Scaling of polynomial eigenvalue problems

Timo Betcke
timo.betcke@maths.man.ac.uk

School of Mathematics
The University of Manchester

March 22, 2007

Why do we scale (balance)?

A is badly scaled:

$$A = \begin{bmatrix} 0 & 2^{-20} & 2^{-5} \\ 2^{20} & 1 & 0 \\ 2^{6} & 2^{-15} & 0 \end{bmatrix}$$

eig(A,'nobalance')
2.170086486635807
-1.481194304285061
0.311107817428706 0

After scaling:

$$D^{-1}AD = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 2 & 1 & 0 \end{bmatrix}$$

eig(A,'nobalance')
2.170086486626031
-1.481194304092015
0.311107817465981 7

$$D = \begin{bmatrix} 2^{-20} & 1 & 2^{-15} \end{bmatrix}$$

Scaling of polynomial eigenvalue problems

Define the matrix polynomial

$$P(\lambda) := \lambda^m A_m + \cdots + \lambda A_1 + A_0$$

Our aim: Find diagonal scaling matrices D_1, D_2 such that the eigenvalues of

$$D_1 P(\lambda) D_2 = \lambda^m D_1 A_m D_2 + \cdots + \lambda D_1 A_1 D_2 + D_1 A_0 D_2$$

are less sensitive to perturbations.

- How can we rigorously define the aim of scaling?
- What is the influence of scaling on condition number, forward error, backward error?
- Scaling and linearizations of matrix polynomials
- Other ways of improving computed eigenvalues

Two examples

A case, where scaling improves the forward error

Define the quadratic eigenvalue problem $P(\lambda) = A_0 + \lambda A_1 + \lambda^2 A_2$ by

$$A_0 = 10^{-8} \begin{bmatrix} -59990 & 50009990 & -60 \\ -12982 & 10002982 & -28 \\ -96984 & 90006984 & -106 \end{bmatrix}, \quad A_1 = 10^{-8} \begin{bmatrix} -44975 & 35009975 & -60 \\ -9955 & 7002955 & -52 \\ -69960 & 63006960 & -103 \end{bmatrix}$$

$$A_2 = 10^{-8} \begin{bmatrix} 14985 & 4980015 & 10 \\ 4973 & 994027 & 26 \\ 4976 & 8986024 & 15 \end{bmatrix}.$$

Two examples...

Before scaling

λ	$\tilde{\lambda}$	$\eta(\tilde{\lambda})$	$\kappa(\lambda)$
$-1/3$	-0.33333333562194	$9.22 \cdot 10^{-17}$	$7.45 \cdot 10^{6}$
1	$+1.00000000001247$	$7.41 \cdot 10^{-17}$	$1.68 \cdot 10^{4}$
2	$+2.00000000584293$	$1.05 \cdot 10^{-16}$	$2.79 \cdot 10^{6}$

After scaling

λ	$\tilde{\lambda}$	$\eta(\tilde{\lambda})$	$\kappa(\lambda)$
$-1/3$	-0.333333333333333	$3.72 \cdot 10^{-17}$	26.9
1	$+1.000000000000006$	$2.08 \cdot 10^{-16}$	28.9
2	$+2.000000000000000$	$1.38 \cdot 10^{-17}$	10.9

- Condition numbers reduced by scaling
- Backward error has only changed little
- Gain of up to 6 digits of accuracy

Two examples...

Consider the generalized eigenvalue problem $P(\lambda) = A_0 - \lambda A_1$ defined by

$$A_0 = \begin{bmatrix} -2.977448532913490e-01 & 4.452263778189542e-01 & -1.373875559192313e+00 \\ -7.863347359785547e+03 & 1.612185635982964e+01 & -6.817404683656819e-02 \\ 3.479541188173907e+01 & -4.950564712472924e-02 & 2.320917894721326e-03 \end{bmatrix}$$

$$A_1 = \begin{bmatrix} -2.167083034887365e-01 & 3.265327471226953e-01 & -8.537224431099331e-01 \\ -1.045204352731886e+04 & 2.142639766742700e+01 & -1.672334249618942e-01 \\ 4.625880080141054e+01 & -7.899480493461045e-02 & 2.159236789260356e-03 \end{bmatrix}$$

$\tilde{\lambda}$	$\eta(\tilde{\lambda})$	$\kappa(\lambda)$
$7.523262370299634e-01$	$6.35 \cdot 10^{-17}$	4.80
$1.378777970940714e+00$	$6.14 \cdot 10^{-22}$	$1.02 \cdot 10^{6}$
$1.602702387792619e+00$	$9.00 \cdot 10^{-22}$	$3.84 \cdot 10^{5}$

$\tilde{\lambda}$	$\eta(\tilde{\lambda})$	$\kappa(\lambda)$
$7.523262370299637e-01$	$1.98 \cdot 10^{-16}$	3.94
$1.378777970940470e+00$	$6.42 \cdot 10^{-17}$	$2.77 \cdot 10^{3}$
$1.602702387792630e+00$	$1.76 \cdot 10^{-18}$	$4.08 \cdot 10^{3}$

Although the condition numbers decrease, the forward errors increase after scaling!

Optimal scalings

Let λ be an eigenvalue of $P(\lambda)$ with left and right eigenvectors y and x. The *normwise* condition number of λ is given as ([Tisseur, 00])

$$\kappa(\lambda) = \frac{\|y\|_2\|x\|_2\ \alpha}{|\lambda||y^H P'(\lambda)x|}, \quad \alpha = \sum_{k=0}^{m} |\lambda|^k \|A_k\|_2.$$

After scaling this becomes

$$\kappa_{D_1,D_2}(\lambda) = \frac{\|D_1^{-1}y\|_2\|D_2^{-1}x\|_2\ \alpha_{D_1,D_2}}{|\lambda||y^H P'(\lambda)x|}, \quad \alpha = \sum_{k=0}^{m} |\lambda|^k \|D_1 A_k D_2\|_2.$$

Scale y and x such that $|y^H P'(\lambda)x| = 1$ to obtain

$$\kappa_{D_1,D_2}(\lambda) = \frac{1}{|\lambda|}\|D_1^{-1}y\|_2\|D_2^{-1}x\|_2\ \alpha_{D_1,D_2}$$

Optimal scalings...

Definition: A matrix polynomial $P(\lambda)$ is optimally scaled for an eigenvalue λ if

$$\kappa(\lambda) = \inf_{\det(D_1 D_2)\neq 0} \frac{1}{|\lambda|} \|D_1^{-1} y\|_2 \|D_2^{-1} x\|_2 \ \alpha_{D_1,D_2},$$

where x and y are chosen such that $|y^H P'(\lambda) x| = 1$.

Note, that an optimal scaling depends on the eigenvalue λ.

How can we characterize optimal scalings?

The componentwise condition number

The componentwise condition number of an eigenvalue is defined as

$$\text{cond}(\lambda) := \lim_{\epsilon \to 0} \sup \left\{ \frac{|\Delta\lambda|}{\epsilon|\lambda|} \; : \; (P(\lambda+\Delta\lambda) + \Delta P(\lambda+\Delta\lambda))(x+\Delta x) = 0, \right.$$
$$\left. |\Delta A_k| \leq \epsilon |A_k|) \right\}.$$

(see [Higham/Higham, 98] for generalized case).

We have

$$\text{cond}(\lambda) = \frac{|y^H|A|x|}{|\lambda||y^H P'(\lambda)x|}, \quad A := \sum_{k=0}^{l} |\lambda|^k |A_k|.$$

Componentwise vs. normwise condition numbers

Theorem: Let all entries of y and x be nonzero. Then

$$\frac{1}{n}\mathrm{cond}(\lambda) \leq \inf_{\det(D_1 D_2)\neq 0} \kappa_{D_1,D_2}(\lambda) \leq n\ \mathrm{cond}(\lambda).$$

An almost optimal scaling is given by

$$D_1 = \mathrm{diag}(|y_i|), D_2 = \mathrm{diag}(|x_i|)$$

For this scaling we have

$$\kappa_{D_1,D_2}(\lambda) \leq n\ \mathrm{cond}(\lambda).$$

The influence on the backward error

The normwise backward error of an approximate eigenpair $(\tilde{x}, \tilde{\lambda})$ is given as [Tisseur, 00]

$$\eta(\tilde{x}, \tilde{\lambda}) = \frac{\|P(\tilde{\lambda})\tilde{x}\|_2}{\|\tilde{x}\|_2 \ \alpha}.$$

Under scaling this becomes

$$\eta_{D_1,D_2}(\tilde{\lambda}) = \frac{\|D_1 P(\tilde{\lambda})\tilde{x}\|_2}{\|D_2^{-1}\tilde{x}\|_2 \ \alpha_{D_1,D_2}}.$$

Theorem: Let $(\tilde{x}, \tilde{\lambda})$ be an approximate eigenpair. Then

$$\sup_{D_1,D_2} \eta_{D_1,D_2}(\tilde{x}, \tilde{\lambda}) \leq n^{\frac{3}{2}}\omega(\tilde{\lambda}, \tilde{x}),$$

where $\omega(\tilde{x}, \tilde{\lambda})$ is the componentwise backward error.

Obtaining good scalings

Use an extension of the scaling of Lemonnier/Van Dooren to matrix polynomials. Find D_1, D_2 such that

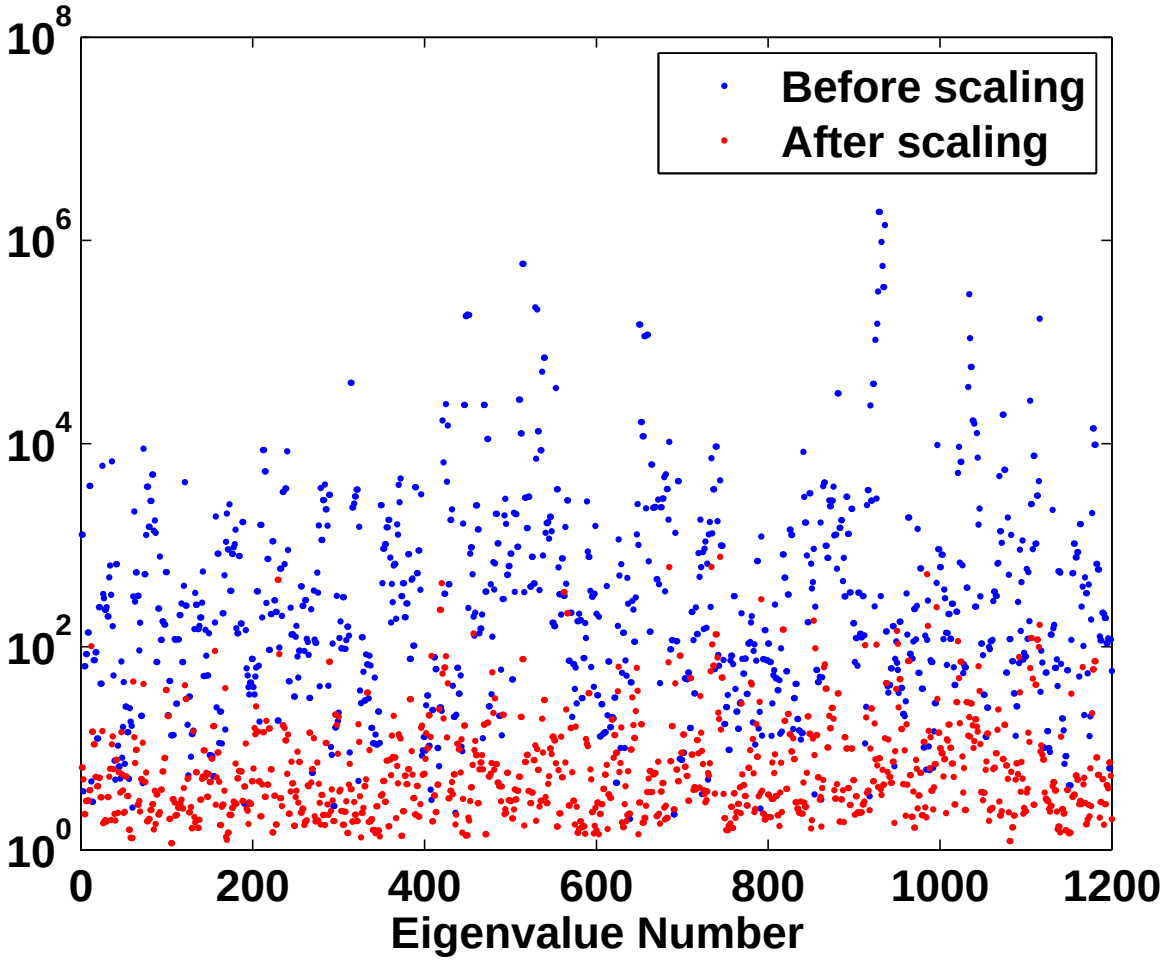
$$\sum_{k=0}^{m} \|D_1 A_k D_2 e_i\|^2 = \sum_{k=0}^{m} \|e_j^H D_1 A_k D_2\|^2 = 1, \;\; i,j = 1,\ldots,n.$$

Heuristic idea: If rows and columns are normalized, the entries of the eigenvectors are expected to be of similar magnitude if $P(\lambda)$ does not have special structure.

- Implementation only a minor modification of the generalized case described in [Lemonnier/Van Dooren, 2006].
- If magnitude of the wanted eigenvalues approximately known it is slightly favorable to use the condition

$$\sum_{k=0}^{m} |\lambda|^{2k} \|D_1 A_k D_2 e_i\|^2 = \sum_{k=0}^{m} |\lambda|^{2k} \|e_j^H D_1 A_k D_2\|^2 = 1, \;\; i,j = 1,\ldots,n.$$

Numerical results



The ratio of componentwise and normwise condition number before and after scaling for 100 random 6×6 quadratic polynomials.

Scalings and linearizations

Should we first scale and then linearize or first linearize and then scale ?
Example:

$$\lambda \begin{bmatrix} D_1 A D_2 & 0 \\ 0 & -D_1 C D_2 \end{bmatrix} + \begin{bmatrix} D_1 B D_2 & D_1 C D_2 \\ D_1 C D_2 & 0 \end{bmatrix} = (I \otimes D_1) \left(\lambda \begin{bmatrix} A & 0 \\ 0 & -C \end{bmatrix} + \begin{bmatrix} B & C \\ C & 0 \end{bmatrix} \right) (I \otimes D_2)$$

vs.

$$\begin{bmatrix} \tilde{D}_{11} & \\ & \tilde{D}_{12} \end{bmatrix} \left(\lambda \begin{bmatrix} A & 0 \\ 0 & -C \end{bmatrix} + \begin{bmatrix} B & C \\ C & 0 \end{bmatrix} \right) \begin{bmatrix} \tilde{D}_{21} & \\ & \tilde{D}_{22} \end{bmatrix}$$

Theorem [Higham/Mackey/Tisseur,06]: Let $L(\lambda) \in \mathbb{DL}(P)$. Then

$$\kappa_L(\lambda; v) = \frac{|\lambda| \|X\|_2 + \|Y\|_2}{|p(\lambda; v)|} \frac{\|\Lambda\|_2^2 \|y\|_2 \|x\|_2}{|\lambda| |y^H P'(\lambda) x|},$$

where $\Lambda := [\lambda^{m-1}, \lambda^{m-2}, \ldots, 1]^T$ and $p(\lambda; v) = \Lambda^T v$.

Scaling leads to

$$\kappa_{L_{D_1, D_2}}(\lambda; v) = \frac{|\lambda| \|\tilde{X}\|_2 + \|\tilde{Y}\|_2}{|p(\lambda; v)|} \frac{\|\Lambda\|_2^2 \|D_1^{-1} y\|_2 \|D_2^{-1} x\|_2}{|\lambda| |y^H P'(\lambda) x|}.$$

Other possible transformations...

Theorem: [Higham/Mackey/Tisseur, 06]
Let λ be a simple, finite, nonzero eigenvalue of P. Then

$$\left(\frac{2\sqrt{m}}{m+1}\right)\frac{1}{\rho} \leq \frac{\inf_v \kappa_L(\lambda; v)}{\kappa_P(\lambda)} \leq m^2\rho,$$

where $\rho = \frac{\max_i \|A_i\|_2}{\min(\|A_0\|_2, \|A_m\|_2)}$.

[Fan/Lin/Van Dooren, 04] For QEP $P(\lambda) = \lambda^2 A + \lambda B + C$ substitute $\lambda = \alpha\mu$. Solve

$$\min_{\alpha,\beta} \max\{|\beta\alpha^2\|A\|_2 - 1|, |\beta\alpha\|B\|_2 - 1|, |\beta\|C\|_2 - 1|\}.$$

Define $\rho(\alpha) = \frac{\max_i \alpha^i \|A_i\|_2}{\min(\|A_0\|_2, \alpha^m\|A_m\|_2)}$. Now minimize $\rho(\alpha)$.

Other possible transformations...

Theorem: For $\alpha > 0$ the global minimizer of $\rho(\alpha)$ is $\alpha = (\|A_0\|_2/\|A_m\|_2)^{1/m}$.

- For $m = 2$ (QEP case) this is identical to the choice of α proposed by Fan/Lin/Van Dooren.
- Note, that $\kappa_P(\lambda)$ is invariant under this scaling.
- For palindromic problems we have $\alpha = 1$. Hence, this approach has no effect on these problems.

Other possible transformations...

Let $P(\lambda) = \lambda^2 A + \lambda B + A^H$ with $B = B^H$. We obtain

$$\rho = \max\{1, \|B\|_2/\|A\|_2\}.$$

If $\|B\|_2 \gg \|A\|_2$ linearization might be badly conditioned.

Try structure preserving transformation of λ that reduces ρ. Set

$$\lambda = \frac{\mu - \alpha}{1 - \overline{\alpha}\mu}, \quad \alpha \in \mathbb{C}.$$

This transformation preserves palindromic structure. We obtain

$$\rho(\alpha) = \max\left(1, \frac{\| - 2\alpha A + (1 + |\alpha|^2)B - 2\overline{\alpha}A^H\|}{\|A - \overline{\alpha}B + \overline{\alpha}^2 A^H\|}\right).$$

It follows that $\rho(1) = 2$. But $\kappa_P(\mu) \to \infty$ as $\alpha \to 1$.