On the Eigenpair Problem

Felipe Cucker

City University of Hong Kong

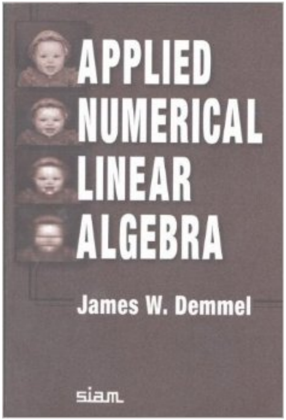
Joint work with D. Armentano, C. Beltrán, P. Bürgisser, and M. Shub

Let $A \in \mathbb{C}^{n \times n}$

We say that $\lambda \in \mathbb{C}$ is an eigenvalue of A when there exists $v \in \mathbb{C}^{n \times n} \setminus \{0\}$ such that

$$Av = \lambda v.$$

In this case we say that v is an eigenvector.



APPLIED NUMERICAL LINEAR ALGEBRA

James W. Demmel

siam

So the problem of devising an algorithm [for the eigenvalue problem] that is numerically stable and globally (and quickly!) convergent remains open. [p. 139]

Current Algorithms

- *Compute the characteristic polynomial* χ_A *of* A *and then compute (i.e., approximate) its zeros.*

Current Algorithms

- *Compute the characteristic polynomial χ_A of A and then compute (i.e., approximate) its zeros.*

[*** Numerically unstable in practice ***]

Current Algorithms

- *Compute the characteristic polynomial* χ_A *of* A *and then compute (i.e., approximate) its zeros.*

[*** Numerically unstable in practice ***]

- *Variations of the QR algorithm.*

Current Algorithms

- *Compute the characteristic polynomial χ_A of A and then compute (i.e., approximate) its zeros.*

[*** Numerically unstable in practice ***]

- *Variations of the QR algorithm.*
  - The unshifted QR algorithm terminates with probability 1 but is probably infinite average cost if approximations to the eigenvectors are to be output.

Current Algorithms

- *Compute the characteristic polynomial χ_A of A and then compute (i.e., approximate) its zeros.*

[*** Numerically unstable in practice ***]

- *Variations of the QR algorithm.*
  - The unshifted QR algorithm terminates with probability 1 but is probably infinite average cost if approximations to the eigenvectors are to be output.
  - The QR algorithm with Rayleigh Quotient shift fails for open sets of real input matrices.

Current Algorithms

- *Compute the characteristic polynomial χ_A of A and then compute (i.e., approximate) its zeros.*

[*** Numerically unstable in practice ***]

- *Variations of the QR algorithm.*
  - The unshifted QR algorithm terminates with probability 1 but is probably infinite average cost if approximations to the eigenvectors are to be output.
  - The QR algorithm with Rayleigh Quotient shift fails for open sets of real input matrices.
  - It is unknown whether the Francis (double) shift algorithm converges generally on real or complex matrices.

Current Algorithms

- *Compute the characteristic polynomial χ_A of A and then compute (i.e., approximate) its zeros.*

[*** Numerically unstable in practice ***]

- *Variations of the QR algorithm.*
  - The unshifted QR algorithm terminates with probability 1 but is probably infinite average cost if approximations to the eigenvectors are to be output.
  - The QR algorithm with Rayleigh Quotient shift fails for open sets of real input matrices.
  - It is unknown whether the Francis (double) shift algorithm converges generally on real or complex matrices.

[*** No theoretical understanding ***]

Theorem We exhibit an algorithm which on input a matrix A with complex Gaussian entries generates (with probability 1) an "excellent" approximation to all the (eigenvalue, eigenvector) pairs of A. Moreover, the running time of this algorithm is polynomial in n on the average.

This algorithm is numerically stable.

Three kinds of approximation ($\widetilde{\zeta}$ for $A \mapsto \zeta$):

Three kinds of approximation ($\widetilde{\zeta}$ for $A \mapsto \zeta$):

- *Backward approximation.* $\widetilde{\zeta}$ is the solution of a datum $\widetilde{A}$ close to A. Given $\varepsilon > 0$, we say that $\widetilde{\zeta}$ is an ε-backward approximation when $\|A - \widetilde{A}\| \leq \varepsilon$.

Three kinds of approximation ($\widetilde{\zeta}$ for $A \mapsto \zeta$):

- *Backward approximation.* $\widetilde{\zeta}$ is the solution of a datum $\widetilde{A}$ close to A. Given $\varepsilon > 0$, we say that $\widetilde{\zeta}$ is an ε-backward approximation when $\|A - \widetilde{A}\| \leq \varepsilon$.
- *Forward approximation.* $\widetilde{\zeta}$ is close to ζ. Given $\varepsilon > 0$, we say that $\widetilde{\zeta}$ is an ε-forward approximation when $|\zeta - \widetilde{\zeta}| \leq \varepsilon$.

Three kinds of approximation ($\widetilde{\zeta}$ for $A \mapsto \zeta$):

- *Backward approximation.* $\widetilde{\zeta}$ is the solution of a datum $\widetilde{A}$ close to A. Given $\varepsilon > 0$, we say that $\widetilde{\zeta}$ is an ε-backward approximation when $\|A - \widetilde{A}\| \leq \varepsilon$.
- *Forward approximation.* $\widetilde{\zeta}$ is close to ζ. Given $\varepsilon > 0$, we say that $\widetilde{\zeta}$ is an ε-forward approximation when $|\zeta - \widetilde{\zeta}| \leq \varepsilon$.
- *Approximation à la Smale.* An appropriate version of Newton's iteration, starting at $\widetilde{\zeta}$, converges immediately, quadratically fast, to ζ.

Newton's method

We define a Newton map associated to A

$$N_A : \mathbb{C} \times (\mathbb{C}^n \setminus \{0\}) \to \mathbb{C} \times (\mathbb{C}^n \setminus \{0\}).$$

Let $\mathcal{V} := \{(A, \lambda, v) \mid (A - \lambda \mathrm{Id})v = 0\}$.

Newton's method

We define a Newton map associated to A

$$N_A : \mathbb{C} \times (\mathbb{C}^n \setminus \{0\}) \to \mathbb{C} \times (\mathbb{C}^n \setminus \{0\}).$$

Let $\mathcal{V} := \{(A, \lambda, v) \mid (A - \lambda \mathrm{Id})v = 0\}$.

Given $(A, \lambda, v) \in \mathcal{V}$ we say that $(\zeta, w) \in \mathbb{C} \times (\mathbb{C}^n \setminus \{0\})$ is an *approximate eigenpair* of A with associated eigenpair (λ, v) when for all $k \geq 1$ the kth iterate $N_A^k(\zeta, w)$ of the Newton map at (ζ, w) is well defined and satisfies

$$\mathrm{dist}\big((N_A^k(\zeta, w)), (\lambda, v)\big) \leq \left(\frac{1}{2}\right)^{2^k - 1} \mathrm{dist}\big((\zeta, w), (\lambda, v)\big).$$

Well-posedness

A triple (A, λ, v) is well-posed when λ is a simple eigenvalue of A. We write $(A, \lambda, v) \in \mathcal{W}$. Otherwise, it is said to be ill-posed.

Proposition Ill-posed eigenpairs have no approximate eigenpairs à la Smale.

Proposition The set Σ of matrices A with multiple eigenvalues has (real) codimension 2 in $\mathbb{C}^{n \times n}$.

Given $(A, \lambda, v) \in \mathbb{C}^{n\times n} \times \mathbb{C} \times \mathbb{P}(\mathbb{C}^n)$, we let $A_{\lambda,v} : T_v \to T_v$ be

$$A_{\lambda,v} := P_{v^\perp} \circ (A - \lambda \mathsf{Id})|_{T_v}.$$

We define the condition number of the triple (A, λ, v) by

$$\mu(A, \lambda, v) := \|A\|_F \, \|A_{\lambda,v}^{-1}\|,$$

Theorem Let $A \in \mathbb{C}^{n\times n}$ with $\|A\|_F = 1$ and $(\lambda, v), (\lambda_0, v_0) \in \mathbb{C} \times (\mathbb{C}^n \setminus \{0\})$. If (λ, v) is a well-posed eigenpair of A and

$$\mathsf{dist}\big((\lambda, v), (\lambda_0, v_0)\big) < \frac{c_0}{\mu(A, \lambda, v)}$$

then (λ_0, v_0) is an approximate eigenpair of A with associated eigenpair (λ, v). One may choose $c_0 = 0.2881$.

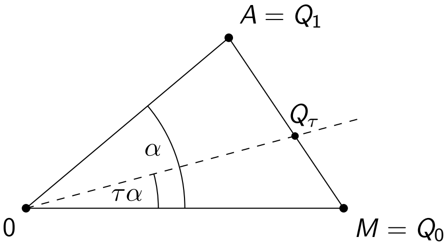
The Algorithm

We are given $A \in \mathbb{C}^{n\times n}$ and an initial triple (M, λ_0, v_0) in $\mathcal{W}$.

Consider the line segment $[M, A]$ in $\mathbb{C}^{n\times n}$ with endpoints M and A

$$[M, A] = \{Q_\tau \in \mathbb{C}^{n\times n} \mid \tau \in [0, 1]\}$$

with Q_τ being the only point in $[M, A]$ such that $d_{\mathbb{S}}(M, Q_\tau) = \tau\alpha$.



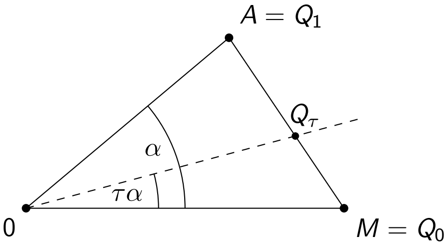
The Algorithm

We are given $A \in \mathbb{C}^{n\times n}$ and an initial triple (M, λ_0, v_0) in $\mathcal{W}$.

Consider the line segment $[M, A]$ in $\mathbb{C}^{n\times n}$ with endpoints M and A

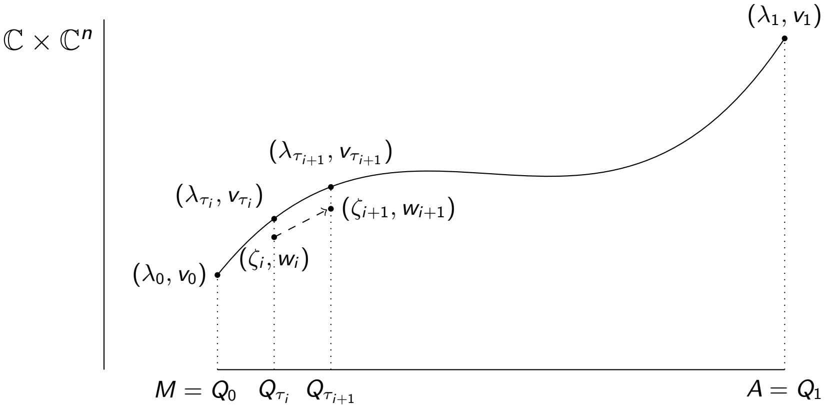
$$[M, A] = \{Q_\tau \in \mathbb{C}^{n\times n} \mid \tau \in [0, 1]\}$$

with Q_τ being the only point in $[M, A]$ such that $d_{\mathbb{S}}(M, Q_\tau) = \tau\alpha$.



When τ moves from 0 to 1 the eigenpair (λ_τ, v_τ) of Q_τ moves from (λ_0, v_0) to an eigenpair (λ_1, v_1) of $A = Q_1$.

We want to "follow" the curve (λ_τ, v_τ).



```
Path-follow

Input: $A \in \mathbb{C}^{n\times n}$ and $(M, \lambda_0, v_0) \in \mathcal{W}$

$\alpha := d_{\mathbb{S}}(M, A)$,  $r := \|A\|_F$,  $s := \|M\|_F$
$\tau := 0$,  $Q := M$,  $(\zeta, w) := (\lambda_0, v_0)$
repeat
    $\Delta\tau := \frac{0.001461}{\alpha\,\mu^2(Q,\zeta,w)}$
    $\tau := \min\{1, \tau + \Delta\tau\}$
    $t := \frac{s}{s+r(\sin\alpha\cot(\tau\alpha)-\cos\alpha)}$
    $Q := tA + (1-t)M$
    $(\zeta, w) := N_Q(\zeta, w)$
until  $\tau = 1$
return  $(\zeta, w)$

Output: $(\zeta, w) \in \mathbb{C} \times \mathbb{C}^n$, approximate eigenpair of $A$.
```

Complexity

Cost of `Path-follow` on input (A, M, λ_0, v_0):

$\mathrm{cost}(A, M, \lambda_0, v_0) =$ Number of iterations × cost of each iteration

Complexity

Cost of `Path-follow` on input (A, M, λ_0, v_0):

$$\text{cost}(A, M, \lambda_0, v_0) = \text{Number of iterations} \times \text{cost of each iteration}$$

$$\nearrow \qquad\qquad\qquad\qquad \nwarrow$$

$$K(A, M, \lambda_0, v_0) \qquad\qquad\qquad\qquad \mathcal{O}(n^3)$$

Complexity

Cost of `Path-follow` on input (A, M, λ_0, v_0):

$$\text{cost}(A, M, \lambda_0, v_0) = \text{Number of iterations} \times \text{cost of each iteration}$$

$$\nearrow \qquad\qquad\qquad\qquad \nwarrow$$

$$K(A, M, \lambda_0, v_0) \qquad\qquad\qquad\qquad \mathcal{O}(n^3)$$

`Path-follow` terminates (i.e., $K(A, M, \lambda_0, v_0) < \infty$) iff $(Q_\tau, \lambda_\tau, v_\tau) \in \mathcal{W}$ for all $\tau \in [0, 1]$.

Theorem

$$K(A, M, \lambda_0, v_0) \leq C \int_0^1 \mu^2(Q_\tau, \lambda_\tau, v_\tau) d\tau.$$

Given $A \in \mathbb{C}^{n\times n}$, we ran `Path-follow` with:

- M a (particular) diagonal matrix

Given $A \in \mathbb{C}^{n\times n}$, we ran `Path-follow` with:

- M a (particular) diagonal matrix
- λ_0 the ith diagonal entry (i choosen at random)

Given $A \in \mathbb{C}^{n \times n}$, we ran `Path-follow` with:

- M a (particular) diagonal matrix
- λ_0 the ith diagonal entry (i choosen at random)
- $v_0 = e_i$

Given $A \in \mathbb{C}^{n \times n}$, we ran `Path-follow` with:

- M a (particular) diagonal matrix
- λ_0 the ith diagonal entry (i choosen at random)
- $v_0 = e_i$

Theorem We have

$$\underset{A \sim N(0,\mathrm{Id})}{\mathbb{E}} \; \underset{i \sim \{1,\dots,n\}}{\mathbb{E}} K(A, M, \lambda_0, v_0) = \mathcal{O}(n^5)$$

and, consequently,

$$\underset{A \sim N(0,\mathrm{Id})}{\mathbb{E}} \; \underset{i \sim \{1,\dots,n\}}{\mathbb{E}} \mathrm{cost}(A, M, \lambda_0, v_0) = \mathcal{O}(n^8).$$

One can also (deterministically) compute **all** the eigenpairs of A. The average total cost is $\mathcal{O}(n^9)$.

A finer randomization?

Theorem We have

$$\mathop{\mathbb{E}}_{A\sim N(0,\mathrm{Id})} \mathop{\mathbb{E}}_{M\sim N(0,\mathrm{Id})} \mathop{\mathbb{E}}_{i\sim\{1,\ldots,n\}} K(A, M, \lambda_0, v_0) = \mathcal{O}(n^3).$$

A finer randomization?

Theorem We have

$$\underset{A\sim N(0,\mathrm{Id})}{\mathbb{E}} \ \underset{M\sim N(0,\mathrm{Id})}{\mathbb{E}} \ \underset{i\sim\{1,\ldots,n\}}{\mathbb{E}} K(A, M, \lambda_0, v_0) = \mathcal{O}(n^3).$$

This result raises the need of an efficient algorithm to draw, given n, a triple (M, λ, v) satisfying:

A finer randomization?

Theorem We have

$$\mathop{\mathbb{E}}_{A\sim N(0,\mathrm{Id})} \mathop{\mathbb{E}}_{M\sim N(0,\mathrm{Id})} \mathop{\mathbb{E}}_{i\sim\{1,\ldots,n\}} K(A, M, \lambda_0, v_0) = \mathcal{O}(n^3).$$

This result raises the need of an efficient algorithm to draw, given n, a triple (M, λ, v) satisfying:

- M is drawn from $N(0, \mathrm{Id})$;

A finer randomization?

Theorem We have

$$\mathop{\mathbb{E}}_{A\sim N(0,\mathrm{Id})} \mathop{\mathbb{E}}_{M\sim N(0,\mathrm{Id})} \mathop{\mathbb{E}}_{i\sim\{1,\ldots,n\}} K(A, M, \lambda_0, v_0) = \mathcal{O}(n^3).$$

This result raises the need of an efficient algorithm to draw, given n, a triple (M, λ, v) satisfying:

- M is drawn from $N(0, \mathrm{Id})$;
- (λ, v) is drawn from the (discrete) uniform distribution among the n eigenpairs of M.

A finer randomization?

Theorem We have

$$\underset{A\sim N(0,\mathrm{Id})}{\mathbb{E}}\ \underset{M\sim N(0,\mathrm{Id})}{\mathbb{E}}\ \underset{i\sim\{1,\ldots,n\}}{\mathbb{E}} K(A, M, \lambda_0, v_0) = \mathcal{O}(n^3).$$

This result raises the need of an efficient algorithm to draw, given n, a triple (M, λ, v) satisfying:

- M is drawn from $N(0, \mathrm{Id})$;
- (λ, v) is drawn from the (discrete) uniform distribution among the n eigenpairs of M.

To do so, it is enough to draw a diagonal matrix whose diagonal elements $(\lambda_1, \ldots, \lambda_n)$ follow the spectrum law of the Ginibre ensemble.

Can we do this in $\mathcal{O}(n^6)$ operations?