

Fast nonstationary preconditioned iterative methods for ill-posed problems, with application to image deblurring

Marco Donatelli* Martin Hanke†

July 8, 2013

Abstract

We introduce a new iterative scheme for solving linear ill-posed problems, similar to nonstationary iterated Tikhonov regularization, but with an approximation of the underlying operator to be used for the Tikhonov equations. For image deblurring problems such an approximation can be a discrete deconvolution that operates entirely in the Fourier domain. We provide a theoretical analysis of the new scheme, using regularization parameters that are chosen by a certain adaptive strategy. The numerical performance of this method turns out to be superior to state of the art iterative methods, including the conjugate gradient iteration for the normal equation, with and without additional preconditioning.

1 Introduction

We consider the iterative solution of ill-posed equations

$$Tx = y, \tag{1.1}$$

where $T : \mathcal{X} \rightarrow \mathcal{Y}$ is a linear operator between two Hilbert spaces \mathcal{X} and \mathcal{Y} . We assume that y is attainable, i.e., that problem (1.1) has a solution x^\dagger of minimal norm. In the present context, the phrase *ill-posed* is used to indicate that the (Moore-Penrose) generalized inverse operator of T is not bounded, and hence, problem (1.1) has to be regularized for a numerical solution.

As far as iterative regularization methods are concerned, these methods typically suffer under one of the following two shortcomings: Either they are extremely slow like, for example, the so-called Landweber iteration, or they are reasonably fast but may deteriorate if not terminated appropriately, like the conjugate gradient iteration (CGLS). We refer to [9, 17] for a comprehensive discussion of these and further properties of iterative regularization methods for linear ill-posed problems. Preconditioners can be used to accelerate the convergence, cf. [4, 8, 14, 16], but an imprudent choice of preconditioner may spoil the achievable quality of the computed restorations.

*Dipartimento di Scienza e Alta Tecnologia, Università dell'Insubria, 22100 Como, Italy

†Institut für Mathematik, Johannes Gutenberg-Universität Mainz, 55128 Mainz, Germany

While both, the Landweber iteration as well as CGLS are based on the normal equation $T^*Tx = T^*y$ associated with (1.1), we propose here a new iterative scheme that uses some sort of (nonstationary) preconditioning applied to the original problem (1.1). In fact, we do not require T^* for our method, and this can be a benefit in some applications. Under certain assumptions that allow the construction of an appropriate approximation C of T the new scheme turns out to be even faster than CGLS, while at the same time being much more stable with respect to the associated termination criterion.

Our scheme can be viewed as a variant of nonstationary iterated Tikhonov regularization, to the effect that the approximation C enters into the residual correction steps, when linear equations with $T^*T + \alpha I$ or $TT^* + \alpha I$, $\alpha \in \mathbb{R}$, are too expensive to solve. Of course, in that case efficient solvers for Tikhonov type equations with C need to be available to qualify a potential candidate for being a good choice of C , compare (2.3) below. At the same time, C should be close to T in an appropriate sense, see (2.8) below; this closeness property then rules the adaptive selection of the individual Tikhonov regularization parameters.

We provide an abstract convergence analysis of the new iterative scheme, and illustrate its numerical performance for some model problems from image deblurring. We start in Section 2 by providing some first preliminary motivation and formulating the final overall algorithm. Subsequently, in Section 3, convergence and regularization properties of the new scheme will be derived. Then we turn to image deblurring as the basic application we have in mind. A brief description of the corresponding setting is the subject of Section 4, and after that we discuss some numerical examples in Section 5. Finally, we conclude with a short summary of our results.

2 A nonstationary preconditioned iteration

Throughout this paper we assume that, instead of the exact data $y \in \mathcal{Y}$ of (1.1), we are only given approximate data $y^\delta \in \mathcal{Y}$ with

$$\|y^\delta - y\| \leq \delta, \quad (2.1)$$

where $\delta \geq 0$ is the corresponding noise level, with the understanding that the data are exact when $\delta = 0$.

The algorithm we propose has the following form (see Algorithm I below for the final version, with all parameters specified): Starting with an initial guess x_0 of x^\dagger we compute, for $n = 0, 1, 2, \dots$,

$$h_n = C^*(CC^* + \alpha_n I)^{-1} r_n, \quad r_n = y^\delta - Tx_n, \quad (2.2a)$$

and set

$$x_{n+1} = x_n + h_n. \quad (2.2b)$$

Note that the linear equation (2.2a) is equivalent to minimizing the Tikhonov functional

$$\|Ch_n - r_n\|^2 + \alpha_n \|h_n\|^2 \longrightarrow \min. \quad (2.3)$$

over $h_n \in \mathcal{X}$, where C is the aforementioned approximation of T , and α_n is the associated regularization parameter. In the literature on iterative solvers for (usually well-posed) problems of the form (1.1) the operator $P = C^*(CC^* + \alpha_n I)^{-1}$ in (2.2a) would be called *preconditioner*. As

we will select different regularization parameters α_n in each iteration (see below), it is therefore appropriate to call the new scheme (2.2) a nonstationary preconditioned iteration.

Before we come to the selection of the regularization parameters $(\alpha_n)_n$ we list three different motivations for the scheme (2.2).

- If the operator C in (2.2a) or (2.3), respectively, were replaced by T then (2.2) would be the (nonstationary) iterated Tikhonov method investigated in [13]. To achieve rapid convergence, a recommended choice for the parameter sequence $(\alpha_n)_n$ is the geometric sequence

$$\alpha_n = \alpha q^n, \quad n = 0, 1, 2, \dots, \quad (2.4)$$

where $\alpha > 0$ and $0 < q \leq 1$. The classical stationary iterated Tikhonov method is obtained for $q = 1$. As mentioned before the iterated Tikhonov method would require an efficient solver for linear systems with operators $T^*T + \alpha_n I$. Variant (2.3) will be a preferred alternative if systems with $C^*C + \alpha_n I$ can be solved more easily.

- The iterated Tikhonov method can be interpreted as an iterative refinement procedure, where the update h_n is obtained by applying Tikhonov regularization to the error equation $Te_n \approx r_n$, where

$$e_n = x^\dagger - x_n \quad (2.5)$$

is the error after n steps. Note that in view of (2.1) this error equation is (only) correct up to the perturbation in the data. Taking this into account one may as well consider instead the “model equation”

$$Ce_n \approx r_n, \quad (2.6)$$

possibly tolerating a slightly larger misfit. Solving (2.6) by means of Tikhonov regularization then corresponds to (2.3), resp. (2.2a). With this point of view the choice of the regularization parameters should also reflect how much we trust in the model (2.6).

- This leads us to our third motivation. The Levenberg-Marquardt iteration is a means to solve nonlinear equations, where in each iteration a Tikhonov-type functional (2.3) is minimized, with C being an approximate derivative of the respective nonlinear operator, and the parameter α_n , again, depends on how much one trusts in the corresponding linearized model. For ill-posed (nonlinear) problems a variant of this scheme was proposed in [12], and there it has been suggested to choose the parameter α_n such that the model equation (2.6) is only solved up to a certain relative amount, i.e., such that

$$\|r_n - Ch_n\| = q_n \|r_n\|, \quad (2.7)$$

where q_n is smaller than one, but not too small. This is the choice that we adopt here for our purposes.

To derive a sophisticated value for the parameter q_n in (2.7) we impose a closeness assumption on the given approximation C of T , namely that

$$\|(C - T)z\| \leq \rho \|Tz\|, \quad z \in \mathcal{X}, \quad (2.8)$$

for some $0 < \rho < 1/2$. We mention that such a condition may be hard to satisfy for a specific problem, as it implies – using the triangle inequality – that

$$(1 - \rho)\|Tz\| \leq \|Cz\| \leq (1 + \rho)\|Tz\| \quad (2.9)$$

for all $z \in \mathcal{X}$, i.e., that C and T are spectrally equivalent. However, in the context of image deblurring an assumption like (2.8) may not be far fetched, see Section 4.

Proposition 1 *Assume that (2.8) is satisfied for some $0 < \rho < 1/2$, and let $\tau_* = (1 + \rho)/(1 - 2\rho)$. Then, if $\tau_n = \|r_n\|/\delta > \tau_*$, it follows that*

$$\|r_n - Ce_n\| \leq \left(\rho + \frac{1 + \rho}{\tau_n}\right)\|r_n\| < (1 - \rho)\|r_n\|. \quad (2.10)$$

Proof: We rewrite

$$r_n - Ce_n = y^\delta - Tx_n - Ce_n = y^\delta - y + (T - C)e_n,$$

and use (2.1) and (2.8) to estimate

$$\begin{aligned} \|r_n - Ce_n\| &\leq \|y^\delta - y\| + \rho\|Te_n\| \leq \|y^\delta - y\| + \rho(\|r_n\| + \|y^\delta - y\|) \\ &\leq (1 + \rho)\delta + \rho\|r_n\|. \end{aligned}$$

By assumption, we have $\delta = \|r_n\|/\tau_n$, and this then yields the first inequality in (2.10). The second inequality follows from

$$\rho + \frac{1 + \rho}{\tau_n} < \rho + \frac{1 + \rho}{\tau_*} = 1 - \rho.$$

□

Given the assumption (2.8) Proposition 1 provides a justification of the model equation (2.6) as long as the size of the residual r_n is not too close to the noise level. Now we can summarize our nonstationary preconditioned iterative scheme as follows:

Algorithm I *Let $x_0 \in \mathcal{X}$ be given, and set $r_0 = y^\delta - Tx_0$. Choose $\tau = (1 + 2\rho)/(1 - 2\rho)$ with ρ from (2.8), and fix $q \in (2\rho, 1)$.*

While $\|r_n\| > \tau\delta$, let $\tau_n = \|r_n\|/\delta$, and compute

$$h_n = C^*(CC^* + \alpha_n I)^{-1}r_n, \quad (2.11a)$$

where α_n is such that

$$\|r_n - Ch_n\| = q_n\|r_n\|, \quad q_n = \max\{q, 2\rho + (1 + \rho)/\tau_n\}, \quad (2.11b)$$

and update

$$x_{n+1} = x_n + h_n, \quad r_{n+1} = y^\delta - Tx_{n+1}. \quad (2.11c)$$

The parameter q in Algorithm I is meant as a safeguard to prevent that the residual decreases too rapidly. Our theoretical results do not utilize this parameter. However, if ρ happens to be too small, or if (2.8) is only satisfied approximately, then error components may build up strongly; see also the discussion at the end of Section 5.1.

We mention that, by construction, and by virtue of Proposition 1, there is a unique positive regularization parameter α_n that determines h_n in the prescribed manner, cf., e.g., Groetsch [11]. This parameter can be computed with few steps of an appropriate Newton scheme, cf., e.g., [9, Prop. 9.8]. Accordingly, Algorithm I is well defined; moreover, as we will see in the following section, if $\delta > 0$, then this algorithm will terminate after $n = n_\delta \geq 0$ iterations with

$$\|r_{n_\delta}\| \leq \tau\delta < \|r_n\|, \quad n = 0, 1, \dots, n_\delta - 1. \quad (2.12)$$

This stopping rule is known as *discrepancy principle*. When $\delta = 0$, on the other hand, i.e., if the data are exact, the iterates x_n of the algorithm will converge to an exact solution as $n \rightarrow \infty$, cf. Theorem 4. To simplify the formulation of subsequent results, we formally set $n_\delta = +\infty$ in this case.

3 Convergence analysis of the nonstationary iteration

We now turn to the theoretical properties of Algorithm I, following mostly the line of argument developed in [12] for the analysis of the aforementioned regularizing Levenberg-Marquardt scheme for nonlinear ill-posed problems.

Proposition 2 *Under assumption (2.8) the norm of the iteration error $e_n = x^\dagger - x_n$ of Algorithm I decreases monotonically for $n = 0, 1, \dots, n_\delta - 1$:*

$$\|e_n\|^2 - \|e_{n+1}\|^2 \geq 2\rho \|(CC^* + \alpha_n I)^{-1} r_n\| \|r_n\|. \quad (3.1)$$

Proof: From (2.11) follows

$$\begin{aligned} \|e_n\|^2 - \|e_{n+1}\|^2 &= 2\langle e_n, h_n \rangle - \|h_n\|^2 \\ &= 2\langle Ce_n, (CC^* + \alpha_n I)^{-1} r_n \rangle - \langle r_n, CC^*(CC^* + \alpha_n I)^{-2} r_n \rangle \\ &= 2\langle r_n, (CC^* + \alpha_n I)^{-1} r_n \rangle - \langle r_n, CC^*(CC^* + \alpha_n I)^{-2} r_n \rangle \\ &\quad - 2\langle r_n - Ce_n, (CC^* + \alpha_n I)^{-1} r_n \rangle \\ &\geq 2\langle r_n, (CC^* + \alpha_n I)^{-1} r_n \rangle - 2\langle r_n, CC^*(CC^* + \alpha_n I)^{-2} r_n \rangle \\ &\quad - 2\langle r_n - Ce_n, (CC^* + \alpha_n I)^{-1} r_n \rangle \\ &= 2\alpha_n \langle r_n, (CC^* + \alpha_n I)^{-2} r_n \rangle - 2\langle r_n - Ce_n, (CC^* + \alpha_n I)^{-1} r_n \rangle \\ &\geq 2\alpha_n \langle r_n, (CC^* + \alpha_n I)^{-2} r_n \rangle - 2\|r_n - Ce_n\| \|(CC^* + \alpha_n I)^{-1} r_n\| \\ &= 2\|(CC^* + \alpha_n I)^{-1} r_n\| \left(\|\alpha_n (CC^* + \alpha I)^{-1} r_n\| - \|r_n - Ce_n\| \right). \end{aligned}$$

Now, since $\alpha_n (CC^* + \alpha_n I)^{-1} r_n = r_n - Ch_n$ by virtue of (2.11a), we have

$$\|e_n\|^2 - \|e_{n+1}\|^2 \geq 2\|(CC^* + \alpha_n I)^{-1} r_n\| \left(\|r_n - Ch_n\| - \|r_n - Ce_n\| \right).$$

Inserting (2.11b) and the estimate (2.10) we thus conclude that

$$\begin{aligned}\|e_n\|^2 - \|e_{n+1}\|^2 &\geq 2\|(CC^* + \alpha_n I)^{-1}r_n\| \left(q_n \|r_n\| - \left(\rho + \frac{1+\rho}{\tau_n}\right) \|r_n\| \right) \\ &\geq 2\rho\|(CC^* + \alpha_n I)^{-1}r_n\| \|r_n\|,\end{aligned}$$

as has been claimed. \square

From inequality (3.1) of Proposition 2 we can draw further important conclusions.

Corollary 3 *With the assumption and notation of Proposition 2 there holds*

$$\|e_0\|^2 \geq 2\rho \sum_{n=0}^{n_\delta-1} \|(CC^* + \alpha_n I)^{-1}r_n\| \|r_n\| \geq c \sum_{n=0}^{n_\delta-1} \|r_n\|^2 \quad (3.2)$$

for some constant $c > 0$, depending only on ρ of (2.8) and on the parameter q of Algorithm I.

Proof: The first inequality simply follows by taking the sum of (3.1) from $n = 0$ up to $n = n_\delta - 1$:

$$\|e_0\|^2 \geq 2\rho \sum_{n=0}^{n_\delta-1} \|(CC^* + \alpha_n I)^{-1}r_n\| \|r_n\|. \quad (3.3)$$

To proceed further we need to derive lower bounds for $\|(CC^* + \alpha_n I)^{-1}r_n\|$, when $0 \leq n < n_\delta$. We first estimate α_n . For $\alpha > q_n \|C\|^2 / (1 - q_n)$ and $0 \leq \lambda \leq \|C\|^2$ there holds

$$\frac{\alpha}{\lambda + \alpha} \geq \frac{\alpha}{\|C\|^2 + \alpha} = (1 + \|C\|^2 / \alpha)^{-1} > q_n,$$

and hence,

$$\|\alpha(CC^* + \alpha I)^{-1}r_n\| > q_n \|r_n\|,$$

as $\|r_n\| > 0$ for $n < n_\delta$. Since the residual norm of the Tikhonov approximation is strictly increasing with α (cf. [11]) this implies that the parameter α_n of (2.11b) satisfies $\alpha_n \leq q_n \|C\|^2 / (1 - q_n)$, and therefore

$$\|(CC^* + \alpha_n I)^{-1}r_n\| = \frac{1}{\alpha_n} \|r_n - Ch_n\| = \frac{q_n}{\alpha_n} \|r_n\| \geq \frac{1 - q_n}{\|C\|^2} \|r_n\|.$$

Now, according to the choice of parameters in Algorithm I, $1 - q_n = \min\{1 - q, 1 - 2\rho - (1 + \rho)/\tau_n\}$, and

$$1 - 2\rho - (1 + \rho)/\tau_n = \frac{1 + 2\rho}{\tau} - \frac{1 + \rho}{\tau_n} > \frac{1 + 2\rho}{\tau} - \frac{1 + \rho}{\tau} = \rho/\tau.$$

Therefore, there exists $c > 0$, depending only on ρ and q such that $1 - q_n \geq c \|C\|^2 / (2\rho)$, and

$$\|(CC^* + \alpha_n I)^{-1}r_n\| \geq \frac{c}{2\rho} \|r_n\| \quad \text{for } n = 0, 1, \dots, n_\delta - 1.$$

Inserting this into (3.3) the second assertion follows as well. \square

The implications of Corollary 3 are two-fold: First, it follows from the outer inequality in (3.2) that the sum of squares of the residual norms is bounded, and hence, if $\delta > 0$ there must be a first integer $n_\delta < \infty$ such that (2.12) is fulfilled. In other words: Algorithm I terminates after finitely many iterations, when $\delta > 0$.

On the other hand, if the data are exact ($\delta = 0$) then it can be deduced from Corollary 3 that the iterates of Algorithm I converge to an exact solution of (1.1) as $n \rightarrow \infty$. This we show next.

Theorem 4 *Assume that the data are exact, i.e., $\delta = 0$, and that x_0 is no solution of problem (1.1). Then the sequence $(x_n)_n$ converges as $n \rightarrow \infty$ to the solution of (1.1) that is closest to x_0 .*

Proof: If $\delta = 0$ then the stopping criterion (2.12) can only be satisfied with $n = n_\delta$ for a solution x_n of (1.1). In this case, if $n > 0$ then h_{n-1} must coincide with e_{n-1} up to an element in the null space of T , that is, in the null space of C , by virtue of (2.9). Accordingly, it follows from (2.11b) and Proposition 1 that

$$q_{n-1}\|r_{n-1}\| = \|r_{n-1} - Ch_{n-1}\| = \|r_{n-1} - Ce_{n-1}\| \leq \left(\rho + \frac{1+\rho}{\tau_{n-1}}\right)\|r_{n-1}\|.$$

This, however, contradicts the definition (2.11b) of q_{n-1} . Therefore, the iteration does not terminate for exact data, unless x_0 is already a solution of (1.1).

We show next that $(x_n)_n$ is a Cauchy sequence. To this end, consider for $m > l$

$$\begin{aligned} \|x_m - x_l\|^2 &= \|e_m - e_l\|^2 = \|e_m\|^2 - \|e_l\|^2 - 2\langle e_l, e_m - e_l \rangle \\ &= \|e_m\|^2 - \|e_l\|^2 + 2\langle e_l, x_m - x_l \rangle. \end{aligned}$$

Inserting (2.11c) and (2.11a) it follows that

$$\begin{aligned} \|x_m - x_l\|^2 &= \|e_m\|^2 - \|e_l\|^2 + 2 \sum_{k=l}^{m-1} \langle e_l, h_k \rangle \\ &= \|e_m\|^2 - \|e_l\|^2 + 2 \sum_{k=l}^{m-1} \langle Ce_l, (CC^* + \alpha_k I)^{-1} r_k \rangle \\ &\leq \|e_m\|^2 - \|e_l\|^2 + 2 \sum_{k=l}^{m-1} \|Ce_l\| \|(CC^* + \alpha_k I)^{-1} r_k\|. \end{aligned}$$

Likewise, we obtain for $l \geq n$ that

$$\begin{aligned}
\|x_l - x_n\|^2 &= \|e_n\|^2 - \|e_l\|^2 - 2\langle e_l, e_n - e_l \rangle = \|e_n\|^2 - \|e_l\|^2 - 2 \sum_{k=n}^{l-1} \langle e_l, h_k \rangle \\
&= \|e_n\|^2 - \|e_l\|^2 - 2 \sum_{k=n}^{l-1} \langle Ce_l, (CC^* + \alpha_k I)^{-1} r_k \rangle \\
&\leq \|e_n\|^2 - \|e_l\|^2 + 2 \sum_{k=n}^{l-1} \|Ce_l\| \|(CC^* + \alpha_k I)^{-1} r_k\|.
\end{aligned}$$

Together with (2.9) it thus follows that for general $m > n$ and any $l \in \{n, \dots, m-1\}$ there holds

$$\begin{aligned}
\|x_m - x_n\|^2 &\leq 2\|x_m - x_l\|^2 + 2\|x_l - x_n\|^2 \\
&\leq 2\|e_m\|^2 + 2\|e_n\|^2 - 4\|e_l\|^2 + 4 \sum_{k=n}^{m-1} \|Ce_l\| \|(CC^* + \alpha_k I)^{-1} r_k\| \\
&\leq 2\|e_m\|^2 + 2\|e_n\|^2 - 4\|e_l\|^2 + 4(1 + \rho) \sum_{k=n}^{m-1} \|r_l\| \|(CC^* + \alpha_k I)^{-1} r_k\|.
\end{aligned}$$

Now we choose l to be that particular iteration index in the interval $\{n, \dots, m-1\}$, for which $\|r_l\|$ becomes minimal, so that

$$\|x_m - x_n\|^2 \leq 2\|e_m\|^2 + 2\|e_n\|^2 - 4\|e_l\|^2 + 4(1 + \rho) \sum_{k=n}^{m-1} \|r_k\| \|(CC^* + \alpha_k I)^{-1} r_k\|.$$

Since the sequence $(\|e_k\|)_k$ is monotonically decreasing and nonnegative, it is convergent with some limit $\epsilon \geq 0$, and thus the above right-hand side becomes arbitrarily small for n and m sufficiently large, as its last term is the partial sum of a converging series, see Corollary 3. This means that $(x_n)_n$ is a Cauchy sequence with limit $x \in \mathcal{X}$, say. Accordingly, the residues $r_n = y - Tx_n$ converge to $y - Tx$, while, by virtue of Corollary 3 again, r_n converges to zero at the same time. This shows that $Tx = y$, i.e., that x solves (1.1).

By construction every iterate x_n satisfies

$$x_n - x_0 = \sum_{k=0}^{n-1} h_k \in \mathcal{R}(C^*) = \mathcal{N}(C)^\perp,$$

and hence, $x - x_0$ is orthogonal to the null space of C . This null space, however, coincides with the null space of T by virtue of (2.9). Therefore, $x - x_0 \in \mathcal{N}(T)^\perp$, i.e., x is the particular solution of (1.1) that is closest to x_0 in the norm of \mathcal{X} . \square

For inexact data, on the other hand, Algorithm I is a regularization method, provided that (2.8) holds true. The precise statement is as follows.

Theorem 5 Assume that (2.8) holds for some $0 < \rho < 1/2$, and let $\delta \mapsto y^\delta$ be a function from \mathbb{R}^+ to \mathcal{Y} such that (2.1) holds true for all $\delta > 0$. For fixed parameters τ and q denote by n_δ the corresponding stopping indices of Algorithm I, and by x^δ the resulting approximations. Then, as $\delta \rightarrow 0$, x^δ converges to the solution of (1.1) that is closest to x_0 in the norm of \mathcal{X} .

We omit the proof, because it can be copied from the proof of Theorem 2.3 in [12]; see also [9, Thm. 11.5]. Its essential ingredients are the monotonicity property established in Proposition 2, and the continuous dependence of the n th iterate x_n on the data y^δ .

4 Image deblurring

As an application we now turn to image deblurring problems. Such problems are usually formulated as large linear systems of the form

$$\mathbf{T}\mathbf{x} = \mathbf{y} \tag{4.1}$$

where the vector \mathbf{x} represents the pixel values of an unknown true image, \mathbf{y} contains the pixel values of the observed (blurred) image, and the matrix \mathbf{T} describes the blurring phenomenon, cf., e.g., Bertero [3]. Again, in practice, noisy data \mathbf{y}^δ are given instead of \mathbf{y} , with

$$\|\mathbf{y}^\delta - \mathbf{y}\| \leq \delta,$$

where, often, the noise level δ is known or can be estimated. For present day imaging purposes, the vectors \mathbf{x} and \mathbf{y} (resp. \mathbf{y}^δ) can be of dimension 10^6 and higher, and accordingly, the matrix \mathbf{T} is extremely large. Although, strictly speaking, its Moore-Penrose inverse is continuous (as \mathbf{T} is acting in finite dimensional spaces), the formulation (4.1) should be viewed as a discretization of an underlying first kind integral equation (1.1) that is ill-posed in the sense used before. Accordingly, (4.1) is very ill-conditioned and, following the terminology introduced by Hansen [15], we call (4.1) a *discrete ill-posed problem*.

In this paper we confine ourselves to the most elementary situation, and assume that the blur of the image can be described by a space invariant point spread function (PSF). While the corresponding integral operator is then of convolution type, the discrete problem also has to take into account the limited field of view, which requires the incorporation of appropriate boundary conditions into the model.

The resulting matrices are usually hard to invert explicitly, and for the same reason, Tikhonov-type functionals like

$$\|\mathbf{y}^\delta - \mathbf{T}\mathbf{x}\|^2 + \alpha \|\mathbf{x}\|^2$$

are difficult to minimize. In the past these discrete problems have therefore been solved by Fourier filtering techniques (discrete deconvolution, cf., e.g., Andrews and Hunt [1]), but it is now commonly recognized that these Fourier techniques introduce annoying artefacts into the reconstructions when the image is not periodic outside the field of view. Today iterative methods are the preferred option for the solution of (4.1), cf., e.g., Berisha and Nagy [2] for a recent survey on these matters.

Fourier filtering techniques can, however, be exploited for the purpose of preconditioning the problem. A first proposal of this sort was made in [14]; however, the corresponding preconditioned conjugate gradient iteration (PCGLS) has stability problems, as the choice of the free parameters is somewhat tricky. As a side comment we also mention that the known preconditioners have been designed for the normal equations system associated with (4.1), and therefore do not apply to all boundary conditions equally well; see Sect. 5.3 below.

Here we make a different suggestion and recommend to apply Algorithm I to the discrete problem (4.1), where the coefficient matrix \mathbf{T} that carries the “correct” boundary conditions of the model is approximated by the doubly circulant matrix \mathbf{C} that corresponds to the discrete Fourier deconvolution (that is, to periodic boundary conditions) of a suitably truncated PSF (where necessary). With this choice of \mathbf{C} the computation of (2.11a), but also the adaptive choice of the regularization parameters α_n in order to fulfill (2.11b) can be realized very efficiently in the Fourier domain, such that the evaluation of the residual \mathbf{r}_n remains the major computational burden in each iteration.

The suggestion of approximating \mathbf{T} by this particular matrix \mathbf{C} goes back to Strang [21]. As far as our assumption (2.8) concerning the closeness of this approximation is concerned, we start by mentioning that the boundary conditions have a very local effect on the overall system, i.e., the approximation error $\mathbf{T} - \mathbf{C}$ can generically be decomposed as

$$\mathbf{T} - \mathbf{C} = \mathbf{E} + \mathbf{R}, \quad (4.2)$$

where \mathbf{E} is a matrix of small norm (it can be zero, if the PSF is compactly supported), and \mathbf{R} is a matrix of small rank (compared to the dimension of the problem), cf., e.g., Ng [19]. Images in the range of \mathbf{R} are usually zero except for small neighborhoods of the boundary, but we have to admit that we do not know whether a rigorous estimate of the form (2.8), i.e.,

$$\|(\mathbf{T} - \mathbf{C})\mathbf{z}\| \leq \rho \|\mathbf{T}\mathbf{z}\|$$

will hold for all (relevant) vectors \mathbf{z} and some parameter $\rho < 1/2$. However, cf. Figure 4 (b) below, our numerical results confirm that the model equation (2.6) holds to good relative accuracy, in that the estimate (2.10), which is one of the major consequences of (2.8), is nicely satisfied with an appropriate (small) choice of ρ of a few percent. An explanation may be that for (2.6) to be an appropriate model, accurate boundary conditions are not that important, since the error is more or less well distributed over the domain – at least in the earlier stages of the iteration.

As we will see in the numerical results of the following section our new nonstationary preconditioned iteration is faster than any other iterative regularization method for this kind of problems, but still very robust with respect to the termination criterion.

5 Numerical results

The following experiments have been carried out in Matlab 7.12 using Version 2 of the Matlab Toolbox `RestoreTools` [18], see also [2]. For the nonstationary preconditioned iteration (2.2) we compare the geometric sequence (2.4) of parameters α_n (labeled “Geometric” in the plots) with their adaptive selection as defined in Algorithm I (“Algor. I”). In both variants the parameter q determines how fast the sequence $(\alpha_n)_n$ decreases, and we fix $q = 0.7$ in either case. (See the end

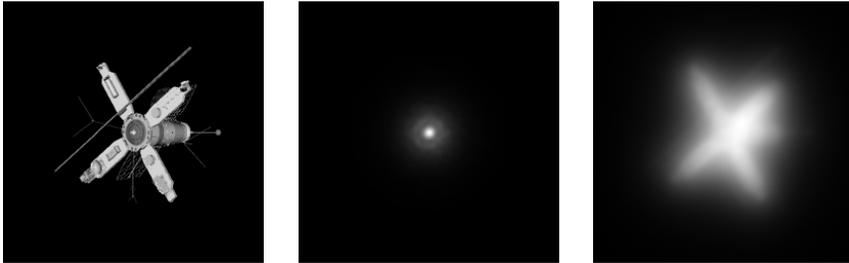


Figure 1: Example 1 – from left to right: true image, PSF, and observed image with $\nu = 0.5\%$.

of Sect. 5.1 for a discussion of this particular choice.) Moreover, for the geometric sequence (2.4) we start with $\alpha_0 = \alpha = 0.5$ because the PSF has mean one and is nonnegative, hence $\|\mathbf{T}\|_\infty \approx 1$. The two nonstationary preconditioned algorithms are compared with CGLS and PCGLS, where the latter uses the preconditioner from [14] provided by the toolbox. For every iterative method we choose the given data as initial approximation \mathbf{x}_0 .

For our simulations the exact data \mathbf{y} have been perturbed by synthetically generated noise vectors with normally distributed entries with zero mean. The variance of the entries is chosen so as to achieve different noise levels δ , and below we refer to

$$\nu = \frac{\delta}{\|\mathbf{y}\|}$$

as the corresponding relative amount of noise. All iterative methods are terminated according to the discrepancy principle (2.12); while Algorithm I requires to choose $\tau = (1 + 2\rho)/(1 - 2\rho)$, we use $\tau = 1.01$ for the other three algorithms. To compare the quality of the restorations, we evaluate their relative restoration errors (RRE), i.e.,

$$\text{RRE} = \frac{\|\mathbf{x} - \mathbf{x}^\dagger\|}{\|\mathbf{x}^\dagger\|},$$

where \mathbf{x} is the computed solution. Since the relevance of this number is often criticized, we also display the corresponding restorations to allow for a visual comparison.

5.1 Example 1

We start with a test problem of size 256×256 , shown in Figure 1, that has often been used as a benchmark in the literature, cf., e.g. [14, 18], with a PSF designed to model atmospheric blur, and $\nu = 0.5\%$ noise. This example employs zero Dirichlet boundary conditions, and hence \mathbf{T} is a block Toeplitz matrix with Toeplitz blocks. Recall that we use for \mathbf{C} the block circulant matrix with circulant blocks obtained by imposing periodic boundary conditions. For this example we have found that we can work with a small value $\rho = 10^{-3}$ to satisfy (2.8) and (2.10).

We refer to Figures 2 and 3 for the iteration history and the restored images of the four different methods, respectively. These results highlight the effectiveness of our nonstationary preconditioned method, both for the geometric sequence and the adaptive choice of the regularization parameters. In particular, both algorithms exhibit fast convergence without losing in

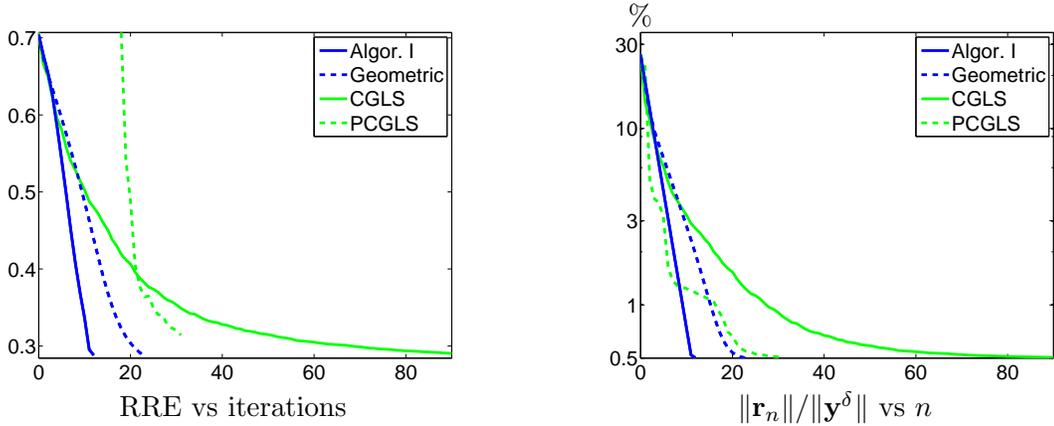


Figure 2: Example 1 – RRE and relative norm of the residual at each iteration.

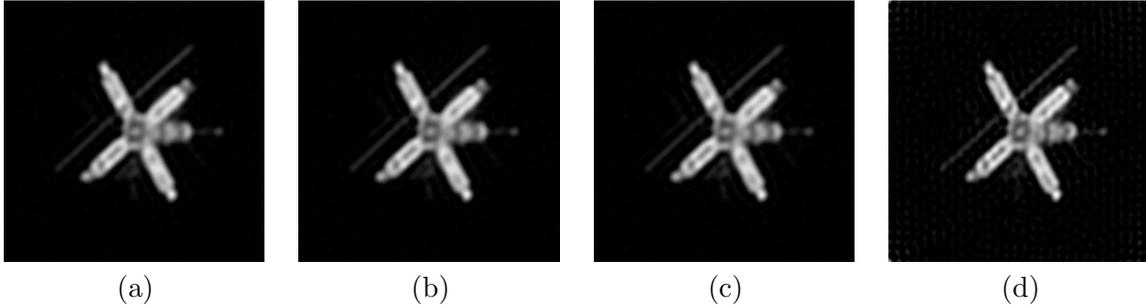


Figure 3: Example 1 – restored images.

- (a) Algorithm I with adaptive choice: RRE = 0.288, it. 12;
- (b) Nonstationary iteration (2.2) with geometric sequence (α_n) : RRE = 0.287, it. 23.;
- (c) CGLS: RRE = 0.290, it. 90; (d) PCGLS: RRE = 0.314, it. 31.

quality of the restoration or stability, as it sometimes happens with PCGLS. In particular, in this example PCGLS exhibits a strong increase of the error norm in the very first iteration.

Figure 4 (a) depicts the regularization parameters $(\alpha_n)_n$ in logarithmic scale. We note that in the very last iteration of Algorithm I the regularization parameter is increasing in order to satisfy (2.11b), because then also the value of q_n increases, compare Figure 4 (b). This latter plot not only shows q_n as a function of n , but also exhibits the quality of the model fit $\|\mathbf{r}_n - \mathbf{C}\mathbf{e}_n\|/\|\mathbf{r}_n\|$ for each iteration. Thus we can check that for our choice of ρ the quantity q_n is indeed an upper bound for this model fit as predicted by Proposition 1. As a consequence the monotonic decrease of the iteration error of Algorithm I is in accordance with Proposition 2.

We hasten to add that for this particular example Algorithm I is not very sensitive to the choice of ρ . Indeed, for $\rho = 10^{-4}$ we obtain RRE = 0.287 after 13 iterations, while for $\rho = 10^{-2}$ we have RRE = 0.295 after 11 iterations. Likewise, Algorithm I is robust also with respect to the parameter q : Figure 5 shows the values of q_n and the quality of the model fit $\|\mathbf{r}_n - \mathbf{C}\mathbf{e}_n\|/\|\mathbf{r}_n\|$ when taking $q = 0.6$ and $q = 0.8$, respectively. With $q = 0.6$ Algorithm I terminates after 10 iterations with RRE = 0.287, while for $q = 0.8$ it terminates after 18 iterations with RRE =

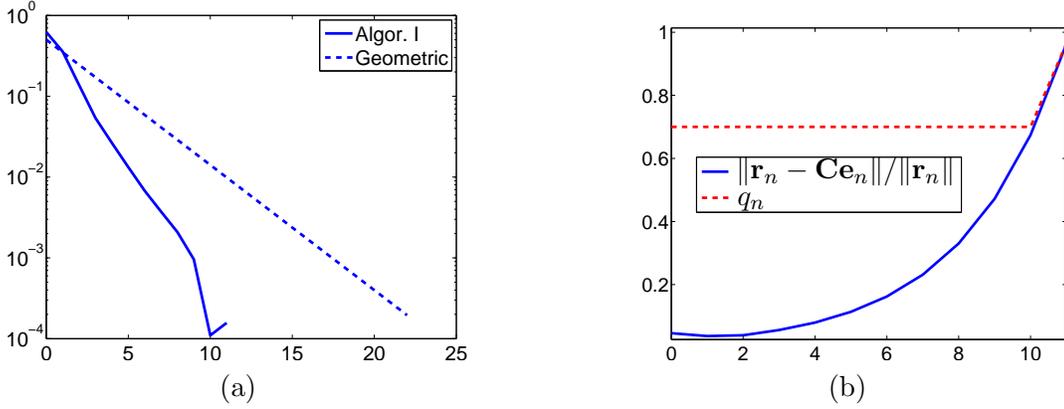


Figure 4: Example 1 – (a) α_n vs n ; (b) verification of (2.10)

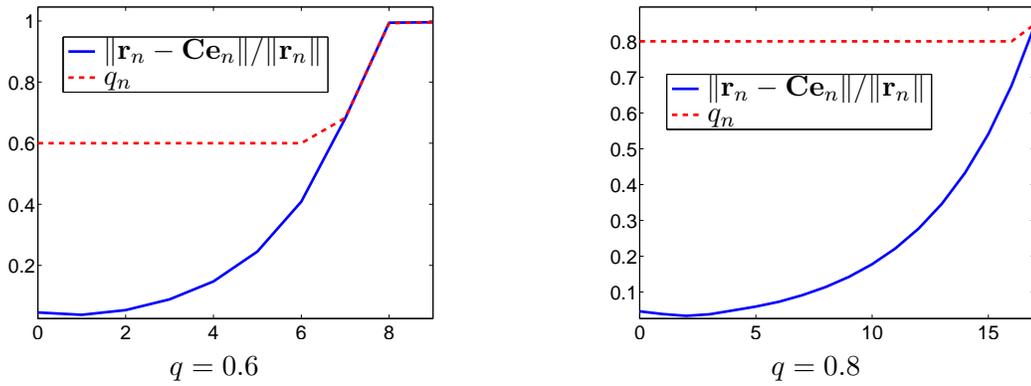


Figure 5: Example 1 – values of q_n and verification of (2.10) for $q = 0.6$ and $q = 0.8$.

0.288. We do not recommend, however, to choose this parameter much smaller than $q = 0.6$, as for smaller values of q and small noise the algorithm with the geometric sequence $(\alpha_n)_n$ may fail to meet the stopping criterion. The adaptive strategy of Algorithm I appears to be more robust in that sense, but still, the residual norms can start oscillating when the regularization parameters are forced to be too small too early in the iteration, and then the convergence history may become somewhat erratic; see the discussion of Figure 9 in Sect. 5.2.

5.2 Example 2

Next we consider the well-known Shepp-Logan phantom displayed in Figure 6, cropped down to size 237×237 , and degraded by a (nonsymmetric) Gaussian blur defined in the file `Gaussian-Blur422.mat` included in Version 2 of `RestoreTools`. Due to the black (i.e., zero) background we again set up \mathbf{T} by imposing zero Dirichlet boundary conditions and choose the same closeness parameter $\rho = 10^{-3}$ as before. For this second example PCGLS, as provided by the toolbox, fails in determining an appropriate threshold parameter (called τ in [14, 18]) to construct the corre-

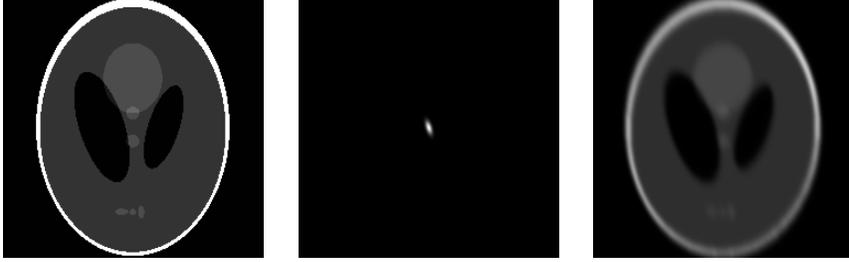


Figure 6: Example 2. From left to right: true image, PSF, and blurred image without noise.

ν	Algorithm I	Iteration (2.2) with geometric sequence	CGLS	PCGLS
0.5 %	0.281 (13)	0.282 (18)	0.284 (55)	0.289 (18)
1 %	0.293 (11)	0.296 (14)	0.296 (32)	0.303 (17)
3 %	0.318 (7)	0.317 (10)	0.321 (13)	0.318 (9)

Table 1: Example 2: RREs for different noise levels with the number of iterations in parentheses.

sponding preconditioner, and so we fix this parameter by hand to be equal to 3ν . This choice does accelerate the convergence of CGLS without losing in quality of the restoration, although it may not be the optimal parameter, which is hard to estimate a priori.

In view of the results for different noise levels listed in Table 1 we emphasize the robustness of the nonstationary preconditioned method, both for the geometric sequence and the adaptive choice of $(\alpha_n)_n$: For each noise level the nonstationary preconditioned methods compute restorations with RREs that are comparable to the ones of CGLS, but with much fewer iterations. Moreover, we recall that matrix–vector products with $\mathbf{C}^*(\mathbf{C}\mathbf{C}^* + \alpha_n\mathbf{I})^{-1}$ are computationally cheaper than those with \mathbf{T}^* , and hence the overall costs of the new iterative schemes are significantly smaller. Figures 7 and 8 depict the iteration history and the reconstructions for the case of 3 % noise.

In this example, when the noise gets small, e.g., for $\nu = 0.1\%$, the convergence history of Algorithm I becomes somewhat erratic if we stick to the small value $\rho = 10^{-3}$ chosen above. This can be detected from the residuals $\|\mathbf{r}_n\|$ as well as from the chosen regularization parameters α_n , as is illustrated in Figure 9: When the estimate (2.10) is violated, and the algorithm attempts to invert noise components, then the residual norms increase, and thereafter the regularization parameters and the residual norms start to zigzag up and down. Interesting enough, the algorithm often recovers, and in the particular case shown in Figure 9 the iteration terminates after 24 steps with a reasonable restoration (RRE = 0.254).

This instability can be fixed by increasing ρ . Figure 9 also displays the history of the residuals and regularization parameters for $\rho = 10^{-2}$. In this case the zigzagging disappears, and Algorithm I terminates after 17 iterations with RRE = 0.261. Note that, according to (2.11b) a different value of ρ only affects the final stage of the iteration; in this particular case the two sequences $(\alpha_n)_n$ start to differ from $n = 15$ onwards.

We mention that larger values of ρ also give rise to larger values of τ (by virtue of the definition of τ in Algorithm I), to satisfy our theory. In practice, however, the latter may result in a too early termination of the iteration according to (2.12). In that case it may pay to

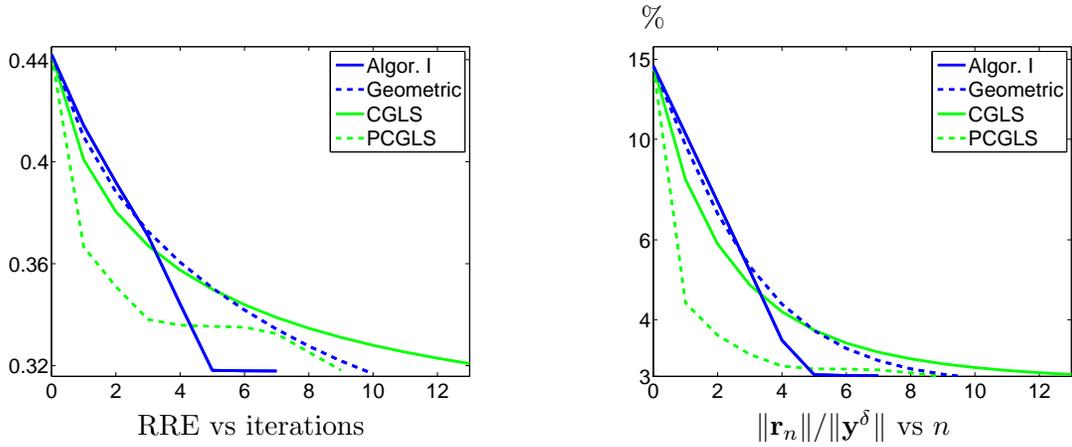


Figure 7: Example 2 – RRE and relative norm of the residual at each iteration ($\nu = 3\%$).

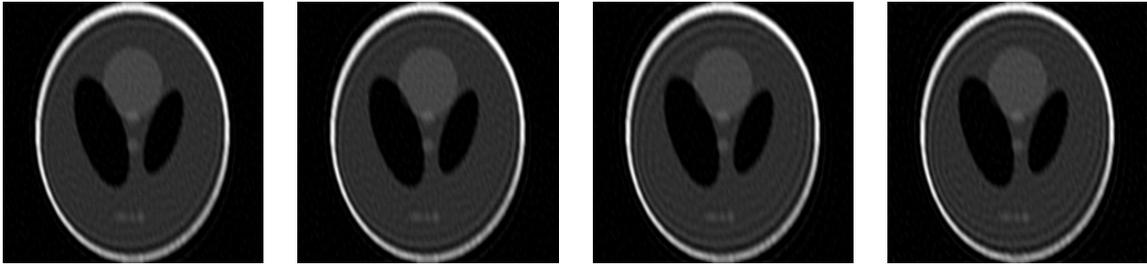


Figure 8: Example 2 – restored images for $\nu = 3\%$.
From left to right: Algorithm I, iteration (2.2) with geometric parameter sequence, CGLS, PCGLS.

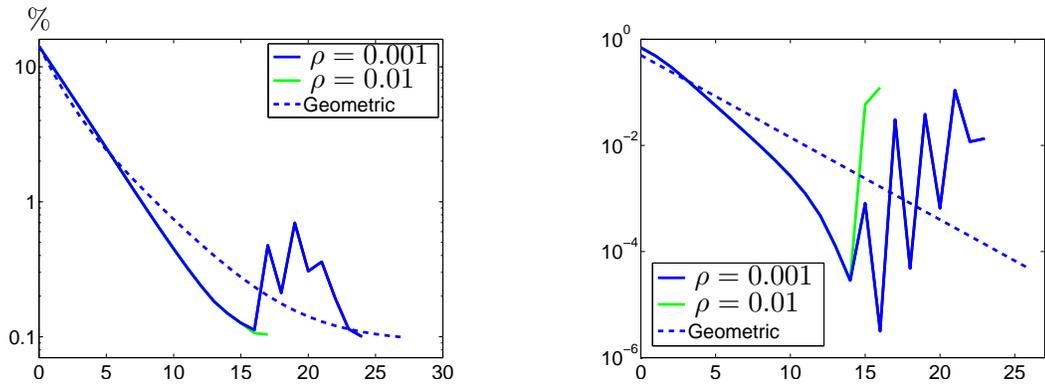


Figure 9: Example 2 with $\nu = 0.1\%$ – $\|r_n\|/\|y^\delta\|$ (left) and α_n (right) vs n .

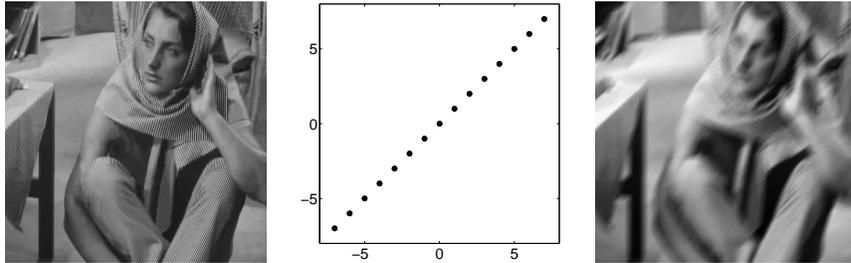


Figure 10: Example 3. From left to right: true image, sketch of the PSF, and blurred image without noise.

cautiously continue the iteration somewhat further while monitoring the residual norms. As mentioned before, an increase of the residual norm is often a reasonable indicator for some trouble going on.

5.3 Example 3

Our third test problem is the image “Barbara” of size 452×452 , corrupted by a 15 pixels diagonal motion blur and $\nu = 1\%$ noise, see Figure 10; the sketch of the PSF displays the locations of the corresponding nonzero pixel entries on the antidiagonal (like the “flip matrix” of order 15), the respective numbers being equal to $1/15$, each. Since the image is part of a larger scene we impose so-called antireflective boundary conditions as suggested by Serra-Capizzano [20]. The corresponding matrix \mathbf{T} has a Toeplitz plus Hankel plus low rank structure, both at the block level and within the blocks. Although antireflective boundary conditions yield a small modelling error, the normal equation system should be avoided for that approach, as the corresponding restorations in the range of \mathbf{T}^* come with unwanted boundary artefacts; compare [5, 7]. One possibility is to use a reblurring approach instead, cf. [6], in which case \mathbf{T}^* is replaced by a so-called reblurring matrix \mathbf{T}' , but then CGLS and PCGLS will fail, as $\mathbf{T}'\mathbf{T}$ is no longer symmetric. In fact, numerical results by Fan and Nagy [10] show that the reblurring conjugate gradient iteration with antireflective boundary conditions fails for this image “Barbara” and a diagonal motion blur, and they suggested so-called “synthetic” boundary conditions instead to fix this problem. Alternatively, one can stay with the simpler antireflective model and use the scheme (2.2) in lieu of the conjugate gradient iteration, because there the preconditioner is applied to the original system (1.1) and not to the normal equation system, and hence, neither the adjoint of \mathbf{T} nor any substitute are needed for the implementation.

We apply Algorithm I with parameter $\rho = 10^{-2}$ for the closeness estimate (2.8) to solve this problem, and compare the results to the nonstationary preconditioned iteration (2.2) with the geometric sequence $(\alpha_n)_n$, and with the reblurring CGLS variant. As expected, CGLS fails to converge and does not meet the stopping criterion, hence we terminate this method when its minimum RRE is reached (this, however, is not possible in practice); even with this beneficial selection the CGLS reconstruction is clearly worse than the ones of the nonstationary schemes, while the number of iterations is comparable. See Figure 11 for the iteration history, and Figure 12 for the corresponding restorations.

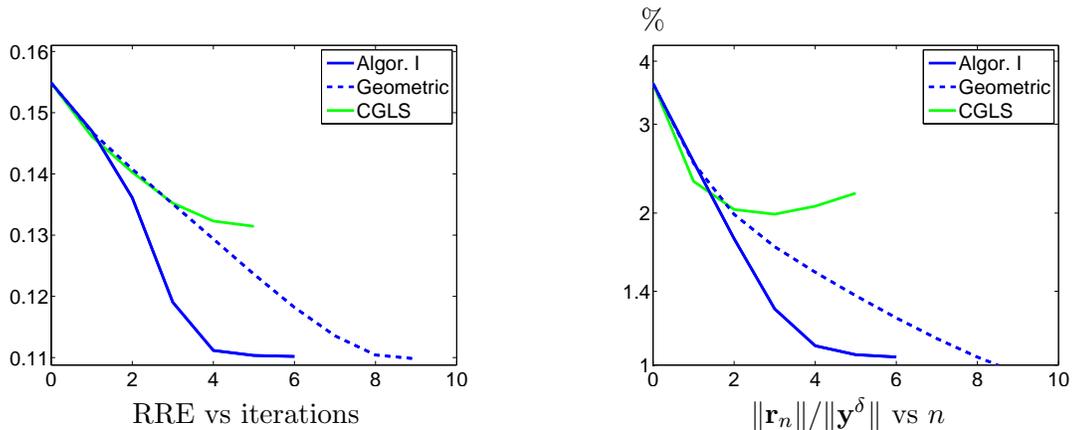


Figure 11: Example 3 – RRE and relative norm of the residual at each iteration.

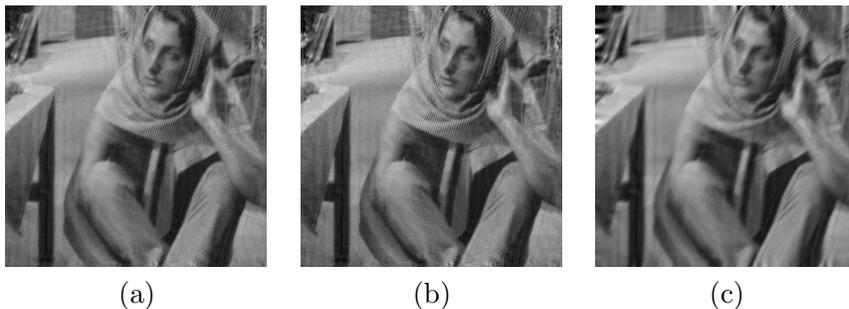


Figure 12: Example 3 – restored images.

- (a) Algorithm I: RRE = 0.110, it. 6;
- (b) Iteration (2.2) with geometric sequence: RRE = 0.110, it. 9;
- (c) CGLS: RRE = 0.131, it. 5;

6 Summary

We have developed a new nonstationary preconditioned iterative scheme for solving linear ill-posed problems which is motivated by the nonstationary iterated Tikhonov method. The new algorithm applies well to image restoration problems with arbitrary boundary conditions, using a preconditioner that operates in the Fourier domain.

The rapid convergence of the iteration is determined by the corresponding sequence of regularization parameters $(\alpha_n)_n$ of the intermediate Tikhonov problems. We propose an adaptive choice introduced earlier for a nonlinear Levenberg-Marquardt scheme, which can be justified theoretically under suitable assumptions, cf. (2.8). Alternatively, a geometric parameter sequence, $\alpha_n = \alpha q^n$, may also be suitable in practice, but for this choice we have no rigorous convergence analysis, and in our numerical experiments this variant exhibited a tendency to require more iterations to achieve the same accuracy.

Our numerical results show that the new method outperforms the standard conjugate gradient iteration of the normal equation, and also its preconditioned counterpart. At the same time

the method appears to be much more stable and also to be robust with respect to appropriate stopping rules.

Acknowledgement

This work has been accomplished while the second author was visiting the Università dell'Insubria; the hospitality he experienced while being in Como is gratefully acknowledged.

References

- [1] H.C. ANDREWS AND B.R. HUNT, *Digital Image Restoration*, Prentice Hall, Englewood Cliffs, NJ, 1977.
- [2] S. BERISHA AND J.G. NAGY, *Iterative methods for image restoration*, to appear as chapter published by Elsevier, in a volume edited by H. J. Trussell.
- [3] M. BERTERO AND P. BOCCACCI, *Introduction to Inverse Problems in Imaging*, Inst. of Physics Publ., Bristol, 1998.
- [4] P. BRIANZI, F. DI BENEDETTO, AND C. ESTATICO, *Improvement of space-invariant image deblurring by preconditioned Landweber iterations*, SIAM J. Sci. Comp., 30 (2008), 1430-1458.
- [5] M. CHRISTIANSEN AND M. HANKE, *Deblurring methods using antireflective boundary conditions*, SIAM J. Sci. Comput., 30 (2008), pp. 855–872.
- [6] 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.
- [7] M. DONATELLI AND S. SERRA-CAPIZZANO, *Anti-reflective boundary conditions and re-blurring*, Inverse Problems, 21 (2005), pp. 169–182.
- [8] H. EGGER AND A. NEUBAUER, *Preconditioning Landweber iteration in Hilbert scales*, Numer. Math. 101 (2005), pp. 643-662.
- [9] H.W. ENGL, M. HANKE, AND A. NEUBAUER, *Regularization Methods for Inverse Problems*, Kluwer, Dordrecht, 1996.
- [10] Y.W. FAN AND J.G. NAGY, *Synthetic boundary conditions for image deblurring*, Linear Algebra Appl., 434 (2011), pp. 2244–2268.
- [11] C.W. GROETSCH, *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*, Pitman, Boston, 1984.
- [12] M. HANKE, *A regularizing Levenberg-Marquardt scheme, with applications to inverse groundwater filtration problems*, Inverse Problems, 13 (1997), pp. 79–95.

- [13] M. HANKE AND C.W. GROETSCH, *Nonstationary iterated Tikhonov regularization*, J. Optim. Theory Appl., 98 (1998), pp. 37–53.
- [14] M. HANKE, J.G. NAGY, AND R.J. PLEMMONS, *Preconditioned iterative regularization for ill-posed problems*, in L. Reichel, A. Ruttan, and R.S. Varga, editors, *Numerical Linear Algebra*, de Gruyter, Berlin, 1993, pp. 141–163.
- [15] P.C. HANSEN, *Rank-Deficient and Discrete Ill-Posed Problems*, SIAM, Philadelphia, 1997.
- [16] M.E. KILMER, *Cauchy-like preconditioners for two-dimensional ill-posed problems*, SIAM J. Matrix Anal. Appl. 20 (1999), pp. 777–799.
- [17] A. KIRSCH, *An Introduction to the Mathematical Theory of Inverse Problems*, 2nd ed., Springer, New York, 2011.
- [18] J.G. NAGY, K. PALMER, AND L. PERRONE, *Iterative methods for image deblurring: A Matlab object oriented approach*, Numer. Algorithms, 36 (2004), pp. 73–93. See also: <http://www.mathcs.emory.edu/~nagy/RestoreTools>.
- [19] M.K. NG, *Iterative Methods for Toeplitz Systems*, Oxford University Press, Oxford, 2004.
- [20] S. SERRA-CAPIZZANO, *A note on anti-reflective boundary conditions and fast deblurring models*, SIAM J. Sci. Comput. 25 (2003), pp. 1307–1325.
- [21] G. STRANG, *A proposal for Toeplitz matrix calculations*, Stud. Appl. Math. 74 (1986), pp. 171–176.