On Lanczos based methods for the regularization of discrete ill-posed problems

Martin Hanke *

September 20, 2000

Abstract

We study hybrid methods for the solution of linear ill-posed problems. Hybrid methods are based on the Lanczos process, which yields a sequence of small bidiagonal systems approximating the original ill-posed problem. In a second step, some additional regularization, typically the truncated SVD, is used to stabilize the iteration. We investigate two different hybrid methods and interpret these schemes as well-known projection methods, namely least-squares projection and the dual least-squares method. Numerical results are provided to illustrate the potential of these methods. This gives interesting insight into the behavior of hybrid methods in practice.

Keywords: 65F10, 65F22

1 Introduction

We consider the solution of a linear system $Ax = b$ obtained by appropriate discretization of some ill-posed operator equation. For simplicity we shall assume throughout that $A \in \mathbb{R}^{m\times n}$ has full column rank, i.e., $m \geq n$ and A is injective. However, all the results extend to general matrices A and even to linear operators in Hilbert spaces.

Linear systems of this sort share a number of remarkable properties:

- The singular values of A cluster at zero giving rise to huge condition numbers

*Fachbereich Mathematik, Johannes Gutenberg-Universität, D–55099 Mainz, Germany (hanke@mathematik.uni-mainz.de)

- The exact right-hand side b satisfies a so-called *discrete Picard condition*, i.e., in the singular value expansion the magnitude of a component of b in the direction of a left singular vector of A is connected to the magnitude of the corresponding singular value. This means that despite its huge condition number the linear system is *effectively well-conditioned* according to a notion by Chan and Foulser [4].

- Reality is different, though, due to data perturbations, which effect all components of b in much the same way.

As a matter of fact, stable approximations of the true solution $x^\dagger$ can only be computed if the linear system is properly *regularized*. For example, one can use *Tikhonov regularization* or the *truncated singular value decomposition*. However, these two methods are very expensive to implement for large-scale applications.

Another option is the *conjugate gradient iteration* CGLS applied to the normal equation system $A^*Ax = A^*b$. It is well-known that CGLS *semiconverges* for this kind of problems, i.e., the iterates first seem to converge to the true solution vector x before they are misled by noisy components in the data and subsequently deteriorate. Thus, the conjugate gradient iteration will only give useful results if terminated early. We emphasize that this is an amenable feature in terms of work load.

For this reason, the conjugate gradient iteration is probably the most efficient tool for solving large-scale discrete ill-posed problems. In fact, it has been shown before and we will provide additional numerical evidence below, that the accuracy obtained with CGLS compares well with other competing algorithms – provided the iteration is properly terminated.

Such *stopping rules* are not as easy to find, though. The so-called *discrepancy principle* is probably the most reliable rule, but it requires knowledge of the noise level in the data. Other stopping rules that have been suggested in the literature are working for some examples, but fail for many others.

Some conjugate gradient type methods are intimately connected to the *Lanczos process*. With the Lanczos process the linear system is projected onto increasing Krylov subspaces, and approximate solutions in these subspaces are computed by subsequently solving small-dimensional linear systems. Being small, however, these systems allow for some additional regularization without much computational overhead.

In this work we will reinvestigate some of these methods within the framework of projection methods. It is known that projection methods have a regularizing side-effect, cf., e.g., [6] or [15, Chapter 17], because they restrict the number of possible degrees of freedom. We will show that CGLS

and another conjugate gradient method (CGME) can be viewed as a least-squares projection and as a dual least-squares method, respectively. These interpretations extend to hybrid variants based on the truncated singular value decomposition.

2 Least-squares projection

A common way of projecting a large-scale linear system $Ax = b$ onto a small-dimensional problem is by choosing subspaces $\mathcal{V}_k \subset \mathbb{R}^n$ of dimension k, say, and solving the least-squares problem

$$\|b - Ax\|_2 \longrightarrow \min \qquad \text{over} \quad x \in \mathcal{V}_k \,.$$

This method is called *least-squares projection.*

Examples include the truncated singular value decomposition and the conjugate gradient iteration CGLS. We start with the former and denote by $\{\, \sigma_j,\ u_j,\ v_j \ :\ j = 1, \ldots, n \,\}$ the singular value decomposition of A, where $\{v_j\}$ is the associated orthonormal basis of $\mathbb{R}^n$ and $\{u_j\}$ the corresponding orthonormal basis of $\mathcal{R}(A)$. By our assumption on A the singular values σ_j are positive and we assume them to be in nonincreasing order.

We shall use the notation

$$AV_k^s = U_k^s \Sigma_k \,, \qquad A^* U_k^s = V_k^s \Sigma_k \,,$$

for the *truncated singular value decomposition* (TSVD), where $V_k^s = [v_1, \ldots, v_k] \in \mathbb{R}^{n \times k}$ and $U_k^s = [u_1, \ldots, u_k] \in \mathbb{R}^{m \times k}$ are orthogonal matrices made up by the first $k \leq n$ singular vectors of A, and $\Sigma_k \in \mathbb{R}^{k \times k}$ is a diagonal matrix with the k largest singular values on its diagonal. The TSVD defines regularized approximations x_k^s of the true solution $x^\dagger$ via

$$x_k^s \ = \ \sum_{j=1}^{k} \frac{u_j^* b}{\sigma_j} \, v_j \ \in \ \mathcal{V}_k^s \ = \ \mathcal{R}(V_k^s) \,. \tag{1}$$

Because of the orthogonality of the vectors $\{u_j\}$ it is easy to see that the Euclidean norm of the residual

$$b - Ax_k^s \ = \ \sum_{j=k+1}^{m} (u_j^* b)\, u_j$$

for $x = x_k^s$ is smaller than for any other vector $x \in \mathcal{V}_k^s$.

We now turn to the *conjugate gradient iteration* CGLS, a detailed exposition of which can be found in the books [2] or [9][1]. CGLS has the well-known optimality property that its kth iterate x_k^{CGLS} minimizes the residual $b - Ax$ among all vectors x from the Krylov subspace

$$\mathcal{V}_k^{\mathcal{K}} \;=\; \mathcal{K}_k(A^*A, A^*b) \;=\; \mathrm{span}\{A^*b, A^*AA^*b, \ldots, (A^*A)^{k-1}A^*b\}\,. \qquad (2)$$

Hence, CGLS is another example of a least-squares projection method.

In computational terms the CGLS iterates are much cheaper to compute than the TSVD approximations. On the other hand, theoretical investigations predict similar error bounds for the optimal errors of TSVD and CGLS (see [6]), and these bounds can be confirmed numerically. In fact, we have seen numerical examples where the optimal reconstruction of TSVD is superior to CGLS, and vice versa. Still, it is common belief that TSVD is the most appropriate tool for the solution of ill-posed problems. We actually consider this to be a *misbelief*, and shall explain this by investigating a representative numerical example.

The example we choose is a standard test problem, `deriv(100,3)`, from Hansen's toolbox [12]. For our computations we take the dimension of the system to be $n = 100$ and we add 0.5% white noise (a Gaussian random vector with zero mean and an identity covariance matrix) on top of the data relative to the Euclidean vector norm. We only present numerical results for this particular problem and one particular noise sample, but the results are rather representative for the general situation.

Figure 1 contains the relative errors of the TSVD and CGLS approximations versus the dimension k of the respective subspaces $\mathcal{V}_k$. The two error curves are very similar qualitatively. In either case the error decays down to the minimal value (which is comparable for the two methods, namely `0.0194` vs. `0.0211`), before it increases eventually when the approximations start to diverge. The interesting observation, however, is that CGLS requires a *lower dimensional* subspace to achieve the same accuracy as TSVD. This observation can be supported to some extent by theory.

Example. Consider an infinite-dimensional problem $Ax = b$ in real Hilbert spaces, where the operator A has singular values

$$\sigma_j = j^{-s}\,, \qquad j \in \mathbb{N}\,,\; s > 0\,, \qquad (3)$$

and the solution $x^\dagger$ satisfies

$$\langle\, v_j, x^\dagger \,\rangle \sim j^{-q}\,, \qquad j \in \mathbb{N}\,,\; q > 1/2\,. \qquad (4)$$

[1] In [9] the method is called CGNE (*c*onjugate *g*radient iteration for the *n*ormal *e*quation system).

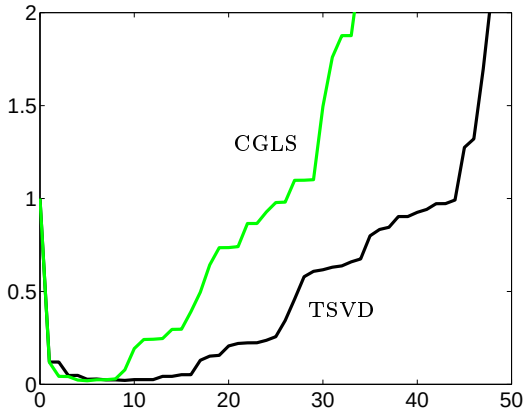

Figure 1: Reconstruction error versus k for TSVD and CGLS.

Here and in the sequel $\langle\,\cdot,\cdot\,\rangle$ and $\|\,\cdot\,\|$ denote the inner products and the associated norms in the respective Hilbert spaces.

It follows that for exact data $b = Ax^\dagger$ we have $\langle\,u_j, b\,\rangle = \sigma_j \langle\,v_j, x^\dagger\,\rangle \sim j^{-q-s}$, and it is easy to see from (1) that

$$\|x^\dagger - x_k^s\|^2 \;=\; \sum_{j=k+1}^{\infty} \langle\,v_j, x^\dagger\,\rangle^2 \;\sim\; \int_k^\infty t^{-2q}\,dt \;\sim\; k^{1-2q}\,. \tag{5}$$

For CGLS an upper bound for $\|x^\dagger - x_k^{\text{CGLS}}\|^2$ can be obtained with the results in [10, Sect. 4]. Theorem 4.1 in [10] provides a bound for the decay of the residuals $b - Ax_k^{\text{CGLS}}$. Referring to the notation in [10] we have to estimate the Christoffel functions for a measure $d\alpha$ over $\mathbb{R}^+$ with jumps at $\sigma_j^2 = j^{-2s}$ of height $\langle\,u_j, b\,\rangle^2 \sim j^{-2q-2s}$. Thus, it follows from [10, Theorem 4.1] that

$$\|b - Ax_k^{\text{CGLS}}\|^2 \;\le\; O(k^{-4\nu(s+1)})\,,$$

where ν can be any positive number up to $(2q+2s-1)/(4s)$, at most. Then, using Eq. (4.2) in [10] we obtain a similar bound for the error norm, except that ν has to be replaced by $\nu - 1/2$, i.e.,

$$\|x^\dagger - x_k^{\text{CGLS}}\|^2 \;\le\; O(k^{-(s+1)(4\nu-2)})\,.$$

Taking the range for ν into account this can be rewritten as

$$\|x^\dagger - x_k^{\text{CGLS}}\|^2 \;=\; O(k^{\frac{s+1}{s}(1-2q)+\epsilon})$$

for any arbitrarily small positive ϵ (in fact, for certain values s and particular $x^\dagger$ examplified in [10] one can even take $\epsilon = 0$). Comparing this with (5) we conclude as a rule of thumb that

$$\|x^\dagger - x_k^{\text{CGLS}}\| \lesssim \|x^\dagger - x_k^s\|^{(s+1)/s},$$

which suggests that the CGLS superiority is more pronounced for ill-posed problems with more rapidly decaying singular values.

These results may be interpreted as follows: The singular value expansion determines subspaces $\mathcal{V}_k^s$ which are suitable for *any* possible right-hand side vector of the linear system; on the other hand, the subspaces $\mathcal{V}_k^{\mathcal{K}}$ chosen by CGLS are taylored for the *particular* right-hand side, providing a more rapid convergence.

3 The dual least-squares method

For ill-posed problems the least-squares projection method in its general form has a certain shortcoming: An example of Seidman [19] shows that in infinite dimensions the subspaces $\mathcal{V}_k$ can be such that even for exact data the approximations x_k diverge as $k \to \infty$.

Therefore, Natterer [16] suggested the dual least-squares method for the regularization of linear ill-posed problems. The dual least-squares method is based on an increasing sequence of subspaces $\mathcal{U}_k \subset \mathbb{R}^m$, where we shall again assume that $\dim \mathcal{U}_k = k$. The corresponding approximation x_k is taken as the least-squares solution of minimal norm of the linear system

$$P_k A x = P_k b\,, \tag{6}$$

where P_k is the orthogonal projector onto $\mathcal{U}_k$. It is known, cf., e.g., [6, Theorem 3.24], that x_k is the orthogonal projection of $A^\dagger b$ onto the subspace $A^*(\mathcal{U}_k)$. For this reason x_k always converges to $x^\dagger = A^\dagger b$ as $k \to \infty$ when the data are exact, even in infinite dimensional Hilbert spaces.

Again, the truncated singular value decomposition is an example for this scheme with $\mathcal{U}_k = \mathcal{U}_k^s = \mathcal{R}(U_k^s)$, i.e., the span of the first k left singular vectors of A. Then it is obvious that x_k^s is the solution of (6) of minimal norm, because x_k^s is the orthogonal projection of $A^\dagger b$ onto $A^*(\mathcal{U}_k^s) = \mathcal{V}_k^s$.

The conjugate gradient method applied to $AA^*y = b$, $x = A^*y$, sometimes called *Craig's method* [5] or CGME [9], is another example for the dual least-squares method. Here one has to choose for $\mathcal{U}_k$ the Krylov subspace

$$\mathcal{U}_k^{\mathcal{K}} = \mathcal{K}_k(AA^*, b) = \text{span}\{b, AA^*b, \ldots, (AA^*)^{k-1}b\}\,,$$

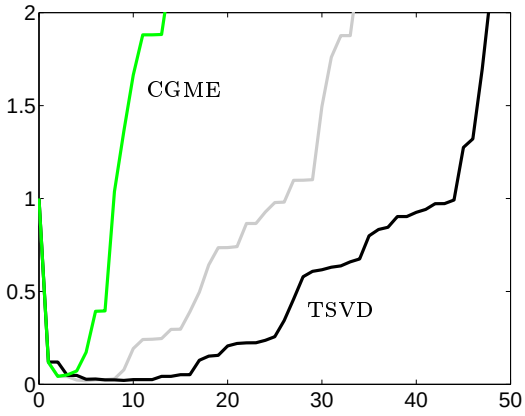

Figure 2: Reconstruction error versus k for TSVD and CGME (and CGLS).

since the kth iterate x_k^{CGME} of CGME is known to minimize the error $\|A^\dagger b - x\|_2$ within the Krylov space $\mathcal{V}_k^{\mathcal{K}}$ of (2), i.e., x_k^{CGME} is the orthogonal projection of $A^\dagger b$ onto $\mathcal{V}_k^{\mathcal{K}} = A^*(\mathcal{U}_k^{\mathcal{K}})$.

The regularizing properties of CGME have been studied in detail in [9]. Because of its optimality property, CGME will converge even faster than CGLS in general, although the optimal iterate is usually inferior to CGLS in the presence of noise. For the model problem from the previous section this is examplified in Figure 2 (the lighter shaded curve shows the CGLS history from Figure 1).

4 Hybrid methods

Both, the CGLS and the CGME iterate from $\mathcal{V}_k^{\mathcal{K}}$ can be computed via the Lanczos process. This has been realized by Paige and Saunders [18] and implemented in their code LSQR. The Lanczos process, when started with the right-hand side vector b, provides cheap recursions which generate factorizations

$$AV_k = U_{k+1} J_k^e\,, \qquad k = 1, 2, \ldots\,, \tag{7}$$

with orthogonal matrices $U_k \in \mathbb{R}^{m\times k}$ and $V_k \in \mathbb{R}^{n\times k}$ whose columns span $\mathcal{U}_k^{\mathcal{K}}$ and $\mathcal{V}_k^{\mathcal{K}}$, respectively, and where J_k^e is a lower bidiagonal $(k+1)\times k$ dimensional matrix. We denote by $J_k^o \in \mathbb{R}^{k\times k}$ the first k rows of J_k^e; J_k^o is a nonsingular matrix.

We do not require more specific properties of the factors U_k, V_k and J_k^e, except for the fact that U_k consists of the first k columns of U_{k+1}; in particular, the first column of U_k is a multiple of b, i.e.,

$$U_k e_1 = b/\beta \qquad \text{with} \qquad \beta = \|b\|_2$$

and $e_1 = [1, 0, \ldots, 0]^T$ being the first Cartesian basis vector in $\mathbb{R}^k$.

Now the CGLS iterate x_k^{CGLS} is obtained from solving the $(k+1) \times k$ dimensional least-squares problem

$$\|\beta e_1 - J_k^e z_k^{\text{CGLS}}\|_2 \;\longrightarrow\; \min, \tag{8}$$

and setting $x_k^{\text{CGLS}} = V_k z_k^{\text{CGLS}}$. Similarly, the CGME iterate x_k^{CGME} is obtained from the solution of the $k \times k$ dimensional nonsingular linear system

$$J_k^o z_k^{\text{CGME}} = \beta e_1 \tag{9}$$

via

$$x_k^{\text{CGME}} = V_k z_k^{\text{CGME}}. \tag{10}$$

Using the implementation of LSQR the computation of x_k^{CGLS} or x_k^{CGME} via (8) or (9), respectively, is only little more expensive than the corresponding conjugate gradient implementation. On the other hand, several authors (e.g., Björck [1] and O'Leary and Simmons [17]) have envisioned the possibility of using the above approach to stabilize the conjugate gradient iteration and overcome its semiconvergence. In fact, since V_k is an orthogonal matrix the divergence of x_k is equivalent to the divergence of the corresponding $\|z_k\|_2$; in other words, the divergence of the conjugate gradient method sets in when the projected problems (8) or (9) have become too ill-conditioned.

The cure of the problem might be a regularization of the small bidiagonal systems, the most promising tool being the truncated singular value decomposition (which is cheap to compute for the small dimensional problems). Because Lanczos projection and TSVD then go hand in hand, such schemes have been called *hybrid methods* in [11].

4.1 Lanczos and least-squares projection

When the least-squares problem (8) is solved with TSVD the resulting approximation is again a least-squares projection of the original system $Ax = b$. To see this we denote by

$$J_k^e V_l^e = U_l^e \Sigma_l^e, \qquad J_k^{e*} U_l^e = V_l^e \Sigma_l^e, \tag{11}$$

the truncated singular value decomposition of J_k^e, where $l \leq k$ is the truncation parameter, $U_l^e \in \mathbb{R}^{(k+1)\times l}$ and $V_l^e \in \mathbb{R}^{k\times l}$ are orthogonal matrices, and $\Sigma_l^e \in \mathbb{R}^{l\times l}$ is a diagonal matrix containing the l largest singular values of J_k^e.

The corresponding approximation x_{kl}^e of x_k^{CGLS} belongs to the subspace $\mathcal{V}_{kl}^e = \mathcal{R}(V_k V_l^e)$, namely

$$x_{kl}^e = V_k z_{kl}^e\,, \qquad z_{kl}^e = V_l^e(\Sigma_l^e)^{-1}U_l^{e*}(\beta e_1)\,. \tag{12}$$

Any vector $x \in \mathcal{V}_{kl}^e$ can be written in the form $x = V_k V_l^e w$ for some $w \in \mathbb{R}^l$, and satisfies in view of (7) and (11)

$$\begin{aligned}\|b - Ax\|_2 \;&=\; \|b - AV_kV_l^e w\|_2 \;=\; \|b - U_{k+1}J_k^eV_l^e w\|_2\\ &=\; \|U_{k+1}(\beta e_1 - U_l^e\Sigma_l^e w)\|_2 \;=\; \|\beta e_1 - U_l^e\Sigma_l^e w\|_2\,.\end{aligned}$$

It follows that the residual $b - Ax$ is minimized over $\mathcal{V}_{kl}^e$ for $x = V_k V_l^e w$ with $w = (\Sigma_l^e)^{-1}U_l^{e*}(\beta e_1)$, i.e., for the TSVD approximation in (12).

We have therefore shown that the TSVD regularized approximation x_{kl}^e of x_k^{CGLS} is a least-squares projection over $\mathcal{V}_{kl}^e$ of the original problem $Ax = b$.

4.2 Lanczos and the dual least-squares method

Now, let

$$J_k^o V_l^o = U_l^o\Sigma_l^o\,, \qquad J_k^{o*}U_l^o = V_l^o\Sigma_l^o\,, \tag{13}$$

be the corresponding truncated singular value decomposition of J_k^o. Then the TSVD approximation x_{kl}^o of x_k^{CGME} is given by

$$x_{kl}^o = V_k z_{kl}^o\,, \qquad z_{kl}^o = V_l^o(\Sigma_l^o)^{-1}U_l^{o*}(\beta e_1)\,. \tag{14}$$

x_{kl}^o belongs to the subspace $\mathcal{V}_{kl}^o = \mathcal{R}(V_k V_l^o)$. We now denote by $Q_{kl} = V_k V_l^o V_l^{o*} V_k^*$ the orthogonal projector onto $\mathcal{V}_{kl}^o$ and by $Q_k = V_k V_k^*$ the orthogonal projector onto $\mathcal{V}_k^{\mathcal{K}} = A^*(\mathcal{U}_k^{\mathcal{K}})$. We recall from Section 3 that

$$x_k^{\text{CGME}} \;=\; Q_k A^\dagger b \;=\; V_k V_k^* A^\dagger b\,.$$

On the other hand we have from (10) that $x_k^{\text{CGME}} = V_k z_k^{\text{CGME}}$, and since V_k has full column rank, this implies

$$z_k^{\text{CGME}} = V_k^* A^\dagger b\,.$$

In view of (14) and (9) we therefore conclude that

$$Q_{kl}A^\dagger b \;=\; V_k V_l^o V_l^{o*} V_k^* A^\dagger b \;=\; V_k V_l^o V_l^{o*} z_k^{\mathrm{CGME}} \;=\; V_k z_{kl}^o \;=\; x_{kl}^o\,.$$

This means that x_{kl}^o is the orthogonal projection of $A^\dagger b$ onto $\mathcal{V}_{kl}^o$, and hence, x_{kl}^o is a dual least-squares approximation of $A^\dagger b$ provided that $\mathcal{V}_{kl}^o = A^*(\mathcal{U}_{kl}^o)$ for some subset $\mathcal{U}_{kl}^o$.

To prove this latter assertion we let $\mathcal{U}_{kl}^o = \mathcal{R}(U_k U_l^o)$ and consider the matrix $A^*U_kU_l^o$ whose columns span the range $A^*(\mathcal{U}_{kl}^o)$. Since $\mathcal{V}_k^{\mathcal{K}} = A^*(\mathcal{U}_k^{\mathcal{K}})$ and the columns of U_k and V_k span $\mathcal{U}_k^{\mathcal{K}}$ and $\mathcal{V}_k^{\mathcal{K}}$, respectively, multiplication of $A^*U_kU_l^o$ with the orthogonal projector $V_kV_k^*$ onto $\mathcal{V}_k^{\mathcal{K}}$ yields

$$A^*U_kU_l^o \;=\; V_kV_k^*A^*U_kU_l^o\,. \tag{15}$$

Furthermore, we have from (7) that $U_k^*AV_k = U_k^*U_{k+1}J_k^e$; since U_{k+1} is an orthogonal matrix and U_k consists of the first k columns of U_{k+1}, there holds $U_k^*U_{k+1}J_k^e = J_k^o$, and hence, $U_k^*AV_k = J_k^o$. Inserting this into (15) we arrive at

$$A^*U_kU_l^o \;=\; V_kJ_k^{o*}U_l^o\,,$$

and therefore it follows from (13) that

$$A^*U_kU_l^o \;=\; V_kV_l^o\Sigma_l^o\,.$$

Since Σ_l^o is a nonsingular matrix this shows that $A^*(\mathcal{U}_{kl}^o) = \mathcal{R}(V_kV_l^o) = \mathcal{V}_{kl}^o$, as was to be shown.

5 Numerical results

Recent work on hybrid methods ([3, 14], see also [8, 13, 7]) has mostly been focused on the choice of appropriate regularization parameters, i.e., the stopping index k for the Lanczos process and the truncation index l for the embedded TSVD regularization. While this question is certainly the most important one for practical purposes, one should investigate first the *potential* of these methods.

This is what we shall do in the sequel by computing all respective approximations numerically for our model problem. Figure 3 illustrates in a gray-scale plot the relative errors of the least-squares projections x_{kl}^e and the dual least-squares approximations x_{kl}^o, respectively, in dependence on $k = 0, 1, 2, \ldots$ and $0 \le l \le k$. The black dots in this figure highlight the optimal value $l(k)$ for which the error $\|x^\dagger - x_{kl}\|_2$ is minimized for fixed k. From this we can distinguish three phases of the iteration:

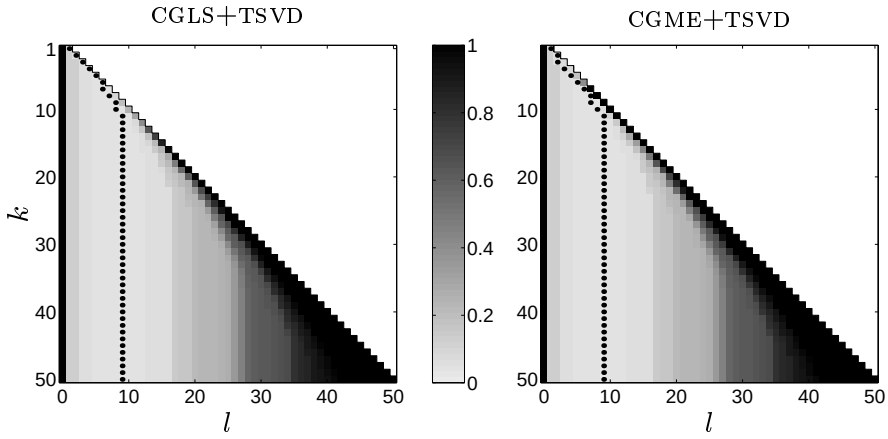

Figure 3: Relative errors of x_{kl}^e (left) and x_{kl}^o (right) versus k and l.

1. In the early stage of the iteration, i.e., for $0 \le k \le k_1$ we have $l(k) = k$, i.e., the optimal reconstruction of the hybrid scheme is the correponding conjugate gradient iterate. In our example, k_1 is also the point where the conjugate gradient iteration attains the minimal error ($k_1 = 6$ for CGLS and $k_1 = 2$ for CGME).

2. Then there is a transient region $k_1 < k < k_2$ where $l(k)$ is nondecreasing but remains below k, i.e., $k_1 = l(k_1) \le l(k) < k$.

3. In the final regime $k \ge k_2$ the parameter $l(k)$ remains constant; for both schemes this happens when $k_2 = 11$ with $l(k) = 9$ thereafter. As one might expect, the truncation parameter $l = 9$ is also optimal for TSVD applied to the full problem $Ax = b$, cf. Figure 1.

Figure 4 compares the minimal errors $\|x^\dagger - x_{k,l(k)}\|_2$ of the hybrid schemes with the errors $\|x^\dagger - x_k\|_2$ of the corresponding conjugate gradient iteration. As desired, the hybrid scheme "regularizes" the plain conjugate gradient method and maintains the error at the optimal level when the conjugate gradient iteration starts to diverge.

Due to so-called "ghosts", i.e., singular values of J_k^o or J_k^e which are doubled in the course of the iteration because of round-off errors and loss of orthogonality in the matrices U_k and V_k of (7), the behaviour of the hybrid methods is somewhat different in practice. To avoid this shortcoming we performed a full reorthogonalization of the Lanczos vectors for the preparation of these figures whenever necessary.

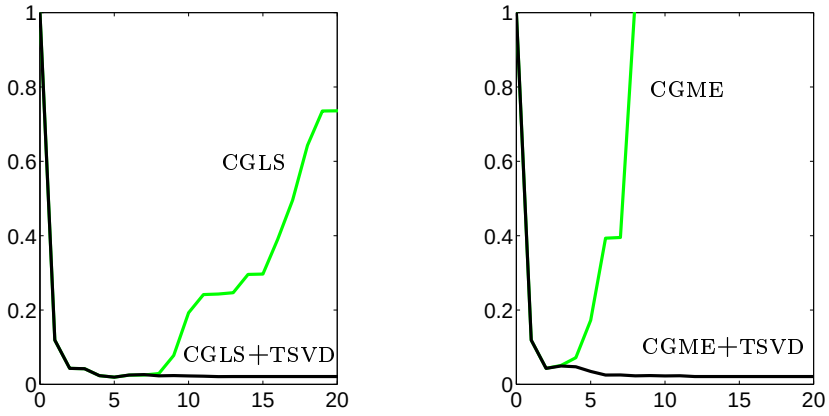

Figure 4: Relative errors of $x_{k,l(k)}$ and x_k versus k.

We mention, though, that in general the loss of orthogonality occurring in LSQR does not affect the final accuracy of the approximations, but merely slows down the iteration process; every doubling of a singular value will roughly cause one additional iteration.

6 Conclusion

We have studied hybrid methods based on the Lanczos process and TSVD regularization in terms of projection methods for ill-posed problems. The methods based on least-squares projection had already been introduced in the literature; the dual least-squares approximations have not yet been introduced, except for the basic Krylov scheme which corresponds to Craig's method CGME.

It has been shown by numerical examples (partly supported by theory) that the Krylov subspaces are more appropriate to use for a projection method than the subspaces associated with the singular value decomposition of the orginal matrix.

Further numerical results indicate that the hybrid methods can have the desired effect of stabilizing the iteration error at its minimum even when the conjugate gradient type method is diverging. At this point the optimal truncation index is close to the optimal truncation index of TSVD for the orginal problem. This indicates that parameter choice strategies which work for TSVD should also work for the hybrid scheme, provided that the Krylov subspace is so large that the minimal error is attainable.

References

[1] Å. Björck. A bidiagonalization algorithm for solving large and sparse ill-posed systems of linear equations. *BIT*, 28:659–670, 1988.

[2] Å. Björck. *Numerical Methods for Least Squares Problems.* SIAM, Philadelphia, 1996.

[3] Å. Björck, E. Grimme, and P. Van Dooren. An implicit shift bidiagonalization algorithm for ill-posed systems. *BIT*, 34:510–534, 1994.

[4] T. F. Chan and D. E. Foulser. Effectively well-conditioned linear systems. *SIAM J. Sci. Stat. Comput.*, 9:963–969, 1988.

[5] E. J. Craig. The n-step iteration procedures. *J. Math. Phys.*, 34:64–73, 1955.

[6] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems.* Kluwer Academic Publishers, Dordrecht, Boston, London, 1996.

[7] A. Frommer and P. Maaß. Fast CG-based methods for Tikhonov-Phillips regularization. *SIAM J. Sci. Comput.*, 20:1831–1850, 1999.

[8] G. H. Golub and U. von Matt. Generalized cross-validation for large-scale problems. *J. Comput. Graph. Statist.*, 6:1–34, 1997.

[9] M. Hanke. *Conjugate Gradient Type Methods for Ill-Posed Problems.* Longman Scientific & Technical, Harlow, Essex, 1995.

[10] M. Hanke. Asymptotics of orthogonal polynomials and the numerical solution of ill-posed problems. *Numer. Algorithms*, 11:203–214, 1996.

[11] M. Hanke and P. C. Hansen. Regularization methods for large-scale problems. *Surveys Math. Indust.*, 3:253–315, 1993.

[12] P. C. Hansen. Regularization tools: A MATLAB package for analysis and solution of discrete ill-posed problems. *Numer. Algorithms*, 6:1–35, 1994.

[13] L. Kaufman and A. Neumaier. Regularization of ill-posed problems by envelope guided conjugate gradients. *J. Comput. Graph. Statist.*, 6:451–463, 1997.

[14] M. E. Kilmer and D. P. O'Leary. Choosing regularization parameters in iterative methods for ill-posed problems. Manuscript, 2000.

[15] R. Kress. *Linear Integral Equations.* Springer-Verlag, New York, 2nd edition, 1999.

[16] F. Natterer. Regularisierung schlecht gestellter Probleme durch Projektionsverfahren. *Numer. Math.*, 28:329–341, 1977.

[17] 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.*, 2:474–489, 1981.

[18] C. C. Paige and M. A. Saunders. LSQR: An algorithm for sparse linear equations and sparse least squares. *ACM Trans. Math. Software*, 8:43–71, 1982.

[19] T. I. Seidman. Nonconvergence results for the application of least-squares estimation to ill-posed problems. *J. Optim. Theory Appl.*, 30:535–547, 1980.