# Iterative Regularization Techniques in Image Reconstruction

Martin Hanke

Fachbereich Mathematik, Johannes-Gutenberg-Universität Mainz, D–55099 Mainz, Germany.

**Abstract.** In this survey we review recent developments concerning the efficient iterative regularization of image reconstruction problems in atmospheric imaging. We present a number of preconditioners for the minimization of the corresponding Tikhonov functional, and discuss the alternative of terminating the iteration early, rather than adding a stabilizing term in the Tikhonov functional. The methods are examplified for a (synthetic) model problem.

## 1 Introduction

Atmospheric turbulences are the reason for severe problems in ground based astronomical imaging. On the passage through the atmosphere, light waves are scattered because of temperature fluctuations both in space and time, which lead to strong aberrations of astronomical images taken by a telescope on the surface of the Earth.

In principle, if a sophisticated model of the scattering process is available, the true image can be reconstructed from the photo by solving the associated inverse problem. Such models, however, are very difficult to derive, because atmospheric turbulences are hard to predict and can currently only be accessed via stochastical processes.

In this survey we shall focus on the inverse problem using a very simple model according to which the observed image, $y$, is a linear convolution of the true image (a nonnegative function $x$ of two variables) with a certain convolution kernel $k$,

$$y(\xi,\eta) \;=\; \int k(\xi-\xi',\eta-\eta')x(\xi',\eta')\,d(\xi',\eta')\,. \tag{1}$$

The function $k$ is known as *point spread function*: it is nonnegative and its $\mathcal{L}^1$-norm equals one; this refers to conservation of energy in the imaging process. The model (1) is quite appropriate for a long-time exposure of incoherent light waves. It is based on the assumption that the way a point source in the sky is mapped onto its image point and the neighbouring points on the photo is space-invariant. More sophisticated models also take space dependency into account. Long-time exposures usually lead to rather wide-spread point spread functions and thus to a significant loss of high-frequent and small detail information. As a consequence, imaging models for coherent light waves are currently under development in

order to deal with short-time exposures, cf., Roggeman and Welsh [21]. We shall not consider these models in the present survey.

Part of the modeling process is the selection of a realistic point spread function $k$ to be used in (1). Based on stochastic reasoning, simple Gaussian kernels were a common choice for $k$ in the early days of ground-based astronomical imaging, cf., e.g., Lagendijk and Biemond [17]. More recently, a method known as guide star imaging has become popular: This refers to a photo of a bright light source, which can be a known star or a so-called artificial beacon, i.e., the backscatter from a laser beam. According to the space-invariance of the imaging process a guide star image is essentially the convolution of $k$ with a delta distribution, and therefore provides an approximation of the values of $k$. Such a (simulated) guide star image is shown in Fig. 1.

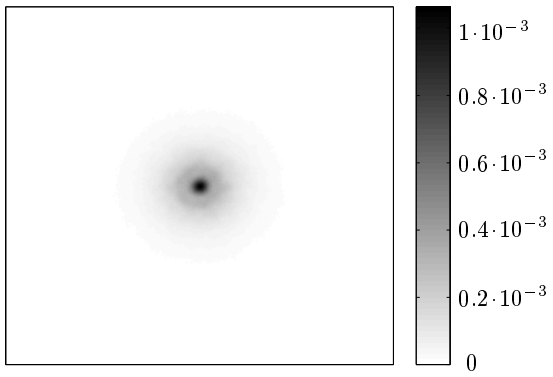



**Fig. 1.** Point spread function

Another option, which is currently under investigation, consists in reconstructing the point spread function *and* the original image simultaneously. This problem is known as *blind deconvolution* and is formally strongly underdetermined; to improve the setting a series of images can be taken within a short time interval, or additional physics and known a priori constraints can be incorporated to make the problem better determined.

## 2 Tikhonov Regularization

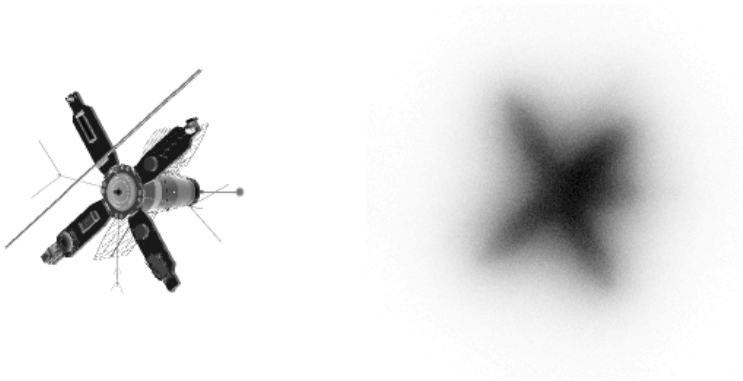
In the hardware (gray-scale) photos are encoded as two-dimensional arrays of pixel values, i.e., integers between 0 and 255 describing the darkness of the corresponding pixel. For numerical computations the integer assumption is usually dropped, and pixel values are allowed to take any real value, preferrably nonnegative reals. We denote by $\boldsymbol{x}$ and $\boldsymbol{y}$ the vector of all pixel values (in a row-wise

ordering) corresponding to the functions $x$ and $y$ of the continuous model (1); given $N \times N$ pixels for each image, the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ have dimension $N^2$.

Using the midpoint rule for the discretization of the convolution (1) we then end up with a finite dimensional linear system of equations,

$$T\boldsymbol{x} = \boldsymbol{y}\,. \tag{2}$$

The matrix $T$ is an $N \times N$ block matrix, with each block being itself an $N \times N$ matrix corresponding to one pair of pixel rows of the two images encoded in $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively. A careful inspection of the quadrature process reveals that the matrix $T$ has additional structure in that, first of all, each of its $N^2$ blocks is a Toeplitz matrix, i.e., its entries do not change along each individual diagonal, and second, the blocks on each block-diagonal of the entire matrix are all the same. We therefore call $T$ a block Toeplitz matrix with Toeplitz blocks (BTTB). We mention that for current images $N$ ranges from 256, say, up to 1024 and more; already for $N = 256$ this yields a dimension of $65536 \times 65536$ for matrix $T$. This is also the size of our numerical test problem which is used as example throughout this survey: This is a test problem from the Phillips Laboratory at Kirtland US Air Force Base, New Mexico (see [21]). The corresponding point spread function $k$ is the one from Fig. 1; the test image and its blurred photo are plotted in Fig. 2.



(a) *original image* (b) *blurred image*

**Fig. 2.** Model problem.

In many cases of interest, in particular for the aforementioned model problem the matrix $T$ is symmetric or close to a symmetric matrix. A symmetric matrix would correspond to a convolution kernel which is symmetric with respect to the origin, i.e.,

$$k(s,t) = k(-s,-t)\,,$$

an assumption which is likely to hold (at least approximately) if the light waves propagate in a normal direction to the surface of the Earth. Whenever appropri-

ate we will limitate our discussion to this symmetric case for simplicity, although similar considerations apply to the general case.

A symmetric BTTB matrix is completely determined by its first column (just as a self adjoint convolution operator (1) is determined by the values of the kernel function $k(\xi, \eta)$ for all nonnegative arguments $\xi$ and $\eta$). We denote the entries of $T$ by $t_{\nu,\mu}$ where the first index counts the block and the second one the index relative to the first entry of this block; it is convenient to start counting by zero so that $0 \le \nu, \mu < N$. In this case it follows for the midpoint quadrature rule that the $(\nu N + \mu, \nu' N + \mu')$ index of the symmetric BTTB matrix $T$ is given by

$$t_{|\nu-\nu'|,|\mu-\mu'|} \;=\; \Delta\, k\big((\nu-\nu')\Delta, (\mu-\mu')\Delta\big)\,,$$

where $\Delta$ is the mesh width.

To analyze the spectrum of $T$ the functions

$$f_N(\phi,\theta) \;=\; \sum_{1-N}^{N-1} t_{\nu,\mu} e^{i(\nu\phi+\mu\theta)}\,, \qquad -\pi \le \phi, \theta \le \pi\,, \tag{3}$$

play a prominent role. In general, $f_N$ is a smooth real-valued function which is essentially zero except for a neighborhood of the origin. Moreover, the distribution of the eigenvalues of $T$ is related to the distribution of the values of $f_N$, which implies that the spectrum of $T$ usually clusters at the origin. The function $f_N$ corresponding to the point spread function of Fig. 1 is shown as a logarithmic gray scale image in Fig. 3; it is obvious that $f_N$ is essentially zero for all angles $\phi$ and $\theta$ with $|\phi|, |\theta| > \pi/3$.

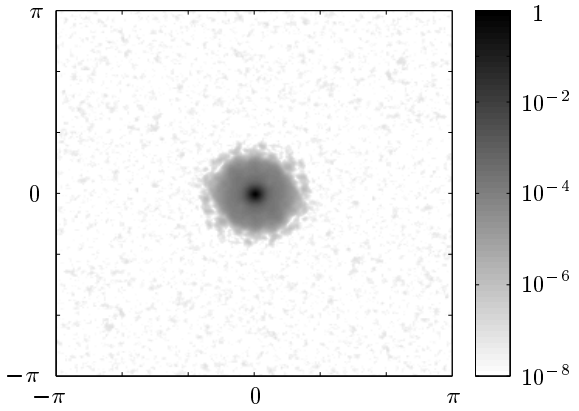



**Fig. 3.** A logarithmic plot of $f_N$.

As we have seen, the eigenvalues of $T$ cluster at the origin so that $T$ has a really large condition number in general. As a matter of fact, the solution $\boldsymbol{x}$ of the

linear system (2) is very sensitive to measurement errors in the right-hand side $\boldsymbol{y}$ resulting from the imaging process. To overcome this ill-conditioning regularization techniques have to be employed, among which Tikhonov regularization has outstanding popularity, cf. Groetsch [9]. In Tikhonov regularization, the goal is to minimize $\|\boldsymbol{y} - T\boldsymbol{x}\|_2$ subject to a constraint on the size or the smoothness of $x$, i.e., a bound for the norm $\|L\boldsymbol{x}\|_2$ for some given matrix $L$. This leads to the minimization problem

$$\|\boldsymbol{y} - T\boldsymbol{x}\|_2^2 \; + \; \alpha \, \|L\boldsymbol{x}\|_2^2 \; \longrightarrow \; \min \, . \tag{4}$$

Here $\alpha$ is some positive parameter, the *regularization parameter*, and the matrix $L$ is often chosen to be the identity matrix, either for simplicity, or for the lack of more sophisticated alternatives. In image restoration, penalty terms $\|L\boldsymbol{x}\|_2^2$ approximating a total variation functional of $x$ have also received increasing interest recently, cf., e.g., Vogel and Oman [24]. The regularization parameter $\alpha$ can be viewed as an a posteriori tuning parameter: theoretically, decreasing $\alpha$ should give higher resolution, but in practice rather leads to increasingly strong artefacts because of the influence of high-frequent noise in the data; with increasing $\alpha$ such artefacts can be reduced but then the details of the reconstructions get smeared. The optimal balance between the two extremes is a very delicate issue, and the costs for tuning the parameter are so high (usually, the code is restarted for every new value $\alpha$) that in practice $\alpha$ is chosen a priori, on the basis of preliminary experiments and experience.

Since the above Tikhonov functional is quadratic, the minimization process is equivalent to solving a linear system, namely

$$(T^*T + \alpha L^*L)\boldsymbol{x} \; = \; T^*\boldsymbol{y} \, . \tag{5}$$

Because of the tremendous size of matrix $T$ this system cannot be solved by direct methods. In the engineering literature it is therefore commonly recommended to replace the doubly Toeplitz matrix $T$ by a doubly circulant one, i.e., a block circulant matrix $S$ with circulant blocks (BCCB), which coincides with $T$ in all central block diagonals and the central diagonals of all blocks. In the important case $L = I$ the resulting linear system

$$(S^*S + \alpha I)\tilde{\boldsymbol{x}} = S^*\boldsymbol{y} \tag{6}$$

has again a BCCB coefficient matrix and can explicitly be solved with only two 2D-FFTs. Moreover, if the original image $x$ is only nonzero in its inner quarter and the point spread function is sufficiently narrow, then the reconstructions $\boldsymbol{x}$ and $\tilde{\boldsymbol{x}}$ coincide. This assumption is essentially satisfied for the satellite image in Fig. 2 (a) because there is a zero boundary layer of roughly 64 pixels width around the satellite. As a consequence, the reconstruction $\tilde{\boldsymbol{x}}$ obtained from (6) is pretty good for this particular model problem. We refer to Figs. 8 (b) and (c) for the two reconstructions obtained from (5) and (6)$^\star$.

$^\star$ The reconstruction for (5) corresponds to the case $L = I$. The optimal regularization parameters have been determined to be $\alpha = 2.3 \cdot 10^{-4}$ and $\alpha = 2.2 \cdot 10^{-4}$ for problems (5) and (6), respectively.

The approximation $T \approx S$ can also be interpreted in terms of the function $f_N$ of (3). In fact, the eigenvalues of a BCCB matrix are given by a 2D-FFT of its first column, and the eigenvectors are the two-dimensional Fourier vectors. By construction of $S$ the eigenvalues $\lambda_{\nu,\mu}$ of $S$ corresponding to the trigonometric monomials

$$p_{\nu,\mu}(\phi,\theta) \;=\; e^{i(\nu\phi+\mu\theta)}\,, \qquad -N/2 < \nu,\mu \le N/2\,,$$

are the values of $f_{N/2}$ at equidistant mesh points,

$$\lambda_{\nu,\mu} \;=\; f_{N/2}(2\nu\pi/N, 2\mu\pi/N)\,, \qquad -N/2 < \nu,\mu \le N/2\,.$$

In particular, high frequent monomials correspond to eigenvalues close to $f_{N/2}(\pm\pi,\pm\pi)$ and are therefore close to zero, cf. Fig. 3.

## 3 Iterative Minimization of the Tikhonov Functional

### 3.1 The Conjugate Gradient Iteration

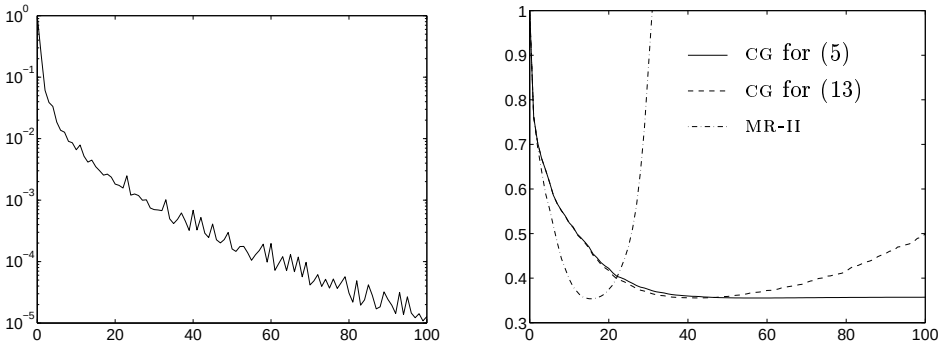
Aside of using the BCCB approximation $S$ instead of $T$, one can use iterative methods rather than direct methods to minimize (4), most notably by applying the well-known *conjugate gradient method* (CG) to (5). The CG-method is certainly one of the most efficient methods to solve large-scale linear systems of equations with a positive definite coefficient matrix (cf.,e.g., Golub and Van Loan [7]). What is important to mention is that although BTTB matrices cannot be inverted by use of FFTs, a matrix-vector multiplication with $T$ can be implemented using FFTs by imbedding $T$ into a BCCB matrix of four times the size of $T$. This well-known fact leads to an operation count of roughly 16 2D-FFTs (of length $N \times N$) per iteration of the conjugate gradient method.

It follows that the CG iteration for solving (5) will be a competitive algorithm provided that the number of required iterations is small. Unfortunately, the rate of convergence of the CG iteration depends significantly on the condition number of the linear system (5), which is still pretty large despite the regularizing term: To illustrate this fact we fix without loss of generality $\|T\|_2 = 1$ and $\|\boldsymbol{x}\|_2 = 1$, in which case it is also reasonable to expect $\|\boldsymbol{y}\|_2 \approx 1$; given a small norm $\delta \ll 1$ for the data error in $\boldsymbol{y}$ we quote from [5] that a reasonable regularization parameter $\alpha$ will typically be of the order of $\delta^2$ but not less – although it might be as large as $O(\delta^{2/3})$ in very special circumstances. From this we conclude the bound

$$\mathrm{cond}(T^*T + \alpha I) \;\le\; \frac{1+\alpha}{\alpha} \;=\; O(\delta^{-2})\,,$$

which turns out pretty sharp in practice. Since the spectrum of $T^*T$ is fairly densely distributed in the interval $[0,1]$ there are no or little additional spectral properties that the CG-method can use with advantage. As a consequence the CG iteration will only be moderately fast in general, requiring about a hundred

iterations or so to converge**. This is illustrated in Fig. 4 which shows the iteration history of CG applied to (5) with $L = I$ and with the optimal regularization parameter $\alpha = 2.3 \cdot 10^{-4}$: the left-hand plot (a) shows the norm of the relative residual of (5) versus the iteration count; the solid line in the right-hand side plot (b) corresponds to the relative errors of the reconstructions as compared to the true image; the dashed and dash-dotted lines in plot (b) will be referred to in Sect. 4.




(a) *relative residual norms* (b) *relative error norms*

**Fig. 4.** Convergence history: CG applied to problem (5) (solod lines).

### 3.2 Circulant Preconditioners

On the other hand there are good news in that several preconditioners are meanwhile available for our problem. In fact, whereas the solution of the BCCB system (6) may often not be an acceptable reconstruction in itself, the matrix $S$ provides a useful approximation of $T$ which can be used for preconditioning the system (5). It has been shown by R. Chan and others (see [2] for a survey of these results) that a variety of BCCB matrices $C$ as well as some other structured matrices approximate $T$ in that

$$C - T = R + E\,, \tag{7}$$

where $R$ is a matrix of small rank and $E$ is a matrix of small norm (depending on the rank of $R$). From this follows, cf. Chan, Nagy, and Plemmons [3], that the preconditioned matrix

$$(C^*C + \alpha I)^{-1}(T^*T + \alpha I) \tag{8}$$

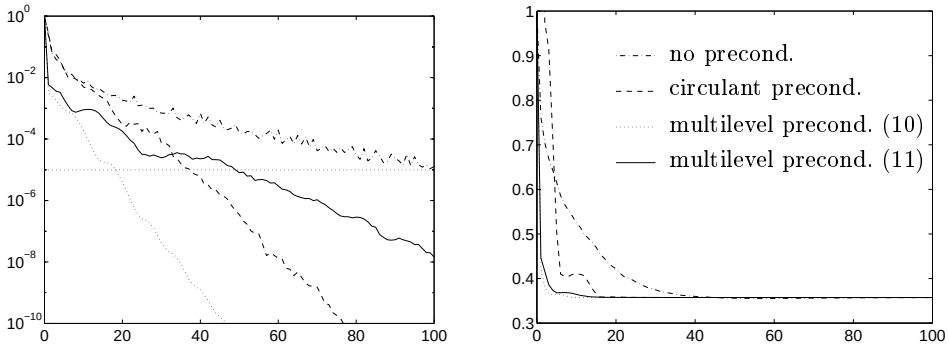
** It must be mentioned, though, that high accuracy of the solution of (5) is typically not an important requirement because the optimal regularization parameter $\alpha$ is not exactly known anyway.

contains relatively few eigenvalues which do not cluster around $\lambda = 1$. This is a situation which is ideally suited for the conjugate gradient method, and which often reduces the number of CG iterations by one order of magnitude.

For our test problem we have used as $C$ in (8) the BCCB matrix which is closest to $T$ in the Frobenius norm (called Level-2 preconditioner in [3]), i.e.,

$$\|C - T\|_F \longrightarrow \min_{\text{BCCB } C};$$

the results (see the dashed line in Fig. 5) are somewhat better than using $C = S$ in (8). We observe, cf. Fig. 5 (b), that the preconditioning deteriorates the iteration in the first few steps before (a much more rapid) convergence to the Tikhonov solution sets in; in fact, the error almost doubles in the first iteration before it drops down to the limit at iteration twenty. With preconditioning the relative residual norm reaches the level of $10^{-5}$ after roughly fourty iterations which corresponds to slightly less than about half the iterations as compared to the unpreconditioned scheme. It should be mentioned that the implementation of this preconditioner requires only two 2D-FFTs, i.e., less than 15% extrawork per iteration, so that the preconditioned CG iteration is really significantly faster than the preconditioned one.




(a) *relative residual norms* (b) *relative error norms*

**Fig. 5.** Convergence history of CG for (5) with various preconditioners.

### 3.3 Multilevel Preconditioners

Another preconditioner has recently been suggested in [23]. It is based on certain multigrid ideas first proposed by Hackbusch [10] for general integral equations of the second kind. For this preconditioner we introduce a second (coarse) grid of $16 \times 16$ pixels with each pixel representing itself $16 \times 16$ pixels from the original fine grid; we denote by $h$ and $H = 16h$ the mesh-widths of the fine and coarse grid, respectively. We require the usual intergrid transfer operations, i.e.,

orthogonal restriction ($I_h^H$) and associated prolongation operators ($I_H^h = I_h^{H*}$) such that $I_H^h I_h^H$ is the orthogonal projector onto the coarse grid functions. The corresponding projector onto the orthogonal complement is denoted by $Q = I - I_H^h I_h^H$. We use the coarse grid piecewise constant basis functions and the Haar tensor-wavelet basis for the orthogonal complement on the fine grid to represent $A = T^*T + \alpha I$ as a two by two block matrix

$$A \sim \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

where, for example, $A_{11} = I_h^H A I_H^h$ is a representation for the restriction $A_H$ of $A$ to the coarse grid (more details are provided in [23]).

After approximating $A_{22} \approx \alpha I$ the preconditioner is defined as the symmetric Gauß-Seidel preconditioner with respect to this block representation/approximation of $A$. With $M = -\alpha^{-1/2} A_{11}^{-1} A_{12}$ the preconditioned matrix $WA$ has the block representation

$$WA \sim \begin{bmatrix} A_{11}^{-1/2} & M \\ 0 & \alpha^{-1/2} I \end{bmatrix} \begin{bmatrix} A_{11}^{-1/2} & 0 \\ M^* & \alpha^{-1/2} I \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}.$$

After some further manipulations we obtain a representation $W$ of the preconditioner in the original fine grid basis, namely

$$W = A_H^\dagger + \frac{1}{\alpha} (A_H^\dagger T^*T - I) Q (T^*T A_H^\dagger - I).$$

Here, $A_H^\dagger$ is defined to be the generalized inverse of the coarse grid blurring matrix, i.e.,

$$A_H^\dagger = I_H^h A_{11}^{-1} I_h^H = I_H^h \big(I_h^H (T^*T + \alpha I) I_H^h\big)^{-1} I_h^H. \tag{9}$$

Even for the $16 \times 16$ dimensional coarse grid the (exact) construction of $A_{11} \in \mathbb{R}^{N \times N}$ is prohibitive because this would require the amount of work of approximately $N = 256$ iterations of the unpreconditioned CG iteration. Instead, Riley and Vogel [20] approximate $A_{11}$ by a BTTB matrix $B_{11}$ of size $256 \times 256$ (the blocks have size $16 \times 16$) with a kernel function obtained from a restriction of the original kernel $k$ to the coarse grid. The Cholesky factor of $B_{11}$ can be computed with approximately $2/3\, N^3$ operations prior to the iteration so that matrix vector multiplications with

$$B_H^\dagger = I_H^h B_{11}^{-1} I_h^H$$

using this Cholesky factorization are negligible extrawork. We denote the corresponding preconditioner by

$$W_1 = B_H^\dagger + \frac{1}{\alpha} (B_H^\dagger T^*T - I) Q (T^*T B_H^\dagger - I). \tag{10}$$

As shown in [20] the preconditioner $W_1$ outperforms the BCCB preconditioners for image restoration problems in terms of iteration counts. However, the

preconditioner $W_1$ as it stands requires two further multiplications with $T^*T$ in each iteration, so that the total amount of work per iteration is *tripled* as compared to the unpreconditioned or the circulant preconditioned iteration. We therefore proceed with a second level of approximation in that we replace $T$ in (10) once again by its optimal BCCB approximation $C$; by this we finally end up with the preconditioner

$$W_2 \;=\; B_H^\dagger \;+\; \frac{1}{\alpha}\,(B_H^\dagger C^*C - I)\,Q\,(C^*CB_H^\dagger - I)\,. \tag{11}$$

The implementation of $W_2$ takes four 2D-FFTs and four triangular solves with the Cholesky factors of $B_{11}$. It is therefore only marginally more expensive than the implementation of a circulant preconditioner.

The solid and dotted lines in Fig. 5 illustrate the performance of the two multilevel preconditioners $W_1$ and $W_2$ for our model problem. In agreement with [20] we observe that $W_1$ is most effective in speeding up the asymptotic rate of convergence; on the other hand, the relatively high costs per iteration almost compensate for this advantage. The approximation $W_2$ of (11) is inferior as far as the asymptotic rate of convergence is concerned, but the initial error history is almost the same with $W_1$ and $W_2$ as preconditioner. Because of its low cost, $W_2$ therefore seems to be the most efficient preconditioner for this particular problem.

In summary, there are quite a few alternative preconditioners for the Tikhonov regularized problem (5). With any of these it should be possible to minimize the Tikhonov functional in a fairly small number of iterations, i.e., in only $O(N^2 \log N)$ operations. But still, even under the assumption that it is possible to solve the regularized problem in only $O(N^2 \log N)$ operations, the aforementioned open problem of choosing an appropriate regularization parameter $\alpha$ remains a crucial issue to deal with. Some attempts to solve this problem can be found in the literature (cf., e.g., [1, 6, 8, 15, 18]) but more research is still necessary.

## 4 Regularization by Iteration

As an alternative we advocate the possibility of *regularizing by iteration*. Historically, this technique originated with the Landweber iteration, which can be viewed as a variant of the steepest-descent method applied to the least-squares functional $\|\boldsymbol{y} - T\boldsymbol{x}\|_2^2$. An intuitive understanding of its performance can be summarized as follows: As long as the current approximation $\boldsymbol{x}_k$ is not too close to the true solution, the residual $\boldsymbol{y} - T\boldsymbol{x}_k$ of the linear system is quite large and the data error in the right-hand side is negligible as compared to the size of the residual. It follows that the negative gradient of the least squares functional, i.e., $T^*(\boldsymbol{y} - T\boldsymbol{x}_k)$, essentially points to the right descent direction. As the residual shrinks, the data error component becomes increasingly important before it eventually dominates the objective function. This is the turning point at which the iterates start to lose orientation and eventually diverge, a phenomenon entitled *semiconvergence* in the literature.

It should be clear from this introduction that the regularization parameter (the tuning parameter) of the Landweber iteration is the iteration index, which should be chosen in such a way that the the residual and the errors in the data are essentially of the same size; a very nice proof of this conclusion was given by Defrise and de Mol [4]. The choice of an optimal regularization parameter therefore amounts to choosing an appropriate stopping rule, and the stopping rule that we have outlined above is the so-called *discrepancy principle*: Find the smallest nonnegative integer $k$ for which

$$\|\boldsymbol{y} - T\boldsymbol{x}_k\|_2 \approx \delta\,. \tag{12}$$

Instead of the Landweber iteration one can again apply the CG-method to the normal equation system

$$T^*T\boldsymbol{x} = T^*\boldsymbol{y}\,. \tag{13}$$

When $T$ is symmetric (not necessarily positive definite) then it is also possible to apply a variant of the CG method to the minimization of $\|\boldsymbol{y} - T\boldsymbol{x}\|_2$ without forming the normal equation system. This latter scheme, called MR-II in [11], has been applied successfully in [12] to our atmospheric imaging model problem after a suitable approximation of the point spread function in Fig. 1 by a symmetric one.

Although much more difficult to establish rigorously, the results for the Landweber iteration extend to these much faster iteration schemes, cf. [5, 11]: It turns out that one can "naively" apply the CG or MR-II iteration, i.e., without any prior regularization as in (5), provided that the iteration is ultimately stopped as soon as the residual first satisfies the discrepancy principle (12).

Interestingly, for the CG method it is *not* the condition number (which is huge) but only the noise level that determines the number of iterations to satisfy this stopping criterion. In fact, as shown by Plato [19] the total number $k$ of CG iterations to meet (12) is always bounded by

$$k = O(\delta^{-1})\,, \tag{14}$$

and this bound is the sharper the more dense the spectrum of $T^*T$ is. For symmetric and positive definite matrices $T$ the MR-II iteration requires at most $O(\delta^{-1/2})$ iterations.

The iteration histories for (plain) CG and MR-II when applied to our model problem are included in Fig. 4 (b): the dashed line refers to CG, the dash-dotted line shows the performance of MR-II. Note that one iteration of MR-II is only half as expensive as one CG iteration because MR-II does not refer to the normal equation system. From this follows that the sixteen MR-II iterations to find the optimal reconstruction of $\boldsymbol{x}$ require less than a sixth of the total work for the fifty or so iterations of the conjugate gradient method applied to (5) to obtain the same accuracy. Furthermore, the MR-II iteration requires no a posteriori selection of a suitable regularization parameter $\alpha$. As can be seen from Figs. 8 (b) and (d) the reconstructions with and without prior regularization are very similar.

As compared to (14), there are significantly smaller bounds for the number of iterations to achieve (12) when the eigenvalues of $T^*T$ decay rapidly to

zero. One of these improved bounds is stated in [5, Theorem 7.15] and concerns the case when the spectrum of the normal equation operator clusters around $\lambda = 1$ *and* $\lambda = 0$, except for only a few eigenvalues outside these two clusters. Then, as shown in [5], it is this number of outlying eigenvalues which essentially determines the number of CG iterations. This result makes use of two major assumptions:

- The first assumption relates the size of the eigenvalue cluster at the origin to the noise level in the data. The less noise there is, the smaller this cluster should be. In fact, during the early stage of the CG iteration this eigenvalue cluster is pretty much ignored – which is good because the corresponding eigenvector components of the right-hand side can be considered hidden by noise. We therefore call this span of eigenvectors the *noise subspace*. If, for some reason, this eigenvalue cluster is larger, then the CG iteration needs to pick up information from the corresponding eigenvectors. This slows down the convergence in a second stage of the iteration (we call this the *transient phase*) before the iteration eventually diverges.
- The other assumption is more significant: it requires that the true image is essentially a linear combination of the eigenvectors which do *not* belong to the noise subspace, i.e., the eigenvectors corresponding to the eigenvalues near $\lambda = 1$ and the eigenvalues which do not belong to either cluster. We call this the *signal subspace*.

It is instructive to interprete these remarks in the context of the approximating BCCB system

$$C\boldsymbol{x} = \boldsymbol{y}\,.$$

To this end, we plot in Fig. 6 (left) the eigenvalues of $C$ (smooth line) and the absolute values of the corresponding Fourier coefficients of $\boldsymbol{y}$ versus the eigenvalue count. The right-hand side plot is a zoom onto the interesting part above the noise level. Under the assumption that the noise level in the Fourier coefficients is about the same for all eigenvalues, this plot leads to the dashed line as a first guess on how to separate signal and noise subspace (more sophisticated algorithms for a separation of the two subspaces are given in [12]). Since, as we have seen in the previous section, the very small eigenvalues of $C$ correspond to high-frequent basis images it follows that this distinction between signal and noise agrees well with the intuition that low and high frequencies are associated with signal and noise, respectively.

## 5 Preconditioning

We stress that the introduction of signal and noise subspace (and their separation) is merely a means for understanding the different stages of the CG iteration. There is no need to explicitly determine appropriate subspaces in order to run the MR-II iteration. This gets different, though, if one is interested in speeding up the iteration by means of preconditioning.

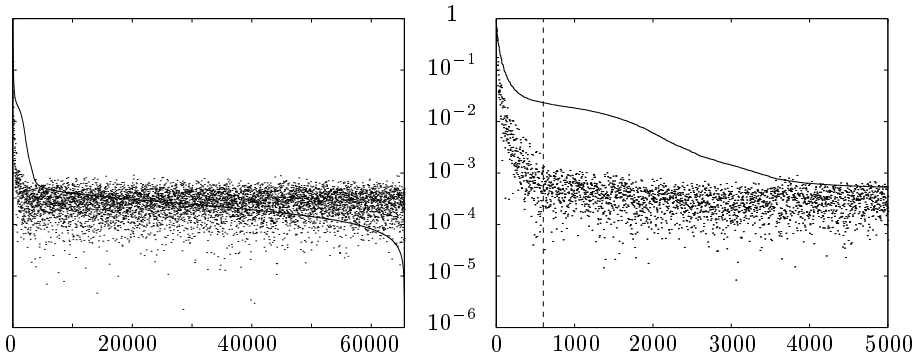



**Fig. 6.** Eigenvalues of $C$ and Fourier coefficients of $\boldsymbol{y}$.

To this end we turn back to the aforementioned theorem ([5, Theorem 7.15]) on the number of iterations of the conjugate gradient iteration. The two assumptions of this theorem are satisfied, for example, for the Gerchberg-Papoulis algorithm from signal processing, cf. [13]. This is a method to extend a given band-limited signal from its values in the time interval $[-T, T]$ to all times $t \in \mathbb{R}$. As was shown by Landau, Pollak, and Slepian in a sequel of papers in the early sixties (cf. [22] for a survey), the operator $T$ of the underlying linear equation $Tx = y$ has a spectrum which consists of an eigenvalue cluster at $\lambda = 1$ corresponding to low-frequent eigenfunctions (the signal subspace), and another cluster at $\lambda = 0$ with high-frequent eigenfunctions (the noise subspace). In image deblurring problems, on the other hand, such a clustering will hardly occur; in Fig. 6, for instance, one can see that the signal subspace corresponds to some hundred eigenvalues which are fairly well distributed in an interval $[0.02, 1]$.

A preconditioning of the original problem may change the situation, though. Limiting our discussion once again to symmetric problems and the MR-II iteration only, we refer by preconditioning to a transformation of the linear system $T\boldsymbol{x} = \boldsymbol{y}$ into

$$M^{1/2} T M^{1/2} \boldsymbol{z} = M^{1/2} \boldsymbol{y}\,, \qquad \boldsymbol{x} = M^{1/2} \boldsymbol{z}\,, \tag{15}$$

for some symmetric and positive definite matrix $M$. While in well-posed problems, the preconditioner is chosen so as to reduce the condition number of (15), the aim in ill-posed problems – according to the above discussion – is to cluster the eigenvalues corresponding to the signal subspace.

In other words, we are searching a matrix $M$ such that $M^{1/2} T M^{1/2}$ has an eigenvalue cluster at $\lambda = 1$ (and possibly another cluster at $\lambda = -1$ if $T$ is indefinite) in the signal subspace, and an eigenvalue cluster at the orgin in the noise subspace. Because of this we restrict our attention to matrices $M$ with

$$\begin{array}{ll} M \approx |T|^{-1} & \text{on the signal subspace, and} \\ M \approx I & \text{on the noise subspace.} \end{array}$$

Here, $|T| = (T^*T)^{1/2}$ is the modulus of $T$. Since the Level-2 BCCB approximation $C$ of $T$ is a useful approximation of $T$ throughout, cf. (7), and since the eigenstructure of $C$ is easily available and allows a plausible separation into signal and noise subspaces of 2D Fourier vectors, cf. Fig. 6, the choice

$$\begin{array}{ll} M = |C|^{-1} & \text{on the signal subspace, and} \\ M = I & \text{on the noise subspace} \end{array} \tag{16}$$

becomes a natural candidate for $M$. More precisely, if $P_S$ and $P_N$ are the orthogonal projectors onto the 2D-Fourier vectors corresponding to signal and noise subspaces, respectively, then we take

$$M \;=\; P_S|C|^{-1}P_S \;+\; P_N\,.$$

It was shown in [12] that the approximation property (7) implies that

$$M^{1/2}TM^{1/2} - P_{S+} + P_{S-} \;=\; E + F + R\,, \tag{17}$$

where $P_{S\pm}$ are the orthoprojectors onto signal vectors corresponding to positive and negative eigenvalues of $C$; furthermore, $E$ and $F$ are matrices of small norm, and $R$ is a matrix of small rank. (17) implies that the eigenvalues of $M^{1/2}TM^{1/2}$ cluster around $\lambda = \pm 1$ and at the origin, except for a few outliers.

In this approach the distinction between signal and noise subspace can be based on the magnitude of the eigenvalues $\lambda_{\nu,\mu}$ of $C$: the signal subspace consists of all eigenvectors of $C$ corresponding to eigenvalues $|\lambda_{\nu,\mu}| > \tau$ (low frequencies, cf. Fig. 3); the noise subspace is the complementary space.

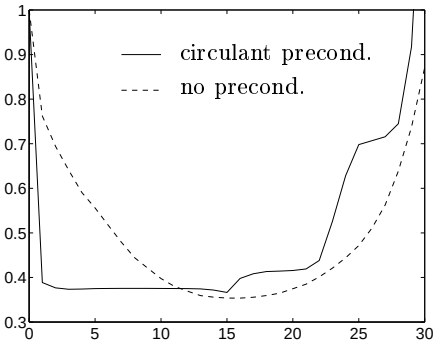



**Fig. 7.** Convergence history of MR-II with/without preconditioner.

For the performance of this preconditioner (with $\tau = 0.0137$) we refer to Fig. 7. Note that the error of the preconditioned MR-II iteration drops rapidly (in three iterations) down to a level which is only marginally above the best possible error attained in iteration fivteen. Still, this reconstruction (shown in Fig. 8 (e)) is slightly worse than what is obtained in sixteen iterations without

preconditioner. The reason is that it is not possible to *completely* separate noise and signal in the preconditioning process: Some information in the right-hand side must have been destroyed by the preconditioner.

Better results might be achieved with a different choice of the tolerance parameter $\tau$ that is used to separate noise and signal subspace. In this example $\tau$ has been selected with the L-curve criterion as described in detail in [12]. This can be considered as a black box criterion for constructing the preconditioner and, as every black box criterion, might not be the most sophisticated.

The above choice for $M$ is not the only possible, although there are certainly not many strategies for preconditioning $T$ such that all aforementioned requirements on $M$ are satisfied. For example, in [12] the preconditioner $M$ was chosen as a modification of the BCCB extension $S$ of $T$ and not as a modification of the optimal BCCB approximation $C$. Meanwhile, other useful choices for $M$ have been suggested, cf. [13, 16].

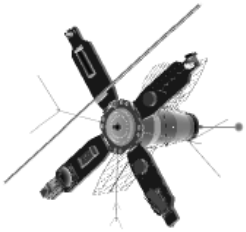
## 6 Summary

We have outlined a number of iteration methods to regularize 2D convolution problems. Our presentation is focusing on a particular model problem arising in astronomical imaging. Those images are affected by the scattering of light due to atmospheric turbulences on their passage to the surface of the Earth.

The methods that we have presented differ in the way regularization is incorporated into the scheme. One option is to first regularize (e.g., by using Tikhonov regularization) before starting the conjugate gradient method to compute the regularized approximation; another option consists in using CG for the 'plain' original problem, with early termination of the iteration to incorporate a different type of regularization. Whereas the former approach requires an a posteriori selection of the regularization parameter (and thus, in principle, multiple restarts), the latter approach appeals because the optimal iterate can be selected interactively, e.g., by monitoring the iterates on the screen.

Either iterative scheme can be accelerated by appropriate preconditioning, but this is more subtle in the non-regularized case. A couple of preconditioners and their performance on a model problem have been investigated above. For a comparison of the various algorithms we refer to Fig. 8 for the corresponding reconstructions of the satellite image. In these images negative pixel values have been set to zero, and the gray levels in all images are the same as in the original image (top left); error numbers refer to the Euclidean norm, which – as is well known – may not to be the optimal measure for a comparison of images.

We remark that for the regularized problem (5) the preconditioner (if any) does not affect the final reconstruction (Fig. 8 b) but only the costs for its computation; reconstruction (c) is the corresponding solution of problem (6) after replacing the BTTB matrix $T$ by its BCCB approximation $S$. For the iterative schemes based on the plain, unregularized equation (2) the situation is somewhat different in that the quality of the reconstruction *does* depend on the preconditioner. This can be seen by comparing the reconstructions (e) and (d) ob-



(a) *original image*

(b) *Tikhonov reconstruction*
*error:* 0.3575

(c) *circulant approximation*
*error:* 0.3549

(d) *plain* MR-II ($16^{th}$ *iterate*)
*error:* 0.3536

(e) *prec.* MR-II ($3^{rd}$ *iterate*)
*error:* 0.3736

(f) *nonneg. Tikhonov reconstruction*
*error:* 0.3340

**Fig. 8.** Various reconstructions.

tained with MR-II, with and without preconditioner: The preconditioner reduces the computational costs, but the reconstruction is slightly inferior. It should be emphasized that those two reconstructions use a modified, symmetrized point spread function because the original one from Fig. 1 is not symmetric. Nonetheless, the reconstruction obtained after 44 CG iterations using the normal equation system (13) without preconditioner and with the original point spread function (compare the dashed line in Fig. 4 (b)) is not much different from the image shown in Fig. 8 (d).

It should be noted that the reconstructions in (b) and (d), and to some minor extent also reconstruction (e) reveal distinct artefacts near the edges of the image. Similar artefacts are missing for the circulant reconstruction (c) because the true image is zero near its boundary; if this were not the case the ciculant reconstruction would probably suffer much stronger under artificial boundary artefacts.

We finally mention that in astronomical imaging it is fairly common to incorporate known a priori constraints on the reconstruction, e.g., nonnegativity. In the reconstructions (b) through (e) the nonnegativity constraint only enters via the plot routine where all negative entries are set to zero. A more sophisticated use of this constraint is very difficult without losing the advantages of the CG iteration.

Still, to illustrate its (potential) advantage we also computed the solution of the nonnegatively constrained Tikhonov minimization problem (4), i.e.,

$$\text{minimize} \quad \|\boldsymbol{y} - T\boldsymbol{x}\|_2^2 + \alpha \, \|\boldsymbol{x}\|_2^2 \qquad \text{subject to } \boldsymbol{x} \geq 0 \,.$$

The corresponding image is shown in Fig. 8 (f); it differs distinctively from the other reconstructions. While the other reconstructions suffer from speckles in those areas where they should be identically zero, the constrained reconstruction (f) does not exhibit such artefacts. This may also be the reason why the Euclidean error of the constrained reconstruction is somewhat smaller than the errors of the other reconstructions although it is not necessarily better from a visual point of view: for example, the details of the antenna of the satellite are more smeared than in the unconstrained reconstructions.

## References


1. Å. Björck, *A bidiagonalization algorithm for solving large and sparse ill-posed systems of linear equations*, BIT, 28 (1988), pp. 659-670.
2. R. Chan and M. K. Ng, *Conjugate gradient methods for Toeplitz systems*, SIAM Rev., 38 (1996), pp. 427-482.
3. R. H. Chan, J. G. Nagy and R. J. Plemmons, *FFT-based preconditioners for Toeplitz-block least squares problems*, SIAM J. Numer. Anal., 30 (1993), pp. 1740–1768.
4. M. Defrise and C. De Mol, *A note on stopping rules for iterative regularization methods and filtered SVD*, in: P. C. Sabatier, ed., Inverse Problems: An Interdisciplinary Study, Academic Press, London, Orlando, San Diego, New York, 1987, 261–268.

5. H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996.
6. A. Frommer and P. Maass, *Fast CG-based methods for Tikhonov regularization*, SIAM J. Sci. Comput. (1999), to appear.
7. G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd. ed., John Hopkins Univ. Press, Baltimore, 1996.
8. G. H. Golub and U. von Matt, *Generalized Cross-Validation for large-scale problems*, J. Comput. Graph. Statist., 6 (1997), pp. 1-34.
9. C. W. Groetsch, *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*, Pitman, Boston, 1984.
10. W. Hackbusch, *Integral Equations. Theory and Numerical Treatment*, Birkhäuser Verlag, Basel, 1995.
11. M. Hanke, *Conjugate Gradient Type Methods for Ill-Posed Problems*, Longman Scientific & Technical, Harlow, Essex, 1995.
12. M. Hanke and J. G. Nagy, *Restoration of atmospherically blurred images by symmetric indefinite conjugate gradient techniques*, Inverse Problems, 12 (1996), pp. 157-173.
13. ———, *Inverse Toeplitz preconditioners for ill-posed problems*, Linear Algebra Appl., 284 (1998), pp. 137-156.
14. M. Hanke, J. G. Nagy, and R. J. Plemmons, *Preconditioned iterative regularization*, in *Numerical Linear Algebra*, L. Reichel, A. Ruttan, and R. S. Varga, eds., de Gruyter, Berlin, 1993, pp. 141-163.
15. L. Kaufman and A. Neumaier, *Regularization of ill-posed problems by envelope guided conjugate gradients*, J. Comput. Graph. Statist., 6 (1997), pp. 451-463.
16. M. E. Kilmer, *Cauchy-like preconditioners for 2-dimensional ill-posed problems*, manuscript, 1997.
17. R. L. Lagendijk and J. Biemond, *Iterative Identification and Restoration of Images*, Kluwer, Boston, 1991.
18. D. P. O'Leary and J. A. Simmons, *A bidiagonalization-regularization procedure for large scale discretizations of ill-posed problems*, SIAM J. Sci. Statist. Comput., 2 (1981), pp. 474-489.
19. R. Plato, *Die Effizienz des Diskrepanzverfahrens für Verfahren vom Typ der konjugierten Gradienten*, in: E. Schock, ed., Beiträge zur Angewandten Analysis und Informatik, Shaker Verlag, Aachen, 1994, pp. 288-297. In German.
20. K. Riley and C. R. Vogel, *Preconditioners for linear systems arising in image reconstruction*, in: F. T. Luk, ed., Advanced Signal Processing Algorithms, Architectures, and Implementations VIII, Proceedings of SPIE Vol. 3461 (1998), pp. 372-380.
21. M. C. Roggemann and B. Welsh, *Imaging Through Turbulence*, CRC Press, Boca Raton, Florida, 1996.
22. D. Slepian, *Some comments on Fourier analysis, uncertainty and modelling*, SIAM Rev., 25 (1985), pp. 379-393.
23. C. R. Vogel and M. Hanke, *Two-level preconditioners for regularized inverse problems II: Implementation and numerical results*, manuscript, 1998.
24. C. R. Vogel and M. E. Oman, *Fast, robust total variation-based reconstruction of noisy, blurred images*, IEEE Trans. Image Proc., 7 (1998), pp. 813-824.