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 9, 2019

Contents

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 basics 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 $v \in \mathbb{C}^n$ such that

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}\,. \tag{1.1}$$ {eq:eigvalprobl

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

Rewriting (1.1) gives $\mathbf{A}\mathbf{v} - \lambda\mathbf{v} = 0$ or $(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = 0$. That is, λ is an eigenvalue of $\mathbf{A}$ if, and only if, the matrix $\mathbf{A} - \lambda\mathbf{I}$ is signular.

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. $\mathbf{A} = \overline{\mathbf{A}}^{\mathsf{T}}$, 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 $\mathbf{A} = \overline{\mathbf{A}}^{\mathsf{T}} \in \mathbb{C}^{n\times n}$ be a Hermitian matrix. Denote by $\lambda_1, \lambda_2, \ldots, \lambda_n$ the eigenvalues[1] of $\mathbf{A}$ with associated eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$.*

(i) All eigenvalues of $\mathbf{A}$ are real.

(ii) There exists an orthonormnal 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 $\lambda_1^{-1}, \ldots, \lambda_n^{-1}$ with eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$.

(iv) Let $\sigma \in \mathbb{R}$ an arbitrary scalar. Then the eigenvalues of $\mathbf{A} - \sigma\mathbf{I}$ are $\lambda_i - \sigma$ with eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$.

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].

[1] Of course, the eigenvalues need not to be distinct. But since the eigenvalues of $\mathbf{A}$ are the roots of the n-degree polynomial $\det(\mathbf{A} - t\mathbf{I})$, when counting these roots with their multiplicity, this polynomial has n roots over $\mathbb{C}$. Thus, we can label the eigenvalues from 1 to n.

(iii) Suppose $\mathbf{A}$ is invertible and let $(\lambda, \mathbf{v})$ 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 $\lambda_1, \ldots, \lambda_n$ and $\mathbf{v}_1, \ldots, \mathbf{v}_n$, respectively. We assume that the eigenvalues are normalised w. r. t. the Euclidean norm, i. e.

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

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

$$|\lambda_1| \leqslant |\lambda_2| \leqslant \ldots \leqslant |\lambda_n|\,.$$

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

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

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

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

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*.[2] 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 $\mathbf{LU}$ 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 $\mathbf{LU}$ decomposition $\mathbf{A} - \sigma\mathbf{I} = \mathbf{LU}$
for $k = 1, 2, \ldots$ *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}

This algorithm will always converge to the eigenvector that corresponds to the eigenvalue closest to σ and convergence is also quadratic.

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

Example 1.5. Consider the matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 8 \end{pmatrix}.$$

Example of Power method and inverse iteration

Since at each step, better approximations for the target eigenvector are computed, one could try to obtain approximations to associated eigenvalue and replace the initial shift with this approximations.

One way to obtain such eigenvalue approximations are introduced in the next section.

1.3.3 Rayleigh Quotient Iteration

Here, we only briefly introduce the Rayleigh quotient and the Rayleigh Quotient Iteration. A rigorous motivation and study is given in Chapter 2.

Definition 1.6 (Rayleigh quotient)**.** Let $\mathbf{A} \in \mathbb{C}^{n\times n}$. The mapping

$$\mathcal{R}_{\mathbf{A}} : \mathbb{C}^n \setminus \{\mathbf{0}\} \to \mathbb{C}, \qquad \mathbf{x} \mapsto \frac{\mathbf{x}^H \mathbf{A}\mathbf{x}}{\mathbf{x}^H \mathbf{x}}$$

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

1.4 History

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

2 Classic Rayleigh Quotient Iteration

{chapter:rqi}

In this chapter we thouroughly introduce the Rayleigh Quotient Iteration (or RQI, for short) and present important contributions regarding its analyis. 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 $\mathbf{A} = \mathbf{A}^\intercal \in \mathbb{R}^{n\times n}$ 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

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

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* $f : \mathbb{R} \to \mathbb{R}$ *by*

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

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

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

Proof. First, note that

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

Differentiating this function gives

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

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

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

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). □

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 $\|\boldsymbol{A}\mathbf{x} - \lambda \mathbf{x}\|_2$. To make this claim more quantitative, first notice that if $\mathbf{x}$ is an eigenvector, then $\mathcal{R}_{\boldsymbol{A}}(\mathbf{x}) = \lambda$ 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

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

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

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

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

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

3 Complex Rayleigh Quotient Iteration

Recall, that since the eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of $\mathbf{A}$ form an orthonormal basis of $\mathbb{R}^n$, we can write every $\mathbf{u} \in \mathbb{R}^n$ as

$$\mathbf{u} = \sum_{i=1}^{n} \alpha_i \mathbf{v}_i$$

for certain $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$. Suppose now, that $\mathbf{u}$ is a good approximation for one of the eigenvectors, say for $\mathbf{v}_k$, i. e.

$$\alpha_k \approx 1 \qquad \text{and} \qquad \alpha_j \approx 0 \text{ for } j \neq k\,.$$

Hence, due to the pairwise orthogonality of the eigenvectors

$$\mathbf{u}^\intercal \mathbf{v}_j \approx \begin{cases} 0 & \text{if } j \neq k\,, \\ 1 & \text{if } j = k\,. \end{cases} \tag{3.1}$$ {eq:guess_ortho

As we have seen before, even good approximations of eigenvectors can lead to convergence to the wrong eigenpair when the gap between the wanted eigenvalue and eigenvalues nearby is very small. We can use (3.1) to alter the spectrum of $\mathbf{A}$ in such a way that the eigenvalue of $\mathbf{A}$ corresponding

to $\mathbf{v}_k$ is "raised" out of the spectrum, increasing the distance to adjacent eigenvalues.

First, observe that given a vector $\mathbf{x}$, the projection of $\mathbf{x}$ onto than span of a unit vector $\mathbf{u}$ is

$$(\mathbf{u}^\mathsf{T}\mathbf{x})\mathbf{u} = \mathbf{u}(\mathbf{u}^\mathsf{T}\mathbf{x}) = (\mathbf{u}\mathbf{u}^\mathsf{T})\mathbf{x}\,,$$

Hence the orthogonal complement is

$$\mathbf{x} - (\mathbf{u}^\mathsf{T}\mathbf{x})\mathbf{u} = \mathbf{x} - (\mathbf{u}\mathbf{u}^\mathsf{T})\mathbf{x} = (\mathbf{I} - \mathbf{u}\mathbf{u}^\mathsf{T})\mathbf{x}\,.$$

Now, consider the matrix $\tilde{\mathbf{A}} \coloneqq \mathbf{A} - \gamma i(\mathbf{I} - \mathbf{u}\mathbf{u}^\mathsf{T})$ where $\gamma > 0$ is an arbitrary real number and i denotes the imaginary unit. If $\mathbf{u} = \mathbf{v}_k$ was an eigenvector of $\mathbf{A}$ we would have

$$\begin{aligned}\tilde{\mathbf{A}}\mathbf{v}_j &= \left(\mathbf{A} - \gamma i(\mathbf{I} - \mathbf{v}_k\mathbf{v}_k^\mathsf{T})\right)\mathbf{v}_j \\ &= \mathbf{A}\mathbf{v}_j - \gamma i\mathbf{v}_j + \gamma i\mathbf{v}_k\mathbf{v}_k^\mathsf{T}\mathbf{v}_j \\ &= (\lambda_j - \gamma i)\mathbf{v}_j + \gamma i\delta_{kj}\mathbf{v}_k\,,\end{aligned}$$

where δ_{kj} denotes the Kronecker delta. In other words, if $j \neq k$, we have

$$\tilde{\mathbf{A}}\mathbf{v}_j = (\lambda_j - \gamma i)\mathbf{v}_j$$

and if $j = k$ we have

$$\tilde{\mathbf{A}}\mathbf{v}_j = \tilde{\mathbf{A}}\mathbf{v}_k = (\lambda_k - \gamma i)\mathbf{v}_k - \gamma i\mathbf{u} = \lambda_k\mathbf{v}_k\,.$$

Therefore, if $\mathbf{u}$ is not just an approximation of an eigenvector but we have $\mathbf{u} = \mathbf{v}_k$ for some k we can conclude that $\tilde{\mathbf{A}}$ has the same set of eigenvectors as $\mathbf{A}$ with corresponding eigenvalues $\lambda_j - \gamma i$ if $j \neq k$ and λ_j if $j = k$. In other words, we "raised" all eigenvalues corresponding to eigenvectors that are orthogonal to $\mathbf{u}$ from the real line into the complex plane such that

their imaginary parts are equal to γ.

Lemma 3.1. *Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be a real symmetric matrix and let $\mathbf{u} \in \mathbb{R}^n$ an arbitrary unit vector. We define the matrices*

$$\mathbf{B} \coloneqq (\mathbf{A} - (\sigma - \gamma \mathrm{i})\mathbf{I}) \tag{3.2}$$ {eq:rayleigh_qu

and

$$\mathbf{C} \coloneqq \left(\mathbf{A} - \sigma\mathbf{I} + \gamma\mathrm{i}(\mathbf{I} - \mathbf{u}\mathbf{u}^{\intercal})\right) \tag{3.3}$$ {eq:rayleigh_qu

where i is the imaginary unit and $\sigma, \gamma > 0$ are arbitrary positive real numbers. Then

$$\mathcal{R}_{\mathbf{A}}(\mathbf{B}^{-1}\mathbf{u}) = \mathcal{R}_{\mathbf{A}}(\mathbf{C}^{-1}\mathbf{u}),$$

given that the inverses exist.

Proof. Without loss of generality, we can assume $\sigma = 0$. Otherwise, set $\tilde{\mathbf{A}} = \mathbf{A} - \sigma\mathbf{I}$ and use this matrix instead of $\mathbf{A}$ ($\tilde{\mathbf{A}}$ is obviously still real and symmetric). □

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.