ON THE BEHAVIOR OF THE CONJUGATE-GRADIENT METHOD ON ILL-CONDITIONED PROBLEMS

Anders FORSGREN*

Technical Report TRITA-MAT-2006-OS1
Department of Mathematics
Royal Institute of Technology
January 2006

Abstract

We study the behavior of the conjugate-gradient method for solving a set of linear equations, where the matrix is symmetric and positive definite with one set of eigenvalues that are large and the remaining are small. We characterize the behavior of the residuals associated with the large eigenvalues throughout the iterations, and also characterize the behavior of the residuals associated with the small eigenvalues for the early iterations. Our results show that the residuals associated with the large eigenvalues are made small first, without changing very much the residuals associated with the small eigenvalues. A conclusion is that the ill-conditioning of the matrix is not reflected in the conjugate-gradient iterations until the residuals associated with the large eigenvalues have been made small.

Key words. conjugate-gradient method, symmetric positive-definite matrix, ill-conditioning

AMS subject classifications. 65F10, 65F22, 65K05

1. Introduction

A fundamental problem in linear algebra is the solution of a system of linear equations on the form $Ax = b$, where A is an $n \times n$ symmetric positive definite matrix and b is an n-dimensional vector. From an optimization perspective, an equivalent problem may be formulated as

$$\underset{x \in I\!\!R^n}{\text{minimize}} \quad \tfrac{1}{2} x^T H x + c^T x, \tag{1.1}$$

where H is an $n \times n$ symmetric positive definite matrix and c is an n-dimensional vector. The unique solution to (1.1) is given by $Hx = -c$. Hence, by identifying $A = H$ and $b = -c$, the problems are equivalent.

If $x^* = -H^{-1}c$ denotes the optimal solution to (1.1), we may write $\frac{1}{2}x^T Hx + c^T x = \frac{1}{2}(x - x^*)^T H(x - x^*) - \frac{1}{2}x^{*T} Hx^*$. If we let $\xi = x - x^*$ and consider a rotation

*Optimization and Systems Theory, Department of Mathematics, Royal Institute of Technology, SE-100 44 Stockholm, Sweden (andersf@kth.se). Research supported by the Swedish Research Council (VR).

of the variables so that the Hessian becomes diagonal, i.e., $H = \text{diag}(\lambda)$, where λ is the vector of eigenvalues of H, (1.1) may equivalently be rewritten as

$$\underset{\xi \in I\!R^n}{\text{minimize}} \quad \tfrac{1}{2} \sum_{i=1}^{n} \lambda_i \xi_i^2, \tag{1.2}$$

where the constant term $\frac{1}{2} x^{*T} H x^*$ has been ignored. Of course, x^* is not known, and hence ξ is not known. However, we will discuss properties of ξ, and the construction is therefore convenient. We will throughout let the eigenvalues of H be ordered such that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n > 0$.

The conjugate-gradient method is a well-known iterative method for solving (1.2). At iteration k, it computes an approximate solution $\xi^{(k)}$ to (1.2) as the minimizer of the objective function of (1.2) subject to the constraint that $\xi - \xi^{(0)}$ belongs to the Krylov subspace spanned by $H\xi^{(0)}, \ldots, H^k \xi^{(0)}$. There is a convenient recurrence formula for recurring $\xi^{(k+1)}$ from $\xi^{(k)}$. For a thorough discussion on the conjugate gradient method, see, e.g., Golub and Van Loan [10, Sections 10.2–10.3] or Luenberger [16, Chapter 8]. The method was originally proposed by Hestenes and Stiefel [15]. There is a rich literature on conjugate-gradient methods, see, e.g., Hestenes [13], Axelsson and Barker [2], Golub and O'Leary [9], Axelsson [1] and Saad [18]. Further references are Faddeev and Faddeeva [6], Dahlquist, Eistenstat and Golub [4], Hestenes [12], and Hestenes and Stein [14].

We are particularly interested in the situation where H has r large eigenvalues and the remaining $n - r$ eigenvalues are small. Our motivation is twofold: first, interior methods [7, 8], where infinitely ill-conditioned matrices arise, and second, radiation therapy optimization [3], where ill-conditioned systems arising from discretized Fredholm equations of the first kind arise. The conjugate-gradient method is known to behave in a regularizing manner on such ill-conditioned linear equations, see, e.g., Squire [19], Hanke [11] and Vogel [20]. This is also related to partial least-squares methods, see, e.g., Wold et al. [21] and Eldén [5].

We show that the components of the iterates associated with the the large eigenvalues, $\xi_i^{(k)}$, $i = 1, \ldots, r$, are close to the iterates that are obtained if the conjugate-gradient method is applied to the r-dimensional problem where only the residuals associated with the r large eigenvalues are considered. In addition, we show that the components of the early iterates associated with the small eigenvalues, $\xi_i^{(k)}$, $i = r+1, \ldots, n$, are close to the corresponding initial residual $\xi_i^{(0)}$, $i = r+1, \ldots, n$. Broadly speaking, this means that the path of iterates is close to first satisfying the partial least-squares problem associated with the large residuals without significantly moving the small residuals, and then moving to reducing the small residuals. An implication of this result is that if the large eigenvalues are of comparable magnitude, the ill-conditioning of the problem caused by the small eigenvalues does not appear in the early iterations. Although there is a rich literature on conjugate-gradient methods, we are not aware of an analysis along the lines presented in our paper. We allow the initial residuals to appear in our expressions, but consider exact polynomials. This can be contrasted to "classical" bounds, e.g., the bounds given in Axelsson and Barker [2, Chapter 1], based on Chebyshev polynomials.

The paper is organized as follows: In Section 2 we give a brief background on the conjugate-gradient method. Section 3 contains a review of relevant properties of the polynomials associated with the conjugate-gradient method. In Section 4, we define the conjugate-gradient problem associated with the large eigenvalues only. Section 5 contains the main results of the paper, the characterization of the iterates of the conjugate-gradient method. In Section 6 we give a brief relation to the steepest-descent method, and finally a summary is given in Section 7.

2. Background

After k steps of the conjugate-gradient method, we obtain $\xi^{(k)}$ and $\alpha^{(k)}$ from

$$\begin{array}{ll} \underset{\xi\in I\!R^n,\alpha\in I\!R^k}{\text{minimize}} & \frac{1}{2}\sum_{i=1}^{n}\lambda_i\xi_i^2 \\ \text{subject to} & \xi_i=\xi_i^{(0)}+\sum_{l=1}^{k}\lambda_i^l\xi_i^{(0)}\alpha_l, \quad i=1,\dots,n, \end{array} \tag{2.1}$$

as optimal solution and optimal value, respectively. This formulation is a convex quadratic program, where the Krylov vectors $\xi^{(0)}, H\xi^{(0)}, \dots, H^k\xi^{(0)}$ appear explicitly.

Alternatively, the constraints $\xi_i = \xi_i^{(0)} + \sum_{l=1}^{k}\lambda_i^l\xi_i^{(0)}\alpha_l$, $i = 1,\dots,n$, may be viewed as to say that $\xi_i = P_k(\lambda_i)\xi^{(0)}$, where $P_k(\lambda)$ is a kth degree polynomial in λ such that $P_k(0) = 1$. This polynomial may be characterized in terms of its zeros $\zeta \in I\!R^k$ as $Q_k(\lambda,\zeta)$, with

$$Q_k(\lambda,\zeta)=\prod_{l=1}^{k}\left(1-\frac{\lambda}{\zeta_l}\right), \tag{2.2}$$

where we will assume that ζ is ordered such that $\zeta_1 \ge \zeta_2 \ge \dots \ge \zeta_k$. Then, (2.1) may equivalently be rewritten as

$$\begin{array}{ll} \underset{\xi\in I\!R^n,\zeta\in I\!R^k}{\text{minimize}} & \frac{1}{2}\sum_{i=1}^{n}\lambda_i\xi_i^2 \\ \text{subject to} & \xi_i=Q_k(\lambda_i,\zeta)\xi_i^{(0)}, \quad i=1,\dots,n, \end{array} \tag{2.3}$$

where the optimal solution is denoted by $\xi^{(k)}$ and $\zeta^{(k)}$. Note that the formulations (2.1) and (2.3) are equivalent, and we will make use of both.

We will denote by n_n the number of iterations it takes for the conjugate-gradient method to solve (1.2). The number n_n equals the number of distinct eigenvalues of H with nonzero initial residuals. Without loss of generality, the reader may consider $n_n = n$, but we use n_n for the sake of completeness. Problems (2.1) and (2.3) have unique solutions for $1 \le k \le n_n$, with the ordering of ζ in (2.3) such that $\zeta_1 \ge \zeta_2 \ge \dots \ge \zeta_k$.

As mentioned in the introduction, we are interested in the situation when H has r large eigenvalues and $n - r$ small eigenvalues. We will throughout quantify this situation by a scalar ϵ, $\epsilon \in (0, 1]$ such that $\lambda_{r+1} \le \epsilon\lambda_r$. We will be interested in the case when $\epsilon \ll 1$, and our analysis applies when $\epsilon r \le 1$.

	λ	$\xi^{(0)}$	$\xi^{(1)}$	$\xi^{(2)}$	$\xi^{(3)}$	$\xi^{(4)}$	$\xi^{(5)}$
$i=1$	2.0000	1.0000	-0.1733	0.0217	-0.0049	0.0000	0.0000
$i=2$	1.5000	1.0000	0.1201	-0.0702	0.0236	-0.0000	-0.0000
$i=3$	1.0000	1.0000	0.4134	0.0623	-0.0413	0.0001	-0.0000
$i=4$	0.1000	1.0000	0.9413	0.8659	0.7647	-0.0108	0.0000
$i=5$	0.0100	1.0000	0.9941	0.9862	0.9747	0.8793	-0.0000

Table 1: Iterates $\xi^{(k)}$, $k=0,\ldots,5$, for problem with $\lambda=(2,1.5,1,0.1,0.01)^T$ and $\xi^{(0)}=(1,1,1,1,1)^T$.

	λ	$\xi^{(0)}$	$\xi^{(1)}$	$\xi^{(2)}$	$\xi^{(3)}$	$\xi^{(4)}$	$\xi^{(5)}$
$i=1$	2.0000	1.0000	-0.1717	0.0181	-0.0001	0.0000	0.0000
$i=2$	1.5000	1.0000	0.1212	-0.0642	0.0003	-0.0000	0.0000
$i=3$	1.0000	1.0000	0.4141	0.0720	-0.0006	0.0000	0.0000
$i=4$	0.0100	1.0000	0.9941	0.9864	0.9784	-0.0001	0.0000
$i=5$	0.0001	1.0000	0.9999	0.9999	0.9998	0.9898	-0.0000

Table 2: Iterates $\xi^{(k)}$, $k=0,\ldots,5$, for problem with $\lambda=(2,1.5,1,0.01,0.0001)^T$ and $\xi^{(0)}=(1,1,1,1,1)^T$.

As a small illustrative example, consider the case where $\lambda=(2,1.5,1,0.1,0.01)^T$ and $\xi^{(0)}=(1,1,1,1,1)^T$. Here, we have three "large" eigenvalues, and then a gap between eigenvalue 3 and 4, and similarly a gap between eigenvalue 4 and 5. Table 1 shows the iterates $\xi^{(k)}$, $k=0,\ldots,5$. The numerical results of this table, as well as those presented in other tables and figures of this paper, have been obtained in Matlab using double precision arithmetic. We notice that the first three iterations are spent making the residuals associated with the three large eigenvalues small, whereas that residuals associated with the small eigenvalues are not changed very much. There is also a hierarchy here, in that we may consider four eigenvalues large compared to the fifth one. Hence, the fourth iteration is spent making the fourth residual small, without decreasing the fifth residual much.

If the gap between the large and the small eigenvalues is increased, the behavior observed above is manifested more clearly. The iterates for the case when $\lambda=(2\ 1.5\ 1\ 0.01\ 0.0001)^T$ and $\xi^{(0)}=(1\ 1\ 1\ 1\ 1)^T$, are given in Table 2. We notice that

	λ	$\xi^{(0)}$	$\xi^{(1)}$	$\xi^{(2)}$	$\xi^{(3)}$
$i=1$	2.0000	1.0000	-0.1717	0.0180	0.0000
$i=2$	1.5000	1.0000	0.1212	-0.0641	0.0000
$i=3$	1.0000	1.0000	0.4141	0.0721	-0.0000

Table 3: Iterates $\xi^{(k)}$, $k=0,\ldots,3$, for problem with $\lambda=(2,1.5,1)^T$ and $\xi^{(0)}=(1,1,1)^T$.

the first three components of the residuals, corresponding to large eigenvalues, in Tables 1 and 2 are similar. The tendency not to reduce the residuals corresponding to the small eigenvalues is increased in Table 2, when the gap has been increased.

The case when the residuals associated with the small eigenvalues are ignored entirely may be viewed as the limiting case when there is an infinite gap. If we in the above example ignore the two smallest eigenvalues, we obtain a three-dimensional problem with $\lambda = (2\ 1.5\ 1)^T$ and $\xi^{(0)} = (1\ 1\ 1)^T$. The iterates for this problem are given in Table 3. Note that during the first three iterations, the first three components of the iterates in Tables 1 and 2 are "close" to the iterates of Table 3, whereas the last two components of the iterates in Tables 1 and 2 not reduced very much.

	λ	$\xi^{(0)}$	$\xi^{(1)}$	$\xi^{(2)}$	$\xi^{(3)}$	$\xi^{(4)}$	$\xi^{(5)}$
$i=1$	2.0000	1.0000	-0.3242	0.3162	-0.0964	0.0003	0.0000
$i=2$	1.5000	1.0000	0.0068	-0.5540	0.4646	-0.0020	0.0000
$i=3$	1.0000	1.0000	0.3379	-0.7300	-0.8113	0.0052	-0.0000
$i=4$	0.1000	10.0000	9.3379	7.0207	1.4241	-0.1078	-0.0000
$i=5$	0.0100	10.0000	9.9338	9.6896	9.0454	8.7927	0.0000

Table 4: Iterates $\xi^{(k)}$, $k = 0, \ldots, 5$, for problem with $\lambda = (2, 1.5, 1, 0.1, 0.01)^T$ and $\xi^{(0)} = (1, 1, 1, 10, 10)^T$.

	λ	$\xi^{(0)}$	$\xi^{(1)}$	$\xi^{(2)}$	$\xi^{(3)}$	$\xi^{(4)}$	$\xi^{(5)}$
$i=1$	2.0000	1.0000	-0.1733	0.0225	-0.0072	0.0000	-0.0000
$i=2$	1.5000	1.0000	0.1200	-0.0712	0.0341	-0.0000	-0.0000
$i=3$	1.0000	1.0000	0.4133	0.0605	-0.0577	0.0000	0.0000
$i=4$	0.0100	10.0000	9.9413	9.8614	9.7316	-0.0010	-0.0000
$i=5$	0.0001	10.0000	9.9994	9.9986	9.9973	9.8978	0.0000

Table 5: Iterates $\xi^{(k)}$, $k = 0, \ldots, 5$, for problem with $\lambda = (2, 1.5, 1, 0.01, 0.0001)^T$ and $\xi^{(0)} = (1, 1, 1, 10, 10)^T$.

Note that the data for problem (1.2) is the eigenvalue vector λ and the initial residual vector $\xi^{(0)}$. Hence, $\xi^{(k)}$ depends on λ as well as on $\xi^{(0)}$. The impact of a small eigenvalue λ_i, $i = r+1, \ldots, n$, is also affected by the size of the initial residual $\xi_i^{(0)}$. In the examples above, we have chosen all residuals equal. If the initial residual associated with a small eigenvalue is large, compared to the residuals associated with the large eigenvalues, we envisage the impact of such a small eigenvalue/initial residual pair to be larger. This is illustrated in Table 4, where the data is identical to that of Table 1, except that the residuals associated with the two smallest eigenvalues are made ten times larger. We see that the iterates in Table 4 do not have the features of those of Table 1. The first three components of the iterates of Table 4 are not particularly close to those of Table 3, and the last two components of the iterates of Table 4 are reduced significantly also in the early iterations. If the gap between the large and small eigenvalues is increased, the behavior observed in Tables 1 and 2 is restored. This is demonstrated in Table 5, where the data is identical to that of Table 2, except that the residuals associated with the two smallest eigenvalues are made ten times larger. The purpose of this paper is to quantify this meaning of closedness and non-reduction.

Rather than looking at a table of residuals, we may view the iterates in terms of the polynomials $Q^{(k)}(\lambda,\zeta^{(k)})$. Figure 1 shows to the left the polynomials $Q^{(k)}(\lambda,\zeta^{(k)})$, $k=1,\ldots,5$, for the example problem with $\lambda=(2,1.5,1,0.1,0.01)^T$ and $\xi^{(0)}=(1,1,1,1,1)^T$. The right part of Figure 1 shows the polynomials $Q^{(k)}(\lambda,\zeta^{(k)})$, $k=1,\ldots,3$, for the problem with $\lambda=(2,1.5,1)^T$ and $\xi^{(0)}=(1,1,1)^T$. Note

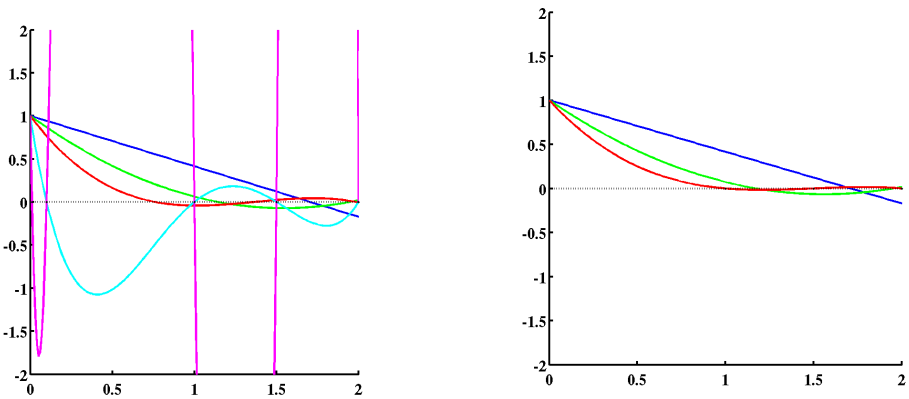

Figure 1: Left: Polynomials $Q^{(k)}(\lambda,\zeta^{(k)})$, $k=1,\ldots,5$, as a function of λ, for problem with $\lambda=(2,1.5,1,0.1,0.01)^T$ and $\xi^{(0)}=(1,1,1,1,1)^T$. Right: Polynomials $Q^{(k)}(\lambda,\zeta^{(k)})$, $k=1,\ldots,3$, as a function of λ, for problem with $\lambda=(2,1.5,1)^T$ and $\xi^{(0)}=(1,1,1)^T$.

that the small eigenvalues have little effect on the three first polynomials in the five-dimensional example, whereas polynomials four and five have increasing oscillations and amplitude. The fifth polynomial does not even fit in the window. As we would expect from the discussion above, the first three polynomials to the left and to the right are "similar". Note that the ill-conditioning of the five-dimensional problem does not appear in the first three iterations, since the iterates are close to those of the well-conditioned three-dimensional example problem.

3. Properties of the polynomials

In this section, we review some well-known properties of the polynomials $Q^{(k)}(\lambda,\zeta)$ and their zeros $\zeta^{(k)}$ that will be useful in our analysis. The first lemma shows how $\zeta_l^{(k)}$, $l=1,\ldots,k$, may be expressed as a convex combination of the eigenvalues λ_i, $i=1,\ldots,n$.

Lemma 3.1. *Let* $\xi^{(k)}$ *and* $\zeta^{(k)}$ *denote the optimal solution to* (2.3)*. Then, for all* k*, it holds that*

$$\zeta_l^{(k)} = \sum_{i=1}^{n} \frac{(\lambda_i \xi_i^{(0)}(Q_{k\setminus l}(\lambda_i, \zeta^{(k)})))^2}{\sum_{j=1}^{n}(\lambda_j \xi_j^{(0)}(Q_{k\setminus l}(\lambda_j, \zeta^{(k)})))^2}\lambda_i, \quad l = 1, \ldots, k,$$

where

$$Q_{k\setminus l}(\lambda, \zeta) = \prod_{\substack{m=1\\ m\neq l}}^{k} \left(1 - \frac{\lambda}{\zeta_m}\right).$$

In particular, $\lambda_n \le \zeta_l^{(k)} \le \lambda_1$*,* $l = 1, \ldots, k$*.*

Proof. We may eliminate ξ from (2.3) and write the objective function as

$$f(\zeta) = \tfrac{1}{2}\sum_{i=1}^{n} \lambda_i (Q_k(\lambda_i, \zeta)\xi_i^{(0)})^2. \tag{3.1}$$

Since $\zeta^{(k)}$ is the global minimizer, which is guaranteed to exist by the equivalence to the quadratic program (2.1), it must hold that $\partial f(\zeta^{(k)})/\partial \zeta_l = 0$, $l = 1, \ldots, k$. Differentiation of (3.1) gives

$$\frac{\partial f(\zeta)}{\partial \zeta_l} = -\frac{1}{\zeta_l^2}\sum_{i=1}^{n}\left(1 - \frac{\lambda_i}{\zeta_l}\right)\lambda_i^2 (Q_{k\setminus l}(\lambda_i, \zeta)\xi_i^{(0)})^2, \tag{3.2}$$

Hence, by the condition $\partial f(\zeta^{(k)})/\partial \zeta_l = 0$, (3.2) gives

$$\zeta_l^{(k)} = \sum_{i=1}^{n} \frac{(\lambda_i \xi_i^{(0)}(Q_{k\setminus l}(\lambda_i, \zeta^{(k)})))^2}{\sum_{j=1}^{n}(\lambda_j \xi_j^{(0)}(Q_{k\setminus l}(\lambda_j, \zeta^{(k)})))^2}\lambda_i, \quad l = 1, \ldots, k, \tag{3.3}$$

giving the required expression for $\zeta^{(k)}$. Since (3.3) gives $\zeta_l^{(k)}$ as a convex combination of the λ_is, it follows that $\lambda_n \le \zeta_l^{(k)} \le \lambda_1$. ∎

The convex combination provided by Lemma 3.1 is not very helpful in general, since the weights for one zero involves the other zeros. However, we get an explicit expression for $\zeta_1^{(1)}$.

Corollary 3.1. *Let* $\xi^{(1)}$ *and* $\zeta^{(1)}$ *denote the optimal solution to* (2.3) *for* $k = 1$*. Then,*

$$\zeta_1^{(1)} = \frac{\sum_{j=1}^{n}\lambda_j^3(\xi_j^{(0)})^2}{\sum_{j=1}^{n}\lambda_j^2(\xi_j^{(0)})^2}, \quad \text{and} \quad \xi_i^{(1)} = \left(1 - \frac{\lambda_i\sum_{j=1}^{n}\lambda_j^2(\xi_j^{(0)})^2}{\sum_{j=1}^{n}\lambda_j^3(\xi_j^{(0)})^2}\right)\xi_i^{(0)}, \quad i = 1, \ldots, n.$$

The behavior of $Q_k(\lambda, \zeta)$ as a function of λ, when λ is smaller than all the zeros $\zeta_l^{(k)}$, $l = 1, \ldots, k$, is of fundamental importance in our analysis. The following lemma gives the required properties, decreasing and convex. See Figure 1 for illustrative examples of polynomials.

Lemma 3.2. *For a fixed $\zeta \in I\!R^k$, with $\zeta_1 \geq \zeta_2 \geq \ldots \geq \zeta_k > 0$, let $Q_k(\lambda,\zeta)$ be defined by* (2.2)*. Then, $Q_k(\lambda,\zeta)$ is convex and decreasing as a function of λ for $\lambda \in [0,\zeta_k]$. In particular, for $\lambda \in [0,\zeta_k]$,*

$$1 - \lambda \sum_{l=1}^{k} \frac{1}{\zeta_l} \leq Q_k(\lambda,\zeta) \leq 1 - \frac{\lambda}{\zeta_k}.$$

Proof. We have

$$\frac{\partial Q_k(\lambda,\zeta)}{\partial \lambda} = -\sum_{l=1}^{k} \frac{1}{\zeta_l} \prod_{\substack{m=1\\ m\neq l}}^{k} \left(1 - \frac{\lambda}{\zeta_m}\right), \tag{3.4a}$$

$$\frac{\partial^2 Q_k(\lambda,\zeta)}{\partial \lambda^2} = \sum_{l=1}^{k} \sum_{m=1}^{k} \frac{1}{\zeta_l \zeta_m} \prod_{\substack{p=1\\ p\neq l,m}}^{k} \left(1 - \frac{\lambda}{\zeta_p}\right). \tag{3.4b}$$

It follows from (3.4b) that $\partial^2 Q_k(\lambda,\zeta)/\partial\lambda^2$ is nonnegative for $\lambda \in [0,\zeta_k]$, and hence $Q_k(\lambda,\zeta)$ is convex as a function of λ for $\lambda \in [0,\zeta_k]$. For $\lambda \in [0,\zeta_k]$, this convexity implies that

$$\begin{aligned} Q_k(\lambda,\zeta) &= Q_k\left(\left(1 - \frac{\lambda}{\zeta_k}\right)0 + \frac{\lambda}{\zeta_k}\zeta_k, \zeta\right) \\ &\leq \left(1 - \frac{\lambda}{\zeta_k}\right) Q_k(0,\zeta) + \frac{\lambda}{\zeta_k} Q_k(\zeta_k,\zeta) = \left(1 - \frac{\lambda}{\zeta_k}\right), \end{aligned}$$

where the identities $Q_k(0,\zeta) = 1$ and $Q_k(\zeta_k,\zeta) = 0$ have been used, thereby verifying the upper bound on $Q_k(\lambda,\zeta)$ for $\lambda \in [0,\zeta_k]$. In addition, the convexity implies that

$$Q_k(\lambda,\zeta) \geq Q_k(0,\zeta) + \frac{\partial Q_k(0,\zeta)}{\partial \lambda}\lambda = 1 - \lambda \sum_{l=1}^{k} \frac{1}{\zeta_l},$$

giving the required lower bound on $Q_k(\lambda,\zeta)$ for $\lambda \in [0,\zeta_k]$. ∎

4. A relaxed problem for the early iterations

The basis for our analysis is to consider the conjugate-gradient problem that arises when only the r large eigenvalues are considered and eigenvalues $r+1$ through n are disregarded. For iteration k, this means considering the optimization problem

$$\begin{array}{ll} \displaystyle\operatorname*{minimize}_{\xi \in I\!R^r, \alpha \in I\!R^k} & \frac{1}{2} \displaystyle\sum_{i=1}^{r} \lambda_i \xi_i^2 \\ \text{subject to} & \xi_i = \xi_i^{(0)} + \sum_{l=1}^{k} \lambda_i^l \xi_i^{(0)} \alpha_l, \quad i = 1,\ldots,r, \end{array} \tag{4.1}$$

where we denote the optimal solution by $\bar{\xi}_i^{(k)}$, $i = 1,\ldots,r$, and $\bar{\alpha}^{(k)}$. We will denote by n_r the first iteration k for which (4.1) has optimal value zero. We have previously talked about *early iterations*. This can now be made precise, and we will refer to iterations k, for which $0 \le k \le n_r$, as early iterations.

Given $\bar{\xi}_i^{(k)}$, $i = 1, \dots, r$, and $\bar{\alpha}^{(k)}$, that solve (4.1) for a given k, $0 \le k \le n_r$, we may define

$$\bar{\xi}_i^{(k)} = \xi_i^{(0)} + \sum_{l=1}^{k} \lambda_i^l \xi_i^{(0)} \bar{\alpha}_l^{(k)}, \quad i = r+1, \dots, n, \tag{4.2}$$

so as to obtain $\bar{\xi}^{(k)}$ as an n-dimensional vector. Equivalently, we may consider the optimization problem

$$\begin{array}{ll} \underset{\xi \in I\!R^n, \alpha \in I\!R^k}{\text{minimize}} & \frac{1}{2} \sum_{i=1}^{r} \lambda_i \xi_i^2 \\ \text{subject to} & \xi_i = \xi_i^{(0)} + \sum_{l=1}^{k} \lambda_i^l \xi_i^{(0)} \alpha_l, \quad i = 1, \dots, n, \end{array} \tag{4.3}$$

which is equivalent to solving (4.1) and then using (4.2). Hence, we obtain $\bar{\xi}^{(k)}$ and $\bar{\alpha}^{(k)}$ as the optimal solution of (4.3). Equivalently, we may write

$$\begin{array}{ll} \underset{\xi \in I\!R^n, \zeta \in I\!R^k}{\text{minimize}} & \frac{1}{2} \sum_{i=1}^{r} \lambda_i \xi_i^2 \\ \text{subject to} & \xi_i = Q_k(\lambda_i, \zeta) \xi_i^{(0)}, \quad i = 1, \dots, n, \end{array} \tag{4.4}$$

where we analogously denote the optimal solution by $\bar{\xi}^{(k)}$ and $\bar{\zeta}^{(k)}$. We prefer the n-dimensional formulation given by (4.3), rather than combining (4.1) and (4.2), since (2.1) and (4.3) have the same feasible sets. For a given feasible point to (2.1) and (4.3), the objective function value of (2.1) is at least as large as the objective function value of (4.3), and this means that (4.3) is a *relaxation* of (2.1).

In order to quantify how close the iterates of the initial problem (2.1) are to the iterates of the relaxed problem (4.3), we start by showing that the difference between the residuals associated with the small eigenvalues, $\bar{\xi}_i^{(k)}$, and the initial residual $\xi_i^{(0)}$, $i = r+1, \dots, n$, is small. This is a consequence of only considering the large eigenvalues in the minimization problem. The following lemma shows that if $\lambda_{r+1} \le \epsilon \lambda_r$, then the difference between $\bar{\xi}_i^{(k)}$ and $\xi_i^{(0)}$ is bounded by $k\epsilon$ for $k = 1, \dots, n_r$ and $i = r+1, \dots, n$. This is a key result to showing properties of $\xi^{(k)}$ later in the paper.

Lemma 4.1. *Assume that $\lambda_{r+1} \le \epsilon \lambda_r$, and let $\bar{\xi}^{(k)}$ together with $\bar{\zeta}^{(k)}$ be optimal solution to* (4.4)*. Then, for $k = 1, \dots, n_r$ and $i = r+1, \dots, n$,*

$$\begin{aligned} \xi_i^{(0)} \ge \bar{\xi}_i^{(k)} \ge (1 - k\epsilon) \xi_i^{(0)}, \quad &\textit{if } \xi_i^{(0)} \ge 0, \\ \xi_i^{(0)} \le \bar{\xi}_i^{(k)} \le (1 - k\epsilon) \xi_i^{(0)}, \quad &\textit{if } \xi_i^{(0)} \le 0. \end{aligned}$$

Proof. Let k be an iteration such that $k \le n_r$. Since (4.4) is a conjugate-gradient problem that only concerns eigenvalues λ_i, $i = 1, \dots, r$, and $k \le n_r$, it follows that

$\bar{\zeta}^{(k)}$ is unique, given the ordering $\bar{\zeta}_1^{(k)} \geq \bar{\zeta}_2^{(k)} \geq \ldots \geq \bar{\zeta}_k^{(k)}$, and Lemma 3.1 shows that $\bar{\zeta}_l^{(k)} \geq \lambda_r$, $l = 1, \ldots, k$. Moreover,

$$\bar{\xi}_i^{(k)} = Q_k(\lambda_i, \bar{\zeta}^{(k)})\xi_i^{(0)}, \quad i = 1, \ldots, n.$$

In particular, for $i = r+1, \ldots, n$, $\lambda_i \leq \lambda_r \leq \bar{\zeta}_l^{(k)}$. Let i be an index such that $r+1 \leq i \leq n$. Then, Lemma 3.2 gives

$$1 \geq Q_k(\lambda_i, \bar{\zeta}^{(k)}) \geq \left(1 - \lambda_i \sum_{l=1}^{k} \frac{1}{\zeta_l^{(k)}}\right) \geq \left(1 - k\frac{\lambda_i}{\lambda_r}\right) \geq (1 - k\epsilon). \tag{4.5}$$

Consequently, since $\bar{\xi}_i^{(k)} = Q_k(\lambda_i, \bar{\zeta}^{(k)})\xi_i^{(0)}$, (4.5) gives $\xi_i^{(0)} \geq \bar{\xi}_i^{(k)} \geq (1 - k\epsilon)\xi_i^{(0)}$ if $\xi_i^{(0)} \geq 0$ and $\xi_i^{(0)} \leq \bar{\xi}_i^{(k)} \leq (1 - k\epsilon)\xi_i^{(0)}$ if $\xi_i^{(0)} \leq 0$, as required. ∎

5. A characterization of the iterates

This characterization of the residuals $\bar{\xi}_i^{(k)}$ associated with the small eigenvalue indices $i = r+1, \ldots, n$ allows us to give an explicit bound on the difference between $\xi_i^{(k)} - \bar{\xi}_i^{(k)}$ for the large eigenvalue indices $i = 1, \ldots, r$. This shows how the iterates $\xi_i^{(k)}$ follow the iterates $\bar{\xi}_i^{(k)}$ for $i = 1, \ldots, k$ for the early iterations, and then remain small for the remaining iterations.

Theorem 5.1. *Assume that $\lambda_{r+1} \leq \epsilon\lambda_r$ and $\epsilon n_r \leq 1$. Let $\xi^{(k)}$ together with $\zeta^{(k)}$ be optimal solution to* (2.3) *and, for $k \leq n_r$, let $\bar{\xi}^{(k)}$ together with $\bar{\zeta}^{(k)}$ be optimal solution to* (4.4). *Then,*

$$\sum_{i=1}^{r} (\xi_i^{(k)} - \bar{\xi}_i^{(k)})^2 \leq \epsilon \sum_{i=r+1}^{n} (\xi_i^{(0)})^2, \quad k = 1, \ldots, n_r,$$

$$\sum_{i=1}^{r} (\xi_i^{(k)})^2 \leq \epsilon \sum_{i=r+1}^{n} (\xi_i^{(0)})^2, \quad k = n_r + 1, \ldots, n_n.$$

Proof. Note initially that $\xi_i^{(k)}$, $i = 1, \ldots, n$, is uniquely determined for $k \leq n_n$, since it is the solution to a conjugate-gradient quadratic program. For the same reason $\bar{\xi}_i^{(k)}$, $i = 1, \ldots, r$, is uniquely determined for $k \leq n_r$. First let $k \leq n_r$. Note that (4.3) is a relaxation of (2.1), and that (2.1) and (4.3) have the same feasible sets. Hence, since $\xi^{(k)}$ is optimal to (2.1) and $\bar{\xi}^{(k)}$ is optimal to (4.3), we conclude that

$$\tfrac{1}{2}\sum_{i=1}^{r} \lambda_i(\bar{\xi}_i^{(k)})^2 \leq \tfrac{1}{2}\sum_{i=1}^{r} \lambda_i(\xi_i^{(k)})^2 \leq \tfrac{1}{2}\sum_{i=1}^{n} \lambda_i(\xi_i^{(k)})^2 \leq \tfrac{1}{2}\sum_{i=1}^{n} \lambda_i(\bar{\xi}_i^{(k)})^2. \tag{5.1}$$

Consequently, Lemma 4.1 in conjunction with $\lambda_i \leq \epsilon\lambda_r$, $i = r+1, \ldots, n$, applied to (5.1), give

$$\tfrac{1}{2}\sum_{i=1}^{r} \lambda_i(\bar{\xi}_i^{(k)})^2 \leq \tfrac{1}{2}\sum_{i=1}^{r} \lambda_i(\xi_i^{(k)})^2 \leq \tfrac{1}{2}\sum_{i=1}^{r} \lambda_i(\bar{\xi}_i^{(k)})^2 + \tfrac{1}{2}\epsilon\lambda_r \sum_{i=r+1}^{n} (\xi_i^{(0)})^2. \tag{5.2}$$

A Taylor-series expansion of the objective function of (4.3) around $\bar{\xi}^{(k)}$ gives

$$\tfrac{1}{2}\sum_{i=1}^{r}\lambda_i(\xi_i^{(k)})^2 = \tfrac{1}{2}\sum_{i=1}^{r}\lambda_i(\bar{\xi}_i^{(k)})^2 + \sum_{i=1}^{r}\lambda_i\bar{\xi}_i^{(k)}(\xi_i^{(k)} - \bar{\xi}_i^{(k)}) + \tfrac{1}{2}\sum_{i=1}^{r}\lambda_i(\xi_i^{(k)} - \bar{\xi}_i^{(k)})^2. \quad (5.3)$$

Since (4.3) is an equality-constrained quadratic program to which $\bar{\xi}^{(k)}$ is optimal and $\xi^{(k)}$ is feasible, we conclude that

$$\sum_{i=1}^{r}\lambda_i\bar{\xi}_i^{(k)}(\xi_i^{(k)} - \bar{\xi}_i^{(k)}) = 0. \quad (5.4)$$

Consequently, a combination of (5.2), (5.3) and (5.4) gives

$$\sum_{i=1}^{r}\lambda_i(\xi_i^{(k)} - \bar{\xi}_i^{(k)})^2 \le \epsilon\lambda_r \sum_{i=r+1}^{n}(\xi_i^{(0)})^2. \quad (5.5)$$

Since $\lambda_i \ge \lambda_r$, $i = 1, \ldots, r$, it follows from (5.5) that

$$\sum_{i=1}^{r}(\xi_i^{(k)} - \bar{\xi}_i^{(k)})^2 \le \epsilon \sum_{i=r+1}^{n}(\xi_i^{(0)})^2, \quad (5.6)$$

as required.

Now let $n_r + 1 \le k \le n_n$. Then, upon observing that the conjugate-gradient method yields decreasing values of the objective function and $\bar{\xi}_i^{(n_r)} = 0$, $i = 1, \ldots, r$, we obtain the inequalities analogous to (5.1) as

$$\tfrac{1}{2}\sum_{i=1}^{r}\lambda_i(\xi_i^{(k)})^2 \le \tfrac{1}{2}\sum_{i=1}^{n}\lambda_i(\xi_i^{(k)})^2 \le \tfrac{1}{2}\sum_{i=1}^{n}\lambda_i(\xi_i^{(n_r)})^2 \le \tfrac{1}{2}\sum_{i=r+1}^{n}\lambda_i(\bar{\xi}_i^{(n_r)})^2. \quad (5.7)$$

In addition, we apply Lemma 4.1 in conjunction with $\lambda_i \ge \lambda_r$, $i = 1, \ldots, r$, and $\lambda_i \le \epsilon\lambda_r$, $i = r+1, \ldots, n$, to (5.7), which gives

$$\tfrac{1}{2}\lambda_r\sum_{i=1}^{r}(\xi_i^{(k)})^2 \le \tfrac{1}{2}\epsilon\lambda_r\sum_{i=r+1}^{n}(\xi_i^{(0)})^2. \quad (5.8)$$

Consequently, (5.8) gives

$$\sum_{i=1}^{r}(\xi_i^{(k)})^2 \le \epsilon\sum_{i=r+1}^{n}(\xi_i^{(0)})^2, \quad (5.9)$$

as required. The complete result is now given by (5.6) and (5.9). ∎

Similarly, we may now obtain a result that bounds the difference $\xi_i^{(k)} - \bar{\xi}_i^{(k)}$ also for the small eigenvalue indices $i = r+1, \ldots, n$ during the early iterations. This bound, however, is not as explicit as the bound for the large eigenvalue indices.

Theorem 5.2. *Assume that* $\lambda_{r+1} \le \epsilon\lambda_r$ *and* $\epsilon n_r \le 1$. *Let* $\xi^{(k)}$ *together with* $\zeta^{(k)}$ *be optimal solution to* (2.3) *and, for* $k \le n_r$, *let* $\bar\xi^{(k)}$ *together with* $\bar\zeta^{(k)}$ *be optimal solution to* (4.4). *Further, let* $\Xi_L^{(0)} = \mathrm{diag}(\xi_1^{(0)}, \dots, \xi_r^{(0)})$, *let* $\xi_S^{(0)} = (\xi_{r+1}^{(0)}, \dots, \xi_n^{(0)})^T$, *and let* $V_L^{(k)}$ *be the* $r \times k$ *matrix with element* ij *given by*

$$(V_L^{(k)})_{ij} = \left(\frac{\lambda_i}{\lambda_1}\right)^j.$$

Then, for $k = 1, \dots, n_r$,

$$|\xi_i^{(k)} - \bar\xi_i^{(k)}| \le \frac{k^{1/2}\epsilon^{3/2}\|\xi_S^{(0)}\|}{\sigma_k(\Xi_L^{(0)}V_L^{(k)})}|\xi_i^{(0)}|, \quad i = r+1, \dots, n,$$

where $\sigma_k(\Xi_L^{(0)}V_L^{(k)})$ *denotes the* k*th singular value of the matrix* $\Xi_L^{(0)}V_L^{(k)}$.

Proof. A combination of (2.1) and (4.3) gives

$$\xi_i^{(k)} - \bar\xi_i^{(k)} = \xi_i^{(0)}\left(\sum_{l=1}^k \lambda_i^l(\alpha_l^{(k)} - \bar\alpha_l^{(k)})\right), \quad i = 1, \dots, n. \tag{5.10}$$

We may normalize (5.10) so that

$$\xi_i^{(k)} - \bar\xi_i^{(k)} = \xi_i^{(0)}\left(\sum_{j=1}^k \left(\frac{\lambda_i}{\lambda_1}\right)^j (\lambda_1^j\alpha_j^{(k)} - \lambda_1^j\bar\alpha_j^{(k)})\right), \quad i = 1, \dots, n. \tag{5.11}$$

If we let $\beta^{(k)} \in I\!R^k$ have components $\beta_j^{(k)} = (\lambda_1^j\alpha_j^{(k)} - \lambda_1^j\bar\alpha_j^{(k)})$, $j = 1, \dots, k$, we may rewrite (5.11) as

$$\xi_i^{(k)} - \bar\xi_i^{(k)} = \xi_i^{(0)}\left(\sum_{j=1}^k \left(\frac{\lambda_i}{\lambda_1}\right)^j \beta_j^{(k)}\right), \quad i = 1, \dots, n. \tag{5.12}$$

Written in block form for components $i = 1, \dots, r$, (5.12) takes the form

$$\xi_L^{(k)} - \bar\xi_L^{(k)} = \Xi_L^{(0)}V_L^{(k)}\beta^{(k)}, \tag{5.13}$$

where $\xi_L^{(k)}$ is the k-dimensional vector with components $\xi_i^{(k)}$, $i = 1, \dots, k$. Taking norms in (5.13) gives

$$\|\xi_L^{(k)} - \bar\xi_L^{(k)}\| \ge \sigma_k(\Xi_L^{(0)}V_L^{(k)})\|\beta^{(k)}\|. \tag{5.14}$$

Note that $\sigma_k(\Xi_L^{(0)}V_L^{(k)}) > 0$, since $k \le n_r$, and hence (5.14) gives an upper bound for $\beta^{(k)}$. Furthermore, if we use (5.14) for $i = r+1, \dots, n$ in (5.12), upon observing that $\lambda_i/\lambda_1 \le \epsilon \le 1$, taking norms and using the Cauchy-Schwartz inequality, we obtain

$$|\xi_i^{(k)} - \bar\xi_i^{(k)}| \le |\xi_i^{(0)}|k^{1/2}\epsilon\|\beta^{(k)}\| \le \frac{k^{1/2}\epsilon\|\xi_L^{(k)} - \bar\xi_L^{(k)}\|}{\sigma_k(\Xi_L^{(0)}V_L^{(k)})}|\xi_i^{(0)}|. \tag{5.15}$$

Theorem 5.1 gives $\|\xi_L^{(k)} - \bar{\xi}_L^{(k)}\| \le \epsilon^{1/2}\|\xi_S^{(0)}\|$, which inserted into (5.15) gives

$$|\xi_i^{(k)} - \bar{\xi}_i^{(k)}| \le \frac{k^{1/2}\epsilon^{3/2}\|\xi_S^{(0)}\|}{\sigma_k(\Xi_L^{(0)}V_L^{(k)})}|\xi_i^{(0)}|,$$

as required. ∎

Note that the matrix $\Xi_L^{(0)}V_L^{(k)}$ of Theorem 5.2 is nonsingular for $k \le n_r$, but since $V_l^{(k)}$ has Vandermonde structure, we expect the smallest singular value $\Xi_L^{(0)}V_L^{(k)}$ to become small as k increases.

We may now combine Lemma 4.1 and Theorem 5.2, to show that $\xi_i^{(k)}$ is close to $\xi_i^{(0)}$ for $k = 1, \ldots, n_r$ and $i = r+1, \ldots, n$ for ϵ sufficiently small.

Corollary 5.1. *Assume that $\lambda_{r+1} \le \epsilon\lambda_r$ and $\epsilon n_r \le 1$. Let $\xi^{(k)}$ together with $\zeta^{(k)}$ be optimal solution to (2.3). Further, let $\Xi_L^{(0)} = \mathrm{diag}(\xi_1^{(0)}, \ldots, \xi_r^{(0)})$, let $\xi_S^{(0)} = (\xi_{r+1}^{(0)}, \ldots, \xi_n^{(0)})^T$, and let $V_L^{(k)}$ be the $r \times k$ matrix with element ij given by*

$$(V_L^{(k)})_{ij} = \left(\frac{\lambda_i}{\lambda_1}\right)^j.$$

Then, for $k = 1, \ldots, n_r$ and $i = r+1, \ldots, n$,

$$\left(1 + \frac{k^{1/2}\epsilon^{3/2}\|\xi_S^{(0)}\|}{\sigma_k(\Xi_L^{(0)}V_L^{(k)})}\right)\xi_i^{(0)} \ge \xi_i^{(k)} \ge \left(1 - k\epsilon - \frac{k^{1/2}\epsilon^{3/2}\|\xi_S^{(0)}\|}{\sigma_k(\Xi_L^{(0)}V_L^{(k)})}\right)\xi_i^{(0)}, \quad \text{if } \xi_i^{(0)} \ge 0,$$

$$\left(1 + \frac{k^{1/2}\epsilon^{3/2}\|\xi_S^{(0)}\|}{\sigma_k(\Xi_L^{(0)}V_L^{(k)})}\right)\xi_i^{(0)} \le \xi_i^{(k)} \le \left(1 - k\epsilon - \frac{k^{1/2}\epsilon^{3/2}\|\xi_S^{(0)}\|}{\sigma_k(\Xi_L^{(0)}V_L^{(k)})}\right)\xi_i^{(0)}, \quad \text{if } \xi_i^{(0)} \le 0,$$

where $\sigma_k(\Xi_L^{(0)}V_L^{(k)})$ denotes the kth singular value of the matrix $\Xi_L^{(0)}V_L^{(k)}$.

This means that we have characterized $\xi_L^{(k)}$ as close to $\bar{\xi}_L^{(k)}$ for $k = 1, \ldots, n_r$, and $\xi_L^{(k)}$ as close to zero for $k = n_r+1, \ldots, n_n$, where $\xi_L^{(k)} = (\xi_1^{(k)}, \ldots, \xi_r^{(k)})^T$ and $\bar{\xi}_L^{(k)} = (\bar{\xi}_1^{(k)}, \ldots, \bar{\xi}_r^{(k)})^T$. In addition, we have characterized $\xi_i^{(k)}$ as close to $\xi_i^{(0)}$ for $i = r+1, \ldots, n$ and $k = 1, \ldots, n_r$. This gives the desired characterization of initially minimizing the residuals associated with the large eigenvalues whereas not decreasing much the residuals associated with the small eigenvalues.

6. Relationship to the steepest-descent method

As a remark, we also briefly review the steepest descent method in a polynomial framework. Steepest descent uses a more "greedy" approach, in that a minimization over a one-dimensional subspace is carried out at each iteration. The steepest-descent method may be viewed as a conjugate-gradient method which is restarted every iteration. Here we obtain $\xi^{(k)}$ and ζ^k as the optimal solution to the problem

$$\begin{array}{ll} \underset{\xi \in I\!R^n, \zeta \in I\!R}{\text{minimize}} & \frac{1}{2}\sum_{i=1}^n \lambda_i \xi_i^2 \\ \text{subject to} & \xi_i = \left(1 - \frac{\lambda_i}{\zeta}\right)\xi_i^{(k-1)}, \quad i = 1, \ldots, n, \end{array} \tag{6.1}$$

which has the closed-form solution

$$\zeta^{(k)} = \frac{\sum_{j=1}^{n} \lambda_j^3 (\xi_j^{(k-1)})^2}{\sum_{j=1}^{n} \lambda_j^2 (\xi_j^{(k-1)})^2}, \quad \text{and} \tag{6.2a}$$

$$\xi_i^{(k)} = \left(1 - \frac{\lambda_i}{\zeta^{(k)}}\right) \xi_i^{(k-1)} = \prod_{l=1}^{k} \left(1 - \frac{\lambda_i}{\zeta^{(l)}}\right) \xi_i^{(0)}, \quad i = 1, \ldots, n, \tag{6.2b}$$

see Corollary 3.1. This means that steepest descent forms polynomials by adding one zero at the time, and not changing the zeros that have already been obtained. Hence, (6.2) generates polynomials in the same fashion as the conjugate-gradient method, but they are only optimal in this “greedy” sense. Consequently, finite termination is not obtained, in general. Figure 2 shows the first ten polynomials generated by steepest descent for the example problem with $\lambda = (2, 1.5, 1, 0.1, 0.01)^T$ and $\xi^{(0)} = (1, 1, 1, 1, 1)^T$.

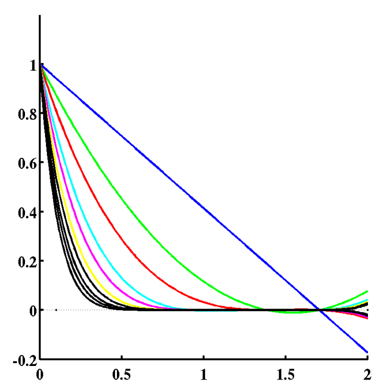

Figure 2: First ten polynomials generated by the steepest-descent method for problem with $\lambda = (2, 1.5, 1, 0.1, 0.01)^T$ and $\xi^{(0)} = (1, 1, 1, 1, 1)^T$.

7. Summary and discussion

We have characterized the path of iterates for the conjugate-gradient method applied to a system of linear equations when the $n \times n$ positive-definite symmetric matrix involved is ill-conditioned in the sense that it has r large eigenvalues and the remaining $n - r$ eigenvalues small. The components associated with the large eigenvalues, $\xi_i^{(k)}$, $i = 1, \ldots, r$, are close to the iterates that are obtained if the conjugate-gradient method is applied to the r-dimensional problem where only the residuals associated with the r large eigenvalues are considered. In addition, we have shown that the components of the early iterates associated with the small eigenvalues, $\xi_i^{(k)}$, $i = r+1, \ldots, n$, are close to the corresponding initial residual $\xi_i^{(0)}$, $i = r + 1, \ldots, n$. An implication of this result is that if the large eigenvalues are of comparable magnitude, the ill-conditioning of the problem caused by the small eigenvalues does not appear in the early iterations.

Further research would be directed towards solving ill-conditioned systems arising in interior methods using preconditioned conjugate-gradient methods. Also, we are interested in quasi-Newton methods for nonlinear optimization problems with ill-conditioned Hessians. The research presented in this paper is of interest for quasi-Newton methods, since quasi-Newton methods are equivalent to a conjugate-gradient method when solving (1.1) if exact linesearch is used [17].

Acknowledgement

I thank Fredrik Carlsson, Lars Eldén, Philip Gill, Axel Ruhe and Anders Szepessy for helpful discussions on conjugate-gradient methods.

References

[1] O. Axelsson. *Iterative solution methods.* Cambridge University Press, Cambridge, 1994.

[2] O. Axelsson and V. A. Barker. *Finite element solution of boundary value problems*, volume 35 of *Classics in Applied Mathematics.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001.

[3] F. Carlsson and A. Forsgren. Iterative regularization in intensity-modulated radiation therapy optimization. *Med. Phys.*, 33(1):225–234, 2006.

[4] G. Dahlquist, S. C. Eisenstat, and G. H. Golub. Bounds for the error of linear systems of equations using the theory of moments. *J. Math. Anal. Appl.*, 37:151–166, 1972.

[5] L. Eldén. Partial least-squares vs. Lanczos bidiagonalization. I. Analysis of a projection method for multiple regression. *Comput. Statist. Data Anal.*, 46(1):11–31, 2004.

[6] D. K. Faddeev and V. N. Faddeeva. *Computational methods of linear algebra.* Translated by Robert C. Williams. W. H. Freeman and Co., San Francisco, 1963.

[7] A. Forsgren, P. E. Gill, and J. D. Griffin. Iterative solution of augmented systems arising in interior methods. Report TRITA-MAT-2005-OS3, Department of Mathematics, Royal Institute of Technology, Stockholm, Sweden, 2005.

[8] A. Forsgren, P. E. Gill, and M. H. Wright. Interior methods for nonlinear optimization. *SIAM Rev.*, 44(4):525–597 (electronic) (2003), 2002.

[9] G. H. Golub and D. P. O'Leary. Some history of the conjugate gradient and Lanczos algorithms: 1948–1976. *SIAM Rev.*, 31(1):50–102, 1989.

[10] G. H. Golub and C. F. Van Loan. *Matrix Computations.* The Johns Hopkins University Press, Baltimore, Maryland, third edition, 1996. ISBN 0-8018-5414-8.

[11] M. Hanke. *Conjugate gradient type methods for ill-posed problems*, volume 327 of *Pitman Research Notes in Mathematics Series.* Longman Scientific & Technical, Harlow, 1995.

[12] M. R. Hestenes. Iterative methods for solving linear equations. *J. Optimization Theory Appl.*, 11:323–334, 1973.

[13] M. R. Hestenes. *Conjugate-Direction Methods in Optimization.* Springer-Verlag, Berlin, Heidelberg and New York, 1980.

[14] M. R. Hestenes and M. L. Stein. The solution of linear equations by minimization. *J. Optimization Theory Appl.*, 11:335–359, 1973.

[15] M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *J. Research Nat. Bur. Standards*, 49:409–436 (1953), 1952.

[16] D. G. Luenberger. *Linear and Nonlinear Programming*. Addison-Wesley Publishing Company, Reading, second edition, 1984. ISBN 0-201-15794-2.

[17] L. Nazareth. A relationship between the BFGS and conjugate gradient algorithms and its implications for new algorithms. *SIAM J. Numer. Anal.*, 16:794–800, 1979.

[18] Y. Saad. *Iterative methods for sparse linear systems*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2003.

[19] W. Squire. The solution of ill-conditioned linear systems arising from Fredholm equations of the first kind by steepest descents and conjugate gradients. *Internat. J. Numer. Methods Engrg.*, 10(3):607–617, 1976.

[20] C. R. Vogel. *Computational methods for inverse problems*, volume 23 of *Frontiers in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2002.

[21] S. Wold, A. Ruhe, H. Wold, and W. J. Dunn III. The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM J. Sci. Stat. Comp*, 5:735–743, 1984.