Electronic Transactions on Numerical Analysis.
Volume 31, pp. 204-220, 2008.

ISSN 1068-9613.

ETNA
Kent State University
http://etna.math.kent.edu

NOISE PROPAGATION IN REGULARIZING ITERATIONS FOR IMAGE DEBLURRING*

PER CHRISTIAN HANSEN† AND TOKE KOLDBORG JENSEN‡

Abstract. We use the two-dimensional discrete cosine transform to study how the noise from the data enters the reconstructed images computed by regularizing iterations, that is, Krylov subspace methods applied to discrete ill-posed problems. The regularization in these methods is obtained via the projection onto the associated Krylov subspace. We focus on CGLS/LSQR, GMRES, and RRGMRES, as well as MINRES and MR-II in the symmetric case. Our analysis shows that the noise enters primarily in the form of band-pass filtered white noise, which appears as "freckles" in the reconstructions, and these artifacts are present in both the signal and the noise components of the solutions. We also show why GMRES and MINRES are not suited for image deblurring.

Key words. Image deblurring, regularizing iterations, Krylov subspaces, CGLS, LSQR, GMRES, MINRES, RRGMRES, MR-II.

AMS subject classifications. 65F22, 65F10.

1. Introduction.

Iterative solvers based on Krylov subspaces are important for image deblurring problems [21], and when combined with good preconditioners [7, 24] these methods are favorable alternatives to the classical FFT-based algorithms. The Matlab packages RESTORE TOOLS [23] and REGULARIZATION TOOLS [13] provide implementations of many of these methods.

In the setting of matrix computations, the model for the blurring of an image is $A\,x = b$, where the vectors x and b represent the sharp and blurred images, and the matrix A represents the blurring process. Since image deblurring is a discrete ill-posed problem, it is necessary to use regularization in order to compute stable solutions [12, 16]. Moreover, it is often advantageous to impose boundary conditions on the reconstruction, which is achieved by a simple modification of the coefficient matrix [16, 23, 25].

One of the main goals of any regularization method is to suppress, as much as possible, the noise in the reconstruction coming from noise in the data, while at the same time computing a good approximate solution. Hence, for a given regularization method it is important to understand its approximation properties as well as how it suppresses or filters the noise. In this paper we will perform a computational study of these properties, with an emphasis on how the noise from the data propagates to the reconstruction.

We focus on *regularizing iterations*, where we apply a Krylov subspace method to the un-regularized problem $\min \|A\,x - b\|_2$ or $A\,x = b$. The regularization comes from the restriction of the solution to the Krylov subspace associated with the method, and the number of iterations plays the role of the regularization parameter. By means of preconditioning techniques one can modify the Krylov subspace in such a way that a general smoothing norm $\|L\,x\|_2$ is incorporated; see [11, 12, 14].

Objective assessment of the perceived quality of images is a difficult task [27]. In this paper we use the two-dimensional discrete cosine transform (DCT) to perform a spectral analysis of the solutions to the image deblurring problem computed by means of regularizing

*Received November 28, 2007. Accepted August 18, 2008. Published online on March 5, 2008. Recommended by Zdeněk Strakoš. This work was carried out as part of the project CSI: Computational Science in Imaging, supported by grant no. 274-07-0065 from the Danish Agency for Science Technology and Innovation.

†Department of Informatics and Mathematical Modelling, Technical University of Denmark, Building 321, DK-2800 Lyngby, Denmark (pch@imm.dtu.dk).

‡Department of Management Engineering, Technical University of Denmark, Building 424, DK-2800 Lyngby, Denmark (tkj@imm.dtu.dk).

ETNA
Kent State University
http://etna.math.kent.edu

iterations, and we focus on CGLS/LSQR and GMRES, and the variants MINRES, RRGMRES, and MR-II. In particular, we are interested in how the filtered noise from the data enters the reconstruction. While error propagation studies have been carried out before (see, e.g., [3]), we are not aware of studies of the spectral properties of the reconstructions and the errors for regularizing iterations.

Our paper is organized as follows. Section 2 gives brief descriptions of the image deblurring problem and regularization in the SVD basis, and in Section 3 we study the spectral properties of the coefficient (blurring) matrix. Sections 4 and 5 contain the main analysis of the iterative methods via a careful study of the Krylov subspaces, as well as a splitting of the solutions into their signal and noise components. Finally, in Section 6 we explain why the perceived quality appears to be different for low noise levels. Two appendices describe the blurring used throughout the paper, and the characteristics of band-pass filtered white noise.

2. The image deblurring problem and the SVD basis. Underlying the image deblurring problem is a 2D Fredholm integral equation of the first kind, whose kernel is the *point spread function* (PSF) for the blurring; see Appendix A. For simplicity of our analysis and notation, we consider $n \times n$ images and PSFs that are spatially invariant and separate in the variables[1]. Discretization of the integral equation then leads to the model

$$A_{\mathrm{c}}\, X\, A_{\mathrm{r}}^T = B, \qquad B = B^{\mathrm{exact}} + E\,, \tag{2.1}$$

where X is the reconstructed image, B^{exact} is the blurred noise-free image, E represents the noise, and the two $n \times n$ Toeplitz matrices A_{c} and A_{r} represent blurring in the direction of the columns and rows of the image, respectively. Moreover, we assume that the elements of the noise matrix E are statistically independent, uncorrelated with X, and coming from a normal distribution with zero mean and standard deviation η.

By introducing the vectors $x = \mathrm{vec}(X)$ and $b = \mathrm{vec}(B)$, where $\mathrm{vec}(\cdot)$ stacks the columns of the matrix, we can rewrite the above system in the "usual" form $A\,x = b$, in which the *PSF matrix* A is the $n^2 \times n^2$ Kronecker product $A = A_{\mathrm{r}} \otimes A_{\mathrm{c}}$. The Kronecker form of the PSF matrix lets us compute the SVD of large matrices, due to the fact that given the SVDs of the two matrices A_{c} and A_{r},

$$A_{\mathrm{c}} = U_{\mathrm{c}}\, \Sigma_{\mathrm{c}}\, V_{\mathrm{c}}^T, \qquad A_{\mathrm{r}} = U_{\mathrm{r}}\, \Sigma_{\mathrm{r}}\, V_{\mathrm{r}}^T,$$

we can write the SVD of the PSF matrix $A = A_{\mathrm{r}} \otimes A_{\mathrm{c}}$ as

$$A = U\, \Sigma\, V^T = \big((U_{\mathrm{r}} \otimes U_{\mathrm{c}})\, \Pi\big)\big(\Pi^T (\Sigma_{\mathrm{r}} \otimes \Sigma_{\mathrm{c}})\, \Pi\big)\big((V_{\mathrm{r}} \otimes V_{\mathrm{c}})\, \Pi\big)^T. \tag{2.2}$$

The $n^2 \times n^2$ permutation matrix Π ensures that the diagonal elements of $\Pi^T (\Sigma_{\mathrm{r}} \otimes \Sigma_{\mathrm{c}})\, \Pi$ appear in non-increasing order. We emphasize that our analysis of the iterative methods is not restricted to Kronecker products, but it holds for all PSF matrices.

Many regularization methods, including regularizing CGLS iterations, lead to regularized solutions x_{reg} which take the form

$$x_{\mathrm{reg}} = \sum_{k=1}^{n^2} f_k\, \frac{u_k^T b}{\sigma_k}\, v_k. \tag{2.3}$$

For Tikhonov regularization we have $f_k = \sigma_k^2/(\sigma_k^2 + \lambda^2)$, while the filter factors are 0 or 1 for the truncated SVD (TSVD) method. These methods are sometimes referred to as spectral

[1] More general PSFs can be studied by the same approach, but the separable PSFs suffice to illustrate our points.

ETNA
Kent State University
http://etna.math.kent.edu

filtering methods; in this paper we avoid this term, in order not to confuse it with our DCT-based spectral analysis. With the convention $x_{\rm reg} = \mathrm{vec}(X_{\rm reg})$, equation (2.3) immediately leads to the expression

$$X_{\rm reg} = \sum_{k=1}^{n^2} f_k \, \frac{u_k^T b}{\sigma_k} \, V^{[k]}, \tag{2.4}$$

where f_k are the filter factors, σ_k are the singular values of A, u_k are the left singular vectors, and $V^{[k]}$ are $n \times n$ matrices such that $v_k = \mathrm{vec}(V^{[k]})$ are the right singular vectors. This relation shows that we can express the regularized solution $X_{\rm reg}$ as a weighted sum over the *basis images* $V^{[k]}$.

With the Kronecker-product form of the PSF matrix A, equation (2.2) shows that there are simple expressions for the singular values and vectors. If $\sigma_{{\rm r}i}$ and $\sigma_{{\rm c}j}$ are the singular values of $A_{\rm r}$ and $A_{\rm c}$, then their products are the singular values of A. Moreover, if $u_{{\rm r}i}$, $u_{{\rm c}j}$, $v_{{\rm r}i}$ and $v_{{\rm c}j}$ are the left and right singular vectors of $A_{\rm r}$ and $A_{\rm c}$, then the left and right singular vectors of A are $u_{{\rm c}j} \otimes u_{{\rm r}i}$ and $v_{{\rm c}j} \otimes v_{{\rm r}i}$, respectively. Then (2.4) takes the form

$$X_{\rm reg} = \sum_{i=1}^{n} \sum_{j=1}^{n} f_{ij} \, \frac{u_{{\rm r}i}^T B \, u_{{\rm c}j}}{\sigma_{{\rm r}i} \, \sigma_{{\rm c}j}} \, V^{[ij]}$$

where f_{ij} is the filter factor associated with the product $\sigma_{{\rm r}i} \, \sigma_{{\rm c}j}$, and the basis images are given by $V^{[ij]} = v_{{\rm r}i} \, v_{{\rm c}j}^T$.

For two-dimensional problems we often observe a slow decay of the singular values, and occasionally we have multiple singular values. For the Kronecker product case, the slow decay can be explained by the fact that the singular values of A are the products $\sigma_{{\rm r}i} \, \sigma_{{\rm c}j}$ of the singular values of $A_{\rm r}$ and $A_{\rm c}$. Even if we have a rather fast decay of the singular values of each of these matrices, their products decay much slower, but the matrix A is still very ill-conditioned, because $\mathrm{cond}(A) = \mathrm{cond}(A_{\rm c}) \cdot \mathrm{cond}(A_{\rm r})$. We also see that if $A_{\rm c} = A_{\rm r}$ then we must have many double singular values.

3. The DCT and spectral properties of the PSF matrix. The two-dimensional discrete cosine transform is a simple frequency transform that is often used in image processing. If X is an $n \times n$ image, then the transformed image is

$$\check{X} = \texttt{dct2}(X) = \mathcal{C} \, X \, \mathcal{C}^T,$$

where $\mathcal{C} \in \mathbb{R}^{n \times n}$ is an orthogonal matrix that represents the one-dimensional DCT[2] [26]. The elements of $\mathcal{C}$ are given by

$$\mathcal{C}_{ij} = \begin{cases} \sqrt{1/n} & i = 1 \\ \sqrt{2/n} \, \cos\big(\pi(i-1)(2j-1)/(2n)\big), & i > 1. \end{cases}$$

The DCT-transformed image $\check{X}$ provides a frequency representation of the image X, where each element $\check{X}_{ij}$ is the coefficient to a specific basis image; see, e.g., [17, p. 136]. The element $\check{X}_{11}$ represents a constant, and the elements $\check{X}_{1j}$ and $\check{X}_{i1}$ correspond to simple cosine waves of varying frequency in horizontal or vertical directions over the entire image. The remaining elements of $\check{X}$ represent combinations of frequencies in the two directions. The lowest spatial frequencies are represented in the top left corner, and the highest in the opposite corner; Figure 3.1 illustrates this.

[2]There are several types of DCT transforms. The one adopted here is the most commonly used, which is named DCT-2 in [26].

ETNA
Kent State University
http://etna.math.kent.edu

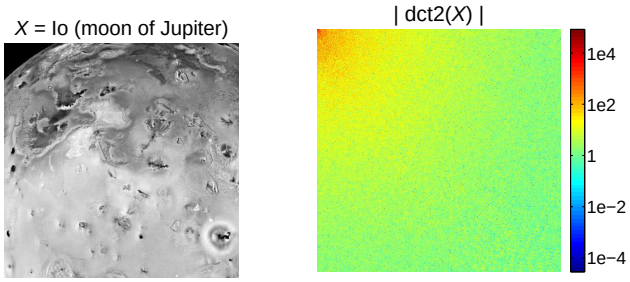

FIGURE 3.1. *The DCT transform shows that the image is dominated by low-frequency information.*

For discrete ill-posed problems arising from first-kind Fredholm integral equations, we know from the analysis in [15] that the singular vectors u_i and v_i of the coefficient matrix A tend to have an increasing number of sign changes in their elements as the index i increases. That is, the smaller the singular value σ_i, the more high-frequent the appearance of the corresponding singular vectors u_i and v_i.

For two-dimensional problems we expect a similar behavior, but the concept of frequency is more complicated because the singular vectors now correspond to two-dimensional basis images. The correct way to study the spectral behavior of the singular vectors is therefore to study the two-dimensional spectral behavior of the basis images $V^{[k]}$ or $V^{[ij]} = v_{\mathrm{r}i}\, v_{\mathrm{c}j}^T$, e.g., by means of the DCT. We need to sort the latter basis images according to decreasing singular values $\sigma_i \sigma_j$ using the permutation matrix Π from equation (2.2); the sorted basis images are then equal to $V^{[k]}$.

To illustrate this we construct two coefficient matrices A and $\overline{A}$ as Kronecker products of 32×32 Toeplitz matrices. Referring to Appendix A, the matrix A describes isotropic blurring with $s_{\mathrm{c}} = s_{\mathrm{r}} = 3$ and $\alpha = 0$, and $\overline{A}$ describes non-isotropic blurring with $s_{\mathrm{c}} = 3$, $s_{\mathrm{r}} = 5$, and $\alpha = 5$. The first four basis images $V^{[1]}$ to $V^{[4]}$ of A are shown in Figure 3.2, together with their DCTs $|\check{V}^{[1]}|$ to $|\check{V}^{[4]}|$. The rightmost picture in Figure 3.2 shows the "accumulated DCT" $\left(\sum_{i=1}^{150} |\check{V}^{[i]}|^2\right)^{1/2}$, which collects all the dominating spectral components in the first 150 basis images. We see that the main contributions lie inside a disc in the upper left corner of the spectrum, showing that all the first basis images are low-frequent.

Figure 3.3 shows the first four basis images for $\overline{A}$. We see that some frequencies in one direction appear before the corresponding frequency in other directions in the image. This is clearly seen in the fourth basis image $V^{[4]}$, which is dominated by the DCT component $\check{V}^{[4]}_{3,1}$, while in the symmetric case $V^{[4]}$ is dominated by $\check{V}^{[4]}_{2,2}$. We also see from the plot of the "accumulated DCT" $\left(\sum_{i=1}^{150} |\check{V}^{[i]}|^2\right)^{1/2}$ that the frequency contents are no longer located primarily inside a disc, but the main contribution is still low-frequent.

The above study corresponds to using zero boundary conditions, but precisely the same behavior is observed if we use other boundary conditions in (2.1). We conclude that, in analogy with 1D problems, we can think of the SVD as a 2D spectral decomposition of the problem.

4. Krylov subspace methods. Given a matrix M and a vector v, the Krylov subspace of dimension k is defined as $\mathcal{K}_k(M, v) \equiv \operatorname{span}\left\{v, Mv, M^2v, \ldots, M^{k-1}v\right\}$. Krylov subspaces have been studied extensively (see, e.g., [8, 22]) and they are necessary for investigating the regularizing properties of the methods in consideration.

4.1. The methods studied here. The two algorithms CGLS and LSQR, which are designed to solve the least squares problem $\min \|A\,x - b\|_2$, are mathematically equivalent.

ETNA
Kent State University
http://etna.math.kent.edu

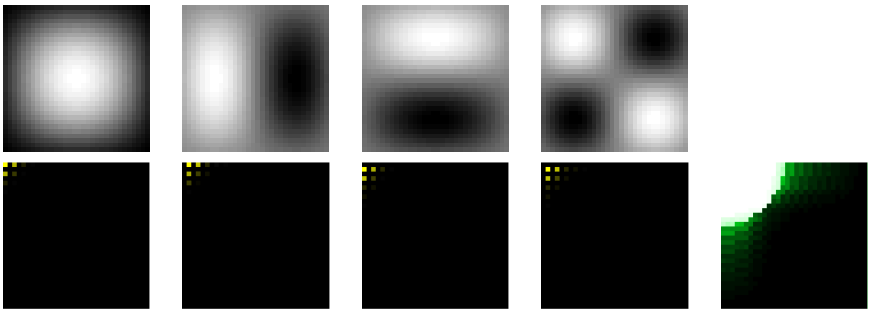

FIGURE 3.2. *Basic images* $V^{[i]}$, $i = 1, 2, 3, 4$ *(top) for the symmetric coefficient matrix* A *and theirs DCTs* $|\check{V}^{[k]}|$ *(bottom), together with the "accumulated DCT"* $(\sum_{i=1}^{150} |\check{V}^{[i]}|^2)^{1/2}$.

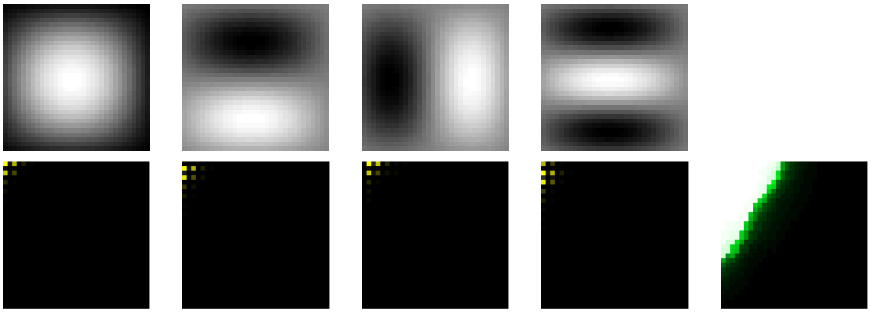

FIGURE 3.3. *Similar to Figure 3.2, but for the non-symmetric coefficient matrix* $\overline{A}$.

They work implicitly with the normal equations $A^T A\, x = A^T b$ and are thus based on the Krylov subspace $\mathcal{K}_k(A^T A, A^T b)$. LSQR constructs basis vectors for this Krylov subspace via the Lanczos bidiagonalization algorithm, which gives the partial decomposition

$$A\, W_k = U_{k+1} B_k, \tag{4.1}$$

where W_k is a matrix with k orthonormal columns that span the Krylov subspace. The matrix U_{k+1} has $k+1$ orthonormal columns, and its first column is chosen as $u_1 = b/\|b\|_2$, which simplifies the implementation considerably. The matrix B_k is a $(k+1) \times k$ lower bidiagonal matrix. The LSQR iterate $x^{(k)}$ minimizes the 2-norm of the residual in the Krylov subspace, i.e., $x^{(k)} = \text{argmin}_x \|A\, x - b\|_2$ such that $x \in \mathcal{K}_k(A^T A, A^T b)$, and it follows that

$$x^{(k)} = W_k\, \xi_k, \qquad \xi_k = \text{argmin}_\xi \|B_k\, \xi - \rho\, e_1\|_2,$$

where e_1 is the first canonical unit vector in $\mathbb{R}^{k+1}$ and $\rho = \|b\|_2$. This algorithm can be implemented using short recurrences, and thus one can avoid storing the partial decomposition (4.1).

The GMRES algorithm is based on the Arnoldi process, that constructs an orthonormal basis for the Krylov subspace $\mathcal{K}_k(A, b)$ leading to the partial decomposition

$$A\, \widehat{W}_k = \widehat{W}_{k+1} \widehat{H}_k, \qquad \widehat{W}_{k+1} = (\, \widehat{W}_k\, ,\, \widehat{w}_{k+1}\,), \tag{4.2}$$

where the k columns of $\widehat{W}_k$ provide an orthonormal basis for the Krylov subspace, and $\widehat{H}_k$ is a $(k+1) \times k$ upper Hessenberg matrix. The first column in $\widehat{W}_k$ is again chosen as $\widehat{w}_1 = b/\|b\|_2$.

ETNA
Kent State University
http://etna.math.kent.edu

FIGURE 4.1. *Image deblurring examples. Left to right: true image X, blurred image B due to the symmetric PSF matrix A, and blurred image $\overline{B}$ due to the nonsymmetric PSF matrix $\overline{A}$.*

If A is symmetric then $\widehat{H}_k$ reduces to tridiagonal form, and the construction of the partial decomposition (4.2) can be done by a three-term recurrence as implemented in MINRES. In this case the solution can be updated without explicitly storing the partial decomposition. In the general case, no such short recurrence exists and GMRES needs to store all the constructed Krylov vectors.

The GMRES/MINRES iterate $\hat{x}^{(k)}$ minimizes the residual norm with respect to the above Krylov subspace, i.e., $\hat{x}^{(k)} = \operatorname{argmin}\|A\,x - b\|_2$ such that $x \in \mathcal{K}_k(A, b)$, which leads to the relation

$$\hat{x}^{(k)} = \widehat{W}_k\,\widehat{\xi}_k, \qquad \widehat{\xi}_k = \operatorname{argmin}_{\widehat{\xi}}\|\widehat{H}_k\,\widehat{\xi} - \rho\,e_1\|_2,$$

where again e_1 is the first canonical unit vector and $\rho = \|b\|_2$.

There exists a variant of GMRES that uses Ab as the starting vector for the Krylov subspace, instead of b, leading to the "shifted" Krylov subspace $\mathcal{K}_k(A, Ab)$. In the nonsymmetric case the algorithm is called RRGMRES [4, 5], and the algorithm MR-II [9, 10] is an efficient short-term recurrence implementation of this method for symmetric matrices. The partial decomposition in RRGMRES/MR-II is written as

$$A\,\widetilde{W}_k = \widetilde{W}_{k+1}\widetilde{H}_k, \qquad \widetilde{W}_{k+1} = (\,\widetilde{W}_k\,,\,\widetilde{w}_{k+1}\,).$$

The RRGMRES/MR-II iterate is then $\tilde{x}^{(k)} = \widetilde{W}_k\,\widetilde{\xi}_k$ with $\widetilde{\xi}_k = \operatorname{argmin}_{\widetilde{\xi}}\|\widetilde{H}_k\,\widetilde{\xi} - \widetilde{W}_{k+1}^T b\|_2$. These two methods are now available in REGULARIZATION TOOLS [13].

4.2. Examples of iterates. We illustrate the typical behavior of the iterative methods using two examples. The true image X is a 175×175 sub-image of the image "Barbara" [2]. Two PSF matrices are used: a symmetric A and a nonsymmetric $\overline{A}$ with the following parameters (see Appendix A for details)

$$A\ :\ s_{\mathrm{c}} = 4, \quad s_{\mathrm{r}} = 4, \quad \alpha = 0; \qquad \overline{A}\ :\ s_{\mathrm{c}} = 8, \quad s_{\mathrm{r}} = 10, \quad \alpha = 4. \tag{4.3}$$

In both cases, we add Gaussian white noise E scaled such that $\|E\|_{\mathrm{F}}/\|B\|_{\mathrm{F}} = 0.05$. The true image and the two blurred and noisy images are shown in Figure 4.1.

Figure 4.2 shows the LSQR, MINRES, and MR-II solutions after 5, 10, and 25 iterations for the symmetric PSF matrix, and we see that the algorithms give very different solutions. The LSQR solutions slowly improve, but after 25 iterations some noise has appeared as small circular "freckles" (see Appendix B for a characterization of these freckles). The MINRES solutions get dominated very fast by high-frequency noise; with the noise level defined above, the iterates are strongly deteriorated by noise after only 5 iterations. The MR-II solutions show artifacts similar to the LSQR solutions, but the convergence seems faster as the solution after 10 MR-II iterations is visually similar to the LSQR solution after 25 iterations.

ETNA
Kent State University
http://etna.math.kent.edu

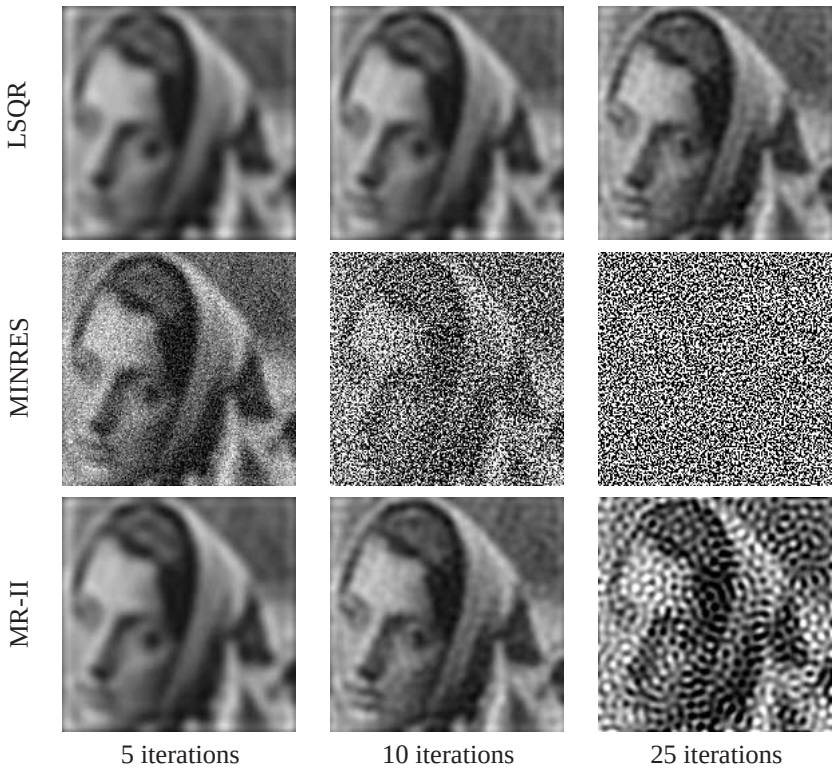

FIGURE 4.2. *LSQR, MINRES, and MR-II iterates* $x^{(k)}$, $\hat{x}^{(k)}$ *and* $\tilde{x}^{(k)}$ *with the symmetric PSF matrix* A, *for* $k = 5$, *10, and 25 iterations.*

The solutions for the non-symmetric PSF matrix are shown in Figure 4.3. Here, the LSQR algorithm is again seen to generate some structured artifacts, observed after 25 iterations. The artifacts are no-longer circular, due to the non-isotropic blurring, but they are still band-limited and certainly not high-frequent as the noise seen in the GMRES solutions. The RRGMRES solutions again show artifacts similar to the artifacts for LSQR. But the difference in convergence speed is not as large as for LSQR and MR-II in Figure 4.2.

4.3. The power basis of the Krylov subspace. The regularizing properties of the Krylov subspace methods come from the fact that for discrete ill-posed problems, the right-hand side b is rich in SVD components corresponding to the large singular values. For this reason, the basis vectors for the Krylov subspaces are also rich in these directions. We illustrate this in Figure 4.4, which shows the coefficients of the first five Krylov vectors in the SVD basis (the columns of V) for all methods.

Consider first the Krylov vectors in the left part of Figure 4.4, corresponding to the symmetric matrix A. For all methods, it is evident that all the Krylov vectors are rich in right singular vectors corresponding to the largest singular values, and that the coefficients are damped according to the multiplication with $A^T A$ or A.

However, the fact that the noisy right-hand side b is present in the MINRES basis has a dramatic impact, because the noise component $e = \mathrm{vec}(E)$ in b is present in the Krylov subspace, leading to a large amount of white noise in the first Krylov vector, and therefore in all the MINRES solutions. While the remaining Krylov vectors are identical to those of MR-II, we see that it is crucial to avoid the noisy vector b in the Krylov subspace.

ETNA
Kent State University
http://etna.math.kent.edu

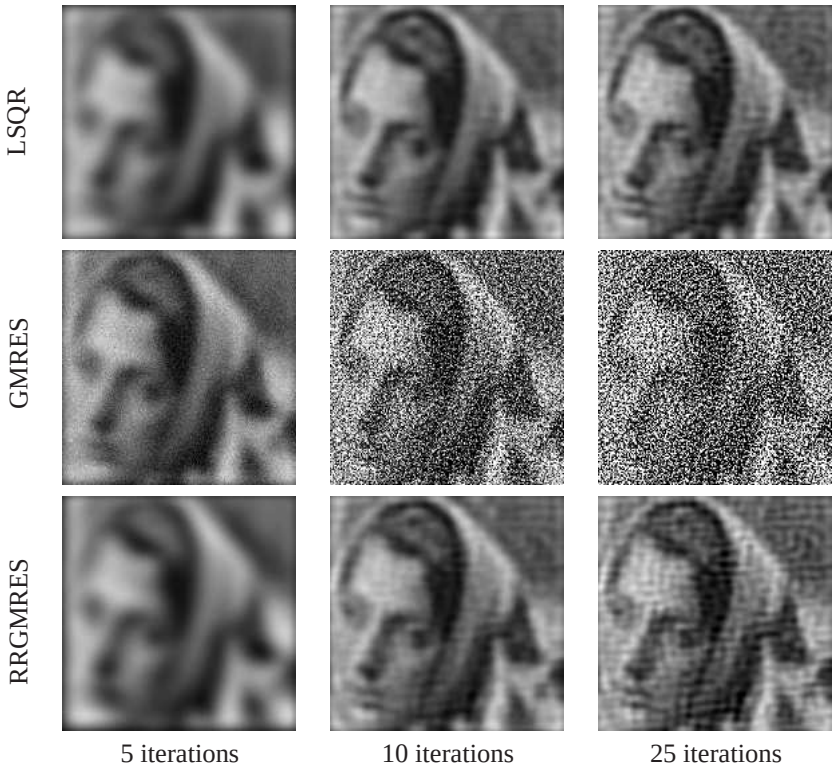

FIGURE 4.3. *LSQR, GMRES, and RRGMRES iterates $x^{(k)}$, $\hat{x}^{(k)}$ and $\tilde{x}^{(k)}$ with the nonsymmetric PSF matrix $\overline{A}$, for $k = 5$, 10, and 25 iterations.*

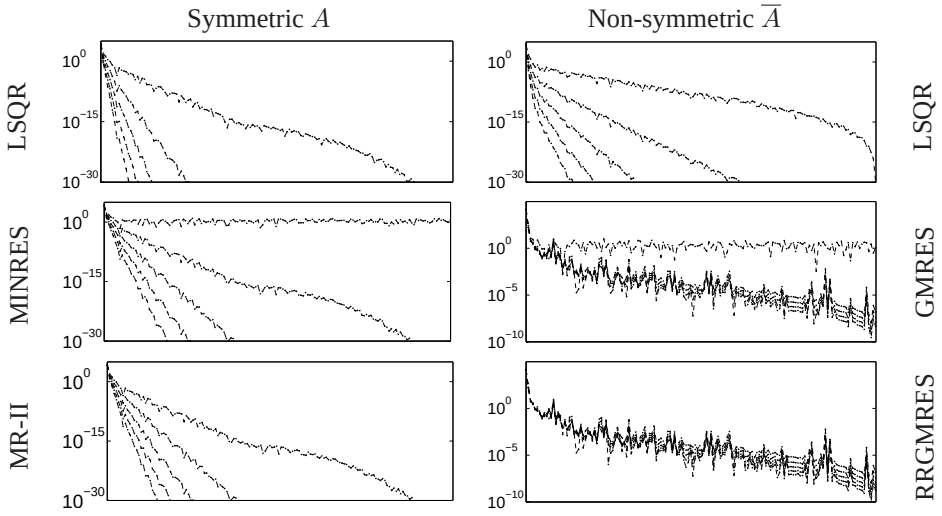

FIGURE 4.4. *The coefficients in the SVD basis of the first five Krylov subspace vectors for the different algorithms. The first vector is always the one on top. Note that the axes for GMRES and RRGMRES are scaled differently from the rest.*

ETNA
Kent State University
http://etna.math.kent.edu

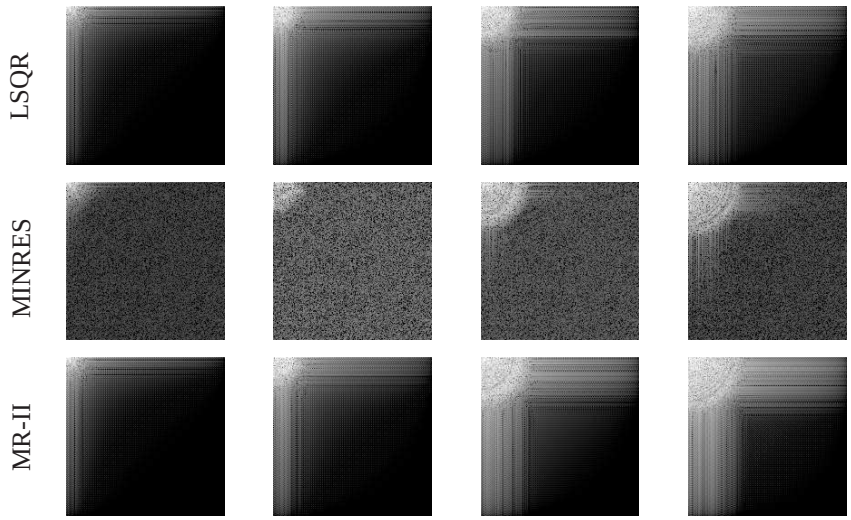

FIGURE 4.5. *DCTs associated with selected Krylov and Arnoldi vectors, i.e., columns of W_k, $\widehat{W}_k$, and $\widetilde{W}_k$ (converted to images), for $k = 40$ iterations. From left to right, we show columns 1, 3, 15, and 40.*

The SVD coefficients for the first five Krylov vectors for the non-symmetric matrix $\overline{A}$ are shown in the right part of Figure 4.4. The CGLS/LSQR Krylov vectors exhibit a behavior similar to the symmetric case where the coefficients are damped by the multiplication with $\overline{A}^T\overline{A}$.

For GMRES and RRGMRES the situation is different. For GMRES the noise components $e = \mathrm{vec}(E)$ is again clearly present in the first Krylov vector, and for both methods the vectors do not exhibit much damping from one vector to the next. As explained in [18], this is caused by a mixing of the SVD components when multiplying with $\overline{A}$, where each SVD component in the kth Krylov vectors is a linear combination of (in principle) all SVD components in the previous iteration.

4.4. Two orthonormal bases of the Krylov subspace. Another perspective on the Krylov subspaces is provided by the orthonormal basis vectors that span these subspaces. Obvious choices of these vectors are the Lanczos and Arnoldi vectors, i.e., the columns of the matrices W_k, $\widehat{W}_k$, and $\widetilde{W}_k$ generated by the three methods. Figure 4.5 shows the DCTs associated with selected columns of these matrices for $k = 40$ iterations. It is evident that we include higher frequency components in the bases as we take more iterations. The white-noise component, arising from e, is clearly visible in the GMRES/MINRES basis.

However, there are other sets of bases that provide important information. If the SVD of the bidiagonal matrix B_k in (4.1) is given by

$$B_k = P_k \, \Sigma_k \, Q_k^T,$$

then we can write the CGLS/LSQR iterates as $x^{(k)} = (W_k \, Q_k) \, \Sigma_k^{-1} (\rho \, P_k^T e_1)$, and the orthonormal columns of the matrix $W_k \, Q_k$ provide an alternative basis for the CGLS/LSQR Krylov subspace. Similarly, with the SVDs of the Hessenberg matrices $\widehat{H}_k$ and $\widetilde{H}_k$ from GMRES and RRGMRES, we have

$$\widehat{H}_k = \widehat{P}_k \, \widehat{\Sigma}_k \, \widehat{Q}_k^T, \qquad \widetilde{H}_k = \widetilde{P}_k \, \widetilde{\Sigma}_k \, \widetilde{Q}_k^T,$$

ETNA
Kent State University
http://etna.math.kent.edu

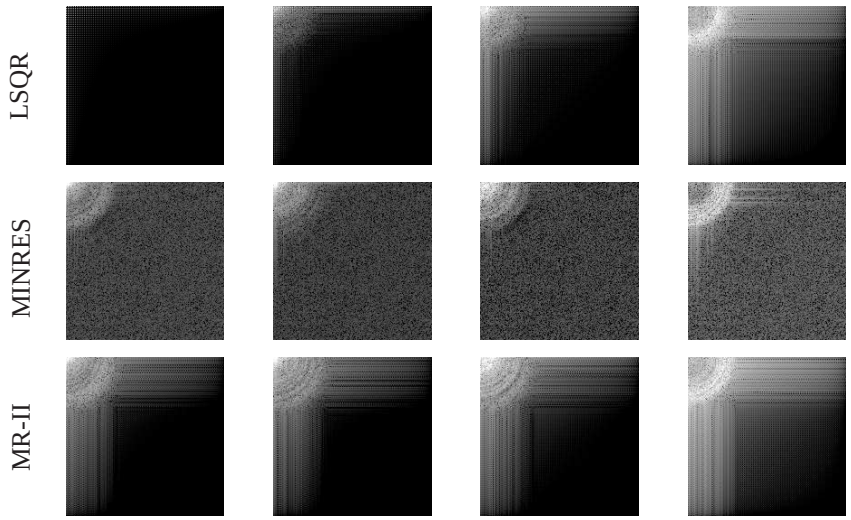

FIGURE 4.6. *DCTs associated with selected columns of $W_k\,Q_k$, $\widehat{W}_k\,\widehat{Q}_k$, and $\widetilde{W}_k\,\widetilde{Q}_k$ (converted to images) for $k = 40$ iterations. From left to right, we show columns 1, 3, 15, and 40.*

and thus the columns of $\widehat{W}_k\,\widehat{Q}_k$ and $\widetilde{W}_k\,\widetilde{Q}_k$ provide orthonormal bases for the Krylov subspaces of GMRES/MINRES and RRGMRES/MR-II, respectively.

Figure 4.6 shows the DCTs associated with selected columns of the matrices $W_k\,Q_k$, $\widehat{W}_k\,\widehat{Q}_k$, and $\widetilde{W}_k\,\widetilde{Q}_k$ for $k = 40$ iterations. As we take more iterations, the dominating spectral components in these basis vectors are band-limited and lie in a fairly narrow band of spatial frequencies (showing up as freckles in the solutions). These DCTs thus confirm that, for each new iteration, we tend to keep the existing low-frequency information in the iteration vector, and add components with slightly higher frequencies. Again, the white noise component is clearly present in all the GMRES/MINRES basis vectors.

5. Study of signal and noise components. We now study in more detail how the noise propagates to the solutions, by considering how the three methods treat, on one hand, the wanted signal contents from the exact component $b^{\text{exact}} = \text{vec}(B^{\text{exact}})$ and, on the other hand, the contents from the noise component $e = \text{vec}(E)$. This analysis is, of course, only possible when the noise is explicitly known.

The LSQR solution, given by $x^{(k)} = W_k\,B_k^{\dagger}\,U_{k+1}^T b$, can be split into the signal and noise components $x^{(k)} = x_{b^{\text{exact}}}^{(k)} + x_e^{(k)}$, with

$$x_{b^{\text{exact}}}^{(k)} = W_k\,B_k^{\dagger}\,U_{k+1}^T b^{\text{exact}} \qquad \text{and} \qquad x_e^{(k)} = W_k\,B_k^{\dagger}\,U_{k+1}^T e.$$

Similarly, for the GMRES and RRGMRES methods we have the splitting

$$\hat{x}^{(k)} = \hat{x}_{b^{\text{exact}}}^{(k)} + \hat{x}_e^{(k)} \qquad \text{and} \qquad \tilde{x}^{(k)} = \tilde{x}_{b^{\text{exact}}}^{(k)} + \tilde{x}_e^{(k)}.$$

The Lanczos bidiagonalization process is generated by b (and not b^{exact}), and therefore $x_{b^{\text{exact}}}^{(k)}$ differs from the LSQR solution produced with b^{exact} as starting vector. The same is true for GMRES and RRGMRES. This situation, where the signal component depends on the noise in the data, is unique for regularizing iterations, due to the dependence of the Krylov subspace on $b = b^{\text{exact}} + e$. In regularization methods, such as Tikhonov and TSVD, the filter factors f_k

ETNA
Kent State University
http://etna.math.kent.edu

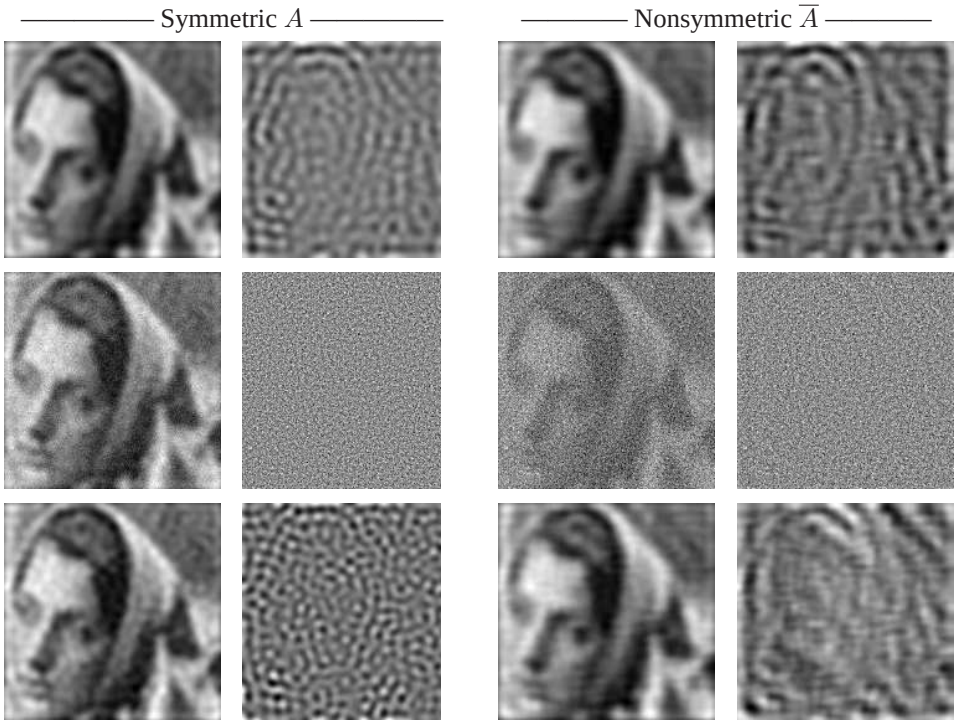

FIGURE 5.1. *Splitting of the solutions by LSQR (top), GMRES (middle), and RRGMRES (bottom) after $k = 10$ iterations, in the signal and noise components.*

in (2.3) depend only on A, and thus the signal component

$$\sum_{k=1}^{n^2} f_k \, \sigma_k^{-1} u_k^T b^{\text{exact}} \, v_k$$

is independent on e.

The above splitting is often used in studies of regularization methods, where one writes the reconstruction error $x^{(k)} - x^{\text{exact}}$ as a sum of the regularization error $x_{b^{\text{exact}}}^{(k)} - x^{\text{exact}}$ and the perturbation error $x_e^{(k)}$ [12]. For example, it was used in [3] and [20] to study error propagation and ringing effects for Tikhonov image deblurring. The situation is more complex here, because the signal and noise components are coupled differently for different Krylov subspace methods and, in fact, for different right-hand sides. The purpose of this section is to illustrate this aspect.

Figure 5.1 illustrates the splitting for LSQR, GMRES, and RRGMRES for the "Barbara" image. Both the symmetric and the non-symmetric PSF matrices are studied. We see how the noise propagates very differently in the three methods, due to the differences in the associated Krylov subspaces.

The LSQR algorithm produces low-frequent ringing effects in the signal component $x_{b^{\text{exact}}}^{(k)}$ for both the symmetric and the nonsymmetric coefficient matrix, and no freckles are present in this component. In accordance with the observations in the previous section, the noise component $x_e^{(k)}$ consists of bandpass-filtered noise in the form of freckles, and the shape of the freckles depends on the shape of the point-spread function. It is interesting to

ETNA
Kent State University
http://etna.math.kent.edu

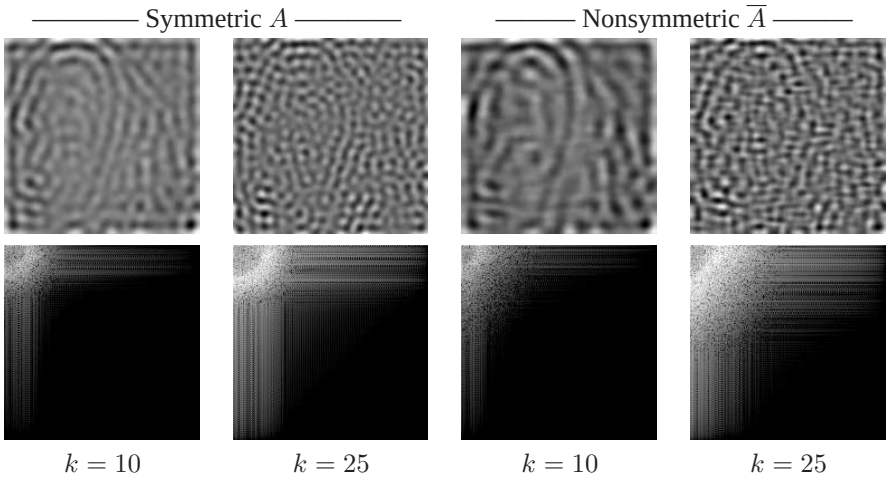

FIGURE 5.2. *The LSQR noise components $x_e^{(k)}$ and the corresponding DCTs.*

see how both the ringing in the signal component and the freckles in the noise component are correlated with the contours of the image, caused by the specific Krylov subspace.

As we have already observed, MINRES and GMRES propagate a white-noise component in the signal component $\hat{x}_{b^{\text{exact}}}^{(k)}$, caused by the explicit presence of the noise in the basis vectors for the Krylov subspace. The white-noise component is particularly pronounced in the GMRES signal component. This example clearly illustrates why MINRES and GMRES cannot be recommended as general regularization methods, a point that is substantiated further in [18].

The RRGMRES and MR-II signal components behave much like the LSQR signal components, except that they tend to carry more details after the same number of iterations. The noise components resemble those of the LSQR method. For the symmetric matrix, the freckles are smaller in diameter than for LSQR, and they are more visible in the signal component. For the nonsymmetric matrix, both components are quite similar to the LSQR components. Avoiding the noisy vector b in the RRGMRES/MR-II subspace gives a huge improvement of the quality of the solutions.

To study the freckles in more detail, we consider the LSQR method and compute the DCT of the noise components $x_e^{(k)}$ for iterations $k = 10$ and $k = 25$, as shown in Figure 5.2 for both the symmetric and the nonsymmetric coefficient matrix. This figure confirms that the freckles are indeed bandlimited noise, because they correspond to a bandlimited ring of frequencies in the DCT domain, and the ring moves towards higher frequencies as the number of iterations increases.

A closer analysis of the LSQR noise component $x_e^{(k)}$ reveals that it tends to be dominated by contributions from the last few columns in the matrix $W_k\, Q_k$, and similarly the RRGMRES noise component tends to be dominated by the last few columns of $\widetilde{W}_k\, \widetilde{Q}_k$. These vectors represent the highest spatial frequencies in the reconstruction, thus explaining the presence of the freckles. A simple mechanism for removing (some of) the freckles is thus to remove these particular components in the solution via a truncated SVD solution to the "projected problems" $\min \|B_k\, \xi - \rho\, e_1\|_2$ and $\min \|\widetilde{H}_k\, \widetilde{\xi} - \widetilde{W}_{k+1}^T b\|_2$, using a truncation parameter close to k. This use of regularization applied to the projected problem is advocated in [11, 19], and the resulting method is sometimes referred to as a *hybrid method*.

ETNA
Kent State University
http://etna.math.kent.edu

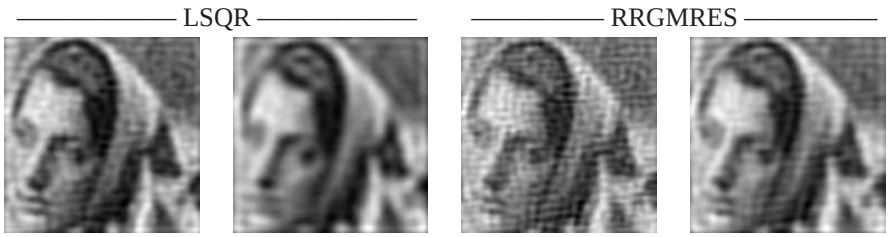

FIGURE 5.3. *Comparison of LSQR and RRGMRES solutions for $k = 25$ with those obtained by TSVD regularization of the projected problem. The removal of the smallest SVD component dramatically reduces the amount of freckles.*

Figure 5.3 compares the standard and regularized LSQR and RRGMRES solutions for $k = 25$; the removal of the smallest SVD components in the projected problem clearly removes a substantial amount of freckles. At the same time, the modified solutions appear less sharp, because the highest frequencies (that gave rise to the freckles) are explicitly removed.

6. Low noise levels are different. For a very low noise level, GMRES was reported in [6] to exhibit faster convergence than LSQR and produce "better" solutions. This seems contradictory to our results, and the purpose of this final section is to study the quality of the reconstructions for low noise levels. We use the nonsymmetric PSF matrix, and the noise level is reduced to $\|E\|_{\mathrm{F}}/\|B^{\mathrm{exact}}\|_{\mathrm{F}} = 5 \cdot 10^{-4}$.

The two top rows in Figure 6.1 show the LSQR, GMRES, and RRGMRES solutions after 30 iterations, together with their DCTs. From the top row, it appears that the GMRES solution has the highest amount of visual detail, and it seems to be superior to the RRGMRES and LSQR solutions. But from the second row we also see that the GMRES solution carries a much larger amount of white noise than the LSQR and RRGMRES solutions.

The two bottom rows show the corresponding noise components and their DCTs. As before, the noise component appears as freckles, and the location of the freckles is correlated with the contours in the image. For all iterations $k \leq 30$ the norm of the noise component is much smaller than the norm of the solution component. Hence, the error in the reconstructions is primarily due to the error in the solution component.

For low noise levels, the GMRES solution appears to be visually superior to the other solutions because the freckles are not as visible as for higher noise levels. The freckles are very disturbing to the human eye, while this is not the case for the propagated white noise in the signal component. In fact, the white noise in $\hat{x}^{(k)}$ "masks" the freckles and creates an illusion of improved resolution through higher spatial frequencies, even though no information can be reliably reconstructed beyond the spatial frequencies associated with the freckles.

The LSQR and RRGMRES solutions, on the other hand, do not carry a large white-noise component that "masks" the freckles, and hence these solutions behave similar to the case of larger noise levels, and they appear as inferior to the GMRES solution. However, all reconstructions have errors of the same magnitude, the errors just appear in different incarnations with very different spectral properties.

7. Conclusion. We used the two-dimensional DCT to study the properties of regularizing iterations for image deblurring. First, we showed that the SVD provides a two-dimensional frequency decomposition similar to the one-dimensional case. Turning to the algorithms, we show that all the Krylov subspace methods produce reconstructions that can be considered as low-pass filtered versions of the "naive" reconstruction (similar to Tikhonov and TSVD). The main difference is that, for regularizing iterations, noise and artifacts in the

ETNA
Kent State University
http://etna.math.kent.edu

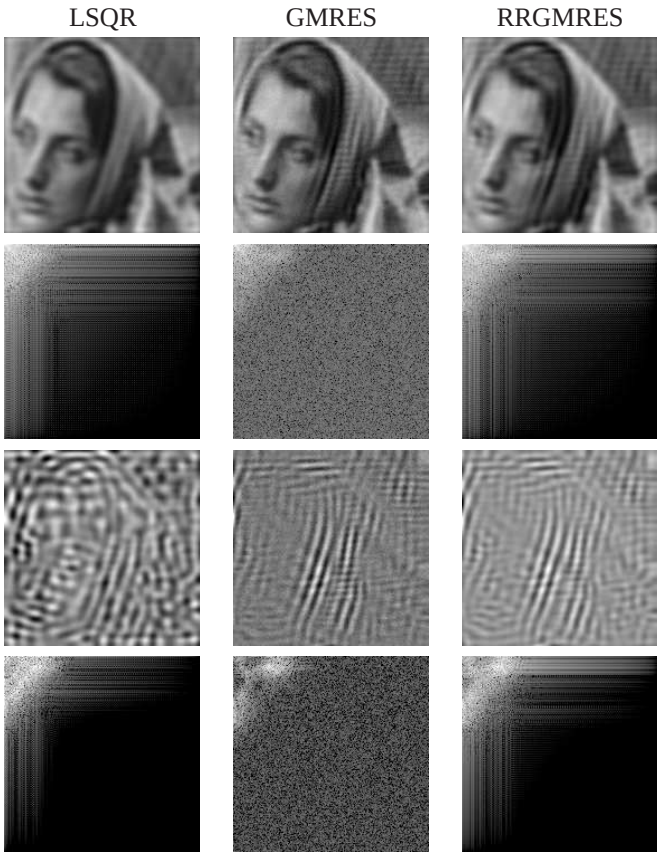

FIGURE 6.1. *Top rows: LSQR, GMRES, and RRGMRES solutions after 30 iterations and their DCTs, for the nonsymmetric PSF matrix $\overline{A}$ with a very low noise level $\|E\|_{\mathrm{F}}/\|B^{\mathrm{exact}}\|_{\mathrm{F}} = 5 \cdot 10^{-4}$. Bottom rows: the corresponding noise components and their DCTs. The GMRES solution is the visually most pleasing.*

form of "freckles" are present in both the signal and noise components. CGLS/LSQR and RRGMRES/MR-II are superior to GMRES/MINRES for regularization problems because they provide better suppression of the noise; for our examples there is no clear winner among the two methods.

All our Matlab codes used in the examples in this work are available from the home page `http://www.imm.dtu.dk/~pch/NoisePropagation.html`.

Appendix A. Isotropic and non-isotropic blurring matrices. Underlying the image deblurring problem is a 2D Fredholm integral equation of the first kind

$$\int_0^1 \int_0^1 K(x', y', x, y)\, F(x, y)\, dx\, dy = G(x', y'), \qquad 0 \le s, t \le 1, \tag{A.1}$$

where F and G represent the exact and blurred images, and the kernel K is the *point spread function* (PSF) for the blurring. Our work here is restricted to the case where the PSF is spatially invariant and separates in the variables, i.e., $K(x', y', x, y) = \kappa_{\mathrm{c}}(x' - x)\, \kappa_{\mathrm{r}}(y' - y)$, where κ_{c} and κ_{r} are given functions.

When we discretize (A.1) by means of the midpoint quadrature rule with collocation in the $n \times n$ grid of midpoints of the pixels, then we obtain two matrices X and B which are

ETNA
Kent State University
http://etna.math.kent.edu

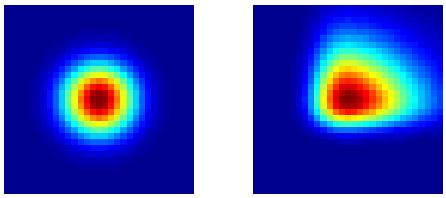

FIGURE A.1. *The isotropic (left) and non-isotropic (right) PSFs used in our numerical experiments, obtained by means of the parameters in equation (4.3).*

samples of the images F and G. Moreover, the matrices A_{c} and A_{r} in equation (2.1) are Toeplitz matrices whose elements are samples of the functions κ_{c} and κ_{r}.

For our numerical experiments we need both isotropic and non-isotropic PSFs, and we use a generalization of the normal distribution. Let

$$\phi(x) = \frac{1}{s\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2s^2}\right) \qquad \text{and} \qquad \Phi(x) = \int_{-\infty}^{x} \phi(t)\, dt$$

be the standard Gaussian density function and its distribution function, and define

$$\mathcal{F}_\alpha(x) = 2\,\phi(x)\,\Phi(\alpha x).$$

Then $\mathcal{F}_\alpha$ is the density function for the *skew-normal distribution* [1], in which α is the skewness parameter. The scale parameter s controls the width of $\mathcal{F}_\alpha$, and by a proper choice of the location parameter μ we can ensure that the maximum of the density function $\mathcal{F}_\alpha$ is centered at $x = 0$.

By setting the functions κ_{c} and κ_{r} equal to $\mathcal{F}_\alpha$ for different choices of α and s, we obtain point-spread functions that are, in general, non-isotropic. The pseudo-Matlab code for generating the Toeplitz matrices A_{c} and A_{r} takes the form

```
vecAc = skewnormal(-n+1:n-1, μc, sc, α);
Ac = toeplitz(vecAc(n:-1:1),vecAc(n:end));
vecAr = skewnormal(-n+1:n-1, μr, sr, α);
Ar = toeplitz(vecAr(n:-1:1),vecAr(n:end));
```

where the function `skewnormal`, which computes the skew-normal density function $\mathcal{F}_\alpha$, is given by

```
function y = skewnormal(x, s, alpha)
mu = fminsearch(@SN,0,[ ],0,s,alpha);
y = -SN(x,mu,s,alpha);
function y = SN(x, mu, s, alpha) % Subfunction.
y = -2*normpdf(x+mu,0,s).*normcdf(alpha*(x+mu),0,s);
```

Here we use functions from Matlab's STATISTICS TOOLBOX to generate the Gaussian density and distribution functions. Figure A.1 shows two PSFs generated by this approach.

Appendix B. "Freckles" are band-pass filtered white noise. A white-noise random image X is characterized by having pixels whose values are statistically uncorrelated and have the same variance. Since the DCT is an orthogonal transformation, it follows that the DCT image $\check{X}$ is also a white-noise image.

A filtered random image is obtained by filtering its spectral components. In particular, if `F` is a matrix of zeros and ones, then we can generate an image of filtered Gaussian white noise by means of X_{filt} = `idct2(F.*randn(n))`.

ETNA
Kent State University
http://etna.math.kent.edu

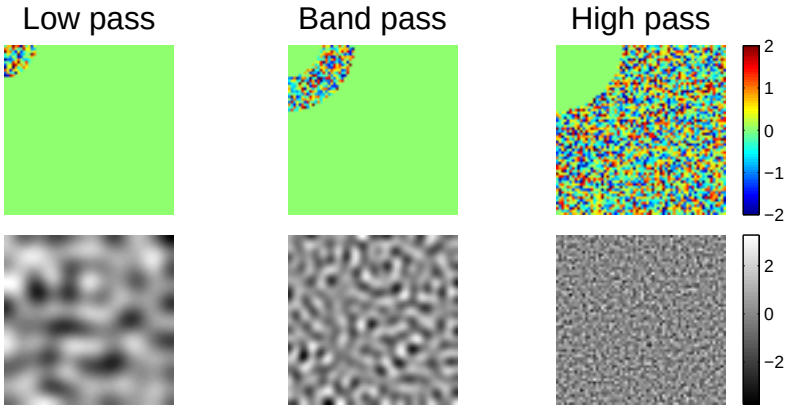

FIGURE B.1. *Top: the filtered DCT images* `idct2(F.*randn(n))`. *Bottom: the corresponding filtered images* X_{filt} = `idct2(F.*randn(n))`. *Band-pass filtered white noise appears as "freckles".*

The choice of the locations of the ones in F determines the type of filtering. If $\mathtt{F}_{ij} = 1$ for $i^2+j^2 < k^2$ for some $k < \sqrt{2}\,n$, then we obtain low-frequency noise, while high-frequency noise is obtained for $i^2+j^2 > k^2$. If the indices for $\mathtt{F}_{ij} = 1$ satisfy $k_{\text{lo}} < i^2+j^2 < k_{\text{hi}}$, then we keep only spectral components in a range of spatial frequencies, and we say that X_{filt} is *band-pass filtered white noise*. This kind of noise has the characteristic visual appearance of the "freckles" that we often see in the reconstructions. See Figure B.1 for examples of filtered noise.

REFERENCES

[1] A. AZZALINI AND A. D. VALLE, *The multivariate skew-normal distribution*, Biometrika, 83 (1996), pp. 715–726.
[2] BARBARA, *test image.* Available from, e.g., `http://www.cs.tut.fi/~foi/GCF-BM3D/BM3D_images.zip`.
[3] J. BIEMOND, R. L. LAGENDIJK, AND R. M. MERSEREAU, *Iterative methods for image deblurring*, Proc. IEEE, 78 (1990), pp. 856–883.
[4] D. CALVETTI, B. LEWIS, AND L. REICHEL, *GMRES-type methods for inconsistent systems*, Linear Algebra Appl., 316 (2000), pp. 157–169.
[5] ———, *On the choice of subspace for iterative methods for linear discrete ill-posed problems*, Int. J. Appl. Math. Comput. Sci., 11 (2001), pp. 1069–1092.
[6] ———, *GMRES, L-curves, and discrete ill-posed problems*, BIT, 42 (2002), pp. 44–65.
[7] R. H. CHAN AND X.-Q. JIN, *An Introduction to Iterative Toeplitz Solvers*, SIAM, Philadelphia, 2007.
[8] A. GREENBAUM, *Iterative Methods for Solving Linear Systems*, SIAM, Philadelphia, 1997.
[9] M. HANKE, *Conjugate Gradient Type Methods for Ill-Posed Problems*, Longman Scientific & Technical, 1995.
[10] ———, *The minimal error conjugate gradient method is a regularization method*, Proc. Amer. Math. Soc., 123 (1995), pp. 3487–3497.
[11] M. HANKE AND P. C. HANSEN, *Regularization methods for large-scale problems*, Surv. Math. Industry, 3 (1993), pp. 253–315.
[12] P. C. HANSEN, *Rank-Deficient and Discrete Ill-Posed Problems*, SIAM, Philadephia, 1998.
[13] ———, *Regularization Tools version 4.0 for Matlab 7.3*, Numer. Algorithms, 46 (2007), pp. 189–194.
[14] P. C. HANSEN AND T. K. JENSEN, *Smoothing-norm preconditioning for regularizing minimum-residual methods*, SIAM J. Matrix Anal. Appl., 29 (2006), pp. 1–14.
[15] P. C. HANSEN, M. E. KILMER, AND R. H. KJELDSEN, *Exploiting residual information in the parameter choice for discrete ill-posed problems*, BIT, 46 (2006), pp. 41–59.
[16] P. C. HANSEN, J. G. NAGY, AND D. P. O'LEARY, *Deblurring Images: Matrices, Spectra, and Filtering*, SIAM, Philadelphia, 2006.
[17] A. K. JAIN, *Fundamentals of Digital Image Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1989.

[18] T. K. JENSEN AND P. C. HANSEN, *Iterative regularization with minimum-residual methods*, BIT, 47 (2007), pp. 103–120.
[19] M. E. KILMER AND D. P. O'LEARY, *Choosing regularization parameters in iterative methods for ill-posed problems*, SIAM J. Matrix Anal. Appl., 22 (2001), pp. 1204–1221.
[20] R. L. LAGENDIJK AND J. BIEMOND, *Iterative Identification and Restoration of Images*, Kluwer Academic Publishers, Boston, MA, 1991.
[21] ———, *Basic methods for image restoration and identification*, in Handbook of Image and Video Processing, A. Bovik, ed., Academic Press, 2005, pp. 167–181.
[22] G. MEURANT AND Z. STRAKOŠ, *The Lanczos and conjugate gradient algorithms in finite precision arithmetic*, Acta Numer., 15 (2006), pp. 471–542.
[23] J. G. NAGY, K. P. LEE, AND L. PERRONE, *Iterative methods for image deblurring: A Matlab object oriented approach*, Numer. Algorithms, 36 (2004), pp. 73–93.
[24] M. K. NG, *Iterative Methods for Toeplitz Systems*, Oxford University Press, 2004.
[25] M. K. NG, R. H. CHAN, AND W.-C. TANG, *A fast algorithm for deblurring models with Neumann boundary conditions*, SIAM J. Sci. Comput., 21 (1999), pp. 851–866.
[26] G. STRANG, *The discrete cosine transform*, SIAM Rev., 41 (1999), pp. 135–147.
[27] Z. WANG, A. C. BOVIK, AND L. LU, *Why is image quality assessment so difficult?*, in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing 2002 (ICASSP '02), vol. 4, New York, 2002, IEEE Press, pp. 3313–3316.