A Class of Modifications for Kovarik's Method

H. Esmaeili

Abstract

The approximate orthogonalization method for a finite set of linearly independent vectors was introduced by Z. Kovarik in which it is necessity to compute explicitly the inverse of a matrix in every iteration. It is proved that Kovarik's method converges quadratically. Several modifications have been proposed for Kovarik's method, all of which try to eliminate the necessity of explicit computation of the inverse. Most of these methods are linear convergent. The best modification, with a good convergent behaviour, is Petcu and Popa's, although they did not express any satisfactory reason for the origin of this modification. In this paper, we present a class of modifications for Kovarik's method which consists of Petcu and Popa's method. We prove that the methods from this class are, generally, linear convergent, while, only for the special case of the Petcu and Popa's method, it is quadratic convergent. Therefore, we show that Petcu and Popa's method, in contrast with their claim, is not linear but quadratic convergent, turning it into an optimal method in this class.

1 Introduction

Z. Kovarik [2] proposed his algorithm for approximate orthogonalization of a finite linearly independent set of vectors from a Hilbert space. His algorithm is some kind of iterative version of the classical Gram-Schmidt one and also some of its direct applications have been derived for variational finite element formulation of elliptic problems and least squares. Kovarik showed that the approximate orthogonalization method has quadratic convergence. The main difficulty with this method is the

Received by the editors April 2007.
Communicated by A. Bultheel.
1991 *Mathematics Subject Classification :* 65F20, 65F25.
Key words and phrases : Approximate Orthogonalization Method, Linear and Quadratic Convergence.

necessity of computing the inverse of a matrix explicitly in every iteration. Many years after Kovarik, Popa [4] adapted and extended his algorithm for a set of arbitrary vectors in $\mathcal{R}^n$, and proved that the transformed matrix columns, in addition to rows, are "quasi-orthogonal".

Suppose that $m \leq n$ and A is a $m \times n$ matrix of rank r. Kovarik's method is given by the following iterations:

$$A_0 = A, \quad K_k = (I - A_k A_k^T)(I + A_k A_k^T)^{-1}, \quad A_{k+1} = (I + K_k)A_k, \quad k \geq 0. \tag{1}$$

We note that if

$$\|A_k A_k^T\|_2 < 1, \tag{2}$$

then the matrix $I + A_k A_k^T$ will be invertible and vice versa [1]. In particular, for $k = 0$, (2) is equivalent to:

$$\|AA^T\|_2 < 1. \tag{3}$$

If (3) is true, then it can be proved that (2) satisfies, for all $k \geq 1$ [4]. Moreover, assumption (3) is not restrictive and can be obtained by an appropriate scaling of matrix A. Therefore, without loss of generality, we assume that A satisfies in (3).

The following result was proved in [2]:

Theorem 1. If the rows of A are linearly independent, and if

$$A_\star = \left[(AA^T)^{1/2}\right]^{-1} A,$$

then,

(a) the matrix $A_\star$ has mutually orthogonal rows;

(b) the sequence $\{A_k\}$ defined by (1) converges to $A_\star$. Moreover,

$$\|K_0\|_2 < 1$$

and

$$\|A_\star - A_k\|_2 \leq \|K_0\|_2^{2^k}, \qquad \forall\ k \geq 1. \tag{4}$$

Relation (4) tells us that the sequence generated by Kovarik's method has quadratic convergence.

We note that since the rows of A are linearly independent, the associated Gram matrix AA^T is symmetric and positive definite, so $A_\star$ is well defined. On the other hand, if the rows of A are not linearly independent, the matrix $(AA^T)^{1/2}$ still exists, but is no longer invertible. Thus, instead of $A_\star$, we have to consider its "natural" generalization A_∞ defined by

$$A_\infty = \left[(AA^T)^{1/2}\right]^{+} A,$$

where B^+ is the Moore-Penrose pseudoinverse of B (see [1]). Popa [4] proved that in this case the sequence $\{A_k\}$ converges to A_∞ and the rows of A_∞ are "quasi-orthogonal".

Despite the quadratic convergence of Kovarik's algorithm, there is a difficult computational aspect related to the matrix inversion in (1) at each of its iterations.

Several modifications have been proposed for Kovarik's method, all of which try to eliminate the necessity to explicitly compute the inverse. These are upon using some approximations for $(I + A_k A_k^T)^{-1}$, which are based on Taylor's series of particular functions and are at least linear convergent [3,4]. Specially, in [3] Petcu and Popa introduced a modification with a good convergent behaviour and showed that it is linear convergent. Using numerical tests, they showed that their method converges rapidly. In addition, they showed, computationally, that their method is superior to the other ones in cost and the number of iterations. Unfortunately, they did not express any satisfactory reason for the origin of their method.

In this paper, we introduce a class of modifications for Kovarik's method that includes also Petcu and Popa's method. This class is based on a special quadratic interpolation (and in a special case, linear interpolation). We will prove in what follows that linear convergence is achieved in general, and, moreover, quadratic convergence, in the special case of the method proposed by Petcu and Popa. Therefore, we show that Petcu and Popa's method, in contrast with their claim, is not linearly but quadratically convergent, turning it into an optimal method in this class.

2 A Class of Modifications for Kovarik's Method

In this section, we describe a class of modifications for Kovarik's method which contains the Petcu and Popa's method. To this end, it is necessary to consider the convergence behaviour of Kovarik's method. The examination of the convergence of Kovarik's method leads to the examination of the convergence of a real numbers sequence

$$\sigma_j^{(k+1)} = \left[1 + \frac{1 - (\sigma_j^{(k)})^2}{1 + (\sigma_j^{(k)})^2}\right] \sigma_j^{(k)}, \tag{5}$$

(see [4]). Here, $\sigma_j^{(k)}$, $j = 1, \ldots, r$, $k \geq 0$, are the singular values of the matrix A_k. Various modifications of Kovarik's method are obtained by some approximations of $1/(1 + (\sigma_j^{(k)})^2)$ (and therefore, $(I + A_k A_k^T)^{-1}$). For example, in [3] the quantity $1/(1 + (\sigma_j^{(k)})^2)$ was approximated by $1 - 0.5(\sigma_j^{(k)})^2$, that leads to the sequence

$$\sigma_j^{(k+1)} = \left[1 + \left(1 - (\sigma_j^{(k)})^2\right)\left(1 - 0.5(\sigma_j^{(k)})^2\right)\right] \sigma_j^{(k)} \tag{6}$$

and the corresponding variation of the Kovarik's method:

$$K_k = (I - A_k A_k^T)(I - 0.5 A_k A_k^T), \quad A_{k+1} = (I + K_k) A_k, \quad k \geq 0. \tag{7}$$

However, there is no satisfactory reason for the origin of this choice. It has been shown that (7) is linear convergent but we will prove in what follows that the order of convergence is, indeed, quadratic.

We present now a class of modifications for Kovarik's method, based on a special quadratic interpolation (or linear interpolation, in a special case). To clarify the issue, we simplify (5) as the following:

$$\sigma_j^{(k+1)} = \frac{2\sigma_j^{(k)}}{1 + (\sigma_j^{(k)})^2}. \tag{5'}$$

Consider the function

$$f(t) = \frac{1}{1+t}, \quad 0 \le t \le 1.$$

We know that $f(0) = 1$ and $f(1) = 0.5$. We are going to approximate this function with an as possible as "good" quadratic polynomial passing through the points (0,1) and (1,0.5). Suppose that this polynomial is

$$p(t) = a_0 + a_1 t + a_2 t^2.$$

Since $p(0) = 1$ and $p(1) = 0.5$, we will have

$$a_0 = 1, \quad a_2 = -0.5 - a_1.$$

Therefore, $p(t)$ is

$$p(t) = 1 + a_1 t - (a_1 + 0.5)t^2,$$

where a_1 is a parameter. Due to the existence of the parameter a_1 in $p(t)$, a class of approximations for $f(t)$ is obtained. Different choices for a_1 lead to different modifications for Kovarik's method. For example, if we choose a_1 so that

$$\int_0^1 f(t)\,dt = \int_0^1 p(t)\,dt,$$

then we will obtain $a_1 \approx -0.841$, leading to the sequence

$$\begin{aligned} \sigma_j^{(k+1)} &= 2p((\sigma_j^{(k)})^2)\sigma_j^{(k)} \\ &= \left[2 - 1.682(\sigma_j^{(k)})^2 + 0.682(\sigma_j^{(k)})^4\right]\sigma_j^{(k)} \\ &= \left[1 + \left(1 - (\sigma_j^{(k)})^2\right)\left(1 - 0.682(\sigma_j^{(k)})^2\right)\right]\sigma_j^{(k)} \end{aligned} \tag{8}$$

and the corresponding variation of the Kovarik's method:

$$K_k = (I - A_k A_k^T)(I - 0.682 A_k A_k^T), \quad A_{k+1} = (I + K_k)A_k, \quad k \ge 0. \tag{9}$$

Also, we can choose a_1 in such a way that least squares error

$$\int_0^1 \left(f(t) - p(t)\right)^2 dt$$

is minimal. With some calculations, we obtain $a_1 \approx -0.848$, leading to the sequence

$$\begin{aligned} \sigma_j^{(k+1)} &= 2p((\sigma_j^{(k)})^2)\sigma_j^{(k)} \\ &= \left[2 - 1.696(\sigma_j^{(k)})^2 + 0.696(\sigma_j^{(k)})^4\right]\sigma_j^{(k)} \\ &= \left[1 + \left(1 - (\sigma_j^{(k)})^2\right)\left(1 - 0.696(\sigma_j^{(k)})^2\right)\right]\sigma_j^{(k)} \end{aligned} \tag{10}$$

and the corresponding variation of the Kovarik's method:

$$K_k = (I - A_k A_k^T)(I - 0.696 A_k A_k^T), \quad A_{k+1} = (I + K_k)A_k, \quad k \ge 0. \tag{11}$$

The convergence of these modifications will be examined later.

To make $p(t)$ a good approximation for $f(t)$, we choose the parameter a_1 such that $p(t)$ is near $l(t)$, where

$$l(t) = 1 - 0.5t$$

is the chord connecting the points $(0, 1)$ and $(1, 0.5)$. Since $f(t) \leq l(t)$, for all $t \in [0, 1]$, and since $p(t)$ must be a good approximation for $f(t)$, we have to impose $p(t) \leq l(t)$. Therefore,

$$|l(t) - p(t)| = l(t) - p(t) = (a_1 + 0.5)(t^2 - t), \quad \forall\ t,\ 0 \leq t \leq 1.$$

On the other hand, since $t^2 \leq t$, for all $t \in [0, 1]$, we must have $a_1 + 0.5 < 0$ or $a_1 < -0.5$.

The special choice of $a_1 = -0.5$ leads $p(t)$ to decrease to $l(t)$, so that the iterative scheme

$$\sigma_j^{(k+1)} = 2\left[1 - (\sigma_j^{(k)})^2\right]\sigma_j^{(k)}$$

and the following sequence

$$A_{k+1} = 2(I - A_k A_k^T)A_k, \quad k \geq 0$$

is obtained. The above mentioned method is the same as that obtained by using the first two terms of Neumann's series [3] (also, see (5′))

$$(I + A_k A_k^T)^{-1} = I - A_k A_k^T + (A_k A_k^T)^2 - (A_k A_k^T)^3 + \cdots .$$

Knowing the restrictions on the parameter a_1 in $p(t)$ (namely, a_1 must satisfy $a_1 + 0.5 < 0$), we can formulate the general iterations

$$\begin{aligned} \sigma_j^{(k+1)} &= 2p((\sigma_j^{(k)})^2)\sigma_j^{(k)} \\ &= \left[2 + a_1(\sigma_j^{(k)})^2 - 2(a_1 + 0.5)(\sigma_j^{(k)})^4\right]\sigma_j^{(k)} \\ &= \left[1 + \left(1 - (\sigma_j^{(k)})^2\right)\left(1 - \alpha(\sigma_j^{(k)})^2\right)\right]\sigma_j^{(k)} \end{aligned} \tag{12}$$

leading to the modified class of methods

$$K_k = (I - A_k A_k^T)(I - \alpha A_k A_k^T), \quad A_{k+1} = (I + K_k)A_{k+1}, \quad k \geq 0 \tag{13}$$

that are similar to Kovarik's method. Here, $\alpha = -2a_1 - 1$. Since $a_1 + 0.5 < 0$, then $\alpha > 0$. We note that the choices of $\alpha = 0.5$, $\alpha = 0.682$, and $\alpha = 0.696$, give us methods (7), (9), and (11), respectively. Therefore, there is a family of modifications for Kovarik's method including the variant proposed by Petcu and Popa.

In the next section, we show that if the parameter α is chosen in a special interval, then the class (13) is always convergent and, in general, the order of its convergence is linear. However, only in the special case of Petcu and Popa's method the order of convergence is quadratic. Therefore, we show that the Petcu and Popa's method is an optimal method in the class of (13).

3 Study of the Convergency

To determine the optimal value of $\alpha > 0$ in the sense of convergence, we should first examine the convergence of the sequence

$$x_{k+1} = h(x_k), \quad k \geq 0 \tag{14}$$

where

$$h(x) = \Big(1 + (1 - x^2)(1 - \alpha x^2)\Big)\, x.$$

The above sequence starts from an initial approximation $x_0 \in (0, 1]$. If there is an x^* such that $x^* = \lim_{k\to\infty} x_k$, then x^* is a fixed point of function $h(x)$ and we have

$$x^* = \Big(1 + (1 - (x^*)^2)(1 - \alpha(x^*)^2)\Big)\, x^*.$$

Therefore,

$$x^*(1 - (x^*)^2)(1 - \alpha(x^*)^2)) = 0$$

which results in

$$x^* \in \{0,\, \pm 1,\, \pm 1/\sqrt{\alpha}\}.$$

First of all, we impose some conditions on α under which the sequence (14) is convergent. For this, we find an interval $[0, b]$, including $(0, 1]$, for which $h : [0, b] \longrightarrow [0, b]$ and the sequence (14) is convergent to $x^* = 1$ for any approximation value $x_0 \in (0, b]$.

For $h(b) \leq b$, we will have $1-(1+\alpha)b^2+\alpha b^4 \leq 0$, which is a quadratic polynomial in terms of b^2 with the following roots:

$$b^2 = \frac{(1 + \alpha) \pm |1 - \alpha|}{2\alpha}.$$

Now one can simply find that

$$\begin{cases} b \in [1,\, 1/\sqrt{\alpha}] & 1 - \alpha \geq 0 \\ b \in [1/\sqrt{\alpha},\, 1] & 1 - \alpha < 0. \end{cases}$$

On the other hand, $[0, b]$ must include $(0, 1]$ and, moreover, $x^* = 1$ is the unique fixed point of $h(x)$ in $(0, b)$; therefore, we must choose the case $b \in (1,\, 1/\sqrt{\alpha})$, corresponding to $1 - \alpha > 0$, which means α must belong to $(0, 1)$.

For the above mentioned values of α and b, it is clear that

$$\begin{cases} h(x) > x & \forall\, x \in (0, 1) \\ h(x) < x & \forall\, x \in (1, b) \end{cases} \tag{15}$$

and also $|h'(1)| = |2\alpha - 1| < 1$, so $h(x)$ is a contraction mapping in a vicinity of $x^* = 1$.

In what follows, we show that the sequence (14) is convergent to $x^* = 1$ for all $x_0 \in (0, b]$ (and hence, for all $x_0 \in (0, 1]$). In order to do so, we should show that $|x_{k+1} - 1| < |x_k - 1|$, for all $k \geq 0$. Suppose that $x_0 \in (0, b]$ is an arbitrary value. According to (15), the sequence $\{x_k\}$ is strictly increasing on $(0, 1)$ and

strictly decreasing on $(1, b)$. Hence, if $x_k, x_{k+1} \in (0, 1)$ or, $x_k, x_{k+1} \in (1, b)$, then $|x_{k+1} - 1| < |x_k - 1|$. Therefore, it is only sufficient to consider the case $x_k \in (0, 1)$ and $x_{k+1} \in (1, b)$ (the second case in which $x_k \in (1, b)$ and $x_{k+1} \in (0, 1)$ gives similar results). It is clear that

$$|x_{k+1} - 1| = |x_k + x_k(1 - x_k^2)(1 - \alpha x_k^2) - 1| = |x_k - 1|\,|1 - x_k(1 + x_k)(1 - \alpha x_k^2)|.$$

If, for different values of α, one plots the function $f(x) = |1 - x_k(1 + x_k)(1 - \alpha x_k^2)|$ on $[0, 1]$ or $[1, 1/\sqrt{\alpha}]$ separately, then it is observed that $f(x) < 1$, whenever $\alpha \in [0.21, 1)$. As a result, if $\alpha \in [0.21, 1)$, then $|x_{k+1} - 1| < |x_k - 1|$, $\forall\, k \geq 0$ and $\forall\, x_0 \in (0, b]$.

Now we can summarize our results into the following theorem:

Theorem 2. *If $\alpha \in [0.21, 1)$ and $b \in (1, 1/\sqrt{\alpha})$, then $x^* = 1$ is the unique fixed point of $h(x)$ in $(0, b)$, and the sequence (14) converges to $x^* = 1$, for any $x_0 \in (0, b]$ (and hence, for any $x_0 \in (0, 1]$).*

The above theorem shows that the class of modifications (13) is convergent only for $\alpha \in [0.21, 1)$.

Now, we consider the rate of convergence for sequence (14). Let $e_m = x_m - 1$ denote the error in the mth iteration. For simplicity, let $\overline{x} = x_{m+1}$, $x = x_m$, and also $\overline{e} = e_{m+1}$, $e = e_m$. Based on the relation

$$\begin{aligned} \overline{x} &= (1 + (1 - x^2)(1 - \alpha x^2))\, x \\ &= 2x - (\alpha + 1)x^3 + \alpha x^5, \end{aligned}$$

we have

$$\begin{aligned} \overline{e} &= 2e - (\alpha + 1)(e + 1)^3 + \alpha(e + 1)^5 \\ &= (-1 + 2\alpha)e + (7\alpha - 3)e^2 + (9\alpha - 1)e^3 + (5\alpha)e^4 + (\alpha)e^5. \end{aligned} \tag{16}$$

Therefore,

$$\lim_{k \to \infty} \frac{|\overline{e}|}{|e|} = |2\alpha - 1|. \tag{17}$$

Relationship (17) shows that the class of methods (13) is, in general, linearly convergent with asymptotic error constant $|2\alpha - 1|$. In this case, convergence is rapid when $|2\alpha - 1|$ is small.

Previously, we saw that the special choice $\alpha = 0.5$ is the same as that in the method of [3] for which $2\alpha - 1 = 0$. Hence, according to (16),

$$\lim_{k \to \infty} \frac{|\overline{e}|}{|e|^2} = |7\alpha - 3| = 0.5. \tag{18}$$

This shows that the method of [3] is quadratically convergent, and it is the only one in the class (13) with this property. Then, we can say that the method of [3] is optimal in our class with respect to convergence.

The following theorem can be proved immediately:

Theorem 3. *Any method of the class (13) is linearly convergent, for all α, with $0.21 \leq \alpha < 1$. For $\alpha = 0.5$, we can obtain Petcu and Popa's method which has quadratic convergence. This is the only method with this property in this class and in this sense is optimal.*

4 Conclusion

In this paper we represented a single-parameter class of modifications for Kovarik's method, which includes Petcu and Popa's modification [3], based on a special quadratic interpolation (with linear interpolation as a special case). We proved that for a parameter α in (13) belonging to $[0.21, 1)$, the above class would be generally linearly convergent. In addition, we proved that Petcu and Popa's method is the only convergent method of second order in this class. This also proves that Petcu and Popa's method has a quadratic convergence, in contrast to their claim, and is an optimal method in this class.

Acknowledgment. I would like to express deep gratitude to referee whose guidance was crucial for the successful completion of this paper.

References

[1] G. H. Golub, C. F. van Loan, *Matrix computations*, The Jhon's Hopkins University Press, Baltimore, 1983.

[2] Z. Kovarik, *Some iterative methods for improving orthogonality*, SIAM J. Numer. Anal. 7, 1970, 386-389.

[3] D. Petcu, C. Popa, *A new version of Kovarik's approximate orthogonalization algorithm without matrix inversion*, International J. Computer Mathematics 82, 2005, 1235-1246.

[4] C. Popa, *A method for improving orthogonality of rows and columns of matrices*, International J. Computer Mathematics 77, 2001, 469-480.

Department of Mathematics,
Bu-Ali Sina University
Hamedan, Iran
email:esmaeili@basu.ac.ir