Relations between Rayleigh Quotient Iteration and the Opitz-Larkin Method

Jens-Peter M. Zemke
zemke@tu-harburg.de

Institut für Numerische Simulation
Technische Universität Hamburg-Harburg

2010/06/28, 9:50–10:15

Outline

Rayleigh Quotient Iteration

John William Strutt's RQI

Wielandt's Inverse Iteration

"Modern" RQI

The Opitz-Larkin Method

Classical Root Finding

Schröder's and König's Methods

The Opitz-Larkin Method

The Hessenberg-Matrix Point Of View

. . . and what about Jenkins-Traub?

Outline

Original RQI

In the second edition of the first volume of his book "The Theory of Sound" (Strutt, 1894), John William Strutt, 3rd Baron Rayleigh, included on page 110 the following passage:

Original RQI

In the second edition of the first volume of his book “The Theory of Sound” (Strutt, 1894), John William Strutt, 3rd Baron Rayleigh, included on page 110 the following passage:

> The stationary property of the roots of Lagrange's determinant (3) § 84, suggests a general method of approximating to their values. Beginning with assumed rough approximations to the ratios $A_1 : A_2 : A_3 \ldots\ldots$ we may calculate a first approximation to p^2 from
>
> $$p^2 = \frac{\frac{1}{2}\, c_{11} A_1^{\,2} + \frac{1}{2}\, c_{22} A_2^{\,2} + \ldots + c_{12} A_1 A_2 + \ldots}{\frac{1}{2}\, a_{11} A_1^{\,2} + \frac{1}{2}\, a_{22} A_2^{\,2} + \ldots + a_{12} A_1 A_2 + \ldots} \quad \ldots\ldots \text{(3).}$$
>
> With this value of p^2 we may recalculate the ratios $A_1 : A_2 \ldots$ from any $(m-1)$ of equations (5) § 84, then again by application of (3) determine an improved value of p^2, and so on.]

Original RQI

In modern notation, Lord Rayleigh starts with an approximate eigenvector $\mathbf{v}_k$, $k = 0$, of a Hermitean matrix (Hermitean pencil), computes its Rayleigh quotient

$$\rho(\mathbf{v}_k) := \frac{\mathbf{v}_k^{\mathsf{H}} \mathbf{A} \mathbf{v}_k}{\mathbf{v}_k^{\mathsf{H}} \mathbf{v}_k},$$

Original RQI

In modern notation, Lord Rayleigh starts with an approximate eigenvector $\mathbf{v}_k$, $k=0$, of a Hermitean matrix (Hermitean pencil), computes its Rayleigh quotient

$$\rho(\mathbf{v}_k) := \frac{\mathbf{v}_k^{\mathsf{H}}\mathbf{A}\mathbf{v}_k}{\mathbf{v}_k^{\mathsf{H}}\mathbf{v}_k},$$

and iterates for some suitably chosen $j \in \{1,2,\ldots,n\}$,

$$\mathbf{v}_{k+1} = \frac{(\mathbf{A}-\rho(\mathbf{v}_k)\mathbf{I}_n)^{-1}\mathbf{e}_j}{\|(\mathbf{A}-\rho(\mathbf{v}_k)\mathbf{I}_n)^{-1}\mathbf{e}_j\|}, \quad k=0,1,\ldots$$

where j may vary, depending on the computed approximate eigenvector.

Original RQI

In modern notation, Lord Rayleigh starts with an approximate eigenvector $\mathbf{v}_k$, $k = 0$, of a Hermitean matrix (Hermitean pencil), computes its Rayleigh quotient

$$\rho(\mathbf{v}_k) := \frac{\mathbf{v}_k^{\mathsf{H}} \mathbf{A} \mathbf{v}_k}{\mathbf{v}_k^{\mathsf{H}} \mathbf{v}_k},$$

and iterates for some suitably chosen $j \in \{1, 2, \ldots, n\}$,

$$\mathbf{v}_{k+1} = \frac{(\mathbf{A} - \rho(\mathbf{v}_k)\mathbf{I}_n)^{-1}\mathbf{e}_j}{\|(\mathbf{A} - \rho(\mathbf{v}_k)\mathbf{I}_n)^{-1}\mathbf{e}_j\|}, \quad k = 0, 1, \ldots$$

where j may vary, depending on the computed approximate eigenvector.

The Rayleigh quotient uniquely solves the least squares problem

$$\rho(\mathbf{v}_k) = \operatorname{argmin}_{\rho \in \mathbb{C}} \|\mathbf{A}\mathbf{v}_k - \mathbf{v}_k \rho\|.$$

Inverse Iteration

Closely connected to RQI is inverse iteration (Wielandt, 1944).

Inverse Iteration

Closely connected to RQI is inverse iteration (Wielandt, 1944). In its most basic variant the shift τ is never updated, but the right-hand side is replaced by the latest approximate eigenvector:

$$\mathbf{v}_{k+1} = \frac{(\mathbf{A} - \tau\mathbf{I}_n)^{-1}\mathbf{v}_k}{\|(\mathbf{A} - \tau\mathbf{I}_n)^{-1}\mathbf{v}_k\|}, \quad k = 0, 1, \ldots$$

Inverse Iteration

Closely connected to RQI is inverse iteration (Wielandt, 1944). In its most basic variant the shift τ is never updated, but the right-hand side is replaced by the latest approximate eigenvector:

$$\mathbf{v}_{k+1} = \frac{(\mathbf{A} - \tau \mathbf{I}_n)^{-1} \mathbf{v}_k}{\|(\mathbf{A} - \tau \mathbf{I}_n)^{-1} \mathbf{v}_k\|}, \quad k = 0, 1, \ldots$$

The shift can be updated by using the approximate eigenvalues obtained by the shift update strategy

$$\tau_{k+1} := \tau_k + \frac{1}{\mathbf{e}_j^\mathsf{T} (\mathbf{A} - \tau_k \mathbf{I}_n)^{-1} \mathbf{v}_k}.$$

Inverse Iteration

Closely connected to RQI is inverse iteration (Wielandt, 1944). In its most basic variant the shift τ is never updated, but the right-hand side is replaced by the latest approximate eigenvector:

$$\mathbf{v}_{k+1} = \frac{(\mathbf{A} - \tau \mathbf{I}_n)^{-1}\mathbf{v}_k}{\|(\mathbf{A} - \tau \mathbf{I}_n)^{-1}\mathbf{v}_k\|}, \quad k = 0, 1, \ldots$$

The shift can be updated by using the approximate eigenvalues obtained by the shift update strategy

$$\tau_{k+1} := \tau_k + \frac{1}{\mathbf{e}_j^\mathsf{T}(\mathbf{A} - \tau_k \mathbf{I}_n)^{-1}\mathbf{v}_k}.$$

The latter variant is described in (Wielandt, 1944, Seite 9, Formel (20)) and converges locally quadratically.

Modern variants of RQI

Combination gives (symmetric/Hermitean) RQI:

$$\mathbf{v}_{k+1} = \frac{(\mathbf{A} - \rho(\mathbf{v}_k)\mathbf{I}_n)^{-1}\mathbf{v}_k}{\|(\mathbf{A} - \rho(\mathbf{v}_k)\mathbf{I}_n)^{-1}\mathbf{v}_k\|}, \quad k = 0, 1, \ldots$$

This iteration is also used for nonsymmetric $\mathbf{A}$.

Modern variants of RQI

Combination gives (symmetric/Hermitean) RQI:

$$\mathbf{v}_{k+1} = \frac{(\mathbf{A} - \rho(\mathbf{v}_k)\mathbf{I}_n)^{-1}\mathbf{v}_k}{\|(\mathbf{A} - \rho(\mathbf{v}_k)\mathbf{I}_n)^{-1}\mathbf{v}_k\|}, \quad k = 0, 1, \ldots$$

This iteration is also used for nonsymmetric $\mathbf{A}$.

Crandall was the first who investigated the three variants (the original Rayleigh quotient iteration; inverse iteration with fixed shift; symmetric RQI), see (Crandall, 1951).

Modern variants of RQI

Combination gives (symmetric/Hermitean) RQI:

$$\mathbf{v}_{k+1} = \frac{(\mathbf{A} - \rho(\mathbf{v}_k)\mathbf{I}_n)^{-1}\mathbf{v}_k}{\|(\mathbf{A} - \rho(\mathbf{v}_k)\mathbf{I}_n)^{-1}\mathbf{v}_k\|}, \quad k = 0, 1, \ldots$$

This iteration is also used for nonsymmetric $\mathbf{A}$.

Crandall was the first who investigated the three variants (the original Rayleigh quotient iteration; inverse iteration with fixed shift; symmetric RQI), see (Crandall, 1951).

Ostrowski proved that unsymmetric RQI still has a quadratic convergence rate, (Ostrowski, 1959e). In (Ostrowski, 1959c), he devised two-sided RQI:

$$\rho(\mathbf{w}_k, \mathbf{v}_k) := \frac{\mathbf{w}_k^{\mathsf{H}}\mathbf{A}\mathbf{v}_k}{\mathbf{w}_k^{\mathsf{H}}\mathbf{v}_k}, \qquad \begin{aligned} \mathbf{v}_{k+1} &= (\mathbf{A} - \rho(\mathbf{w}_k, \mathbf{v}_k)\mathbf{I}_n)^{-1}\mathbf{v}_k, \\ \mathbf{w}_{k+1} &= (\mathbf{A} - \rho(\mathbf{w}_k, \mathbf{v}_k)\mathbf{I}_n)^{-\mathsf{H}}\mathbf{w}_k, \end{aligned} \qquad k = 0, 1, \ldots$$

Modern variants of RQI

Combination gives (symmetric/Hermitean) RQI:

$$\mathbf{v}_{k+1} = \frac{(\mathbf{A} - \rho(\mathbf{v}_k)\mathbf{I}_n)^{-1}\mathbf{v}_k}{\|(\mathbf{A} - \rho(\mathbf{v}_k)\mathbf{I}_n)^{-1}\mathbf{v}_k\|}, \quad k = 0, 1, \ldots$$

This iteration is also used for nonsymmetric $\mathbf{A}$.

Crandall was the first who investigated the three variants (the original Rayleigh quotient iteration; inverse iteration with fixed shift; symmetric RQI), see (Crandall, 1951).

Ostrowski proved that unsymmetric RQI still has a quadratic convergence rate, (Ostrowski, 1959e). In (Ostrowski, 1959c), he devised two-sided RQI:

$$\rho(\mathbf{w}_k, \mathbf{v}_k) := \frac{\mathbf{w}_k^{\mathsf{H}}\mathbf{A}\mathbf{v}_k}{\mathbf{w}_k^{\mathsf{H}}\mathbf{v}_k}, \qquad \begin{aligned} \mathbf{v}_{k+1} &= (\mathbf{A} - \rho(\mathbf{w}_k, \mathbf{v}_k)\mathbf{I}_n)^{-1}\mathbf{v}_k, \\ \mathbf{w}_{k+1} &= (\mathbf{A} - \rho(\mathbf{w}_k, \mathbf{v}_k)\mathbf{I}_n)^{-\mathsf{H}}\mathbf{w}_k, \end{aligned} \qquad k = 0, 1, \ldots$$

This trick recovers the cubic convergence rate of RQI at the expense of an additional system.

Modern variants of RQI

Combination gives (symmetric/Hermitean) RQI:

$$\mathbf{v}_{k+1} = \frac{(\mathbf{A} - \rho(\mathbf{v}_k)\mathbf{I}_n)^{-1}\mathbf{v}_k}{\|(\mathbf{A} - \rho(\mathbf{v}_k)\mathbf{I}_n)^{-1}\mathbf{v}_k\|}, \quad k = 0, 1, \ldots$$

This iteration is also used for nonsymmetric $\mathbf{A}$.

Crandall was the first who investigated the three variants (the original Rayleigh quotient iteration; inverse iteration with fixed shift; symmetric RQI), see (Crandall, 1951).

Ostrowski proved that unsymmetric RQI still has a quadratic convergence rate, (Ostrowski, 1959e). In (Ostrowski, 1959c), he devised two-sided RQI:

$$\rho(\mathbf{w}_k, \mathbf{v}_k) := \frac{\mathbf{w}_k^{\mathsf{H}}\mathbf{A}\mathbf{v}_k}{\mathbf{w}_k^{\mathsf{H}}\mathbf{v}_k}, \qquad \begin{aligned} \mathbf{v}_{k+1} &= (\mathbf{A} - \rho(\mathbf{w}_k, \mathbf{v}_k)\mathbf{I}_n)^{-1}\mathbf{v}_k, \\ \mathbf{w}_{k+1} &= (\mathbf{A} - \rho(\mathbf{w}_k, \mathbf{v}_k)\mathbf{I}_n)^{-\mathsf{H}}\mathbf{w}_k, \end{aligned} \qquad k = 0, 1, \ldots$$

This trick recovers the cubic convergence rate of RQI at the expense of an additional system. Parlett's alternating RQI preserves monotonicity.

Outline

Classical methods

Methods for the computation of a root of a rational function

$$f : \mathbb{C} \to \mathbb{C}, \quad f(z) := \frac{p(z)}{q(z)}, \quad p, q \in \mathbb{P}_m$$

include Newton's method

$$z_{k+1} = z_k - \frac{f(z_k)}{f'(z_k)}$$

and the secant method:

$$z_{k+1} = z_k - \frac{f(z_k)}{[z_k, z_{k-1}]f}.$$

Classical methods

Methods for the computation of a root of a rational function

$$f : \mathbb{C} \to \mathbb{C}, \quad f(z) := \frac{p(z)}{q(z)}, \quad p, q \in \mathbb{P}_m$$

include Newton's method

$$z_{k+1} = z_k - \frac{f(z_k)}{f'(z_k)}$$

and the secant method:

$$z_{k+1} = z_k - \frac{f(z_k)}{[z_k, z_{k-1}]f}.$$

The secant method has R-order of convergence given by the golden ratio

$$\phi := \frac{1+\sqrt{5}}{2} \approx 1.618.$$

Classical methods

Methods for the computation of a root of a rational function

$$f : \mathbb{C} \to \mathbb{C}, \quad f(z) := \frac{p(z)}{q(z)}, \quad p, q \in \mathbb{P}_m$$

include Newton's method

$$z_{k+1} = z_k - \frac{f(z_k)}{f'(z_k)}$$

and the secant method:

$$z_{k+1} = z_k - \frac{f(z_k)}{[z_k, z_{k-1}]f}.$$

The secant method has R-order of convergence given by the golden ratio

$$\phi := \frac{1+\sqrt{5}}{2} \approx 1.618.$$

Two steps of the secant method are as costly as one step of Newton's method. This makes the secant method the winner:

$$\phi^2 = \phi + 1 \approx 2.618 > 2.$$

Schröder's and König's methods

Newton's method has been generalized to incorporate higher order derivatives and to exhibit a higher order of convergence. Well-known generalized Newton's methods are Halley's and Laguerre's methods.

Schröder's and König's methods

Newton's method has been generalized to incorporate higher order derivatives and to exhibit a higher order of convergence. Well-known generalized Newton's methods are Halley's and Laguerre's methods.

In 1870 E. Schröder from Pforzheim came up with two infinite families of generalizations (Schröder, 1870).

Schröder's and König's methods

Newton's method has been generalized to incorporate higher order derivatives and to exhibit a higher order of convergence. Well-known generalized Newton's methods are Halley's and Laguerre's methods.

In 1870 E. Schröder from Pforzheim came up with two infinite families of generalizations (Schröder, 1870). In 1884 Julius König proved a theorem on the limiting behavior of certain ratios of Taylor coefficients (König, 1884), enabling a simpler derivation of Schröder's family A_ω^λ with $\lambda = 0$.

Schröder's and König's methods

Newton's method has been generalized to incorporate higher order derivatives and to exhibit a higher order of convergence. Well-known generalized Newton's methods are Halley's and Laguerre's methods.

In 1870 E. Schröder from Pforzheim came up with two infinite families of generalizations (Schröder, 1870). In 1884 Julius König proved a theorem on the limiting behavior of certain ratios of Taylor coefficients (König, 1884), enabling a simpler derivation of Schröder's family A_ω^λ with $\lambda = 0$.

This family is nowadays known as "König's method":

$$z_{k+1} = z_k + s \frac{(1/f)^{(s-1)}(z_k)}{(1/f)^{(s)}(z_k)}, \quad s = 1, 2, \ldots$$

Schröder's and König's methods

Newton's method has been generalized to incorporate higher order derivatives and to exhibit a higher order of convergence. Well-known generalized Newton's methods are Halley's and Laguerre's methods.

In 1870 E. Schröder from Pforzheim came up with two infinite families of generalizations (Schröder, 1870). In 1884 Julius König proved a theorem on the limiting behavior of certain ratios of Taylor coefficients (König, 1884), enabling a simpler derivation of Schröder's family A_ω^λ with $\lambda = 0$.

This family is nowadays known as "König's method":

$$z_{k+1} = z_k + s\frac{(1/f)^{(s-1)}(z_k)}{(1/f)^{(s)}(z_k)}, \quad s = 1, 2, \ldots$$

König's method for $s = 1$ is Newton's method,

$$z_{k+1} = z_k + \frac{(1/f)(z_k)}{(1/f)'(z_k)} = z_k - \frac{1/f(z_k)}{f'(z_k)/(f(z_k))^2} = z_k - \frac{f(z_k)}{f'(z_k)}.$$

The Opitz-Larkin method

There is a natural extension of König's method using divided differences in place of the derivatives.

The Opitz-Larkin method

There is a natural extension of König’s method using divided differences in place of the derivatives. This natural extension (without the connection to König’s method) was published in 1958 by Günter Opitz in a two-page article in ZAMM.

The Opitz-Larkin method

There is a natural extension of König's method using divided differences in place of the derivatives. This natural extension (without the connection to König's method) was published in 1958 by Günter Opitz in a two-page article in ZAMM.

He published few additional papers on the subject (including his most famous "Steigungsmatrizen" paper). A more complete presentation can be found in his "Habilitationsschrift". There, he even pointed out the connection to König's method.

The Opitz-Larkin method

There is a natural extension of König's method using divided differences in place of the derivatives. This natural extension (without the connection to König's method) was published in 1958 by Günter Opitz in a two-page article in ZAMM.

He published few additional papers on the subject (including his most famous "Steigungsmatrizen" paper). A more complete presentation can be found in his "Habilitationsschrift". There, he even pointed out the connection to König's method.

Independently, 23 years later F. M. Larkin re-developed Opitz' method, see (Larkin, 1981) and the predecessor (Larkin, 1980).

The Opitz-Larkin method

There is a natural extension of König's method using divided differences in place of the derivatives. This natural extension (without the connection to König's method) was published in 1958 by Günter Opitz in a two-page article in ZAMM.

He published few additional papers on the subject (including his most famous "Steigungsmatrizen" paper). A more complete presentation can be found in his "Habilitationsschrift". There, he even pointed out the connection to König's method.

Independently, 23 years later F. M. Larkin re-developed Opitz' method, see (Larkin, 1981) and the predecessor (Larkin, 1980).

We will refer to this method as the Opitz-Larkin method. The Opitz-Larkin method is based on iterations of the form

$$x_{k+1} = z_k + \frac{[z_1, z_2, \ldots, z_{k-1}](1/f)}{[z_1, z_2, \ldots, z_{k-1}, z_k](1/f)}.$$

The Opitz-Larkin method

Mostly, the z_i are all distinct and the next iterate is used as new evaluation point $z_{k+1} = x_{k+1}$,

$$z_{k+1} = z_k + \frac{[z_1, z_2, \ldots, z_{k-1}](1/f)}{[z_1, z_2, \ldots, z_{k-1}, z_k](1/f)}.$$

The Opitz-Larkin method

Mostly, the z_i are all distinct and the next iterate is used as new evaluation point $z_{k+1} = x_{k+1}$,

$$z_{k+1} = z_k + \frac{[z_1, z_2, \ldots, z_{k-1}](1/f)}{[z_1, z_2, \ldots, z_{k-1}, z_k](1/f)}.$$

This variant of the Opitz-Larkin method converges with R-order 2.

The Opitz-Larkin method

Mostly, the z_i are all distinct and the next iterate is used as new evaluation point $z_{k+1} = x_{k+1}$,

$$z_{k+1} = z_k + \frac{[z_1, z_2, \dots, z_{k-1}](1/f)}{[z_1, z_2, \dots, z_{k-1}, z_k](1/f)}.$$

This variant of the Opitz-Larkin method converges with R-order 2.

Frequently, the Opitz-Larkin method is used with truncation:

$$z_{k+1} = z_k + \frac{[z_{k-p}, \dots, z_{k-1}](1/f)}{[z_{k-p}, \dots, z_{k-1}, z_k](1/f)},$$

see (Opitz, 1958, Seite 277, Gleichung (9)) and (Larkin, 1981, Section 4, pages 98–99).

The Opitz-Larkin method

It is possible to use confluent divided differences, i.e., multiple points of evaluation, i.e., higher order derivatives of $1/f$.

The Opitz-Larkin method

It is possible to use confluent divided differences, i.e., multiple points of evaluation, i.e., higher order derivatives of $1/f$.

When we use only confluent divided differences in the truncated Opitz-Larkin method with truncation parameter $p = s$, we recover König's method:

$$z_{k+1} = z_k + \frac{[\overbrace{z_k, \dots, z_k}^{s}](1/f)}{[\underbrace{z_k, \dots, z_k, z_k}_{s+1}](1/f)}$$

$$= z_k + \frac{(1/f)^{(s-1)}(z_k)/(s-1)!}{(1/f)^{(s)}(z_k)/s!} = z_k + s\frac{(1/f)^{(s-1)}(z_k)}{(1/f)^{(s)}(z_k)}.$$

The Opitz-Larkin method

Truncated Opitz-Larkin with $p = 1$ is the secant method,

$$\begin{aligned} z_{k+1} &= z_k + \frac{[z_{k-1}](1/f)}{[z_{k-1}, z_k](1/f)} \\ &= z_k + \frac{1}{f(z_{k-1})} \cdot \frac{z_{k-1} - z_k}{1/f(z_{k-1}) - 1/f(z_k)} \\ &= z_k + \frac{f(z_k)f(z_{k-1})}{f(z_{k-1})} \cdot \frac{z_{k-1} - z_k}{f(z_k) - f(z_{k-1})} \\ &= z_k - \frac{f(z_k)}{[z_{k-1}, z_k]f}. \end{aligned}$$

The Opitz-Larkin method

Truncated Opitz-Larkin with $p = 1$ is the secant method,

$$\begin{aligned} z_{k+1} &= z_k + \frac{[z_{k-1}](1/f)}{[z_{k-1}, z_k](1/f)} \\ &= z_k + \frac{1}{f(z_{k-1})} \cdot \frac{z_{k-1} - z_k}{1/f(z_{k-1}) - 1/f(z_k)} \\ &= z_k + \frac{f(z_k)f(z_{k-1})}{f(z_{k-1})} \cdot \frac{z_{k-1} - z_k}{f(z_k) - f(z_{k-1})} \\ &= z_k - \frac{f(z_k)}{[z_{k-1}, z_k]f}. \end{aligned}$$

Confluent truncated Opitz-Larkin with $p = 1$ is Newton's method.

The Opitz-Larkin method

In general, the Opitz-Larkin method is closely connected to rational interpolation of the inverse function (Larkin, 1981, Theorem 1, page 96):

The Opitz-Larkin method

In general, the Opitz-Larkin method is closely connected to rational interpolation of the inverse function (Larkin, 1981, Theorem 1, page 96):

Theorem (Larkin 1981)

If, for any integer $k > 1$*, there exists a rational function of the form*

$$r_k(z) = \frac{q_d(z)}{z - \alpha}, \quad \forall\, z,$$

where q_d *is a polynomial of degree* $d \leqslant k - 2$*, such that* $q_d(\alpha) \neq 0$ *and*

$$r_k(z_j) = f(z_j)^{-1}, \quad j = 1, 2, \ldots, k,$$

then

$$z_k + \frac{[z_1, z_2, \ldots, z_{k-1}](1/f)}{[z_1, z_2, \ldots, z_{k-1}, z_k](1/f)} = \alpha.$$

Outline

Simplification

By the implicit Q-Theorem we obtain a unique Hessenberg matrix given nonderogatory $\mathbf{A} \in \mathbb{C}^{n\times n}$ and $\mathbf{q} \in \mathbb{C}^n$ if we fix the signs of the elements in the lower diagonal, e.g., to be non-negative real.

Simplification

By the implicit Q-Theorem we obtain a unique Hessenberg matrix given nonderogatory $\mathbf{A} \in \mathbb{C}^{n \times n}$ and $\mathbf{q} \in \mathbb{C}^n$ if we fix the signs of the elements in the lower diagonal, e.g., to be non-negative real.

We use the implicit Q-Theorem to unitarily transform the pair $(\mathbf{A}, \mathbf{q})$ with $\|\mathbf{q}\|_2 = 1$ to the pair $(\mathbf{H}_n, \mathbf{e}_1)$, where $\mathbf{H}_n$ is upper Hessenberg and $\mathbf{e}_1$ denotes the first standard unit vector.

Simplification

By the implicit Q-Theorem we obtain a unique Hessenberg matrix given nonderogatory $\mathbf{A} \in \mathbb{C}^{n\times n}$ and $\mathbf{q} \in \mathbb{C}^n$ if we fix the signs of the elements in the lower diagonal, e.g., to be non-negative real.

We use the implicit Q-Theorem to unitarily transform the pair $(\mathbf{A}, \mathbf{q})$ with $\|\mathbf{q}\|_2 = 1$ to the pair $(\mathbf{H}_n, \mathbf{e}_1)$, where $\mathbf{H}_n$ is upper Hessenberg and $\mathbf{e}_1$ denotes the first standard unit vector.

The following Matlab-code gives the transformed pair:

```
[Q,R] = qr(q);
[P,H] = hess(Q'*A*Q);
```

Simplification

By the implicit Q-Theorem we obtain a unique Hessenberg matrix given nonderogatory $\mathbf{A} \in \mathbb{C}^{n \times n}$ and $\mathbf{q} \in \mathbb{C}^n$ if we fix the signs of the elements in the lower diagonal, e.g., to be non-negative real.

We use the implicit Q-Theorem to unitarily transform the pair $(\mathbf{A}, \mathbf{q})$ with $\|\mathbf{q}\|_2 = 1$ to the pair $(\mathbf{H}_n, \mathbf{e}_1)$, where $\mathbf{H}_n$ is upper Hessenberg and $\mathbf{e}_1$ denotes the first standard unit vector.

The following Matlab-code gives the transformed pair:

```
[Q,R] = qr(q);
[P,H] = hess(Q'*A*Q);
signs = sign(diag(H,-1));
S = diag(cumprod([1;signs]));
H = S'*H*S;
```

Simplification

By the implicit Q-Theorem we obtain a unique Hessenberg matrix given nonderogatory $\mathbf{A} \in \mathbb{C}^{n \times n}$ and $\mathbf{q} \in \mathbb{C}^n$ if we fix the signs of the elements in the lower diagonal, e.g., to be non-negative real.

We use the implicit Q-Theorem to unitarily transform the pair $(\mathbf{A}, \mathbf{q})$ with $\|\mathbf{q}\|_2 = 1$ to the pair $(\mathbf{H}_n, \mathbf{e}_1)$, where $\mathbf{H}_n$ is upper Hessenberg and $\mathbf{e}_1$ denotes the first standard unit vector.

The following Matlab-code gives the transformed pair:

```
[Q,R] = qr(q);
[P,H] = hess(Q'*A*Q);
signs = sign(diag(H,-1));
S = diag(cumprod([1;signs]));
H = S'*H*S;
```

Any left vector is modified accordingly.

Simplification

We set ${}^{z}\mathbf{H}_n := (z\mathbf{I}_n - \mathbf{H}_n)$. By the first resolvent identity (Chatelin, 1993)

$$({}^{z_1}\mathbf{H}_n)^{-1}({}^{z_2}\mathbf{H}_n)^{-1} = (z_1\mathbf{I}_n - \mathbf{H}_n)^{-1}(z_2\mathbf{I}_n - \mathbf{H}_n)^{-1} \tag{1a}$$

$$= \frac{({}^{z_1}\mathbf{H}_n)^{-1} - ({}^{z_2}\mathbf{H}_n)^{-1}}{z_2 - z_1} = -[z_1, z_2]({}^{z}\mathbf{H}_n)^{-1}. \tag{1b}$$

Simplification

We set ${}^{z}\mathbf{H}_n := (z\mathbf{I}_n - \mathbf{H}_n)$. By the first resolvent identity (Chatelin, 1993)

$$({}^{z_1}\mathbf{H}_n)^{-1}({}^{z_2}\mathbf{H}_n)^{-1} = (z_1\mathbf{I}_n - \mathbf{H}_n)^{-1}(z_2\mathbf{I}_n - \mathbf{H}_n)^{-1} \tag{1a}$$

$$= \frac{({}^{z_1}\mathbf{H}_n)^{-1} - ({}^{z_2}\mathbf{H}_n)^{-1}}{z_2 - z_1} = -[z_1, z_2]({}^{z}\mathbf{H}_n)^{-1}. \tag{1b}$$

The first resolvent identity is based on the trivial observation that

$$(z_2\mathbf{I}_n - \mathbf{H}_n) - (z_1\mathbf{I}_n - \mathbf{H}_n) = (z_2 - z_1)\mathbf{I}_n.$$

Simplification

We set ${}^{z}\mathbf{H}_n := (z\mathbf{I}_n - \mathbf{H}_n)$. By the first resolvent identity (Chatelin, 1993)

$$({}^{z_1}\mathbf{H}_n)^{-1}({}^{z_2}\mathbf{H}_n)^{-1} = (z_1\mathbf{I}_n - \mathbf{H}_n)^{-1}(z_2\mathbf{I}_n - \mathbf{H}_n)^{-1} \tag{1a}$$

$$= \frac{({}^{z_1}\mathbf{H}_n)^{-1} - ({}^{z_2}\mathbf{H}_n)^{-1}}{z_2 - z_1} = -[z_1, z_2]({}^{z}\mathbf{H}_n)^{-1}. \tag{1b}$$

The first resolvent identity is based on the trivial observation that

$$(z_2\mathbf{I}_n - \mathbf{H}_n) - (z_1\mathbf{I}_n - \mathbf{H}_n) = (z_2 - z_1)\mathbf{I}_n.$$

Generalization (see also (Dekker and Traub, 1971)):

$$\prod_{i=1}^{k}({}^{z_i}\mathbf{H}_n)^{-1} = (-1)^{k-1}[z_1, \ldots, z_k]({}^{z}\mathbf{H}_n)^{-1}. \tag{2}$$

Simplification

We set ${}^{z}\mathbf{H}_n := (z\mathbf{I}_n - \mathbf{H}_n)$. By the first resolvent identity (Chatelin, 1993)

$$({}^{z_1}\mathbf{H}_n)^{-1}({}^{z_2}\mathbf{H}_n)^{-1} = (z_1\mathbf{I}_n - \mathbf{H}_n)^{-1}(z_2\mathbf{I}_n - \mathbf{H}_n)^{-1} \tag{1a}$$

$$= \frac{({}^{z_1}\mathbf{H}_n)^{-1} - ({}^{z_2}\mathbf{H}_n)^{-1}}{z_2 - z_1} = -[z_1, z_2]({}^{z}\mathbf{H}_n)^{-1}. \tag{1b}$$

The first resolvent identity is based on the trivial observation that

$$(z_2\mathbf{I}_n - \mathbf{H}_n) - (z_1\mathbf{I}_n - \mathbf{H}_n) = (z_2 - z_1)\mathbf{I}_n.$$

Generalization (see also (Dekker and Traub, 1971)):

$$\prod_{i=1}^{k}({}^{z_i}\mathbf{H}_n)^{-1} = (-1)^{k-1}[z_1, \ldots, z_k]({}^{z}\mathbf{H}_n)^{-1}. \tag{2}$$

Confluent divided differences are well-defined.

Simplification

For simplicity we assume that $\mathbf{H}_n$ is unreduced.

Simplification

For simplicity we assume that $\mathbf{H}_n$ is unreduced. We denote products of sub-diagonal elements of the unreduced Hessenberg matrices $\mathbf{H}_n \in \mathbb{C}^{n\times n}$ by

$$h_{i:j} := \prod_{\ell=i}^{j} h_{\ell+1,\ell}.$$

Simplification

For simplicity we assume that $\mathbf{H}_n$ is unreduced. We denote products of sub-diagonal elements of the unreduced Hessenberg matrices $\mathbf{H}_n \in \mathbb{C}^{n \times n}$ by

$$h_{i:j} := \prod_{\ell=i}^{j} h_{\ell+1,\ell}.$$

Polynomial vectors $\boldsymbol{\nu}$ and $\check{\boldsymbol{\nu}}$ are defined by

$$\boldsymbol{\nu}(z) := \left(\frac{\chi_{j+1:n}(z)}{h_{j:n-1}}\right)_{j=1}^{n} \quad \text{and} \quad \check{\boldsymbol{\nu}}(z) := \left(\frac{\chi_{1:j-1}(z)}{h_{1:j-1}}\right)_{j=1}^{n}. \tag{3}$$

Simplification

For simplicity we assume that $\mathbf{H}_n$ is unreduced. We denote products of sub-diagonal elements of the unreduced Hessenberg matrices $\mathbf{H}_n \in \mathbb{C}^{n\times n}$ by

$$h_{i:j} := \prod_{\ell=i}^{j} h_{\ell+1,\ell}.$$

Polynomial vectors $\boldsymbol{\nu}$ and $\check{\boldsymbol{\nu}}$ are defined by

$$\boldsymbol{\nu}(z) := \left(\frac{\chi_{j+1:n}(z)}{h_{j:n-1}}\right)_{j=1}^{n} \quad \text{and} \quad \check{\boldsymbol{\nu}}(z) := \left(\frac{\chi_{1:j-1}(z)}{h_{1:j-1}}\right)_{j=1}^{n}. \tag{3}$$

The elements are $\nu_j(z)$ and $\check{\nu}_j(z)$, $j = 1, \ldots, n$. Observe that $\nu_n \equiv 1 \equiv \check{\nu}_1$.

Simplification

For simplicity we assume that $\mathbf{H}_n$ is unreduced. We denote products of sub-diagonal elements of the unreduced Hessenberg matrices $\mathbf{H}_n \in \mathbb{C}^{n\times n}$ by

$$h_{i:j} := \prod_{\ell=i}^{j} h_{\ell+1,\ell}.$$

Polynomial vectors $\boldsymbol{\nu}$ and $\check{\boldsymbol{\nu}}$ are defined by

$$\boldsymbol{\nu}(z) := \left(\frac{\chi_{j+1:n}(z)}{h_{j:n-1}}\right)_{j=1}^{n} \quad \text{and} \quad \check{\boldsymbol{\nu}}(z) := \left(\frac{\chi_{1:j-1}(z)}{h_{1:j-1}}\right)_{j=1}^{n}. \tag{3}$$

The elements are $\nu_j(z)$ and $\check{\nu}_j(z)$, $j = 1, \ldots, n$. Observe that $\nu_n \equiv 1 \equiv \check{\nu}_1$.

The polynomials $\chi_{i:j}$ are the characteristic polynomials of submatrices of $\mathbf{H}_n$,

$$\chi_{i:j}(z) := \det\left({}^{z}\mathbf{H}_{i:j}\right) = \det\left(z\mathbf{I}_{j-i+1} - \mathbf{H}_{i:j}\right).$$

Simplification

For z in the resolvent set

$$({}^{z}\mathbf{H}_{n})\boldsymbol{\nu}(z) = \frac{\chi(z)}{h_{1:n-1}}\mathbf{e}_1 \quad \Leftrightarrow \quad \frac{\boldsymbol{\nu}(z)h_{1:n-1}}{\chi(z)} = ({}^{z}\mathbf{H}_{n})^{-1}\mathbf{e}_1, \tag{4a}$$

$$\check{\boldsymbol{\nu}}(z)^{\mathsf{T}}({}^{z}\mathbf{H}_{n}) = \mathbf{e}_n^{\mathsf{T}}\frac{\chi(z)}{h_{1:n-1}} \quad \Leftrightarrow \quad \frac{h_{1:n-1}\check{\boldsymbol{\nu}}(z)^{\mathsf{T}}}{\chi(z)} = \mathbf{e}_n^{\mathsf{T}}({}^{z}\mathbf{H}_{n})^{-1}. \tag{4b}$$

Simplification

For z in the resolvent set

$$({}^{z}\mathbf{H}_n)\boldsymbol{\nu}(z) = \frac{\chi(z)}{h_{1:n-1}}\mathbf{e}_1 \quad \Leftrightarrow \quad \frac{\boldsymbol{\nu}(z)h_{1:n-1}}{\chi(z)} = ({}^{z}\mathbf{H}_n)^{-1}\mathbf{e}_1, \tag{4a}$$

$$\check{\boldsymbol{\nu}}(z)^{\mathsf{T}}({}^{z}\mathbf{H}_n) = \mathbf{e}_n^{\mathsf{T}}\frac{\chi(z)}{h_{1:n-1}} \quad \Leftrightarrow \quad \frac{h_{1:n-1}\check{\boldsymbol{\nu}}(z)^{\mathsf{T}}}{\chi(z)} = \mathbf{e}_n^{\mathsf{T}}({}^{z}\mathbf{H}_n)^{-1}. \tag{4b}$$

The repeated application of resolvents to $\mathbf{e}_1$ results in

$$\left(\prod_{i=1}^{k}({}^{z_i}\mathbf{H}_n)^{-1}\right)\mathbf{e}_1 = (-1)^{k-1}[z_1,\dots,z_k]({}^{z}\mathbf{H}_n)^{-1}\mathbf{e}_1 \tag{5}$$

$$= (-1)^{k-1}[z_1,\dots,z_k]\frac{\boldsymbol{\nu}(z)h_{1:n-1}}{\chi(z)}. \tag{6}$$

Simplification

For z in the resolvent set

$$({}^{z}\mathbf{H}_n)\boldsymbol{\nu}(z) = \frac{\chi(z)}{h_{1:n-1}}\mathbf{e}_1 \quad \Leftrightarrow \quad \frac{\boldsymbol{\nu}(z)h_{1:n-1}}{\chi(z)} = ({}^{z}\mathbf{H}_n)^{-1}\mathbf{e}_1, \tag{4a}$$

$$\check{\boldsymbol{\nu}}(z)^{\mathsf{T}}({}^{z}\mathbf{H}_n) = \mathbf{e}_n^{\mathsf{T}}\frac{\chi(z)}{h_{1:n-1}} \quad \Leftrightarrow \quad \frac{h_{1:n-1}\check{\boldsymbol{\nu}}(z)^{\mathsf{T}}}{\chi(z)} = \mathbf{e}_n^{\mathsf{T}}({}^{z}\mathbf{H}_n)^{-1}. \tag{4b}$$

The repeated application of resolvents to $\mathbf{e}_1$ results in

$$\left(\prod_{i=1}^{k}({}^{z_i}\mathbf{H}_n)^{-1}\right)\mathbf{e}_1 = (-1)^{k-1}[z_1,\ldots,z_k]({}^{z}\mathbf{H}_n)^{-1}\mathbf{e}_1 \tag{5}$$

$$= (-1)^{k-1}[z_1,\ldots,z_k]\frac{\boldsymbol{\nu}(z)h_{1:n-1}}{\chi(z)}. \tag{6}$$

Note that $z\mathbf{I}_n - {}^{z}\mathbf{H}_n = z\mathbf{I}_n - (z\mathbf{I}_n - \mathbf{H}_n) = \mathbf{H}_n$, i.e., $\mathbf{H}_n({}^{z}\mathbf{H}_n)^{-1} = z({}^{z}\mathbf{H}_n)^{-1} - \mathbf{I}_n$.

Simplification

For the sake of eased understanding, we look at inverse iteration with a two-sided Rayleigh quotient where the left vector is the last standard unit vector $\mathbf{e}_n^{\mathsf{T}}$.

Simplification

For the sake of eased understanding, we look at inverse iteration with a two-sided Rayleigh quotient where the left vector is the last standard unit vector $\mathbf{e}_n^\mathsf{T}$. For this method we have the iterates

$$\mathbf{v}_{k+1} = \Big(\prod_{i=1}^{k} ({}^{z_i}\mathbf{H}_n)^{-1}\Big)\mathbf{e}_1, \quad x_{k+1} = \frac{\mathbf{e}_n^\mathsf{T}\mathbf{H}_n\mathbf{v}_{k+1}}{\mathbf{e}_n^\mathsf{T}\mathbf{v}_{k+1}},$$

Simplification

For the sake of eased understanding, we look at inverse iteration with a two-sided Rayleigh quotient where the left vector is the last standard unit vector $\mathbf{e}_n^\mathsf{T}$. For this method we have the iterates

$$\mathbf{v}_{k+1} = \Big(\prod_{i=1}^{k} ({}^{z_i}\mathbf{H}_n)^{-1}\Big)\mathbf{e}_1, \quad x_{k+1} = \frac{\mathbf{e}_n^\mathsf{T}\mathbf{H}_n\mathbf{v}_{k+1}}{\mathbf{e}_n^\mathsf{T}\mathbf{v}_{k+1}},$$

and thus the approximate eigenvalues are given by the Opitz-Larkin method:

$$x_{k+1} = \frac{\mathbf{e}_n^\mathsf{T}\mathbf{H}_n\big(\prod_{i=1}^{k}({}^{z_i}\mathbf{H}_n)^{-1}\big)\mathbf{e}_1}{\mathbf{e}_n^\mathsf{T}\big(\prod_{i=1}^{k}({}^{z_i}\mathbf{H}_n)^{-1}\big)\mathbf{e}_1} = \frac{\mathbf{e}_n^\mathsf{T}(z_k\mathbf{I}_n - ({}^{z_k}\mathbf{H}_n))\big(\prod_{i=1}^{k}({}^{z_i}\mathbf{H}_n)^{-1}\big)\mathbf{e}_1}{\mathbf{e}_n^\mathsf{T}\big(\prod_{i=1}^{k}({}^{z_i}\mathbf{H}_n)^{-1}\big)\mathbf{e}_1} \tag{7a}$$

$$= z_k - \frac{\mathbf{e}_n^{\mathsf{T}\,z_k}\mathbf{H}_n\big(\prod_{i=1}^{k}({}^{z_i}\mathbf{H}_n)^{-1}\big)\mathbf{e}_1}{\mathbf{e}_n^\mathsf{T}\big(\prod_{i=1}^{k}({}^{z_i}\mathbf{H}_n)^{-1}\big)\mathbf{e}_1} = z_k - \frac{\mathbf{e}_n^\mathsf{T}\big(\prod_{i=1}^{k-1}({}^{z_i}\mathbf{H}_n)^{-1}\big)\mathbf{e}_1}{\mathbf{e}_n^\mathsf{T}\big(\prod_{i=1}^{k}({}^{z_i}\mathbf{H}_n)^{-1}\big)\mathbf{e}_1} \tag{7b}$$

$$= z_k + \frac{[z_1,\dots,z_{k-1}](1/\chi)}{[z_1,\dots,z_{k-1},z_k](1/\chi)}. \tag{7c}$$

Simplification

When we update the shifts by choosing $z_{k+1} = x_{k+1}$ we obtain the standard variant of the Opitz-Larkin method. This method has asymptotically second order convergence against the roots of the characteristic polynomial χ.

Simplification

When we update the shifts by choosing $z_{k+1} = x_{k+1}$ we obtain the standard variant of the Opitz-Larkin method. This method has asymptotically second order convergence against the roots of the characteristic polynomial χ.

Inverse iteration with fixed shift $\tau = z_1 = z_2 = \ldots = z_k$ results in the recurrence

$$x_{k+1} = \tau + \frac{[\tau, \ldots, \tau](1/\chi)}{[\tau, \ldots, \tau, \tau](1/\chi)} = \tau + k\frac{(1/\chi)^{(k-1)}(\tau)}{(1/\chi)^{(k)}(\tau)}. \tag{8}$$

Simplification

When we update the shifts by choosing $z_{k+1} = x_{k+1}$ we obtain the standard variant of the Opitz-Larkin method. This method has asymptotically second order convergence against the roots of the characteristic polynomial χ.

Inverse iteration with fixed shift $\tau = z_1 = z_2 = \ldots = z_k$ results in the recurrence

$$x_{k+1} = \tau + \frac{[\tau, \ldots, \tau](1/\chi)}{[\tau, \ldots, \tau, \tau](1/\chi)} = \tau + k\frac{(1/\chi)^{(k-1)}(\tau)}{(1/\chi)^{(k)}(\tau)}. \quad (8)$$

Inverse iteration with fixed shift performs one step of König's method. Restarting inverse iteration every s steps with updated shift given by the current eigenvalue approximation converges with order s (divided by steps: linearly).

Simplification

When we update the shifts by choosing $z_{k+1} = x_{k+1}$ we obtain the standard variant of the Opitz-Larkin method. This method has asymptotically second order convergence against the roots of the characteristic polynomial χ.

Inverse iteration with fixed shift $\tau = z_1 = z_2 = \ldots = z_k$ results in the recurrence

$$x_{k+1} = \tau + \frac{[\tau, \ldots, \tau](1/\chi)}{[\tau, \ldots, \tau, \tau](1/\chi)} = \tau + k \frac{(1/\chi)^{(k-1)}(\tau)}{(1/\chi)^{(k)}(\tau)}. \tag{8}$$

Inverse iteration with fixed shift performs one step of König's method. Restarting inverse iteration every s steps with updated shift given by the current eigenvalue approximation converges with order s (divided by steps: linearly).

Symmetric RQI is very pleasant to analyze, likely-wise is two-sided RQI, but unsymmetric RQI (and thus, the QR algorithm) and alternating RQI do not fit into the picture.

Simplification

The original Rayleigh quotient iteration (Strutt, 1894) with the symmetric Rayleigh quotient and, because of the symmetry, a tridiagonal Hermitean Hessenberg matrix $\mathbf{H}_n$, gives the update

$$z_{k+1} = \frac{\mathbf{e}_1^{\mathsf{T}} ({}^{z_k}\mathbf{H}_n)^{-\mathsf{H}} \mathbf{H}_n ({}^{z_k}\mathbf{H}_n)^{-1} \mathbf{e}_1}{\mathbf{e}_1^{\mathsf{T}} ({}^{z_k}\mathbf{H}_n)^{-\mathsf{H}} ({}^{z_k}\mathbf{H}_n)^{-1} \mathbf{e}_1} = \frac{\mathbf{e}_1^{\mathsf{T}} \mathbf{H}_n ({}^{z_k}\mathbf{H}_n)^{-2} \mathbf{e}_1}{\mathbf{e}_1^{\mathsf{T}} ({}^{z_k}\mathbf{H}_n)^{-2} \mathbf{e}_1} \tag{9a}$$

$$= \frac{\mathbf{e}_1^{\mathsf{T}} (z_k \mathbf{I}_n - {}^{z_k}\mathbf{H}_n) ({}^{z_k}\mathbf{H}_n)^{-2} \mathbf{e}_1}{\mathbf{e}_1^{\mathsf{T}} ({}^{z_k}\mathbf{H}_n)^{-2} \mathbf{e}_1} \tag{9b}$$

$$= z_k - \frac{\mathbf{e}_1^{\mathsf{T}} ({}^{z_k}\mathbf{H}_n)^{-1} \mathbf{e}_1}{\mathbf{e}_1^{\mathsf{T}} ({}^{z_k}\mathbf{H}_n)^{-2} \mathbf{e}_1} = z_k + \frac{[z_k](\chi_{2:n}/\chi)}{[z_k, z_k](\chi_{2:n}/\chi)} \tag{9c}$$

$$= z_k - \frac{r(z_k)}{r'(z_k)}, \qquad r(z) := \frac{\chi(z)}{\chi_{2:n}(z)}. \tag{9d}$$

Simplification

The original Rayleigh quotient iteration (Strutt, 1894) with the symmetric Rayleigh quotient and, because of the symmetry, a tridiagonal Hermitean Hessenberg matrix $\mathbf{H}_n$, gives the update

$$z_{k+1} = \frac{\mathbf{e}_1^{\mathsf{T}}({}^{z_k}\mathbf{H}_n)^{-\mathsf{H}}\mathbf{H}_n({}^{z_k}\mathbf{H}_n)^{-1}\mathbf{e}_1}{\mathbf{e}_1^{\mathsf{T}}({}^{z_k}\mathbf{H}_n)^{-\mathsf{H}}({}^{z_k}\mathbf{H}_n)^{-1}\mathbf{e}_1} = \frac{\mathbf{e}_1^{\mathsf{T}}\mathbf{H}_n({}^{z_k}\mathbf{H}_n)^{-2}\mathbf{e}_1}{\mathbf{e}_1^{\mathsf{T}}({}^{z_k}\mathbf{H}_n)^{-2}\mathbf{e}_1} \quad \text{(9a)}$$

$$= \frac{\mathbf{e}_1^{\mathsf{T}}(z_k\mathbf{I}_n - {}^{z_k}\mathbf{H}_n)({}^{z_k}\mathbf{H}_n)^{-2}\mathbf{e}_1}{\mathbf{e}_1^{\mathsf{T}}({}^{z_k}\mathbf{H}_n)^{-2}\mathbf{e}_1} \quad \text{(9b)}$$

$$= z_k - \frac{\mathbf{e}_1^{\mathsf{T}}({}^{z_k}\mathbf{H}_n)^{-1}\mathbf{e}_1}{\mathbf{e}_1^{\mathsf{T}}({}^{z_k}\mathbf{H}_n)^{-2}\mathbf{e}_1} = z_k + \frac{[z_k](\chi_{2:n}/\chi)}{[z_k, z_k](\chi_{2:n}/\chi)} \quad \text{(9c)}$$

$$= z_k - \frac{r(z_k)}{r'(z_k)}, \qquad r(z) := \frac{\chi(z)}{\chi_{2:n}(z)}. \quad \text{(9d)}$$

This is Newton's method on the meromorphic function r. As the poles of this meromorphic function are the eigenvalues of a submatrix, they interlace by Cauchy's interlace theorem the roots, which are the eigenvalues.

Simplification

Symmetric RQI for Hermitean matrices gives the update

$$z_{k+1} = z_k + \frac{[z_1, z_1, \ldots, z_{k-1}, z_{k-1}, z_k](\chi_{2:n}/\chi)}{[z_1, z_1, \ldots, z_{k-1}, z_{k-1}, z_k, z_k](\chi_{2:n}/\chi)}. \quad (10)$$

Simplification

Symmetric RQI for Hermitean matrices gives the update

$$z_{k+1} = z_k + \frac{[z_1, z_1, \ldots, z_{k-1}, z_{k-1}, z_k](\chi_{2:n}/\chi)}{[z_1, z_1, \ldots, z_{k-1}, z_{k-1}, z_k, z_k](\chi_{2:n}/\chi)}. \quad (10)$$

This update has by a result of Tornheim asymptotically a cubic convergence rate. We have to compute the limit of the real root of the equations

$$x^k - 2x^{k-1} - 2x^{k-2} - \cdots - 2 = 0, \quad k = 1, \ldots$$

Simplification

Symmetric RQI for Hermitean matrices gives the update

$$z_{k+1} = z_k + \frac{[z_1, z_1, \ldots, z_{k-1}, z_{k-1}, z_k](\chi_{2:n}/\chi)}{[z_1, z_1, \ldots, z_{k-1}, z_{k-1}, z_k, z_k](\chi_{2:n}/\chi)}. \tag{10}$$

This update has by a result of Tornheim asymptotically a cubic convergence rate. We have to compute the limit of the real root of the equations

$$x^k - 2x^{k-1} - 2x^{k-2} - \cdots - 2 = 0, \quad k = 1, \ldots$$

This is the maximal eigenvalue of a Hessenberg matrix with one in the lower diagonal and two in the last column. The approximate eigenvector of all ones to the approximate eigenvalue 3 gives the backward error $1/\sqrt{k}$ and the only positive real eigenvalue of the matrix is well separated, the other eigenvalues lie close to a circle of radius one around zero.

Simplification

The picture changes if we apply the special inverse iteration to a general unreduced Hessenberg matrix, not necessarily Hermitean or symmetric.

Simplification

The picture changes if we apply the special inverse iteration to a general unreduced Hessenberg matrix, not necessarily Hermitean or symmetric.

If we take another standard unit vector $\mathbf{e}_\ell$ as left vector, we obtain the Opitz-Larkin method applied to the meromorphic function

$$m_\ell(z) = \frac{\chi(z)}{h_{1:\ell-1}\chi_{1+\ell:n}(z)}. \tag{11}$$

Simplification

The picture changes if we apply the special inverse iteration to a general unreduced Hessenberg matrix, not necessarily Hermitean or symmetric.

If we take another standard unit vector $\mathbf{e}_\ell$ as left vector, we obtain the Opitz-Larkin method applied to the meromorphic function

$$m_\ell(z) = \frac{\chi(z)}{h_{1:\ell-1}\chi_{1+\ell:n}(z)}. \tag{11}$$

If we take an arbitrary left vector $\mathbf{y}$, we obtain the Opitz-Larkin method applied to the meromorphic function

$$r(z;\mathbf{y}) = \frac{\chi(z)}{\sum_{i=1}^{n} y_i h_{1:i-1}\chi_{1+i:n}(z)} = \frac{\chi(z)}{p(z;\mathbf{y})}, \quad p(z;\mathbf{y}) \in \mathbb{P}_{<n}. \tag{12}$$

Simplification

The picture changes if we apply the special inverse iteration to a general unreduced Hessenberg matrix, not necessarily Hermitean or symmetric.

If we take another standard unit vector $\mathbf{e}_\ell$ as left vector, we obtain the Opitz-Larkin method applied to the meromorphic function

$$m_\ell(z) = \frac{\chi(z)}{h_{1:\ell-1}\chi_{1+\ell:n}(z)}. \quad (11)$$

If we take an arbitrary left vector $\mathbf{y}$, we obtain the Opitz-Larkin method applied to the meromorphic function

$$r(z;\mathbf{y}) = \frac{\chi(z)}{\sum_{i=1}^{n} y_i h_{1:i-1}\chi_{1+i:n}(z)} = \frac{\chi(z)}{p(z;\mathbf{y})}, \quad p(z;\mathbf{y}) \in \mathbb{P}_{<n}. \quad (12)$$

The polynomials $\chi_{1+i:n}$ have degree $\deg(\chi_{1+i:n}) = n - i$ and leading coefficient one, thus they form a basis of the space of polynomials of degree less n.

Simplification

The two-sided RQI variant corresponds to a confluent Opitz-Larkin method with double nodes. In this method the left vector determines a polynomial, which is formed as a linear combination of characteristic polynomials of trailing submatrices.

Simplification

The two-sided RQI variant corresponds to a confluent Opitz-Larkin method with double nodes. In this method the left vector determines a polynomial, which is formed as a linear combination of characteristic polynomials of trailing submatrices.

In single-sided RQI for non-Hermitean matrices, we change the vector $\mathbf{y}$ that determines the denominator polynomial of the rational function

$$r(z;\mathbf{y}) = \frac{\chi(z)}{p(z;\mathbf{y})}$$

in every step and apply one step of the Opitz-Larkin method without confluent nodes.

Simplification

The two-sided RQI variant corresponds to a confluent Opitz-Larkin method with double nodes. In this method the left vector determines a polynomial, which is formed as a linear combination of characteristic polynomials of trailing submatrices.

In single-sided RQI for non-Hermitean matrices, we change the vector $\mathbf{y}$ that determines the denominator polynomial of the rational function

$$r(z;\mathbf{y}) = \frac{\chi(z)}{p(z;\mathbf{y})}$$

in every step and apply one step of the Opitz-Larkin method without confluent nodes.

Convergence of $\mathbf{y}$ indicates that we might arrive at second order convergence. One multi-shift does not change $\mathbf{y}$ compared to several consecutive single shifts. Multiple multi-shifts are locally favourable in the Opitz-Larkin context.

Jenkins-Traub

The Jenkins-Traub algorithm is related to a generalized RQI. Thus, it should fit into the Opitz-Larkin framework.

Jenkins-Traub

The Jenkins-Traub algorithm is related to a generalized RQI. Thus, it should fit into the Opitz-Larkin framework. By inspection, the three stages correspond to the following variants of Opitz-Larkin:

Jenkins-Traub

The Jenkins-Traub algorithm is related to a generalized RQI. Thus, it should fit into the Opitz-Larkin framework. By inspection, the three stages correspond to the following variants of Opitz-Larkin:

First stage: compute iterates using König's method at zero,

$$x_{k+1} = 0 + \frac{[0,\dots,0](1/\chi)}{[0,\dots,0,0](1/\chi)} = (k+1)\frac{(1/\chi)^{(k)}(0)}{(1/\chi)^{(k+1)}(0)}, \quad k = 0,\dots,p-1. \quad (13)$$

Jenkins-Traub

The Jenkins-Traub algorithm is related to a generalized RQI. Thus, it should fit into the Opitz-Larkin framework. By inspection, the three stages correspond to the following variants of Opitz-Larkin:

First stage: compute iterates using König's method at zero,

$$x_{k+1} = 0 + \frac{[0, \dots, 0](1/\chi)}{[0, \dots, 0, 0](1/\chi)} = (k+1)\frac{(1/\chi)^{(k)}(0)}{(1/\chi)^{(k+1)}(0)}, \quad k = 0, \dots, p-1. \tag{13}$$

Second stage: select a fixed shift $s \in \mathbb{C}$, compute

$$x_{k+1} = s + \frac{[0, \dots, 0, s, \dots, s](1/\chi)}{[0, \dots, 0, s, \dots, s, s](1/\chi)}, \quad k = p, \dots, q-1. \tag{14}$$

Third stage: Set the starting value z_0 to the one obtained by rational interpolation of $1/\chi$ at 0 and s, i.e.,

$$z_0 := x_q = s + \frac{[0,\dots,0,s,\dots,s](1/\chi)}{[0,\dots,0,s,\dots,s,s](1/\chi)}. \tag{15}$$

Repeat

$$z_{k+1} = z_k + \frac{[0,\dots,0,s,\dots,s,z_0,z_1,\dots,z_{k-1},z_k](1/\chi)}{[0,\dots,0,s,\dots,s,z_0,z_1,\dots,z_{k-1},z_k,z_k](1/\chi)}, \quad k = 0,\dots \tag{16}$$

Third stage: Set the starting value z_0 to the one obtained by rational interpolation of $1/\chi$ at 0 and s, i.e.,

$$z_0 := x_q = s + \frac{[0, \dots, 0, s, \dots, s](1/\chi)}{[0, \dots, 0, s, \dots, s, s](1/\chi)}. \tag{15}$$

Repeat

$$z_{k+1} = z_k + \frac{[0, \dots, 0, s, \dots, s, z_0, z_1, \dots, z_{k-1}, z_k](1/\chi)}{[0, \dots, 0, s, \dots, s, z_0, z_1, \dots, z_{k-1}, z_k, z_k](1/\chi)}, \quad k = 0, \dots \tag{16}$$

This proves amongst others the well-known fact that stage three of Jenkins-Traub, if it converges, does so with R-order $\phi^2 = \phi + 1 \approx 2.618$.

Third stage: Set the starting value z_0 to the one obtained by rational interpolation of $1/\chi$ at 0 and s, i.e.,

$$z_0 := x_q = s + \frac{[0,\dots,0,s,\dots,s](1/\chi)}{[0,\dots,0,s,\dots,s,s](1/\chi)}. \tag{15}$$

Repeat

$$z_{k+1} = z_k + \frac{[0,\dots,0,s,\dots,s,z_0,z_1,\dots,z_{k-1},z_k](1/\chi)}{[0,\dots,0,s,\dots,s,z_0,z_1,\dots,z_{k-1},z_k,z_k](1/\chi)}, \quad k = 0,\dots \tag{16}$$

This proves amongst others the well-known fact that stage three of Jenkins-Traub, if it converges, does so with R-order $\phi^2 = \phi + 1 \approx 2.618$.

Thus, Jenkins-Traub is a special form of Opitz-Larkin with, at first glance, rather strange evaluation scheme. This scheme is natural in view of the companion matrix interpretation given in (Jenkins and Traub, 1970).

Conclusions and Outlook

- We have presented the less well-known Opitz-Larkin method, which is a generalization of König's method using divided differences.

Conclusions and Outlook

- We have presented the less well-known Opitz-Larkin method, which is a generalization of König's method using divided differences.
- We have shown that many variants of inverse iteration and Rayleigh quotient iteration correspond to variants of the Opitz-Larkin method on certain rational functions with the characteristic polynomial as the numerator.

Conclusions and Outlook

- We have presented the less well-known Opitz-Larkin method, which is a generalization of König's method using divided differences.
- We have shown that many variants of inverse iteration and Rayleigh quotient iteration correspond to variants of the Opitz-Larkin method on certain rational functions with the characteristic polynomial as the numerator.
- We have indicated why non-symmetric RQI and thus the QR algorithm are not that easily analyzed using this "missing link".

Conclusions and Outlook

- We have presented the less well-known Opitz-Larkin method, which is a generalization of König's method using divided differences.
- We have shown that many variants of inverse iteration and Rayleigh quotient iteration correspond to variants of the Opitz-Larkin method on certain rational functions with the characteristic polynomial as the numerator.
- We have indicated why non-symmetric RQI and thus the QR algorithm are not that easily analyzed using this "missing link".
- We have shown that the well-known Jenkins-Traub method is a special instance of a Opitz-Larkin method.

Conclusions and Outlook

- We have presented the less well-known Opitz-Larkin method, which is a generalization of König's method using divided differences.
- We have shown that many variants of inverse iteration and Rayleigh quotient iteration correspond to variants of the Opitz-Larkin method on certain rational functions with the characteristic polynomial as the numerator.
- We have indicated why non-symmetric RQI and thus the QR algorithm are not that easily analyzed using this “missing link”.
- We have shown that the well-known Jenkins-Traub method is a special instance of a Opitz-Larkin method.
- Next, we want to take a closer look at the global behaviour of these methods using the Opitz-Larkin framework.

Conclusions and Outlook

- We have presented the less well-known Opitz-Larkin method, which is a generalization of König's method using divided differences.
- We have shown that many variants of inverse iteration and Rayleigh quotient iteration correspond to variants of the Opitz-Larkin method on certain rational functions with the characteristic polynomial as the numerator.
- We have indicated why non-symmetric RQI and thus the QR algorithm are not that easily analyzed using this "missing link".
- We have shown that the well-known Jenkins-Traub method is a special instance of a Opitz-Larkin method.
- Next, we want to take a closer look at the global behaviour of these methods using the Opitz-Larkin framework.
- The local link between one step of Opitz-Larkin and shifts in the QR algorithm should enable a better understanding of multi-shift strategies and the development of new ones.

Thank you for your attention.

Chatelin, F. (1993).
Eigenvalues of matrices.
John Wiley & Sons Ltd., Chichester.
With exercises by Mario Ahués and the author, Translated from the French and with additional material by Walter Ledermann.

Crandall, S. H. (1951).
Iterative procedures related to relaxation methods for eigenvalue problems.
Proceedings of the Royal Society. London. Series A. Mathematical, Physical and Engineering Sciences, 207:416–423.

Dekker, T. J. and Traub, J. F. (1971).
The shifted QR algorithm for Hermitian matrices.
Linear Algebra and Appl., 4:137–154.

Jenkins, M. A. and Traub, J. F. (1970).
A three-stage variable-shift iteration for polynomial zeros and its relation to generalized Rayleigh iteration.
Numer. Math., 14(3):252–263.

König, J. (1884).
Ueber eine Eigenschaft der Potenzreihen.
Mathematische Annalen, 23:447–449.
Scanned article online available at the Göttinger Digitalisierungszentrum (GDZ): http://resolver.sub.uni-goettingen.de/purl?GDZPPN002248026.

Larkin, F. M. (1980).
Root-finding by fitting rational functions.
Mathematics of Computation, 35(151):803–816.

Larkin, F. M. (1981).
Root finding by divided differences.
Numer. Math., 37(1):93–104.

Opitz, G. (1958).
Gleichungsauflösung mittels einer speziellen Interpolation.
Z. Angew. Math. Mech., 38:276–277.

Opitz, G. (1959).
Ein Verfahren zur gleichzeitigen Bestimmung einer beliebigen Zahl von Nullstellen von Funktionen einer Veränderlichen.
Habilitationsschrift, TH Dresden.

Opitz, G. (1961).
Zum Vergleich von Iterationsverfahren zur Gleichungsauflösung.
Z. Angew. Math. Mech., 41:T48–T50.

Opitz, G. (1964).
Steigungsmatrizen.
Z. Angew. Math. Mech., 44:T52–T54.

Ostrowski, A. M. (1958).
On the convergence of the Rayleigh quotient iteration for the computation of the characteristic roots and vectors, I.
Arch. Rational Mech. Anal., 1:233–241.

Ostrowski, A. M. (1959b).
On the convergence of the Rayleigh quotient iteration for the computation of the characteristic roots and vectors, II.
Arch. Rational Mech. Anal., 2:423–428.

Ostrowski, A. M. (1959c).
On the convergence of the Rayleigh quotient iteration for the computation of the characteristic roots and vectors. III. (Generalized Rayleigh quotient and characteristic roots with linear elementary divisors).
Arch. Rational Mech. Anal., 3:325–340.

Ostrowski, A. M. (1959d).
On the convergence of the Rayleigh quotient iteration for the computation of the characteristic roots and vectors. IV. (Generalized Rayleigh quotient for nonlinear elementary divisors).
Arch. Rational Mech. Anal., 3:341–347.

Ostrowski, A. M. (1959e).
On the convergence of the Rayleigh quotient iteration for the computation of the characteristic roots and vectors. V. Usual Rayleigh quotient for non-Hermitian matrices and linear elementary divisors.
Arch. Rational Mech. Anal., 3:472–481.

Ostrowski, A. M. (1959a).
On the convergence of the Rayleigh quotient iteration for the computation of characteristic roots and vectors. VI. (Usual Rayleigh quotient for nonlinear elementary divisors).
Arch. Rational Mech. Anal., 4:153–165.

Schröder, E. (1870).
Ueber unendlich viele Algorithmen zur Auflösung der Gleichungen.
Math. Ann., 2(2):317–365.

Strutt, J. W. (1894).
The Theory of Sound, volume one.
Macmillan, London, second edition.

Tornheim, L. (1964).
Convergence of multipoint iterative methods.
J. Assoc. Comput. Mach., 11:210–220.

Wielandt, H. (1944).
Beiträge zur mathematischen Behandlung komplexer Eigenwertprobleme, Teil V: Bestimmung höherer Eigenwerte durch gebrochene Iteration.
Bericht B44/J/37, Aerodynamische Versuchsanstalt Göttingen.

Work relating to the eigenvector polynomials . . .

Ericsson, T. (1990).
On the eigenvalues and eigenvectors of Hessenberg matrices.
Numerical Analysis Group, Göteborg, Report 10, Chalmers University of Technology and the University of Göteborg, Department of Computer Sciences, S-412 96 Göteborg, Sweden.
Online available at
http://www.cs.chalmers.se/pub/num_analysis/reports/.

Franklin, J. N. (1968).
Matrix theory.
Prentice-Hall Inc., Englewood Cliffs, N.J.

Schweins, F. (1825).
Theorie der Differenzen und Differentiale.
Verlag der Universitäts-Buchhandlung von C. F. Winter, Heidelberg.
Digitized by Google, online available at
http://books.google.de/books?id=dntNAAAAMAAJ.

Zemke, J.-P. M. (2006).
Hessenberg eigenvalue-eigenmatrix relations.
Linear Algebra Appl., 414(2-3):589–606.

Zemke, J.-P. M. (2007).
Abstract perturbed Krylov methods.
Linear Algebra Appl., 424(2-3):405–434.