

RUPRECHT-KARLS-UNIVERSITÄT HEIDELBERG
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE
INSTITUTE OF APPLIED MATHEMATICS

The Complex Rayleigh Quotient Iteration

Bachelor Thesis

Author: NILS FRIESS

Supervisor: PROF. DR. ROBERT SCHEICHL

Submitted: November 3, 2019

Contents

1	Introduction	3
1.1	Motivation	3
1.2	General problem	3
1.3	Iterative methods for eigenvalue problems	6
1.3.1	Power method	6
1.3.2	(Shifted) Inverse Iteration	6
1.3.3	Rayleigh Quotient Iteration	7
1.4	History	8
2	Classic Rayleigh Quotient Iteration	9
2.1	The Rayleigh Quotient	9
3	Complex Rayleigh Quotient Iteration	12
A	Proofs	13

1 Introduction

In this chapter, we introduce the general problem by discussing an example that naturally arises in different physical applications. We will then define that basic notion of the Rayleigh quotient and derive the Rayleigh quotient iteration (or RQI, for short). Finally, we provide an overview of the historic developments of RQI.

1.1 Motivation

1.2 General problem

We will now formulate the abstract problem that was derived in the last section. To that end, we will give a few basic facts and definitions from linear algebra.

Definition 1.1. Let $\mathbf{A} \in \mathbb{C}^{n \times n}$ be a matrix. A scalar $\lambda \in \mathbb{C}$ is called *eigenvalue* of \mathbf{A} if there exists a nonzero vector $\mathbf{v} \in \mathbb{C}^n$ such that

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v} . \tag{1.1} \quad \text{\texttt{\{eq:eigvalprobl}}$$

The vector \mathbf{v} is called an *eigenvector* of \mathbf{A} associated with λ . The tuple (λ, \mathbf{v}) is called an *eigenpair*. The set of all eigenvalues of \mathbf{A} is called the *spectrum* and is denoted by $\sigma(\mathbf{A})$.

In the following proposition we combine some basic facts on eigenvalues and eigenvectors. The results are shown under the assumption that $\mathbf{A} \in \mathbb{C}^{n \times n}$ is a complex Hermitian matrix, i. e. $\mathbf{A} = \overline{\mathbf{A}}^\top$, where the bar denotes the complex conjugate. If \mathbf{A} is a real matrix, we have $\overline{\mathbf{A}} = \mathbf{A}$ and thus the following facts hold in particular for real symmetric matrices.

{prop:eigval:fa

Proposition 1.2. *Let $\mathbf{A} = \overline{\mathbf{A}}^\top \in \mathbb{C}^{n \times n}$ be a Hermitian matrix. Denote by $\lambda_1, \lambda_2, \dots, \lambda_n$ the eigenvalues of \mathbf{A} with associated eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_n$.*

- (i) *All eigenvalues of \mathbf{A} are real.*
- (ii) *There exists an orthonormal basis of \mathbb{C}^n consisting of eigenvectors of \mathbf{A} .*
- (iii) *If \mathbf{A} is non-singular the eigenvalues of \mathbf{A}^{-1} are given by $\lambda_1^{-1}, \dots, \lambda_n^{-1}$ with eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_n$.*
- (iv) *Let $\sigma \in \mathbb{R}$ an arbitrary scalar. Then the eigenvalues of $\mathbf{A} - \sigma \mathbf{I}$ are $\lambda_i - \sigma$ with eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_n$.*

Proof. The proofs for (i) and (ii) can be found in most standard linear algebra literature, for example in [Fis13, Lemma and Theorem 5.6.2, p. 312].

- (iii) Suppose \mathbf{A} is invertible and let (λ, \mathbf{v}) be an eigenpair of \mathbf{A} (note that since \mathbf{A} is non-singular we have $\lambda \neq 0$). Then

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v} \quad \Leftrightarrow \quad \mathbf{A}^{-1}\mathbf{A}\mathbf{v} = \lambda\mathbf{A}^{-1}\mathbf{v} \quad \Leftrightarrow \quad \lambda^{-1}\mathbf{v} = \mathbf{A}^{-1}\mathbf{v},$$

i. e. $(\lambda^{-1}, \mathbf{v})$ is an eigenpair of \mathbf{A}^{-1} .

Finish
proof

□

In the following we restrict our attention to the *symmetric eigenvalue problem*, i. e. we want to find solutions of equation (1.1) given $\mathbf{A} \in \mathbb{R}^{n \times n}$ is a real symmetric matrix. The eigenvalues and associated eigenvectors of \mathbf{A} are denoted by $\lambda_1, \dots, \lambda_n$ and $\mathbf{v}_1, \dots, \mathbf{v}_n$, respectively. We assume that the eigenvalues are normalised w. r. t. the Euclidean norm, i. e.

$$\|\mathbf{v}_i\| = \sqrt{\mathbf{v}_i^T \mathbf{v}_i} = 1 \quad \text{for all } i = 1, \dots, n.$$

Note, that due to Proposition 1.2 (ii) we have

$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle = \mathbf{v}_i^T \mathbf{v}_j = 0 \quad \text{for } i \neq j,$$

where $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product on \mathbb{R}^n . Since all eigenvalues are real we can label them in increasing order of magnitude

$$|\lambda_1| \leq |\lambda_2| \leq \dots \leq |\lambda_n|.$$

The eigenvalues λ_1 and λ_n are called *extreme* eigenvalues. The remaining eigenvalues $\lambda_2, \dots, \lambda_{n-1}$ are called *interior* eigenvalues. Sometimes we will make the assumption that \mathbf{A} is positive definite, i. e. that

$$\mathbf{x}^T \mathbf{A} \mathbf{x} > 0 \quad \text{for all } \mathbf{x} \in \mathbb{R}^n.$$

Given an eigenvalue λ and an associated eigenvector \mathbf{v} we then have

$$\lambda = \mathbf{v}^T \lambda \mathbf{v} = \mathbf{v}^T \mathbf{A} \mathbf{v} > 0.$$

i. e. eigenvalues of symmetric positive definite matrices are positive.

1.3 Iterative methods for eigenvalue problems

{sec:iterative_

In this section we introduce the most simple iterative methods for computing eigenpairs of symmetric matrices.

1.3.1 Power method

The power method is based on generating the sequence $\mathbf{A}^k \mathbf{v}_0$ where \mathbf{v}_0 is a non-zero unit vector. Of course, \mathbf{A}^k does not have to be computed explicitly since

$$\mathbf{A}^k \mathbf{x} = \mathbf{A}(\mathbf{A}(\dots \mathbf{A}(\mathbf{A}\mathbf{x}) \dots)).$$

The sequence \mathbf{v}_k as generated in Algorithm 1.3 converges to the eigenvector associated with the eigenvalue λ_n (under the assumption that λ_n is the unique dominant eigenvalue, cf. [Par98, Theorem 4.2.1]). The power method

Algorithm 1.3: Power method

Input: Nonzero unit vector \mathbf{v}_0 **for** $k = 1, 2, \dots$ *until convergence* **do** $\tilde{\mathbf{v}}_k \leftarrow \mathbf{A}\mathbf{v}_{k-1}$
 $\mathbf{v}_k \leftarrow \tilde{\mathbf{v}}_k / \|\tilde{\mathbf{v}}_k\|$

{alg:power:method}

converges linearly with convergence factor $|\lambda_{n-1}|/|\lambda_n|$.

1.3.2 (Shifted) Inverse Iteration

The *inverse iteration* is the power method applied to \mathbf{A}^{-1} (given \mathbf{A} is non-singular). Due to Proposition 1.2 (iii) this will produce a sequence of vectors \mathbf{v}_k converging to the eigenvector associated to the smallest eigenvalue λ_1 .

Combining this idea with Proposition 1.2 (iv) produces the *shifted inverse iteration*.¹ There, the iterates are defined by

$$\mathbf{v}_k = (\mathbf{A} - \sigma \mathbf{I})^{-1} \mathbf{v}_{k-1},$$

where we have omitted the normalisation step. The scalar $\sigma \in \mathbb{R}$ is often called the *shift* since it shifts the eigenvalues of \mathbf{A}^{-1} by σ on the real axis. Of course, we do not have to compute the inverse explicitly. Instead, before the loop we can compute the LU decomposition of $\mathbf{A} - \sigma \mathbf{I}$ and solve the system

$$(\mathbf{A} - \sigma \mathbf{I}) \mathbf{v}_k = \mathbf{v}_{k-1}$$

for \mathbf{v}_k reducing the complexity from $\mathcal{O}(n^3)$ to $\mathcal{O}(n^2)$ at each step. We summarise the results in Algorithm 1.4.

Algorithm 1.4: Shifted inverse iteration

Input: Nonzero unit vector \mathbf{v}_0 , shift $\sigma \in \mathbb{R}$

Compute LU decomposition $\mathbf{A} - \sigma \mathbf{I} = \mathbf{L}\mathbf{U}$

for $k = 1, 2, \dots$ *until convergence* **do**

 Solve $\mathbf{v}_{k-1} = (\mathbf{A} - \sigma \mathbf{I}) \tilde{\mathbf{v}}_k$ for $\tilde{\mathbf{v}}_k$
 $\mathbf{v}_k \leftarrow \tilde{\mathbf{v}}_k / \|\tilde{\mathbf{v}}_k\|$

{alg:sii}

1.3.3 Rayleigh Quotient Iteration

Definition 1.5 (Rayleigh quotient). Let $\mathbf{A} \in \mathbb{R}^{n \times n}$. The mapping

$$\mathcal{R}_{\mathbf{A}} : \mathbb{R}^n \setminus \{\mathbf{0}\} \rightarrow \mathbb{R}, \quad \mathbf{x} \mapsto \frac{\mathbf{x}^\top \mathbf{A} \mathbf{x}}{\mathbf{x}^\top \mathbf{x}}$$

is called the *Rayleigh quotient* corresponding to the matrix \mathbf{A} .

¹Note that some authors call this the inverse iteration and do not make a distinction of a shift being used or not.

1.4 History

In this section we present several publications that contributed towards the Rayleigh Quotient iteration.

Formulierung

2 Classic Rayleigh Quotient Iteration

In this chapter we thoroughly introduce the Rayleigh Quotient Iteration (or RQI, for short) and present important contributions regarding its analysis. Among others, we give a proof of the cubic convergence of RQI. We provide numerical examples for the presented results and discuss disadvantages and shortcomings that will directly lead us the main part of the thesis, discussed in the next chapter.

2.1 The Rayleigh Quotient

Recall the general problem: Given approximations of an eigenvector of a symmetric matrix $\mathbf{A} = \mathbf{A}^\top \in \mathbb{R}^{n \times n}$ we want to compute the exact eigenvector and associated eigenvalue. In Section 1.3 we briefly introduced iterative methods for computing eigenpairs, including the Rayleigh Quotient Iteration. There, we defined the *Rayleigh Quotient* (w. r. t. the matrix \mathbf{A}) as a function that maps a vector $\mathbf{x} \in \mathbb{R}^n$ to the scalar

$$\frac{\mathbf{x}^\top \mathbf{A} \mathbf{x}}{\mathbf{x}^\top \mathbf{x}},$$

to replace the fixed shift in the shifted inverse iteration. This is motivated by the following fact.

Lemma 2.1. *Let $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ and define the function $f : \mathbb{R} \rightarrow \mathbb{R}$ by*

$$f(\lambda) := \|(\mathbf{A} - \lambda \mathbf{I})\mathbf{x}\|_2^2.$$

The function f becomes minimal for $\lambda = \mathcal{R}_{\mathbf{A}}(\mathbf{x})$ with value

$$f(\mathcal{R}_{\mathbf{A}}(\mathbf{x})) = \|\mathbf{A}\mathbf{x}\|_2^2 - \mathcal{R}_{\mathbf{A}}(\mathbf{x})^2 \|\mathbf{x}\|_2^2. \quad (2.1) \quad \{\text{eq:minmal_func}$$

Proof. First, note that

$$f(\lambda) = \|\mathbf{A}\mathbf{x} - \lambda\mathbf{x}\|_2^2 = \|\mathbf{A}\mathbf{x}\|_2^2 - 2\lambda\mathbf{x}^\top \mathbf{A}\mathbf{x} + \lambda^2 \|\mathbf{x}\|_2^2.$$

Differentiating this function gives

$$f'(\lambda) = -2\mathbf{x}^\top \mathbf{A}\mathbf{x} + 2\lambda \|\mathbf{x}\|_2^2$$

and letting $f'(\lambda) = 0$ we obtain the only root

$$\lambda = \frac{\mathbf{x}^\top \mathbf{A}\mathbf{x}}{\|\mathbf{x}\|_2^2} = \frac{\mathbf{x}^\top \mathbf{A}\mathbf{x}}{\mathbf{x}^\top \mathbf{x}} = \mathcal{R}_{\mathbf{A}}(\mathbf{x}).$$

Since $f''(\lambda) = 2\|\mathbf{x}\|_2^2 > 0$ this is a minimum and inserting it into f yields the value as given in (2.1). \square

This motivates us to interpret the Rayleigh Quotient as the value that “acts most like an eigenvalue” for \mathbf{x} in the sense of minimizing $\|\mathbf{A}\mathbf{x} - \lambda\mathbf{x}\|_2$. To make this claim more quantitative, first notice that if \mathbf{x} is an eigenvector, then $\mathcal{R}_{\mathbf{A}}(\mathbf{x}) = \lambda$ is the corresponding eigenvalue. Taylor expansion of

\mathcal{R}_A around $\mathbf{x} = \mathbf{v}_j$, where \mathbf{v}_j is the eigenvector associated with the j -th eigenvalue, yields

$$\mathcal{R}_A(\mathbf{x}) = \mathcal{R}_A(\mathbf{q}_j) + (\mathbf{x} - \mathbf{q}_j)^\top \nabla \mathcal{R}_A(\mathbf{q}_j) + \mathcal{O}(\|\mathbf{x} - \mathbf{q}_j\|_2^2). \quad (2.2) \quad \{\text{eq:rq:taylor}\}$$

To compute the gradient $\nabla \mathcal{R}_A(\mathbf{x})$ we compute the partial derivatives w. r. t. to the coordinates x_i , $i = 1, \dots, n$, of \mathbf{x} using the quotient rule

$$\partial_i \mathcal{R}_A(\mathbf{x}) = \partial_i \left(\frac{\mathbf{x}^\top \mathbf{A} \mathbf{x}}{\mathbf{x}^\top \mathbf{x}} \right) = \frac{\partial_i (\mathbf{x}^\top \mathbf{A} \mathbf{x}) \mathbf{x}^\top \mathbf{x} - \mathbf{x}^\top \mathbf{A} \mathbf{x} \partial_i (\mathbf{x}^\top \mathbf{x})}{(\mathbf{x}^\top \mathbf{x})^2},$$

where we used the abbreviation $\partial_i = \frac{\partial}{\partial x_i}$. The first part of the nominator can be computed using the product rule

$$\begin{aligned} \partial_i (\mathbf{x}^\top \mathbf{A} \mathbf{x}) &= \partial_i (\mathbf{x}^\top) \mathbf{A} \mathbf{x} + \mathbf{x}^\top \partial_i (\mathbf{A} \mathbf{x}) \\ &= \mathbf{e}_i^\top \mathbf{A} \mathbf{x} + \mathbf{x}^\top \mathbf{A} \mathbf{e}_i \\ &= \mathbf{e}_i \mathbf{A} \mathbf{x} + \mathbf{A}^\top \mathbf{x} \mathbf{e}_i \\ &= 2 \mathbf{A} \mathbf{x} \mathbf{e}_i = 2 (\mathbf{A} \mathbf{x})_i. \end{aligned}$$

3 Complex Rayleigh Quotient Iteration

A Proofs

{AppendixA}

Bibliography

- [Fis13] Gerd Fischer. *Lineare Algebra - Eine Einführung für Studienanfänger*. 18. Aufl. Berlin Heidelberg New York: Springer-Verlag, 2013.
- [Par98] Beresford N. Parlett. *The Symmetric Eigenvalue Problem*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1998.