FRIEDRICH-ALEXANDER-UNIVERSITÄT ERLANGEN-NÜRNBERG
INSTITUT FÜR INFORMATIK (MATHEMATISCHE MASCHINEN UND DATENVERARBEITUNG)

Lehrstuhl für Informatik 10 (Systemsimulation)

Numerical Solution of Least-Squares Problems by a Modified Kovarik Algorithm

M. Mohr, C. Popa and U. Rüde

Technical Report 03-7

Abstract

In a previous paper we presented a modified version of Kovarik's approximate orthogonalisation algorithm for arbitrary symmetric matrices. In the present one we apply this algorithm for numerical solution of linear least-squares problems with a symmetric problem matrix. The basic idea is to modify also the right hand side of the problem during the transformation of the matrix. We prove that the sequence of vectors generated in this way converges to the minimal norm solution of the problem. Some applications are also presented for a collocation discretisation of a "model problem" first kind integral equation.

AMS Subject Classification: 65F10, 65F20, 65R20

Key words and phrases: linear least-squares problems, minimal norm solution, orthogonalisation algorithms, symmetric matrices, collocation, first kind integral equations.

1 Modified Kovarik algorithm for symmetric matrices

Let A be an $n \times n$ symmetric matrix, $(A)_i$ its i-th row and $A^\dagger$ its Moore-Penrose pseudoinverse (see [2]). By $gk_2(A)$ we shall denote its generalised spectral condition number defined as the square root of the ratio between the largest and smallest singular value. $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ will be the Euclidean scalar product and norm on some space $\mathbb{R}^q$ and P_S will denote the orthogonal projection onto a vector subspace $S \subset \mathbb{R}^q$. For a square matrix B, $\sigma(B)$, $\rho(B)$ and $\| B \|$ will denote its spectrum, spectral radius and spectral norm respectively. In our previous paper [7] we considered the following version of Kovarik's algorithm (B) from [6].

Algorithm 1 (KOBS) *Let $A_0 = A$ be a general symmetric matrix. Then for $k = 0, 1, \ldots,$ we construct a sequence (A_k) via*

$$K_k = (I - A_k)(I + A_k)^{-1} \ , \quad A_{k+1} = (I + K_k)A_k \ . \tag{1}$$

Let E be the subset of real numbers defined by

$$E := \left\{ -\frac{1}{\alpha_p} \ \middle| \ \alpha_0 = 1, \alpha_{p+1} = 2\alpha_p + 1, p \geqslant 0 \right\} \tag{2}$$

The following results concerning the convergence properties of the above algorithm were proved in [7].

Theorem 1 *Let us suppose that none of the eigenvalues of the symmetric matrix A lies in the set E from (2). Then the sequence of matrices $(A_k)_{k \geqslant 0}$ generated by the above algorithm **KOBS** converges and*

$$\lim_{k \to \infty} A_k = A^\dagger A. \tag{3}$$

Corollary 1 *Under the conditions of the above theorem algorithm **KOBS** exhibits the following properties.*

(i) If

$$\sigma(A) \cap (0, 1) = \emptyset, \tag{4}$$

then it exists an integer $k_0 \geqslant 0$ such that

$$\| A_k - A^\dagger A \| \ \leq \ \left(\frac{1}{2}\right)^{k-k_0} \| A_{k_0} - A^\dagger A \|, \ \forall k \geqslant k_0. \tag{5}$$

(ii) If

$$\sigma(A) \subset [0, 1] \tag{6}$$

and $\delta > 0$ is given by

$$\delta = \frac{1}{1 + \lambda_{min}(A)}, \tag{7}$$

where $\lambda_{min}(A) \in (0, 1)$ is the smallest positive eigenvalue of A, then

$$\| A_k - A^\dagger A \| \ \leq \ \delta^k, \ \forall k \geqslant 0. \tag{8}$$

Remark 1 *The above corollary tells us about the linear convergence rate of the algorithm* ***KOBS****. Moreover, in the situation that A is positive semi-definite and has eigenvalues both smaller and larger than 1, the latter will converge much more quickly than the former. This aspect might be of interest in so called discrete ill-posed problems coming from the discretisation of ill-posed inverse problems, since it constitute an implicit regularisation of a sort (see [5] in this respect).*

Remark 2 *For the sequence $(A_k)_{k\geqslant 0}$ generated with the algorithm* ***KOBS*** *it can be proved that*

$$\lim_{k\to\infty} gk_2(A_k) = gk_2(A^\dagger A) = 1, \tag{9}$$

i.e. it acts as an "iterative preconditioner" for A. From this point of view we can "combine" it with some other direct or iterative solvers for a problem in which A is the "system matrix" (see in this sense [3]). However, as we shall see in the next section of the paper, we can also directly use the ***KOBS*** *algorithm for the iterative numerical solution of a large class of problems – linear least-squares problems with a symmetric matrix.*

2 Numerical solution of least-squares problems

For a symmetric $n \times n$ matrix A and a given vector $b \in I\!R^n$ we consider the linear least-squares problem: find $x^* \in I\!R^n$ such that

$$\| Ax^* - b \| = \min! \tag{10}$$

Let $LSS(A;b)$ be the set of all least-squares solutions of (10) and x_{LS} the unique minimal norm one. In the case that (10) is consistent we shall denote $LSS(A;b)$ by $S(A;b)$. In order to apply the above algorithm **KOBS** for the numerical solution of (10) we consider the following formulation of it, in which also the right hand side b is modified.

Algorithm 2 (KOBS with rhs) *Set $A_0 = A$ and $b^0 = b$. Now for $k = 0, 1, \ldots$ we construct two sequences (A_k) and (b^k) via*

$$K_k = (I - A_k)(I + A_k)^{-1},\ A_{k+1} = (I + K_k)A_k,\ b^{k+1} = (I + K_k)b^k\ . \tag{11}$$

For the rest of the paper we will assume that none of the eigenvalues of A lies in the above set, i.e. $\sigma(A) \cap E = \emptyset$. Under this assumption the following lemma can be proved.

Lemma 1 *For any $k \geqslant 0$ the matrix A_k is symmetric.*

Proof: Observe firstly that for any $k \geqslant 0$ the matrix $I + A_k$ is invertible, because -1 is not among the eigenvalues of A. See in this sense (2) and the proof of Theorem 2 in [7]. Thus, we have the equalities

$$I + K_k = I + (I - A_k)(I + A_k)^{-1} = [(I + A_k) + (I - A_k)]\,(I + A_k)^{-1} = 2(I + A_k)^{-1} \tag{12}$$

and

$$A_k(I + A_k) = (I + A_k)A_k \quad \Leftrightarrow \quad (I + A_k)^{-1}A_k = A_k(I + A_k)^{-1}\ . \tag{13}$$

Since the left hand side of (13) is always true, we get by combining (12) and (13)

$$A_kK_k = K_kA_k\ ,\quad \forall k \geqslant 0\ . \tag{14}$$

We shall now prove the lemma by mathematical induction. The initial matrix $A_0 = A$ is symmetric by our general hypothesis. Assume then that A_k is symmetric for all $0 \leqslant k \leqslant j$. From (12) we obtain that K_j is symmetric and then, by also using (11) and (14), we get

$$A_{j+1}^T = A_j^T(I + K_j^T) = A_k(I + K_j) = (I + K_j)A_k = A_{j+1} \tag{15}$$

which means that A_{j+1} is also symmetric and the proof is complete.

□

We are now able to prove the two main results of the paper.

Theorem 2 *If the problem (10) is consistent, i.e.*

$$b \in R(A) \tag{16}$$

then the sequence $(b^k)_{k\geqslant 0}$ *from (11) converges and*

$$\lim_{k\to\infty} b^k = A^\dagger b = x_{LS} \ . \tag{17}$$

Proof: From (16) it results that

$$b = Ax \ , \tag{18}$$

for some $x \in I\!R^n$. Then, by also using (11), we obtain

$$b^1 = (I + K_0)b^0 = (I + K_0)A_0x = A_1x$$

and by an induction argument

$$b^k = A_kx \ , \quad \forall k \geqslant 0 \ . \tag{19}$$

Then we get from (19), (18) and (3)

$$\lim_{k\to\infty} b^k = \left(\lim_{k\to\infty} A_k\right) x = A^\dagger Ax = A^\dagger b = x_{LS}$$

and the proof is complete.

□

Theorem 3 *In the case that (10) is inconsistent we have*

$$\lim_{k\to\infty} A_kb^k = x_{LS} \ . \tag{20}$$

Proof: Since A is symmetric we have $N(A^T) = N(A)$. Thus,

$$b = P_{R(A)}(b) + P_{N(A)}(b) \ , \tag{21}$$

with

$$P_{R(A)}(b) = Ax \tag{22}$$

for some $x \in I\!R^n$. Let $r = \text{rank}\,(A) < n$ and Q an $n \times n$ orthogonal matrix such that

$$A = Q\,\text{diag}(\lambda_1^{(0)}, \ldots, \lambda_r^{(0)}, 0, \ldots, 0)\,Q^T \ , \tag{23}$$

where $\lambda_i^{(0)}$ are the nonzero eigenvalues of A. Then, as in [7], we obtain

$$A_k = Q\,\text{diag}(\lambda_1^{(k)}, \ldots, \lambda_r^{(k)}, 0, \ldots, 0)\,Q^T \ .$$

Combining this with (12) we see that

$$I + K_k = Q\,\text{diag}\left(\frac{2}{1+\lambda_1^{(k)}}, \ldots, \frac{2}{1+\lambda_r^{(k)}}, 2, \ldots, 2\right) Q^T \ . \tag{24}$$

On the other hand we know that

$$A^\dagger = Q\,\text{diag}\left(\frac{1}{\lambda_1^{(0)}}, \ldots, \frac{1}{\lambda_r^{(0)}}, 0, \ldots, 0\right) Q^T \ , \tag{25}$$

and thus, see e.g. [2], we have

$$P_{N(A)} = I - P_{R(A)} = I - A^\dagger A = Q\,\text{diag}(0, \ldots, 0, 1, \ldots, 1)\,Q^T. \tag{26}$$

From (26) and (24) we get that

$$(I + K_k)P_{N(A)} = Q\,\text{diag}(0, \ldots, 0, 2, \ldots, 2)\,Q^T = 2P_{N(A)} \ . \tag{27}$$

Now, from (11), (21), (22) and (27) we get

$$\begin{aligned} b^1 &= (I+K_0)b^0 = (I+K_0)b \\ &= (I+K_0)P_{R(A)}(b) + (I+K_0)P_{N(A)}(b) \\ &= (I+K_0)Ax + 2P_{N(A)}(b) \ . \end{aligned}$$

By a recursive argument and also considering (24), (26), (18), (19) (for $P_{R(A)}(b)$ instead of b) and (22) we obtain

$$b^k = A_k x + 2^k P_{N(A)}(b) \ , \quad \forall k \geqslant 0 \ . \tag{28}$$

But from (11) and the invertibility of the matrix $I + A_k$ (see Lemma 1) we have that

$$N(A_k) = N(A) \ , \quad \forall k \geqslant 0 \ ,$$

which together with (28) gives us

$$A_k b^k = A_k A_k x = A_k^2 x \ , \quad \forall k \geqslant 0 \ . \tag{29}$$

From (22), (3), (29) and the definition of the Moore-Penrose pseudoinverse $A^\dagger$ (see e.g. [2]) we obtain

$$\begin{aligned} \lim_{k\to\infty} A_k b^k &= \lim_{k\to\infty} A_k A_k x \\ &= (A^\dagger A)(A^\dagger A)x \\ &= (A^\dagger A A^\dagger)(Ax) \\ &= A^\dagger A x \\ &= A^\dagger P_{R(A)}(b) \\ &= x_{LS} \end{aligned}$$

and the proof is complete.

□

Remark 3 *The result in (20) does not contradict Theorem 1. Indeed, if (10) is consistent, both relations (17) and (20) hold.*

Remark 4 *In the case that (10) is inconsistent relation (17) is no longer true. Indeed, if* $P_{N(A)}(b) \neq 0$*, we obtain from (28)*

$$\lim_{k\to\infty} \| \, b^k \, \| = \infty \ . \tag{30}$$

We shall finish this section of the paper with some computational considerations concerning the above algorithm **KOBS with rhs** (11). Let $k \geqslant 0$ be arbitrary, but fixed. In this case we get from (12) that $A_k = 2(I + K_k)^{-1} - I$. Thus, by also using (11), we have

$$\begin{aligned} I + K_{k+1} &= 2(I + A_{k+1})^{-1} = 2[I + (I + K_k)A_k]^{-1} \\ &= 2[I + (I + K_k)(2(I + K_k)^{-1} - I)]^{-1} = 2(2I - K_k)^{-1} \ . \end{aligned}$$

Then, by eliminating A_k, the algorithm (11) can be written as follows.

Algorithm 3 (KOBSb_con) *Set* $b^0 = b$ *and* $K_0 = 2(I + A)^{-1} - I$*. For* $k = 0, 1, 2, \ldots$ *construct a sequence* (b^k) *via*

$$b^{k+1} = (I + K_k)b^k \ , \quad K_{k+1} = 2(2I - K_k)^{-1} - I \ . \tag{31}$$

It immediately results from Theorem 2 that in the consistent case (16) the sequence $(b^k)_{k\geqslant 0}$ generated by (31) converges to x_{LS}.

Such a transformation is especially helpful in the inconsistent case since it allows to avoid potential computational problems due to the property (30), if the sequence $(b^k)_{k\geqslant 0}$ is generated in the iterations (11). We start from the version (31) and, using (20), define the following algorithm.

Algorithm 4 (KOBSb_incon) *Set* $\beta^0 = Ab$ *and* $K_0 = 2(I + A)^{-1} - I$. *For* $k = 0, 1, 2, \ldots$ *construct a sequence of vectors* β^k *via*

$$\beta^{k+1} = (I + K_k)^2 \beta^k \ , \quad K_{k+1} = 2(2I - K_k)^{-1} - I \ . \tag{32}$$

In order to see that the above algorithm solves the inconsistent least-squares problem, note that, from (11) and (14) we obtain

$$A_{k+1} b^{k+1} = (I + K_k) A_k (I + K_k) b^k = (I + K_k)^2 A_k b^k \ ,$$

which tells us that the sequence $(\beta^k)_{k \geqslant 0}$ from (32) satisfies

$$\beta^k = A_k b^k \ , \quad \forall k \geqslant 0 \ . \tag{33}$$

Thus, from Theorem 3 we obtain that the sequence $(\beta^k)_{k \geqslant 0}$ generated by (32) converges to x_{LS} in the inconsistent case for (10).

3 Numerical experiments

We considered in our tests the following first kind integral equation: for a given function $y \in L^2([0,1])$, find $x^* \in L^2([0,1])$ such that

$$\int_0^1 k(s,t) x(t) dt \ = \ y(s), \ s \in [0,1], \tag{34}$$

with

$$k(s,t) = \frac{1}{1 + |s - 0.5| + t}, \ y(s) = \begin{cases} \ln \frac{2.5 - s}{1.5 - s}, & s \in [0, 0.5) \\ \\ \ln \frac{1.5 + s}{0.5 + s}, & s \in [0.5, 1] \end{cases} \tag{35}$$

Remark 5 *The right hand side* y *was computed as in (35) such that the equation (34) has the solution* $x(t) = 1, \forall t \in [0,1]$.

We discretised (34)-(35) by the collocation algorithm from [8], with the collocation points

$$s_i = (i-1) \frac{1}{n-1}, \ i = 1, 2, \ldots, n. \tag{36}$$

Thus, we obtained the symmetric system

$$Ax = b, \tag{37}$$

with the $n \times n$ matrix A and $b \in I\!R^n$ given by

$$A_{ij} \ = \ \int_0^1 k(s_i, t) k(s_j, t) dt = \begin{cases} \frac{1}{\alpha_i (1 + \alpha_i)}, & \text{if } \alpha_i = \alpha_j, \\ \\ \frac{1}{\alpha_i - \alpha_j} \ln \frac{(1 + \alpha_j) \alpha_i}{(1 + \alpha_i) \alpha_j}, & \text{if } \alpha_i \neq \alpha_j, \end{cases} \tag{38}$$

$$b_i \ = \ y(s_i) \ , \tag{39}$$

where for the index pair (i, j) the two values α_i and α_j are given by

$$\alpha_p = 1 + |s_p - \frac{1}{2}| \ . \tag{40}$$

Remark 6 *For $n \geqslant 3$ we observe that the matrix A from (38) is positive semi-definite with*

$$rank(A) = \begin{cases} \frac{n+1}{2} \ , & \text{if } n \text{ is odd} \\ \\ \frac{n}{2} \ , & \text{if } n \text{ is even} \ . \end{cases} \tag{41}$$

Thus

$$\lim_{n \to \infty} rank(A) = \infty \tag{42}$$

and the least-squares solution $x^n_{LS} = (x^n_{LS,1}, \ldots, x^n_{LS,n})^T \in I\!R^n$ of (37) approximates the (continuous) least-squares solution of the initial integral equation (34), denoted here by $X_{LS} \in L^2([0,1])$, according to the extension that was obtained in [9]. More precisely, if we denote by $X^n_{app,LS}$ the approximate least-squares solution of (34) defined by

$$X^n_{app,LS}(t) = \sum_{j=1}^{n} x^n_{LS,j} \ k(s_j, t) \ , \tag{43}$$

then the following convergence result was proved in [9] (by also taking into account (42))

$$\lim_{n \to \infty} \parallel X_{LS} - X^n_{app,LS} \parallel_{L^2([0,1])} = 0 \ , \tag{44}$$

where $\parallel f \parallel_{L^2([0,1])} = \sqrt{\int_0^1 |f(t)|^2 dt}$.

First of all we have to observe that, because the problem (34)- (35) is consistent (see Remark 5) and by using Lemma 2 in [9] it results that the system (37) is also consistent. We then applied the algorithm (11), for different values of n with the "residual" stopping rule (see Theorem 2)

$$\parallel Ab^k - b \parallel \leqslant 10^{-5} \ . \tag{45}$$

The corresponding numbers of iterations are described in Table 1, in which we have also indicated the absolute (abserr) and relative (relerr) errors (with respect to the exact minimal norm solution of (37) x_{LS}), defined by

$$\text{abserr} = \parallel b^k - x_{LS} \parallel \ , \quad \text{relerr} = \frac{\text{abserr}}{\parallel b^k \parallel} \ . \tag{46}$$

We then considered a perturbation of the right hand side b of (37) of the form

$$\tilde{b} = b + \delta b \ , \tag{47}$$

with $\delta b \in I\!R^n$ a randomly generated vector such that $\parallel \delta b \parallel = 5\% \parallel b \parallel$. We then considered the (inconsistent) least-squares formulation

$$\parallel Ax - \tilde{b} \parallel = \min! \tag{48}$$

and applied to it the algorithm (11), for the same values of n but with the "normal equation" stopping test (according to Theorem 3)

$$\parallel A \left(A(A_k b^k) - \tilde{b} \right) \parallel \leqslant 10^{-5}. \tag{49}$$

The results are given in Table 2.

Remark 7 *According to the results in Tables 1 and 2 we have to observe the "mesh independent" behaviour of the algorithm (11). Moreover, in Table 2 the absolute error abserr has larger values than in Table 1. This is due to the fact that in the inconsistent case the ill-conditioning and large "rank-deficiency", see (41), of the matrix A have a bigger influence on the computed solution. Work is in progress on analysing this in detail for the algorithm* ***KOBS with rhs*** *and will be reported in a follow-up paper.*

Table 1. Results for the consistent system (37)			
n	Iter. for (45)	*abserr*	*relerr*
8	18	10^{-5}	10^{-5}
16	18	10^{-4}	10^{-4}
32	19	10^{-4}	10^{-4}
64	19	10^{-4}	10^{-4}
128	20	10^{-4}	10^{-4}

Table 2. Results for the inconsistent problem (48)			
n	Iter. for (49)	*abserr*	*relerr*
8	20	0.00001	10^{-5}
16	22	0.00022	10^{-5}
32	23	0.00072	10^{-5}
64	25	0.03254	10^{-3}
128	27	0.51348	10^{-3}

Note. All the computations were made with the software package OCTAVE, freely available under the terms of the GNU General Public License, see `www.octave.org`.

Acknowledgements

The paper was supported by NATO through collaborative linkage grant PST.CLG.977924 and by the DAAD via a grant that one of the authors had as a visiting professor at the Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany, in the period October 2002-August 2003.

References

[1] Björck, A., *Numerical methods for least squares problems*, SIAM, Philadelphia, 1996.

[2] Boullion, L. T. and Odell, P. L., *Generalized inverse matrices*, Wiley-Interscience, New York, 1971.

[3] Evans, D. J. and Popa, C. , *Projections and preconditioning for inconsistent least-squares problems*, Intern. J. Computer Math., **78(4)**(2001), 599-616.

[4] Golub, G. H. and van Loan, C. F., *Matrix computations*, The John's Hopkins Univ. Press, Baltimore, 1983.

[5] Hansen, P. C., *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*, SIAM Monographs on Mathematical Modeling and Computation, SIAM, 1998.

[6] Kovarik, S., *Some iterative methods for improving orthogonality*, SIAM J. Num. Anal., **7(3)**(1970), 386-389.

[7] Mohr, M., Popa C., Rüde U., *An Iterative Algorithm for Approximate Orthogonalisation of Symmetric Matrices*, Lehrstuhlbericht **03-2**, Lehrstuhl für Informatik 10 (Systemsimulation), Friedrich-Alexander-Universität Erlangen-Nürnberg.

[8] Nashed, M. Z. and Wahba, G,, *Convergence rates of approximate least squares solutions of linear integral and operator equations of the first kind*, Math. of Comput., **28(125)**(1974), 69-80.

[9] Pelican, E., Popa C., *Some Remarks on a Collocation Method for First Kind Integral Equations*, Lehrstuhlbericht **03-1**, Lehrstuhl für Informatik 10 (Systemsimulation), Friedrich-Alexander-Universität Erlangen-Nürnberg.