Matrix balancing for field of values type inclusion regions

Student Krylov Day 2015

Ian Zwaan

Joint work with Michiel Hochstenbach

Eindhoven Univeristy of Technology

February 02, 2015

The challenge

Given large sparse A or routine `mv(A,x)`.

Want:

- high quality inclusion region;
- fast.

Applications:

- eigenvalue localization;
- stability;
- convergence behavior linear solvers;
- etc.

Motivation

Inclusion regions

Largest eigenvalue of `A = grcar(10000) + 5I`:

- `eigs`: fail
- `krylov_schur` (Stewart 2001): $\mathcal{O}(5000)$ MVs, accuracy?
- Good FoV based eigenvalue inclusion region: $10 - 20$ MVs.

Motivation

Field of values

Field of values:

- cheap to approximate numerically;
- convex;
- guaranteed to contain all eigenvalues;
- often relatively tight around eigenvalues, but not always!

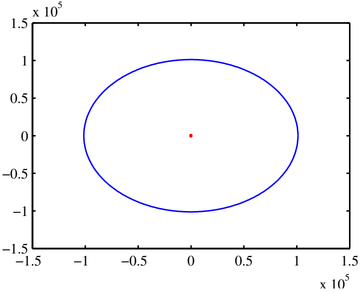


Figure: Tolosa before balancing

Motivation
Balancing

Balancing badly scaled A may:

- dramatically improve quality of FoV as inclusion region;
- improve accuracy of eigenvalue computations;
- improve convergence behaviour.

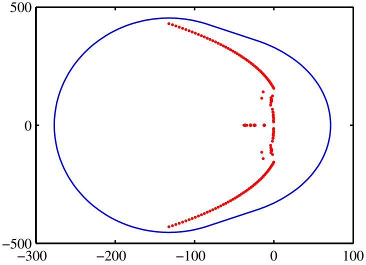


Figure: Tolasa after balancing

Definitions

Field of values:

$$W(A) \equiv \{\mathbf{x}^* A \mathbf{x} : \|\mathbf{x}\|_2 = 1\}$$

Spectral radius:

$$\rho(A) \equiv \max_{\lambda \in \Lambda(A)} |\lambda|$$

Numerical radius:

$$r(A) \equiv \max_{z \in W(A)} |z|$$

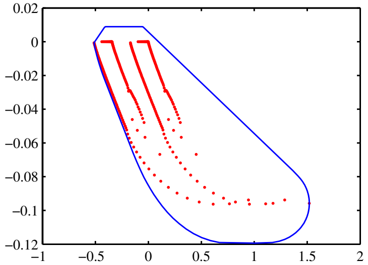


Figure: Quantum Chemistry

Math

Ideally:

$$\frac{r(A)}{\rho(A)} \approx 1$$

Unbalanced:

$$\frac{r(A)}{\rho(A)} \gg 1$$

Balancing:

Replace A by $D^{-1}AD$, usually: $\|D^{-1}AD\| \ll \|A\|$.

Effect:

$\frac{1}{2}\|A\| \leq r(A) \leq \|A\|$ (Theorem)

$$\frac{r(D^{-1}AD)}{\rho(D^{-1}AD)} = \frac{r(D^{-1}AD)}{\rho(A)} \ll \frac{r(A)}{\rho(A)}$$

$W(A) \equiv \{\mathbf{x}^* A \mathbf{x} : \|\mathbf{x}\|_2 = 1\}, \qquad \rho(A) \equiv \max_{\lambda \in \Lambda(A)} |\lambda|, \qquad r(A) \equiv \max_{z \in W(A)} |z|$

Results

Quebec Hydroelectric Power System

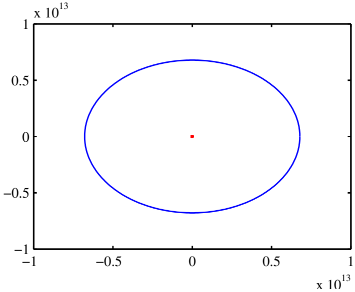


Figure: Unbalanced

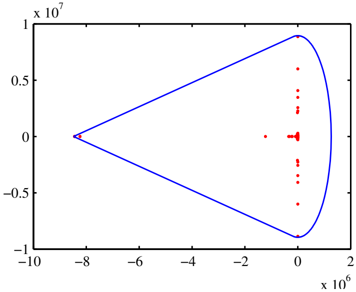


Figure: Balanced

Results

Table: $r(A)/\rho(A)$ for different matrices and methods.

Matrix	spbal	K&B	Az	ATz	Cutoff
af23560	1.01	1.02	—	1.08	1.29
bcsstk29	1.00	1.00	2.32	—	1.00
cry10000	1.00	1.00	3.87	1.53	1.00
dw8192	1.00	1.00	4.64	1.03	1.12
e40r0000	1.00	1.00	—	1.00	1.00
grcar10000	1.08	1.08	3.88	—	2.52
memplus	1.00	1.00	3.36	—	1.00
olm5000	1.00	1.01	—	1.64	1.54
rw5151	0.98	0.98	—	9.79	2.22
sherman3	1.00	1.00	1.06	1.00	1.00
tols4000	1.01	3.27	—	2.55	5.08
utm5940	0.95	0.94	7.20	—	1.36

`spbal`, Az, ATz, Cutoff by (Chen, Demmel 2000)

$W(A) \equiv \{\mathbf{x}^* A \mathbf{x} : \|\mathbf{x}\|_2 = 1\}, \qquad \rho(A) \equiv \max_{\lambda \in \Lambda(A)} |\lambda|, \qquad r(A) \equiv \max_{z \in W(A)} |z|$

Conclusions

Field of values:

- simple;
- cheap;
- often tight.

Balancing:

- simple (e.g., `B = spbalance(A)`);
- fast (e.g., 0.25ms for `A = tols4000`);
- it's super effective!

"Krylov & Balancing" (K&B) shows high quality inclusion regions can be created for large sparse A with only a few (e.g., 20) MVs.

Computing the Field of Values

Small Dense Matrices

(Johnson 1978)

Step 1

For a few $\alpha \in [0, 2\pi)$ compute

$$\lambda_\alpha = \frac{1}{2}\lambda_{\max}(e^{i\alpha}A + (e^{i\alpha}A)^*)$$

and the corresponding eigenvector $\mathbf{x}_\alpha$.

Step 2

Compute the convex hull of

$$\{\mathbf{x}_\alpha^* A \mathbf{x}_\alpha : \alpha\}.$$

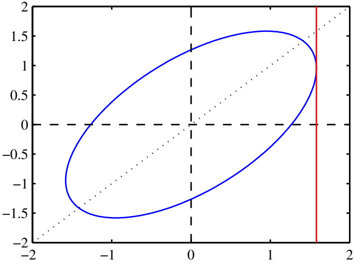


Figure: Example where $\alpha = \pi/4$.

$$W(A) \equiv \{\mathbf{x}^* A\mathbf{x} : \|\mathbf{x}\|_2 = 1\}, \qquad \rho(A) \equiv \max_{\lambda \in \Lambda(A)} |\lambda|, \qquad r(A) \equiv \max_{z \in W(A)} |z|$$

Computing the Field of Values

Large Sparse Matrices

Step 1

Create basis V_k for Krylov subspace

$$\mathrm{span}\{\mathbf{v}_1, A\mathbf{v}_1, \ldots, A^{k-1}\mathbf{v}_1\}$$

and compute $H_k = V_k^* A V_k$.

Step 2

Compute $W(H_k)$. Why?
Ritz values approximate exterior eigenvalues well and

$$W(H_k) \subseteq W(H_{k+1}) \subseteq W(A).$$

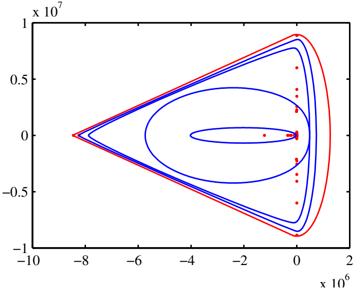


Figure: $W(H_2), \ldots, W(H_5)$.

$W(A) \equiv \{\mathbf{x}^* A\mathbf{x} : \|\mathbf{x}\|_2 = 1\}$, $\rho(A) \equiv \max_{\lambda \in \Lambda(A)} |\lambda|$, $r(A) \equiv \max_{z \in W(A)} |z|$

The Krylov and Balancing Method

Easy as π!

Step 1

Compute $H_k = V_k^* A V_k$ for, say, $k = 20$.
(Remember: need H_k anyway for $W(A)$.)

Step 2

Balance H_k (e.g. using Matlab's `balance`).

Step 3

???

Step 2

Profit!!!

$$W(A) \equiv \{\mathbf{x}^* A \mathbf{x} : \|\mathbf{x}\|_2 = 1\}, \qquad \rho(A) \equiv \max_{\lambda \in \Lambda(A)} |\lambda|, \qquad r(A) \equiv \max_{z \in W(A)} |z|$$

Dealing with outliers

Recall

$$\lambda_\alpha = \frac{1}{2}\lambda_{\max}(e^{i\alpha}A + (e^{i\alpha}A)^*)$$

Assume:

$$\lambda_{\max} = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n = \lambda_{\min}$$

With outliers

$$\lambda_\alpha = \frac{1}{2}\lambda_2(e^{i\alpha}A + (e^{i\alpha}A)^*)$$

Or λ_3, λ_4, etc.

$$W(A) \equiv \{\mathbf{x}^* A\mathbf{x} : \|\mathbf{x}\|_2 = 1\}, \qquad \rho(A) \equiv \max_{\lambda \in \Lambda(A)} |\lambda|, \qquad r(A) \equiv \max_{z \in W(A)} |z|$$

Example

```
A = rand(1000);
```

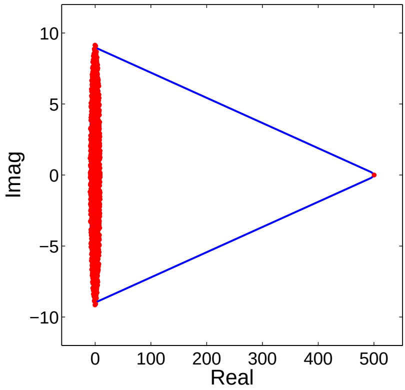


Figure: Original

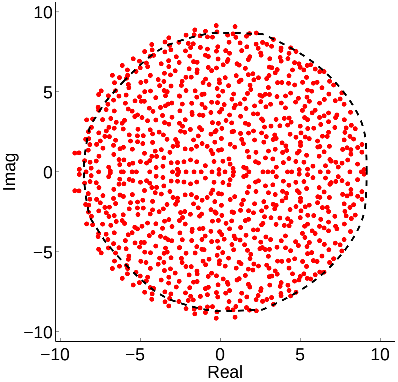


Figure: Higher-order FoV

Exclusion regions

Idea

Shift-and-invert: eigenvalues of $(A - \tau I)^{-1}$ are $(\lambda - \tau)^{-1}$ and

$$\Lambda(A) = \bigcap_{\tau \in \mathbb{C} \setminus \Lambda(A)} \frac{1}{W((A - \tau I)^{-1})} + \tau$$

(Hochstenbach, Singer, Zachlin 2008, 2013)

Example

- Incl. region $W(H_k)$
- Excl. region $1/W((H_k - \tau_j I)^{-1}) + \tau_j$
- Random starting on S^{n-1}
- Krylov dimension $k = 20$

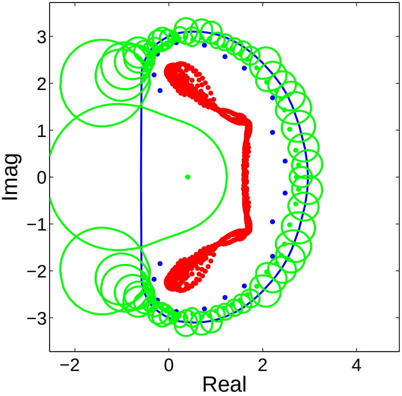


Figure: grcar

$$W(A) \equiv \{\mathbf{x}^* A \mathbf{x} : \|\mathbf{x}\|_2 = 1\}, \qquad \rho(A) \equiv \max_{\lambda \in \Lambda(A)} |\lambda|, \qquad r(A) \equiv \max_{z \in W(A)} |z|$$

spbalance

Parlett–Reinsch (1969):

- gebal (LAPACK);
- balanc (EISPACK);
- balance (MATLAB).

Chen–Demmel (2000):

- spbalance.

Sparse Parlett–Reinsch with improved permutation phase.

spbalance

Permutation phase

What?

Find P such that

$$P^T A P = \begin{bmatrix} X_{11} & \dots & X_{1b} \\ 0 & \ddots & \vdots \\ 0 & 0 & X_{bb} \end{bmatrix}$$

is block upper triangular.

Why?

$$\Lambda(A) = \bigcup_i \Lambda(X_{ii})$$

spbalance

Scale phase

What?

Compute (diagonal) D_i such that

$$D_i^{-1} X_{ii} D_i$$

is balanced.

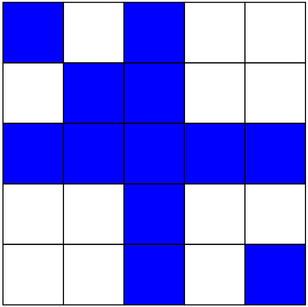
Why?

$$B = D^{-1} P^T A P D = \begin{bmatrix} D_1^{-1} X_{11} D_1 & \dots & D_1^{-1} X_{1b} D_b \\ 0 & \ddots & \vdots \\ 0 & 0 & D_b^{-1} X_{bb} D_b \end{bmatrix}$$

spbalance

Scale phase cont'd

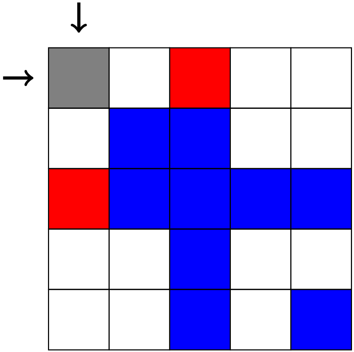
1. Compute $C_j \equiv \|X_{ii}(:,j)\|_1$, $R_j \equiv \|X_{ii}(j,:)\|_1$, exclude $X_{ii}(j,j)$.
2. While $C_j < R_j$ ($C_j > R_j$) increase (decrease) $D_i(j,j)$.
3. Iterate over j until convergence.



spbalance

Scale phase cont'd

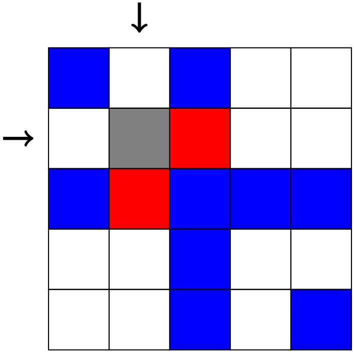
1. Compute $C_j \equiv \|X_{ii}(:,j)\|_1$, $R_j \equiv \|X_{ii}(j,:)\|_1$, exclude $X_{ii}(j,j)$.
2. While $C_j < R_j$ ($C_j > R_j$) increase (decrease) $D_i(j,j)$.
3. Iterate over j until convergence.



spbalance

Scale phase cont'd

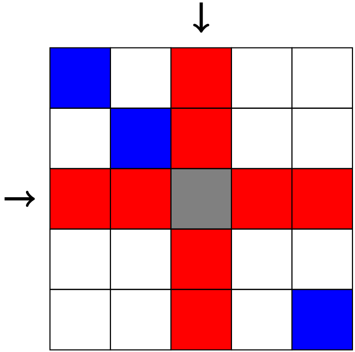
1. Compute $C_j \equiv \|X_{ii}(:,j)\|_1$, $R_j \equiv \|X_{ii}(j,:)\|_1$, exclude $X_{ii}(j,j)$.
2. While $C_j < R_j$ ($C_j > R_j$) increase (decrease) $D_i(j,j)$.
3. Iterate over j until convergence.



spbalance

Scale phase cont'd

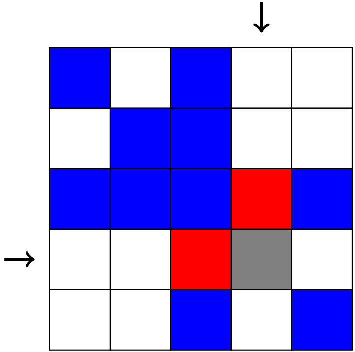
1. Compute $C_j \equiv \|X_{ii}(:,j)\|_1$, $R_j \equiv \|X_{ii}(j,:)\|_1$, exclude $X_{ii}(j,j)$.
2. While $C_j < R_j$ ($C_j > R_j$) increase (decrease) $D_i(j,j)$.
3. Iterate over j until convergence.



spbalance

Scale phase cont'd

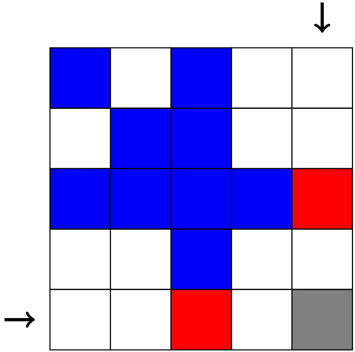
1. Compute $C_j \equiv \|X_{ii}(:,j)\|_1$, $R_j \equiv \|X_{ii}(j,:)\|_1$, exclude $X_{ii}(j,j)$.
2. While $C_j < R_j$ ($C_j > R_j$) increase (decrease) $D_i(j,j)$.
3. Iterate over j until convergence.



spbalance

Scale phase cont'd

1. Compute $C_j \equiv \|X_{ii}(:,j)\|_1$, $R_j \equiv \|X_{ii}(j,:)\|_1$, exclude $X_{ii}(j,j)$.
2. While $C_j < R_j$ ($C_j > R_j$) increase (decrease) $D_i(j,j)$.
3. Iterate over j until convergence.



Summary

1
FoV and balancing $\rightsquigarrow$ high quality inclusion region.

2
Computing FoV for large sparse $A \rightsquigarrow$ K&B.

2
Many extensions.

3
Balancing with Parlett–Reinsch.