Estimating the Condition Number of $f(A)b$

Edvin Deadman
School of Mathematics
The University of Manchester

October 26, 2015

Numerical Algorithms, 70(2), 2015, 10.1007/s11075-014-9947-4

http://eprints.ma.man.ac.uk/2200/
http://github.com/edvindeadman/fAbcond

What goes in the yellow box?

A condition number algorithm for $f(A)$b ***must not***:

- explicitly require A,
- store dense $O(n^2)$ quantities e.g. $f(A)$,
- use dense $O(n^3)$ techniques such as Schur decomposition.

Compute condition number using only matrix-vector products Av ***and*** A^*v**.**

We will assume we already have an algorithm for computing $f(A)b$.

Condition Number for $f(A)$

Definition

$$\operatorname{cond}(f, A) := \lim_{\epsilon \to 0} \sup_{\|\Delta A\| \le \epsilon \|A\|} \frac{\|f(A + \Delta A) - f(A)\|}{\epsilon \|f(A)\|}.$$

To obtain more useful formula, use the ***Fréchet derivative***:

Linear mapping $L_f(A) : \mathbb{C}^{n \times n} \to \mathbb{C}^{n \times n}$

$$E \mapsto L_f(A, E)$$

where $f(A + E) - f(A) = L_f(A, E) + o(\|E\|)$.

Condition Number for $f(A)$

Operator norm of the Fréchet derivative:

$$\|L_f(A)\| := \max_{E \neq 0} \frac{\|L_f(A, E)\|}{\|E\|}.$$

Lemma

$$\operatorname{cond}(f, A) = \frac{\|A\| \|L_f(A)\|}{\|f(A)\|}.$$

Estimate $\operatorname{cond}(f, A)$ ***by estimating*** $\|L_f(A)\|$**.**

Estimating cond(f, A)

Kronecker form, $K_f(A) \in \mathbb{C}^{n^2 \times n^2}$ defined via

$$K_f(A) vec(E) = vec(L_f(A, E)).$$

Lemmas

$\|K_f(A)\|$ and $\|L_f(A)\|$ are closely related:

$$\|K_f(A)\|_2 = \|L_f(A)\|_F,$$
$$\frac{\|L_f(A)\|_1}{n} \leq \|K_f(A)\|_1 \leq n\|L_f(A)\|_1.$$

Estimate cond(f, A) ***by estimating*** $\|K_f(A)\|$**.**

(Use power method or `normest1`.)

Condition Number for $f(A)b$

Definition

$$\operatorname{cond}(f, A, b) := \lim_{\epsilon \to 0} \sup_{\substack{\|\Delta A\| \le \epsilon \|A\| \\ \|\Delta b\| \le \epsilon \|b\|}} \frac{\|f((A + \Delta A))(b + \Delta b) - f(A)b\|}{\epsilon \|f(A)b\|}$$

... do some algebra ... $\implies$

$$\frac{1}{\|f(A)b\|} \max \left(\|A\| \max_{\|\Delta A\|=1} \|L_f(A, \Delta A)b\|, \|f(A)\| \|b\| \right)$$
$$\le \operatorname{cond}(f, A, b)$$
$$\le \frac{1}{\|f(A)b\|} \left(\|A\| \max_{\|\Delta A\|=1} \|L_f(A, \Delta A)b\| + \|f(A)\| \|b\| \right).$$

How to Estimate $\max_{\|\Delta A\|=1} \|L_f(A, \Delta A)b\|$

Kronecker form $K_f(A, b) \in \mathbb{C}^{n\times n^2}$ defined via:

$$K_f(A, b)vec(E) = L_f(A, E)b.$$

Depending on norm:

$$\max_{\|\Delta A\|=1} \|L_f(A, \Delta A)b\| \quad \text{“}=\text{”} \quad \|K_f(A, b)\|$$

Estimate cond(f, A, b) ***by estimating*** $\|K_f(A, b)\|$**.**

(Use `normest1` or power method.)

Why not use `normest1`?

`normest1` (Higham & Tisseur, 2000):

- estimates matrix 1-norm,
- requires several products $y = Av$ and $x = A^* w$.

Apply to $K_f(A, b) \in \mathbb{C}^{n \times n^2}$:

$$y = K_f(A, b)v, \quad v = vec(E) \text{ for some } E \in \mathbb{C}^{n \times n}$$

$$x = K_f(A, b)^* w$$

Dense $O(n^2)$ ***quantities!!***

Power Method for $\|K_f(A, b)\|_2$

Obtain an estimate γ of $\|K_f(A, b)\|_2$

```
Choose a unit nonzero starting vector z_0 (= vec(Z_0))
for k = 0: ∞
    w_{k+1} = K_f(A, b) z_k = L_f(A, Z_k) b
    z_{k+1} = K_f(A, b)^* w_{k+1}, via Z_{k+1} = L_f^⋆(A, w_{k+1} b^*)
    γ_{k+1} = ‖z_{k+1}‖_2 / ‖w_{k+1}‖_2
    if converged, γ = γ_{k+1}, quit, end
    z_{k+1} = z_{k+1} / ‖z_{k+1}‖_2
end
```

The z_k are still dense and $O(n^2)$!

Altered Power Method for $\|K_f(A,b)\|_2$

Obtain an estimate γ of $\|K_f(A,b)\|_2$

```
Choose a unit nonzero starting vector y_0 ∈ C^n
for k = 0: ∞
    y_{k+1} = L_f(A, L_f^⋆(A, y_k b^*))b    (= K_f(A,b) K_f^*(A,b) y_k)
    γ_{k+1} = sqrt(‖y_{k+1}‖_2)
    if converged, γ = γ_{k+1}, quit, end
    y_{k+1} = y_{k+1}/‖y_{k+1}‖_2
end
```

$$y_{k+1} = L_f(A, L_f^\star(A, y_k b^*))b \quad (= K_f(A,b)K_f^*(A,b)y_k)$$

$$\gamma_{k+1} = \sqrt{\|y_{k+1}\|_2}$$

$$y_{k+1} = y_{k+1}/\|y_{k+1}\|_2$$

Need to compute $L_f(A, L_f^\star(A, y_k b^*))b$ ***using only matrix-vector multiplications.***

How to Compute $L_f(A, L_f^\star(A, y_k b^*))b$

Exploit

$$f\left(\begin{bmatrix} A & E \\ 0 & A \end{bmatrix}\right) = \begin{bmatrix} f(A) & L_f(A, E) \\ 0 & f(A) \end{bmatrix}.$$

$L_f(A, L_f^\star(A, y_k b^*))b$ given by top n entries of

$$f\left(\begin{bmatrix} A & L_f^\star(A, y_k b^*) \\ 0 & A \end{bmatrix}\right)\begin{bmatrix} 0 \\ b \end{bmatrix}.$$

$L_f^\star(A, y_k b^*)q$ given by top n entries of

$$\bar{f}\left(\begin{bmatrix} A^* & y_k b^* \\ 0 & A^* \end{bmatrix}\right)\begin{bmatrix} 0 \\ q \end{bmatrix}.$$

If** $f(A)b$ **requires** Π **matmuls then** $L_f(A, L_f^\star(A, y_k b^*))b$ **requires** $2\Pi(\Pi + 1)$ **matmuls.

Application to $e^A b$

Al-Mohy & Higham (2011):

- Use $e^A b = e^{(A/s)} \dots e^{(A/s)} b$.
- $e^{(A/s)} b$ computed using m^{th} degree Taylor polynomial.
- $O(sm)$ matmuls for $e^A b$.
- Can extend to $e^A B$ and use level-3 BLAS.

Apply to condition number:

- Choose s and m for half-precision.
- $O(s^2 m^2)$ matmuls for cond(exp, A, b).
- Extension to $e^A B$ still works.