Wilkinson's bus: Weak condition numbers, with applications

Vanni Noferini, Aalto University (Finland)

Joint with Martin Lotz (Warwick)

J.H.Wilkinsons 100th birthday, Manchester

May 30th, 2019

Wilkinson's bus

Wilkinson on the rarity of worst-case scenarios, speaking of Gaussian Elimination:

Wilkinson's bus

Wilkinson on the rarity of worst-case scenarios, speaking of Gaussian Elimination:

Anyone that unlucky has already been run over by a bus.

(I learnt the quote by reading N. Trefethen, The Smart Money's on Numerical Analysts, SIAM News 45(9), 2012)

Classical condition number

$$f : \mathcal{V} \to \mathcal{W}$$

Ideas by Von Neumann, Turing: a single number κ in $[0, \infty]$ that measures the worst-case sensitivity of f with respect to perturbations in the input.

Classical condition number

$$f : \mathcal{V} \to \mathcal{W}$$

Ideas by Von Neumann, Turing: a single number κ in $[0, \infty]$ that measures the worst-case sensitivity of f with respect to perturbations in the input.

- If f is differentiable then κ is the norm of the Frechet derivative of f

Classical condition number

$$f : \mathcal{V} \to \mathcal{W}$$

Ideas by Von Neumann, Turing: a single number κ in $[0, \infty]$ that measures the worst-case sensitivity of f with respect to perturbations in the input.

- If f is differentiable then κ is the norm of the Frechet derivative of f
- Higham's rule of thumb:

$$\text{forward error} \lesssim \kappa \cdot (\text{backward error})$$

Classical condition number

$$f : \mathcal{V} \to \mathcal{W}$$

Ideas by Von Neumann, Turing: a single number κ in $[0, \infty]$ that measures the worst-case sensitivity of f with respect to perturbations in the input.

- If f is differentiable then κ is the norm of the Frechet derivative of f
- Higham's rule of thumb:

$$\text{forward error} \lesssim \kappa \cdot (\text{backward error})$$

- Morally κ should predict (in a worst case sense) the accuracy of computations in a fixed finite precision arithmetic setting

A generalized eigenproblem

$$L(x) = \begin{bmatrix} -1 & 1 & 4 & 2 \\ -2 & 3 & 12 & 6 \\ 1 & 3 & 11 & 6 \\ 2 & 2 & 7 & 4 \end{bmatrix} x + \begin{bmatrix} 2 & -1 & -5 & -1 \\ 6 & -2 & -11 & -2 \\ 5 & 0 & -2 & 0 \\ 3 & 1 & 3 & 1 \end{bmatrix}$$

A generalized eigenproblem

$$L(x) = \begin{bmatrix} -1 & 1 & 4 & 2 \\ -2 & 3 & 12 & 6 \\ 1 & 3 & 11 & 6 \\ 2 & 2 & 7 & 4 \end{bmatrix} x + \begin{bmatrix} 2 & -1 & -5 & -1 \\ 6 & -2 & -11 & -2 \\ 5 & 0 & -2 & 0 \\ 3 & 1 & 3 & 1 \end{bmatrix}$$

$L(x)$ is a singular pencil whose only eigenvalue is 1; $\kappa = \infty$.

A generalized eigenproblem

$$L(x) = \begin{bmatrix} -1 & 1 & 4 & 2 \\ -2 & 3 & 12 & 6 \\ 1 & 3 & 11 & 6 \\ 2 & 2 & 7 & 4 \end{bmatrix} x + \begin{bmatrix} 2 & -1 & -5 & -1 \\ 6 & -2 & -11 & -2 \\ 5 & 0 & -2 & 0 \\ 3 & 1 & 3 & 1 \end{bmatrix}$$

$L(x)$ is a singular pencil whose only eigenvalue is 1; $\kappa = \infty$.

MATLAB R2016a's solution:

```
» eig(L0,-L1)

ans =

        -138.1824366539536
        -0.674131242894470
         1.000000000000000
         0.444114486065683
```

Observations

Observations

- QZ is not a structured algorithm such as e.g. staircase! As a consequence it is blind to the fact that $L(x)$ has only one eigenvalue, because almost all 4×4 pencils have four.

Observations

- QZ is not a structured algorithm such as e.g. staircase! As a consequence it is blind to the fact that $L(x)$ has only one eigenvalue, because almost all 4×4 pencils have four.
- Hence it is expected that out of the four computed eigenvalues three are just noise. This cannot be avoided (although one can check reliability a posteriori... more to come)

Observations

- QZ is not a structured algorithm such as e.g. staircase! As a consequence it is blind to the fact that $L(x)$ has only one eigenvalue, because almost all 4×4 pencils have four.
- Hence it is expected that out of the four computed eigenvalues three are just noise. This cannot be avoided (although one can check reliability a posteriori... more to come)
- The problem is ill posed (=discontinuous): plenty of pencils arbitrarily close to $L(x)$ whose eigenvalues are *all* nowhere near 1: definitely $\kappa = \infty$, and arguably of the worst kind.

Observations

- QZ is not a structured algorithm such as e.g. staircase! As a consequence it is blind to the fact that $L(x)$ has only one eigenvalue, because almost all 4×4 pencils have four.
- Hence it is expected that out of the four computed eigenvalues three are just noise. This cannot be avoided (although one can check reliability a posteriori... more to come)
- The problem is ill posed (=discontinuous): plenty of pencils arbitrarily close to $L(x)$ whose eigenvalues are *all* nowhere near 1: definitely $\kappa = \infty$, and arguably of the worst kind.
- Yet the eigenvalue 1, in spite of its infinite condition, is computed with full accuracy!

Observations

- QZ is not a structured algorithm such as e.g. staircase! As a consequence it is blind to the fact that $L(x)$ has only one eigenvalue, because almost all 4×4 pencils have four.
- Hence it is expected that out of the four computed eigenvalues three are just noise. This cannot be avoided (although one can check reliability a posteriori... more to come)
- The problem is ill posed (=discontinuous): plenty of pencils arbitrarily close to $L(x)$ whose eigenvalues are *all* nowhere near 1: definitely $\kappa = \infty$, and arguably of the worst kind.
- Yet the eigenvalue 1, in spite of its infinite condition, is computed with full accuracy!
- How come?

Stochastic condition number

Wilkinson's intuition: worst-case can be overly pessimistic, probabilities matter.

Stochastic condition number

Wilkinson's intuition: worst-case can be overly pessimistic, probabilities matter.

In work by Weiss, Stewart, Armentano et al. the effect of perturbations is studied on average rather than in the worst case. This leads to an alternative notion of condition κ_s. Average case vs. worst case.

Stochastic condition number

Wilkinson's intuition: worst-case can be overly pessimistic, probabilities matter.

In work by Weiss, Stewart, Armentano et al. the effect of perturbations is studied on average rather than in the worst case. This leads to an alternative notion of condition κ_s. Average case vs. worst case.

Important: perturbations are not random! But it may make sense to model our ignorance by randomness.

Stochastic condition number

Wilkinson's intuition: worst-case can be overly pessimistic, probabilities matter.

In work by Weiss, Stewart, Armentano et al. the effect of perturbations is studied on average rather than in the worst case. This leads to an alternative notion of condition κ_s. Average case vs. worst case.

Important: perturbations are not random! But it may make sense to model our ignorance by randomness.

For differential problems in finite dimension there is a $\sqrt{m}$ factor between κ and κ_s (Armentano, 2010) where m is the real dimension of $\mathcal{V}$. (Often not a terrific gain!)

Stochastic condition number

Wilkinson's intuition: worst-case can be overly pessimistic, probabilities matter.

In work by Weiss, Stewart, Armentano et al. the effect of perturbations is studied on average rather than in the worst case. This leads to an alternative notion of condition κ_s. Average case vs. worst case.

Important: perturbations are not random! But it may make sense to model our ignorance by randomness.

For differential problems in finite dimension there is a $\sqrt{m}$ factor between κ and κ_s (Armentano, 2010) where m is the real dimension of $\mathcal{V}$. (Often not a terrific gain!)

For our $L(x)$: $\kappa_s =$

Stochastic condition number

Wilkinson's intuition: worst-case can be overly pessimistic, probabilities matter.

In work by Weiss, Stewart, Armentano et al. the effect of perturbations is studied on average rather than in the worst case. This leads to an alternative notion of condition κ_s. Average case vs. worst case.

Important: perturbations are not random! But it may make sense to model our ignorance by randomness.

For differential problems in finite dimension there is a $\sqrt{m}$ factor between κ and κ_s (Armentano, 2010) where m is the real dimension of $\mathcal{V}$. (Often not a terrific gain!)

For our $L(x)$: $\kappa_s = \infty$. Average case analysis does not explain why QZ computes 1 so well.

First ingredient to go beyond

Directional sensitivity:

$$\sigma_E = \lim_{\epsilon \to 0} \frac{\|f(D + \epsilon E) - f(D)\|}{\epsilon \|E\|}$$

- Ratio of forward and backward errors for a particular direction of the backward error
- If f is differentiable, $\|E\|^{-1}$ times the norm of the Gateaux derivative

Weak condition number

We want to do some probability, so we fix a probability space $(\Omega, \Sigma, \mathbb{P})$ and a random variable $E : \Omega \to \mathcal{V}$.

Weak condition number

We want to do some probability, so we fix a probability space $(\Omega, \Sigma, \mathbb{P})$ and a random variable $E : \Omega \to \mathcal{V}$.

Given $0 \leq \delta < 1$, the δ-weak worst-case condition is

$$\kappa_w(\delta) = \inf_{\substack{S \in \Sigma, \\ |S| \geq 1-\delta}} \sup_{\omega \in S} \sigma_E(\omega)$$

Weak condition number

We want to do some probability, so we fix a probability space $(\Omega, \Sigma, \mathbb{P})$ and a random variable $E : \Omega \to \mathcal{V}$.

Given $0 \leq \delta < 1$, the δ-weak worst-case condition is

$$\kappa_w(\delta) = \inf_{\substack{S \in \Sigma, \\ |S| \geq 1-\delta}} \sup_{\omega \in S} \sigma_E(\omega)$$

- In practice, "$\kappa_w(\delta)$ is bounded above by $b(\delta)$" means: with probability $1 - \delta$, the forward error is bounded by $b(\delta)$ times the backward error.
- The probability distribution can be seen as a parameter of this definition, and can be made more concrete according to context
- δ can be seen as a very concrete parameter (confidence level) to be input by the user (engineer, scientist, mathematician)
- We do not wish to model rounding errors probabilistically, but to argue that the set of bad perturbations may be so small that algorithms would need a good reason to stumble on it

Back to $L(x)$

Theorem (Lotz, VN)

With respect to uniformly distributed real perturbations on the unit sphere, the weak condition of the eigenvalue 1 *of* $L(x)$ *is bounded by*

$$\kappa_w(\delta) \leq \max\{12.16, \frac{2.149}{\delta}\}.$$

Back to $L(x)$

Theorem (Lotz, VN)

With respect to uniformly distributed real perturbations on the unit sphere, the weak condition of the eigenvalue 1 *of* $L(x)$ *is bounded by*

$$\kappa_w(\delta) \leq \max\{12.16, \frac{2.149}{\delta}\}.$$

(In the paper: way more general results for any simple eigenvalue of any matrix polynomial, singular or regular, for both real and complex perturbations.)

Is this practical?

Yes. The general result depend on a parameter γ_P that generalizes Tisseur's formula for eigenvalue condition of regular matrix polynomials, but is trickier to compute exactly for singular matrix polynomials because eigenvectors are only defined in certain quotient spaces (as opposed to traditional vector spaces).

Is this practical?

Yes. The general result depend on a parameter γ_P that generalizes Tisseur's formula for eigenvalue condition of regular matrix polynomials, but is trickier to compute exactly for singular matrix polynomials because eigenvectors are only defined in certain quotient spaces (as opposed to traditional vector spaces).

However γ_P can be cheaply estimated in practice. All we need is the computed eigentriple from QZ. No more expensive than computing traditional condition numbers for regular polynomial eigenvalue problems, and this can tell reliable eigenvalues from rubbish (remember the 3 spurious computed eigenvalues).

Is this practical?

Yes. The general result depend on a parameter γ_P that generalizes Tisseur's formula for eigenvalue condition of regular matrix polynomials, but is trickier to compute exactly for singular matrix polynomials because eigenvectors are only defined in certain quotient spaces (as opposed to traditional vector spaces).

However γ_P can be cheaply estimated in practice. All we need is the computed eigentriple from QZ. No more expensive than computing traditional condition numbers for regular polynomial eigenvalue problems, and this can tell reliable eigenvalues from rubbish (remember the 3 spurious computed eigenvalues).

For more details:
`arxiv.org/pdf/1905.05466.pdf`