := n_1n_2 \). By the multi-indices \( i = (i_1, i_2) \) and \( j = (j_1, j_2) \) we label a single entry of \( M_n \), located at the inner \( i_2, j_2 \) position of the block having the indices \( i_1, j_1 \) at the outer level. The indices \( i_l, j_l \) at each level \( (l = 1, 2) \) range from 1 to the respective dimension \( n_l \).

The same notation applies to the preconditioners \( D_n \) and to the rectangular Toeplitz blurring matrices \( A_{m,n} \in \mathbb{R}^{N(m) \times N(n)} \) too. This way we may consider a (double) sequence \( \{M_n\}_{n \in \mathbb{N}^2} \) and adjust the concept of equal distribution to our two-level setting, according to the following new definition.

**Definition 4.2** Consider a sequence \( \{B_n\}_{n \in \mathbb{N}^2} \) of two-level matrices as described before, and a sequence of unitary vectors

\[
\left\{ q_k^{(n)} : \quad n = (n_1, n_2) \in \mathbb{N}^2, \; k \in \{1, \ldots, n_1\} \times \{1, \ldots, n_2\} \right\} \subseteq \mathbb{C}^{N(n)} .
\]

We say that the eigenvectors of \( B_n \) are distributed like the sequence \( \{q_k^{(n)}\} \) if the discrepancies \( r_k^{(n)} \) as in Definition 4.1 satisfy

\[
\forall \epsilon > 0 : \# \left\{ k \in \{1, \ldots, n_1\} \times \{1, \ldots, n_2\} : |r_k^{(n)}| > \epsilon \right\} = o(N(n)) .
\]
Our vectors $q_k^{(n)}$ will be the two-dimensional Fourier vectors, indexed in such a way that they correspond to the columns of the two-level Fourier matrix

$$F = \left( \frac{1}{\sqrt{n_1}} \exp \left\{ i \frac{2\pi(i_1-1)(j_1-1)}{n_1} \right\} \right)_{i_1,j_1=1}^{n_1} \otimes \left( \frac{1}{\sqrt{n_2}} \exp \left\{ i \frac{2\pi(i_2-1)(j_2-1)}{n_2} \right\} \right)_{i_2,j_2=1}^{n_2},$$

where $\otimes$ denotes the Kronecker (tensor) product and $i$ is the imaginary unit.

**Lemma 4.1** For a sequence $\{T_n(a)\}_{n \in \mathbb{N}_2}$ of two-level Toeplitz matrices generated [29] by the $L^2$ bivariate function $a(x)$, the eigenvectors are distributed like the Fourier vectors, in the sense of Definition 4.2.

**Proof.** The same statement has been proved in [32] for the one-level case; the proof was a direct consequence of the estimate

$$\|T_n(a) - C_n(a)\|^2_F = o(n),$$

where $C_n(a)$ is the optimal circulant preconditioner of $T_n(a)$, denoted as $B_{opt}$ in (7). The last result has been extended to the two-level version

$$\|T_n(a) - C_n(a)\|^2_F = o((N(n))$$

in the paper [29], so that the proof can be fully generalized by following the same steps as in [32], without any substantial change.

The one-level equidistribution result has been further extended to $L^1$ generating functions in [33]; since the proof no longer uses (17), we cannot say that Lemma 4.1 holds if $a \in L^1$ too. The result is probably true, but is not of interest in image deblurring, where generating functions are related to the PSFs and hence they have a high degree of regularity.

In order to extend the equidistribution result to our matrices $M_n = D_nA_{m,n}^*A_{m,n}$, we must overcome two difficulties: the presence of a first factor $D_n$ (this will not be a great problem, since the Fourier vectors are exactly the eigenvectors of $D_n$), and the loss of Toeplitz structure in forming the normal equation matrix product $A_{m,n}^*A_{m,n}$. The next lemma takes care of this second topic.

**Lemma 4.2** [10] Let $\{A_{m,n}\}$ be a sequence of two-level Toeplitz matrices generated by a continuous bivariate function $a(x)$. Then for every $\epsilon > 0$ there exist two sequences $\{R_n\}, \{E_n\}$ of $N(n) \times N(n)$ matrices and a constant $s$ such that

$$A_{m,n}^*A_{m,n} = T_n(b) + R_n + E_n \quad \forall n, m,$$

where $b(x) = |a(x)|^2$, $\|E_n\|_2 < \epsilon$ and each $R_n$ is a low-rank matrix having the following multilevel pattern:

$$R_n = (R_{i_1,j_1})_{i_1,j_1=1}^{n_1}, \quad R_{i_1,j_1} = (R_{i_2,j_2})_{i_2,j_2=1}^{n_2}$$

where each element $R_{i,j}$ is nonzero only if all the indices $i_1, i_2, j_1, j_2$ take the first or the last $s$ values.

The continuity assumption for $a(x)$ is stronger than $L^2$ hypothesis used in Lemma 4.1, but again is not restrictive in the image deblurring context.

We end this section by proving the equal distribution for the eigenvectors of our preconditioned matrices with respect to Fourier vectors.

**Theorem 4.1** Under the same assumptions as in Lemma 4.2, the eigenvectors of the matrices $M_n$ are distributed like the Fourier vectors, in the sense of Definition 4.2.
Proof. First we will show that the equidistribution property holds for normal system matrices $A_{m,n}^* A_{m,n}$ too. Notice that $\forall B \in \mathbb{C}^{N \times N}$ and $\forall q \in \mathbb{C}^N$, as observed e.g. in [32],

$$\langle q, Bq \rangle = \arg \min_{\lambda \in \mathbb{C}} \|Bq - \lambda q\|_2,$$

so that in order to prove an equidistribution result for the sequence $\{B_n\}$ it suffices to show a $o(N(n))$ bound for vectors of the form $Bq_k^{(n)} - \lambda q_k^{(n)}$ for any suitable $\lambda$ in place of the discrepancies $r_k^{(n)}$.

Now let $\lambda := \langle q_k^{(n)}, T_n(f)q_k^{(n)} \rangle$ where $q_k^{(n)}$ is a Fourier vector and $f$ is given by Lemma 4.2. Then for every $\epsilon > 0$

$$\|A_{m,n}^* A_{m,n} q_k^{(n)} - \lambda q_k^{(n)}\|_2 = \|T_n(f)q_k^{(n)} + R_n q_k^{(n)} + E_n q_k^{(n)} - \lambda q_k^{(n)}\|_2 \leq r_k^{(n)} + \|R_n q_k^{(n)}\|_2 + \|E_n\|_2 \|q_k^{(n)}\|_2 < \|R_n q_k^{(n)}\|_2 + 2\epsilon \quad (18)$$

except for $o(N(n))$ multi-indices $k$, where we have denoted by $r_k^{(n)}$ the discrepancy referred to the Toeplitz matrix $T_n(f)$, and we have applied Lemmata 4.1 and 4.2.

In order to manage the product $R_n q_k^{(n)}$, we must take into account the sparsity pattern of $R_n$ given by Lemma 4.2 and the following general expression for the $j = (j_1, j_2)$-th entry of the Fourier vector $q_k^{(n)}$, deduced by (16):

$$(q_k^{(n)})_j = \frac{1}{\sqrt{N(n)}} \exp \left\{ 2\pi i \left[ \frac{(j_1 - 1)(k_1 - 1)}{n_1} + \frac{(j_2 - 1)(k_2 - 1)}{n_2} \right] \right\}. \quad (19)$$

The structure of $R_n$ implies that the product with the Fourier vector involves just the first and the last $s$ values of the indices $j_1$ and $j_2$; moreover, only such entries of the product are nonzero. More precisely, by exploiting the usual multi-index notation we obtain

$$\|R_n q_k^{(n)}\|_2^2 = \sum_{I_1 \times I_2} \left\| \sum_{j \in I_1 \times I_2} R_{i,j}(q_k^{(n)})_j \right\|^2,$$

where $I_l = \{1, \ldots, s\} \cup \{n_l - s + 1, \ldots, n_l\}$, $l = 1, 2$. Setting $c := \max_{i,j} |R_{i,j}|$ and using the expression (19), we have the bound

$$\left| \sum_{j \in I_1 \times I_2} R_{i,j}(q_k^{(n)})_j \right| \leq c \sum_{j \in I_1 \times I_2} |(q_k^{(n)})_j| = \frac{c}{\sqrt{N(n)}} \#(I_1 \times I_2) = \frac{4cs^2}{\sqrt{N(n)}},$$

whence

$$\|R_n q_k^{(n)}\|_2^2 \leq \#(I_1 \times I_2) \frac{16c^2 s^4}{N(n)} = \frac{64c^2 s^6}{N(n)},$$

which can be made less than $\epsilon$ provided that $N(n)$ is large enough. Substituting into (18) and observing that $\epsilon$ is arbitrarily small, we have proved the equal distribution between the eigenvectors of $A_{m,n}^* A_{m,n}$ and the Fourier vectors.

Concerning the preconditioned matrices, it suffices to observe that for $\mu := \lambda \cdot \lambda_k^D$

$$\|M_n q_k^{(n)} - \mu q_k^{(n)}\|_2 = \|D_n[A_{m,n}^* A_{m,n} q_k^{(n)} - \lambda q_k^{(n)}]\|_2 \leq \|D_n\|_2 \epsilon \quad (20)$$

except for $o(N(n))$ values of $k$, where we have used the property that Fourier vectors are also eigenvectors of $D_n$. Since $\|D_n\|_2$ equals the maximal eigenvalue of $D_n$, having a uniform upper bound for regularizing preconditioners (see §2.1), the inequality (20) proves the equal distribution result for the matrices $\{M_n\}$. 

\[\blacksquare\]
5 Numerical results

In this section we provide some numerical experiments illustrating the effectiveness of the preconditioned Landweber method, both in the basic version and in the projected variant. In particular, the deblurring capabilities of the method will be tested both through synthetic and widely used experimental data. The analysis of §3 is a useful starting point for the appropriate choice of all the parameters of the algorithm. Along this direction, the aim of the present section is to compare the results related to different settings. As already shown (see §2), our discrete model of image formation is the image blurring with shift invariance, and it can be simply written as follows

\[ g = Af_* + \omega, \]

where \( g \) is the blurred data, \( A \) is the matrix version of the PSF, \( f_* \) is the input true object and \( \omega \) is the noise which arises in the process. The deblurring problem is to recover a suitable approximation \( f \) of the true object \( f_* \), by means of the knowledge of \( g \), \( A \), and some statistical information about the noise \( \omega \).

We consider two different test examples.

T1 In the first one, the true image \( f_* \) is the 256 × 256 brain section of Fig. 1 (left), and the convolution operator \( A \) is a Gaussian PSF with standard deviation \( \sigma = 5 \) shown in Fig. 1 (center), with an estimated condition number of \( 1.2 \cdot 10^{20} \). In this first example, we compare the restorations corresponding to different levels of artificial Gaussian white noise \( \omega \), where the relative data error \( \|\omega\|/\|Af_*\| \) ranges from 3% to 30%, being \( \|\cdot\| \) the vector 2-norm.

T2 In the second example, the object \( f_* \) to be recovered is the 256 × 256 satellite image of Fig. 4 (left), while the blurring operator and the blurred data are experimental and plotted in Fig. 4 (center, right). These test data have been developed by the US Air Force Phillips Laboratory, Lasers and Imaging Directorate, Kirtland Air Force Base, New Mexico, and they are widely used in the literature [26]. This time \( A \) is just moderately ill-conditioned (\( \text{cond}(A) \approx 1.3 \cdot 10^6 \)), but the data image \( g \) is corrupted by a noise of unknown distribution, corresponding to \( \|\omega\|/\|Af_*\| \approx 5\% \).

According to Table 1 of §2.1, we test the following four filters \( F_\alpha \):

F1 \( F_\alpha \) is the Tikhonov filter I;

F2 \( F_\alpha \) is the Low Pass filter II;

F3 \( F_\alpha \) is the Hanke, Nagy and Plemmons' filter III;

F4 \( F_\alpha \) is the \( p \)-Polynomial Vanishing filter V, with \( p = 1 \).

The convergence parameter \( \tau \) of the method is set to 1, providing a good compromise between fast convergence and noise filtering for the preconditioned Landweber method, as discussed in §3.

We stress that the choice of the regularization parameter \( \alpha \) is not a simple task, so that we attempt several values in order to select the best one. In the test T2 we try to adopt the strategy proposed by Hanke, Nagy and Plemmons [19] recalled in §3, which is based on an appropriate estimate of the noise space, and we suggest some improvements and remarks.

The number of iterations can be well controlled by means of the discrepancy principle, even though can be underestimated with respect to the optimal one; therefore we prefer to present the best achievable results within the first 200 iterations.

All the experiments have been implemented on Matlab 6.1 and performed on IBM PC, with floating-point precision of \( 10^{-16} \).
5.1 Test 1 (synthetic data with different levels of noise)

We test the projected variant of the Landweber method, where each iteration \(f_k\) is projected onto the nonnegative cone (see end of §2), using the four filters for \(\alpha\) ranging from 0.005 to 0.1, and we take the best restoration among the first 200 iterations. Table 2 shows the values of the minimal relative restoration error (RRE) \(\|f_k - f_*\|/\|f_*\|\) and the corresponding iteration number \(k\). The left side of the table reports the results with 3\% of relative data error, the right side with 8\%.

First of all, we point out that the non-preconditioned Landweber method is really slower than the preconditioned version. After 200 iterations, the RRE without preconditioning gives the same accuracy provided by the Tikhonov filter F1 within at most 20 iterations (see the columns F1 of Table 2 relative to the values \(\alpha = 0.05\) for 3\% of noise and \(\alpha = 0.1\) for 8\% of noise).

As expected, when the noise is higher the filtering parameter \(\alpha\) has to be larger and the method has to be stopped earlier. Indeed, both \(\alpha\) and \(k\) play together the role of regularization parameter, as it can be observed by comparing the results on left and right sides of the table. In particular, a good choice of \(\alpha\) should be the compromise between noise filtering and fast convergence.

Small values of \(\alpha\) yield low noise filtering: in the first rows of the table, the noise dominates on the restoration process and the RREs are larger. If this is the case, it is interesting to notice that the filter F4 outperforms the others, because the problem stays ill-conditioned and F4 “cuts” very much on the noisy components. On the other hand, too large values of \(\alpha\) do not speed up the convergence, as shown by the latter rows, especially for filters F3 and F4.

The best restorations for this example are given by the Tikhonov filter F1, provided that a good value of \(\alpha\) has been chosen. The corresponding convergence histories, that is, the values of all RREs versus the first 200 iterations, are shown in Fig. 2. In this graph, it is quite evident that small filtering parameters \(\alpha\) provide fast convergence and low accuracy, while large filtering parameters \(\alpha\) provide high accuracy and low convergence speed. The graph of Fig. 2 confirms how the improvement provided by the preconditioned Landweber method is high with respect to the non-preconditioned case.

We remark that the low pass filter F2 always gives worse results. The reason is that this filter neglects all the eigenvalues lower than \(\alpha\): in this case, the corresponding components are completely lost, and the reconstruction is done without those pieces of information. All the other filters keep into account such components, although the restoration is slow therein, and hence the results are more accurate.

A comparison between projected and non-projected versions is shown in Table 3, related to filters F1 and F4. The left side of each table concerns the Landweber method with projection on
Figure 2: Test 1 - RREs versus the first 200 iterations; noise 3%, filter F1.

Figure 3: Test 1 - Best reconstructions with $\alpha = 0.03$, for relative noise 8%, within 200 iterations. Top: with projection on positives. Bottom: without projection.
Table 2: Test 1 - Best Relative Restoration Errors and number of iterations for 3% and 8% noise

<table>
<thead>
<tr>
<th>data</th>
<th>3% relative noise</th>
<th>8% relative noise</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.1794</td>
<td>0.1855</td>
</tr>
<tr>
<td>No prec.</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td>α</td>
<td>F1 F2 F3 F4</td>
<td>F1 F2 F3 F4</td>
</tr>
<tr>
<td>0.005</td>
<td>0.2028 0.1900 0.1898 0.1908</td>
<td>0.2107 0.1968 0.1967 0.2031</td>
</tr>
<tr>
<td></td>
<td>1 3 3 2</td>
<td>1 2 2 2</td>
</tr>
<tr>
<td>0.01</td>
<td>0.1890 0.1979 0.1969 0.1849</td>
<td>0.2016 0.2024 0.2018 0.1930</td>
</tr>
<tr>
<td></td>
<td>2 7 7 4</td>
<td>2 3 3 3</td>
</tr>
<tr>
<td>0.02</td>
<td>0.1854 0.2012 0.1847 0.1862</td>
<td>0.1925 0.2074 0.1967 0.1946</td>
</tr>
<tr>
<td></td>
<td>3 199 199 13</td>
<td>2 19 139 7</td>
</tr>
<tr>
<td>0.03</td>
<td>0.1823 0.2055 0.1798 0.1844</td>
<td>0.1907 0.2090 0.1886 0.1949</td>
</tr>
<tr>
<td></td>
<td>7 200 200 43</td>
<td>5 179 200 15</td>
</tr>
<tr>
<td>0.05</td>
<td>0.1782 0.2172 0.1796 0.1801</td>
<td>0.1896 0.2188 0.1867 0.1914</td>
</tr>
<tr>
<td></td>
<td>16 123 200 200</td>
<td>9 200 200 116</td>
</tr>
<tr>
<td>0.1</td>
<td>0.1752 0.2325 0.1794 0.1861</td>
<td>0.1881 0.2335 0.1858 0.1910</td>
</tr>
<tr>
<td></td>
<td>49 200 200 200</td>
<td>20 200 200 200</td>
</tr>
</tbody>
</table>

positives, while the right side concerns the method without projection. Notice that the positivity constraint improves the results, since actually the projected variant always leads to the minimal RREs and the maximal convergence speed. We can say that the projection on positives acts as regularizer, since it reduces the instability due to the noise on the data. In addition, the higher regularizing properties of the projected variant allow us to adopt smaller values of the regularization parameter α, which accelerates the method.

As graphical examples of restorations, the two images on the left of Figure 3 are the best restorations after 200 iterations of the non-preconditioned Landweber method, with projection onto positive values (top), and without projection (bottom), for input data with 8% of relative noise. The images on the center are the best restorations with filter F1 and α = 0.03: 5 iterations with projection on positives on the top, and 3 iterations without projection on the bottom. The images at the right are the best restorations with filter F4, again with α = 0.03: 15 iterations with projection on positives on the top, and 32 iterations without projection on the bottom.

5.2 Test 2 (experimental data)

The experimental data (blurring operator and blurred image with about 5% of noise) are used here (see Fig. 4). Table 4 shows the best RREs \( \|f_k - f_*\|/\|f_*\| \) and the corresponding iteration \( k \) obtained by using the four preconditioners and several thresholds \( \alpha \) with the projected variant of the Landweber method.

In this second test, we try to adopt the strategy proposed by Hanke, Nagy and Plemmons for the choice of the regularization parameter \( \alpha \). Basically, the Fourier spectral components of the ("unregularized") optimal preconditioner \( B^*B \) are compared with the Fourier spectral components of the blurred and noisy image \( g \), in order to estimate the components of \( g \) where the noise dominates on the signal. We summarize the procedure by the following three steps: (i) collect in decreasing order the eigenvalues (computed by 2d-FFT) of the circulant optimal preconditioner \( B^*B \); (ii) look at the Fourier components of the blurred image \( g \) with respect to the same ordering of step
<table>
<thead>
<tr>
<th>noise</th>
<th>With Positivity</th>
<th>Without Positivity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3%  8%  30%</td>
<td>3%  8%  30%</td>
</tr>
<tr>
<td>No prec.</td>
<td>0.1794 0.1855 0.2185</td>
<td>0.1817 0.1928 0.2263</td>
</tr>
<tr>
<td></td>
<td>200 200 29</td>
<td>200 156 15</td>
</tr>
</tbody>
</table>

Filter F1

<table>
<thead>
<tr>
<th>noise</th>
<th>With Positivity</th>
<th>Without Positivity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3%  8%  30%</td>
<td>3%  8%  30%</td>
</tr>
<tr>
<td>$\alpha = 0.01$</td>
<td>0.1890 0.2016 0.1866</td>
<td>0.2091 0.2142 0.2083</td>
</tr>
<tr>
<td></td>
<td>2   2   2</td>
<td>1    1   1</td>
</tr>
<tr>
<td>$\alpha = 0.03$</td>
<td>0.1823 0.1907 0.2276</td>
<td>0.1906 0.1977 0.2361</td>
</tr>
<tr>
<td></td>
<td>7   5   1</td>
<td>4    3   1</td>
</tr>
<tr>
<td>$\alpha = 0.1$</td>
<td>0.1752 0.1881 0.2219</td>
<td>0.1763 0.1929 0.2268</td>
</tr>
<tr>
<td></td>
<td>49  20  3</td>
<td>69   16  2</td>
</tr>
</tbody>
</table>

Filter F4

<table>
<thead>
<tr>
<th>noise</th>
<th>With Positivity</th>
<th>Without Positivity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3%  8%  30%</td>
<td>3%  8%  30%</td>
</tr>
<tr>
<td>$\alpha = 0.01$</td>
<td>0.1849 0.1930 0.2642</td>
<td>0.1891 0.1964 0.2886</td>
</tr>
<tr>
<td></td>
<td>4   3   1</td>
<td>3    2   1</td>
</tr>
<tr>
<td>$\alpha = 0.03$</td>
<td>0.1844 0.1949 0.2301</td>
<td>0.1851 0.1949 0.2395</td>
</tr>
<tr>
<td></td>
<td>43  15  2</td>
<td>48   32  2</td>
</tr>
<tr>
<td>$\alpha = 0.1$</td>
<td>0.1861 0.1910 0.2240</td>
<td>0.1895 0.1956 0.2293</td>
</tr>
<tr>
<td></td>
<td>200 200 8</td>
<td>200  200  5</td>
</tr>
</tbody>
</table>

Table 3: Test 1 - Best Relative Restoration Errors and number of iterations for filters F1 and F4

<table>
<thead>
<tr>
<th>No prec.</th>
<th>0.4689</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>200</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.5176</td>
<td>0.4837</td>
<td>0.4351</td>
<td>0.5347</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>200</td>
<td>200</td>
<td>3</td>
</tr>
<tr>
<td>0.02</td>
<td>0.4924</td>
<td>0.5449</td>
<td>0.4568</td>
<td>0.5132</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>200</td>
<td>200</td>
<td>17</td>
</tr>
<tr>
<td>0.03</td>
<td>0.4249</td>
<td>0.5581</td>
<td>0.4606</td>
<td>0.4146</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>200</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td>0.04</td>
<td>0.3929</td>
<td>0.5752</td>
<td>0.4630</td>
<td>0.4340</td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>200</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td>0.05</td>
<td>0.3758</td>
<td>0.5911</td>
<td>0.4642</td>
<td>0.4496</td>
</tr>
<tr>
<td></td>
<td>62</td>
<td>200</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td>0.06</td>
<td>0.3623</td>
<td>0.6171</td>
<td>0.4677</td>
<td>0.4626</td>
</tr>
<tr>
<td></td>
<td>99</td>
<td>200</td>
<td>200</td>
<td>200</td>
</tr>
</tbody>
</table>

Table 4: Test 2 - Best Relative Restoration Errors and number of iterations.
(i), and choose the first index, say \( r \), such that all the following harmonics have small widths of approximately constant size; (iii) \( \lambda_r(B^*B) \) is the truncation parameter \( \alpha \) for the filtered optimal preconditioner \( D_\alpha \).

A graphic example of the three steps is given in Fig. 5. The graph on the left shows all the 256\(^2\) Fourier components, the graph on the right is the corresponding zoom onto the first 500. As already mentioned, the procedure is not simple to apply. Indeed, there is a large set of “reasonable” indices where the Fourier components of \( g \) start to be approximately constant, which gives rise to very different truncation parameters \( \alpha \). In the figure, the value \( \alpha \approx 10^{-3} \) seems to be a good “stagnating” value, corresponding to a noise space generated by the Fourier components of index greater than about 350. Unfortunately, this choice gives a too small parameter, as it can be observed in Table 4, and the results are unsatisfactory due to the high contribution of the noise in the restoration process. In our test, good values of \( \alpha \) are contained between \( 0.4 \cdot 10^{-1} \) and \( 10^{-1} \), corresponding to a noise space generated by the Fourier components more or less after the first 30 – 70. Hence, in order to avoid noise amplification, it seems to be better to overestimate the noise space by choosing a truncation parameter \( \alpha \) higher than the one located by steps (i) – (iii). In this way, although the convergence may be slower, the restoration is generally good.
Figure 6: Test 2 - Restorations of the Landweber method with positivity, filter F1.
Since it is easy to miss the proper value of $\alpha$, it is worth to notice that “sharp” filters are not appropriate for such situations, due to their discontinuity with respect to this parameter. This is confirmed by the column F4 of Table 4: there is a remarkable gap in the results between the second and the third row (in particular, a dramatic change occurs from 17 iterations for $\alpha = 0.02$ to 200 iterations for $\alpha = 0.03$). Similarly, other experiments not reported in the table show very irregular performances for the filter F3 when $\alpha$ goes from 0.0087 (RRE of 0.5734 in 3 iterations) to 0.06 (RRE of 0.4351 in 200 iterations).

Concerning the quality of the restorations, we can say that after 200 iterations the RRE without preconditioning is still very high (RRE=0.4689) and the reconstruction is unsatisfactory, as shown on the top-left of Fig. 6. A similar accuracy is provided within about 10 iterations by the Tikhonov filter, with $\alpha = 0.03$ as instance. The Tikhonov filter F1 gives again the best results, and for this filter the choice of the regularization parameter is simpler since it is not very sensitive with respect to non-optimal choices. The other choices, especially F2 and F3, do not provide good results, since they bring too much regularization, that is, the convergence speed is too slow along the components related to the smallest eigenvalues.

Notice that the restorations with F1 are good for any $\alpha$ between 0.03 and 0.06. Within few iterations, we obtain very fast and sufficiently good restorations with the low values $\alpha = 0.03$ and $\alpha = 0.04$ (see the corresponding images on the top-right and the center-left of Fig. 6). On the other hand, if we are more interested in low RREs than in fast computation, we can adopt a larger $\alpha$. As instance, by using $\alpha = 0.06$, within about 100 iterations the details of the restored image are accurate, as shown by the image on the center-right of Fig. 6.

In addition, we remark that the projected Landweber method with regularizing preconditioners presented here allows us to recover better restorations than other widely used deblurring algorithms: with a very large value of the regularization parameter the convergence is not fast, but the restored image is very good (RRE=0.3510 at iteration 158 with $\alpha = 0.07$; RRE=0.3438 at iteration 200 with $\alpha = 0.08$). In this case, the method could be very favorable when the speed is not crucial. The image on the bottom-left of Fig. 6 relative to $\alpha = 0.08$, where RRE=0.3438, seems better than others provided in the previous literature on the same example [18, 20, 23].

6 Conclusions

In this paper, the Landweber iterative method with positive constraints on the iterations has been considered. On account of its quite good properties of reliability and easiness of implementation, we have proposed a regularizing preconditioned version of the algorithm for space-invariant image deblurring, which speeds up the convergence without losing the regularization effectiveness.

Our theoretical contribution has proceeded along two main directions.

First, a basic analysis of the method in a simplified setting related to periodic boundary conditions has been given. This way, the convergence properties of the algorithm have been easily addressed to provide practical rules about the choice of all the several parameters of the method.

Secondly, Dirichlet boundary conditions have been considered as a case study of other settings. In particular, by proving that the eigenvectors of the preconditioned matrix are distributed like the Fourier vectors, we have shown that the algorithm is able to speed up the convergence among all the eigendirections related to the signal space only. These arguments extend the results of [19] and the subsequent literature, where signal and noise spaces were described in terms of low and high frequencies, regardless the eigenvectors of the preconditioned matrix.

The numerical results have confirmed the main properties of the method, that is, robustness
and flexibility. On these grounds, the method may become a valid tool for the solution of inverse problems arising in real applications [2, 5, 4, 25, 31]. Indeed, other widely used strategies, such as CGLS, GMRES methods, or the MRNSD method [31], are less reliable: although much faster, these methods may lead to less accuracy in the restorations with respect to the proposed method, if the signal to noise ratio is not large or in severely ill-posed problems [16]. As instance, in these faster methods it is essential to choose the filtering preconditioner with very high accuracy, and stop the iterations really very close to the optimal point, and we know how both of these tasks are so crucial as difficult to solve. On the contrary, the regularizing preconditioned Landweber method with positivity allows us to operate a higher safety, since it is much less sensitive to non-optimal settings of the parameters.

References


